

UC Berkeley

Research Reports

Title

Distributed Surveillance and Control on Freeways

Permalink

<https://escholarship.org/uc/item/2wx1d9ck>

Author

Coifman, Benjamin

Publication Date

2004-11-01

CALIFORNIA PATH PROGRAM
INSTITUTE OF TRANSPORTATION STUDIES
UNIVERSITY OF CALIFORNIA, BERKELEY

Distributed Surveillance and Control on Freeways

Benjamin Coifman
Ohio State University

**California PATH Research Report
UCB-ITS-PRR-2004-43**

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Final Report for Task Order 4133

November 2004

ISSN 1055-1425

Distributed Surveillance and Control on Freeways,

Assistance to Caltrans for Assessment of Communications Needs and Solutions, TO 4133 Final Report,

Benjamin Coifman, PhD
Assistant Professor, Civil and Environmental Engineering and Geodetic Science
Assistant Professor, Electrical Engineering
Ohio State University
470 Hitchcock Hall
2070 Neil Ave
Columbus, OH 43210
Coifman.1@OSU.edu
<http://www-ceg.eng.ohio-state.edu/~coifman>
614 292-4282

ABSTRACT

Efficient management of a road network requires continuous decision-making based on conditions on the network and an understanding of the impacts of the decisions made. These conditions are usually measured with fixed-point surveillance systems, most of which are deployed in such a manner as to require communication links that are always connected and are polled at regular intervals. All of the sensor data are typically sent to a Traffic Management Center (TMC) for assessment, yet most of the time no action is taken in response to the data, leading to unnecessarily high communication costs.

To reduce communication costs without a significant loss in the quality of information received at the TMC, this report lays the foundation for an event driven communication system. The analysis considers many related aspects, first by examining the information value of specific events at the detectors. This procedure can then be used to transfer a portion of the decision making from the TMC to the field controllers, which would make the initial valuation of information and only send data that might elicit a control response or benefit comparative decisions between detector stations. In other words, rather than relying on the conventional, centrally polled communication system, these events could be used to initiate communication from the field when the potential value outweighs the cost per communication. The information value could also lead to better data handling for decision-making or archiving in a conventional, polled communication system. We develop the methodology by deconstructing several incidents on a freeway and identify the observable events at a pair of detector stations that may be upstream, downstream or straddle the incident. This analytical process could be repeated for any other condition of interest.

Key to the distributed surveillance scheme is reliable measurements in the field, so next this report presents an analytical methodology to increase the accuracy of speed estimates from freeway traffic detectors by integrating information across lanes. The approach could also prove beneficial for conventional surveillance systems relying on fixed polling periods. Many traffic-monitoring applications only require a single measurement per direction at a detector station. Earlier efforts have focused on improving speed estimates of the individual lanes and then using a simple average to arrive at a single measure of speed at the detector station. This work improves speed estimates on a lane-by-lane basis using conventional aggregated flow and occupancy data obtained from single loop detectors. It then goes further by exploiting the information in the adjacent lanes to eliminate noise. In the course of analysis, data cleaning tools have been developed to identify and exclude malfunctioning detectors and transient errors, improving the validity of traffic data before estimating speed. These data cleaning tools consist of threshold value tests and basic traffic flow theory principles. In the presence of noise, our goal is to produce an estimate of speed across all lanes at the detector station that is within 5 mph of true measure. The single measure of speed at a detector station is obtained by taking the median of the improved estimates across lanes in each direction. The methodology is then extended to clean noisy speeds measured from dual loop detectors and other sensors that mimic single loops.

The next section looks at the mechanics and algorithms necessary to realize a distributed surveillance system that aims at reducing the combined surveillance and communication costs by pre-filtering data in the field. To pre-filter data, distributed control algorithms for five communication modes were developed. These five communication modes exploit the fact that free-flow speeds are relatively stable and constant. Thus, one does not need to transfer this information, and lack of communication can be reliably interpreted as the continuation of the last

reported state. Key to this event-driven approach is the recent emergence of new wireless communication technologies where the communication cost is usage based. The data communication structure was developed for several contemporary wireless technologies available in the United States. A communication protocol was developed that aimed at minimizing the message size as well as maximizing the amount of information that can be transmitted within the limited data size. The protocol also ensured that certain reliability checks are included in the message so that the TMC can identify whenever a message is lost or delayed. Finally, to compute the recurring communication cost, a cost equation was developed that took the number of transmissions per station as the input and the performance of each algorithm was verified on several freeway segments.

The decentralized surveillance structure indeed promises a cost-effective approach that could potentially expand the coverage of a surveillance system to remote locations and help improve the decision making process of the operating agencies. As such this report specifically addresses one of the original impetus for this research, Caltrans having insufficient funds to communicate with all of the operational freeway detector stations.

1 INTRODUCTION

Surveillance of transportation networks is important for efficient operation, effective management and safety of the users. For instance, Freeway Management Systems (FMS) help improve mobility by employing traffic detectors to support the traffic management strategies such as ramp meters, traveler information and incident detection [1]. Typically data are collected about the operational condition of the network using spatially distributed sensors and transmitted to a central system for decision-making process. Conventional highway surveillance systems usually employ a fixed-time communication structure, each sensor is polled by a central server to transmit the current state of the facility. Most of the time the response to the data is simply that no action is necessary.

As the freeway network becomes more congested the option of building more freeways is prohibitively expensive both in terms of monetary constraints and social barriers. By adding a relatively inexpensive amount of intelligence one can increase the effective capacity of the freeway system considerably [2]. But this added intelligence has a price. These freeway traffic management systems most commonly collect traffic data using inductive loops. A loop detector station costs approximately \$50,000 - \$60,000 with an additional \$2,000 per year in communication and maintenance costs [3]. Loops are usually installed at 1/3 mile to 1-mile intervals to ensure sufficient coverage and response time. Other surveillance systems that promise to be cheaper than loops have been proposed or are in development. These include the use of radar, microwaves or video sensors that have installation costs in the range of \$12,000 to \$30,000 per location [3]. These systems still require a communications link back to a central system and roadside maintenance, so their ongoing operating costs are similar to loops.

In an attempt to reduce significantly the required communications network and transmission costs per station, this work develops the new approach of a decentralized, event-driven surveillance system in which the first step of decision-making process is distributed from the central system to the field sensors. In this scheme, the sensors in the field transmit data only after an event occurs in the field that may elicit a response from the central server. The central server would then assess information from multiple stations to decide whether to take action (potentially including polling other stations that have not yet initiated communication) or updating the status of the facility. Whether loop or emerging technology, both detection systems are potentially compatible with either centralized or decentralized surveillance.

Through the use of emerging wireless communication technologies, many of which have cost structures that charge per the unit of data transmitted or the amount of time connected, the pre-filtering in the field reduces unnecessary transmissions when no action is warranted; thereby potentially reducing operational costs and improving the benefits from transmitted data without sacrificing the performance of the applications. The savings could then be used to expand coverage of the network and provide surveillance in areas that do not experience recurring events but may have infrequent but severe non-recurring events. Although it is a relatively new area, there has been at least one field operational test along these lines, but the results were poor due to hardware integration problems [4, 5]. Unlike the earlier effort, this report focuses on the communications methodology and quantifies the potential costs using empirical data.

1.1 Centralized versus Decentralized Structure

In the traditional centrally controlled system, the communication network must support a continual communication link between many field controllers and the central server, constantly moving both data and control commands. The local processing power and memory in the field may be minimal, only enough to carry out the commands issued by the central facility and store recent sensor data collected for transmission, while large volumes of raw, unprocessed data would typically be transmitted over the telecommunication network from the field [6]. In a fully distributed system the reverse is true; control is totally localized, with information and control messages transmitted only when major deviations are detected or when special needs exist for changes to the field devices. The decentralized systems rely more extensively on processing resources in the field with greater capabilities filter and analyze the data, thereby reducing the demands on the telecommunications network. Currently such devices are being used for applications like managing vending machines and communications with the central facility is less frequent.

Centralized systems offer conceptually simpler corridor or regional management, since extensive information is available both temporally and spatially for processing. Provided care is taken to ensure that the set of feasible states are known in the absence of communication, the amount of information available from a decentralized communication system can approach that of a centralized system. A decentralized system can also potentially be made to operate like a centralized system when needed to deal with extreme conditions.

The underlying principle in decentralized surveillance structure is that communication is needed only when the data transfer has informational value and can provide additional knowledge about the system. This information is valuable if knowledge of it could result in a response by the central server, most likely located in a traffic management center (TMC), to a change in the current state of the system. Thus, in a relatively simple architecture, when no data are transferred from the field devices, the TMC could assume that the current state is similar to the last reported or expected state of the system. This assumption about the current state can be erroneous when the field device is malfunctioning or when the communication link to the center is broken. A structure where the center polls the field device at a lower frequency when there is no news from the field sensor can help eliminate this uncertainty. Similarly, the field sensors may send data on a lower frequency without polling even when no control actions are warranted. These periodic connections can also be used to transmit summary data, allowing the field detector to support planning applications as well.

1.2 Impact of Decentralized Structure on Traffic Management Applications

The types of data and resolution required for traffic management applications depend on the application and data processing algorithm. The set includes real time applications like freeway incident detection, freeway ramp metering, traveler information system, as well as offline applications like traffic data collection for planning, system performance evaluation and archival or historical purposes [7]. Conventionally these applications are developed and deployed for a centralized surveillance structure. The proposed decentralized surveillance structure for traffic management provides limited information based on events, and the impacts of decentralized structure on the applications is briefly discussed below.

Offline applications usually require aggregated temporal and spatial traffic data. The decentralized communication algorithm can be designed to collect and store data locally and then transmit it to center at low frequency and coarse aggregation. While at the same time, these transmissions can be used by the TMC to verify a station is operational. Hence, provided care is taken, the quality of offline applications is less likely to be affected by decentralized structure.

Real time applications like freeway incident management, traveler information systems, and local freeway metering systems will not be affected significantly if the design of the decentralized surveillance and communication structure is done appropriately. These applications require either event specific data or data based on local conditions and the decentralized structure can be designed to accommodate them. In fact this report will develop and test several algorithms and communication scenarios that are capable of supporting these real time applications. Other real time applications are not likely to function well under distributed surveillance. Coordinated ramp metering requires data about an extended section of freeway to adjust the metering rate at each ramp while the relevant information is not likely to be apparent at the local detector station level. At the extreme, vehicle re-identification algorithms require data from each individual vehicle, both temporal and spatially distributed, to calculate link travel time. The decentralized structure developed in this paper will not be able to support these applications and other applications that explicitly require continuous temporal and spatial data. However, additional processing power in the field coupled with peer-to-peer communication could potentially be able to support the data needs of such applications.

1.3 Data Sets

The data used for development, testing and analysis of methodologies developed in this report were primarily obtained from six freeway corridors as described below. The data from each corridor were collected using inductance loop detectors. Flow, speed, and occupancy data were aggregated over 30-second intervals for our analysis. The diverse set of freeways, from rural and urban locations, allowed us to make the methodology robust and widely applicable.

- I-70/71_CMH: The Interstate-70/71 corridor in Columbus, OH is instrumented with single and dual loop detectors at 1/3 mile spacing over approximate 14 miles. Data were collected from 40 detector stations during the months of January and February 2002. A few days in this data set were affected by severe weather conditions and this fact allowed us to test our methodology for the resulting impacts of anomalous days.
- I-80_BHL: The segment of Interstate-80 through Berkeley and Emeryville located just north of Oakland, CA is known as Berkeley Highway Laboratory (BHL) [8]. The BHL has eight dual loop detector stations along 2.7 miles of freeway with five lanes in each direction. Data collected from four detector stations, eastbound and westbound, during August 1998 were used for our analysis. This urban corridor experiences extended periods of recurring congestion during weekdays.

The following four freeway corridors were also selected from the California Department of Transportation (Caltrans) Districts 3, 4, 7, and 11. The data for these districts were collected via Freeway Performance Measurement System (PeMS) [9].

- SR-51_D3: The corridor selected for District 3 is State Route 51 between absolute post-miles¹ 3 and 8 through Sacramento. Data were collected from eight detector stations in each direction (northbound and southbound) for five weekdays from 4th August 2003 to 8th August 2003. The majority of this corridor spans a rural area.
- SR-101_D4: The corridor selected for District 4 is State Route 101 between absolute post-miles 27 and 112. Data were collected from 97 detector stations between San Jose, CA and Novato, CA (north of San Francisco) in each direction (northbound and southbound) for five weekdays between 10th March 2003 and 14th March 2003. A few stations on this corridor observe extended periods of congestion.
- I-405_D7: The Interstate 405 through Los Angeles/Venture in District 7 between absolute post-miles 25.28 and 64.33 has 53 detector stations. Data on this urban corridor were collected on five weekdays between 31st July 2003 and 4th August 2003.
- I-15_D11: The corridor selected for District 11 is Interstate 15 between absolute post-miles 4.6 and 26 through San Diego County. Data were collected from 18 detector stations on each directional freeway (northbound and southbound) for five weekdays between 10th March 2003 and 14th March 2003. This corridor observes a short period of recurring congestion during morning and evening peak hours.

1.4 Overview

To reduce communication costs without a significant loss in the quality of information received at the TMC, Chapter 2 lays the foundation for an event driven communication system by examining the information value of specific events at the detectors. This procedure can then be used to transfer a portion of the decision making from the TMC to the field controllers. We develop the methodology by deconstructing several incidents on a freeway and identify the observable events at a pair of detector stations that may be upstream, downstream or straddle the incident. This analytical process could be repeated for any other condition of interest. Chapter 3 explains the importance of reliable detectors and clean data for distributed surveillance. It is followed by a brief literature review of current detector diagnostic and data cleaning practice. Detector diagnostic, data cleaning and speed estimation algorithms are then developed and analyzed for freeway detectors. These advances can improve the performance of conventional surveillance system as well as enabling efficient distributed surveillance. Chapter 4 shows that the communication needs are a function of application and location. It then develops algorithms for different communication scenarios followed by comparison, usefulness and implementation of each scenario in the field. The need for a reliable communication structure is also explained. Chapter 5 provides an introduction to different modes of communication and the currently available technology. It then develops a communication protocol and cost structure for each mode. This cost structure is applied to various communication scenarios. This discussion is followed by a discussion of the application of surveillance algorithms and cost structure to different freeways and segments. Chapter 6 summarizes the findings and conclusions of this work.

¹ The PeMS system converts the Caltrans' post-miles into absolute post-miles. The absolute post-mile is the actual centerline distance down the freeway from the district boundary.

2 DISTRIBUTED SURVEILLANCE ON FREEWAYS WITH AN EMPHASIS ON INCIDENT DETECTION AND VERIFICATION

As urban roadways become more congested, greater emphasis is being placed on active monitoring of the transportation network. Typically these surveillance systems use loop detectors or similar sensors to monitor traffic flow at discrete points in the network. The data from these sensors are normally transmitted to a Traffic Management Center (TMC) at regular intervals for assessment. Most of the time no action is taken in response to the data, leading to unnecessarily high communication costs. Deviating from this centrally polled communications architecture, we consider information value from the standpoint of distributed decision making, e.g., having the remote detector stations initiate communications when local conditions indicate the potential need for a response from the TMC. By knowing the information value of events in the field, one could develop a distributed decision making process and greatly reduce the amount of data transfer needed to operate the system and potentially reduce communications costs. Such an approach has been employed in many other disciplines, but has seen little use in highway traffic monitoring. Perhaps more importantly, by establishing the value of information, the approach presented herein could allow for more efficient data handling for decision-making processes and archiving in a conventional, centrally polled communication system. Though incident detection on freeways using loop detectors is emphasized in this chapter, the principles can be applied to other monitoring applications and sensors. The main motivation of this research is to reduce the frequency of communications (and thus, potentially cost) without significant impacts to the decision making process.

This chapter presents a simple methodology to identify observable events arising from an incident at an unknown location on a homogeneous freeway with two loop detector stations. After accommodating for noise and measurement uncertainty, e.g., as will be presented in Chapter 3, these events can be used to filter the data at the field instead of the TMC, and preempt the transmission of non-informative data. Naturally, this 'filtering' of data to determine the benefit requires intelligence in the field to understand the decision making process at the TMC. Implicit in the distributed communication approach is that "no news is good news," so periodic transmissions will be necessary to confirm the operational status of a station that does not see any critical events. Such transmissions can also be used to transmit summary data at a coarse aggregation, thereby allowing the station to provide a temporally comprehensive overview of conditions at the location without a continuous communication link. This latter point is significant, since it will enable planning applications that require continuous monitoring, e.g., average daily travel (ADT).

There are a few existing examples of distributed surveillance applied to traffic monitoring, the most common one being the use of cellular phones for incident detection. Many operating agencies have developed call-in programs to accept cellular phone calls by motorists to report incidents. These wireless devices act as probes in the field, communicating information to a central decision making system on an event driven basis. But coverage can be uneven, since this information is voluntarily provided by the motorists a given incident may be reported multiple times or not at all. It may also be difficult to establish precisely the location or magnitude of the event.

There have been a few applications of conventional detectors in a distributed surveillance system, e.g., [10] deals with anomaly detection as opposed to incident detection to lower the

communication costs. The authors proposed that loop detectors in the field communicate to a TMC only when an anomaly is detected and then, transfer data at a higher frequency to aid the center in making a decision about the traffic states. Their theoretical model includes a modified McMaster incident detection algorithm [11] with an online calibration technique based on spatial-based averages. There has also been at least one field operational test but the results were poor due to hardware integration problems [4, 5]. Unlike these earlier efforts, this chapter focuses on identifying the key events that would facilitate decision-making.

2.1 Distributed Surveillance

This section both provides motivation for the distributed surveillance concepts and a description of the analytical methodology in the context of detecting incidents. It begins with a brief discussion on the value of information in a general context, providing a framework for decision-making later in the chapter. Next, the discussion turns to incident detection and observable events both before and after an incident occurs. Some simplifications are necessary to emphasize the critical events in the process, e.g., traffic only assumes a finite number of traffic states. These simplifications do not preclude practical application, rather, they bypass otherwise prohibitive nuances. The entire set of observable traffic states at a pair of detectors adjacent to an incident on a homogeneous segment is then examined, along with the temporal progression. This information is used to develop distributed decision-making to determine what information is important and should be transmitted. The work is then extended to inhomogeneous segments with a recurring bottleneck. To control for the additional source of queuing, this section will present a discussion to identify and locate recurring bottlenecks and guidelines on measuring bottleneck capacity. After completing this development, the impacts of the initial assumptions are considered before presenting the conclusions of this work.

2.1.1 Value of Information

One must know the value of information before deploying a distributed surveillance system. This knowledge is vital in deciding the benefits resulting from actions taken at the TMC in response to the events transmitted from the field and thus, the utility of transmitting specific data. In centralized surveillance many of the data transfers result in no action, i.e., the additional knowledge gained due to the data from the field does not result in a change in the decision about system.

Consider the benefit of additional information from the field to the TMC over the existing knowledge at the center (historical trends from previous days, recently reported values or combination of both), in the perspective of decision-making process. *Information* refers to the reduction in uncertainty about the state of an event after a message has been sent relative to the uncertainty about the state of the event before the message was sent [12, 13]. Along the same lines, *Information value* refers to the difference in what one can gain by action taken knowing the state of the event relative to what one can gain by action taken without such knowledge [14].

To understand the characteristics of the value of information, which can take the form of a benefit function, consider a simple one-detector problem. Given a sensor whose finest sampling rate is T and can report to TMC in nT , where n is a positive integer. Let the measurement be $Ma(t)$ and for simplicity, assume that the expected value at the TMC in absence of more recent

measurements is just that of the last transmitted observation at $Ma(t-kT)$. If the TMC were to employ more sophisticated techniques to better estimate $Ma(t)$ since the last report, $Ma(t-kT)$, the system need only be modified to include these estimation processes in the field controller as well. The discussion is deliberately abstract, allowing for the possibility that $Ma(t)$ may be designed to include additional information, e.g., statistical properties of the recent measurements. Figure 2.1 illustrates the value of information in the difference between the observed value at the field and the known value at the TMC. The error bound, $eb(Ma(t-kT),k,\dots)$, is a measurable function that maps the uncertainty for every feasible measurement point. An $Ma(t)$ outside of the error bound would indicate that conditions on the roadway have changed since the last report to the TMC. Hypothetical action-bounds, $a_i(k,\dots)$, are shown in this plot, indicating when $Ma(t)$ would dictate some remedial action or confirm that the state has reached some other critical point. Anything outside the action bound results in no action, by definition.

2.1.2 Incident detection and observable events

Two signals emanate from an incident, a backward moving shock wave and a forward moving drop in flow [15]. Reliable detection of an incident using point measurements can happen only when both of the signals have been received at detector stations straddling the incident. The first stage of a distributed surveillance algorithm must nonetheless operate in the field using data from a single detector station. No communication would be made until some anomaly is observed, at which point communication would be made with the TMC to allow it to integrate information from adjacent stations and localize problems, e.g., distinguishing between recurring congestion from a known bottleneck and an incident. In the absence of such events, the TMC and field controller would share a common history of transmitted data and the field controller may mimic subsequent processing conducted at the TMC.

This section develops a representation of detectable events arising from an incident to aid in understanding the incident detection decision-making process. The underlying concept is a simple, yet effective model of representing the progression of the traffic state over time. Using this concept, we represent the progression of an incident on a homogenous freeway segment on a 'time progression chart', as defined below. This chart gives insights to develop a set of feasible controls to aid the decision-making process at the TMC, representing graphically how the traffic states evolve over time. The development is limited to isolated incidents on a freeway segment without ramps, but the procedure could easily be expanded to include such features.

In this construction, first consider a traditional time-space diagram depicting the evolution of traffic states on a homogeneous freeway segment, showing both what is observable at the detector stations and what is not. The presentation is made with the simplifying assumption that the observed traffic states on the freeway segment are discrete, stable and can be distinguished from one another. In this fashion, the problem is rarefied down to the fundamentally observable events, with the real measurement uncertainty being captured by the error bound discussed earlier. This presentation is followed by the approach to construct time progression charts to track the evolution of the traffic states at adjacent stations in presence or absence of an incident. Finally, the chapter closes with further discussion on accommodating the noise in real detector data.

Figure 2.2 shows the time-space diagram corresponding to an isolated incident on the homogenous segment and the corresponding flow-density diagram. The interface between two

traffic states in the time-space diagram is modeled using the theories of Lighthill, Whitham, and Richards [16, 17]. For convenience in the following discussion, the flow-density relationship is assumed triangular [18]. The results still apply if the relationship takes another shape, but the signals will propagate at different speeds. It is assumed that there are five discrete states of flow: free flow (F0, F1, and F2), congested (X1), and capacity (C). The flows at states F0, F1, and F2 increase in that order, with F0 is taken as the initial demand of the freeway segment. The flow at states F1 and X1 are equal to the capacity of the incident, corresponding respectively to conditions immediately upstream and downstream of it. Finally, flow at state F2 represents an arbitrary increase in demand from flow at F0 in the homogenous freeway segment prior to the incident. The focus is on delay-inducing incidents, so the capacity of the incident is chosen to be below the demand in the freeway segment at the time and location of its occurrence. If there is no flow restriction due to an incident, then it is unlikely that it will disrupt traffic enough to be easily discernable, but similarly, if it does not restrict flow, it will cause little traveler delay.

The time-space diagram of Figure 2.2 shows an incident at (t_i, d_i) . After being in state F2, the service is restricted by the incident capacity at F1 and this state is subsequently observed downstream of the incident. A queue forms upstream of the resulting bottleneck in state X1, the rate the queue grows depends on the severity of the incident, i.e., capacity of the incident relative to the demand on the segment. For simplicity of presenting the relationships, the incident is assumed to have a constant capacity (flow at state F1) and the demand (flow at state F2) does not change after the incident occurs.

To capture all possible relative relationships between a pair of detectors and an incident, as will be discussed shortly, four possible detector locations are considered and they are represented in Figure 2.2A as Detectors A-D. Far upstream of the queue vehicles remain free flowing at F2, and in this case demand exceeds incident capacity, X1. The backward moving shockwave traces the tail of the queue, in which flow drops and density increases. The queue grows upstream with an interface velocity corresponding to the slope of the line segment F2-X1 in the flow-density diagram, represented as dashed lines connecting the two traffic states. The incident starves demand downstream, causing both flow and density to drop, and this forward moving interface travels with a velocity equal to the slope of the line segment F2-F1 in the flow-density diagram. Once the incident is cleared, the queue discharges at the maximum service rate, the capacity of the freeway segment. This interface velocity corresponds to the slope of the line segment X1-C. Similarly, the interface velocity representing the raise in flow downstream of the incident corresponds to the slope of the line segment F1-C. Once the entire queue is discharged, traffic returns to the pre-incident states.

2.1.3 Time progression chart for an isolated incident on a freeway

Since the detectors on the freeway segment are located at discrete points, the entire time-space diagram cannot be observed or inferred from a single detector station's data alone. Furthermore, an incident's location relative to any given pair of detectors cannot be known a priori. Instead, we consider the progression of traffic states for any pair of detectors, A-D, from Figure 2.2A to facilitate this understanding, i.e., both stations may be upstream or downstream of the incident, or the pair of stations may straddle it. We develop a time progression chart to describe the traffic state evolution for all six combinations of detector and incident locations and aid the decision-making process. The time progression chart further condenses the time-space diagram to the states observed by *any* pair of fixed-point detectors in the field. Consider the time progression

chart shown in Figure 2.3, which illustrates the possible scenarios that any pair of detectors will observe if an incident happens on the homogenous freeway segment. Each unique combination of the states observed by any pair of detectors is shown in a separate box, denoting the (upstream detector traffic state/downstream detector traffic state). Arrows indicate the transition from one pair of states to another with all feasible combinations considered. So in this case, the bottom left-most box contains (F0/F0), which indicates traffic state F0 is observed initially at both the upstream and downstream detector for any pair of detector present in the homogeneous freeway segment. Then due to raise in demand, the upstream detector rises first, as shown in the adjacent box (F2/F0), and the signal propagates with the flow of traffic, impacting the downstream detector in the next box, (F2/F2). For any pair of detectors in Figure 2.2A, these transitions occur prior to time t_1 .

There are four arrows emanating from (F2/F2). If there is no incident, the state remains (F2/F2) and when the demand drops, the state changes to (F0/F2) as shown with the dashed arrow. If an incident occurs upstream of both the detectors, then the state progresses from (F2/F2) to (F1/F2) as the drop in demand propagates downstream from the incident, i.e., detector locations C and D in Figure 2.2. But again, the incident location or even its occurrence is not known a priori. So next we consider the possibility that the incident occurs downstream of both detectors, then the state progresses from (F2/F2) to (F2/X1), as the queue grows upstream from the incident, i.e., detector locations A and B in Figure 2.2A. If the incident is severe enough that the end of queue reaches the further upstream detector, the state changes from (F2/X1) to (X1/X1). On the other hand, until the shockwave reaches the downstream detector location, no state change from (F2/F2) can be observed from this pair. Finally, if the detectors straddle the incident, e.g., detector locations B and C in Figure 2.2A, then the state is shown as progressing from (F2/F2) to (F2/F1) since the slow moving shockwave will usually reach the upstream station after the fast moving drop in flow reaches the downstream station. This exercise can be continued to obtain the recovery states on the right hand side of Figure 2.3.

The chart captures the history leading to a given pair of traffic states, rather than just the current traffic state. In addition, some paths are highlighted with solid squares indicating where historic information about the freeway segment and traffic states (e.g., characteristics of the particular freeway segment, temporal variations of traffic like time-of-day, day-of-week etc.) is particularly useful to distinguish between alternatives. Similarly, stars highlight paths where historic information is particularly helpful to identify the fact that the state has progressed from one box to the next, e.g., to differentiate flow downstream of an incident from the normal drop in flow at the end of the peak period. Long-lived traffic state combinations that may be sustained for an extended period are shown with bold outlines while short-lived states, such as F2/F0 during the transition from F0/F0 to F2/F2, are shown with thinner outlines. Naturally, the precise time for the signal to reach a detector depends on the relative location of an incident to the detector location as well as the severity of an incident.

2.1.4 Decision making under distributed surveillance

The time progression chart captures the observable traffic states. After this calibration one would use the chart to determine the freeway conditions from observations at each consecutive pair of detector stations. Under distributed surveillance, intelligence is required in the field to make decisions on communicating the observed data to the TMC, including what data to send and when to initiate communication. The time progression chart can be used to help answer these

questions by constructing a feasible set of control points, as will be defined below, under various cost scenarios. The set of feasible controls acts as a proxy for the value of information of observed data and any control has a set of decision states, i.e., the response by the traffic control system to observed events. The initial decisions must occur in the field in a distributed surveillance system. To this end, each box in Figure 2.3 is grouped into one of five decision states: No Action, Suspect Incident, Confirm Incident, Recovery, and Post-recovery, as noted at the bottom of each box. Consider the progression from (F0/F0) to (F2/F0) on the left hand side of the figure, since this change simply shows a rise in demand, No Action is taken. On the other hand, the progression from (F2/F2) to (F2/X1) indicates that the downstream detector observes queued traffic from an obstruction further downstream, and hence the decision state indicates Confirm Incident. Similar reasoning is applied to the remaining boxes and reflects the typical actions appropriate for the states that are observed in the field. The assignment of decision states includes the history of feasible paths, i.e., the value of the current state depends on the previous states.

After establishing the decision states, control points are the set of measurements observed by the field that are chosen to trigger a shift of decision-making from the field to the TMC. They are characterized by the deviation from the expected set of measurements, indicating an anomaly in the field (though not necessarily an incident). The control points can be established by tracing the feasible paths in time progression chart while considering the traffic state observed at only *one* detector at a time. In this work, control points are denoted by a set of traffic states at a given detector, [expected state, measured state], e.g., if the historical trends or the last reported traffic state is F2 and the observed traffic state in field becomes X1, the control point is denoted by [F2, X1]. Figure 2.4 is analogous to Figure 2.1, only now it shows the control points based on their observed field values $M_a(t)$ and expected measurement values $M_a(t-kT)$ from the time progression chart, Figure 2.3. These transition points would occur relatively infrequently since successive observations of the traffic state will normally remain at the same state, i.e., most data will fall on the line at 45 degrees. In practice, one would likely want to extend the progression chart and include control points at [not X1, X1] to capture queues arising from incidents when demand is below F2 and perhaps [X1, not X1] to capture an upstream drop in demand.

Depending on the cost of a single transmission, control actions can be chosen selectively from the feasible set of controls. Table 2.1 lists the control points when the time progression chart is reduced to observations from a single station. A sub-set of control points could be chosen depending on the cost per transmission incurred. When the cost of initiating individual connections dominate the cost per unit of data sent, any intermediate states from Table 2.1 that are not transmitted could be sent to the center along with the next state when the next transmission occurs.

2.1.5 Incidents in the Presence of Recurring Bottlenecks

Thus far, this chapter has considered a homogeneous freeway segment, with no other capacity constraints beyond an incident. Although queuing is a key indicator of an incident, obviously queuing can also arise upstream of a recurring bottleneck on inhomogeneous segments. Any practical solution must be able to differentiate between recurring congestion and an incident. This section examines the interaction between a recurring bottleneck and an incident. Building off of the simplified model used above, it is assumed that an inhomogeneous segment can be modeled with a single recurring, point bottleneck limiting flow in the absence of any incident.

Now it is assumed that there are seven discrete states of flow: free flow (F0, F1, F3, and F4), congested (X1 and X3), and capacity (C). Once more, flow increases with increasing number, and the flow is the same for a given number whether free flow (F) or congested (X), e.g., the states F3 and X3 correspond to the recurring bottleneck capacity. An incident must restrict flow below this level to impact the segment throughput and we arbitrarily denote such a restriction with F1 and X1, corresponding respectively to conditions immediately upstream and downstream of the incident. Flow at state F4 represents an arbitrary increase in demand above the capacity of the recurring bottleneck, prior to the incident. Finally, note that state F2 is omitted to avoid any confusion with the arbitrary increase in flow used in the homogeneous example.

In the absence of an incident, when demand climbs to state F4, a queue would build up behind the recurring bottleneck until demand drops below that of state F3, at which point, the queue would recede back towards the bottleneck and eventually dissipate. In the presence of queuing, conditions downstream of the recurring bottleneck would be expected to be at F3. The relationship becomes more complicated when an incident limits flow below that of state F3. Figure 2.5 shows the time-space diagram corresponding to an incident upstream of a recurring bottleneck at t_1 . The construction of the diagram follows the same process used in Figure 2.2. Notice how the incident starves demand at the recurring bottleneck when it limits downstream flow to that of state F1. Given the two restrictions, an arbitrary pair of detector stations could both be downstream, between, or upstream of both restrictions; or, the detector stations might straddle one or both of the restrictions. To capture all of these combinations, six possible detector locations are shown in the figure, denoted E-J. Alternatively, an incident may occur downstream of the recurring bottleneck, as shown at t_1 in the time-space diagram of Figure 2.6. Once more, one needs to consider six detector station placements to capture all possible relationships between an arbitrary pair of detectors and the two restrictions. Highlighting the fact that a given recurring bottleneck may be downstream of one incident and upstream of another, this figure repeats the detector station labels G-J with the same relative positions to the recurring bottleneck from Figure 2.5, while adding stations K and L downstream of the incident.

As with the homogeneous example, this section develops time progression charts for an incident at an arbitrary location relative to any pair of detector stations in Figure 2.5 and 2.6. Presumably, the recurring bottleneck location would be known to an operating agency, so the analysis is split across three different time progression charts based on the detector locations relative to the recurring bottleneck. First, consider the case where both of the detector stations are upstream of the recurring bottleneck, i.e., any two stations from E-H in Figure 2.5 and 2.6. This progression chart is shown in Figure 2.7 and uses information from the two time-space diagrams. For example, if it turned out that both detectors were downstream of the incident (stations G and H in Figure 2.5), the combined traffic state could progress as follows: (F0/F0) as the initial state, (F4/F0) and (F4/F4) as demand rises, then (F1/F4) as the incident starves supply, and finally settling in (F1/F1) until the incident is cleared. The drop in flow suggests that an incident may have occurred, but given the fact that traffic remained free flowing at both stations, the possibility that demand simply waned cannot be eliminated. Thus, the decision state is denoted as Suspect Incident. On the other hand, if the incident occurred downstream of the recurring bottleneck, the combined traffic state for the same two stations (this time as shown in Figure 2.6) could progress as follows: (F0/F0) as the initial state, (F4/F0) and (F4/F4) as demand rises, then (F4/X3) as the queue from the recurring bottleneck overruns station H, and finally passing to (F4/X1) as the queue from the incident overruns station H. Provided the capacity of the bottleneck is known, traffic state X1 with a lower flow would clearly indicate the incident,

hence, this decision state is denoted as Confirm Incident. This process is repeated for each of the detector station pairs upstream of the recurring bottleneck to complete the progression chart in Figure 2.7. The procedure is then repeated for the case where the detectors straddle the recurring bottleneck, Figure 2.8, and when both are downstream of it, Figure 2.9, in each case combining information from Figure 2.5 and 2.6.

Once more, after establishing the decision states, the control points can be chosen to trigger a shift of decision-making from the field to the TMC. As before, they are established by tracing the feasible paths in time progression charts while considering the traffic state observed at only *one* detector at a time and depending on the cost of a single transmission, control actions could be chosen selectively from the feasible set of controls. Combining information from all three of the progression charts for the inhomogeneous segment, Table 2.2 lists the control points when the time progression chart is reduced to observations from an individual station upstream of the recurring bottleneck, and Table 2.3 repeats the process for an individual station downstream of the bottleneck.

2.1.6 Identifying Recurring Bottlenecks

Distinguishing recurring congestion from an incident as well as the location of the incident using detector data can be difficult. To this end it is important to know the location of the bottlenecks and their capacity before applying the time progression charts to an inhomogeneous segment. Historical traffic data can be used to identify and locate the existing bottlenecks, e.g., a contour plot of velocity over time and space. Time-series plots of occupancy and velocity can also be used to localize the bottlenecks. It is important to analyze the traffic conditions over several days to establish a trend and confirm the existence of the bottlenecks in the segment [19]. If there are no existing loop detectors present on the freeway segment under consideration, one can employ probe vehicles to identify and locate the bottlenecks in the system.

Once the presence of recurring bottlenecks and their location is established, the capacity of a given bottleneck can be approximated with historic cumulative arrival curves from the adjacent detector stations. The slope of this curve is the instantaneous flow and an extended sampling period can be used to reduce the impact of transient fluctuations. When the bottleneck is active, the maximum observed flow at the downstream station is its capacity. Care should be taken to use the station downstream of the bottleneck or restrict analysis to queued conditions at a station upstream of the bottleneck, since a station upstream can observe sustained flows in excess of the bottleneck capacity prior to the onset of queuing, e.g., state F4 at detector E in Figure 2.5 has a flow in excess of the bottleneck capacity. Naturally, multiple days should be studied to reduce the chance that incident conditions are accidentally included in this calibration.

2.1.7 Practical Issues

There are several practical issues that must be considered for a distributed surveillance system, including: the benefit of integrating information across adjacent stations, relaxing the assumptions used in this work, and accommodating noisy measurements. Stepping through these issues one at a time, information from adjacent detectors is useful in the decision-making process at the TMC, it complements the temporal information from a single station with the spatial knowledge along the freeway. This information is useful to confirm an anomaly observed at a

single station, to predict a trend or anticipate the state of the larger system. In the proposed distributed surveillance system, a detector communicates with the TMC when it observes a control point and thereby, transfers the decision making process from the field to the TMC. The TMC must integrate the information from several detector stations before reaching a conclusive decision on the course of action. Here, information from adjacent detectors (upstream and downstream) is useful to get a better understanding of what is happening, e.g., the information is typically valuable in confirming incidents and distinguishing them from recurring congestion or a drop in demand. In a centrally polled communication system, this information from adjacent detector stations is readily available, allowing for easy comparison of input-output on a link, as well as velocity differences and occupancy differences between stations. In a distributed surveillance system, the initial decisions must be made without such comparisons between stations. However, the performance of the distributed surveillance can approach that of centralized surveillance if after one detector station initiates communication with the TMC, the TMC can poll the adjacent stations to collect additional information.

In the construction of a time progression chart, it was assumed that the traffic states observed were discrete and stationary in nature. However, traffic states are a continuum. The characteristics that a given traffic state exhibit in this analysis, (e.g., F1 in Figure 2.2) may be representative of a continuous range of traffic states. For decision-making, however, it remains important to distinguish those traffic states that can be differentiated from other traffic states by the sensors in the field. This concept plays a vital role, since the field measurements are noisy due to transient events, caused by either inherent traffic characteristics or detector errors. Discernable traffic states play a key role in assessing the current status of the freeway without ambiguity, thereby leading to decisions to return conditions to the desired status. The discrete traffic states used in a time progression chart can potentially be a sub-set of the discernable traffic states. It is possible to empirically establish the bounds between some of the discrete traffic states, e.g., F0 and C in Figure 2.2, while others may be infeasible given the noise and transient errors in the field measurements, e.g., F1 and F2 in Figure 2.2 may be too close on the continuum. However, the key problem of distinguishing between free flow and queued states is generally feasible with most traffic detectors. Although this task is complicated by the typically noisy measurements, several approaches have been developed for reducing the impact in aggregate measurements, including: diagnosing detectors for malfunctions [20, 21], validating speed estimates using flow-occupancy relationships, exploiting the redundant information from adjacent lanes information (as will be discussed in Chapter 3).

2.1.8 Effect of Distributed Surveillance on Time Critical Applications

In time critical applications like Incident Detection and Management, the effectiveness of the application depends on the time taken to respond to events of interest. For incident management, response time is defined as the time elapsed from the incident (or anomaly) in the field to the time some remedial action is taken to rectify the environment. In other words, response time is the sum of time taken to detect and confirm the incident on the freeway (detection time), and time taken to clear the incident that is restricting traffic conditions (clearance time). Consider a network of loop detectors on a freeway segment. In the ideal case, without detector errors, response time includes the time for the signal to reach a detector, time to reach a decision to transmit, and time for the remedial action after the decision is made. Relaxing the idealization by introducing noise and detector errors, consider two modes of surveillance - Centralized

Surveillance and Distributed Surveillance. In the centralized surveillance, field measurements are polled from a central decision maker at fixed interval of time (often 30-sec data). The central decision-maker obtains data from all the detectors in the field and then processes the information to see if there is any anomaly or suspicious trend in the field. Here response time can be defined as,

$$RT_c = T_w + T_c + T_n + T_d + T_r \quad (2.1)$$

Where,

- RT_c = response time under central decision making
- T_w = time taken for the signals to reach the detector. T_w is a function of the wave speed, spacing of the detector, location of incident with respect to the detector and sensitivity of the detector.
- T_c = time lost in communicating from the field to the center. In a typical polling system, on average it will be the half of the fixed time interval.
- T_n = Imposed delay at the TMC to establish a trend to detect events of interest from noise aided by the spatial knowledge of the adjacent detectors.
- T_d = time spent in decision-making process to confirm an anomaly (or incident) using past historic trends or other surveillance devices.
- T_r = time taken for the actual remedial process once the decision is made

In central decision-making system, T_w and T_c occur in the field and the other T 's happen in the central decision maker. In addition, we assume that computational delay is negligible.

Under distributed surveillance, Response time can be given as,

$$RT_d = T_w + T_c + T'n + T_d + T_r \quad (2.2)$$

Where,

- RT_d = response time under distributed surveillance
- T_d = time spent in decision-making process to confirm an anomaly (or incident) in distributed surveillance.
- $T'n$ = Imposed delay in the field to establish a trend to detect events of interest from noise and is dependent on the thresholds to detect an anomaly in the field.

In Equations 2.1 and 2.2, T_w is decision independent, and is inherent to the nature of the anomaly and the detector characteristics at the field. T_c in centralized surveillance will be the half of the fixed time interval (on average). T_c under distributed surveillance may go to zero in many cases. However, T_c may take the form of time spent in the TMC communicating with adjacent detector stations in distributed surveillance. For simplicity, T_c under both modes of surveillance is assumed the same. Similarly, if the same response is given by either mode then, T_r , the time to respond as the decision is made, is likely to be independent of communication protocol. If we assume that, other surveillance techniques available at the center, (e.g., video camera) and human expertise is same in both of the decision processes, T_d will also be the same in both decision processes. Hence,

$$RT_d - RT_c = T'n - T_n \quad (2.3)$$

$T'n$ is the time spent (or imposed) to identify discernable data from the noise at the *field*, usually about few samples. T_n contains the amount of imposed delay (or wait) at the TMC to detect events of interest from noise and depends on the method of data aggregation and accepted

percentage of false calls. However, the main difference is that during the waiting time (T_n), the TMC has spatial knowledge. If a trend can be established due to the availability of the spatial knowledge, T_n will be smaller than $T'n$. If the conservation of flow exists in the freeway segment under consideration, using cumulative arrivals, one can quickly establish the trend indicating an onset of congestion. Therefore, any decision that can benefit from spatial spread of knowledge is likely to be made quicker in the central surveillance scheme, since no time is spent on waiting for $T'n$ at the field to differentiate events of interest from noise.

