Title
Highway Traffic Data Sensitivity Analysis

Authors
Lu, Xiao-Yun
Coifman, Benjamin

Publication Date
2007

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

Highway Traffic Data Sensitivity Analysis

Xiao-Yun Lu,
Benjamin Coifman

California PATH Research Report
UCB-ITS-PRR-2007-3

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.

Final Report for Task Order 5320

January 2007

ISSN 1055-1425

Highway Traffic Data Sensitivity Analysis

Final Report

Project Number: TO5320

Sponsored Through California PATH

By

Caltrans

Project Manager: Dr. Xiao-Yun Lu
Associate Research Engineer
California PATH, University of California at Berkeley
Richmond Field Station, Bldg. 452
1357 S. 46th Street, Richmond, CA 94804-4648
Tel: (510) 665-3644
Email: xylu@path.berkeley.edu
http://path.berkeley.edu/~xylu/xyl_home.html

Associate Professor Benjamin Coifman
Department of Civil and Environmental Engineering and Geodetic Science
Department of Electrical and Computer Engineering
The Ohio State University
Hitchcock Hall 470
2070 Neil Ave
Columbus, OH 43210
Tel: (614) 292-4282
Email: Coifman.1@osu.edu
http://www.ceegs.ohio-state.edu/~coifman/

Key Words

traffic parameter estimation error, sensitivity, loop data, Sensys data, traffic parameter estimation, point estimation, link estimation, travel time delay, shockwave, NGSIM data, vehicle trajectory, shockwave propagation speed, numerical algorithm, congestion onset (incident) detection

Acknowledgements

The work presented in this report was performed as part of the California PATH Program of the University of California under project TO5320, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation (Caltrans). 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.

The guidance, management and support from Mr. Fred Dial of Caltrans Traffic Operations and Mr. Joel Palen of Caltrans Division of Research and Innovation (DRI), are gratefully acknowledged. We also appreciate the contributions and input from Mr. Michael Mauch of DKS Consulting (formerly of PATH and originally an investigator on this study). Finally, we also appreciate the help and contribution from the several students who have contributed to this work, chief among them being Ho Lee.

Contents

Page

Executive Summary

This report is divided into two parts:

- Part I - Highway Traffic Data Sensitivity Analysis in Systems Approach. This work has been conducted by Dr. Xiao-Yun Lu at California PATH, U. C. Berkeley.

- Part II - Empirical and Analytical Results for Sensitivity to Loop Station Spacing. This work has been conducted by Benjamin Coifman in Civil Engineering, Ohio State University.

Since the aspects of the problem considered and approaches adopted by the two are different, it is thus necessary to separate the report into two parts. Each part is self-content.

Part I. Highway Traffic Data Sensitivity Analysis in System Approach

This part describes a *system approach* used for highway traffic data sensitivity analysis. It looks at the input, the output, and the processes between the input and the output and identifies the key factors in the input and processes that would affect the output.

Traffic parameter estimation in surveillance is critical to traffic management and control and traveler's information. On one hand, those applications require high precision estimation of traffic parameters such as mean speed or link travel time. On the other hand, most links do not have continuous measurement like video camera plus real-time image processing. Instead, loop stations, as point measurement sensors, are popularly used on freeways as fundamental measurement. The problem arises naturally: how one could obtain link (continuous) estimation with required estimation error and time delay based on point measurement? It is obvious that, several key factors would affect the estimation error: (a) sensor measurement noise, estimation error and time delay at each station (individual measurement point); (b) data aggregated: method of aggregation and time interval for aggregation; (c) method to get link estimation from point estimation; and (d) the density of the point-measurement stations (loop station distance). Under the assumption that (i) all the estimation methods adopted for a point measure produce the minimum error and time delay; and (ii) sub-second data are used without aggregation, it becomes a sensor configuration (location and density) problem. Suppose loop location problem is solved based on the knowledge of traffic flow. Then the problem is further reduced to: what is the loop density in main lanes to satisfy required link estimation error and time delay?

Highway traffic flow can be divided into four phases from a macroscopic viewpoint: freeway flow, congestion onset and building up, congested static state, and recovering from congestion. The phase that affects the traffic parameter estimation error most is the phase of congestion onset and building up since the traffic flow speed is subjected to fluctuations. Collectively, this presents a wave phenomenon – shockwave. Wave propagation speed is one of the important characteristics which could be used to estimate the delay error from point estimation to link estimation. Based on macroscopic and microscopic observation and analysis of the vehicle-by-vehicle trajec-

tory plots from NGSIM [19] data, a numerical algorithm to identify the wave propagation speed has been developed. The algorithm is created from the following facts: the speed-time trajectory is of 2nd order parabolic type and the corresponding distanced-time curve is of 3rd curves for a each vehicle trajectory. A single shockwave in speed-time or distance-time space is presented as the collection of such curves. Multiple shockwaves are combination of 2nd order parabolic in speed-time space and 3rd curves in distance-time space. Thus the shockwave propagation speed can be identified as the traveling speed of the local minimum of the speed-time trajectories. The algorithm has been implemented in *Matlab* and has been used for processing all the NGSIM [19] data. As the result, an interesting and surprising result shows that all the significant shockwaves have similar propagation speed about 5.1[m/s], or 11.4mph which is independent of the traffic flow speed prior to congestion for *saturated traffic* (flow close or above the capacity or density is high enough). This may suggest tat shockwave propagation speed is mainly determined by driver behavior: perception of time to collision, or speed dependent following distance. This also coincides with intuitive observation from trajectory plots with NGSIM data.

An immediate application of this result to the determination of the upper bound for delay error for link travel time estimation is also considered. This result also presents a way to validate traffic flow theory: the shockwave propagation speed deduced from models should have similar propagation speed for saturated traffic. One possible application is to the Cell Transmission Model [13, 14].

In summary, the main contribution of this part of work includes: (1) looking at point estimation error using sub-second data without aggregation and comparing with ground truth from independent sensors: Sensys and vehicle-by-vehicle trajectory tracking by video camera; (2) dividing the highway traffic into four phases and selecting the most critical one that would affect the traffic parameter estimation error most for consideration; (3) a new congestion onset detection method is proposed compared to all the previous work, which is expected to produce less time delay than previous methods; and (4) based on the new congestion onset detection, a simple relationship between the upper bound of travel time estimation delay error and sensor density has been established, which could be used by traffic engineers to decide how dense the loop station should be on freeways to satisfy the required time delay error threshold on link travel time estimation.

Part II. Empirical and Analytical Results for Sensitivity to Loop Station Spacing

Intelligent Transportation Systems (ITS) rely on sensors to monitor roadway conditions These active controls need to determine reliably whether a facility is free flowing or congested and respond rapidly when facilities become congested. Automatic vehicle detectors only monitor discrete points along a freeway and cannot observe the entire facility (both loop detectors and non-invasive vehicle detectors like radar). Conditions between detector stations have to be inferred from data observed at the stations. So a trade off arises between the spacing between detector stations and the latency resulting from the detector network. Deploying a suboptimal traffic monitoring system could unnecessarily increase the cost or it could lead to a system that cannot respond reliably. This part examines the trade off in traffic control applications by calculating response time and traveler delay. In particular, this study seeks to find how big of an event can be missed and at what cost for a given deployment. We examine several bottlenecks using real

detector data and consider the effect of detector station spacing, including: present spacing, half spacing, and double spacing. Next, since conventional detectors only monitor conditions at discrete locations, the influence area or link assigned to a detector station is examined. Consideration is given to different boundary definitions between links, e.g., a detector may be considered to be at the start of a link or the middle of a link. Different definitions will impact the nature of errors in estimating traveler delay though it should not impact the overall responsiveness to detecting events. Costs are presented in terms of time (vehicle hours of delay per lane) and as a function of spacing. This general approach allows the reader to experiment with different dollar values for time and to use the actual deployment costs in their service area. Since there are numerous applications that use detector data, each with their own specific needs, the flexibility of the metrics is intended to help operating agencies make smarter decisions regarding the deployment of detection options, improving the cost-benefit ratio for detection investments using off-the-shelf products while also guiding development of future detection technologies. Without addressing this problem, operating agencies will continue to deploy vehicle detection following ad hoc guidelines that were typically developed out of convenience rather than any attempt to optimize performance.

The response time is basically a function of the incident or bottleneck location and the speed at which the queue grows. For slower growing queues (2.5 mph) with detector stations spaced 2 miles apart, the average time to detection is over 20 minutes. While faster growing queues (10 mph) with detector stations every third of a mile have a time to detection on the order of a minute. Several incidents were examined in detail to establish the impacts of various parameters on delay estimation errors. While the magnitude of the errors might change from one facility to the next, the relative impacts should stay the same, i.e., absolute errors in delay roughly doubles as one doubles the spacing between detector stations. If one allowed overestimation errors to cancel underestimation errors, the net result would be much smaller. In other words, the simple assumption that conditions at a detector station are representative of an entire link is inherently prone to errors. The analysis suggests a slightly more sophisticated model could be used to better quantify delay if higher precision were desired.

While the response time is independent of the link location, the magnitude of the delay estimation errors definitely does depend on where the links are defined. Putting the boundary between two links halfway between detector stations will generally minimize the expected error, though as discussed herein, some special cases could modify that condition.

Part I

Highway Traffic Data Sensitivity Analysis

in Systems Approach

conducted by

Dr. Xiao-Yun Lu at California PATH
University of California, Berkeley

Chapter 1

Introduction

1.1 Parameter Requirement

For traffic management, control and driver information, there are certain acceptable thresholds for error tolerance. One may give a specification of the tolerance based on the performance expectation. In general, traffic management and control, such as ramp metering, requires more parameters and more stringent threshold than traveler information. According to Caltrans *Traffic Management System Detection Plan (TMSDP)*, Table 1.1 is a list of required parameters and their thresholds.

Table 1.1: Traffic Control Functional Data Needs

USER NEEDS																
	Traffic data from single point									***Information derived from two or more data points***			***Other Data***			
	Presence	Occupancy	Headway	Volume	Speed	Weight	Height	Classification	Vehicle Location	Incident Detec-	Origin & Desti-	Travel Time	No. of passen-	Emissions	Conditions: fog, dust, weather	Sound
Traffic Control	X	X	X	X	X	X	X	X	X	X	X	X	X			
Traveler Information				X	X				X	X	X	X			X	
Incident Management				X	X	X	X	X	X	X	X	X	X	X	X	

The required accuracies listed are all above 95% for traffic control. Such accuracy, although ideal, may be prohibitively expensive to achieve based on current technologies. The reasons are the following:

(i) Sensors have measurement errors: Usually, measurement error of a sensor such as a loop may be higher than 5% even under ideal conditions;

(ii) Fault in sensors: According to a survey of the largest Caltrans districts, detector station availability ranges from a high of 87% to a low of 32% in daily operations. This means that faulty loops are range 13~68 percent varying from District to District, which can significantly affect traffic parameter estimation error;

(iii) There are complicated processes from the point measure at loop station to traffic parameter estimation, which include

- Data passing from loop station to PeMS and/or TMC
- Data pre-processing such as aggregation
- Traffic parameter estimation

It is thus necessary to look at the problem in a system approach, i.e. the input, the output, the processes involved to generate the outputs from the input, and the key factors that would affect the output.

1.2 Objectives

The objectives of this research are of two folds:

(1) To find out what accuracy is required in practical traffic management and control;
(2) To improve in traffic parameter estimation accuracy and to reduce time delay.

Since Caltrans *Traffic Management System Detection Plan (TMSDP)* answers the first question partially in principle, Part I concentrate on the problems related to (2). Furthermore, we are considering general point sensors although inductive loop sensors are often used as examples. As generally recognized, mean speed or travel time are the fundamental traffic parameters. Once either of them has been estimated, other required parameters can be deduced. As is known, loop stations only provide point measurement apart from its measurement accuracy at discrete points. A natural problem arises is how to get better estimation of mean speed or travel time along the link from point measurement, such as loop station. In particular, Part I will present the investigation on the following problems:

(i) The main factors that would affect traffic parameter estimation error/time delay;
(ii) Assuming other factors be fixed, how the loop density would affect the time delay error for link estimation;
(iii) How one could detect congestion onset more quickly based point measurement at discrete loop stations;
(iv) To achieve better accuracy from point measurement to link estimation even in different traffic phases:
  - Free flow
  - Congestion onset and building up
  - Congested steady state
  - Recovering from congestion

Of the four phases, free flow and congested steady state are relatively easier to analyze since the speed distribution along the link can be determined by the speed at the initial point using a dual loop station. However, congestion onset and recovering phases would subject to vehicle speed distribution fluctuation in space along a link and with respect to time. Of the two, the phase of congestion onset and building-up is the more difficult for link speed or travel time estimation. It is noted that, since the link length (the distance between the two dual loop stations) is known, speed distribution along the link and travel time along the link are equivalent concept and will be used interchangeably in the Part I depending on the context.

The estimation error considered in Part I includes two types: absolute error and time delay. As is known, for speed estimation, time delay also contributes to absolute estimation error. This is particularly true if there are speed fluctuations which happen very often for *saturated traffic*. As will be discussed later, the way to properly describe the estimation error should be determined by the traffic situation.

Here, the *saturated traffic* means the traffic flow rate is close to or slightly over the capacity of the highway. Saturated traffic is not necessarily congested traffic, for example, free-flow with density just below the *critical density*.

1.3 Previously Proposed Approach

The proposed study would employ a combination of numerical comparisons of detector data to identify errors, manual validation of detector performance, and simulation to quantify the impacts of errors. It was proposed to work with Caltrans engineers to specifically examine the two applications listed in the RFP, ramp metering and identifying the onset of congestion. As noted in the RFP, the redundant information from adjacent lanes could potentially be used to eliminate or reduce sensitivity to errors. But such an aggregation might obscure features of interest, e.g., if different lanes have significantly different speeds. So the research was proposed to continue the work in this area [6] by quantifying the impacts on operations of an across-lane cleaning. Deviating from the RFP problem statement, it was not assume that a specific detection technology over-counts or under-counts. Most technologies were capable of making either error. Rather, it was suggested to investigate the impacts of these problems independent of the detection technology, and then empirically quantify the performance of specific detectors in the BHL (including an RTMS unit).

After identifying the control conditions for the two core applications, it was suggested to collect archived data from the PeMS and BHL databases. It was also suggested to work with concurrent video and detector data collected in the BHL during the summer of 2003 for detector validation. It was suggested to examine the noise patterns in the data, making use of both manual verification and automated analysis, to quantify the extent that the data are noisy. It was proposed to collect additional field data to supplement the existing resources. After that, it was proposed to use numerical comparisons to examine the sensitivity in terms of absolute error and percent error throughout the range of feasible meas-

urements. After quantifying the measurement errors, the analysis would explicitly examine their impacts on the two core applications of detector data listed in the problem statement.

1.4 Previously Proposed Tasks

Task 1: To discuss with Caltrans engineers to find out their needs. ***Identify the control or decision algorithms***: or more precisely, the input needed for these algorithms (e.g. ramp metering algorithm; onset congestion algorithm) so that the researchers could establish the algorithms' sensitivity to input noise in later tasks.

Task 2: *Establish the frequency, severity and impact of Measurement Errors*: process data from PeMS and BHL to quantify the frequency, severity and impact of measurement errors. It might include manual comparison of synchronized video data and loop data.

Task 3: *Establish the Sensitivity to Measurement Errors*: addresses the fundamental question about how good data need to be, e.g., variability in the real traffic flow conditions may outweigh the impacts of measurement errors. To obtain sensitivity of the parameter w. r. t. measurement error without analysis, empirical method has to be used. Sensitivity analysis may be more direct but models are necessary.

Task 4: *Establish the Impacts of Measurement Errors*: Explicitly examine their impacts on the two core applications: Ramp metering and onset detection of incident. Investigate what are the error tolerances of those algorithms. E.g. algorithm that seeks to identify the onset of congestion using average vehicle speeds may function just fine in the presence of relatively large measurement errors of flow or vehicle count. Clearly, this is algorithm dependent. By comparison, a density based algorithm that compares the difference in cumulative arrivals at successive detector stations to find some critical vehicular density and identify the onset of congestion are very sensitive to systematic over (or under) counting of vehicular flows.

Method of this task: To use empirical results, deterministic and stochastic queuing models, and micro-simulation to quantify the impacts of various levels of measurement errors on the core applications.

Task 5: Documentation of the methodology through quarterly report and yearly report etc. Gain feedback from Caltrans engineers for practicality and transferability of the algorithms.

1.5 Systems Approach

To achieve the objectives previously proposed, systems approach has been adopted for this project in the last six month for the work in Part I. Specifically, the following aspects have been considered:

- What are the input
- What are the output
- What are the characteristics of the system
- What are the processes involved from the input to the output
- What are the key factors that would affect the input
- What are the key factors that would affect the outputs through the processes
- How to quantitatively describe the key factors and their effect on the output

Those points are briefly described below:

(1) The input: This study is focusing on point sensors, particularly the inductive loops. The input is the raw loop data.

(2) The output: The output is the list of required traffic parameters. Particularly, link travel time, flow, speed, and volume, …

(3) The characteristics of the system: Highway traffic may be divided into four phases: (a) free flow; (b) congestion onset and building-up; (c) congested steady state; and (d) recovering from congestion. Traffic parameters estimation in phases (a) and (c) are not difficult for analysis because the traffic flow speed has not much variation. The main difficulty is in phase (b) and (d). Phase (b) is of most interests for traffic management and control and travelers' information.

(4) The processes between the input and the out: To generate the required output, a single or multiple processes are necessary. Those processes are related to the methods (analytical procedures) adopted for measurement, filtering, estimation and prediction of traffic parameters.

(5) The key factors affecting the input: The input data to the system is from the sensors measurement which naturally involves noises and measurement error, which are the key factors that affect the input.

(6) Key factors that would affect the output through the processes: Since the key factors here are related to the methods adopted, the effects of key factors on the output are *method dependent*, which will be investigated in details later.

(7) Quantitatively description of the key factors and their effect on the output: The key factors include those from the input and those in the processes. To understand what their effects are on the output, it is necessary to quantitatively describe such effects. In this way, the sensitivity could be quantified.

The system approach practically adopted is compatible with previously proposed approach in the sense that the systems approach adopted include the tasks previously proposed as a subset.

1.6 Tasks Fulfilled in the Work of Part I

- Identifying the key factors that would affect traffic parameter estimation error using loop detection, which will eventually have significant impact on traffic parameter application: congestion onset detection and ramp metering;

- Generating ground truth by retrieving speed at an arbitrary point from the starting point and within the video camera tracking range for all the lanes based on NGSIM vehicle trajectory data;

- Developing the estimation of instant point speed based on BHL raw loop sub-second data. Here "instant" means no time aggregation has been used. "point" means at the point of sensor (such as loop) location;

- Comparing estimated sub-second loop speed with the generated "semi-ground truth" from NGSIM data;

- Comparing estimated sub-second loop speed with Sensys speed of the some time interval;

- Analyzing shockwave characteristics based on NGSIM data for stretches of freeways Inter-State 80 and US 101 and developing a new algorithm for estimating shockwave propagation speed using the NGSIM data;

- Developing a new congestion onset detection method using speed at dual-loop station, either upstream or downstream, but not necessarily both. This is considered necessary to determine the upper bound for time delay error in link travel time estimation based on terminal detector stations.

It is emphasized that a significant difference of this work compared to previous work is that the speed is estimated directly from sub-second data using filtering techniques instead of aggregating over time. This approach can obviously reduce delay caused by aggregation overtime. For example, several traffic parameter estimation based on loops are aggregated over 5min time interval in [5]. Such an aggregation obviously obscures traffic dynamics and brings time delay to estimation.

The discussion will be spread on in the following three chapters. Chapter Two will discuss from a system point of view the key factors that would affect the output. It is noted that a well established method for sensitivity analysis exists mathematically. However, they rely on a mathematical model from the input to the output and thus will not be adopted here. The comparison of instant point speed with Sensys and video camera trajectory tracking will be presented in this chapter. Chapter Three will be devoted to shockwave analysis, particularly its speed estimation. Chapter four will propose a new approach for Automated Congestion Onset Detect which is essentially based on more accurate instant point speed estimation. This approach is necessary for understanding the link travel time estimation error caused traffic fluctuations presented as shockwaves.

Chapter 2

Highway Traffic Data Sensitivity from Systems Viewpoint

2.1 Document Review

Traffic parameter estimation from sensors is critical for traffic management (such as congestion onset detection), traveler information, performance evaluation (such as travel time) and traffic control (such as ramp metering). Among those traffic parameters, vehicle speed or travel time is the most fundamental parameter. Once vehicle speed and link travel time can be estimated accurately, other parameters can be well deduced if in addition the vehicle count is available. Large quantity of work has been conducted on traffic parameter estimation using loop and other sensors. For loop detection, several problems have been addressed: (1) using single loop and dual loop for mean speed estimation; and (2) loop data cleansing and correction to remedy loop detection or data passing errors.

Speed estimation based on single loop relies on occupancy. An experimental study conducted by Chan and May (1986) showed with field data that the average detector pulse on-times for two longitudinally closely spaced stations could vary by 5-10%, or even higher. However, dual loop station makes it possible to estimate vehicle speed more accurately by using loop on/off time instant. The reports in [26, 24] presented the findings of a comprehensive evaluation of the FSP program at a San Francisco Bay Area freeway sections. More than 276 hours of field loop data were collected “before” and “after” the implementation of FSP. The loop data were processed for the evaluation of the effectiveness of FSP. One of the valuable contributions of this work is very detailed analysis of noise loop dada, its remedy and data mining for traffic parameter estimation. From this work, one could investigate the speed estimation error caused by the noise or error in loop data. Work in [7] considered traffic parameter estimation error based on loop station at fixed points. The error may be caused by many factors, to name a few: loop sensitivity.

The PeMS is the practical implementation of the traffic surveillance which involves all the procedures for sensor measurement, estimation, data correction, analysis and presentation. It also provides a live platform for any level of traffic analysis. As such, large number of references is related to PeMS. Work in [5] presented g-factor method using single loop for mean speed estimation and some deduced traffic parameters for performance measurement using PeMS. Most those efforts on loop detection are essentially to establish or to improve point estimation. Precise estimation of traffic parameters over each link is desirable in practice, however. Parameter estimated at the fixed point need to be extended to link estimation, which is the case for loop stations.

Sun et al [29] considers the estimation of traffic parameters along a link from point measurement such as loop stations using a mixture Kalman filtering which is based on

sequential Monte Carlo method. [9] considers link travel time in two steady state phases. i.e. free-flow, and congested steady state. As discussed before, there is no much speed fluctuations in those two phases, thus vehicle moving distance *x(t)* and collective vehicle speed distribution *v(x)* along the link starting to a point loop station could be resented as

$$\frac{dx(t)}{dt} = v(x)$$

with the initial condition at the starting point to be the point speed estimated for the point loop station. However, for the two non-steady state, i.e. congestion onset and building up and recovering from congestion, this relationship would not hold.

Wang and Liu [30] present some results related to link estimation from point estimation when the traffic is free-flow. As [30] indicated, Adopting higher reference speed at about 60 mph (96 kmph), longer aggregation interval, say 300-second, and shorter loop detector spacing, say 1000-ft. (305 m), would help to produce lower mean absolute percentage error in loop detector-estimated travel time delay along the link, which is as expected. However, once congestion presents, or equivalently, there is speed fluctuations, such link estimation error would be very bad as the authors have recognized. In those cases, simple interpolation of the speed estimation at the two stations does not generate reasonable speed estimation along the link because the nonlinear characteristics of speed distribution along the link.

2.2 Key Factors Affecting Sensitivity

This section is to investigate the key factors that would affect the sensitivity of the traffic parameter estimation error from a system approach. First, the estimation error concerned is defined; second, the estimation error propagating though simple calculation is shown; third, main factors that affect the point measurement error described; finally, the impact of the point estimation error on the link estimation error are investigated.

2.2.1 How to Describe the Estimation Error:

There are basically two ways to look at the estimation error. One is in the sense of statistics approach: (a) assuming it is white noise (zero mean with Gaussian distribution), the magnitude determines its characteristics; and (b) assuming normal distribution, mean and variance determine its characteristics. In this approach, the statistical distribution of the error can be determined by finite approximation of mean and variance based sampling. The other way is to look at the *absolute error* and the *relative error* corresponding to the worst case scenarios. The former is the maximum possible error between the true values and the estimated values evolving overtime; and the latter is simply the ratio of the absolute error and the true value. What we considered in Part I are the absolute error and the relative error. To mention those differences is to keep consistent when we are examining the error propagation from sensor measurement error to traffic parameter estimation error.

Since sensor measurement data are time series, traffic parameters are time dependent as well. Sensor measurement errors may be caused by two factors: absolute measurement error and time delay. Time delay effect is not significant if the measured value is nearly constant; however, it will be more significant if the measured value has fluctuations. Error caused by time delay may be considered as due to phase shift from a frequency domain in signal analysis. For a given time delay threshold, higher frequency and larger magnitude in fluctuation will result in larger estimated error. Time delay may be due to physical characteristics of the sensor, internal signal processing, or external factors such as filtering and estimation process. Data aggregation overtime popularly used for traffic data processing naturally causes time delay although it smoothes out the noise at the same time. What time delay is allowed in practice depends on the usage of traffic parameters. For macroscopic traffic analysis, time period for aggregate could be as large as tens of minutes. For ramp metering, 10~15 seconds may be acceptable. For automatic traffic congestion onset detection, sub-second data and more accurate estimation may be necessary to achieve faster responses than cellular phones.

Traffic flow may be divided into four phases: free flow, congestion onset, congested static state, recovering (from congested to free flow). Since the mean speed trajectory has different characteristics for each phase, the estimation error should be represented in different ways. For free-flow and congested steady state, there is not much speed fluctuations. Thus time delay in the estimation does not play significant role to the error. The absolute speed error – the error between the true speed and that estimated from senor, and relative speed error make more sense. For congestion onset and recovering phases, the traffic flow has negative and positive acceleration presented as flow speed fluctuations. Time delay in the estimation makes a significant contribution to estimation error. Thus, it is more appropriate to use both time delay and absolute error as the indicator. Representing the estimation error in this way is also compatible with travel time estimation.

2.2.2 Error Propagate and Sensitivity

Data are preprocessed, aggregated over time and/or space, used to estimate some traffic parameters such as mean speed, flow, and travel time. Estimation methods vary from one to another. Some traffic parameters are directly estimated from measured data, some are deduced from estimated parameters. In all those processes, the errors, starting from measurement error to estimation error of deduced traffic parameters, are amplified or diminished, depending on the methods adopted. In general, the propagation of the error depends on the following factors:

- Sensor measurement error
- Method used for preprocessing
- Method for data aggregation over time and/or space including selection of time period and space distance for aggregation
- Method used for traffic parameter estimation

Deduced parameter error depends on the way they are calculated from fundamental parameters. Any complicated calculation can be represented or approximated with some of the four basic arithmetic calculations: plus, minus, multiplication and division. It is thus possible to show how the estimation errors of the fundamental parameters propagate in the calculation processes. As an example, let have a look how the estimation error is to propagate through those basic arithmetic calculations and how the sensitivity of the output error can be determined from the input error.

Suppose (x_1, x_2) is measured with error vector $(\Delta x_1, \Delta x_2)$. Then

$$\begin{aligned} \hat{x}_1 &= x_1 \pm \Delta x_1 \\ \hat{x}_2 &= x_2 \pm \Delta x_2 \\ \hat{y} &= \hat{x}_1 \pm \hat{x}_2 \end{aligned} \tag{2.1}$$

Where (x_1, x_2) is input; $y = x_1 \pm x_2$ is output and $\hat{y} = \hat{x}_1 \pm \hat{x}_2$ is the estimation of y.
From (2.1), the estimation of y is

$$\hat{y} = \hat{x}_1 \pm \hat{x}_2 = (x_1 \pm \Delta x_1) + (x_2 \pm \Delta x_2) = (x_1 \pm x_2) + (\Delta x_1 \pm \Delta x_2)$$
$$\hat{y} = \hat{y}_0 + \Delta y$$
$$\Delta y = (\Delta x_1 \pm \Delta x_2)$$

where $\hat{y}_0$ is the estimated value of y. It is clear that the magnitude of the error for the estimation of y can be estimated form that of (x_1, x_2):

$$|\Delta y| = |\Delta x_1 \pm \Delta x_2| \le |\Delta x_1| + |\Delta x_2|$$

What is the sensitivity of output error Δy with respect to the input error $(\Delta x_1, \Delta x_2)$?

$$\frac{\partial(\Delta y)}{\partial(\Delta x_1)} = \frac{\partial(\Delta y)}{\partial(\Delta x_2)} = 1$$

which means that the output error is about the same magnitude as the input error after the processes of addition or subtraction.

Similarly, if $z = x_1 \cdot x_2$ and $w = \frac{x_1}{x_2}$, then

$$\hat{z} = \hat{x}_1 \cdot \hat{x}_2 = (x_1 + \Delta x_1) \cdot (x_2 + \Delta x_2) = (x_1 \cdot x_2) + (x_2 \Delta x_1 + x_1 \Delta x_2 + \Delta x_1 \cdot \Delta x_2)$$

$$\hat{z} = \hat{z}_0 + \Delta z$$

$$\Delta z = (x_2 \Delta x_1 + x_2 \Delta x_1 + \Delta x_1 \cdot \Delta x_2)$$

$$\hat{w} = \frac{\hat{x}_1}{\hat{x}_2} = \frac{(x_1 + \Delta x_1)}{(x_2 + \Delta x_2)} = \frac{x_1 / x_2 + \Delta x_1 / x_2}{1 + \Delta x_1 / x_2}$$

$$= x_1 / x_2 + (x_1 / x_2) \cdot \sum_{i=1}^{\infty} \left(\Delta x_1 / x_2 \right)^i + \frac{\Delta x_1 / x_2}{1 + \Delta x_1 / x_2}$$

$$= \hat{w}_0 + \Delta w$$

$$\Delta w = (x_1 / x_2) \cdot \sum_{i=1}^{\infty} \left(\Delta x_1 / x_2 \right)^i + \frac{\Delta x_1 / x_2}{1 + \Delta x_1 / x_2}$$

where $\hat{z}_0, \hat{w}_0$ is the estimated value of z and w respectively, and $\Delta z, \Delta w$ are corresponding error resulted from the estimation. It is thus possible to determine the magnitude of those error based on the magnitude of $(\Delta x_1, \Delta x_2)$. We can also consider the sensitivity of output errors Δz and Δw with respect to the input errors $(\Delta x_1, \Delta x_2)$.

It is noted that the estimation error propagation is independent of the measurement technology as long as the measurement error can be estimated. The error analysis based on the relationship of parameter calculation, from which, it is possible to figure out the error of the output (estimated value) if the relationship between the fundamental value and estimated value is known.

In general, there is a standard way for conducting sensitivity analysis analytically if the relationship between the input parameter and the output parameter are known. Since this project is to propose an empirical approach, the method for sensitivity analysis will not be discussed further here.

2.2.3 Main Factors Affecting Point Estimation Error

As indicated in the right block of Figure 2.1, the main factors affecting the point estimation error would be: sensor measurement noise and error, estimation method, aggregation method, and data processing method.

2.2.4. Main Factor Affecting Link Estimation Error from Point Estimation

On one hand, sensors used for traffic surveillance are usually point sensors (discrete in space) such as popularly used loop stations in the sense that they provide measurement at a fixed point. Even video camera and radar can only look at a short distance. They are still approximately point sensor compared to the inter-sensor distances on the freeway. On the other hand, the parameters required for traffic management and control such as mean speed are required to be continuous in space. There is a natural problem: how the point measurement or estimation error would affect the estimated parameter error along

the link between the two points. Clearly, the estimation error would also be method dependent. The method used for obtaining the values could be simple (linear or nonlinear) interpolation and/or extrapolation, prediction using Kalman filtering, or other method.

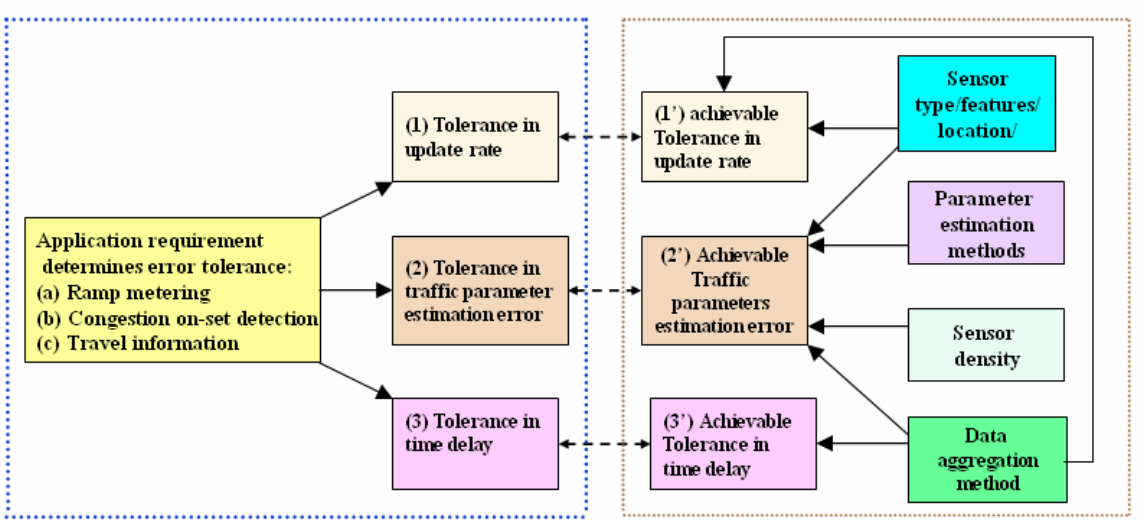


Figure 2.1. Left block: acceptable error tolerance for traffic parameters; Right block: Factors affecting point estimation error

It can be seen that the output error is dependent on the methods used and is evolving with technology development. Currently specified tolerance for one method may not be appropriate for another method or for the technology in the future. From an analytic viewpoint, it is essential to investigate the way the error propagates through all those procedures to obtain a relationship between sensor measurement error and traffic parameter estimation error. Such relationship, or a method to quantify such propagation, once established, can be used for any point estimation methods and new specification of the tolerance. However, this topic is out of the scope of this project. The Figure 2.2 shows qualitatively the relationship.

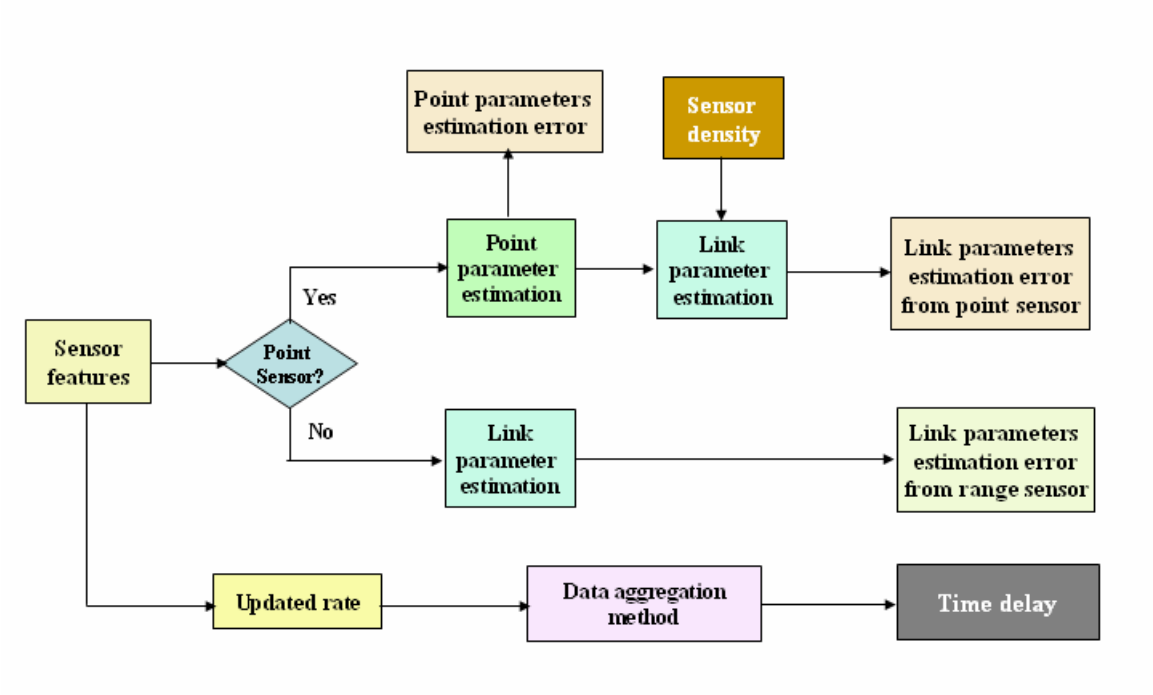


Figure 2.2 Factors Affecting Link Estimation from Point Estimation

2.3 Instant Point Speed Estimation

From the discussion above, it is clear that, estimation error at point sensors make critical contribution to the parameter error. Traditional methods for data processing to estimate traffic parameters are statistical in nature. Aggregation overtime is popularly used for traffic parameter definition. This is mainly inherited from the macroscopic traffic model and concept. Data aggregated overtime has several advantages from statistics viewpoint:

- Less fluctuation
- Smoother
- Closer to statistical mean
- Data error or noise absorbed or reduced through aggregation sine moving average itself is a low pass filter

For application, practical requirement may allow the data to be aggregated over different time periods. Aggregated traffic data over several minutes are adequate for planning, driver information, and operation in the level of performance in the past. However, for high level of performance in traffic management and control, such as congestion onset detection and ramp metering, it is necessary to measure the traffic more accurately with as less time delay as possible since time delay itself bring error to the estimation. This is the impetus to introduce the concept of *instant point speed*.

Instant point speed means that the traffic flow speed is estimated at a given point detect station using sub-second data plus filtering techniques without aggregation overtime. The

estimated single vehicle speed at discrete time points are interpolated-extrapolated to obtain continuous speed curve overtime. Filtering processing with less resulted time delay may be used for smoothing the speed curve interpolated estimated from different vehicles and different time point but the same location in space.

2.3.1 Instant Point Speed Estimation Using Loops

Traffic point speed estimation at dual loop stations has been revisited using sub-second data. The reason to do so is that dual-loop station provides more accurate vehicle speed measure than single-loop station. It is thus possible for aggregation time period to be significant reduced or removed. The objective to use the described method below is to achieve more accurate estimation in the following senses:

- Less time delay
- Lower absolute estimation error

For this purpose, sub-second data has been used without aggregation to reduce time delay. However, data losses in PeMS sub-second data are not necessarily evenly distributed due to several causes such as loop fault and data loss in media from the control cabinet to the PeMS server. The vehicle speed at each station is estimated using received up and down time instants of the two loops upstream-side and downstream-side of the loops respectively. Then they are averaged over the two time points. Because vehicles passing a loop station is a discrete event, the estimated speed at the station are extrapolated over time to get continuous speed-time curve at the station as shown in the following Figure 2.3.

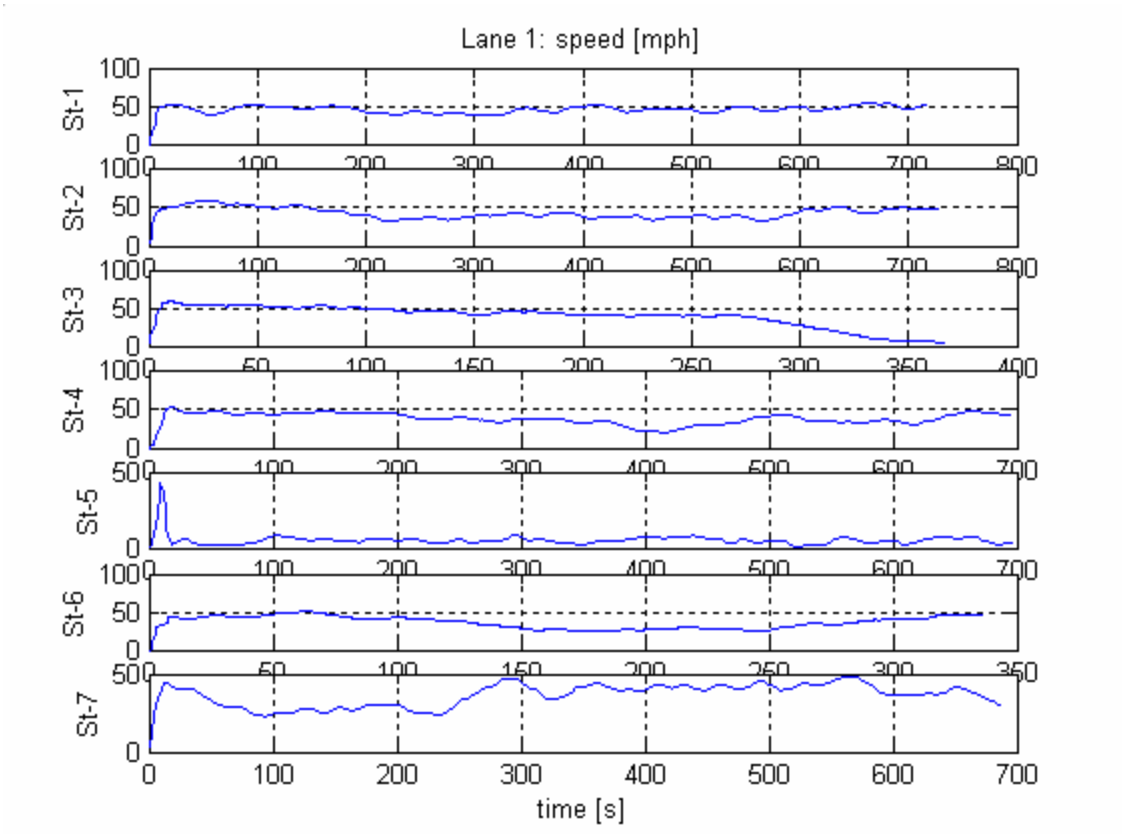


Figure 2.3. Instant point speed estimation at the 7 stations in lane 1 based on sub-second data; Data source: 04/13/05 4:00pm East Bound on I-80 from BHL (Berkeley Highway Lab)

2.3.2. Comparison with NGSIM Data

To understand the error, it is necessary to compare the estimated value with ground truth. Since NGSIM dada provides vehicle-by-vehicle trajectories based on video tracking, it can reasonably represent the ground truth. NGSIM East Bound vehicle trajectory data are used for comparison. The relative location of loop stations and the video camera coverage area is shown in Figure A.5 in Appendix A. The maximum tracking distance is about 518m in the field of view of the camera, which has not reached the Station 7 yet. Instead, loop station 7 is located about 50m downstream of the traffic flow. The speed at the point 500m where the loop is located has been extracted and plotted with speed estimated based on the BHL Loop Station 7 data. One of the trajectory plots for comparison is shown in Figure 2.4 with others listed in Appendix A. Vehicle speed trajectories from video of the NGSIM data are rather noisy. This is because vehicle tracking using video camera is distance based. The speed trajectories are simply calculated for distance trajectories by numerical differentiation. This is the reason why linear filter will be used to smooth them up in the analysis of the next chapter. For lane 1, these two estimates are matched reasonable well. However, we do not calculate the absolute error and relative error since the video speed trajectories need to be improved to be qualified as ground truth for quantitative comparison. This could be a research topic in the future.

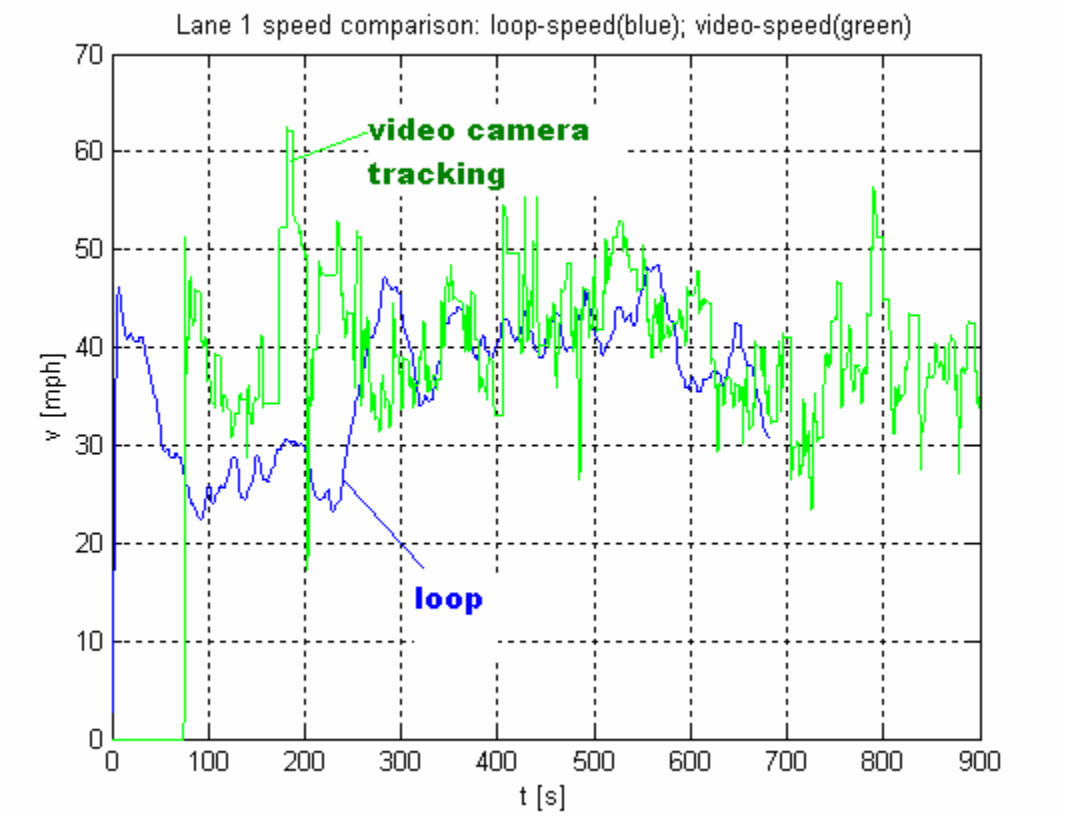


Figure 2.4. Comparison loop speed and video camera speed at station 7 in Lane 1

2.3.3 Comparison with Sensys Data

The speed estimation at loop station has also been compared with that estimated from Sensys sensor which is new magnetic sensor with wireless capability for data passing [31]. Two Sensys studs were put serially with one of them put right in the centers of loop stations in both lane 1 and lane 2.

The comparison of the speed estimation of loops and Sensys sensors show that they are pretty consistent. However, there is a significant time lag for the loop speed compared to Sensys speed. This may be caused by: (a) the time stamp of the loop and that of the Sensys sensors are not well synchronized; (b) the inductive loops may have larger time delay compared to Sensys sensors measurement and estimation; or (c) the combination of (a) and (b).

For the same reason, for the Sensys data to be qualified as the ground truth, it needs to be validated independently.

Figure 2.5. Comparison of speed estimates based on dual loops at Station 7 and dual Sensys sensors; Data source: 12/15/05 9:00am Lane 2 of East Bound I-80.

2.4 Concluding Remarks for Chapter 2

This chapter investigates the sensitivity from systems approach. It identifies the key factors that would affect the estimation error at a point detector station and the key factors that would affect the output error, which are basically process dependent. Those processes are the relationship between the input and the output. As pointed out, although standard sensitivity analysis method exists, they rely on the mathematical relationship between the input and the output. To reduce time delay in traffic speed estimation, a concept of instant point speed has been introduced. The objective is to reduce time delay in parameter estimation by remove data aggregation overtime which itself introduces delay error to the estimation. Qualitative comparison of the estimated instant point speed with both the speed trajectories from NGSIM data (video trajectory tracking) and the estimated trajectories from Sensys sensors have been conducted. Since the later two are just independent sensors which have not been justified as ground truth yet, it would not make sense to calculate the absolute error and relative error at this stage.

The work related to this chapter remains to be done in future phase of the project:

(1). To adopt a sensor fusing approach for highway traffic surveillance: Traffic surveillance is a complicated issue. The requirement for parameters in traffic management, control and traveler information is getting more stringent than ever. Point sensors such as loop detectors and short range detector such as radar and video camera have their own

advantages and disadvantages for traffic surveillance. Sensor fusion approach, which is widely used in defense industry for vehicle/object detection, can be adopted to optimally combine the merits of different sensors in location, position, density to achieve the best performance from the overall system point of view.

(2). To investigate further the time lag between the loop data and that of the Sensys sensors. Since the time lag of the loop data is significant, it likely caused by adrift in clock of the PeMS system since the time stamp is recorded in control cabinet, time used for data passing would not affect the delay at all.

(3) To validate a round truth for comparison, either the Sensys sensors or vehicle speed trajectory tracking using video camera. Once this is done, quantitative comparison of the loop estimation and the ground truth can be conducted.

Chapter 3

Freeway Traffic Shockwave Analysis

3.1 Introduction

Shockwave has been generally recognized as the characteristics of highway traffic flow in the phase of congestion onset and building-up. Due to the practical difficulty in tracking individual vehicles, most previous work conducted on shockwave characteristics concentrated on analysis based on the wave models for highway traffic. Some experimental work reported in [10] was based on tracking using video camera for individual vehicles in two lanes on inter State I-680, but there was no analysis conducted there.

Recently released NGSIM data provide a rich source for highway traffic characteristics analysis. Vehicle trajectories from digital video camera tracking from stretches of freeway at Berkeley Highway Lab on Inter-State I-80 (April 13, 2005, PM peak hour from 4:00pm) and US 101 (June 15, 2005 AM peak hours from 7:50am) are used for shockwave analysis. One of the main contributions of this part of work is to develop a numerical algorithm to identify shockwave propagation speed based on consecutive vehicle time-distance trajectories in each lane. It is not difficult to understand that all the time-distance trajectories are 3rd order curves or combination of such curves. Accordingly, all the time-speed trajectories are 2nd order curves or combination of such curves. Shockwave happen at the collective valley points or local minimum of time-speed curves of multiple vehicles. Shockwave propagation speed can thus be estimated based on those considerations. A numerical algorithm for after processing has been developed to identify the shockwave propagation speed with all the NGSIM data. It has been found that all the significant shockwaves have similar propagation speed which is about 5.1[m/s]. Since all those data were from peak hours, it is suggested that shockwaves for saturated traffic has similar characteristics which is independent of the traffic mean speed. This implies that the characteristics for saturated traffic are dominated by driver behavior: acceptable following distance at different mean traffic speed (or perceived time to collision).

The following nomenclature is used throughout this chapter:

j – vehicle or trajectory index (each vehicle corresponds to a trajectory) in a consecutive order; two vehicles are consecutive when (a) they are in the same lane, and (b) vehicle $j+1$ is behind vehicle j;

i – the index of shockwave: There may be more than one shockwave appeared in one data set; and each vehicle may be involved in more than one shock wave;

t_k - discrete time points for $k = 0,1,2,...\mathrm{K}$; synchronized for all the vehicles;

$\Delta t = t_{k+1} - t_t$ is a constant of 0.1 sec – data sample period;

k - the index of time step;

$t_{k_i}^{(j)}$ - time points corresponding to the local minimum of the speed-time trajectory for vehicle j;

$y_j(t_k)$ - distance of vehicle j at time point t_k ; all the vehicle trajectories starting from the same point;

$v_j(t_k)$ - speed of vehicle j at time point t_k ;

It is noted that, in NGSIM data, since each vehicle as a speed and distance trajectory. Thus vehicle and its trajectory are equivalent in discussion.

3.2 Characteristics of Shockwave

3.2.1 Macroscopic Observation

The following concept can be observed from the speed-time and distance-time plot as in Figures 3.1:

- Shockwave propagation speed: the tangent of the shockwave in distance-time plot
- Different phases of traffic can be observed from the speed envelops and distance-time diagram as shown in Figure 3.1 (Data source: NGSIM, US101 07:15 - 08:05am):
  - Free flow
  - Congestion onset and build up -- **a**
  - Congested static state -- **b**
  - Recovering -- **c**
  - Congestion (or Shockwave propagation) duration **-- d**
  - Congestion (Shockwave propagation) influential distance -- **e**

Comparing the corresponding speed-time and distance-time plots as in Figure 3.1 and Figure 3.2, it can be observed the correspondence between the speed drops in the multi-vehicle speed envelop and the shockwave in the distance-time plot. Besides, the significance of the speed envelop drop (wave energy) is approximately proportional to the Congestion (Shockwave propagation) influential distance.

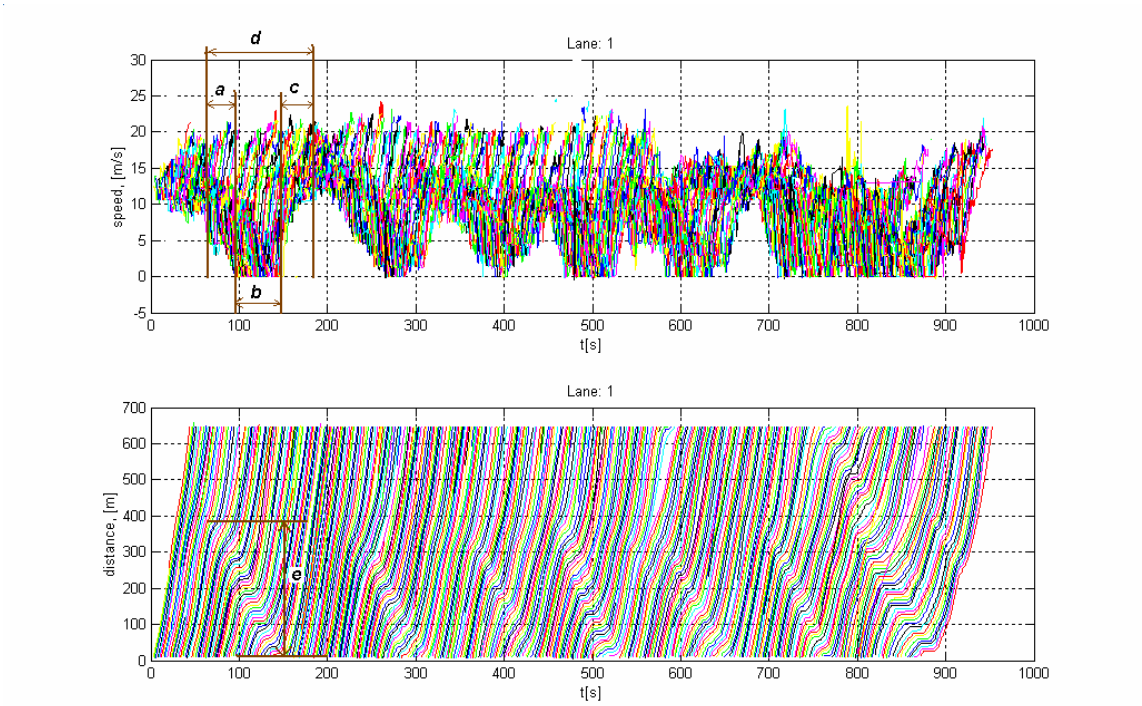


Figure 3.1 Macroscopic shockwave characteristics

3.2.2 Microscopic Observations

Macroscopic Observation:

According to the observation in [10], the shockwave propagation speed is about 12.5 [mph] for the two lanes in consideration during the time period. If we look at all the plots form NGSIM data as in the Appendix, it is observed that, all the significant shockwave have the similar tangent in the distance-time plot. Since all the NGSIM data are collected an AM peak or PM peak periods, this would suggest that **all the significant shockwaves propagate at similar speed** for saturated traffic. However, it would be more rigorous if a numerical algorithm can be developed for identifying the propagation speed, which is the main purpose of this chapter.

Microscopic Observation:

To analyze the characteristics of the wave, it is necessary to look into details of the peed –time and distance-time trajectories. For this purpose, it would be helpful to observe vehicle trajectories related to shockwave. The vehicle speed and distance trajectories related to single wave are shown in Figure 3.2. From those figures, it is clear that

- The speed-time line is 2nd order parabolic type;
- The space-time line is 3rd order curve;

The vehicle trajectory related to multiple waves (Figure 3.3). It can be observed that the speed-time and space-time trajectories for multi-waves correspond to combinations of 2nd order and 3rd order curves respectively.

From Figure 3.2 and Figure 3.3, it could be observed further that:

- Each shockwave leads to a valley for the speed lines: large speed valley corresponding to deep valley and small relative speed reduction corresponds to shallow valley in speed line; The relative depth of the valley represents speed reduction or effects of shockwave; Thus, such depth could be used to classify the shockwave: long term congestion or short term fluctuation.
- In any case, each valley corresponds to at last one local minimum. There could be uncountable local minimum if speed keeps the same in the valley, but there is a unique minimum which corresponds to the earliest time.
- The propagation of those unique minimum determines the shockwave propagation speed.

Based on those observations, the numerical algorithm for identifying the shockwave speed can be describes as follows: The algorithm is to search for those unique local minimum for each speed-time trajectories. Then find out how those local minimums for each trajectory propagate in space-time, which gives wave propagation speed.

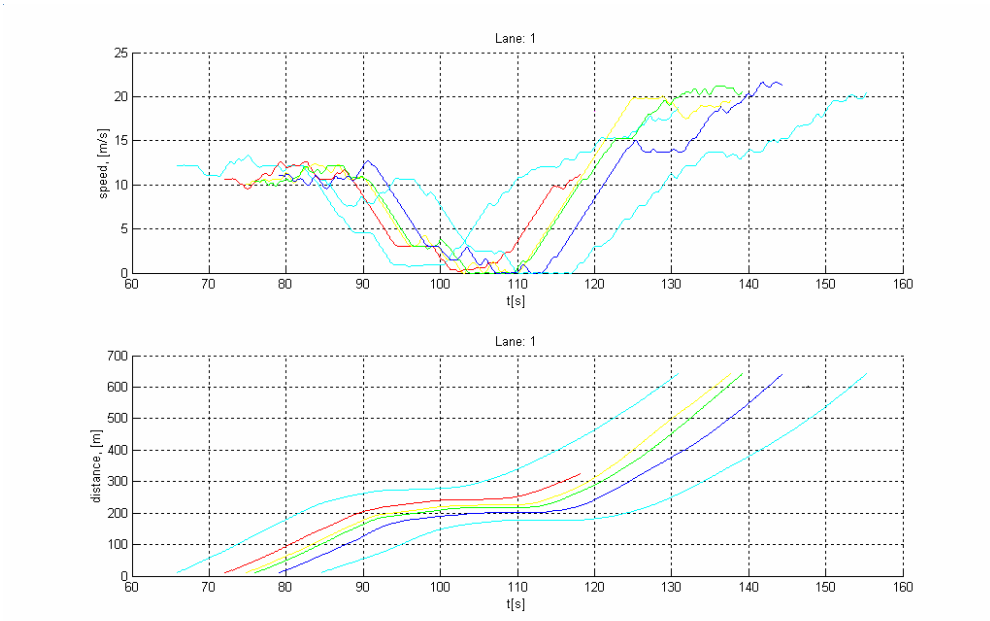


Figure 3.2. NGSIM data, US-101: Single wave propagation

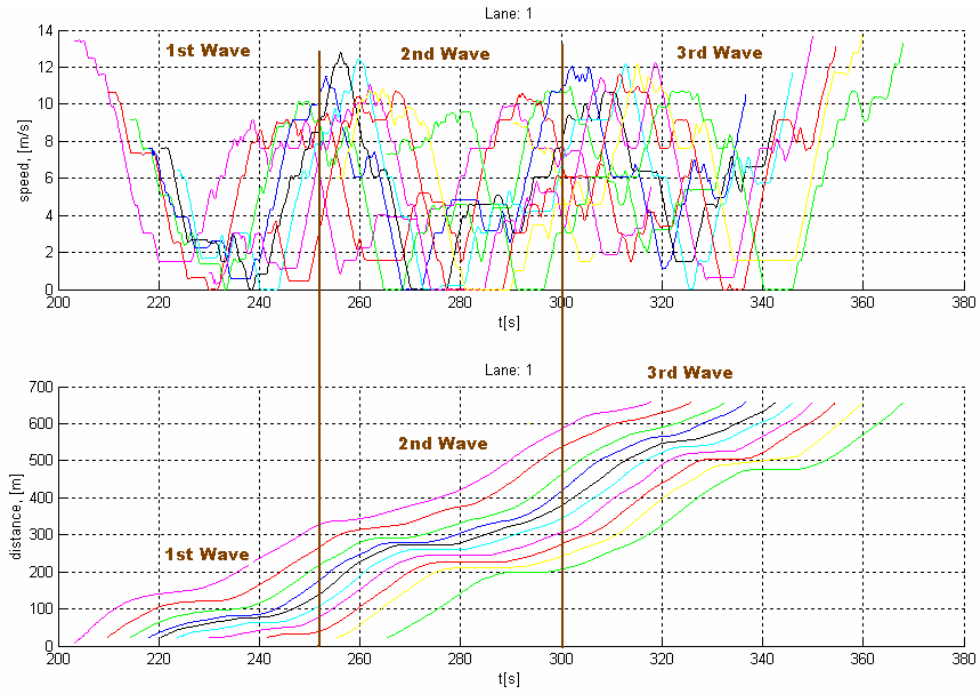


Figure 3.3. NGSIM data, US-101: Combination of 3 waves

3.3 Shockwave Analysis Algorithm

For each vehicle *j*, suppose a time and distance coordinate (in a global inertia coordinate system) for the unique local minimum point corresponding to a speed drop is $\left(t_{k_i}^{(j)}, y_j(t_{k_i}^{(j)})\right)$. The time points for minima of different consecutive trajectories corresponding to the same shockwave are not necessarily equally distributed. i.e. $t_{k_i}^{(j+1)} - t_{k_i}^{(j)}$ is not necessarily constant for all the vehicles involved in the same shockwave. However, there should hold the following relationship for any two consecutive vehicle j and vehicle $j+1$:

$$\begin{aligned} & t_{k_i}^{(j+1)} > t_{k_i}^{(j)} \\ & y_{j+1}(t_{k_i}^{(j+1)}) < y_j(t_{k_i}^{(j)}) \end{aligned} \tag{3.1}$$

which is a necessary condition for a shockwave appearing at congestion onset. This is because the vehicle behind begins to reduce speed at shorter distance relative a common starting point. It is impossible that

$$\begin{aligned} & t_{k_i}^{j+1} > t_{k_i}^{j} \\ & y_{j+1}(t_{k_i}^{(j+1)}) \geq y_j(t_{k_i}^{(j)}) \end{aligned} \tag{3.2}$$

The shockwave propagation speed determined from two consecutive vehicles is estimated as:

$$V_j^{(i)} = \frac{y_{j+1}(t_{k_i}^{(j+1)}) - y_j(t_{k_i}^{(j)})}{t_{k_i}^{(j+1)} - t_{k_i}^{(j)}} < 0 \tag{3.3}$$

where $V_j^{(i)}$ - shockwave propagation speed deduced from trajectories *j and j+1*

The negative sign of $V_j^{(i)}$ indicates that the shockwave is back-propagating. To implement this algorithm, we need not only the vehicle moving distance but also the start tracking point to get the position of the vehicle with respect to the inertial coordinate system at any time.

