STATE OF CALIFORNIA DEPARTMENT OF TRANSPORTATION

TECHNICAL REPORT DOCUMENTATION PAGE

TR0003 (REV. 10/98)

1. REPORT NUMBER CA12-1116	2. GOVERNMENT ASSOCIATION NUMBER	3. RECIPIENT'S CATALOG NUMBER
4. TITLE AND SUBTITLE Deliver a Set of Tools to Resolving Bad Inductive Loops and Correcting Bad Data		5. REPORT DATE April 10, 2012 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Xiao-Yun Lu (Project Manager and Principal Researcher) ZuWhan Kim, Meng Cao, Zhaomiao Guo, Scott Johnston, and John Spring		8. PERFORMING ORGANIZATION REPORT NO.
9. PEFORMING ORGANIZATION NAME AND ADDRESS University of California, Berkeley 2195 Hearst Ave, Rm 1300, Mc 1103 Berkeley, CA 94720		10. WORK UNIT NUMBER 11. CONTRACT OR GRANT NUMBER 65A0335
12. SPONSORING AGENCY AND ADDRESS California Department of Transportation Division of Research and Innovation, MS-83 1227 O Street Sacramento CA 95814		13. TYPE OF REPORT AND PERIOD COVERED Final 14. SPONSORING AGENCY CODE
15. SUPPLEMENTAL NOTES		

16. ABSTRACT

This project prototyped and demonstrated procedures to find and mitigate loop detector errors, and to derive more valuable data from loops. Specifically, methods were developed to find and isolate out loop data which is "bad" or invalid, so that mitigation means, or "fixes" can be implemented. Methods of extracting very accurate speed (+/- 3mph) and vehicle length data (+/- 1meter) from single loop stations were demonstrated to be much more accurate than current Caltrans practice . The validity of these methods were statistically proven using hundreds of thousands of vehicles. Additionally, more accurate and reliable methods of detecting the onset of both recurring or "incident" based congestion were demonstrated.

These methods require access to the unprocessed loop detector card data. This unprocessed data can be acquired from the Log170 program, third party loop readers like the Infotek Wizard, or DRI's ubiquitous "C1 reader". DRI intends to implement many of these methodologies in the C1 reader client software, Videosync.

17. KEY WORDS Loop detector, Videosync, duel loop, incident detection, single loop, traffic detection, Log170, C1 reader, faulty data, detector station tools, speed estimation, congestion onset, event data correction, imputing missed data, cross talk, vehicle length measurement, Portable Loop Fault Detection Tool (PLFDT), vehicle-by-vehicle tracking	18. DISTRIBUTION STATEMENT No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. SECURITY CLASSIFICATION (of this report) Unclassified	20. NUMBER OF PAGES 213	21. PRICE

Reproduction of completed page authorized

Deliver a Set of Tools for Resolving Bad Inductive Loops and Correcting Bad Data

Final Report

Task Order 6327 and Award Number 65A00335

California PATH, ITS

University of California, Berkeley

Project Team:

Xiao-Yun Lu (Project Manager and Principal Researcher)

ZuWhan Kim, Meng Cao, Zhaomiao Guo
Scott Johnston, and John Spring

PIs: Pravin Varaiya and Roberto Horowitz

January 6, 2012

DISCLAIMER STATEMENT

Keywords: *inductive loop detector, loop fault detection, event data, traffic speed estimation, dual loop event data correction/imputation, vehicle length estimation and classification based on single loop event data, congestion onset detection, Active Traffic Management, integrated traffic data system, communication systems*

Abstract

This report documents practical works conducted at California PATH under the following project funded by Caltrans Division of Research and Innovations (DRI- Task Order 6327 and Award Number 65A00335): *Deliver a Set of Tools for Resolving Bad Inductive Loops and Correcting Bad Data*. The following topics have been studied under this project:

(1) Systematic Loop fault detection considerations and a Portable Loop Fault Detection Tool (PLFDT) development: Work in this aspect has considered the possible faults that could appear in loop detector system and the methods to detect them in a microscopic (control cabinet) level. A prototype PLFDT has been developed, which is based on the comparison of loop detector data and that from vehicle-by-vehicle tracking using digital video camera for loop fault diagnosis.

(2) Traffic speed and vehicle length estimation using single inductive loop event data: A new algorithm has been developed for traffic speed estimation based on single loop event data using *mode occupancy* corresponding to *mode vehicle length* which is known as *a priori* to be 15ft in California. Results showed that the estimated speed is very close to that estimated from dual loop data. Individual vehicle length has also been estimated with the known speed. Those algorithms are very simple and efficient and can be used in a control cabinet where even data are available.

(3) Event data correction to provide better quality event data of dual loop detector stations: Dual loop data are good for speed estimation only when they have a good quality. Algorithms have been developed to correct/impute some missed, mismatched, and incorrect time sequence data problems for Upstream-loop and/or downstream loop data.

(4) Congestion onset detection using dual loop event data: Traffic speed estimated from event data of consecutive dual-loop stations along a freeway corridor has been used to quick detection of traffic congestion onset for both time and location. The algorithm is based on the knowledge of shockwave back-propagation characteristics for saturated traffic. The algorithm can detect congestion onset within 1 min with dual loop stations 500m apart. Shorter distances between dual-loop stations will lead to less time delay.

(5) Developing an integrated traffic data system for future applications: Active traffic Management (ATM) intended to improve mobility and safety and to reduce emission and energy consumption with optimal use of current highway infrastructure. Such application poses some requirement for data system. Considering some traditional use of traffic data in both operation and planning, it makes better sense to develop an integrated traffic data system which could be used for most applications. Such system needs to take into account the impact of new data sources from vehicles through VII (Vehicle Infrastructure Integration) which is fast-growing in technologies and market penetration.

(6) Considerations of communication links in integrated traffic data system: Reliable communication is one of the key factors for ATM and for the Integrated Traffic Data System. Development of such system needs to meet the requirement of future application as well as the current situation with the cost constraint. For current communications used in data passing, old telephone lines and fiber optical cable are three times more reliable than wireless UDPD modem.

Acknowledgements

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 (Caltrans) under TO6327 and Award Number: 65A00335. 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

The project has been executed under the project manager Mr. Joe Palen of Caltrans Division of Research and Innovations. His direction and advice are gratefully acknowledged.

Other software and hardware support from PATH staff Susan Dickey, Thang Lian, and Dave Nelson, are also gratefully acknowledged.

Table of Contents

Appendices

List of Figures and Tables

Page

Figure

Table

Executive Summary

This report documents a practical work conducted at California PATH under the following project funded by Caltrans Division of Research and Innovations (DRI): *Deliver a Set of Tools for Resolving Bad Inductive Loops and Correcting Bad Data*. The following topics have been studied under this project:

(1) Systematic Loop fault detection considerations and a Portable Loop Fault Detection Tool (PLFDT) development

(2) Traffic speed and vehicle length estimation using single inductive loop event data

(3) Event data correction for dual loop detector stations

(4) Congestion onset detection using dual loop event data

(5) Developing an integrated traffic data system for future applications

(6) Considerations of communication links in an integrated traffic data system

Systematic Loop Fault Detection

Inductive Loops are widely used in California for traffic detection and monitoring. However, several faults may appear here and there in a loop detector system including the loop circuit buried in the ground, cable, loop card, and the communication systems used for data passing from the traffic control cabinet to TMC. Any fault may cause faulty data received in TMC, which directly affects the data application. To achieve high performance for highway operation and planning, it is very critical to have high quality data. Thus to maintain a healthy loop detector system is absolutely necessary. Due to a large number of loop detector systems, it is necessary to have a systematic and effective way to diagnose and pin-point the fault(s). Traditional approach for loop fault diagnosis was through aggregated data analysis in two levels: (a) macroscopic level as in TMC (Transportation Management Center) or PeMS (Performance Measurement System) in California, which uses highly aggregated data to look at loop problems related to an area; (b) mesoscopic level, which involves synchronized data for a section of freeways involving several control cabinet such as Berkeley Highway Lab (BHL), where the data used is still aggregated. Those indirect approaches can diagnose certain type of system faults to some extent. Due to the communication error such as packet loss, it was difficult to tell exactly where and what is the fault from those higher levels. This study proposed a combined approach: to find out certain system faults in higher level through data analysis, where one can also identify suspicious loops; and to diagnose faults in the inductive loop system directly at the traffic control cabinet level, which can be called a microscopic approach.

Portable Loop Fault Detection Tool

The objective of this effort is to develop a portable tool to be used at the control cabinet level to accurately diagnose any fault(s) of a loop detection system (including loop circuits, loop cards, cable links, etc.), to check the detection accuracy, to deal with detector card sensitivity, and to correct faulty data. To achieve those functionalities, it is necessary to utilize an independent source as the ground truth to compare against the loop detection system output. Such a comparison also permits an evaluation of the data quality of the loop system. Since multiple-vehicle tracking technologies using digital video camera on freeways have been well-developed and tested at PATH, it is used as the baseline measurement in the portable tool for the loop fault diagnosis.

This portable tool is composed of (a) a tractable pole with maximum height over 50 ft on a mobile trailer; (b) a Pan-Tilt-Zoom (PTZ) camera mounted on top; (c) a computer laptop for real-time image processing for multiple lane vehicle tracking; (d) a computer laptop interfacing with the loop cards in the controller cabinet; (e) wireless communication between the two computers for synchronization and data passing; and (f) a whole set of software to compare the loop detection signal and the vehicle detection signal from video tracking for loop fault detection. This report presents the preliminary development of such a system including the hardware, the software, the data communication method, and the algorithm. Some preliminary consideration has been conducted on lower (control cabinet) level data correction and investigating communication reliability of several Caltrans Districts.

Traffic Speed and Vehicle Length Estimation Using Single Inductive Loop Event Data

Traffic speed estimation based on loop detector occupancy is a long standing problem, which is critical to traffic management and control. Dual loop event data can be used to estimate reasonably accurate speed based on the ON/OFF time instant of the upstream and the downstream loop. However, single loop detector stations are still popular in practice. Previous studies for single loop speed estimation (e.g. the g-factor method) assumed the average vehicle length and aggregated loop data. However, this method did not produce good results sometimes, particularly in traffic transition phases. This work proposes a new approach using the *mode occupancy* with a moving window of fixed number of event occupancy data samples. BHL

(Berkeley Highway Lab) dual loop data which has 60Hz information with 1Hz update rate have been used for algorithm developing and validation. The speed estimated from the corrected dual loop data was used as the ground truth for comparison. Results showed that the developed algorithm generated very satisfactory speed estimation compared to that from the dual loop station. As a direct application, the estimated speed is used for individual vehicle length estimation using the event data. With this, one can conduct length based vehicle classification within a moving time window. The developed algorithms are simple and efficient and can be run in a control cabinet where Log-170 event data of an individual loop is available without need of any hardware modification.

Event Data Correction for Dual Loop Detector Stations

Loop detectors are widely installed nation-wide as the basic traffic detectors. In California, dual loop stations are mandatory in several Caltrans Districts. The advantage of dual loop stations over single loop is the speed trap property with event data. However, correct speed estimation using dual loop event data requires well-matched ON/OFF times from upstream loop (U-loop) and downstream loop (D-loop). As other sensors, dual loop data sometimes may have errors. Those errors include: incomplete pulse of a single loop, pulse breaking of an individual loop, mismatch of pulses between U-loop and D-loop, and improper sensitivities for one or both of them. Loop sensitivity is a generic problem and it affects traffic state parameter estimation too. This part of work has developed algorithms to systematically correct those errors to provide good quality even data for dual loop vehicle detector stations (VDS). . BHL archived event data and filed collected data in this section are used to explain the problem and to develop the algorithm. Admittedly, some mismatch error in BHL archived data might have been caused by data logging or communication processes. Even then , the developed algorithm is still useful to the archived data. The algorithms are simple and can be used in control cabinets in real-time too, if necessary.

Congestion Onset Detection Using Dual Loop Event Data

Fast and accurate detection of traffic congestion along a freeway corridor with heavy traffic is very critical for traffic management and control. For management purposes, an incident or accident must be removed in the shortest time possible so that the traffic can resume normal operation. Results obtained that could help alleviate this problem include: (a) factors that significantly affect traffic parameter estimation error and time delay at point sensor stations; (b) estimation method of traffic parameters using sub-second dual-loop station data with filtering and without aggregation over time; and (c) real-time algorithm using speeds and occupancies at two consecutive stations of a freeway section for quick congestion onset detection with the upper bound of time delay estimated. Two detection methods are presented: one uses the mean speed/occupancy across all the lanes at the two stations; and the other uses the speed/occupancy difference of one lane at an upstream station and the mean speed/occupancy across all the lanes at a downstream station. Method validation using BHL dual loop data with 60Hz information is also presented. It has been shown that for a high traffic flow corridor, the proposed algorithm could detect the location of the bottleneck with about 46s time delay if the loop stations are no more than 350m apart. Shorter distances between VDSs will lead to smaller time delay. The algorithm will be valid if other sensors are used as long as they provide reasonably accurate speed. Although the algorithm development initially relies on a good estimation of speed-based dual loop detectors, if the speed estimation from a single loop detector is adequately accurate (such as those in Chapter 3), the algorithm can also be implemented for a freeway corridor.

Developing an Integrated Traffic Data System for Future Application

In current traffic data systems, most raw sensor data are directly passed to the TMC for processing, archiving and immediate use in applications such as Active Traffic Management (ATMS) and/or traffic management. With the increase in the numbers andtypes of sensors, and traffic data from other resources such as on-vehicle information through VII (Vehicle Infrastructure Integration), passing all the data back to TMS may not be an economical and optimal option. As an example, for a freeway corridor level ATM, the data could be processed locally and used locally with higher resolution, and only some aggregated processed data are passed back to the TMC for higher level coordination. This way will obviously reduce communication overhead, which leads to cost reduction and reliability increase.

The goal of ATM is to improve mobility and safety, and reduce emissions and energy consumption with optimal use of current highway infrastructure. Such an application poses some requirements for a data system. Considering some traditional uses of traffic data in both operation and planning, it makes better sense to develop an integrated traffic data system which could be used for most applications. Such a system needs to take into account the impact of new data sources from vehicles through VII, which is fast-growing in technologies and market penetration. It also necessary to consider the data needs in the application of different levels: the lowest level would be in control cabinets for freeways (section traffic control) or arteries and roadways (intersection signal control). At the highest level data would be for regional level integrated corridor management. Data used in different levels would require aggregation in different time/space ranges.

Communication System Considerations

Reliable communication is one of the key factors for ATM and for the Integrated Traffic Data System. Development of such systems must meet the requirement of future application as well as the current situation and the cost. For current communications used in data passing, the old telephone line and fiber optical cable is three times more reliable than a wireless UDPD modem. The development of communication links needs to consider all possible applications at different levels in the future. The lowest level would be along a freeway and arterial corridor for ATM and ATMS. Although current freeway traffic control such as ramp metering is conducted at the TMC where the integrated VDS data are available, future ATM application may be conducted at the freeway and/or arterial corridor level using a data-control hub. Here, all the event data are processed for higher resolution/accuracy traffic state parameter estimation, which could be directly used as the input to the ATM controller. TMC could just receive processed and aggregated data and monitor the performance of the controller and send higher control commands to the data-control hub. The communication system also needs to possess the following capabilities:

- Link a freeway corridor and its related arterials and surface street to facilitate the coordination and control of all the subsystems belonging to different jurisdictions;
- Reconfigure the physical links of communication if necessary in an event such as incident/accident or evacuation in a natural disaster such as serious earthquake or tsunami. Such links will be critical for traffic routing that may not be used in daily operations.

Wireless communication could be used in such situations where the cables are usually not available. However, current wireless communication needs to improve its reliability in data passing, for instance by using more reliable protocol such as TCP instead of UDP, or including a byte in the data packet for automatic communication fault detection.

Chapter 1. Introduction

This final report includes all the studies and results under the following project funded by Caltrans Division of Research and Innovations (DRI): *Deliver a Set of Tools for Resolving Bad Inductive Loops and Correcting Bad Data*

Since the scope of the work includes systematic loop fault detection and data correction at a microscopic level, the following topics have been investigated under this project:

- Systematic Loop fault detection and a Portable Loop Fault Detection Tool (PLFDT) development
- Traffic speed and vehicle length estimation using single inductive loop event data
- Event data correction to provide better quality data for dual loop stations
- Congestion onset detection using dual loop event data
- Data systems requirements for Integrated Active Traffic Management (ATM) and planning
- Communication systems considerations

Each topic will be briefly introduced below.

1.1 Systematic Loop Fault Detection and PLFDT Development

A systematic approach to detecting faulty loops is crucial to traffic operation such as ATM and ATMS. Traffic detection systems are widely used for traffic management and control in California. The statewide *sensor system* consists of over 25,000 sensors located on the mainline and ramps, and grouped into 8,000 vehicle detector stations (VDS). Over 90 percent of the sensors use inductive loops. However, loop data are not reliable. The loop data delivered to TMC may contain errors generated at a point or several points in a loop detector system and between the loop detector and the TMC database, which presents a great challenge to loop fault detection. To solve this problem, it is necessary to take a systematic approach. This approach is composed of three complementary tasks: (a) loop fault detection; (b) faulty loop data correction/imputation; and (c) loop detection system maintenance. In our previous work [38], we categorized the loop fault detection approaches into three levels: (i) Macroscopic Level: such as the TMC/PeMS; (ii) Mesoscopic Level - a stretch of freeway (or a freeway corridor) such as the Berkeley Highway Lab (BHL); and (iii) Microscopic Level at a control cabinet which is connected with VDS in all lanes. The former two are of a high level and the latter is of a low level. Loop fault detection at the high level is done usually through an analysis of aggregated data. Such an approach is indirect, and its shortcomings are obvious: (a) data aggregation in time and space would obscure the fault problem; and (b) a communication fault may cause data error/ loss, which makes it impossible to isolate the data error/loss from the loop fault detection problem. Only the detection at the control cabinet level can directly detect the fault(s) in hardware and software, ,isolate them from the communication fault and correct them permanently. In [38] we conducted (a) Systematic review of previous loop fault detection and data correction methods; and (b) Systematic classification of possible faults and causes at the three levels described above.

The loop faults to be diagnosed at the microscopic level include problems of hardware, software, installation, and loop card faults. Those faults can be roughly classified as: *mis-assignment, temporary data missing, crosstalk, absence of data or constant data for a period of time, broken cable, chattering, broken card, card sensitivity being too high or too low, broken pulse, mismatch of ON/OFF time instances between upstream and downstream loops for dual loop stations*.

Only at the microscopic level (control cabinet) one could conduct direct loop fault detection and isolate the loop faults from other system faults. Data in this level can either be from the output of a loop detector card, which will be loop ON/OFF time instances or occupancies, or the raw loop pulse signal before the loop card obtained by wiring to the back of card cage. Those data are isolated from the communication system, and are accessible in real-time. It is also possible, only at this level, to identify all the loop detector system faults and their exact causes.

To directly detect loop system faults at the control cabinet level, a prototype Portable Loop Fault Detection Tool (PLFDT) has been developed in this project. It is composed of a mobile trailer, a retractable pole with a video camera mounted on the top to look at the suspicious loop detector on the ground, a computer running a vehicle tracking algorithm, with which the video camera is connected, and another computer at the control cabinet interfacing with a loop detector card though an RS232 serial port to receive loop data. Both computers also run IEEE 802.11b wireless for information passing and synchronization. In video tracking, a vehicle passing over a loop can also be considered as a vehicle activating a virtual loop. The virtual loop information is passed from the computer running vehicle tracking with digital video camera, to the cabinet computer through the wireless with UDP protocol. Information from the virtual and the physical loops are then compared for loop fault detection, which can be viewed on a visual display. Clearly, it is implicitly assumed that vehicle tracking through digital video should be reliable for this purpose.

1.2 Traffic Speed and Vehicle Length Estimation Using Single Inductive Loop Event Data

Single loop detector stations can provide occupancy and vehicle count (or flow) directly as raw measurement data. Occupancy is roughly equivalent to traffic density. With those two measurements, it is possible to estimate traffic speed, which is essential to many applications in traffic management. However, for single loop detectors, the two problems of finding the speed (individual vehicle speed or time mean/distance mean speed) and finding the vehicle length (individual vehicle length or average length) are equivalent. Algorithms used in practical applications such as the *g-factor* method usually assume a known average vehicle length to find mean speed if aggregated data are available. It is obvious that vehicle length varies significantly in time, location, and between lanes. Therefore, practical speed estimations from single loop detectors were not satisfactory. Although dual loop detector stations are better for traffic speed and length estimation, most vehicle detector stations (VDS) are single loops. To update from single loop to dual loop is very costly. Therefore, it is imperative to develop an efficient algorithm which can provide good speed and vehicle length estimation. It is obvious that working with aggregated data will not be able to achieve this because the characteristics of individual vehicles cannot be distinguished.

The work in this part of the project proposes to use event data for better speed and vehicle length estimation. The event data considered are those collected in higher frequency such that the activation of individual vehicles can be captured. Based on the vehicle size, loop detector size in vehicle moving direction, and possible vehicle speed range, if the data contains 60Hz information, it will satisfy this requirement. The Berkeley Highway Lab (BHL) is the unique data system in the world to provide such a data source. All the VDS in BHL are of dual loops which facilitate the development and validation of the speed estimation algorithm. The main idea is to extract the *mode occupancy* in a moving window with a fixed number of samples. It is known that, for a short enough time, the *mode occupancy* corresponds to those vehicles with the *mode length* which is known to be 15ft in California. This length has been used as a known vehicle length in the speed estimation. It has been proved to be much more accurate than the length assumptions in any other speed estimation algorithms through extensive data analysis. The algorithm has been applied to BHL data with the estimated speed compared with that from the corresponding dual loop stations. Root Mean Square Error shows quantitatively that the two estimations are very close.

Is timely event data acquisition and analysis possible? The answer is ―Yes!”, since it is not necessary to pass all the event data to the TMC. Instead, the speed estimation process can be easily done at the control cabinet with Log-170 or 2070 processor, due to its simplicity. Then only a minimum set of traffic state parameters (including occupancy, traffic count and the estimated speed) is passed to a higher level such as TMC. with lower update rate such as 1Hz.

With the speed estimation available, individual vehicle length can also be estimated with event data at the same time as a byproduct. It is well-known that vehicle length detection is very essential in goods movement, maintenance and planning. The estimated vehicle length from single loop VDS has also been compared with that estimated from the dual loop. The results showed that RMSE of the two was within 1m in most cases. Beside the length estimation, we also showed how to conduct length-based vehicle classification in a moving time window in real-time.

1.3 Event Data Correction to Provide Better Quality Data

Traffic data are applied in many areas and in multiple levels including Active Traffic Management, and Advanced Driver Information System. Quality of data is crucial for transportation systems. The quality of the data at the lowest (sensor) level is fundamental to all the higher level applications since the latter is aggregated from the former. High quality data at the sensor level, plus reliable communication systems for data passing, will save significant effort at all the higher levels for data processing and for relevant traffic state parameter estimation.

Former studies have been conducted extensively in aggregated traffic data cleansing, correction, imputation and mining, but very little such work has been done to lower level event data. This was partly due to lower level event data not being generally available in most data systems, partly due to the fact that interfacing with a loop detector (Reno 222) or popularly used 170 traffic controller to obtain such event data is difficult, partly due to the fact that the most popular statistical methods are naturally in favor of aggregated data, and most importantly due to the fact that the application of the traffic data in traffic management, particularly in planning, in the past did not require high resolution traffic state parameters. However, this situation has been changing very rapidly in recent years with the fast development of Active Traffic Management. In fact, if the raw loop data, basically occupancy and vehicle count, are processed at the lowest level to get traffic speed, and we only pass the processed traffic state parameters (speed, occupancy and vehicle count) at the finest resolution, most application requirements will be satisfied. Therefore, it makes better sense to develop the tools with the following capabilities:

- Detect fault in loop system to pin-point the problem: loop circuit, wiring between the circuit and the control cabinet, loop detector card, or software problem such as loop map, which is isolated from the communication problem;

- Event data cleaning and correction at control cabinet level;

- Traffic speed estimation of event data – with speed, occupancy and flow (traffic count in certain time interval) forming a complete set of traffic state parameters;

Dual loop event data from log-170 are not perfect in the sense that, even if the two loops in the dual-loop station work normally, measurement errors and noise still exist, which may be caused by the geometric shape of the vehicles, the distance of vehicle chassis to the ground, and the number axles and different materials. Besides, the U-loop/D-loop may miss some detection, and/or the data logging process and communication link may miss some data. The actual data may contain some mismatches of U-loop/D-loop ON/OFF times. Such mismatches would cause problems if they are used for individual vehicle speed or length estimation. This has been witnessed through the analysis of BHL archived data. Of course, the data loss in this data set might also be caused by miscommunication of data, passing from the loop detect station to BHL database. In any case, the development of systematic data correction, particularly for dual-loop stations, is useful even just for using the archived data. The main ideas are (a) to pair U-loop/D-loop data; (b) to check and correct the duration of U-loop and D-loop events; and (c) to streamline the time sequences: U-loop ON (OFF) time should be earlier than D-loop ON (OFF) time respectively. Such a correction process is critical to the improvement of accuracy and reduction of noise in speed and length estimation using dual loop event data. The accuracy improvement will be important for future real-time application at the control cabinet level.

1.4 Application of Dual-Loop Speed Estimation for Quick Traffic Congestion Detection

Modern ITS applications require highly accurate real-time estimation of traffic state parameters with minimum time delay. Loop stations, as point sensors, are popularly used on freeways as primary traffic measurement. An immediate application is to use the estimated parameters for quick automatic congestion onset (location and time) detection. Fast and accurate detection of traffic congestion along a freeway corridor with heavy traffic is very critical for traffic management and control. For management purpose, the incident/accident could be removed in a shortest time so that the traffic could be recovered for normal operation. For ATM, if the congestion location and time is quickly detected, the traffic upstream could be diverted through an alternative route such as a parallel arterial/freeway, and other traffic control measures such as Variable Speed Limit (VSL) and Coordinated Ramp Metering (CRM) can be adjusted to maximize the flow through the bottleneck caused by the incident/accident. If the real-time traffic speed estimations are available at consecutive stations along a freeway corridor, and if the overall system is coordinated, it is possible to use the integrated information for traffic congestion detection including the location, time and impact on the traffic. The main idea is to use the shockwave characteristics for saturated traffic and its effect on speed estimated at each station. Although the algorithm development here used the estimation of speed based dual loop detectors, if the speed estimation from single loop detector is adequately accurate (such as those in Chapter 3), the algorithm can also be implemented.

1.5 Towards an Integrated Hierarchical Traffic Data System

In current traffic data systems, most raw sensor data are directly passed to TMC for processing, archiving and application such as ATMS and/or traffic management (e. g. ramp metering). With the increased number of sensors and traffic data from other resources such as on-vehicle information through VII (Vehicle Infrastructure Integration), passing all the data to TMS may be neither economical nor optimal. As an example, for a freeway corridor level ATM, the data could be processed and used locally in a data hub with higher resolution. Only some aggregated essential traffic state parameters are passed back to TMC for higher level applications. This way will obviously reduce communication overhead, which leads to cost reduction and reliability increase. Since the purpose of the data system development is for application to traffic management and planning, it is necessary to look ahead to what is needed in the future, and also how new technologies, concepts, andapplications could impact traffic data systems. Based on current knowledge and fast development of ATM and VII (Vehicle Infrastructure Integration), a roadmap for the development of an integrated traffic data system is proposed. Suggestions of low goal and high goal for the traffic data system to be used in ATM have been made. The roadmap is intended to fully use current traffic data system infrastructure, while gradually incorporating other traffic data resources while exploiting the increased market penetration of VII. A gradual process has also been sketched for improving traffic data quality and resolution, and establishing a hierarchical data system with the dynamic reconfiguration capabilities required by ATM in the future. It is also necessary to think about how to acquire and reduce data for all possible applications.

1.6 Building Reliable Communication Links in the Data System

Communication for data passing is essential for practical applications (ATM and ATMS) in an integrated hierarchical traffic data system. High performance traffic management systems would require high performance communication links. The performance of the communication system could include, but not be limited to, the following factors: (a) availability of physical links; (b) reliability of the links; (c) packet size to be carried and its update rate; and (d) cost. The overall communication links will depend on the overall structure of the integrated traffic data system. In the future, local communication system configuration will depend on the data needs for the Traffic Management of freeway and arterial corridors (as an integrated system).This would most likely require a local data hub. Medium/long distance communication links may only be used for the link between the local data hub and TMC/PeMS for a higher level application such as regional level traffic management and planning.

Although current freeway traffic control such as ramp metering is conducted at TMC where the integrated VDS data are available, future ATM may be conducted at the freeway and arterial corridor level using a data-control hub. Here, all the event data are processed for higher resolution traffic state parameter estimation, which could be directly used as an input to the ATM controller. TMC could just receive processed and aggregated data and monitor the performance of the controller while sending higher control commands to the data-control hub.

The communication system needs to have the following capabilities:

- Link a freeway corridor and its related arterials and surface streets, to facilitate the coordination and control of all the subsystems which may belong to different jurisdictions;

- Reconfigure the physical communication links for a special event such as an incident/accident or evacuation in a natural disaster such as serious earthquake or tsunami. Such links will be critical for traffic routing, although such reconfiguration may not be desirable for daily operations.

To guarantee high quality of data, the communication links must be reasonably reliable. For current communications used in data passing, old telephone lines and fiber optical cables are three times more reliable than wireless UDPD modems. Besides the system itself, some suggestions have been made on communication protocol to avoid data loss, and a simple method for automatic detection of communication faults.

Chapter 2. Loop Fault Detection and Development of a Portable Loop Fault Detection Tool

2.1 Introduction to Chapter 2

Traffic detection systems are widely used for traffic management and control in California. The statewide *senso*r *system* consists of over 25,000 sensors located on the mainline and ramps, which are grouped into 8,000 vehicle detector stations (VDS). Over 90 percent of the sensors use inductive loops. However, loop data are not reliable. The loop data delivered to TMC may contain errors created at a point or several points between the loop detector and the TMC database, which presents a great challenge to loop fault detection. To solve this problem, it is necessary to take a systematic approach. This approach is composed of three complementary tasks: (a) loop fault detection; (b) faulty loop data correction/imputation; and (c) loop detection system maintenance. In our previous work [38], we categorized the loop fault detection approaches into three levels: (i) Macroscopic Level: such as the TMC/PeMS; (ii) Mesoscopic Level - a stretch of freeway such as the Berkeley Highway Lab (BHL); and (iii) Microscopic Level at a control cabinet. Different data are available at different system levels. The former two are of a high level and the latter is of a low level. Loop fault detection at the high level is usually conducted through the analysis of aggregated data. Such an approach is indirect with obvious shortcomings: (a) data aggregation in time and space would smear the fault problem; and (b) communication faults caused by data error/loss make it impossible to isolate the loop fault detection problem. Only detection at the control cabinet level can directly detect the fault(s) in hardware and software, isolate them from the communication fault, and correct the faults permanently.

Many methods have been adopted for loop fault detection and data correction/imputation. Different methods worked on different level of data in different ways. For example,

- Time aggregated data versus sub-second data
- TMC level versus control cabinet level
- Synchronized adjacent lane data versus downstream/upstream data
- Historical data versus real-time data
- Raw loop data versus filtered/aggregated data

- Statistical methods versus deterministic filtering
- Single loop stations versus dual-loop stations

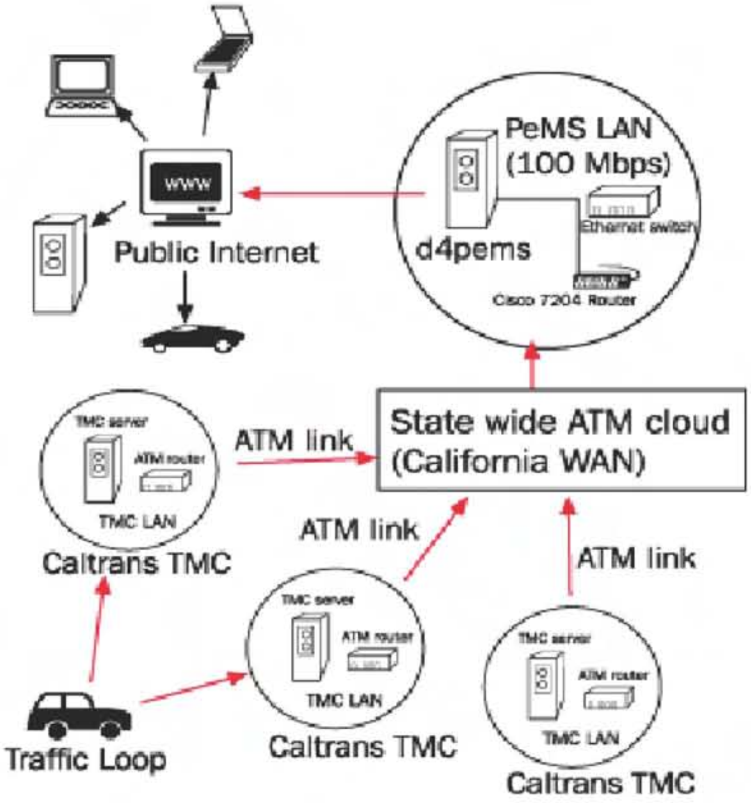


Figure 2.1. PeMS Structure in California

To systematically consider loop fault detection and data correction/imputation, it is necessary to diagnose possible faults at different levels of a traffic monitoring system. The overall picture for the data flow from individual loops to the TMC and PeMS in California can be described as (Figure 2.1): loop ➔ pull-box ➔ control cabinet with 170 + modem (30s data packets)➔ communication link (Old Telephone line up to 20 cabinet share one line, fiber optical cable, UDPD modem) ➔ Front-End-Process (FEPT) of ATMS of District TMC ➔ PeMS. The system can be divided into three levels: (i) Macroscopic Level: such as TMC/PeMS; (ii) Mesoscopic Level - a freeway corridor such as Berkeley Highway Lab (BHL); and (iii) Microscopic Level, i.e. at a control cabinet. Different data are available at different system levels. For example, in California PeMS (Figure 2.1), 2Hz data is available at the TMC and PeMS level, which are further aggregated into 5 minutes data or even longer time thresholds.

Loop fault detection in mesoscopic and microscopic levels are necessary to produce good data quality in all levels, which is required by current and future traffic management and control, particularly the new trend in Integrated Active Traffic and Demand Management along a freeway/arterial corridor. This systematic approach considers the demand from all modes, all the roads, in all the time for a transportation corridor. It is intended to develop a strategy to optimally manage traffic for improved mobility, safety, emission, land use and energy consumption. The strategy could include different levels of management tactics which correspondingly needs different levels of good quality data for support. Therefore, systematic sensor detection and data correction in all the levels are imperative and indispensible.

Systematic fault detection of the traffic monitoring system composed of loops, needs the combination of diagnostics at different levels. As is shown in Figure 2.1, the traffic monitoring system has a hierarchical structure for data collection, processing and passing. Inductive loops and other sensors are in the lower level. Data analysis at any higher level through data analysis can only diagnose the loop fault indirectly. This also suggests that it is necessary

(a) to distinguish data analysis and data correction at higher level from loop fault detection since they are indirect;

(b) to use higher level data analysis to identify suspicious control cabinets which may have potentially faulty loop stations;

(c) to diagnose higher level problems in the communication system, power supply, or data acquisition system or software;

(d) to combine higher level data analysis with lower level (onsite) loop fault detection using the Portable Tool, to pin-point the fault and its causes.

Errors in the loop data obtained at TMC may be caused by faults at any point or several points from the loop to the TMC database: physical loop, connection of the loop to the control cabinet, loop card errors including sensitivity, and communication media between the control cabinet and TMC. It can also be envisaged that data analysis at any point other than the control cabinet can only indirectly diagnose the loop fault in the sense that faults at higher levels could also interfere with fault loop data analysis. Thus those methods claimed for loop fault detection which are based on possibly faulty loop data analysis, are essentially fault diagnosis of the monitoring system, which can be called *indirect methods.* Only those methods or tools used at the control cabinet level can be called *direct methods*. Direct detection must be performed by a *portable tool* at the control cabinet level, either online or off-line using collected data, which needs to have the following functions:

(a) generate ground truth based on some independent sensor;

(b)synchronize the detection of the loops connected to the control cabinet with the ground truth detection;

(c) compare the loop data with ground truth for diagnosis.

According to the report on Detector Fitness Program [56], loop detector system healthy conditions varies significantly from Caltrans district to district. The following faults often cause faulty traffic data from the viewpoint of higher level traffic monitoring system for a large area:

- Communication failure
- Systematic failures: Systematic differences in failure rates by freeway and by lane, which could be affected by vehicle types;
- Electrical failures such as splicing problems or detector card faults;
- Synchronous failures: District-wide synchronous failures; e.g., unusually many loops in a District fail on the same day;

The following faults may appear at a mesoscopic level such as on a stretch of freeway. A typical example is the Berkeley Highway Lab.

- Communication Down: No samples were received for the loop between 5:00 am to 10:00 pm;
- Mis-assignment: Mismatch between the real location and the location assigned in the map in control cabinet;
- Insufficient Data: PeMS receives too few samples to determine the loop health;
- Card Off: Too many samples have zero occupancy;
- High Occupancy: Too many samples with occupancy above 70%;
- Intermittent: Too many samples with zero flow and non-zero occupancy;
- Constant: The loop is stuck on a particular value;
- Feed Unstable: The detector failed in the past, and its current status cannot be determined due to problems in the data feed;
- Correcting the data.

The problems to be looked at for an on-site fault diagnosis tool at the control cabinet level are:

- No loop data;
- Chattering and misfiring;
- Cross-talk;
- Pulse duration error;
- Pulse breaking;
- Temporary inductance variation;
- Mis-assignment;
- Sensitivity problem;
- Loop detector card broken.

Based on the higher level diagnosis, the Field Tool would check suspicious loops by comparing loop data with ground truth from independent sensors at the control cabinet level. This will be able to exactly identify the problem.

2.2 Literature Review

This section focuses on the following points: (a) Systematic review of previous loop fault detection and data correction methods; and (b) Systematic classification of possible faults and causes in different levels. Although this review did not exhaust all the publications in this area, the reader could trace other publications further from the literatures reviewed. The objective this review is to find merits and weaknesses of those methods which will be used as the basis for the development of this project.

Literature review of the project is related to several areas investigated:

- Loop Fault Detection
- Microscopic Level Data Correction
- Congestion Onset Detection
- Traffic Speed Estimation Using Dual and Single Loop Detector Data

This section classifies and reviews previous studies on faulty loop data analysis and loop fault detection at different levels of the system, which corresponds to using different levels of aggregated data. Since data correction/cleansing and imputation are usually closely related to data analysis and detection, they will be briefly reviewed and classified in parallel. Previous work on loop fault detection and data correction/imputation can be divided according to the data levels: macroscopic, mesoscopic and microscopic.

A systematic literature review on loop fault detection through faulty data at different aggregation levels was conducted in [38]. The characteristic of this approach is to apply various statistical analysis methods to the aggregated loop data to find out possible faults in the loop detection system. Since most previous approaches are indirect, the faults that can be detected are usually large scale problems such as electric or communication system faults, which cannot pinpoint the problem and the exact location in the loop detection system. This is one of the limitations to the macroscopic and mesoscopic approaches. Besides, communication faults are tangled with the loop detection system faults.