In distributed surveillance, the first steps of the decision-making process, i.e., detecting the discernable events of interest, is achieved to an extent at the field. However, that decision process is location-specific and does not incorporate the spatial knowledge about the adjacent detector locations. Therefore, we have to integrate the spatial knowledge at the center. In case of an anomaly (or incident), the detector immediately upstream will be the first to detect a discernable event, as the queued state will be first formed at the upstream detector location of the incident, thereby reducing the dependence on integrating the spatial knowledge.

When considering only a single station, the response times under both centralized and distributed surveillance should be similar. Across multiple stations, distributed surveillance can approach the performance of centralized surveillance by passing control to the TMC with a lower threshold, so that the TMC can poll adjacent stations.

Now, let us consider $T'n$, which is the time spent at the field to identify discernable events before communicating with the central decision-maker. This waiting incurs a cost when there really is an event of interest and can be associated with the rate of change of delay. The rate of change of delay can be calculated from a cumulative arrival curve, total number of vehicles passing a given location as a function of time, as shown in Figure 2.10.

Let the capacity of the freeway location be f_c and the initial demand on the freeway be f_e . Consider an incident of capacity f_m and duration t . We are interested in the rate of change of delay since if we detect the incident Δt earlier, then we should be able to clear it Δt earlier because of all the other time components in Equation 2.1 and 2.2 being fixed. From queuing theory, the delay, $D(t)$, incurred by a total clearance time t , as shown in Figure 2.10, is given by,

$$D(t) = \frac{t^2(f_e - f_m)(f_c - f_m)}{2(f_c - f_e)} \quad (2.4)$$

Similarly, delay incurred if the incident is removed at time $t + \Delta t$ is given by,

$$D(t + \Delta t) = \frac{(t + \Delta t)^2(f_e - f_m)(f_c - f_m)}{2(f_c - f_e)} \quad (2.5)$$

For comparison purposes, if demand and capacity of the incident remain constant over the extended time Δt , we can calculate the rate of change of delay, $\Delta D / \Delta t$, as,

$$\Delta D / \Delta t = (2t + \Delta t) \frac{(f_e - f_m)(f_c - f_m)}{2(f_c - f_e)} \quad (2.6)$$

From Equation 2.6, it can be seen that as the incident duration increases delay incurred increases at a higher rate due to the presence of Δt in the equation. This result illustrates the need for quick identification and response to events that cause unexpected delay in the system, like

incidents. Note that lower communication costs could allow a higher density of detectors and thereby reduce the average T_w in Equation 2.2. In addition, the above equation gives a quantitative approach towards establishing the acceptable waiting time T_n at the field before initiating the communication with the center for distributed surveillance. Of course, a proactive response could reduce the demand and thus, the delay. Such second order effects make it impossible to do a perfect one-on-one comparison, but the reader can adjust accordingly.

2.1.9 Effect of Distributed Surveillance on Off-line Applications

The impacts to off-line applications based on data from distributed surveillance will typically be less than the on-line applications. Typical offline applications generally require only coarse aggregates of traffic estimates from the field. These data can be collected via summary reports sent in a concise form to the TMC on a regular basis (say daily, or if need be, over several hours). As noted earlier, sending summary data also provides a check to verify that the detection system is fully functional, especially at detector stations that do not observe any events that would otherwise be reportable and would trigger the station to initiate communication. Resolution of the archived data from the field depends on the desired application, e.g., for measuring ADT on a freeway segment, coarse aggregate of traffic volume over a day would suffice. If the application were to observe the flow on the exit ramp feeding to an arterial during the peak hour, higher resolution data would be required during the periods of interest. Coarse estimates like VMT, ADT will be the same irrespective of the mode of the surveillance used.

2.2 Summary

Most traffic surveillance systems have been deployed with simple communication network architecture, detector stations are polled regularly from a TMC. This report proposes an alternative approach that would start evaluating data in the field, before making a decision to transmit data. In the short term, such a process could allow communication using limited bandwidth, wireless connections. Because communications costs often dominate other operating costs for traffic surveillance systems, this approach could be used to inexpensively extend monitoring to freeway links that do not normally see congestion and are not cost effective to monitor using conventional strategies.

Given the rapidly improving wireless communication networks, in the long run perhaps the cost benefit will disappear and the biggest impact of this research will be the process of identifying key events in traffic data using a finite number of states to reduce the impact of noisy data on the analysis, aiding the study of the temporal progression in a meaningful manner that facilitates an understanding of the system. This fact, in turn, could be used to streamline many control processes even in a conventional, centrally polled surveillance system. The fundamental elements of this process can be broken down into three steps, first establish the space of traffic states on a time space diagram using only enough discrete traffic states as necessary while adding enough detectors to capture all feasible sets of states observed from any pair of detector stations. In more complicated scenarios, such as an incident in the presence of a recurring bottleneck, one may have to develop several time space diagrams to capture the entire set of observable states, e.g., Figure 2.5 and 2.6. Second, trace each pair of detectors to identify the evolution of the traffic states and record this information on a time progression chart. The chart should be constructed to clearly delineate events by information value, e.g., 'No Action' versus

'Confirm Incident'. Finally, control points can be identified from the time progression chart and the responsiveness can be determined as a function of communication cost. If these costs change, the degree of responsiveness can be adjusted accordingly. Of course this final step requires accommodating the continuous spectrum of traffic states and noisy measurements. The impact of the latter point can be reduced through data cleaning techniques, e.g., as will be discussed in Chapter 3, while both points are addressed with the error bounds. Obviously, if there were more than two detector stations on the freeway segment, one would rely on whichever successive pair provided the fastest response. One could then either suppress a subsequent response from the other detector stations if communications costs are high, or anticipate subsequent notification from them for verification.

By assuming discrete traffic states, it is possible to simplify the otherwise difficult problem and gain a better understanding of how the traffic state evolves over time in the presence of an incident. In closing, although the methodology was developed by deconstructing an incident on a freeway, this analytical process could be repeated for any other condition of interest on the freeway, e.g., where to locate detectors on a freeway, or even applied to many other problems necessitating information extraction from a large quantity of data.

Table 2.1, Control Points for an individual detector station on a homogeneous freeway based on time progression chart.

Control Points	Possible Action Set	Cost per Transmission				TMC Action
		Low	Mid	High	Very High	
[F0,F2], [C,F2]	No Action	X				Update trends
[X1,C], [F1,C]	Detect Recovery	X	X			Update trends
[F2,F1], [F2,F0]	Suspect Incident	X	X	X		Poll adjacent detector stations
[F2,X1]	Confirm Incident	X	X	X	X	Take remedial measures, e.g., dispatch tow truck.

Table 2.2, Control Points for an individual detector station upstream of a recurring bottleneck based on the time progression charts.

Control Points	Possible Action Set	Cost per Transmission				TMC Action
		Low	Mid	High	Very High	
[F0,F4], [F4,F3], [C,F4], [F4,X3], [C,X3], [X3,F0]	No Action	X				Update trends
[X1,X3], [X1,C], [F1,C]	Detect Recovery	X	X			Update trends
[F4,F1], [F4,F0], [X3,F1]	Suspect Incident	X	X	X		Poll adjacent detector stations
[F4,X1], [X3,X1]	Confirm Incident	X	X	X	X	Take remedial measures, e.g., dispatch tow truck.

Table 2.3, Control Points for an individual detector station downstream of a recurring bottleneck based on the time progression charts.

Control Points	Possible Action Set	Cost per Transmission				TMC Action
		Low	Mid	High	Very High	
[F0,F3], [C,F3]	No Action	X				Update trends
[F1,F3], [F1,C], [X1,C]	Detect Recovery	X	X			Update trends
[F3,F1], [F3,F0]	Suspect Incident	X	X	X		Poll adjacent detector stations
[F3,X1]	Confirm Incident	X	X	X	X	Take remedial measures, e.g., dispatch tow truck.

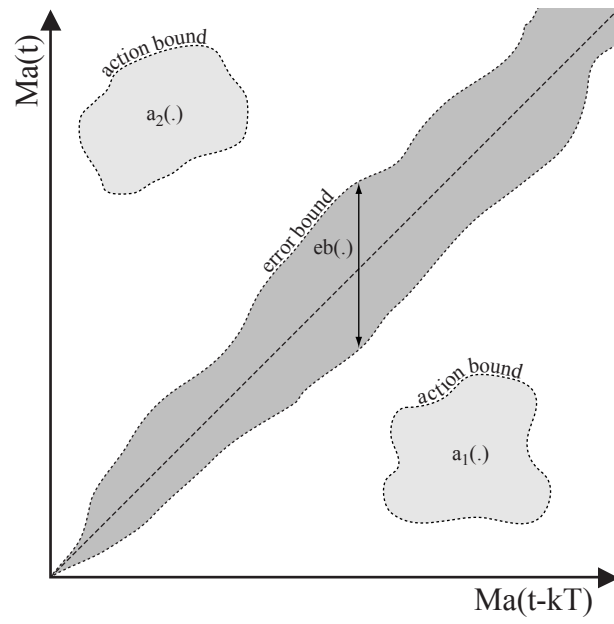


Figure 2.1, Parameter plot of a hypothetical benefit function.

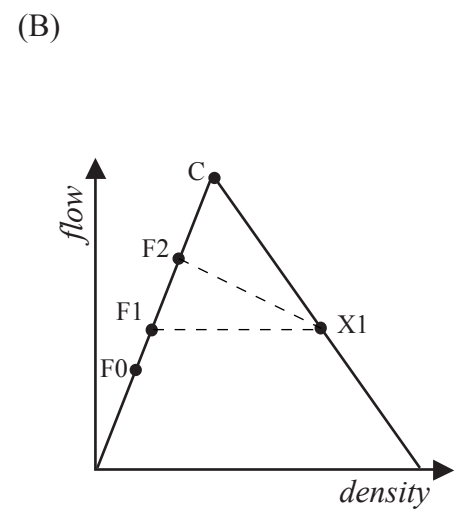
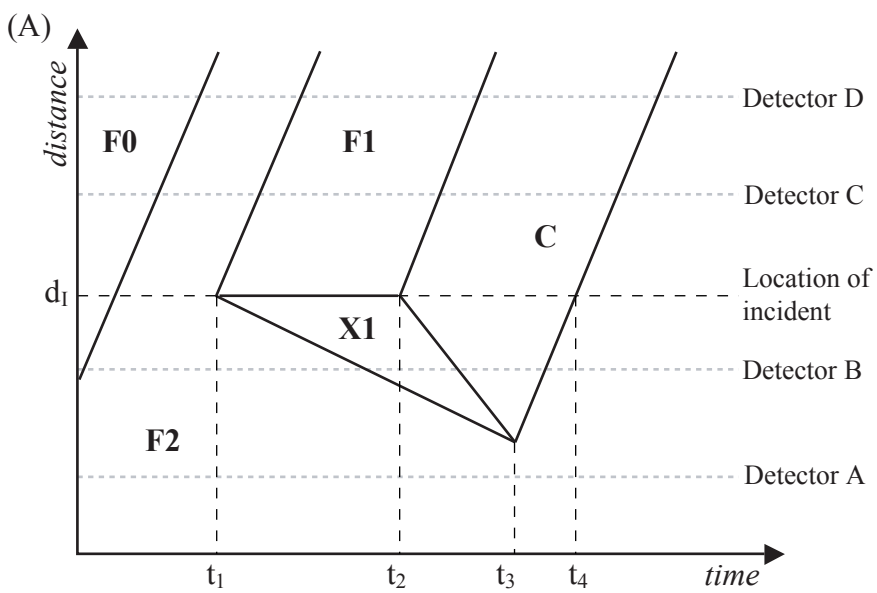


Figure 2.2, (A) Time-space diagram for an isolated incident on a homogenous freeway segment, (B) the corresponding flow-density relationship.

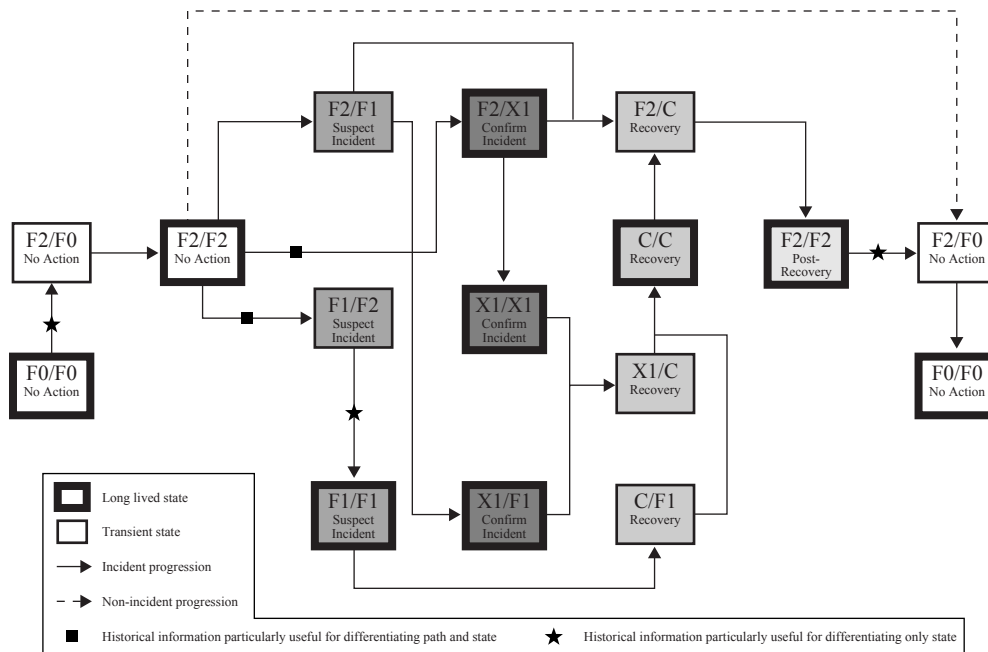


Figure 2.3, Time progression chart (us/ds) for an isolated incident on a homogenous section

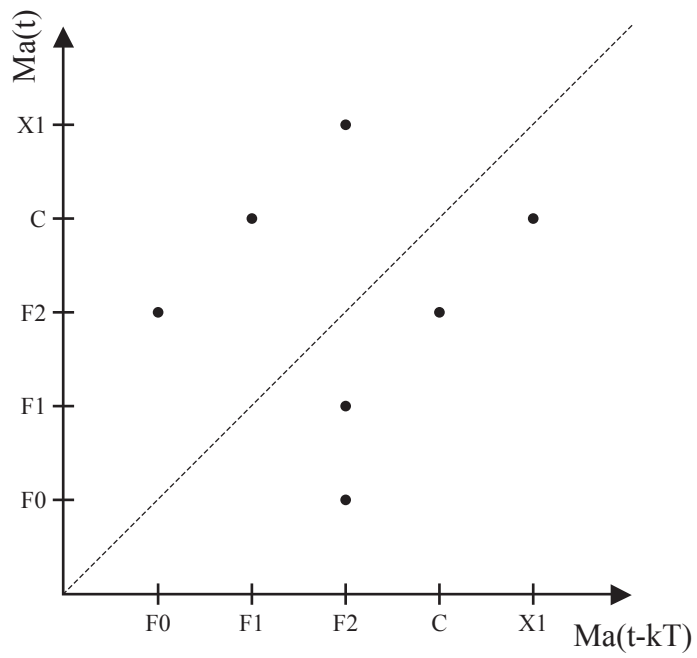


Figure 2.4, Controls for the Time Progression Chart in Figure 2.3, points denote transitions when transmissions would be initiated are denoted with points (compare to Figure 2.1).

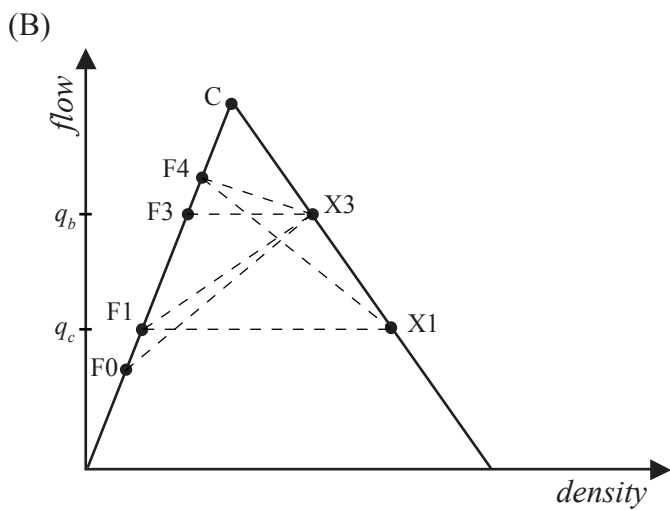
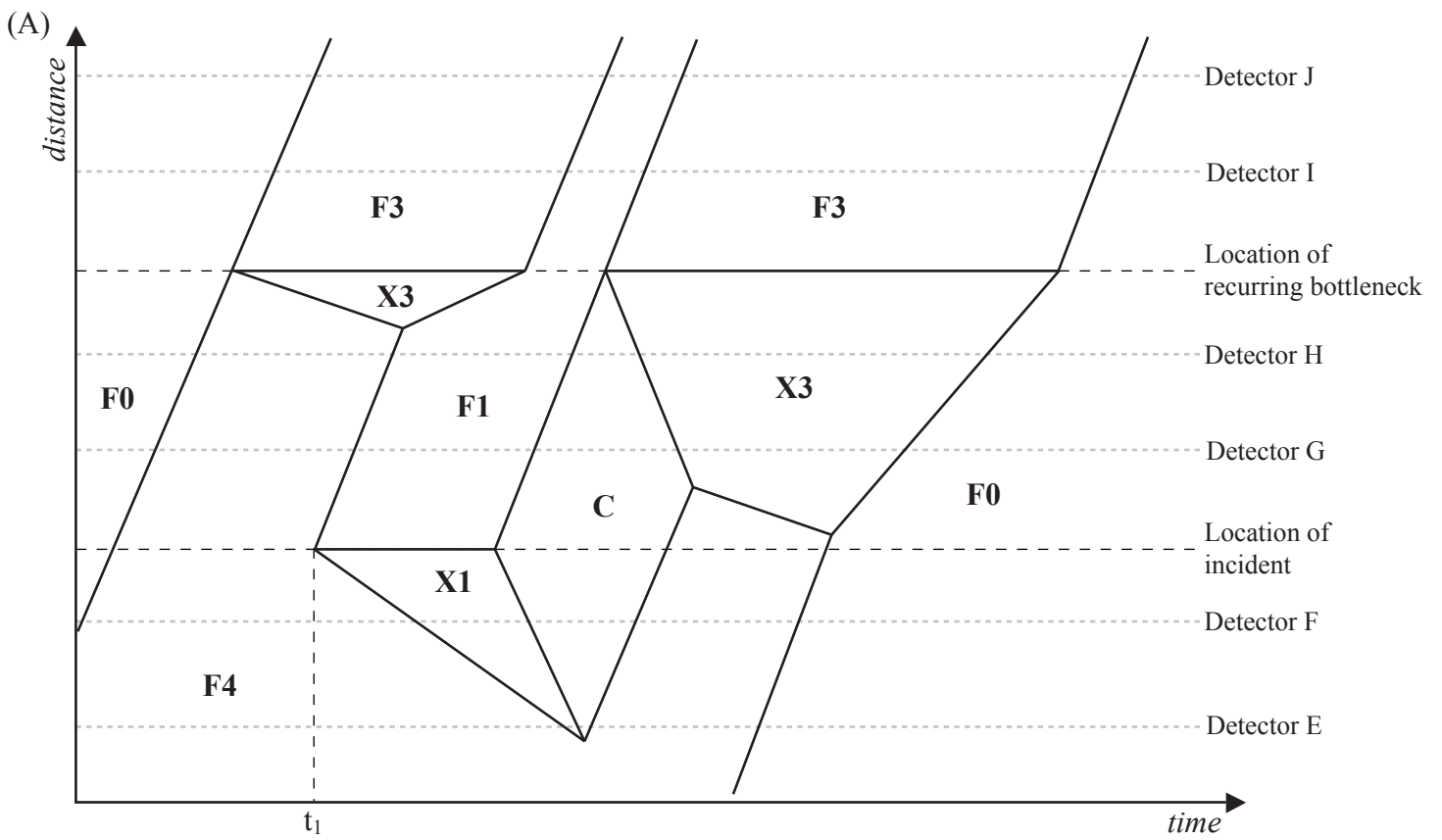


Figure 2.5, (A) Time-space diagram for an incident upstream of a recurring bottleneck, (B) the corresponding flow-density relationship.

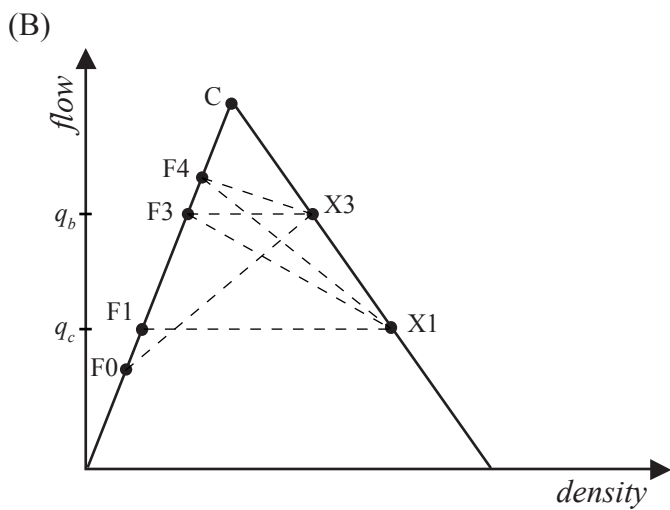
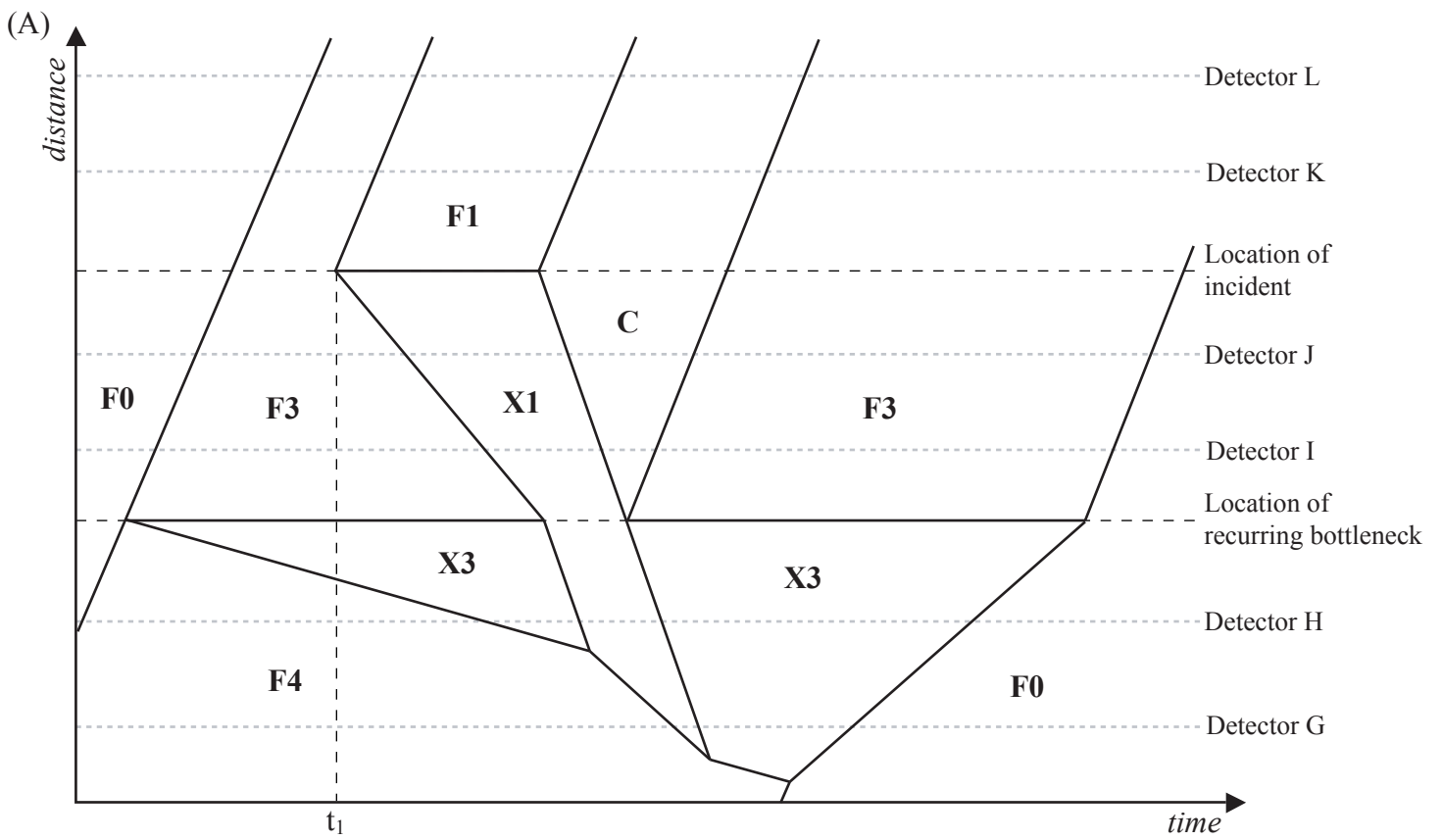


Figure 2.6 (A) Time-space diagram for an incident downstream of a recurring bottleneck, (B) the corresponding flow-density relationship.

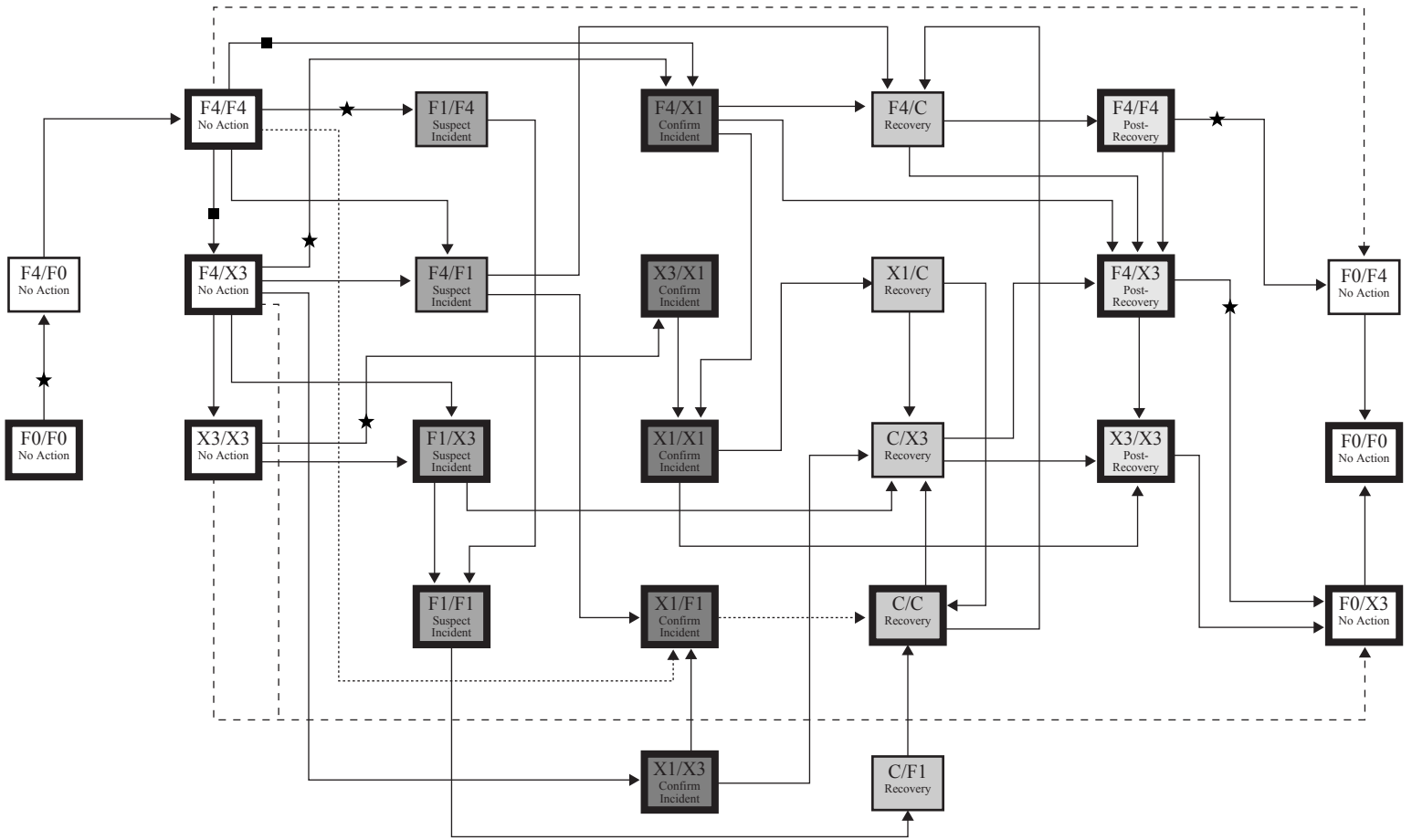


Figure 2.7, Time progression chart (us/ds) for an incident as observed by a pair of detectors upstream of a recurring bottleneck. The legend is given in Figure 2.3. Note that the incident may be downstream of the bottleneck.

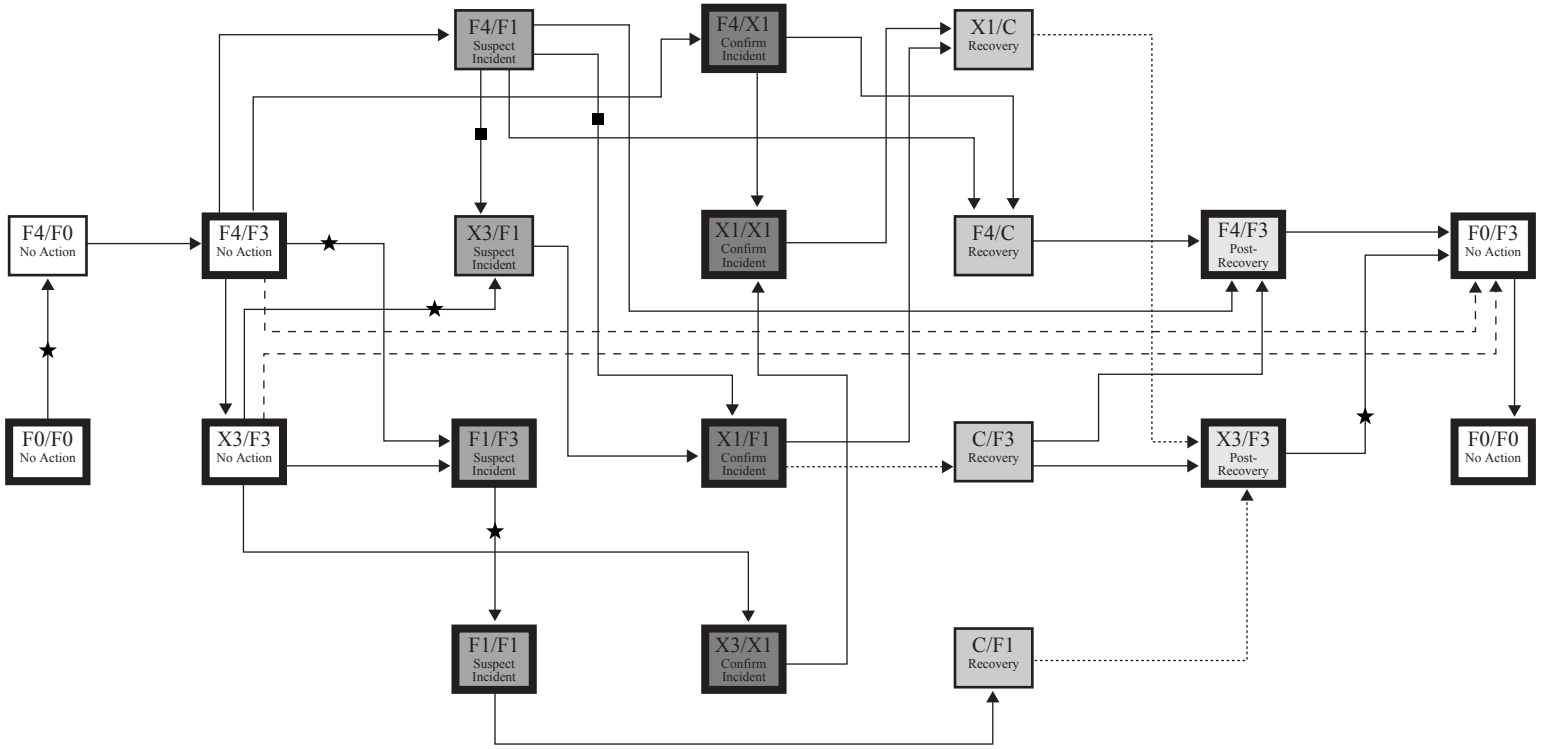


Figure 2.8, Time progression chart (us/ds) for an incident as observed by a pair of detectors straddling a recurring bottleneck. The legend is given in Figure 2.3.

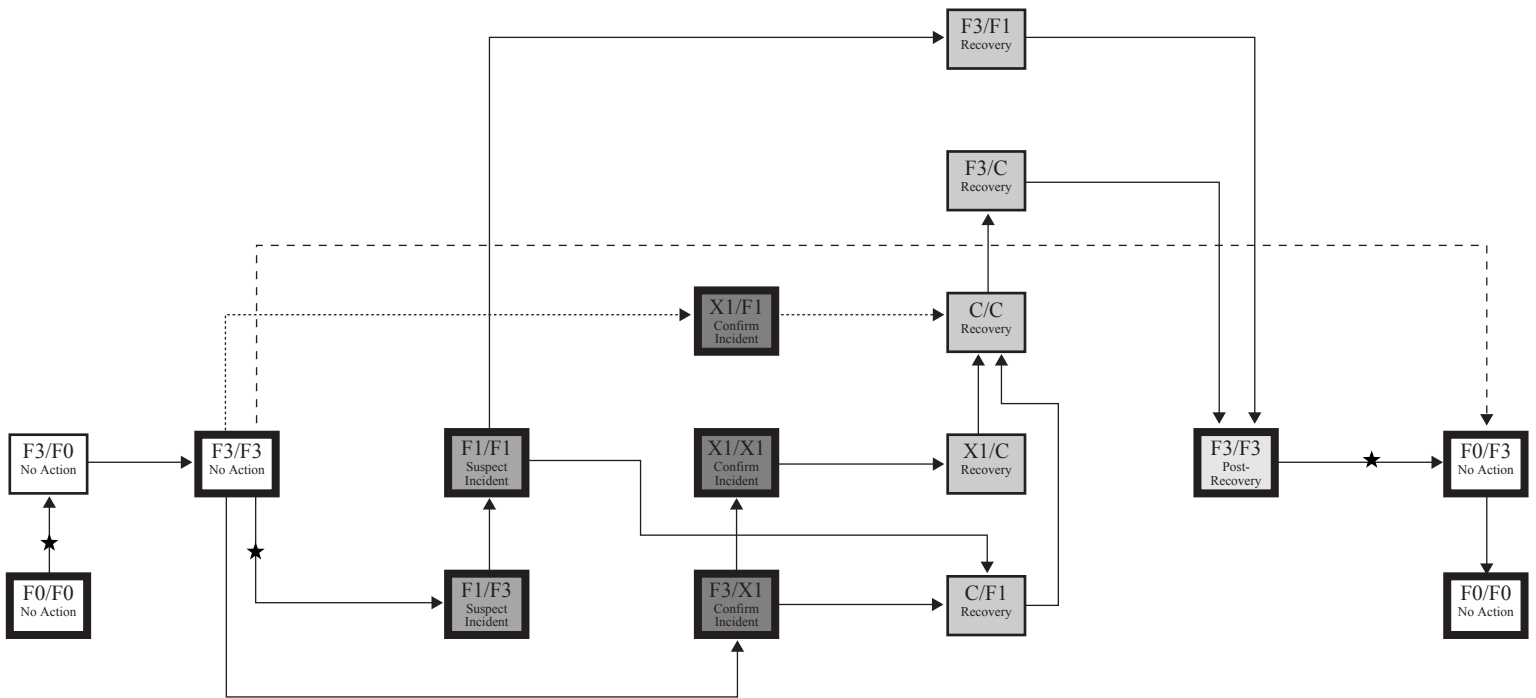


Figure 2.9, Time progression chart (us/ds) for an incident as observed by a pair of detectors downstream of a recurring bottleneck. The legend is given in Figure 2.3. Note that the incident may be upstream of the bottleneck.

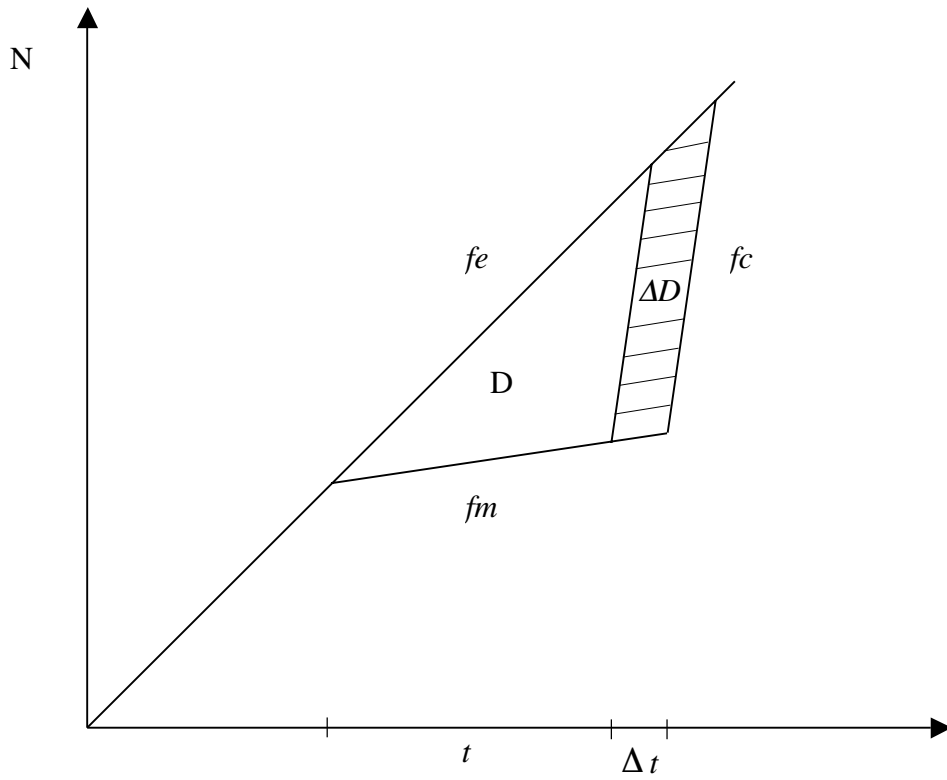


Figure 2.10, Queuing Diagram showing change in delay from ' t ' to ' $t + \Delta t$ '

3 DATA CLEANING AND IMPROVED SPEED ESTIMATES FROM FREEWAY TRAFFIC DETECTORS

The distributed surveillance system requires rigorous detector diagnostic and reliable data because erroneous detector or data might lead to a wrong decision at the field station and that could obscure an event or trigger a false alarm. Clean data are also required to pull out trends and the true relationship among traffic variables thereby simplifying the surveillance algorithms. In this chapter we develop detector diagnostic and data cleaning tools to identify and exclude malfunctioning detectors and transient errors. These tools can help calibrate the surveillance algorithms developed in the following Chapter 4.

This work starts with the premise that for most traffic applications local speed measurements are most important and information becomes valuable when speeds change. Thus, it is important to have accurate local speed measurements/estimates to identify critical events. After validating the traffic data, an analytical methodology is developed to make improved estimates of speed from single loop detector data. This methodology can also help clean noisy speeds measured from dual loop detectors and other sensors that mimic single loops. These advances can improve the performance of conventional surveillance system as well as enabling efficient distributed surveillance.

3.1 Introduction

Vehicle speed is one of the most important measures of freeway service quality. It is a principle indicator of highway performance, such as congestion and mobility, which agencies use to monitor freeways. Many Intelligent Transportation Systems (ITS) applications like traveler information and incident detection use speed as an input. Greater accuracy in speed estimates can improve decisions by agencies. Single loop detectors are the most common sensors used on today's freeways. The single loop detectors can measure flow, q , the number of vehicles that crossed the detector in a given time period, and occupancy, occ , the fraction of time the detector sensed vehicles above it. Most transportation agencies aggregate q and occ over fixed periods ranging from 20 seconds to 5 minutes. Mean speed is typically estimated using the following equation:

$$\hat{V} = (L*q)/(occ*T) \quad (3.1)$$

where

T = Duration of the aggregation period

\hat{V} = Estimated mean speed over T

L = Assumed average vehicle length

L and \hat{V} cannot be measured independently at a single loop so L is normally set to a constant, but the simple application of Equation 3.1 fails to account for the variation in true effective average vehicle length. As a result, small variations in q or occ measurements from sample to sample can cause large errors in \hat{V} . This problem is most pronounced during periods of low q and occ . During congested periods, \hat{V} exhibits smaller absolute errors due to lower speeds and generally a lower frequency of long vehicles. There are several methods proposed to choose L

[22, 23, 24]. These methods seem to work well when the aggregation is over longer periods, like five minutes. As shown by [22], the assumed average vehicle length becomes more representative as the sampling period increases because the number of vehicles increases. When the number of vehicles in a sample is small, a long vehicle can skew occupancy and hence the speed estimation. Recognizing that low occupancy must correspond to free flow conditions, [24-25] developed a methodology where a time dependent, location specific average vehicle length curve is developed using known local free-flow speeds. The L during high occupancy periods is estimated by fitting a smooth regression curve with the assumption that L is independent of occupancy [25]. Historical information and average vehicle length time-series are used to make real time speed estimation for subsequent days. This methodology works fine if the time dependent average vehicle length curve can be developed reliably, the free-flow speeds are known, and there is little variation in vehicle lengths within a sample. Other researchers have produced clean estimates of speed by developing algorithms that use statistical methods, like minimum mean-square error, to estimate speed and assume that flow, occupancy, speed and vehicle length follow a distribution with a mean value and variance [23, 26].