Suppose there are J_i vehicles involved in the *i-th* shockwave. Then the *average shockwave propagation speed* is defined as:

$$V^{(i)} = \frac{1}{J_i - 1} \sum_{j=1}^{J_i - 1} V_j^{(i)} \tag{3.4}$$

where $V^{(i)}$ - the *i-th* shockwave speed identified; It is the average of the $V_j^{(i)}$ overall the vehicles involved.

3.4 Numerical Implementation for Shockwave Algorithm

The implementation of the above described shockwave speed estimation algorithm requires to find appropriate minima of the speed trajectories and their corresponding time and distance points. The trajectories from NGSIM data on I-80 and US-101 are used for this purpose. Both the speed and distance trajectories have been estimated from digital video camera tracking of the vehicles along a stretch of the freeways. They have significant noises which causes difficulties to directly searching the local minima of the speed trajectories. To avoid this difficulty, four steps are adopted to obtain the estimation: (a) filtering to smooth the speed-time trajectories; (b) searching local minimum of the smoothed trajectories; (c) Clustering the time and distance points corresponding to those local minima according to some criteria; and (d) calculating the average shockwave propagation speed from the distance-time points corresponding to the clustered local minima.

3.4.1 Vehicle Trajectory Filtering

The vehicle trajectories from NGSIM data have disturbances which come from measurement noises and estimation error. Those noses are reflected as many artificial local minima which are not true minima to be used to represents the shockwave characteristics.

Numerically, they causes problem for search true minima. To relieve this, a low pass filter and a rate limit filter are used to smooth the trajectory somehow to reduce those irrelevant local minima coming from estimation. The following low pass (Butterworth) linear filter is used for filtering purpose:

$$
\begin{aligned}
x_1(k) &= 0.4320x_1(k-1) - 0.3474x_2(k-1) + \ 0.1210v_{in}(k) \\
x_2(k) &= 0.3474x_1(k-1) + 0.9157x_2(k-1) + \ 0.0294v_{in}(k) \\
v_{out}(k) &= 0.4984x_1(k) + 2.7482x_2(k) + \ 0.0421v_{in}(k) \\
k &= 1, 2, ...
\end{aligned}
$$

where $(x_1(t), x_2(t))$ is the filter state variable. v_{in} is the input speed signal and v_{out} is the output - filtered speed. The initial condition is chosen as $(x_1(0), x_2(0)) = (0, 0)$. Figures 3.2 and 3.3 show smoothed speed-time and distance-time trajectories using the above methods.

3.4.2 Numerical Algorithm for Searching Local Minimum of Speed Trajectories

To study the characteristics of shockwave, it is expected to obtain as many as possible the time and distance points corresponding to the true local minimum which is related to speed drop due to shockwave. However, the speed curve deduced from the distance measurement in tracking using video camera is not smooth at all: noises also produce many smaller local minimum traps. This also suggests the necessity of filtering the speed trajectory before searching the required local minimums. Many methods exist [2, 25] for search local minimum such as Newton Gradient Method, Steepest decent method for continuous functions and Genetic Algorithm for Discontinuous functions or data set. They are not suitable for the purpose here. The numerical method need to consider the following problem characteristics:

(a) It is after processing and thus the information at any time point cold be used;
(b) The algorithm should be able to move out of the traps due to noise in speed trajectory;
(c) It is only necessary to catch the appropriate minimum - speed drop exceeding some threshold. This is because (i) only shockwaves with significant magnitude could affect the traffic flow; and (ii) lower speed drops ,or insignificant shockwave should be filtered out if possible instead of get trapped there;

The algorithm can be called Moving Window Minimum Searching. This algorithm is to search the local minimum of the speed trajectory for any vehicle j which may be related to some shockwave i. If speed is constant for certain period of time, for example the speed is zero, or vehicles are completely stopped, there are uncountable number of minimum. For example, the vehicle speed is zero for a period of time. In this case, the first minimum is registered, which is a unique minimum $v_{k_i}^{(j)}$ for each speed valley. It corresponds to a unique time point $t_{k_i}^{(j)}$ and unique distance point $y_j\left(t_{k_i}^{(j)}\right)$.

The main idea of the algorithm is as follows:

(i) The width L of the window is specified as the number of speed-time points to be searched.
(ii) After each progressive search, moving one step forward;
(iii) Initial search: Searching incremental time intervals before it goes over the whole window width;

The algorithm can be described as follows:

For $j = 1: J$
  While $t_k < T$
  $k=1$
  If $k < L$
    $v_{j_i} = \min\left\{v_j(t_1), v_j(t_2), ..., v_j(t_k)\right\}$
    If t_{k+s} is the first time point among $\{t_k, ..., t_{k+L}\}$ such that $v_{j_1} = v_j(t_{k+s})$, then register t_{k+s} as $t_{k_1}^{(j)}$ and also $y\left(t_{k_1}^{(j)}\right)$
    $i=1+1$
  Else
    $v_{j_i} = \min\left\{v_j(t_k), v_j(t_{k+1}), ..., v_j(t_{k+L})\right\}$
    If t_{k+s} is the first time point among $\{t_k, ..., t_{k+L}\}$ such that $v_{j_1} = v_j(t_{k+s})$, then register t_{k+s} as $t_{k_1}^{(j)}$ and also $y\left(t_{k_1}^{(j)}\right)$
    $i=i+1$
  If Loop End
  $k=k+1$
  While Loop End
  $j=j+1$
For Loop End

It is noted that, if the window width is specified shorter, more local minimum other than true minimum related shockwave will be registered. If it is specified too long, some true local minimum related to shockwave may be missed. In this study, it is specified as 20 [s] or 200 time steps. Physically, this means that shockwave lasts less than 20 [s] may be ignored.

3.4.3 Clustering of Local Minimum

The local minimums registered as above belong to each trajectory. It is necessary to cluster them to form a chain of minimums for a shockwave. It is clear that the clustered minimums should satisfy the following criteria for to be included in a cluster:

- Two consecutive minimums and their time points should have the same sequential order, or:

$t_{k_i}^{(j+1)} - t_{k_i}^{(j)} > \delta$

- Two consecutive minimums should come from different but also consecutive speed trajectories;
- Corresponding distance point should satisfy the condition (1), i.e.

$$y_{j+1}(t_{k_i}^{(j+1)}) < y_j(t_{k_i}^{(j)})$$

- The following conditions are satisfied:

$$-V_{\max} \le \frac{y_{j+1}(t_{k_i}^{(j+1)}) - y_j(t_{k_i}^{(j)})}{t_{k_i}^{(j+1)} - t_{k_i}^{(j)}} \le -V_{\min} \tag{3.5}$$

where $\delta > 0$ is pre-specified time threshold; $V_{\min}, V_{\max}$ are possible minimum and maximum shockwave propagation speed respectively. In the practical calculation, the following values are used:

$$\delta = 0.1[s]$$
$$V_{\min} = 1.0[m/s]$$
$$V_{\max} = 10.0[m/s]$$

The algorithm for building the clusters can be described as follows:

Step 0: Set $j = 1$, $i=1$ and Initialize cluster 1 with $\left(t_{k_i}^{(j)}, y_j(t_{k_i}^{(j)})\right)$

Step 1: *Set* $j = j+1$

Step 2: Take a tentative point $\left(t_{k_0}^{(j)}, y_j(t_{k_0}^{(j)})\right)$ from trajectory j and check if $\left(t_{k_0}^{(j)}, y_j(t_{k_0}^{(j)})\right)$ and $\left(t_{k_i}^{(j-1)}, y_{j-1}(t_{k_i}^{(j-1)})\right)$ satisfy the condition in (1.5) for all the previous clusters set up so far;

If YES: register $\left(t_{k_0}^{(j)}, y_j(t_{k_0}^{(j)})\right)$ as the follower of $\left(t_{k_i}^{(j-1)}, y_{j-1}(t_{k_i}^{(j-1)})\right)$ in the same cluster;

If NO: Initialize a new cluster with the first minimum point of the second trajectory:
Set $i=i+1;$
$\left(t_{k_i}^{(j)}, y_j(t_{k_i}^{(j)})\right)$; Notice that the subscript of k has been increased by *1*;

Step 3: Remove the tentative point $\left(t_{k_i}^{(j)}, y_j(t_{k_i}^{(j)})\right)$ from the minimum set of trajectory j

Step 4: Repeat Step 2

Because there are only finite number of trajectories and each trajectory contains finite number of minimums. This process will stop after all the minimums have been put in some cluster. Each cluster has at least one point.

3.4.4 Shockwave Propagation Speed Estimation.

Once the clustering has been achieved, the length of each cluster may vary. To represent a shockwave, it is required that the cluster length should have a lower bound, say *N*, which is he number of vehicles (trajectories) involved in the same shockwave in deal case. In practice, this number is much smaller and the number of clusters is larger than the number of shockwaves. This happens simply because the trajectories are noisy due to sensor measurement and estimation error. This could be explained as one shockwave has been broken into several pieces. Fortunately, this would not affect much for shockwave speed estimation, however.

Then (1.3) and (1.4) are used calculate the wave speed for each cluster. If the are *N* points in a cluster, there would be *(N-1)* wave speed estimation for the cluster. Average the *(N-1)* to get wave speed estimation with respect to the cluster. This algorithm has been applied to all the NGSIM data sets. The result shows that all the shockwaves have similar propagation speed at 5.1 [m/s] (or 11.4 mph) which is independent of the traffic flow mean speed.

3.5 Concluding Remarks

This chapter considers shockwave characteristics from a microscopic viewpoint using vehicle-by-vehicle speed-time and distance-time trajectories of NGSIM data along two stretches of freeways on I-80 and US-101. Those data allow one to analyze the shockwave characteristics in very much detail. Some basic traffic concept such free-flow, congestion onset, building-up, congested steady state, and recovery from congestion can have a clear view from the trajectory plots. Detailed observation shows that a single shockwave can be described as: (1) a collective speed-time trajectories which are 2^{nd} order (parabolic) curves; or (2) a collective of distance-time trajectories which are 3^{rd} order curves. Consecutive shockwaves are combination of those curves. Based on this intuitive observation, a numerical algorithm for identifying shockwave propagation speed has been developed. This algorithm has been applied to all the NGSIM data sets. The result shows that all the shockwaves have similar propagation speed of 5.1 [m/s] (or 11.4 mph) which is independent of the traffic flow mean speed. This may suggest that shockwave propagation speed for saturated traffic is determined by human driving behavior, or equivalently the speed dependent average safe following distance perceived by the driver. However, this needs to be justified in the future work from human factors perspective.

Macroscopic observation of the shockwave plots from NGSIM data in the Appendix shows that all the shockwaves has similar tangent, or similar propagation speed. This is in agreement with analysis using numerical algorithm developed in this chapter.

This result has been used to estimate upper bound for time delay in travel time estimation caused shockwave characteristics in next chapter. Other application of this result to other areas in traffic analysis needs further investigation. A possible application would be to validate microscopic traffic models incorporated with driver behavior: the shockwave speed deduced from it should be nearly constant and close to 5.1[m/s].

Chapter 4

Congestion Onset Detection and Travel-Time Delay Estimation

4.1 Introduction

An incident may cause traffic congestion but congestion may be caused by other factors such as weather, road geometry and traffic flow rate going over capacity. It has been clamed that cell phone report is faster than automatic congestion onset detection. Of course, this will not apply to the cases when the congestion is not caused by incident. From the author's point of view, the reasons for the delay in traffic congestion onset detection may be of multiple folds: (1) there no efficient method without using aggregated data. Since data aggregation over time naturally leads to time delays, if the detection method can use sub-second data, time delay would be greatly reduced; (2) data or information routing: Cell phone could directly reach CHP (California Highway Patrol) but loop data has to be aggregated and sent to TMC for processing, and then if a congestion is detected, the CHP will be informed. Those processes also causes time delays; (3) For detection method: All the previous methods the authors are aware of need to use both upstream and downstream loop station data, which are usually 500m apart. This means that the shockwave has to reach the upstream station and the discharge wave has to reach the downstream station before it can be detected. This delay plus other delays mentioned before, might be the cause of the actual delay encountered in traffic congestion onset detection. Besides, if the loop station does not work properly which is the case according to [3] faulty loop station is over 60% in some districts. In such a situation, there is no hope to use loop sensors for congestion onset detection.

Here we are not trying to argue that the Automatic Congestion Onset Detection (ACOD) should be developed simply for traffic management purpose. Instead, it is argued that, for link travel time estimation based on the detailed analysis of shockwave propagation, it is necessary to have ACOD at any point/range detect stations. The reason is simple: It would be difficult for *manual* incident report by Cellular phone to be integrated into *automatic* link travel time estimation based on two terminal loop stations. This may be the justification to reconsider the ACOD in new approaches which could potentially lead to more precise congestion onset detection with much less time delay. In fact, with the development of sensor and wireless technologies and detection methods, ACOD will be greatly improved. This is the main impetus for us to discuss this problem from a completely new viewpoint.

This chapter will present an algorithm which is based on either upstream loop station or downstream loop station, but not necessarily both. For reducing induced time delay, sub-second loop data will be used. However, this method needs further validation using appropriate data. Due to the time and resource limit of the project, this cannot be fulfilled in

the current phae of the project. If this method is effective, based on the results about the propagation speed of the shockwave, it is able to deduce the upper bound for possible time delay error in the estimation of link travel time based on two terminal loop stations. An immediate application of this result is determine the loop density based on required threshold of link travel time delay error. If this method is valid, it can be directly used by Caltrans engineers.

4.2 Document Review on Congestion Onset Detection

The work in [27] used Moving-Window Average or Median filtering to smooth the occupancy for specially distributed loop stations and then the smoothed occupancy to detect incident. It is claimed that performance has been improved. The paper of Lin and Daganzo [16, 17] provided a method for traffic congestion detection without complicated traffic flow theory involved. It is completely a scheme using both upstream and downstream loop stations. It is essentially an integration approach which posses some filtering effect naturally. However, the off-set $-\tau$ may cause some problem as well as the difference between upstream and downstream occupancy error. An empirical threshold is used as criterion for congestion.

It was pointed out by Persaud and Hall in [23] that the loop detect o*ccupancy discrepancy* had made the conventional occupancy-based incident detection logic difficult to apply. Furthermore, an experimental study conducted by Chan and May [4] showed with field data that the average detector pulse on-times for two longitudinally closely spaced stations could vary by 5-10% or even higher.

Work in [8] proposed to use travel time of a special vehicle, or vehicle signature, on two consecutive loop stations to detect if there is a congestion onset. This approach may work provided that pecial vehicle signature could be caught by loop station at both upstream and downstream stations. This in fact is an extremely difficult issue for using loop-only detection (even dual loop stations) in practice.

It has been observed on freeway and demonstrated by analysis of appropriate model [20] that the freeway traffic becomes unstable (stop & go; or shockwave) when density increases beyond that corresponding the capacity even if there is no accident. A desirable condition for sustainable operating of a freeway would have the density slightly less than that corresponding to capacity for expected speed. This operating condition provides a safe margin permitting stable operation in which minor disturbances could be absorbed.

Different methods for incident detection have been evaluated and documents extensively reviewed in [18]. It is recommended that the Cellular phone is the most effective and method with the least false alarms. While other methods would have problems in one situation or the other: “The McMaster algorithm was reported to suffer an increase in false alarms during a snowstorm. The Bayesian method is also reported to be sensitive to weather conditions. The SND and DES algorithms can tolerate moderate variations in weather conditions. Image processing technology as it is applied to incident detection also can be affected by weather and lighting conditions.”

In the book by Martin et al [18], the approaches for the incident detection developed so far have been classified into four categories:

- Pattern recognition
- catastrophe theory
- statistical
- artificial intelligence

For the discussion in this report, it is necessary to discuss pattern based and Catastrophe Theory based incident detection algorithms which are related to what we are to propose.

(1) Pattern recognition

Information used include in this approach: both upstream and downstream detected occupancy, traffic volume, and traffic flow by loops or video camera. If one or some of those parameters is outside of the threshold compared to normal case, an incident is announced. Parameters used include at both upstream and downstream detection stations or at downstream station detection only. Threshold calibration depends on different road geometries (i.e. ramps, weaving sections, hills, etc.), which is complicated for large networks. Persistence is applied to reduce false alarm rate by checking for a specified period of time.

Three pattern based algorithms are worth mentioning:

(i) California algorithm is in this category. It has different versions which mainly using upstream and downstream occupancy. This is the most extensively explored algorithm and widely implemented according to [18].

(ii) The APID (All Purpose Incident Detection) algorithm is a combination of the various California algorithms along with a compression wave test routine and a persistence test routine. Unlike the California algorithms, it uses smoothed-occupancy as the detection variable to reduce false-alarm rates. The algorithm's goal was to provide excellent performance under all conditions, thus the "all purpose" acronym.

(iii)*PATREG Algorithm*: Developed in 1979 by the Traffic Road and Research Laboratory (TRRL), the Pattern Recognition Algorithm (PATREG) was designed to work in conjunction with the High Occupancy (HIOCC) algorithm [11]. This algorithm exceptionally uses speed but has not been developed further since then.

(2). Catastrophe Theory:

This theory is based in sudden changes that occur in one variable of interest while other related variables exhibit smooth and continuous change. These variables are speed, flow, and occupancy. When speed drops dramatically without a corresponding increase in occupancy and flow, the alarm sounds**.** The algorithm functions are based on data from a single detector station [1]. The ***McMaster Algorithm*** is the representative [18].

4.3 Congestion Onset Detection Based on Instant Point Speed

Our algorithm is more close to the *McMaster Algorithm* than to the California algorithm for two reasons: (a) We used the idea in Catastrophe Theory that the speed will have sudden change while other parameters may change smoothly; (b) there is no pattern recognition and comparison. Instead, we propose to use the following information for incident detection:

- instant point speed and its changing rate (essentially acceleration)
- instant point occupancy and its changing rate
- detailed analysis on shockwave and its effect on speed detection

The algorithm we propose uses a dual loop station either upstream or downstream which means that each dual loop station is to detect both shockwave and discharging wave (counterpart of shockwave, to be explained later). The reason for using dual loop station is to obtain more accurate vehicle speed with the least aggregation which essentially obscures the sudden change of traffic parameters and bring more time delay in detection. There is also a threshold set up on the speed rate change and persistence detection involved. Only when the speed rate change is over the pre-specified threshold for certain steps (persistence threshold), an incident is announced. At this stage, the algorithm does not claim to distinguish the types of an incident such recurrent or non-recurrent. So it is basically a congestion onset detection algorithm.

This subsection suggests an algorithm which uses only one of the two terminal point stations, but not both, which is the main difference compared to all the previous methods.
In the following discussion, the concept *instant point speed* is used. Its practical meaning can be fund in Chapter 3.

The following notations are used for discussion in this chapter:

L – distance between two loop stations related to the given link

$v_u(t), v_d(t)$ - Upstream and downstream *instant point speed* at time t

$q_u(t)$ - Upstream flow including on-ramp in-flow at time t

$q_d(t)$ - Downstream flow including off-rap out-flow at time t

$o_u(t), o_d(t)$ - Upstream and downstream occupancy at time t

4.3.1 Congestion Onset Algorithm

Consider a pair of loop detector (upstream and downstream) stations in Figure 4.1 with inter-station distance L. Previous work [16, 17] showed that a congestion onset could be detected if the following condition is satisfied:

$$o_u(t) - o_d(t) \geq \delta_o \tag{4.1}$$

where $\delta_o > 0$ is pre-specified thresholds. The occupancy is chosen as test parameter because they can be detected and estimated with dual loop station directly. The synchronized occupancy of upstream and downstream occupancies are compared for such detection.

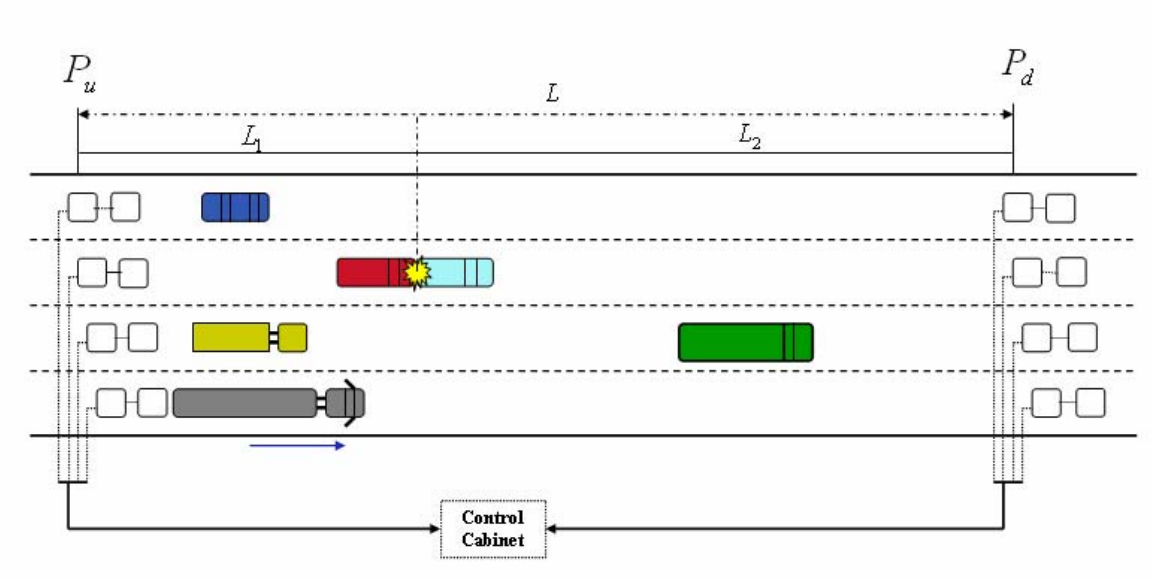


Figure 4.1. Congestion onset location and time delay in detection

It is assumed that both upstream and downstream stations have dual-loop detectors which allow more precise detection of vehicle-by-vehicle instant point speed and occupancy. The instant point speed detection has been described before, which is based on up/down time instant of the dual loops. The occupancy for a dual loop can be taken as the average occupancy of the two. It is worth pointing out that for single loop detect station, speed is deduced from occupancy by assuming an average vehicle length. They are thus practically equivalent. For dual loop stations, the speed estimation is not deduced from occupancy anymore. They are not equivalent anymore although they are not independent conceptually. It is thus makes sense to use them both for detection which implies some level of redundancy.

To reduce time delays in congestion onset detection, let look at the changing rate of the point detection data considered as time series. For convenience of discussion, the following state is called ***discharging phenomenon*** (gas expanding in a closed container) which is opposite to the ***shockwave phenomenon*** (gas compression in a closed container). Its speed wave called ***discharging wave*** as apposed to ***shockwave***. It can be characterize as:

$$\begin{aligned} v_d(t) - v_d(t-1) &\geq \sigma_v^d \\ o_d(t) - o_d(t-1) &\leq -\sigma_o^d \end{aligned} \tag{4.2}$$

where $\sigma_v^d > 0; \sigma_o^d > 0$ are pre-specified thresholds. It can be interpreted as: speed strictly increasing significantly and occupancy strictly decreasing significantly. The discharging

phenomenon appears in the immediate downstream of the bottleneck. If, in addition, the following conditions are satisfied:

$$\begin{aligned} v_u(t) - v_u(t-1) \le 0 \\ o_u(t) - o_u(t-1) \ge 0 \end{aligned} \tag{4.3}$$

it can be concluded that a congestion onset has happened somewhere between the two terminal stations. It is noted that such criterion only relies on the variation a downstream detection point.

Similarly, one can detect congestion onset only based on a shockwave characteristics at the loop station. It can be characterized as

$$\begin{aligned} v_u(t) - v_u(t-1) \le -\eta_v^u \\ o_u(t) - o_u(t-1) \ge \eta_o^u \end{aligned} \tag{4.4}$$

where $\eta_v^u > 0, \eta_o^u > 0$ are properly pre-specified thresholds. It can be interpreted as: speed significantly decreasing and occupancy significantly increasing. If, in addition, the following conditions are satisfied:

$$\begin{aligned} v_d(t) - v_d(t-1) \ge 0 \\ o_d(t) - o_d(t-1) \le 0 \end{aligned} \tag{4.5}$$

it can be concluded that a congestion is building up somewhere between the two terminal stations. It is noted that such criterion only relies on the variation a upstream detection point.

Remark 4.1. In the above detection criteria, all the occupancy part can be dropped out to used speed variation only since point speed and point occupancy are conceptually equivalent although it is possible to estimate them independently.

Remark 4.2. *Persistence checking* is necessary to reduce false alarm rate (FAR), which can be implemented as the congestion is continuously detected for certain time steps. This number needs to be determined through tuning in practice. Obviously, there is trade-off between the time delay caused by persistence checking and the number of FAR.

4.3.2 Algorithm Validation

To validate the algorithm for congestion onset detection based on the instant speed and/or occupancy variations at a point detect station, the data from the pair of loop stations need to satisfy the following requirements:

(1) Both of the terminal stations have dual loop detectors and their data are synchronized;
(2) 5~10 Hz update rate is necessary for data collection;
(3) There was a known incident/accident happened between the two terminal station during the time period of data collection;
(4) Data are to be error-free; otherwise, an effective method is available for data correction and cleansing;

Due to the time and resource limit of the project, this algorithm cannot be validated in current phase of the project.

4.4 Upper Bound for Delay in Link Travel Time Estimation

As discussed in previous chapters, the most serious delay in link travel time estimation would result in the phase of congestion onset and building up, which is characterized by shockwave propagation. This subsection will explain how the shockwave would affect the link travel time estimation and will use the result about shockwave propagation speed of last chapter to deduce an upper bound for the time delay in link travel time estimation. The relationship to be established is independent of methods used for the estimation of such delay. However, it is assumed that it is able to detect the congestion onset only based on point detection at upper stream station or down stream station instead of both as described in previous subsection.

Using the previous congestion onset detection, it is clear that the congestion onset location would affect the time delay error in link travel time estimation. Intuitively, this effect can be described as follows. If the shockwave back propagates to the upstream station, it can be detected; or if the discharging wave is reaching the downstream station, it can be detected as well.

What is the worst case for time delay estimation? Or equivalently, one may ask the following question: where is the bottleneck location that causes the most time delay in congestion detection? If the discharging wave is detected first at downstream station, or the shockwave is detected at upstream station, the time delay will be less than the case when the shockwave is detected at upstream and discharging wave is detected at downstream simultaneously, which produce the largest time delay. In this worst case, let τ_c denote the upper bound for the shockwave to reach upstream detect point and the discharge wave to reach the downstream detect point. The following relationship would hold:

$$L = V_{shock}\tau_c + v_d \cdot \tau_c \tag{4.6}$$

Or equivalently

$$\tau_c = \frac{L}{V_{shock} + v_d} \tag{4.7}$$

Considering the time delay τ_0 used for the detection, which may be caused by data processing such as aggregation over time and/or by persistence checking to reduce false alarm in congestion onset detection. For example, if the data update rate is 0.1s and data processing delay is 0.5s and persistence checking is 10 steps. Then $\tau_0 = 1.5s$. The upper bound for time delay error estimation can be expressed as

$$\tau_{\max} = \tau_c + \tau_0 = \frac{L}{V_{shock} + v_d} + \tau_0$$

Note that L is the distance between the two terminal stations. $V_{shock} \doteq 5.1 m/s$ is the shockwave speed according to previous chapter. v_d is the instant point speed at downstream station before congestion onset, which can be buffered in practical implementation. Now

$$\tau_{\max} = \frac{L}{5.1 + v_d} + \tau_0 \tag{4.8}$$

from which one can determine the upper bound for time delay from the distance between the two terminal stations and the instant point speed at downstream station.

4.5 Loop Density Determination

The discussion in this chapter applies to all point detect stations. The equation in (4.8) could be used for determining the density of the detect stations, such as loops. For example, if the free-flow speed or the speed limit for the highway is *65mph* or τ_d=29.1[m/s], (4.8) establishes a relationship between the upper bound for time delay in link travel-time estimation as follows:

$$\begin{aligned} \tau_{\max} &= \frac{L}{5.1 + 29.1} + \tau_0 \\ \tau_{\max} &= 0.02924 \cdot L + \tau_0 \end{aligned} \tag{4.9}$$

This simple formula can be used by traffic engineer to determine how dense the loop stations should be installed on highways to achieve the acceptable estimation error threshold $\tau_{\max}$.

4.6. Concluding Remarks for Chapter 4

This chapter has suggested a method for quick congestion onset detection. This detection method is based on finer analysis of the changing rate of the instant point speed. It basically says that if the changing rate of the instant speed and or occupancy at on one of the two terminal station is over a certain threshold, it can be determined that a congestion onset is detected. However, the suggested congestion onset detection method needs to be validated with specified traffic data, which can only be achieved in future phase of the project.

It has also showed how this method could be used for determining the upper bound for the delay in link ravel time estimation error. A direct application which can be used by traffic engineers is to determine the density of detect stations such as loops for given acceptable upper bound for the delay in link travel time estimation.

The following are the work remain to be done in the future phase of the project:

(1). To validate the congestion onset detection algorithm proposed in this project, which relies on the variation at a single point detection station only.

(2). If the congestion onset detection algorithm can be validated, the upper bound for the link travel time delay in estimation is validated as well. It can thus be directly used for determining loop density in freeway main lane for given acceptable threshold for link travel time delay error in estimation.

References for Part I

[1] Antoniades, Charalambos N.; Stephanedes, Yorgos J. Single-Station Incident Detection Algorithm (SSID) for Sparsely Instrumented Freeway Sites. *Transportation Engineering,* 1996.

[2] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, Nonlinear Programming Theory and Algorithms, John Wiely & Sons, Inc., 1993.

[3] Caltrans and IBM, *Traffic Management System Detection Plan (TMSDP)*, December 2002.

[4] Chan L. and May A. D. (1986) The use of vehicle detectors for freeway traffic management. UCB-ITS-WP-86-6, Institute of Transportation Studies, University of California at Berkeley, U.S.A.

[5] Chen, C., K. Petty, A. Skabardonis, P. Varaiya, Z. Jia, Freeway Performance Measurement System: mining loop detector data, 80th Annual Meeting, TRB, January 2001, Washington, D.C., July 24, 2000

[6] T. Choe, A. Skabardonis, P. Varaiya, Freeway Performance Measurement Systems (PeMS): An operational analysis tool, TRB 81st Annual Meeting, Washington, D.C., Jan. 2002

[7] Coifman, B., Dhoorjaty, S., "Event Data Based Traffic Detector Validation Tests", *Proc. 81st Annual Meeting Transportation Research Board*, 2002.

[8] Coifman, B., Identifying the Onset of Congestion Rapidly with Existing Traffic Detectors, *Transportation Research*, Part A. , vol 37, no 3, 2003, pp. 277-291.

[9] Coifman,B., Estimating Travel Times and Vehicle Trajectories on Freeways Using Dual Loop Detectors, *Transportation Research Part A*, vol 36, no 4, 2002, pp. 351-364,

[10] Coifman, B., 1997. Time Space Diagrams for Thirteen Shock Waves, PATH Report No. UCB-ITS-PWP-97-1, University of California, Berkeley, CA.

[11] Collins, J.F., Hopkins, C.M., and Martin, J. A. . Automatic Incident Detection TRRL Algorithms HIOCC and PATREG. Supplementary Report 526, Transport and Road Research Laboratory, Crowthorne, England, 1979.

[12] Cook, A. R. and Cleveland D. E. (1974) Detection of freeway capacity-reducing incidents by traffic-stream measurements. ***Transpn Res. Rec.*** 495. l-l I.

[13] Daganzo, C. F., "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic *Transportation Research Part B:*

Methodological, vol. 28, no. 4, p269–287, Aug. 1994.

[14] Daganzo, C. F., "The cell transmission model, part II: Network traffic," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. Apr. 1995.

[15] 15 C. L.. Messer C. J. and Nuckles N. B. (1974) Incident detection on urban freeways. ***Transpn Rev. Rec.*** 495. 12-24.

[16] Lin, W. H. (1995). Incident detection with data from loop surveillance systems: the role of wave analysis. Ph.D Dissertation. UCB-ITS-DS-95-4. Institute of Transportation Studies. University of California at Berkeley, U.S.A.

[17] Lin, W. H. and Carlos F. Daganzo, A simple detection scheme for delay-induce freeway incident, *Transportation Research A,* Vol. 31. No. 2. PP. 141-155. 1997

[18] Martin, P. T., Perrin, J., Blake Hansen, Kump, R., and Moore, D. , *Incident Detection Algorithm Evaluation,* University of Utah, Prepared for Utah Department of Transportation, March 2001; Available at website:
http://www.ndsu.nodak.edu/ndsu/ugpti/MPC_Pubs/html/MPC01-122/index.html

[19] Next Generation Simulation, Cambridge Systematics, http://ngsim.camsys.com./

[20] Payne, H. J., Models of freeway traffic and control, in Mathematical Models of Public Systems (Simulation Council Procedings), Vol. 1. No. chp. 6

[21] Payne, H. J., Goodwin D. N. and Teener M. D. (1975), ***Evaluation of Existing Incident Drtection Algorithm,*** Report No. FHWA-RD-75-39, Federal Highway Administration, Washington DC. Feb. 1975. p. I IS. PB 241883.

[22] Payne, H. J. and Tignor S. C. (1978) Freeway incident detection algorithms based on decision trees with states. Transpn Res. Rec. 682. 30-37.

[23] Persaud, B. and Hall F. L. (1989) Catastrophe theory and patterns in 30-second freeway traffic data – implications for incident detection. Transpn *Res.* 23A. 103-I 13.

[24] Petty, K., Small time scale analysis of loop data, 1995, available at: http://ipa.eecs.berkeley.edu/~pettyk/FSP/

[25] Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. P. Lannery, *Numerical Recipes in C, The Art of Scientific Computing*, 2nd Ed., Cambridge University Press, 1992,

[26] Skabardonis, A., H. Noeimi, K. Petty, D. Rydzewski, P. Varaiya, H. Al-Deek, *Freeway Service Patrol Evaluation*, California PATH Research Report, **UCB-ITS-PRR-95-5**, U. C. Berkeley, 1995

[27] Stephanedes, Y. J. and Chassiakos A. P. **(1993)** Freeway incident detection through

filtering. ***Transpn Res. 1C.*** 219-233.

[29] Sun, X., Munoz, L., and Horowitz, R., Highway traffic state estimation using improved mixture Kalman filters for effective ramp metering control, Proc. 42nd IEEE CDC, Maui, Hawaii UAS, Dec. 2003, p6333-6338

[30] Wang, Z. R. and Chiu Liu, Using loop detectors to estimate travel time delay: How accurate is the method?, Proc. of 12th ITS World Congress, Nov. 6-10, 2005, San Francisco, California

[31] Sensys Networks, http://www.sensysnetworks.com/

Part II

Empirical and Analytical Results for Sensitivity to Loop Detector Station Spacing

conducted by

Prof. Benjamin Coifman's group
at The Ohio State University.

Chapter 5

Empirical and Analytical Results for Sensitivity to Loop Detector Station Spacing

5.1 Introduction

Traffic congestion increases each year on the nation's highways. Intelligent Transportation Systems (ITS) including traffic responsive ramp metering, incident detection and travel time predication have been developed to operate the network more efficiently by responding to conditions. These active controls rely on sensors to monitor what is occurring on the roadways. The majority of data upon which ITS services are based come from inductive loop detectors. For these active controls the most important function of a traffic surveillance system is determining reliably whether a facility is free flowing or congested. The second most important function is responding rapidly when facilities become congested.

It is well known that the currently deployed traffic monitoring technologies only monitor discrete points along a freeway and cannot observe the entire facility. This fact is true for loop detectors, non-invasive vehicle detectors like radar, and even most close circuit television (CCTV) deployments. Limiting the scope to automatic vehicle detectors (loops, and non-invasive detectors), conditions between detector stations have to be inferred from data observed at the stations. So a trade off arises between the spacing between detector stations and the latency resulting from the detector network. Deploying a suboptimal traffic monitoring system could unnecessarily increase the cost or it could lead to a system that cannot respond reliably. The impact of the trade off on traffic control applications remain unquantified and this research begins to address the problem by calculating response time and traveler delay. In particular, this study seeks to find how big of an event can be missed and at what cost for a given deployment. We examine several bottlenecks using real detector data and consider the effect of detector station spacing, including: present spacing, half spacing, and double spacing. Next, since conventional detectors only monitor conditions at discrete locations, the influence area or link assigned to a detector station is examined. Consideration is given to different boundary definitions between links, e.g., a detector may be considered to be at the start of a link or the middle of a link. Different definitions will impact the nature of errors in estimating traveler delay though it should not impact the overall responsiveness to detecting events.

Since the bottlenecks are examined individually the work represents the system performance both for incidents and recurring congestion with following caveat. The work might overestimate errors when an operating agency has pinpointed the precise location of a recurring bottleneck (with an accuracy on the order of 100 m) and deployed detectors carefully and deliberately to respond quickly when it becomes active.

Costs are presented in terms of time (vehicle hours of delay per lane) and as a function of spacing. This general approach allows the reader to experiment with different dollar values for time and to use the actual deployment costs in their service area, e.g., there are several common vehicle detectors each with a different performance range. Similarly, there are numerous applications that use detector data, each with their own specific needs. The flexibility of the metrics is intended to help operating agencies make smarter decisions regarding the deployment of detection options, improving the cost-benefit ratio for detection investments using off-the-shelf products while also guiding development of future detection technologies. Without addressing this problem, operating agencies will continue to deploy vehicle detection following ad hoc guidelines that were typically developed out of convenience rather than any attempt to optimize performance. Without such analysis, an agency risks inappropriate traffic detection investments, potentially deploying an excessive system based on standards that are too strict, or under-designing the system that cannot collect data accurate enough for decision-making. The more informed decision-making could also benefit operations, leading to control algorithms (ramp metering, incident detection, etc.) that are more tolerant to conditions that are observed in detector data.

5.2 Analysis of Response Time

Consider the two hypothetical incidents presented in Figure 5.1, in both cases the incident occurs at some location upstream of the i-th detector station, denoted by IL. A queue grows backward from the incident, represented by the triangular region in the time space plane. The velocity that the queue grows is denoted by v_1. Vehicles accelerate as they pass the fixed bottleneck, as evidenced by the interface that lasts a long duration and remains fixed at IL. One of the goals of a traffic monitoring system, e.g., a sequence of loop detector stations, is to rapidly identify the presence of an incident or any other active bottleneck. The figure shows four detector stations, with station i downstream of the incident. The detection system does not observe the impacts of the queue until it reaches the first station, i+1 in this case. The response time of the detector system is bounded by this travel time, i.e.,

$$response\_time \geq \frac{D_i - IL}{v_1} \tag{5.1}$$

Assuming that IL is uniformly distributed over the entire link, on average,

$$average\_response\_time \geq \frac{1}{2} \cdot \frac{D_i}{v_1} \tag{5.2}$$

As will be illustrated shortly, v_1 was observed empirically to fall between 2 mph and 10 mph. In any event, an operating agency has little control over v_1 and can only affect D_i. Doubling D_i will double the magnitude of the constraint on the expected response time. In practice an additional constant delay will be incurred due to the following two sources. First, the signal will likely arrive in the middle of a sample period, so that sample might not reflect the full impact of the queue, making difficult the task of identifying the queu-

ing in this first sample. Second, most vehicle detectors exhibit errors, so a latency of one or two samples is typically employed by operating agencies to differentiate between transient errors and real events. Collectively, these delays result in an additional delay of 2-3 sample periods above that of Equation 5.2. Using shorter sampling periods would reduce the added delay, but might increase the noise in the measurements, resulting in the need for additional samples to filter out detector errors. In other words, one should only expect marginal improvements to response time based on changes in the sampling period alone. Improved detector performance could help in this regard, but such analysis is beyond the scope of this work and the average response time is simply taken as variable component, i.e., Equation 5.2. Table 5.1 shows the average and maximum response time for various combinations of v_1 and D_i. Finally, note that these response times are independent of where the boundary between links fall, using conventional detectors one has to wait for the signal to reach the first detector station upstream of the incident.

5.3 Analysis – Original Detector Spacing

The analysis employs a standard model of queue growth and decay in the time space plane, e.g., Figure 5.2. Detector stations are superimposed on the queuing phenomena as indicated by the station locations on the right hand side of the figure and D_i on the left hand side of the figure denotes the distance between adjacent stations. Figure 5.2B shows the complete queue, IL defines the location of the bottleneck relative to the next station downstream and EQ defines the end of the queue, again relative to the next station downstream. Because it can be difficult to discern the exact time a disturbance passes a detector station (e.g., because it passes in one lane before it passes in another), we assume that the end of the queue grows with a constant velocity, v_1, from the IL to the last detector station and similarly that it recedes with a constant velocity, v_2. Figure 5.2A shows the queue that is observable from the detector stations. Upstream of the last detector station we model the queue growth and decay also with a constant velocity, but generally different than v_1 and v_2. We use Monte Carlo simulation to estimate IL (and thus the duration, T_{new} of the bottleneck activation) and EQ, assuming that each distance comes independently from a uniform distribution.

As noted above, this analysis uses a combination of hypothetical and real surveillance networks. The latter come from the Columbus Metropolitan Freeway Management System (CMFMS) along I-70 and I-71 in Ohio [1]. The facility includes 45 detector stations as shown in Figure 5.3. Approximately a third of the stations are equipped with dual loop detectors while the remainder has single loop detectors. Stations 8 and 12 are excluded from this study due to preexisting problems at these stations that preclude accurate data measurement. Speed is either measured directly at dual loop detector stations or estimated using the median on-time [2]. The data are aggregated in to 30 sec samples across all lanes and moving median of 11 samples (centered on the given sample) is then applied to eliminate any noise or transients in the speed data. The times that the queue passes a station can then be found in the speed time series, and the v_1 and v_2 can be derived from the times at consecutive stations and D_i. Figure 5.4 shows the moving median loop speeds from eight northbound stations on the I-71 corridor on April 23, 2005. In this case, based on police accident reports, an accident blocked the left lane between stations

5 and 6 around 12:00. Station 7 is furthest downstream and does not appear to see any impact from the bottleneck until after the queue starts to dissipate around 13:00. Station 6 is just downstream of the incident and the speed drops a few mph, presumably due to vehicles accelerating away from the bottleneck. Stations 5, 4, 3, 2 and 1 see the queue in that order and speeds drop significantly at each of the stations, with station 5 having the highest speed within the queue. The queue does not reach station 110, so the end of the queue falls between stations 110 and 1. From the spacing between stations and the time of the speed drop at each station, shown in Figure 5.4, Table 5.2 tabulates the upstream wave velocity across each successive pair of detector stations as the queue grows and similarly the downstream wave velocity as the queue recedes after the accident is cleared.

To reduce sensitivity to the 30 sec sampling period and small traffic fluctuations, we estimate velocities v_1 and v_2 from the furthest stations at which the queue was observed, in this case stations 5 and 1; v_1 and v_2 were –5.3 mph and 11.4 mph, respectively. In fact this queue was used to build the model shown in Figure 5.2.

Looking at the segments between detector stations, there are three distinct types of segments in each queue: head, tail, and body. The head of the queue contains the bottleneck, with the downstream station remaining uncongested and the upstream station records a speed drop. The tail of the queue represents the furthest upstream link, where the upstream station remains uncongested and the downstream station records a speed drop. All intervening segments (if any) fall within the body of the queue. In these body segments the queue grows from the downstream detector to the upstream detector, covering the entire distance. From an operational standpoint each link should contain exactly one detector station, but there is flexibility in defining where the boundary falls between links, e.g., a detector station represents conditions from that station to the next station downstream, or perhaps the boundary between links is placed at the midpoint between two successive stations. LL_i denotes the boundary between two links, where LL_i is the percentage of D_i associated with station i. In this analysis D_i is allowed to vary from location to location, but LL is fixed across all stations. The link boundary defines influence area of the detector station and in turn, affects magnitude of estimation errors arising from the discrepancy between real traffic conditions and estimated traffic conditions. In standard practice, conditions at the detector station are assumed to represent the traffic state on the entire link, but as the queue passes through the link, part of the link will be congested and part of it will be freely flowing. Depending on where the detector is located in the link, at any moment during these transitions standard practice will underestimate or overestimate the true delay in the link. The following analysis separately tallies the underestimation and overestimation, rather than tallying the net delay error in which the two sources can cancel one another. These errors are represented as an "area" of the time-space plane with units of mile-hr.

5.3.1 Empirical Results

Figure 5.1 shows these errors in the three types of segments, head (HQ), body (BQ) and tail (TQ) of a hypothetical queue in the time space plane. As before the detector stations are shown with horizontal dashed lines while the boundaries between segments are

shown with thinner solid lines. Because the empirical data come from detectors, we tally the errors in the segments between detectors rather than adding an additional step of trying to estimate when the queue reaches the unobserved boundary between links. The amount of overestimation and underestimation remains a function of LL and we explicitly account for LL in our analysis[1]. So in Figure 5.1A the bottleneck occurs upstream of the link boundary in HQ and there is a large overestimation when conditions at station i+1 are extrapolated to the entire link, while in Figure 5.1B the bottleneck occurs downstream of the link boundary and there is a large underestimation when conditions at station i are extrapolated to that link. Underestimation and overestimation errors in HQ, BQ, and TQ are functions of LL. Errors HQ also depend on IL, while errors in TQ depend on EQ. To facilitate comparisons, IL and EQ are discretized to 1/20th of the distance between the given stations while LL is discretized to 1/10th of the distance.

Using the terminology defined in terms of Figure 5.1 and detector data from the real incident shown in Figure 5.4, Figure 5.5 shows the estimation errors as a function of LL when IL and EQ are both placed 0.6 away from the downstream detector station in HQ and TQ, respectively. The plots show underestimation, overestimation, and sum of the absolute value of the two errors. Note that the vertical scale changes from one plot to the next. For reference, IL and EQ are marked as dashed vertical lines in Figure 5.5A and 5.5C, respectively. Looking closer at HQ and TQ, Figure 5.6 shows the results when IL and EQ are varied. Since IL=0 and IL=1 are functionally equivalent, the former is excluded from the analysis. Similarly, EQ=0 is functionally equivalent to EQ=1, so in this case the latter is excluded.

From Figure 5.6A the sum of estimation errors in HQ is minimized when LL=IL, and in the absence of more precise positioning information IL is taken to be uniformly distributed between 0 and 1 for incidents[2]. From Figure 5.5B the sum of estimation errors in BQ is minimized when the link boundary is placed midway between stations (recall that the estimation error in BQ is independent of IL and EQ). Finally, from Figure 5.6C the sum of estimation errors in TQ is minimized when LL=0.5*EQ. The results for a particular location might differ given a detailed geometric study, but in the absence of such detailed information EQ is taken to be uniformly distributed between 0 and 1[3].

Holding v_1, v_2, and T constant[4], pairs of IL and EQ were generated at random from a uniform distribution (increments of 0.05) for each of the 11 LL (increments of 0.1). Here, 20 points per LL were used for HQ, BQ and TQ, while 400 points per LL were used for the total combined error across all links. The resulting errors for each of the simulation runs were calculated (sum of the absolute value of overestimates and underestimates), as shown by the points in Figure 5.7 while the curve shows the average at each LL. Once

[1] For example, if the detector is assumed to represent a link strictly downstream of the station such a boundary would result in little overestimation but a large underestimation of delay since queues grow upstream.

[2] IL may be deterministic for recurring bottlenecks, in which case one can select the appropriate subplot from Figure 5.6A.

[3] In this case EQ should be randomly distributed both for recurring and non-recurring bottlenecks, though some geometric features, e.g., a lane drop or diverge, could change the actual distribution.

[4] Thus, in the empirical analysis preserving what was observed at the detector stations. As a result, however, the actual duration of the incident, T_{New}, is a function of the constants.

more, note that the vertical scale changes from plot to plot. The average error is minimized at 0.5 for HQ and BQ, while it is minimized at 0.2 for TQ. Summing across the three types of segments, the lower right hand plot shows that the net error is minimized when LL is 0.3.

5.3.2 Analytical Results

Assuming the distribution of IL and EQ are independent, uniformly distributed, and discretized to 1/20 as above, from Figure 5.6A it is clear that the error in HQ is minimized when LL=IL, where $0 < IL \leq D_i$. So the expected value of IL over many samples is,

$$\overline{IL} = \frac{1}{2} D_i \tag{5.3}$$

and the value of LL that is expected to minimize the error is,

$$LL^* = \overline{IL} = \frac{1}{2} D_i \tag{5.4}$$

i.e., the midpoint between detector stations. However, because the values are discretized to 1/10ths and IL=0 is excluded, the empirical results for the optimal LL in HQ should be slightly larger than 0.5[5].

In BQ, the estimation error from the growing queue, v_1, can be expressed as follows,

$$\begin{aligned} &\left|Overestimation\ error\right| + \left|Underestimation\ error\right| \\ &= \frac{1}{2} \cdot \frac{LL_{i+1}}{V_1} \cdot LL_{i+1} + \frac{1}{2} \cdot \frac{(D_{i+1} - LL_{i+1})}{V_1} \cdot (D_{i+1} - LL_{i+1}) \\ &= \frac{1}{2} \cdot \frac{\alpha_1 D_{i+1}}{V_1} \cdot \alpha_1 D_{i+1} + \frac{1}{2} \cdot \frac{(1-\alpha_1) D_{i+1}}{V_1} \cdot (1-\alpha_1) D_{i+1} \\ &= \frac{{D_{i+1}}^2}{2V_1} \left(2\left(\alpha_1 - \frac{1}{2}\right)^2 + \frac{1}{2} \right) \end{aligned} \tag{5.5}$$

This error is minimized when $LL \leq \alpha_2 \leq 0.5$. Substituting in v_2 when the queue recedes will produce the same result, the error being minimized when LL=0.5. So the optimal link boundary in BQ is at the midpoint between detector stations.

In TQ, if EQ<LL then the net error is all due to over estimation and it is greater than 0.5*t*LL, if EQ>LL, then this error becomes,

[5] Consistent with this analysis, careful inspection of Figure 5.7A will reveal that the average error is slightly smaller at 0.6 than it is at 0.4, and in fact the error at 1.0 is roughly equal to that 0.1 instead of 0. This slight shift is an artifact of the discretization scheme.

$$\begin{aligned} &|Overestimation\ error| + |Underestimation\ error| \\ &= \frac{1}{2} \cdot LL \cdot \frac{t \cdot LL}{EQ} + \frac{1}{2} \cdot (EQ - LL) \cdot \left(t - \frac{t \cdot LL}{EQ} \right) \\ &= \frac{t}{2} \cdot \left(EQ^2 - 2 \cdot LL + 2 \cdot \frac{LL^2}{EQ} \right) \end{aligned} \tag{5.6}$$

where t is the time that detector station i+2 actually sees congested traffic. Taking the first derivative relative to LL and setting it equal to zero yields LL=0.5*EQ, while the second derivative is greater than zero at this point, indicating that it is a minimum. Over many samples where $0 \leq EQ < D_{i+2}$,

$$\overline{EQ} = \frac{D_{i+2}}{2} \tag{5.7}$$

and

$$LL^* = \frac{1}{2}\overline{EQ} = \frac{1}{4} D_{i+2} \tag{5.8}$$

However, because the values are discretized to 1/20ths and EQ=1 is excluded, the empirical results for the optimal LL in TQ should be slightly smaller than 0.25[6].

5.3.3 Vehicle Hours of Delay per Lane

It is straightforward to calculate the size of the regions in the time-space plane where delay is overestimated or underestimated, e.g., the shaded areas of Figure 5.1. In algebraic terms this "area" is simply the size of the region, but care must be taken because this area does not have units of distance squared. Rather, the result will have units of "mile-hour", e.g., Figure 5.7, does not have a direct interpretation. To correct for this problem, the measures are converted to "vehicle-hours" of delay using an estimate of density, k, from flow, q, and speed, v. Namely, during each five minute sample at station i, the per lane density is estimated to be,

$$\hat{k}_i = \frac{q_i}{v_i} \tag{5.9}$$

and the resulting delay per lane is calculated as follows,

$$Delay_{UE(i)} = \hat{k}(A_{UE(i)}) \cdot \left| A_{UE(i)} \right| - \frac{q(A_{UE(i)}) \times \left| A_{UE(i)} \right|}{v_{ff}} \tag{5.10A}$$

[6] Consistent with this analysis, Figure 5.7C reveals that the empirical minimum falls at 0.2 rather than in between 0.2 and 0.3, again, this small shift is an artifact of the discretization scheme.

$$Delay_{OE(i)} = \hat{k}(A_{OE(i)}) \cdot \left|A_{OE(i)}\right| - \frac{q(A_{OE(i)}) \times \left|A_{OE(i)}\right|}{v_{ff}} \quad (5.10B)$$

where,

$\left|A_{UE(i)}\right|$ = Area associated with underestimation error at station *i*
$\left|A_{OE(i)}\right|$ = Area associated with overestimation error at station *i*
$\hat{k}(A_{UE(i)})$ = Median estimated density (from 5 min samples) during the time period of the underestimation error at station i
$\hat{k}(A_{OE(i)})$ = Median estimated density (from 5 min samples) during the time period of the overestimation error at station i
$q(A_{UE(i)})$ = Median flow (from 5 min samples) during the time period of the underestimation error at station i
$q(A_{OE(i)})$ = Median flow (from 5 min samples) during the time period of the overestimation error at station i
v_{ff} = Free flow speed of traffic

As a numerical example of estimation error delay, while Figure 5.7 showed the error in terms of mile-hours of delay, Figure 5.8 shows the corresponding vehicle-hours of delay error per lane using this estimation technique for each of the segment types and the net delay over all segments. Again the points show the results of individual trials while the curve shows the average delay at each LL. The lower right-hand plot in Figure 5.8 shows the error in delay from this incident ranges between 10 and 25 vehicle hours per lane, depending on LL.

5.3.4 Other Incidents

A total of eight incidents are examined in detail in this study, as enumerated in Tables 5.3-5.4 and Appendix C. The corresponding stations and direction of travel are shown in Figure 5.9, asterisks show queue boundaries for HQ and TQ, while circles denote BQ stations. In all cases the direction of travel is from bottom to top, i.e., BQ is at the bottom and HQ is at the top. The vertical axis on this figure is proportional to the distance between stations, so as is clearly evident, the stations are not evenly spaced and the detector spacing impacts the various incidents differently. While incident A was examined in detail above, the corresponding detailed analysis for the remaining incidents is presented in Appendices D and E.

For each incident the resulting optimal link location for the three segment types and the entire queued area are summarized in Table 5.3. In all eight cases the optimal link locations in HQ and BQ are placed on 0.5, and optimal link locations in TQ are 0.2. Aggregating across the entire queued area (i.e., "all"), the optimal link locations across the intervening segments falls in a range of 0.3 to 0.5. This result indicates that the magnitude of estimation error in each of three segment types has an effect on determining optimal link location for the entire queued area, i.e., it is effectively a weighted average of the op-

timal link location from the intervening segments. If the error in TQ is relatively large, it will cause the optimal link location to be smaller.

5.4 Analysis – Half Detector Station Spacing

To analyze the effect of detector station spacing the study considered the error in the event that there were additional detector stations in between the existing stations. Or more generally, what should one expect with a higher density of detector stations. Figure 5.10 shows the scope of queuing for two hypothetical examples, delineating both the existing detector stations and the "New" detector stations on the right-hand side of the figure, while the resulting distance between stations is shown on the left-hand side. In the absence of the availability of more precise positioning information, the shape of the queue is assumed to remain unchanged from the original detector spacing. So v_1 and v_2 still capture the growth and decay of the queue from station i+2 in Figure 5.10, the signals may propagate with a different velocity in the one or two links upstream of this station, e.g., as shown in the left-hand side of the figure.

In contrast to Figure 5.6, Figure 5.11 shows estimation errors for the same incident given half detector station spacing in each of the three segment types as a function of IL and LL. In the case of half detector station spacing, even when new detector stations are added, each of three segments is considered to hold the same boundary with the existing spacing, i.e., the range of the physical location of the incident come from the original spacing. So when a link between two stations is split the analysis allows an incident to fall in either of the new halves. The subplots in Figure 5.11A are indexed at the top by the original IL, while IL relative to the new half spacing is marked with a dashed vertical line in each subplot.[7] Thus, as one progresses through the subplots the IL relative to the new half spacing sweeps from 0 to 1 twice. Likewise, the end of queue comes from the original spacing, the subplots in Figure 5.11C are indexed by the original EQ while the location of EQ under the new half spacing is marked with a dashed vertical line in each subplot. As with the original spacing, the estimation errors in HQ is minimized when LL=IL, the sum of estimation errors in BQ is minimized when LL=0.5, and the sum of estimation errors in EQ is minimized when LL=0.5*EQ, where IL and EQ are now measured relative to the new half spacing.

Repeating the analysis from Figure 5.7-5.8, only with the new half spacing between detectors, the points in Figures 5.12-5.13 show the resulting error for each LL for a given simulation run (sum of the absolute value of overestimates and underestimates) while the curve shows the average at each LL. Again note that the vertical scale changes from plot to plot. Once more, the average error is minimized at 0.5 for HQ and BQ. There is a slight change in TQ and net error, the average error is minimized at 0.3 for TQ and summing across the three types of segments, the lower right hand plot shows that the net error is minimized when LL is 0.4.

[7] So an incident with IL=0.1 or IL=0.6 in the original spacing will both result in an IL=0.2 in the new half spacing, but the two cases are shown in separate subplots.

5.4.1 Analytical Results for Half Detector Station Spacing

When the original HQ is split, IL can be upstream or downstream of the new detector station. When IL is downstream of the new station, there are two links over the section of freeway comprising the original HQ, i.e., the links associated with D_i' and D_i'' in Figure 5.14. The upstream of the two stations (D_i'') is simply BQ in the present spacing. As already shown in both HQ and BQ, the optimal LL is 0.5. In other words,

$$LL^* = \overline{IL} = \frac{1}{2} D_i' \tag{5.11}$$

Once more, because the values are discretized to 1/20ths and IL=0 is excluded, the empirical results for the optimal LL in HQ should be slightly larger than 0.5. On the other hand, when IL is upstream of the new detector station, once more there are two links over the section of freeway comprising the original HQ as shown in Figure 5.15. By definition the downstream of the two links is beyond the influence of the incident and thus, is excluded. For the new upstream link the optimal LL remains 0.5. In other words,

$$LL^* = \overline{IL} = \frac{1}{2} D_i'' \tag{5.12}$$

In either case, the optimal LL in the HQ under half spacing is half of the new detector spacing. As already noted, the analysis in BQ does not change, only the distances of the links is reduced and thus, the magnitude of the expected error.

When the original TQ is split, EQ can be upstream or downstream of the new detector station. When EQ is downstream of the new detector station, as shown in Figure 5.16, the upstream link does not contribute to the error and the analysis for the downstream link remains the same as in the original spacing analysis, only the distances of the link is reduced and thus, the optimal LL is,

$$LL^* = \frac{1}{2}\overline{EQ} = \frac{1}{4} D_{i+2}' \tag{5.13}$$

On the other hand, EQ could fall upstream of he new detector station, as shown in Figure 5.17. Because the shape of the queue is assumed to remain unchanged from the original detector spacing,

$$\begin{aligned} &\left|Overestimation\ error\right| + \left|Underestimation\ error\right| \\ &= \frac{t}{2} \cdot \frac{1}{1+EQ} \cdot \left(2 \cdot LL^2 + (1-LL)^2 + (EQ-LL)^2\right) \\ &= \frac{t}{2} \cdot \frac{1}{1+EQ} \cdot \left(4 \cdot LL^2 - 2 \cdot (1-EQ) \cdot LL + 1 + EQ^2\right) \end{aligned} \tag{5.14}$$

where t is the time that detector station i+2 actually sees congested traffic. Taking the first derivative relative to LL and setting it equal to zero yields LL=0.25*(1+EQ), while

the second derivative is greater than zero at this point, indicating that it is a minimum. Over many samples,

$$\overline{EQ} = \frac{D_{i+2}^{''}}{2} \tag{5.15}$$

and

$$LL^{*} = \frac{1}{4}\left(1 + \overline{EQ}\right) = \frac{3}{8} \cdot D_{i+2}^{''} \tag{5.16}$$

Finally, dropping D from the equation, the average across the two types of links is,

$$LL^{*} = \frac{1}{2} \cdot \left(\frac{1}{4} + \frac{3}{8}\right) = 0.31 \tag{5.17}$$

5.4.2 Extending to Other Incidents

Once more, the analysis is repeated for all eight incidents listed in Table 5.3, with the corresponding detailed analysis for the remaining incidents is presented in Appendices E and F. For each incident the resulting optimal link location for the three segment types and the entire queued area in the half spacing case are summarized in Table 5.5. In all eight cases the optimal link locations in HQ and BQ are placed on 0.5 as before, but now the optimal link locations in TQ are 0.3, consistent with Equation 5.17. This shift is small and the fact that it exists at all appears to be an artifact from the way we handle TQ in the half spacing. Aggregating across the entire queued area (i.e., "all"), the optimal link locations across the intervening segments now falls in a smaller range of 0.4 to 0.5.

5.5. Analysis – Double detector Station Spacing

While the previous section considered half spacing between detector stations, this section examines the impacts of doubling the spacing between the existing detector spacing, i.e., performance with a lower density of detectors. Once more the size and shape of the actual queue is taken from the original detector spacing. So the location of HQ and TQ will be bound by the complete set of stations, but now, the error is quantified as if every-other detector station was not available. Provided the queue is long enough to span three or more stations there will be two possibilities, either there will be an even number of stations in the original BQ (i.e., normal spacing) or there will be an odd number in BQ. If it is even, either HQ or TQ (but not both) will start at a new location. If there are an odd number of stations in the original BQ either, HQ and TQ will start at a new location, or both will start at the same location as in the original set of stations. The condition of the start of HQ is used to separate each of two cases into two cases, either a New Head

Boundary (NHB) or keep the Present Head Boundary (PHB).[8] Figure 5.18 shows queue boundary of present spacing, even BQ double spacing, and odd BQ double spacing.