2.2.1 Loop Fault Detection at Macroscopic Level

A typical example is the PeMS level or Caltrans District TMC level, which provide 30 second and 5 minute aggregated data. Each Caltrans District is composed of multiple highway corridors. The main characteristics of those data are that (a) they are the data practically used for traffic management such as ATMS and ramp metering; (b) heavy data aggregation are usually involved; (c) those data usually need to pass through long distance communication media to reach PeMS/TMC; and (d) the data will be subject to small time delay due to data processing and passing through the communication.

PeMS data DSA (Daily Statistics Algorithm) checking for data errors [3]:

- The number of samples in a day that have zero occupancy must be less than a certain threshold;
- The number of samples in a day that have occupancy greater than zero and flow equal to zero must be less than a certain threshold;
- The number of samples in a day that have occupancy greater than a certain value (PeMS uses 35%) must be less than a certain threshold;
- The entropy of occupancy samples must be greater than a certain threshold.
- The definition of entropy is:

$$E = \sum_{x:p(x)>0} p(x) \times \log\left(p(x)\right)$$

The idea is that constant value of flow will lead to low entropy. Thus entropy could be used to detect if the detector has constant value consistently.

[57] used adjacent loop point flow for comparison to detect possible erroneous data. It used the ratio of flows of upstream and downstream stations as the measure for test. The reason is

that: for some time t, the upstream and downstream have completely different clusters of vehicles. For free-flow traffic and 10-minute aggregated data, this makes sense.

[3] is a systematic work in data based fault detection focusing and on how to correct the data for the following two cases: data missing and bad data. It also proposed a method for data correction. [46] used the ARMA model for prediction of loop data, which was over time and could be used to fill in faulty data. But [3] commented that that its response was too fast. It suggested using good neighbor (same location but different lanes, or adjacent locations) data for patching the whole. Averaging or interpolation over space methods were used for filling the whole. The mathematical foundation for this method was that occupancy and flow of neighbor loops were highly correlated. However, if several loop stations were down in a section of freeway, this method would become questionable. The algorithm developed in [3] is called Daily Statistics Algorithm (DSA) since it produces only one result using a whole day's 30s *q* (volume) and *k* (occupancy) data: good or bad on that day. The Detection criterion is based on the value of 4 statistic parameters and the selected threshold. Each statistic parameter targets for one error type.

The main method used [3] for data correction was to look at neighboring loops in adjacent lanes and/or up/downstream as well as historical data:

- Linear interpolation over time of the loop itself
- Linear interpolation over space of neighboring loops
- Averaging over time of the loop itself
- Averaging over space of neighboring loops
- Combinations of them all – in fact, averaging is a special case of interpolation

This method could not distinguish the case of temporal loop failures since the statistic over a whole day will not tell temporal behaviors. The proposed method used thresholds to identify 4 types of loop data errors:

- Occupancy and flow are mostly zero
- Positive occupancy and zero flow
- Very high occupancy
- Constant occupancy and flow

This has been achieved by classifying fixed loop daily data into 4 categories and then aggregating over time. Then thresholds are defined for such error identification, which is based on some common knowledge. These methods could not be used for the following faulty loops:

- Permanent isolated fault loop
- Temporal faulty data, such as those cases which are affected by weather and heavy traffic
- Individual loop faults such as sensitivity, crosstalk, etc.

This algorithm has been used in PeMS for several years. It proved to be reliable and better than other methods for higher level aggregated data for some larger range and or longer time loop problems.

Data correction methods were also proposed in [48], which was basically using historical data as well as adjacent station data for interpolation over distance and time. A Kalman filter is also designed for estimation of lane volume to filter out measurement noise. The filter performance showed that it was unbiased with discrepancy of 300vhr.

In the work of [57], Poisson distribution was used to describe the probability for the number of vehicles counted (flow) at a loop station every 30s interval.

$$p(y) = e^{-\mu} \frac{\mu^{y}}{y!}$$

$y -$ point flow: vehicle count at a given loop station. The probability for n continuous readings of a flow y was:

$$p^{n}(y) = e^{-n\mu} \frac{\mu^{ny}}{(y!)^{n}}$$

Then a threshold was set for data error checking: $p^{n}(y) \le P_{\min} = 0.0005$. An accumulated Poisson distribution should be used to represent the point flows at a loop station.

$$P(0 < y \le x) = \sum_{y=1}^{x} e^{-\mu} \frac{\mu^{y}}{y!}$$

Due to the stochastic property, the point flow y could be quite different for different traffic situations: AM peak, PM peak, off-peak, congested and non-congested cases. This idea is quite different from the entropy test of PeMS where constant flow will lead to very low entropy. This means that low entropy corresponds to invariance of traffic flow, which can happen only if the loop has faulty reading.

Time-of-day flow and occupancy ratio were used to reflect vehicle types such as trucks and passenger cars [57, 17]:

- This ratio could assume any value;
- Trucks correspond to low flow and high occupancy
- Passenger cars the other way around in the same time period
- Low flow and high occupancy may indicate congested traffic in another time period (caused by AM peak, PM peak and incident/accident)

[61] used loop data to calculate average vehicle length: 2.7m~18.0m. This threshold is used for data error checking. It is obvious that a such check can only tell if the data is reasonable or not. It could not tell what was wrong exactly with the system.

The Detector Fitness Program (DFP) [56] looked at the loop station in three Caltrans Districts: D4, D7 and D11. It defined some measurement parameters. The study proposes and calculates three metrics of system performance: *productivity* is the fraction of days that sensors provide reliable measurements; *stability* is the frequency with which sensors switch from being reliable to becoming unreliable; and *lifetime and fixing time* — the number of consecutive days that sensors are continuously working or failed, respectively. Productivity measures the performance of the sensor system; stability measures the reliability of the communication network; lifetime and fixing time provide more detailed views of both components of the sensor network. The evaluation method first uses PeMS 30s data. The second data set comprises records from the Detector Fitness Program (DFP) for Districts 4 and 7. These records were created by crews following a field visit to a loop. *Fault States* looked at included: *line down, controller down, no data, insufficient data, card off, high value, intermittent, constant value,* and *feed unstable*. Detection methods involved was mainly *Data Threshold Checking*. This work also looked at the possible higher level fault caused by communication systems involved in data passing for TMC/PeMS, which include: Caltrans owned fiber optics, wireless GPRS modem (UDP, TCP), telephone line and wireless cell-phone lines. The main idea is to tell if the communication system is healthy from the status of all the loop data related to the same communication system such as those belonging to the same control cabinet.

Summary: The problems to be looked at for macroscopic data analysis are:

- Communication Down: No samples were received for the loop between 5:00 am to 10:00 pm;
- Insufficient Data: PeMS receives too few samples to determine the loop health;
- Card Off: Too many samples have zero occupancy;

- High Occupancy: Too many samples with occupancy above 70%;
- Intermittent: Too many samples with zero flow and non-zero occupancy;
- Constant: The loop is stuck on a particular value;
- Feed Unstable: The detector failed in the past, and its current status cannot be determined due to problems in the data feed;
- Systematic failures: Systematic differences in failure rates by freeway and by lane which could be affected by vehicle types;
- Electrical failures such as splicing problems or detector card faults;
- Synchronized failures: District-wide synchronized failures; e.g., unusually many loops in a District fail on the same day;
- Identifying suspicious loops

Methods used at this level for direct loop fault detection include: (a) statistical, (b) entropy, (c) threshold checking based on some known physical limits and empirical values, and (d) comparson with neighboring (adjacent lanes, upstream/downstream) stations.

Methods used at this level for data correction/cleansing/imputation include: omitting unreasonable data based on some threshold; linear interpolation or moving window averaging over time, space (adjacent lanes, upstream and downstream)

2.2.2 Loop Fault Detection at Mesoscopic Level

The testystem in this level involved a section of freeway which has more than one control cabinet with multiple loops. The characteristics in this level are:

- Sub-second data of each are available;
- Loops connected with the same control cabinet are time synchronized;
- Loops connected with different control cabinets are time synchronized;
- Minor communication system is involved in data synchronization and data passing.

Thus the communication system fault can be easily determined by some simple ad hoc method such as check sum. In this way, the communication system fault could be isolated from the loop fault. Berkeley Highway Lab (BHL) is a typical example of such a system. BHL has 9 loop stations with 164 loop detectors for both sides of Interstate I-80 between Gilman St. and Powell St. (Figure 2.2).

Work in [43, 44] considered loop fault detection systematically based on the BHL system. A two-level, nine-diagnostic scheme has been developed including dynamic diagnostics based on speed and vehicle composition. The developed algorithms were implemented in software and are currently running in the BHL system. This work separated detector *deficiencies* and detector *faults*. The fault detection system used 1/60s data from loops, which were basically some threshold tests:

- activity test: test criterion: continuous 15 minute constant signal;
- Minimum on-time test for at least 100 vehicles (failure criterion: 5% vehicles occupancy < 8/60s);
- Maximum on-time test for at least 100 vehicles (failure criterion: 5% vehicles occupancy > 600/60s);
- Dynamic Minimum/ Maximum on-time test: similar to minimum/maximum adjust those time interval test threshold based on speed and vehicle length;
- Minimum Off-Time – If 5% or more of the off-times in a sample of 100 vehicles are less than 25/60 seconds, the test fails;
- Dynamic Maximum Off-Time – This is one of the new diagnostics. If 5% or more of the off-times in a sample of 100 vehicles are greater than a threshold value, which is a variable depending on the calculated average time headway, the test fails;
- Mode on-time test: test for 1000 vehicles; Test criterion: calculate mode of the distribution is outside of the interval [10/60s, 16/60s];
- Dual loop on-time difference test: test for 1000 vehicles; Test criterion: if the difference between the upstream and downstream loop is outside the time interval [-3.5s, +3.5s]; it is only valid in free-flow condition; not well-designed yet;
- Refining those tests in two aspects:
    - Predicting that the detector passes the tests when in fact the detector data is not good
    - Predicting that the detector fails the tests when in fact the detector data is good.

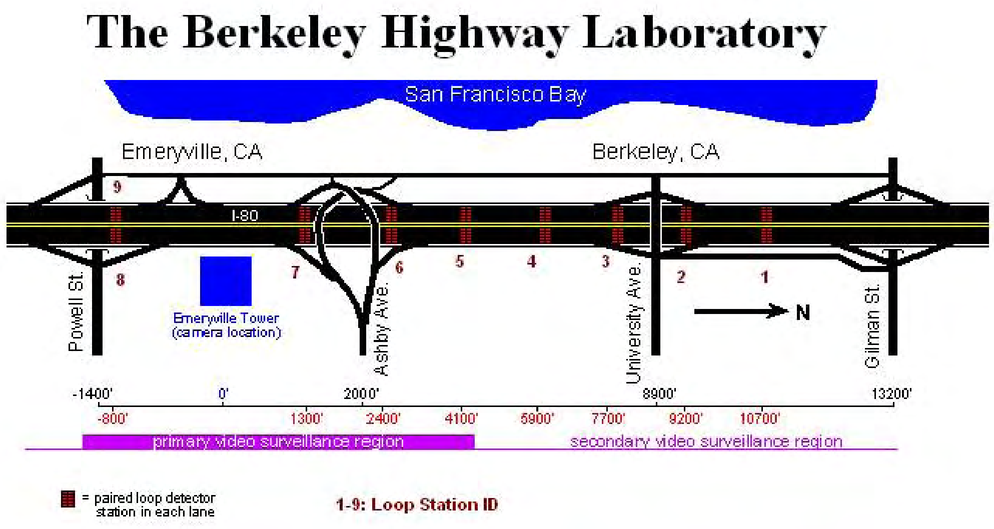


Figure 2.2. Berkeley Highway Lab

The test will need to account for the situations when there is little traffic such as in the early morning. This study also identified that some data problems are due to the Verizon CDPD modem network connection instead of loop stations faults, which means that the communication fault could not fully separated. It indicated the necessity of direct loop fault detection at control cabinet level.

This work recognized the importance of using low level sub-second data instead of aggregated data. Conventional traffic monitoring aggregates the event data to fixed period samples of flow, velocity and occupancy before transmitting the data to the Transportation Management Center (TMC). The sampling period is typically on the order of 30 sec or 5 min. This relatively coarse aggregation can obscure features of interest and is vulnerable to noise. Both of these factors delay the identification of resolvable events, the former due to the need to wait until the end of a given sample period and the latter due to the need to wait for multiple sample periods to exclude transient errors. The Nyquist sampling criteria from basic signal processing dictates that one can only resolve features that last two sampling periods and the need to tolerate noise in the measurements further reduces the response time. As such, it is necessary to have a trade-off between cost and data passing frequency. This study also suggested passing all the event (low level sub-second) data to MTC as well as all the data processing. It mentioned that link travel time for BHL is based on vehicle re-identification.

A methodology for substituting for missing data (imputation) was also developed in [44]. The missed data is imputed based on the data of adjacent lanes using interpolation.

Work in [49] also looked at 20s and 5minute data. The work used reasonable interval for flow density and speed to test if the data were reasonable: if they fell into the interval, then they are considered good data. Otherwise, they were considered bad data. The thresholds of those intervals were specified based on experiences on historical data. Similar idea was used for *k-q* plane for specifying a criterion region by [30]. The boundary of the region is determined by some parameters which need to be calibrated according to the site situation. This idea is slightly better because the relationship between *k* and *q* is taken into consideration. However, they did not take the advantage of using historical data as well temporal data relationships in detection and correction. [3] indicated that those methods were difficult to use in practice since the thresholds were difficult to calibrate. Due to those factors, several situations were incorrectly detected: false positives and false negatives happened.

[48] uses FSP data which is composed of three parts: loop detector data, probe vehicle data and incident data of approximately two months. The loop detector data includes 30s data and 5 min aggregated data for data error checking. The loop locations are divided into mainline, HOV lane and on-ramp, which have different traffic characteristics. 14 error checking criteria based on the two types of data sets are proposed. Parameters taken into consideration are volume, occupancy and average speed. The data needs pass 10 consecutive tests. Those checks include bounding checking – traffic parameters must be within certain physical bounds; contradictory check – two traffic parameters such as occupancy and value, occupancy and speed from the same loop station must be consistent. The seriousness of erroneous data have been analyzed according percent of time in malfunction, percent of station and percent of time in malfunction, etc. It has been found that data missing is the most significant error which appeared for blocks are sensors/stations. This may suggest that such error is caused by data transmission or the communication system. It was found that for I-880 FSP data, malfunctioning stations are about 21% of the whole on average, even if the stations are well-maintained for proper function.

<u>Summary:</u> Highway Section/Corridor: Typical example is the data from Berkeley Highway Lab. In this level, 60Hz data is available every second. The characteristics of those data are that (a) sub-second data could be obtained at this level; (b) time synchronized sub-second data are available for loop stations on a stretch of highways; (c) only a short distance communication system is involved; (d) the detection could be near real-time in the sense that the time delay for data passing was at the level of a few seconds. Besides hardware and software problems, other loop faults looked at this level include: *mis-assignment, temporary data missing, crosstalk, no data or constant for a period of time, broken cable, chattering, card broken, card sensitivity too high or too low, pulse broken, mismatch of ON/OFF time instant between upstream* and *downstream loops for dual loop stations*, *and identifying suspicious loops.*

Methods used for direct loop fault detection include: analyzing sub-second data, threshold checking, and vehicle re-identification. Methods used at this level for data correction/imputation include: linear interpolation or moving window averaging over time, space (adjacent lanes, upstream and downstream). It is noted that even at this system level, some detailed loop faults still cannot be detected. The advantage for such a system is that one could compare the synchronized upstream station and downstream station data for diagnosis and data correction, which could not be achieved at the control cabinet level.

2.2.3 Loop Fault Detection at Microscopic Level

Many operating agencies use specialized loop testers to assess the quality of the wiring [31, 28], but these tools bypass the controller and loop sensors; thus, they do not analyze the entire detector circuit, nor do they analyze the circuit in operation. To this end, most operating agencies employ simple heuristics methods such as whether the loop sensor indicator lights turn on as a vehicle passes. Such tests are typically employed when the loops are installed close to the control cabinet. Many practitioners and some researchers [30, 13, 46] have worked to formalize the latter heuristic by looking at if the time series 30 second average flow and occupancy within statistical tolerance.

Low level loop data correction could be traced back to the Freeway Service Patrol study in the 1990s [51, 59]. It looked at the transition times in sub-second sampling of dual loop stations with 20ft distance between upstream and downstream loops. It noticed some problems in low level data including:

- data missing;
- mis-matching of those data which results in unreasonable occupancy and speed;
- on-time and off-time are not always related;
- no-flow and no-speed but with positive occupancies;
- existing pulses in both up and down streams

The author mentioned that some of the phenomena could be explained as caused by a vehicle changing lanes. However, there is no systematic diagnosis in [51] for loop faults, nor systematic methods for lower level data correction.

In [4], Chen and May considered fault detection problems for a single loop. It used the number of pulses as vehicle counts to verify loop data. If a pulse were broken, it would cause the data problem. They developed an automated loop fault detection system which uses aggregated data. They must accept a large sample variance and potentially miss problems altogether. For example, the systems have to tolerate a variable percentage of long vehicles in the sample population. Their methodology examines the distribution of detector *on-time*, i.e., the time the detector is occupied by a vehicle. Unlike conventional aggregate measures, their approach is sensitive to errors such as "pulse breakups", where a single vehicle registers multiple detections because the sensor output flickers off and back on. This is the main disadvantage to using vehicle count from a single loop for fault detection: one cannot isolate other loop faults from the pulse flickering problem.

Studies in [14] use dual loop information for comparison to detect loop faults. It focused on evaluation of loop sensors and detection of cross-talk. It was developed for off-line data analysis but could possibly be used for on-line in the future. It can be summarized in three steps:

(i) Record a large number of vehicle actuations during free flow traffic;

(ii) For each vehicle, match actuations between the upstream and downstream loops in the given lane;

(iii) Take the difference between matched upstream and downstream on-times and examine the distribution on a lane-by-lane basis. Assuming the loops are functioning properly, only a small percentage of the differences should be over 1/30 seconds. Otherwise, –Cross-talk” fault is inferred.

Using dual loop speed traps to identify detector errors is another approach conducted in [14]. At free flow, on-time difference and off-time difference should be the same if there is no hardware problem. So if they are not the same, there may be a hardware and/or software problem. But this is not true if it is not free-flow speed.

For traffic data processing and loop fault detection at microscopic level, it is necessary to understand the physical principle of a loop detector system. The physical link of a loop circuit buried in the ground and the control cabinet is shown in

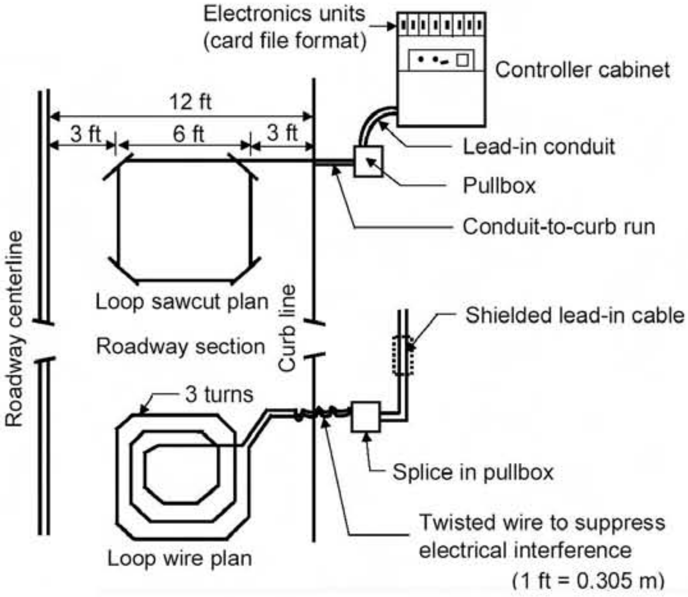


Figure 2.3. Loop detector system: loop circuit and connection with Controller Cabinet

About Loop Data Correction/Cleansing in Microscopic Level

In [51], Karl Petty analyzed the situations of data loss of both upstream and downstream detectors and the mismatch between them which leads to errors in vehicle counts, speed and occupancy. Some preliminary correction methods for post-processing were proposed for vehicle counts and occupancy. The method was to use the data for least square fitting to get the occupancy trajectory over time. Then the incorrect or missed occupancy value could be inferred from the Least-Square fitted trajectory. For count correction, it is a high level approach by using the law of conservation of vehicle numbers in main lanes, on-ramps and off-ramps.

[21] considered event based traffic data validation for 5 different loop detector cards. The purpose is to compare the performance of those cards under similar circumstances using event data including measurement accuracy and flaws such as data error caused by cross-talk. The method is to set up eight criteria for testing the data over 24 hours. Those criteria are based on some common sense of vehicle behavior, loop detection system characteristics, and traffic state parameters. Of the eight test criteria, five apply to single loop detectors and all of them apply to dual loop detectors. Here event data means the lower level data from sensors without aggregation as those of BHL data which has 1 minute update rate with *60Hz* information.

The Advanced Loop Event Data Analyzer (ALEDA) system developed by the Smart Transportation Applications and Research Laboratory (STAR Lab) of the UW, is a plug and play system for detecting and correcting dual-loop sensitivity problems based on loop event data and has been applied for improving dual-loop data [8, 9]. The specific problem focused on there is loop sensitivity. It was claimed that an adaptive method has been developed for changing the loop card sensitivity. In fact, the sensitivity problem only exists for Reno 222 detector card which needs manual adjustment of the sensitivity to different levels by traffic engineers. For other smart cards such as 3M Canoga car or IST card, the sensitivity is automatically adjusted. This has been proved by some experiments conducted in this project.

Summary: This is the only level at which traffic engineers could conduct direct loop fault detection and isolate the loop faults with possibly other system faults such as data loss/pollution through wireless communication. Data in this level can be either data processed by a loop detector card, which will be loop ON/OF time instant or occupancy; or the raw loop pulse signal before the loop card. The main characteristics of those data are that (a) they do not pass any communication media and thus there is no possibility of a communication fault which usually pollutes or loses the data stream; (b) all the raw information is available if a proper interface with the control cabinet is available; (c) real-time data are available; and (d) most importantly, ground truth could be obtained at this level. Thus loop fault detection could be conducted through comparison between the loop detector reading and the ground truth.

Loop faults to be look at the microscopic level include any loop card faults: *mis-assignment, temporary data missing, crosstalk, no data or constant for a period of time, broken cable, chattering, card broken, card sensitivity too high or too low, pulse broken, mismatch of ON/OFF time instant between upstream* and *downstream loops for dual loop stations*.

Methods used at this level for direct loop fault detection include: using 60 Hz data, using pulse signals which bypass the loop card. Systematic data correction/imputation in this level has not been well-developed and documented yet.

It is our opinion that high quality of data is absolutely necessary for all levels of application in Active Traffic Management and Planning. Such data would require high quality basic sensor data and a reliable communication system. Data quality from sensors depends on two factors:

- sensor detection system working condition
    - normal mode

    - error mode
        - persistent error caused by sensor system fault
        - intermittent error caused by sensor measurement noise
        - external uncertainty such as vehicle relative location to the sensor
- sensor characteristics
    - sensor measurement noise/error compared to ground truth
    - physical mechanism and its limit
    - environmental condition effect

To have an accurate and reliable fault detection system, it is imperative to develop and build the following functions in the system:

- automatically detect/isolate the detector fault and its location as quickly as possible
- automatically correct the error by some software approach if possible
- establish a systematic sensor detection system maintenance regulation and implementation mechanism
- field operation to detect and correct those errors that could not be detected, isolated, or corrected only from data side and by software.

Besides, it is also noted that any data processing method for estimating the relevant traffic parameters from the basic sensor data also affect the estimation error, which traffic engineers should also pay attention to.

In this sense, data quality largely depends on the quality of lower level sensor data. If the basic sensor data reading has a high quality, and if the data processing for traffic state parameter estimation in all the application levels are appropriate, high quality application data can be generated. The sensor detection and data correction strategies developed in this project are for this purpose.

2.3 Systematic Loop Fault Detection

2.3.1 Some Consideration on Loop Fault Detection at Macroscopic Level

The PeMS system devotes a large amount of effort to identifying bad detectors and then presenting those detectors to users. The algorithm that is used here has its historical roots in the work of [3]. The basic idea of the diagnostic scheme is to identify bad detectors based on the observed measurements of volume and occupancy over an entire day. This is done by computing summary statistics from the flow and occupancy measurements every day.

It is important to note that we are attempting to identify bad detectors, not bad data samples, which is the intention of higher level detection. PeMS performs a number of simple filters on individual data samples just to make sure that they make sense as a measurement (e.g., that there aren't any values less than zero, that the values can fit into database fields, etc.). However, it doesn't perform real-time checks to verify that each individual data sample conforms to traffic flow theory (e.g., that the combination of measured flow and occupancy fit into some space on the Fundamental Diagram). This approach is based on the observation that when detectors are broken they stay broken.

In principle, the diagnostic tests can be conducted on all detectors in the PeMS system including on and off ramps. The types of tests are different for each type of detector. For example, most mainline detectors report flow and occupancy whereas ramp detectors usually only report flow. Only a subset of the tests can be conducted for ramp loops.

The tests that are applied to the data of each detector at the end of the day are listed in Table 2.1. The tests are conducted in sequence as in the list. Most of these tests are applied to the 30-second data but some are applied to the 5-minute data. For all of these tests we test whether there are too many samples that match a certain criterion. We always define *too many* relative to the maximum number of samples collected by any detector during the day. This way if the data feed cuts off in the middle of the day then we'll still be able to apply the tests consistently. We indicate whether the test is applied to mainline, ML, or ramp detectors, RM, (by *ramp* detectors we mean all non-mainline, non-HOV detectors).

Table 2.1 Higher Level Loop Fault Detection Test at PeMS Level Using Aggregated Data to find Suspicious VDS

Test Num	Det Types	Condition	Description	Diagnostic Test	Data Used	Diagnostic State
1	ML, RM	Never receive any data samples	We break down this condition into three bins based on the communication infrastructure. The first bin indicates that none of the detectors on the same communication line as the selected detector are reporting data. Note that information about communication lines is not always available. In this case, this test is omitted.	Number of samples received is equal to zero for all detectors on the same communication line.	30-sec	Line Down
			The second bin indicates that none of the detectors attached to the same controller as the selected detector are reporting data. This probably indicates no power at this location or the communication link is broken.	Number of samples received is equal to zero for all detectors attached to the controller. If communication line information is available, then at least one other controller on the same line is reporting data.	30-sec	Ctlr Down
			The third bin indicates that the individual detector is not reporting any data, but other detectors on the same controller are sending samples. This most likely indicates a software configuration error or bad wiring.	Number of samples received is equal to zero, but other detectors on the same controller are reporting data.	30-sec	No Data
2	ML, RM	Too few data samples	We received some samples but not enough to perform our diagnostic tests. Other detectors reported more samples (so the data feed didn't die).	# of samples < 60% of the max collected samples during the test period.	30-sec	Insufficient Data

3	ML, RM	High values	There are too many samples with either occupancy above 0% (ML only) or flow above veh/30-sec (ramps only). The detector is probably stuck on.	ML: # high occ samples > % of the max collected samples during the test period. RM: # high flow samples > % of the max collected samples during the test period.	30-sec	Card Off
4	ML, RM	Zero occ or flow	There are too many samples with an occupancy (ML only) or flow (RM only) of zero. We're suspecting that the detector card (in the case of standard loop detectors) is off.	ML: # zero occ samples > % of the max collected samples during the test period. RM: # zero flow samples > % of the max collected samples during the test period.	30-sec	High Val
5	ML	Flow-Occ mismatch	There are too many samples where the flow is zero and the occupancy is non-zero. This could be caused by the detector hanging on.	# flow-occ mismatch samples > % of the max collected samples during the test period.	30-sec	Intermittent
6	ML	Occupancy is constant	The detector is stuck at some value for some reason. We know that occupancy should have some variation over the day. We count the number of times that the occupancy value is non-zero and repeated from the last sample (is exactly the same as the last sample).	# repeated occupancy values > 5-min samples.	5-min	Constant

The procedures for practical implementation of the diagnostic algorithms are as follows:

- We compute the statistics needed for the above tests over the time period from 5am until 10pm (we don't want to capture the time period when there are very few vehicles on the freeway anyway) every day;
- For each of the above tests, we check the statistics against the predefined thresholds. If any of the tests for a detector fails, then the detector is declared bad and we stop testing;
- We record that the detector is declared as bad in the database. Users can subsequently view tables of bad loops;
- Once a mainline or HOV detector is identified as bad, we impute data in order to fill in for the bad detector the next day. Note that we do not impute for any ramp detectors;
- It is important to note that we do not identify individual data samples as bad or good. We make this determination on a detector-by-detector basis each day;
- When the data feed fails in the middle of the day, or it has failed for a number of days in a row, and we can't collect enough data to run the diagnostic algorithms, then we copy the detector diagnostics from the previous day. For the detectors that used to be good we simply mark them as good for this day as well. For the detectors that used to be bad we keep them marked as bad but we change their status to *Feed Unstable*. Hence when the feed returns, we are assuming that the previously good detectors are still good and that the previously bad detectors are still bad - the feed failure had nothing to do with the health of each individual detector. We change the status to *Feed Unstable* because we can't collect enough samples to verify the different types of error conditions that were previously assigned to the detectors. So we don't want to incorrectly declare a reason for failure.

2.3.2. Systematic Loop Fault Detection

The enclosed table (Table 2.2) contains some preliminary thoughts about loop fault detection. The table can still be refined and improved gradually in the future. More loop system faults, their detection strategies and algorithms and software to prevent event data quality drop will be added to it with further development of the project. Although those strategies are developed for inductive loop detectors, some are applicable to other traffic detectors.

Table 2.2 Systematic Loop Fault Detection Strategy

Fault Type (Symptom)	**Possible Causes and Detection Methods**	**Recommended** **Algorithm and Software for Er Prevention and Data Correctio**
No/Insufficient data in a region or freeway section	Communication error; Line broken, Packet loss; Detection: TMC/PeMS level aggregated data analysis to investigate: (a) communication problem; (b) to locate a suspicious controller cabinet; If communication error is excluded, field testusing PLFDT at control cabinet;	Automatic communication error detection – adding real-time coun in data packet; Event data quality checking for occupancy and flow;
Synchronized error • District • Freeway corridor	Communication or power outage Detection: TMC/PeMS level aggregated data analysis to investigate: (a) communication problem; (b) power system failure;	Automatic communication error detection – adding real-time coun in data packet;
Inconsistent speed/ occupancies for adjacent lanes persistently, even in night hours	Card quality or different types of card working in the same environment; loop circuit partially damaged; Detection: locate suspicious loops for field visit and investigate with PLFDT at control cabinet	Algorithm and software for data correction taking into account tim day traffic characteristics; Event data quality checking for occupancy and flow based on Fundamental Diagram

No data from some individual loops	Open loop/circuit, wiring, power; missing parts, or disconnected by road service; Detection: locate suspicious loops and field visit to detect at control cabinet using PLFDT to compare loop data with video data; Compare with adjacent loop data;	After connection, event data quality checking for occupancy and flow;

Insufficient data from certain loops	Communication problem; Loop detector problem; or Card sensitivity problem; or loop circuit partially damaged; Detection: TMC/PeMS level aggregated data analysis to locate suspicious loops and to exclude communication error; Using PLFDT at control cabinet to exclude card sensitivity/error/quality problem;	Automatic communication error detection – adding real-time counting data packets; Event data quality checking for occupancy and flow;
Mis-assignment in (a) highway direction; (b) between lanes; (c) upstream and downstream or a dual loop	Wrong internal loop map use in loop card data decoding; Incorrect wiring Detection: TMC/PeMS level aggregated data analysis to locate suspicious loops; Field visit using PLFDT: persistence check of flow from loops in adjacent lanes;	Re-arranging loop map assignment in data reading and logging; Event data quality checking for occupancy and flow;
Cross-talk	Sensitivity; Interference between loop detector cards; Detection at control cabinet using PLFDT: Compare neighbor loop signals	Algorithm to prevent bad effects on data; Event data quality checking for occupancy and flow;
Pulse flickering, Chattering, Misfiring	Loop card might be in ―pulse” mode, ; Loop circuit connection problem; Card quality problem; improper sensitivity setup Detection: Field visit or off-line data analysis using PLFDT for adequately long period of time;	Software prevention: Control cabinet filtering using low pass filter; prediction and duration bound checking; interpolation and extrapolation to smooth up; Event data quality checking for occupancy and flow; automatic sensitivity adjustment;
Inconsistent data quality problem from time to time, and from lane to lane without communication problem	Loop circuit problem: number of rounds may not be consistent (happened at Station 5 of BHL section) Card sensitivity; Card quality problem Detection: at Controller cabinet level use PLFDT; or through offline data analysis; replacing with a good card;	Smart card has solved the sensitivity problem already; Algorithm and software to prevent dat quality drop through prediction/correction/imputation; Event data quality checking for occupancy and flow;

Mismatch between upstream & downstream loops signal for dual loop stations	Temporary mismatch may be caused by: transition, vehicle lane changing; Persistent mismatch may be caused by: card signal reliability problem, sensitivity of U-loop and D-loop may not be consistent; and loop installation difference; loop circuit partially damaged or installed improperly; Detection at control cabinet using PLFDT to exclude the possibility of such mismatch from software (data reading and logging) problem	Real-time algorithm for data correction/prediction/imputation; Event data quality checking for occupancy and flow; To pair U-loop and D-loop ON/OFF times, and to streaming up time sequence of dual loops

Occupancy/duration and flow problem: • Occupancy and flow are mostly zero • Very high occupancy • Constant occupancy and flow	*If temporarily* ➔ Large vehicle tracking*;* Heavy Traffic/incident caused stop *If persistently*➔ Loop card sensitivity, or loop card broken, or loop circuit damaged; Detection at control cabinet using PLFDT to investigate: (a) software (data reading and logging) problem; (b) card quality problem; (c) problem of pulse signal directly from the loop circuit;	Event data quality checking for occupancy and flow; Software to compare with adjacent lan data, and to interpolate between upstream/downstream and/or adjacent lane data to reduce data error; Automatic sensitivity adjustment;

PLFDT - Portable Loop Fault Detection Tool which will be discussed next.

2.4 Developing Portable Loop Fault Detection Tool

This part presents some preliminary results in the research and development of a Portable Loop Fault Detection Tool for use at the control cabinet level. This work is complementary to most previous work focusing on macroscopic faulty loop data. Part of the project is to develop a real-time vision-based vehicle tracking for freeways to be used as baseline measurements to compare with the lower level loop signal for direct loop fault detection. The system is primarily developed for both freeways and arterials. It is composed of a mobile trailer, a retractable pole with a video camera mounted on top to look at the suspicious loop detector on the ground, a computer running the vehicle tracking algorithm, and another computer at the control cabinet interfacing with a loop detector card though its RS232 serial port. Both computers run IEEE 802.11b wireless for information passing and synchronization. A small data packet of the virtual loop information passes from the trailer computer to the cabinet computer through a UDP protocol. In video tracking, a vehicle passing a loop can be considered as the vehicle activating a virtual loop. Information from the virtual and the physical loops are then compared which can be viewed on a visual display. Preliminary tests have been conducted and the results are analyzed.

2.4.1 Development of Portable Loop Fault Detection Tool (PLFDT)

The purpose for developing PLFDT is for systematic loop fault detection and providing high quality event (low level) data. To guarantee high quality event data, the system needs to have the following functionalities:

- To automatically detect and report some higher level problem such as communication and power outages;
- To automatically detect, isolate and report the type of the data error and pinpoint its causes;
- To use developed software to avoid/prevent event data quality drop caused by some intermittent fault(s).

The PLFDT is depicted in Figure 2.4. A loop detector(s) could be identified as being *suspicious* from a higher level data analysis in TMC/PeMS. The suspicious loops will then be diagnosed further using the portable tool. The tool will enable the operator to use independent streams of traffic measurements for comparing with the suspicious loop detector data. This portable tool has been designed to achieve the following objectives:

- determination of the exact fault type and causes in the detection system

- on-site diagnosis of faults including:
    - mis-assignment
    - cross-talk
    - malfunctioning such as misfiring or pulse broken
    - inappropriate card sensitivity settings
    - inconsistence data (occupancy and vehicle count)
    - broken loop circuit due to
        - improper installation
        - road surface maintenance
        - other factors such as fatigue
    - mismatch of U-loop and D-loop data for dual loop stations
- facilitating on-site detector precision evaluation and calibration.

2.4.2 Overall System Structure of PLFDT

The overall system structure is depicted in Figure 2.4.

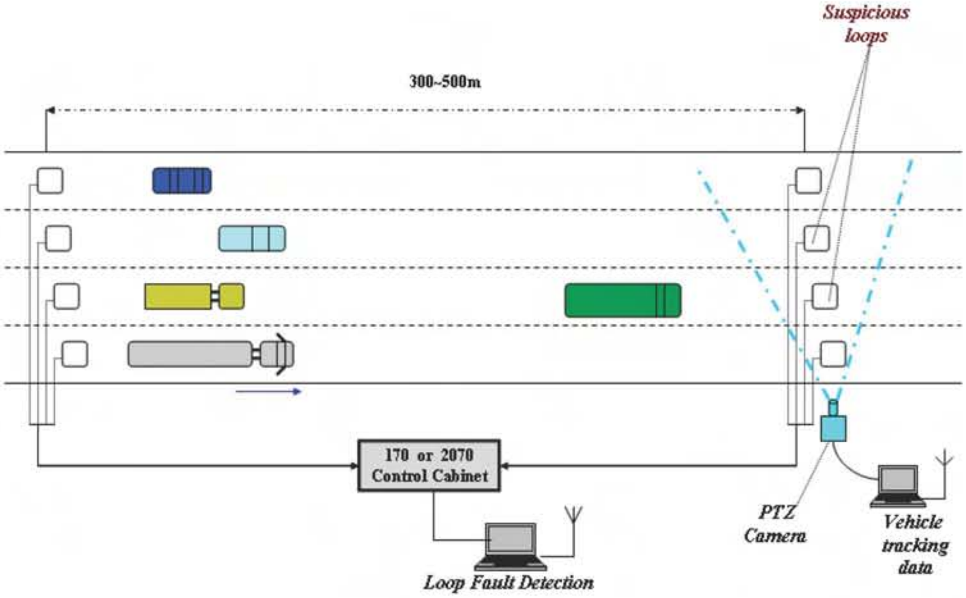


Figure 2.4. Overall PLFDT system structure

The system development includes hardware, software and algorithm development. Our hardware setup consists of the following components (Figure 2.4):

- Mobile trailer which can be towed to the site near the suspicious loop location;
- Retractable pole with a PTZ (Pan-Tilt-Zoom) camera mounted on top;
- Two laptop computers with the Linux operating system;

- *Computer A*: to conduct vehicle-by-vehicle tracking using digital imaging, processed in real-time; and to send out virtual loop activation information via IEEE 802.11b wireless;
- *Computer B*: to interface with a loop detector card for retrieving the event data, receive the video processing data from Computer A through IEEE 802.11b wireless, and compare the synchronized signals for loop fault detection;

2.4.3 Mobile Pole for Roadside Video Camera Mounting

A mobile pole for the roadside camera setup has been developed (Figure 2.4).

Figure 2.5. Video Camera Mounting on Mobile Trailer: Left: Mobile retractable pole; Upper right: PTZ camera on top for looking at the loop and for vehicle tracking to obtain baseline data; Lower right: video computer also run IEEE 802.11b wireless communication via USB port.