Some researchers have focused on improving data quality using adjacent lane information for detecting bad data samples and for imputing missing or bad samples, e.g., [20]. Combining data across lanes increases the number of vehicles during shorter aggregation periods, since the adjacent lanes provide redundant information about the traffic state [22]. In many traffic-monitoring applications, operators are only interested in a single estimate of speed per direction at a detector station. Conventionally, this aggregation is obtained by taking the arithmetic mean of \hat{V} across the individual lanes. But the accuracy of \hat{V} is degraded by the fact that the true vehicle length spans a large range, some vehicles are four times longer than the average vehicle [22]. This fact affects the estimates by Equation 3.1 as it uses fixed L . The median value is less sensitive to outliers when there are more than two lanes and is a better central measure for this application.

This chapter describes a data cleaning and speed estimation methodology that exploits the redundant information from adjacent lanes and applies basic traffic flow theory principles to 30 second aggregated single loop detector data. The methodology is generic and could be easily be modified for other sampling periods (though it might require some changes to thresholds). The methodology was validated by comparing with measured speeds from dual loop detectors (a pair of detectors in each lane at the detector station, which can measure speed directly from the difference in actuation times between the two loops) to the estimates from a single loop of the pair using data from I-71 in Columbus, Ohio and I-80 in Berkeley, California [8]. This methodology is transferable to other sensors that mimic single loops, e.g., the Remote Traffic Microwave Sensor (RTMS), and speeds measured from dual loop detectors. The chapter is organized into four major sections: data cleaning tools and types of tests to identify malfunctioning detectors are given first. Next, an overview of the development of the speed estimation procedure, including detailed descriptions of the steps and application of the procedure to the data is explained. An analysis and discussion of the results is then presented, followed by conclusions.

3.2 Identifying Malfunctioning Detectors and Erroneous Data

Before estimating speed from q and occ , it is important to identify erroneous data reported by the detectors. The most common errors arise from a malfunctioning detector or data lost during transmission. Often the malfunctioning detector can be identified easily by visualizing the data. The time-series plots of measurements from a chronically abnormal detector are strikingly different from the typical time-series. Figure 3.1 compares the time-series of q and occ as reported by two adjacent detector stations on I-71 southbound in Columbus, Ohio. The malfunctioning detector station exhibits flows between 2800 vehicles per hour per lane (vphpl) to 3800 vphpl between 13:00 and 15:00, compared to flows below 2000 vphpl at the working station just 590 feet downstream. Flows as high as 3100 vphpl have been reported in the literature [27], but such high flows are not sustained for long periods and in general, the maximum sustainable flow is expected to be in the range of 2500 vphpl. The shaded region of Figure 3.1C-D shows the periods when either the detector did not report any data or reported unusually high values of flow. It can be seen that by this criterion, more than half of data reported from this station are erroneous on this day.

To identify if the detector is malfunctioning, several threshold value tests are applied to the data by comparing measurements against maximum and minimum acceptable values [20]. For example, a test may consider q greater than 3000 vphpl as erroneous; if the frequency of such records is greater than an acceptable value then the detector is noted as malfunctioning and ignored by later steps in our speed estimation method. The thresholds for various tests described below can be set by looking at historical data and data from reliable detectors in that lane. For this chapter, the threshold values were calculated using data for 5 weekdays at 10 single loop detector stations obtained via Performance Measurement System (PeMS) [9] from the Main Line loops along 58 miles (mile-marker 52 to 110) of SR-101 in the San Francisco Bay area of California. SR-101 was chosen because data for several consecutive stations and days were available and there were many detectors with noisy and erroneous data. The threshold values for the tests described below were established empirically such that only data from severely problematic detectors are discarded. These results were tested with data from other stations on SR-101 and nearby I-80 yielding similar results.

The following threshold value detector diagnostic tests are designed to be applied to each day of detector data to decide if the detector can be trusted on the following day. If a detector fails any of the threshold value tests then data from that detector are not used for speed estimation until it passes all of the tests.

Test 1:

Samples that show zero flow and zero occupancy: During a day there are periods when samples show zero flow and occupancy. The frequency of zero flow-occupancy measurements is a function of time of the day and we can expect to see many such samples during late night and early morning periods. But for the rest of the day, we do not expect to see many samples with zero flow and occupancy. By counting the total number of zero flow and occupancy samples in the period between 05:00 a.m. and 10:00 p.m. and comparing with a set threshold, we can identify whether the detector is stuck or not. The threshold value for this test is set at 50 percent of the samples, if exceeded, the detector is considered to be malfunctioning.

Test 2:

Samples that show zero flow and non-zero occupancy: A sensor reporting zero flow and non-zero occupancy over a 30-second period can be viewed potentially malfunctioning. However, one such sample cannot be deemed faulty. A vehicle is counted in q at a single instant, e.g., upon entering the detector, but may contribute to occ of more than one sample if it remains over the detector at the end of a sample. As a result, a vehicle stopped above the sensor for the entire 30-second period can result in zero flow and 100 percent occupancy, or a vehicle begins the 30-second period positioned above the sensor, then moves away from the sensor without another vehicle passing the sensor in the 30-second period, resulting in zero flow and nonzero occupancy. This second scenario tends to occur in very heavy traffic or in very light traffic. If the traffic is known to be light, then we do not expect occupancy to be greater than 3 percent as typically a free flow vehicle takes less than 1 second to cross the detector. If the number of samples between 5:00 a.m. and 10:00 p.m. with zero flow and non-zero occupancy exceeds a threshold value of 25 percent, then detector is considered to be malfunctioning. When the period of aggregation is increased to five minutes these events become far less likely and this type of validity check becomes more indicative for longer aggregation intervals.

Test 3:

Samples that show non-zero flow and zero occupancy: Any vehicle that passes the sensor must register a positive value of occupancy during the sample it contributes to flow. The threshold value for this test is set at 25 percent of the samples from the whole day. If during a day, the number of samples with zero occupancy and non-zero flow is greater than the threshold value, the detector is assumed to be malfunctioning.

Test 4:

Samples that show abnormally high flow or occupancy: Chattering and pulse break up at the detector can cause the flow in the sampling period to be abnormally high. Samples with flow as high as 3100 vphpl (25 vehicles per 30 seconds) were found in the test data from SR-101. Setting a threshold at a conservative 3100 vphpl [27], a sample above this value is considered suspicious. Once more, the number of samples that exceed this threshold is counted and if greater than 25 percent of the samples for the day, the detector is considered to be malfunctioning. However, if there are only a few such samples, then only those samples might be erroneous. A similar test with occupancy data can also be performed to check the validity of the data. An occupancy above 35 percent is considered high, arising from congestion. We assume that any freeway should be uncongested at least 60 percent of the time [25], thus if number of samples with occupancy exceeding 50 percent for a day, the detector is considered to be malfunctioning.

Test 5:

Samples that show constant occupancy and flow: Sometimes the field controller gets stuck and reports the same flow and occupancy for successive samples. This error can be identified by taking the difference of successive flow measurements and again for occupancy for entire day's data. The presence of a long series of zeroes (3 hours in our case) verifies that the detector is stuck and such data should be discarded.

3.3 Macroscopic Speed Estimation and Filter Rules

After excluding all malfunctioning detectors based on the preceding tests, a clean estimate of vehicle speeds is obtained by following the steps 1-6 below. As previously noted, the conventional speed estimation from Equation 3.1 is noisy. These raw speed estimates are refined on an individual lane basis by passing them through a series of filters, steps 2-6, which use data from adjacent lanes and apply basic traffic flow theory principles to identify erroneous speed estimates. It is important to follow Steps 1 and 2 in order followed by Steps 3-5, not necessarily in the given order, and concluded with Step 6. The reason for this ordering is because the earlier steps affect the thresholds in the subsequent steps, as will become more apparent in the following section.

Unless otherwise stated, the data for illustrating the filtering process were obtained from detector station 4 on I-80 westbound at Berkeley Highway Laboratory (BHL), California [8]. This station has five lanes in each direction and data were aggregated to 30-second samples such that each day's data had 2880 samples per lane. The data from PeMS were not used because of the poor quality of dual loop speed measurements, as mentioned in white paper at PeMS [9] and verified by our empirical analysis. Meanwhile, the data from SR 101 were used to develop thresholds in Section 3.2, as there are only 8 dual-directional detector stations in the BHL.

Step 1:

Identify erroneous samples: In this step, erroneous data samples are identified and are either discarded or replaced by an estimate. The methodology to identify erroneous data samples is similar in principal to the detector diagnostic discussed in Section 3.2. Samples with $q = 0$ and $occ > 3$ percent, or $q > 0$ and $occ = 0$ are suspect (refer to Test 2 and 3). Also successive samples with the exact same flow, occupancy, or both are suspect. The suspect samples for individual lanes are validated by looking at the samples in their "neighborhood"; where the neighborhood consists of adjacent lanes during the same sample period and the three most recent valid measurements in the subject lane within previous five minutes, if they exist. If the suspect sample is consistent with the traffic state indicated by neighboring samples, then it is retained as a valid sample. The remaining suspect samples are discarded at this stage so that the errors introduced by imputing bad samples do not affect estimated speeds. If one wishes to eliminate holes in the data due to presence of erroneous samples, they could be filled by an estimate from the average q and occ in the adjacent lanes. If the samples in the adjacent lanes are not reliable then the median of the samples from the past 3 valid measurements can be used as an estimate. If the last three valid samples in the subject lane are older than 10 samples, then the estimation can be done by combining data from the valid samples from all lanes over the last five minutes. This data estimation methodology is applicable to short-term malfunctions only. If a malfunction persists for several minutes, the traffic measurements may no longer apply to the current traffic situation, resulting in inaccurate estimates. Furthermore, this mechanism is applicable only to locations where individual lane data are available and there are at least two reliable lanes. If this condition is no longer valid then the estimate might be as poor as the erroneous measurement and in such case we would be better off simply flagging the sample as invalid.

The scatter plot in Figure 3.2A compares \hat{V} , for an individual lane, at the end of Step 1 plotted against measured speed. The data were collected on August 12, 1998. Notice the considerable amount of scatter.

Step 2:

Raw estimate of speed: First a target period of the day is chosen which is defined as a period when the traffic is expected to be free flowing and the flow far from zero. When the traffic is free flowing, flow and occupancy measurements are expected to be in phase, i.e., increase or decrease together through small fluctuations. The correlation among q and occ measured in successive 5-minute samples in the target period is calculated and if most of these calculations are positive, then q and occ are in phase verifying free-flow. Usually, the period between the morning and evening peaks or after the evening peak is a good choice for the target period. A ratio of measured flow over occupancy is taken. Ordinarily, this ratio would be multiplied by L to estimate \hat{V} and we initially set $L=20$ feet. We then estimate the "raw speed" using Equation 3.1. A correction factor is then calculated by taking ratio of the median of raw speed during the free-flowing target period and the assumed free-flow speed for that station and lane. This assumed free-flow speed might come from the posted speed limit, as in this chapter, or direct measurement, e.g., from a radar gun. More precise measurements of free-flow speed are desirable but that would require more resources for an operating agency and the assumption is similar to what is being practiced with existing single loop detector data. The correction factors are calculated for all lanes at every station and in subsequent steps, the raw speed estimates are multiplied by corresponding correction factor.

If non free-flow conditions were present during the target period then the computed correction factor would be erroneous. The median of the estimated speeds after applying the correction factor to entire day's data should be close to the assumed free-flow speed thereby validating the correction factor and the presence of free-flow conditions in the target period. When the phase data indicate congested conditions, the correction factors can be computed by using recent historical data from last few days. Taking the moving median of the recent correction factors can eliminate error due to undetected congestion and reduce fluctuations due to daily changes in the traffic patterns. Using data from 15 consecutive days, over the period between 19:00 and 23:00 the daily correction factor for each lane was calculated and as expected, there was little variation from day to day (Figure 3.3). In this manner, these factors can be developed for every station and help correct for any error due to sensitivity of the detector. The scatter plot in Figure 3.2B shows \hat{V} at the end of Step 2.

Step 3:

Speed-Flow filter: A typical curve representing the relationship between speed and flow is shown in the right half of Figure 3.4. Using this relation one can define threshold values for free flow speed corresponding to different flows. Flow-Speed pairs that are outside these threshold boundaries either fall in an infeasible region or are from congested conditions. The left hand side of Figure 3.4 shows the monotonically decreasing relationship between speed and occupancy. The exact shapes of these curves are not important; we will develop our thresholds to be loose enough to accommodate any reasonable approximations. In Figure 3.4, we do not expect a flow-speed pair to fall

in shaded region 'A' unless it is from congested conditions. But such measurements should have high occupancy if valid. Given the maximum speed (V_{\max}) of region 'A', the valid data belonging to region 'A' must correspond to occupancy higher than threshold (Occ_{\min}) on left of region 'B'. If a measurement falls in both region 'A' and 'B' then it is erroneous. For our test data, threshold values were established such that a sample would belong to an infeasible region if $q < 1000$ vphpl, $\text{occ} < 15$ percent and $\hat{V} < 50$ mph. When an estimate fails this test, it is compared to the sample neighborhood and if it is similar to the other estimates it is retained; otherwise it is flagged as erroneous and re-estimated as follows. The speed is estimated by taking the median of the speeds in the adjacent lanes. If the samples in the adjacent lanes are not reliable then the median of the samples from the past 3 valid measurements can be used as an estimate. If the last three valid samples in the subject lane are older than 10 samples, then the estimation can be done by combining data from the valid samples from all lanes over the last five minutes. If valid data are not available then the sample is marked as erroneous and no estimation is made. In this process, care was taken to use the median rather than mean to reduce sensitivity to outliers. The improvement in speed estimates at the end of this step is shown in Figure 3.2C.

Step 4:

Speed-Occupancy filter: Again, using the argument presented in Step 3 and the relation between flow, occupancy and speed, one can define the threshold values for speed corresponding to different flow-occupancy values. As has been observed, both in our research and by other researchers [28], capacity flow corresponds to occupancy between 10-25 percent and below this free-flow conditions exist. We set a conservative threshold for free-flow occupancy at 8 percent, thus samples below this threshold and $\hat{V} < 50$ mph are suspicious, e.g., region B in Figure 3.4. These suspect estimates are compared to their neighborhood samples to verify they are not feasible. Also, the current traffic state (free-flowing or congested) is verified using the phase information between flow and occupancy measurements of the previous few samples. If the suspect sample's speed is found to be inconsistent with the current traffic state then the speed for the sample is re-estimated using the process outlined in Step 3. Now, however, we do not expect an occupancy-speed pair to fall in region B. In this case, the measurements with $\hat{V} < 50$ mph, $\text{occ} < 8$ percent and $q < 840$ vphpl are replaced with the assumed local free-flow speed. The scatter in \hat{V} at the end of this step is shown in Figure 3.2D.

Step 5:

Speed filter: In this step, samples with $\hat{V} > 90$ mph are identified as erroneous since we do not expect the speeds to be this high. This error could be either due to transient error from a short on time or due to the overestimation of L. If most of the vehicles in the sample are truly shorter than assumed L, then we will overestimate the speed. The phase information and information from adjacent lanes are used to determine if traffic is free-flowing or congested. If the traffic is free flowing then these samples are scaled down to the assumed free-flow speed, otherwise these samples are discarded altogether. Figure 3.2E shows \hat{V} at the end of Step 5.

Step 6:

Moving median of three samples: Equation 3.1 is sensitive to small variations in q and occ . This sensitivity is inversely proportional to the number of vehicles observed. When only a few vehicles pass the detector in a given time interval, i.e., light traffic, the true average vehicle length is more likely to differ substantially from L , and \hat{V} will be noisy [28]. The errors in speed estimation might not be eliminated if the sample falls in the feasible regions of the previous steps. In order to reduce error from high variance in L we take a moving median of three samples. However, we restrict the moving median to only estimates with $\hat{V} > 25$ mph as L exhibits smaller variance during lower speeds. The speed estimates at the end of filtering Step 6 are shown in Figure 3.2F.

Finally, a single accurate estimate of speed per direction at a detector station is obtained by taking the median of the improved individual lanes' speed estimates. The lane-by-lane \hat{V} after filtering is shown in Figure 3.5A-E and median \hat{V} across all lanes is shown in Figure 3.5F.

3.4 Analysis and Discussion

Figure 3.6A-F shows several scatter plots of different estimates of speed versus measured speed for a day of data from BHL station 4. The left column shows conventional estimates from Equation 3.1 while the right column shows the new estimates from the Section 3.3. This figure shows (A-B) estimates from a single lane, (C-D) mean across all lanes and (E-F) median across all lanes. The mean and median come from those lanes that were not eliminated by the filters. In each row, the circles are placed at the same location in the plot pair to highlight the difference between the estimations. If the estimations were perfect a given plot would show all the points on the straight line inclined at 45-degrees. In all three cases the speed estimation using conventional method shows greater deviation from the straight line than the new methodology. The estimation for individual lanes is not as good as that across lanes, which is expected because there are fewer vehicles in the individual lane samples. This drop in quality can be observed by comparing the scatter around the 45-degree straight line in plots B, D and F.

Figure 3.7 compares the time series of estimated speeds after Step 2 and those after Step 6. As can be observed from the plot, there is high scatter in the raw speeds estimated by Equation 3.1, especially during early morning periods. For the same data, when the speeds are passed through the filters developed above, the scatter is reduced significantly without degrading the estimates. Figure 3.7A compares speed for an individual lane, Figure 3.7B compare the mean speed across all lanes and Figure 3.7C compares median speed across all lanes. Observe that the mean and median across all lanes provides a cleaner estimate than a single lane for both the raw and processed data.

It is important to ensure that the speed estimates after filtering are representative of the traffic state. The estimates should not indicate congested conditions when it is actually free-flow and vice versa. Though we do not expect the estimates to perfectly match the true measurements, they should be as close as possible. To check the quality of estimates, the number of estimates with $\hat{V} > 50$ mph and measured speed less than 35 mph (Type-1 error) and $\hat{V} < 30$ mph and measured speed greater than 50 mph (Type-2 error) per day were counted. The quality of the estimates is good if the number of estimates with each error type is zero or close to zero. One

month of data from BHL station 4 were used. Figure 3.8A shows the cumulative distribution function (cdf) for the median error of each error type across the lanes. As can be observed from the plot, the number of samples with error type-1 is zero for almost 90 percent of the time and few samples (15 out of 2880) exhibit error type-2, which likely is acceptable for most applications.

A considerable amount of error in raw speed estimates is simply due to bias in L . The bias in L was removed by calculating the average representative vehicle length for every two hours using measured speed. Unbiased speeds were estimated using Equation 3.1 with new average vehicle lengths, representing the best possible performance from conventional practice. Figure 3.8B shows the cdf of absolute error in speed estimation for an individual lane data, Figure 3.8C for the median across lanes and Figure 3.8D for the mean across all lanes. Each plot shows the cdf of error for the cleaned estimate from the filters, raw unbiased estimates and raw estimates. Notice that simply by un-biasing the error, we get significant improvement in individual lane speed estimates and similar performance when combining data across lanes, though the proportional improvement is smaller than the single lane. For the raw estimates, the median across all lanes performs better than mean, which can be observed by comparing Figure 3.8C-D. Also the plots in Figures 3.6C-F and 3.7B-C highlight that median is less sensitive to outliers.

As observed by comparing Figure 3.8B with 3.8C-D, the absolute error decreases by combining data across lanes. In all cases, the error after cleaning the data is smaller than unbiased speed estimates error. The mean absolute error of cleaned speeds is less than less than 7 mph for individual lanes and is less than 3 mph when data are combined across lanes.

The algorithm developed in Section 3.3 using single loop detector data can also be extended and applied to speeds measured from dual loop detectors that are noisy. For cleaning noisy speed measurements, only the part of the Step 2, which estimates raw speed using Equation 3.1, needs to be omitted and the rest of the algorithm can be applied without change. For dual loop detectors, there is no need to correct for error in L as speed is calculated by measuring the time taken to travel between the two loop-detectors. In this case, the correction factors from Step 2 correct for any error in detector spacing (which is constant, though potentially unknown).

3.4.1 Analysis of RTMS data:

This algorithm was also tested with data obtained from an RTMS in side fire mode. The RTMS is functionally very similar to a single loop detector station since the unit measures flow and occupancy and then estimates speed. The RTMS incorporates the functionality of the traffic controller, but it also provides contact closure outputs that can be input to a conventional traffic controller (henceforth referred to RTMS-via-controller). For this study, concurrent RTMS, RTMS-via-controller and dual loop detector data were collected on May 26, 2000 at eastbound BHL station 7 and aggregated to 30-second sample.

Figure 3.9 shows several scatter plots comparing median estimated and measured speed across all lanes. The left hand column shows the raw estimates and the right hand column shows the corresponding cleaned estimates after applying the filters. The first row, A-B, show the speed estimates reported by the RTMS's internal controller. The second row, C-D, show the speed estimates after applying Equation 3.1 to the measured q and occ reported directly by the RTMS. The third row, E-F, show the speed estimates from Equation 3.1 applied to q and occ for RTMS-via-controller. For reference, the fourth row shows the speed estimates from the loop q and occ

(the downstream loop detector was within 20 ft of the RTMS detection zone). Quantifying these errors, Figure 3.10 shows the distribution of absolute and percentage error for each of the rows from Figure 3.9. Note that our filtering improves the performance in all four cases and similar results were observed on an individual lane basis, as shown in Table 3.1. Among the raw data, the direct RTMS estimates (Figure 3.9A) show the best performance because the RTMS applies its own filters to the estimation process while the estimates from q and occ are worst for the direct RTMS measurements (Figure 3.9C) because it only reports integer occupancy. Except for these rounding errors, Figure 3.10 shows that after applying the filter from this chapter, the other three estimates exhibited similar performance (Figure 3.10A, E, G).

3.5 Summary

Many researchers have sought better estimates of speed from single loop detectors. The earlier works have emphasized statistical techniques to reduce the bias from long vehicles in mean speed estimates on a lane-by-lane basis. This chapter has taken a different approach, it applies threshold value tests and traffic flow theory principles to identify erroneous speed estimates. These erroneous estimates are re-estimated using redundant information from adjacent lanes and recent history. These individual lane estimates were shown to have improved by as much as 70 percent as compared to unbiased conventional estimates and by as much as 125 percent as compared to conventional raw estimates. It was also shown that the directional estimate of speed obtained by averaging the improved individual lane estimates across all lanes is more accurate than conventional estimates.

In the course of developing the algorithm, data validation and detector diagnostic tools were developed to eliminate malfunctioning detectors and transient errors. These data validation tools can help improve the data quality, which in turn would improve speed estimates and decision-making process by agencies. The methodology was shown to be less sensitive to site-specific characteristics of the traffic flow and also applicable to noisy traffic data collected by other methods.

Table 3.1, Absolute mean error and absolute percentage mean error for RTMS estimates.

	Lane 1		Lane 2		Lane 3		Lane 4		Lane 5		All lanes combined	
	Abs Error (mph)	Abs % Error	Abs Error (mph)	Abs % Error	Abs Error (mph)	Abs % Error	Abs Error (mph)	Abs % Error	Abs Error (mph)	Abs % Error	Mean Abs Error (mph)	Mean Abs % Error
RTMS Measured	9.95	18.68	6.19	15.79	5.6	16.15	3.97	10.08	6.4	14.45	3.91	6.27
RTMS Measured Clean	10.29	16.46	4.47	11.71	3.44	11.2	3.15	8.77	5.2	12.48	2.57	3.36
RTMS (q, occ) Raw	23.27	39.27	12.5	23.42	10.61	22.29	11.41	21.54	32.74	55.93	10	15.02
RTMS (q, occ) Clean	12.11	18.35	8.43	16.13	7.53	18.18	8.27	19	8.7	16.43	6.28	11.38
RTMS-via-controller (q, occ) Raw	15.62	24.32	13.19	25.1	16.57	32.55	14.85	27.17	7.77	14.95	12.95	22.79
RTMS-via-controller (q, occ) Clean	9.47	13.82	4.82	9.61	5.55	14.74	6.37	14.42	5.69	11.84	3.35	5.49
d/s Loop (q, occ) Raw	5.47	11.07	2.06	5.47	7.96	16.26	10.3	19.22	12.26	22.07	6.41	10
d/s Loop (q, occ) Clean	3.06	6.03	1.87	5.14	5.57	12.18	5.94	11.43	4.89	8.3	2.68	3.85

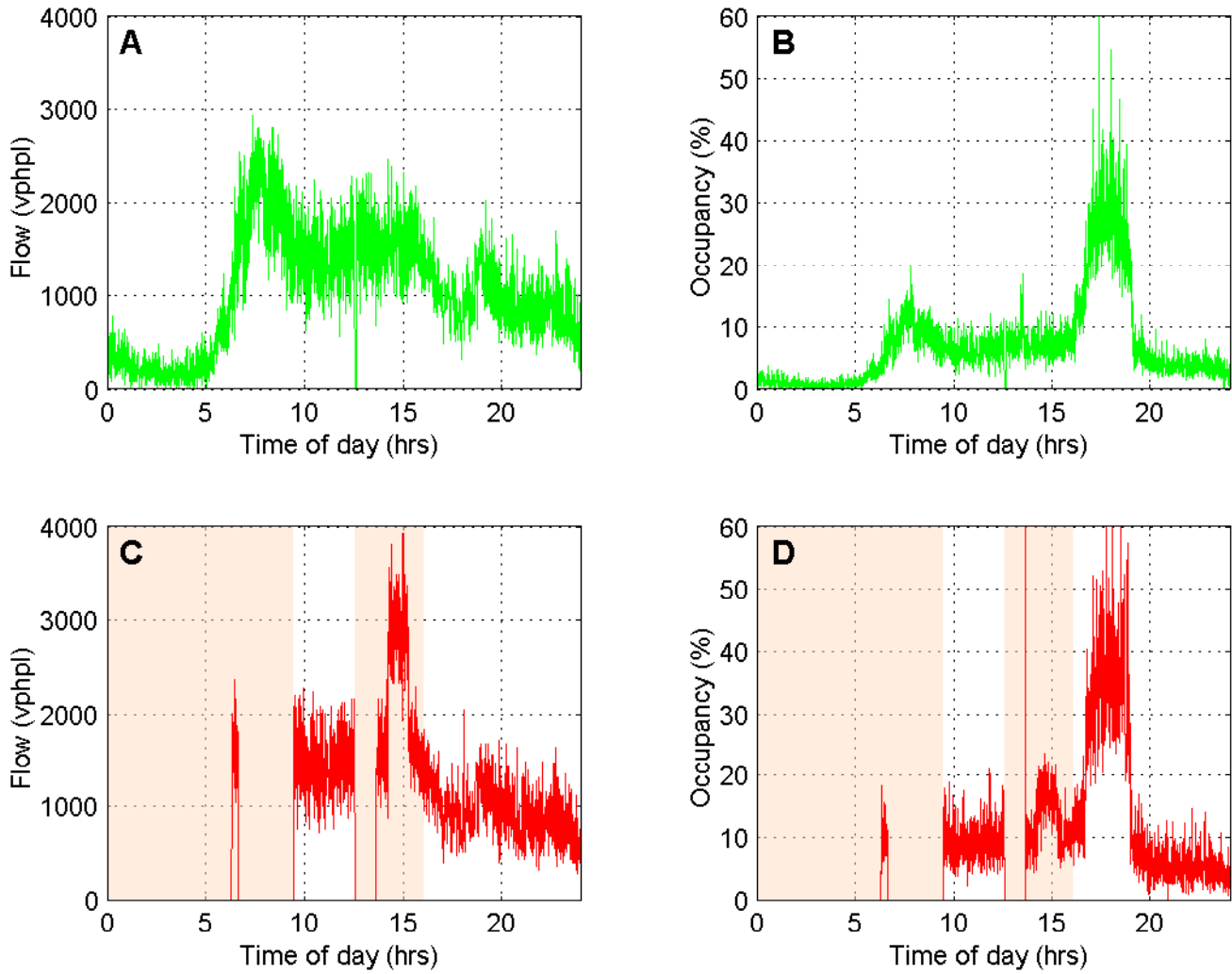


Figure 3.1 Stations exhibiting typical (A & B) and abnormal (C & D) flow and occupancy measurements. The plots show average flow and occupancy data across southbound lanes from two adjacent stations (St7: A & B, St8: C & D) at I-71, Columbus Ohio on 4th October 2002. The shaded region in plots C & D shows the period when either the detector did not report any data or reported unusually high values of flow

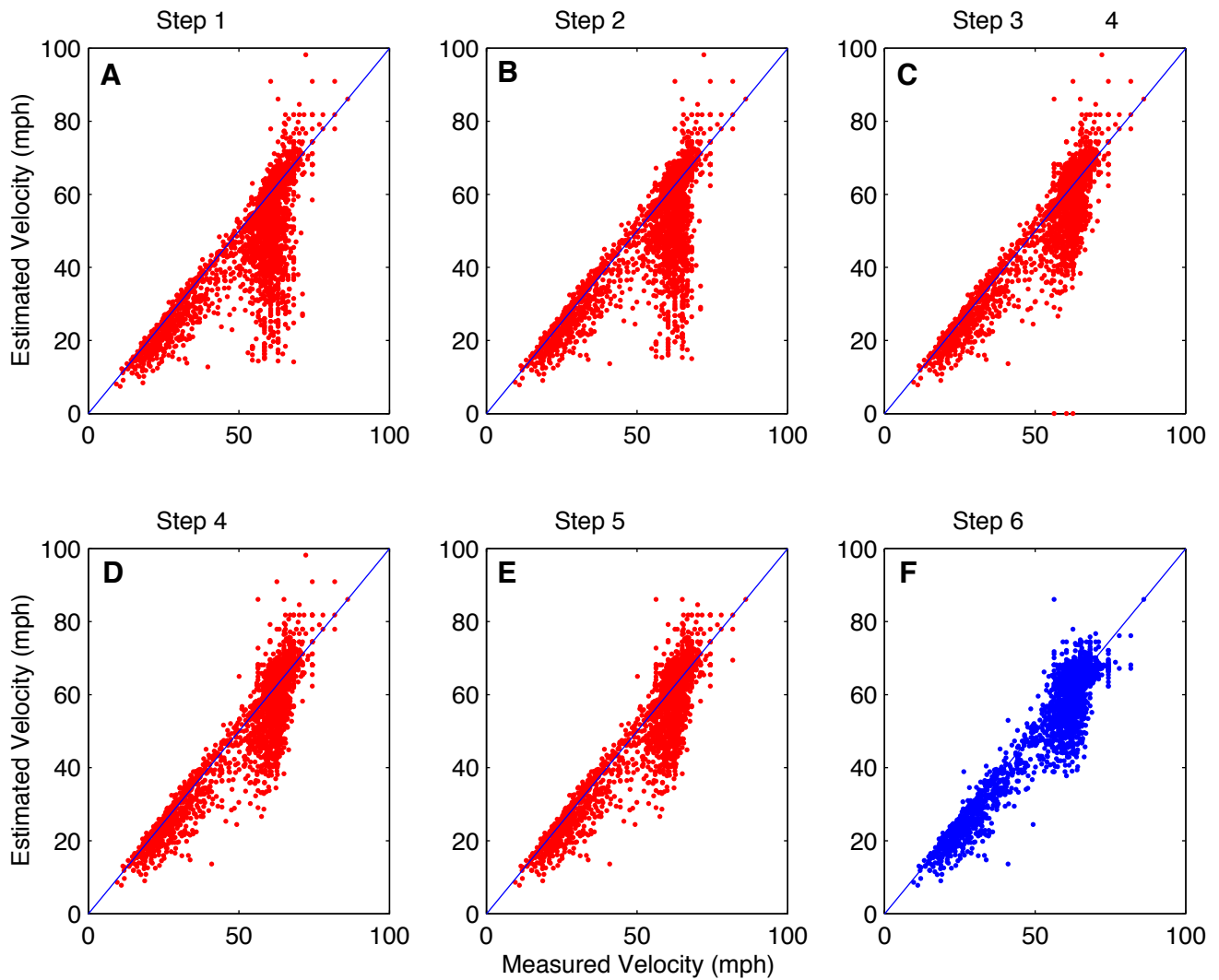


Figure 3.2 Plot showing progression of improvement in speed estimates for lane 4 on BHL station 4 westbound on 12th August 1998. The plot A shows the estimates at the end of Step 1, B shows the estimates at the end of Step 2, C shows the estimates at the end of Step 3, D shows the estimates at end of Step 4, E at the end of Step 5 and F at the end of Step 6.

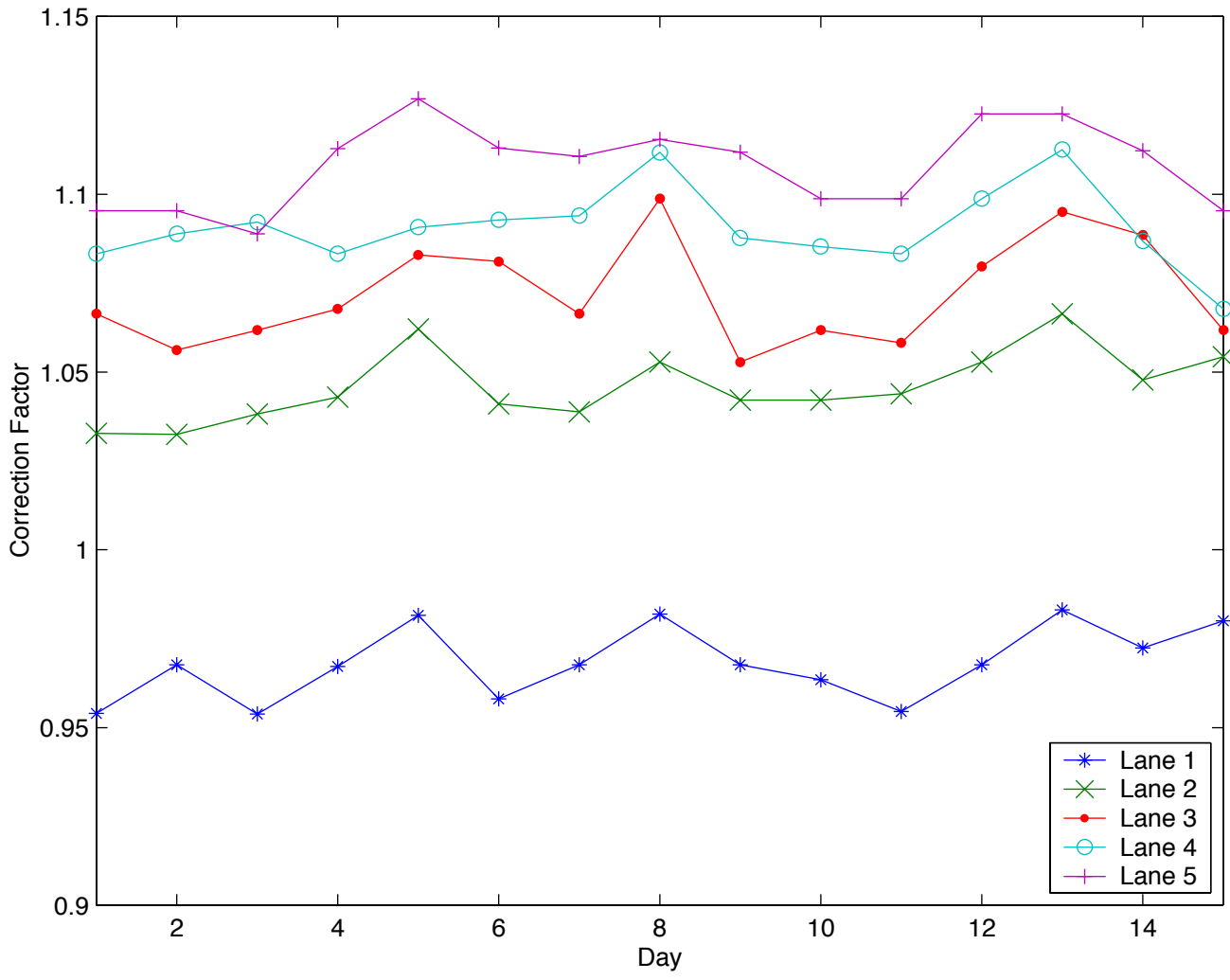


Figure 3.3 Variation in correction factors across days for 15 days data collected at BHL station 4 westbound.

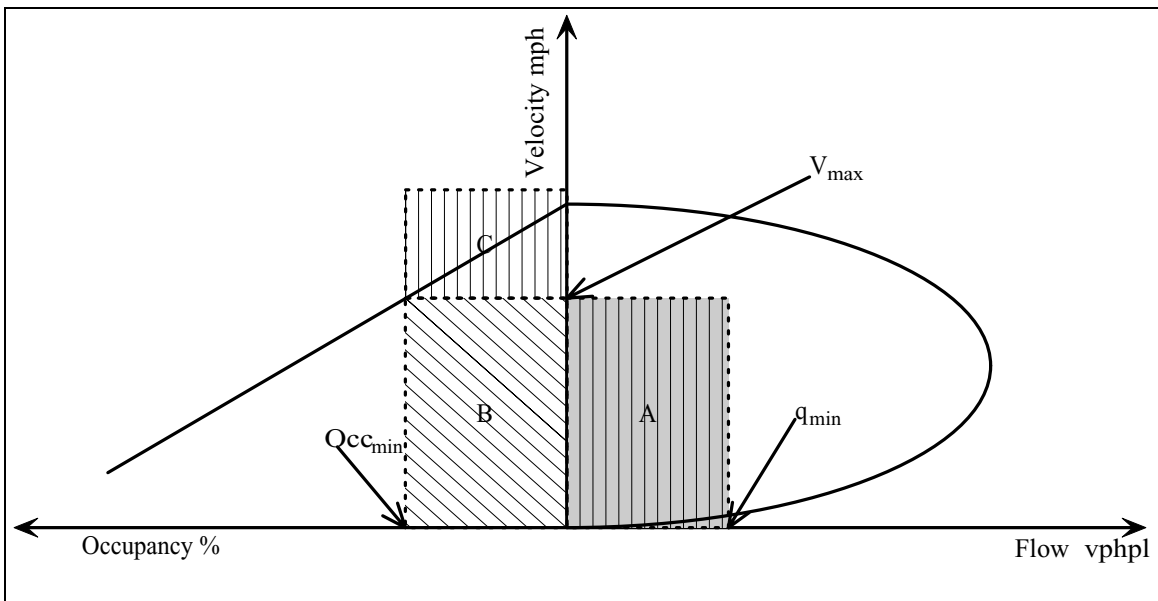


Figure 3.4, Plot showing relationship between flow-speed and occupancy-speed

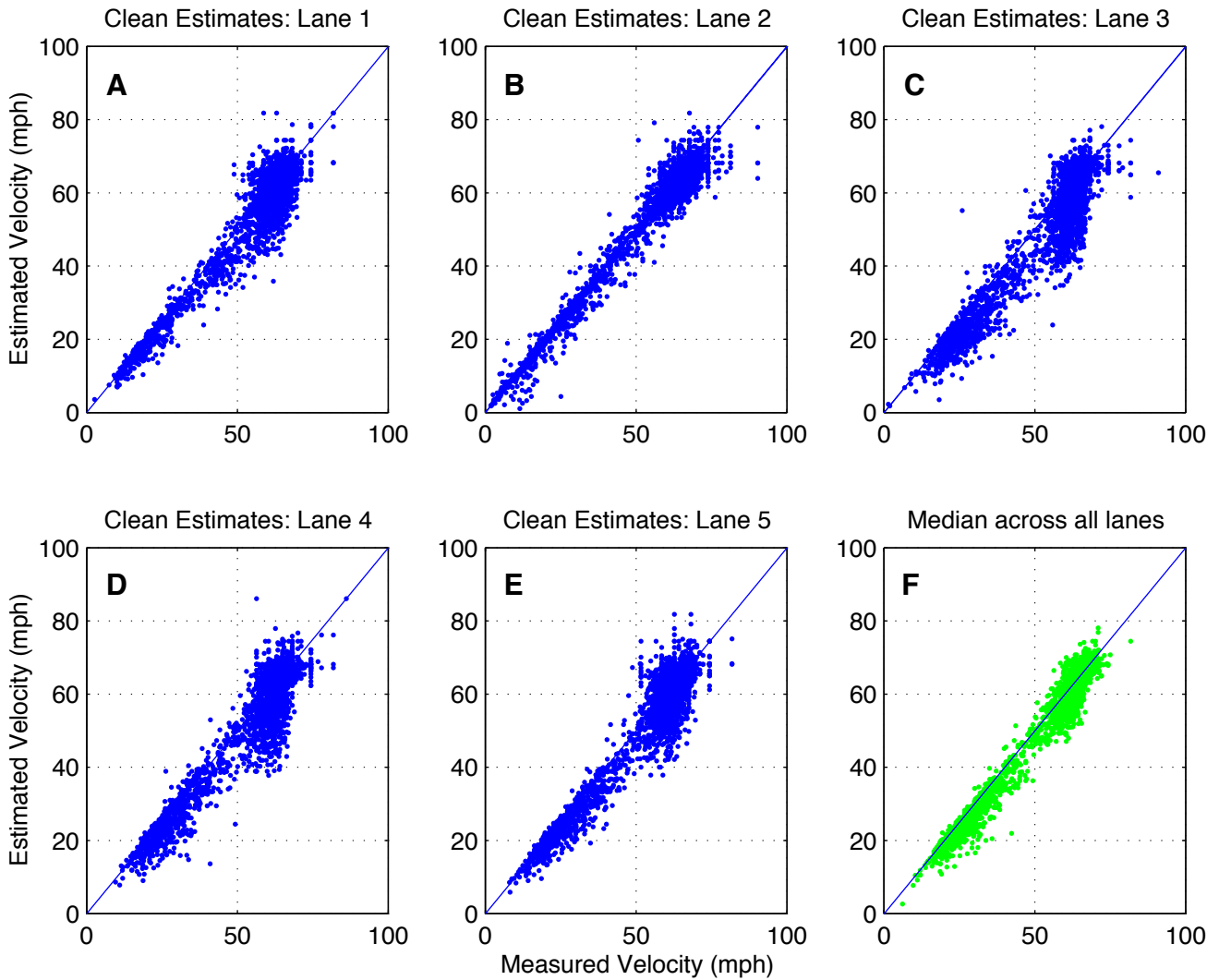


Figure 3.5, Plot showing improved speed estimates at the end of filtering for individual lanes (A-E: lanes 1-5) and improved median estimate across all lanes (F). Note that lane 1 is restricted to High Occupancy Vehicles (HOV) and we do not use adjacent lanes data to estimate speeds for this lane because its time-series is different from the other lanes.