Figures 5.19-5.20 show the HQ estimation errors in the presence of NHB and PHB, respectively. The two left-hand plots in each figure show the errors under current spacing while the two right-hand plots show the errors at double spacing (dashed lines indicate LL). Figure 5.21 shows the BQ estimation error, while Figures 5.22-5.23 show the TQ estimation errors with the present tail boundary (PTB) and new tail boundary (NTB)[9], which in practice would be a function of even or odd and NHB or PHB. Note how there are two link boundaries shown in the left-hand side for Figures 5.20-5.22, while in the other two figures there is only one due to the fact that the given link is expanded beyond the queue. In all five figures the right-hand side shows a single boundary after merging two links.

In contrast to Figure 5.6, Figure 5.24 shows estimation errors for the same incident given double detector station spacing, NHB, in each of the three segment types as a function of IL and LL. As before, the range of the physical location of the incident comes from the original spacing. The subplots in Figure 5.24A are indexed at the top by the original IL, while IL relative to the new double spacing is marked with a dashed vertical line in each subplot. In the case of NHB, since the size and shape of the assumed queued region does not change, since the original HQ fell in the upstream of the two links being merged, by definition the queue has to start in the upstream end of the new HQ, e.g., the new IL will be at least 0.5 if the two links are equal length, as reflected in Figure 5.24A. Since the queue resulting from the subject incident covers an odd number of stations, if there is a NHB, then there also must be a NTB, and the results of Figure 5.24C reflect this fact.[10] If the two links were equal length, then EQ would have to be less than 0.5 in the new link, however, in this case, from Figure 5.9, it is clear that the old TQ was in a link that was almost three times as long as the adjacent links (see Table 5.6). So in this case the EQ ranges between 0 and 0.7. The process is repeated in Figure 5.25 for the same incident with PHB and thus, PTB. The range of IL and EQ are constrained in a similar manner, albeit on the opposite half of the given link (now EQ is constrained between 0.3 and 1 due to the long original TQ). As before, in both figures the HQ error is minimized when LL=IL, BQ error is minimized when LL=0.5 and TQ error is minimized when LL=0.5*EQ.

Repeating the analysis from Figure 5.7-5.8 only with the new double spacing between detectors, the points in Figures 5.26-5.31 show the resulting error for each LL for a given simulation run (sum of the absolute value of overestimates and underestimates) while the curve shows the average at each LL. Once more, note that the vertical scale changes from plot to plot. Taking the average across NHB and PHB, the error is minimized at 0.7 for HQ, 0.5 for BQ, and 0.3 for TQ (from Figures 5.30-5.31).

[8] So the downstream end of HQ is new.
[9] So the upstream end of TQ is new.
[10] For NHB with PTB, or PHB with NTB, see incidents B, C, and D.

5.5.1 Analytical Results for Double Detector Station Spacing

When the original HQ is merged with an adjacent station, IL will fall in the downstream half if PHB and upstream half if NHB. Because the incident location is constrained in our model to fall between the original detector stations, in the case of NHB it would fall in the upstream half. Given unequal distances between detector stations, in reference to the example in Figure 5.18A,

$$LL^* = \overline{IL} = D_{i-1} + \frac{1}{2} D_i \tag{5.18}$$

where D_{i-1} and D_i are relative to the original detectors before combining (or doubling) the links. On the other hand, in the case of PHB, the incident falls in the downstream half, i.e.,

$$LL^* = \overline{IL} = \frac{1}{2} D_i \tag{5.19}$$

These results are consistent with the top left plots in Figures 5.26-5.31. In either case, the analysis in BQ does not change, only the distances of the links is increased and thus, the magnitude of the expected error.

When the original TQ is merged, EQ will be constrained by PTB or NTB, i.e., the original EQ will fall in the upstream or downstream half, respectively, of the combined link. For NTB, LL=0.5*EQ still holds, and in the case of Figure 5.18A (right-most plot),

$$\overline{EQ} = \frac{D_{i+4}}{2} \tag{5.20}$$

where D_{i+4} is relative to the original detectors before combining the links. For PTB the calculus needs to be revisited because EQ now includes a link that was formerly BQ, e.g., Figure 5.18A (center plot). First consider the case where the new LL falls in the same original link containing EQ, as shown in Figure 5.32. Now there is a fixed overestimation error due to the former downstream link, i.e., regions (1) and (2). The other overestimation error is also larger, as shown by the now trapezoidal regions (3) and (4). Finally, the underestimation region, (5), retains the same triangular form. Ignoring the constant overestimation error from (1) and (2), the sum of the remaining errors is:

$$\begin{aligned} &\left|Overestimation\ error\right| + \left|Underestimation\ error\right| \\ &= (b-a)h_1 + \frac{a}{2}\frac{h_1^2}{h} + \frac{a}{2}\frac{(h-h_1)^2}{h} \\ &= a\frac{h_1^2}{h} + (b-2a)h_1 + a\frac{h}{2} \end{aligned} \tag{5.21}$$

Taking the first derivative relative to h_1 and setting equal to zero, yields,

$$h_1 = h \cdot \left(1 - \frac{b}{2a}\right) \tag{5.22}$$

while the second derivative is greater than zero at this point, indicating that it is a minimum. If $1 < \frac{b}{a} < 2$,

$$LL^* = \bar{h}_1 + D_{i+3} = \frac{D_{i+4}}{2} \cdot \left(1 - \frac{b}{2a}\right) + D_{i+3} \tag{5.23}$$

but if $\frac{b}{a} > 2$, then LL^* falls in the downstream half of the link, as shown in Figure 5.33. Regions (1) and (2) still represent an over estimation, but now the net area is a function of h_1. While the underestimation error is split in to two parts, region (3) which varies with h_1 and region (4) which is fixed. Ignoring the constant region (4), the sum of the remaining errors is:

$$\begin{aligned} &|Overestimation\ error| + |Underestimation\ error| \\ &= \frac{1}{2}(b-a)\left(\frac{h_1^2}{h} + \frac{(h-h_1)^2}{h}\right) + a \cdot (h - h_1) \\ &= (b-a)\frac{h_1^2}{h} - b \cdot h_1 + (b+a)\frac{h}{2} \end{aligned} \tag{5.24}$$

Taking the first derivative relative to h_1 and setting equal to zero, yields,

$$h_1 = h \cdot \left(\frac{b}{2(b-a)}\right) \tag{5.25}$$

while the second derivative is greater than zero at this point, indicating that it is a minimum, and

$$LL^* = D_{i+3} \cdot \left(\frac{b}{2(b-a)}\right) \tag{5.26}$$

So because the geometry of the original queue is preserved, the optimal link location in PTB depends on a ratio of queue duration in the last two loop detector stations, i.e., b/a. Figure 5.34 shows this ratio for the eight incidents first presented in Table 5.3. Only incidents D and E have the ratio above 2. Using this analytical method the optimal link locations for the eight incidents are calculated and shown in Table 5.7.

5.5.2 Application of Double Detector Station Spacing

Once more, the analysis is repeated for all eight incidents listed in Table 5.3, with the corresponding detailed analysis for the remaining incidents is presented in Appendices H and I. Figure 5.35 shows the detector stations affected by the queue under original conditions, NHB and PHB. Five incidents have an odd number of original stations (A, E, F, G, and H) while the other three have an even number. Figure 5.35A includes all of the stations that are ultimately evaluated, extending one station past what was shown previously in Figure 5.9. Triangles represent new head boundary and tail boundary for double spacing. Asterisks show head boundary and tail boundary in present spacing, and circles are stations included in body in present spacing. The vertical axis represents detector station number and again the spacing is shown to scale. Figures 5.35B and 5.35C show the detector stations that are retained under NHB and PHB, respectively. Note that the stations are not equally spaced, in particular, stations 8 and 12 were excluded from the start due to chronic problems, as a result, their absence creates two larger than average gaps between stations.

After the simulations were conducted for each incident[11], the optimal link locations for each of the segment types and the sum over the entire queue are summarized in Table 5.8. The results are based on mile-hours in the queue and are presented based on NHB or PHB. These simulation results are consistent with the analytical results from Table 5.7. Table 5.9 shows the optimal link locations from NHB (e.g., Figure 5.26), PHB (e.g., Figure 5.28) and combined (e.g., Figure 5.30) for each of the segment types. In HQ the optimal LL is close to 0.5 for six of the eight incidents, with the remaining two (A and E) having the largest discrepancy between NHB and PHB. In BQ, all of the incidents had an optimal LL at 0.5. Finally, in TQ, all of the incidents had an optimal link location close to 0.3.

5.6 Analysis – Hypothetical Example

This section replicates the work above as applied to a hypothetical example where the original distance between detector stations is fixed at 1 mile apart. Using Figure 5.2 for reference, the queue is assumed to start between stations 5 and 6 (but not at station 6), and grows upstream, ending somewhere between stations 1 and 110 (but not at station 110). Based on the empirically observed wave speeds from the incidents, v_1 is set to -10 mph and v_2 to 10 mph. Finally, T is held constant at 60 min. As has already been done in the previous sections, IL and EQ are estimated assuming their respective locations are uniformly distributed between the two bounding detector stations.

5.6.1 Present Loop Spacing

Figure 5.36 shows the estimation errors for the incident in each of the three segment types as a function of IL and LL. For reference, IL and EQ are marked as dashed vertical

[11] IL and EQ were selected relative to the segments bounding the original HQ and TQ, respectively, i.e., the feasible range is the same as was used earlier when considering the original spacing. As before, however, there were 220 simulation runs for HQ BQ, and TQ, while there were 4,400 simulation runs for the total combined queue.

lines in Figure 5.36A and C. As with the empirical example, the sum of estimation errors in HQ is minimized when LL=IL, the sum of estimation errors in BQ is minimized when the link boundary is placed midway between stations, and the sum of estimation errors in TQ is minimized when LL=0.5*EQ. Holding v_1, v_2, and T constant, once more, 220 pairs of IL and EQ were generated in each of the three segment types, while another 4,400 pairs were generated for the combined error across all links. The resulting errors for each of the simulation runs were calculated (sum of the absolute value of overestimates and underestimates), as shown by the points in Figure 5.37 while the curve shows the average at each LL. Once more, note that the vertical scale changes from plot to plot. The average error is minimized at 0.5 for HQ and BQ, while it is minimized at 0.2 for TQ. Summing across the three types of segments, the lower right hand plot shows that the net error is minimized when LL is 0.45, reflecting a smaller influence from TQ in the sum across all segments. This deviation from the empirical incidents arises in part from the fact that all links are now held constant, while in roughly half of the empirical incidents TQ was twice as long as the other links.

5.6.2. Half Loop Detector Station Spacing

Figure 5.38 shows the estimation errors for the same incident given half detector station spacing in each of the three segment types as a function of IL and LL. As before the range of the physical location of the incident come from the original spacing. The subplots in Figure 5.38A are indexed at the top by the original IL, while IL relative to the new half spacing is marked with a dashed vertical line in each subplot. Thus, as one progresses through the subplots the IL relative to the new half spacing sweeps from 0 to 1 twice. Likewise, the end of queue comes from the original spacing, the subplots in Figure 5.38C are indexed by the original EQ while the location of EQ under the new half spacing is marked with a dashed vertical line in each subplot. As with the original spacing, the estimation errors in HQ is minimized when LL=IL, the sum of estimation errors in BQ is minimized when LL=0.5, and the sum of estimation errors in EQ is minimized when LL=0.5*EQ, where IL and EQ are now measured relative to the new half spacing. Repeating the analysis from Figure 5.37 only with the new half spacing between detectors, the points in Figure 5.39 shows the resulting error for each LL for a given simulation run (sum of the absolute value of overestimates and underestimates) while the curve shows the average at each LL. Once more, the average error is minimized at 0.5 for HQ and BQ. There is a slight change in TQ and net error, the average error is minimized at 0.3 for TQ and summing across the three types of segments, the lower right hand plot shows that the net error is minimized when LL is 0.5.

5.6.3. Double Loop Detector Station Spacing

Figure 5.40 shows the estimation errors for the same incident given double detector station spacing, NHB, in each of the three segment types as a function of IL and LL. As before, the range of the physical location of the incident comes from the original spacing. The subplots in Figure 5.40A are indexed at the top by the original IL, while IL relative to the new half spacing is marked with a dashed vertical line in each subplot. As with the empirical analysis the new IL must be at least 0.5 for NHB. Since the queue resulting from the subject incident covers an odd number of stations, if there is a NHB, then there

also must be a NTB, and the results of Figure 5.40C reflect this fact, i.e., that EQ must be less than 0.5 in the new link. The process is repeated in Figure 5.41 for the same incident with PHB and thus, PTB. The range of IL and EQ are constrained in a similar manner, albeit on the opposite half of the given link. As before, in both figures the HQ error is minimized when LL=IL, BQ error is minimized when LL=0.5 and TQ error is minimized when LL=0.5*EQ. Figures 5.42-5.43 show the resulting error for each LL for a given simulation run (sum of the absolute value of overestimates and underestimates) while the curve shows the average at each LL. Once more, note that the vertical scale changes from plot to plot. Taking the average across NHB and PHB, the error is minimized at 0.5 for HQ and BQ, 0.3 for TQ.

5.6.4 Vehicle Hours of Delay

Given the fact that this incident is a hypothetical example, it becomes more difficult to translate mile-hours into vehicle-hours per lane. With a moderate density of 50 vehicles per mile per lane (jam density is around 200 vehicles per mile per lane), the results of Figure 5.37 would result in an error between 25 and 50 vehicle-hours of delay per lane. This expected error drops roughly by a factor of two for half detector station spacing and increases roughly by a factor of two for double detector station spacing. In all three cases choosing LL at 0.5 either minimizes the expected error or it is very close to this optimal location.

5.7 Discussion and Conclusion

This work examined the response time to detecting incidents or the activation of any bottleneck and the errors in estimating traveler delay in the resulting queue. This research focused exclusively on queues that once they began discharging, receded back to the bottleneck, e.g., Figure 5.1. With only modest effort, the reader could employ the principles presented herein and extend the work to other queues, e.g., one in which an obstruction is cleared and the queue dissipates upstream, e.g., Figure 5.44.

The response time is basically a function of IL, v_1, and any latency in the detection system (e.g., one may have to wait one or two sample periods to differentiate between noise and a real event). The empirically observed queues studied in this work grew with velocities ranging between 2 and 10 mph. For slower growing queues (2.5 mph) with detector stations spaced 2 miles apart, the average time to detection is over 20 minutes. While faster growing queues (10 mph) with detector stations every third of a mile have a time to detection on the order of a minute. These results are independent of LL, i.e., they depend only on how long it takes the queue to reach the first detector station upstream of the bottleneck.

Several incidents were examined in detail to establish the impacts of IL, EQ, LL, detector spacing, etc. The left-hand column of Figure 5.45 shows the average estimation error delay (vehicle hours per lane) for each incident, under the three conditions, original spacing, half spacing, and double spacing. The right-hand column shows the corresponding ratio relative to the original spacing. The error drops roughly by 50 percent when going to half spacing and almost doubles when going to double-spacing. As discussed above,

the fact that the error does not quite double at the double-spacing may be an artifact of our modeling assumptions. Figure 5.46B takes the average ratio across the 11 LL considered in the right-hand column of Figure 5.45 while Figure 5.46A repeats the exercise using mile-hours. Again, we see that at double-spacing, most of the incidents see slightly less than twice the error under original conditions. The one notable exception is incident E, in this case the queue boundary is station 11, when adding another link we skipped over the problematic station 12 and went to station 13, the larger than average jump also resulted in a larger error in Figure 5.46A. Tables 5.10-5.12 tabulate the minimum and maximum error (in vehicle hours per lane) for the eight incidents under the original spacing, half spacing and double spacing. Finally, Table 5.13 shows that the magnitude of the estimation error ranges between 20 and 150 percent of the queued region.

While the magnitude of the errors might change from one facility to the next, the relative impacts should stay the same, i.e., absolute errors in delay roughly doubles as one doubles the spacing between detector stations. If one allowed overestimation errors to cancel underestimation errors, the net result would be much smaller. In other words, the simple assumption that conditions at a detector station are representative of an entire link is inherently prone to errors. The analysis suggests a slightly more sophisticated model could be used to better quantify delay if higher precision were desired, e.g., replicating our efforts to measure v_1 and v_2 in real time and then modeling the queue as being triangular in the time space plane rather than "stair stepped."

While the response time is independent of LL, the magnitude of the delay estimation errors definitely does depend on LL. Putting LL at 0.5, i.e., defining the boundaries between links to be at the midpoint between detector stations, either minimizes the error or comes very close to the minimum estimation error in all of the cases examined. Given the fact that incident locations are not known a priori, it does not make sense to try to pull marginal performance gains by splitting the links a small distance away from the midpoint. Furthermore, since many links are bidirectional, it becomes difficult to select a link boundary other than at the midpoint or at one of the detector stations. As shown above, placing LL at the detector stations actually maximizes the net error. Though from Figures like 5.6, it should be evident that if an agency wished to minimize overestimation or underestimation in a specific link (e.g., HQ) at the expense of the net error, placing the division between links at the detector stations could be optimal under these special conditions.

References for Part II

[1] Coifman, B., *The Columbus Metropolitan Freeway Management System (CMFMS) Effectiveness Study: Part 2 - The After Study*, Ohio Department of Transportation, 2006.

[2] Coifman, B., Dhoorjaty, S., 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.

Tables for Chapter 5

v_1 (mph)	Time for queue to grow one mile (min)	Average response time (min) for detector station spacing of				Maximum response time (min) for detector station spacing of			
		1/3 mi	1/2 mi	1 mi	2 mi	1/3 mi	1/2 mi	1 mi	2 mi
2.5	24	4	6	12	24	8	12	24	48
5	12	2	3	6	12	4	6	12	24
10	6	1	1.5	3	6	2	3	6	12

Table 5.1, Average and maximum response time for a queue to reach a detector station for the given queue growth velocity and detector spacing.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
6	-	-	24594	-	-	Free Flow
5	12:01:00	13:11:30	22819	-	-	
4	12:02:30	13:10:00	20829	-15.1	11.3	St 5 – St 4
3	12:05:30	13:09:00	19114	-6.5	19.5	St 4 – St 3
2	12:13:00	13:06:00	17114	-3.0	7.6	St 3 – St 2
1	12:16:00	13:04:30	15766	-5.1	10.2	St 2 – St 1
110	-	-	12055	-	-	Free Flow

Table 5.2, Wave speed from incident data on April 23, 2005 (NB). T_1 represents an observed time that speed drops below 45 mph. T_2 represents an observed time that speed is over 45 mph.

Index	Date	mile-hr			
		HQ	BQ	TQ	All
A	April 23, 2005	0.5	0.5	0.2	0.3
B	September 17, 2005	0.5	0.5	0.2	0.3
C	May 17, 2005	0.5	0.5	0.2	0.3
D	October 12, 2004	0.5	0.5	0.2	0.4
E	July 23, 2004	0.5	0.5	0.2	0.5
F	September 23, 2005	0.5	0.5	0.2	0.5
G	January 3, 2005	0.5	0.5	0.2	0.5
H	March 26, 2005	0.5	0.5	0.2	0.5

Table 5.3, Optimal link location at head of queue (HQ), body of queue (BQ), and tail of queue (TQ) for present spacing

Wave speed (mph)	A	B	C	D	E	F	G	H
Queue growth	-5.3	-2.8	-10.5	-3.9	-3.4	-2.9	-5	-8.7
Queue recession	11.4	10.5	10.5	6.1	19.2	9.6	8.2	5.4

Table 5.4, Wave speeds from the eight incidents for queue growth and recession

Index	Date	mile-hr			
		HQ	BQ	TQ	All
A	April 23, 2005	0.5	0.5	0.3	0.4
B	September 17, 2005	0.5	0.5	0.3	0.4
C	May 17, 2005	0.5	0.5	0.3	0.4
D	October 12, 2004	0.5	0.5	0.3	0.5
E	July 23, 2004	0.5	0.5	0.3	0.5
F	September 23, 2005	0.5	0.5	0.3	0.5
G	January 3, 2005	0.5	0.5	0.3	0.5
H	March 26, 2005	0.5	0.5	0.3	0.5

Table 5.5, Optimal link location at head of queue (HQ), body of queue (BQ), and tail of queue (TQ) for half detector station spacing

Station		Distance (ft)
From	To	
2	1	1348
1	110	3711
110	109	1398

Table 5.6, Distance between stations in TQ of incident A.

Index	New Head Boundary (NHB)			Present Head Boundary (PHB)		
	HQ	BQ	TQ	HQ	BQ	TQ
A (odd)	0.81	0.50	0.18*	0.24	0.50	0.43
B (even)	0.74	0.50	0.41	0.27	0.50	0.18*
C (even)	0.74	0.50	0.42	0.27	0.50	0.18*
D (even)	0.72	0.50	0.34	0.25	0.50	0.18*
E (odd)	0.86	0.50	0.13*	0.24	0.50	0.44
F (odd)	0.68	0.50	0.12*	0.36	n/a	0.48
G (odd)	0.74	0.50	0.15*	0.27	0.50	0.47
H (odd)	0.72	0.50	0.15*	0.24	0.50	0.53

Table 5.7, Analytical results for optimal link location under double loop spacing. Data associated with new tail boundary (NTB) are marked as asterisks.

Index	New Head Boundary (NHB)				Present Head Boundary(PHB)			
	HQ	BQ	TQ	All	HQ	BQ	TQ	All
A (odd)	0.8	0.5	0.1*	0.7	0.2	0.5	0.4	0.3
B (even)	0.7	0.5	0.4	0.6	0.3	0.5	0.1*	0.2
C (even)	0.7	0.5	0.4	0.6	0.3	0.5	0.1*	0.2
D (even)	0.7	0.5	0.3	0.5	0.3	0.5	0.1*	0.3
E (odd)	0.9	0.5	0.1*	0.8	0.2	0.5	0.4	0.3
F (odd)	0.7	0.5	0.1*	0.6	0.4	n/a	0.5	0.4
G (odd)	0.7	0.5	0.1*	0.6	0.3	0.5	0.5	0.4
H (odd)	0.7	0.5	0.1*	0.6	0.2	0.5	0.5	0.3

Table 5.8, Experimental results for optimal link location under double loop spacing based on mile-hours in the queue. Data associated with new tail boundary (NTB) are marked as asterisks.

Index	Head			Body			Tail			Whole boundary		
	NHB	PHB	All	NHB	PHB	All	NHB	PHB	All	NHB	PHB	All
A (odd)	0.8	0.2	0.7	0.5	0.5	0.5	0.1*	0.4	0.3	0.7	0.3	0.4
B (even)	0.7	0.3	0.5	0.5	0.5	0.5	0.4	0.1*	0.3	0.6	0.2	0.4
C (even)	0.7	0.3	0.5	0.5	0.5	0.5	0.4	0.1*	0.3	0.6	0.2	0.4
D (even)	0.7	0.3	0.5	0.5	0.5	0.5	0.3	0.1*	0.3	0.5	0.3	0.4
E (odd)	0.9	0.2	0.8	0.5	0.5	0.5	0.1*	0.4	0.3	0.8	0.3	0.6
F (odd)	0.7	0.4	0.6	0.5	n/a	0.5	0.1*	0.5	0.3	0.6	0.4	0.5
G (odd)	0.7	0.3	0.5	0.5	0.5	0.5	0.1*	0.5	0.3	0.6	0.4	0.5
H (odd)	0.7	0.2	0.5	0.5	0.5	0.5	0.1*	0.5	0.4	0.6	0.3	0.5

Table 5.9, Experimental results for optimal link location under double loop spacing based on mile-hours in the queue, averaging across NHB and PHB. Data associated with new tail boundary (NTB) are marked as asterisks.

Index	Head		Body		Tail		Whole	
	Min	Max	Min	Max	Min	Max	Min	Max
A	3.7	7.5	1.3	2.7	3.6	16.1	10.0	25.6
B	2.7	5.4	1.3	2.7	2.5	11.0	7.6	18.3
C	1.1	2.2	0.3	0.7	0.3	1.3	1.9	4.1
D	2.7	5.2	1.2	2.5	0.7	3.1	5.0	10.3
E	1.5	3.1	2.4	4.9	0.1	0.4	4.1	8.1
F	6.7	11.9	0.7	1.6	0.5	2.3	8.4	14.9
G	3.7	7.5	2.6	6.0	0.1	0.3	6.5	13.1
H	3.9	7.7	1.5	3.3	0.1	0.3	5.5	10.6
average	3.3	6.3	1.4	3.1	1.0	4.3	6.1	13.1

Table 5.10, Min and max estimation error delay (veh-hr per lane) in present loop detector station spacing

Index	Head		Body		Tail		Whole	
	Min	Max	Min	Max	Min	Max	Min	Max
A	1.9	4.0	0.7	1.3	2.4	7.0	5.3	11.7
B	1.4	3.0	0.6	1.3	1.6	4.8	4.0	8.5
C	0.6	1.2	0.2	0.3	0.2	0.6	1.0	2.0
D	1.4	2.9	0.6	1.2	0.5	1.4	2.6	5.1
E	0.8	1.7	1.2	2.4	0.1	0.2	2.1	4.2
F	3.9	7.3	0.4	0.8	0.3	1.0	4.6	8.5
G	1.9	4.1	1.3	2.8	0.0	0.1	3.3	6.6
H	2.1	4.3	0.8	1.6	0.0	0.1	2.9	5.8
average	1.7	3.6	0.7	1.5	0.6	1.9	3.2	6.5

Table 5.11, Min and max estimation error delay (veh-hr per lane) in half loop detector station spacing

Index	Head		Body		Tail		Whole	
	Min	Max	Min	Max	Min	Max	Min	Max
A	3.7	30.4	1.2	4.4	3.6	24.1	10.9	42.5
B	2.8	14.2	1.2	2.9	2.5	16.4	8.2	31.6
C	1.2	5.9	0.5	1.2	0.3	2.0	2.5	8.3
D	2.7	13.1	1.8	4.5	0.7	8.7	8.0	22.3
E	1.6	16.9	2.1	7.7	0.1	1.2	5.3	24.7
F	6.7	22.2	0.0	4.6	0.5	5.7	10.5	27.5
G	3.8	19.7	4.9	13.1	0.1	0.7	10.8	32.9
H	3.9	20.5	1.6	8.0	0.1	0.5	8.7	26.4
average	3.3	17.9	1.7	5.8	1.0	7.4	8.1	27.0

Table 5.12, Min and max estimation error delay (veh-hr per lane) in double loop detector station spacing

"area" of queue (mi-hr)	Error over queue (mi-hr)		Error over queue (%)	
	Min	Max	Min	Max
1.3247	0.27	0.69	20%	52%
0.7511	0.24	0.56	32%	75%
0.3356	0.11	0.25	32%	75%
0.4563	0.15	0.32	34%	71%
0.7255	0.15	0.29	20%	40%
0.2969	0.24	0.43	80%	146%
1.095	0.18	0.35	17%	32%
0.7084	0.17	0.33	24%	47%

Table 5.13, Min and max estimation error delay (mi-hr) in present loop detector station spacing

Figures for Chapter 5

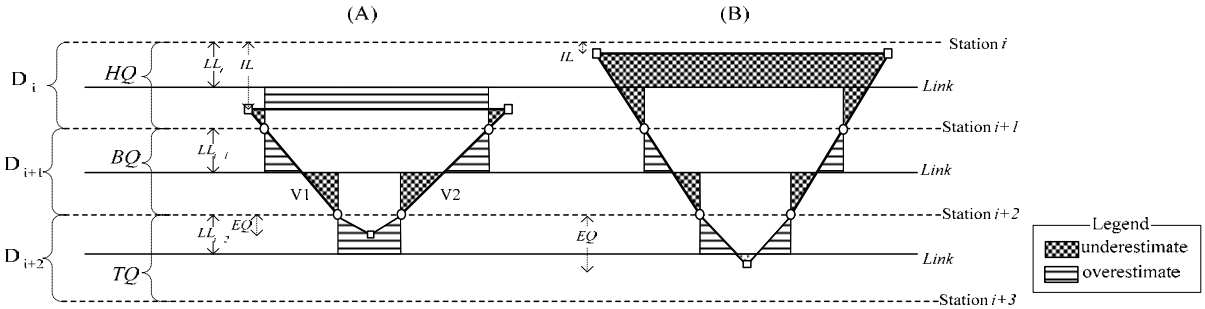


Figure 5.1, Queue boundary and estimation errors in general case

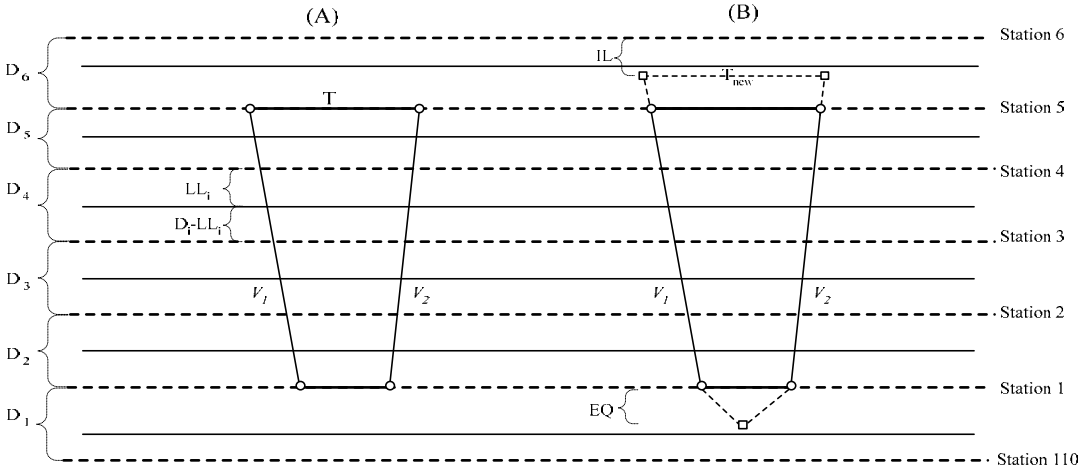


Figure 5.2, A) Queue boundary in given average wave speed and queue duration time, B) extended queue boundary with location of incident and end of queue

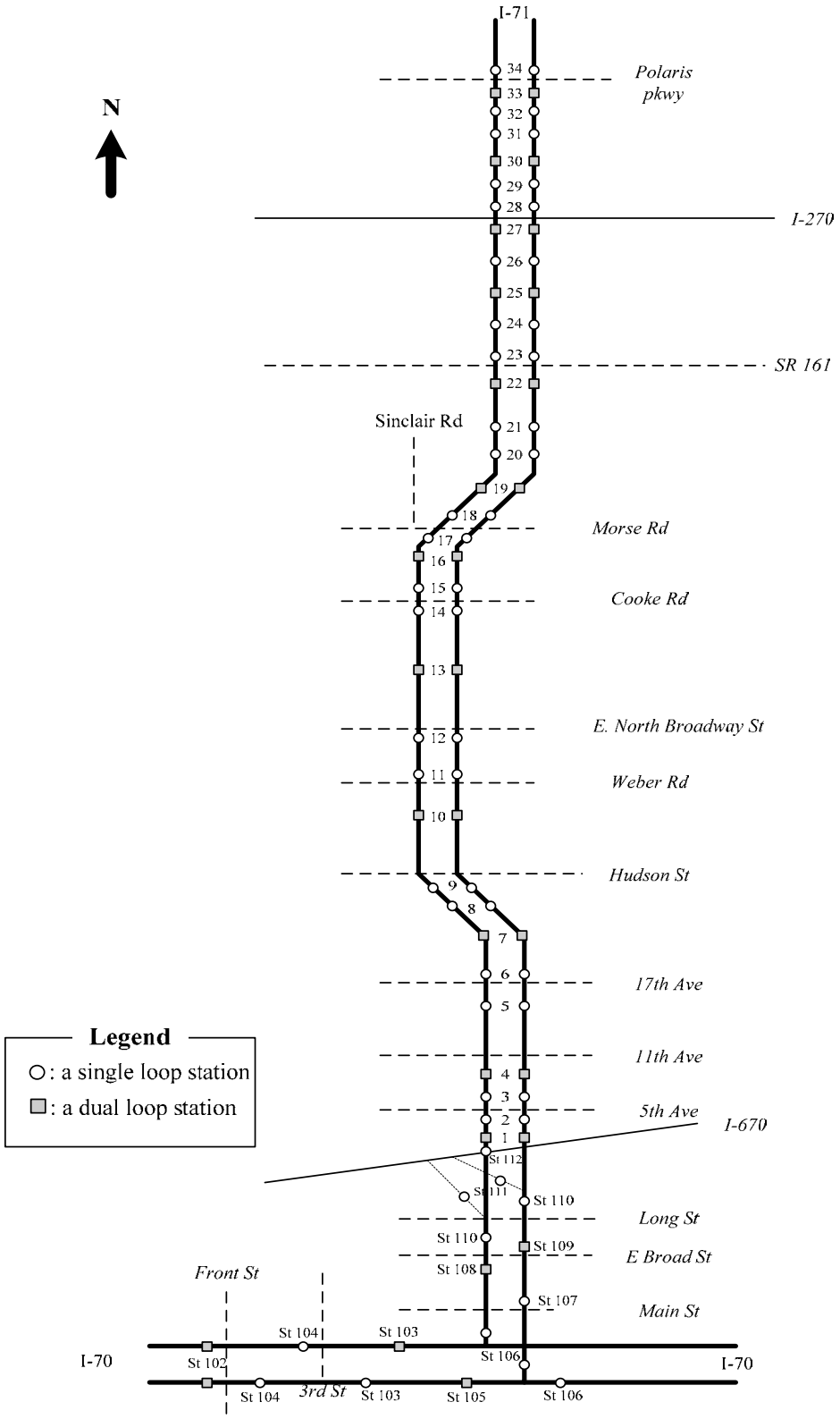


Figure 5.3, A schematic of the I-71 corridor

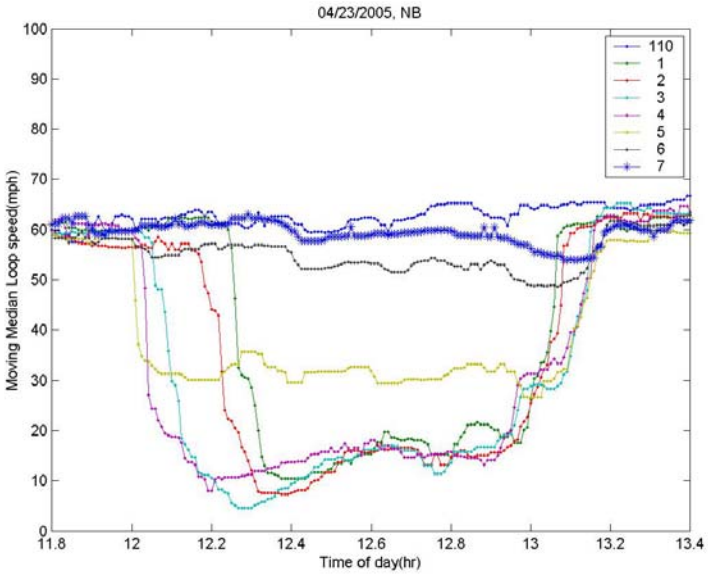


Figure 5.4, Moving median loop speed trends over short time periods.

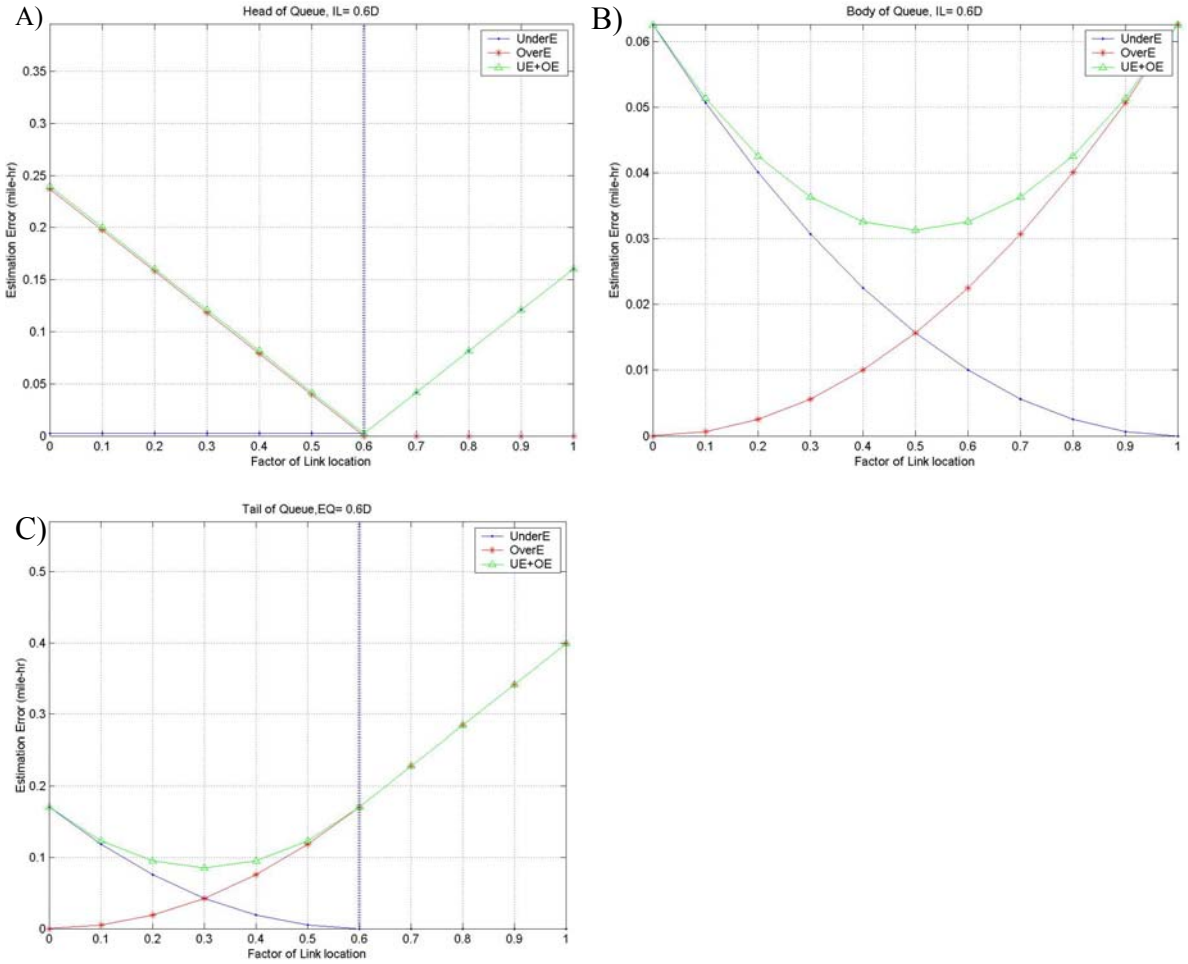


Figure 5.5, Estimation errors (mile-hr) in each of three segments when a ratio of incident or end of queue and detector spacing is 0.6. A) Head of queue, B) Body of queue, C) Tail of queue.

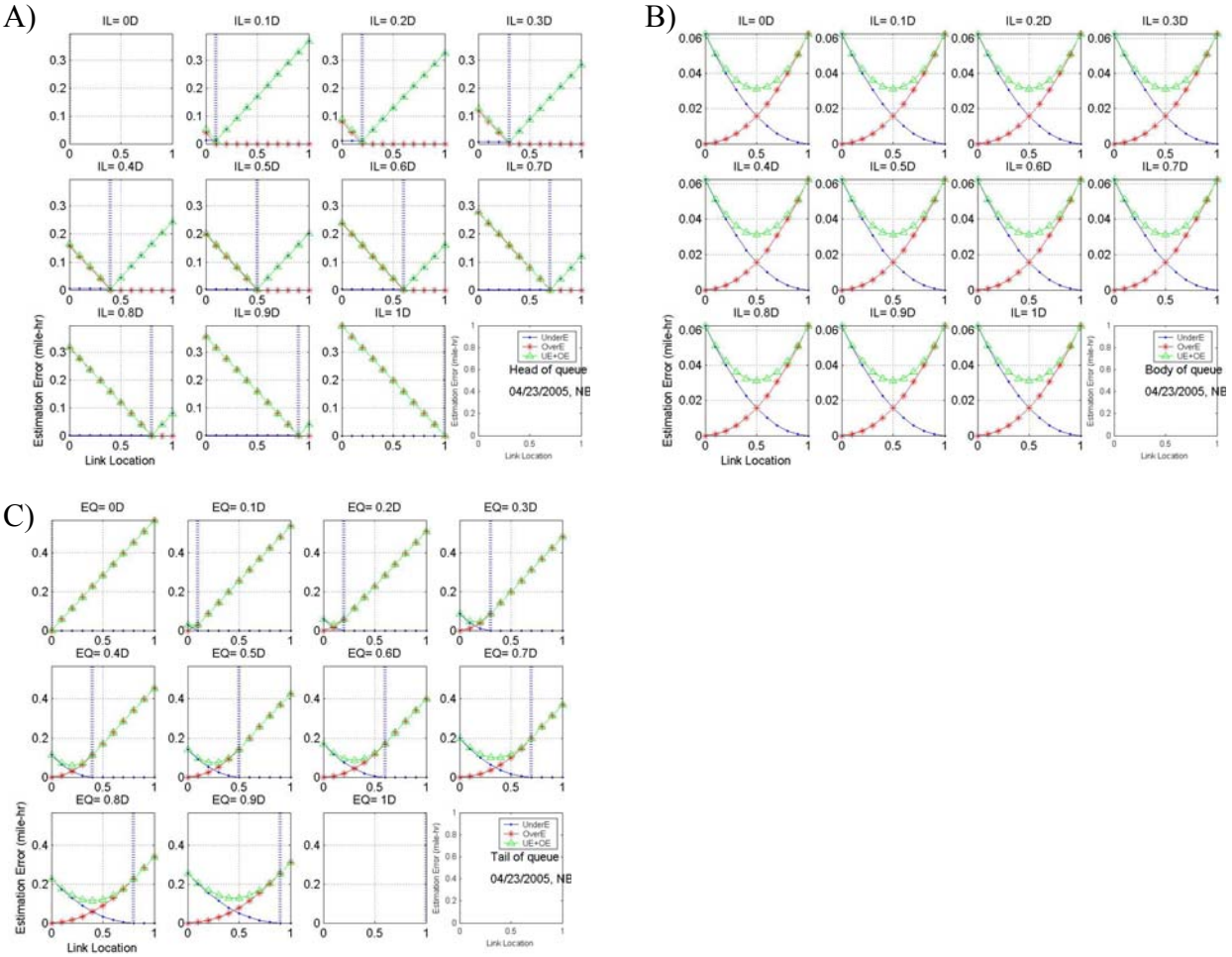


Figure 5.6, Estimation error in each of three segments by link locations, given location of incident and end of queue A) Head of queue, B) Body of queue, C) Tail of queue

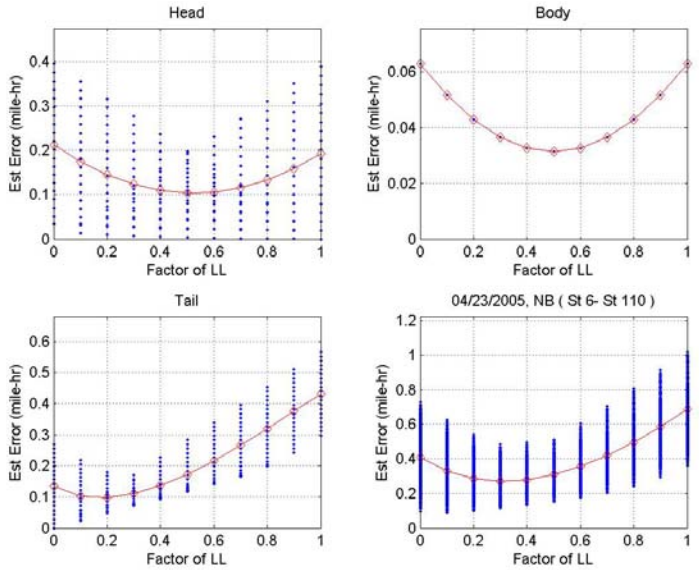


Figure 5.7, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours).

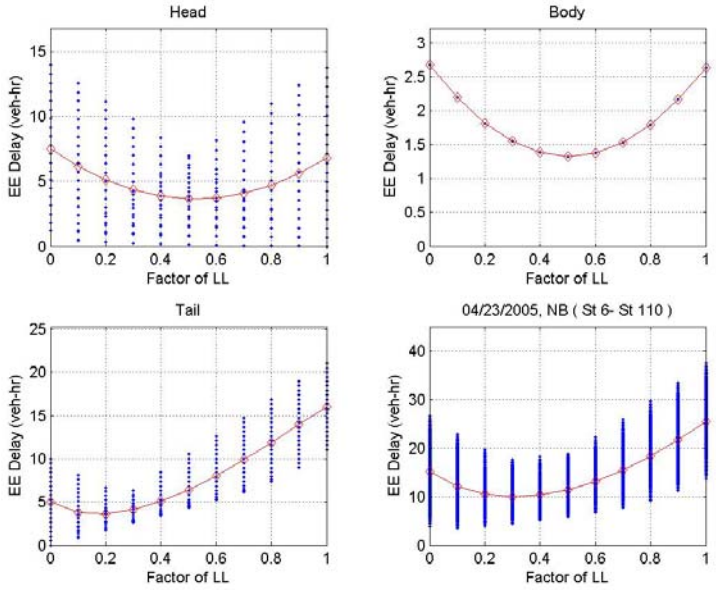


Figure 5.8, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in vehicle-hours per lane).

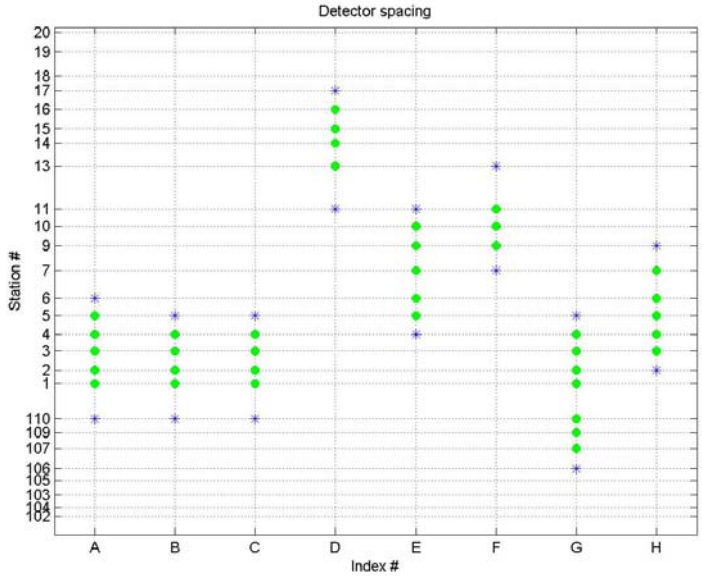


Figure 5.9, Location of detector stations

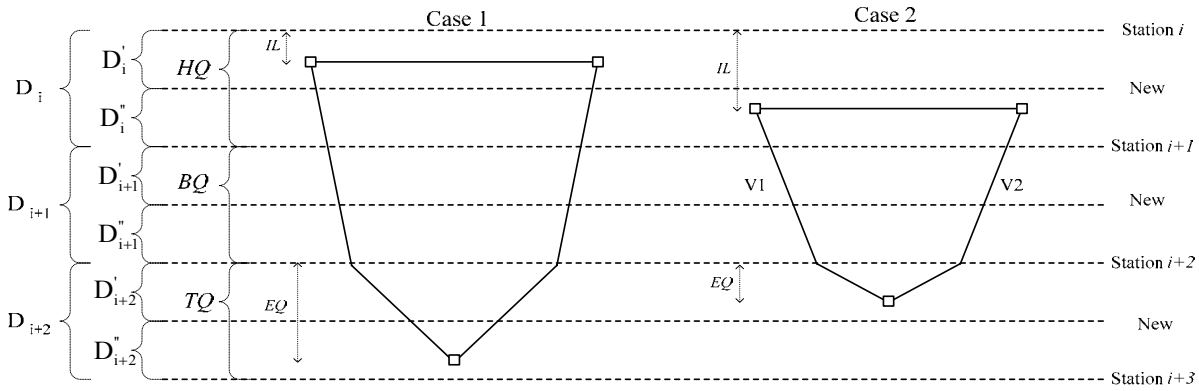


Figure 5.10, Queue boundary in half detector station spacing

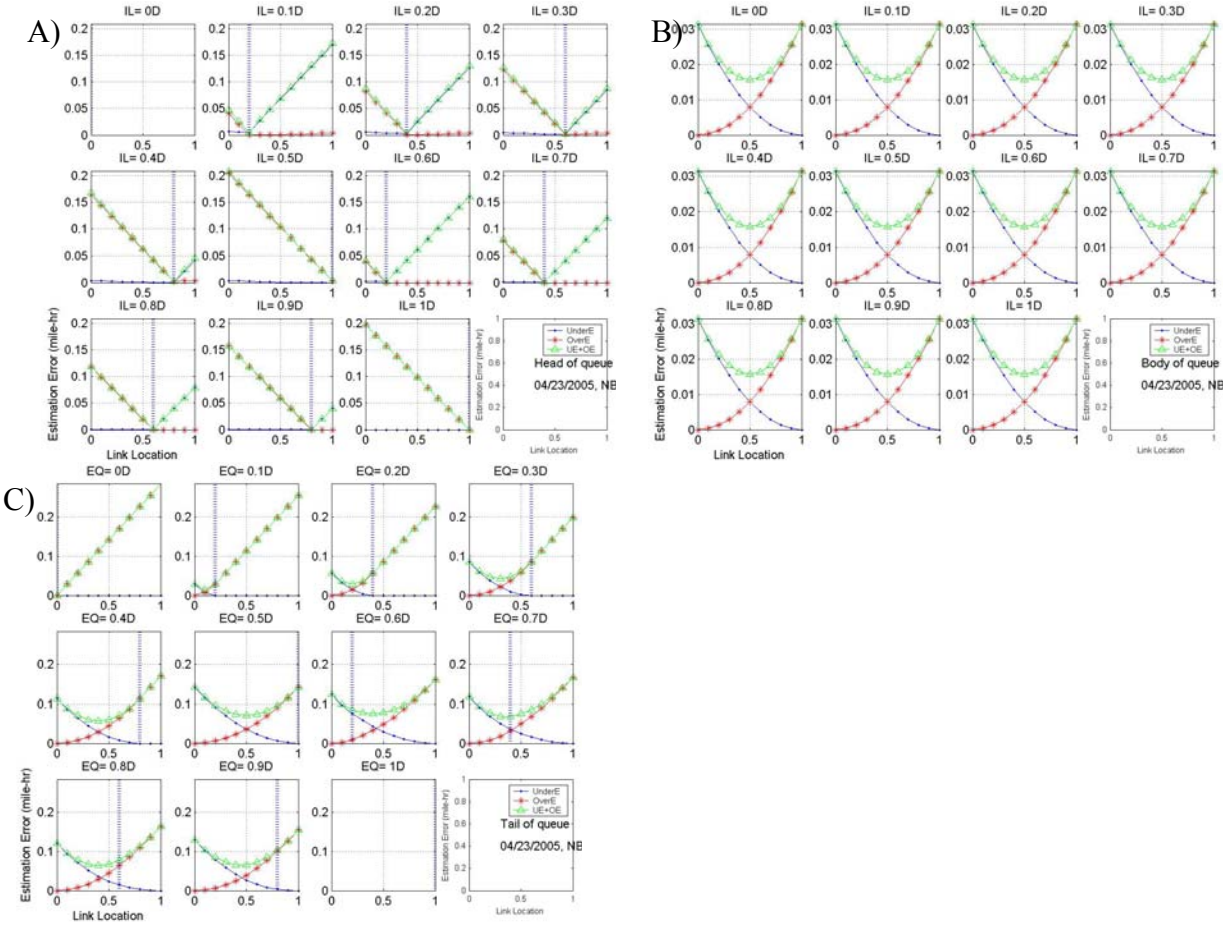


Figure 5.11, Estimation error in each of three segments by link locations for half detector station spacing, given location of incident and end of queue A) Head of queue, B) Body of queue, C) Tail of queue

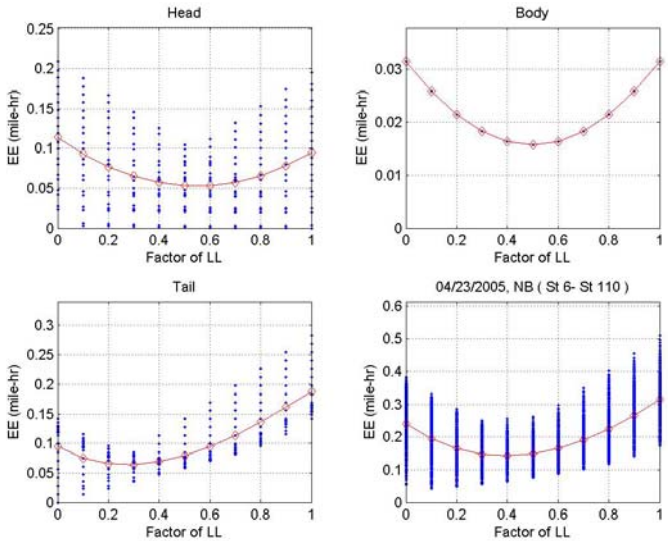


Figure 5.12, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours) for half detector station spacing.

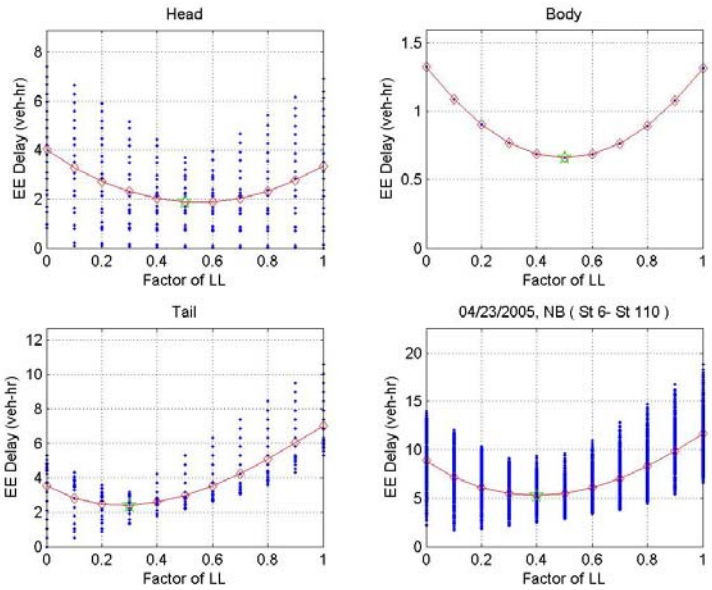


Figure 5.13, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in vehicle-hours per lane) for half detector station spacing.

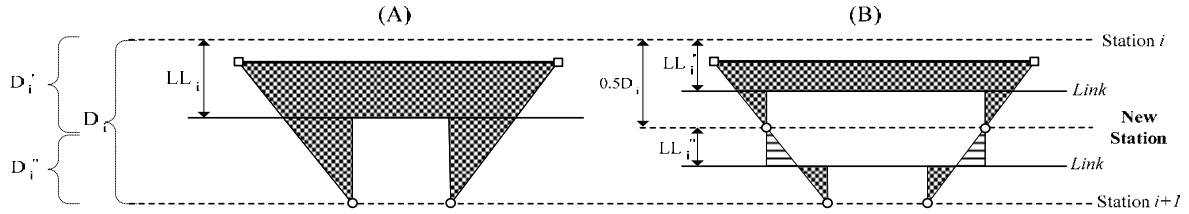


Figure 5.14, (A) Original detector station spacing, (B) half detector spacing for the same IL, with the incident downstream of the new detector station.

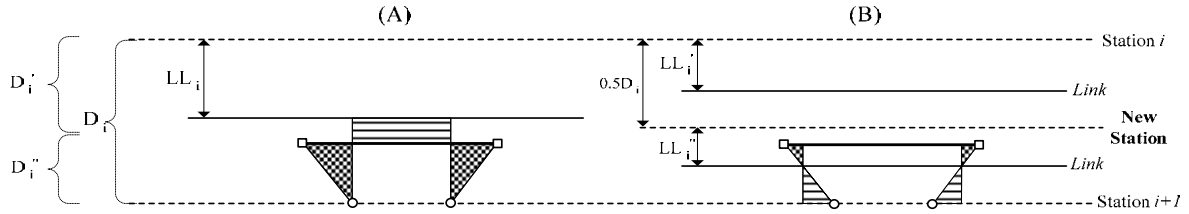


Figure 5.15, (A) Original detector station spacing, (B) half detector spacing for the same IL, with the incident upstream of the new detector station.

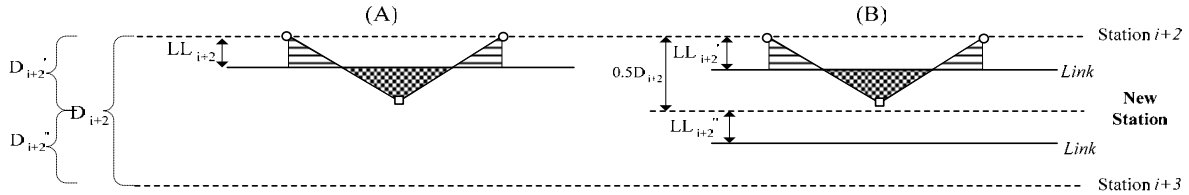


Figure 5.16, (A) Original detector station spacing, (B) half detector spacing for the same TQ, with the incident downstream of the new detector station.

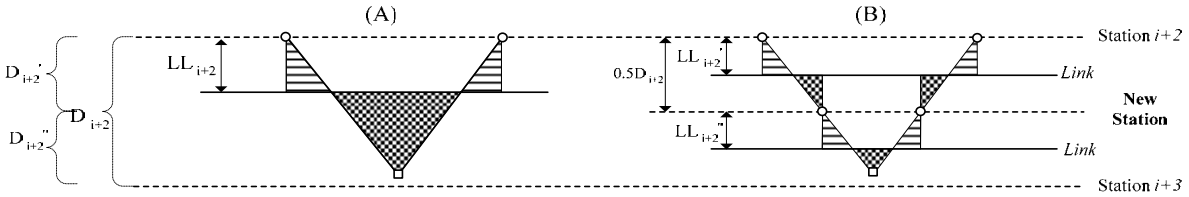


Figure 5.17, (A) Original detector station spacing, (B) half detector spacing for the same TQ, with the incident upstream of the new detector station.