The mobile trailer has four retractable folding legs for supporting the platform for leveling. It has several extra supports for robustness if necessary. The mast on the mobile platform can be retracted and folded for easy movement of the trailer. The pole can reach up to 60ft high (Figure 2.5). On the side of a freeway, a camera mounted on top can have a good view angle over 6-lane freeway traffic.

The Pan-Tilt-Zoom parameters can be controlled using the remote controller or using control software running under Microsoft Windows System through the RS 232 serial port interface. This setup process is necessary for the camera to view the loops on the ground and to display on the computer screen so that a virtual loop can be overlaid on the actual loop.

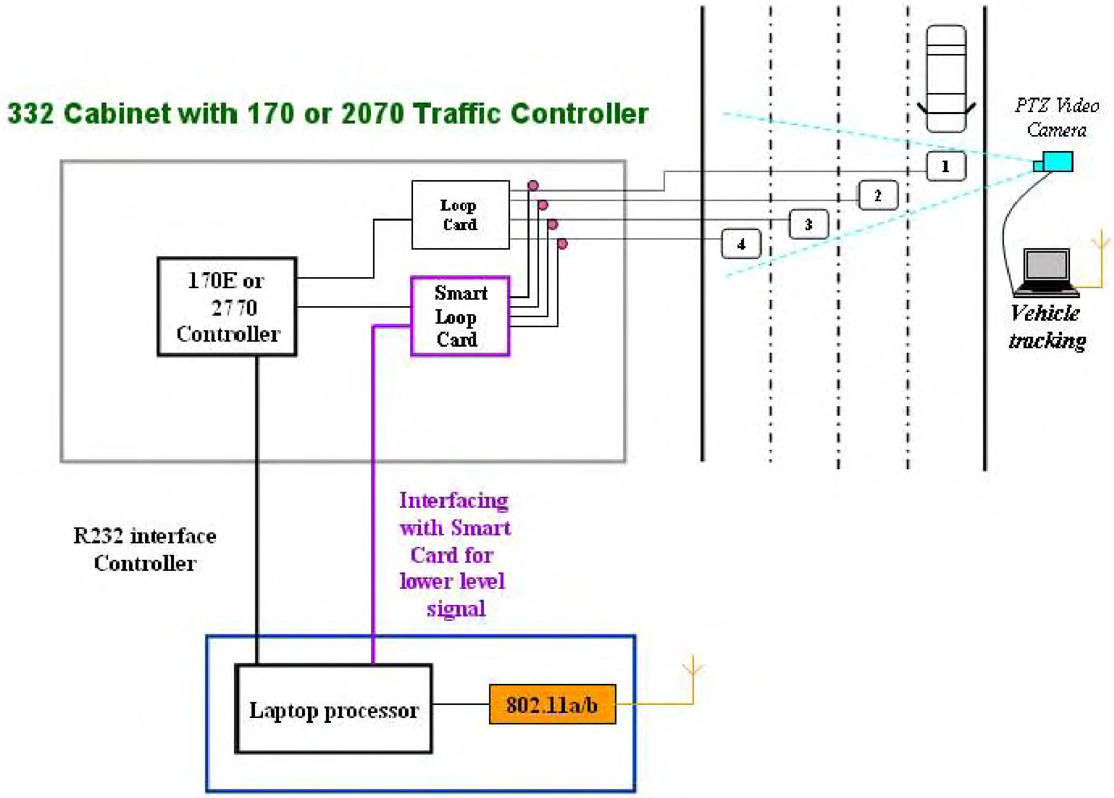


Figure 2.6. Interfacing with control cabinet and smart card

2.4.4 Interface with Control Cabinet

A loop card receives raw analog signals from each loop circuit, processes them with a physical oscillator and amplifier, and outputs traffic signals. Loop cards can be divided into two types: single-layer and multi-layer output cards (Figure 2.6). Single-layer output cards have only two outputs -- the vehicle count (volume) and occupancy – which are results of processing the input signal from the loop circuit. For example, Sarasota GP5 and Reno 222 cards are single-layer output ones that are widely used. There is no direct interface port with these cards. Instead, their signals are directly fed into the controller. The output from the card to the controller is either 1 or 0 without the lower level analog signal available. The low level signal is more attractive than the binary data for several reasons: (a) it can be used for extracting vehicle signatures for re-identification; (b) it tells if the sensitivity of the card is properly adjusted; (c) it tells if any algorithm in the card has a flaw; and, (d) most importantly, we can remove time delay incurred in the traffic controller. Multi-layer output or *Smart* Cards, such as the 3M Canoga and IST cards, have multi-level output information including the start detection times, the occupancy, the vehicle count, fault status, and even the inductance intensity signals calculated from the frequency. A smart card has a built-in RS232 interface port, and thus lower level signals can be obtained. We chose a smart card, 3M Canoga C922, which is compatible with 332 Traffic Control Cabinets and both 170 and 2070 controllers, for our current development. The update rate for data logging from the 3M Canoga C922 card was first obtained as 13Hz, and then at 27Hz after some modifications of software.

A 3M Canoga 922 card was connected to a 322 traffic control cabinet to read the raw loop inductance data directly from the physical loop, as shown in Figure 2.7. It also transfers this loop information to the Laptop through the RS-232 serial cable as shown in Figure 2.8. The 3M Canoga 922 card could read at most two physical loops at the same time.

Figure 2.7. Laptop using RS232 serial interface with C922 3M Canoga Card

Figure 2.8. Laptop interfacing with C922 3M Canoga Card also runs IEEE802.11b wireless communication handshaking with the laptop computer running video camera

2.4.5 Computer Vision System

On the other side, the vision system was set up as illustrated in Figure 2.4. The camera was mounted on the top of the trailer pole to look downward towards the loop detectors at the RFS test intersection. The camera's intrinsic parameters were estimated by using the Camera Calibration Toolbox for Matlab® (http://www.vision.caltech.edu/bouguetj/calib_doc/). The extrinsic parameters are estimated by a simple external calibration algorithm which uses a single rectangle [33]. USB 300mW Wi-Fi adapter with 9dBi and 5dBi antennas were used for reliable wireless communication with about 800m distance coverage.

2.4.6 Real-Time Multi-lane Vehicle Tracking Algorithm

We have designed a computer vision system to obtain baseline measurements to compare. The whole system consists of three parts:

- a camera (Canon VC50i pan-tilt-zoom communication camera);
- a moving platform; and
- a Linux-based video processing software

The Canon VC50i camera provides a wide range of view by panning through a broad fan of 200 degrees, tilting of 120 degrees, as well as a 26x optical zooming. It provides superior camera optics such that good quality images can be obtained even with challenging illumination conditions, such as strong shadow cast that causes too high image contrast. An Intel Pentium Core laptop computer equipped with a USB frame grabber was used for video data processing.

There are many commercial vision-based vehicle detection systems (―virtual loop detectors”) also available. However, most virtual loop detectors are based on the background subtraction algorithm. They normally use frontal-view video images to avoid difficulties caused by occlusions and to get better lane positioning. Since our application requires the camera on the roadside, it is difficult to adopt those systems.

A video processing algorithm has been developed to detect and track vehicles in multiple lanes at the same time. The algorithm combines the background subtraction algorithm with a feature tracking and grouping algorithm to better handle the occlusion problem. The developed algorithm is more robust to shadow and occlusions than conventional virtual loop detectors and, thus, better separates between lanes. An example detection result is shown in Figure 2.9. We see that the upper-left background subtraction result cannot separate the vehicles in multiple lanes but the newly developed algorithm can correctly localizes them. An example placement of a ―virtual loop” over the loop mark on the ground is shown in Figure 2.10. The trajectories of all the vehicles in the image are estimated, and the virtual loop is triggered by analyzing the trajectories.

The software was developed under Linux environment using the OpenCV library. The algorithm runs on a Pentium Core processor (1.83GHz) in real-time at 10 fps. The details of the image processing algorithm are described in [32].

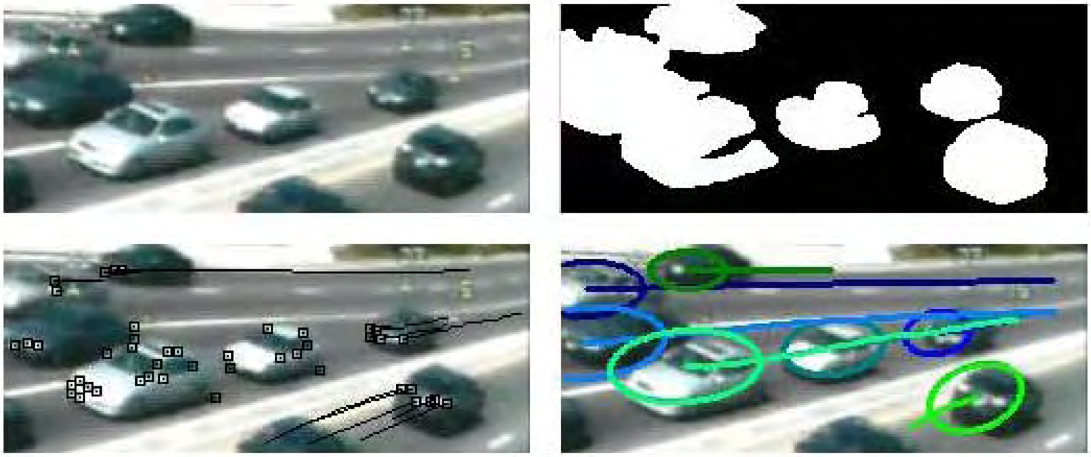


Figure 2.9: An example feature tracking. Upper-left: the original video image; Upper-right: the 'background subtraction' cue where the four vehicles in the left are detected as one big region; Lower-left: the feature detection and tracking cue; Lower-right: result by combining the background subtraction cue and the feature detection and tracking cue.

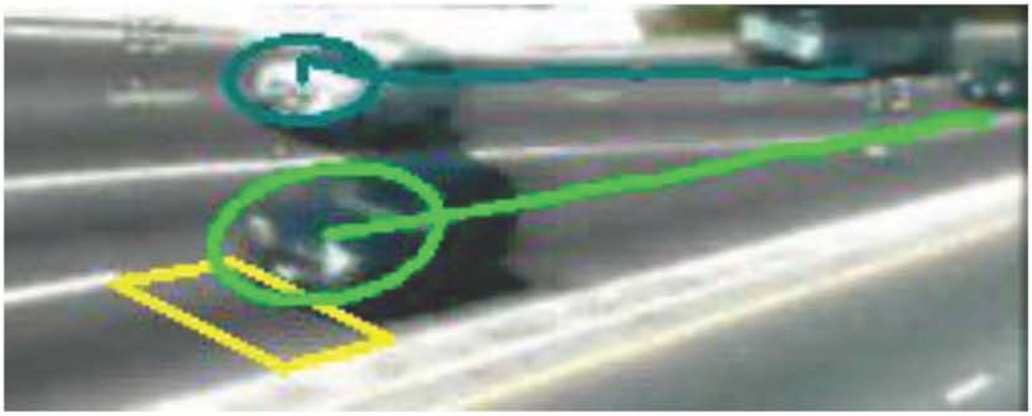


Figure 2.10: The virtual loop is drawn over the physical loop on the scene with video camera calibration tool. Flashing of the virtual loop is triggered when a vehicle trajectory reaches it.

The 10Hz update rate is frequent enough to avoid any missing vehicle count due to high vehicle speed. For example, when a vehicle moves at 70mph or 31.3 m/s, and loop length 2m and vehicle length 4m, the total crossing length for a vehicle starting on upstream edge and leave at downstream edge is: $2+4=6m$. Thus the expected duration is $10.0/31.3=0.192s$. If the video camera update rate is 10Hz, there are at least one or two frames of the video where the vehicle is on the virtual loop. In addition, even if a fast vehicle passes through the virtual loop in between the frames, we can still infer vehicle's passage by analyzing the continuous trajectory that the vision algorithm provides.

2.4.7 Detection Software Development

The software has the following three components:

- High precision synchronization of the timers on the two computers through wireless communication
- Real-time multi-lane and multi-vehicle tracking using the video camera, and
- Matching signals from the two data streams,

which are described respectively in this section.

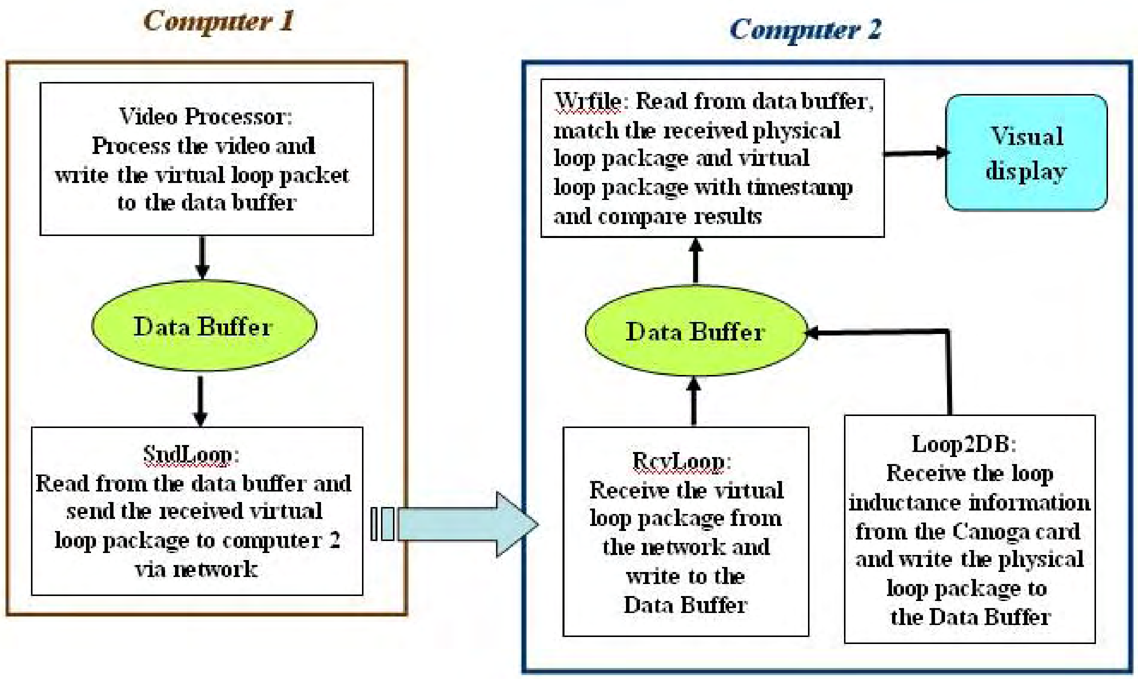


Figure 2.11. Software structure and interaction for the two computer Laptops

2.4.8 Synchronization of the Two Computers with Wireless Communication

The two data streams (from the loop and the video camera) include timestamps for matching. Potential faults are diagnosed by comparing the matched data pairs. The video processing data and the loop data are collected with time stamps in two different computers. Therefore, computer system time synchronization is critical.

We use wireless-based (UDP) synchronization tool developed by the California PATH to synchronize the two computers within 1 millisecond difference. The procedure is described as follows:

Step 1: *Computer A* send a signal packet, *MSG1*, to *Computer B* containing its current system time, say *Start_TIME*.

Step 2: When *Computer B* receives the signal packet *MSG1*, it immediately sends back the acknowledge signal packet, *MSG2*, to *Computer A*, which contains:

a) the *Start_TIME* from the packet it received

b) the current system time on *Computer B* at the time of receiving the packet from *Computer A*, say *Rcv_TIME*

Step 3: *Computer A* gets the acknowledge message, *MSG2*, and marks the current system time after receiving *MSG2* from *Computer B*. Then, the round trip time *R1* of data passing is calculated by subtracting the Start_TIME from the current system time. The clock skew between the two computers are then estimated by comparing *Rcv_TIME* with (*Start_TIME*+0.5**R1*).

Step 4: *Computer A* sends a time setting packet to *Computer B* with the clock skew and *Computer B* adjusts its system time accordingly.

The above process is iterated for 100 times and the average round trip time is used to estimate the clock skew. According to our experiment, the resulting clock skew is far less than 1 millisecond. It is a much more accurate and reliable way to synchronize the two computers than other affordable methods, such as using GPS units.

2.4.9 Comparison of Physical Loop and Virtual Loop

Figure 2.11 illustrates the system structure developed to graphically monitor and compare the loop information. The instantaneous physical loop information and virtual loop information packets are processed and formatted as follows:

Loop Information Package

```
typedef struct{
   double timestamp;
   double Inductance[Max_Loops];
}Loops_TYPE
```

Virtual Loop information package

```
typedef struct {
double timestamp;
double On[Max_Loops];
}Virtual_Loops_TYPE
```

Thus the physical loop packet and virtual loop packet are matched based on the time stamp. However, the packet update rates from the vision system and that for Canoga card are different. The update of information packet for a vehicle over the virtual loop in the vision system is about 10 fps. The maximum update rate from the Canoga card is around 13 Hz. So this is not a one-to-one matching. On the other hand, the messages from both sides could possibly have some delay due to wireless communication or some other unknown reasons. To solve this problem, two Fist-In-First-Out (FIFO) buffers were built on the cabinet computer. One is used to store virtual loop packets from vision system and the other is used to store the physical loop packets from the Canoga card. With those two buffers, two initially synchronized computers can work independently as long as the data is time stamped. Each packet from the video computer (which has a lower update rate) is matched with the physical loop packet which has the closest time stamp by looking up their buffer. This approach significantly increases the reliability of the system.

Currently, the Loop Information Packet includes: time stamp and inductance. Occupancy can be deduced for a given sensitivity threshold. One of the research topics in the near future is to develop an adaptive sensitivity to address the inductance fluctuation caused by changes in temperature and humidity over the loop the road surface.

2.5 Preliminary Experimental Data Analysis

Tests were first conducted at the Experimental Intersection in PATH Headquarters, RFS, U. C. Berkeley, and then on Freeway I-80 at VDS Station 5 and Station 4 of BHL. Figure 2.12-2.13 show the vehicle detection and tracking process. A virtual loop is turned on (as highlighted) when the vehicle ellipse hits the loop rectangle in the world coordinates in the vision system.

We tested the system for around one and half hours at the RFS and all of the vehicles passed through the intersection have been detected from both virtual loop system and physical loop system based on the observation. Note that it is a particularly difficult environment for video processing due to heavy moving shadow of trees shaken by a strong wind. The packet buffer described in the previous section makes the synchronization only have an average error of 0.0436 seconds, which means an error of 1.16 meters in space if the vehicle runs at 60 mps. Figure 2.14 - (a-e) shows the comparison of detections from virtual loops and physical loops when the vehicle is running at different speeds. The red bars represent the virtual loops' on/off

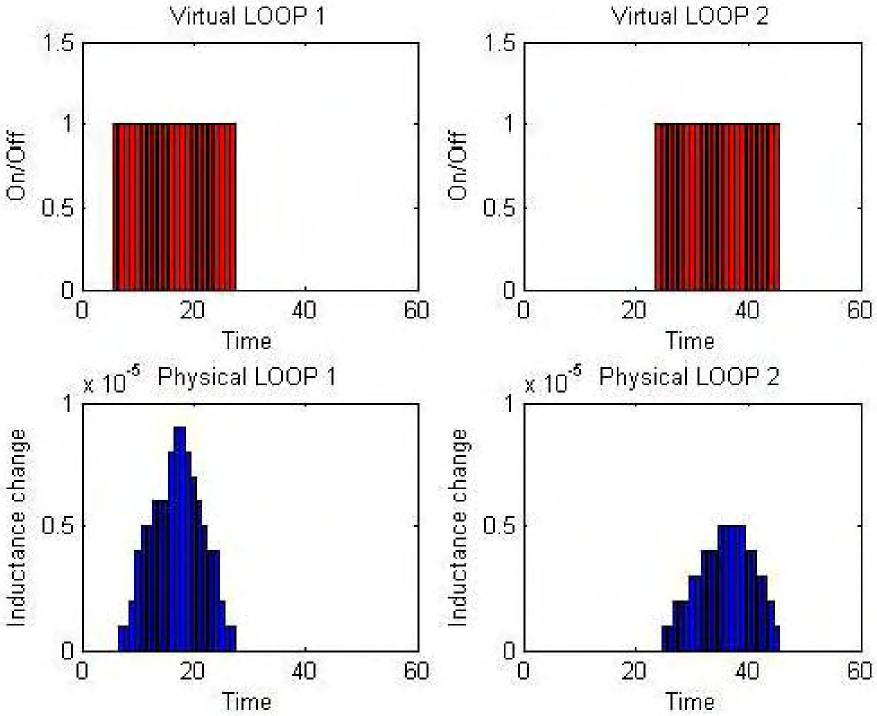
information and the blue bars indicate the inductance changes of the physical loops. Each column shows a pair of results for the corresponding virtual and physical loop. The reference origin time in the four sub-figures is exactly the same. The x axis represents the time domain with the number of packets as a unit. Since the vision algorithm works at around 10Hz, each packet is about 0.1s long. In our experiment, two loops were monitored. The exact physical loop size was 2 meters in width by 1.8 meters in length. Considering the vehicle's physical length, the efficient length of a regular sedan is around 6 meters.

Figure 2.12. Vehicle tracking with virtual loop matched with real loop on the ground

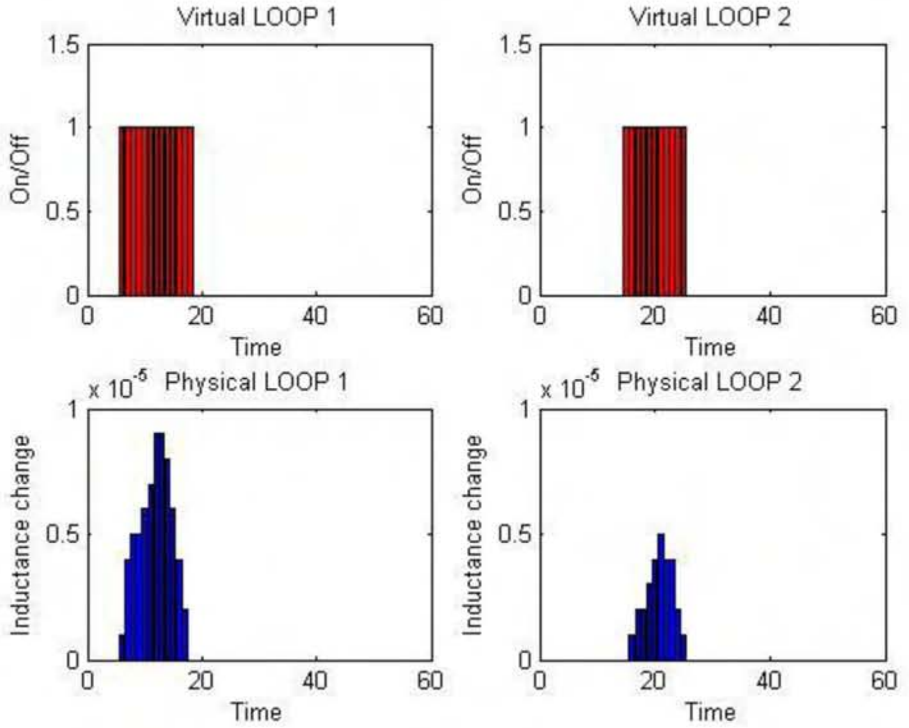
Figure 2.13. Vehicle tracking with virtual loop matched with real loop on the ground

In Figure 2.14 (a-e), loop inductance for different vehicle speeds from the same vehicle are depicted. It can be observed that at a low speed, such as 5 mph in (a), 15 mph in (b) and 25 mph in (c), the physical loop data is bell-shaped, while at high speed, such as 45 mph in (d) and 50 mph in (e), it is shown as a signal pulse. The vertical axis is the inductive intensity variation, calculated based on variation of the pulse frequency of the loop as a vehicle passes over, relative to the inductance in the absence of vehicles. It can be observed that as the vehicle speed increases, the occupancy time decreases. At the speed of 50mph, the duration only lasted for two time steps. Even at higher speed, it is expected that vehicle presence over the loop can still be caught due to high frequency pulse signal of the loop circuit. For the vehicle tracking over virtual loop with video, the time instant for virtual loop ON (with a over it) can still be guaranteed due to continuous tracking in advance.

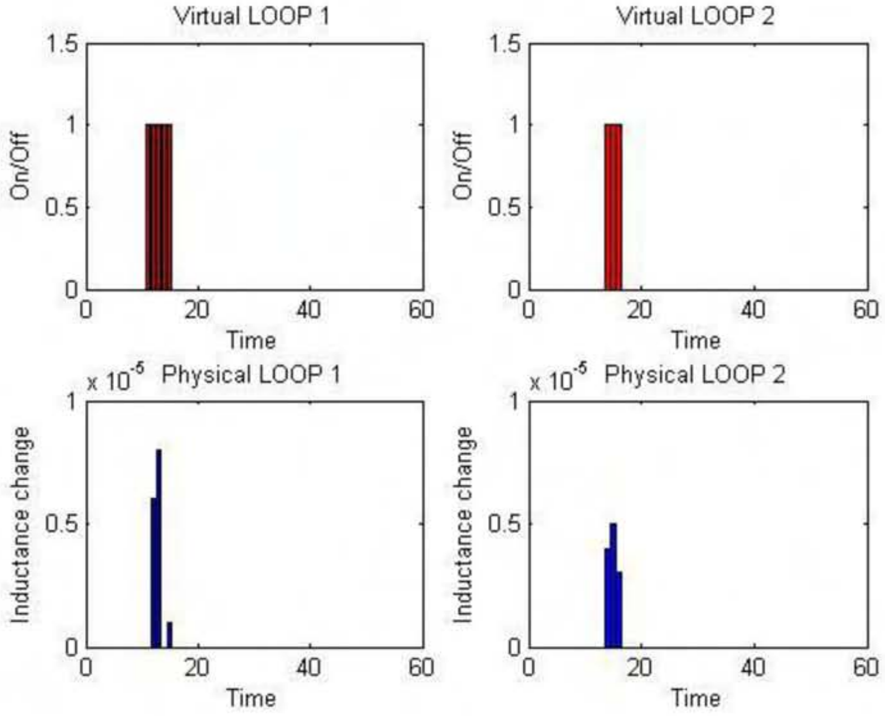
Note that the inductance intensities of the two nearby loops are different even for the same vehicle passing at the same speed. This implies a practical challenge in directly using the inductance data as the only vehicle signature for re-identification over different loops.



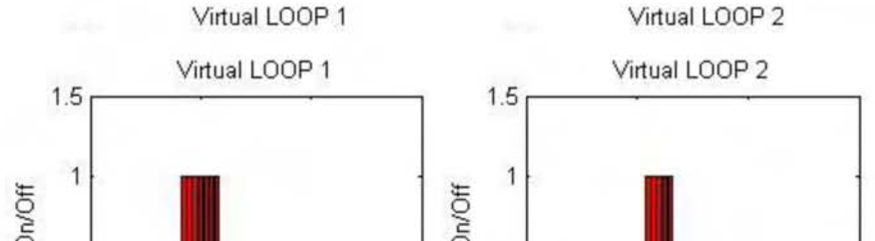
(a) vehicle speed at 5mph

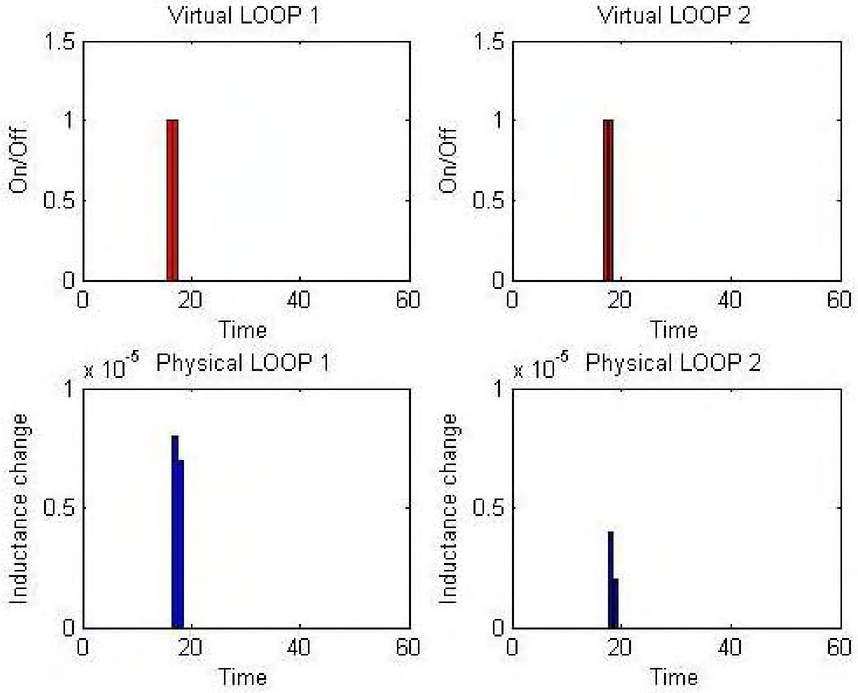


(b) vehicle speed at 15mph
(c) vehicle speed at 25mph



(d) vehicle speed at 45 mph





(e) vehicle speed at 50mph

Figure 2.14 (a-e) Comparison of the virtual loops and physical loops data when the vehicle is running at different speeds. The time duration of the signal in each plot is the duration of the vehicle practically over the loop.

2.6 Fault Detection with Higher Frequency Data

For lower level loop fault detection and vehicle characteristics analysis such as re-identification, it is desirable to obtain higher frequency loop data. The data we obtained from the 3M Canoga 922 Card was about 13 Hz at the very beginning. Only 2~4 sample points are obtained for individual vehicles. With such data, it is difficult to analyze more details regarding the loop detector signal characteristics with respect to the vehicle (length, distance to the ground, number of axles, etc.). Although, some other data logging commercially available tools are available, we intended to develop something which is self-consistent and also maximally uses the information from the loop detector card. Besides, training graduate students to understand the loop detector system and data logging is another purpose. Therefore, we have modified the data reading code which allows us to obtain data with 27Hz eventually, which is much better. With such capability, we conducted two data collection activities in BHL section at Station 5 and Station 4 respectively on June 14 2011 and June 21 2011. The data on June 14th did not have a good quality for both loop data and video data. After modification and preparation, the video data on June 21 had better quality. Through the experiments, it was observed that several loop circuits of Station 4 and Station 5 were damaged. This was confirmed by swapping different 3M Canoga detector cards that were known to be healthy. Besides, further data analysis of Station 4 and Station 5 BHL for March and April in 2011 lead to the same conclusion.

However, due to the leave of PATH staff (ZuWhan Kim) on computer vision, work in this aspect did not go much further than previously planned. However, we did not waste the resources. We have used the resources (1) to develop further the C1 Connector interface with the control cabinet for real-time information retrieving and data logging, (2) to develop more extensive in traffic speed estimation using mode occupancy with event data, and (3) to use the estimated traffic speed for individual vehicle length estimation which can be directly used for vehicle length based classification. The latter two parts will be reported in detail in Chapter 2.

2.7 Development of Interface with Control Cabinet Using C1 Connector

However, most loop detector cards used on freeways are Reno 222 cards. Those cards are designed very simply without an interface port. To solve this problem, a C1 Connector has been modified to retrieve information from the 170 Traffic Controller as shown in Figure 2.15. The C1 Connector has 104 pins which pass signals between the 170 Traffic Control and the cabinet, including processed traffic state parameters such as occupancy and traffic light control signals. However, the raw signals from the loop circuit are not connected to those pins. This cable could also be used in the control cabinet with the 2070 traffic controller if direct interfacing with the controller is prohibited. Through this cable, it is possible to obtain the following information:

- Output of loop detector card: occupancy and flow (vehicle count);
- Output of traffic controller – the control signal to the traffic light of all phases and all movements;
- All the other signals through this link.

We have developed a PC-104 computer with an analog-to-digital (A/D) card to collect all the data and save them in real-time. Through the development of such a system, we understand the input/output signals (analog or digital) and voltages, and how to adapt the A/D card for collecting the data. This system could be used in future phases of the project and other projects when it is necessary to interface with the control cabinet.

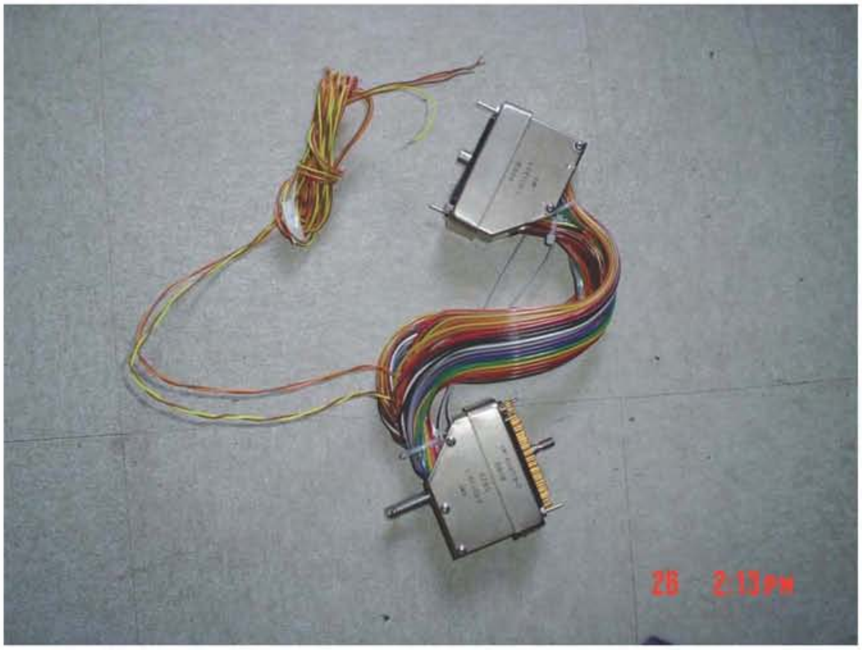


Figure 2.15. Interface control cabinet and traffic controller through the C1 Connector

2.8 Loop Sensitivity Factors Consideration

Besides the sensitivity setting of the loop card, it was believed that vehicle detection may be affected by several other factors: the installation of the loop circuit, road surface temperature and humidity, vehicle types (basically the height of the axles and locations), vehicle speed over the loop circuit. To investigate the functionality of those factors on vehicle detection, we designed some test scenarios and tested using the loop and control cabinet facility at the Research Intersection of PATH at Richmond Field Station.

Tests were conducted to analyze the sensitivity of the inductive loops and the 3M 922 Canoga Card:

- Four vehicle types were used for the tests including:
    - full size passenger car
    - SUV
    - full size van
    - the tractor of Class-8 commercial heavy-duty truck with three axles
- each vehicle type was run when the road surface over the inductive loop was dry and wet (splashed with water while vehicles is running) respectively

- each combination above has been run at speeds from 10mph to 40mph with an increment of 5mph
- each run was conducted in two opposite directions over the loop

Detailed test scenarios are listed in the Table in the Appendix A.

Preliminary analysis showed that, except fotr vehicle speed affecting the occupancy of the loop signal, vehicle types, the number of axles, road surface temperature and humidity did not affect the magnitude and shape of the inductive signal significantly. This finding was rather surprising, which may be due to built-in filtering and/or the capability of adaptive sensitivity adjustment of the 3M Canoga card.

2.9 Concluding Remarks and Further Work in this Direction

We have presented the research and development of a portable tool for systematic loop fault detection at the control cabinet level. Experimental tests up to 50mph vehicle speed demonstrated that this concept is feasible in operating in real-time. Continuous tracking of the vehicles from further the upstream of the loop guaranteed that the vehicle is reliably caught over a loop even at a high speed. An effective and reliable synchronization and data communication scheme was presented. Experimental results showed that the matching of the two sensors was reliable -- it never missed for over 30 tests.

The next step of the research will be in three directions: (a) interfacing with the 170 controller for practical application of PLFDT; (b) testing and improving for multi-lane and multi-vehicle tracking algorithm; and (c) developing output display for automatic and systematic loop fault detection; and (d) implementing a data correction and cleansing algorithm in 170/2070 controller at the control cabinet level.

Chapter 3. Traffic Speed and Vehicle Length Estimation Using Single Loop Event Data

3.1 Introduction to Chapter 3

Traffic speed estimation is critical to traffic operations: Active Traffic Management (ATM) and Advanced Traveler's Information Systems (ATMS) [41]. Traffic speed, flow and density are the fundamental traffic state parameters for most applications. ATM needs an accurate and reliable speed estimation to determine control strategies. ATMS needs link speed to estimate travel time etc. Traffic speed estimation using healthy dual loop detectors can be accurate if event data are used for this purpose, since individual vehicle speed can be obtained based on ON and OFF times of the upstream and downstream loops [15]. However, most loop stations in traffic operation in California are single loops. Updating them to dual loops would be very costly. Therefore, improving traffic speed estimation based on single loops is most important. Although several algorithms have been developed for this purpose and reported in literature, single loop speed estimation is still not good enough. This includes the g-factor method used in PeMS system [29]. Practical application requirement is the motivation of this work.

It is recognized from previous research that vehicle lengths on the highways have special distributions, i.e., most vehicles are passenger cars which have average length of 14~17 ft. Therefore, they generate *mode occupancy (or duration)* on a fixed loop over all the vehicles passing by. An idea that comes up naturally is to use this mode occupancy characteristic for traffic speed estimation. The critical issue is how to properly extract the mode occupancy (or equivalently the mode vehicle length plus the loop length) in a moving window.

The main characteristics and contributions of this study include:

- Directly using the 60Hz information in the event data of Berkeley Highway Lab (BHL) [16] without aggregation, which implies much less time delay is induced;
- Assuming that the mode vehicle length for each polling period is known or can be calibrated for specific road section, which is 15 ft. in California;
- Using variable length time window for mode vehicle data selection to capture traffic transition dynamics and to reduce induced time delay;

- Using mode vehicle length and mode occupancy for speed estimation is much less sensitive to the presence of long vehicles than using the mean values;
- Using the speed estimated from the corresponding dual loop station of BHL for method validation;
- Root Mean Square Error (RMSE) is used to quantitatively measure the discrepancies.

Based on data analysis and comparison with speed estimation from dual loop event data, it has been observed that the performance of the proposed method is about the same level. Quantitatively, RMSE is in the level of 2~3 mph for a healthy station. This indicates that the method could be applied directly to all the single loop detector stations over the highways as long as real-time event data are available without need of any hardware modification.

3.2 Literature Review

There are much literature on traffic speed estimation based on loop detectors. This section only reviews a small portion of it, which are most relevant to this work.

The physical principle of inductive loop detectors is introduced in [31], which is very useful for understanding the sensitivity setting of a loop. Physical characteristics have been discussed in detail in the work of [12] for several loop shapes. The main factors which affect the inductance of the loop include loop shape and size, vehicle shape and size, height from the ground, number of turns of the circuits, and location with respect to the loop. Those factors and the variation of the inductance are very useful for the determination of the sensitivity of a loop. Those facts are important for analyzing the loop data, both single loop and dual loop data for traffic state parameter estimation. Vehicles with high ground clearance (such as trailers) present challenges for accurate detection because loop sensitivity decreases by the inverse square of the distance between the vehicle undercarriage and the loop face. This means that the noise-caused improper sensitivity setting needs to be filtered out in speed estimation based on a single loop. We will show later that this is not a problem for dual loops.