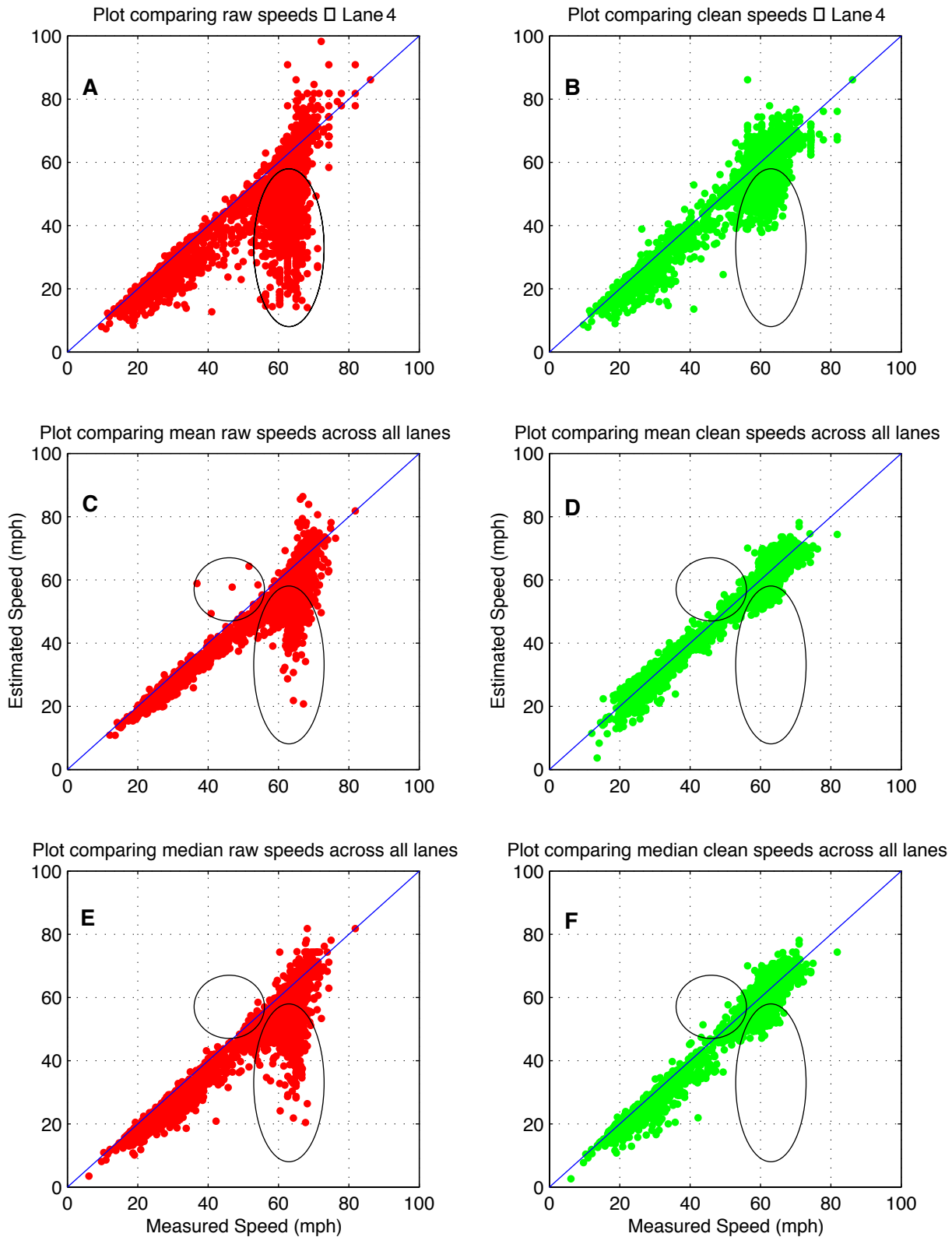


Figure 3.6, This figure shows (A-B) estimates from a single lane, (C-D) mean across all lanes and (E-F) median across all lanes. The left column shows conventional estimates from Equation 3.1 while the right column shows the new estimates from the methodology developed in this paper. In each row, the circles are placed at the same location in the plot pair to highlight the difference between the estimations. Observe that mean across lanes shows greater scatter than median across lanes for raw estimates (C and E). After cleaning the estimates using the filters, both mean and median across lanes show similar performance.

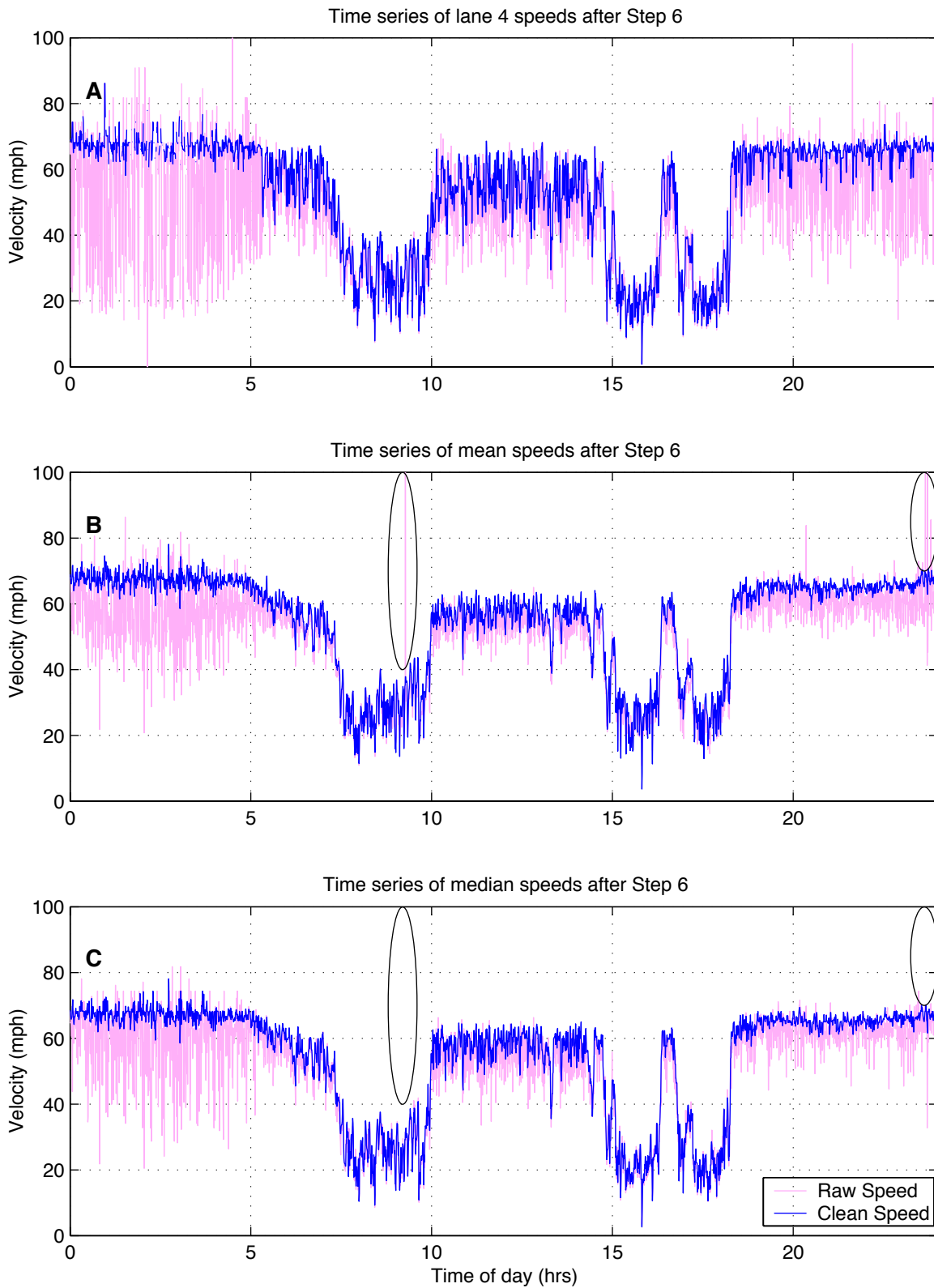


Figure 3.7 Time series of speeds after Step 6 (Clean) plotted over the speeds at the end of Step 2 (Raw). (A) Speeds for Lane 4, (B) mean speeds across lanes and (C) median speed across lanes. The circles are placed to illustrate the effect of outlier in a lane. The mean value is more sensitive to outliers than median

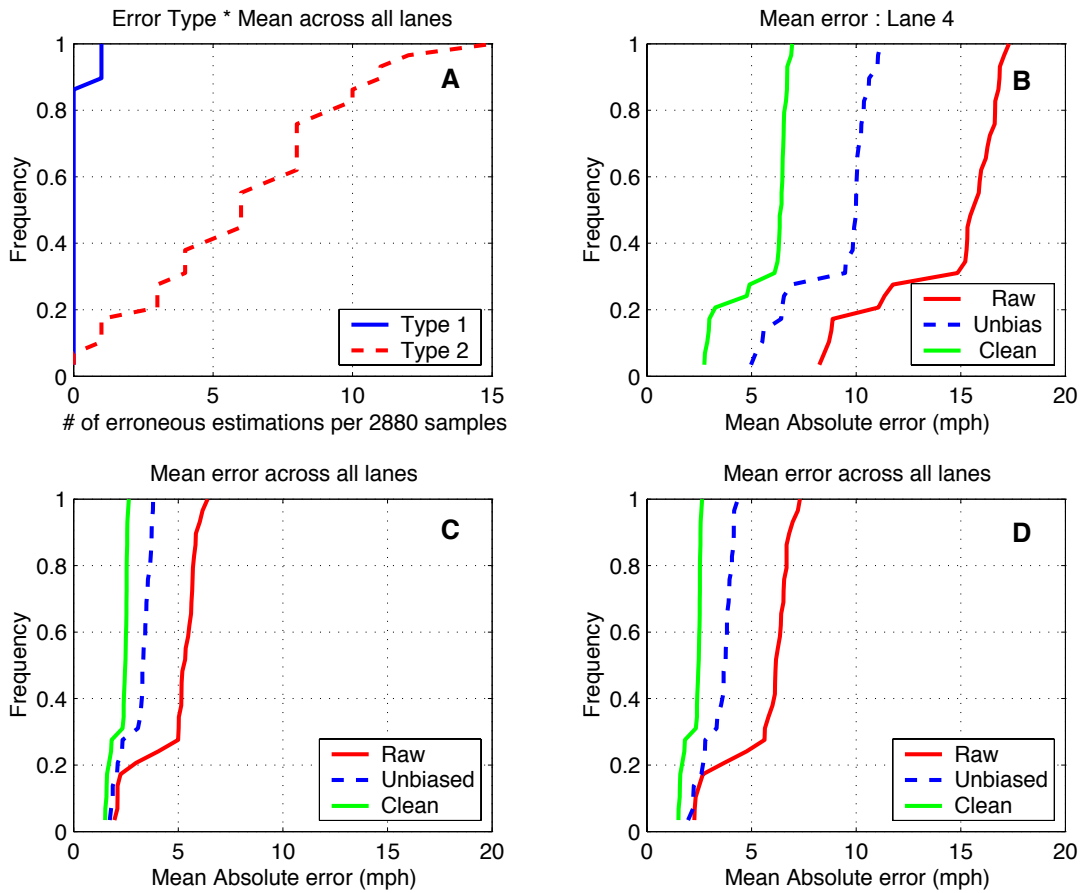


Figure 3.8, Cumulative Distribution Function (CDF) of (A) number of erroneous samples out of 2880 30-second samples for error types 1-2. (B) CDF of absolute mean error for Lane 4, (C) absolute mean error for median of speeds across all lanes and (D) absolute mean error for mean of speeds across all lanes for Raw and Unbiased estimates. For reference, (D) shows CDF of errors for median across lanes for clean speeds.

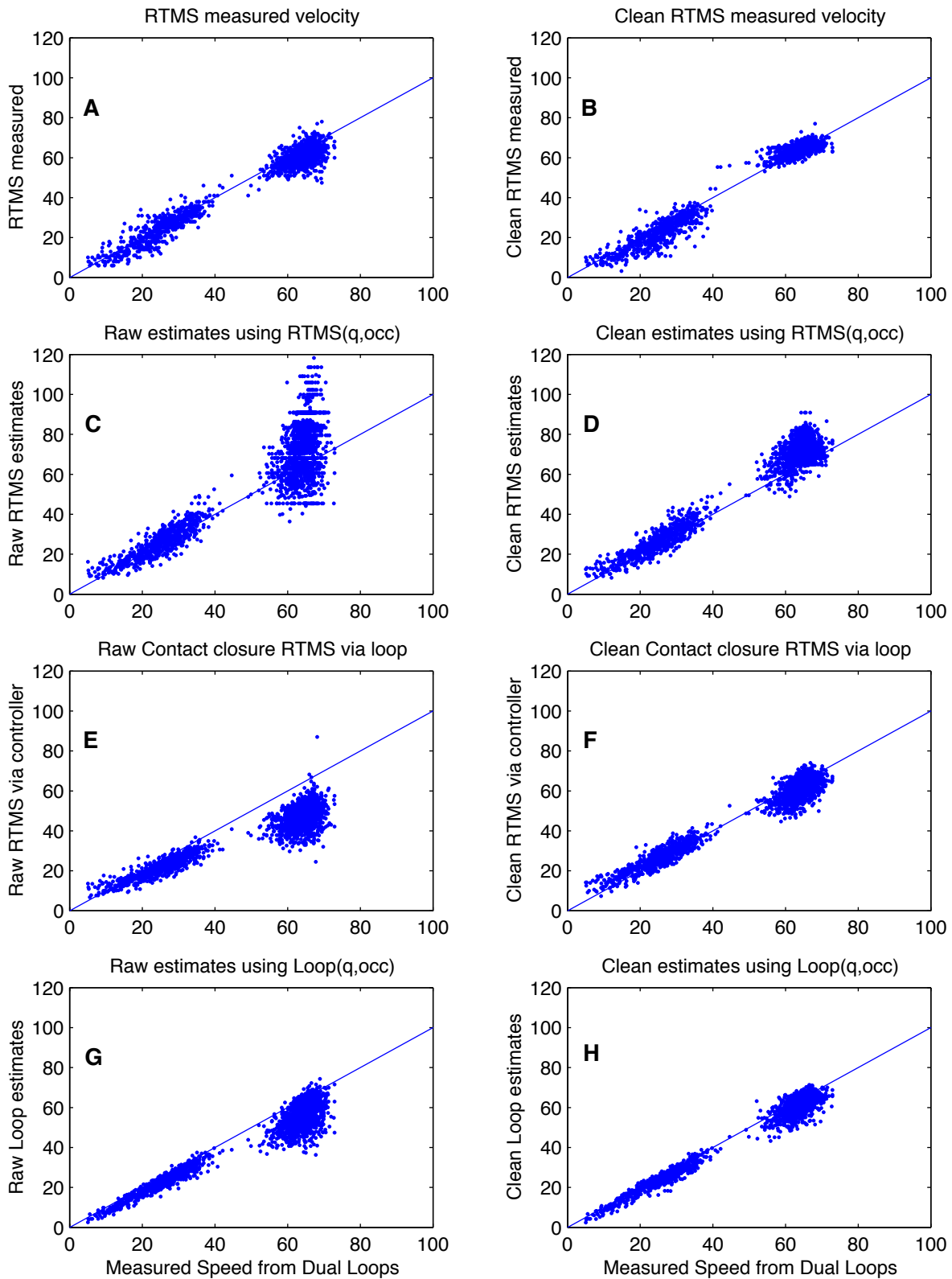


Figure 3.9, Scatter plots comparing median estimated and measured speed across all lanes. The left hand column shows the raw estimates and the right hand column shows the corresponding cleaned estimates after applying the filters. The first row, A-B, shows the speed estimates reported by the RTMS. The second row, C-D, shows the speed estimates after applying Equation 3.1 to the measured q and occ reported directly by the RTMS. The third row, E-F, shows the speed estimates from Equation 3.1 applied to q and occ for RTMS-via-controller. The fourth row shows the speed estimates from the loop q and occ .

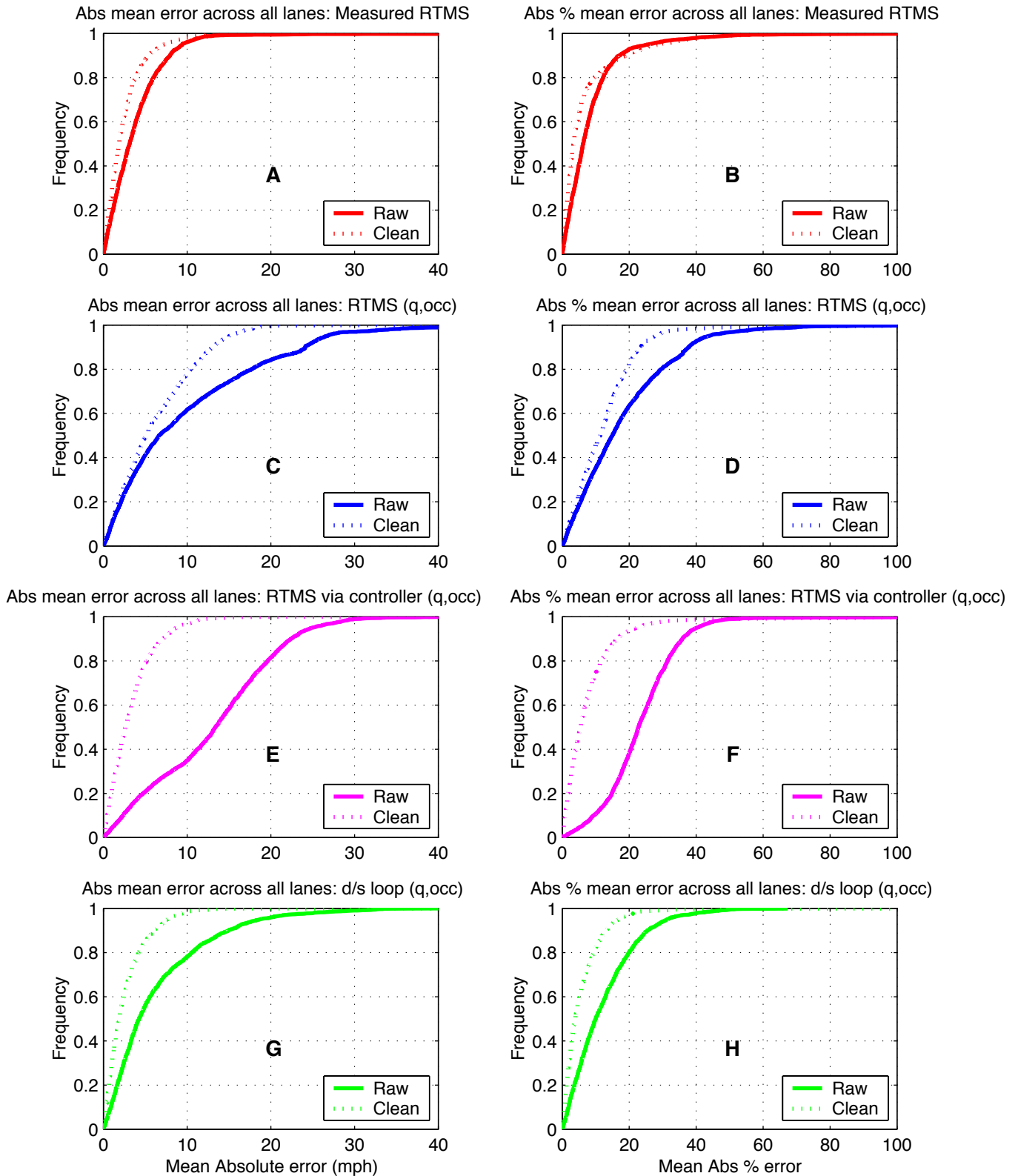


Figure 3.10, Distribution of absolute and percentage error for median estimated and measured speed across all lanes. The left hand column shows the distribution of absolute mean error and the right hand column shows the distribution for absolute percentage error. The first row, A-B, shows the speed estimates reported by the RTMS. The second row, C-D, shows the speed estimates after applying Equation 3.1 to the measured q and occ reported directly by the RTMS. The third row, E-F, shows the speed estimates from Equation 3.1 applied to q and occ for RTMS-via-controller. The fourth row shows the speed estimates from the loop q and occ .

4 DECENTRALIZED SURVEILLANCE AND CONTROL ALGORITHMS

As previously noted, many conventional traffic surveillance systems use dedicated communications links to report data from a field station to a centralized control system at a TMC. Considerable resources are devoted to transmitting these data. Yet most of the time, the response to the real-time data is simply that no action is necessary. This results in a high cost of communication, which limits the coverage area of surveillance technologies. Researchers and operating agencies have both realized that driver based cellular phone reporting provides an alternative means of incident detection, which is effectively an event-based surveillance system. The success is based on the fact that drivers replace the need for static detectors and assume part of the communication costs. But one cannot completely rely on cellular phones for surveillance due to the wide variance of response rates, potentially high for some incidents and low for others, the imprecision of driver's observations, and the need for human operator's to answer the calls. Furthermore, cellular phones might be effective in incident detection in congested corridors but may not be useful for other ITS applications like traveler information and ramp metering. Deriving an analogy from cellular-phone based event driven surveillance system, we can develop surveillance algorithms that meet the needs of traffic applications and provide expanded coverage at a lower cost without sacrificing the performance of other applications.

Different ITS applications require differing rates of data transfer, for instance, incident management requires data transfer at a higher frequency during incident detection and clearance with lower rates other times, traveler information may only require data transfer when the operating speeds change significantly. The amount of data transferred also varies from location to location and one might expect a large amount of data transferred from a location that experiences high traffic volume and longer periods of congestion as compared to a remote location that rarely sees congestion. Moreover, in such remote locations, the need and operating cost of dedicated communication link often is not justified.

In the following section we present several algorithms for different communication scenarios that pre-filter data in the field and only initiate communication when the information might benefit decision-making. These algorithms differ primarily in the criteria for the pre-filtering decisions, subject to the goals of reducing surveillance costs and improving the benefits from transmitted data. The algorithms are developed based on the assumption that a communication technology is available that charges per unit of data transferred. For such an event-based surveillance system, the cost of communication is directly proportional to the amount of data transferred, as will be discussed in Chapter 5.

4.1 Algorithms for Different Communication Scenarios

Below we present a conceptual overview of the communication modes developed for decentralized surveillance in the increasing order of information transmitted. The algorithms apply threshold value tests to 30-second aggregated detector data by comparing measurements against maximum and minimum acceptable values to validate the data as well as make a decision about the need for communication. The methodology developed is generic and could be modified for other sampling periods, possibly with some changes to thresholds. To identify the trends and relationship between traffic variables and to make the decision to transmit, the algorithms use a single measure of flow, occupancy and speed obtained by combining aggregated data from all

lanes and taking the mean of flow, mean of occupancy, and median of speed. Care is taken to use only the lanes that have non-zero flow during the sampling period and are not marked as malfunctioning. The data are cleaned using the macroscopic data-cleaning algorithm presented in Chapter 3, which allows us to pull out trend and relationships among traffic variables thereby simplifying the communication algorithms.

Although the algorithms use data combined across lanes to make decisions, they also track anomalies in an individual lane's behavior and lane blockages to diagnose the detectors, i.e., whether a lane reports speeds that are consistently higher or lower than what adjacent lanes experience or a lane shows zero flow when the adjacent lanes experience flows close to capacity. The former potentially indicating that either the loop calibration is erroneous and the latter potentially indicating that the lane is blocked or the detector is malfunctioning. Similarly, if a lane or set of lanes that had been blocked then starts to flow, it might indicate that the block has been removed.

When the algorithms contact the TMC for transmitting speed information, diagnostic information is also transmitted. Finally, to verify that an otherwise silent station is functioning properly, to evaluate and calibrate the thresholds used in the algorithms (as discussed in Section 4.1.2), and to collect data for planning applications all of the algorithms send summary data on a lower frequency (once a day) that includes Average Daily Traffic (ADT) counts, average daily profile of flow occupancy, speed and daily number of transmissions made. ADT counts can help measure Annual Average Daily Traffic (AADT) that provides an estimate of typical daily traffic on a road segment for all days and a quick indication of average usage of roadway.

4.1.1 *Communication Modes*

The following descriptions provide an overview of the five communication modes, with a more detailed description given in Appendix A.

Mode 1: Free-flow or congested

From an operations standpoint, the most important information for a traffic surveillance system is to determine reliably and quickly whether the facility is free flowing or congested. The Communication Mode 1 seeks to minimize the number of transmissions while conveying this state change by transmitting information only when the traffic condition changes between free-flow and congested. The algorithm identifies the traffic state locally and initiates communication only when the speeds at the detector station drop below a threshold, e.g., 50 mph on a 65 mph facility, indicating that the facility is likely congested and then again when the speed rises above the threshold, indicating that the facility has returned to free flowing.

The algorithm combines data across lanes using the methodology of Chapter 3 to reduce a considerable amount of the noise normally observed in freeway data. In order to further make the algorithm tolerant to noise, the decision to transmit information is triggered when a set of conditions, as described in Appendix A, are satisfied, e.g., in order to reduce the error due to traffic speeds/volume flickering in and out of free-flow conditions, the algorithm waits for few consecutive samples to be above the threshold before deciding that the traffic is free-flowing. The flowchart in Figure 4.1 shows the logic for Communication Mode 1.

Mode 2: Multiple thresholds drop

In the previous Communication Mode, only information about a change in traffic speed from free-flow to congested and vice-versa is transmitted. No further information about variation in traffic speed is transmitted. Often operators are also interested in quantifying the magnitude of congestion and need more information to compute statistics like traveler delay and travel time. To this end a communication scenario that goes a step further in providing data resolution at a slightly greater communication cost than Mode 1 is developed. Usually the speeds on a freeway are observed to deteriorate slowly, especially during the onset of recurring congestion, and then remain at a particular level for extended period. The level at which the speeds become stable might vary from day to day and is influenced by demand, location, and weather. Knowing this level on a given day may help applications, e.g., ATIS. The communication scenario in Mode 2 transmits information about the traffic speed dropping below each of multiple thresholds when the facility becomes congested until the speeds become stable or reach a lower bound, as discussed below, after which the information is transmitted only when the facility returns to free-flow conditions. The different threshold ranges for a freeway with posted free-flow speed of 65 mph could be set at 50 mph, 40 mph, 30 mph and 20 mph, as the speed between consecutive samples during a stable traffic state is observed to vary within the range of 10 mph. A different threshold range, either smaller or larger than 10 mph, can be used depending upon the resolution desired. The flowchart in Figure 4.2 shows the logic for Communication Mode 2.

Mode 3: Multiple thresholds drop and recovery

The communication scenario in Mode 2 follows dropping speeds but does not report any improved conditions until free-flow is observed. In Mode 3, the algorithm is similar to Mode 2, transmitting data when the traffic drops below several thresholds but it also reports recovery slowly tracking any trends of increasing speeds. The rising speed thresholds are followed because data transmitted in Mode 2 may not be sufficient for some applications like travel time information. The criteria for transmitting information to report the decreasing speeds are same as for Mode 2 and the slow recovery in this algorithm follows a similar process except it uses a larger number of samples to make the decision to transmit for increasing speeds. The flowchart in Figure 4.3 shows the logic for Communication Mode 3.

Mode 4: Significant deviation from last reported state

The communication scenarios in Modes 1-3 check whether the traffic state falls in one of the pre-defined threshold ranges. The information is transmitted whenever a set of conditions for a given range is satisfied. This approach may not be able to capture a sharp variation in speeds that does not last for many samples. For applications like incident management, operators are interested in identifying and localizing the source of delay as quickly as possible. The short and sharp changes in traffic speed help identify and localize sources of delay quickly. Also the algorithms in Mode 1-2 fail to transmit information about partial recoveries that do not lead to free-flow conditions. To address these omissions, a communication mode that can provide precise measurements and quicker response without increasing significantly the number of transmissions, and thereby the communication cost, is developed.

As noted earlier, the maximum benefit from a piece of information is derived when it is used for making a decision or when it can help the decision making process. Criteria are developed for this Communication Mode to pre-filter data and only transmit information whenever the current traffic state changes significantly from the expected or the last reported state. This transmission decision includes situations when a piece of information can help decrease the uncertainty, confirm a state and assist in decision-making beyond the station. In order to help the TMC make a better estimate of traffic speed during the periods when no data are transmitted, the algorithm in this Mode identifies the recent trend of speed measurements and computes the rate of change of speed, as defined below, which is transmitted to the TMC along with the current speed. The trend is defined as the general direction in which speed along successive samples change and the traffic speed measurements are categorized in three possible trends: negative trend, positive trend, and stable trend.

Mode 5: Highest frequency when congested

The detector data are potentially of greatest interest to the TMC when the conditions in the field are not free flowing. The objective of Communication Mode 5 is to minimize the number of transmissions during free-flowing periods, as the freeway speeds can easily be estimated during these periods provided the freeway is known to be free-flow. If the facility is not free flowing then the algorithm transmits data to the TMC at the highest frequency, i.e., every sample, until the facility returns to free-flow conditions. Since most facilities are uncongested for the majority of the time, this approach can reduce the number of transmissions without reducing speed precision. The flowchart in Figure 4.4 shows the pre-filtering logic for Communication Mode 5.

4.1.2 Calibration of the Algorithms

As noted earlier, the surveillance above algorithms apply threshold value tests to identify trends, relationship between traffic variables and to make a decision about the need to transmit information to the TMC. The threshold values can be location and communication mode specific. In some cases the algorithms will have to be recalibrated and thresholds reset to account for the specific geometrical and physical nature of the roadway being monitored or the detectors being used, e.g., respectively a detector station located near a diverge might require different threshold set or a loop that differs from 6 ft. by 6 ft. In such cases, if the historical data are available, then these data can be used to set the thresholds and calibrate the algorithm. In the absence of historical data, the algorithms can be deployed with the thresholds developed in this work and allowed to calibrate themselves for few days, as outlined below, before the output can be used reliably and confidently. The communication algorithms are intended to send a daily profile of flow, occupancy and speed with the summary data, which can be used to establish long-term trends at a field station. The thresholds are re-calibrated based on long-term trends, e.g., a given facility should be free-flowing for the majority of a 24 hour period therefore the time series of speed estimated using the information transmitted by the algorithms should have most of the samples above the free-flow threshold and less than some maximum feasible value. Similarly if the long-term profile indicates that the station exhibits high degree of flicker, i.e., the traffic speed frequently and quickly changing between free-flow and congestion, then the time for which the algorithm waits to identify a given trend has to be increased. This mechanism also helps the algorithms adapt to time-of-day, day-of-week, and seasonal trends. Also, the data-

cleaning algorithm discussed in Chapter 3 calibrates the algorithms for detector spacing by computing a correction factor for each location.

The threshold values used in this work for each mode were developed empirically using historical data collected from 40 detector stations on I-70/71 corridor in Columbus, OH and 97 detector stations on SR-101 through San Francisco, CA. These thresholds were then tested with data collected on different days from I-70/71 and from four other freeways/highways on a diverse set of locations in California (4 detector stations on I-80 in Berkeley, 8 detector stations on SR-51 through Sacramento, 18 detector stations on I-15 through San Diego and 40 detector stations on I-405 through Los Angeles). The performance measures, i.e., the numbers of transmissions, mean absolute error between measured and estimated speeds, mean percentage absolute errors between measured and estimated speed are computed in Section 4.4. These measures exhibit similar trends on each of the corridors indicating that the thresholds developed and used in the algorithms presented above are transferable.

4.2 Other features

To minimize false alarms and avoid obscuring real events, the decentralized surveillance structure requires a rigorous detector diagnostic and data validation process, a synchronized controller clock with the TMC, and reliable communication structure. The detector diagnostic and data validation tools developed in Chapter 3 address long duration errors, e.g., if a lane or set of lanes at a detector station observe noisy measurements for majority of the day. The data validation algorithm identifies the erroneous lanes using the relationship between the variables flow, occupancy and speed. The field controller algorithm ignores the malfunctioning lanes until manually reset by the TMC. To address the short term transient and local errors, e.g., a detector station exhibiting high degree of noise and flicker, the algorithms keep track of the frequency and number of samples that passed between changes of state from free-flow to congested or back. If the duration of time that a given state persists is short (less than 3 minutes in our case) and if this change of state happens frequently (more than twice within 15 minutes in our case), then the algorithms shift to a new set of thresholds which have longer waiting period for making a decision to transmit. In our case, the new thresholds are set at two samples (one minute) above the original threshold values. The algorithm returns back to original thresholds if the noise and flicker in the data is observed to have subsided, i.e., frequent change of state is not observed for more than a half hour period.

Further detector diagnostic and performance evaluation can be done at the TMC by using the summary information sent by the algorithms on a lower frequency (once a day) which includes ADT counts, average daily profile of flow occupancy, speed and daily number of transmissions made. The TMC can compare temporal and spatial trends in ADT across stations to assess detector performance. If the ADT values differ considerably between two adjacent stations on a day or across days at a station then the detector station might need field maintenance. Similarly, to verify that individual detectors are functioning properly and identify if individual lanes deviate from expected patterns, the field controller algorithms build a profile of transitions between free-flow and non free-flow state across days. This profile is expected to remain stable across similar days. If the algorithm suspects that a particular detector is malfunctioning, then it can send a request for maintenance to the TMC. If the surveillance system has duplex communication, then the TMC could remotely respond to the field initiated maintenance request, e.g., asking the field station to ignore specific detectors until further notice.

The issue of synchronizing controller clocks can be addressed by installing an inexpensive Global Positioning System (GPS) receiver at the field station. This GPS receiver would also allow the field station to correct for daylight saving since the GPS signal includes the date. Finally, a reliable communication network is critical for decentralized surveillance, both to reduce the cost and to minimize the missed events. The communication structure developed in the following Chapter 5 addresses this issue in detail.

The algorithms can be designed to be location independent and flexible by developing a site-specific configuration file and deploying the file on the field and at the TMC. The configuration file would contain information about the data structure to be used during transmission and any modification to be made in the algorithm when the data formats are modified. The configuration file would also contain pertinent information about the geometry of the location, e.g., presence of High Occupancy Vehicle (HOV) lanes, ramp information and information about any special events (major events, national holidays etc.) that are likely to alter the traffic flow pattern. To respond to these events the algorithm might be required to shift to a different communication mode.

Often a TMC is only staffed for a part of the day and the data transmitted from the field stations are only used in real-time while the TMC is occupied. If the surveillance system is designed with the possibility of two-way communication between the field device and the TMC, then the TMC could potentially ask the field stations to stop sending data during specific hours or days. In this case, the field stations would only transmit the end of the day summary information during the "off" hours. Two-way communication also allows for the possibility of the TMC playing a proactive role in the surveillance, for instance, the TMC could inform the upstream stations to expect congested conditions due to a known lane blockage at a downstream station. This communication option also allows the TMC to request recent historical data from adjacent stations when an incident is suspected. If the field controller suspects the requested information may contain a transient error, it would also inform the TMC of this fact.

The algorithms provide the flexibility of changing between different communications scenarios and mode to suit the changing cost of communication, demand on the facility and data needs of the traffic application. For instance, by tracking the number of transmissions at the TMC, a station that uses event driven transmission on a per-transmission cost basis can be migrated to flat rate communication as demand increases. Similarly, if a new traffic application that needs more data is installed at the TMC, e.g., incident management, then the communication scenarios in Mode 4 or 5 might be chosen as they transfer more information compared to other modes.

4.3 Constructing synthetic time series of speed

Many existing traffic applications use time series speed as an input. Conventionally all the data from field stations are transferred at a fixed rate and the time series of speed, flow, and occupancy can be generated directly in the TMC. In the event-based system developed in this report, such a time series would have to be estimated using the limited amount of information transmitted. In the rest of this section, we describe the methodology for estimating time series speed by using limited information transmitted for each Communication Mode presented in Section 4.1.1.

For Mode 1, the synthetic time series is constructed by assuming that the sample speed observed at the time of communication with the TMC is representative of the speed in the future samples

until more information is transmitted from the field. For Modes 2 and 3 the time series is constructed by assuming that the mean of the threshold range is representative of the speed until the speeds move to a different threshold range or returns to free-flow state. In Mode 4 the algorithm monitors whether the speeds are dropping, recovering or are stable. When the speeds are observed to be stable, the rate of change of speeds is zero and the median of the five most recent sample speeds is reported as the operating speed. For all five modes the synthetic time series can be estimated by assuming that speed stays at the last reported value until such time that the field algorithm sends more information. In case the field algorithm observes that the speeds are dropping or recovering, the rate of change of speed is computed using recent historical information, as described earlier. The time series is constructed by assuming that samples in the successive future samples follow a linear profile with the reported rate of change until more information is transmitted from the field. In Mode 5, data is transmitted at the highest frequency during congestion so the time series needs to be estimated only for the free-flow periods. The estimates for the free-flow period are assumed to be constant, e.g., the posted free flow speed.

To aid such a process, a more sophisticated algorithm could be developed where the field sends measure of confidence with the rate of change of speed. Coupling these measures of confidence with long term trends extracted from historical data, one can possibly obtain better estimates. Other traffic state estimation algorithms and time series forecasting algorithms that employ extrapolative forecasting methods may help improve the accuracy of the synthetic time series, but were not considered in this research.

4.4 Analysis of the algorithms and their usefulness

In this section we evaluate the performance of various surveillance algorithms presented in Section 4.1.1 by computing and analyzing performance measures like the ratio of the number of transmissions to total number of samples observed in the field, mean absolute error between estimated and measured time series speed, and mean absolute percentage error between estimated and measured time series speed. The average error is then split between free-flowing and non free-flowing periods to analyze the distribution of error between two traffic states. In section 4.4.1 we analyze the effectiveness of decentralized surveillance algorithm for real time traffic applications with a sample case study for traveler information systems. Finally, in Section 4.4.2 we evaluate the performance and usefulness of lower grade surveillance algorithms in Modes 1-3, which aim at minimizing the number of transmissions, for remote locations. The detailed analysis presented here is for data collected at 40 detector stations during February 2002 on I-70/71 corridor in Columbus, OH. Similar analysis was performed for 4 other corridors, as described in Section 1.3, in California and their results are presented in Appendix B.

The plots in Figure 4.5 compare the time-series speed for a northbound station (Station 107) at I-71 in Columbus, OH, on a randomly chosen weekday. The plots show the time series speed after passing through the data-cleaning algorithm developed in Chapter 3 and the synthetic time series constructed using the methodology described in Section 4.3. The plots compare cleaned times series because most stations examined in this study exhibited noisy data and thus, the algorithm uses the clean time series as the input for determining the trends and deciding the transfer points when the algorithm decides to send information to the TMC (as evident in the figure). The plots A-E are for communication Modes 1-5 respectively. The horizontal axis shows the 30-second sample number starting at 1 from 12:00 a.m. to illustrate that several samples are observed

between two successive transfer points. A second horizontal axis at the bottom of the plot E shows the time of the day. As can be observed from the plots, the synthetic time series follows more closely the cleaned time series with the increasing Mode number. This is not surprising as the communication scenarios in Mode 1-5 are presented in increasing order of information transmitted. The communication scenario in Mode 4 follows the cleaned time series closely without requiring many transfer points because of the pre-filtering logic used for the transmission decision. The synthetic time series in Mode 2 and 3 deviate considerably from the input series because the algorithm in Mode 2 and 3 transmit information at discrete thresholds separated by 10 mph. The estimates for Mode 3 are slightly better than those for Mode 2 during congestion (as highlighted by dashed ovals) because the algorithm in Mode 3 follows slowly recovering speeds during congestion.

Figure 4.6 shows the average number of transmissions made by stations during one month, expressed as a percentage of the daily samples collected in the field on I-70/71 corridor. The daily numbers of transmissions at a station are summed up for both northbound and southbound direction lanes. The plots A-E show the percentage of transmissions for Mode 1-5 respectively. Although not shown in the figure, the average numbers of transmissions on weekend day were observed to be lower than those for weekdays. A horizontal axis at the bottom of the figure shows the location of Central Business District (CBD) and the relative distance of the detector stations from the CBD expressed in miles. As expected, the number of transmissions is minimum for Mode 1 and maximum for Mode 5 (note the different vertical scale for plot E). The transmission frequency for Mode 1-4 is comparable to each other. As observed in the figure, the numbers of transmissions for all communication modes are highest for stations close to the CBD, while the remote stations require few or no transmissions. This distribution is not surprising because the stations near the CBD observe longer periods of congestion and higher traffic volumes compared to remote stations. The fewer number of transmissions for remote stations further justifies the need for a decentralized event-based surveillance structure where the cost of communication is usage based. Table 4.1 summarizes the results of Figure 4.6 showing the mean and maximum number of transmissions across the 40 stations. To get the percentages shown in the figure, one should divide the values in the table by the maximum feasible number of transmissions, 2880 samples per day.

The plots in Figures 4.7-4.11 show the error distribution of synthetic time series speed estimates as compared to the clean speeds for Modes 1-5 respectively. In each figure, A1 shows the surface plot of daily mean absolute error between synthetic and clean time series speed at each station for a month's data, A2 shows the mean absolute error at each station averaged across days for the same data, and A3 shows the daily mean absolute error averaged across 40 detector stations. The plots B1, B2 and B3 show the average percentage absolute error similar to A1, A2 and A3 respectively. As can be observed from the plots A2 and A3 of Figures 4.7-4.11, the mean absolute error for each station and day is less than 5 mph for all modes. Similarly, as observed from plots B2 and B3, the average percentage absolute error is less than 8 percent for all modes.