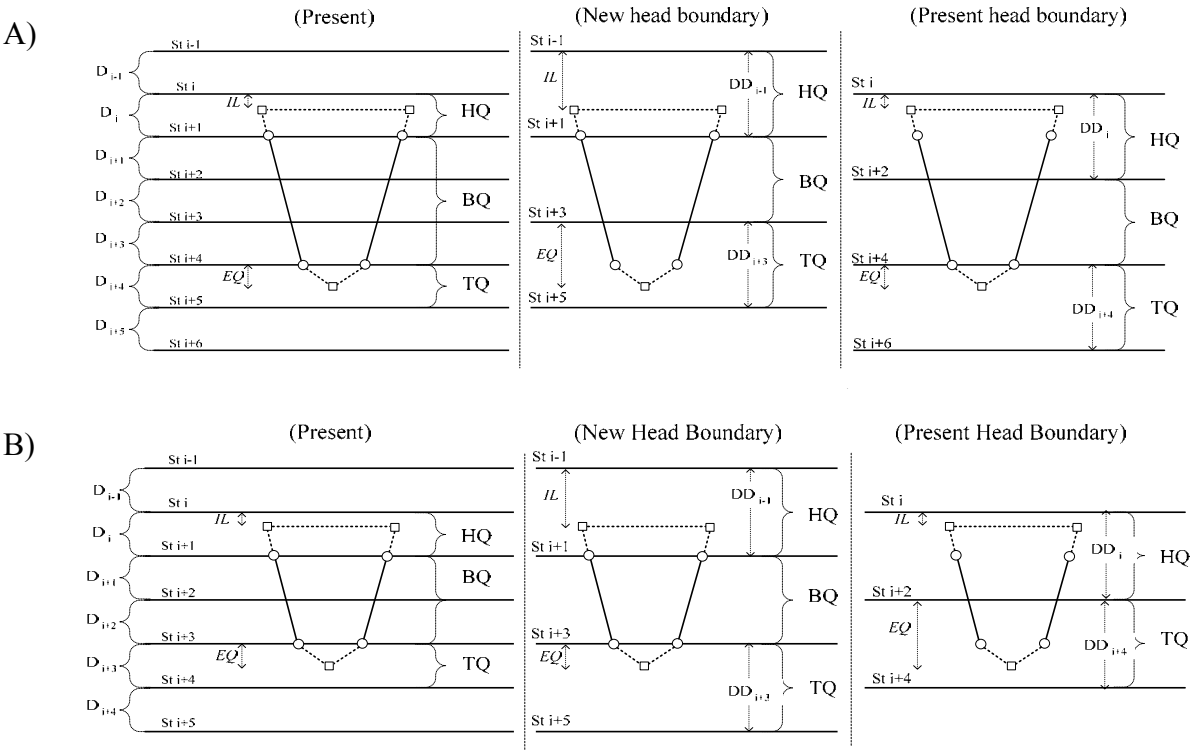


Figure 5.18, Queue boundary in double detector spacing A) Even case, B) Odd case

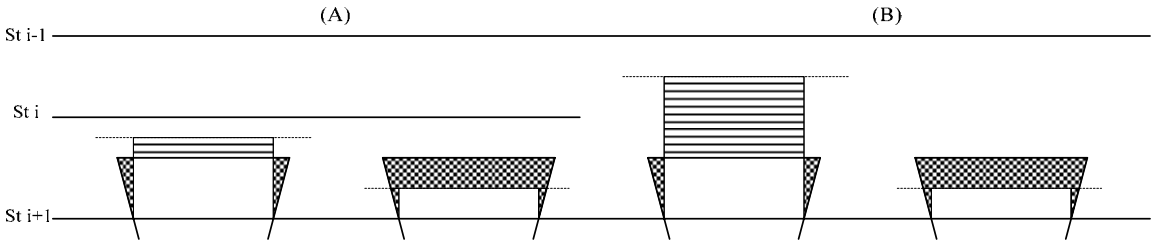


Figure 5.19, Estimation error in head of queue, A) present spacing, B) double spacing with new head boundary

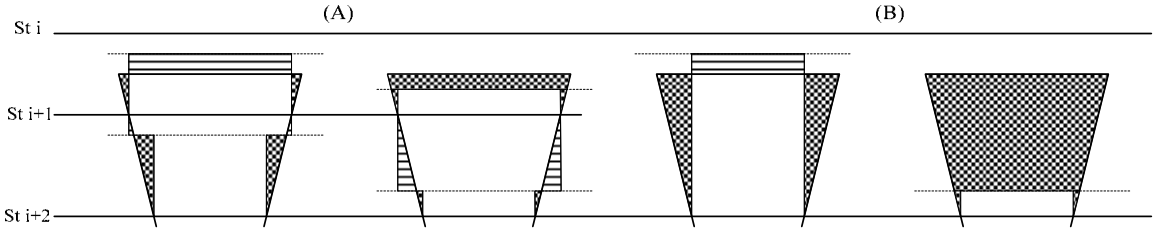


Figure 5.20, Estimation error in head of queue, A) present spacing, B) double spacing with present head boundary

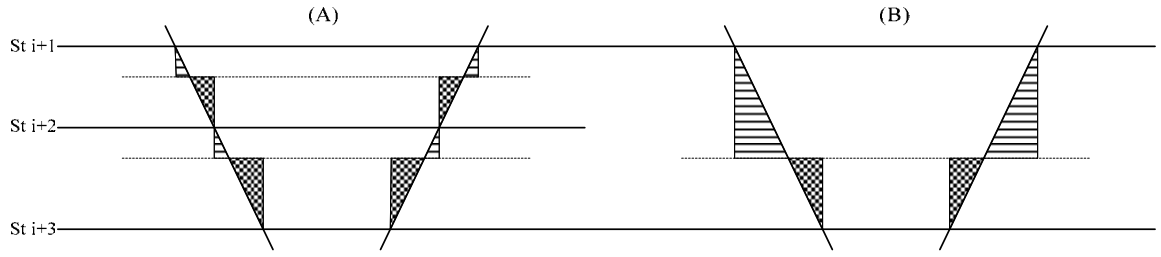


Figure 5.21, Estimation error in body of queue, A) present spacing, B) double spacing

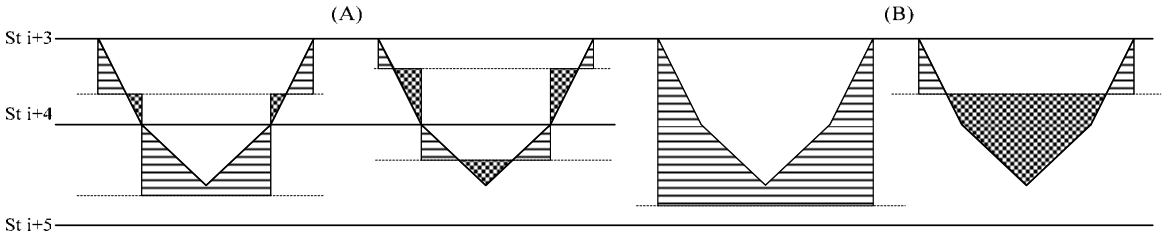


Figure 5.22, Estimation error in tail of queue, A) present spacing, B) double spacing with present tail boundary

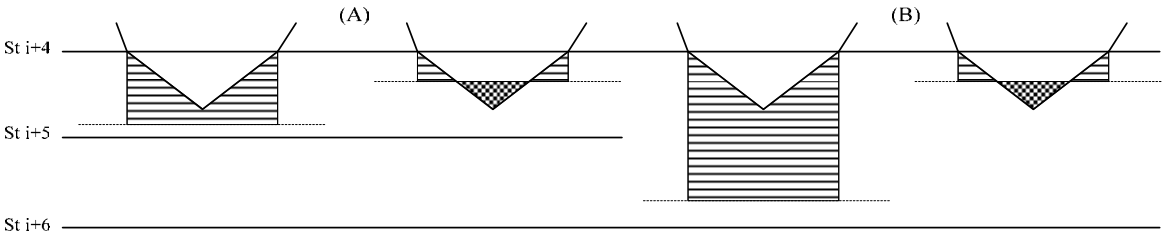


Figure 5.23, Estimation error in tail of queue, A) present spacing, B) double spacing with new tail boundary

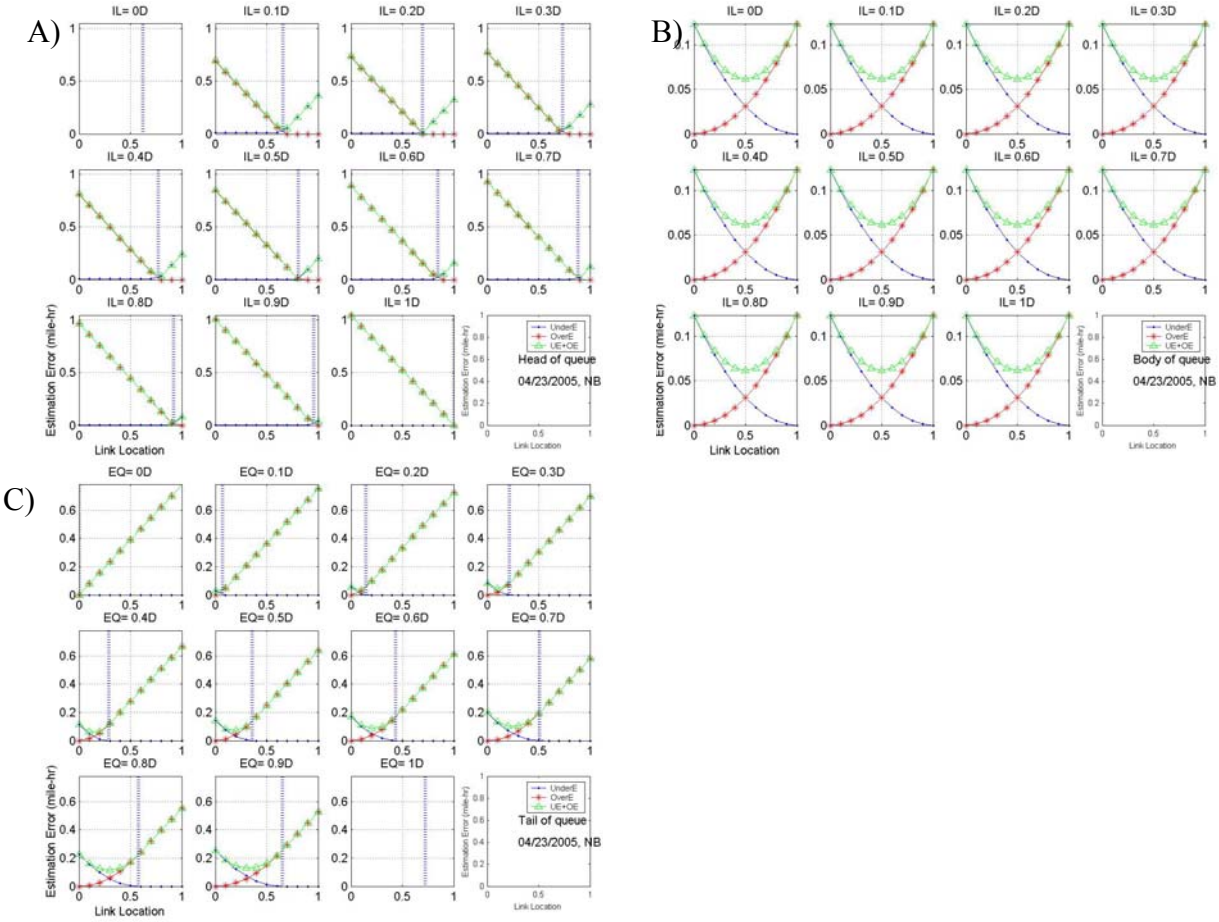


Figure 5.24, The sample queue covers an odd number of stations. Estimation error in each of three segments by link locations, given location of incident and end of queue for NHB A) Head of queue, B) Body of queue, C) Tail of queue

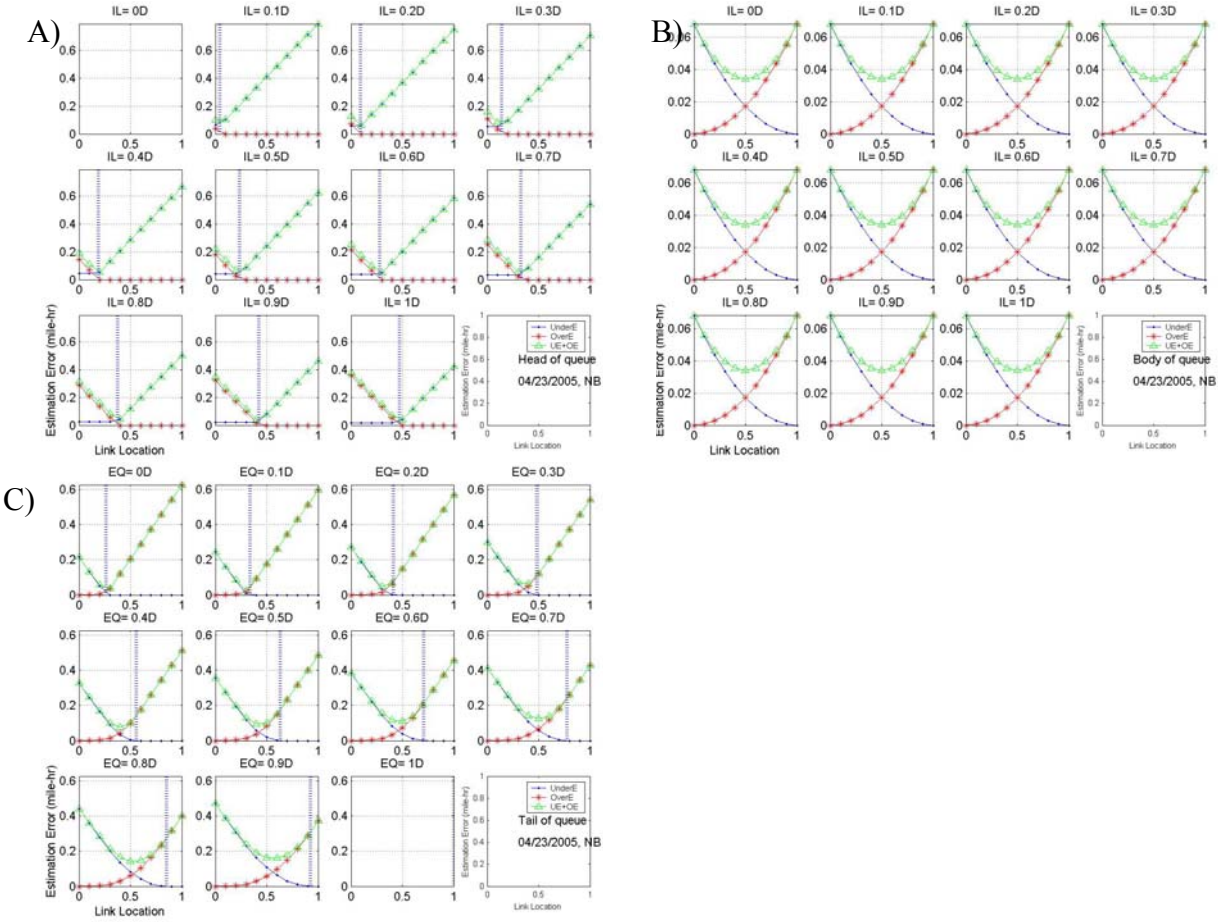


Figure 5.25, The sample queue covers an odd number of stations. Estimation error in each of three segments by link locations, given location of incident and end of queue for PHB A) Head of queue, B) Body of queue, C) Tail of queue

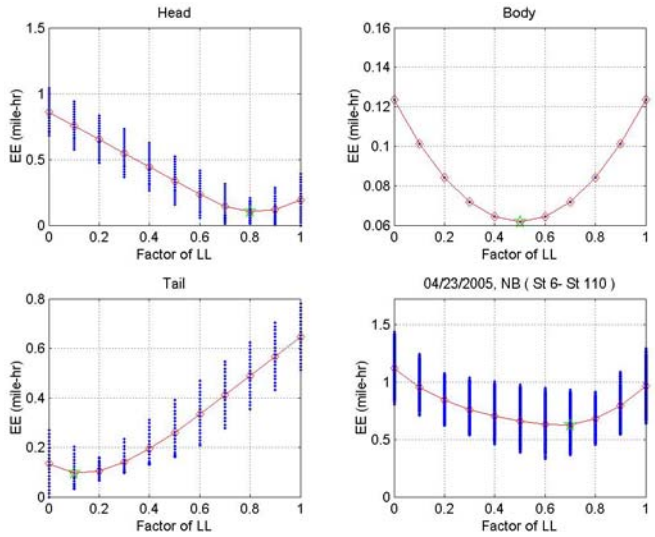


Figure 5.26, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours) for double detector station spacing, NHB and NTB.

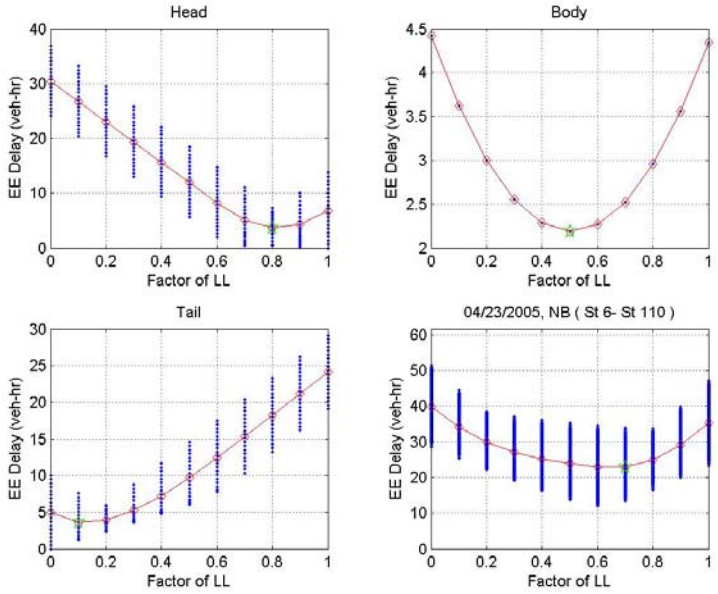


Figure 5.27, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in vehicle-hours per lane) for double detector station spacing, NHB and NTB.

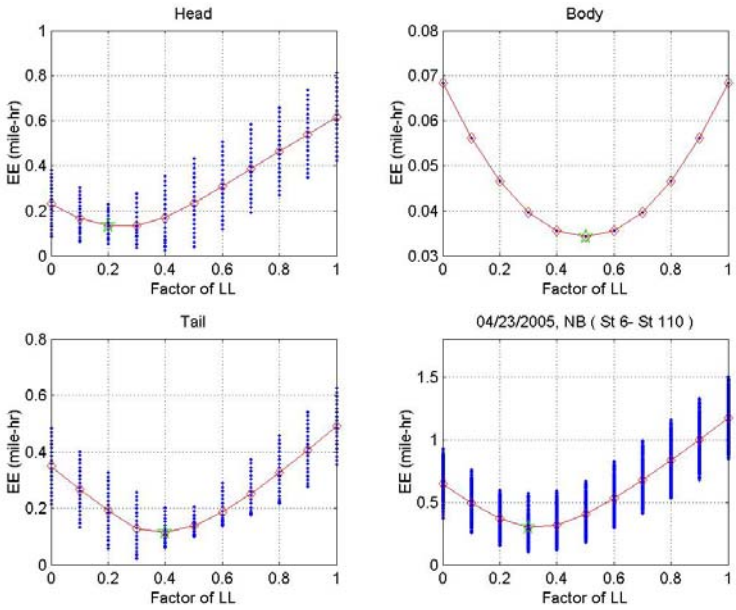


Figure 5.28, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours) for double detector station spacing, PHB and PTB.

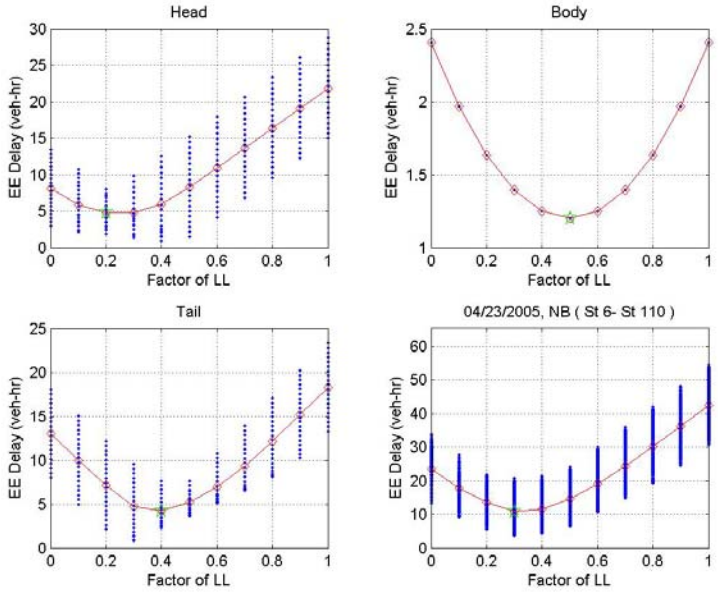


Figure 5.29, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in vehicle-hours per lane) for double detector station spacing, PHB and PTB.

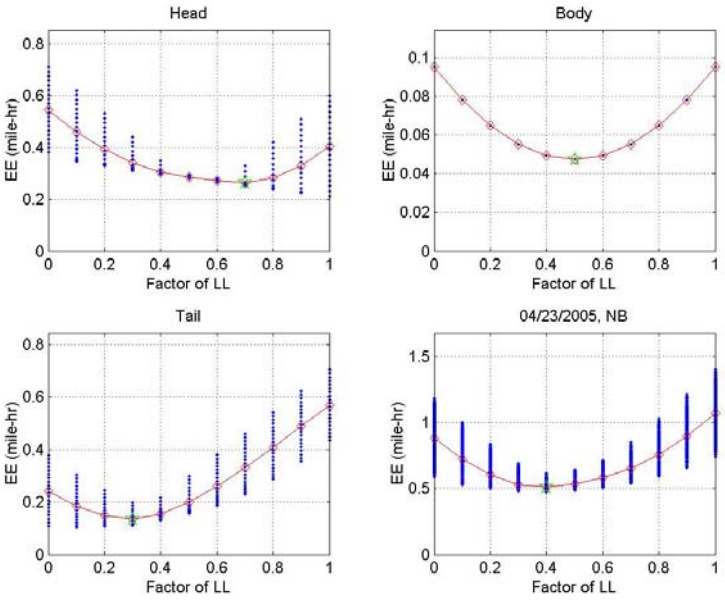


Figure 5.30, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours) for double detector station spacing, PHB and NHB, PTB and NTB.

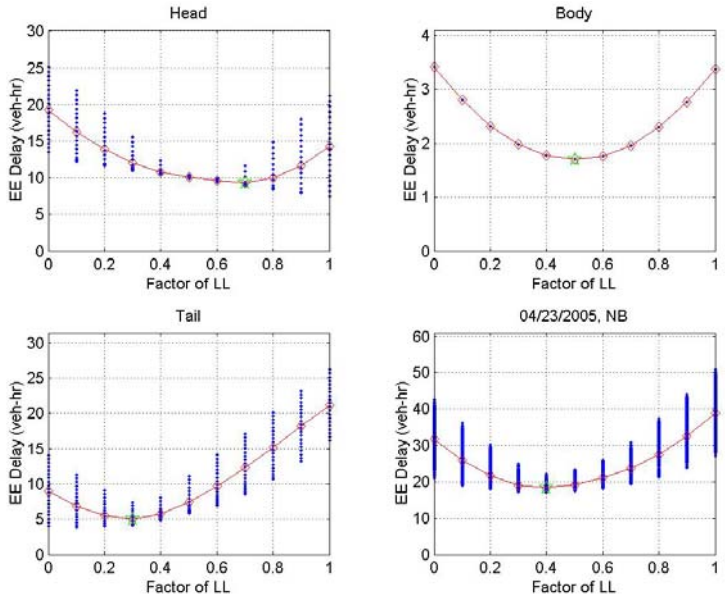


Figure 5.31, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in vehicle-hours per lane) for double detector station spacing, PHB and NHB, PTB and NTB.

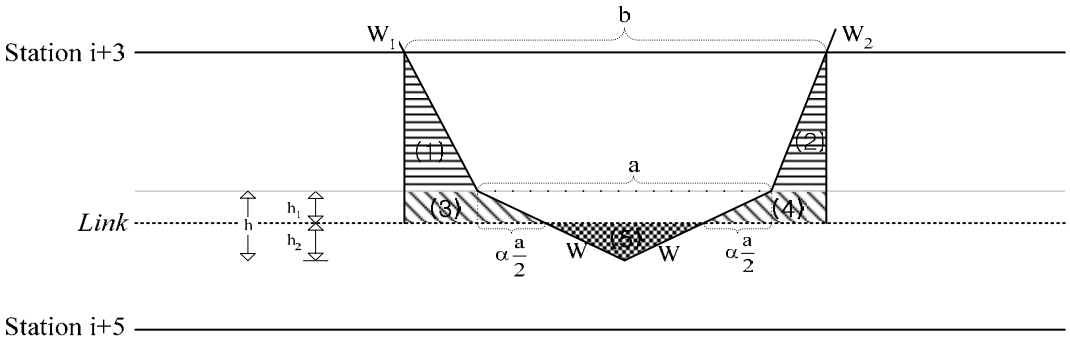


Figure 5.32, Estimation errors in present tail boundary when a link is placed on the last link segment

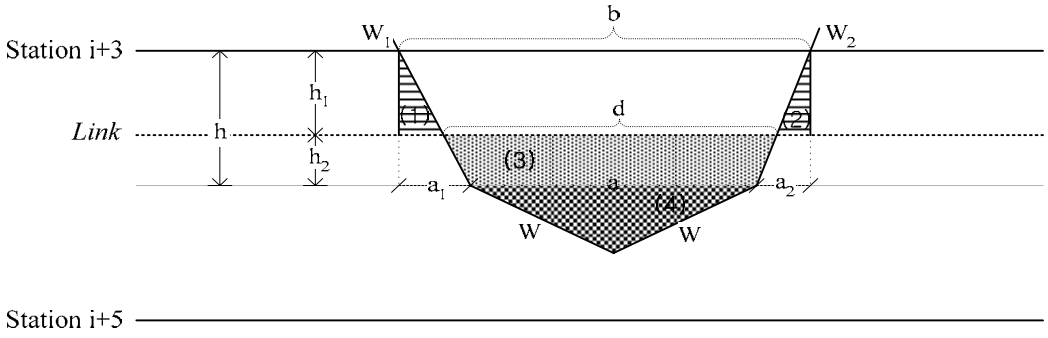


Figure 5.33, Estimation errors in present tail boundary when a link is placed on the second last link segment

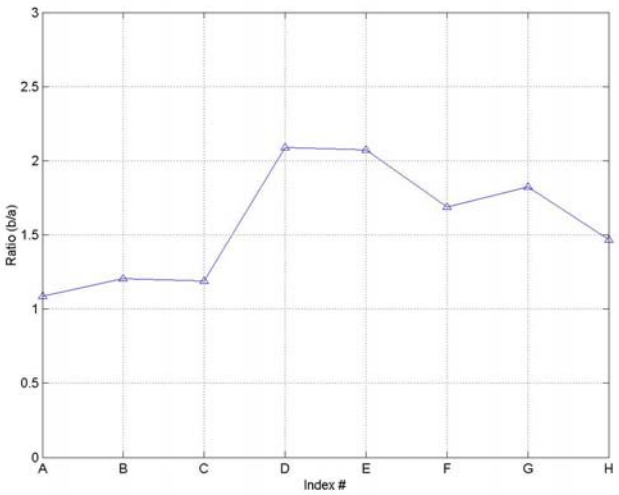


Figure 5.34, The ratio of queue duration time in last two loop detector stations in case of present tail boundary

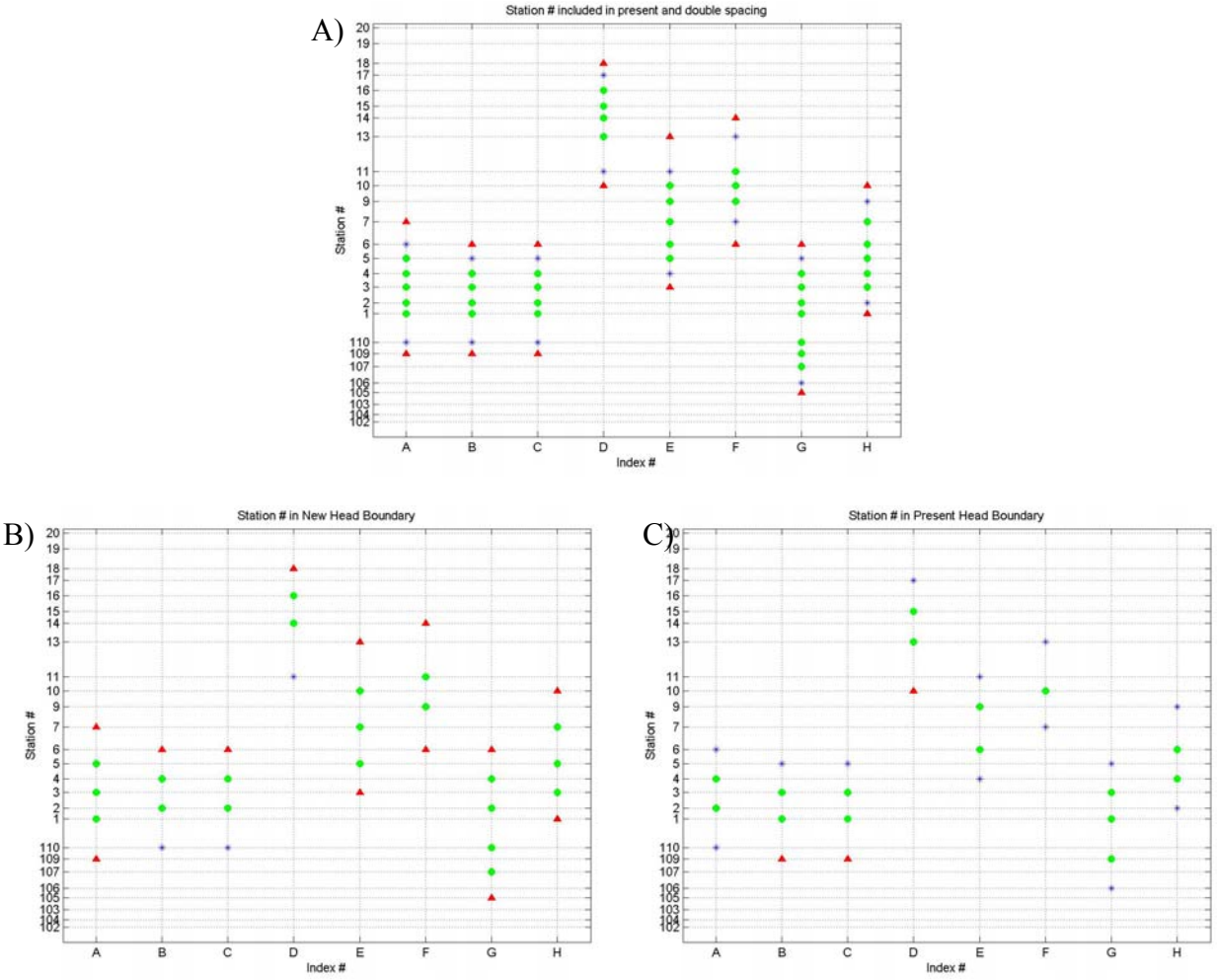


Figure 5.35, Station location in our analysis A) original with all stations, B) double spacing under NHB, C) double spacing under PHB.

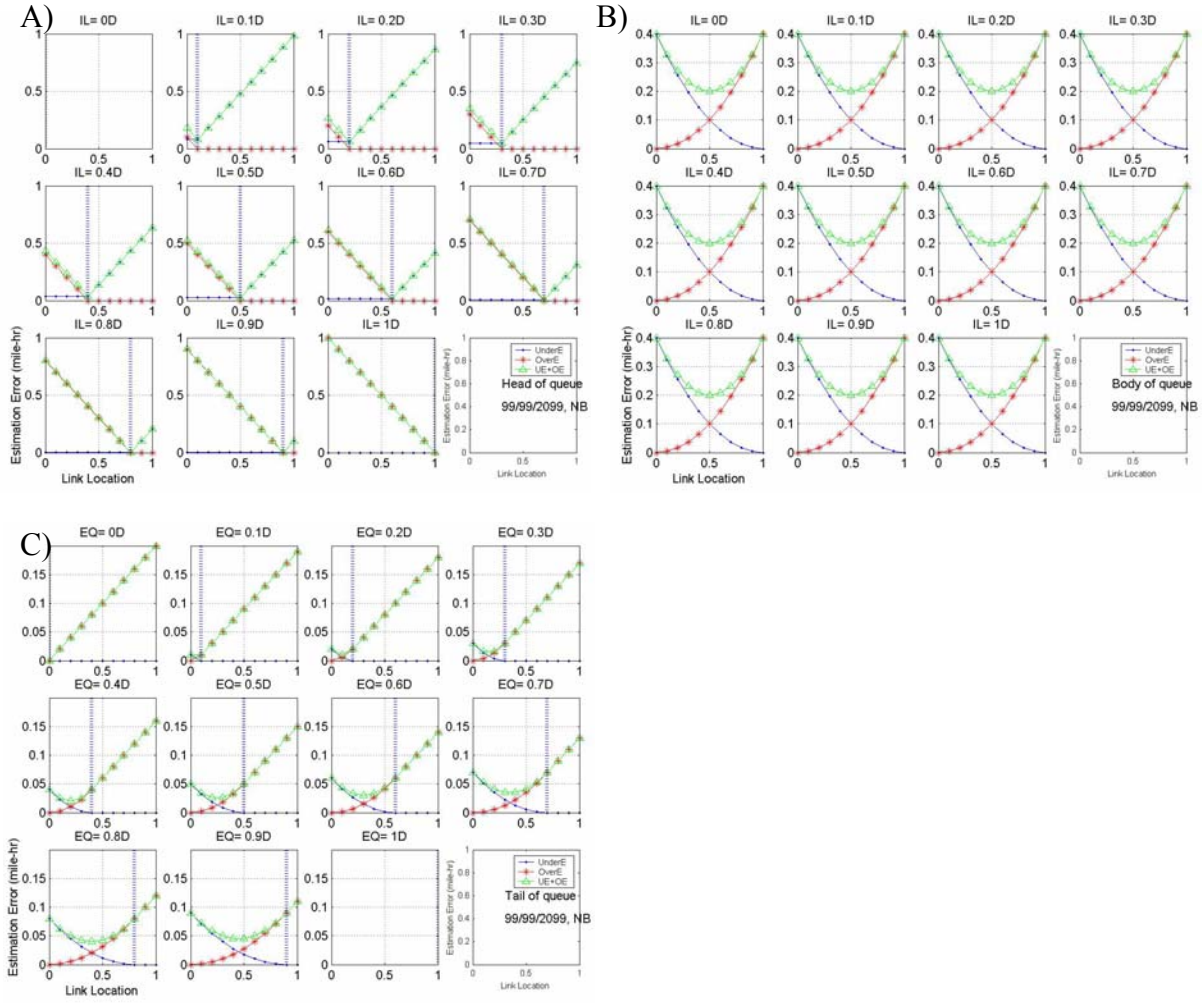


Figure 5.36, Estimation error in each of three segments by link locations, given location of incident and end of queue A) Head of queue, B) Body of queue, C) Tail of queue

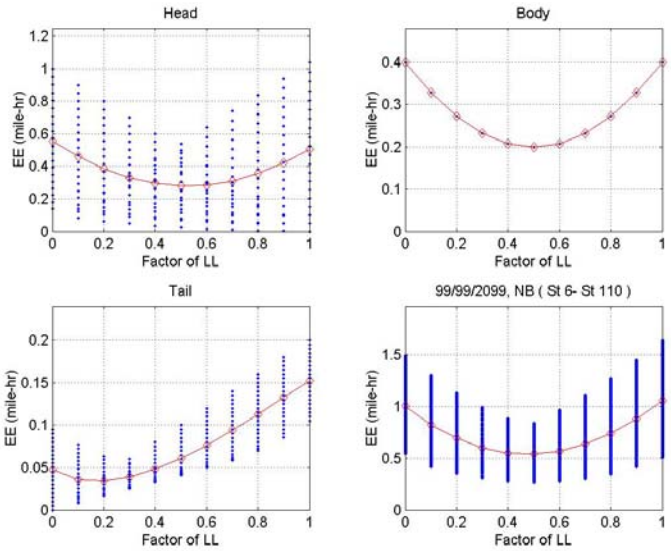
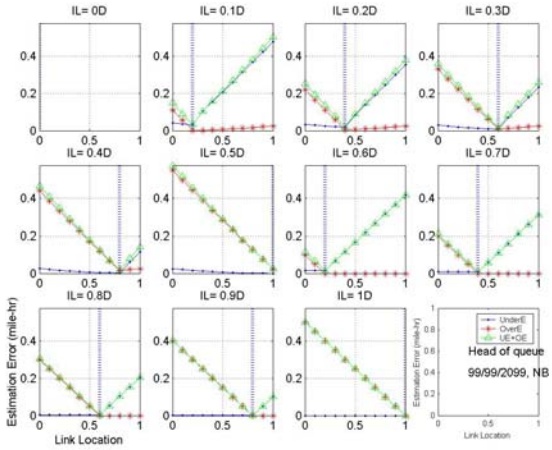
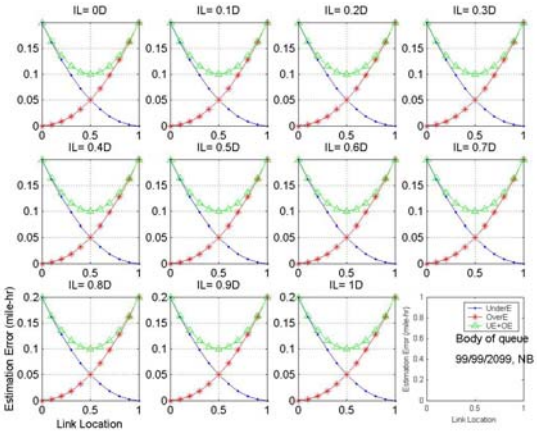


Figure 5.37, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours).

A)



B)



C)

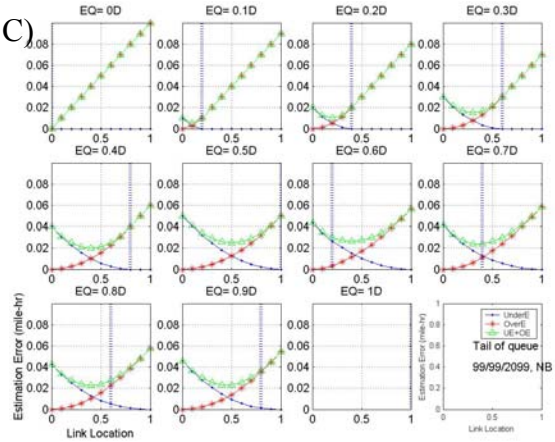


Figure 5.38, Estimation error in each of three segments by link locations for half detector station spacing, given location of incident and end of queue A) Head of queue, B) Body of queue, C) Tail of queue

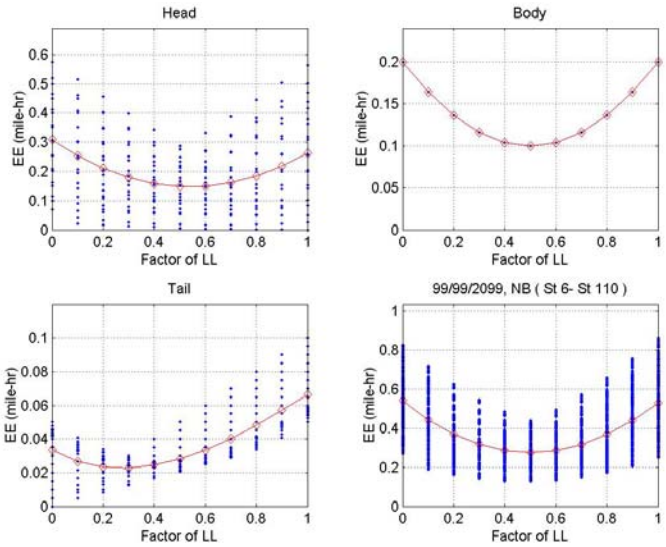


Figure 5.39, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours) for half detector station spacing.

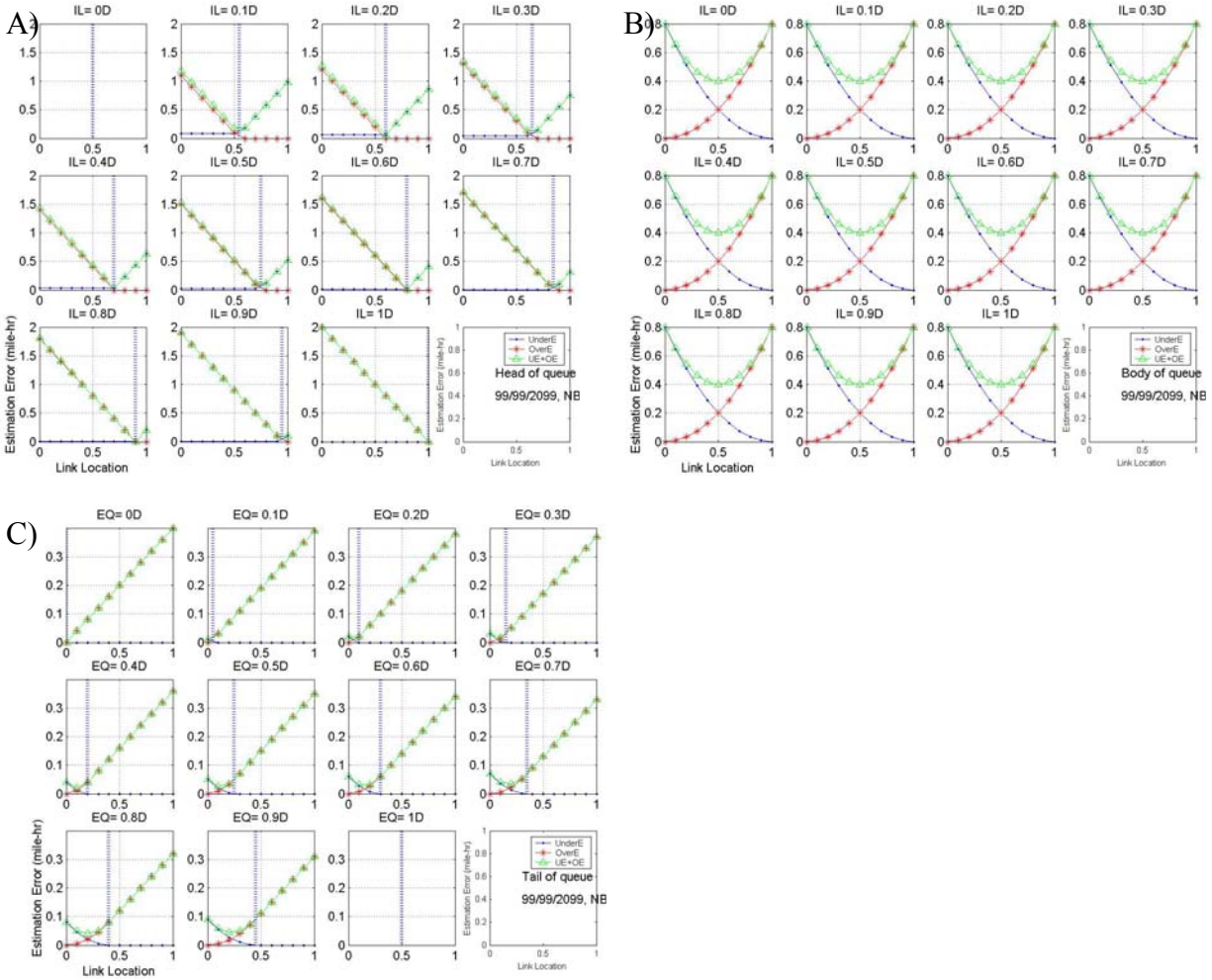


Figure 5.40, The sample queue covers an odd number of stations. Estimation error in each of three segments by link locations, given location of incident and end of queue for NHB A) Head of queue, B) Body of queue, C) Tail of queue

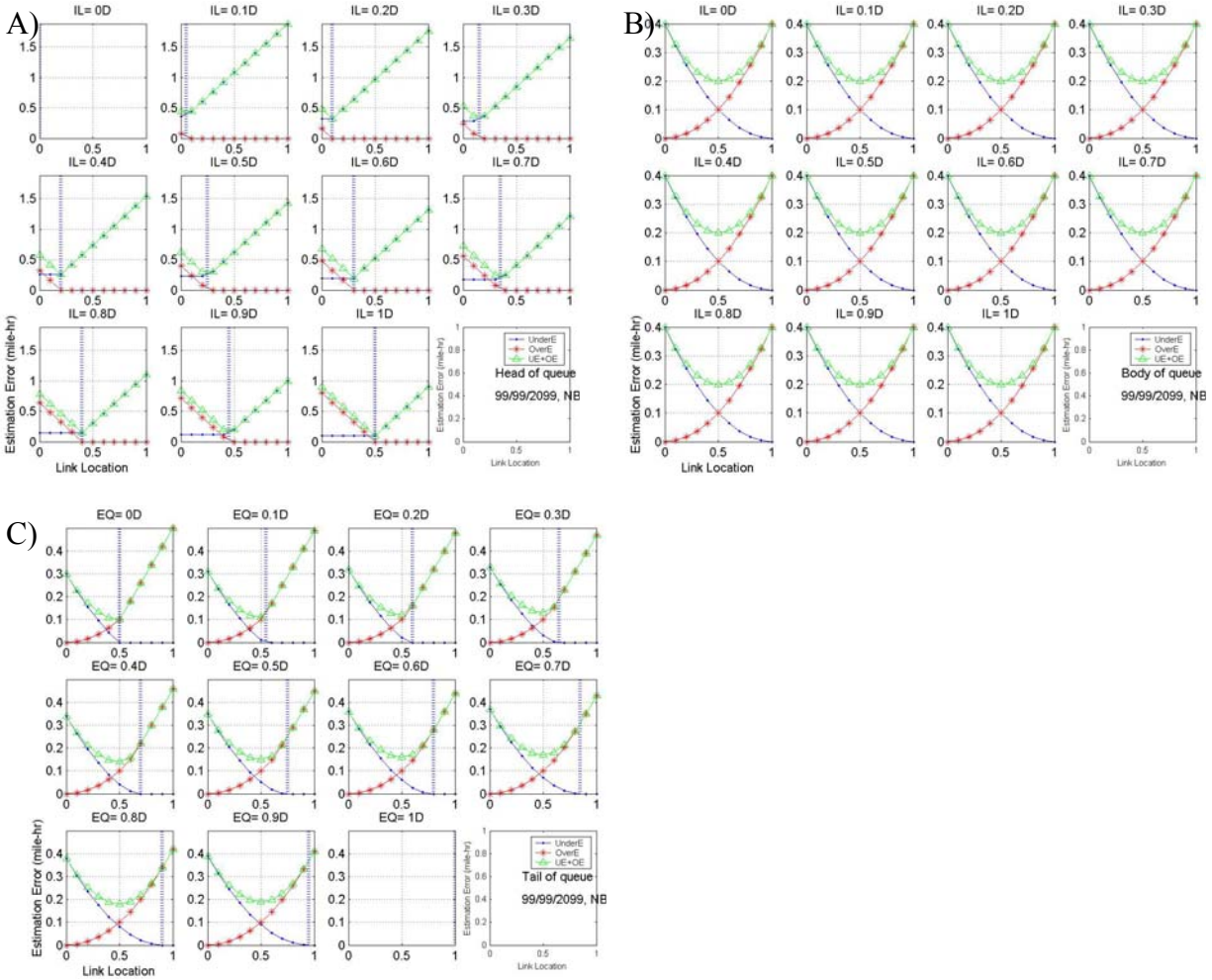


Figure 5.41, The sample queue covers an odd number of stations. Estimation error in each of three segments by link locations, given location of incident and end of queue for PHB A) Head of queue, B) Body of queue, C) Tail of queue

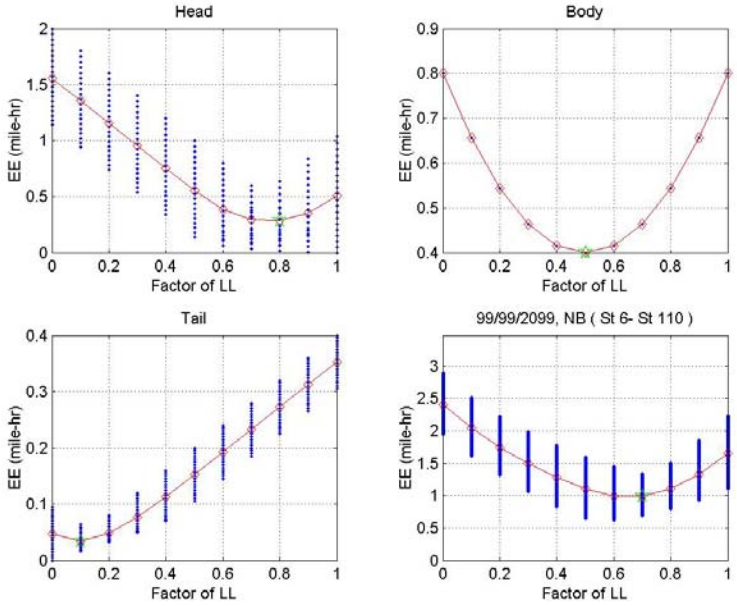


Figure 5.42, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours) for double detector station spacing, NHB and NTB.

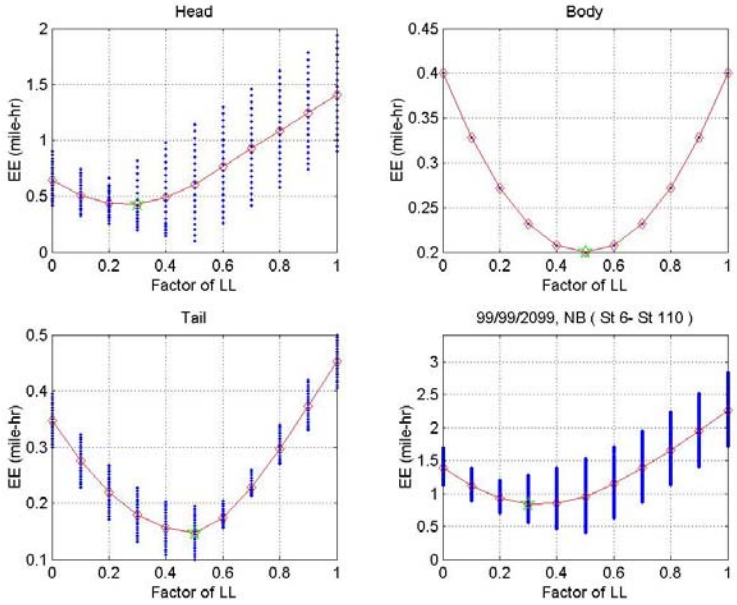


Figure 5.43, Absolute estimation error in the three segment types (HQ, BQ, and TQ), and the sum across all segments (in mile-hours) for double detector station spacing, PHB and PTB.

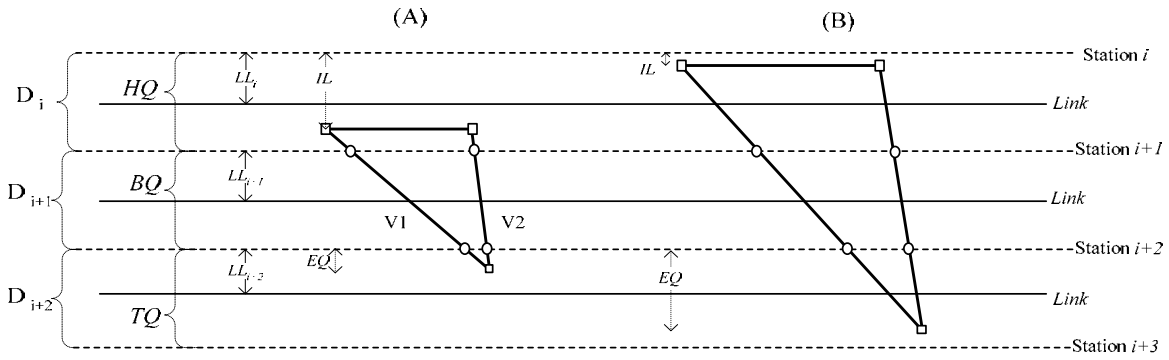


Figure 5.44, Another type of queue where v_2 propagates upstream from the bottleneck

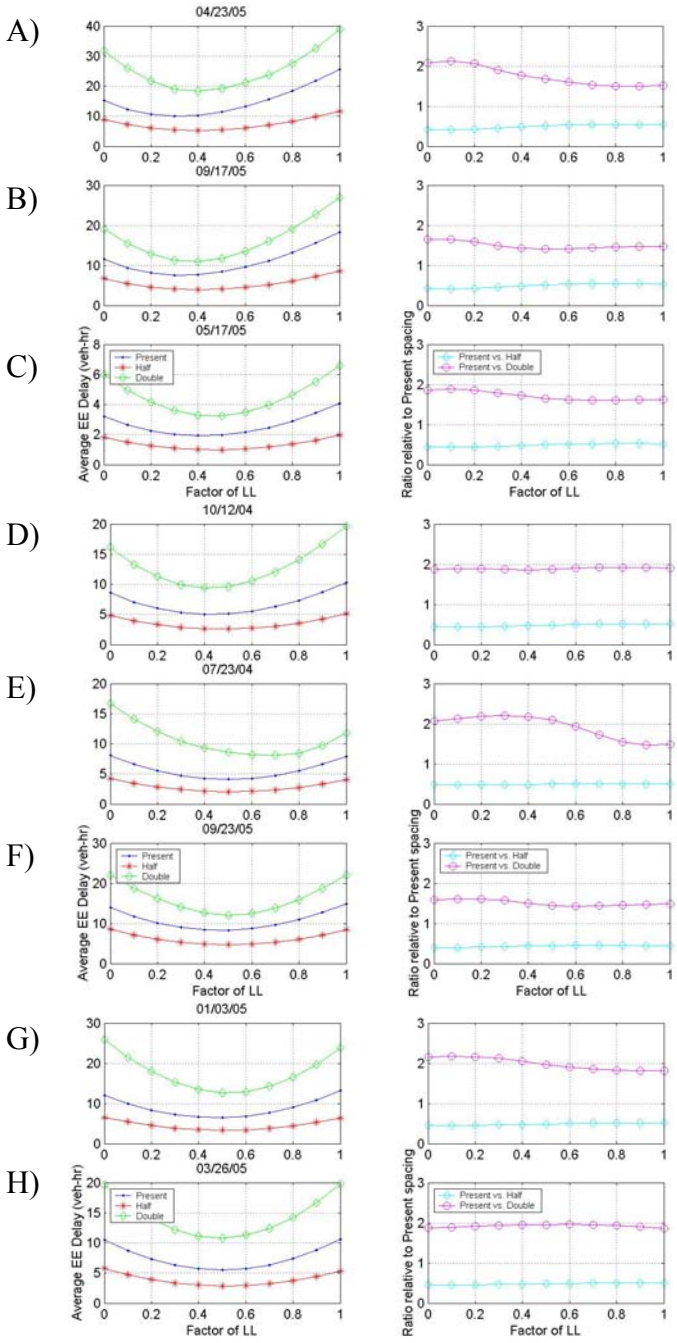


Figure 5.45, Average estimation error delay in three case and ratio relative to present spacing over the eight study incidents, respectively.

A) B)

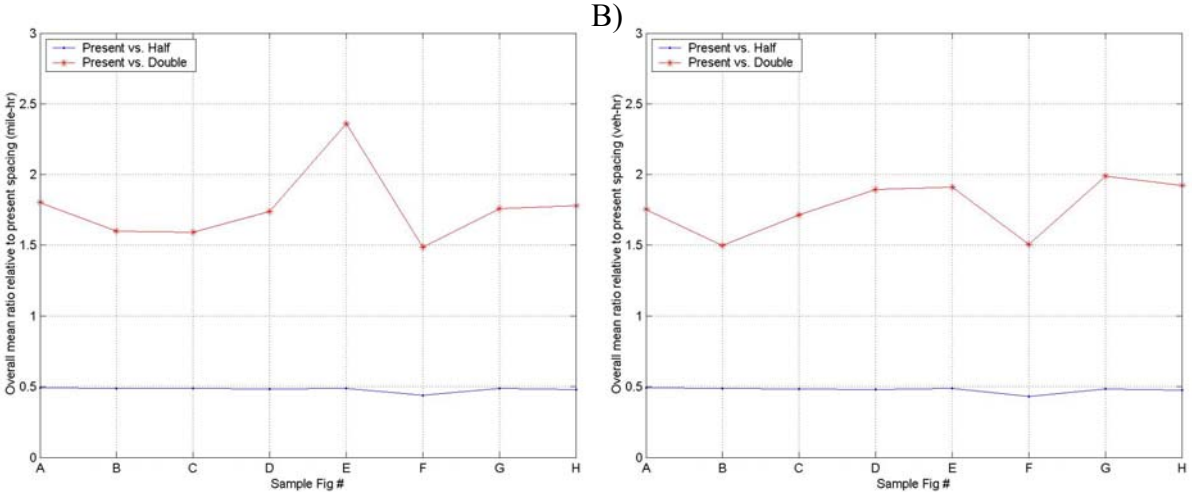


Figure 5.46, Overall mean ratios relative to present spacing, A) mile-hr, B) veh-hr per lane

Appendix A.

Instant Point Speed Estimation and Comparison

A.1 Point Estimation Using Loops

Traffic point speed estimation at dual loop stations has been revisited using sub-second data. The objective is to achieve more accurate estimation in the following senses:

- Less time delay
- Lower estimation error
- Data source: 04/13/05 4:00pm

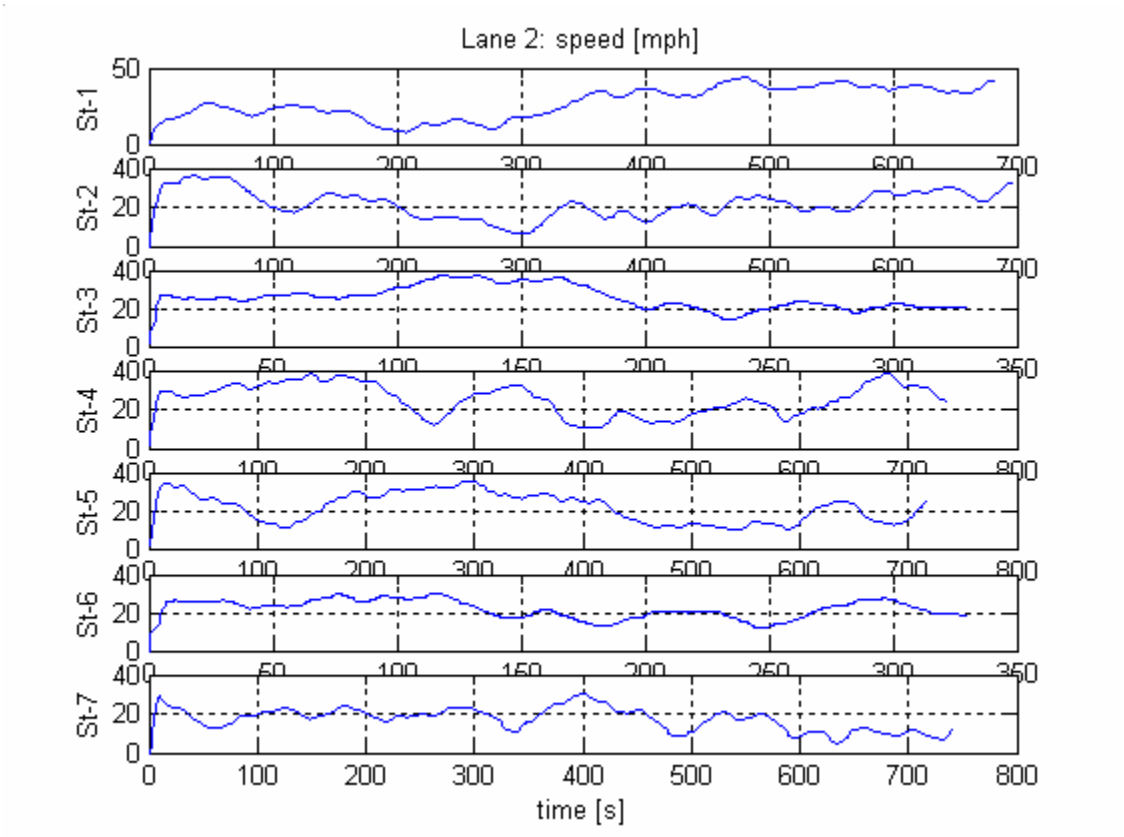


Figure A.1. Sub-second speed estimation at the 7 stations in lane 2

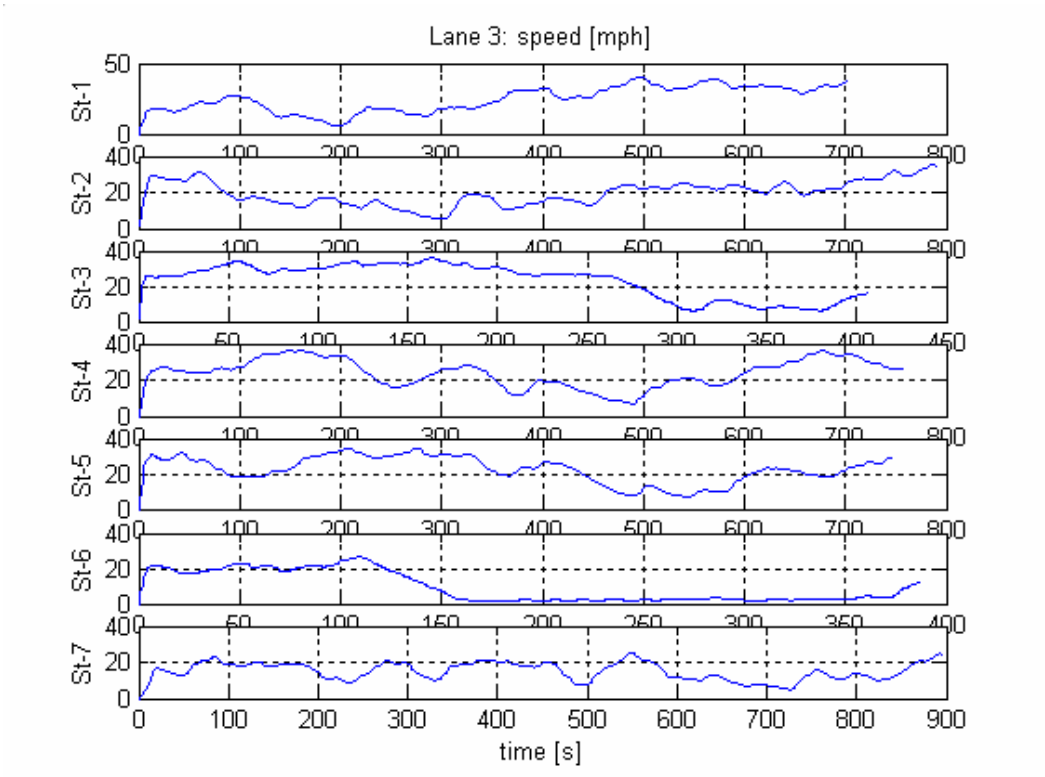


Figure A.2. Sub-second speed estimation at the 7 stations in lane 3

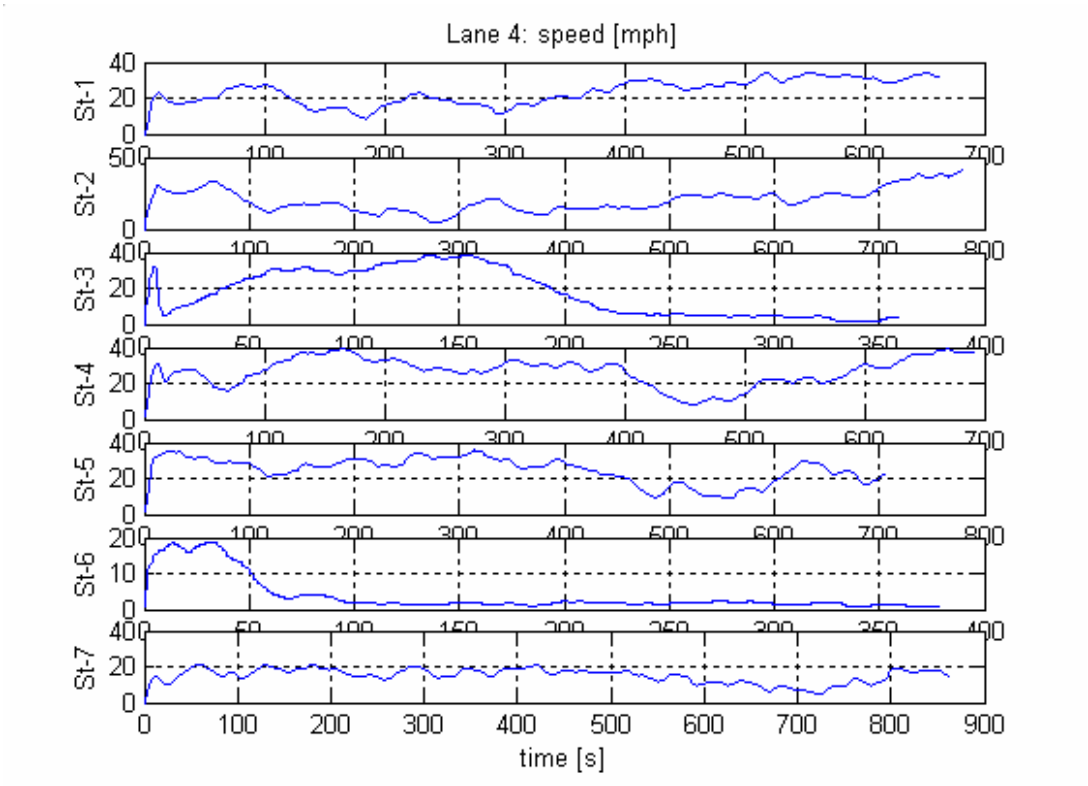


Figure A. 3. Sub-second speed estimation at the 7 stations in lane 4

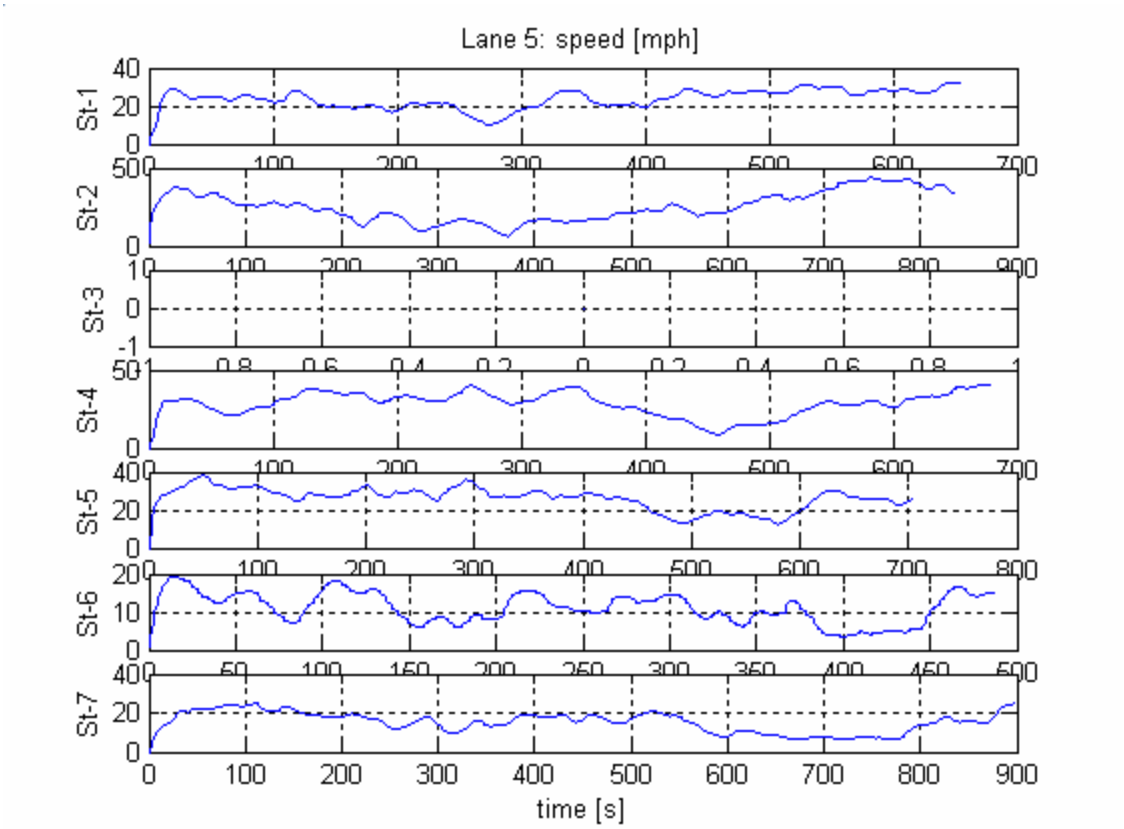


Figure A. 4. Sub-second speed estimation at the 6 stations in lane 5; Station 3 has not data

Problem: Occupancy estimation tend to be higher which may be due to data extracting from the raw loop data which needs improving in next step.