Preliminary time mean speed estimation was proposed by Dailey [22]. A continuous flow model is used to predict the auto and cross-correlation function that will be used to estimate delay time between loops half mile apart. This could be considered as distance mean speed estimation. The work in [15] uses dual loop event data for speed estimation, which has obvious advantage since the on-off time instants of each loop are available. It also implies the advantage of using event data for speed estimation. The work of [23] proposed a statistical analysis method by introducing explicitly the variance of the measurement while deducing the distance mean speed from flow and occupancy. With such a formulation, a Kalman filter is further used to smooth the time mean speeds over time. Coifman [17] intended to improve the effective length for improving the distance mean speed estimation. It noticed that for the same section of the road, different times of day would have different estimation of the length. Work in [6] used aggregated single loop detector for traffic time mean speed estimation.

Hellinga [27] indicated that the former approaches for traffic speed estimation over single loop had RMSE about 23% which was significant. This work intended to improve time mean speed estimation accuracy using real-time estimated effective vehicle length during each polling interval. This work proposed an approach for speed estimation in a situation where single loop and dual loops are mixed. It expected to use the effective vehicle length estimated from dual loop station for single loop speed estimation. The results indicated significant improvement on the speed estimation accuracy by 41% after applying exponential smoothing to flow. It is clear that, it implicitly assumed that upstream and downstream stations had the same effective vehicle length estimation. This is true if the vehicle types are evenly distributed along the link. Otherwise, even with FIFO assumption, a time delay is implicitly introduced since the platoon of vehicles passing the upstream loop will take some time (depending on traffic speed again) to reach the downstream loop.

Wang and Nihan [62] used single-loop measurements to provide speed and vehicle classification estimates. Vehicles were divided into two types (short and long) according to the vehicle length with 11.89m as the threshold. Then the algorithm consists of three steps: to find out intervals (dwell-time of vehicles over a loop in 20s) of long vehicles; to use measurements of short vehicle time intervals for speed estimation with constant vehicle length; and to identify the volumes for the time intervals of the long vehicles using the estimated speed.

Coifman and Dhoorjaty [19] examined new aggregation methods to reduce the estimation errors. The computed the speed from the median occupancy to improve speed estimation to reduce variance of measurement caused by a wide range of vehicle lengths in practical traffic. The work showed that using 5min aggregated data would produce satisfactory results. Hazelton

[26] estimated traffic speed using single loop detector aggregated vehicle count and occupancy with a statistical method.

Rakha and Zhang [55] cleared up some confusion in time mean speed at fixed loop location and space mean speed and provided the relationship between the two. Those differences are important for the type of sensors used for traffic state estimation and also the specific application in modeling, simulation and traffic management and control. This paper also reviewed several other approaches for speed estimation.

All the previous work indicated or implied that vehicle effective length is critical in speed estimation from single loop detector data if occupancy and vehicle count in the polling time interval measurement are reasonably accurate. Therefore, improving the estimation of the effective vehicle length would improve the speed estimation. The problems still remain if we assume the traffic is free-flow and/or most vehicles have the same length for some time interval, since different vehicle types are mixed up in practice, particularly in heavy traffic corridors such as Inter-state 80 in San Francisco Bay Area. How to extract effective vehicle length for more accurate estimation still remains a problem.

3.3 Speed Estimation Algorithm

To address the difficulty that all the vehicle types may appear all the time and traffic flow changes significantly around peak hours, this chapter propose a new approach for speed estimation based on single loop data which has two distinctive characteristics: (a) using event data which distinguish individual vehicles; and (b) using mode occupancy (or duration) in the polling interval since the mode occupancy of CA is known to be 15 ft.. Clearly, only event data could provide mode occupancy. By comparison of the speed estimation from single loop and that from dual loop, the results showed that errors are in the level of 2~4 [mph] in most cases. After analyzing multiple-day data, the results seemed consistent and reliable. This may be partially due to another benefit of using mode occupancy - the variation and noise of the data have been effectively reduced and/or filtered out.

3.3.1 Basic Concepts

The following basic concepts have been used for discussion throughout this chapter:

- Dwell time (duration – ON time –minus OFF time) of individual vehicles over a loop: this is available if the data logging rate from the loop detector is high enough such as that in the BHL archived data which has 60Hz information;
- Moving Window: Traffic changes could be ignored within the time interval of the moving window. Therefore, the moving window length is directly related to the time delay induced in the algorithm: longer window length in time will induce more time delay in the process.
- Mode Dwell Time
  - Put all the dwell times of individual vehicles obtained in the moving window into N_b evenly distributed bins;
  - Choose the bin with the largest number of deposits, which corresponds to the mode duration;
  - Average the dwell times of all the deposits in the bin to produce the mode dwell time;
- Mode Occupancy: This is defined as the ratio of the mode dwell time and the sample time interval which is 1 [s] for most BHL data.

3.3.2 Algorithm Development Strategies

BHL archived data were from dual loop stations which could be used for speed estimation as the reference. The data are processed and used in the following way:

- Cleaning and repairing archived dual loop BHL data to match the U-loop and D-loop pairs for each recorded dwell time (duration);
- Estimate speed and flow based on cleaned dual loop data;
- Using estimated speed from dual loop data as the ground truth;
- Using single loop occupancy data to develop the speed estimation algorithm with some unknown parameters;
- Calibrating the parameter by matching the estimated speed with the ground truth speed estimated from dual loop data;

3.3.3 Algorithm Development

The algorithm development process includes several steps: preliminary data selection; extracting vehicle mode dwell time (or mode occupancy) from a proper moving window; filtering of the mode occupancy to reduce noise; estimating the speed; and capturing transition phase dynamics of traffic for better estimation.

(1) Speed Estimation Algorithm

The g-factor algorithm below is used, which calculates essentially the space mean speed as indicated in [17]. The g-factor is essentially an average of vehicle length. Using static relationships among distance mean speed, density (or occupancy) and flow does not need the assumption of the Fundamental Diagram relationship. The g-factor method is as follows:

$$v(t) = g(t) \times \frac{c(t)}{o(t) \times T} \tag{3-1}$$

T – - polling time interval

$v(t)$ - space mean speed

$g(t)$ - effective vehicle length plus loop length

$g_m(t) = 21[ft]$ - mode vehicle length plus loop length

$c(t)$ - mode vehicle count during time interval *T*, which is the flow at the loop detector

$o(t)$ - mode occupancy: the fraction of time during this period that the detector senses a vehicle

$dur(t)$ - mode dwell time in time interval *T*

For event data (for individual vehicle) the parameters in (3-1) are replaced with its mode value, and $c(t) = 1$. (3-1) becomes

$$v(t)=\frac{g_m(t)}{o(t)\times T}$$

(3-2)

Since the mode occupancy $o(t)$ and mode dwell time (duration) $dur(t)$ have the following relationship:

$$o(t)=\frac{dur(t)}{T}$$

An equivalent formula to calculate speed from (3-1) by model dwell time can be obtained as follows:

$$v(t)=\frac{g_m(t)}{dur(t)}$$

(3-3)

The formula for point mean speed estimation based on mode occupancy (3-1) and mode dwell time (3-2) are summarized as the following:

$$v(t)=\eta\frac{g_m(t)}{o(t)\times T}$$
$$v(t)=\eta\frac{g_m(t)}{dur(t)}$$

(3-4)

where η is a sensitivity parameter. Using (3-1) to estimate speed would require counting the number of vehicles during (moving window) time interval T, while (3-4) would not have such a requirement. However, it needs event data, i.e. the sampling rate of the loop detector should be fast enough to capture individual vehicle activation. Introducing η as a parameter is based on the consideration that the dwell time (or equivalently the occupancy) depends on the sensitivity selection of the loop detector card and also the installation of the loop. Data analysis showed that the variation of η is between $0.9 \sim 1.0$.

(2) Preliminary Data Selection

The maximum and minimum dwell time (duration) will change according to the traffic and the vehicle type. Due to sensor measurement noise and errors, it is necessary to preliminarily correct or eliminate data error by some simple sensible bound checking. The following upper bound and lower bound for vehicle dwell time over a loop are estimated:

$$\begin{aligned} T_{on}^{\min} &= \left(L_{loop} + L_{veh}^{\min}\right)/V_{\max} = \left(3 + 1.83\right)/31.29 = 0.13[s] \\ T_{on}^{\max} &= \left(L_{loop} + L_{veh}^{\max}\right)/V_{\min} = \left(18.5 + 1.83\right)/2.24 = 9.1[s] \end{aligned} \tag{3-5}$$

The estimation is based on the following considerations: Loop length is $L_{loop} = 1.83[m]$ (or 2 [ft.]) fixed; the minimum vehicle length is assumed to be $L_{veh}^{\min} = 3.0[m]$, and the maximum speed is assumed to be $31.29[m/s]$ or 70mph. the maximum vehicle length is 18.5[m] corresponding to Class A Trucks with a trailer. The minimum speed is assumed to be 5[mph] or 2.24[m/s]. With such estimation, the dwell time is set to the upper/lower bound value if it is over/below it.

(3) To Extract Mode Occupancy

To extract *mode occupancy* or *mod dwell time* (*duration*) from a data within a moving window, it is necessary to determine the following factors:

- How to determine the moving window;
- Size of the bin for mode occupancy selection
- Range of the bin for efficient location and data selection

which are discussed below.

(a) Variable Length Moving Window

Traffic other than Stop& Go: Variable Length Moving Window with Fixed Number of Samples: This moving window is used under the situation other than the congested static state. Such a traffic situation is characterized by relatively low occupancy. Most moving window approach in former works used a fixed time length for data processing. However, for traffic data at a fixed location, the vehicle arrival is stochastic and vehicle arrival at higher frequencies for heavier traffic on average. If a fixed time window is used, it will end up with too few samples for off-peak hours and too many samples for peak hours. The former with too few samples will not be good for mode selection. To avoid this, we propose to select the window based on a fixed number of samples with flexible length in time stretch. e. g., a moving window under consideration is defined as having 100 vehicles passed the loop detector. One could adjust the number in practice.

Another less obvious advantage of using such a variable length moving window is that higher flow will lead to a shorter time window for a fixed number of vehicle passing by. From a filtering viewpoint, moving window with shorter time length would lead to less time delay in speed estimation, which reduces the estimation error caused by the delay when speed changes. Accordingly, it makes the algorithm more sensitive to high flow traffic, which often happens just before breaking down.

Heavily Congested Traffic – Stop&Go: Such traffic characteristics can be identified as very high occupancy and very low flow - the congested static state. It could take a long time for a fixed number of vehicles to pass a loop station. Under such a situation, one could still use Variable Length Moving Window with Fixed Number of Samples as before, or using Moving Window with Fixed Time Length until traffic begins to recover, which is characterized by occupancy decreasing. Then the former method would be more appropriate.

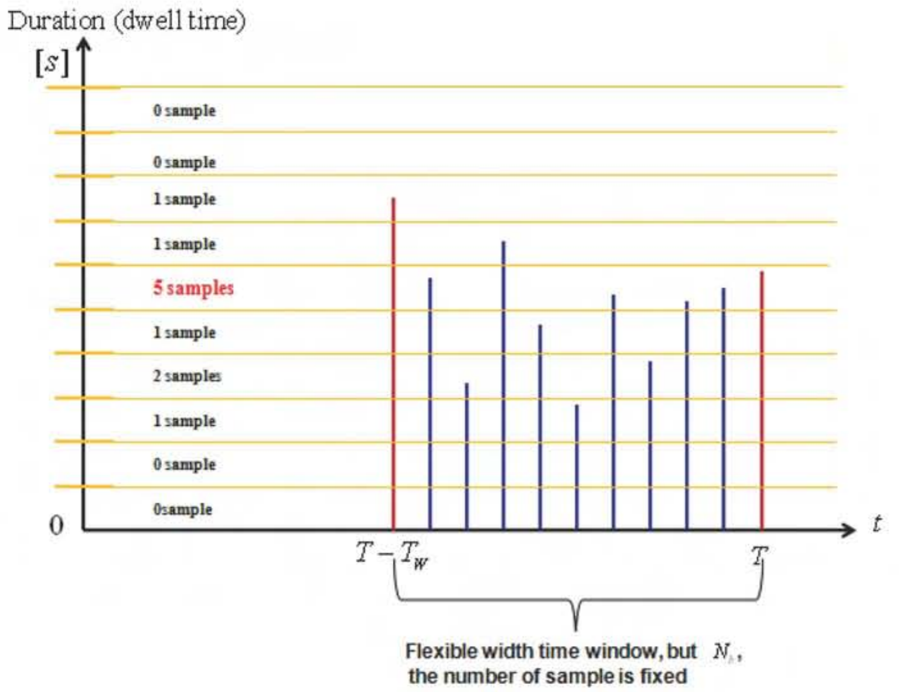
(b) Range of the Bin

It is also necessary to determine the range of the bins for the vehicle dwell time over a loop. It is obvious that such a range should be between the lower and upper bounds identified in each time interval. This dynamic range identification is necessary since the mode dwell time varies according the traffic situation. Within this range, evenly distributed bins are set up to select the mode occupancy.

(c) Size of the Bin

The moving window needs to be divided into bins for mode occupancy selection. The question is how many bins should be used for the moving window. Although the moving window is determined with a fixed number of samples, the size of the bin should also be elastically stretched or reduced accordingly. This can be achieved by evenly divide the moving window along the time horizon. If we decide to have 10 bins, then the number of samples in the bin corresponding to the mode dwell-time should have more than 10 samples. Clearly, too few bins would lead to more coarse selection, but too many bins would not get a significant portion of samples in the mode bin. In practice, bin size could be a considered as a tuning number. Our data processing showed that 20~30 samples/bin will be adequate.

Figure 3.1. Mode dwell time extraction: N_b evenly distributed bins (between horizontal lines); averaging the dwell times (durations) in the bin with the largest number (5 samples in the Figure) produces the mode dwell time (duration).



(4) Filtering the Mode Duration

At each time step, there is a mode occupancy selected. All the mode occupancies together compose a time series. We can remove spikes and noise using rate limiting with a low pass filter. This is necessary to smooth up the estimated speed.

(5) To Capture Traffic Transition Phase Dynamics

In traffic transition phases (with significant speed changes) such as congestion onset with shockwave, the mode occupancy may need to provide quick enough information to determine the speed changing, which is necessary for traffic congestion detection, or there may not be mode occupancy extracted at all from the time sequence.

To capture the transient traffic dynamics, the following 3 values are also monitored from the time sequence data:

- Mean occupancy for all the samples within the moving window
- Mean occupancy for all the samples within the first (earlier) half of the moving window
- Mean occupancy for all the samples within the second (latter) half of the moving window

(6) The Selection of g_m

It is clear that the selection of g_m is very import to the accuracy of the speed estimation. Coifman [17] found that approximately 85% of the individual vehicle lengths observed at one detector station were between 15 [ft] and 22 [ft], and some vehicles were as long as 85 [ft], or roughly four times the median length. For example, on the I-80 BHL section, since it leads to the Port of Oakland, many heavy-duty trucks would travel on that section in Lane 3 and Lane 4. However, we still use the CA State Wide mode vehicle length 15 [ft] in speed estimation.

3.4 Algorithm Validation Using BHL Event Data

Multiple days of BHL-archived 24hr data have been used for algorithm validation. The daily traffic in this section is extremely high in peak hours. After validation, two months of data have been used for reliability test of the algorithm. Data on 04/13/2005 at several stations (1, 4, 6, 7 as shown in Figure 2.2) and data on 03/01/2011 in both directions (EB and WB) are used for error analysis. Each station has 5 lanes. Lane 1 is the HOV lane. Those stations have been selected because the data quality was mostly good except Station 1 Lane 2. Besides, the selection also considers the traffic volume and the recurrent bottleneck in WB at McArthur Maze. Station 7 is close to it.

The fixed number of data samples and the number of bins in the moving window are $N_v = 140,\ N_b = 20$ respectively. The selection of number of bins depends on the duration range in consideration. If the preliminary data selection range is tight, the number of bins could be reduced.

3.4.1 Dual Loop Speed Estimation

Dual-loop speed trap characteristics are used as reference. To obtain accurate speed estimation from the archived dual loop data using the ON/OFF times *(2)* of the upstream and downstream loop, it is necessary to clean up the data first, fill in the gaps with interpolation over time and/or over distance where some upstream and/or downstream data were missing, and then match upstream and downstream data in pairs. One thing worth noting, is that speed estimation with dual loop data is independent from the detector sensitivity level as long as it is the same for U-loop and D-loop.

3.4.2 Discussion

It is interesting to investigate the shift of mode bins for the same station. We use Station 7 WB as an example. For Lane 1 - HOV (Figure 3.2), the Bin 1 always corresponds to the mode duration for low to high traffic volume but not congested, and Bin2 is mode bin for congested traffic, which may indicate vehicle type changes in PM Peak hours. This is similar for Lane 3 (Figure 3.3) except that most mode durations fall into Bin 2. But some fall into Bin 3 in PM Peak hours.

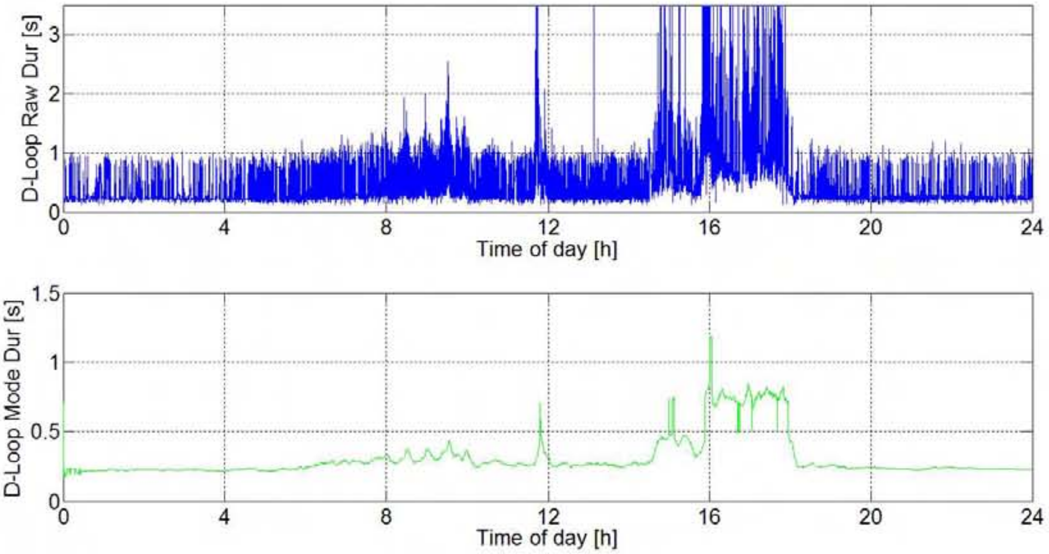


Figure 3.2. BHL data on 03/01/2011, St-7 WB Lane 3 D-loop: raw and mode durations

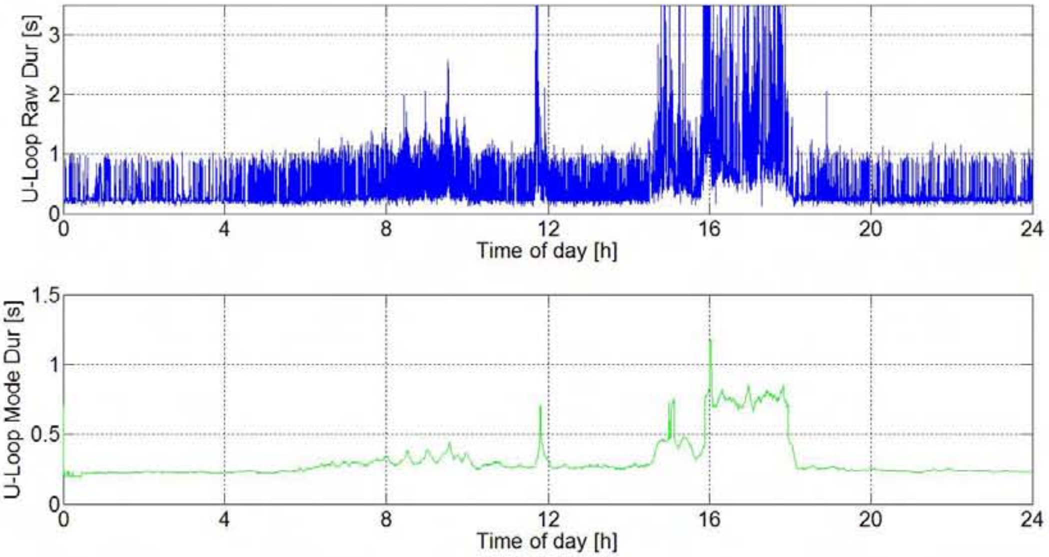


Figure 3.3. BHL data on 03/01/2011, St-7 WB Lane 3 U-loop: raw and mode durations

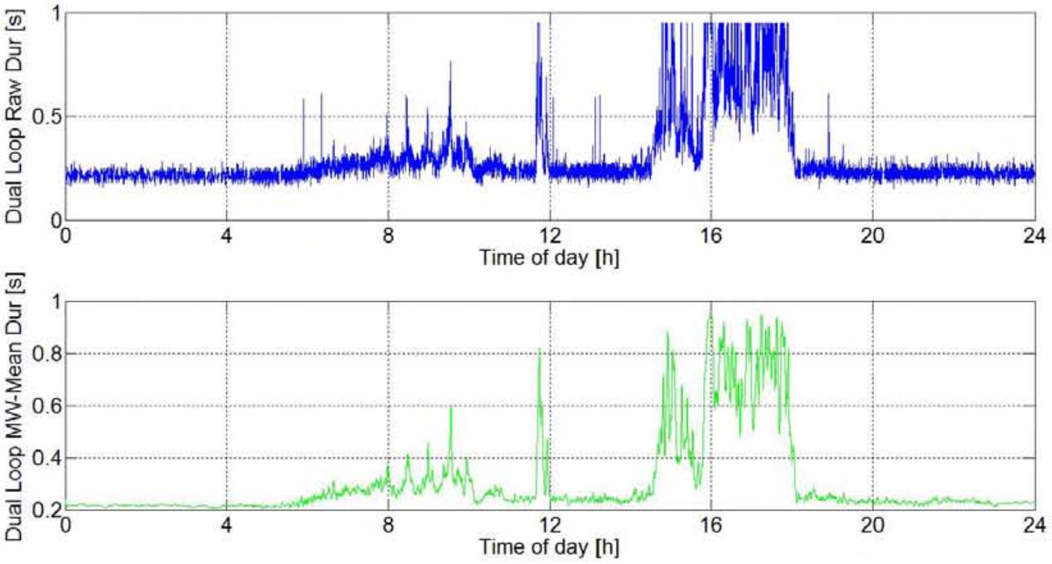


Figure 3.4. BHL data on 03/01/2011, St-7 WB Lane 3 Dual-loop: raw and moving window averaged durations

Accordingly, Figure 3.2, Figure 3.3 and Figure 3.4 show the raw duration and model duration extracted of lane 1, 3 and 5 with respect to the time. It can be observed that they have little fluctuation except in in PM Peak congested hours.

The corresponding speed estimations are in Figure 3.5, Figure 3.6 and Figure 3.7. It can be observed that the speeds estimated for both upstream and downstream single loops are well-matched with that obtained from the dual loop. This is the case even for traffic transition phases and congested situations. It is noted that the speed estimated is not aggregated over time. The same time interval as the event data still remains in the corresponding speed estimated.

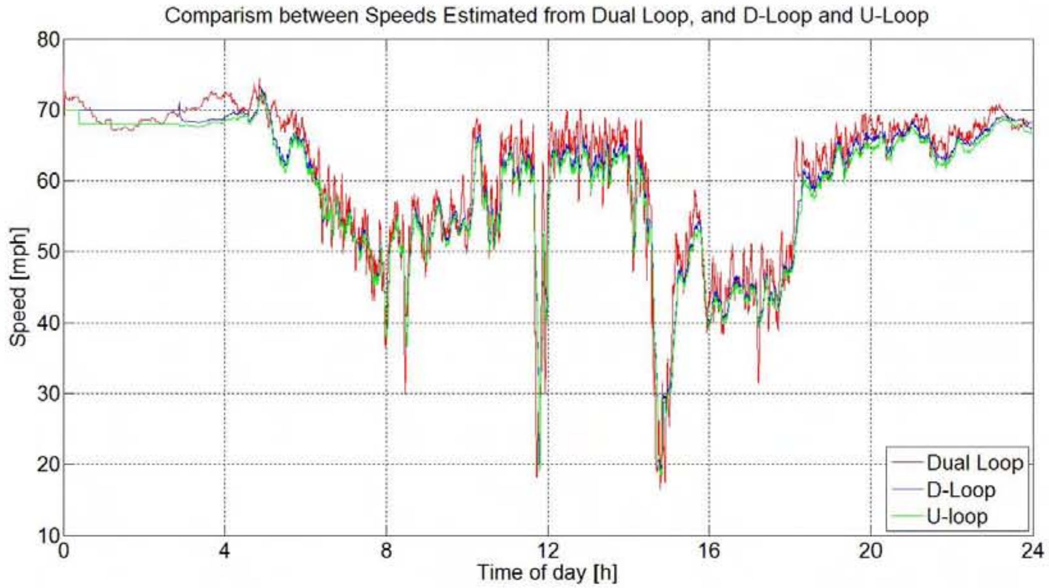


Figure 3.5. BHL data on 03/01/2011, St-7 WB Lane 1: estimated speed comparison

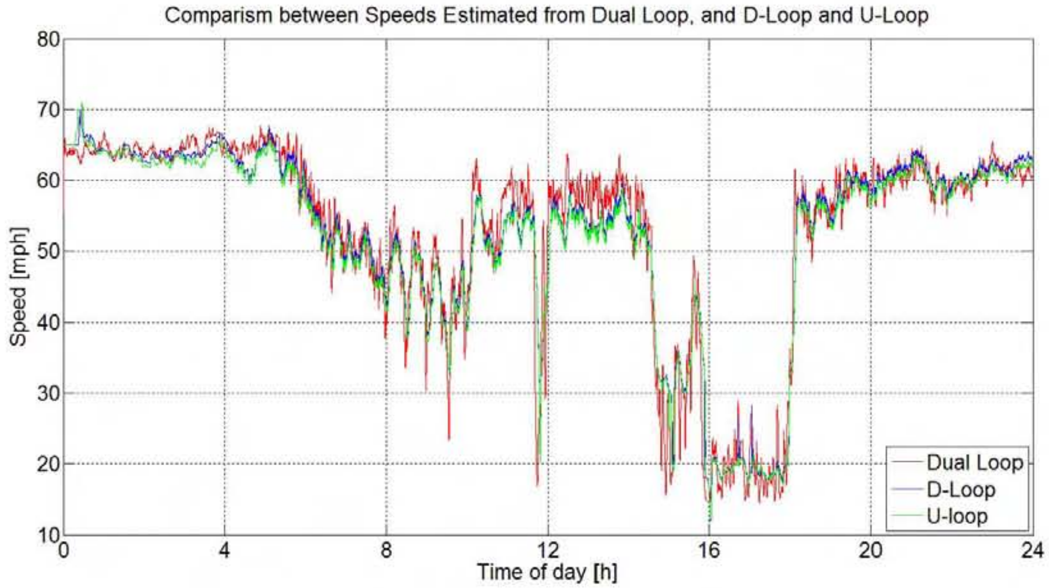


Figure 3.6. BHL data on 03/01/2011, St-7 WB Lane 3: estimated speed comparison

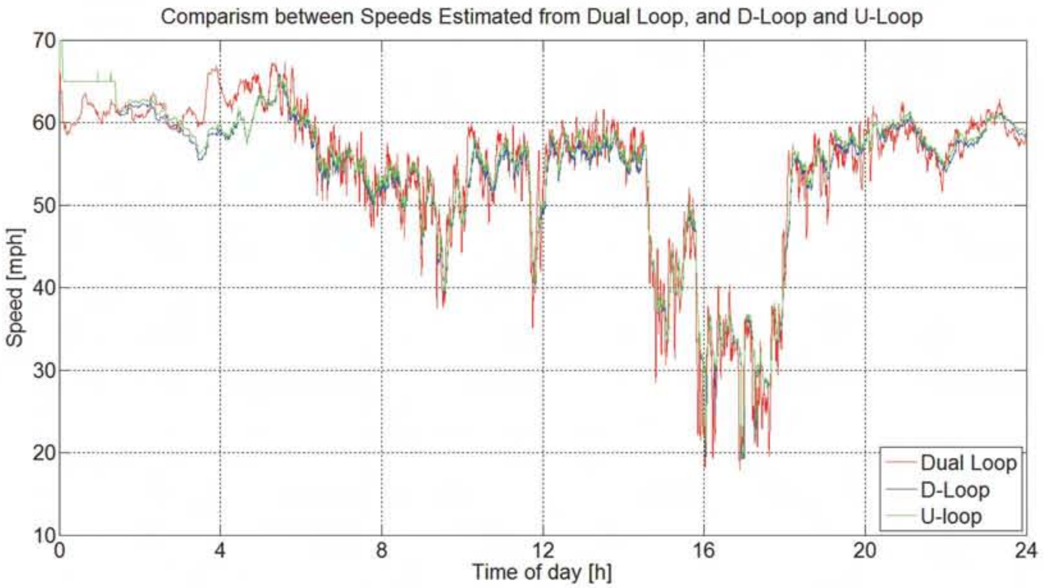


Figure 3.7. BHL data on 03/01/2011, St-7 WB Lane 5: estimated speed comparison

3.4.3 Estimation Error Analysis

Root Mean Square Error in Table 3.1 and Table 3.2 are used to quantify the discrepancy between speeds estimated from upstream/downstream single loops and that from the dual loop. The data analysis has been conducted for 4 stations in both East Bound and West Bound over 24 hours. It can be observed from the table that the RMSE is between 1.73~5.39 [mph], and in most stations/lanes, it is about 2~4 [mph]. The sensitivity parameter $\eta : 0.95 \sim 1.0$. For consistency, we did not adjust the parameter with respect to each lane at a station, but the results seem to be reasonably good. This can be explained as that the loop circuits installation for all the lanes at a station are very similar and all the detector cards of all the lanes at the same station are likely to be set at the same sensitivity level.

3.5 Vehicle Length Estimation

It is well-known that individual vehicle length estimation and speed estimation are equivalent. If one is known, the other is known. Therefore, as a by-product, we can use the speed estimated from the mode occupancy to estimate individual vehicle length based on raw event occupancy data. This process is depicted in Figure 3.8.

Table 3.1. RMSR of upstream/downstream loop single and dual loop speed, [mph]

Data Set and Station	**Dir**	η	**Lane 1**		**Lane 2**		**Lane 3**		**Lane 4**		**Lane 5**	
			D	**U**	**D**	**U**	**D**	**U**	**D**	**U**	**D**	**U**
04-13-2005	**E**	**0.95**	2.52	2.51	2.70	2.49	2.73	2.56	2.91	3.56	3.49	2.41
St-1 (Wed)	**W**	**0.95**	4.14	3.77	**Bad**	**data**	3.15	2.81	2.37	2.40	2.74	2.41
04-13-2005	**E**	**1.0**	3.05	2.94	3.44	3.12	3.96	2.77	2.56	3.85	2.70	2.44
St-4 (Wed)	**W**	**0.95**	2.19	2.12	3.69	3.48	2.22	2.22	2.13	**1.73**	1.79	1.71
04-13-2005	**E**	**0.95**	4.66	2.71	**5.39**	3.01	2.88	2.67	3.57	2.72	3.70	3.03
St-6 (Wed)	**W**	**0.95**	2.65	1.80	3.25	3.49	3.43	2.93	3.97	2.39	3.94	2.71
04-13-2005	**E**	**0.95**	3.10	2.96	6.71	3.32	3.54	2.80	2.66	2.64	3.08	2.72
St-7 (Wed)	**W**	**1.0**	2.64	2.45	4.27	3.70	2.87	2.57	2.95	2.61	2.28	2.50

Table 3.2. RMSR of upstream/downstream loop single and dual loop speed [mph]

Data Set and Station	**Dir**	η	**Lane 1**		**Lane 2**		**Lane 3**		**Lane 4**		**Lane 5**	
			D	**U**	**D**	**U**	**D**	**U**	**D**	**U**	**D**	**U**
03-01-2011	**E**	0.985	2.88	3.03	3.13	3.07	3.23	3.16	3.39	3.49	3.69	3.65
St-1 (Wed)	**W**	0.975	3.95	4.06	**4.58**	**6.49**	3.91	3.97	3.69	3.68	4.00	4.17
03-01-2011	**E**	0.945	4.66	6.06	3.88	3.95	3.73	3.73	3.88	3.83	4.06	4.21
St-2 (Wed)	**W**	0.925	3.83	3.97	4.11	4.21	3.83	3.85	3.80	3.74	3.35	3.39
03-01-2011	**E**	0.900	3.64	3.84	4.11	3.97	4.08	4.00	3.74	3.73	3.39	3.97
St-3 (Wed)	**W**	0.900	3.57	3.53	**Bad data**	**Bad data**	4.12	4.19	4.17	4.09	4.21	4.18
03-01-2011	**E**	0.980	3.67	3.66	3.58	3.55	3.54	3.63	3.61	3.58	3.02	3.093
St-6 (Wed)	**W**	0.950	4.93	4.58	4.38	4.32	3.85	3.84	3.57	3.18	2.83	2.72
03-01-2011	**E**	1.000	2.86	3.31	3.57	3.93	3.26	3.47	2.75	2.75	2.76	2.73
St-7 (Wed)	**W**	1.000	3.76	4.01	4.41	4.23	3.83	3.97	3.08	3.04	2.96	2.97

3.5.1. Algorithm for Length Estimation

The estimation process is in Figure 3.8. The traffic speed is estimated from the mode occupancy extracted from a moving window up to the current time point. The estimated traffic speed has similar time resolution as the event data. Now, we consider the estimated speed as that of every vehicle passing the loop detector. With this in mind, it is possible to estimate individual vehicle speed in principle with the raw occupancy data with the following algorithm.

$$L_s = v_s \cdot dur_s - L_l \tag{3-6}$$

L_s – vehicle length estimated based single loop event data; $s = U, D$ means U-loop or D-loop

v_s – traffic speed considered as individual vehicle speed; $s = U, D$ means U-loop or D-loop

dur_s – raw loop on time duration activated by vehicle i

L_l – length of loop detector in vehicle moving direction

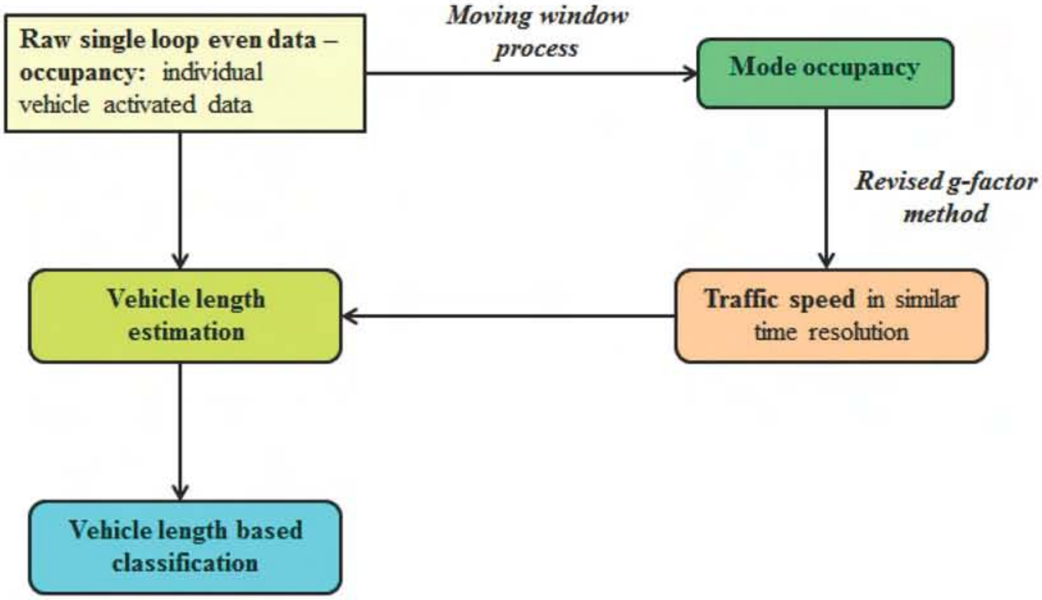


Figure 3.8 Vehicle length estimation and length-based classification using single loop event data

The raw loop on time duration is subject to similar bounding as in (3-5) to reduce measurement noises. Otherwise, there is no other filtering process used.

Using a similar principle, vehicle length is also calculated from the traffic speed estimated from dual loop event data. The algorithm used is as follows.

$$\begin{aligned} L_{dual}^{s} &= v_{dual} \cdot dur_s - L_l, \quad s = U, D \\ L_{dual} &= \frac{L_{dual}^{U} + L_{dual}^{D}}{2} \end{aligned} \tag{3-7}$$

L_{dual}^{s} – vehicle length estimated from single loop event data with dual loop speed; $s = U, D$ means U-loop or D-loop

v_{dual} – traffic speed estimated from dual loop considered as individual vehicle speed

dur_s – raw loop on time duration activated by vehicle; $s = U, D$ means U-loop or D-loop

L_l – length of loop detector in vehicle moving direction

Vehicle length estimation has been conducted for U-loop and D-loop as single loop and for dual loop station. RMSE is used to describe the discrepancy between length estimations of U-loop and D-loop with respect to the dual loop. The results are listed in the following Table 3.3.

Table 3.3. RMSR of Speed Estimation U/D Loop and Dual Loop for Each Lane; [m]

Data Set and Station	**Dir**	**Lane 1**		**Lane 2**		**Lane 3**		**Lane 4**		**Lane 5**	
		D	**U**	**D**	**U**	**D**	**U**	**D**	**U**	**D**	**U**
03-01-2011	**E**	**0.39**	**0.39**	0.56	0.56	0.63	0.65	0.72	0.74	0.71	0.72
St-1 (Wed)	**W**	0.68	0.69	**1.30**	**1.19**	0.71	0.73	0.62	0.69	0.77	0.8
03-01-2011	**E**	0.56	0.67	0.78	0.81	0.77	0.78	0.85	0.86	1.16	1.12
St-2 (Wed)	**W**	0.52	0.53	0.69	0.72	0.76	0.76	0.71	0.71	0.63	0.61
03-01-2011	**E**	0.50	0.46	0.82	0.86	0.80	0.82	0.69	0.70	0.54	0.52
St-3 (Wed)	**W**	0.43	0.41	**Bad data**	**Bad data**	0.79	0.80	0.77	0.84	0.87	0.87
03-01-2011	**E**	0.68	0.68	0.89	0.89	0.88	0.90	0.91	0.96	0.94	0.99
St-6 (Wed)	**W**	0.95	0.97	1.06	1.09	0.89	0.88	0.70	0.73	0.55	0.56
03-01-2011		0.48	0.49	0.96	0.97	0.93	0.92	0.75	0.79	0.74	0.80
St-7 (Wed)		0.75	0.77	1.11	1.16	0.99	1.03	0.68	0.68	0.68	0.66

It can be observed that the RMSE is within 1 [m] in most cases. With such estimation, the estimated three vehicle lengths (U-loop, D-loop and dual-loop) are classified into six categories based on the following length (Table 3.4).

Table 3.4 Thresholds for Vehicle Length Based Classification

Vehicle Class	1	2	3	4	5	6
Length [m]	$1.5 \le L_s < 4$	$4 \le L_s < 7$	$7 \le L_s < 10$	$10 \le L_s < 13$	$13 \le L_s < 16$	$16 \le L_s < 22$

However, such thresholds are quite arbitrary. One could choose other lengths and classes. As an example, such classifications for Station 7, Lane 3, WB (24 hours BHL data on 03/01/2011) are shown in Figure 3.9, Figure 3.10 and Figure 3.11. More such plots for other lanes are listed in Appendix B.