The errors shown in Figures 4.7-4.11 are averaged across entire day's data, both free-flow and congestion period's combined. Several stations on the I-70/71 corridor, especially remote stations, experience little or no congestion. In Figure 4.12 we split the error between congested and free-flow periods to analyze the error distribution between two traffic states. If an individual speed measurement of the cleaned time series lies below 50 mph then it is considered as congested. Again, the plots A-E show errors for Mode 1-5 respectively. Each plot shows

combined mean absolute error of all data, free-flow periods mean absolute error and congested periods mean absolute error for 40 stations on I-70/71 northbound corridor. As can be observed from the plots, the combined error and free-flow period's error are less than 5 mph for all modes. The non free-flow periods mean absolute error is observed to be as high as 30 mph for all modes. The high value of error during congested conditions is also seen for communication Mode 5 that transmits at the highest frequency during congestion. Most of the error can be attributed to the fact that the algorithm waits for few samples to confirm the congested traffic state before transmitting the information. The remote stations exhibit high error during congestion because there are few congested samples observed. The stations near the CBD observe many congested samples causing the average error to reduce. The plots in Figure 4.13 show the mean absolute percentage error broken down between free-flow and congested periods and they exhibit similar features as observed for mean absolute error.

Table 4.2 summarizes three performance measures for various freeway segments analyzed in this work: mean absolute error between the synthetic time series and measured speed, mean percentage absolute errors for the synthetic time series and measured speed, and the ratio of the number of transmissions per number of samples at a given station. The table shows the maximum across stations of the daily mean during the periods specified in Section 1.3. Although the thresholds used by each Communication Mode were developed using only data from I-70/71, the thresholds were the same for each of the freeway segments analyzed. The similarity on performance measures for all freeways suggests that the surveillance algorithms are portable and the thresholds developed in this work could be widely applicable. Sub-Table A summarizes measures for I-70/71 corridor in Columbus, OH, Sub-Table B for I-80 in Berkeley, CA, Sub-Table C for SR-51 through Sacramento, Sub-Table D for I-15 through San Diego and Sub-Table E for I-405 thorough Los Angeles.

4.4.1 Case study: traveler information systems

To illustrate the performance of decentralized surveillance algorithms on existing traffic applications, consider estimated travel times on a segment of freeway computed using data collected from conventional surveillance system and that from data generated using information transmitted by decentralized surveillance system developed in this report. As noted earlier, many existing traffic applications use time series speed as an input and in the event-based system, such a time series is estimated using the limited amount of information transmitted and methodology outlined in Section 4.3.

To calculate the travel times on a segment of freeway with several detector stations, the methodology outlined in [29] is followed where the freeway is divided into segments with each segment starting midway between two adjacent detector stations. The time space plane is segmented into a grid, with each cell taking the speed measurement corresponding to that time and station. The travel time between two locations is computed by moving a virtual vehicle in the time space plane with a velocity determined by its current time and segment location and the travel time on the freeway is estimated as the sum of the individual segment's travel time. By repeating this process for entire day, one can estimate the travel-time time series for the freeway. By dividing the travel-time time series with the length of the freeway, link velocity time series can be obtained. We computed the travel-time and link-velocity using clean data and synthetic time series of each Communication Mode. The plots in Figure 4.14 compare the travel-time and link velocity using all data (conventional method), Mode 1 output, Mode 4 output and Mode 5

output. The plots A and B are for a randomly chosen weekday's data and the plots C and D are for the average of one month's data, all collected on a 4.4 mile long section of I-70/71 northbound in Columbus, OH. Only three communication modes are shown for clarity. Mode 1 and 5 represent the lower and upper bound for amount of information transmitted and Mode 4 is of special interest due to its accuracy and quick response time in transmitting change of speed while maintaining a low transmission frequency. As can be observed from the plots, the estimates of Mode 5 are closest to the conventional methodology and that of Mode 1 deviate the most. The estimates of Mode 4 also follow the conventional methodology closely, which is encouraging considering the significant difference in the number of transmissions between Modes 4 and 5. The time series plots of travel-time and link-velocity appear smooth because combining data across several stations reduces much of the noise. This smoothing of travel time is acceptable as people usually travel long distances on the freeways. The plots are further smoothed out when average of a month's data are taken. Figure 4.15 shows several scatter plots comparing average estimated travel times and link velocities between all data and limited data transfer of Modes 1, 4 and 5 for the same data set as in Figure 4.14C-D. The plots in first column compare travel time estimates and the second column compare link velocity estimates. Plot A, B are for Mode 1; C, D are for Mode 4; E, F are for Mode 5. Once more Mode 5 performs the best and Mode 1 the worst. Notice that the Mode 1's performance deteriorates considerably in the presence of larger travel times, i.e., congestion. Mode 4 shows mixed results and Mode 5 performs well during congestion as well as free-flowing periods. The link velocity estimates by Mode 4 and 5 are comparable.

Table 4.3 summarizes several performance measures including mean absolute travel time difference between conventional methodology's estimate and synthetic time series speeds estimate for the communication Modes 1, 4 and 5. Similarly, maximum absolute travel time difference, mean absolute link velocity difference and maximum absolute link velocity differences are summarized. As with Figure 4.14, these measures are computed for a day's data and for average of one month's data. Again the algorithm in Mode 1 performs worst and that in Mode 5 performs the best.

4.4.2 Performance evaluation of lower grade algorithms

As noted earlier, the operating costs of a dedicated communication network may not be justified for remote locations that do not see frequent congestion or that have low traffic volumes. The objective of the lower grade surveillance algorithms in Mode 1-3 is to minimize the communication cost by reducing the number of transmissions, sending a limited amount of information. These algorithms could potentially increase the surveillance coverage in remote locations at low communication costs. To evaluate the usefulness and performance of these lower grade surveillance algorithms in remote locations, a case study was done that measured the detection rate and false alarm rate, as follows. The detection rate is defined as the number of times the algorithm reports a need for transmission when the speeds at the detector station deviate significantly from the estimated time series. This measure can help us compute the frequency of missed calls. The false alarm rate measures the number of incorrect detection calls. For this study, data collected during February 2002 from 30 semi-remote detector stations on I-70/71 corridor were used. These remote stations do not see recurring congestion and required few or no transmissions, as evident in the right hand 3/4ths of Figure 4.6. To compute the number of missed calls, the synthetic time series from each algorithm were compared to the

corresponding complete speed data on the same plot. The numbers of missed calls were then manually counted by visual inspection. Table 4.4 summarizes the total number of missed calls for 30 stations and 26 days of data for northbound and southbound traffic. Detection by the algorithm was suspected to be incorrect if a station reported speed drop and the adjacent stations did not report any drop during the same time window bounded by reasonable travel time. The suspected drops were verified for their validity by comparing the individual lanes speeds. If all the lanes reported a drop in speed then the call was considered to be valid, otherwise it was considered as false alarm. Total numbers of such false alarms were counted and the results are summarized in Table 4.4.

The missed calls by each mode were then further analyzed to determine the kind of features that were missed by the algorithms. It was observed that algorithms in Mode 2 and 3 missed a call frequently whenever the drop in speed was local and short lived. An example of this phenomenon can be observed in Figure 4.5 where the missed call by Mode 2 and 3 is highlighted by a solid oval. The algorithm in Mode 1 also missed a few calls when the drop in speed was transient and short lived. Whether or not these errors are significant depends on the applications that will use the data, but they are included for completeness.

4.5 Variants of the Surveillance algorithms

In this section we analyze two different variants of the algorithms developed in Section 4.1.1. In the first variant, an average daily profile, as described below, of traffic parameters flow, occupancy and speed is developed and used for decision-making process. In the second variant, data from more than one directional highway are analyzed simultaneously by the algorithm to determine the extent of overlap in the need for data transmission.

An average daily profile at a detector location can be developed for every day of the week and season of the year using historical data. For instance, the historical speed data collected for previous 10 Monday's are averaged and then a moving median of 5 samples taken to develop the average profile of speed for a Monday which can be used the following Monday. The daily profiles for all weekdays were found to be similar and the profiles for weekend were observed to be different from weekdays. Thus, for a season of the year two types of profiles were developed, weekday and weekend. In this variant of surveillance algorithm, the TMC assumes that the current traffic state is similar to the average profile unless stated otherwise from the field station. The field algorithm transmits data to the TMC whenever the observed state deviates from the average profile. Communication is initiated again when the traffic state return back to the average profile. Once the traffic state deviates from the average profile, the pre-filtering logic presented in Section 4.1.1 algorithms is followed. The communication algorithm in Mode 4 was modified to include the average profile in the pre-filtering logic and the numbers of daily transmissions, with and without the daily profile included, were compared. There was not a significant change in the number of transmissions for both cases. For some days, the number of transmissions was higher when the algorithm employed the daily average profile. This increase in number of transmissions was caused by the additional need for informing the TMC whether the traffic state is following the average profile or not and for it to transmit the average daily profile at the end of the day. Considerable variation was observed in the daily onset and termination of the recurring congestion that resulted in reduced benefits from the daily profile. Nevertheless, the average profile could be useful for constructing the synthetic time-series as it

can reduce the uncertainty about the estimates during the periods when no transmissions are made from the field.

The second variant of the algorithms analyzes data from all directions at a station simultaneously. Although the traffic in one direction should not influence another direction, the need for transmission by a direction might overlap another direction. Thus, when a decision to transmit information is made by traffic in any one direction, the data is transmitted about the current state of all directions monitored. The maximum overlap in the need for data transmission would be observed if the recurring congestion periods overlap for the different directions. Figure 4.16 compares the average daily number of transmissions on I-70/71 corridor for the two different algorithms. In one algorithm, the number of transmission needed for each direction are simply summed up for all directions where as in the second algorithm the traffic for all directions is monitored simultaneously and the overlap in need for transmissions is considered. As can be observed from the figure, when both directions are considered together the number of transmissions is reduced for all the stations. The maximum reductions in the number of daily transmissions were observed for Mode 5, at stations that observed extended congestion periods.

Table 4.1, Estimated maximum and mean number of daily transmissions per station on I-70/71 corridor in Columbus, OH, across 40 stations for February 2002 data

Communication Mode	Number of transmissions	
	Maximum	Mean
1	28	6
2	25	5
3	26	7
4	30	9
5	413	71

Table 4.2, Summary of performance measures for different corridors analyzed in this work. Table A summarizes measures for I-70/71 corridor in Columbus, OH, Table B for I-80 in Berkeley, CA, Table C for SR-51 through Sacramento, Table D for I-15 through San Diego and Table E for I-405 through Los Angeles

Sub-Table A	I-70/71_CMH		
Communication Mode	Max across stations of the Mean daily Abs Err	Max across stations of the Mean daily % Abs Err	Max across stations of the Mean daily % of Transmissions
1	3.97	10.28	0.97
2	4.15	8.63	0.87
3	3.98	8.21	0.9
4	3.94	7.55	1.04
5	3.18	5.65	14.34
Sub-Table B	I-80_BHL		
Communication Mode	Max across stations of the Mean daily Abs Err	Max across stations of the Mean daily % Abs Err	Max across stations of the Mean daily % of Transmissions
1	5.75	22.29	0.39
2	4.84	12.49	0.5
3	4.48	11.58	0.66
4	4.06	11.84	0.82
5	3.07	7.63	25.17
Sub-Table C	SR-51_D3		
Communication Mode	Max across stations of the Mean daily Abs Err	Max across stations of the Mean daily % Abs Err	Max across stations of the Mean daily % of Transmissions
1	6.5	26.76	0.52
2	5.67	13.69	0.56
3	5.45	13.32	0.59
4	4.94	13.75	0.77
5	5.42	13.69	15.89
Sub-Table D	I-15_D11		
Communication Mode	Max across stations of the Mean daily Abs Err	Max across stations of the Mean daily % Abs Err	Max across stations of the Mean daily % of Transmissions
1	6.36	34.09	0.33
2	6.74	11.65	0.36
3	5.53	10.98	0.43
4	4.84	10.06	0.7
5	4.44	9.83	11.88
Sub-Table E	I-405_D7		
Communication Mode	Max across stations of the Mean daily Abs Err	Max across stations of the Mean daily % Abs Err	Max across stations of the Mean daily % of Transmissions
1	8.05	38.29	0.6
2	7.61	27.88	0.69
3	7.58	26.43	0.78
4	7.4	25.06	0.82
5	7.32	24.24	21.05

Table 4.3, Summary of performance measures for travel time estimation and link velocity estimation

	Average of one month's data			Single weekday's data		
	Mode 1	Mode 4	Mode 5	Mode 1	Mode 4	Mode 5
Max TT Diff (sec)	57.77	35.67	17.50	183.11	31.66	90.00
Max Link Velocity Diff (mph)	-6.50	-4.57	-2.12	-20.04	-6.78	-13.07
Mean abs TT Diff (sec)	6.14	3.91	4.09	8.60	5.30	3.85
Mean abs Link Vel. Diff (mph)	1.12	0.78	0.83	1.48	1.04	0.88

Table 4.4, Total number of false alarms and missed calls by algorithms in Mode 1-3 for data collected during February 2002 at remote detector stations on I-70/71 corridor in Columbus, OH.

	False Alarm	Missed Calls
Mode 1	14	5
Mode 2	13	7
Mode 3	13	7

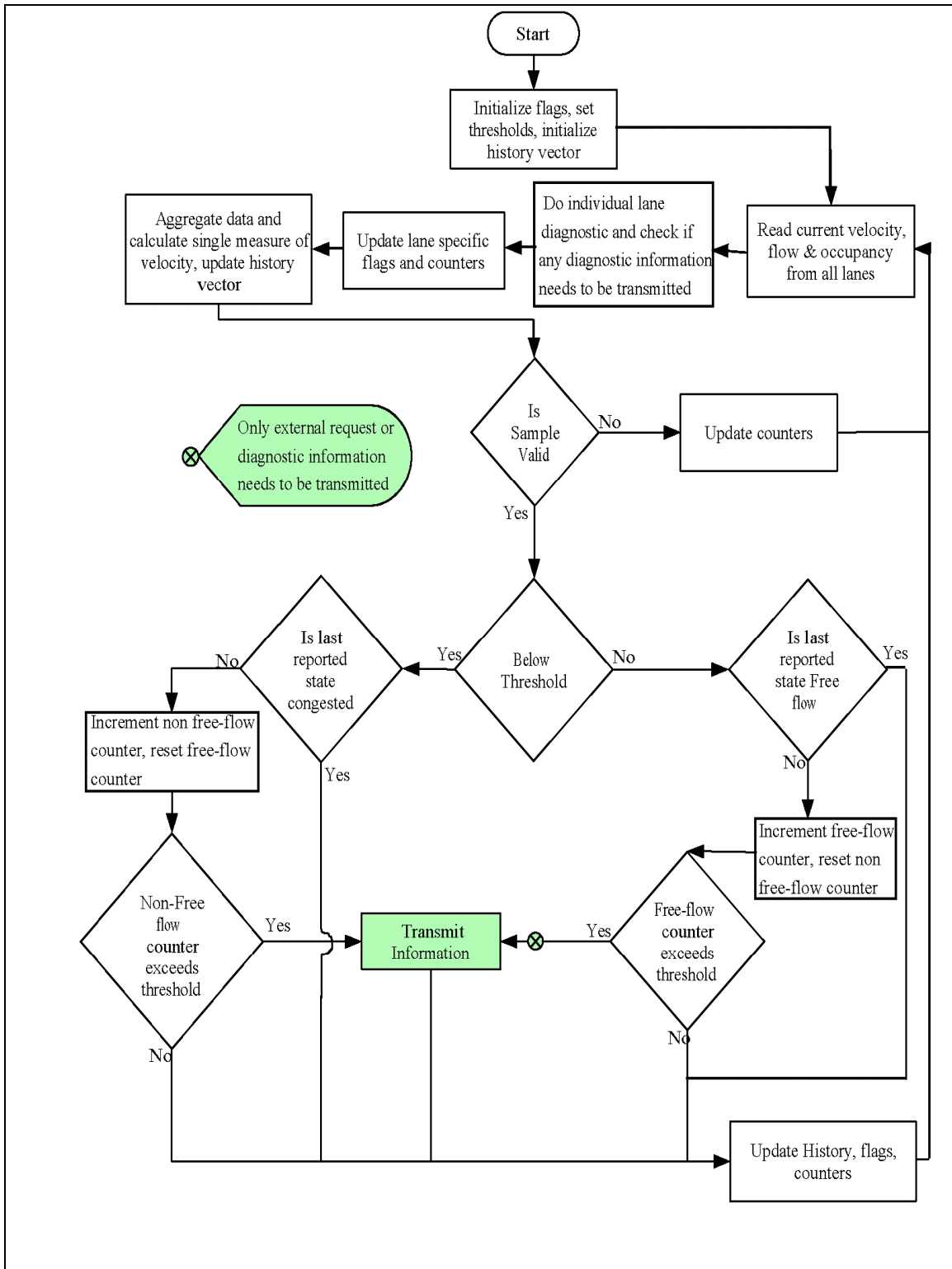


Figure 4.1, Flow chart showing the pre-filtering logic for Communication Mode 1 where the data are transmitted only when the traffic state changes between free-flow and congested.

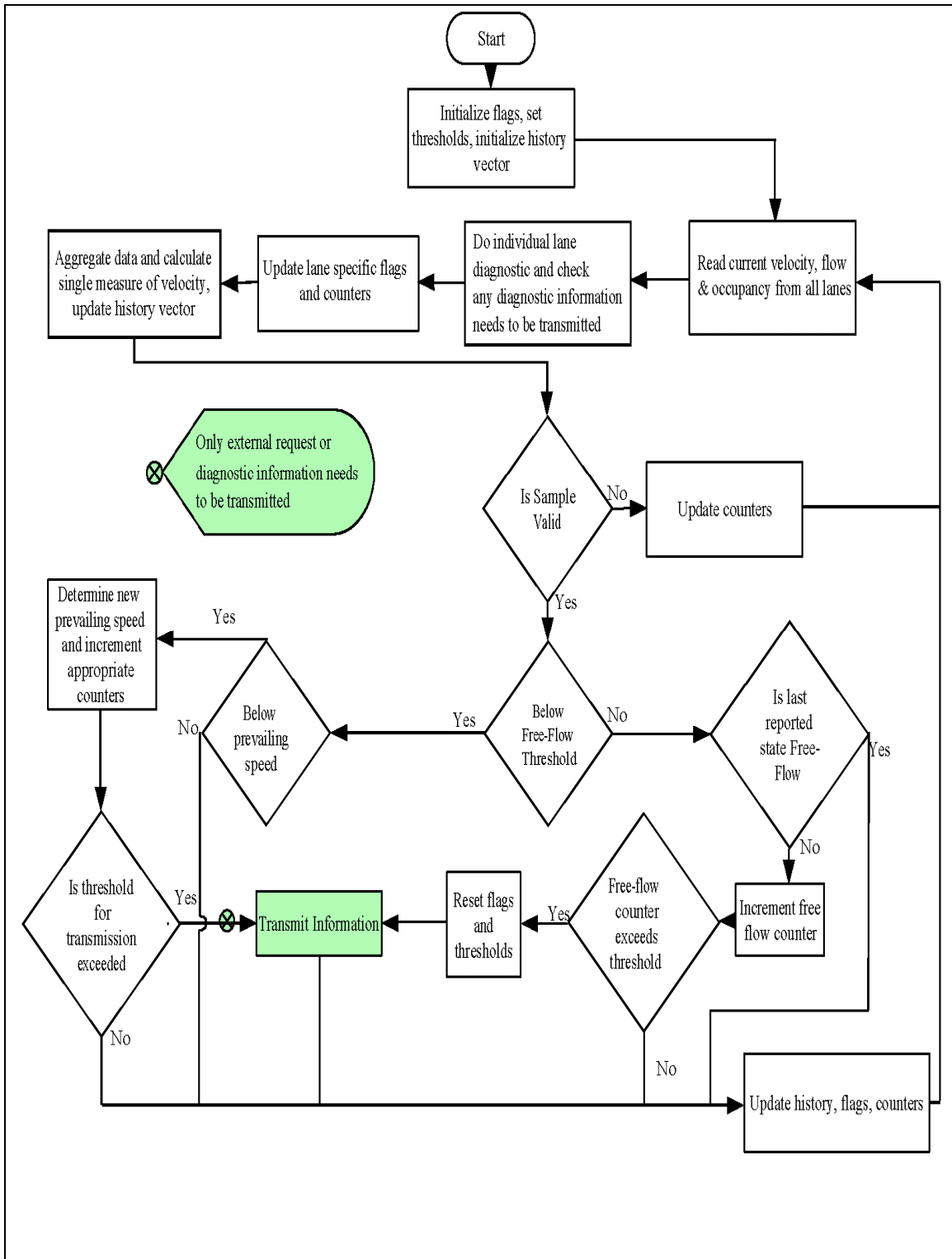


Figure 4.2, Flow chart showing the pre-filtering logic for Mode 2 where data are transmitted when traffic speed changes between multiple thresholds and again when it return to free-flow.

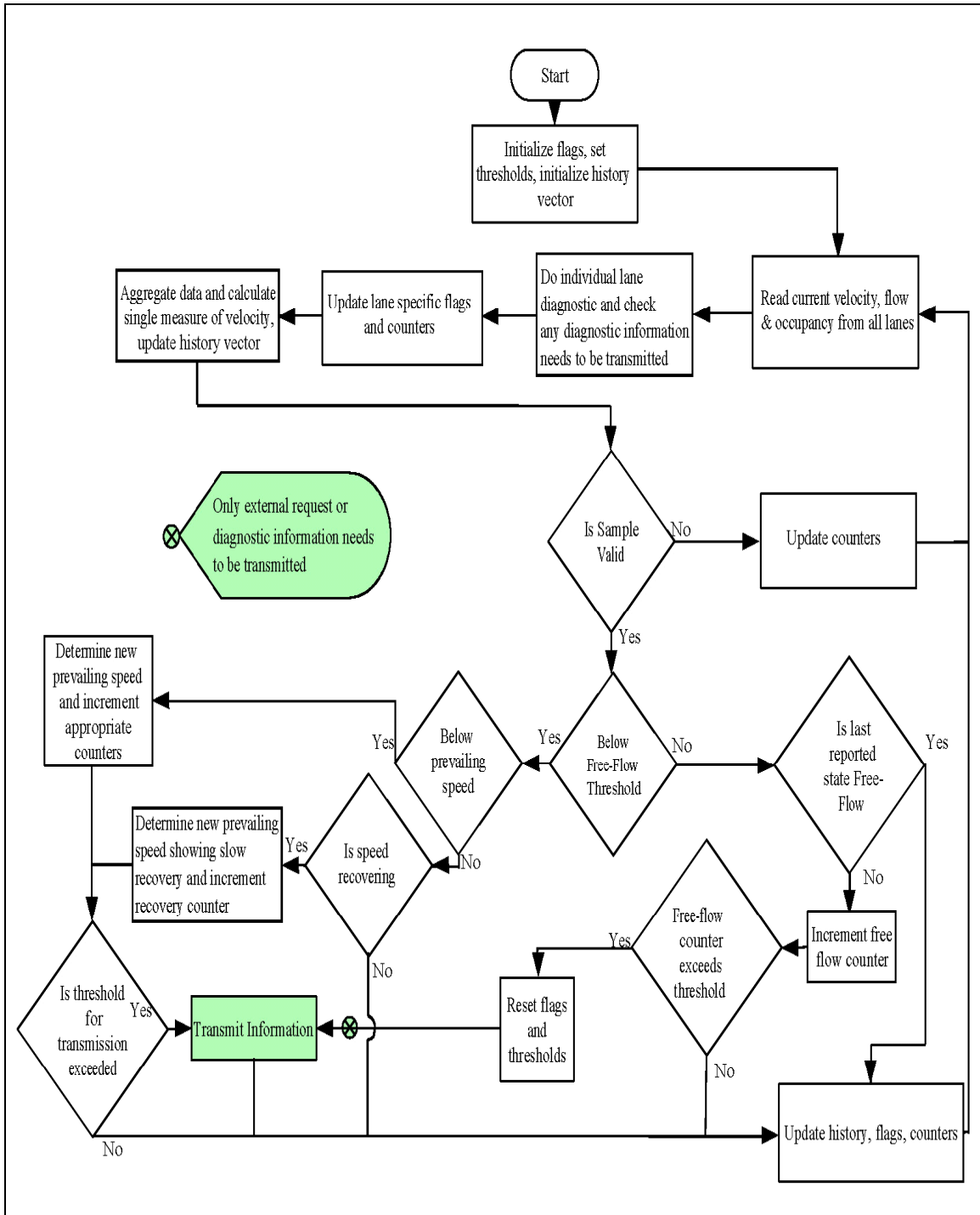


Figure 4.3, Flow chart showing the pre-filtering logic for Mode 3 where data is transmitted when the traffic speeds drop below several thresholds and also to report slow recovery by tracking any trends of increasing speeds.

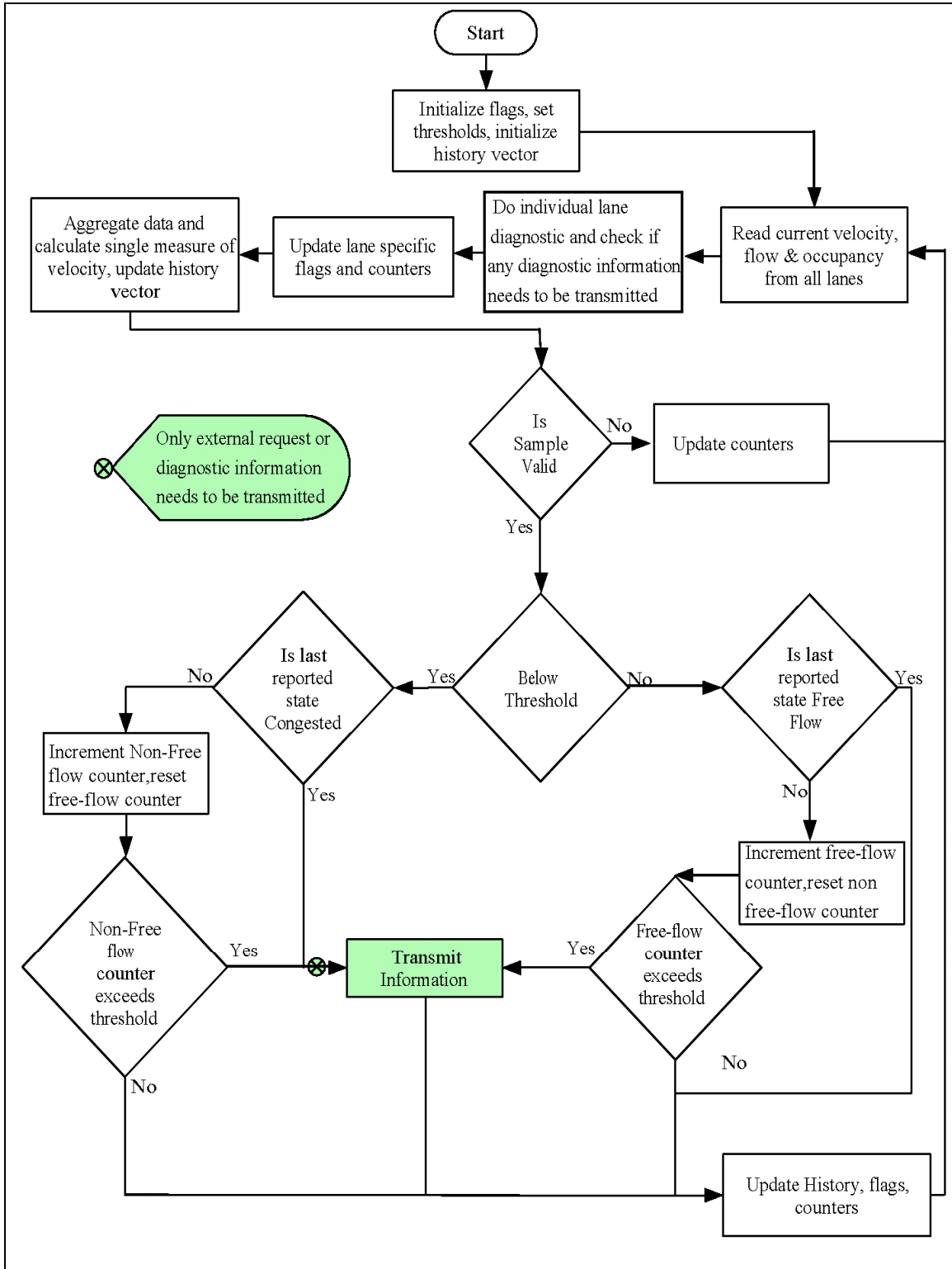


Figure 4.4, Flow chart showing the pre-filtering logic for Mode 5 where data are transmitted at the highest frequency when traffic is congested.

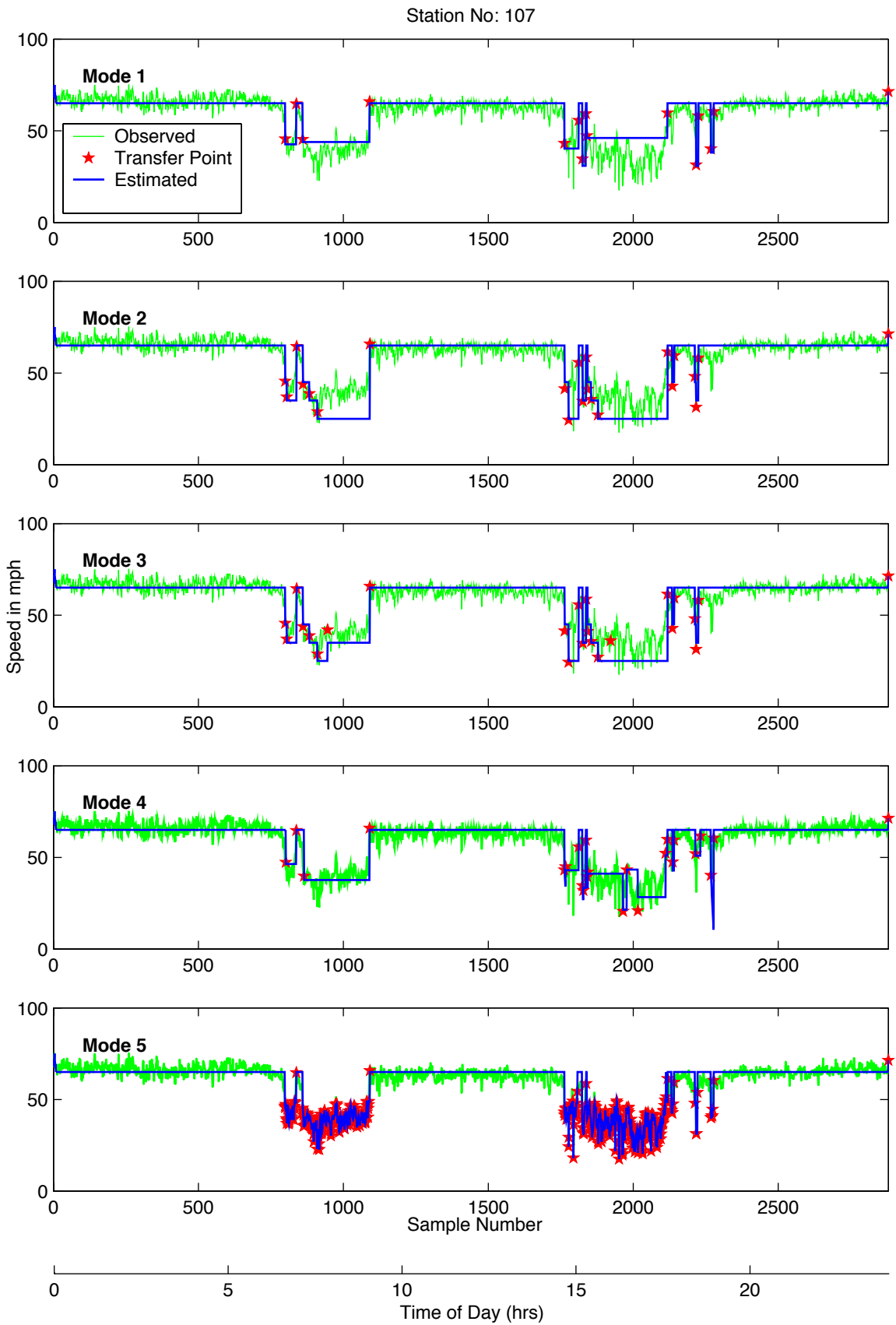


Figure 4.5, Time series of clean speeds, transfer points and synthetic speed series at Station 107 Northbound on I-71 for a randomly selected weekday. Plots A-E is for mode's 1-5 respectively.

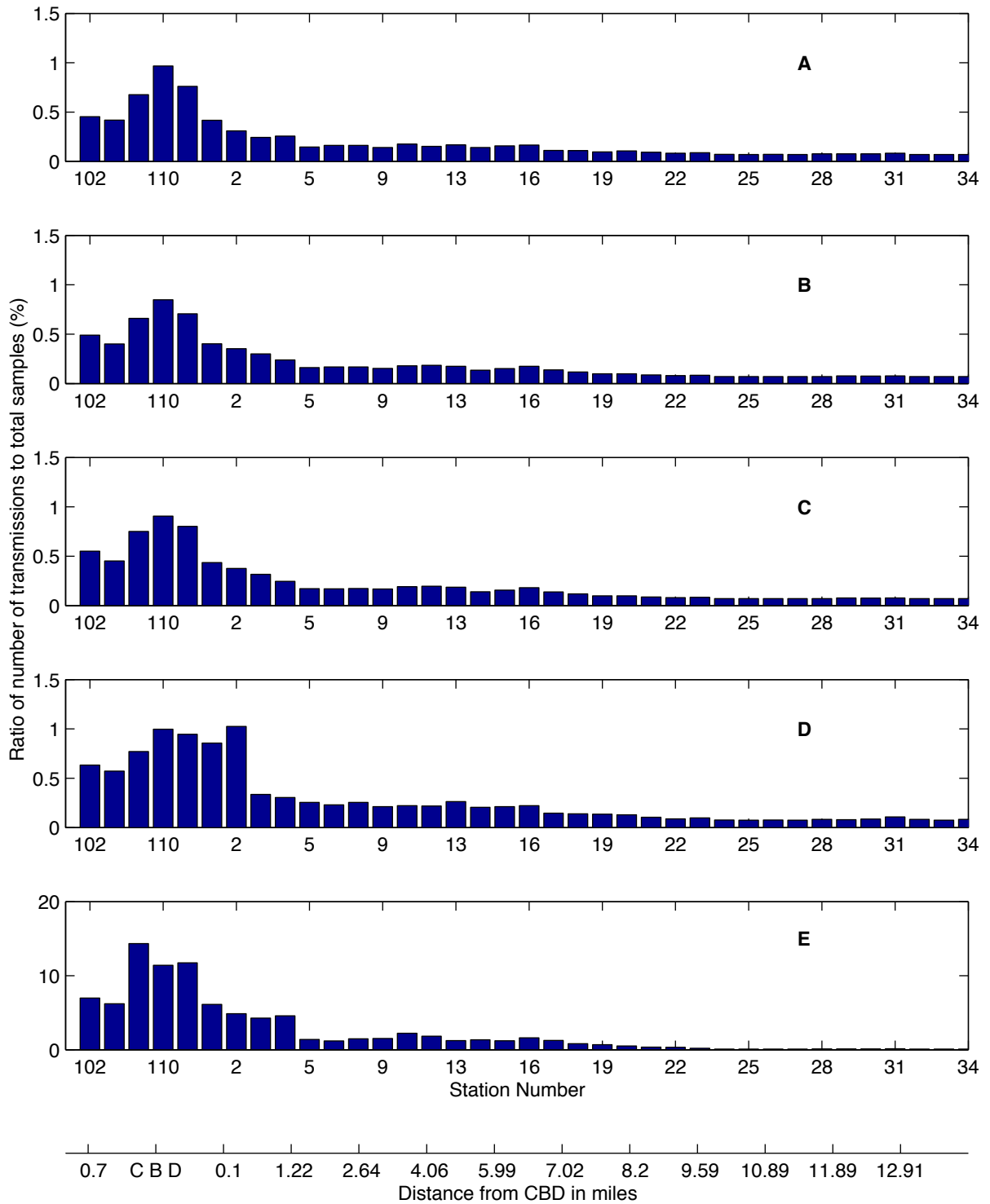


Figure 4.6, Estimated average number of calls made by station, expressed as a percentage of samples in the field on I-70/71 corridor at Columbus, OH in February 2002. A-E shows percentage of transmissions for Mode 1-5 respectively. Notice the different vertical scale on plot E for Mode 5.

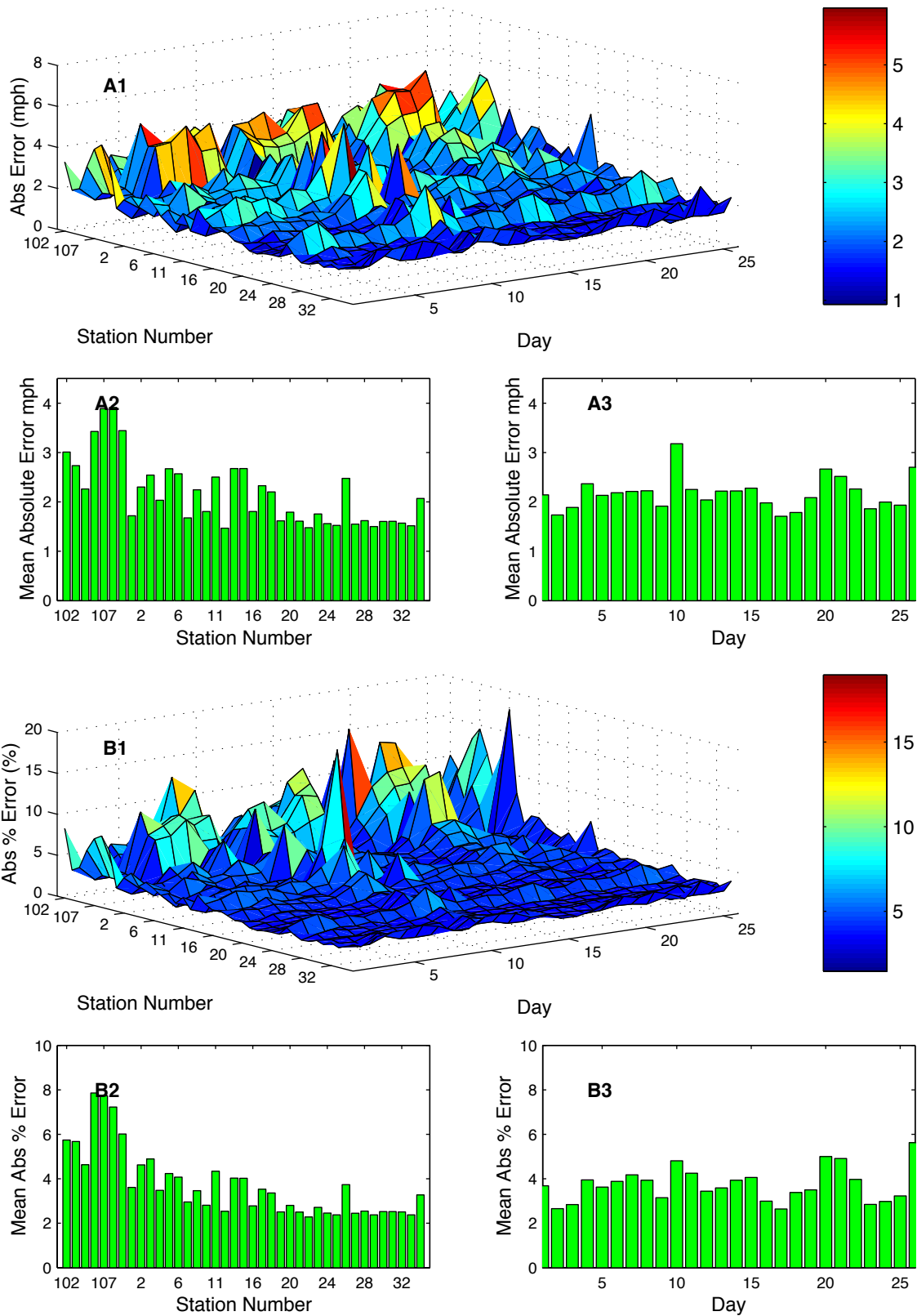


Figure 4.7, Average error distributions of synthetic time series speed as compared to the clean speeds for Mode 1. A1 shows the surface plot for daily mean absolute error at each station for a month's data collected from 40 stations on I-70/71 northbound during February 2002. A2 shows the mean absolute error at each station for the same data averaged across days and A3 shows the daily mean absolute error averaged across 40 detector stations. B1, B2 and B3 show the mean percentage error similar to A1, A2 and A3 respectively.