A.2. Comparison of loop data with NGSIM Data

04/13/05, 16:00-16:15 NGSIM East Bound vehicle trajectory data are used for comparison. The maximum tracking distance is about 518m in NGSIM trajectory which has not reached the Station 7 yet. The speed at the point 500m has been extracted and plotted with speed estimated based on the BHL Loop Station 7 data. The location of loop station 7 is located about 50m downstream of the traffic flow. The comparison is as follows:

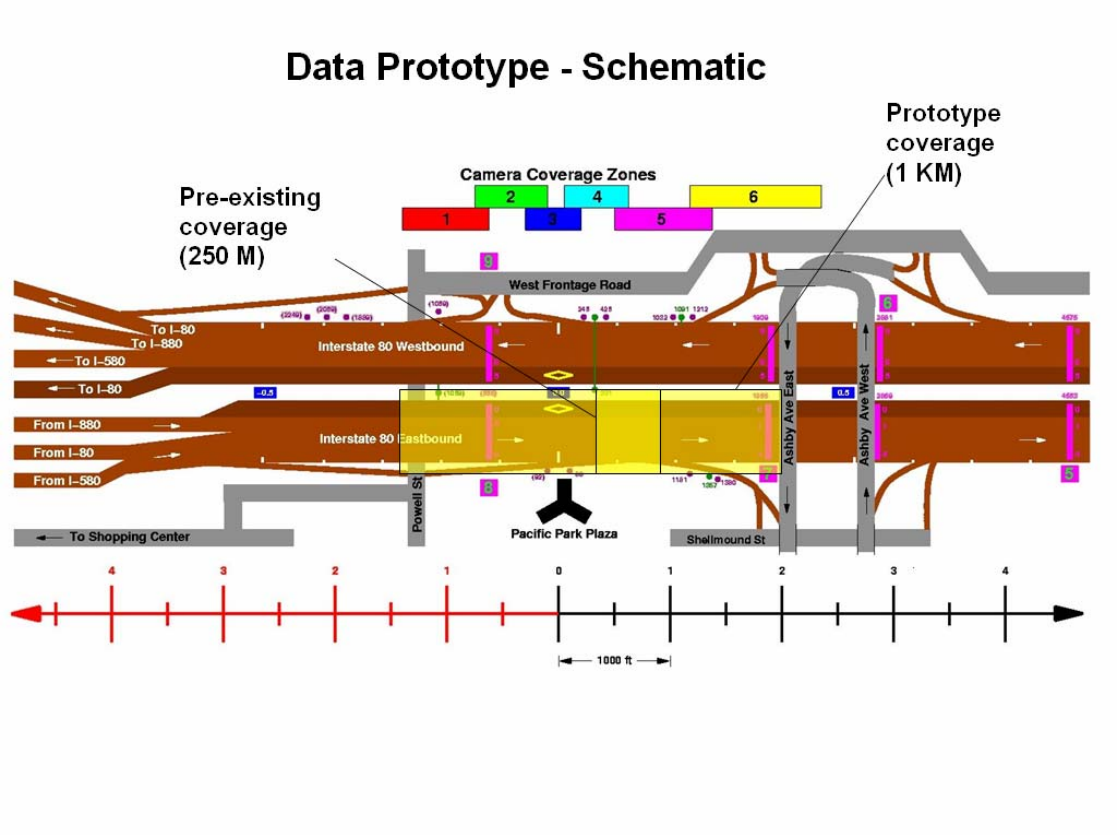


Figure A. 5. Video Camera and loop station locations of BHL stretch on I-80

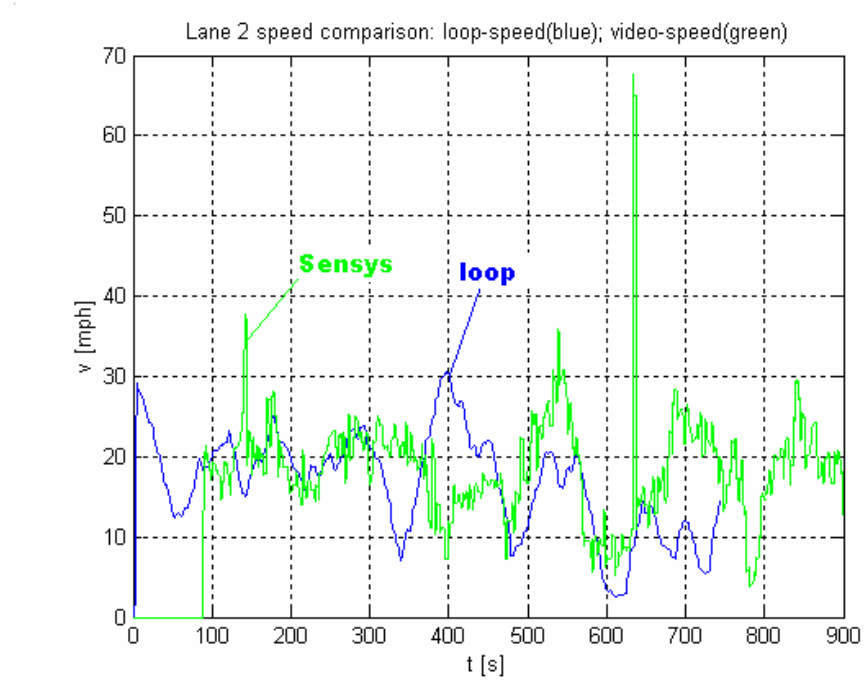


Figure A. 6. Comparison loop speed and video camera speed at station 7 in Lane 2

A. 3. Comparison with Sensys Data (12/15/05 Data used)

Data update rate: 1[s]; Average speed is used for comparison

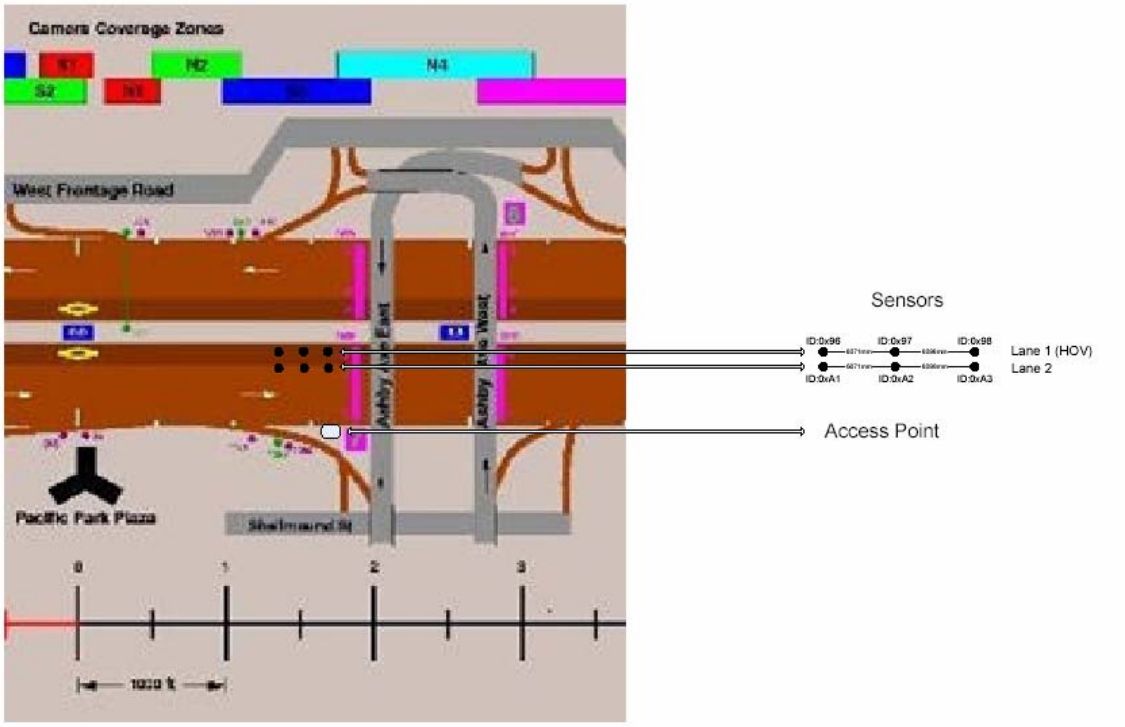


Figure A. 7. Sensys Location: About 6.00m upstream of Station 7 in Lane 1, 2

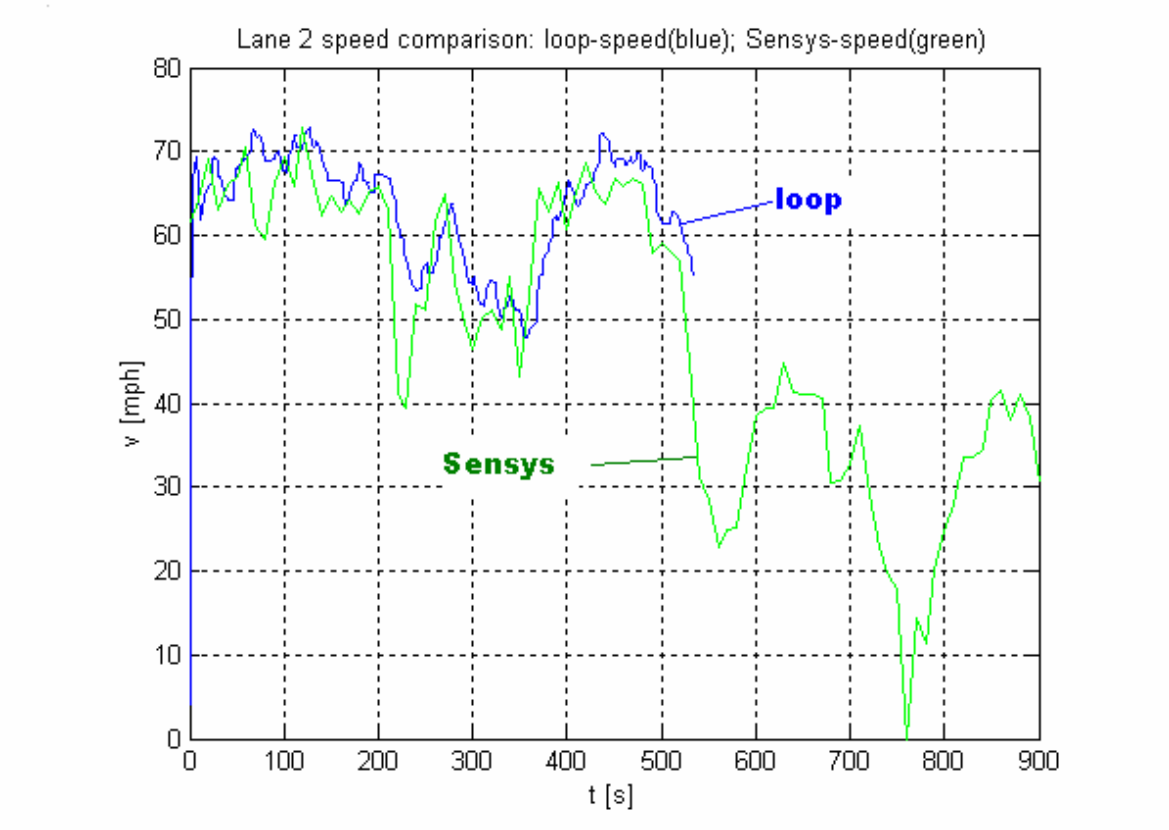


Figure A. 8. 8:00am Lane 2

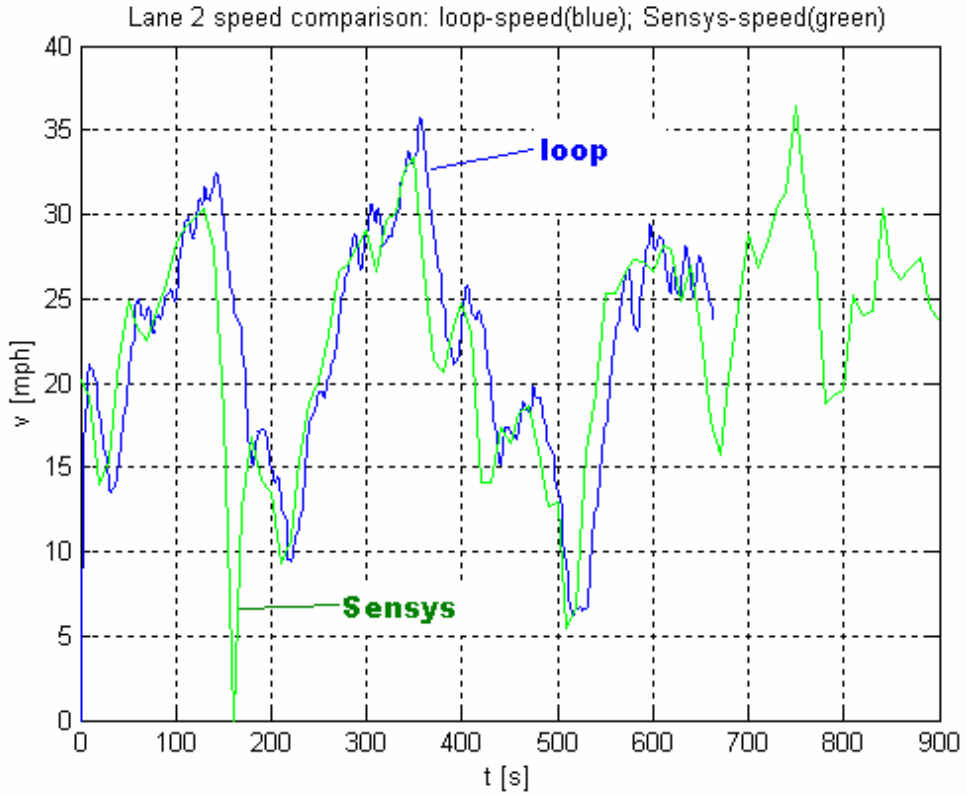


Figure A. 9. 4:00pm Lane 2

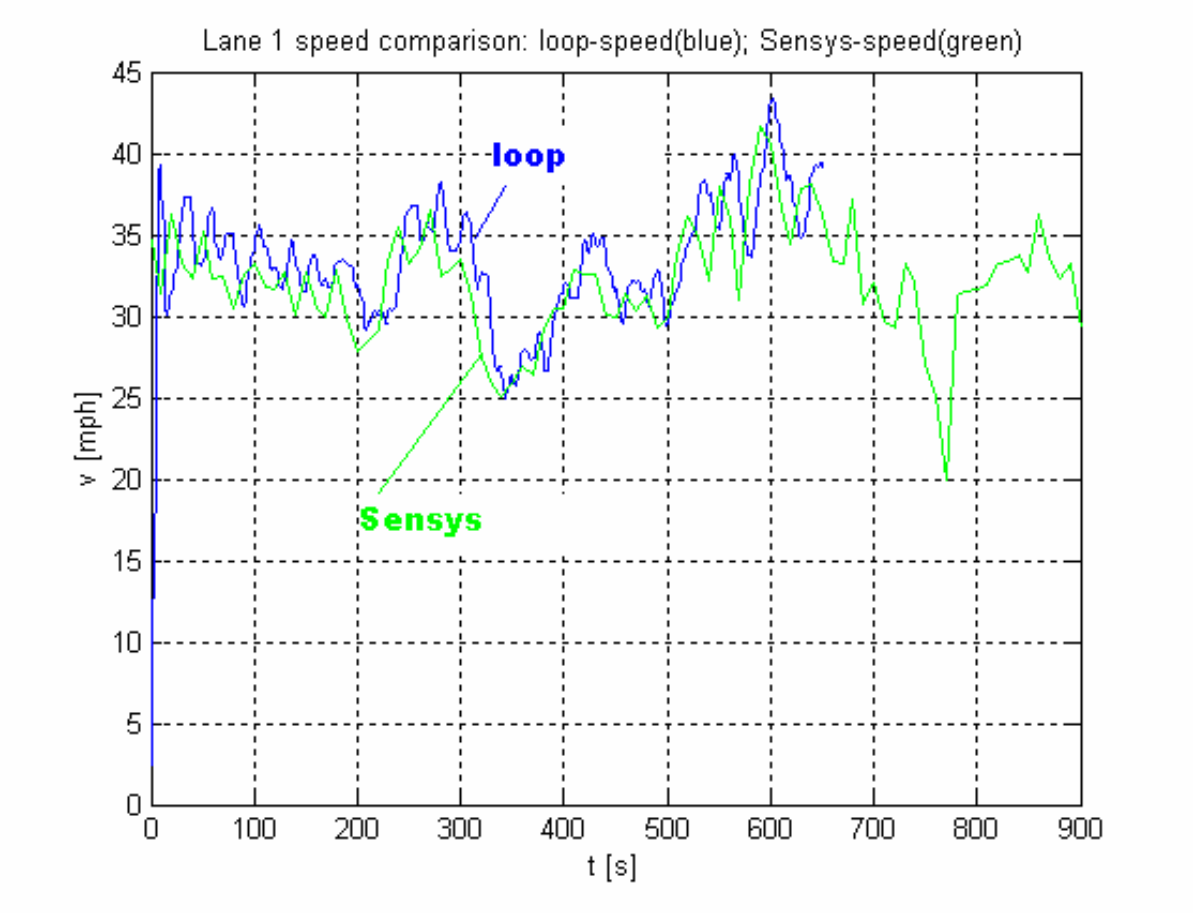


Figure A. 10. 5:00pm Lane 1 (HOV)

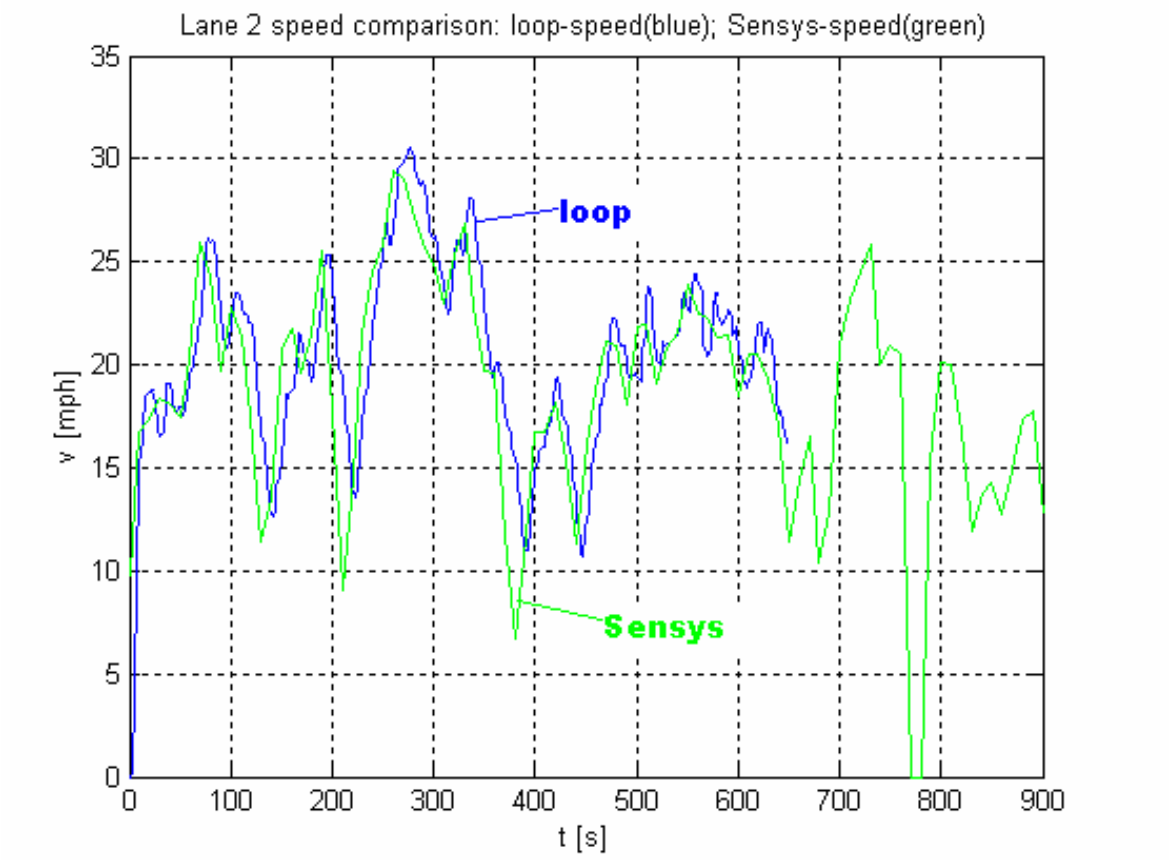


Figure A. 11. 5:00pm Lane 2

Appendix B

A Complete List of Vehicle Trajectory Plot of NGSIM Data

The following are plot of speed-time and distance-time trajectories of all the NGSIM data. Time frame: Inter-State I-80 data are obtained for April 13, 2005 PM peak hours from 4:00pm; US 101 data are obtained for June 15, 2005 AM peak hours from 7:50.
Prehistoric observation shows that all the significant shockwaves have similar tangent: they propagate at similar speed.