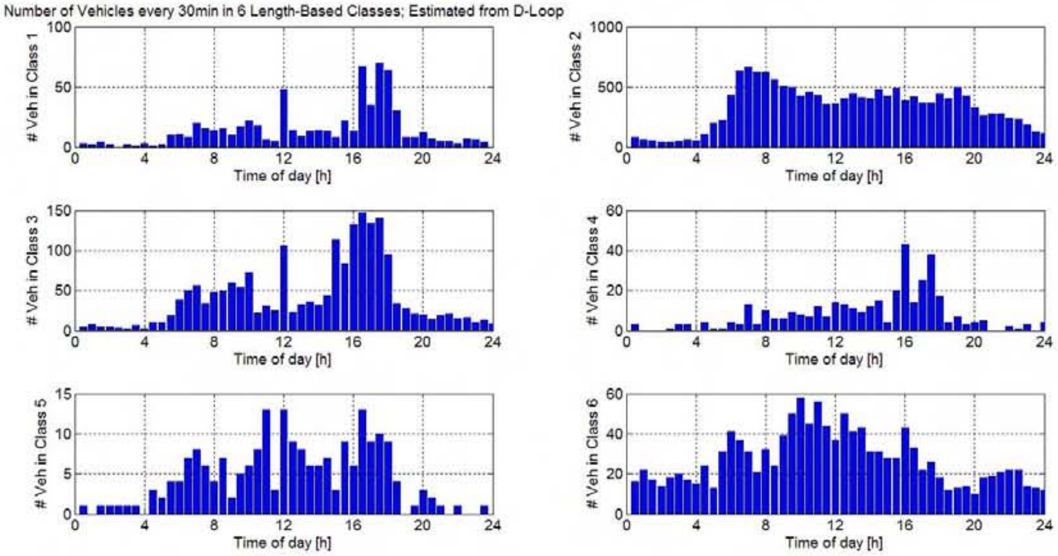


Figure 3.9. BHL data on 03/01/2011, St-7 WB Lane 3, D-loop, vehicle length based classification: # vehicles with every 30min in 6 classes vs. time of day

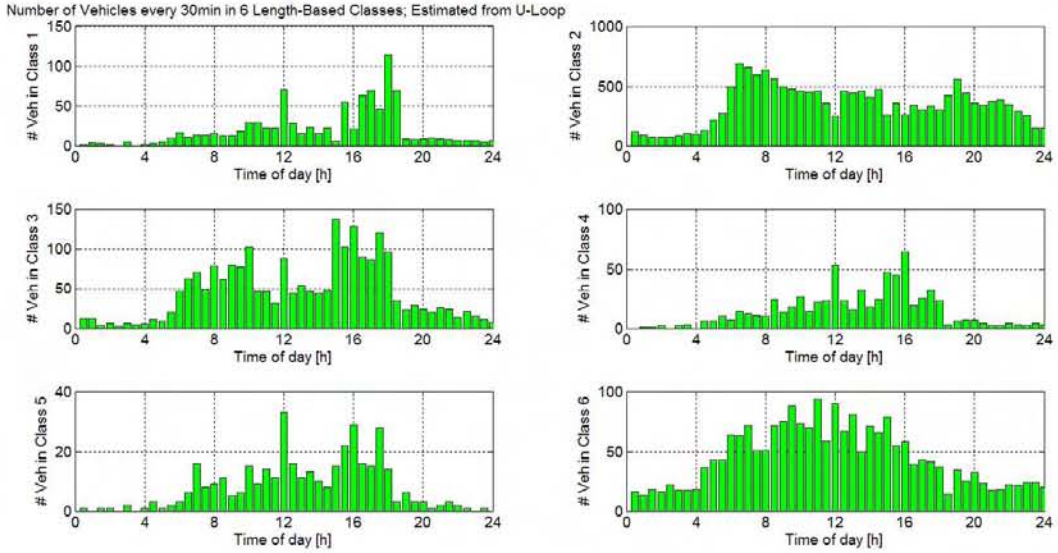


Figure 3.10. BHL data on 03/01/2011, St-7 WB Lane 3, U-loop, vehicle length based classification: # vehicles with every 30min in 6 classes vs. time of day

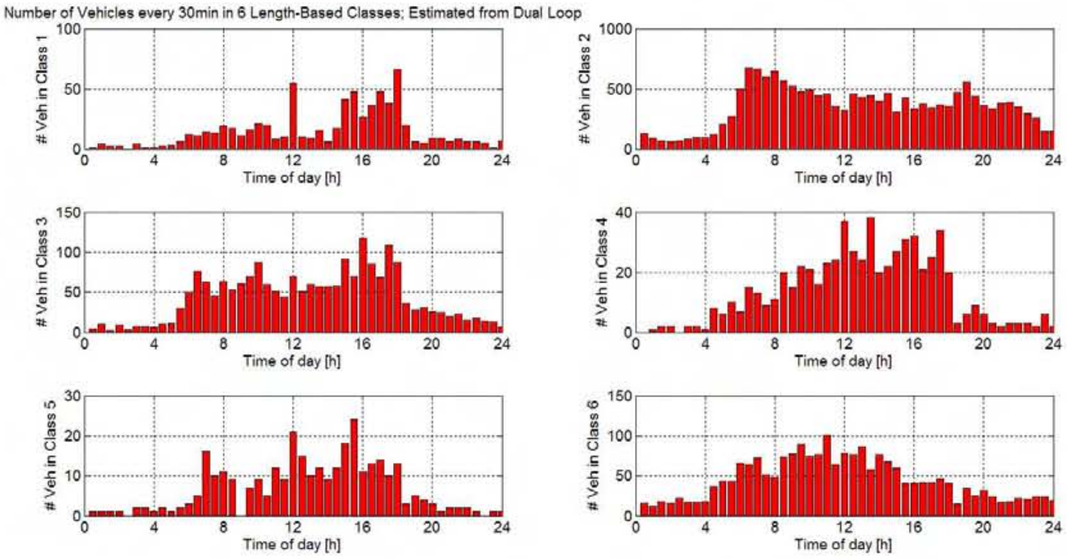


Figure 3.11. BHL data on 03/01/2011, St-7 WB Lane 3, Dual-loop, vehicle length based classification: # vehicles with every 30min in 6 classes vs. time of day

3.6 Concluding Remarks

This work presented in this chapter uses single inductive loop event data with 1Hz update rate and 60Hz information for speed estimation with the g-factor algorithm. The key points are to properly select the g-factor as the mode vehicle length (fixed for specific road) and to extract the mode occupancy (or dwell-time) of individual vehicles over the loop. To capture traffic transition dynamics, for traffic other than congested static state, a moving window with variable time length but a fixed number of samples is used for mode occupancy determination; while for a congested static state, a fixed time length moving window is used.

Multiple-day BHL event data have been used for algorithm validation. Since all the loop stations have dual loops, the speed estimated with the dual loop is used as the ground truth for comparison in evaluating the performance of the estimation with single loop information. The dual loop data have been cleaned up and match the upstream and downstream ON/OFF times before being used for speed estimation. All the speed estimation did not conduct any aggregation which is believed to artificially introduce time delay and therefore estimation error. It showed that the result from a single loop is almost identical to that from the dual loop estimation. RMSE (Root Mean Square Error) has been used to measure discrepancy error.

With event data available, vehicle speed and length are equivalent. Therefore, a direct application of the estimated speed is to calculate the corresponding vehicle length. This has been done for both D-loop and U-loop as single loops and also for the dual loop. The results are compared with RMSE, which is within 1 [m] for most cases. The estimated vehicle length is also classified into six categories in a 30min time window.

Chapter 4. Systematic Event Data Correction for Dual Inductive Loop Station

4.1 Introduction to Chapter 4

Loop detectors currently serve as a major data source for Advanced Traffic Management Systems (ATMS) and the Advanced Traveler Information Systems (ATIS). For example, loop data have been used by transportation engineers to

- measure and enhance freeway performance
- detect and manage incidents for real time operation
- alleviate delays from freeway accidents
- provide necessary information to develop freeway traffic control such as monitoring ramp-meter timing
- arterial and surface street intersection adaptive and coordinated traffic signal control
- classifying vehicles and measuring freight movements for planning
- strategic planning
- tunnel and toll plaza operation

Therefore, loop data accuracy is a key requirement for successful ATMS and ATIS (Figure 4.1).

However, for Active Traffic Management such as demand management, traffic control (Variable Speed Limit or VSL and Ramp Metering or RM) [41], it is necessary to process the data online. For practical use, it is necessary to develop an algorithm that is applicable to such situations. i. e. using the current and historical information from all the inductive loops connected with one control cabinet. Besides, if the traffic data are corrected at the event data level, then the archived data or the data passed to the TMC (Traffic Management Center) will be of higher quality. Therefore, all the data processing for application will naturally be simpler and more reliable. Since the archived data are usually for multiple applications such as traffic planning, higher level operation planning, traveler's information, to name a few, data correction at the event data level will save data correction effort of all the applications and also improves the performance. It is therefore a great advantage to have correct data at the event data level. If the data sent out from the detector terminal is guaranteed healthy, the only possible error or data loss that could be caused is in the communication system for data passing. This, however, can be detected and avoided by adding some simple error checking techniques and communication packet and to adopting more reliable communication protocol such as TCPIP instead of UDP, which will acknowledge the receipt in the receiver end, and if necessary, request for resend.

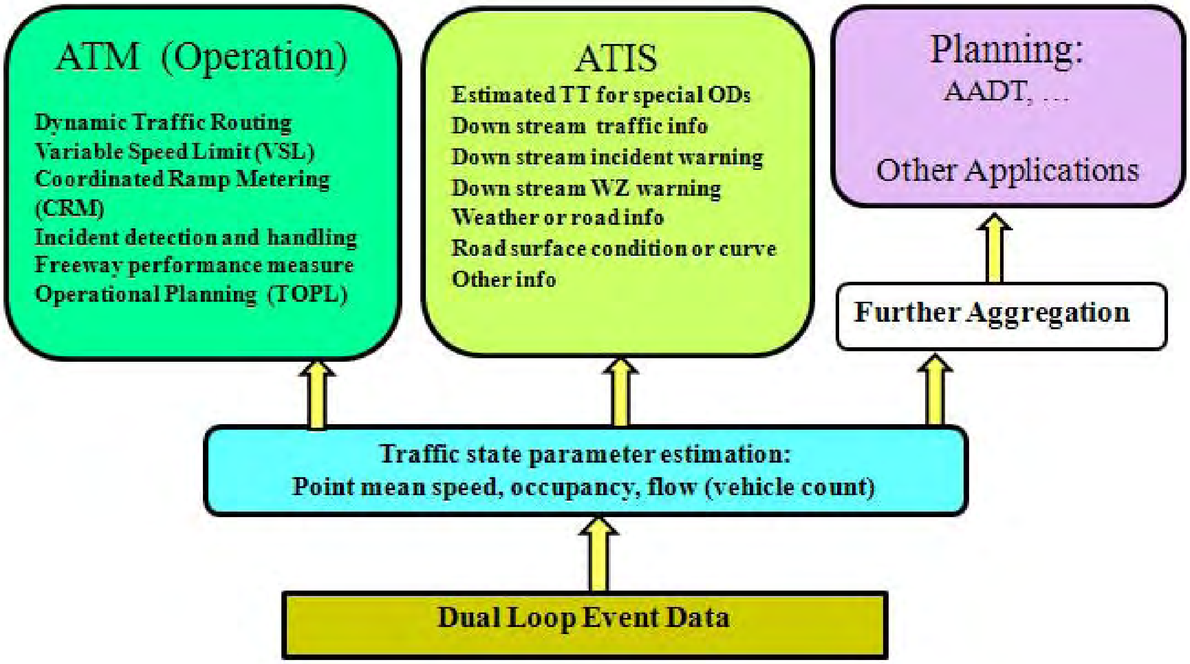


Figure 4.1. Loop Data Application

Inductive loops are the most popularly used traffic detectors in US highways, particularly in California. This is due to the following reasons: (a) historically, it was the earliest available and affordable traffic detector; (b) it met the most requirements of traditional traffic management needs; and (c) it works reasonably reliably if it is installed properly. With the development of traffic management, it is required to have more accurate traffic state parameter estimation. More and more dual loop stations are installed in highways. Accordingly, loop detector cards are also developing into higher level with smarter functions such as flexibility of polling data, multiple choice of direct interface such as RS232 serial connection or Ethernet cable connection etc. However, the data from the loop detector cards is the interpretation of the raw loop signal of the installed loop detector card. Different cards may have different interpretations depending mainly on sensitivity to the pulse variation and sensitivity threshold level setting.

As in other sensors, dual loop data sometimes may have error. Those errors include: incomplete pulse of a single loop, pulse breaking of individual loop, mismatch of pulses between U-loop and D-loop, and improper sensitivities for one or both of them. Loop sensitivity is a generic problem and it affects traffic state parameter estimation. Those problems have encountered and observed in the analysis of BHL archived data as well the collected field data by directly interfacing with the detector card.

This chapter develops algorithms to systematically correct those errors to provide good quality even data. Berkeley Highway Lab (BHL) archived event data and filed collected data in this section are used to explain the problem and to develop the algorithm. The proposed algorithms are simple and therefore easy to implement for processing both archived data and real-time data [42].

4.2 Literature Review

Petty first considered the low level data problem for dual loop stations in [52]. The first problem is the mismatch between upstream and downstream ON/OFF time instant, which the author called ―Transition times”. This problem is shown in the following.

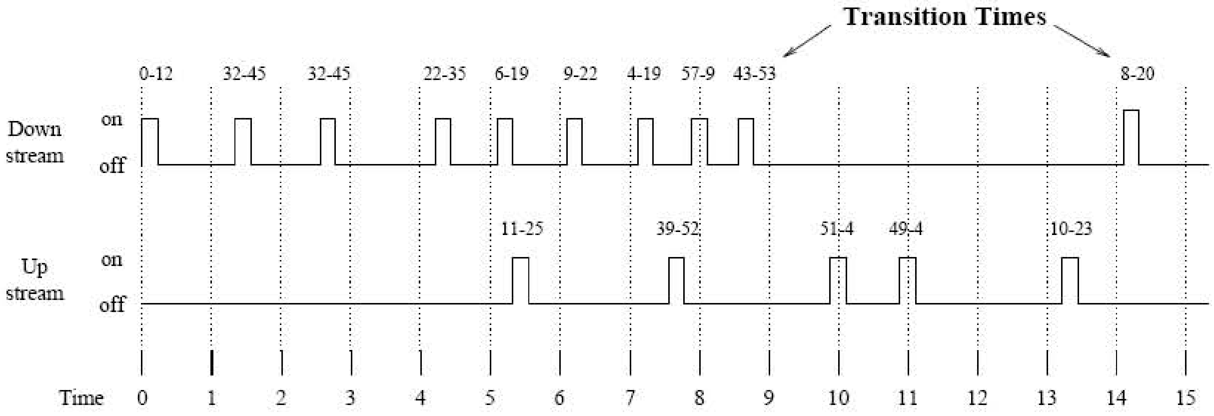


Figure 4.2 Mismatch between upstream and downstream ON/OFF time instant

In Figure 4.2, the controller clock time is divided in second separated with dotted lines and numbered in the lower row. The upstream and downstream ON/OFF time durations are shown with bars. It is obvious that with such mismatch, it would be very noisu using the ON/OFF time instant of the dual loop for speed estimation. This phenomenon was also discussed in [14] where the author suggested a method which matches each downstream pulse to the most recent upstream pulse. With such an assumption, a loop fault cross talk was detected.

As discussed before, some data correction methods proposed in [54] for vehicle count, point speed, and occupancy are from a macroscopic viewpoint and for post-processing instead of real-time. The algorithm presented here can be used for real-time processing since it only uses the information up to current time point.

□Provided some preliminary thought about how to match the upstream and downstream data streams into pairs;

□For free flow traffic, the difference between on-times added to a running distribution for the given lane;

□It recognized the difficulty of matching the data when traffic is not in free-flow;

□It pointed out that Cross-talk could be one reason causing such mismatch.

□This work also proposed a very rough sketch, but not details, as to how to correct the dual loop station data;

The work of Ben Coifman in [14] provided some preliminary thought about how to match the upstream and downstream data streams into pairs. If the vehicle is free flowing, the difference between on-times is added to a running distribution for the given lane. Otherwise, the difference is not included because acceleration, which cannot be measured, becomes a significant factor in the difference. This work recognizes the difficulty of matching the data when traffic is not in free-flow. It pointed out that Cross-talk could be one reason caused such mismatch. This work also proposed a very rough sketch, but not details, as to how to correct the dual loop station data in real-time.

- Free-flow traffic assumed
- Evaluating loop sensor units and detecting crosstalk between sensors: (1) recording a large number of vehicle actuations during free flow traffic, (2) for each vehicle, matching actuations between the upstream and downstream loops in the given lane, (3) taking the

difference between matched upstream and downstream on-times and examining the distribution on a lane-by-lane basis.

- The method matches each downstream pulse to the most recent upstream pulse.

Work in [9] developed an algorithm to identify dual-loop sensitivity problems using individual vehicle data extracted from loop event data

(*a*) removed the sensitivity discrepancy between the two single loops

(*b*) adjusted their sensitivities to the appropriate level

Features of vehicle length distribution are used to find the appropriate sensitivity levels.

[24, 25] presented a detailed physical analysis of four types of loop detector according to their response to different vehicle types, particularly the sensitivity which has been expressed analytically. It helps to understand the physical principle of loop detectors and characteristics of loop detector data. The results showed that the pulse shape of a vehicle depends on several factors: height from the ground, shape of the chassis, material etc.

4.3. Characteristics of Dual Loop Station

4.3.1 Loop Detection Principle, Sensitivity and Sensitivity Level of Detector Card

To understand the characteristics of an inductive loop detector system, it is necessary to know the physical principle of the system. First of all, the loop wires (acting as a solenoid) buried in the ground and the loop detector card form an integrated electric circuit. Without the loop card, the wires buried in the ground cannot do anything. The loop card generates a high frequency signal through the wire in the ground. If a vehicle rides over it, the inductance of the solenoid changes as the permeability of the material for the vehicle, usually metal, is different from that of the air (default – without vehicle over it). Such a change is the detecting principle of the inductive loop detector card. It is clear that the loop detector card determines its output – what the users could see. For Reno 222 card, there is no direct interface with it. To see the output, one have to go through the traffic controller such as 170 or 2070 connected with the control cabinet which is linked with the loop detector card cage. For Canoga C922 or C924 card, one could just directly interface with card through serial interface. With this interface, one can choose to access multi-level signals including the low level analog signal: the frequency and magnitude are measured from the loop wire in the ground, which directly reflects the inductance changes. The detector card circuit design has a sensitivity to such inductance changes, which depends on how the loops are connected with the card. There are two ways to connect: serial connection of the loops is more sensitive to permeability changes over the loop, and parallel connection which will be less sensitive to permeability changes. In principle, more sensitive will lead to higher pulse. The typical shape of a pulse is shown in Figure 4.3 [24].

One could also access the higher level output such as occupancy and vehicle count which are the interpretation of the measured signal by built-in algorithm. Such an interpretation will depend on the sensitivity threshold level (Figure 4.3) setting which needs to be done manually through firmware. How the sensitivity threshold can be set for the loops also depends on how the loops are connected to a card. For a serial connection, the sensitivity threshold setting will be the same level for all the loops. For a parallel connection, the sensitivity threshold can be set differently for individual loops. After manual setting, the loop card does not have the capability to automatically change the sensitivity threshold level.

4.3.2 Data Quality in Sensor Data Level is Essential

For convenience, for a dual loop station, the upstream loop is called U-loop, and downstream loop is called D-loop. Here the sensor level means the output of the data from a detector card. The fundamental data from the loop detector card is the ON and OFF time instant of the U-loop and D-loop. As the BHL section, it is assumed that the sampling rate from the circuit is 60 Hz while the data logging update rate is 1s. Such a sampling rate is believed to be adequate for most traffic applications.

Why Data Correction in Event Data Level?

- All the data applications rely on sensor data quality
- Event data are the closest data and lowest level data from the sensor
- Good event data satisfy the data requirements of all the levels
- Data correction at the event data level
    - reduces data error in higher levels ➔Less burden in data processing
    - Better application parameter estimation
    - Saves resources significantly for all applications

- Systematic data correction needs to be tightly coupled with systematic sensor fault detection

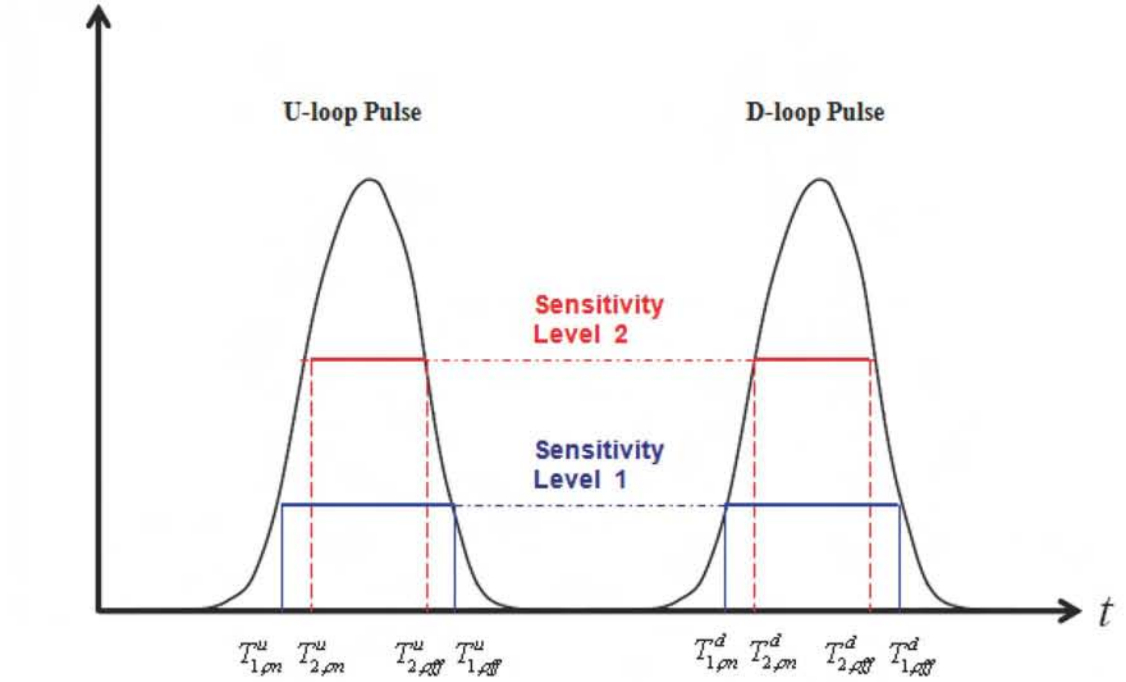


Figure 4.3. Loop signal pulse, sensitivity threshold level setting and its effects on loop ON-time duration and starting time-instant

4.3.3 Main Advantages of Dual Loop Station

To understand the advantages of dual loop stations and limit, it is necessary to know:

- What information the loop detector could provide after detector card
- How the traffic state parameters are estimated from loop data

Event data have the following advantages:

- Speed trap: good speed estimation if sensitivity levels are the same for two
- Redundant flow or vehicle count
- Fault detection of dual loop:

Speed Trap: Dual loop information used for speed estimation is the ON time instant and OFF time instant of the two loops provided that the following conditions are met:

- Each pulse received is complete;
- There is pulse breaking for both U-loop and D-loop signals;
- U-loop pulse and D-loop pulses are well paired;
- Both loop detectors have a similar level of sensitivity;

If the two loops could not provide the ON time instant and OFF time instant information, then speed estimation cannot be conducted. It is clear that, to provide such information, the loop card polling rate needs to be high enough, such as 60Hz for BHL data system. Besides, if the above four conditions are not met, the speed estimation cannot be accurate enough for some applications such as ATM. For dual loop, if the shape of the U-loop is congruent to that of the D-loop, difference between on-times of the two loops will not depend on the sensitivity threshold as long as the sensitivity of the two are set at the same level and are high enough to catch the signal. This fact can be viewed from Figure 4.3. Ideally, assuming the same loop circuit for U-loop and D-loop, then pulse shape and magnitude will be the same for the same vehicle. It is clear that different sensitivity of the detector card will lead to different on-time duration:

$$T^{u}_{1,off} - T^{u}_{1,on} > T^{u}_{2,off} - T^{u}_{2,on}$$

$$T^{d}_{1,off} - T^{d}_{1,on} > T^{d}_{2,off} - T^{u}_{2,on}$$

therefore different occupancy. However,

$$T^{d}_{1,on} - T^{u}_{1,on} = T^{d}_{2,on} - T^{u}_{2,on}$$

$$T^{d}_{1,off} - T^{u}_{1,off} = T^{d}_{2,off} - T^{u}_{2,off}$$

which means that speed estimation is independent from the sensitivity level as long as they are the same for U-loop and D-loop. This is the case since the dual loops are usually connected with a single loop detector card in the field. The only difference is that the speed estimated for a higher sensitivity setting will lead slightly in time over the speed estimated for a lower sensitivity settin from on-time instants of the two loops. We will see later that this condition cannot be met in practice, which means that data correction needs to take into account this factor.

Redundancy Use: It is clear that the vehicle count from the two loops should be very close if not the same. Vehicle count (or flow) is very important for traffic control and ATM, particularly for traffic control. A loop could overcount due to several reasons including: sensitivity setting two low, pulse breaking, and cross-talk. The redundancy could be used for checking such errors and for data correction. Such redundancy can also be used for occupancy estimation. Since density estimation for highway traffic is generically difficult due to low sensor density and a video camera is not always available, occupancy is used as its substitute in traffic control. It is also argued that if vehicle classification is not accurate enough and if the truck percentage is varying, using occupancy for freeway traffic control is more reliable. For dual loop station, the occupancy from the two loops could be used to obtain a better estimation of occupancy similar to the flow estimation. A simple way to estimate one of them is to average the estimation from the two loops.

Fault Detection: Checking each other provides one channel of information for detecting faults from one of the dual loops. Those faults may include pulse breaking and cross-talk, but again, they are sensitivity dependent – it should be set high enough. Those faults could be temporary faults due to system characteristics and flaws or persistent due to some permanent fault or sensitivity settings in the loop.

It is clear that if each loop could only provide 30s data, the dual loop station could not be used as speed trap. Therefore, the advantage of the dual loop is wasted.

The function of dual loops still has limit. As an example, dual loop station still does not bring much more advantage for occupancy and the effective vehicle length estimation due to the following factors [24]:

- Sensitivity of the detector card
- Chassis shape facing the ground;
- The height from the ground
- Algorithm here will not depend on vehicle length assumption

However, the following parameters cannot be accurately inferred from the dual loop still:

- occupancy: average of the two would be slightly better but still depends on sensitivity; if, for speed estimation, sensitivity is set to the same for U-loop and D-loop, then average of the two for occupancy estimation could improve the results;
- average vehicle length: it is also sensitivity dependent, and equivalent to occupancy – if one is determined, the other is also determined;

The occupancy of the effective vehicle length depends on

- Chassis shape facing the ground;
- The height from to the ground
- Sensitivity of the detector card

For vehicle count or flow, too high sensitivity will lead to over count only if the pulse from a single vehicle breaks into more than one. But such over-count would not happen if the signal from one vehicle does not break. This applies to a dual loop station as well.

Most previous work focused on data correction of aggregated traffic state parameter instead of raw data. This approach is good for archived data off-line processing. For archived data, one could use the data from several consecutive stations as long as the data are synchronized such as using the UTC time for the time step.

4.4. Event Data Correction

4.4.1 BHL Archived Data System

To validate the proposed method, the algorithm from Section 3 has been implemented and the practical data from the Berkeley Highway Laboratory (BHL) system (Figure 2.2) have been used. BHL is a test site which covers 2.7 miles of I-80 immediately east of the San Francisco-Oakland Bay Bridge in California. BHL provides event data on individual vehicle actuations, accurate to 1/60th of a second (60 Hz data). Most other loop detector systems collect only aggregated data over periods of 20 seconds or even longer (1 minute, or 5 minutes). Based on the raw BHL event data, accurate aggregated flow and speed information can be extracted straightforwardly. The exact location including longitude and latitude of each loop detector is known, which makes it feasible to know the distance between each loop detector and any vehicle of interest with known location.

4.4.2 Filed Collected Raw Loop Data

A second data set used for algorithm validation was directly collected using the Portable Loop Fault Detection Tool [39] at BHL section Station 5 about 1:00pm on June 9, 2010. The traffic then was almost free-flow. We brought a Canoga 922 loop detector card with a serial interface and inserted it into the control cabinet. The update rate for polling theraw loop signal (frequency and magnitude) was 100Hz. This is the raw information for the pulse generated by the integrated circuit of the loop and the detector card. The data are shown in Figure 4.4. It can be observed that: (1) pulse shape are different from vehicle to vehicle; (2) even serially connected two loop could generate different pulses to the same vehicle; and (3) other temporary data faults may appear such as pulse breaking. It is noted that, all the previous work in [14, 43, 52, 51, 54, 9] did not distinguish the two means of the sensitivity. In fact, those works implicitly assumed that the pulses of the U-loop and D-loop are congruent.

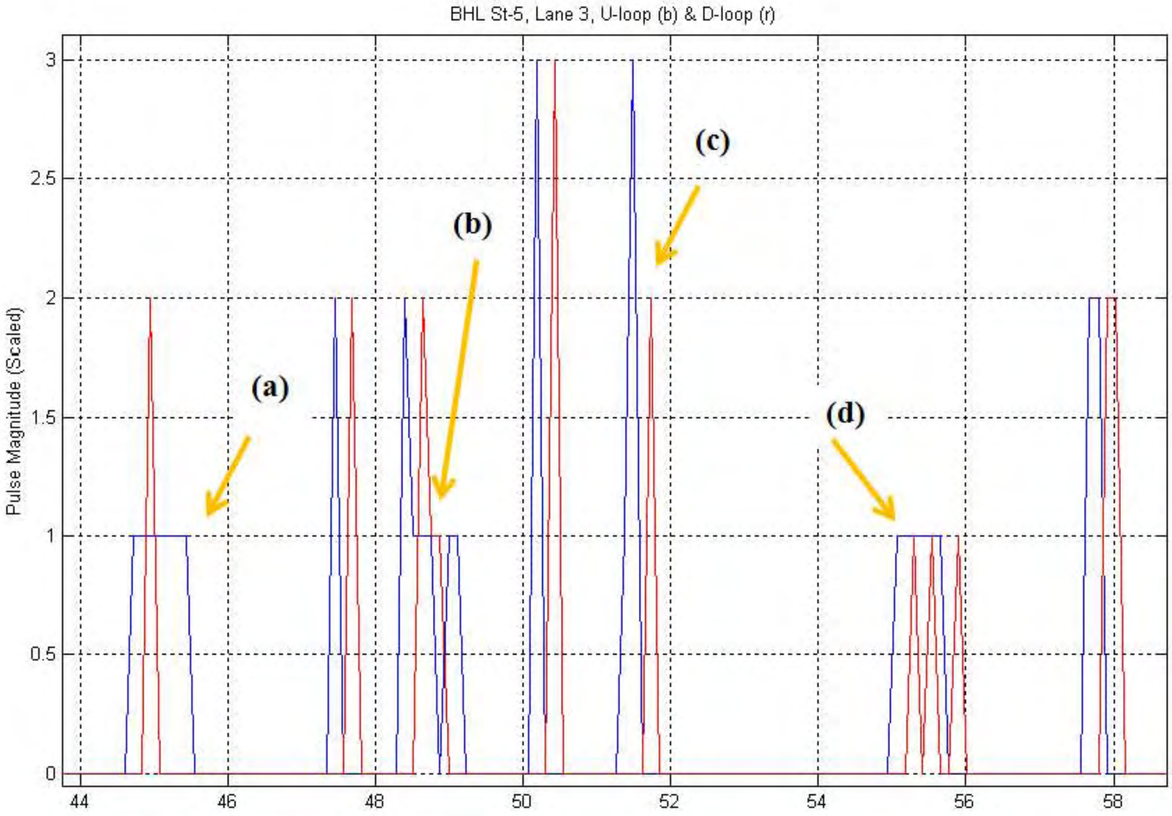


Figure 4.4. (a), (b), & (c) : difference between pulse magnitude of U-loop and D-loop; (b) U-loop pulse breaking; (d) D-loop pulse breaking

4.4.3 Physical Bound Check: for individual loop: upper and lower bounds for pulse length

$$T_{on}^{\min} \le T_{off}^{u} - T_{on}^{u} \le T_{on}^{\max}$$
$$T_{on}^{\min} \le T_{off}^{d} - T_{on}^{d} \le T_{on}^{\max}$$

$$L_{veh}^{\min} = 3m \le L_{veh} \le L_{veh}^{\max} = 18.5$$
$$L_{loop} = 6[ft] = 1.83[m]$$
$$L_{gap} = 26[ft] = 7.94[m]$$
$$V_{\max} = 70mph = 112.63[km/h] = 31.3[m/s]$$
$$V_{\min} = 5[mph] = 8[km/h] = 2.24[m/s]$$

$$T_{on}^{\min} = \left(L_{loop} + L_{veh}^{\min}\right)/V_{\max} = \left(3 + 1.83\right)/31.29 = 0.13[s]$$
$$T_{on}^{\max} = \left(L_{loop} + L_{veh}^{\max}\right)/V_{\min} = \left(18.5 + 1.83\right)/2.24 = 9.1[s] \quad (4\text{-}1)$$

Bound check for whole dual loop station:

It is also possible to consider the dual loop as a whole single loop station with the upstream edge of U-loop and downstream edge of D-loop as the edges. The reason is to fully use the information obtained.

U-loop and D-loop Pulse Pairing Principle: Two pulses satisfying this condition are considered potential pair; otherwise, they are not;

$$T_{St}^{\min} \le T_{off}^{d} - T_{on}^{u} \le T_{St}^{\max}$$
$$T_{St}^{\min} = \left(2L_{loop} + L_{gap} + L_{veh}^{\min}\right)/V_{\max} = \left(2\times 1.83 + 7.94 + 3.0\right)/31.29 = 0.47[s]$$
$$T_{St}^{\max} = \left(2L_{loop} + L_{gap} + L_{veh}^{\max}\right)/V_{\min} = \left(2\times 1.83 + 7.94 + 18.5\right)/2.24 = 13.44[s] \quad (4\text{-}2)$$

4.4.3 Using Progressively Filtered Information:

For a temporary data fault, it is logical to use the filtered information from previous times for data correction. Such information include: on-time difference, off-time difference, pulse width and pulse gap. The justification is that if the traffic is to change gradually including changing from one phase to another, the traffic data should also change gradually. For efficient real-time calculation and minimum computer resource use, a recursive exponential filter is used for the estimation for those parameters. They have the following characteristics:

- Simple recursive algorithm to save memory – only previous measures are used
- Memory of past information diminished exponentially
- Flexible in parameter choice
- Good for traffic state parameter estimation
- Suitable for real-time processing

Filtered average pulse width

$$\begin{aligned}
\Delta_{on}^{u}(k-1) &= \alpha \cdot \Delta_{on}^{u}(k-2) + (1-\alpha)\cdot\left(T_{on}^{u}(k-1) - T_{off}^{u}(k-1)\right)\\
\Delta_{on}^{d}(k-1) &= \alpha \cdot \Delta_{on}^{d}(k-2) + (1-\alpha)\cdot\left(T_{on}^{d}(k-1) - T_{off}^{d}(k-1)\right)\\
\Delta_{on}(k-1) &= 0.5\left(\Delta_{on}^{u}(k-1) + \Delta_{on}^{d}(k-1)\right)
\end{aligned} \tag{4-3}$$

Filtered difference of on-times between U-loop and D-loop

$$\begin{aligned}
\Delta_{on}^{diff}(k-1) &= \alpha \cdot \Delta_{on}^{diff}(k-2) + (1-\alpha)\cdot\left(T_{on}^{d}(k-1) - T_{on}^{u}(k-1)\right)\\
\Delta_{off}^{diff}(k-1) &= \alpha \cdot \Delta_{off}^{diff}(k-2) + (1-\alpha)\cdot\left(T_{off}^{d}(k-1) - T_{off}^{u}(k-1)\right)\\
\Delta^{diff}(k-1) &= 0.5\left(\Delta_{oN}^{diff}(k-1) + \Delta_{off}^{diff}(k-1)\right)
\end{aligned} \tag{4-4}$$

Filtered average pulse gaps

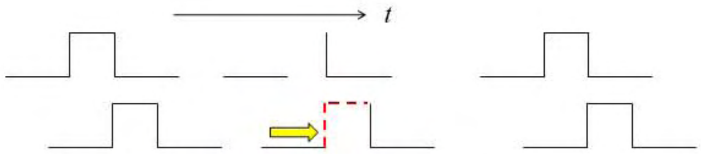
$$\begin{aligned}
\Delta_{gap}^{u}(k-1) &= \alpha \cdot \Delta_{gap}^{u}(k-2) + (1-\alpha)\cdot\left(T_{on}^{u}(k-1) - T_{off}^{u}(k-2)\right)\\
\Delta_{gap}^{d}(k-1) &= \alpha \cdot \Delta_{gap}^{d}(k-2) + (1-\alpha)\cdot\left(T_{on}^{d}(k-1) - T_{off}^{d}(k-2)\right)\\
\Delta_{gap}(k-1) &= 0.5\left(\Delta_{gap}^{u}(k-1) + \Delta_{gap}^{d}(k-1)\right)
\end{aligned} \tag{4-5}$$

Filtered duration of a station: difference between U-loop on-time and D-loop off-time (on-time considering dual loops as a single loop)

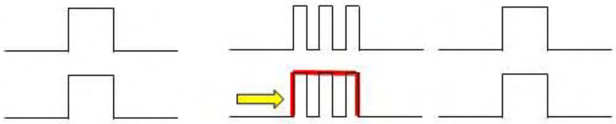
$$\Delta_{on}^{st}(k-1) = \alpha \cdot \Delta_{on}^{st}(k-2) + (1-\alpha)\cdot\left(T_{off}^{d}(k-1) - T_{on}^{u}(k-1)\right) \tag{4-6}$$

4.4.5 Completing ON and OFF Time Pairs for Individual Loops

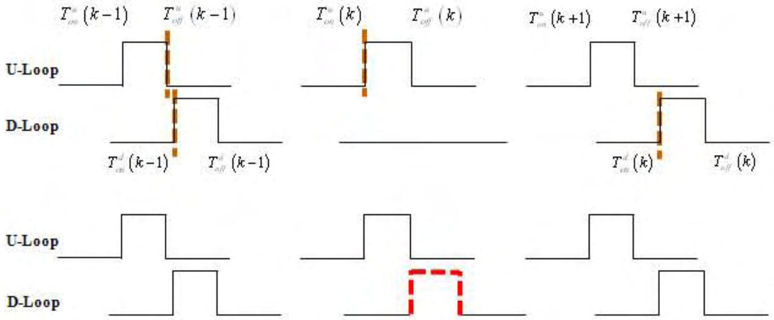
This needs to be conducted for each loop individually. It is necessary to conduct persistent checking for fault detection. If a pulse is over 10% incomplete, it can be concluded that the loop detector card has some loop signal polling problem or data logging problem. The filtered average pulse width is used to complete the pulse as in Figure 4.5 (a).