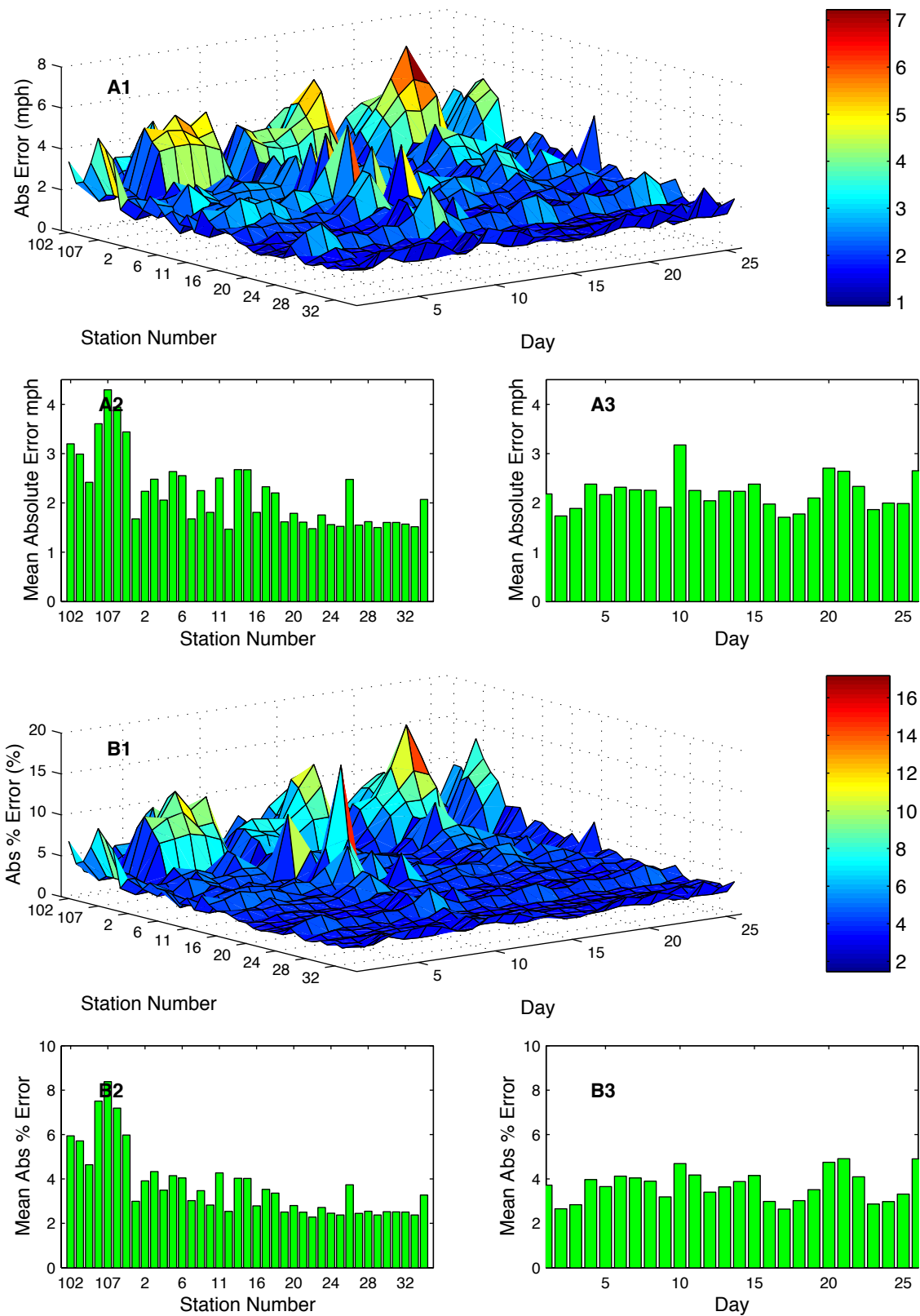


Figure 4.8, Average error distributions of synthetic time series speed as compared to the clean speeds for Mode 2. A1 shows the surface plot for daily mean absolute error at each station for a month's data collected from 40 stations on I-70/71 northbound during February 2002. A2 shows the mean absolute error at each station for the same data averaged across days and A3 shows the daily mean absolute error averaged across 40 detector stations. B1, B2 and B3 show the mean percentage error similar to A1, A2 and A3 respectively.

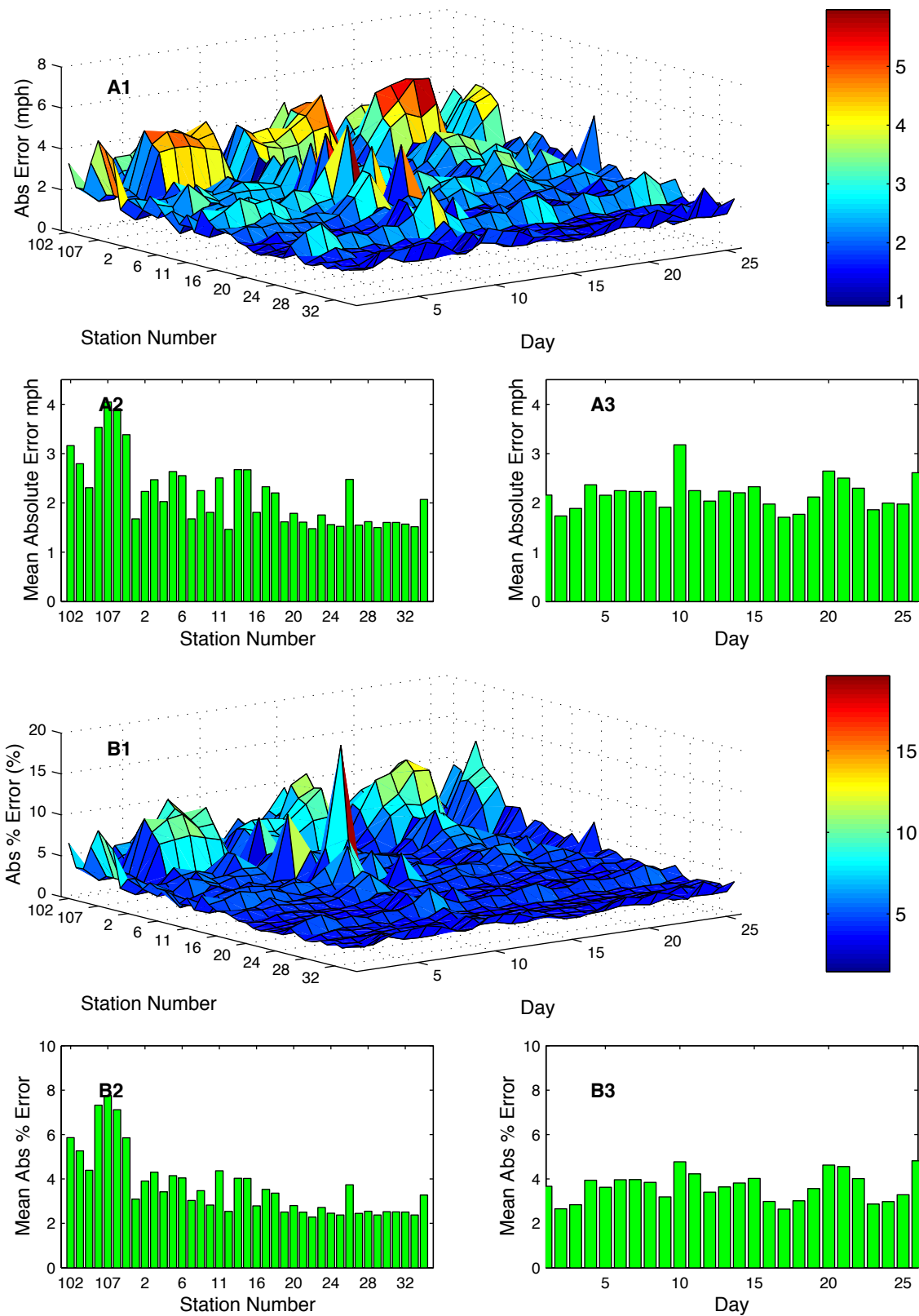


Figure 4.9, Average error distributions of synthetic time series speed as compared to the clean speeds for Mode 3. A1 shows the surface plot for daily mean absolute error at each station for a month's data collected from 40 stations on I-70/71 northbound during February 2002. A2 shows the mean absolute error at each station for the same data averaged across days and A3 shows the daily mean absolute error averaged across 40 detector stations. B1, B2 and B3 show the mean percentage error similar to A1, A2 and A3 respectively.

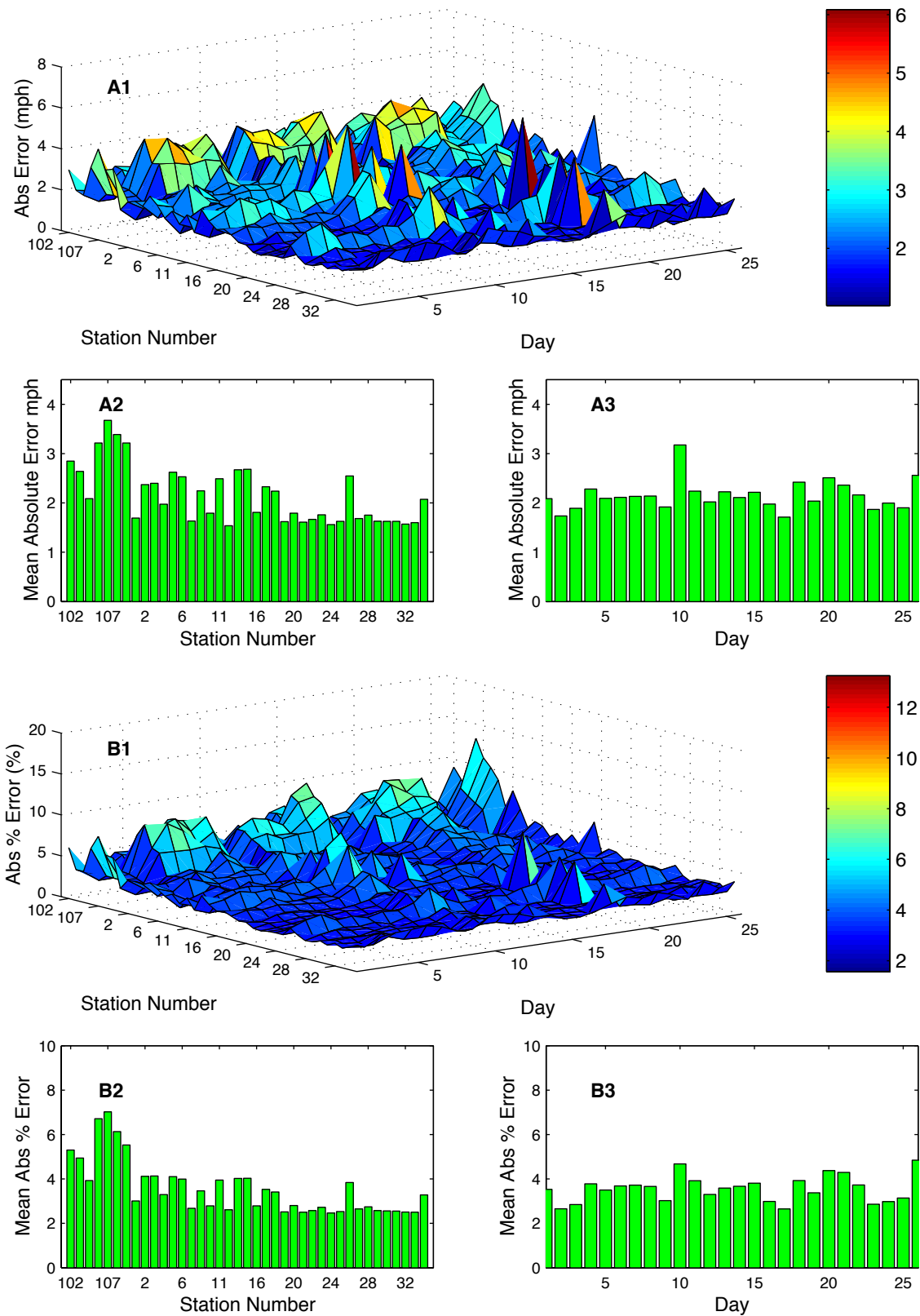


Figure 4.10, Average error distributions of synthetic time series speed as compared to the clean speeds for Mode 4. A1 shows the surface plot for daily mean absolute error at each station for a month's data collected from 40 stations on I-70/71 northbound during February 2002. A2 shows the mean absolute error at each station for the same data averaged across days and A3 shows the daily mean absolute error averaged across 40 detector stations. B1, B2 and B3 show the mean percentage error similar to A1, A2 and A3 respectively.

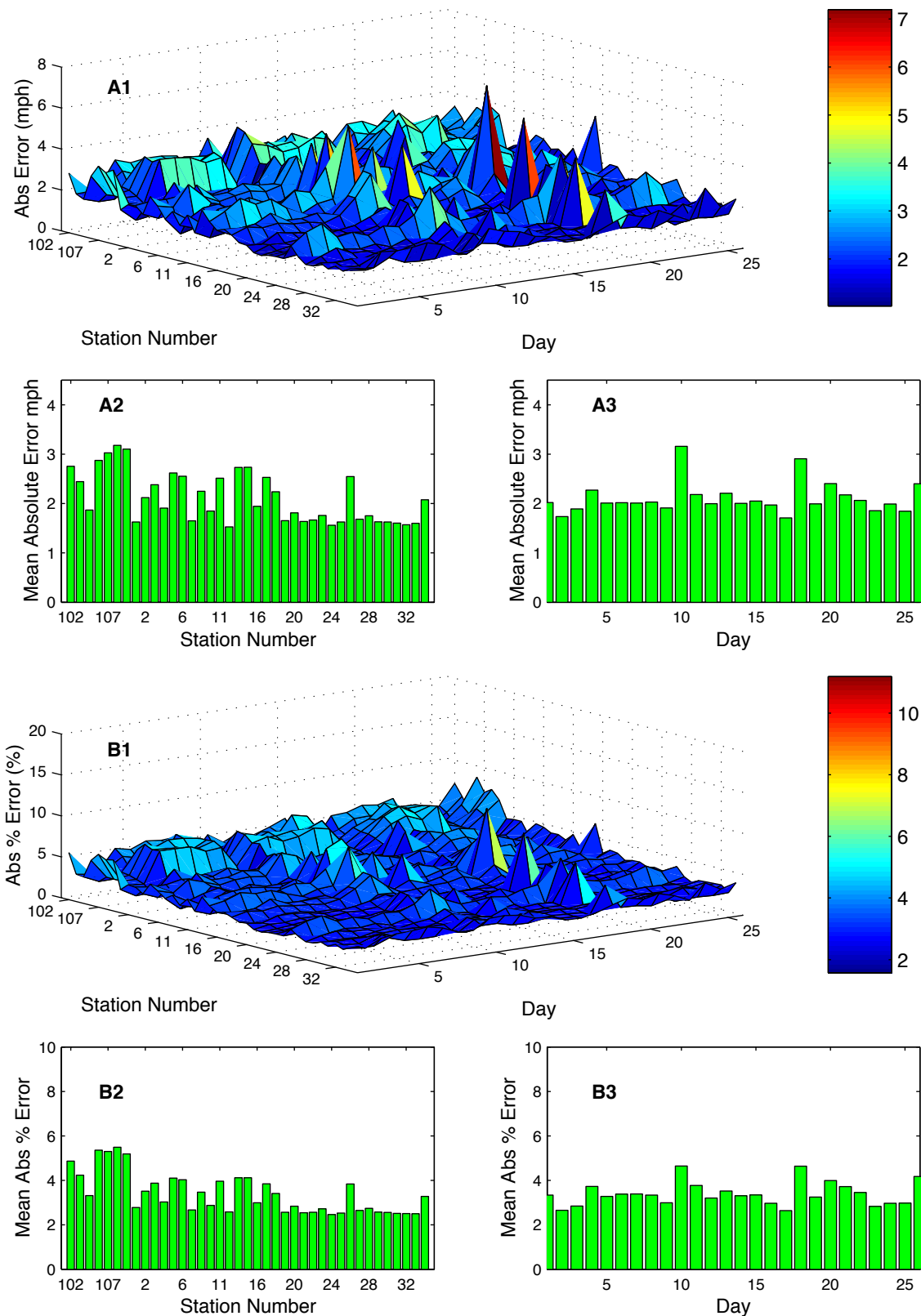


Figure 4.11, Average error distributions of synthetic time series speed as compared to the clean speeds for Mode 5. A1 shows the surface plot for daily mean absolute error at each station for a month's data collected from 40 stations on I-70/71 northbound during February 2002. A2 shows the mean absolute error at each station for the same data averaged across days and A3 shows the daily mean absolute error averaged across 40 detector stations. B1, B2 and B3 show the mean percentage error similar to A1, A2 and A3 respectively.

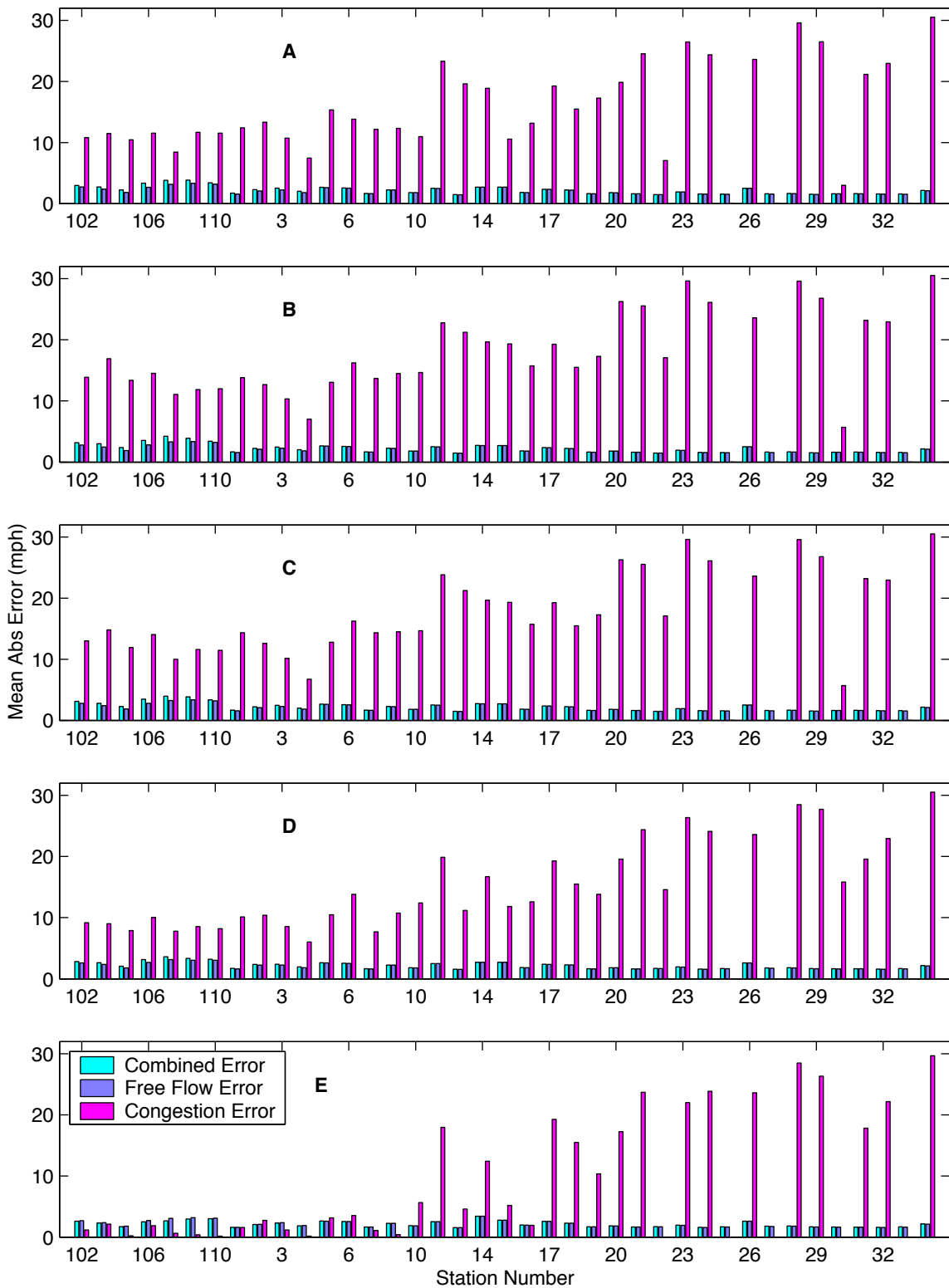


Figure 4.12, Plot showing combined mean absolute error, free-flow periods mean absolute error and congested period's mean absolute error at 40 stations on I-70/71 northbound for February 2002 data. A-E shows errors for Mode 1-5 respectively.

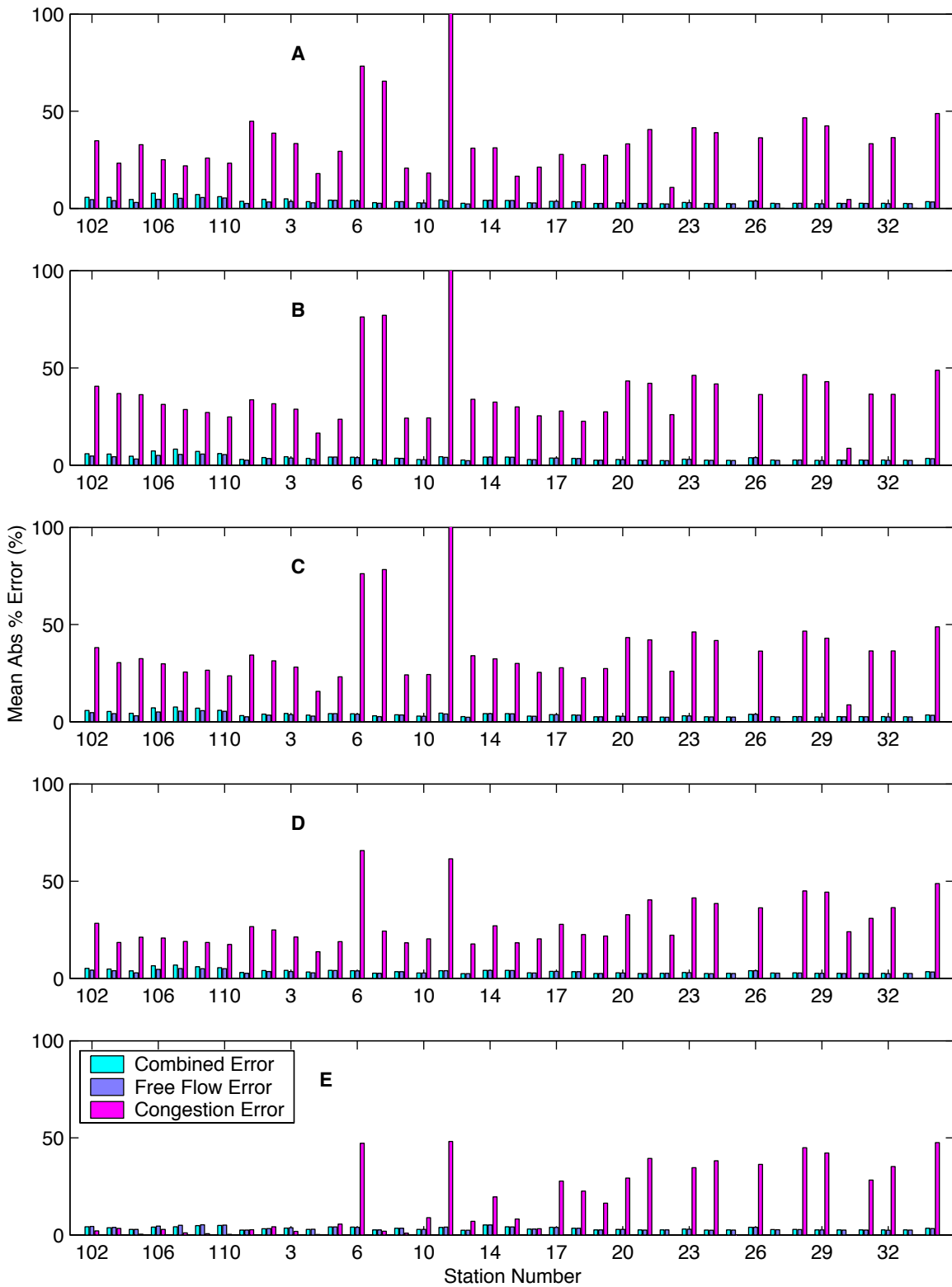


Figure 4.13, Plot showing combined mean absolute percentage error, free-flow periods mean absolute percentage error and congested period's mean absolute percentage error at 40 stations on I-70/71 Northbound for February 2002 data. A-E shows errors for Mode 1-5 respectively.

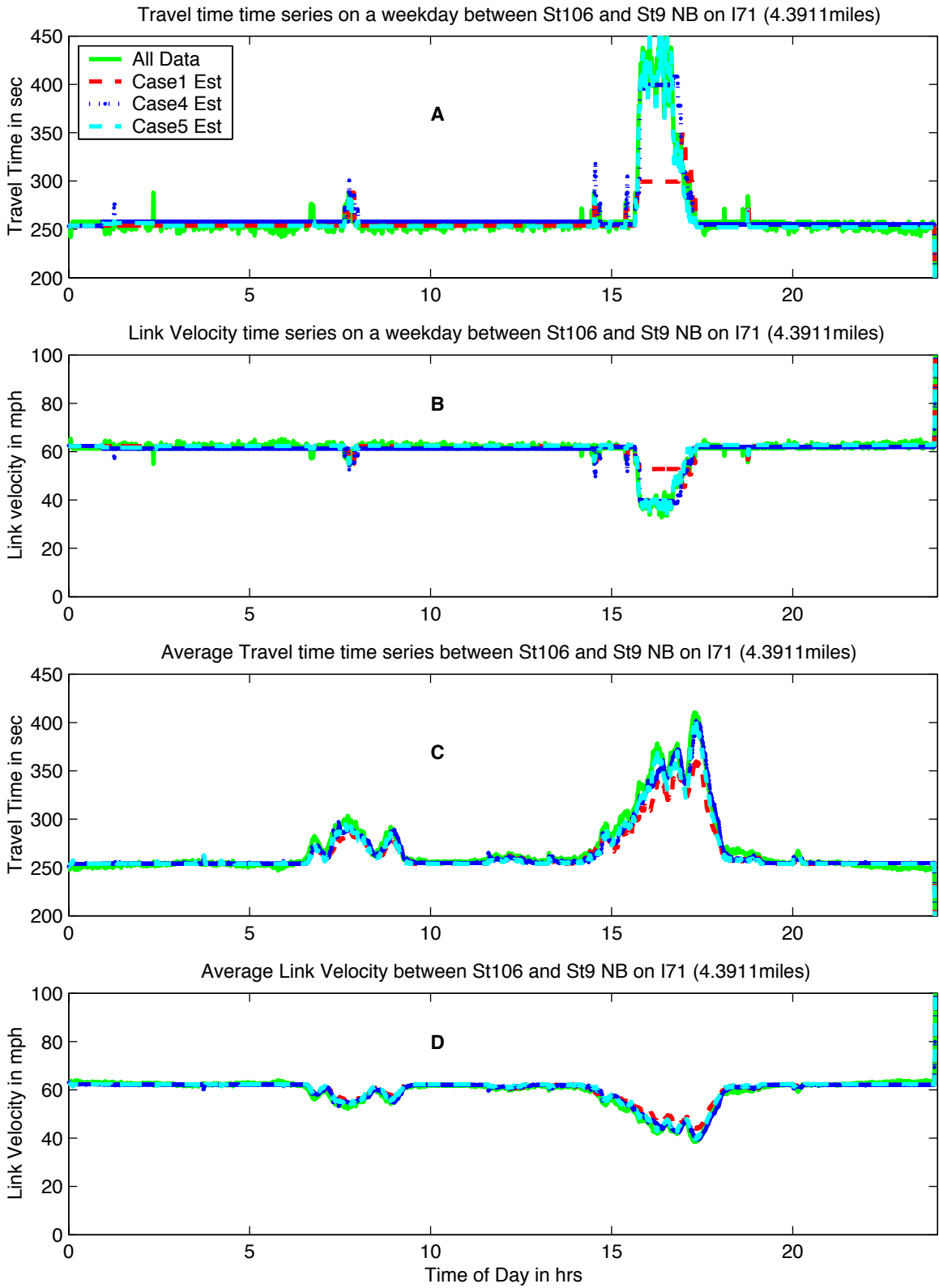


Figure 4.14, Travel time and link velocity estimations using conventional method (all data), Mode 1 data, Mode 4 data and Mode 5 data. Plots A and C compare travel time estimates for one weekday's data and average across one month's data respectively. Plots B and D compare link velocity estimates for one weekday's data and average across one month's data respectively.

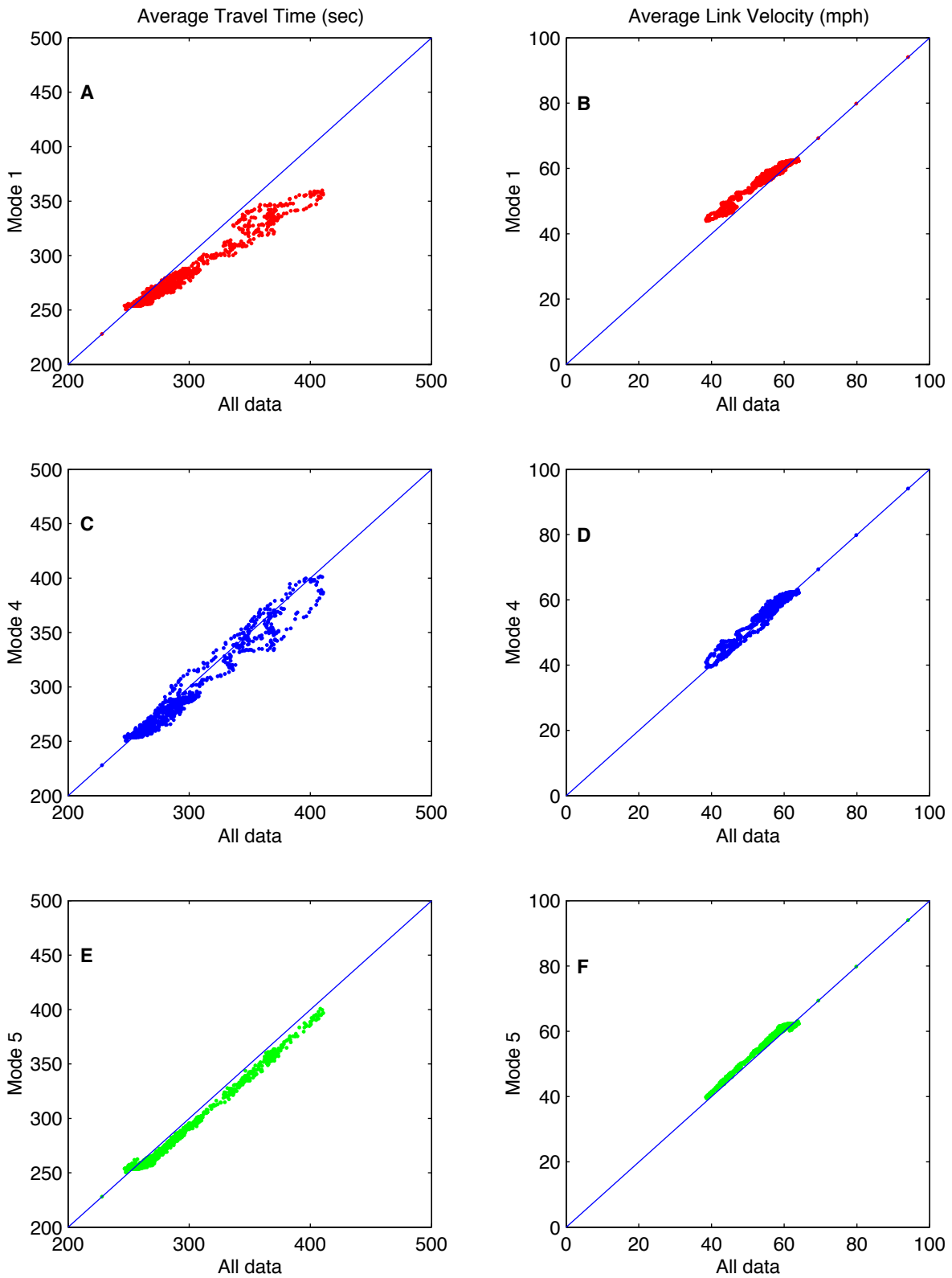


Figure 4.15, Plots showing average travel time and link velocity estimates obtained from Mode 1, 4 and 5 data as compared to the estimates obtained using all data. The first column compares travel time estimates and the second column compares link velocity estimates. A, B are for Mode 1; C, D are for Mode 4; E, F are for Mode 5.

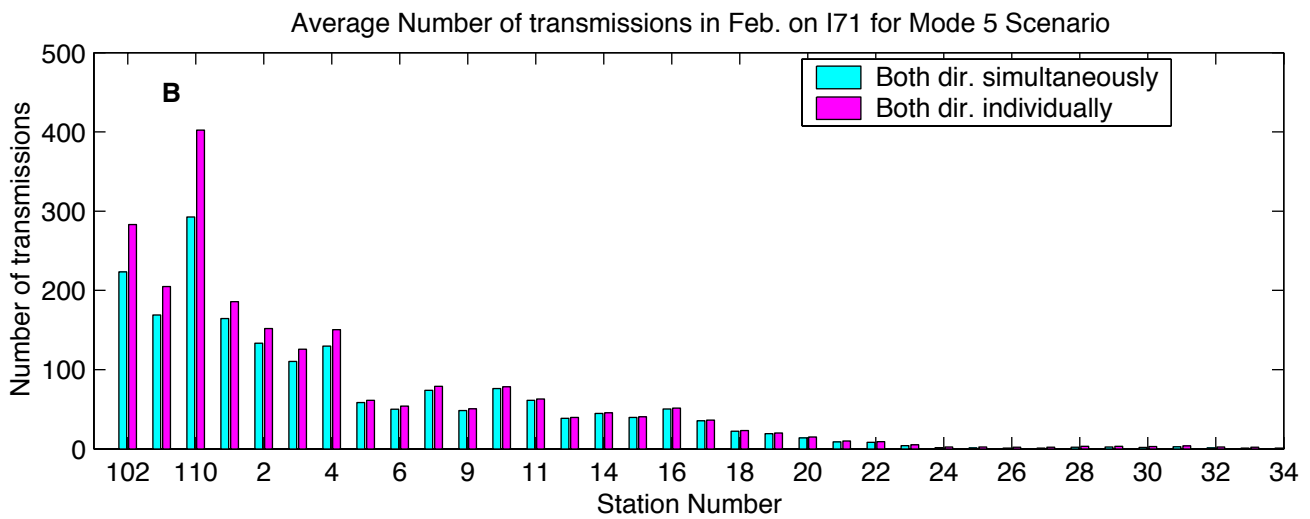
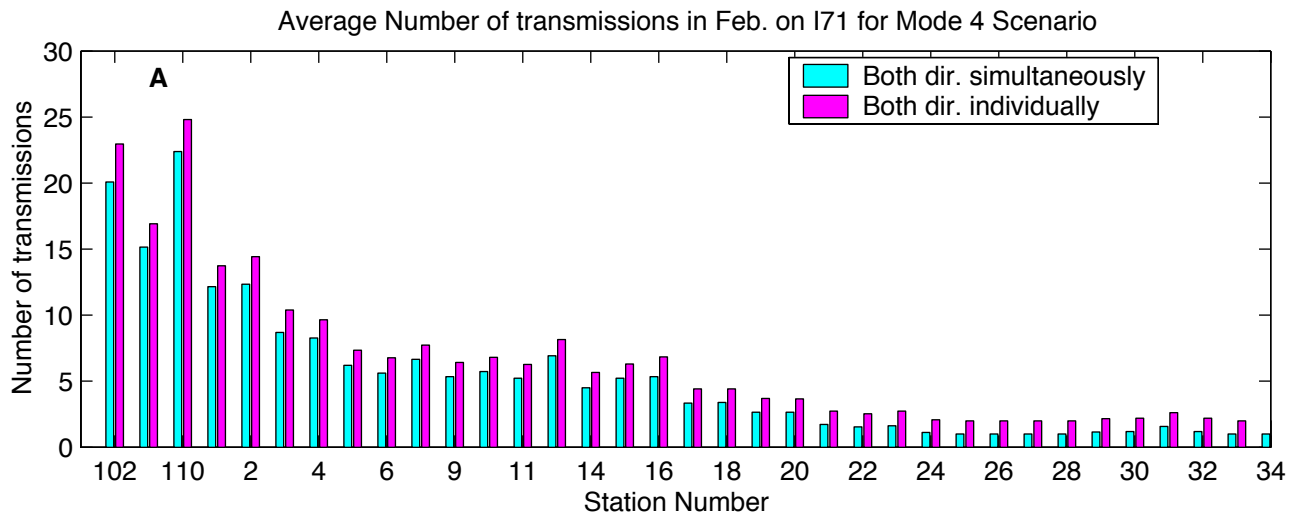


Figure 4.16, Average daily number of transmissions on I-70/71 corridor for the two different algorithms – Both directions simultaneously and individual directions summed up. (A) compares number of transmissions for Mode 4 and (B) compare that for Mode 5.

5 COMMUNICATION STRUCTURE

The decentralized surveillance structure developed in this work assumes that there exists a communication technology where the cost of communication is based on usage. The recent emergence of wireless communication technology provides for such an avenue. Section 5.1 summarizes the properties of some of the popular wireless communication technologies currently available and the criteria for choosing a suitable technology for distributed surveillance. For each technology, the cost structure of the dominant service provider in the United States is presented. The desirable properties of communication protocol for distributed surveillance structure are presented with an example. We then develop the cost equation for distributed surveillance and apply the cost structure of prominent service providers to a number of freeway segments to estimate the average cost under distributed surveillance.

5.1 Wireless Communication

A number of wireless technologies are currently available, as summarized below, and the selection of a particular technology will impact the viability of the communication network. This selection must be based on clear understanding of the costs and other performance parameters of the various technologies available. With this goal, some of the key requirements for judging the effectiveness of different mobile communication technologies are presented in Section 5.1.2. The cost structure of the popular carriers of each technology is summarized in Section 5.1.3. This brief summary would have practical value for managers in identifying the most cost-effective and productive technology for data transmission.

5.1.1 *Wireless Technologies*

This section briefly summarizes the leading wireless data transmission technologies based on [30]. First, General Packet Radio Service (GPRS) is a packet-switched service that allows data to be sent and received over the existing Global System for Mobile (GSM) communications network. The GSM network, upon which GPRS is built, has more coverage than any other cellular technology. In GPRS, data is broken down into small pieces and when the data is ready to be sent, the network assigns time slots on a channel for the transmission. The cost is based on the amount of data sent and received, not on the length of the user's session. The user, therefore, is able to stay connected all day, and is charged fairly for their use of the network. Currently, the leading carriers that use GPRS/GSM technology in the United States are AT&T and T-Mobile

Next, Mobitex is an industry leader for wireless data communications, with networks operating in 22 countries. It is a packet-switched radio technology that provides always on instant two-way messaging and data delivery. It is a narrowband dedicated data-only network that is based on an open international standard. The open standard allows the technology to be customized for the needs of the application being served. Currently, the leading carrier of Mobitex in the United States is Cingular Wireless.

Cellular Digital Packet Data (CDPD) is among the earliest wireless IP packet data standards and is based on one of the most widely accepted networking protocols. CDPD's always on connection to the Internet gives the user instant access to their data. The service pricing plans are usually based on data transferred or as a flat rate rather than the time connected, so the costs are

reasonable. However, the CDPD network is currently being phased out in the United States in favor of the digital systems mentioned above. Since CDPD will be discontinued, it will be excluded from further analysis in this research. A comparison between Mobitex and GPRS/GSM technology for decentralized structure's purpose is presented in Appendix C.

5.1.2 Criteria for selecting wireless technology

Transmission Speeds: Transmission speeds and the throughput of the network (the amount of data that reaches the recipient in a given period of time) are critical to determine the efficacy of time sensitive wireless surveillance structure. Some mobile technologies employ store-and-forward network design where messages are not transmitted immediately but are held in a queue for few seconds to several minutes [31]. Some systems do not adequately support data transmissions and are better equipped for relay of voice communications. Such systems might give preference to voice communication over data and this factor should be accounted for before selecting the technology for data communication.

Reliability: A wireless system that cannot guarantee reliability and consistency in delivering messages will be of limited use for any time sensitive and critical application. Packetizing data into smaller chunks, instead of continuous stream, i.e., circuit switching, makes it easier for equipment to check for transmission errors and prompt retransmissions when some data is missed or corrupted [31]. However a system that requires multiple sends to deliver a single complete message, no matter how fast the transmission speed, will prove to be an inefficient carrier.

Coverage Area: Many popular mobile technologies serve only metropolitan areas and, as noted earlier, the decentralized surveillance structure provides the best opportunity for cost optimization in rural and remote areas. The less frequently traveled stretches of highway often have no available wireless coverage and the requirement of statewide coverage is important for some traffic monitoring applications. The technologies that offer seamless coverage throughout the state are beneficial both for cost optimization and future expansion of surveillance area.

Simplex or Duplex: Simplex services are those that support only one-way transmissions, e.g., radio pagers. Duplex services, such as cellular phones, are able to carry full two-way communication. Not all wireless technologies are duplex because two-way transmissions demand greater bandwidth than one-way transmission, and bandwidth is always in short supply. Also, not every situation demands two-way communication. The operators should identify the present and future need of surveillance structure before deciding upon selecting simplex or duplex technology as the cost and performance of the technology is likely to be different for the two systems.

Costs: The cost of equipment and airtime charges is critical in selecting a technology. The growing competition among service providers is causing the airtime charges to decrease. If the cost of communication by a competitive service provider decreases or the technology evolves, then the cost to replace or upgrade the equipment should not deter the operators from switching to a better and economic option. Likewise, to the extent possible, any surveillance should be made as independent of the specific communication hardware and service provider as possible, e.g., reducing the impact of CDPD discontinuation.

5.1.3 Cost Structure

In this section we summarize the cost structure (tariffs and data plans) for two wireless carriers of Mobitex and GPRS technology. The tariff and data plan information was obtained from the website of the carriers and will undoubtedly change during and after the completion of this work. The tariffs and plans used in this work are for individual users. For large-scale deployment of this communication structure on multiple stations, the agencies might be offered a different cost structure.

Mobitex: Cingular wireless is the leading carrier of Mobitex technology in the United States. Cingular offers six types of data plans as summarized in Table 5.1. Each plan has a fixed monthly cost and included monthly data transfer threshold. When the amount of data transferred exceeds the included threshold, there is a variable cost per additional kilobyte of data transferred [32].

GPRS/GSM: AT&T wireless is the leading carrier that uses GPRS/GSM technology for data transmission. It has several data plans that may or may not be coupled with a voice plan. The Table 5.2 summarizes the data plans and tariffs offered by AT&T wireless [33]. As noted earlier, the data plans offered by the carriers change rapidly with time as the competition increases, technology evolves or federal guidelines change.

5.2 Communication Protocol

As can be observed from Tables 5.1 and 5.2, the included data bandwidth for most data plans is limited and in order to optimize the cost, the size of each transmission has to be minimized. In general the carriers round off the message size to the next higher integer kilobyte unit for every transmission. Also the entire kilobyte is usually not available for data alone as part of the space is lost in encoding or adding the header information to the message. The transmission cost per message is minimized if the entire message packet is kept within one-kilobyte size. In this section, we present a communication protocol that aims at minimizing the message size as well as maximizing the amount of information that can be transmitted within the limited data size. The protocol also ensures that certain reliability checks, as described below, are included in the message so that the TMC can identify whenever a message is lost or delayed.