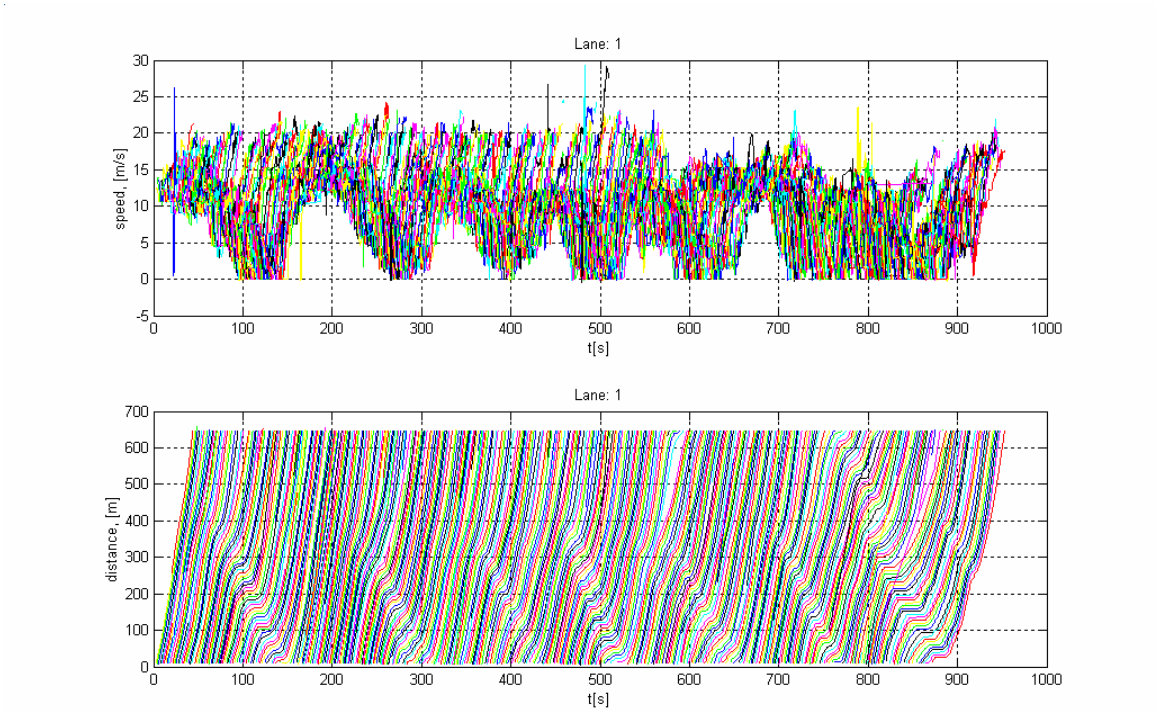


Figure B. 1. US-101, 07:50-08:05am; Lane 1

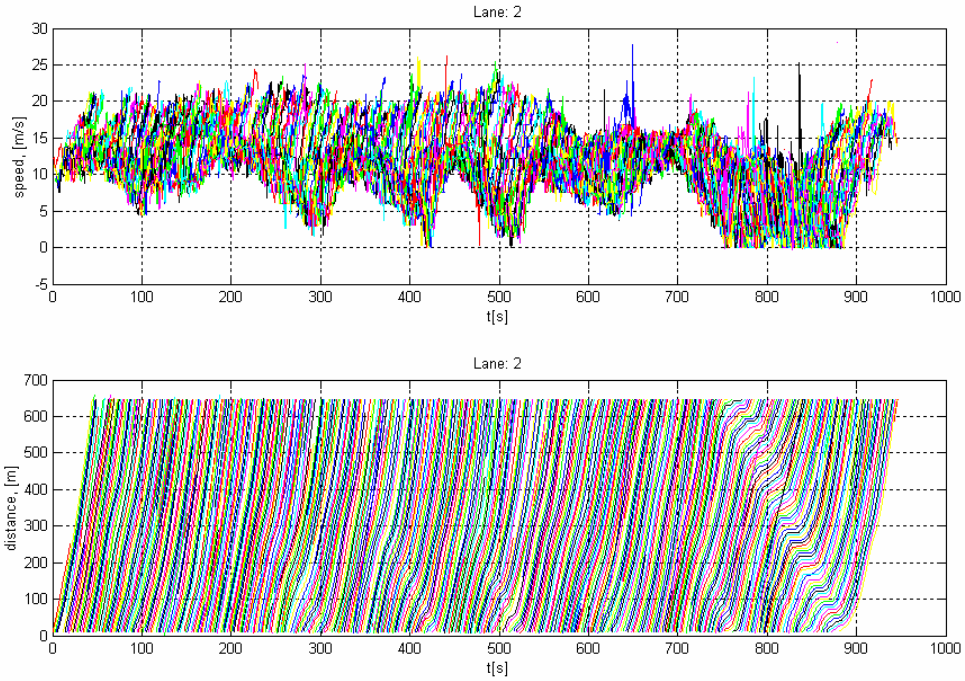


Figure B. 2. US-101, 07:50-08:05am; Lane 2

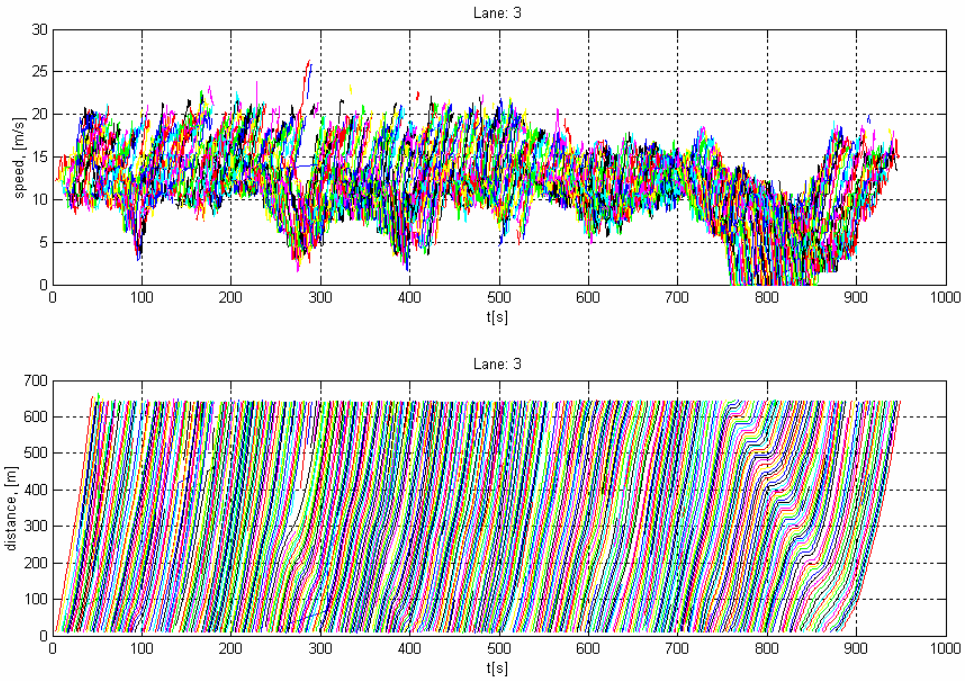


Figure B. 3. US-101, 07:50-08:05am; Lane 3

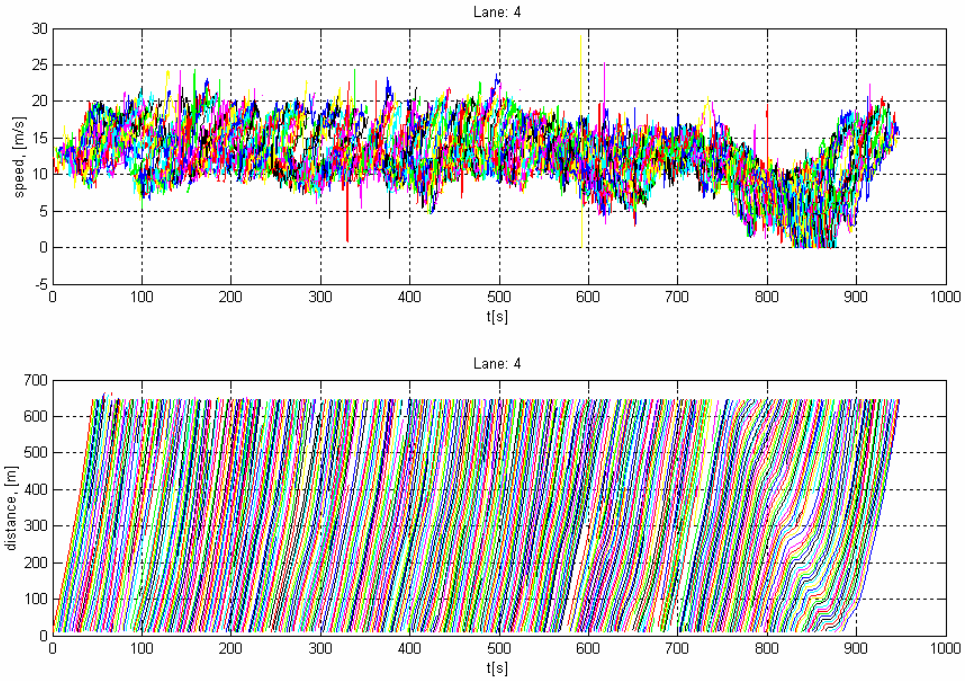


Figure B. 4. US-101, 07:50-08:05am, Lane 4

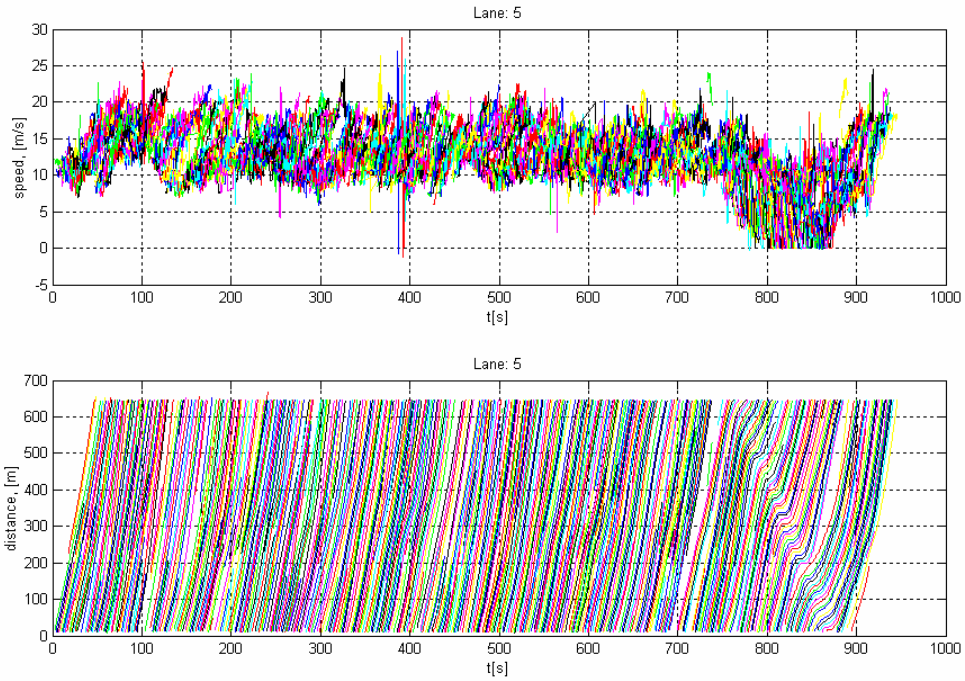


Figure B. 5. US-101, 07:50-08:05am, Lane 5

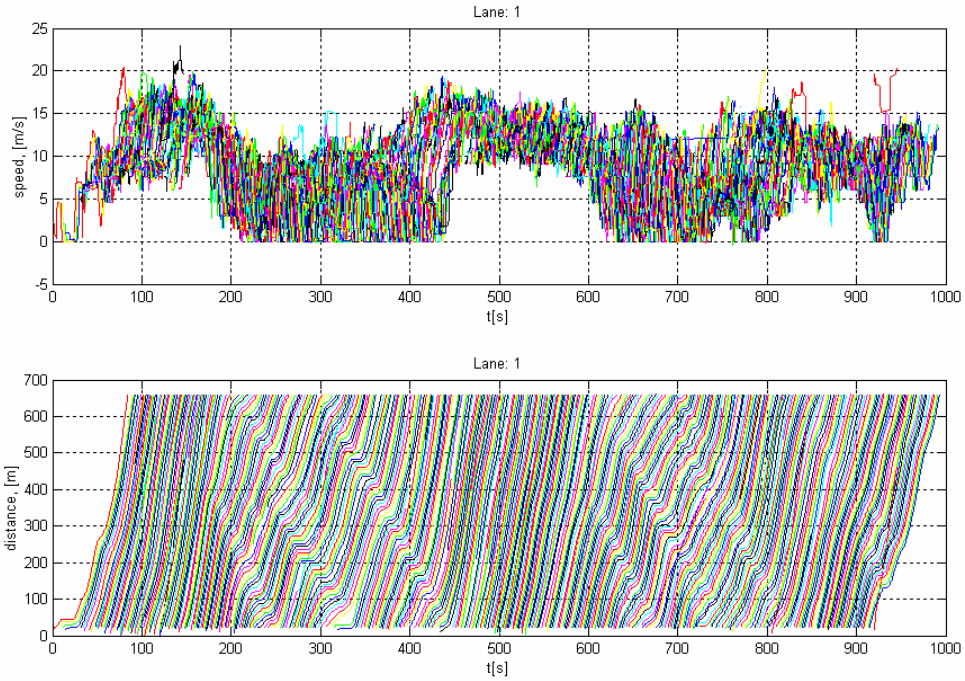


Figure B. 6. US-101, 08:05-08:20am, Lane 1

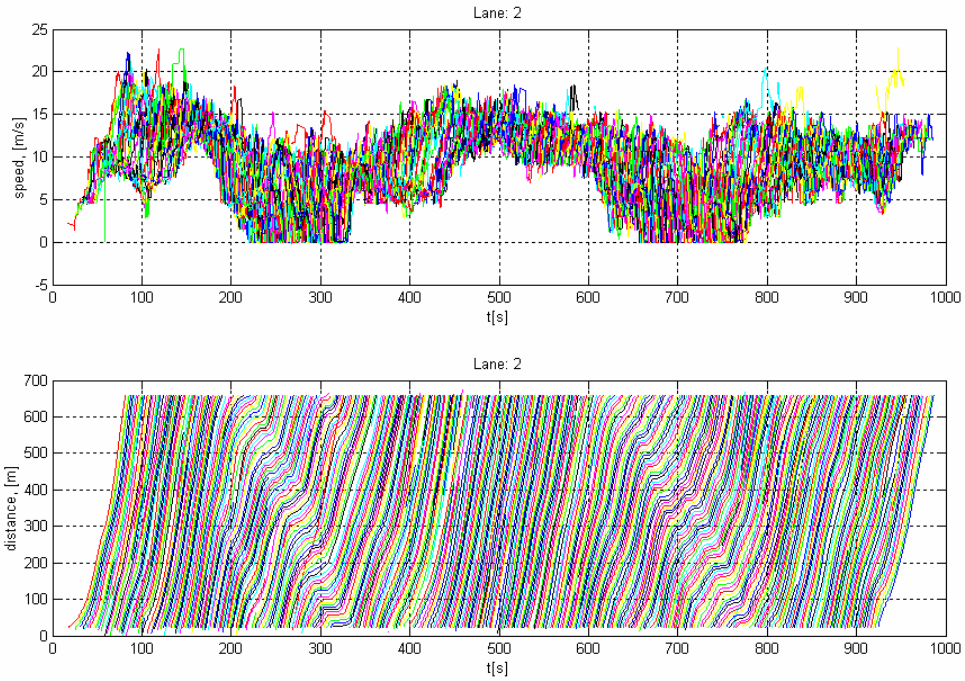


Figure B. 7. US-101, 08:05-08:20am, Lane 2

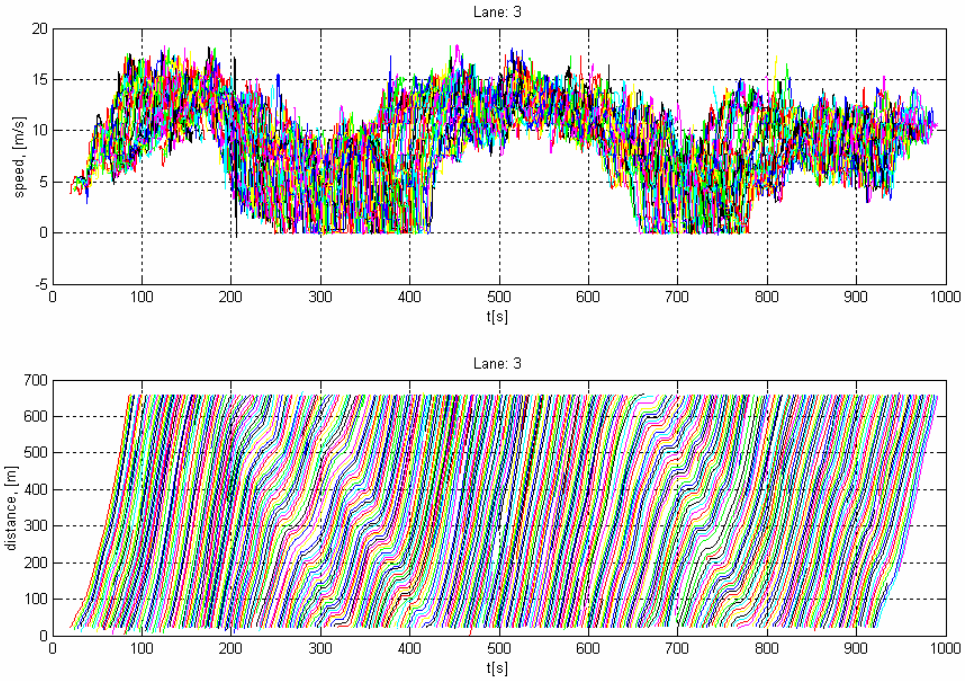


Figure B. 8. US-101, 08:05-08:20am, Lane 3

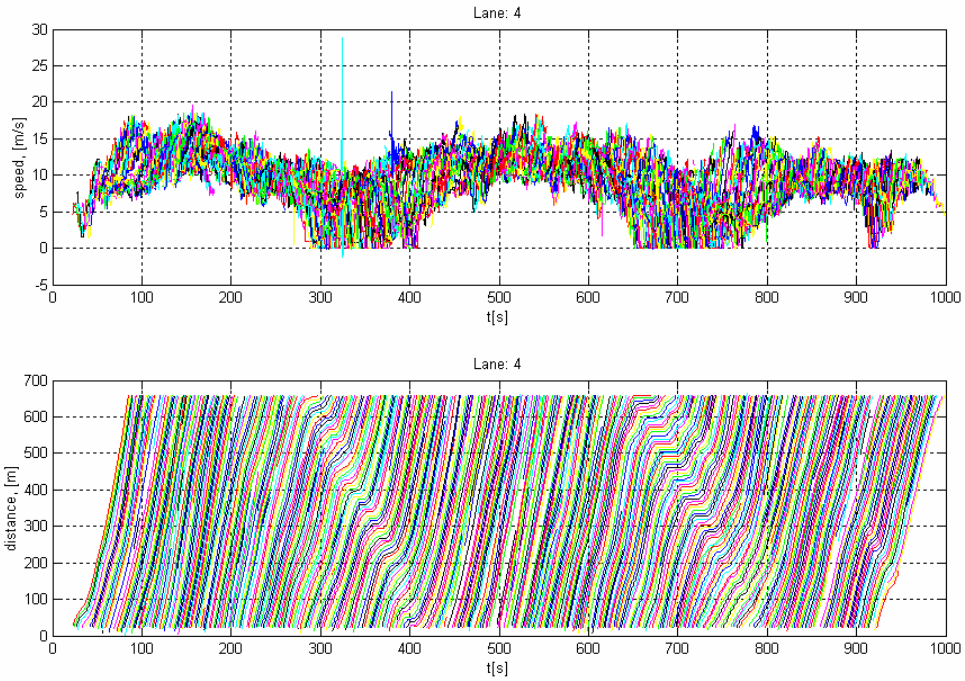


Figure B. 9. US-101, 08:05-08:20am, Lane 4

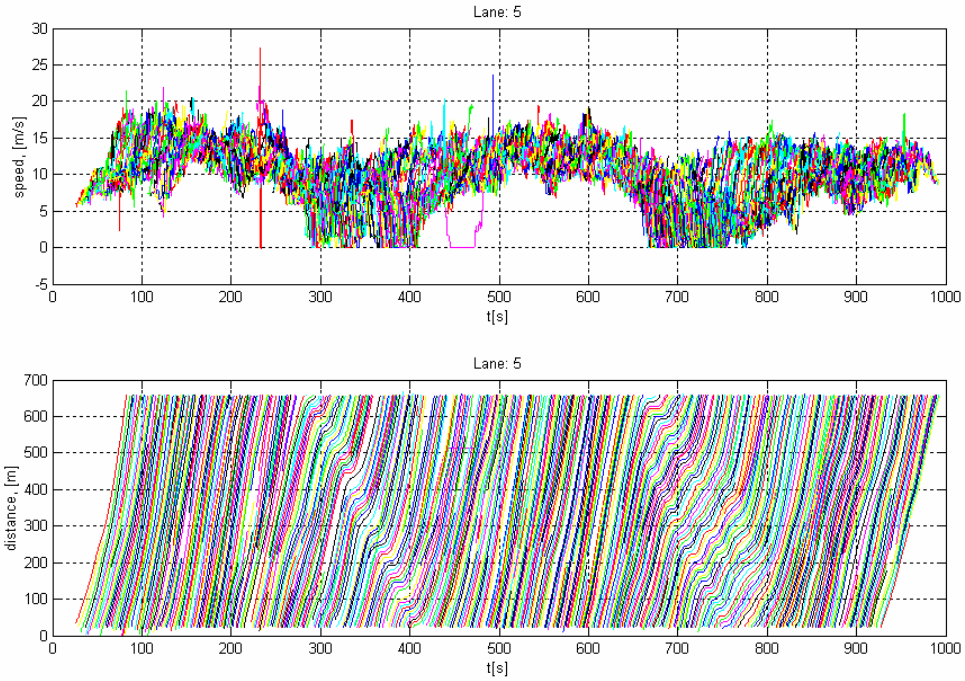


Figure B. 10. US-101, 08:05-08:20am, Lane 5

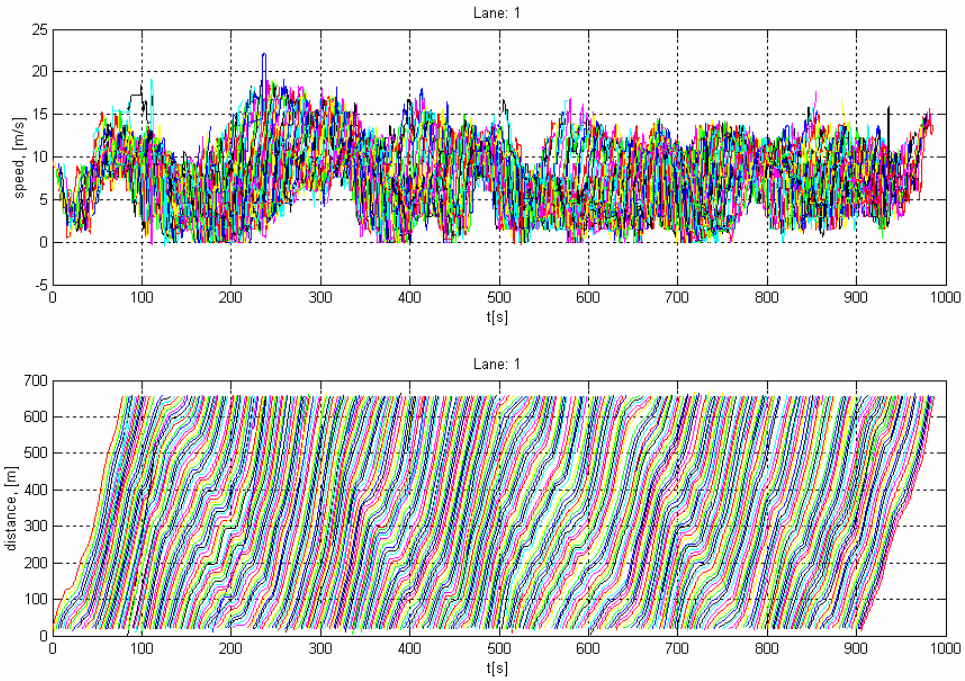


Figure B. 11. US-101, 08:20-08:35am, Lane 1

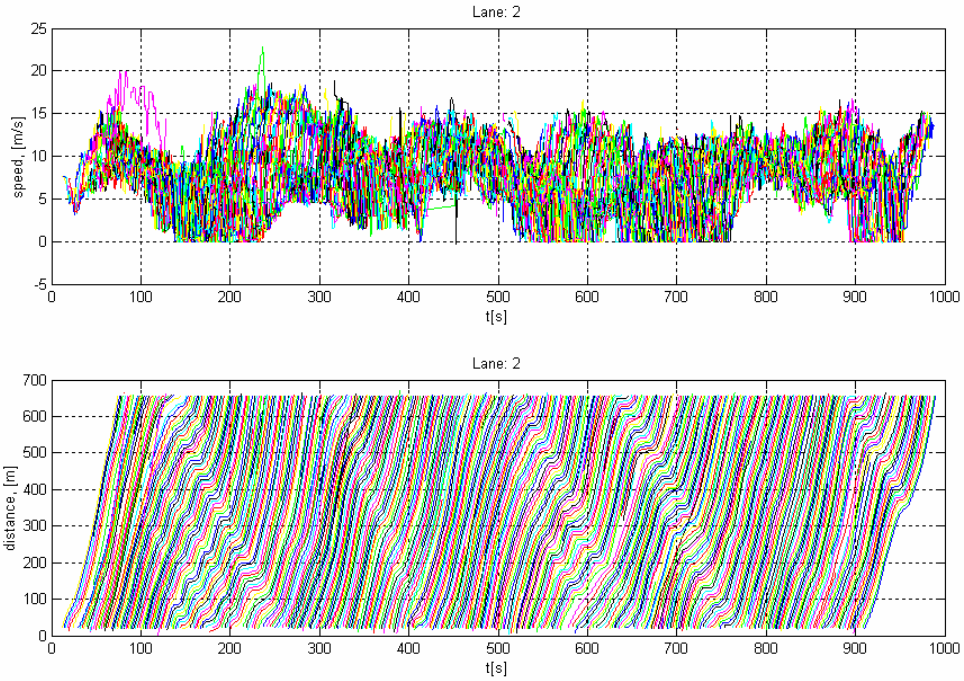


Figure B. 12. US-101, 08:20-08:35am, Lane 2

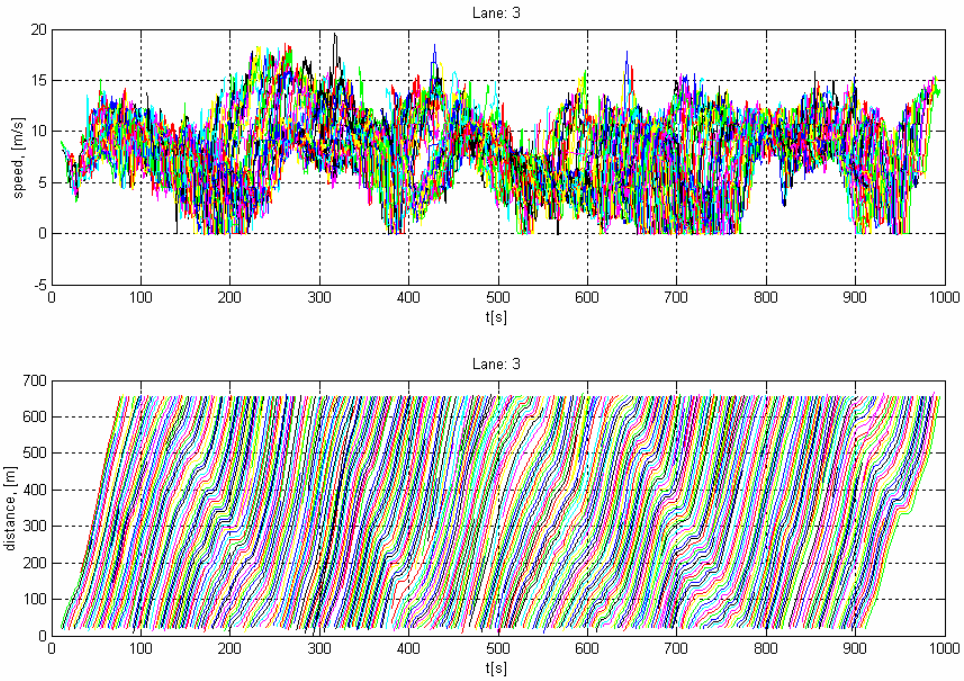


Figure B. 13. US-101, 08:20-08:35am, Lane 3

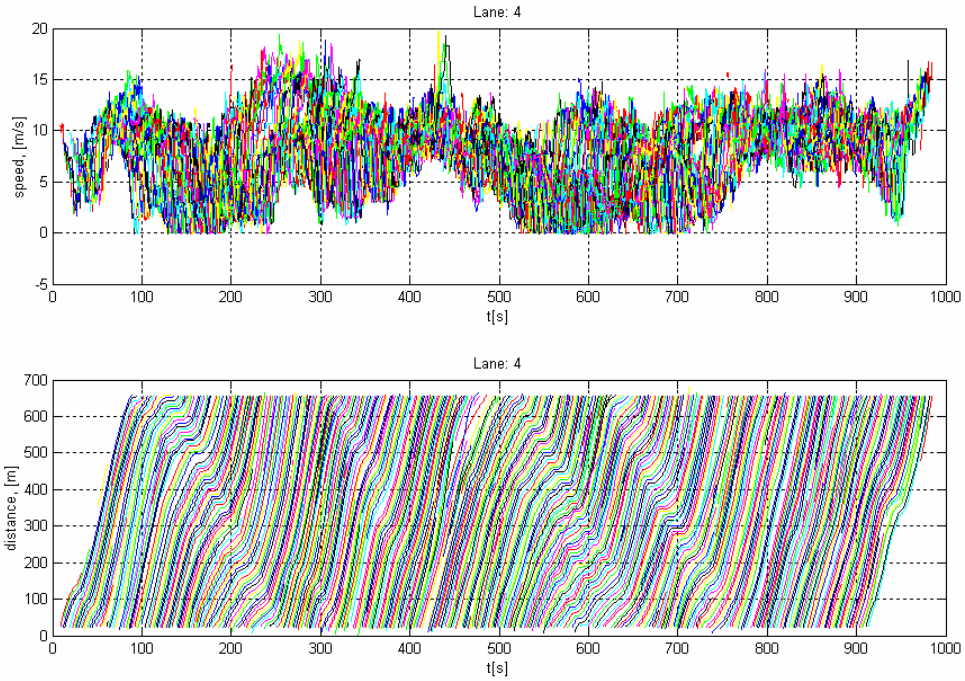


Figure B. 14. US-101, 08:20-08:35am, Lane 4

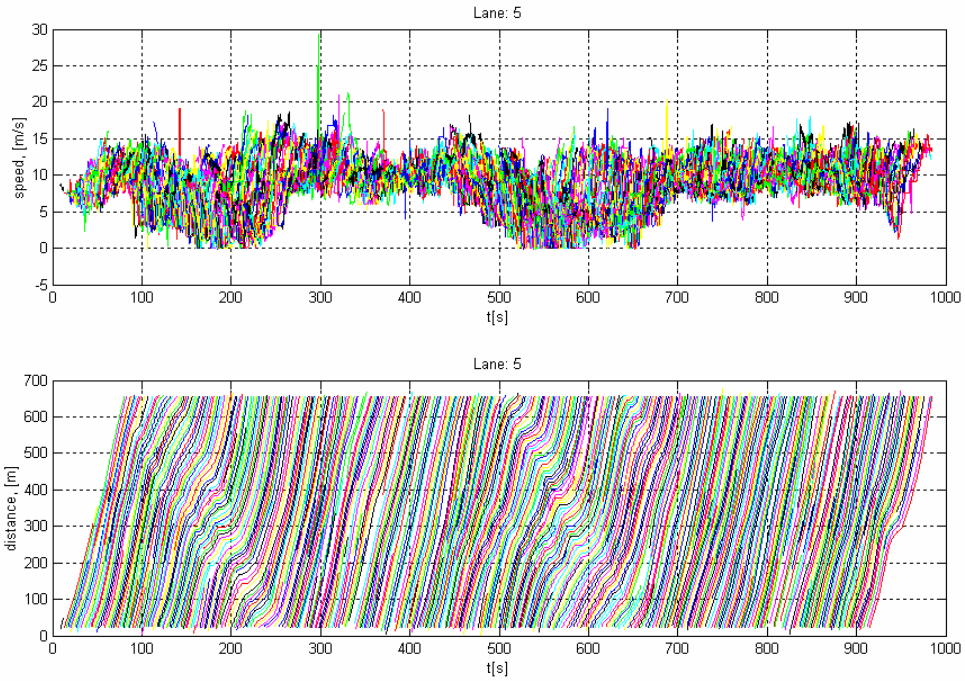


Figure B. 15. US-101, 08:20-08:35am, Lane 5

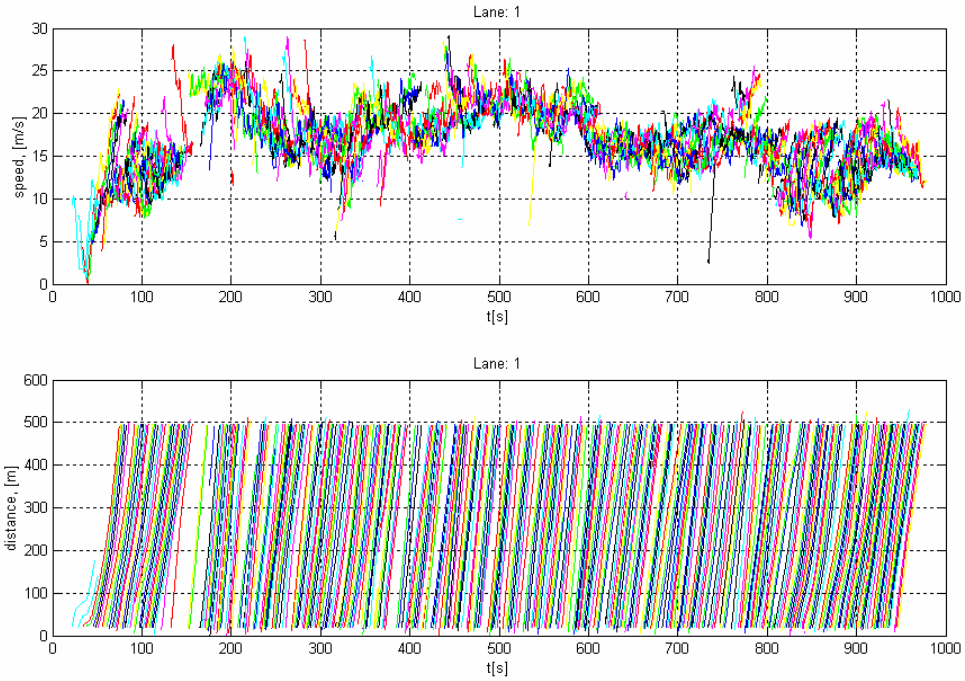


Figure B. 16. I-80, 04:00-04:15 pm, Lane 1

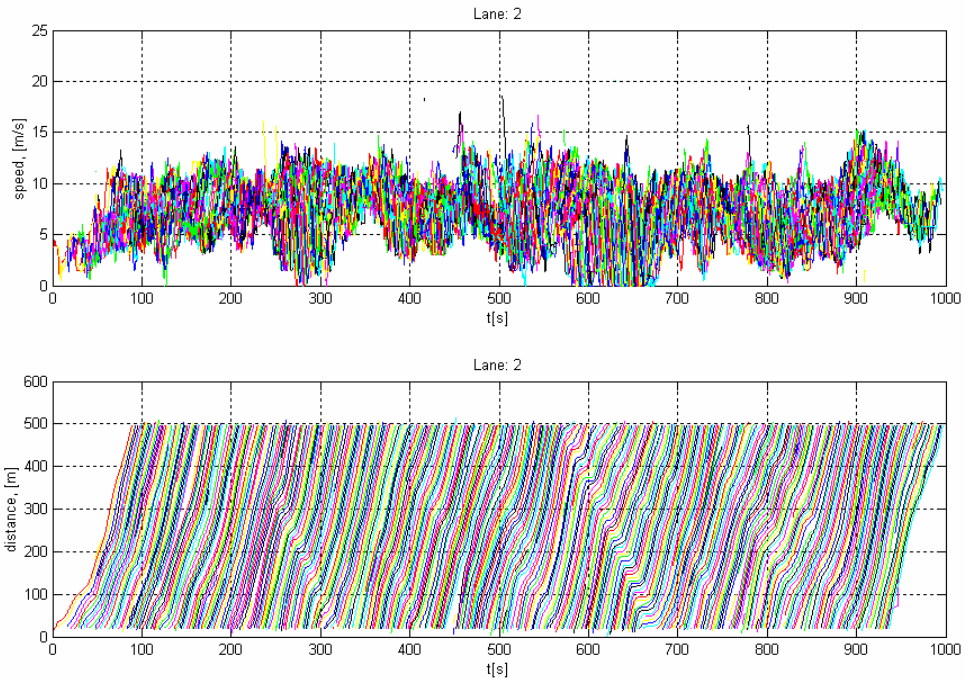


Figure B. 17. I-80, 04:00-04:15 pm, Lane 2

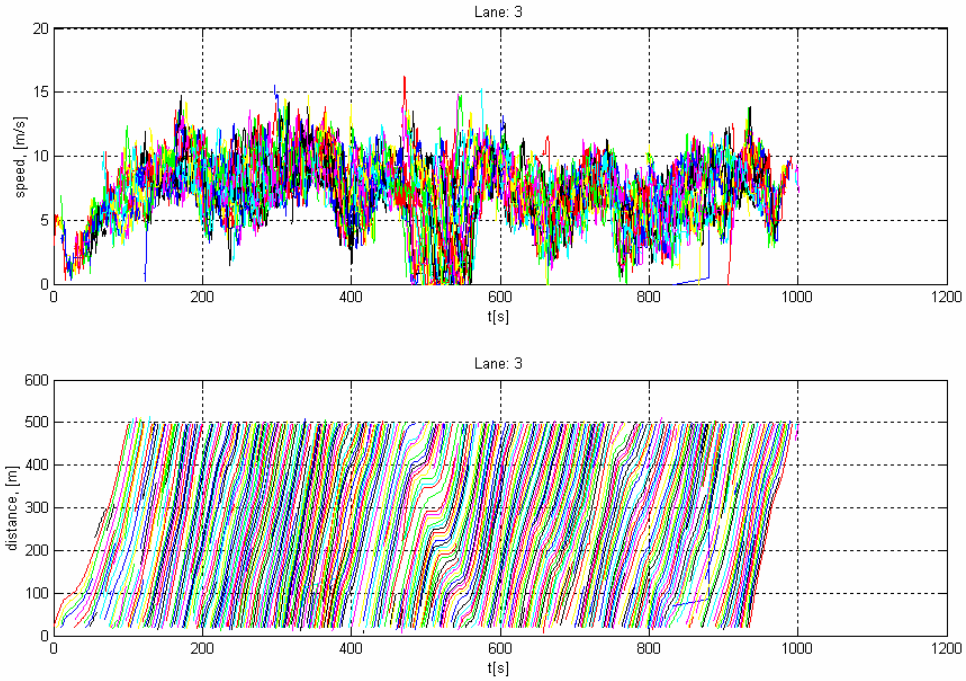


Figure B. 18. I-80, 04:00-04:15 pm, Lane 3

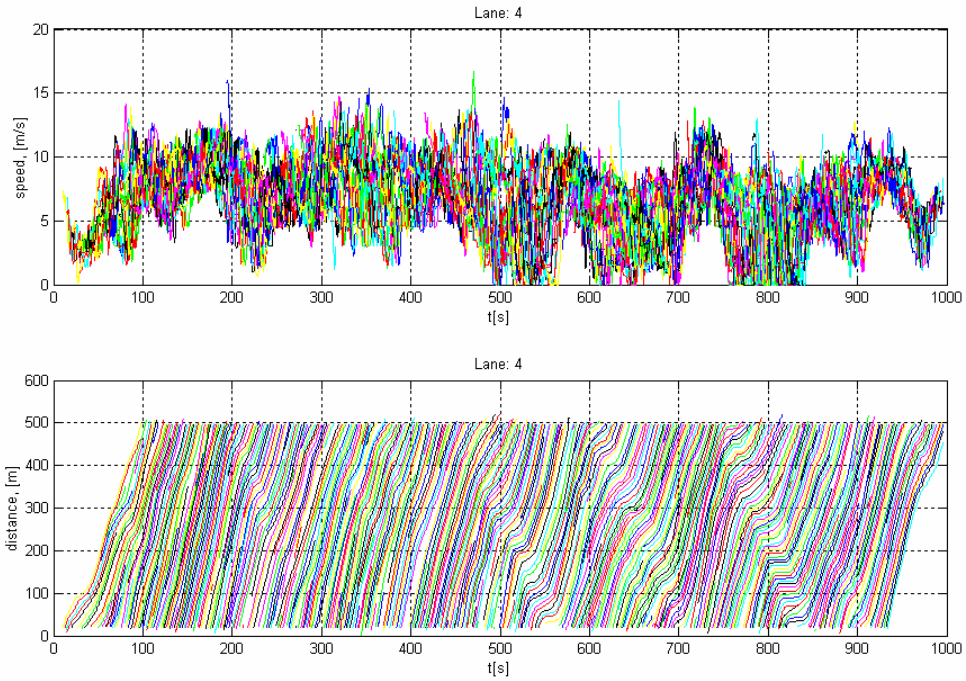


Figure B. 19. I-80, 04:00-04:15 pm, Lane 4

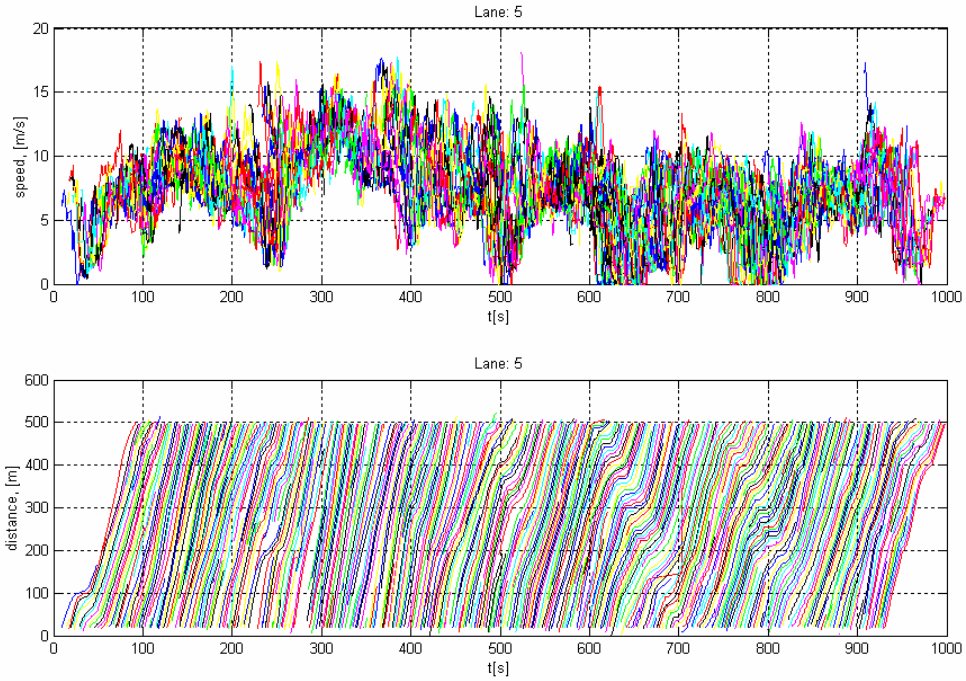


Figure B. 20. I-80, 04:00-04:15 pm, Lane 5

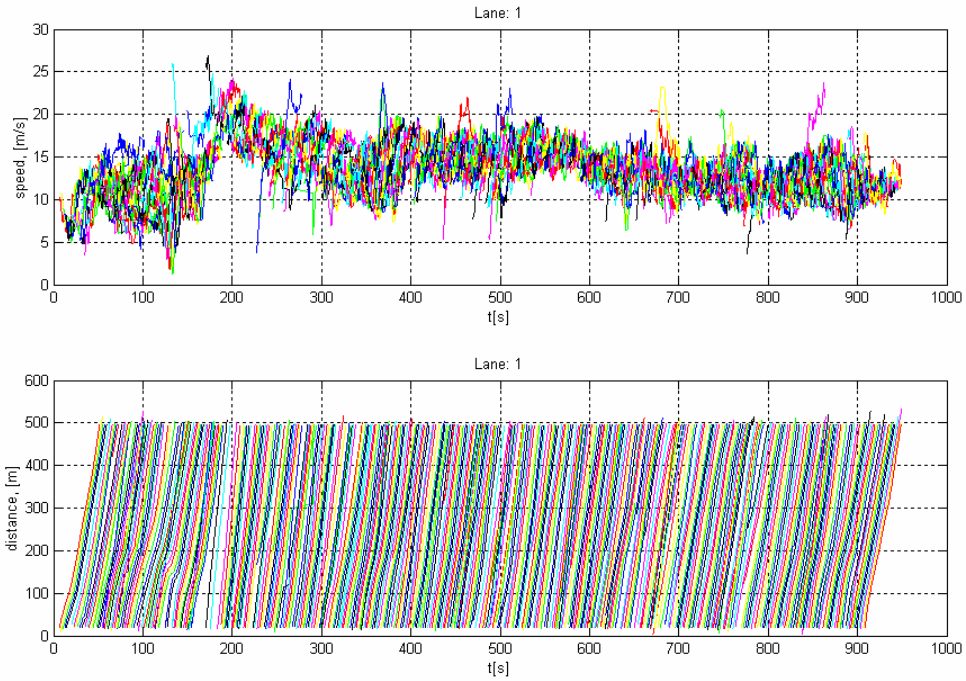


Figure B. 21. I-80, 05:00-05:15 pm, Lane 1

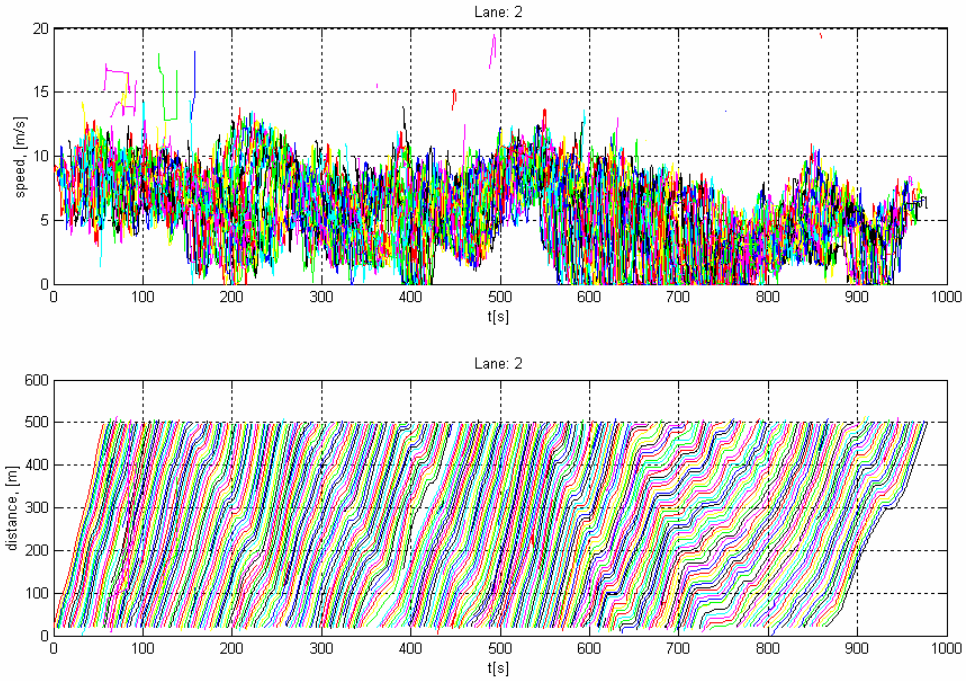


Figure B. 22. I-80, 05:00-05:15 pm, Lane 2

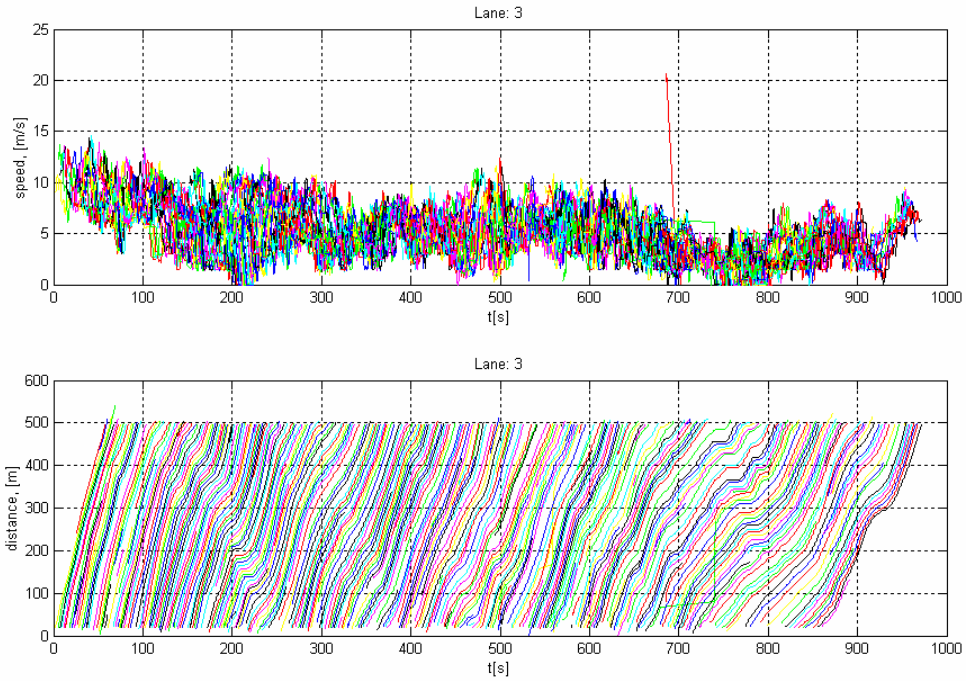


Figure B. 23. I-80, 05:00-05:15 pm, Lane 3

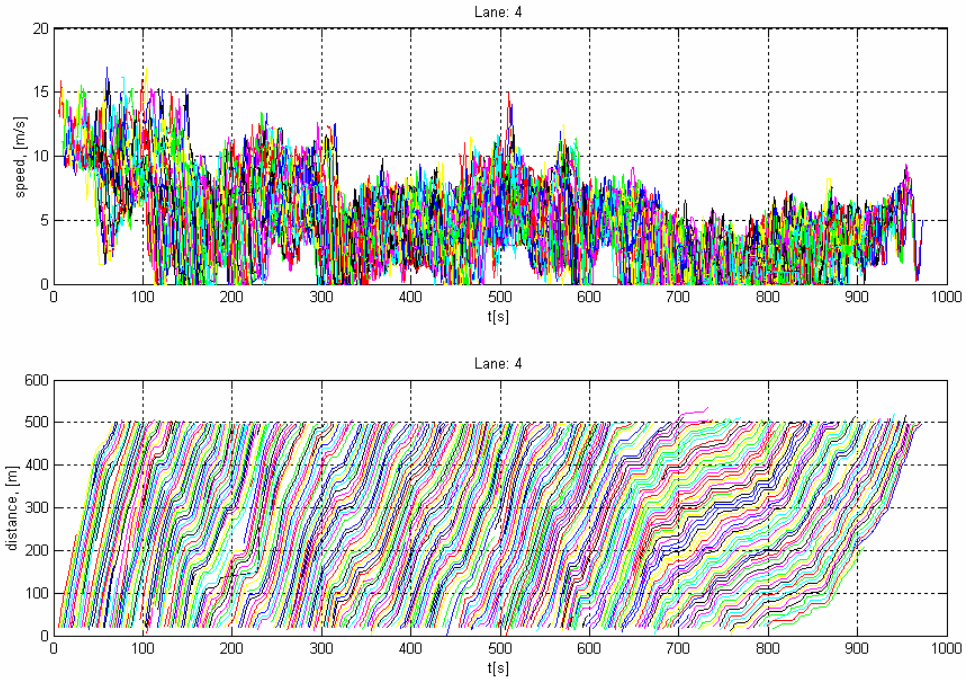


Figure B. 24. I-80, 05:00-05:15 pm, Lane 4

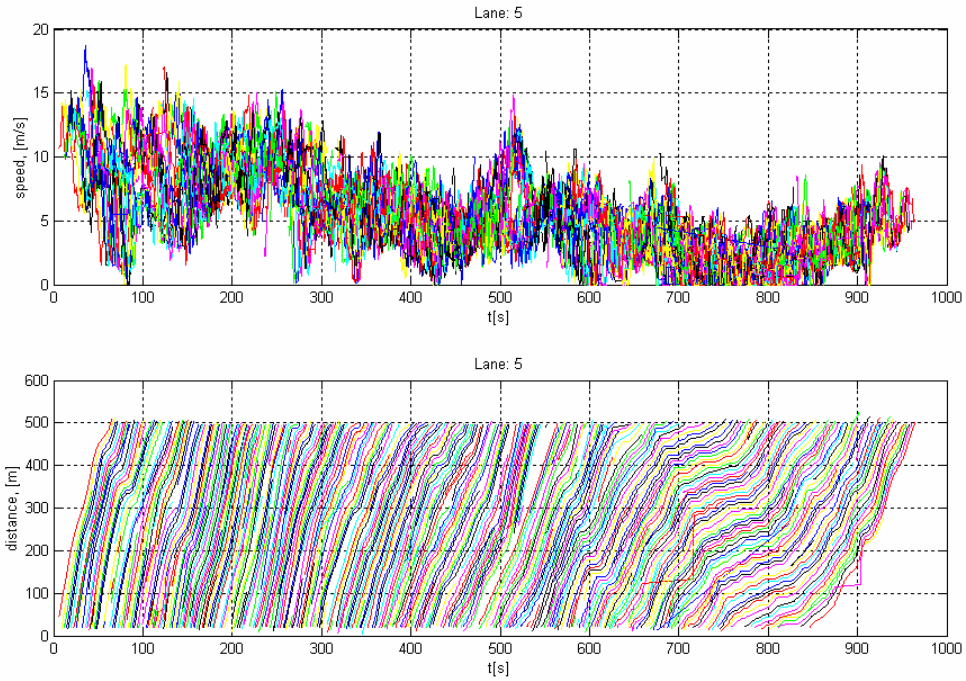


Figure B. 25. I-80, 05:00-05:15 pm, Lane 5

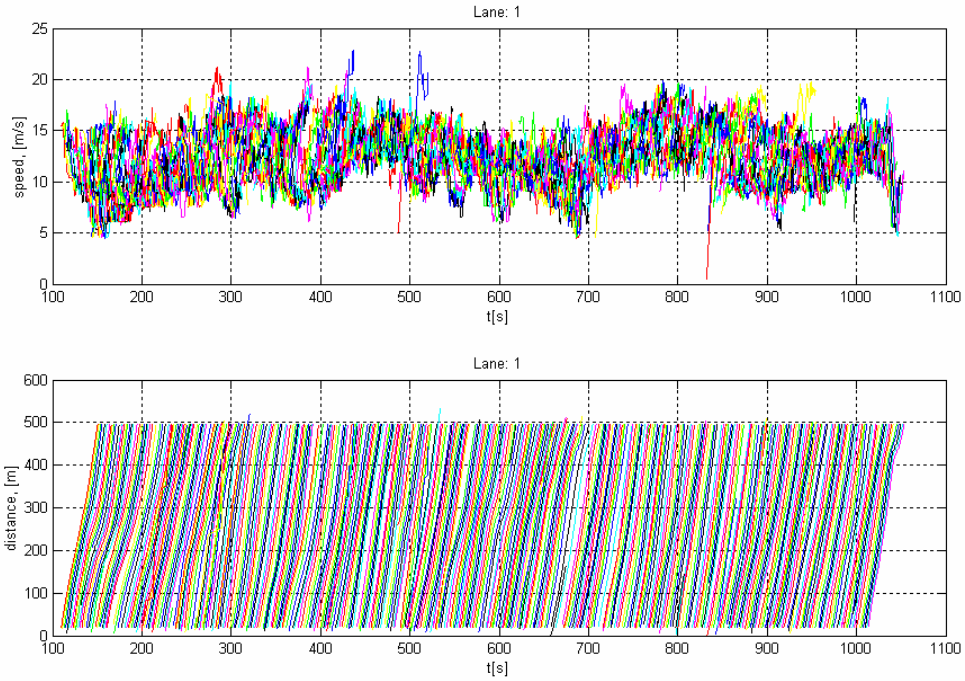


Figure B. 26. I-80, 05:15-05:30 pm, Lane 1

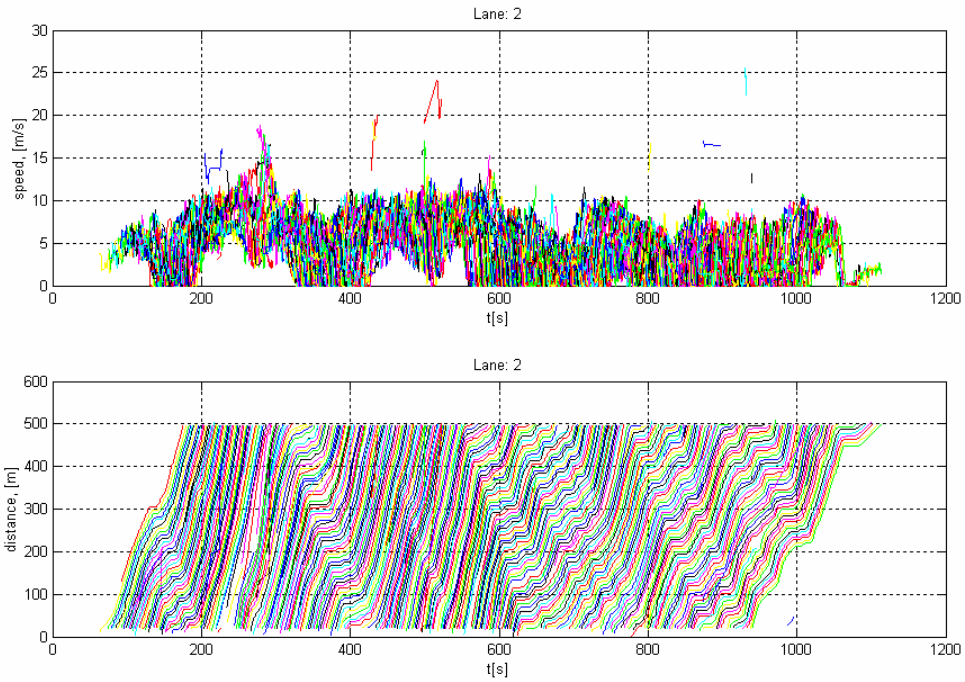


Figure B. 27. I-80, 05:15-05:30 pm, Lane 2

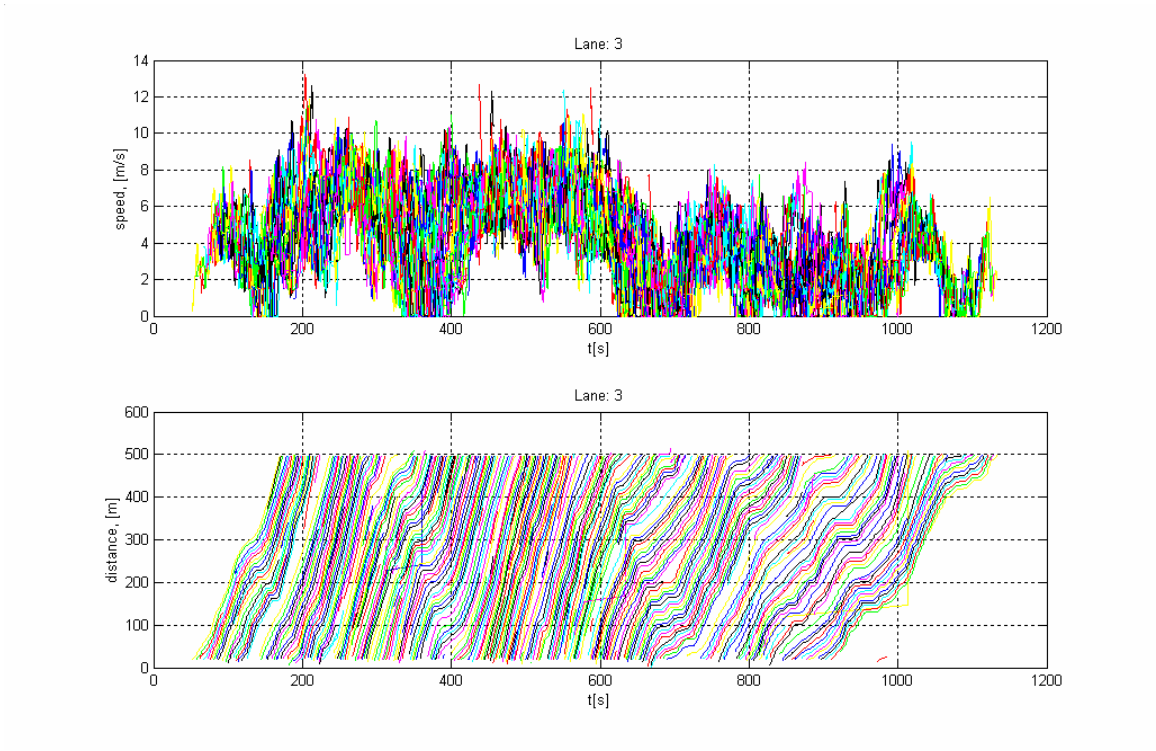


Figure B. 28. I-80, 05:15-05:30 pm, Lane 3

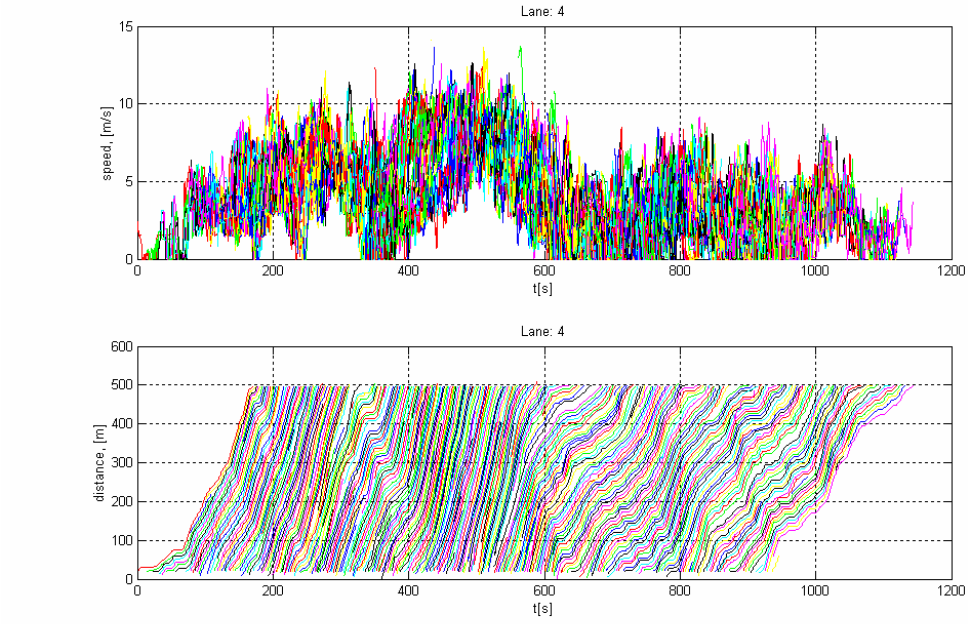


Figure B. 29. I-80, 05:15-05:30 pm, Lane 4

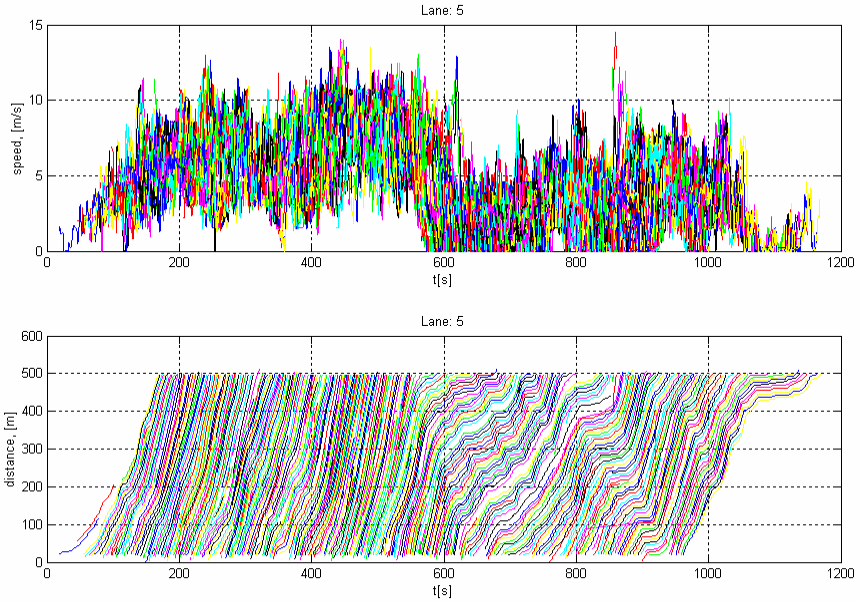


Figure B. 30. I-80, 05:15-05:30 pm, Lane 5

Appendix C

Moving median loop speed trend and average wave speeds

Incident A

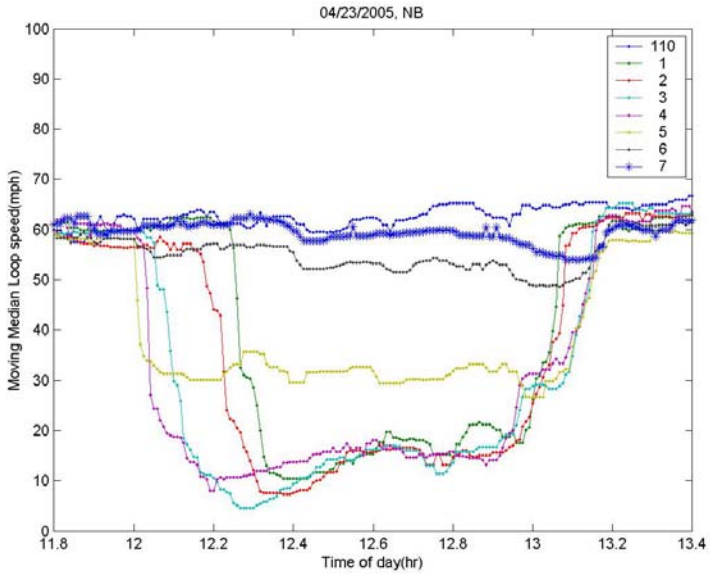


Figure C.1, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
6	-	-	24594	-	-	
5	12:01:00	13:11:30	22819	-	-	
4	12:02:30	13:10:00	20829	-15.1	11.3	St 5 – St 4
3	12:05:30	13:09:00	19114	-6.5	19.5	St 4 – St 3
2	12:13:00	13:06:00	17114	-3.0	7.6	St 3 – St 2
1	12:16:00	13:04:30	15766	-5.1	10.2	St 2 – St 1
110	-	-	12055	-	-	

Table C.1, Wave speed from incident data on April 23, 2005 (NB).

Moving wave speed from station 5 to station 1 is as follows;

Average upstream moving wave speed = -5.3 mph

Average downstream moving wave speed = 11.4 mph.

Incident B

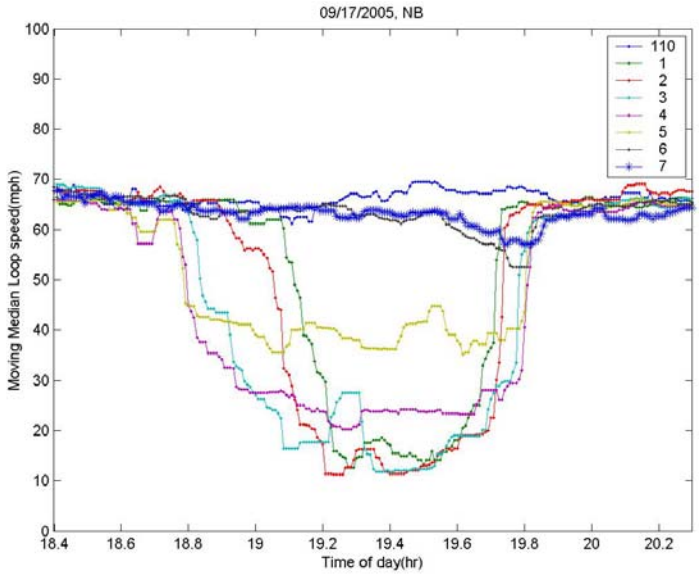


Figure C.2, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
7	-	-	27504	-	-	
6	-	-	24594	-	-	
5	-	-	22819	-	-	
4	18:48:00	19:48:00	20829	-	-	St 5 – St 4
3	18:54:00	19:46:30	19114	-3.2	13.0	St 4 – St 3
2	19:04:00	19:44:00	17114	-2.3	9.1	St 3 – St 2
1	19:08:30	19:42:30	15766	-3.4	10.2	St 2 – St 1
110	-	-	12055	-	-	

Table C.2, Wave speed from real data on September 17, 2005 (NB)

Moving wave speed from station 4 to station 1 is as follows;

Average upstream moving wave speed = -2.8 mph

Average downstream moving wave speed = 10.5 mph

Incident C

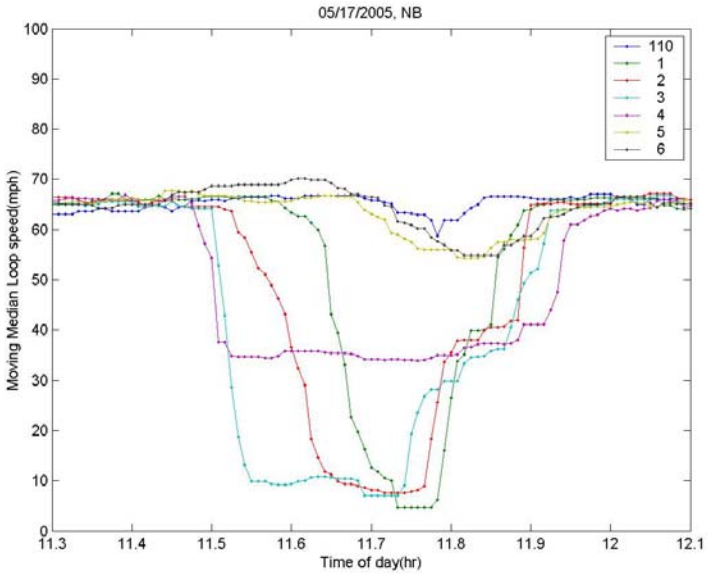


Figure C.3, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
6	-	-	24594	-	-	
5	-	-	22819	-	-	
4	11:28:30	11:55:00	20829	-	-	
3	11:30:00	11:52:30	19114	-13.0	7.8	St 4 – St 3
2	11:31:30	11:51:00	17114	-15.2	15.2	St 3 – St 2
1	11:34:00	11:49:30	15766	-6.1	10.2	St 2 – St 1
110	-	-	12055	-	-	

Table C.3, Wave speed from real data on May 17, 2005 (NB)

Moving wave speed from station 4 to station 1 is as follows;

Average upstream moving wave speed = -10.5 mph

Average downstream moving wave speed = 10.5 mph

Incident D

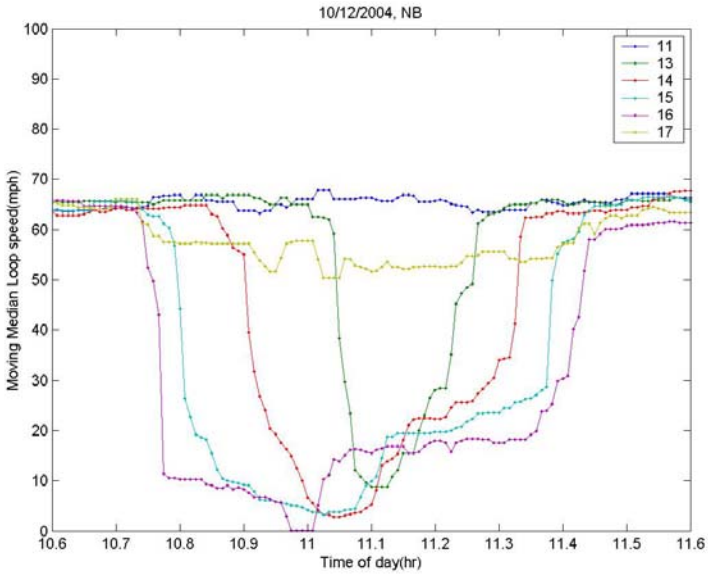


Figure C.4, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
17	-	-	46192	-	-	
16	10:46:00	11:24:30	44252	-	-	
15	10:48:00	11:22:30	42262	-11.3	11.3	St 16 – St 15
14	10:54:30	11:19:30	40762	-2.6	5.7	St 15 – St 14
13	11:03:00	11:13:30	38352	-3.2	4.6	St 14 – St 13
11	-	-	33892	-	-	

Table C.4, Wave speed from real data on October 12, 2004 (NB)

Moving wave speed from station 16 to station 13 is as follows;

Average upstream moving wave speed = -3.9 mph

Average downstream moving wave speed = 6.1 mph

Incident E

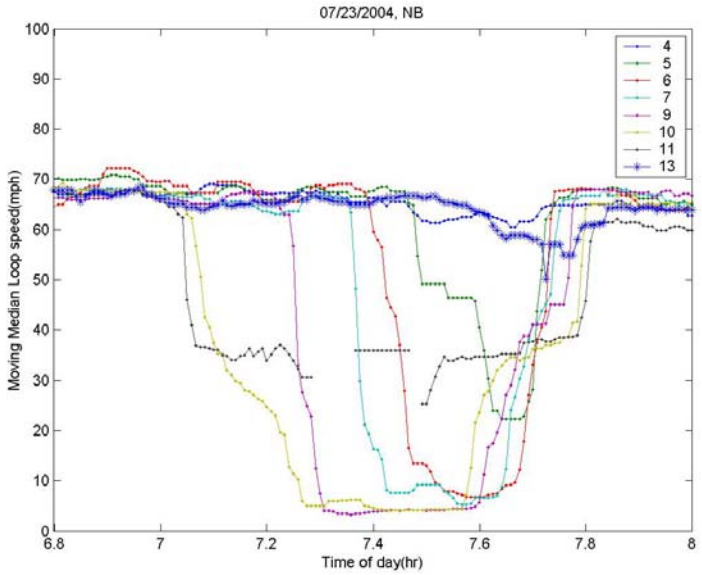


Figure C.5, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
13	-	-	38352	-	-	
11	-	-	33892	-	-	
10	7:05:00	7:48:00	32105	-	-	
9	7:15:30	7:46:30	30092	-2.2	15.3	St 10 – St 9
7	7:22:30	7:45:00	27504	-4.2	19.6	St 9 – St 7
6	7:26:00	7:44:00	24594	-9.4	33.1	St 7 – St 6
5	7:36:00	7:42:30	22819	-2.0	13.4	St 6 – St 5
4	-	-	20829	-	-	

Table C.5, Wave speed from real data on July 23, 2005 (NB)

Moving wave speed from station 10 to station 5 is as follows;

Average upstream moving wave speed = -3.4 mph

Average downstream moving wave speed = 19.2 mph

Incident F

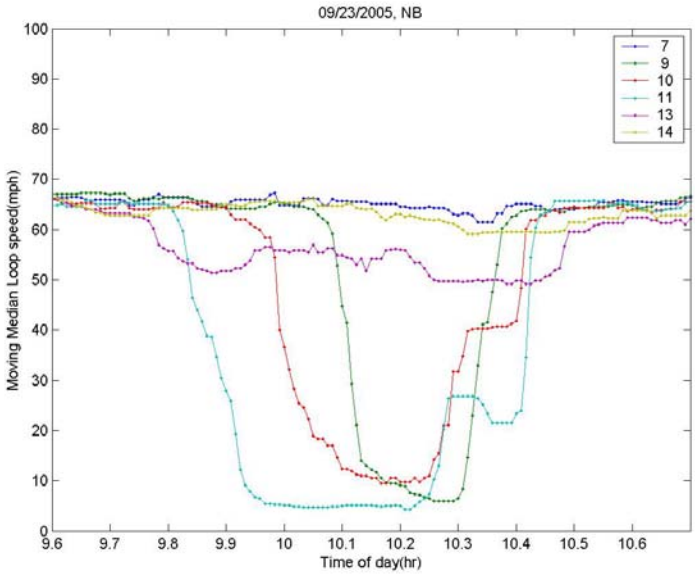


Figure C.6, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
14	-	-	40762	-	-	
13	-	-	38352	-	-	
11	9:51:00	10:25:30	33892	-	-	
10	9:59:30	10:24:00	32105	-2.4	13.5	St 11 – St 10
9	10:06:00	10:21:00	30092	-3.5	7.6	St 10 – St 9
7	-	-	27504	-	-	

Table C.6, Wave speed from real data on September 23, 2005 (NB)

Moving wave speed from station 11 to station 9 is as follows;

Average upstream moving wave speed = -2.9 mph

Average downstream moving wave speed = 9.6 mph

Incident G

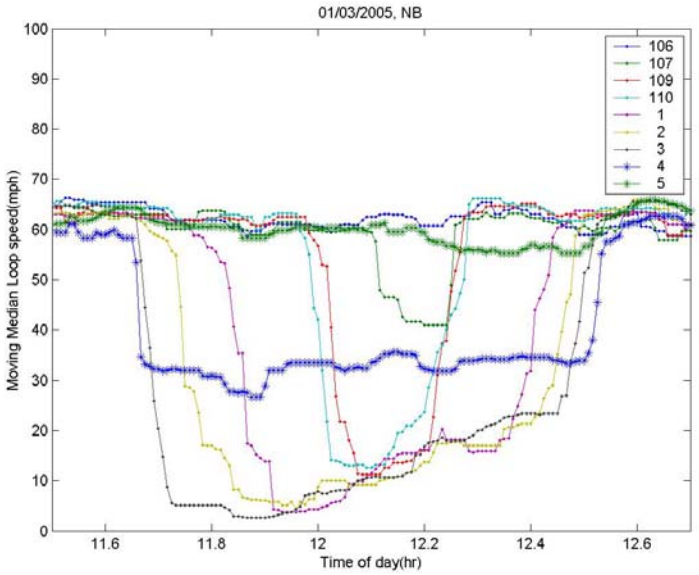


Figure C.7, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
5	-	-	22819	-	-	
4	11:40:00	12:31:00	20829	-	-	
3	11:41:00	12:29:30	19114	-19.5	13.0	St 4 – St 3
2	11:44:30	12:28:00	17114	-6.5	15.2	St 3 – St 2
1	11:50:30	12:24:30	15766	-2.6	4.4	St 2 – St 1
110	11:59:30	12:16:30	12055	-4.7	5.3	St 1 – St 110
109	12:01:30	12:15:30	10655	-8.0	15.9	St 110 – St 109
107	12:07:00	12:14:30	8969	-3.5	19.2	St 109 – St 107
106	-	-	6907	-	-	

Table C.7, Wave speed from real data on January 3, 2005 (NB)

Moving wave speed from station 4 to station 107 is as follows;

Average upstream moving wave speed = -5.0 mph

Average downstream moving wave speed = 8.2 mph

Incident H

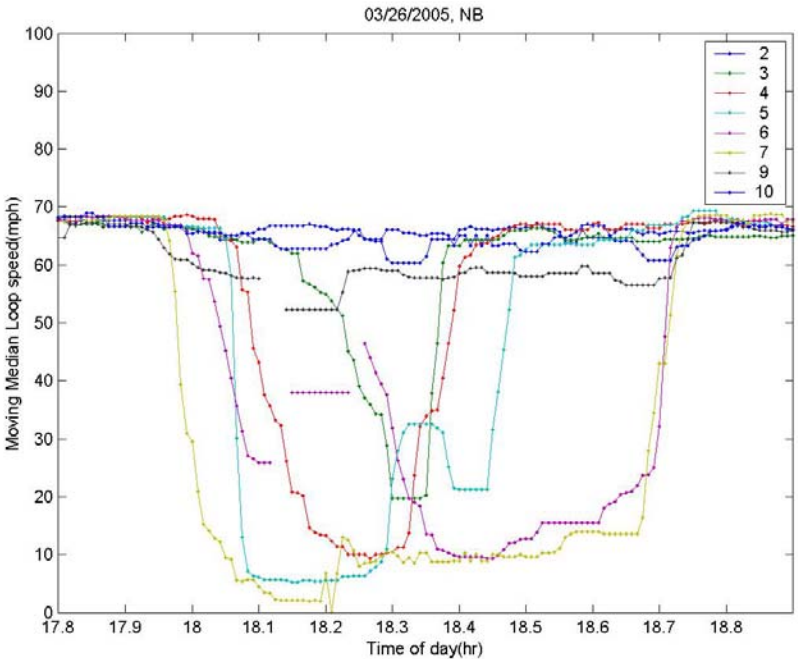


Figure C.8, Moving median loop speed trends over short time periods.

Station	T_1	T_2	Mile Marker (ft)	Upstream wave velocity, v_1 (mph)	Downstream wave velocity, v_2 (mph)	Remarks
9	-	-	30092	-	-	
7	17:59:00	18:40:00	27504	-	-	
6	18:01:00	18:36:30	24594	-16.5	9.4	St 7 – St 6
5	18:03:30	18:26:30	22819	-8.1	2.0	St 6 – St 5
4	18:05:00	18:24:00	20829	-15.1	9.0	St 5 – St 4
3	18:10:00	18:22:30	19114	-3.9	13.0	St 4 – St 3
2	-	-	17114	-	-	

Table C.8, Wave speed from real data on March 26, 2006 (NB)

Moving wave speed from station 7 to station 3 is as follows;

Average upstream moving wave speed = -8.7 mph

Average downstream moving wave speed = 5.4 mph

Appendix D

Estimation errors by link locations for present loop detector station spacing

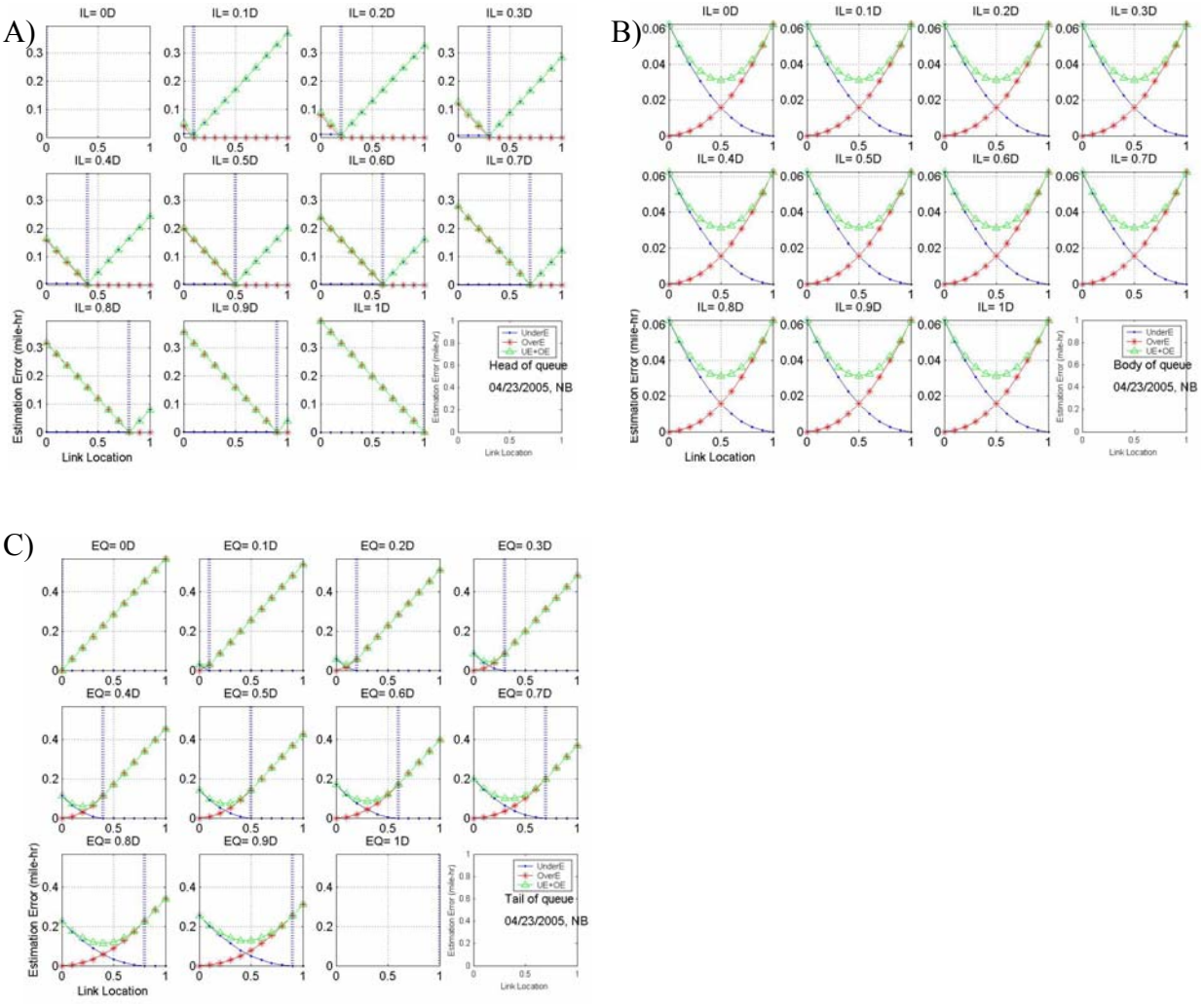
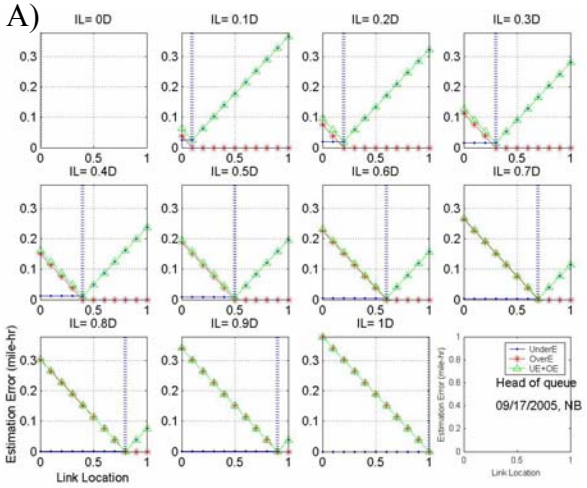
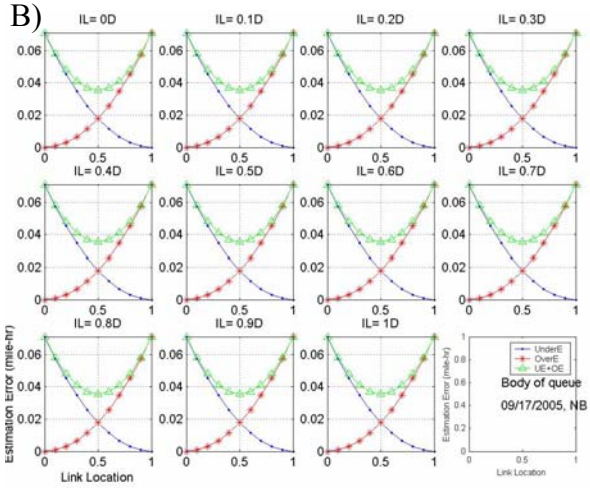


Figure D.1, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident A (04/23/2005)

A)



B)



C)

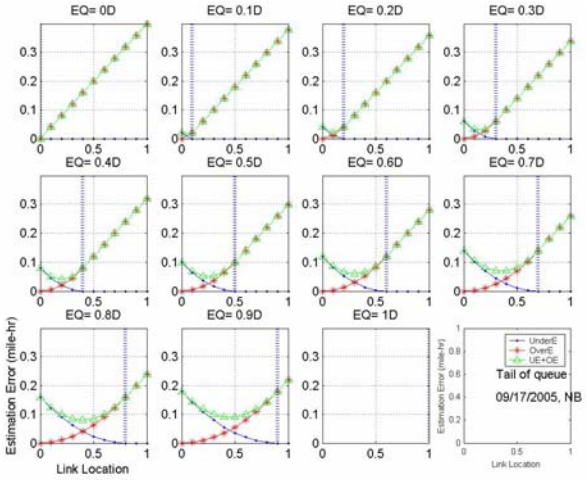


Figure D.2, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident B (09/17/2005)

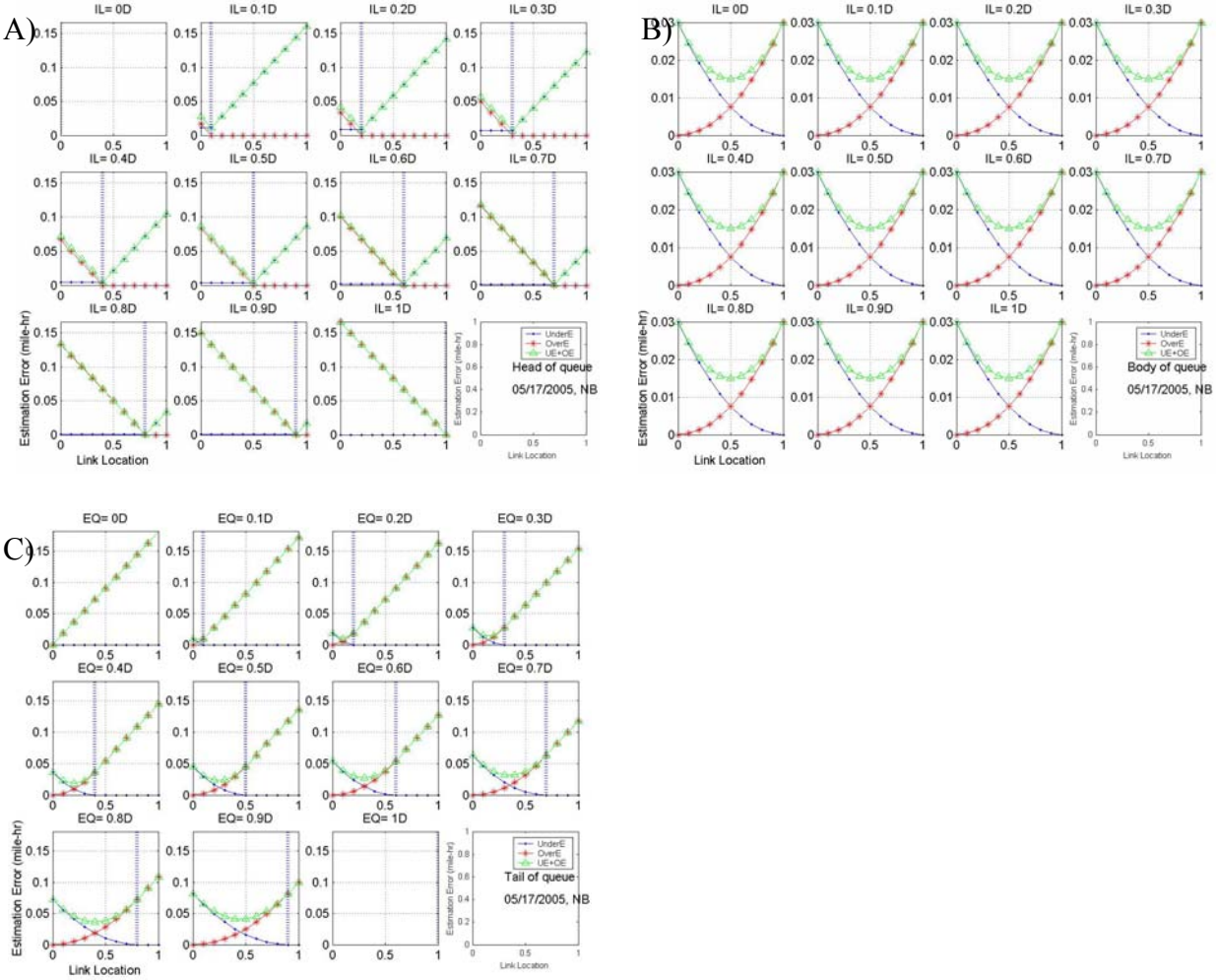


Figure D.3, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident C (05/17/2005)

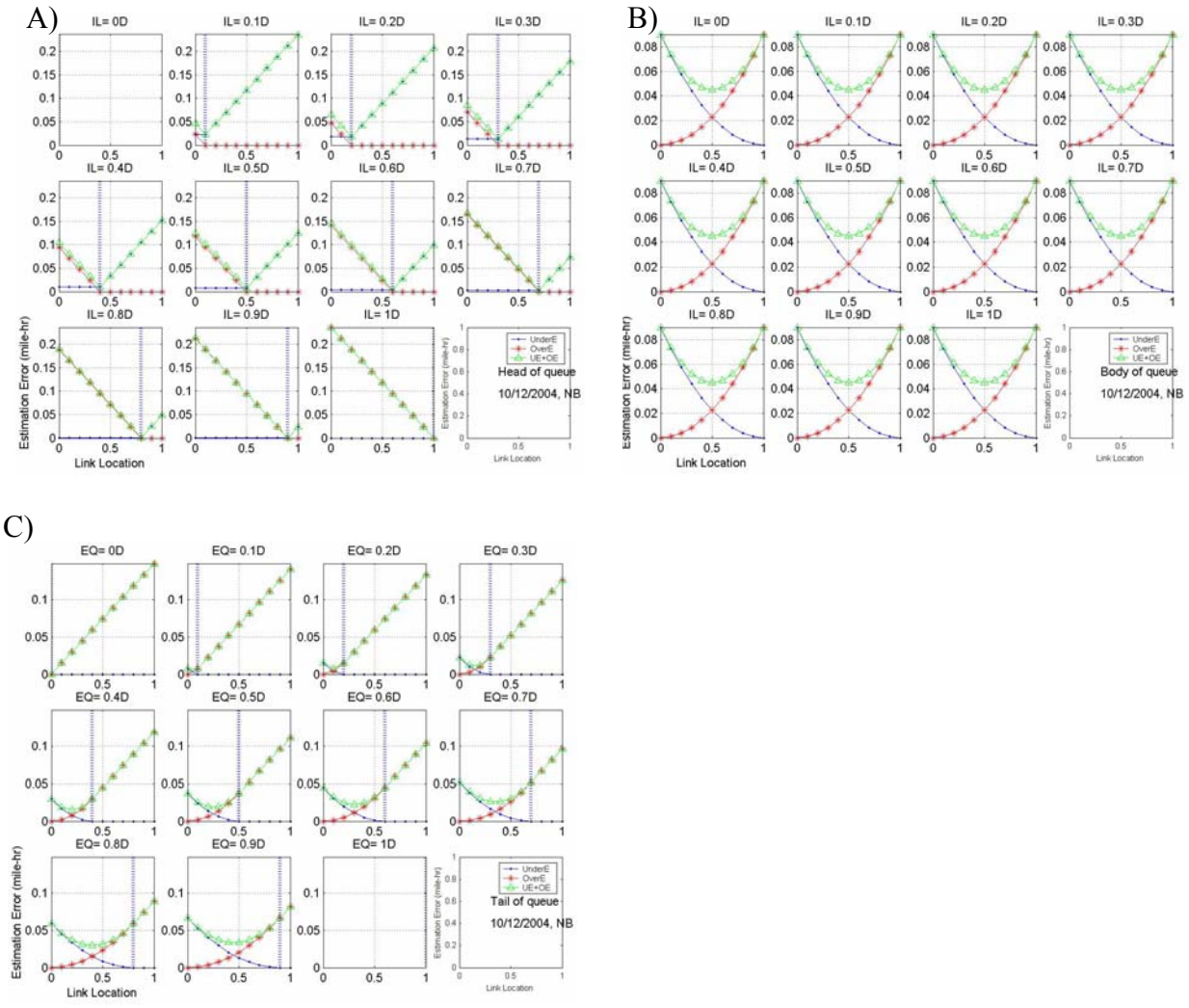
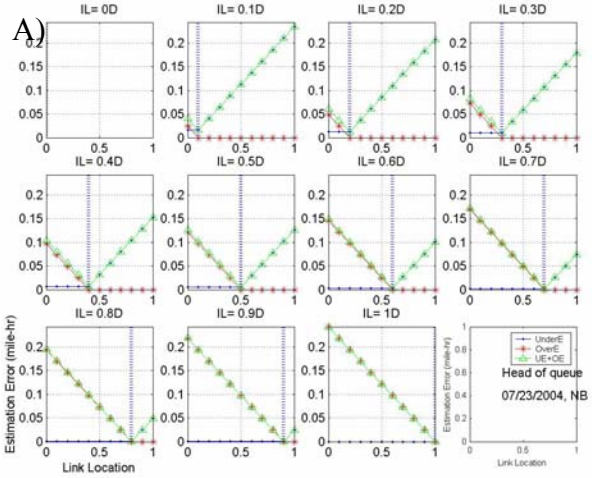
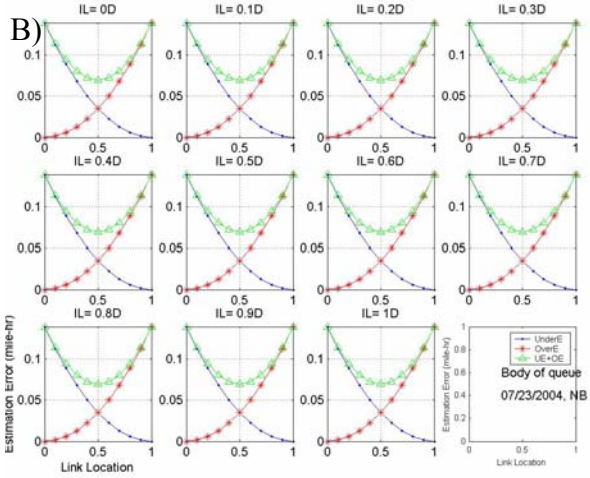


Figure D.4, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident D (10/12/2004)

A)



B)



C)

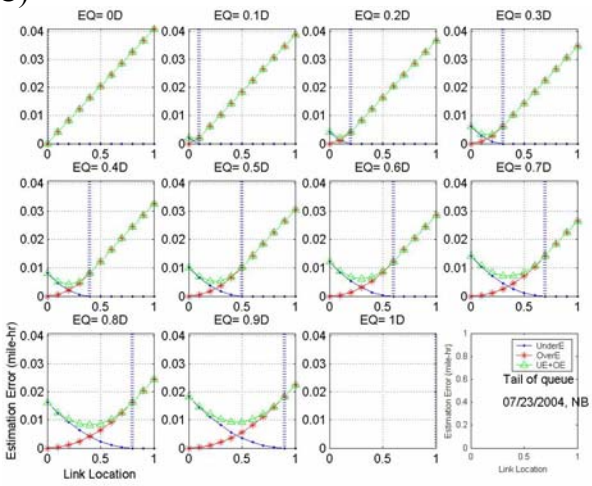


Figure D.5, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident E (07/23/2004)