(a) Incomplete pulse correction for individual Loop



(b) Pulse break correction for individual Loop



(c) Mismatching pulse correction for U-loop and D-loop jointly

Figure 4.5. Schematic plot of data correction

4.4.6 Correcting Pulse Breaking for Individual Loop

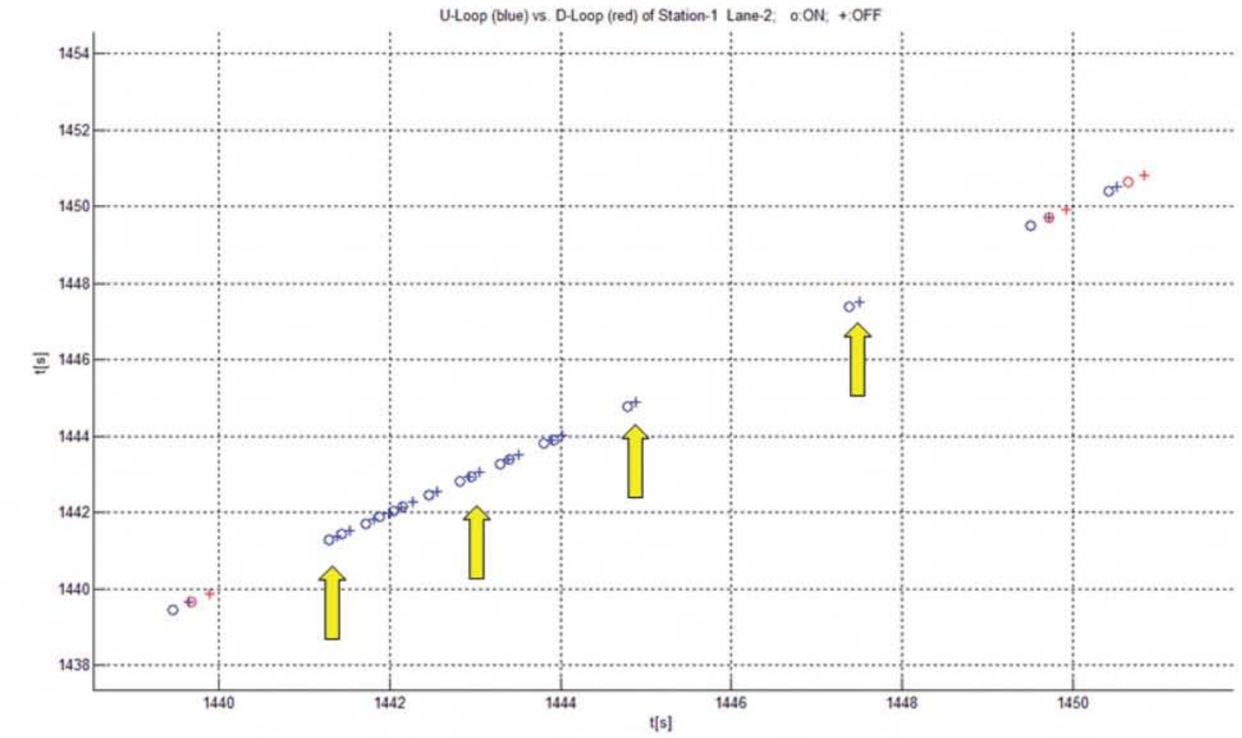
The practical example of pulse breaking is shown in Figure 4.5. It is conducted for each loop individually.

Fault checking: conducting persistent test. If a pulse is less than 10% incomplete, it is considered a temporary fault or loop card flaw and can be corrected. Otherwise, it is considered a persistent fault. In the latter case, loop detector cards and the lower level loop analog signal need to be analyzed to pinpoint the actual fault. The temporary pulse breaking can be characterized as several consecutive significantly short pulses with short gaps. It can be identified by first using the minimum pulse width $T_{on}^{\min}$ as in (4-1). Further consolidated testing would be using the filtered average pulse width (4-3) and pulse gap (4-4) to identify pulse breaking.

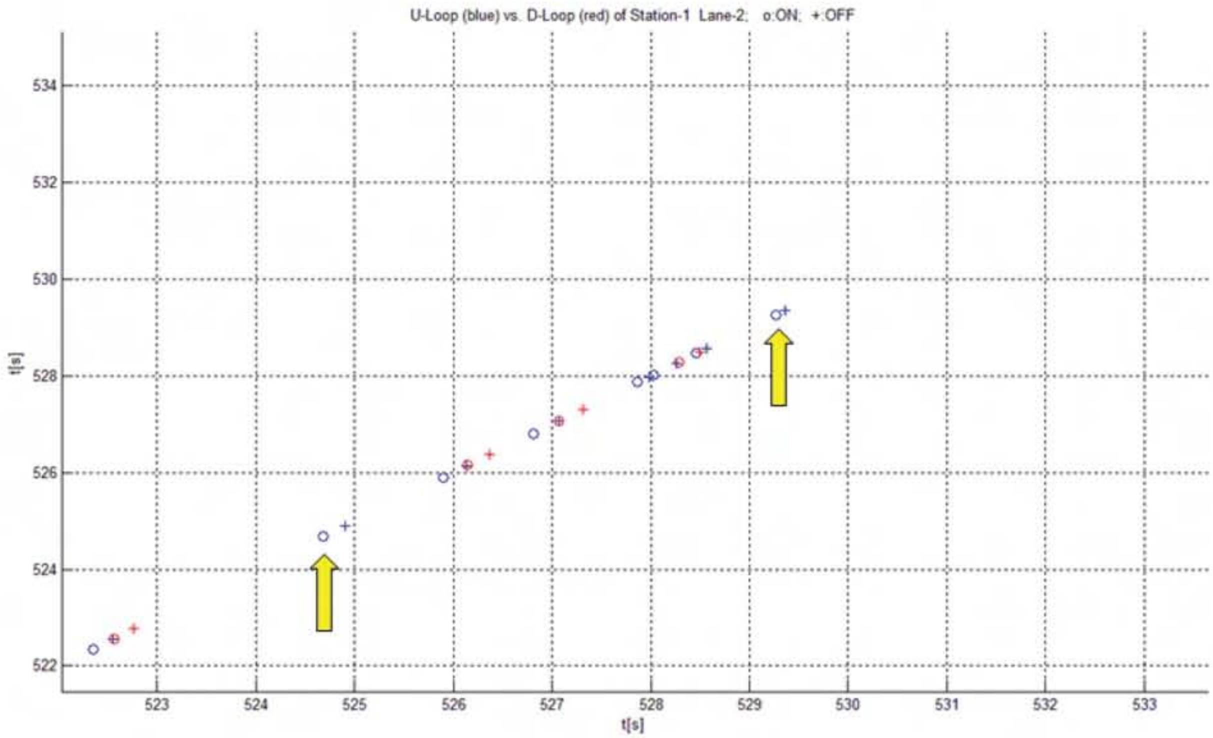
Once this identified, an envelope pulse again based on the filtered pulse width is used to replace the several consecutive broken pulses with a healthy pulse as in Figure 4.5 (b).

4.4.7 Detection and Correct Missing Pulse for Dual Loops

Missing one D-loop pulse or U-loop pulse is another phenomenon observed in the archived BHL event data as shown in Figure 4.6. Such missing data will obviously cause problems for speed estimation. Again, this fault needs to be distinguished as a temporary fault or persistent fault. This could be conducted through persistent test using the 10% threshold. It is noted that such a fault could appear for one of the dual loops for multiple consecutive steps as indicated Figure 4.6 for certain period of time interval but not in other times.



(a)



(b)

Figure 4.6. Data missing example: (a) D-loop pulse missing for certain period of time completely; (b) D-loop pulse missing one pulse occasionally.

To detect such a problem, the following procedure is taken:

(a) check if the loop is reporting a healthy pulse using the minimum and maximum possible pulse width as in (4-1), and the filtered averaged pulse width up to current time step;

(b) Persistent checking: still, 10% missing rate is used as the threshold to distinguish temporary fault and persistent fault;

(c) For real-time processing, one can check the time step of the data buffer to detect if there is pulse missing for U-loop or D-loop: if both do not have new reading it is not considered as missing a pulse;

(d) For archived data, using the following bound checking recursively:

if

$$\begin{aligned} &T_{on}^{d}(k) > T_{off}^{u}(k+2) \\ &or \\ &T_{on}^{d}(k) > T_{off}^{u}(k) + T_{on}^{\max} + \Delta_{gap}(k-1) \end{aligned} \qquad (4\text{-}7)$$

then the D-loop missed a pulse.

If

$$T_{on}^{u}(k) > T_{on}^{d}(k) \tag{4-8}$$

then the U-loop missed a pulse. The first condition in (4-7) and (4-8) are easy to understand. The second condition in (4-7) is based on physical limit of the on-time and filtered time gap.

After identifying a temporary pulse missing, one can fill it in using the filtered pulse width up to previous time step. The trick is to determine the starting time of the pulse. It is recognized statistically that the ON-time instant of the D-loop is very close to the OFF-time instant of the U-loop. This rule is used to determine the location of the missing pulse for both the U-loop and D-loop.

4.4.8 Sensitivity Correction

Loop sensitivity setting is a generic problem. For single loop detector, the speed is inferred from the occupancy. Therefore, sensitivity setting would affect both occupancy and speed. For dual loop station, sensitivity level setting for the U-loop and D-loop would not affect speed estimation if (a) the shape of the U-loop pulse is congruent with that of the D-loop shape; and (b) they are set to the same level which is not too low and not too high. From the discussion in Section 3, one can set the sensitivity level the same from the loop detector card manually for both serial and parallel connection of the loops to the detector card. However, occupancy or density estimation would require sensitivity setting correctly for both U-loop and D-loop, not just the same. If they are not in correct level, it needs to be corrected. Through the raw field data analysis, as indicated in Figure 4.4, the shape of the pulses for U-loop and D-loop may be different even if they are serially connected. As the pair of pulses in legend (a) in Figure 4.4 show, whatever sensitivity level is set in detector card, the duration over the two loops are different.

Therefore, event data level sensitivity correction means two things:

- Setting the sensitivity threshold level properly;

- If one of the On-time durations is significantly differ from the other, trying to find a way to bring them close;

It is suggested that the sensitivity correction is conducted in two steps: *preliminary correction*, and *refined correction*. In the proposed approach, we do not assume the length of vehicles base on classification as that in [9].

Step 1: Preliminary correction: The objective of this step is to match the sensitivity level of two loops and bring them to a reasonable range, but not necessarily the correct level. The preliminary sensitivity correction includes the following steps:

(1)Using the firmware to set the sensitivity of all the loops connected to a card to the same level, which need to be high enough to capture all the pulses triggered by vehicles;

(2)After sensitivity level setting, the ON-time duration of U-loop still different from the D-loop ON-time duration sometimes. This indicates that two loops and the installation are not exactly the same. The following preliminary correction procedure will be used to bring the U-loop pulse width and D-loop pulse width close to each other. This could be done by using the progressively filtered pulse width as shown in Figure 4.7 as long as the center of the pulse can be determined, which can be chosen as the center of the previous pulse.

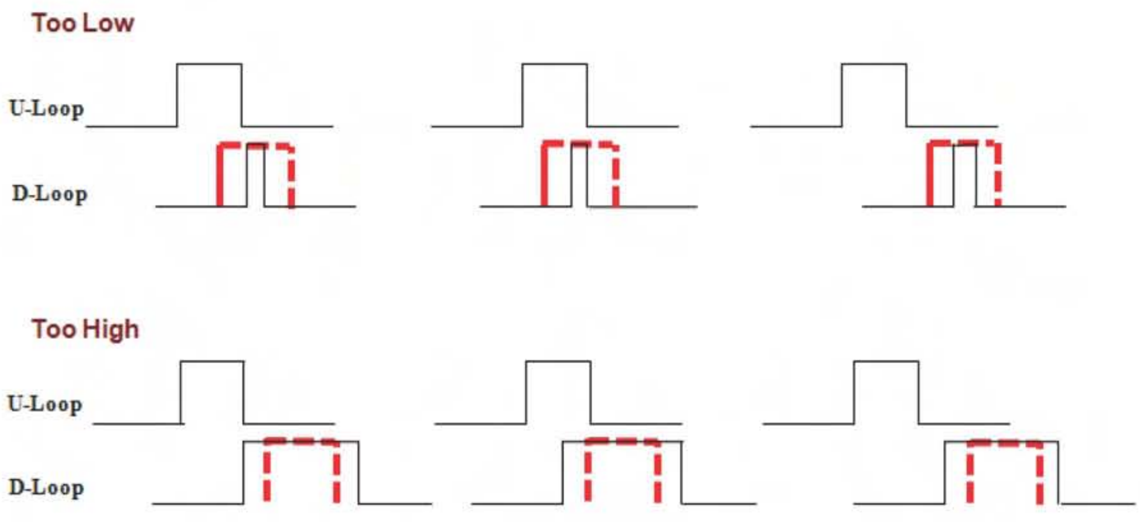


Figure 4.7. Preliminary Sensitivity Correction: To match the sensitivity of U-loop and D-loop to a similar level (This may not be correct yet)

After this step, the speed estimation should be reasonably accurate. However, the sensitivity level may not be correct yet.

Step 2: Refined correction

- Using speed information
- Estimating average vehicle length in free-flow traffic hours
    - Speed has good estimation
    - Regulate Pulse width
    - Estimate vehicle length based on speed and filtered occupancy
- Persistence test
- Refined Correction of Sensitivity
    - Determine sensitivity correction factor

4.5 Conclusion and Remarks

- All the data applications rely on sensor data quality
- Event data are the closest and lowest level data from the sensor
- Data correction at the event data level reduces data error in higher levels
- Data correction for temporary faults
    - Incomplete pulse
    - Pulse breaking
    - Missing pulse
    - Sensitivity (2 steps)
- Using filtered time series information with exponential filter

- Simple algorithms are suitable for archived data and real-time processing
- After data correction, traffic state parameter estimation is straightforward
- Sensitivity correction is still underway

Chapter 5. Using Dual Loop Event Data for Quick Congestion Onset Detection

5.1 Introduction to Chapter 5

Traffic management, control and driver information requires real-time traffic parameter estimation error to be within specified thresholds. In general, traffic management and control, such as ramp metering, requires more parameters and more stringent threshold than traveler information [1]. The required accuracies listed in [1] are all above 95% for traffic control. Such accuracy requirements are great challenges to traffic surveillance because several factors would affect traffic parameter estimation accuracy. To name a few: sensor measurement error (measurement noise and discrepancy), data processing methods, time delays (in sensor measurement and data processing such as filtering and aggregation). Time delay will have more effect on estimation accuracy if traffic speed is varying. Data aggregation is important for understanding traffic flow characteristics for planning since aggregated data usually provides good statistical pattern or trend which has been widely used in traditional traffic data analysis. For real-time operation, if the sensor allows for better traffic parameter estimation, aggregation over time needs to be reduced since it naturally brings time delay and thus error which will hide some important traffic characteristics such as congestion onset time.

It has been claimed that cell phone reporting is faster than automatic congestion onset detection. This may be true for incident management but not necessarily true for automated traffic control, which still needs automatic detection of incidents. Beside incidents, other factors may also cause congestion such as weather, road geometry and upstream and/or onramp traffic flow over capacity. Those situations will not be reported by public drivers except probe vehicles. The reasons for the delay in traffic congestion onset detection may be manifold: (1) No efficient method exists without using traffic data aggregated over several minutes, which naturally leads to time delays and estimation error; (2) driver reports could directly reach CHP (California Highway Patrol) but could not be used for traffic control such as ramp metering along a corridor by TMC (Traffic Management Center), which requires automated incident detection; and (3) loop stations are usually 500m apart. The shockwave has to reach the upstream station and the discharge wave has to reach the downstream station before it can be detected.

The main contributions of this part of the work include: (1) to present the main factors that would affect the traffic parameter estimation error, which has been largely ignored in traditional traffic data analysis; (2) to fully use the merit of dual loop stations for more accurate point speed estimation with sub-second data without aggregation over time; and (3) to apply the obtained speed/occupancy estimation for quick real-time congestion onset detection. Two Berkeley Highway Lab data sets involving incidents have been used for model validation with results analyzed. The presented automatic congestion onset detection method is shown able to detect an onset of an incident/accident within one minute if the two consecutive loops are within $350m$ or less. Main work of this chapter was presented in [40].

5.2 Literature Review

For clarity, we briefly review some previous work on traffic parameter estimation and congestion onset (or bottleneck) detection respectively.

5.2.1 Traffic State Parameter Estimation

This part focuses on estimation of speed/occupancy. Among those traffic parameters, vehicle speed is the fundamental one. Since *time mean speed* and *distance mean speed* can be converted to each other, they are equivalent. The time mean speed is used in this chapter.

There is a large quantity of work on traffic parameter estimation using inductive loops and other sensors, which can only be selectively reviewed. The reports in [52, 58] present the findings of a comprehensive evaluation of the FSP (Freeway Service Patrol) along 10 miles I-880. Several traffic parameters were estimated based on low level loop data with aggregation over 5 min time intervals [5].

The Performance Measurement System (PeMS) [5, 11] processes the data in real-time to fulfill the following tasks: (i) To aggregate 30-second values of counts and occupancy to lane-by-lane, 5-minute values; (ii) To calculate the g-factor of each loop; (iii) To use the g-factor to calculate the speed for each lane; (iv) To aggregate the lane-by-lane value of flow, occupancy, and speed across all lanes at each detector station (one station typically serves the detectors in all the lanes at one location); and (v) To compute the basic performance measures. Many data mining works have been done based on PeMS data.

Dual loop stations have many advantages over single loop. The most important feature is that both vehicle speed and length can be directly detected instead of assuming one of them in the case of single loop station. The detection accuracy and reliability are significantly increased. This merit was used in [10] for improving truck identification.

5.2.2 Congestion Onset Detection

Congestion may be caused by incident/accident, but may also be caused by other factors such as weather, road geometry and traffic flow over capacity. It has been observed on freeway and demonstrated by analysis using a model [47] that the freeway traffic becomes unstable (stop & go; or shockwave) when density increases beyond the capacity even if there is no accident. So, automatic incident detection is absolutely necessary, particularly for traffic control. The work in Stephanedes and Chassiakos [60] used Moving-Window Average or Median filtering to smooth the occupancy for specially distributed loop stations and then used the smoothed occupancy to detect incident. It was claimed that performance has been improved.

Work in [34, 35] proposed a simple method for traffic congestion detection. It was a scheme based on the measurement over the two consecutive loop stations for a section of the freeway. The approach was essentially an integration which has a filtering effect. The detection principle could be described as this: if the difference of occupancies at the two consecutive stations was over some threshold, it was considered as congestion. The measure of magnitude for the congestion was considered, which could be understood as the value in excess of the threshold. An upper bound for congestion onset time delay was further suggested in [35] as:

$$\frac{L \times \rho_j}{\Delta q}$$

where L was the detector spacing; ρ_j was the jam density; and Δq was reduction flow – the flow difference between the two detection points. In later discussion, we will provide an estimation of the time delay based on shockwave speed [37].

It was pointed out by Persaud and Hall [50] that *Occupancy Discrepancy* had made the conventional occupancy-based incident detection logic difficult to apply. An experimental study conducted by Chan and May [2] showed with field data that the average detector pulse on-times for two longitudinally closely spaced stations could vary by 5-10% or higher. Those studies used single loop stations.

Work in [20] proposed to use travel time of a special vehicle, or vehicle signature, on two consecutive loop stations to detect if there was a congestion onset. This approach would work provided that special vehicle signature could be re-identified at both upstream and downstream, which is related to the work of [10].

Work in [7] used loop data aggregated over 5 minutes for sustained bottleneck detection. It could be used to detect recurrent bottlenecks on freeway networks for planning purposes. The data aggregation and detection methods developed there were for sustained congestion instead of temporary slowing down. Using this method for real-time incident detect implied that the time delay would be at least 5 minutes.

Different methods for incident detection were evaluated and documents extensively reviewed in [45]. 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 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.” The McMaster Algorithm used lane data for detection, while most other methods used aggregated directional data. The approaches developed were classified into four categories:

- Pattern recognition
- catastrophe theory
- statistical
- artificial intelligence

It is necessary to discuss pattern based and Catastrophe Theory based incident detection algorithms which are related to what is proposed in this chapter.

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 exceeds the threshold compared to normal case, an incident is announced. Parameters used include those at both upstream and downstream detection stations or at the 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 algorithms in this category are worth mentioning:

(i) California algorithm: It had different versions which mainly use upstream and downstream occupancy. This is the most extensively explored algorithm and widely implemented according to [45].

(ii) The APID (All Purpose Incident Detection) algorithm: It was a combination of the various California algorithms along with a compression wave test routine and a persistence test routine. Unlike the California algorithms, it used 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 [50]. This algorithm exceptionally used speed but has not been developed further since then.

Catastrophe Theory: This theory was based in sudden changes that occur in one of the three variables (speed, flow, and occupancy) while the other two exhibited smooth and continuous changes. When speed drops dramatically without a corresponding increase in occupancy and flow, the alarm sounds**.** The algorithm functions were based on data from a single detector station [1]. The *McMaster Algorithm* is the representative [45].

5.3 Point (Time) Means Speed Estimation

Dual loop station make it possible to estimate vehicle speed using loop on/off time instant based on sub-second loop data [58]. However, noise still exists. An experimental study conducted by Chan and May [2] 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. This implies that care still needs to be taken for using dual loop sub-second data for traffic parameter estimation. Besides, techniques developed in FSP and PeMS for data correction and cleansing as pre-processing [11, 52, 58], and other linear filtering are used to smooth the data in this work.

5.3.1 How to Describe the Estimation Error

The estimation error includes two types: absolute error and time delay. As is known, for speed estimation, time delay also makes some contribution to absolute estimation error. This is particularly true if there is speed fluctuation which happens very often for congested traffic.

Traffic flow may be divided into four phases: free flow, congestion on-set, congested static state, recovering (from congested to free flow). Since the mean speed trajectory has different characteristics for each phase, the estimation error will behave differently. For free-flow and congested steady state which are homogeneous flow, there is not much speed fluctuation. Thus time delay in the estimation does not play a significant role to the error except in the first transient period. For congestion onset and recovering phases, the traffic flow has negative and positive acceleration respectively. Time delay in the estimation makes a significant contribution to estimation error. Intuitively, if one shifts a non-constant signal by a time interval $\Delta T > 0$ and then compares it with the original signal, the vertical difference – the error - would become apparent. This is the reason to reduce time delay to improve traffic parameter estimation.

5.3.2 Time Delay and its Effect on Traffic Parameter Estimation Error

Work in [18] considered traffic parameter estimation error based on a loop station at a fixed point. Beside sensitivity, several other important factors have been identified that would affect the estimation error: sensor measurement error, estimation method, and time delays. Detailed discussions are referred to [36].

In traditional traffic management and planning, data aggregated over time and distance were common practice. To see this more clearly, it is necessary to separate real-time data processing and archived data off-line processing since the data availability and methods used in the two situations are quite different. For example, for real-time processing, data available at time instant t are at most all the data in the past up to current, but not the future data; however, for archived data, one can use data later than current instant t for processing. This is the reason why after-processing archived data can produce much better results than real-time data.

(a) Time Delay Caused Aggregation Method

Let's look at the time delay caused by Moving Window Averaging popularly used in real-time data aggregation. We aggregate every N time intervals Δt, the sampling time interval. Let's estimate o, the occupancy, using data from the last N time steps.. Suppose the method for aggregation is simple averaging:

$$\bar{o} = \frac{1}{N}\sum_{i=1}^{N} o\left(t-(i-1)\Delta t\right)$$

From formal Taylor expansion:

$$\begin{aligned}\bar{o} &\approx \frac{1}{N}\sum_{i=1}^{N}\left(o(t)-\dot{o}(t)(i-1)\Delta t\right)\\ &= \left[\frac{1}{N}\sum_{i=1}^{N} o(t)\right]-\left[\frac{\dot{o}(t)\cdot\Delta t}{N}\sum_{i=1}^{N}(i-1)\right]\\ &= o(t)-\frac{N^2}{2}\frac{\dot{o}(t)\cdot\Delta t}{N} = o(t)-\frac{N\cdot\Delta t}{2}\dot{o}(t)\end{aligned} \tag{5-1}$$

which means the time delay caused by this moving window average in the level of $\frac{N\cdot\Delta t}{2}$.

(b) Estimation Error Caused by Time Delay

If a traffic parameter is a constant or near constant value such as free-flow speed, time delay does not matter much. However, when the traffic parameter varies significantly over time, the error caused by time delay would be more significant as seen from (5-1) that

$$\left|\bar{o}-o(t)\right| \approx \left|\frac{N\cdot\Delta t}{2}\dot{o}(t)\right|$$

which depends on the slope of occupancy $\left|\dot{o}(t)\right|$.

(c) More Attention Needed on Time Delay Needed for ITS Application

The effect of time delay in traditional traffic data analysis was not well-recognized due to several reasons:

- Sensor detection systems were not develope 15 years ago: compared to sensor systems 15 years ago when single loop detection system and probe vehicles were the main tools for traffic data collection; nowadays, dual loop stations, video cameras, microwave/laser/infrared radar systems, cell-phones with GPS, Weigh-in-Motion systems, sensor networks with wireless capabilities such as Sensys Systems are all commercially available for traffic monitoring and detection; those commercially products can provide traffic data from many points of view.

- Communication systems for synchronized data passing were not generally available; today GPS equipment has been widely used. Its UTC time is generally used as a data time stamp, which could be used for real-time processing or in archived data for after-processing;
- Most importantly, traffic management and control needs: with the development of ITS technologies, traffic management has also developed from planning into real-time operation such as traffic signal control and optimization, incident detection and handling, freeway ramp metering, and traffic speed regulation, of which congestion onset is only one important piece. The development also brings new challenges to traffic monitoring and detection. One of the most important changes for traffic monitoring and detection is to provide more accurate estimation of traffic parameters in real-time with delay or hysteresis.

5.3.3 Preliminary Loop Data Processing

The recorded Berkeley Highway Lab (BHL) raw loop data from 170E control cabinets are decoded to give each station (as shown in Figure 2.2) ON and OFF time instant counted as the number of (1/60)s. Thus the information obtained is practically 60Hz. The raw data need to be cleaned, properly matched for upstream-downstream of the same station, and missing data to be imputed. Then speed and occupancy estimation for each lane at each dual loop station are estimated using the method similar to those introduced in [18]. The following 2nd order low pass real-time Butterworth filter is used afterwards to smooth up the occupancy and the speed trajectories with respect to time at each station.

$$
\begin{aligned}
&x(t+1) = A \times x(t) + B \times x_{in}(t+1) \\
&y_{out}(t+1) = C \times x(t) + D \times x_{in}(t+1) \\
&t = 0, 1, 2, \ldots \\
&\mathrm{A}=\begin{bmatrix} 0.2779 & -0.4152 \\ 0.4152 & 0.8651 \end{bmatrix}, \mathrm{B}=\begin{bmatrix} 0.5872 \\ 0.1908 \end{bmatrix} \\
&\mathrm{C}=[0.1468 \quad 0.6594], \quad \mathrm{D}=[0.0675] \\
&x(0) = x_{in}(0)
\end{aligned}
\qquad (5\text{-}2)
$$

where $x(t) = \left[x_1(t), x_2(t)\right]^T$ is the filter state; $x_{in}(t)$ is the input signal; and $y_{out}(t)$ is the filtered output single. Through phase analysis, it can be seen that this filter causes time delay less than

0.5s which can be ignored for traffic control purposes. It is noted that the data are not aggregated over time, nor aggregated over distance. However, it is implicitly assumed that if there is no vehicle passing the given station, the speed will keep the value of the previous time step instead of setting it to zero. The purpose to do so is to avoid unnecessary speed fluctuations.

5.4 Congestion Onset Detection

Congestion onset detection principle can be decided from the following characteristics at point stations:

- Upstream occupancy is significantly higher than that of the downstream;
- Upstream speed at the loop station is significantly lower than the speed at the downstream loop station;
- The detection criterion is quantified by a selected detection threshold;
- The value above the threshold can be used as a measure for the magnitude of the congestion;
- The spatial/temporal effects can be obtained by analyzing consecutive loop stations upstream.

The following methods use the merits of dual loop stations, which is usually good for point speed measurement based on the registered on/off time instants. We still put methods based occupancy there although its estimation is not as good as that of speed even for dual loop stations.

5.4.1 Detection Methods

The detection methods in this chapter are different from those proposed in Lin and Daganzo [34, 10] in three folds:

(1) traffic parameter estimation methods are different: here the speed and occupancy estimation at the two consecutive stations removed the time delay caused by aggregation; filtered (essentially time mean) speed is used;

(2) lane traffic parameters are used due to better estimation from dual loop stations, which leads to higher detection sensitivity;

(3) the traffic situation between the two consecutive stations is taken into account by using the most recent shock-wave back-propagation speed; this will give a quantitative estimation of maximum time delay caused in the detection.

Method 1: Mean Speed or Mean Occupancy Difference

Since significant congestion is a collective behavior for all the lanes, it makes sense to average the speed and occupancy trajectories respectively for all the lanes at a given station. Clearly, this method is less sensitive.

If congestion onsets between two stations or at the upstream site of two stations, a congestion on-set can be identified if either of the following criteria is satisfied:

$$\overline{v}_u(t) - \overline{v}_d(t) \le -\delta_v^{(1)} \tag{5-3}$$

or

$$\overline{o}_u(t) - \overline{o}_d(t) \ge \delta_o^{(1)} \tag{5-4}$$

where

$\overline{v}_u(t), \overline{v}_d(t) \ge 0$ - Upstream and downstream mean speed over all lanes at time t

$\overline{o}_u(t), \overline{o}_d(t) \ge 0$ - Upstream and downstream mean occupancy over all lanes at time t

$\delta_v^{(j)}, \delta_o^{(j)} > 0$ - Pre-specified thresholds for method j to be selected in validation.

Detection Principle 1 for Method 1: If either of the above inequalities (5-3) and (5-4) holds consistently for 90% of the time steps certain time interval $t \in [T_1, T_1 + \Delta T]$, then a congestion onset is detected, which starts at time instant T_1 and the location is between the two consecutive stations or at the upstream of two stations.

Detection Principle 2 for Method 1: If congestion happens at a location further downstream of two stations and there is no detector station further downstream available, the onset can be identified if either of the following criteria (5-5) and (5-6) are satisfied:

$$\overline{v}_u(t) - \overline{v}_d(t) \ge -\delta_v^{(1)} \tag{5-5}$$

or

$$\overline{o}_u(t) - \overline{o}_d(t) \le \delta_o^{(1)} \tag{5-6}$$

If either of the above inequalities (5-5) and (5-6) holds consistently for 90% of the time steps during a certain time interval $t \in [T_1, T_1 + \Delta T]$, then a congestion onset can be claimed, which starts at time instant T_1 and the location is further downstream of the two stations.

This method can be used in the case of significant congestion where most of the lanes at upstream stations are affected noticeably. ΔT is the time threshold for consistency, which is suggested as *15s* in later validations using data.

Method 2: Single Lane Upstream and Average over Lanes Downstream

It is also possible to compare single lane upstream traffic with average traffic across all lanes downstream.

$$v_u^{(i)}(t) - \overline{v}_d(t) \le -\delta_v^{(2)} \tag{5-7}$$

or

$$o_u^{(i)}(t) - \overline{o}_d(t) \ge \delta_o^{(2)} \tag{5-8}$$

where i is the lane index; $v_u^{(i)}(t)$ and $o_u^{(i)}(t)$ are the lane speed and occupancy upstream.

Detection Principle 1 for Method 2: If either of (5-7) or (5-8) holds consistently for 90% of the time steps and for some fixed lane i for certain time interval $t \in [T_1, T_1 + \Delta T]$, then a congestion onset is detected, which starts at time instant T_1 and the location is between the two consecutive stations or at the upstream site of the two stations. Here, $\Delta T = 15s$ is used.

Detection Principle 2 for Method 2: If congestion happens at a location further downstream of two stations and there is no detector station further downstream available, the following criterion (5-9) and (5-10) is used:

$$v_u^{(i)}(t) - \overline{v}_d(t) \ge -\delta_v^{(2)} \tag{5-9}$$

or

$$o_u^{(i)}(t) - \overline{o}_d(t) \le \delta_o^{(2)} \tag{5-10}$$

If either of the above inequalities holds consistently for 90% of the time steps and for some fixed lane i for certain time interval $t \in [T_1, T_1 + \Delta T]$, then a congestion onset is detected, which starts at time T_1 and the location is somewhere downstream of the two stations. This method can be used for non-significant congestion cases where at least one lane at the upstream station is noticeably affected for a certain period of time by the congestion. Clearly, this criterion is more sensitive than the Method 1 above. In this case, if the average value is used as in Method 1, the Mean Difference may not be significant enough to detect the congestion onset.

Threshold Determination: Threshold selection depends on the capacity of the freeway, or the number of lanes: one lane closure would affect a 5-lane freeway section less than affecting a 3-lane freeway section. Thus, the thresholds are suggested as the following:

$$\begin{aligned} \delta_v^{(j)} &= \frac{\sigma_v^{(j)}}{N} \\ \delta_o^{(j)} &= \frac{\sigma_o^{(j)}}{N} \\ N &= \frac{N_U + N_D}{2} \end{aligned} \tag{5-11}$$

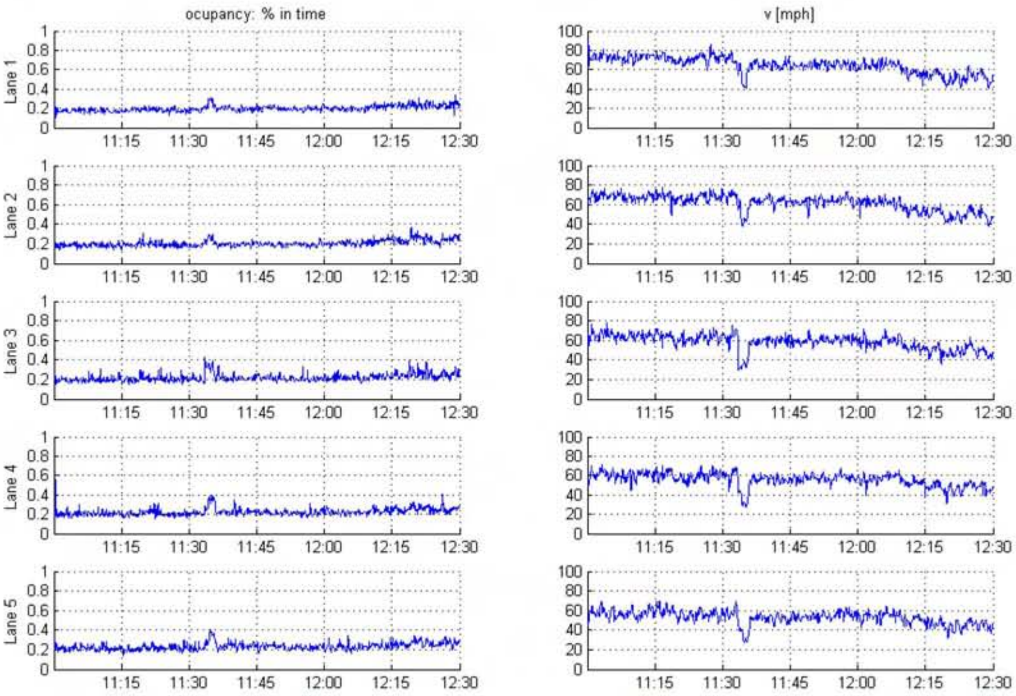
where N_U, N_D are the number of lanes at an upstream station and downstream station respectively. Now the thresholds can be generically expressed as

$$\begin{aligned} \delta_v^{(j)} &= \frac{2 \times \sigma_v^{(j)}}{N_U + N_D} \\ \delta_o^{(j)} &= \frac{2 \times \sigma_o^{(j)}}{N_U + N_D} \\ j &= 1,2 \end{aligned} \tag{5-12}$$

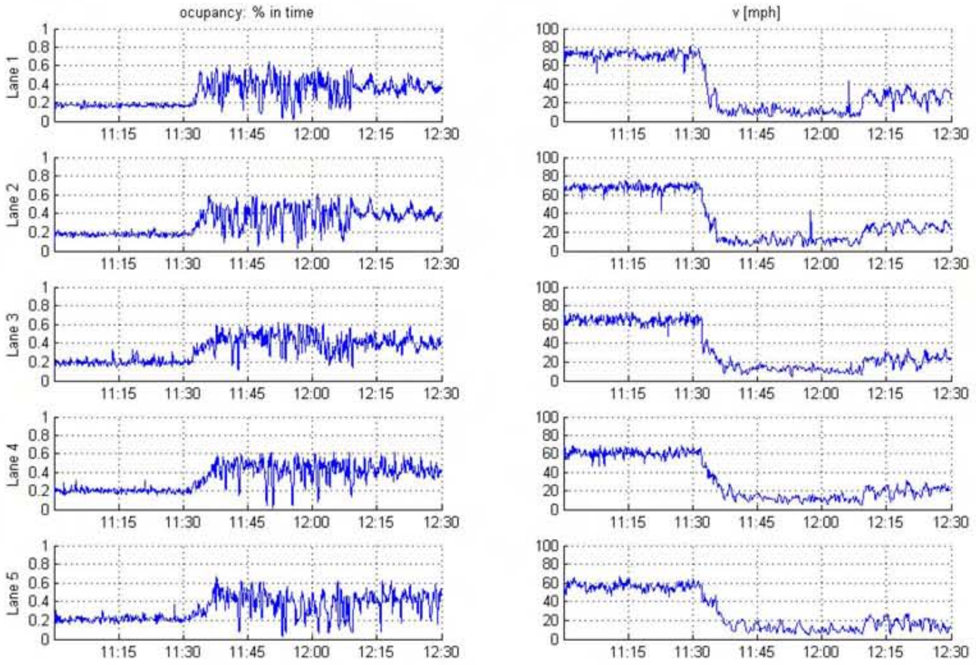
The number $\sigma_v^{(j)}, \sigma_o^{(j)}$ can be calibrated for freeway networks.

5.4.2 Algorithm Validation Using BHL Data

The following two sets of BHL data have been used for the validation work of the proposed methods: I-80 East Bound on 07/01/2007 of all stations between 11:00am and 12:30pm, and I-80 East Bound on 07/17/2008 of all stations between 7:30am and 9:00am. It is noted that the observed congestion in the first dataset started between Station 6 and Station 4, and the incident which caused critical congestion in the second dataset occurred downstream of Station 1. Traffic flow direction is from Station 8 to Station 1 as indicated in Figure 2.2. This is a set of $\frac{1}{60}s$ data in the sense that, although data are recoded every second, loop ON-OFF is counted in $60Hz$. The selection of those data is based on freeway PeMS incident records from California Highway Patrol and the online statistical plot of BHL website. They indicated that there were congestions in the BHL stretch during the selected time periods. The filtered trajectories for speed and occupancy are shown in Figure 5.1 for 07/01/2007 and in Figure 5.2 for 07/17/2008. Station 5 did not have correct data during that period of time and thus was dropped out.