In this protocol, following pieces of information are transmitted with every message:

- **Station Id:** It is a unique identifier for each station in the field
- **Sample number:** The sample number provides a time stamp to the message transmitted. If the sampling rate were 30 seconds then there would be 2880 samples collected every day. The sample numbers are reset at the beginning of each day (00:00 a.m.) and are counted in the intervals of sampling rate. Thus, a sample number 500 with sampling rate of 30 second would be collected at time 06:10:00 a.m.
- **Transmission number:** The transmission number serves as a reliability check for the message. It is the number of transmission from the field since the beginning of the day. If the TMC finds that the transmission number of a message is greater than the number of messages received on that day from a station, it would indicate that the messages are being lost or delayed. The TMC could take appropriate action, e.g., request missing information. At the end of the day, it is envisioned that the field sends the total number of

transmissions for the day along with the summary data for the day, which can help the TMC evaluate the reliability of the communication structure.

- **Direction:** A field station may monitor traffic from more than one directional highway, e.g., opposing directions of traffic or adjacent facilities. The message sent from the field might contain information from a single directional highway or a set of directional highways. A flag is included in the message that provides information about the traffic direction and/or highway, e.g., a flag might have the following values: 0-All directions, 1-East/North, 2-West/South, and 3-Set I etc. These values are mapped from a configuration file that is present both at the TMC and the field station.
- **Status flag:** It defines the type of information being transmitted from the field station. The information type is the function of the current traffic state at the station and we define following types in this protocol:
 - Drop: The status type 'D' is transmitted when the speeds at the station are dropping and the rate of change of speed is negative. The status type might be followed by the rate at which the speeds are dropping which could be used by the TMC to predict speeds in future samples.
 - Stable: The status type 'S' is transmitted when the speeds at the station are stable.
 - Recovery: The status type "R" is transmitted when the speeds at the station are recovering. The rate of recovery is also transmitted with the type and the status can be interpreted as "The speeds are recovering at z mph".
 - Above Threshold: When the speeds return back to free-flow conditions, status type 'A' is transmitted indicating that the speeds are above free flow threshold.
 - Lane data: When diagnostic information about a particular lane or a set of lanes is transmitted, the status type is set at 'L'. A message with status type 'L' could be sent when information like lane's p-q are blocked or are showing reduced speeds needs to be transmitted.
 - Summary Data: The field stations transmit summary data either at fixed low frequency time intervals or at the end of the day. The summary data is potentially the most important and highest priority message from the field. It contains information about the number of transmissions for the day along with other summary information like AADT, profile of speed, flow and occupancy for the day, number of transitions between Free Flow (FF)/Non Free Flow (NFF) conditions, etc. The summary information is used by the TMC to diagnose the performance of detector station and archive the daily observations for planning applications. A status flag 'Z' indicates that summary data is being transmitted. Since this transmission is important and high priority, the communication protocol should have the mechanism to verify that the transmission was successful and without errors. One such reliability check could be that the field station continues to transmit summary information until a confirmation from the TMC is received. This kind of reliability check can also be applied to other important and high priority messages.
 - Miscellaneous: The status flag 'M' is transmitted when information other than listed above is transmitted from the field station. Information like field initiated

maintenance request or a request from TMC to transmit recent historical data are miscellaneous type.

Apart from the five pieces of information listed above that are sent with every transmission from the field, the following type of optional information can be sent in a transmission:

- Current speed, average flow/occupancy across lanes for last few samples
- Flag showing which communication scenario being used
- Flag showing whether mainline or ramp information being transmitted

If during a transmission there is available data bandwidth in the message, then additional information could be transmitted to use up the available space. The additional information could include a summary of the last few transmissions; a summary of individual lane flow; or a profile of recent flow, speed and occupancy. The various flags used in this protocol can be mapped to a configuration file that should be set while deploying the algorithms in the field and at the TMC.

For reduced storage and transmission requirements, lossless data compression techniques, e.g., Probabilistic Coding, Shannon-Fano Algorithm [34], can be employed. In lossless compression, the output exactly matches with the input after a compression/expansion cycle and a decompressor can be installed at the TMC. Space can also be saved by efficiently reporting the measurements, flow value can be transmitted as integer vehicles per 30 sec (rather than hourly count), occupancy less than 10 percent can be rounded off to single decimal place and that above 10 percent can be rounded off to nearest integer value. Similarly the value of speed can be rounded off to integer; summary flow can be sent in the units of vehicle per hour and average profile can be time segmented and transmitted as set of discrete point values in time. A few sample data transmission messages for this communication protocol are shown in Figure 5.1.

If the communication structure supports a duplex system, then the TMC can contact the field station and request or send information from or to the field station. When the TMC contacts the field, the message should contain the information about station number, traffic direction and message type. The message type could be either 'request' or 'inform' as describe below.

1. Request: The message type flag is set as 'q' for requests from the TMC. The request message should include the kind of data required (flow, occupancy, speed), duration of data required (e.g., from sample number 'x' to 'y'), format of data required (e.g., 30 second or 5 minute aggregate, individual lanes or average across lanes). The request can also be made to resend transmissions number 'x' to 'y' if some data is lost during transmission from the field.
2. Inform: The message type flag is set as 'f' when the TMC wants to inform the field station. This type of message is sent when the TMC wants a field station to ignore particular lane's data, or ignore particular period's data. This type of message could also be sent when the TMC wants the field station to stop sending data for specific duration, e.g., during the weekends or holidays when the TMC is not staffed, or after detecting an incident.

5.3 Cost Evaluation

In this section we develop an equation for computing recurring cost for wireless communication. The cost equation can be broken down into two components, as described below, fixed periodic cost and additional overage cost. The fixed cost component has to be paid periodically even if the service is not utilized and it includes certain amount of allowed data transfer. Once the usage exceeds included data transfer threshold, additional overage cost applies. The additional cost value depends upon the amount of exceeded data transfer needed and the carriers usually charge it per additional kilobyte of data transferred. Both the fixed cost and additional cost rate vary according to the data plan and the service provider, e.g., the summary in Section 5.1.3 for two technologies and service providers. The recurring cost is generally calculated on a monthly basis. Equation 5.1 shows the two components of the monthly recurring cost equation.

$$\text{Recurring monthly cost for a station} = C_F + \text{Max} [0, D - D_P] * C_k \quad (5.1)$$

Where

- C_F = Fixed monthly cost (\$)
- D = Monthly data transfer needed (KB)
- D_P = Data transfer allowed by plan (KB)
- C_k = Cost per kilobyte of additional data transferred (\$/KB)

The monthly data transferred, as used in Equation 5.1, is the function of monthly number of transmissions required by a detector station for all the directional highways monitored. We analyze two scenarios in our work, i.e., risky scenario, where mean number of transmissions averaged over the study period are used to calculate the monthly data transfer and conservative scenario, where maximum number of transmissions per day is used to calculate the monthly data transfer. Equation 5.2 computes monthly data transfer.

$$D_{\text{mean}} = N_D * [\{\sum T_{\text{Mean}} * S_T + N_{\text{TMC}} * S_{\text{TMC}}\} * O_f * OL_f + S_D * O_f * R_f] \quad (5.2)$$

Where

- D_{mean} = Mean data transfer per month (KB)
- N_D = Number of days
- T_{Mean} = Mean number of daily transmissions per direction
- S_T = Size of each transmission (KB)
- N_{TMC} = Number of calls from TMC
- S_{TMC} = Size of transmission from TMC
- O_f = Overhead factor, the amount of data space used up by carrier for encoding and adding header information to the message. Its value is ≥ 1 .
- OL_f = Overlap factor, accounts for overlap in the need for transmission by each direction. The value of overlap factor can vary from 0.5, for complete overlap, to 1, no overlap. As can be observed from Figure 4.16, this factor is significant for the communication mode where data is transmitted at the highest frequency during congestion and not very significant for other communication modes.
- S_D = Size of summary data transfer (KB)
- R_f = Redundancy factor, accounts for high priority messages that need verification from the TMC about successful delivery and might need to be re-transmitted for redundancy purpose.

In Equation 5.2, the mean numbers of transmissions were calculated using weekday's data as the number of transmissions for weekends were observed to be smaller than that for weekdays on most of the facilities examined. A similar equation can be developed for calculating data transfer need for a conservative scenario where maximum numbers of transmissions are used. In the following Section 5.4 we apply Equations 5.1 and 5.2 to a number of freeway segments and compute the average surveillance communication costs per station. The following assumptions about Equation 5.2 are made for calculating the amount of data transfer:

- The size of a regular transmission, including the overhead, does not exceed 1 KB. This assumption is reasonable if the communication protocol developed in Section 5.2 is followed
- The size of transmission containing summary data, including overhead, is 2 KB
- The redundancy factor is set to a value of 3
- The transmissions resulting due to requests from TMC is 10 percent of the total transmissions made by station
- Number of days in a month = 30
- Overlap factor is assumed to be equal to 1, implying that there is no overlap in need for transmission. This assumption puts the analysis on a conservative side.

Based on the above assumptions, an equation for computing monthly *Mean data transfer* reduces to Equation 5.3.

$$D_{\text{mean}} = 30 * \{ \text{CEIL} (\text{Total}_{\text{Mean}} * 1.10 * 1 * 1) \} + 180 \text{ KB} \quad (5.3)$$

Where

$$\text{Total}_{\text{Mean}} = \text{Sum of mean number of calls for each direction}$$

Similarly the monthly *Maximum data transfer* reduces to Equation 5.4.

$$D_{\text{max}} = 30 * \{ \text{CEIL} (\text{Total}_{\text{Max}} * 1.10 * 1 * 1) \} + 180 \text{ KB} \quad (5.4)$$

Where

$$\text{Total}_{\text{Max}} = \text{Sum of maximum number of calls for each direction}$$

5.4 Analysis

This section applies the cost structure and equations presented in the earlier sections to the five freeway corridors described in Section 1.3 for the various communication modes presented in Section 4.1.1. The detailed analysis of results for I-70/71 corridor is presented in this section and the results for the remaining corridors are presented in Appendix B.

The cost structure was applied to a month of weekday data collected during February 2002 on all the stations that had both north/eastbound and south/westbound detectors working. Out of 40 detector stations on the corridor, 35 stations had data from both traffic directions. Figure 5.2 shows the total data transferred for the month and the cost using GPRS/GSM (AT&T) and Mobitex (Cingular) technology for each station on the I-70/71 corridor. Plots A-E are for communication Modes 1-5 respectively. The left vertical axis of the plot shows the cost per month from Equation 5.1 and the right vertical axis shows the amount of data transferred in MB

from Equation 5.3. Note the larger data transfer vertical scale for plot E, where data is transmitted at the highest frequency during congestion. The cost of communication and the data transfer demand are highest for stations close to the CBD. The cost per station and amount of data transferred for risky and conservative scenarios, described above, are summarized in Table 5.3. Aggregating data to 30-second samples yields 2880 data packets per day and if all of these packets were sent the amount of data transferred per month would be 93 megabytes (based on Equation 5.3). Based on the cost thresholds from Table 5.2, one would choose the unlimited data transfer plan if the amount of data transferred exceeds 37 MB for AT&T and 13.3 MB if the data carrier is Cingular. Thus, under conventional communication scenario, where all the data is transferred, the communication cost would be \$80 and \$60 per month for AT&T and Cingular respectively. At the current costs of wireless communication, all decentralized modes should be cost effective for the outlying stations on this corridor. Only a few of the stations in the CBD would be more cost effective with unlimited service under scenario 5 with Cingular. These modes would allow lower cost surveillance in lower demand areas. By tracking the number of transmissions at the TMC, a station that uses per-transmission cost plan can be migrated to flat rate unlimited bandwidth plan as demand increases or service costs decrease.

Table 5.1, Cost structure for Mobitex technology by Cingular wireless on November 24, 2003.

Fixed monthly charge (\$/month)	Included data transfer per month (Megabytes)	Cost of additional data transfer (\$/Kilobyte)
\$19.99	0.1 MB	0.20 \$/KB
\$27.99	0.2 MB	0.17 \$/KB
\$29.99	0.5 MB	0.15 \$/KB
\$34.99	1.0 MB	0.15 \$/KB
\$49.99	2.9 MB	0.05 \$/KB
\$59.99	Unlimited	N/A

Table 5.2, Cost structure for GPRS/GSM technology by AT&T wireless on November 24, 2003.

Fixed monthly charge (\$/month)	Included data transfer per month (Megabytes)	Cost of additional data transfer (\$/Kilobyte)
\$29.99	10 MB	0.0030 \$/KB
\$39.99	20 MB	0.0020 \$/KB
\$59.99	40 MB	0.0015 \$/KB
\$74.99	60 MB	0.0013 \$/KB
\$79.99	Unlimited	N/A

Table 5.3, Estimated monthly communication costs per station and amount of data transfer for I-70/71 corridor in Columbus, OH for weekday's data collected during February 2002.

Commu-nication Mode	GPRS/GSM (AT&T)		Mobitex (Cingular)		Average data transferred MB*	Maximum data transferred MB*
	Risky	Conservative	Risky	Conservative		
	Average Cost (\$/month)	Maximum Cost (\$/month)	Average Cost (\$/month)	Maximum Cost (\$/month)		
Mode 1	30	30	31	33	0.31892	0.58008
Mode 2	30	30	31	35	0.31975	0.58761
Mode 3	30	30	31	35	0.32813	0.62695
Mode 4	30	30	31	36	0.39007	0.79939
Mode 5	30	32	43	51	1.762	7.4196

* 1 MB = 1024 KB

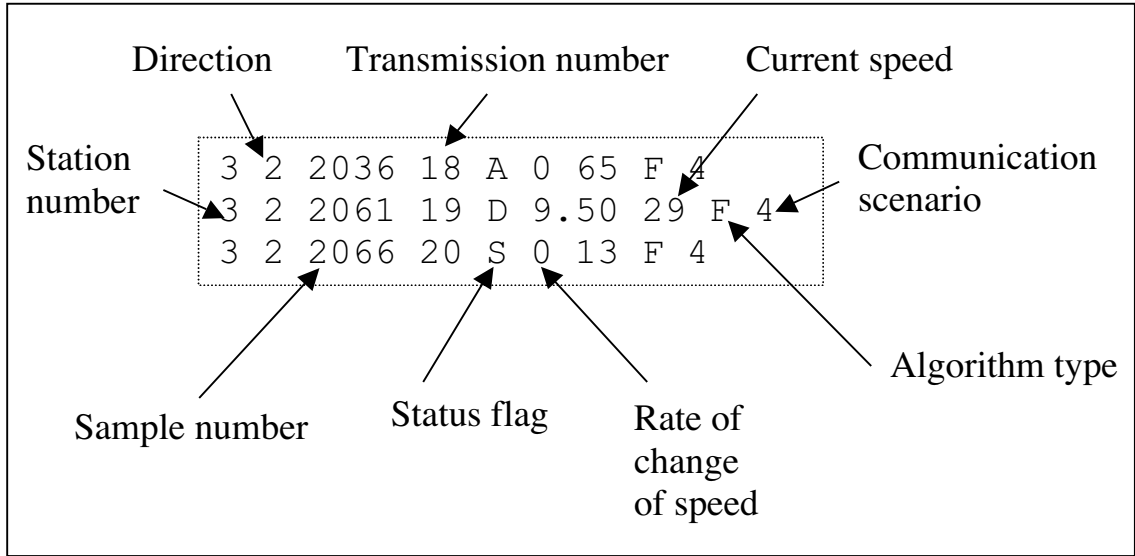


Figure 5.1, Sample of data transmitted from field station for communication scenario in Mode 4.

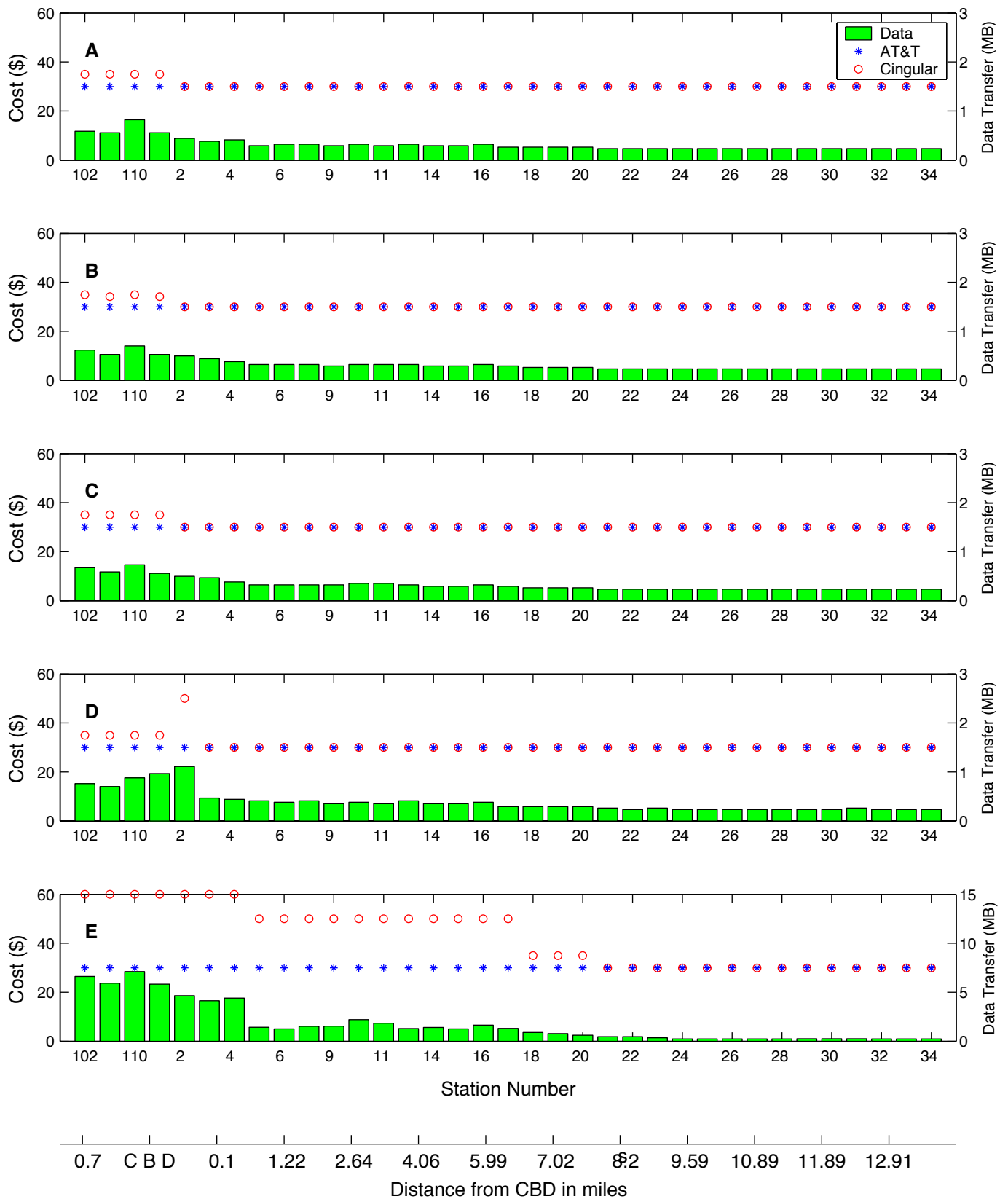


Figure 5.2, Estimated average monthly communication cost structure as applied to a month's weekdays data collected on 35 stations of I-70/71 corridor during February 2002. A-E are for communication Modes 1-5 respectively. The plots show the average amount of data transferred by each station per month, and the per station monthly communication cost when using either GPRS/GSM (AT&T) or Mobitex (Cingular) technology. Note the different data transfer vertical scale for plot E, five times larger than that for plots A-D.

6 SUMMARY, CONCLUSIONS AND FUTURE RESEARCH

Surveillance of many facets of transportation networks requires data about the operational condition and status of the network. Conventionally, the fixed-point surveillance systems are polled at regular intervals from a central location and a given sensor station only responds when polled. These systems require the transmission of all of the data, yet most of the time, the response to the data is simply that no action is necessary. The required communications network and transmission costs can become significant which could limit the surveillance coverage area, especially in remote locations. Decentralized surveillance, control and communication structures have the potential to reduce the combined surveillance and communication costs without sacrificing the performance of existing applications. This cost-effective surveillance structure could potentially allow cost-effective expansion of a surveillance system to cover more remote locations.

This study focused on developing a decentralized surveillance, control and data transmission structures for freeway traffic surveillance; although the concepts can be expanded to other surveillance systems like roadway weather information systems (RWIS). To study the affect of amount of information transmitted on the performance of surveillance application and communication costs distributed surveillance algorithms for five communication modes were developed. The algorithms for these communication modes pre-filter data in the field and differ primarily in the criteria for the pre-filtering decisions. These five communication modes exploit the fact that free-flow speeds are relatively stable and constant. Thus, one does not need to transfer this information, and lack of communication is interpreted as the continuation of the last reported state. Key to this event-driven approach is the recent emergence of new wireless communication technologies where the communication cost is usage based.

The objective of communication Mode 1 is to minimize the number of transmissions and transmit information only when the traffic state changes between free-flow and congestion. Mode 2 goes a step further in providing data improved resolution at a slightly greater communication cost than Mode 1. It transmits information about dropping traffic speeds at multiple thresholds when the facility is becoming congested until the speeds reach a lower bound after which the information is transmitted only when the facility returns to free-flow conditions. In Mode 3, the algorithm is similar to Mode 2, transmitting data when the traffic drops below several thresholds but it also reports recovery, slowly tracking any trends of increasing speeds. The algorithm in Mode 4 combines the features of Modes 1-3 with precise measurements and quicker response by transmitting information whenever the speed changes significantly from the expected or last reported condition. In Mode 5, the algorithm transmits data at the highest frequency when the facility is congested and minimizes the transmissions during free-flow periods.

To make the surveillance system least susceptible to false alarms, errors and to accommodate various communication modes and portability, the algorithms were designed to be self-calibrating with rigorous error diagnostic mechanism for sensor, communication link and data included. The data validation and detector diagnostic tools were developed to eliminate malfunctioning detectors and transient errors. These data validation tools can help improve the data quality, which in turn would improve speed estimates and decision-making process by operating agencies even when using conventional surveillance. In order to pull out the trends and relationships between traffic variables in the presence of noise, clean data are required which

would also simplify the control algorithms. To clean data and improve speed estimations, a methodology was developed which applies threshold value tests and traffic flow theory principles to identify erroneous speed estimates. These erroneous estimates are re-estimated using redundant information from adjacent lanes and recent history.

The data communication structure was developed for several contemporary wireless technologies available in the United States. A communication protocol was developed that aimed at minimizing the message size as well as maximizing the amount of information that can be transmitted within the limited data size. The protocol also ensured that certain reliability checks are included in the message so that the TMC can identify whenever a message is lost or delayed. To compute the recurring communication cost, a cost equation was developed that took the number of transmissions per station as the input.

The data-cleaning, speed estimation, decentralized surveillance, control and communication methodology were tested off-line with 30 second aggregate data collected from a diverse set of freeways and it was shown to be fairly insensitive to site-specific characteristics. The summary of the results follows:

- First, the individual lane speed estimates obtained from the methodology developed in this work were shown to have improved by as much as 70 percent as compared to unbiased conventional estimates and by as much as 125 percent as compared to conventional raw estimates. It was also shown that the directional estimate of speed obtained by averaging the improved individual lane estimates across all lanes is more accurate than conventional estimates.
- Second, the communication Modes 1-4 required transmission of less than 2 percent of the samples on all locations. Mode 5 required transmission of less than 2 percent of samples for remote locations and as high as 21 percent of samples for locations that observe extended congestion periods. The number of transmissions for all communication modes was found to be highest for stations close to the CBD, while the remote stations required few or no transmissions. The fewer number of transmissions for remote stations further illustrates the premise of a decentralized, event-based surveillance structure where the cost of communication is usage based. The average error in speed estimates obtained using information transmitted by each Mode was less than 8 mph for all modes and locations. The time-series speed estimates as compared to measured speed for Modes 4 and 5 showed the best performance.
- Third, at the current costs of wireless communication, all decentralized modes should be cost effective for the outlying stations and only a few of the stations in the CBD would be more cost effective with unlimited service under Mode 5. These modes would allow lower cost surveillance in lower demand areas. By tracking the number of transmissions at the TMC, a station that uses a per-transmission cost plan can be migrated to flat rate unlimited bandwidth plan as demand increases or service costs decrease.

The decentralized surveillance structure developed in this work is not conducive to applications that require data with temporal and spatial breadth, e.g., coordinated ramp metering, vehicle re-identification algorithms and dynamic traffic assignment applications. On the other hand, the cost of wireless communication in the United States was observed to decrease during the course of this work and this trend is expected to continue. The reduced wireless communication costs would reduce the incentive for deploying decentralized structure developed in this work.

Nevertheless, the concepts developed and analyzed in this work can help improve the decision making process of operating agencies. Also, these concepts will provide a greater opportunity for deploying surveillance application in the regions where dedicated communication network costs are large, e.g., developing countries like India where the expressways and ITS applications are at the initial stage of development but the wireless communication technology is more advanced than the United States.

7 ACKNOWLEDGEMENTS

Ramachandran Mallika and Manish Jain contributed significantly to the research presented throughout this report and the documentation thereof. They were both hard workers and I enjoyed working with them.

This material is based upon work supported in part by the National Science Foundation under Grant No. 0127944 and by the California PATH (Partners for Advanced Highways and Transit) Program of the University of California, in cooperation with the State of California Business, Transportation and Housing Agency, Department of Transportation. The Contents of this report reflect the views of the author who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification or regulation.

8 REFERENCES

- [1] Maccubbin, R. P., Staples, B. L., and Mercer, M. R., *Intelligent Transportation Systems Benefits and Costs: 2003 Update*, technical report FHWA-OP-03-075, FHWA, USDOT, Washington, D.C., May 2003.
- [2] Varaiya, P., "California's performance measurement system: Improving freeway efficiency through transportation intelligence," July 2001, *TR News*, pp 218, Jan-Feb 2002.
- [3] May, A., Cayford, R., Coifman, B., Merritt, G., *Automated Diagnostics of Loop Detectors and the Data Collection System in the Berkeley Highway Laboratory, TO 4307 Final Report*, University of California - PATH, 2004.
- [4] Banks, J., Powell, P., "San Diego Field Operational Test of Smart Call Boxes: Institutional Issues," TRB, *Transportation Research Record 1603*, 1997, pp 20-26.
- [5] Banks, J., Powell, P., "San Diego Field Operational Test of Smart Call Boxes: Technical Aspects," TRB, *Transportation Research Record 1603*, 1997, pp 27-33.
- [6] Johnson, C. M., and Thomas, E. L., *Communications for Intelligent Transportation Systems, successful practices*, ITS Joint Program Office, Federal Highway Administration (U.S. Department of Transportation), FHWA-JPO-99-023; FTS-TRI-11-00-02, November 2000
- [7] Margiotta, R., *State of the Practice for Traffic Data Quality*, white paper prepared for Office of Policy Federal Highway Administration Washington, D.C., 2002.
http://www.itsdocs.fhwa.dot.gov/jpodocs/repts_te/13768.html
- [8] Coifman, B., Lyddy, D., and Skabardonis, A., "The Berkeley Highway Laboratory-Building on the I-880 Field Experiment", Proc. *IEEE ITS Council Annual Meeting, IEEE*, 2000, pp 5-10.
- [9] <http://pems.eecs.berkeley.edu/> - Last accessed on 11/24/2003
- [10] Weil, R., Wootton, J., Garcia-Ortiz, A., "Traffic Incident Detection: Sensors and Algorithms," *Mathematical. Computer Modeling*, Vol 27, no 9-11, 1998, pp 257-291.
- [11] Persaud, B., Hall, F., and Hall, L., "Congestion Identification Aspects of the McMaster Incident Detection Algorithm," TRB, *Transportation Research Record 1287*, 1990, pp 167-175.
- [12] Shannon, C., "Communication in the presence of noise," *Proceedings of the Institute of Radio Engineers (IRE)*, vol. 37, no 1, 1949, pp 10-21.
- [13] Sheridan, T., "Reflections on Information and Information Value", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 25, no 1, 1995, pp 194-196.
- [14] Howard, R., "Information Value Theory," *IEEE Trans. Systems Science and Cybernetics*, vol. SSC-2, no 1, 1966, pp 22-26.
- [15] Lin, W., Daganzo, C., "A Simple Detection Scheme for Delay-Inducing Freeway Incidents," *Transportation Research-Part A*, Vol 31A, No 2, 1997, pp 141-155.

- [16] Lighthill, M. and Whitham, G., "On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads," *Proc. Royal Society of London, Part A*, vol 229, no 1178, 1955, pp 317-345.
- [17] Richards, P., "Shock Waves on the Highway", *Operations Research*, 4 (1), 1956, pp 42-51.
- [18] Newell, G., "A simplified theory of kinematic waves in highway traffic i: General theory. ii: Queuing at freeway bottlenecks. iii: Multi-destination flows", *Transportation Research Part B* 27, no 4, 1993, 281-313.
- [19] Varaiya, P., "Freeway Data Collection, Storage, Processing, and Use", Presented at the Workshop on the National Transportation Information Infrastructure, Transportation Research Board, 2002.
- [20] Chen, C., Kwon, J., Skabardonis, A., Varaiya, P., "Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems," TRB, *Transportation Research Record* 1855, 2003, pp 160-167.
- [21] Coifman, B., "Using Dual Loop Speed Traps to Identify Detector Errors," TRB, *Transportation Research Record* 1683, 1999, pp 47-58.
- [22] Coifman, B., Dhoorjaty, S., and Lee, Z. "Estimating Median Velocity Instead of Mean Velocity at Single Loop Detectors", *Transportation Research: Part C*, vol. 11, no 3-4, 2003, pp 211-222.
- [23] Dailey, D., "A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements", *Transportation Research-B*, Vol. 33B, No 5, June, 1999, pp 313-322. Elsevier Science, London
- [24] Jia, Z., Chen, C., Coifman, B., Varaiya, P., "The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors", Presented at *IEEE 4th International ITS Conference*, 2001.
- [25] Van Zwet, E., Chen, C., Jia, Z., Kwon, J., "A statistical method for estimating speed from single loop detectors", Freeway Performance Measurement System (PeMS), March 2003.
- [26] Mikhalkin, B., Payne, H., Isaksen, L., "Estimation of speed from presence detectors." Hwy. Res. Rec. 388, *Highway Research Board, National Research Council*, Washington, D.C., 1972, pp 73-83
- [27] Turochy, R., Smith, B., "A new procedure for Detector Data Screening In Traffic Management Systems", *Transportation Research Record* 1727, 2000, pp127-131
- [28] Coifman, B., "Improved Velocity Estimation Using Single Loop Detectors", *Transportation Research: Part A*, vol. 35, no 10, 2001, pp. 863-880
- [29] Neelisetty, S., *Detector Diagnostics, Data Cleaning and Improved Single Loop Velocity Estimation from Conventional Loop Detectors*, Masters thesis, OSU, 2004.
- [30] http://www.wireless-data.com/wireless_tech.html - Accessed on 11/11/2003
- [31] Elliott, S. D., and Dailey, D. J., *Wireless Communications for Intelligent Transportation Systems*, Artech House Inc., 1995
- [32] http://www.cingular.com/business/calling_plans/ - Accessed on 11/14/2003

- [33] http://www.attwireless.com/mobileinternet/bus_plan.jhtml - Accessed on 11/14/2003
- [34] Howard, P. G., *The design and analysis of efficient lossless data compression systems*, PhD thesis, Brown University, 1993
- [35] http://www.ericsson.com/network_operators/mobitex/subpages/case/mt/mtwea.shtml - Accessed on 11/24/2003

9 APPENDIX A, A DETAILED DESCRIPTION OF THE DECENTRALIZED COMMUNICATION AND SURVEILLANCE ALGORITHMS

In this section we present the detailed methodology outlining the pre-filtering logic and criteria used for making a decision about transmitting data by each Communication Mode developed in Section 4.1.1.

Mode 1: Free-flow or congested

In order to make the algorithm tolerant to noise, the decision to transmit information in Mode 1 is triggered when one of the following condition sets are satisfied. For testing the conditions sets, a single measure of flow, occupancy and speed per direction per sample is obtained by taking the median across each individual lane's measured flow, occupancy and speed. Care is taken to use only the lanes that have non-zero flow during the sampling period and are not marked as malfunctioning. Also, the data were cleaned and validated using the algorithm developed in Chapter 3 before aggregating across lanes. Sets 1-3 are tested when the last transmitted traffic state is free-flow and the speeds might be changing from free-flow to congested. Set 4 is tested when the last transmitted traffic state is congested.

Set 1: The trend in speed is computed by taking the difference in speeds of the previous three consecutive samples. Also the absolute difference of current speed and previous samples speed is computed (henceforth referred to as speed-diff) and compared with a threshold (10 mph in this case) to distinguish a true change in speeds of consecutive samples from expected variation in speeds between samples. When the speeds begin to drop the variance, and hence the standard deviation, between the consecutive measurements increases. The variance increases during recovery also, but not during stable velocities. This property is used to further verify that the traffic state is changing. The standard deviation, 's', of the current sample's speed and last few sample's speed (four samples in this case, i.e., $n = 5$) is calculated using the expression:

$$s = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \quad (\text{A.1})$$

$$\text{where } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

If s is greater than a threshold value, e.g., 4 in our case, then there is considerable variance between recent speeds. One large erroneous sample may cause the variance to be considerably high giving a false indication about the traffic state. But to confirm the change, all the conditions in the set have to be simultaneously satisfied which would reduce the likelihood of the test being influenced by the erroneous sample. Thus if the current state is free-flow, trend in speed of three consecutive samples is negative, speed-diff > 10 mph and $s > 4$, then the current traffic state is considered to be changing from free-flow to congested and decision to transmit this information is made. If any condition in this set is not satisfied then the condition Set 2 is tested.

Set 2: In this set, speed-diff and s are computed as in Set 1. The decision to transmit information is triggered if the last reported traffic state is free-flow, $\text{diff-speed} > 10$, $s > 4$ and number of consecutive samples with speed below free-flow threshold (henceforth referred to as consecutive-below-threshold) is greater than two. If any condition in this set is not satisfied then the conditions in Set 3 are tested.

Set 3: In this set, if the last reported traffic state is free-flow, the consecutive-below-threshold > 3 , and the trend is negative then decision to transmit information is triggered. This set indicates that the traffic state is slowly changing from free-flow to congested.

If an individual measurement is observed to be below the free-flow threshold while the last reported state was free-flow, and none of the condition sets 1-3 are satisfied then a counter consecutive-below-threshold is incremented by one and the algorithm waits for the next sample. This counter is used by the condition sets above in successive samples to identify the change in traffic state. When a sample is observed to be above the free-flow threshold the counter, consecutive-below-threshold, is reset to zero. If a decision to transmit information about a change of state from free-flow to congested is made, the communication with the TMC is initiated and the current traffic state along with information about anomalies in individual lane behavior, if any, is transmitted. The algorithm updates the variable keeping track of the last transmitted current traffic state (henceforth referred to as last-transmitted-state) to congested. The variable counting number of consecutive samples that are above free-flow threshold is reset to zero. This counter is used in the following condition set 4 to identify the free-flow traffic state.

When the last-transmitted-state is congested, the algorithm tests the following condition Set 4 to determine if the traffic state is changing from congested to free-flow.

Set 4: If the current sample's speed is above the free-flow threshold and the number of consecutive samples that are above free-flow threshold is greater than two, then the decision to transmit information is triggered. However, if the current sample's speed is greater than free-flow threshold and a decision to transmit information is not made then the variable counting the number of consecutive sample above free-flow threshold (consecutive-above-threshold) is incremented by one and the algorithm waits for the next sample. This counter is used in Set 4 to minimize the error due to flicker in speeds and is reset to zero if a speed measurement is below the threshold.

If a decision to transmit information about a change of state from congested to free-flow is made, the communication with the TMC is initiated and the current traffic state along with information about anomaly in individual lanes behavior, if any, is transmitted. The algorithm updates the variable last-transmitted-state to free-flow and the process is repeated.

Mode 2: Multiple thresholds drop

The data for Communication Mode 2 are aggregated and validated as in Mode 1. The following condition sets 1-2 are tested sequentially when the current speed is below free flow threshold to make a decision whether or not to transmit data.

Set 1: The trend and median of speed for the current sample and previous four samples is computed and the threshold range within which the current speed measurement lies is determined. If the median of previous few sample's speed lies in or above the same threshold range as the current speed, the trend is negative, and the new threshold range is lower than the last reported range, then the decision to transmit information is made. For example, if the current speed lies between 20 and 30 mph range, the median of previous four measurements and current speed is less than 30 mph, the trend of speeds for last four samples is negative and the last reported threshold range is greater than 30 mph, then a decision to transmit information will be made and the new reported traffic state would be the 20-30 mph range.

Set 2: For this set, the decision to transmit data is triggered when the trend is negative, standard deviation $s > 4$ and number of consecutive samples with speed below last reported threshold range is greater than two.

If the current speed is below the last reported threshold range and neither condition Set 1 nor Set 2 is satisfied then a counter of the number of consecutive samples below last reported threshold range is incremented. This counter is used for testing Set 2 in the consecutive sample. If an individual measurement lies in a range that is above the last transmitted range, then the counter is reset to zero. When the decision to transmit the information is made, the communication with the TMC is initiated and the current traffic state, new threshold range, and the lane diagnostic information, if any, are transmitted. The algorithm updates the variable keeping track of the last transmitted threshold range to the new threshold range.

The condition sets 1 and 2 follow the dropping speeds and the recovery back to free-flow is reported directly if the following condition set 3 is satisfied.

Set 3: If the number of consecutive samples that are above free-flow threshold is greater than two and the trend of speed is positive then the decision to transmit information about return to free-flow state is made.

When condition set 3 is satisfied, the algorithm initiates the communication with the TMC and reports that traffic is free flowing. The algorithm also updates the variable keeping track of the last reported threshold range to free-flow. Thus, in this communication mode, at least two and a maximum of five transmissions (up to four for different threshold range's and one to indicate free-flow condition) are made every time the traffic state changes from free-flow to congested and back to free-flow. The maximum number of possible transmissions depends on the number of threshold ranges desired.

Mode 3: Multiple thresholds drop and recovery

In Mode 3, the algorithm is similar to Mode 2, transmitting data when the traffic drops below several thresholds but it also reports slow recovery at every stage by slowly tracking any trends of increasing speeds. The criteria for transmitting information to report onset of congestion are same as described in Set 1 and 2 of Mode 2. The information about slow recovery is transmitted at every stage, i.e., 20 mph, 30 mph, 40 mph and above free-flow threshold and the criteria for transmission are similar in

principle to that for reporting onset of congestion. Instead of looking for negative trend, the algorithm now looks for positive trend and a larger number of samples (five 30-second samples in our case) to report recovery.

Mode 4: Significant deviation from last reported state

Before describing the condition sets that are tested to make a decision about transmitting data in Mode 4, let us define three mutually exclusive logical flags, i.e., $f_uptrend$, $f_downtrend$, f_stable , that will be often used in the discussion of the sets. These flags can take two values, either 1 (true) or 0 (false). When any one flag is set at value = 1, the remaining flags are set at value = 0, i.e., at any time at most only one flag can be true otherwise all false. These flags, $f_uptrend$, $f_downtrend$, and f_stable , respectively represent three possible trends, positive, negative, and stable, exhibited by the speeds in consecutive samples. For instance, if the speeds are dropping and exhibit a negative trend then the flag $f_downtrend$ would be set at value = 1 and the other two flags would be set at value = 0.

The following condition sets 1-8 are tested for when the current sample is below the free-flow threshold and the condition Set 9 is tested when the last transmitted traffic state is congested and the current sample is above the free-flow threshold. The condition sets 1-5 are tested when any one of the three flags ($f_uptrend$, $f_downtrend$ and f_stable) are true, i.e., the traffic state is different from free flow state and is following a trend. The condition sets 6 - 8 are tested when all three flags have value = 0 indicating that the traffic is changing from free-flow to congested.

Set 1: The standard deviation s and the trend in speed of the previous three consecutive samples and the current sample are computed. Also the difference of current speed and the speed of the sample collected three samples earlier is taken (speed-diff). If the trend in speed of consecutive four samples is positive, $speed-diff > 15$ mph, $s > 4$, the flag $f_uptrend = 0$ and the flag $f_stable = 0$ then the current traffic state is considered to be recovering from the last reported state and decision to transmit this information is made. This situation indicates that the traffic state had earlier shown steep drop and has recovered from that state quickly without moving to a stable state and would be seen as a sharp spike in the time series of speed. At this point the flag $f_uptrend$ is set at value = 1 and the flags $f_downtrend$ is reset to value = 0. The communication with the TMC is initiated and the current traffic state is transmitted. If any condition in this set is not satisfied then the next condition set is tested.

Set 2: The trend, standard deviation and median of speed are computed for measurements of previous three consecutive samples and the current sample. Also, the numbers of consecutive samples that are below free-flow threshold and have absolute "speed-diff" as compared to last transmitted speed greater than 20 mph are computed (henceforth referred to as second-sample-ctr). This counter helps identify whether the current operating speeds differ by at least 20 mph from the last reported or expected speeds. If the trend is absent, i.e., stable, the standard deviation $s < 4$, the counter $second-sample-ctr > 1$, the flag $f_stable = 0$, and median of last four sample speeds is below free-flow threshold then current traffic state is considered to be changing to a stable state and the decision to transmit this

information is made. This situation is observed after the speeds drop or recover sharply for few samples and then become stable, i.e., just after a traffic state transition. The algorithm sets the value of flag f_stable to 1. The flags $f_downtrend$ and f_stable are reset to value = 0. The counter $second_sample_ctr$ is reset to zero and the communication with the TMC is initiated. The current stable traffic state along with information about anomalies in individual lane behavior, if any, is transmitted. If any condition in this set is not satisfied then the condition set 3 is tested.