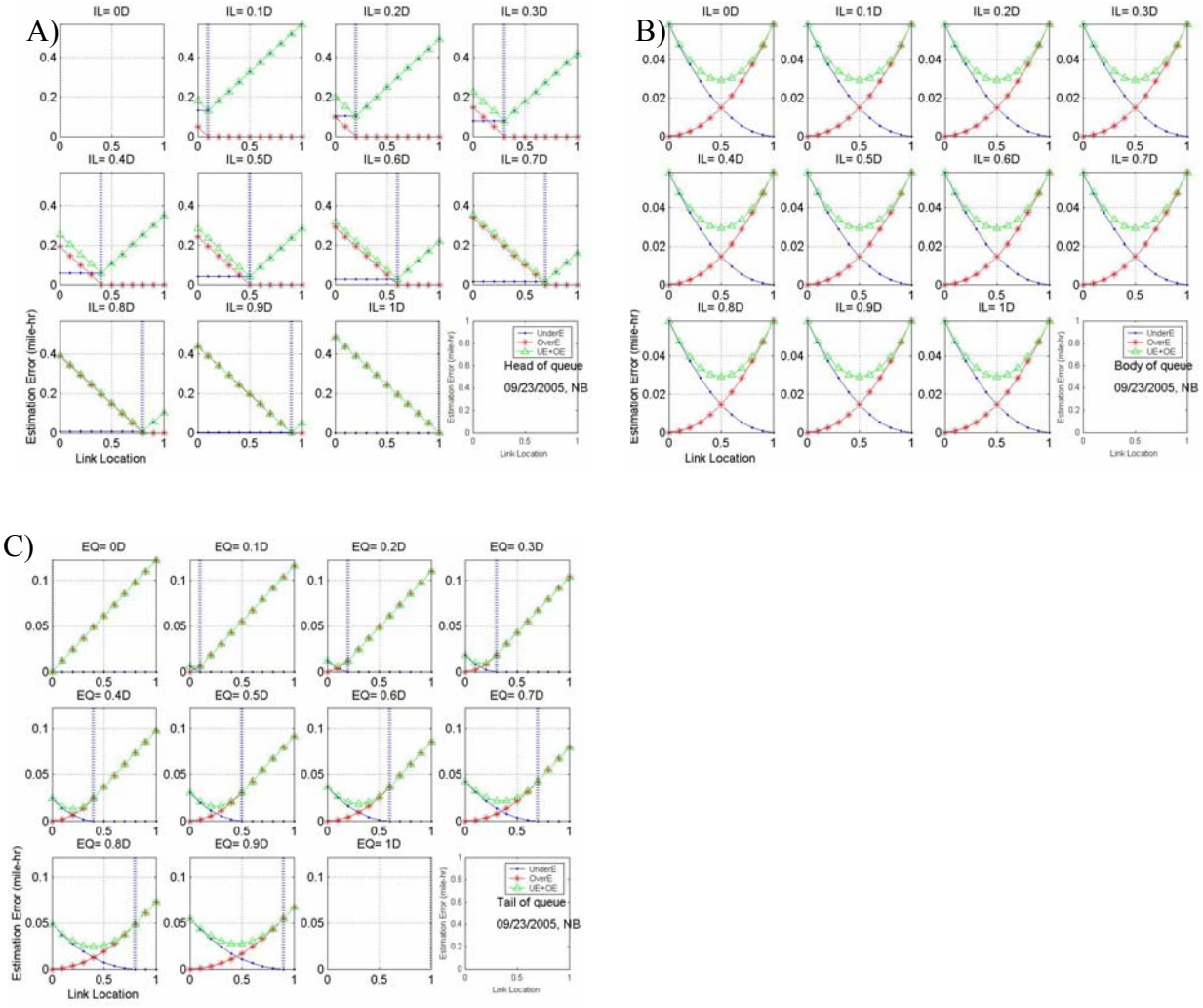


Figure D.6, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident F (09/23/2005)

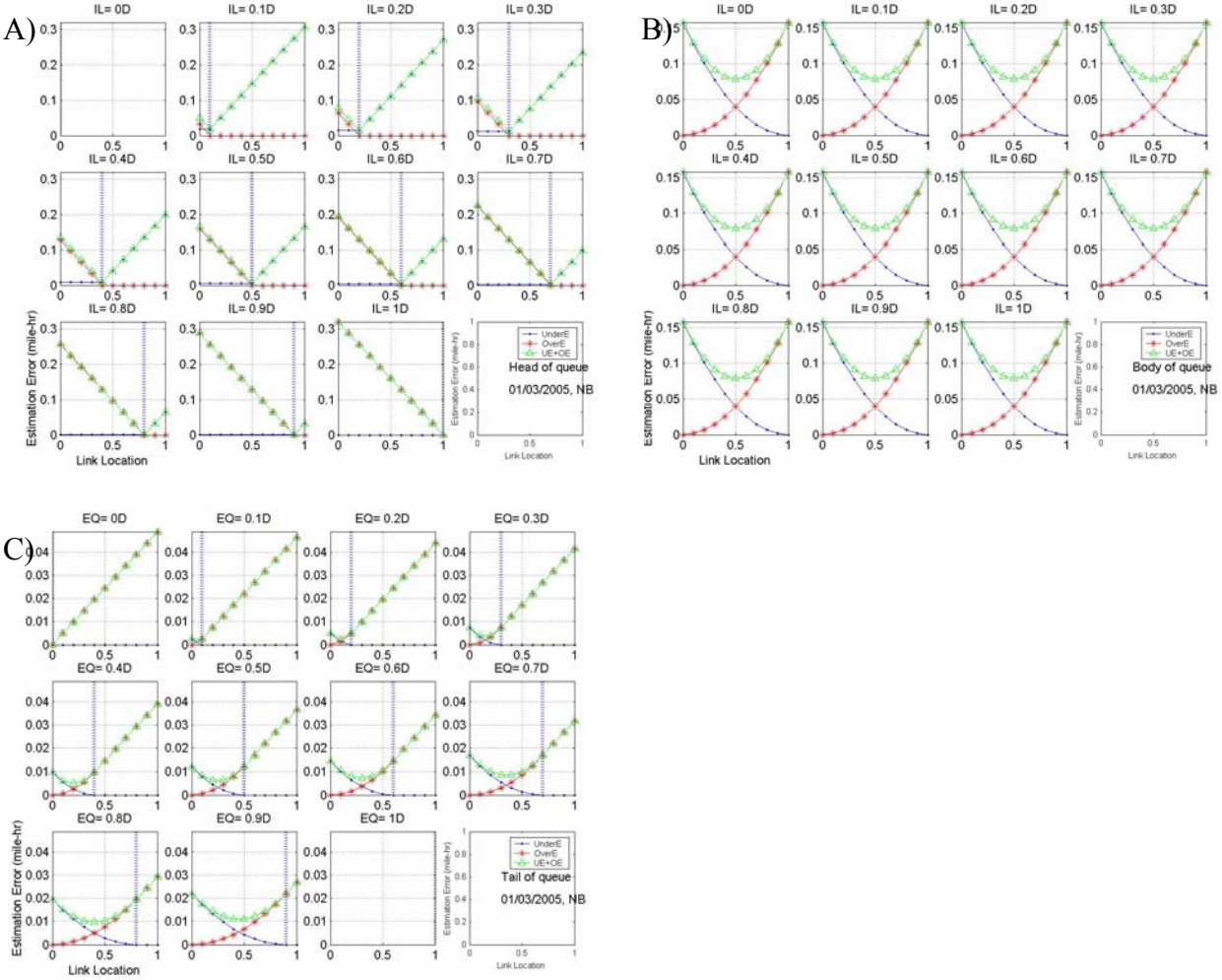


Figure D.7, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident G (01/03/2005)

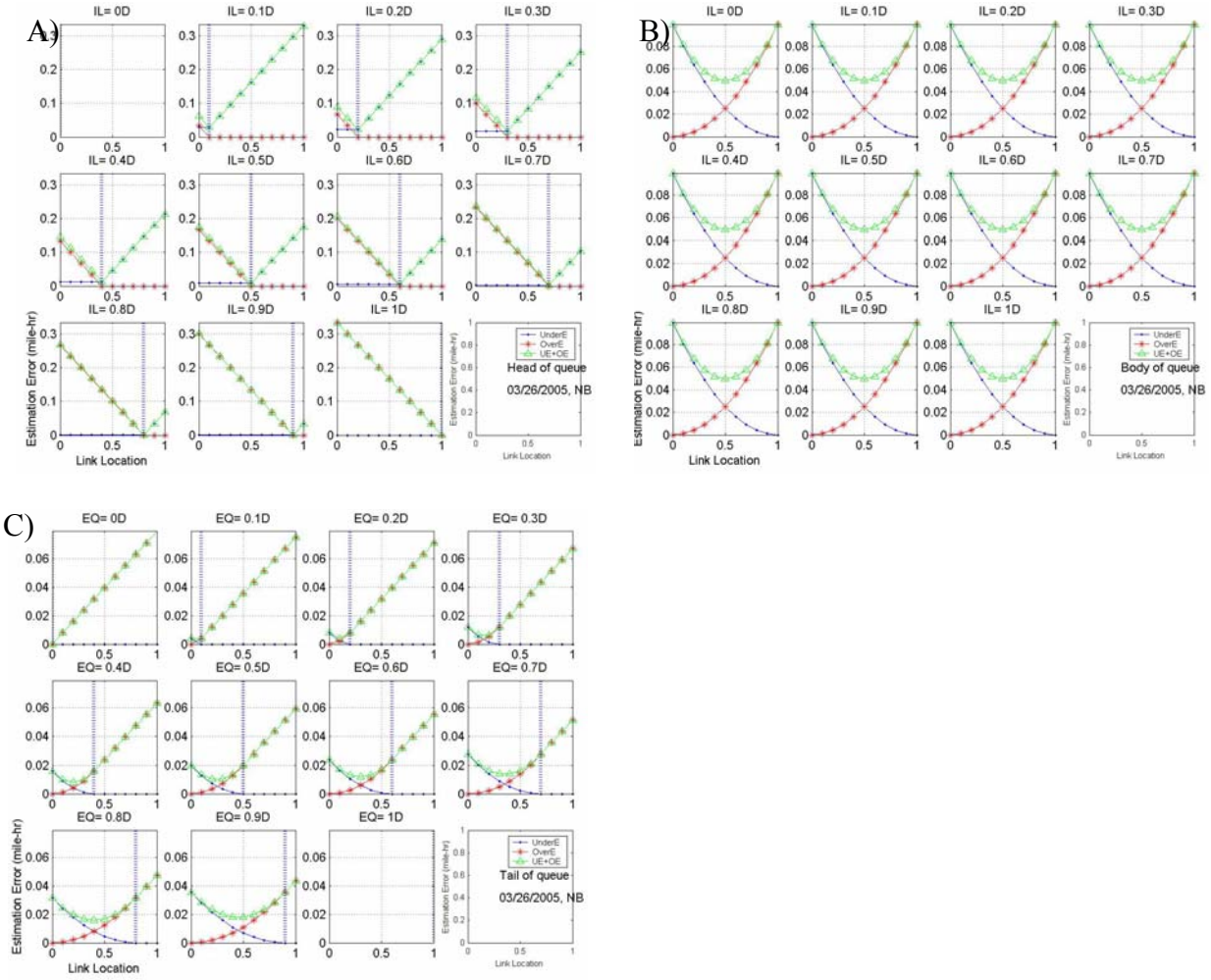


Figure D.8, Estimation error on three segments in present loop detector station spacing (mi-hr) Incident H (03/26/2005)

Appendix E

Average estimation errors by link location for present loop detector station spacing

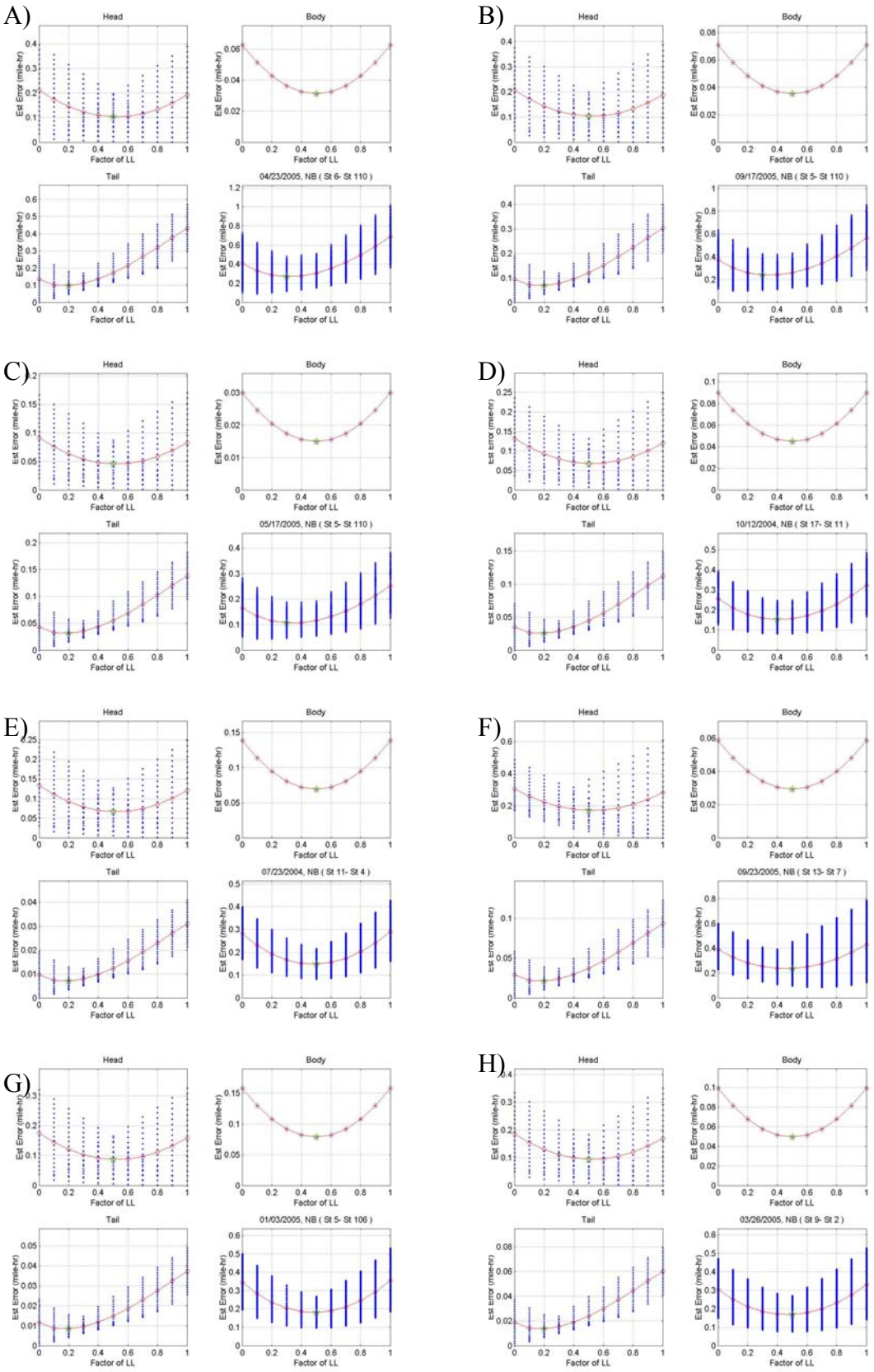


Figure E.1, Average estimation errors in each of three segments and in whole segments by link locations for present loop detector station spacing (mi-hr).

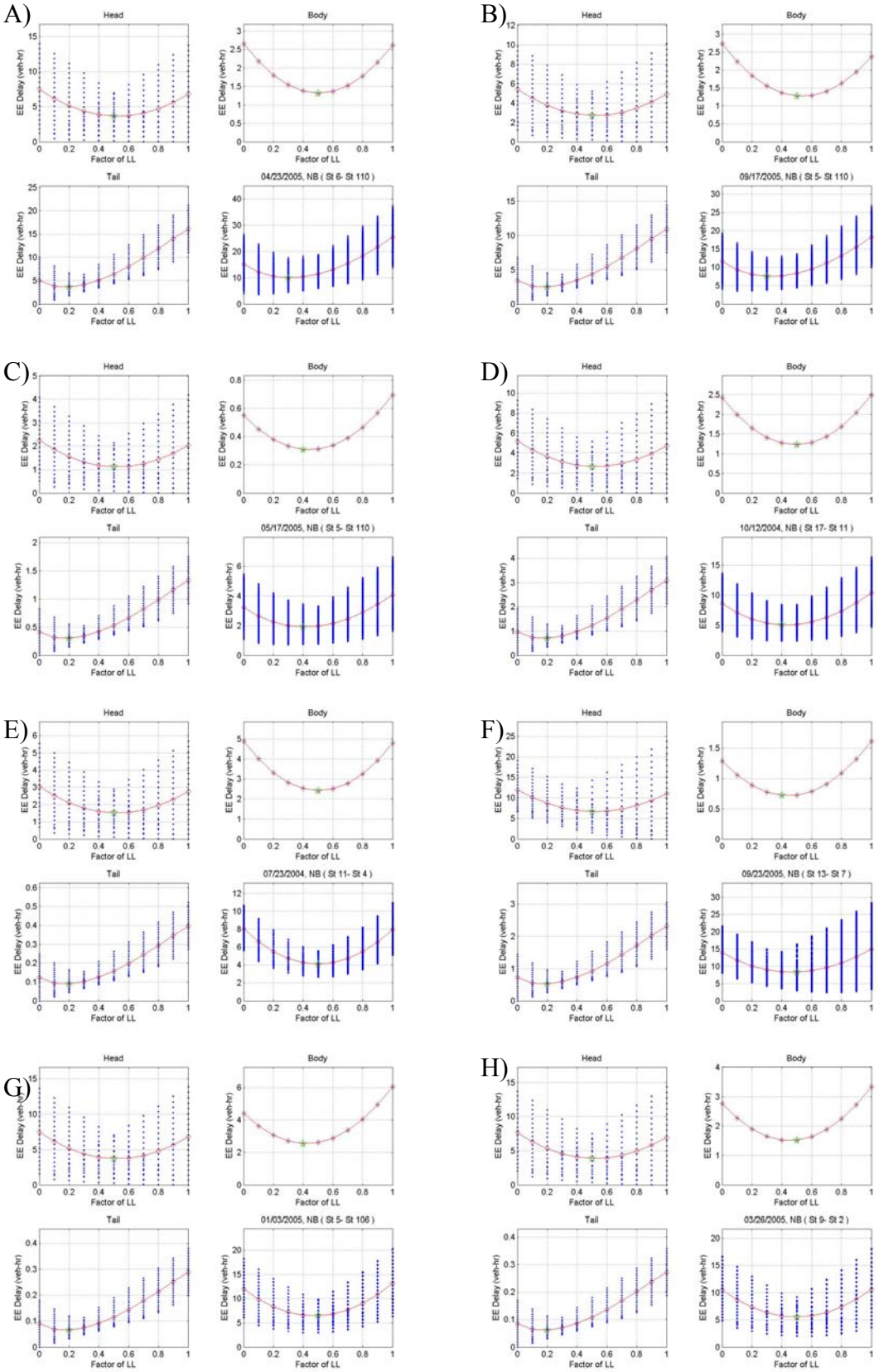


Figure E.2, Average estimation errors in each of three segments and in whole segments by link locations for present loop detector station spacing (veh-hr per lane).

Appendix F

Estimation errors by link locations for half loop detector station spacing

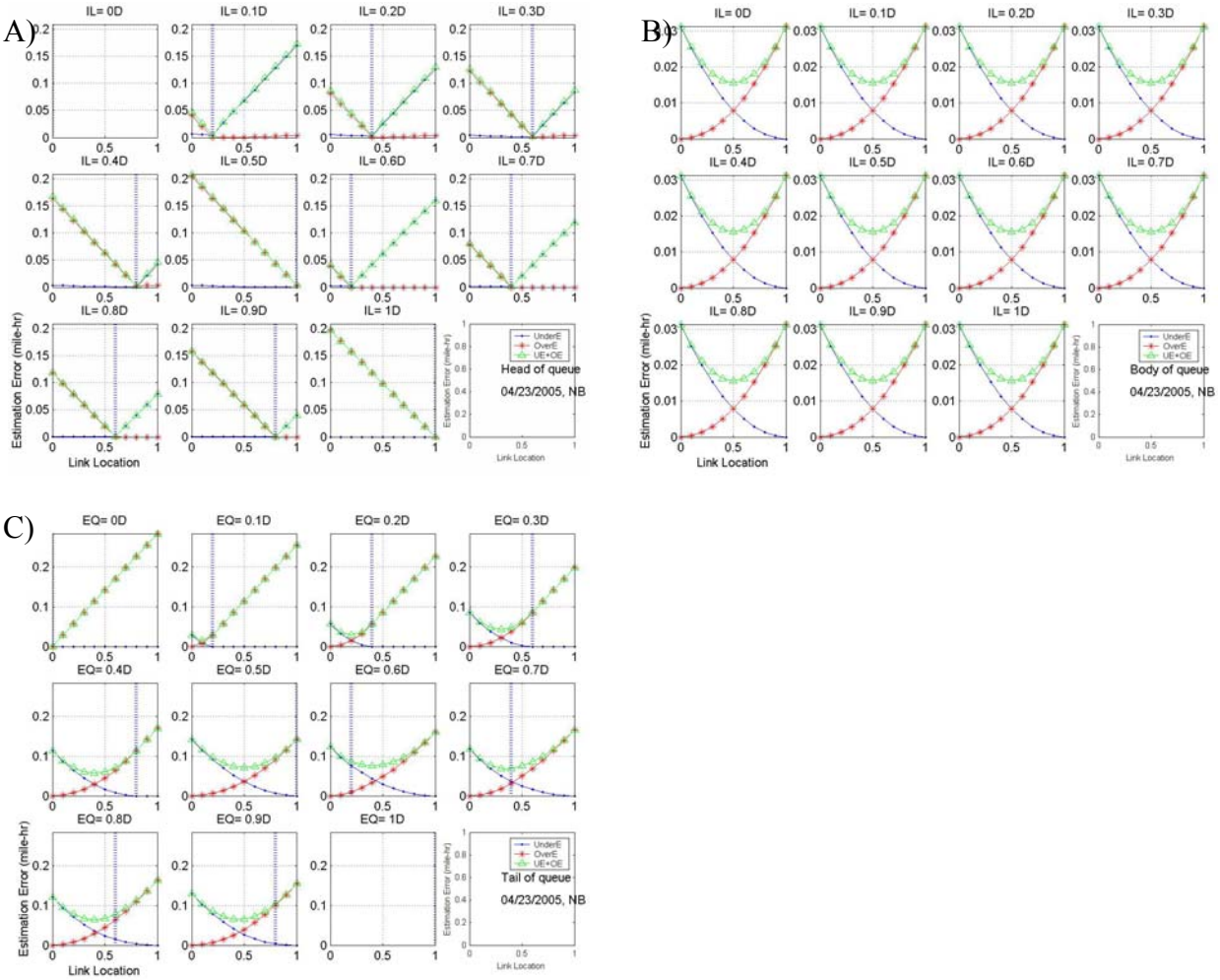
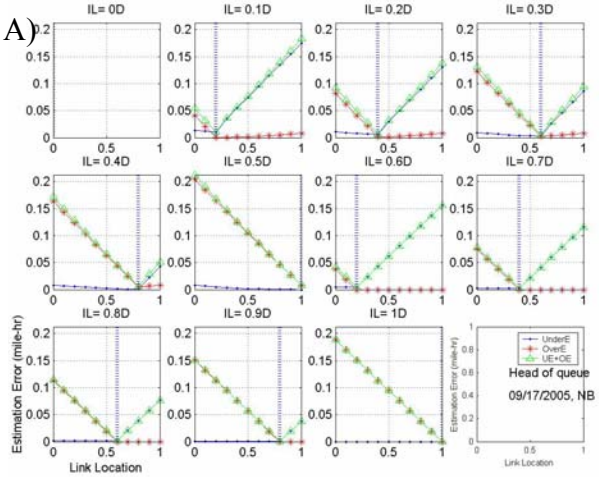
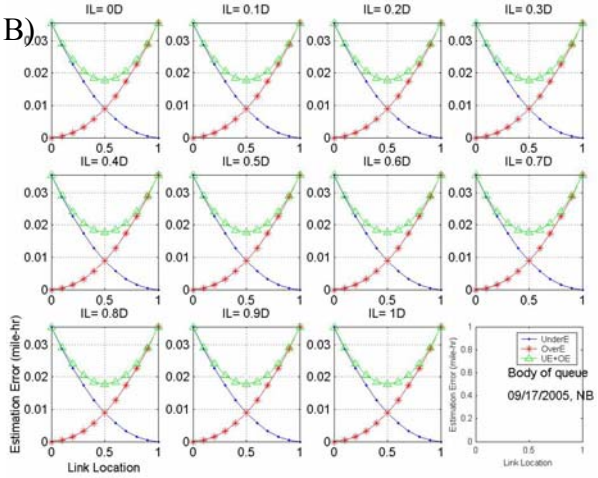


Figure F.1, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident A (04/23/2005)

A)



B)



C)

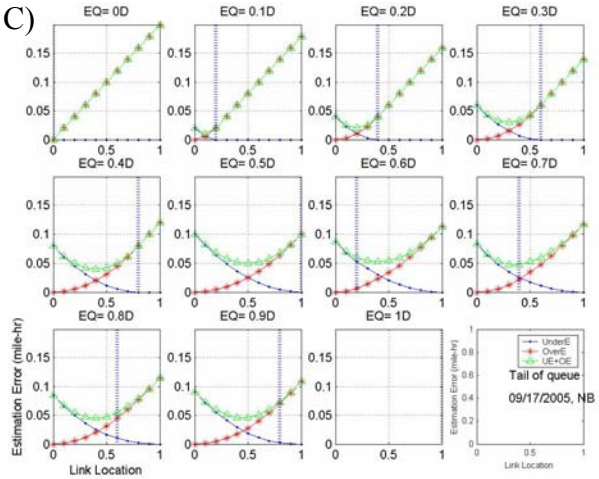
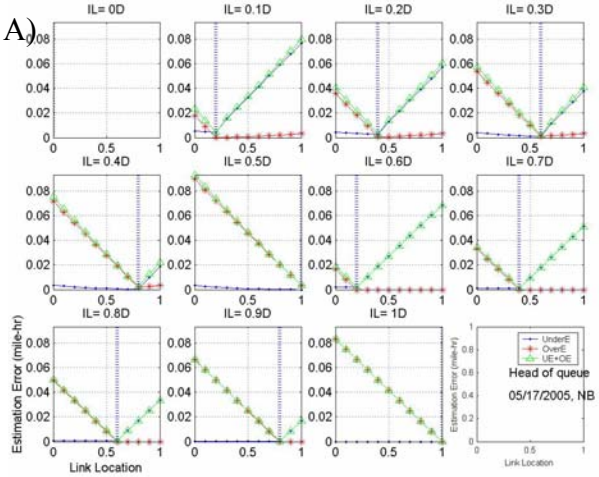
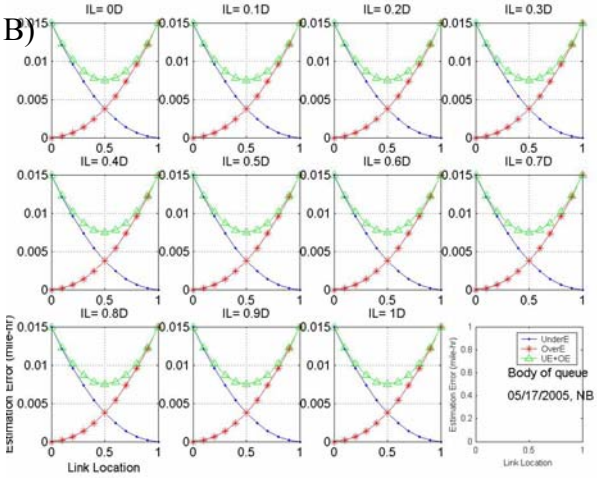


Figure F.2, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident B (09/17/2005)

A)



B)



C)

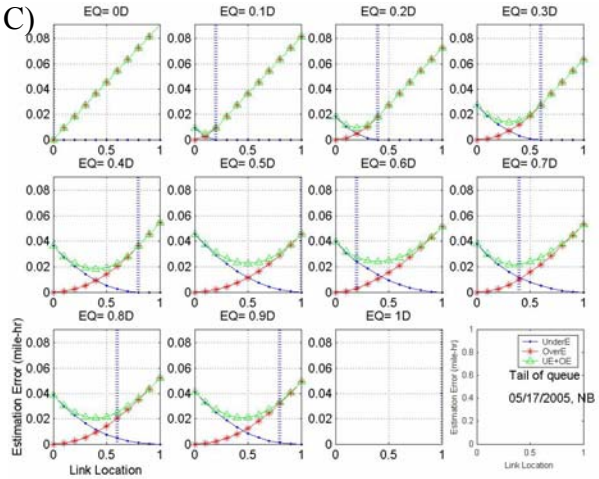


Figure F.3, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident C (05/17/2005)

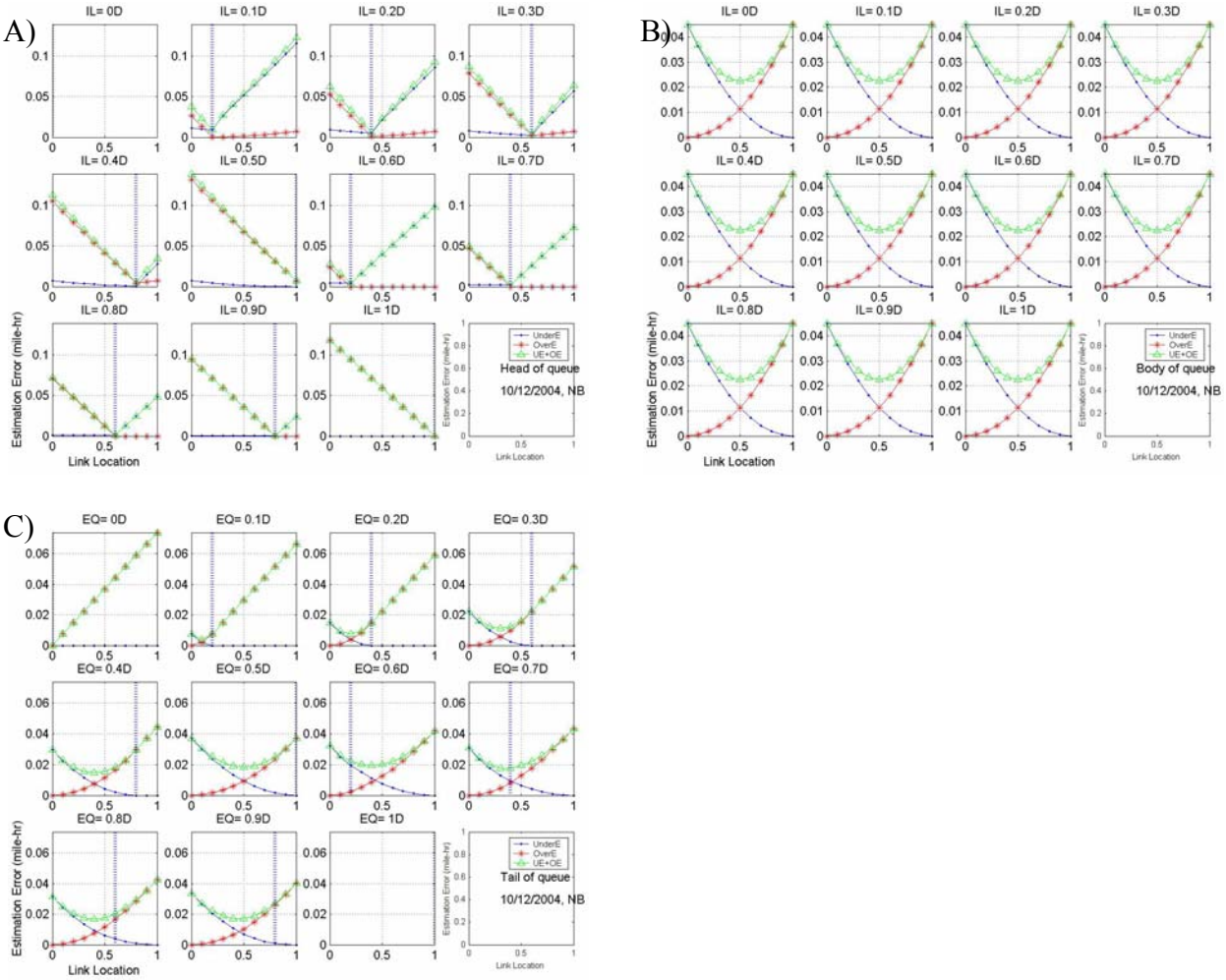


Figure F.4, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident D (10/12/2004)

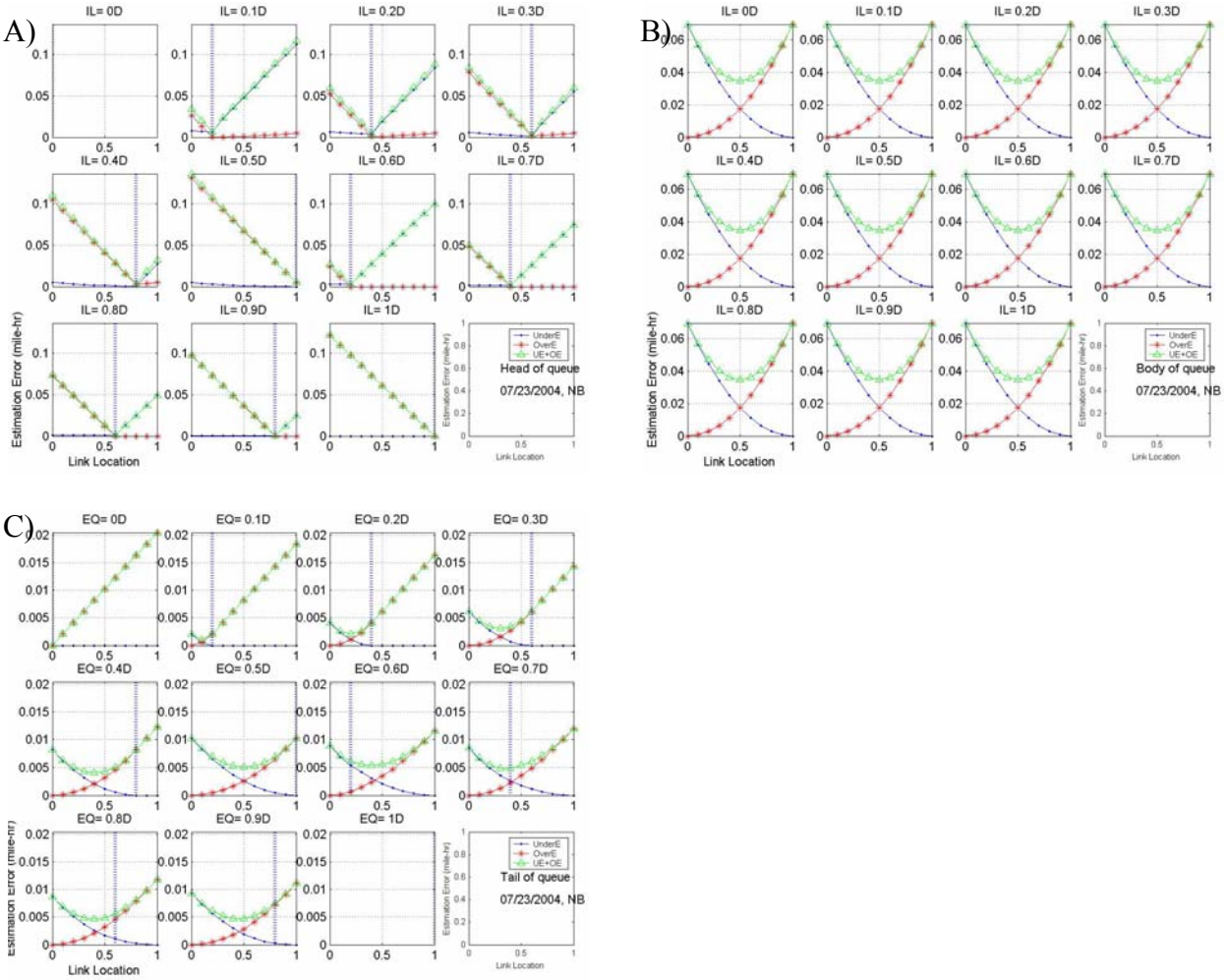
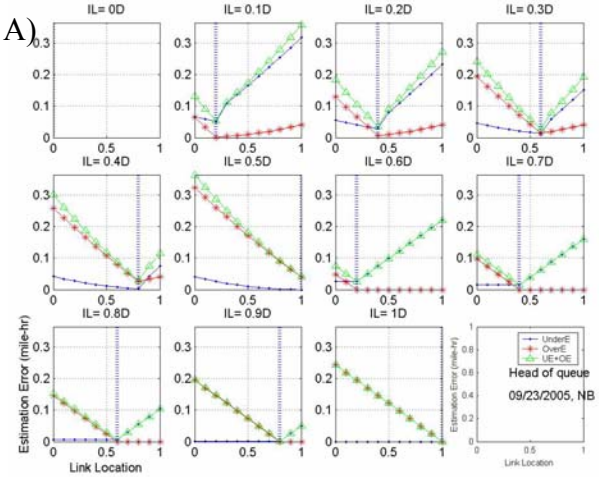
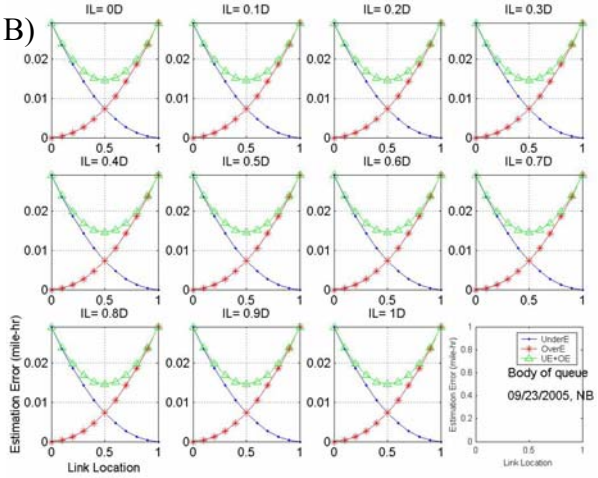


Figure F.5, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident E (07/23/2004)

A)



B)



C)

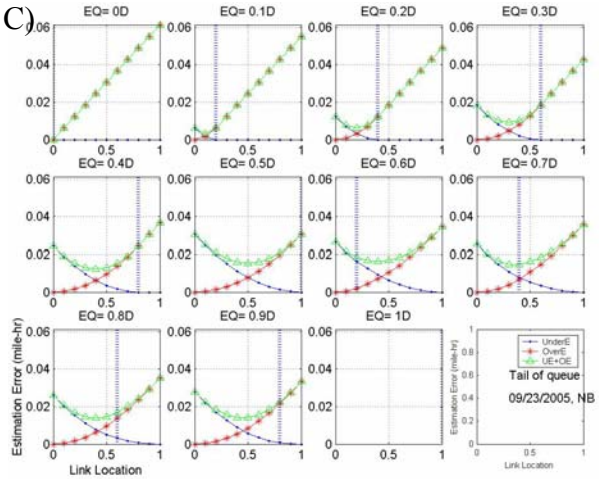


Figure F.6, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident F (09/23/2005)

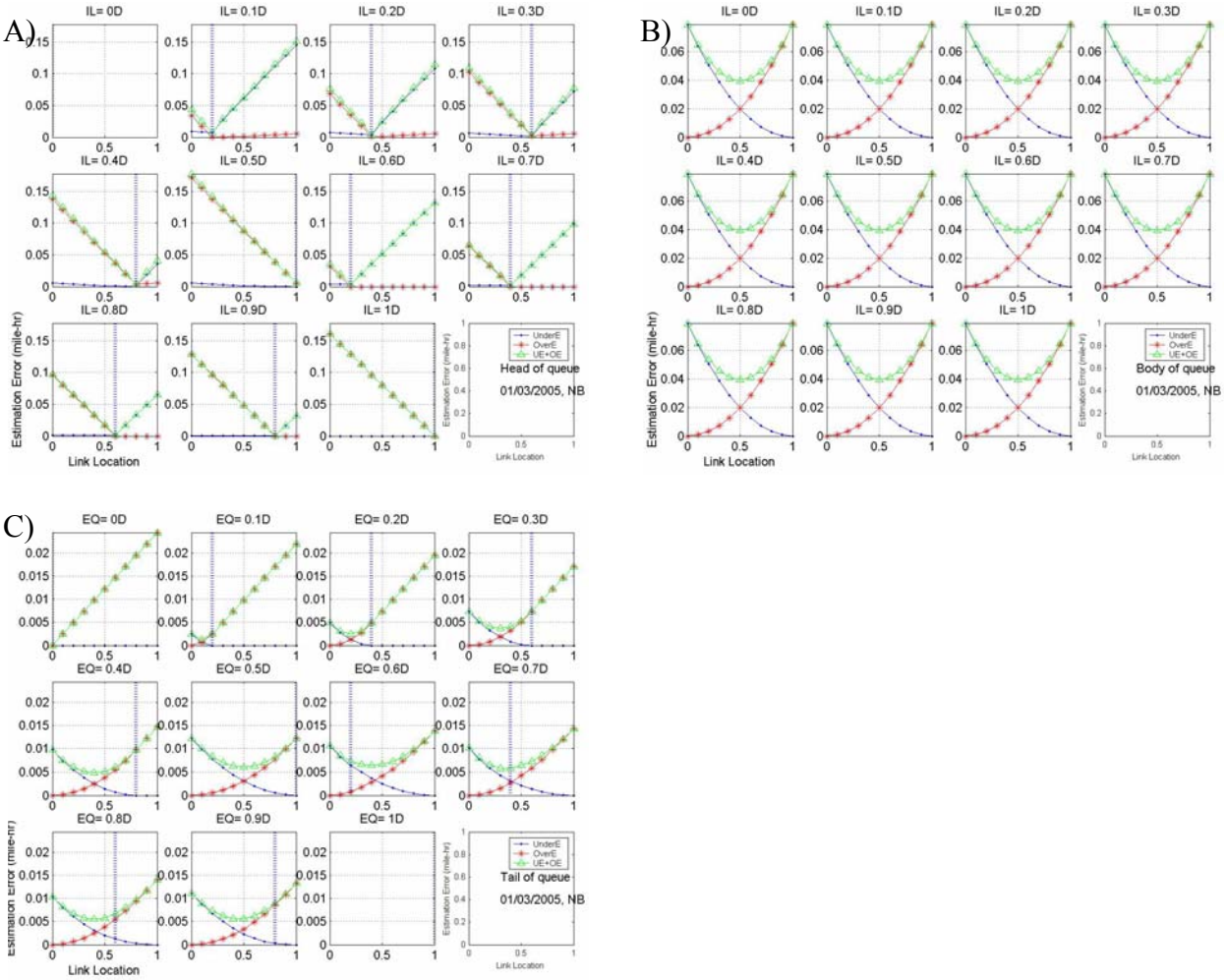


Figure F.7, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident G (01/03/2005)

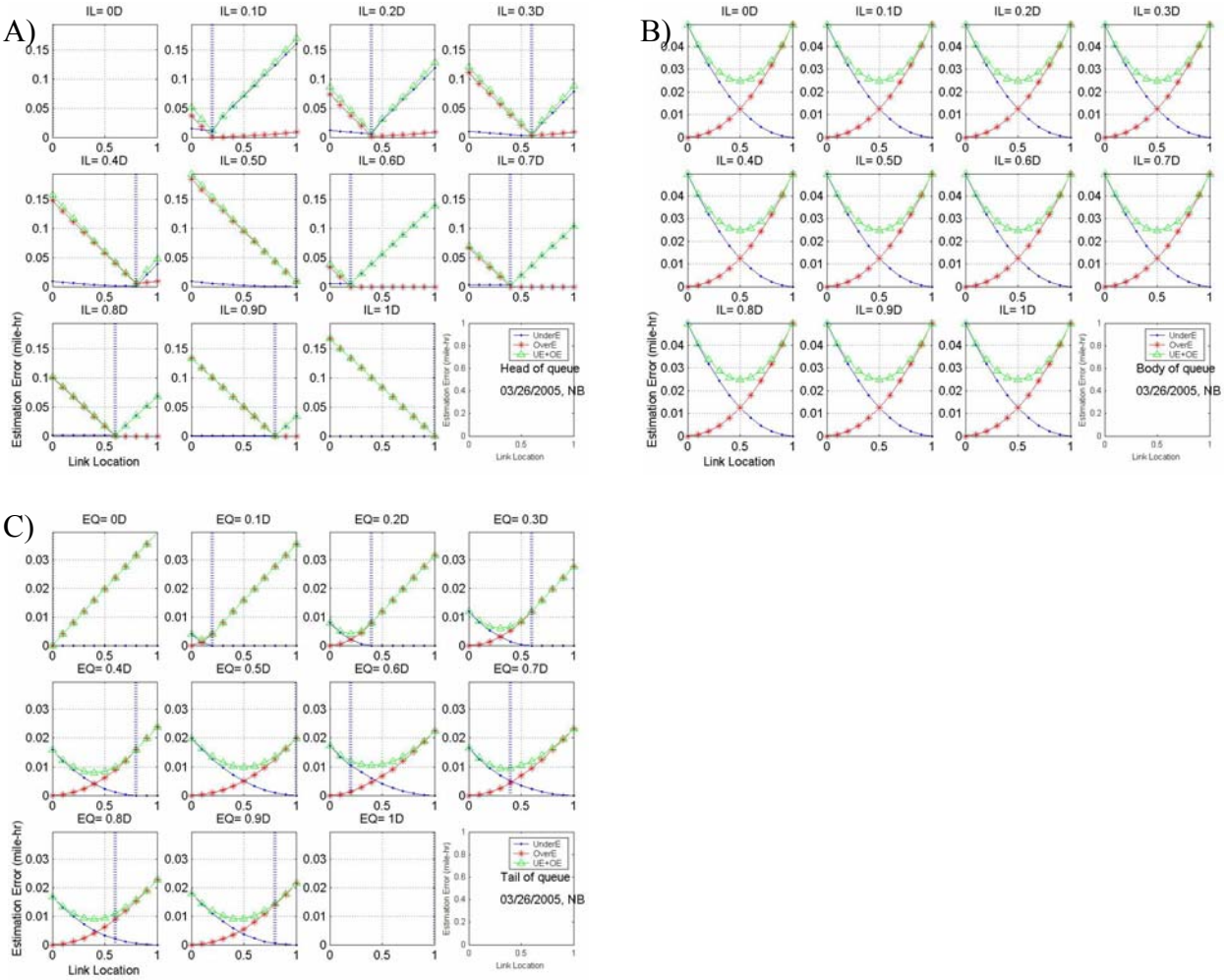


Figure F.8, Estimation error on three segments in half loop detector station spacing (mi-hr) Incident H (03/26/2005)

Appendix G

Average estimation errors by link location for half loop detector station spacing

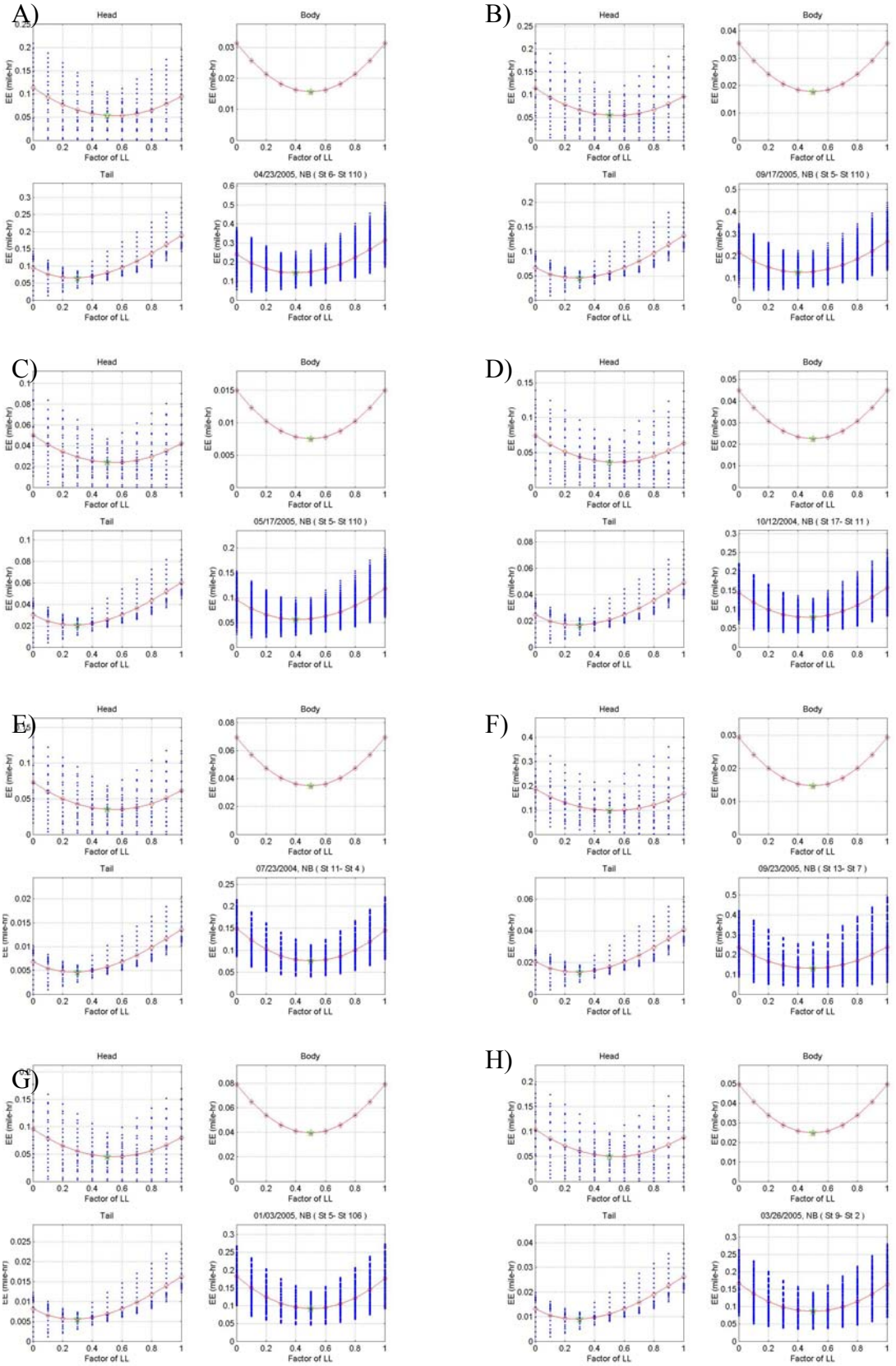


Figure G.1, Average estimation errors in each of three segments and in whole segments by link locations for half loop detector station spacing (mi-hr).

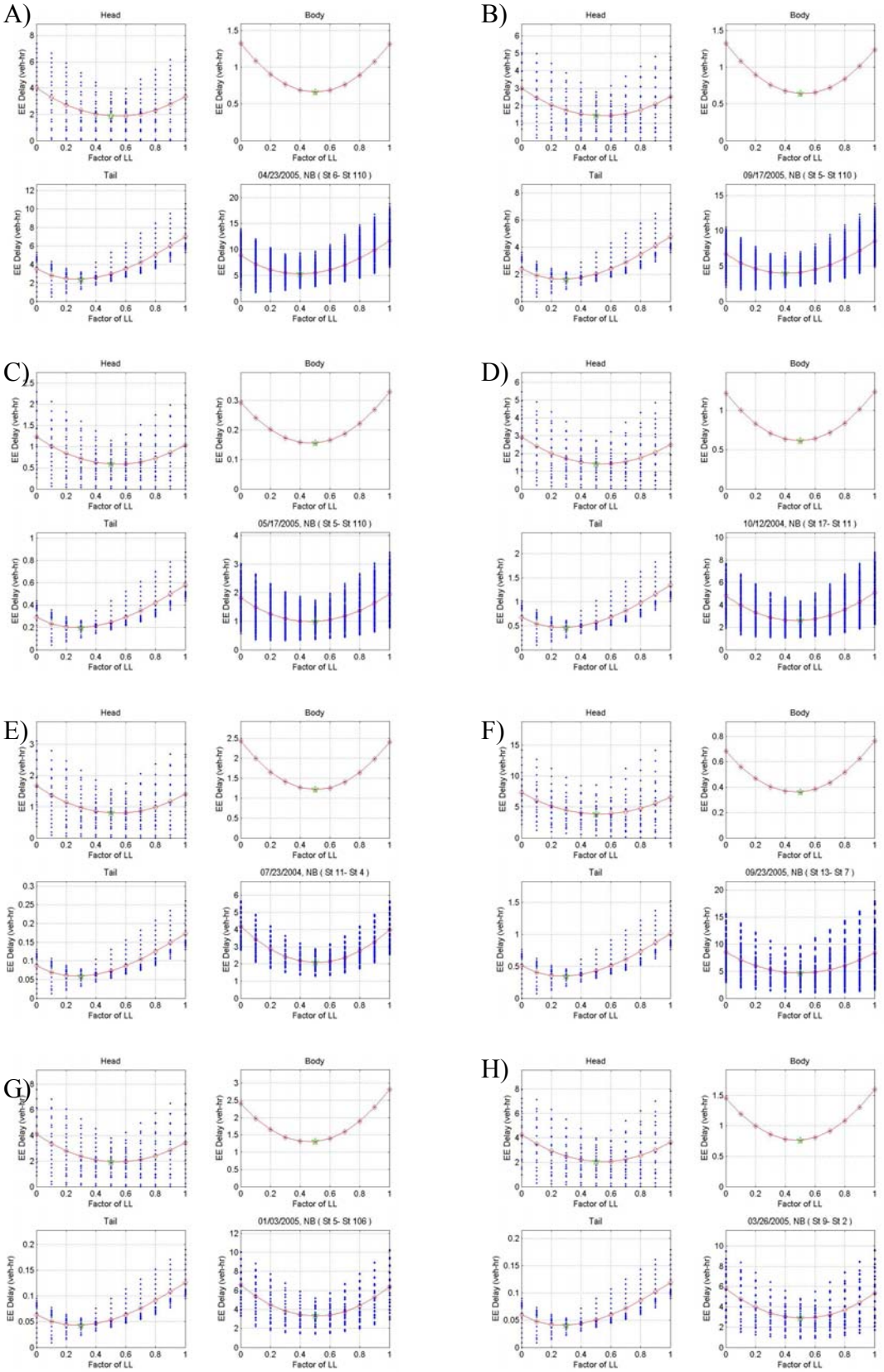


Figure G.2, Average estimation errors in each of three segments and in whole segments by link locations for half loop detector station spacing (veh-hr per lane).