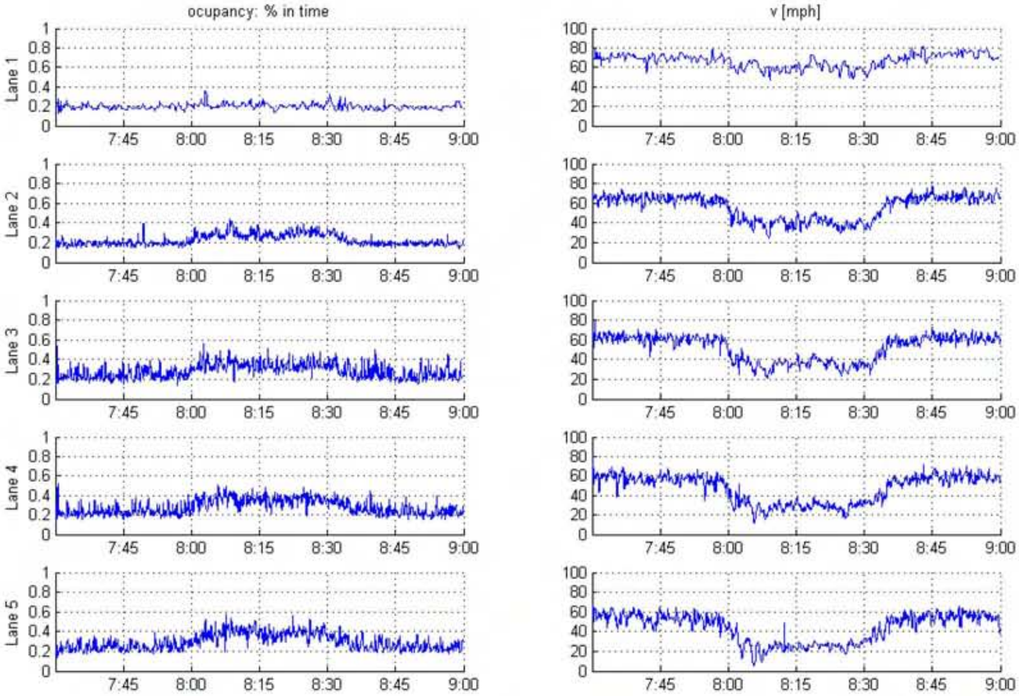


(A). Station 4

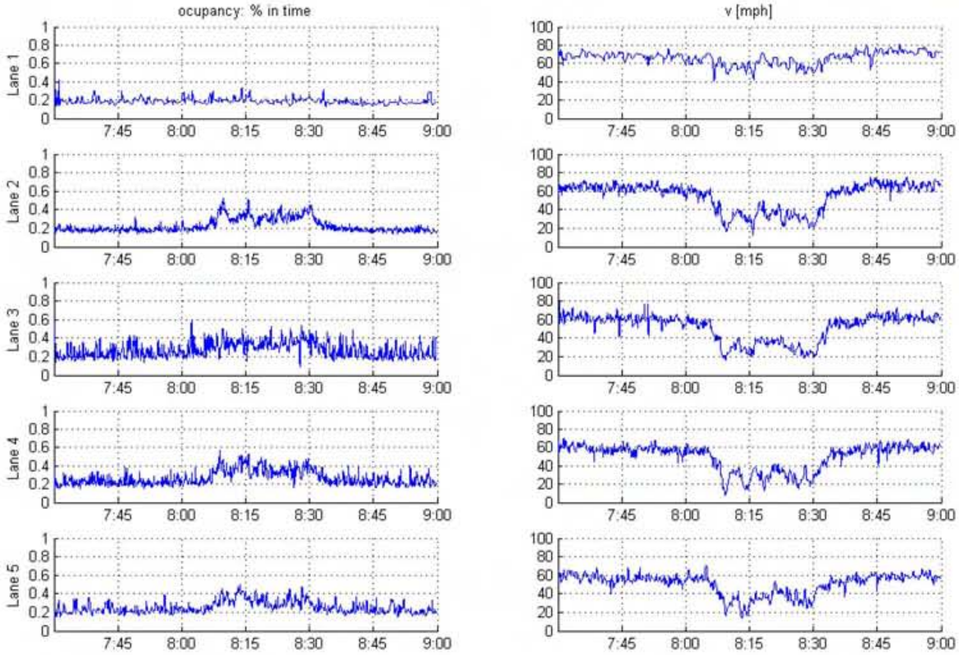


(B). Station 6

Figure 5.1. Occupancy and speed estimation at loop Station 4 and Station 6 on 07/01/2007



(A). Station 1



(B). Station 2

Figure 5.2. Occupancy and speed estimation at loop Station 1 and Station 2 on 07/17/2008

5.4.3 Further Data Processing

To validate both Detection Principle 1 and Detection Principle 2 of Method 1, averaging the speed trajectories across all 5 lanes at a station is necessary. However, it is not straightforward since a vehicle arrival triggers a loop station ON at random time instants. Those time instants for each lane at the same station are independent. Alternatively, the discrete time points for the dual loop in each lane of the same station are NOT synchronized. To overcome this difficulty, a common time interval $\Delta t = 1.0s$ is selected and the following interpolation methods are used for synchronizing all the 5 lane data as follows: Suppose $\left\{t_0^{(i)}, t_1^{(i)}, t_2^{(i)}, \ldots, t_{N_i}^{(i)}\right\}$ are the time points for the discrete speed and occupancy trajectories for Lane i of the given station. Set $N = \min\left\{N_1, N_2, N_3, N_4, N_5\right\}$ and generate a common time sequence $\left\{t_0, t_1, \ldots, t_N\right\}$ as: $t_0 = 0, t_n = n \cdot \Delta t,\ n = 1, \ldots, N.$ Now for each lane i, synchronized speed trajectory $\left\{v^{(i)}(t_0), v^{(i)}(t_1), \ldots, v^{(i)}(t_N)\right\}$ and occupancy trajectory $\left\{o^{(i)}(t_0), o^{(i)}(t_1), \ldots, o^{(i)}(t_N)\right\}$ are constructed through linear interpolation as follows: if $t_j^{(i)} \le t_k \le t_{j+1}^{(i)}$, then

$$v^{(i)}(t_k) = v^{(i)}\left(t_j^{(i)}\right) \times \frac{t_{j+1}^{(i)} - t_k}{t_{j+1}^{(i)} - t_j^{(i)}} + v^{(i)}\left(t_{j+1}^{(i)}\right) \times \frac{t_k - t_j^{(i)}}{t_{j+1}^{(i)} - t_j^{(i)}}$$

$$o^{(i)}(t_k) = o^{(i)}\left(t_j^{(i)}\right) \times \frac{t_{j+1}^{(i)} - t_k}{t_{j+1}^{(i)} - t_j^{(i)}} + o^{(i)}\left(t_{j+1}^{(i)}\right) \times \frac{t_k - t_j^{(i)}}{t_{j+1}^{(i)} - t_j^{(i)}} \tag{4-9}$$

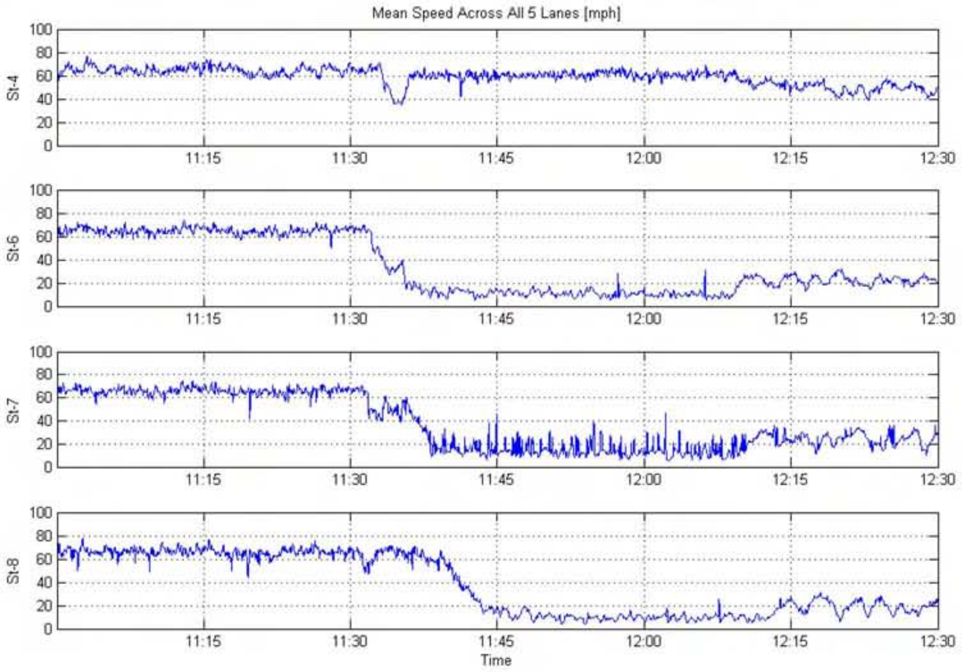
Then the mean value of the synchronized trajectories for all the lanes at a given station can be simply calculated at the synchronized time points t_k as follows:

$$\bar{v}(t_k) = \frac{1}{N} \sum_{i=1}^{N} v^{(i)}(t_k)$$

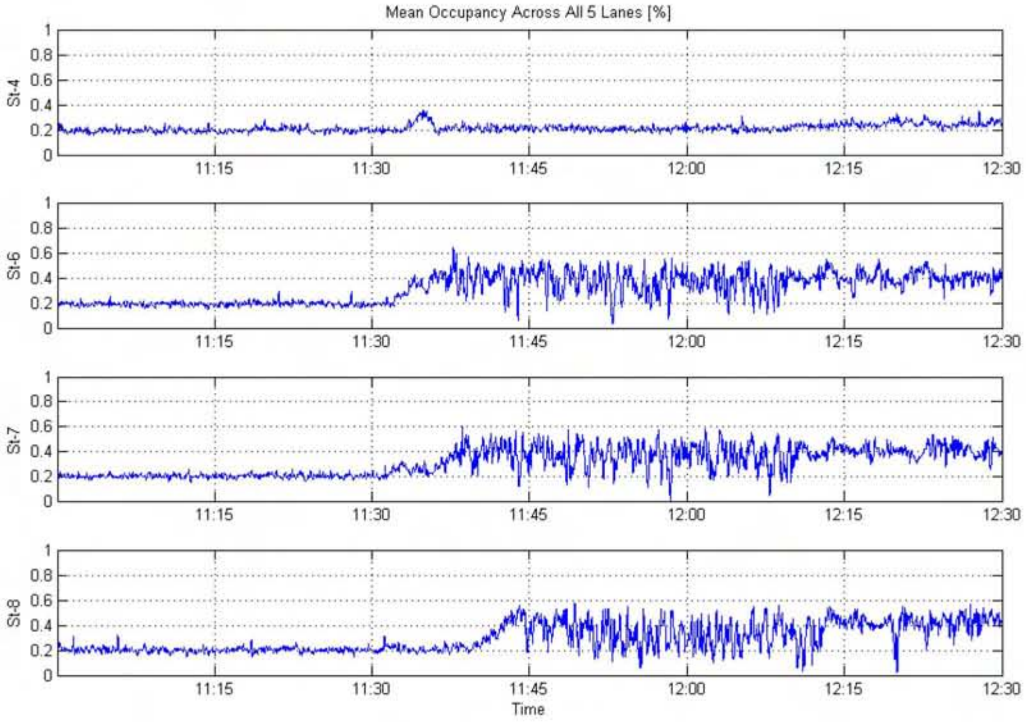
$$\bar{o}(t_k) = \frac{1}{N} \sum_{i=1}^{N} o^{(i)}(t_k) \tag{4-10}$$

where $N = 5$ is the number of lanes. For the selected data sets, the mean values of speed and occupancy trajectories across all the lanes on 07/01/2007 for Station 4, 6, 7, and 8 are shown in

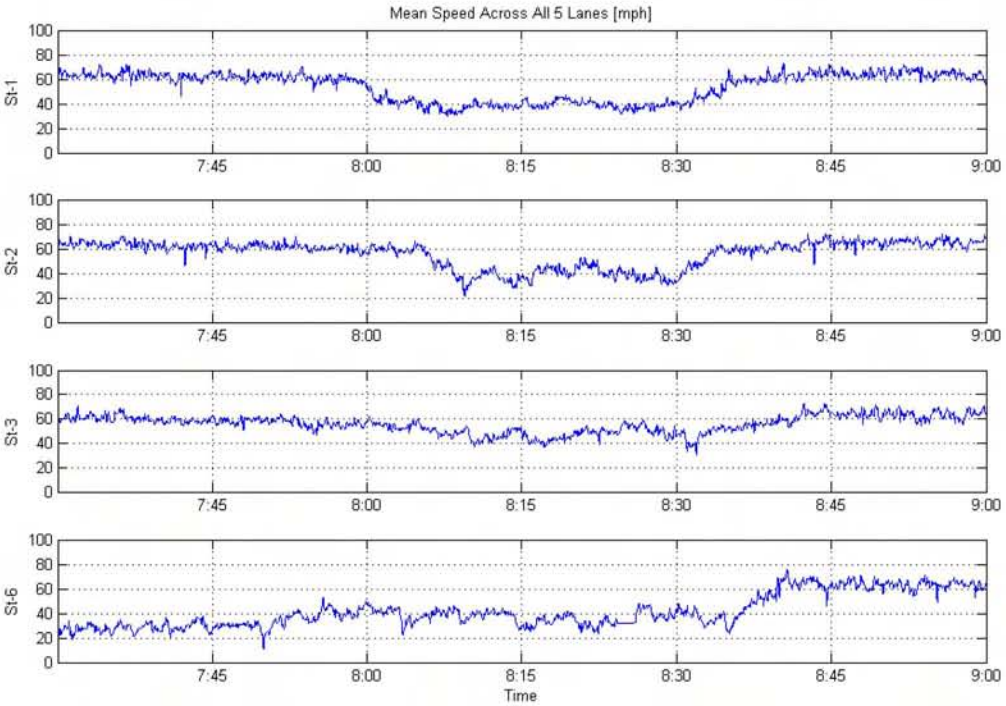
Figure 5.3. A and Figure 5.3 B, and those on 07/17/2008 for Station 1, 2, 3 and 6 are shown in Figure 5.3. C and Figure 5.3 D.



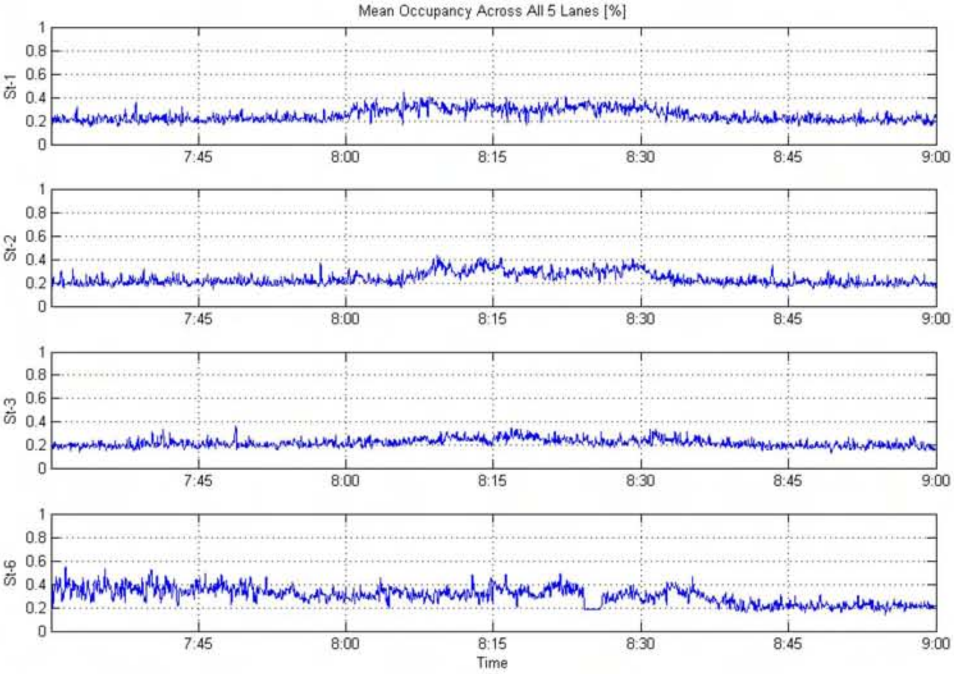
(A). Mean speed *mph* for Station 4, 6, 7, 8 on 07/01/2007



(B). Mean occupancy [xxx% time] for Station 4, 6, 7, 8 6 on 07/01/2007



(C). Mean speed *mph* for Station 1, 2, 3, 6 on 07/17/2008



(D). Mean occupancy [xxx% time] for Station 1, 2, 3, 6 on 07/17/2008

Figure 5.3 Mean speed and occupancy for lane groups

5.4.4 Validation of Congestion Onset Detection Principle 1

We can observe from Figure 5.3.(A) and Figure 5.3. (B) the following facts for the incident on 07/01/2007:

- The distance location for the congestion onset is at some point between Station 4 and Station 6;
- The starting time for the onset is around 11:35am.
- In the time period between 11:30am and 11:35am, a non-significant incident upstream of the Station 4 caused some speed drop below $40mph$ at Station 4; this affected the traffic somehow at Station 6 in a short period of time, but not much to Station 7 and Station 8;
- After 11:36am, Station 4 and downstream appears to be free-flow; the congestion started at some point between Station 4 and Station 6 back-propagated to Station 7 and Station 8, which caused a significant speed drop and occupancy increase;
- Speed estimation has much less noise than occupancy estimation.

Based on this consideration, it is only necessary to focus on the data after 11:05am. The speed difference and occupancy difference between Station 6 and Station 4 are shown in Figure 5.4. It can be seen that if the following thresholds are chosen:

$$\delta_v^{(1)} = 10mph, \delta_o^{(1)} = 0.1, \Delta T = 15s$$

this is equivalent to choosing $\sigma_v^{(1)} = 50.0$ and $\sigma_v^{(1)} = 0.5$. After the onset detection algorithm is implemented, the congestion onset can be detected, and it started from 11:32am. This means the Detection Principle 1 (4-1) and (4-2) hold, and it has validated the algorithm. This data set can also be used for validation of Method 2 since it is a weaker condition than Method 1, which will not be repeated here.

5.4.5 Validation of Congestion Onset Detection Principle 2 of Method 1

Based on the incident record from PeMS and Figure 5.3, the following facts are known for the incident on 07/17/2008:

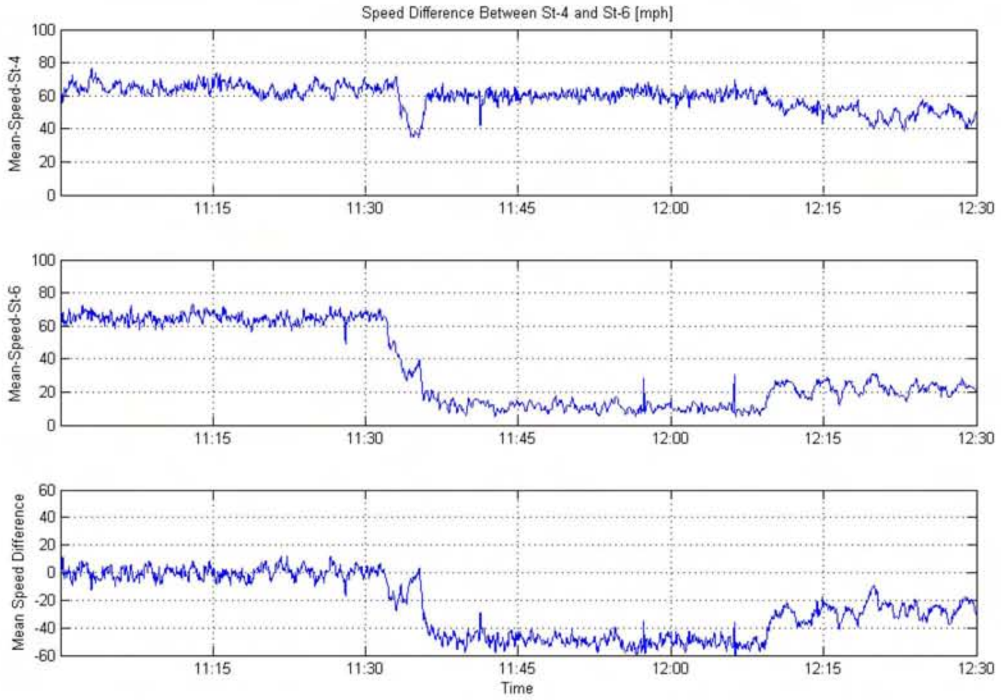
- The distance location for the congestion onset is at some point downstream of Station 1;
- The starting time of the incident is around 7:51am;

- In the time period between 8:00am to 8:05am, a significant incident downstream of Station 1 caused some speed drop below $40mph$ at Station 1; this affected the traffic somehow at Station 2 and Station 3 at around 8:07am and 8:12am respectively, but not much to Station 6 as the speed on it is relatively slow before 8:00am for some reason;
- The incident was cleared up at 8:27am.

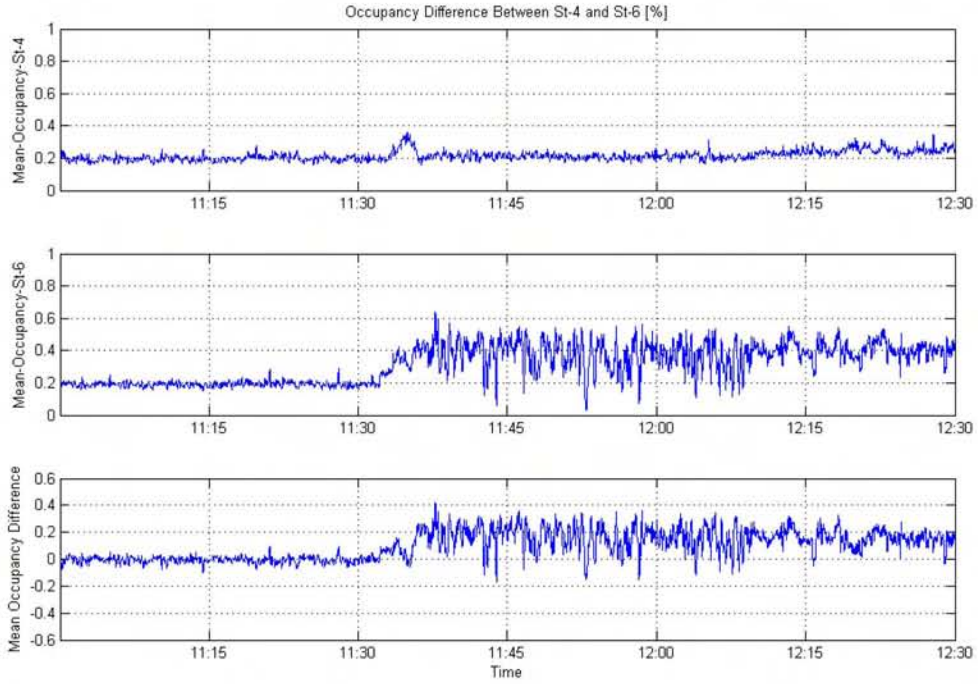
The speed difference and occupancy difference between Station 1 and Station 2 are shown in Figure 5.5. Similar thresholds as above are chosen:

$$\delta_v^{(1)} = 10mph, \delta_o^{(1)} = 0.1, \Delta T = 15s$$

which is equivalent to choosing $\sigma_v^{(1)} = 50.0$ and $\sigma_v^{(1)} = 0.5$. After the onset detection algorithm is implemented, the onset is detected, and it started on 8:00am. This means the Detection Principle 2 (4-3) and (4-4) hold, and it has validated the algorithm.

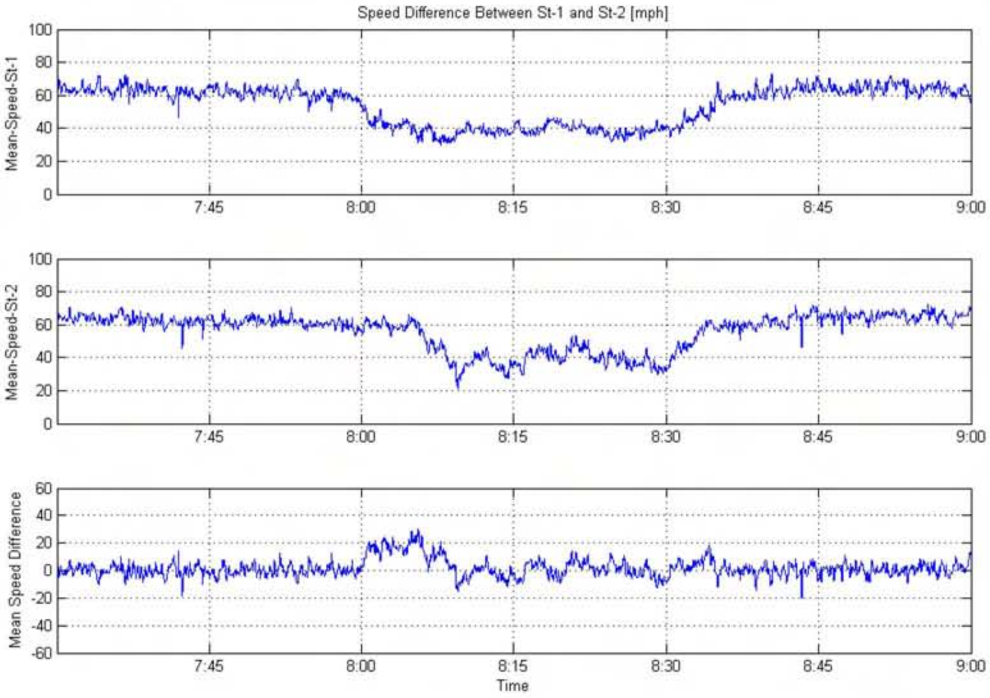


(A) Mean speed, and speed difference between Station 4 and Station 6

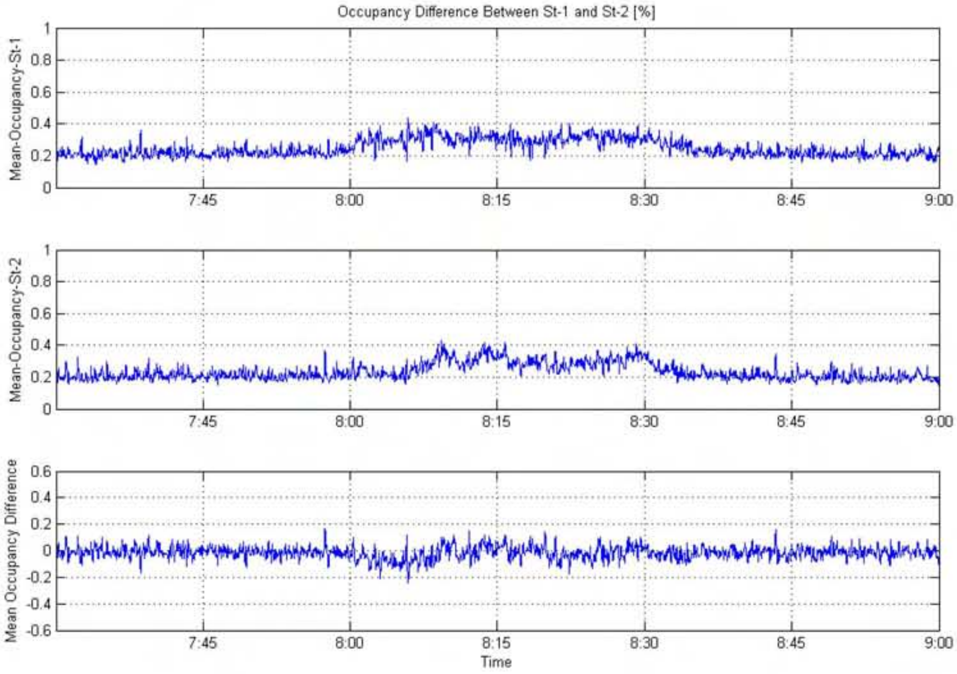


(B) Mean occupancy and occupancy difference between Station 4 and Station 6

Figure 5.4 Mean Speed, Occupancy, and Differences on 07/01/2007



(A) Mean speed, and speed difference between Station 1 and Station 2



(B) Mean occupancy and occupancy difference between Station 1 and Station 2

Figure 5.5. Mean Speed, Occupancy, and Differences on 07/17/2008

5.4.6 Time Delay Estimation for Congestion Onset Detection

It would be useful to understand quantitatively the upper bound for time delay using the above methods. In general, the time delay for congestion onset detection can be expressed as

$$\tau_{delay} = \Delta T_0 + \Delta T + \tau_{\max}$$

Where $\Delta T_0 = T_1 - T_0$ is the time period $\left[T_0, T_1\right]$ for speed difference to reach the threshold in (4-1) ~(4-4); ΔT is the persistent period for the test criteria; and $\tau_{\max}$ is the maximum time period for the shockwave to reach the upstream station or the discharge vehicle stream to reach the downstream station [37]:

$$\tau_{\max} = \frac{L}{5.1 + \overline{v}_d(t)}$$

where L is the distance between two dual loop stations. It is noted that ΔT_0 depends on the gradient of the speed difference reduction which reflects, to some extent, the seriousness of the incident/accident; $\tau_{\max}$ is largely affected by loop density; the $\Delta T = 10 \sim 15s$ will be reasonable. In our case, $L = 1076.6m,\ \overline{v}_d = 26.82m/s$ and thus $\tau_{\max} = 31.5\text{s}$; $\Delta T_0 = 20s;$ and $\Delta T = 15s$. The upper bound of total time delay for congestion onset detection is

$$\tau_{delay} = 20 + 15 + \tau_{\max} = 66.5s$$

However, if the loop distance $L = 350m$ as in most cases on freeways, $\tau_{delay} = 46s$, which means that the congestion can be detected within 1 minute.

5.5 Concluding Remarks

Although data aggregation over time in traffic state parameter estimation is necessary for analysis, it would lead to time delay. Modern traffic control and active traffic management require highly accurate estimation of traffic parameters with minimum time delay. Sub-second dual loop data can lead to accurate speed and occupancy estimation with filtering and without time aggregation.

An immediate application of the speed and occupancy estimation is for quick automated congestion onset detection. Two real-time methods are proposed for the detection: one uses Mean Speeds over all lanes at both upstream and downstream stations, which is a stronger condition; the other uses Means Speed for downstream but a significant lane speed at upstream station, which is a weaker condition but more sensitive. The principle for the detection is to look at speed difference and occupancy difference at the two consecutive loop stations of a freeway section. Once they are above specified threshold for 90% of the time steps and for certain time period consistently, an incident is claimed happening between the two loop stations or further downstream. Two BHL sub-second dual loop station data sets have been used for the validation of the methods. It is also shown how to estimate the upper bound for the time delay in congestion onset detection. The chosen data shows that the onset time and location of a significant congestion could be detected within one minute generally, which is an improvement over previous work [5].

The methods proposed use the merits of dual loop stations. It is expected to achieve quicker and reliable congestion onset detection by more accurate point speed estimation and reduced time delay caused by unnecessary time aggregation.

Related future work would include: validation of the algorithms and calibration of the thresholds through large numbers of data sets; investigation of false alarm rate over large numbers of data sets; establishing a quantitative measure of the seriousness of the congestion.

Although the algorithm relies the good estimation of speed-based dual loop detectors, if the speed estimation from a single loop detector is adequately accurate (such as those in Chapter 3), the algorithm can also be implemented on a freeway corridor with healthy single loop detections.

Chapter 6. Towards Integrated Hierarchical Traffic Data System

1.1 Introduction to Chapter 6

This Chapter presents some preliminary thoughts on (a) traffic data availability with the rapid development of communication systems and on-vehicle electronics which are potential traffic sensors; (b) data needs in the near future for planning and operation; and (c) how the traffic data system should evolve from the current system to a future system with the minimum infrastructure cost.

Highway traffic systems are very complicated because they use related networks and are highly stochastic. Active Traffic Management (ATM) has to be conducted from a systems approach. The system should include

- highway network modeling which captures physical links of the network, demand and capacity;
- traffic management strategies from both planning and operating viewpoint;
- data requirement to support the control strategies from both planning and operating viewpoint;

On one hand, with the development of technologies and their market penetration, more and more data resources are available for traffic planning and operating such as VII (Vehicle Infrastructure Integration including cell phone, Toll Plaza Transponder, Blue Tooth Technology, …) beside road sensors (loop detectors, microwave radar, lidar, video camera, …). Discussion about how to integrate the data sources on a common platform is an active research area in the ITS community. On the other hand, it is necessary to look at what active traffic management strategies are feasible and are likely to be or proved to be effective by other countries such as those in Europe, and what is the minimum data requirement for implementing them.

- Overall picture of near-future Active Traffic Management System;
- What data system is required to support it;
- How to maximally use current traffic data systems in future development;
- Closely related challenging issues.

1.2 Current Traffic Data System

The current traffic data system is depicted in Figure 6.1. The raw data (occupancy and vehicle count) from all sensors, mainly the inductive loop detectors, are directly packed and sent to the TMC of each District (and/or PeMS in CA). As discussed in previous chapters, besides the error in sensor detection and reading, a communication system can also cause data loss. Therefore the data obtained at TMC often has errors or missed data. If such data are processed in TMC for traffic state parameter estimation and then used for free corridor or arterial traffic signal control and their coordination, there could be more problems appearing since the communication system would have to send back the processed data.

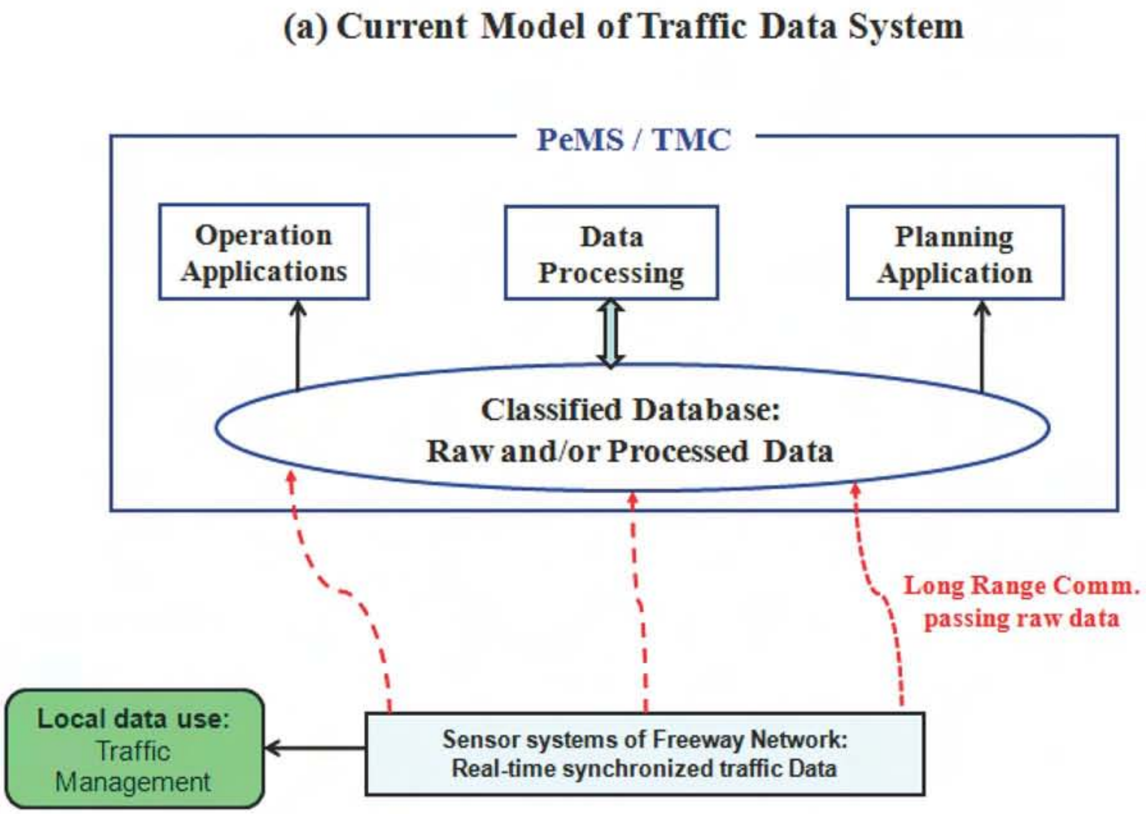


Figure 6.1. Current Traffic Data System

1.3 Integrated Traffic Data System

An integrated traffic data system is indispensible for both corridor management and operation. Traffic data for real-time operation requires more detailed data sets with higher update rate than traffic management such as traveler’s information. Traffic Operation includes ramp metering, VSL (Variable Speed Limit) or speed regulation, and incident/accident (location/time) detection and handling. For high performance of operation and fast response, sub-second data are necessary. For strategic planning, aggregated traffic data will be sufficient, which can be obtained from the operation data. Traffic data integration for corridor management means:

- Integrated data in time (synchronization) and space (converted into unified coordinate system)
- Integrated private and public data
- Integrated and optimal combination of sensor types and locations
- Integrated data from road-side and on-vehicle sensors

- Integrated communication systems for reliable sensor data passing
- Integrated data utility: traffic management, ATIS, strategic planning
- Integrated data analysis: cleansing, correction and imputation
- Integrated data processing: filtering/fusion/traffic parameter estimation
- Integrated systematic sensor/communication fault detection and remedy
- Integrated and classified database at PeMS or TMC level
- Integrated performance measurement

If the traffic data are corrected and processed locally at the control cabinet level directly using the event data as is discussed in Chapter 2, to generate the finest possible traffic state parameters such as speed, and density, (flow and occupancy are directly from the sensor), then the following advantages are obvious:

- those parameters can be directly used locally for traffic control (freeway ramp metering, arterial traffic signal control and the coordination between the two);
- only pass the traffic state parameter aggregated in some level back to District TMC and PeMS;
- if necessary, the raw occupancy and flow data can also be passed to TMC and achieve other uses such as sensor fault monitoring;
- TMC can use the collected data for system level Traffic Management and Operation Planning;
- TMC only needs to send back the system level Traffic Management high level command to local/corridor/arterial for execution;
- TMC/PeMS can further aggregate the traffic data for operational and planning purposes;
- Data loss due to communication would be much less;
- Traffic state parameters estimated using the raw event data at control cabinet level would cause much less time delay and estimation error, which will be more beneficial for Active Traffic Management.

As for the resolution of the data, different applications would have different requirements. For ATM, the following are proposed:

(1) Low Goal for local traffic responsive or coordinated RM and VSL

- Distance mean speed, average density, flow
- Data aggregation level: 30~60s

- Sensor Space: 500~750m

(2) High Goal: Accurate and reliable traffic state parameter estimation over time and space for ATM

- Distance mean speed, average density, flow
- Data aggregation level: 10s ~ 20s
- Sensor Spacing: 150m ~ 200m

This data system can be depicted as in Figure 6.2. It is clear that some communication infrastructure investment would be necessary at freeway corridor and arterial corridor level for coordination and control of the traffic in the corresponding level. However, off-the-shelf technologies are already there and the data packet (basically, processed traffic state parameters) is small.

1.4 Integrated Traffic Data System Maintenance

To keep the integrated traffic data system for healthy operation, it is necessary to maintain the system systematically and update the system with new technologies and products.