- Set 3:** The absolute difference of the last transmitted speed and the current speed is computed. If the absolute difference of speed is greater than 20 mph, the flag $f_stable = 1$, and the $second_sample_ctr > 1$, then the decision to transmit information is made. This set indicates that the speeds have slowly changed from last reported stable state to another stable state. The communication with the TMC is initiated and the information about the new stable state is transmitted.
- Set 4:** The standard deviation and trend in speed are computed for the previous three consecutive samples and the current sample. Also the difference of current speed and the speed of sample collected three samples earlier is taken ($speed_diff$). If the trend in speed of consecutive four samples is negative, $speed_diff > 10$ mph, $s > 4$, and the flag $f_downtrend = 0$ then the traffic state is considered to be deteriorating from the last transmitted state and the decision to transmit is made. This condition set is satisfied when either the traffic state deteriorates from the last transmitted stable state or when the speeds begin to drop again after some recovery. The flag $f_downtrend$ is set to 1, the flags $f_uptrend$ and f_stable are set to zero, and the counter $second_sample_ctr$ is set to zero. The communication is initiated and the current traffic state is transmitted. If any of the conditions in the set are not satisfied then the next set is tested.
- Set 5:** In this set, if the current sample is below free-flow threshold, the absolute difference in speeds of current sample and last transmitted speed is greater than 20 mph and none of the condition sets 1-4 were satisfied then the counter $second_sample_ctr$ is incremented by one and the algorithm waits for the next sample. The algorithm uses this counter in consecutive samples to identify the traffic-state.

As noted earlier, the condition set 1-4 track the change in traffic speed from the last reported congested state and report either recovery, further deterioration or new stable state. In the following condition sets 6-8, the algorithm tracks the change of state from free-flow condition to congested. Here all the three flags ($f_uptrend$, $f_downtrend$, and f_stable) have an initial value = 0.

- Set 6:** The trend, standard deviation s and the $speed_diff$ are computed as before. If the trend is negative, $s > 4$, $speed_diff > 15$ mph, current speed is below free-flow threshold then the traffic state is considered to be changing from free-flow to congested and the decision to transmit is made. The flag $f_downtrend$ is set to true ($f_downtrend = 1$), the communication with the TMC is initiated and the current traffic state is transmitted. If any of the conditions in this set are not satisfied then $consecutive_below_threshold$ is incremented and the algorithm waits for the next sample. On the other hand, if a decision to transmit is made then the counter

consecutive-below-threshold is reset to zero. This counter is used by other condition sets in successive samples.

- Set 7:** In this set, if the counter consecutive-below-threshold > 1 , $s > 4$, and speed-diff > 15 mph, then the decision to transmit information is made. This set represents the situation when the speeds drop sharply and there are not enough samples to establish the trend. Once the decision to transmit information is made, the flag $f_downtrend$ is set to 1, communication is initiated and the information about the changing state is transmitted as before.
- Set 8:** Sometimes the traffic state changes slowly from free flow to a congested state and the dropping speeds do not exhibit any trend. In this set, if the counter consecutive-below-threshold > 5 and $s < 4$ then the traffic state is considered to be slowly changing from free-flow to a stable congested state and the decision to transmit this information is made. The flag f_stable is set to true ($f_stable = 1$) and the communication with the TMC is initiated. The information about current state is transmitted and the counter consecutive-below-threshold is reset to zero. If any of the conditions in this set are not satisfied then the counter consecutive-below-threshold is incremented by one and the algorithm waits for the next sample.

The conditions in set 1-8 track the change of traffic state from free-flow to congested and once congested, the change in states that do not result in complete recovery to free-flow. The following condition Set 9 tracks the change in traffic state that results in complete recovery from congested to free-flow state.

- Set 9:** If the current sample speed is above the free-flow threshold and the counter consecutive-above-threshold > 2 , then the traffic state is considered to have changed from congested to free-flow and a decision to transmit this information is made. The flags ($f_uptrend$, $f_downtrend$ and f_stable) are reset to zero; the counter consecutive-above-threshold is reset to zero, and the counter consecutive-below-threshold is also reset to zero. In order to reduce the error due to noise and flicker, the algorithms waits for 2 consecutive samples that are above free-flow threshold to report the free-flow state. If a decision to transmit information is not made after testing this set, then the counter consecutive-above-threshold is incremented by one and the algorithm waits for the next sample. If the next sample is observed to be below free-flow threshold then the counter is reset to zero. This problem can happen either when the speeds flicker from congested to free-flow and back, or when few samples are observed above free-flow and the traffic state is still congested.

Mode 5: Highest frequency when congested

In Mode 5, the following sets 1 and 2 are tested to identify whether the facility is becoming congested (from free-flow to congested) and the condition Set 3 is tested when the facility is in congested state. The free-flow threshold is set at 50 mph as before.

- Set 1:** The trend in speed is computed by taking the difference in speeds of previous two consecutive samples and the current sample. Next, the standard deviation s and the speed-diff are computed as in Set 1 of Mode 4. If the last transmitted traffic

state is free-flow, trend is negative, $s > 4$, and the speed-diff > 10 mph then the current traffic state is considered to be changing from free-flow to congested and decision to transmit this information is made. If any condition in this set is not satisfied then the conditions in Set 2 are tested.

Set 2: In this set, if the number of consecutive samples that are below free-flow threshold is greater than two, speed-diff > 10 , current sample is below free flow threshold and the last transmitted state is free-flow then the decision to transmit information is made. If at this point, both Set 1 and 2 are not satisfied and the current sample is below free-flow threshold then the counter consecutive-below-threshold is incremented by one and the algorithm waits for the next sample.

If a decision to transmit information is made based on Set 1 and 2, then the variable keeping track of last transmitted state is set to congested and communication with the TMC is initiated. The counter consecutive-below-threshold is reset to zero. The current traffic state and data in the future samples are transmitted until conditions in the following Set 3 are satisfied.

Set 3: If the last transmitted state is congested, the current sample speed is above free-flow threshold and the counter consecutive-above-threshold > 2 then the traffic state is considered to be changing from congested to free-flow and this information is transmitted to the TMC. If the current sample is above free-flow threshold and the decision to transmit information is not made then the counter consecutive-above-threshold is incremented by one and the algorithm waits for the next sample.

If a decision to transmit information is made based on Set 3 then the last transmitted state is updated to free-flow state, the free-flow state is communicated to the TMC. Communication is initiated again if the facility goes back to congested state as indicated by either Set 1 or 2

10 APPENDIX B, ANALYSIS OF ADDITIONAL FREEWAYS

The communication algorithms developed in Chapter 4 and the cost equations developed in Chapter 5 were applied to the following corridors: I-80_BHL, SR-51_D3, I-405_D7, and I-15_D11, as described in Section 1.3. In the following section we present the performance measures and results for these corridors.

The plots in Figures B.1-B.4 show the error distribution of synthetic time series speed estimates as compared to the clean speeds for Communication Modes 1-5 on the four corridors mentioned above. In each figure, Column 1 shows the mean absolute error and Column 2 shows mean absolute percentage error. The rows 1-5 are for Communication Mode 1-5 respectively. Compare to Figures 4.7-4.11 A2 and B2. The performance measures from these plots are also summarized in Table 4.2.

The cost equations developed in Chapter 5 were also applied to the four corridors. Table B.1 summarizes the average cost per station for each Communication Mode on the four corridors. Figure B.5 shows the average cost structure for I-80_BHL corridor, Figure B.6 for SR-51_D3, Figure B.7 for I-405_D7 and Figure B.8 for I-15_D11. Compare to Figure 5.2, once more, the plots A-E on each figure are for Communication Mode 1-5 respectively. The plots show the average amount of data transferred by each station per month, and the cost of communication when using either GPRS/GSM (AT&T) or Mobitex (Cingular) technology. The left vertical axis of the plot shows the cost per month in \$ calculated using Equation 5.1 and the right vertical axis shows the amount of data transferred in MB calculated using Equation 5.3. Note the different data transfer vertical scale for plot E, where data is transmitted at the highest frequency during congestion.

Table B.1, Summary of estimated monthly communication costs per station for each Communication Mode and technology by corridor.

Corridor	Mode 1		Mode 2		Mode 3		Mode 4		Mode 5	
	AT&T	Cingular	AT&T	Cingular	AT&T	Cingular	AT&T	Cingular	AT&T	Cingular
I-80_BHL	\$30	\$35	\$30	\$37	\$30	\$46	\$30	\$42	\$61	\$60
SR-51_D3	\$30	\$37	\$30	\$37	\$30	\$37	\$30	\$39	\$38	\$60
I-405_D7	\$30	\$36	\$30	\$37	\$30	\$42	\$30	\$50	\$58	\$60
I-15_D11	\$30	\$32	\$30	\$33	\$30	\$33	\$30	\$37	\$37	\$57

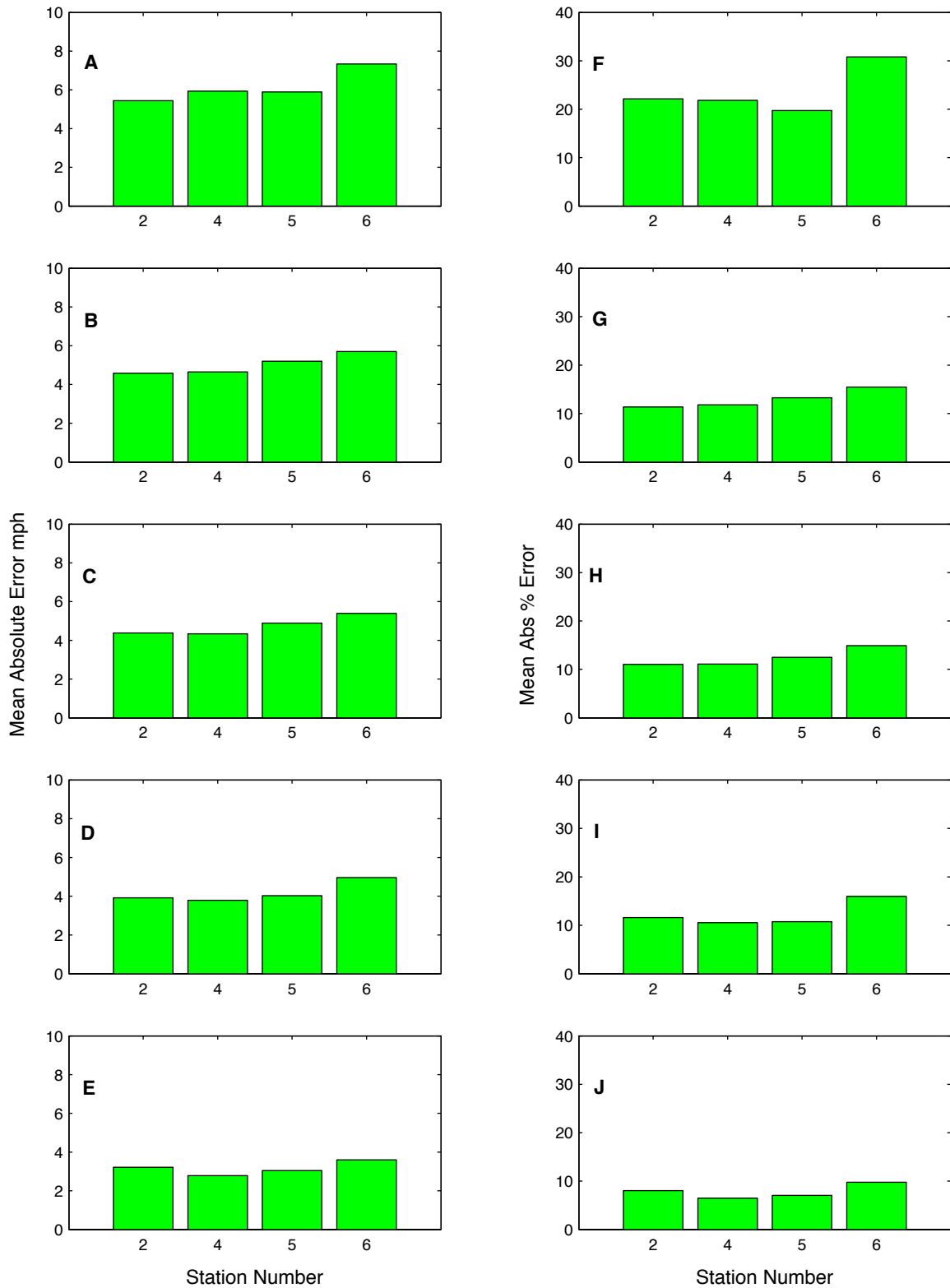


Figure B.1, Performance measures for 4 northbound stations of I-80_BHL corridor. Column 1 shows the mean absolute average error and Column 2 shows mean absolute percentage error. Plots A-E are for Communication Mode 1-5, respectively. Similarly plots F-J are for Communication Mode 1-5.

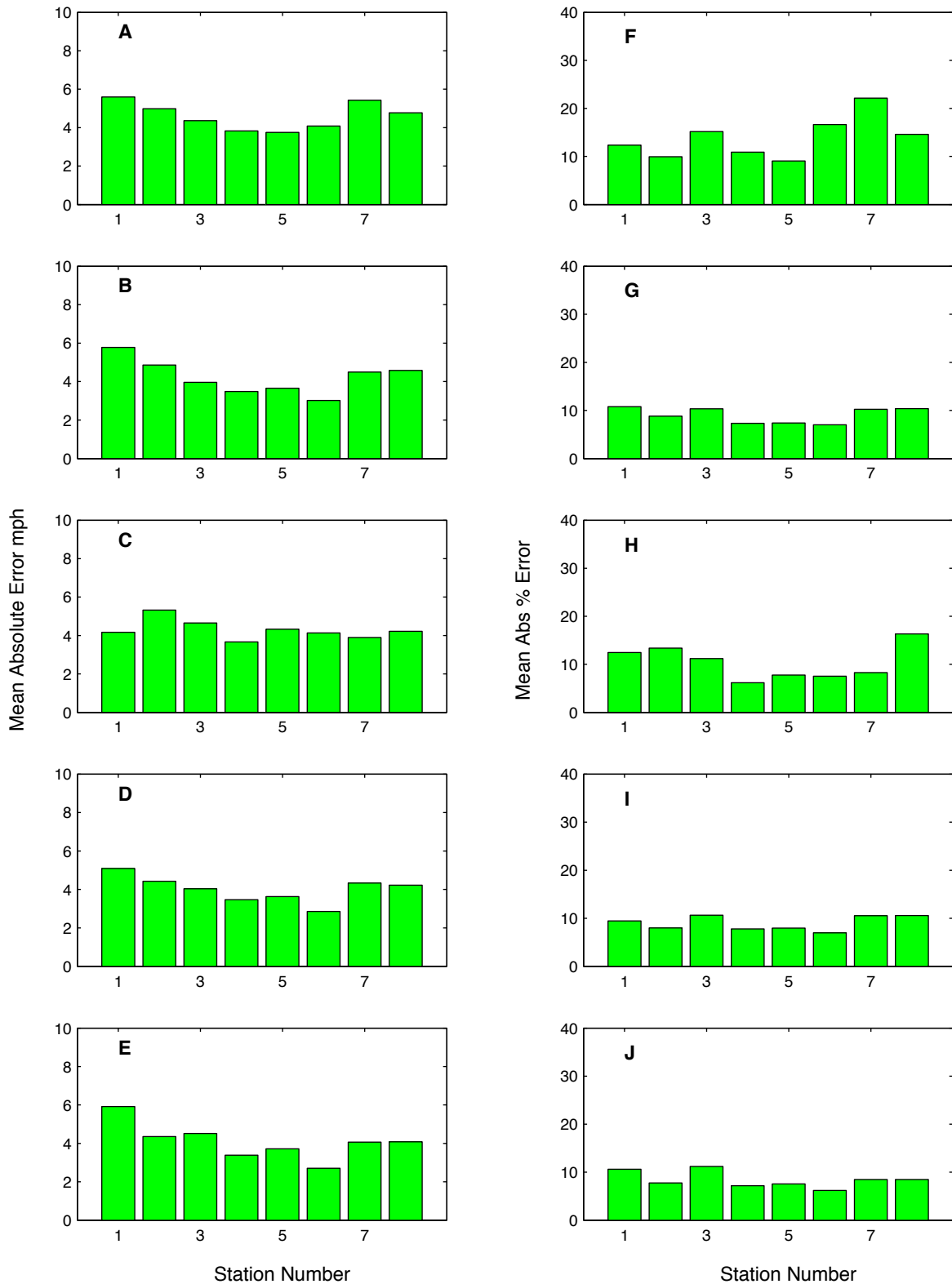


Figure B.2, Performance measures for 8 northbound detector stations on SR-51_D3 corridor. Column 1 shows the mean absolute average error and Column 2 shows mean absolute percentage error. Plots A-E are for Communication Mode 1-5, respectively. Similarly plots F-J are for Communication Mode 1-5.

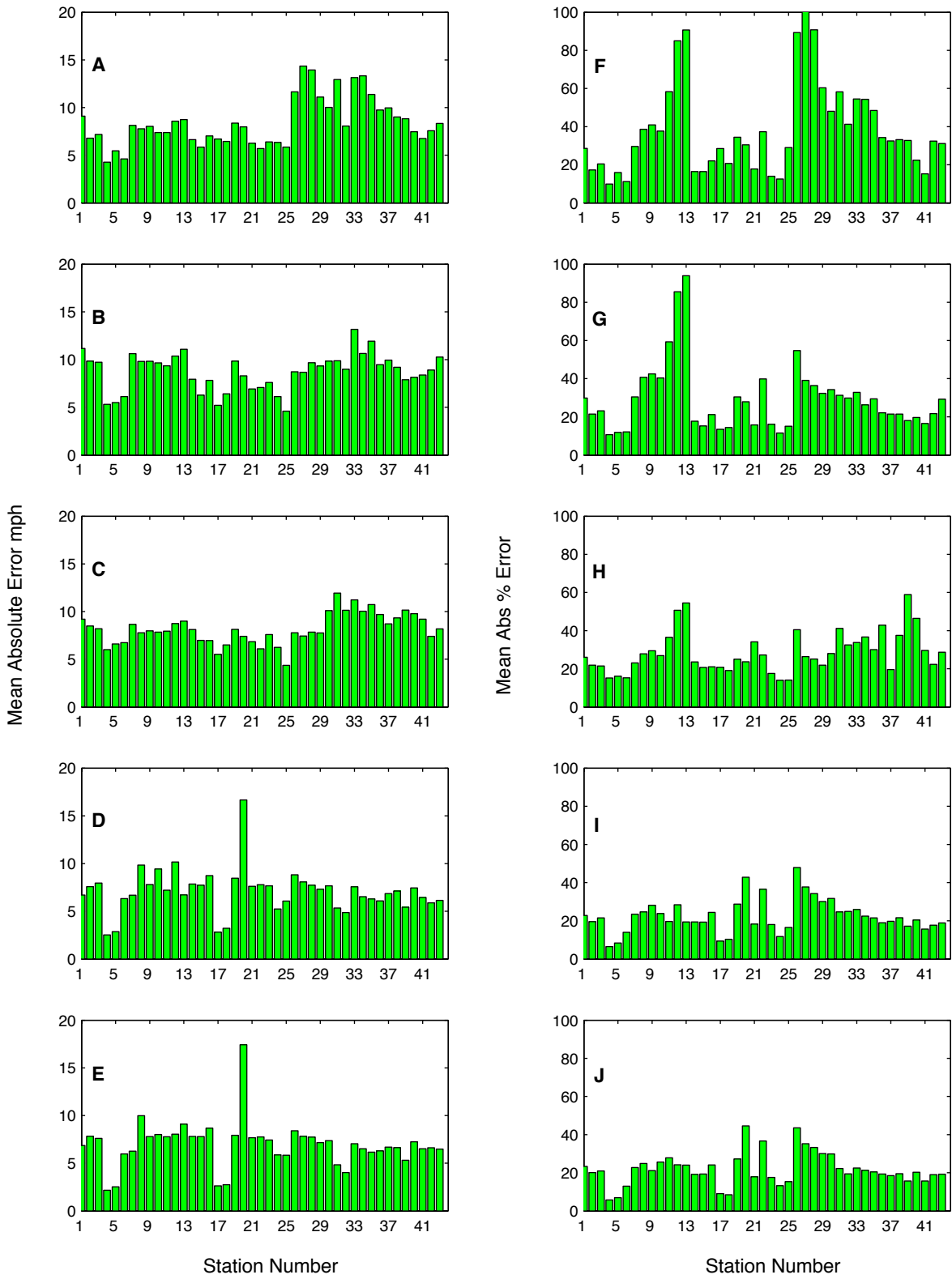


Figure B.3, Performance measures for 43 northbound detector stations on I-405_D7 corridor. Column 1 shows the mean absolute average error and Column 2 shows mean absolute percentage error. Plots A-E are for Communication Mode 1-5, respectively. Similarly plots F-J are for Communication Mode 1-5.

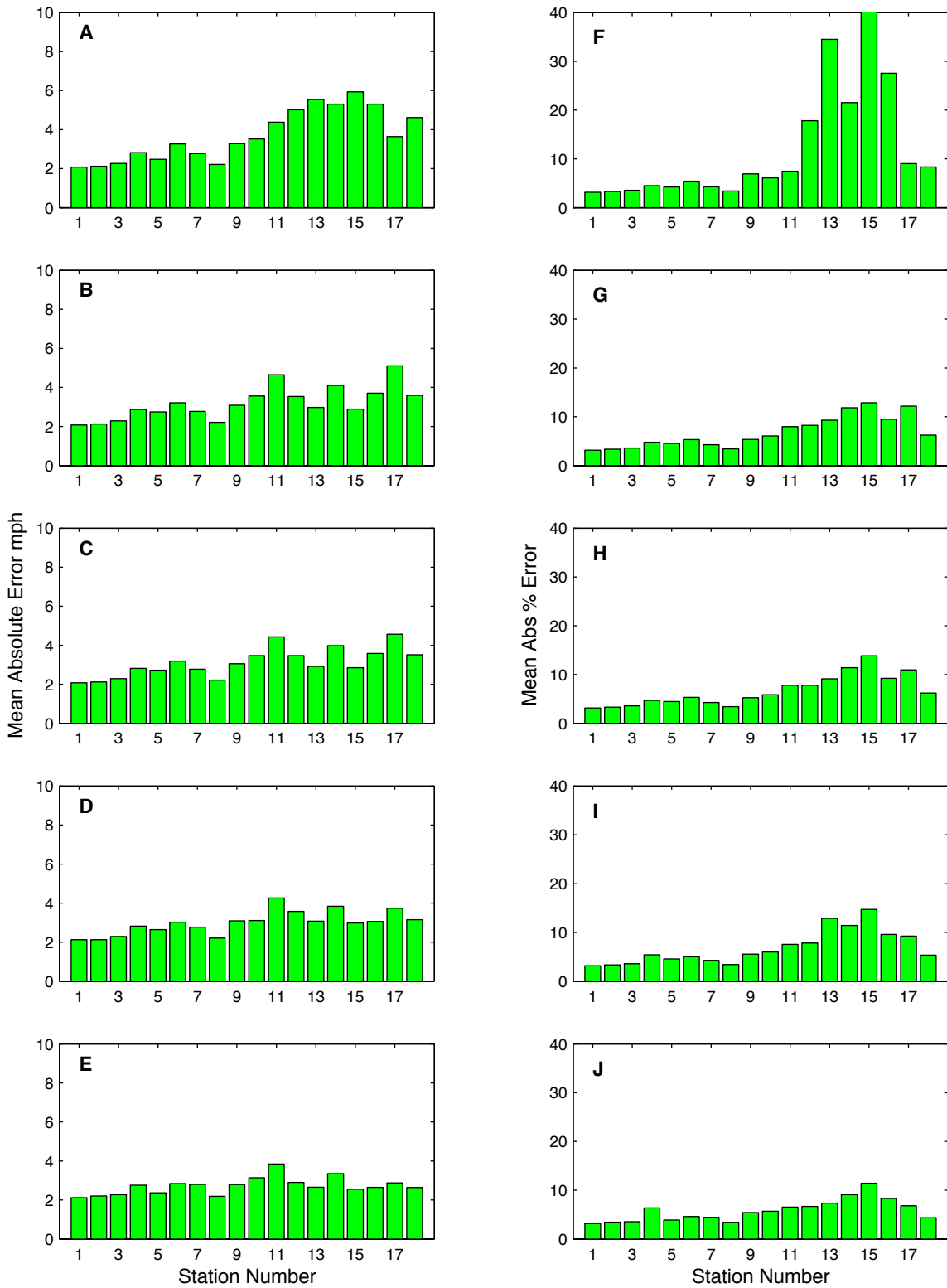


Figure B.4, Performance measures for 18 northbound detector stations on I-15_D11 corridor. Column 1 shows the mean absolute average error and Column 2 shows mean absolute percentage error. Plots A-E are for Communication Mode 1-5, respectively. Similarly plots F-J are for Communication Mode 1-5.

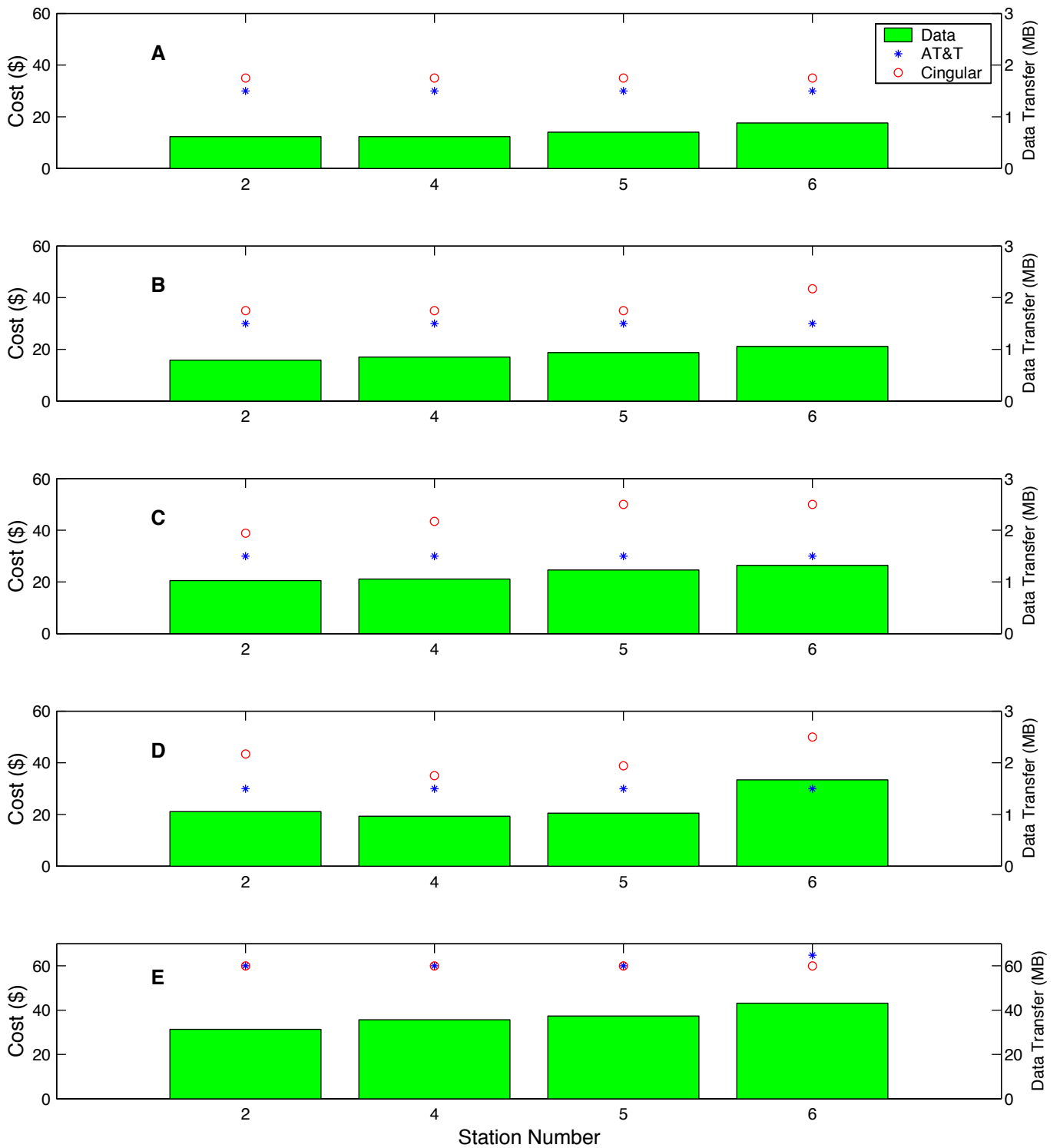


Figure B.5, Estimated average monthly communication cost structure as applied to a month's weekdays data collected on 4 stations of I-80_BHL corridor during August 1998. A-E are for communication modes 1-5, respectively. The plots show the average amount of data transferred by each station per month, and the per station monthly communication cost when using either GPRS/GSM (AT&T) or Mobitex (Cingular) technology. Note the different data transfer vertical scale for plot E (23 times larger than that for plot A-D).

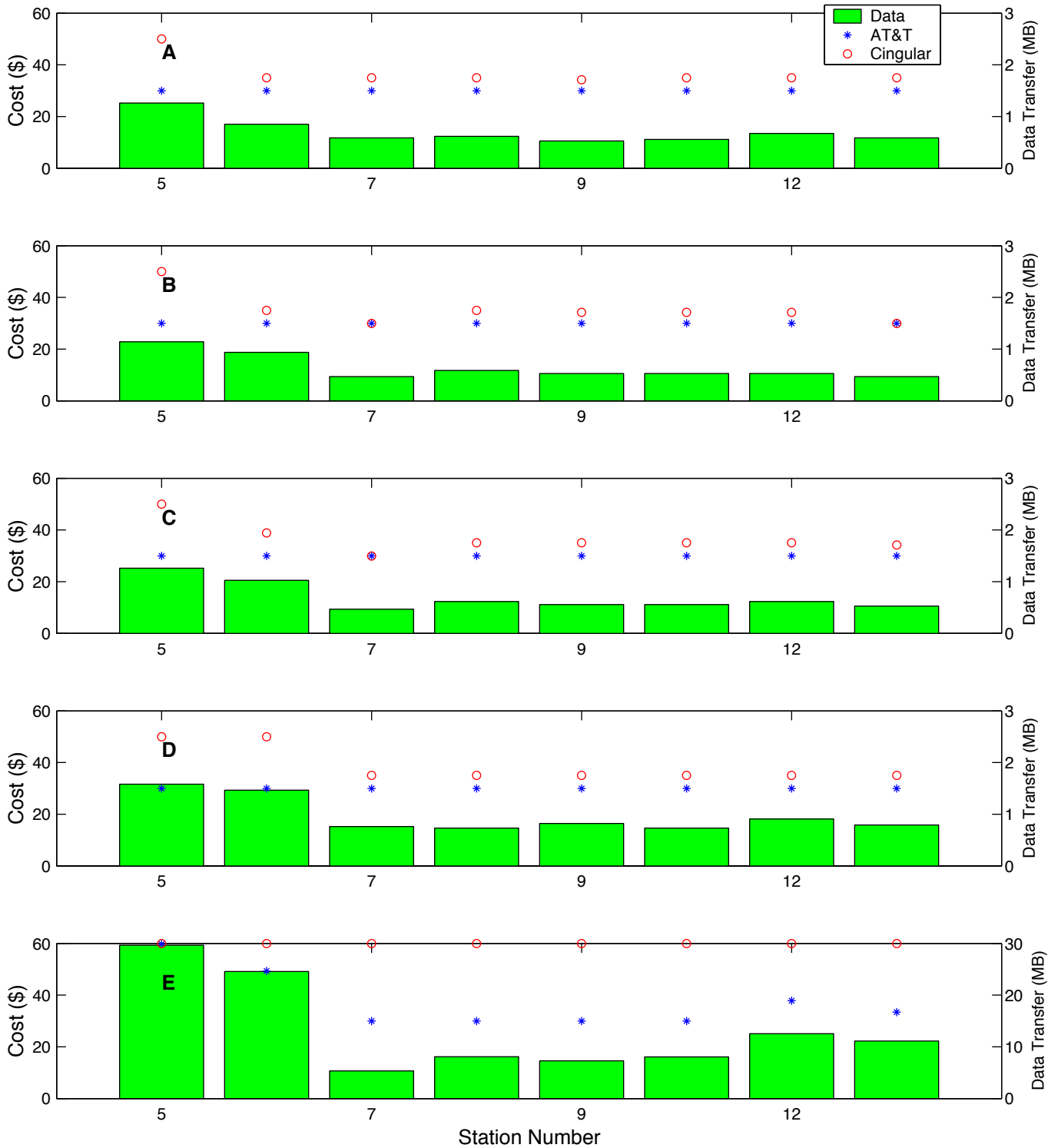


Figure B.6, Estimated average monthly communication cost structure as applied to five weekdays' data collected on 8 stations of SR-51_D3 corridor. A-E are for communication modes 1-5, respectively. The plots show the average amount of data transferred by each station per month, and the per station monthly communication cost when using either GPRS/GSM (AT&T) or Mobitex (Cingular) technology. Note the different data transfer vertical scale for plot E (10 times larger than that for plot A-D).

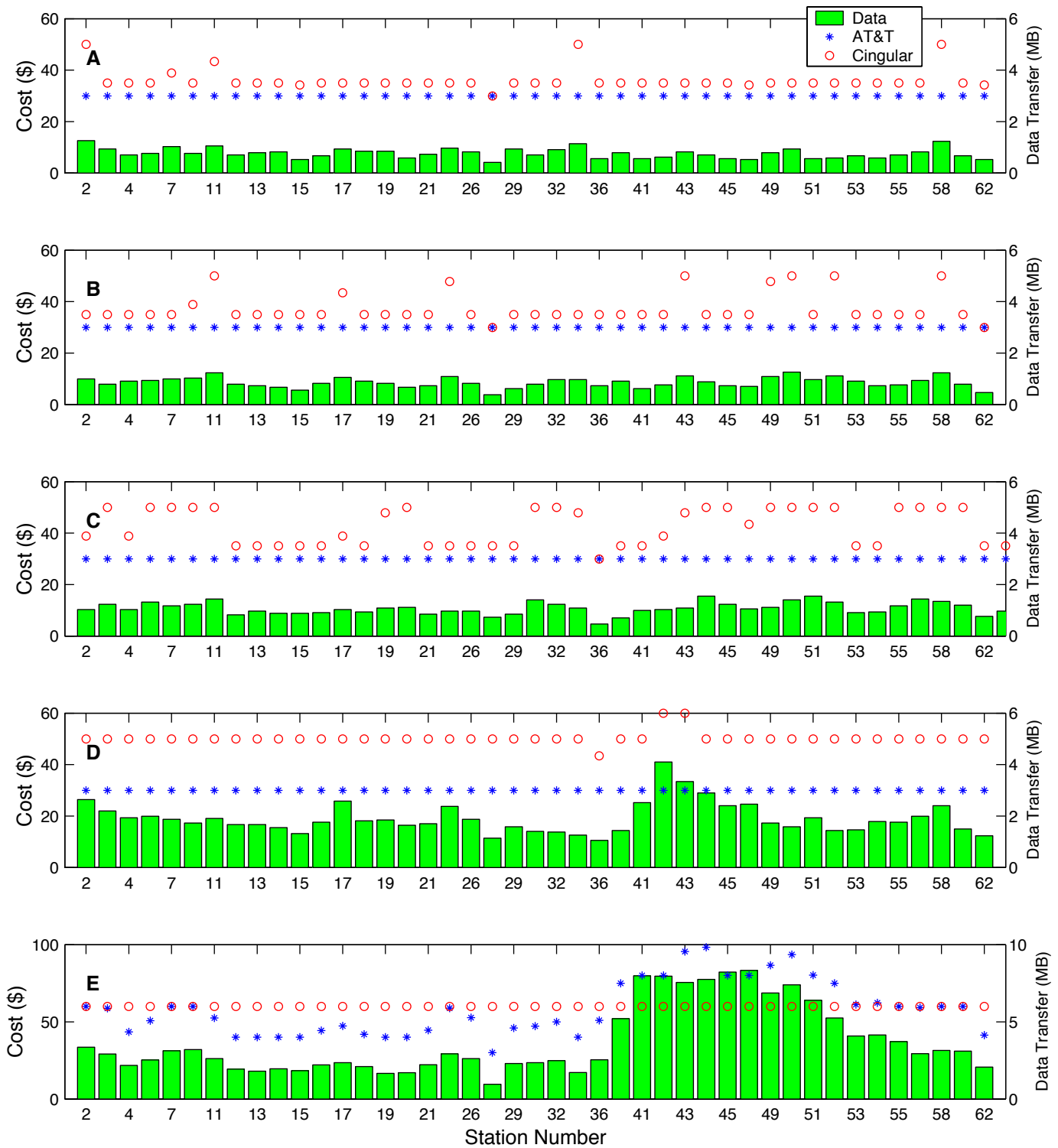


Figure B.7, Estimated average monthly communication cost structure as applied to five weekdays' data collected on 43 stations of I-405_D7 corridor. A-E are for communication modes 1-5, respectively. The plots show the average amount of data transferred by each station per month, and the per station monthly communication cost when using either GPRS/GSM (AT&T) or Mobitex (Cingular) technology. Note the different data transfer vertical scale for plot E (2 times larger than that for plot A-D).

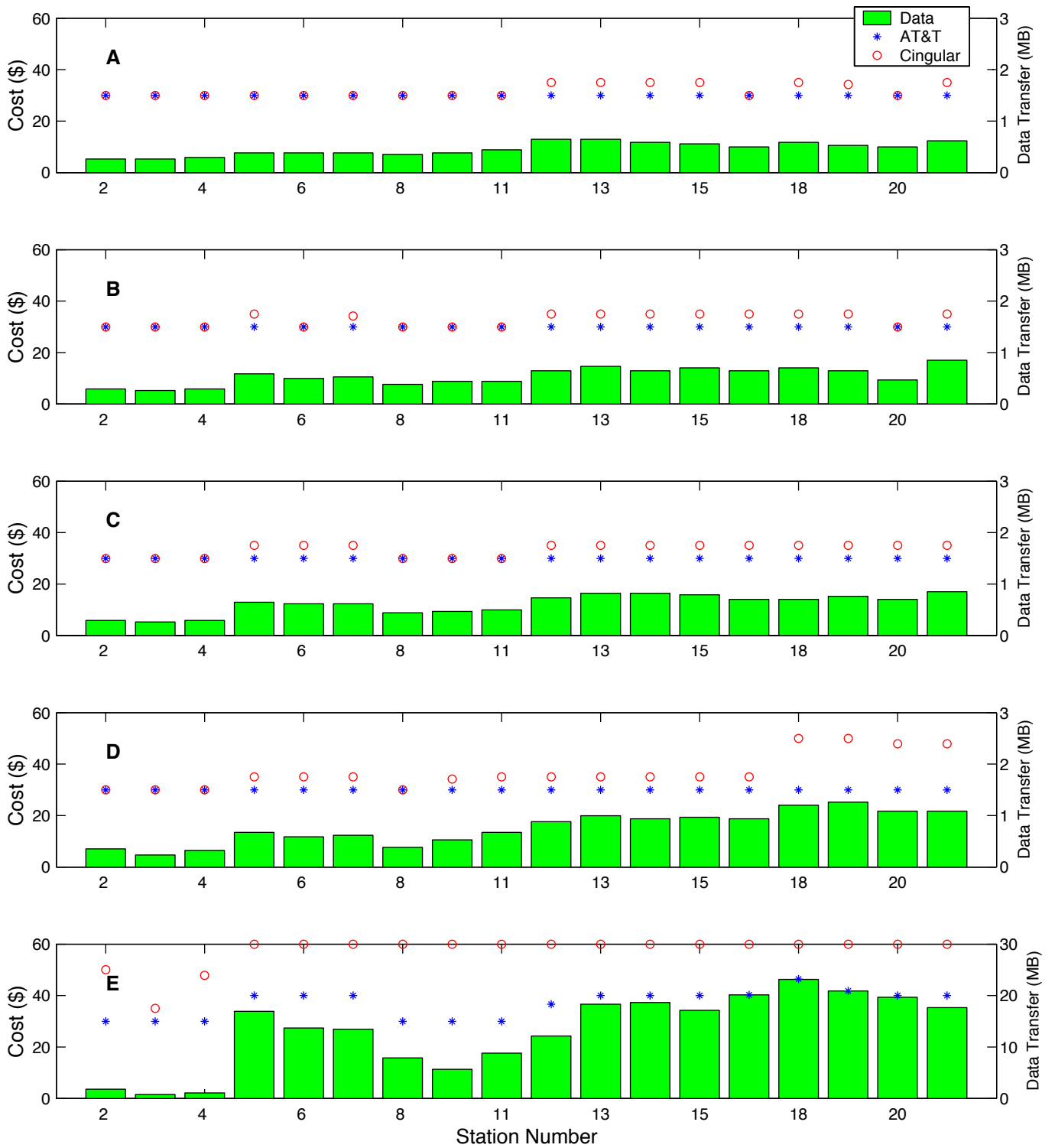


Figure B.8, Estimated average monthly communication cost structure as applied to five weekdays' data collected on 18 stations of I-15_D11 corridor. A-E are for communication modes 1-5, respectively. The plots show the average amount of data transferred by each station per month, and the per station monthly communication cost when using either GPRS/GSM (AT&T) or Mobitex (Cingular) technology. Note the different data transfer vertical scale for plot E (10 times larger than that for plot A-D).

11 APPENDIX C, COMPARISON BETWEEN MOBITEX AND GPRS

According to [30] both GPRS and Mobitex offer "an always on, always connected" service. Mobitex has been specifically designed to carry two-way data, quickly and securely, and not voice. GPRS supports both data and voice transfer as it is overlaid on top of GSM. Mobitex offers a data transfer speed of 8 KB/s. GPRS, on the other hand can offer transfer speeds of up to 38 KB/s (because it operates on four channels). But voice communication requires significantly more bandwidth. So therefore, performance could be constrained by how many users are on the network at any one time. Realistically achievable data transfer speeds could be much less. GPRS is an open technology, whereas Mobitex is a proprietary standard. There seems little difference between Mobitex and GPRS on the pricing side.

Mobitex is particularly well suited for applications where there is need for transferring data from remote sources to central receiving stations. Such applications involve small amounts of data that must be sent at regular intervals or whenever need occurs. This is exactly the traffic pattern for which a packet-switched network such as Mobitex is optimal. Because call set-up time is negligible in a packet-switched network, data transmission in telemetry applications is extremely efficient. Both with Mobitex and GPRS, the user is charged for the amount of data sent and not for connection time.

Mobitex operators provide additional support for telemetry applications. In several countries operators offer special tariffs for applications showing the traffic pattern typical for telemetry applications [35]. For organizations that need to deploy large numbers of radio modems in fixed locations, operators are usually willing to discuss special subscription forms that reduce the fixed costs associated with each station.