Appendix H

Estimation errors by link locations for double loop detector station spacing

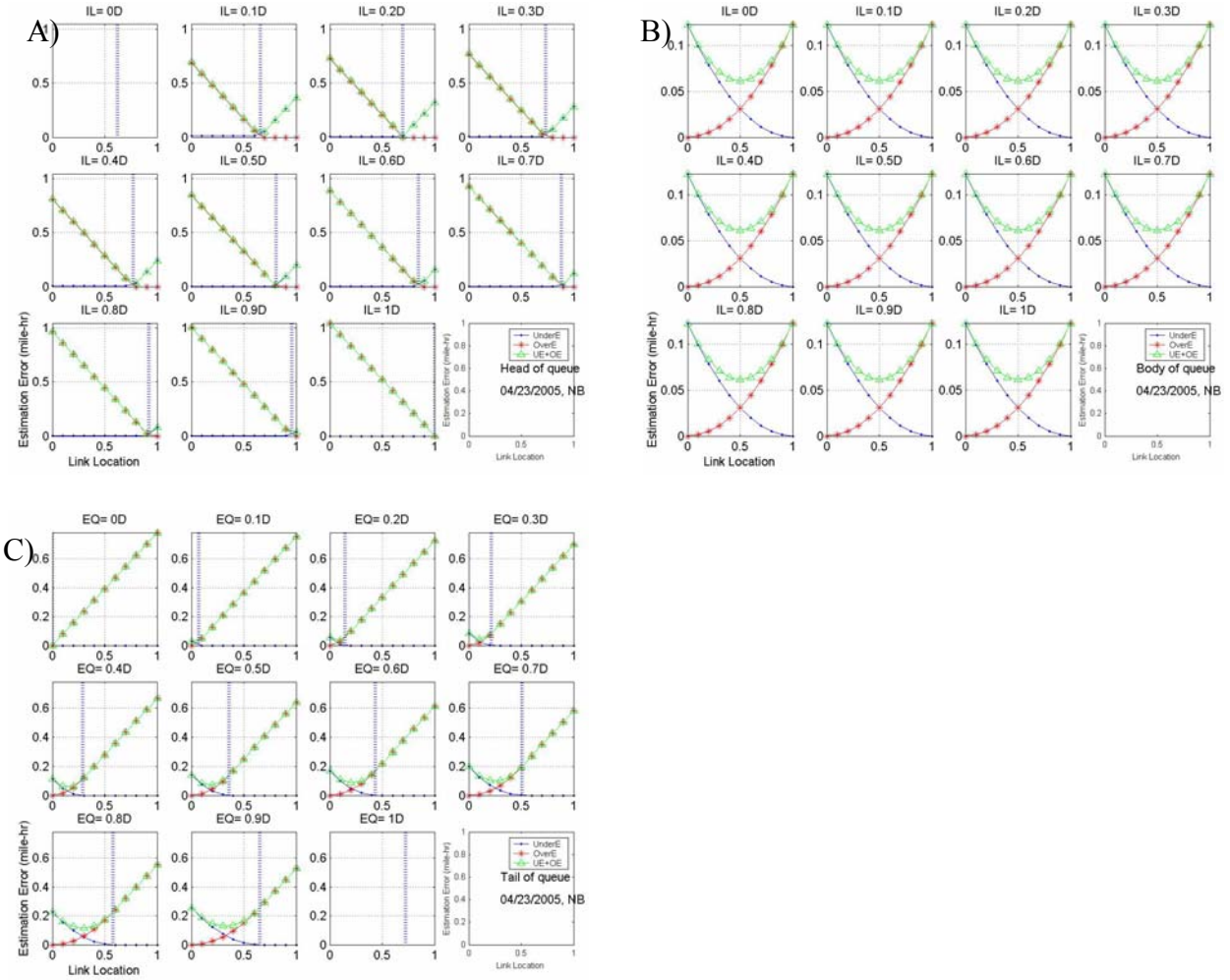
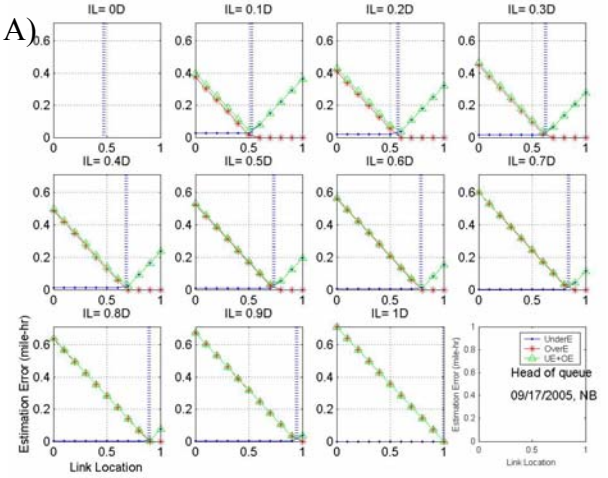
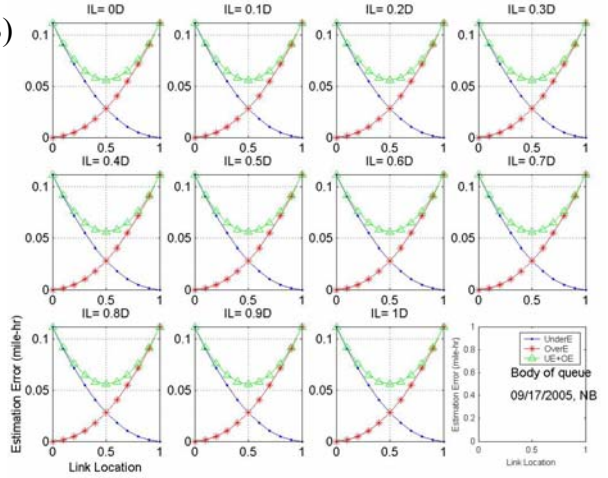


Figure H.1, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident A (04/23/2005)

A)



B)



C)

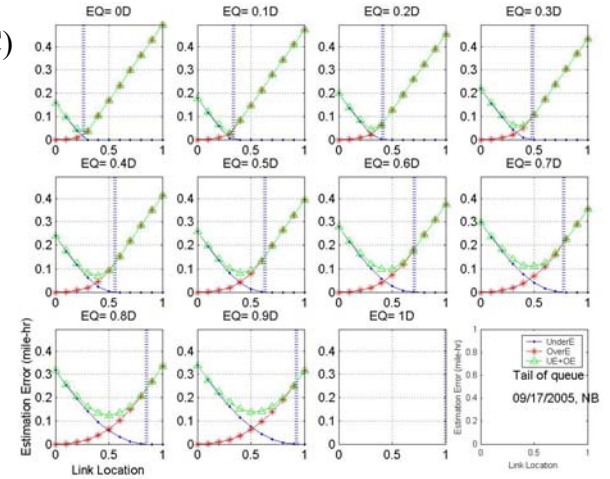
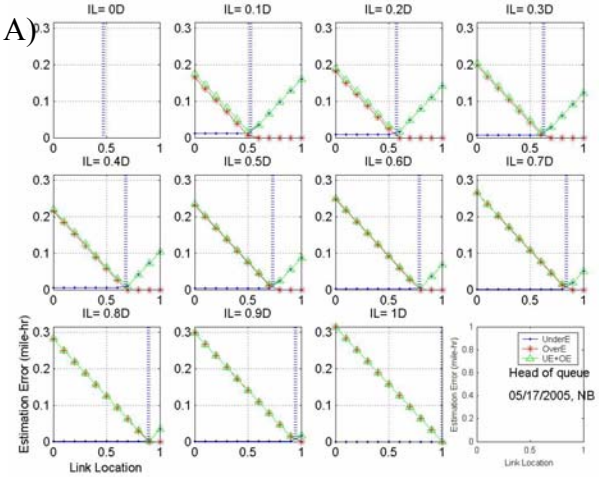
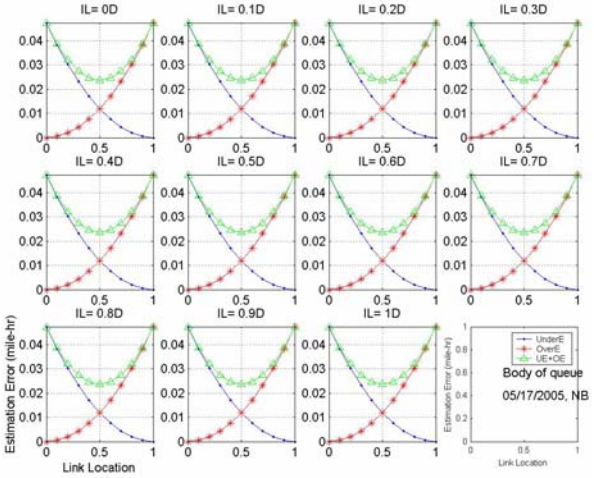


Figure H.2, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident B (09/17/2005)

A)



B)



C)

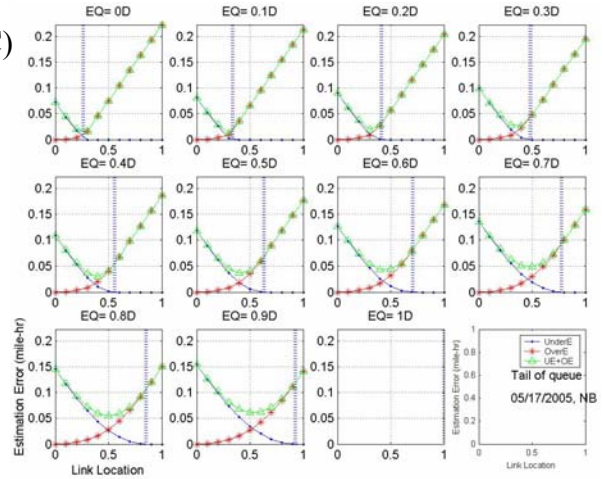
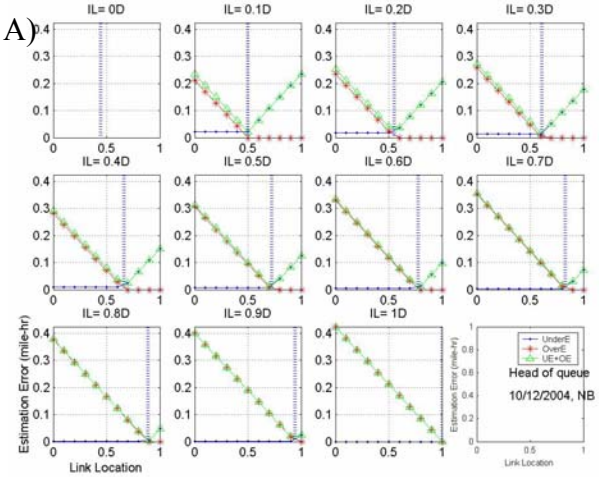
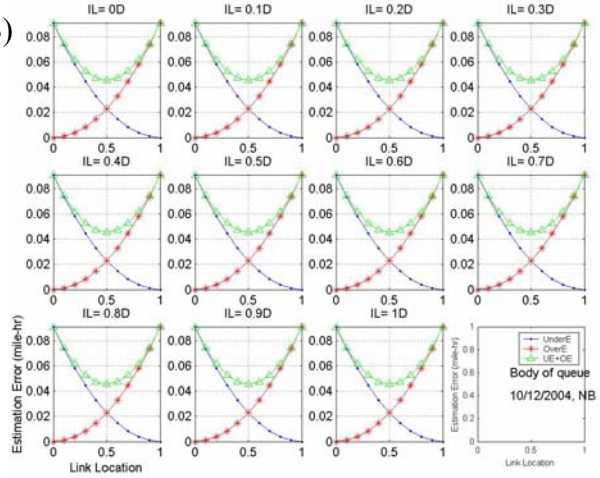


Figure H.3, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident C (05/17/2005)

A)



B)



C)

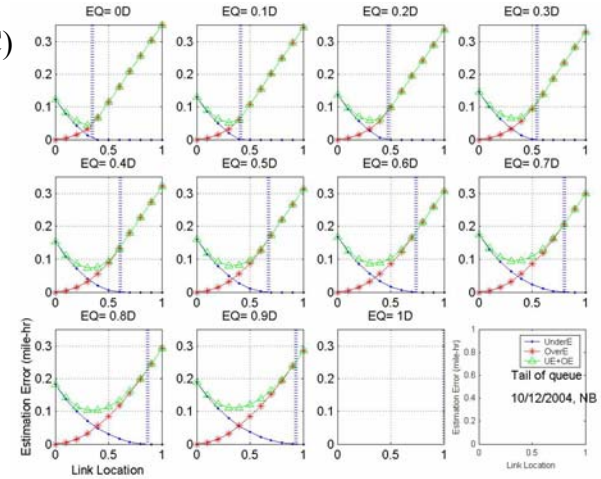


Figure H.4, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident D (10/12/2004)

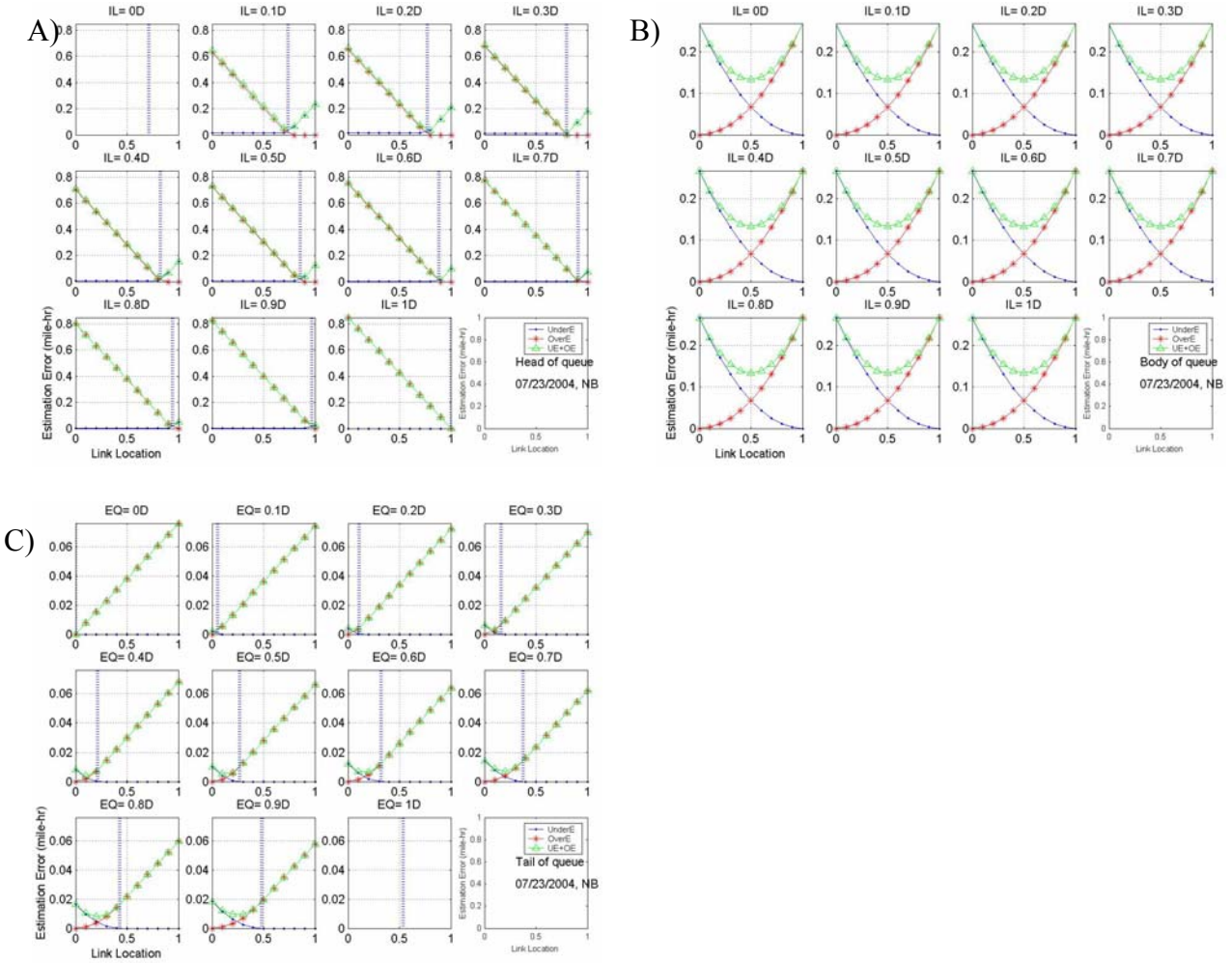


Figure H.5, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident E (07/23/2004)

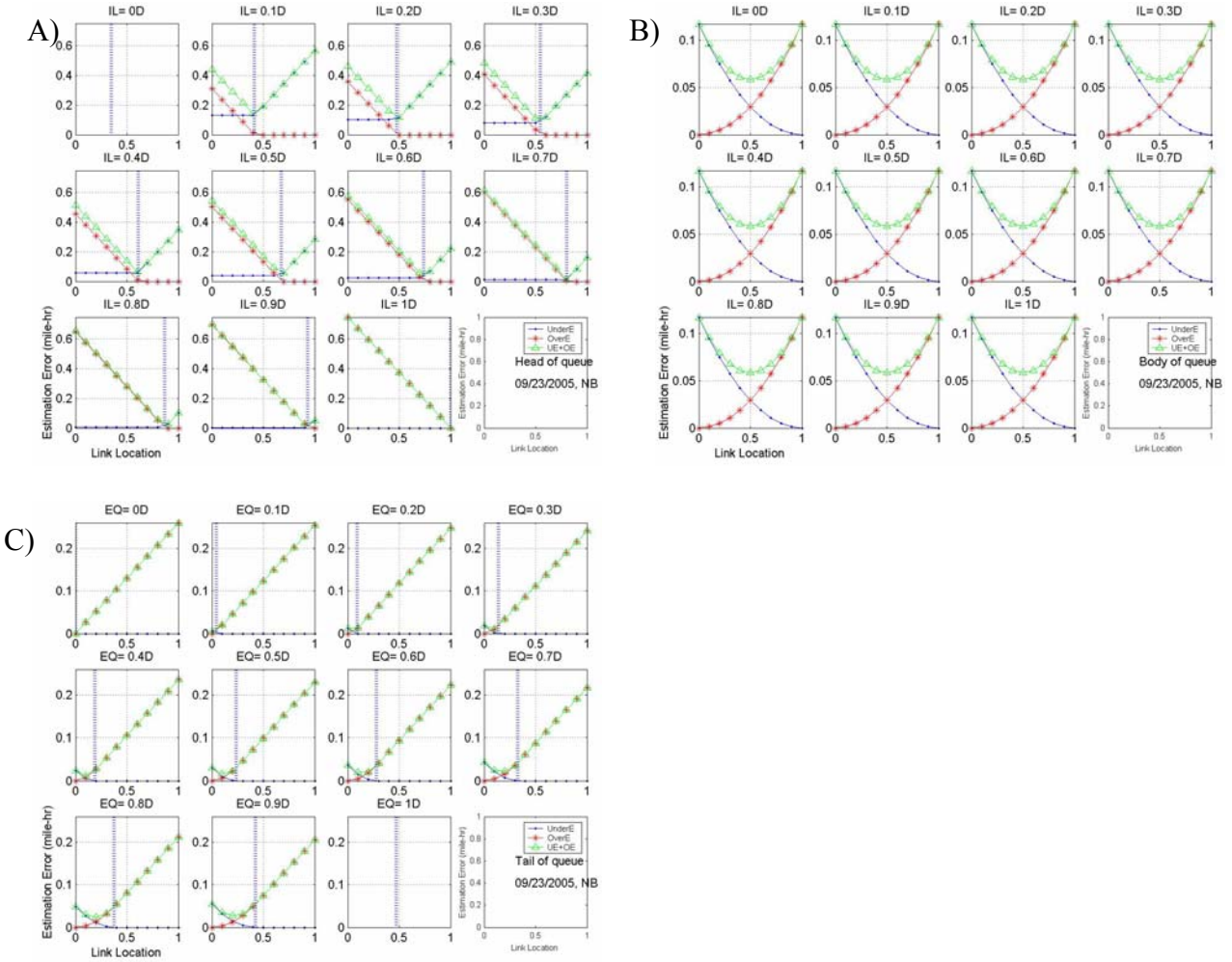
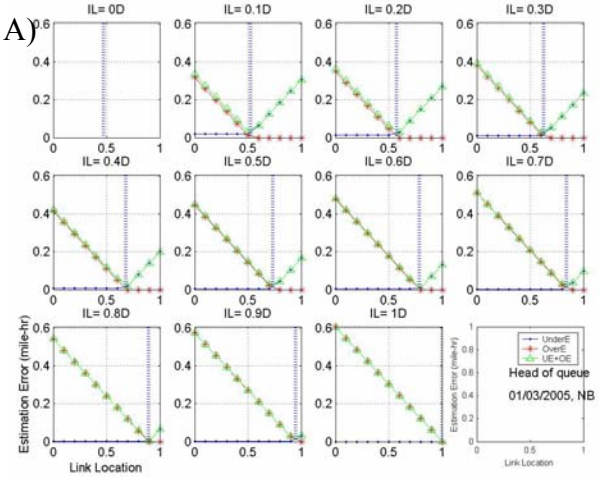
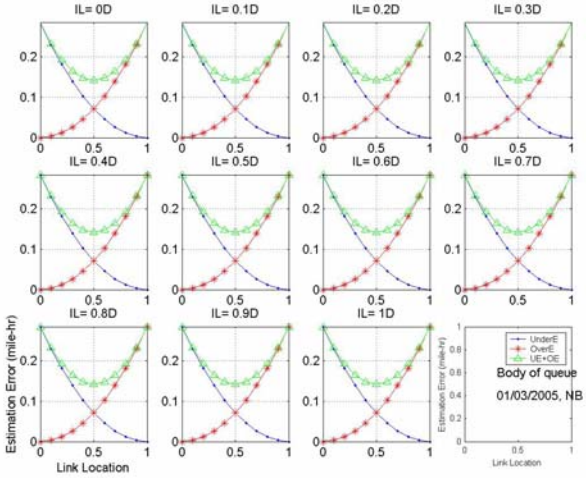


Figure H.6, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident F (09/23/2005)

A)



B)



C)

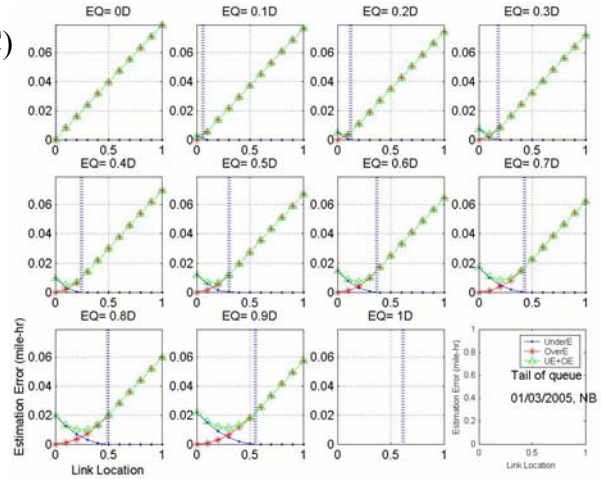


Figure H.7, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident G (01/03/2005)

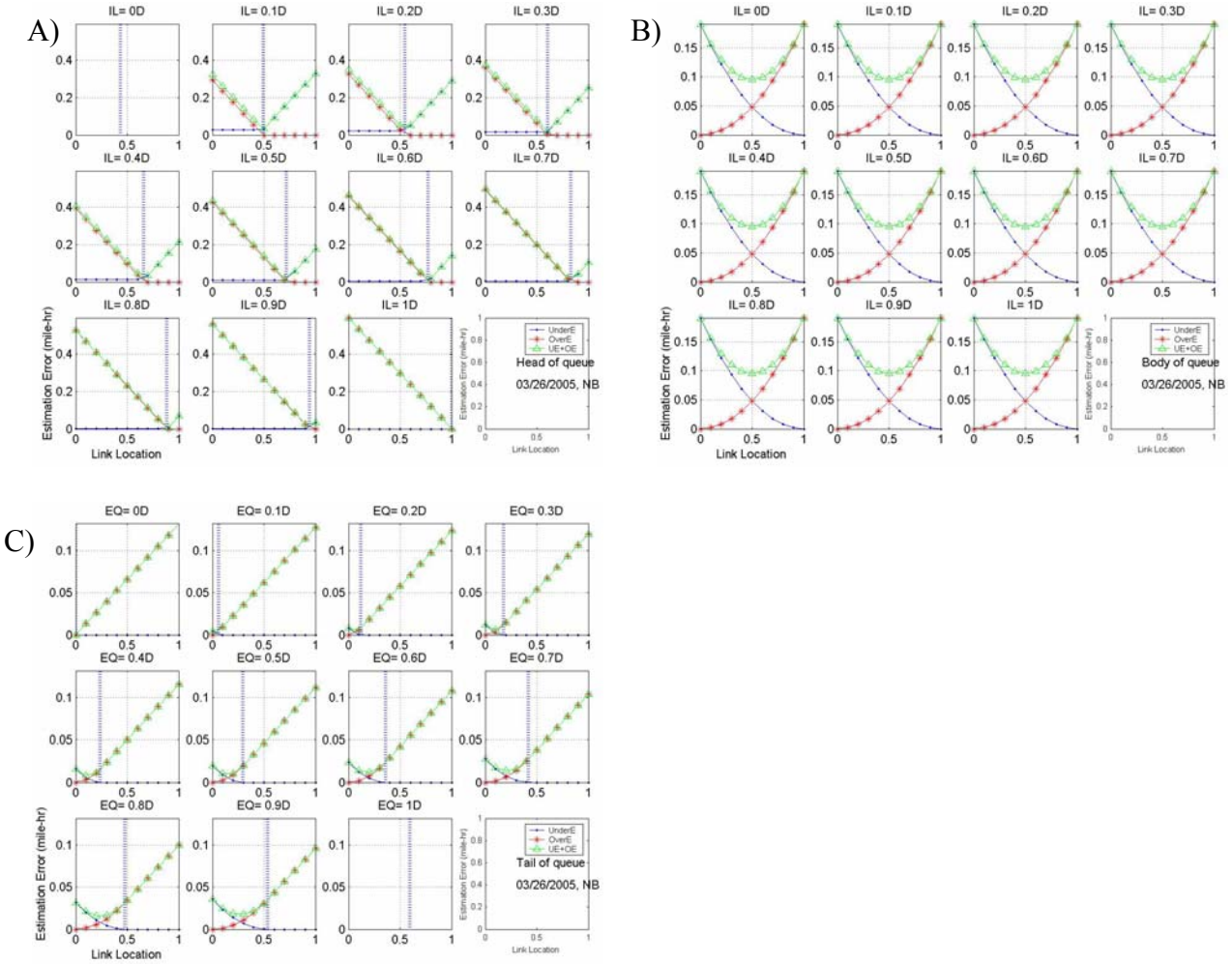


Figure H.8, Estimation error on three segments in double loop detector station spacing, NHB, (mi-hr) Incident H (03/26/2005)

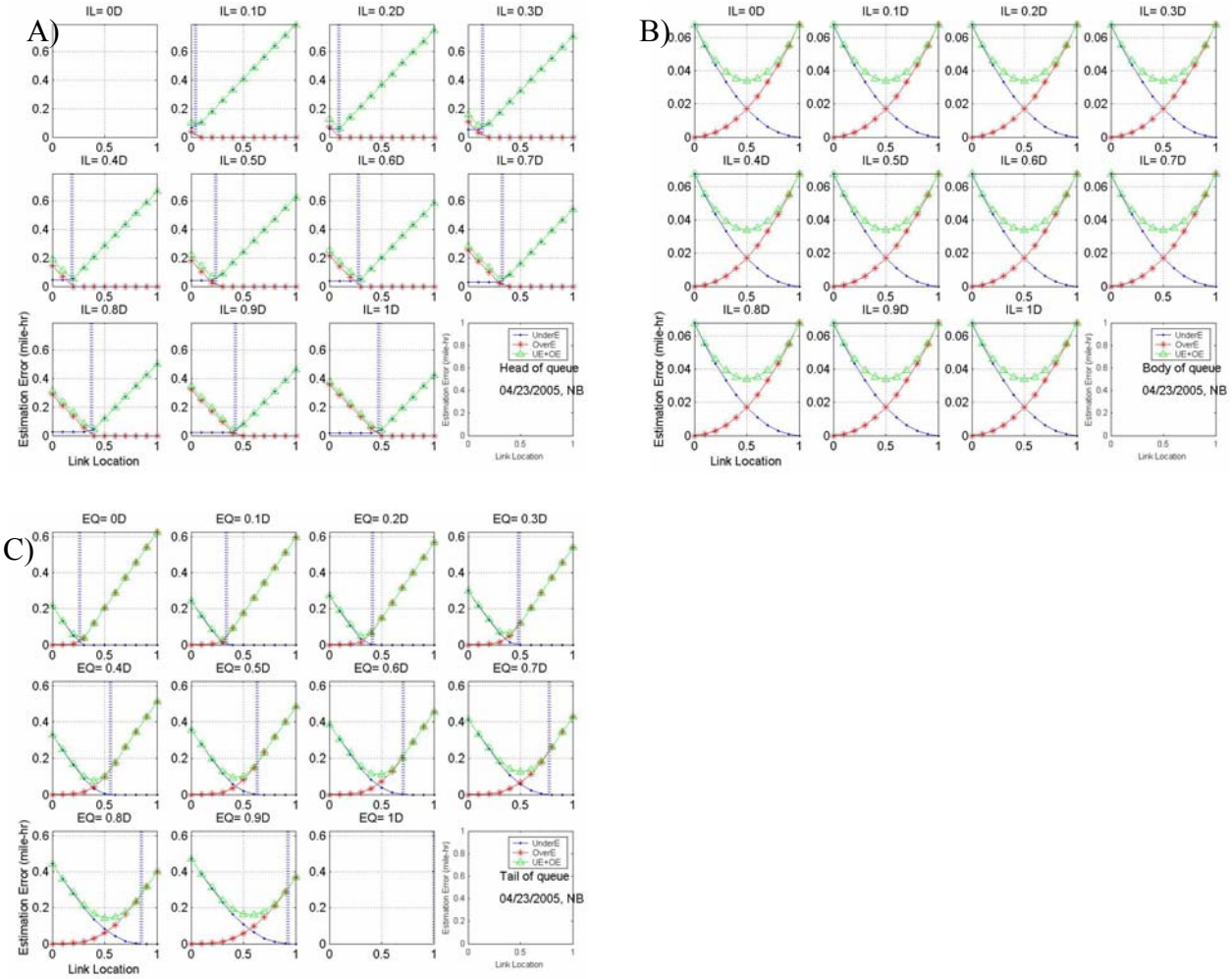
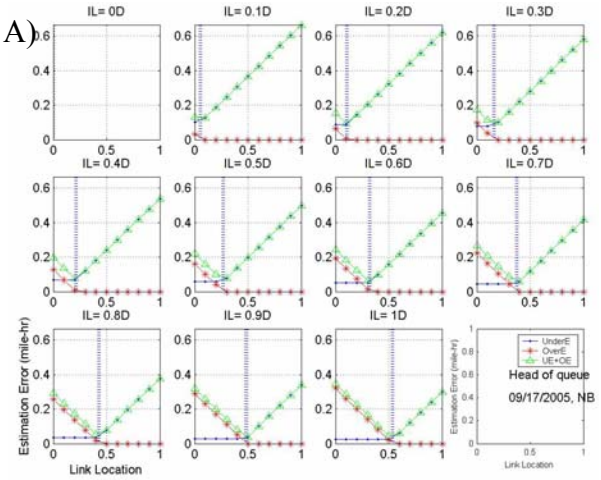
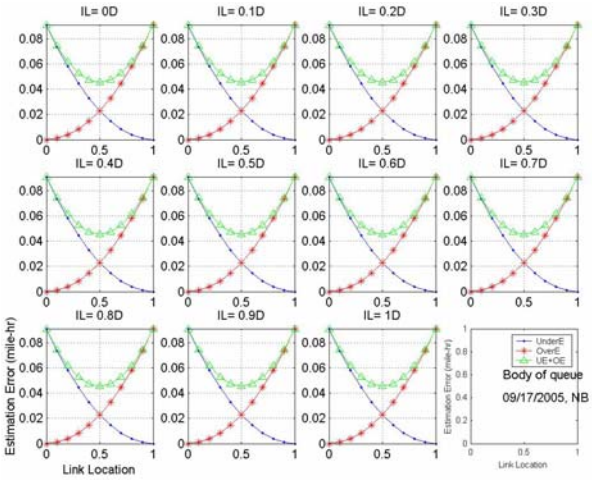


Figure H.9, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident A (04/23/2005)

A)



B)



C)

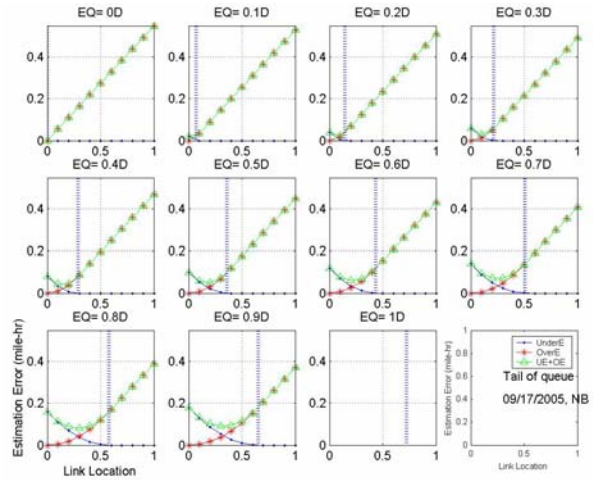
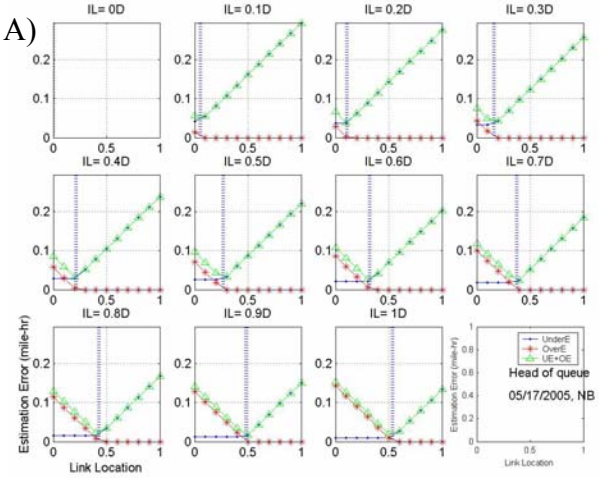
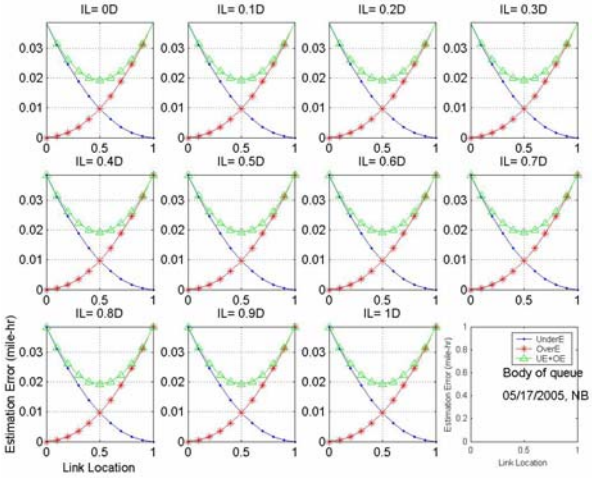


Figure H.10, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident B (09/17/2005)

A)



B)



C)

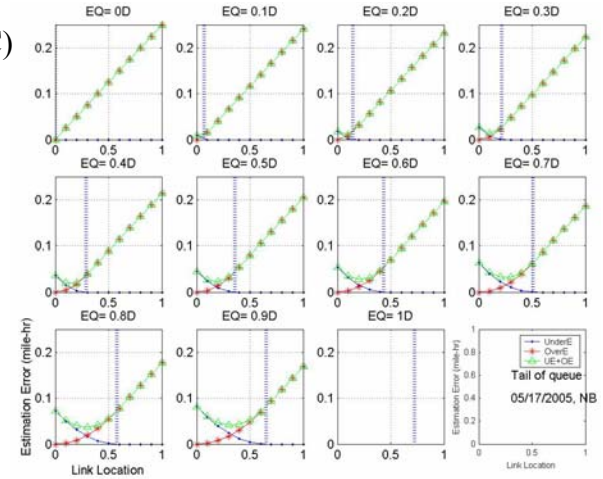


Figure H.11, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident C (05/17/2005)

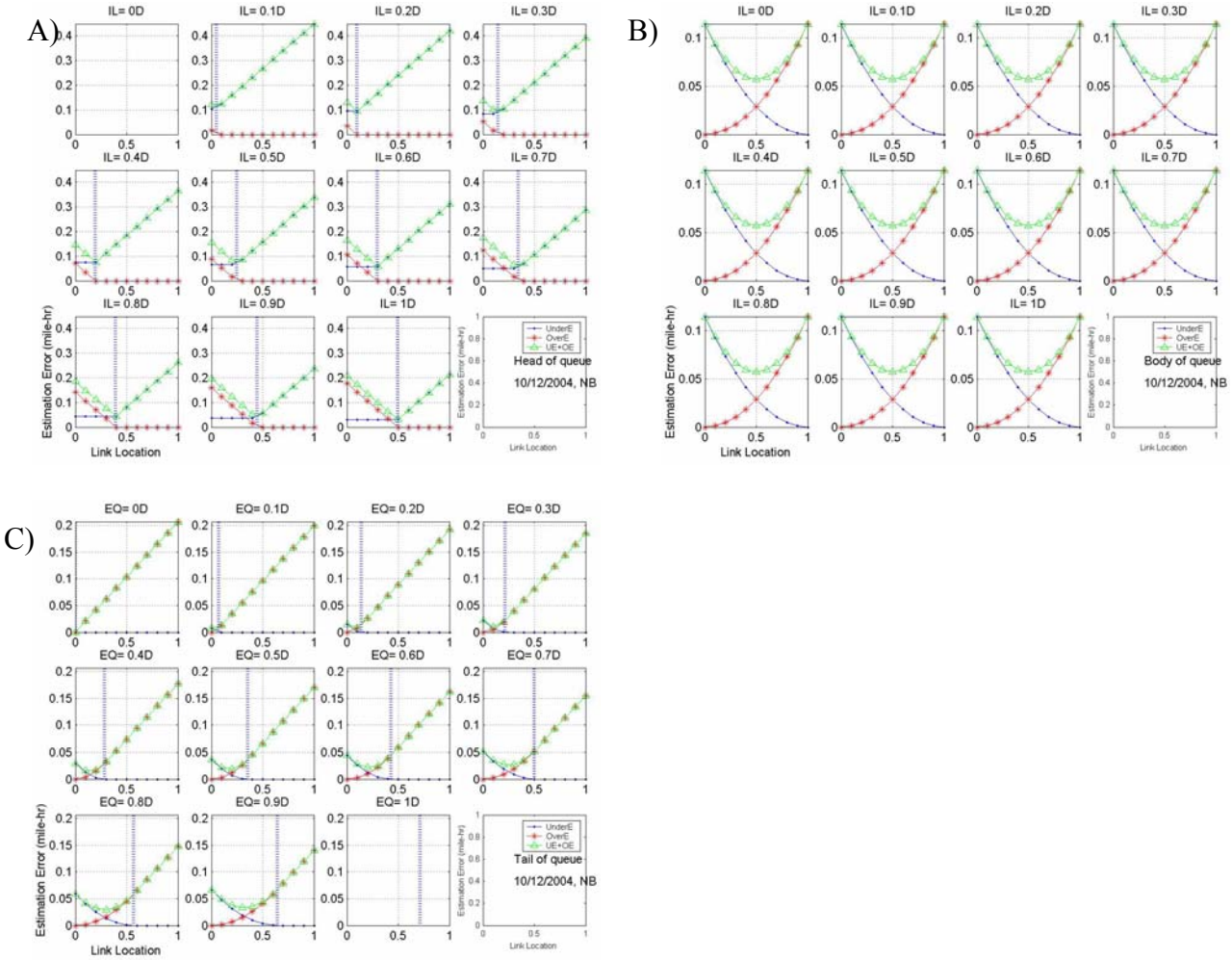


Figure H.12, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident D (10/12/2004)

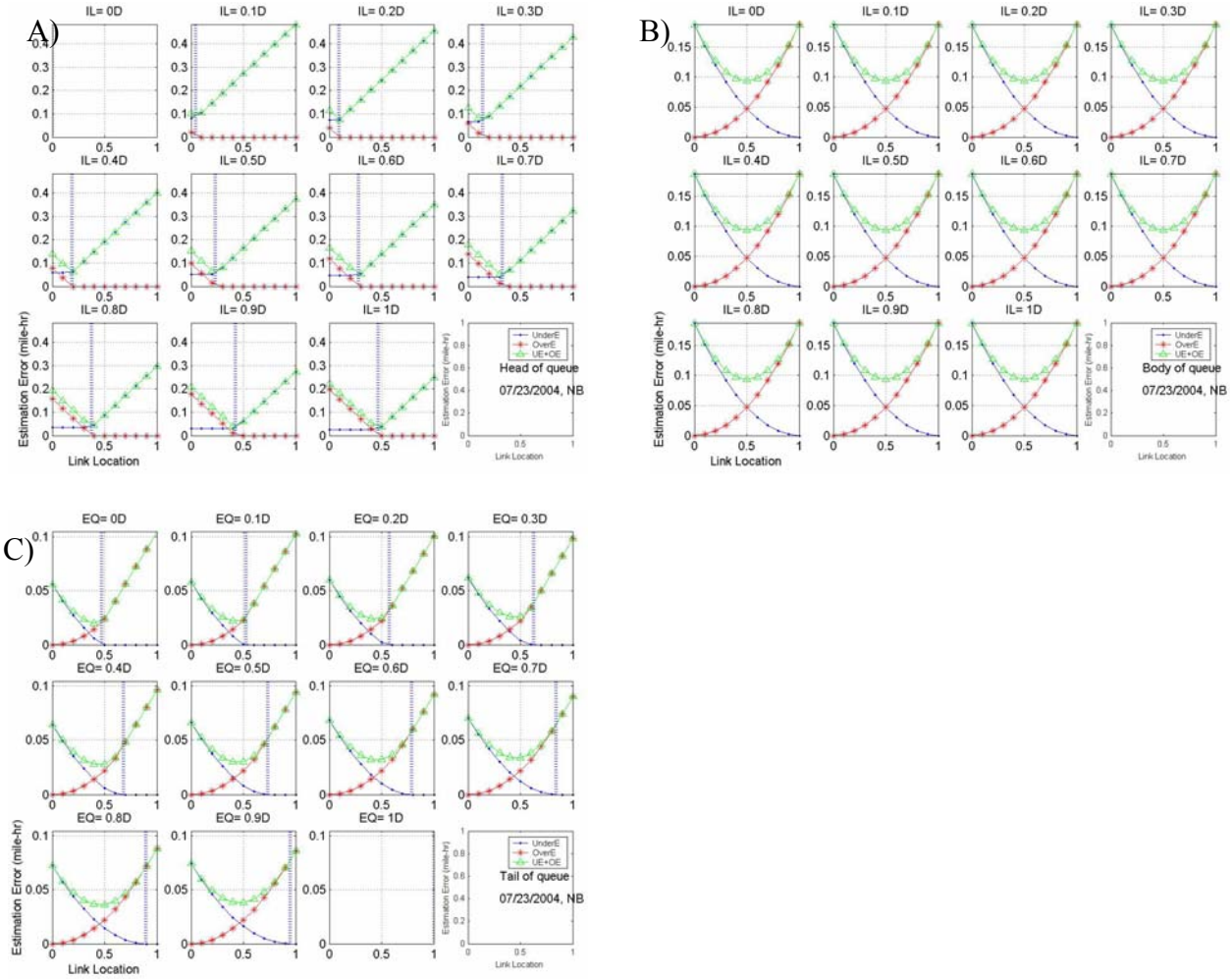


Figure H.13, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident E (07/23/2004)

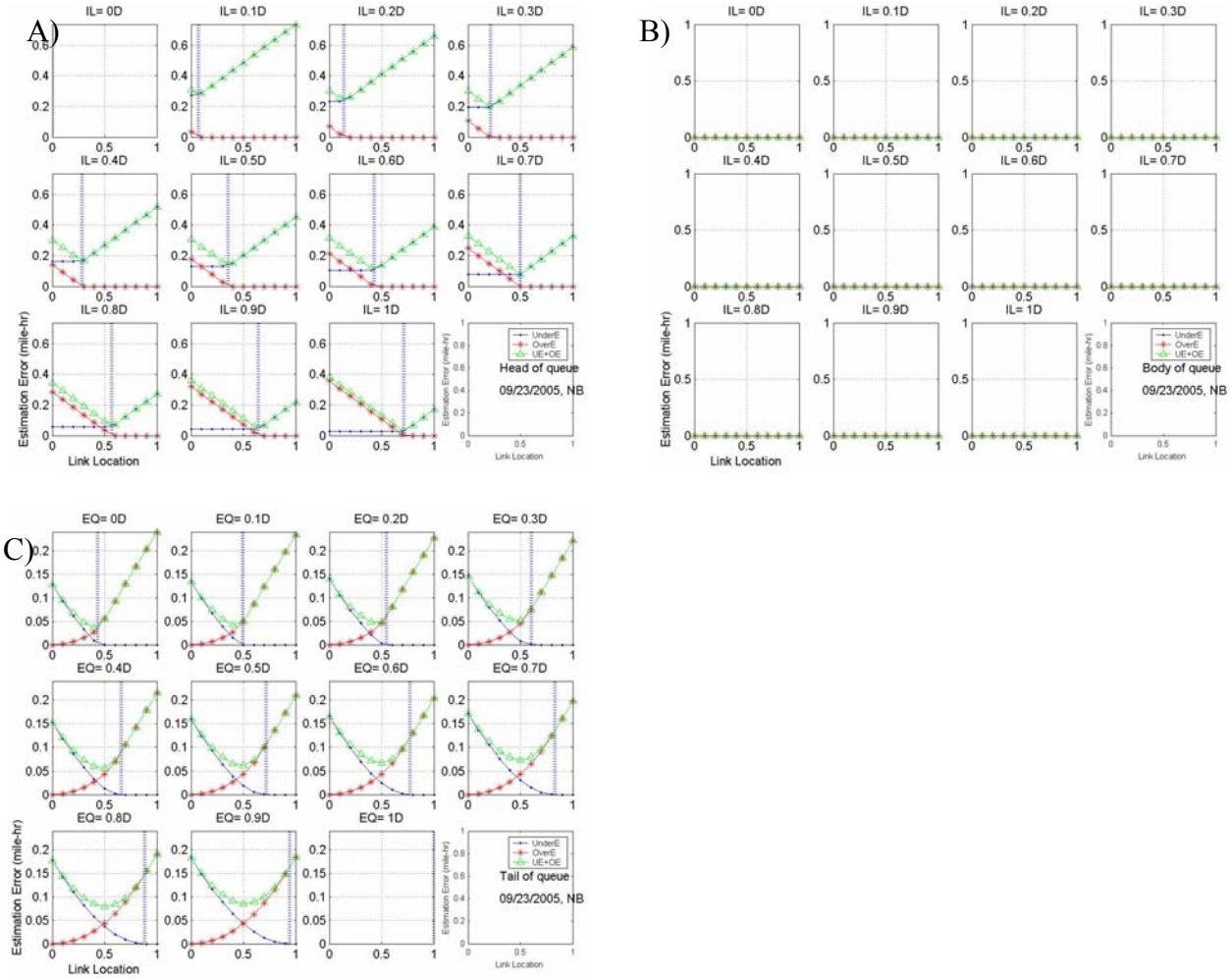
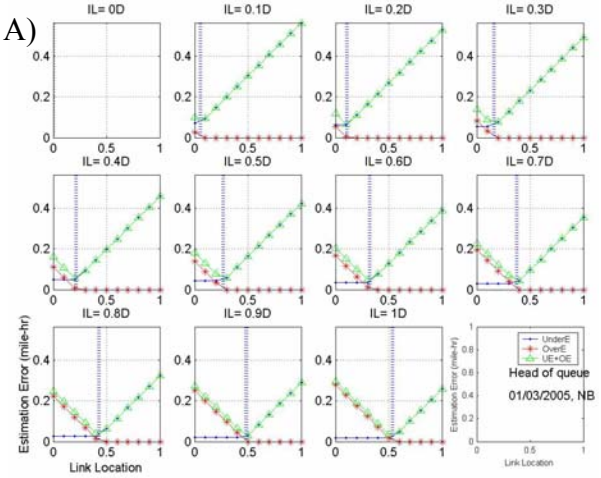
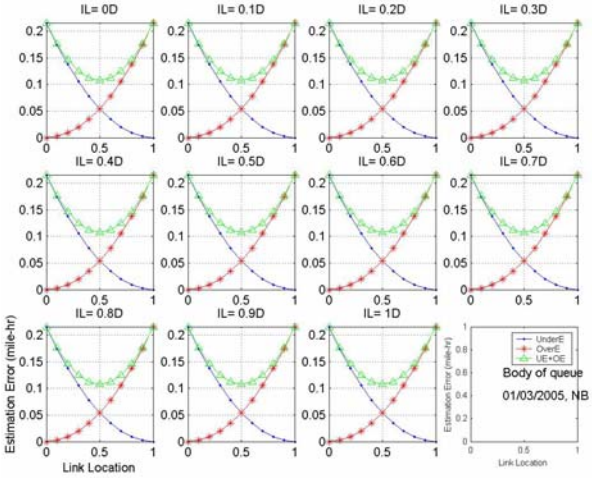


Figure H.14, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident F (09/23/2005)

A)



B)



C)

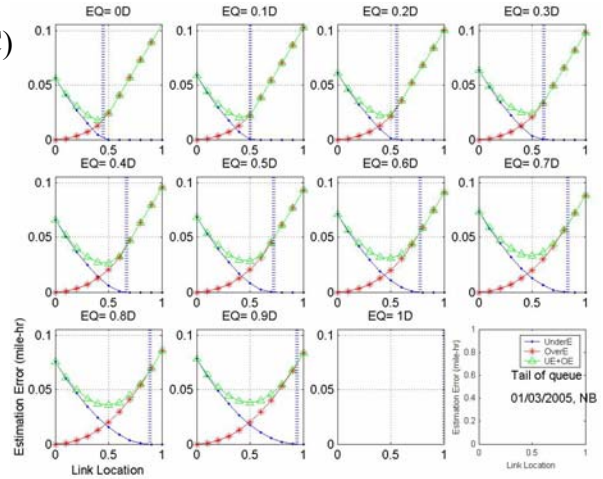


Figure H.15, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident G (01/03/2005)

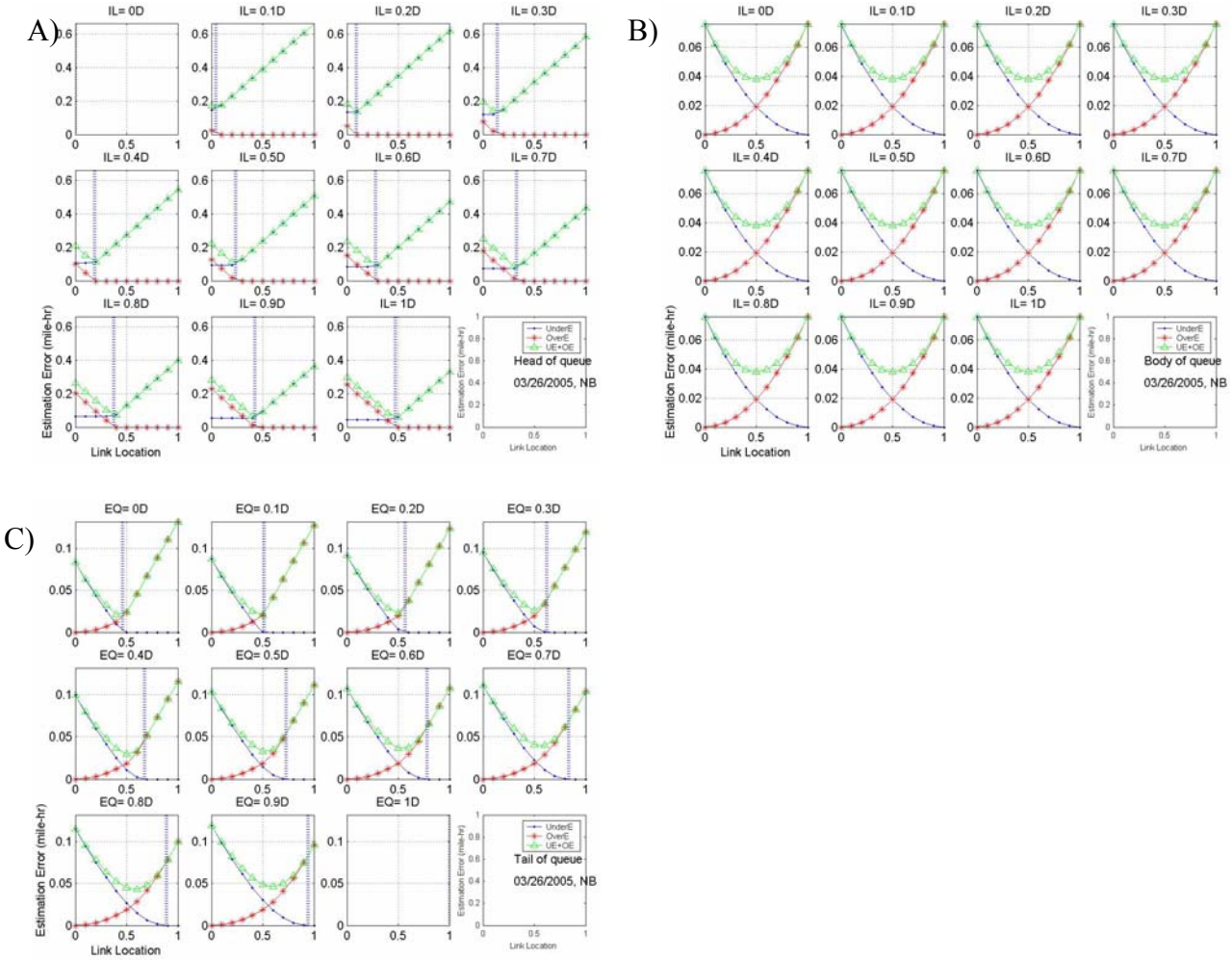


Figure H.16, Estimation error on three segments in double loop detector station spacing, PHB, (mi-hr) Incident H (03/26/2005)

Appendix I

Average estimation errors by link location for double loop detector station spacing

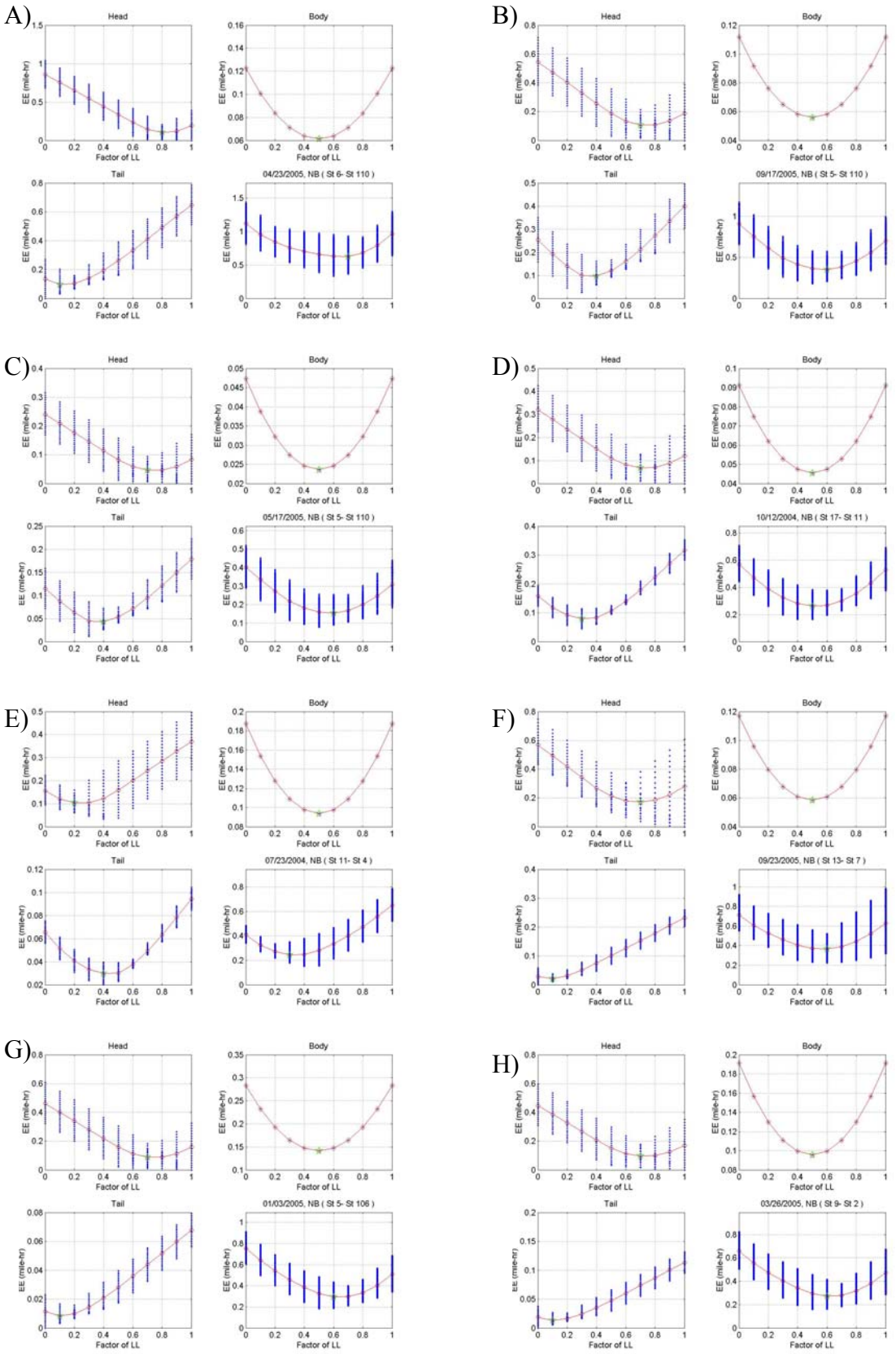


Figure I.1, Average estimation errors in each of three segments and in whole segments by link locations for double loop detector station spacing, NHB, (mi-hr).

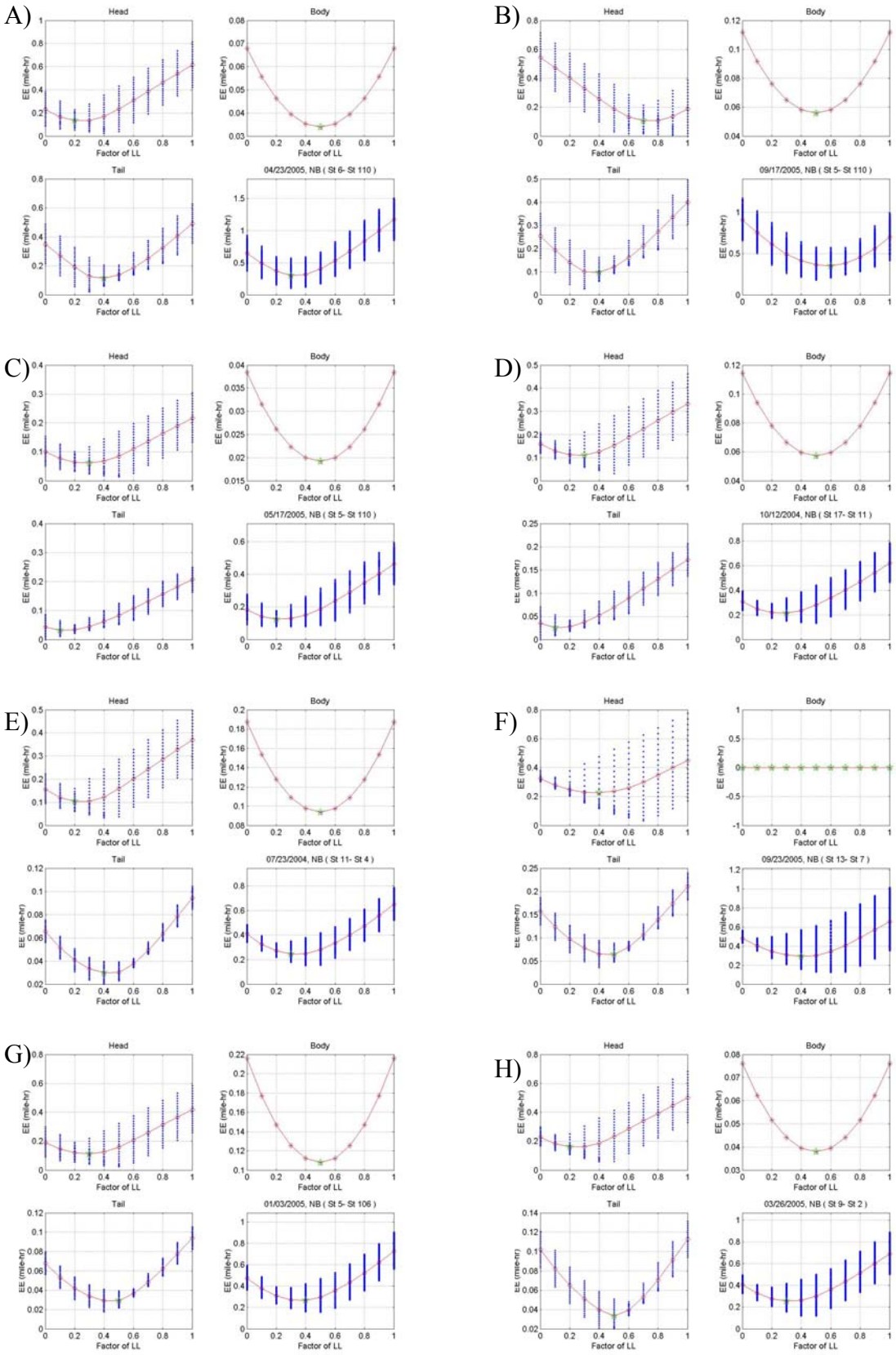


Figure I.2, Average estimation errors in each of three segments and in whole segments by link locations for double loop detector station spacing, PHB, (mi-hr).

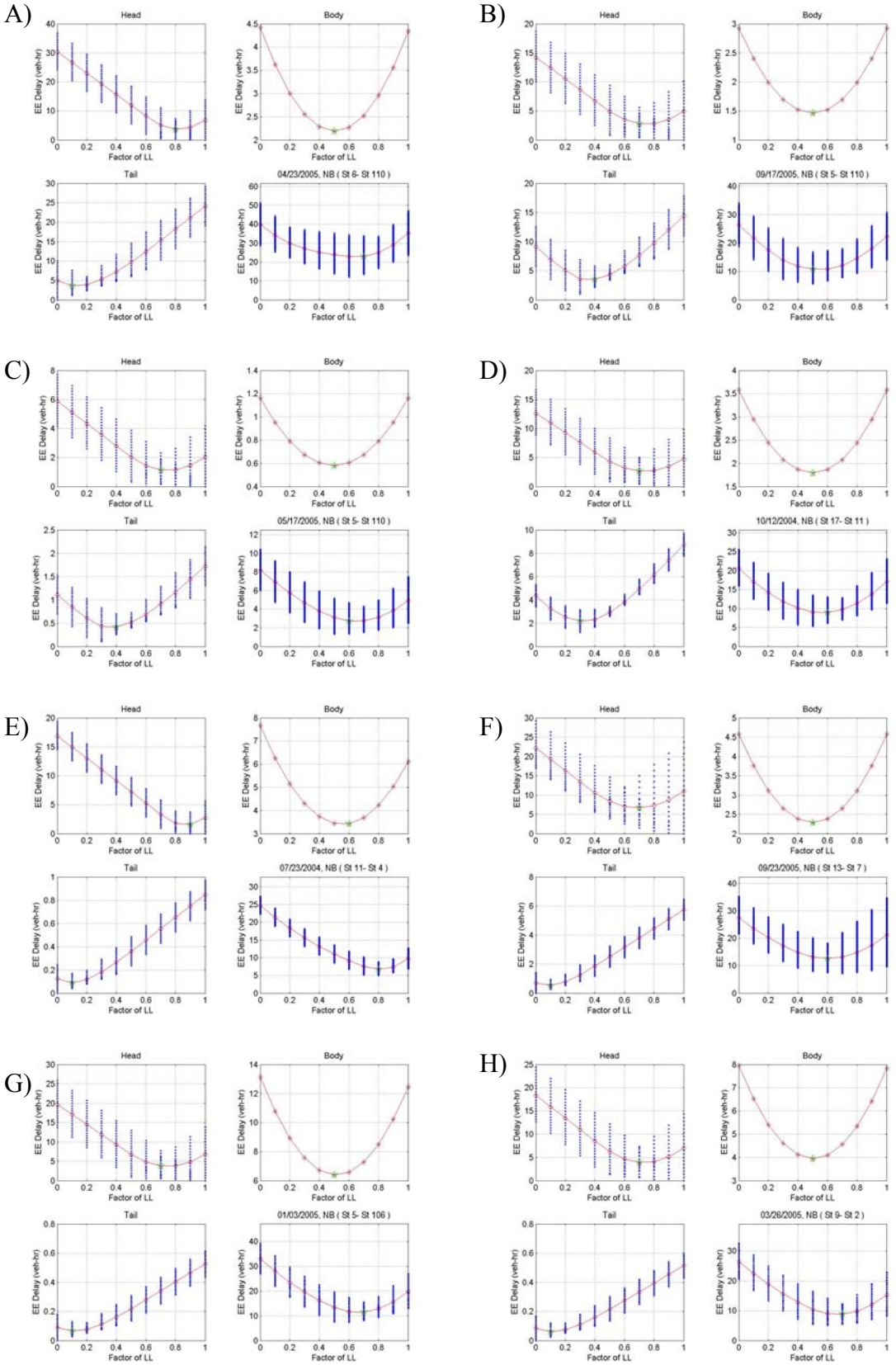


Figure I.3, Average estimation errors in each of three segments and in whole segments by link locations for double loop detector station spacing, NHB, (veh-hr per lane).

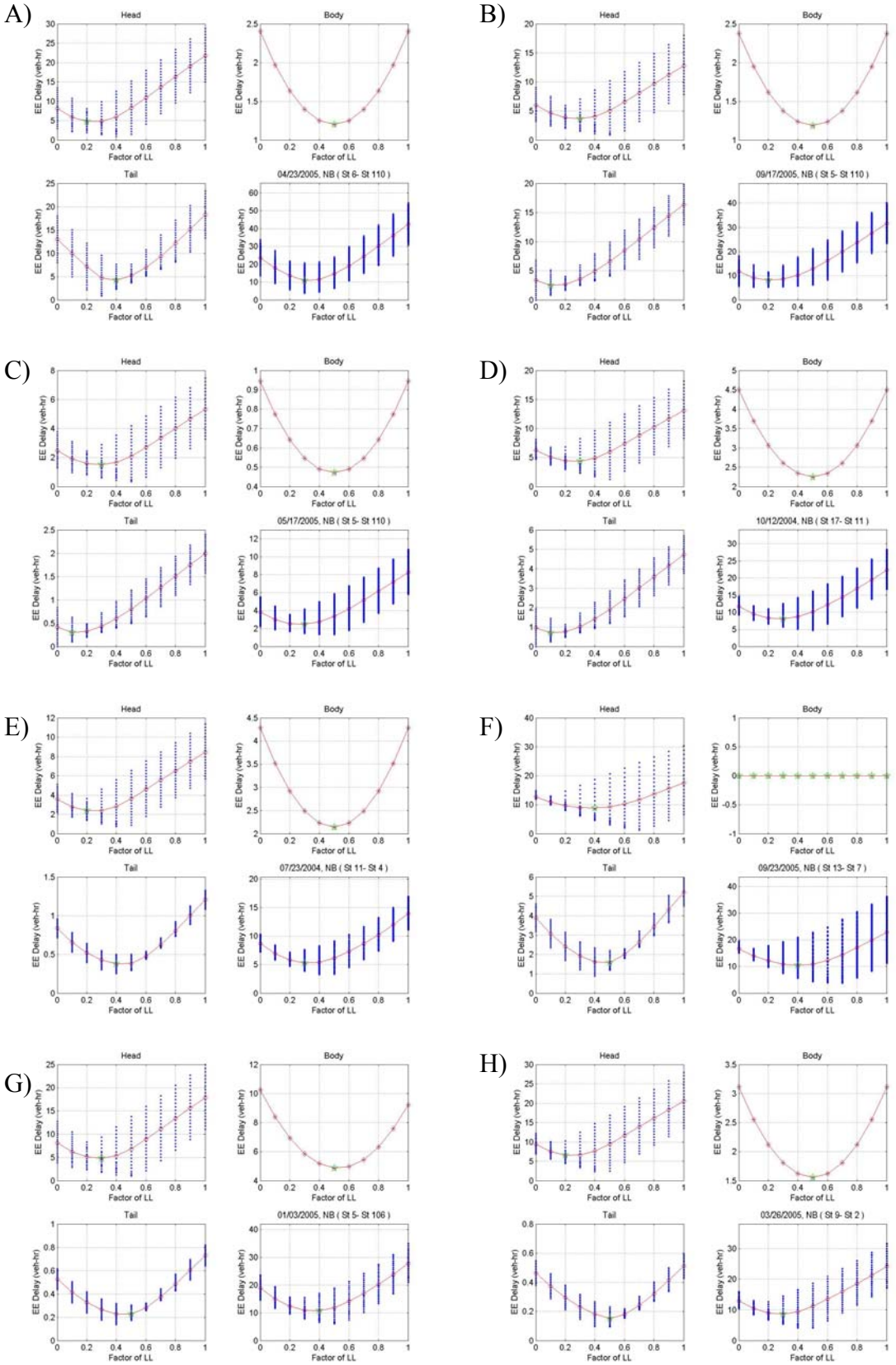


Figure I.4, Average estimation errors in each of three segments and in whole segments by link locations for double loop detector station spacing, PHB, (veh-hr per lane).