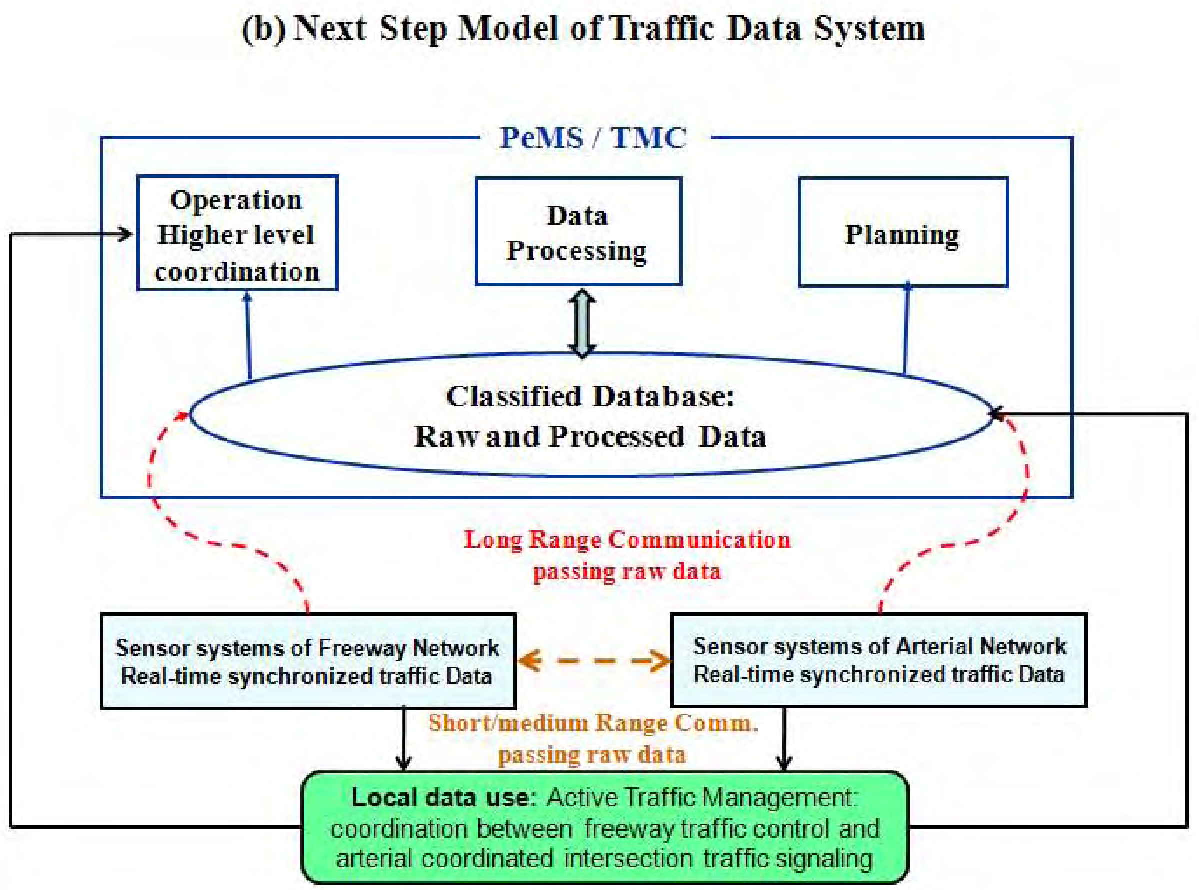


Figure 6.2. Expected Traffic Data System in near future

1.4.1 Systematic Faulty Data Analysis and Sensor Fault Detection

Inductive Loops are the main sensors widely used in California for traffic detection and monitoring. A systematic approach for loop fault detection and retrieving reliable data to support Corridor Management is crucial. However, loop data are not reliable. The error in the loop data obtained at TMC may be caused by fault at any point or several points from the loop to the TMC database: physical loop, connection of the loop and the control cabinet, loop card including sensitivity, communication media between control cabinet and TMC, and other software and hardware. Loop fault detection should be conducted in three levels: macroscopic (TMC/PeMS), mesoscopic (corridor or a stretch of freeway) and microscopic (control cabinet 170/2070). Fault detection at different levels look at the problems from different angles. They are complementary to each other. This approach also applies to other traffic sensors.

1.4.2 Systematic Loop Data Correction/Correction/Imputation

How to process the data, to cleanse measurement noise, to correct faulty data, and to impute filling up missed data to maximally achieve reasonable traffic parameter estimation are very important for applications. Those processes should be accomplished jointly with loop fault detection in the same levels, which include

- *Synchronization:* all the sensors' data time-stamped with GPS UTC
- *Conversion:* all sensor data converted to the same set of traffic parameters with the same metric under a unified coordinated system
- *Cleansing/correction*: removing measurement noise and correcting faults
- *Imputation:* Imputing lost data due to whatever reason if possible
- *Filtering/ aggregation*: Smoothing data to reduce measurement noise (linear, Kalman, or averaging)
- *Fusion:* Fusing data from multiple sensors for more accurate and reliable traffic parameter estimation

1.4.3 Developing Sustainable Communication System

Communication system is indispensible for corridor-wide and/or area-wide data collection, synchronization and integration. Such a system could be established or improved gradually with the development of technologies. Data loss is very common when passing through communication systems. Some suggestions for reliable communication are:

- Using fiber optics or GPRS modem if possible
- Using TCP as communication protocol with resending capability if possible
- Automatic communication fault diagnosis at macroscopic and mesoscopic levels
- Professional staff regularly checking/reporting sensor and communication faults
- Regular and in-time system maintenance
- Current traffic data are directly passed from control cabinet to TMC/PeMS. In the long run, traffic data system may need to be divided into three levels:
- Short/medium range communication in freeway and arterial corridor:
  Sensors and Control Cabinet ←→ Corridor Hub Computer and Database
- Long Range: Corridor Hub Computer and Database ←→ TMC or PeMS

1.5 Impact of VII (Vehicle Infrastructure Integration)

Data from on-vehicle electronics is getting more and more attention. With the development of VII, such information is a great source for traffic management. Here VII vehicles are in a rather general sense: (a) Probe vehicles of transportation agencies; (b) Vehicles equipped with VII DSRC; and (c) Vehicles equipped with combined Cell-phone and GPS. Roadside/inline sensors only provide point or short range measurements which heavily depend on sensor types, locations, density, measurement accuracy, and reliability. Compared to point road sensor measurement, VII vehicles provide continuous measurement over time. A single VII vehicle only provides a narrow moving distance-window measure. A certain percentage of market penetration of VII vehicles provide traffic measurements continuously over time and distance. It is expected that, with the increase of market penetration of VII vehicles and roadside infrastructure setups, merging/fusing point sensor data with VII data would bring enormous benefit for corridor traffic management.

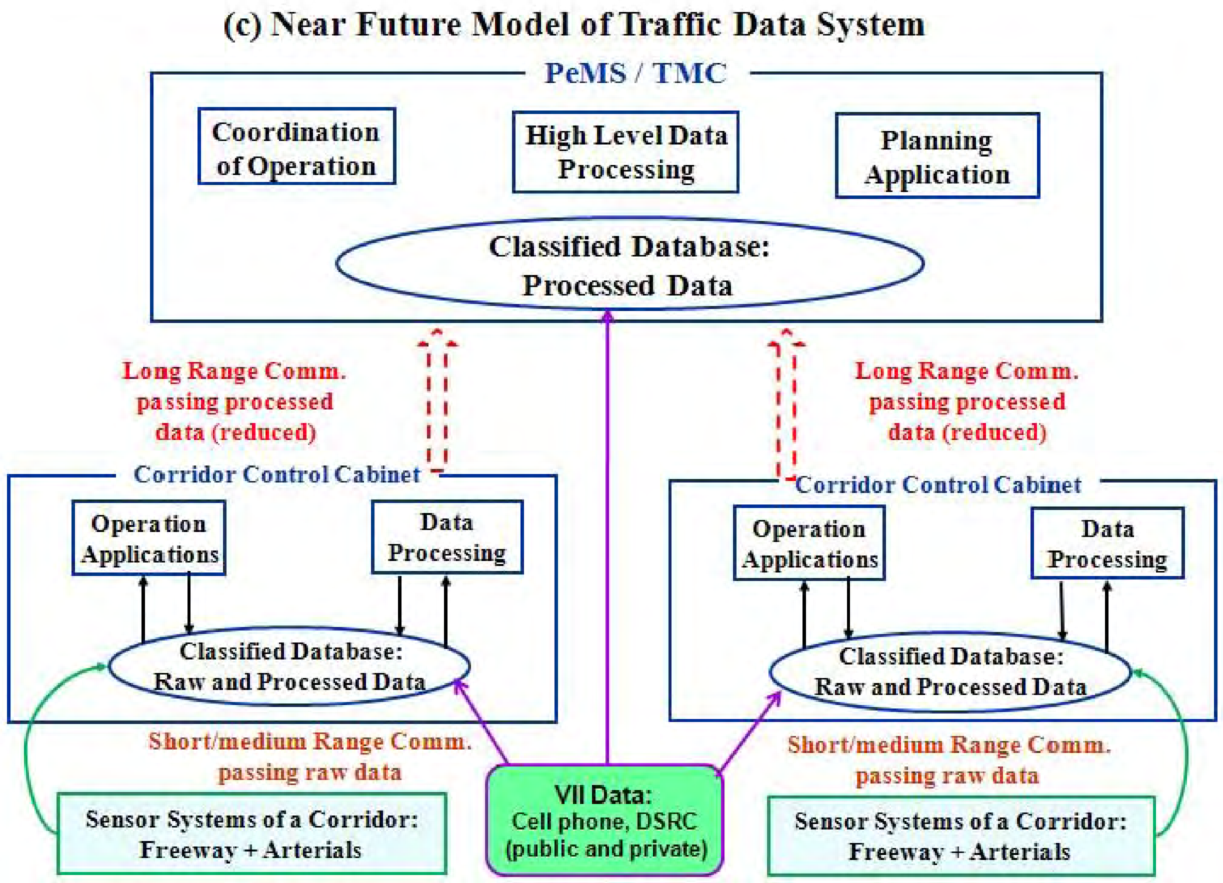


Figure 6.3 Integrated Traffic Data System to include all the sensors and VII data

However, future traffic data systems are likely to include both road sensors and on-vehicle sensors. This is based on the following considerations:

(1) On-vehicle sensors and road sensors are complementary to each other in function: road sensors such as loop detectors directly provide vehicle count or flow; while on-vehicle sensors alone could not provide such information directly unless they provide their position information and communicate with a fixed roadside server. The latter could estimate the flow at a fixed location.

(2) Time delay and cost of on-vehicle sensor data: If the on-vehicle senor data are collected by a private entity, then costs will be incurred. Besides, data passing from vehicle to private entity, then to TMC and then to traffic control system would lead to more time delay which is undesirable.

How to economically and efficiently use VII data for ATM with progressive market penetration of different technologies will be an important research topic for traffic data systems.

Chapter 7. Communication System Reconsiderations

7.1 Introduction to Chapter 7

Different sensor data would have different uses. It is necessary to develop an optimal strategy which can satisfy the data needs of traffic management and control and minimize the cost for data passing.

- If passing processed data, how to process at control cabinet?
- If passing reduced raw data, how to reduce data size such that it still keeps the original functionality for traffic parameter estimation?
- How to combine the data processing with communication system structure?
- Other related issues.

Loop detector data are usually used in different locations from where they were collected. The main uses of traffic data include traffic operation and planning, which are integrated in Active Traffic Management alone a corridor, which include

- Demand Management
- Lane (Capacity) Management
    - Lane changing assistance
- Traffic Management
    - Coordinated Ramp metering and Merging Assistance
    - Variable Speed Limit (VSL)

Communication systems are indispensible for corridor-wide and/or area-wide data collection, synchronization and integration. However, data loss is very common when passing through communication systems. Good communication systems are thus critical for passing the data. Data quality to the end user mainly depends on the following factors:

- Sensor detection error
- The way the data are pre-processed and packed for sending through the communication system
- Reliability of the communication system

There is usually a trade-off between the communication reliability and cost.

Since Phase 1 and 2 of the project were focused on the development of a Portable Loop Fault Detection Tool, the communication problem is only preliminarily considered. This chapter presents some work done in this area. Detailed study and recommendations will be conducted in the next phase of the project. Also, with the research and development of VII (Vehicle-Infrastructure-Integration), which in general sense, will include DSRC (Dedicated Short Range Communication) and Cell Phone, the data collection, processing, passing, storage and use will be affected greatly. We will discuss such impacts briefly in this chapter.

Although current freeway traffic control such as ramp metering is conducted at TMC where the integrated VDS data are available, future ATM application may be conducted at freeway and arterial corridor level using a data-control hub, where all the event data are processed for higher resolution traffic state parameter estimation, which could be directly used as input to the ATM controller. TMC could just receive processed and aggregated data and monitor the performance of the controller and send higher control commands to the data-control hub. The communication system also needs to possess the following capabilities:

- link a freeway corridor and its related arterials and surface streets to facilitate the coordination and control of all the subsystems which may belong to different jurisdictions;

- to reconfigure the physical links of communication if necessary in the event of an incident/accident or evacuation in a natural disaster such as serious earthquake or tsunami. Such links will be critical for traffic routing which may not be used in daily operations.

7.2 Sensor Data Passing with Communication Systems

Since traffic data passing is required for all sensors instead of just inductive loop detectors, they all will need a communication system. The following discussion will apply to all the sensors from which the data need to be passed.

Traffic raw data include those from sensors such as

- inductive loops
- video camera
- Sensys sensor
- radar/lidar

- WIM (Weigh in Motion) system
- Other sensors

Those raw data are massive and thus effective data transfer is directly related to the bandwidth of the communication system, either wireless or cable. High bandwidth costs much more and tends to be more unreliable as complained by Caltrans engineers. If we can reduce the size of the data significantly, Caltrans district TMC will be able to able to achieve the same goal for traffic management, control and traveler’s information with much less cost. The discussion in Section 2 shows that the traffic parameter estimation is very sensitive to update rate, which indicates that it is necessary to come up with an optimal strategy for Caltrans districts to collect and use the traffic data.

7.3 Current Situations of Communication Systems

We have investigated three Caltrans Districts: D4, D7 and D12, regarding the current situation of the available communication system. Our approach is to look at the PeMS data and use data analysis methods to identify possible data loss caused by the communication system. To achieve this, we will need detailed information for communication systems between each loop station to TMC in your District. It would be much appreciated if you could provide us a list of all the loop station numbers used in PeMS system under one of the following communication systems:

- Fiber optics
- Cell phone
- Old telephone line
- GPRS Modem
- CDPD Modem
- Other media (please specify the name)

It is also important for us to know what the communication protocol is: TCP or UDP. The difference between the two:

TCP: send the packet, the other end receives and acknowledgmes. If not received, the sender may resend the data to guarantee data reliability.

UDP: It does not use acknowledgments. It sends the packets without waiting on confirmation of received packet. Thus there is no resend even if data got lost.

7.3.1 Caltrans District 4

Mr. Ray Duschane (Tel: 510-286 5105; Email: Ray_Duschane@dot.ca.gov) is responsible for all the wireless communication systems for data passing in Caltrans D4. According to the discussion with Ray on March 27th 2007, all the loop data are on wireless. All the CCTV data passing with telephone line or fiber optics are independent from the communication system used for loop data. GPRS modems previously used dynamic IP addressing. It was then changed to Persistent IP, which is something between static IP and dynamic IP. The disadvantage of dynamic IP is that if there is problem, it may be necessary to take the modem back to TMC for manually rebooting.

From another discussion later on, all of the mainline inductive loop detectors here in the District 4 provided traffic data to Traffic Management Center (TMC) through wireless GPRS communications. In the field, these GPRS modems resided in cabinets along the side of the freeways. The loop detectors for a given location terminated in these cabinets. The traffic engineers identified each cabinet through a numbering system that begins with the letter DT or E (for example DT864 or E37CM). In PeMs, the cabinet numbers are referred to as MS ID's.

Now each cabinet would normally have one or two sets of mainline loop detectors depending upon whether the cabinet is monitoring one side of the freeway (for example, North or South) or both sides of the freeway (for example, East and West). One side of a freeway was referred as a station. In PeMs, these stations were referred to as VDS ID's. So an MS ID (cabinet) could have one or two VDS IDs (stations) associated with it in District 4.

Unfortunately, the traffic engineers did not know how to extract a list of District 04 MS ID's with associated VDS ID's from PeMs. Also as mentioned previously, for whatever reasons, the MS ID's in PeMs were not all populated with District 4 cabinet numbers. In other words, a lot of times one would find locations with VDS ID's, but the MS ID for that location would be blank. It was suggested to use PeMs to search for District 4 detector stations by routes or counties or etc., and PeMs would provide with all the detector locations for a given route, county, etc..

7.2.2 Caltrans District 7

The person responsible for the communication between 170 Cabinet and TMC in Caltrans District 7 was Mr. Alebachew Bekele (Tel: 323 259 1803; Email: Alebachew_Bekele@dot.ca.gov)

Based on the discussion with Mr. Bekele on March 27, 2007, 40% ~ 45% District wide communication does not work properly for daily operation. The communication system used:

- 98% on telephone and fiber optics
- 1~2% by wireless

District 7 emphasized the problem caused by maintenance and construction. They complained of insufficient resources/engineers to maintain the system. Possible ways for evaluation include:

- Choosing certain locations which are working
- Providing support to District 7 to make the system 90% working before evaluation

District 7 provided a detailed list of loop detector stations and the communication system used to pass the data to TMC.

7.4 Communication Systems Failure Comparison

7.4.1 Decomposing Communications Down Errors

The PeMS diagnostic approach has always been to assign failure modes to the individual loop detectors. When PeMS receives data samples it can test the samples for reasonableness and deduce a number of different types of errors. But when PeMS receives no data samples it's not possible to tell if the detector itself is bad. For example, it could be that we aren't receiving any data samples from any detector simply because the FEP is down. For situations like this we used to mark the detector as bad with a reason of *communication down*. But in reality this failure category has always been a catch-all category that simply represents that we didn't receive any data samples.

It turns out that we can do better than this if we know the physical data collection infrastructure. By physical data collection infrastructure we mean knowing which detectors are connected to which controllers, and which controllers are connected over which lines to the FEP. For example, if we know that all of the detectors that are connected to a single controller are not sending any data samples, in other words, we're not receiving any data from the controller, then it's likely that the controller itself is broken. For example there could be no power at the controller, or it could have been damaged in an accident (or removed during construction). In a similar manner, if we're not receiving any data from any of the controllers on a line then it's likely that the line itself is bad rather than every controller is bad.

In response to this we've taken out the diagnostic state *communications down* and we've substituted in the states of *controller down*, *line down* and *no data*. If we don't receive any samples from a line then the line is considered to be down and all of the detectors on the line are marked as line down. If we receive samples from some controllers on a line but not others then the controllers are considered to be down and the loops belonging to those controllers are marked as controller down. And if we receive samples from some detectors but not others for a single controller then the non-reporting detectors are considered to be down but not the controller and are marked as no data. It should be pointed out that we're assuming that the line and/or the controller is the problem if we don't receive data samples from the detectors underneath them. But it could be that all of the individual loops are broken (we consider this to be unlikely, though).

The locations where this scheme doesn't work are the districts where we don't have the physical topology of the data collection infrastructure, or the data collection infrastructure is just a star. District 4 doesn't provide us with their data collection infrastructure and District 10 has only wireless modems that are sending data back to a centralized point. In the D10 case the rollup algorithm still works but the results are blank for the line level.

The easiest way to detection communication system healthy is to add one more byte in the communication packet, which is the time step count. At each time step of sending, the counter is incremented by 1. The count is reset to 0 once it reaches 127. With this extra byte, the receiver can detect if there is any packet lost at any time step. The receiver can also tell if the loop data is healthy.

7.4.2 Communication Failure Rate Comparison

Based on the D7 communication list, figure out the location of the detector using a specific type of communication method (e.g. Old Telephone Line, Fiber and Temp-wireless (CDPD)). Notice that only Freeway I-5 consists of these three communication methods. Therefore, I-5 is chosen for the analysis. The analysis steps are summarized as follows:

(1). Data source: http://pems.dot.ca.gov , Facilities & Devices > Field Elements > Controllers: go to D7 and choose the freeway I5-SB on 2/25/2011;

(2). Download the LDS information. Filter by Post Mile where the same communication methods are used. Then LDS (Loop Detector Station) using specific communication methods are captured;

(3). Find out the loop status. And pick out the loops which are in ―Ctlr down”. Data source: http://pems.dot.ca.gov , Data Quality > Detector Health > Detail;

(4). This error indicates that none of the detectors attached to the same controller as the selected detector are reporting data. This probably indicates no power at this location or the communication link is broken. Since other station nearby have data, this means that communication down is the most likely problem;

(5). By comparing the percentage Ctlr Down of different communication methods, we can judge which communication method is better. All the percentage is calculated relative to the communication error of the temporary wireless. The results are summarized as follows:

Figure 7.1 Communication Failure Percentage of different media: Old Telephone line and Fiber are less than 1/3 of the Temporary Wireless

Based on the diagram shown above, the Temp-wireless (CDPD modem) is the worst communication method regarding reliability. The old telephone line is the most reliable communication link. Fiber optics cable is similar to the old telephone line with slightly less reliability. It is noticed that here only one day’s data wasused for the analysis. More reliable results should be derived by analyzing data during a larger time span.

7.5 Preliminary Recommendations on Communication System

Some suggestions for reliable communication are:

- Using available old telephone line as much as possible since it is cheap and reliable;
- Using fiber optics is possible and if the budget could afford for medium to long distance communication and data passing, and for heavy traffic corridors if old telephone line is not available;
- Using GPRS modem for short to medium range communication where old telephone line is not available and Fiber optics cable is too expensive;

- Using TCP instead of UDP as communication protocol with resending capability if possible to ensure data logging;
- Automatic communication faults diagnosed at macroscopic and mesoscopic levels
- Professional staff regularly checking/reporting sensor and communication faults
- Regular and in-time system maintenance

Current traffic data are directly passed from control cabinet to TMC/PeMS. In the long run, traffic data system may need to be divided into three levels:

- TMC or PeMS
- Corridor
- Freeway/Arterial Sections

Communication system needs changing accordingly

- Short/medium range communication:
- Sensors and Control Cabinet ←→ Corridor Hub Computer and Database
- Long Range: Corridor Hub Computer and Database ←→ TMC or PeMS

7.5.1 Coordination between Divisions within a Caltrans District

Regular coordination between Divisions of Caltrans Districts and maintenance contractors are necessary. After any road maintenance, Caltrans District electrical engineers responsible for loop detector stations need to check if the engineers for the road maintenance have kept/restored the loop detector stations in working condition. A written report should be provided by the road maintenance contractors regarding the impact of the road maintenance on the loop detection health, before and after the work.

7.5.2 Communication Links and Reconfiguration Capability

Although current freeway traffic control such as ramp metering is conducted at TMC where the integrated VDS data are available, future ATM application may be conducted at freeway and arterial corridor level using a data-control hub where all the event data are processed for higher resolution traffic state parameter estimation, which could be directly used as input to the ATM controller. TMC could just receive processed and aggregated data and monitor the performance of the controller and send higher control commands to the data-control hub. The communication system also needs to possess the following capabilities:

- link a freeway corridor and its related arterials and surface streets to facilitate the coordination and control of all the subsystems which may belong to different jurisdictions;
- to reconfigure the physical link of communication if necessary in the event such as incident/accident or evacuation in a natural disaster such as serious earthquake or tsunami. Such a link will be critical for traffic routing which may not be used in daily operations.

To guarantee high quality of data, the communication must be reasonably reliable. For current communications used in data passing, old telephone lines and fiber optical cable are three times more reliable than wireless UDPD modems. Besides the system itself, some suggestions on communication protocol to avoid data loss, a simple method for automatic detection of communication fault, and the connection to meet system re-configuration capability in ATM have been proposed.

Chapter 8. Recommendations for Future Research

This chapter presents some follow-on research and implementation related to the results in this report.

8.1 Field Test of Traffic Speed and Vehicle Length Estimation Using Single Loop Event Data

The algorithm developed in this study on traffic speed estimation using single loop detector event data has been tested from multiple day archived data, which seem to be reliable. However, the sensitivity parameter needs to be adjusted with respect to a specific loop detector. It is necessary to refine the algorithm to adjust the sensitivity parameter automatically, for example, for a specific time interval when the traffic is known free-flow. Then with a free-flow speed assumed, it is possible to use this speed as the ground truth for sensitivity parameter calibration. This process could be conducted for several days in non-peak hours for consistency. Then it could be fixed for the specific control cabinet.

Since the implementation of the algorithm will have a significant impact on traffic detection and monitoring, it is suggested to have a project on the field test of the algorithm for selected sites where event data are available. This will also include the interface with the traffic controller such as 170 or 2070 and use the controller to run the speed estimation algorithm. It would be more efficient if such a project is linked with an Active Traffic Management (Variable Speed Limit or Coordinated Ramp Metering) project since the results could be directly used for support or as evaluation. This will be an IT related project.

Vehicle length based classification is important to traffic operation, freight movement analysis and planning. This is imperative to some freeway corridors to a freight hub such as a Seaport or a center of warehouses. When the speed estimation from single loop event data is available, vehicle length estimation is also available as a by-product. Those two can be field tested in the same project.

8.2 Further Development on Portable Loop Fault Detection Tool (PLFDT)

Although a prototype of a PLFTD has been developed, convenient access of event data is not available in most cases. It is known that using IST detector cards can retrieve raw loop signal in much higher frequencies, other cards may not have such capability. The most popularly used Reno 222 cards do not even allow an interface. Therefore, higher frequency raw loop data has to be obtained from the traffic controller. Besides, interfacing with the 170 controller is not straightforward. Therefore, an efficient interface method is necessary for higher frequency event data for online or off-line loop fault detection.

This project has used multi-lane-multi-vehicle tracking algorithm. Based on the laptop computer used for the project, it was possible to track vehicles in one lane simultaneously from low to moderately high traffic flow. However, the tracking algorithm needs refining for more reliable tracking since some missed tracking has been observed.

A GUI (Graphical User Interface) needs to be developed to facilitate automatic loop fault detection of multi-lane-multi-loop stations.

8.3 Field Test of Congestion Onset Detection Algorithm

Quick congestion onset detection is critical to traffic operations. The algorithm developed for quick congestion onset detection in Chapter 5 of this report relies on the good estimation of speed based dual loop detectors. However, if the speed estimation from single loop detectors is adequately accurate, the algorithm can also be implemented for a freeway corridor with healthy single loop detections. It is suggested to select a freeway corridor with heavy traffic for limited field test of the algorithm. The freeway corridor may have VDS of dual loop stations or single loop stations. It would be better if the distances between the stations are within 500m. In general, shorter VDS distance would lead to quicker detection. The project would include tasks to interface with all the control cabinets for event data and to communicate with a central computer which will also run the detection algorithm. It is suggested that such a project be coupled with an ATM project for higher efficiency and low cost since the latter requires a similar system setup.

8.4 Pilot Project for Developing Integrated Traffic Data System

Based on the consideration of progressively developing an integrated hierarchical traffic data system in Chapter 6, it would be necessary to select a traffic system with a freeway corridor and related arterial corridor(s) with medium to heavy traffic demand to both. Such a system would be ideal for a pilot project for developing an integrated traffic data system. Here the coordination of all the VDS along a freeway, the coordination of all intersections along the arterial, and a higher level coordination between freeway and arterial(s) are all necessary for integrated traffic control of the overall system. Developing such a system may face problems in data system structure, data resolution and acquisition, communication link, interface with different control cabinets and controllers, different types of VDS sensors, different control system and software, and institutional issues. It is suggested that such a project be coordinated with a project on integrated traffic control of freeway and arterials, or a related ATM project for higher efficiency and lower cost.

References

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

[2] Chan L. and May A. D., 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., 1986.

[3] Chen, C., Kwon, J., Rice, J., Skabardonis, A., and Varaiya, P., Detecting errors and imputing missing data for single loop surveillance systems, 82nd TRB Annual Meeting, January 2003, Washington, D.C.

[4] Chen, L., and May, A., Traffic Detector Errors and Diagnostics, *Transportation Research Record 1132*, TRB, Washington, DC, 1987, pp 82-93.

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

[6] Chen, C., K. Petty, A. Skabardonis, and P. Varaiya. Freeway Performance Measurement System: Mining Loop Detector Data. Transportation Research Record No.1748, Transportation Research Board, Washington, D.C., 2001, pp 96-102

[7]. Chen, C, Skabardonis, A. and Varaiya, P., Systematic identification of freeway bottlenecks, 83rd TRB Annual Meeting, Jan. 2004, Washington, DC.

[8] Cheevarunothai, P., Wang, Y. and Nihan, N.L. (2005). Development of advanced loop event data analyzer (ALEDA) system for dual-loop malfunction identification, The 12th World Congress on ITS in San Francisco.

[9] Cheevarunothai, P., Wang, Y. and Nihan N.L. (2006). Identification and correction of dual-loop sensitivity problems, Transportation Research Record, No. 1945, 73-81.

[10]. P. Cheevarunothai, Y. Wang, and Nancy L. Nihan, Using Dual-Loop Event Data to Enhance Truck Data Accuracy, Transportation Research Record 1993, TRB Washington D. C., 2007, P131-137

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

[12] Christopher M. Day, Thomas M. Brennan, Jr., Matthew L. Harding, Hiromal Premachandra, Allen Jacobs, Darcy M. Bullock, James V. Krogmeier, and James R. Sturdevant, Three-Dimensional Mapping of Inductive Loop Detector Sensitivity with Field Measurement, Transportation Research Record, No. 2128, 2009, p. 35–47.

[13] Cleghorn, D., Hall, F., and Garbuio, D., Improved Data Screening Techniques for Freeway Traffic Management Systems, *Transportation Research Record 1320*, TRB, Washington, DC, 1991, pp 17-31.

[14] Coifman, B., Using Dual Loop Speed Traps to Identify Detector Errors, 78th TRB Annual Meeting, Washington, DC., Jan. 1999

[15] Coifman, B. Using dual loop speed traps to identify detector errors. Transportation Research Record no. 1683, 1999, pp47-58.

[16] Coifman, B., D. Lyddy, A., Skabardonis,The Berkeley Highway Laboratory-building on the I-880 field experiment, Proc. IEEE ITS Conf., Dearborn, MI, 2000

[17] Coifman, B., Improved velocity estimation using single loop detectors, *Transportation Research A*, 35, 863-880, 2001

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

[19] Coifman, B., Dhoorjaty, S., and Lee, Z., Estimating Median Velocity Instead of Mean Velocity at Single Loop Detectors, Transportation Research: Part C, Vol. 11(3-4), 2003, pp 211-222.

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

[21] Coifman, B., Dhoorjaty, S., Event Data Based Traffic Detector Validation Tests, *ASCE Journal of Transportation Engineering,* Vol 130, No 3, 2004, pp 313-321.

[22] Dailey, D. Travel time estimation using cross-correlation techniques. Transportation Research, Part B, Vol. 27, 1992, pp97-107.

[23] Dailey, D. A statistical algorithm for estimating speed from single loop volume and occupancy measurements. Transportation Research, Part B, Vol. 33, 1999, pp313-322.

[24] Day, C. M., T. M. Brennan, T. M., Jr., M. L. Harding, H. Premachandra, A. Jacobs, D. M. Bullock,J. V. Krogmeier, and J. R. Sturdevant, Three-Dimensional Mapping of Inductive Loop Detector Sensitivity with Field Measurement, TRR No. 2128: J of TRB, Washington,

D.C., 2009, pp. 35–47.

[25] FHWA, Traffic Detector Handbook: Third Edition—Volume I, Pub. No. FHWA-HRT-06-108, Oct. 2006

[26] Hazelton, M. L., Estimating vehicle speed from traffic count and occupancy data, J. of Data Science, 2, 2004, pp231-244

[27] Hellinga B. Improving Freeway Speed Estimates from Single-Loop Detectors, Journal of Transportation Engineering, Vol. 128(1), 2002, pp. 58-67.

[28] Ingram, J., *The Inductive Loop Vehicle Detector: Installation Acceptance Criteria and Maintenance Techniques*, California Department of Transportation, Sacramento, CA, 1976.

[29] Jia, Z., C. Chen, B. Coifman, and P. Varaiya. The PeMS Algorithms for Accurate, Real-Time Estimates of g-factors and Speeds from Single-Loop Detectors. Proc. 2001 IEEE ITSC, Oakland, CA, USA, 2001, pp. 536-541.

[30] Jacobson, L, Nihan, N., and Bender, J., Detecting Erroneous Loop Detector Data in a Freeway Traffic Management System, *Transportation Research Record 1287*, TRB, Washington, DC, 1990, pp 151-166.

[31] Kell, J., Fullerton, I., and Mills, M., *Traffic Detector Handbook, Second Edition*, Federal Highway Administration, Washington, DC, 1990.

[32] Kim, Z., Realtime Object Tracking based on Dynamic Feature Grouping with Background Subtraction, Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2008.

[33] Kim, Z., Geometry of Vanishing Points and its Application to External Calibration and Real-time Pose Estimation, *Institute of Transportation Studies, Research Reports*, Paper UCB-ITS-RR-2006-5, July 1, 2006.

[34]. 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.

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

[36]. Lu, X. Y. and Coifman, B., Highway Traffic Data Sensitivity Analysis, California PATH Working Paper – Final Report, UCB-ITS-PRR-2007-3**,** Univ. of California, Berkeley

[37]. Lu, X. Y. and A., Skabardonis, Freeway Traffic Shockwave Analysis: Exploring the NGSIM Trajectory Data, 86th Annual Meeting, Transportation Research Board, Washington,

D.C., #07-3016, January 21-25, 2007.

[38] Lu, X. Y., Varaiya, P., Horowitz, R. and Palen, J., Faulty loop data analysis/correction and loop fault detection, accepted to 15th ITS World Congress, November 16-20, 2008, New York

[39] Lu, X. Y., Kim, Z., Cao, M., Varaiya, P., Horowitz, R., and Palen, J., Portable loop fault diagnosis tool development for control cabinet level, CD ROM of 88th TRB Annual Meeting, Washington D. C., Jan. 2009

[40] Lu, X. Y., A. Skabardonis and T. Z. Qiu, Using Sub-second Dual Loop Station Data for Congestion Onset Detection, CD ROM of 2nd International Symposium on Freeway and Tollway Operations, Honolulu, Hawaii -- June 21-24, 2009

[41] Lu, X. Y., and J. Palen, Integrated Data System Structure for Active Traffic Management - Planning and Operation, NATMEC**,** June 21-24 2010, Seattle, WA.

[42] Lu, X. Y., P. Varaiya, R. Horowitz, Dual loop data correction at microscopic level*,* NATMEC**,** June 21-24 2010, Seattle, WA.

[43] May, A., Coifman, B., Cayford, R., Merritt, G., Automated diagnostics of loop detectors and the data collection system in the Berkeley Highway Laboratory, California PATH Research Report, **UCB-ITS-PRR-2004-13**

[44] May,A., R. Cayford, L. Leiman, G. Merritt, BHL Traffic Detector Analysis, Consolidation of BHL Detector System at CCIT, and Development of Portable Detector Diagnostic Tool, California PATH Research Report, UCB-ITS-PRR-2005-24

[45]. 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

[46] Nihan, N., Aid to Determining Freeway Metering Rates and Detecting Loop Errors, *Journal of Transportation Engineering*, Vol 123, No 6, ASCE, November/December 1997, pp 454-458.

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

[48] Payne, H.J. and Thompson S., Malfunction Detection and Data Repair for Induction-Loop Sensors Using I-880 Data Base, TRB, *Transportation Research Record, 1570*, p.191-201

[49] Payne, H.J., E.D. Helfenbein and H.C. Knobel. Development and testing of incident detection algorithms, FHWA-RD-76-20, Federal Highway Administration, Washington DC 1976.

[50]. Persaud, B. and Hall F. L. (1989) Catastrophe theory and patterns in 30-second freeway traffic data – implications for incident detection. Transpn *Res.* 23A. p103-I 13.

[51] Petty, K., Freeway Service Patrol (FSP) 1.1: The Analysis Software for the FSP Project, California PATH Research Report, UCB-ITS-PRR-95-20

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

[53] Petty, K., Hishan Noeimi, Kumud Sanwal, Dan Rydzewski, Alexander Skabardonis, and Pravin Varaiya, Haithan Al-Deek, The Freeway Service Patrol Evaluation Project: Database, Support Program and Accessibility, California PATH Report,

[54] Petty, K., H. Noeimi, K. Sanwal, D. Rydzewski, A. Skabardonis, and P. Varaiya, and H. Al-Deek, The Freeway Service Patrol Evaluation Project: Database, Support Program and Accessibility, California PATH Report,

[55] Rakha, H., and W. Zhang, Estimating traffic stream space-mean speed and reliability from dual and single loop detectors, Transportation Research Record: Journal of the Transportation Research Board, No. 1925, 2005, pp. 38-47

[56] Rajagopal, R. and Varaiya, P., Health of California's Loop Detector System: Final Report for PATH TO 6300, 2007

[57] Robinson, S. and Polak, J. W., ILD data cleaning treatments and their effect on the performance of Urban Link Travel Time models, 85th TRB Annual Meeting, Jan. 2006

[58]. 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.

[59] Skabardonis, A., Petty, K., Noeimi, H., Rydzewski, D. and Varaiya, P. (1996). I-880 Field Experiment: Data-Base Development and Incident Delay Estimation Procedures, Transportation Research Record 1554 , TRB, pp 204-212

[60]. Stephanedes, Y. J. and Chassiakos A. P., (1993) Freeway incident detection through filtering. *Transpn Res. 1C.* p.219-233.

[61] Turochy, R.E. and B.L. Smith. A New Procedure for Detector Data Screening in Traffic

Management Systems. Transportation Research Record 1727, Transportation Research Board, Washington, D.C., 2000, pp 127-131

[62] Wang Y. and Nihan L. N. Can Single-Loop Detectors Do the Work of Dual-Loop Detectors?, Journal of Transportation Engineering, Vol. 129(2), 2003, pp. 169-176.

Appendix A. Loop Sensitivity Factor Test Record

Run #	**Max Spd**	**Vehicle type**	**Road surface**		**Run #**	**Max Spd**	**Vehicle Type**	**Rod surface**
1	10	Audi Car	dry		21	20		dry
2	10	Audi Car	dry		22	20	Ford WinStar Van	dry
3	15	Audi Car	dry		23	25	Ford WinStar Van	dry
4	15	Audi Car	dry		24	25	Ford WinStar Van	dry
5	20	Audi Car	dry		25	30	Ford WinStar Van	dry
6	20	Audi Car	dry		26	30	Ford WinStar Van	dry
7	25	Audi Car	dry		27	35	Ford WinStar Van	dry
8	25	Audi Car	dry		28	35	Ford WinStar Van	dry
9	30	Audi Car	dry		29	40	Ford WinStar Van	dry
10	30	Audi Car	dry		30	40	Ford WinStar Van	dry
11	35	Audi Car	dry		31	10	Ford WinStar Van	dry
12	35	Audi Car	dry		32	10	Ford WinStar Van	dry
13	40	Audi Car	dry		33	15	Truck tractor	dry
14	40	Audi Car	dry		34	15	Truck tractor	dry
15	45	Audi Car	dry		35	20	Truck tractor	dry
16	45	Audi Car	dry		36	20	Truck tractor	dry
17	10	Ford WinStar Van	dry		37	25	Truck tractor	dry
18	10	Ford WinStar Van	dry		38	25	Truck tractor	dry
19	15	Ford WinStar Van	dry		39	30	Truck tractor	dry

20	15	Ford WinStar Van	dry		40	30	Truck tractor	dry
Run #	**Max Spd**	**Vehicle type**	**Road surface**		**Run #**	**Max Spd**	**Vehicle Type**	**Rod surface**
41	10	Truck tractor	wet		61	35		wet
42	10	Truck tractor	wet		62	35	Ford WinStar Van	wet
43	15	Truck tractor	wet		63	40	Ford WinStar Van	wet
44	15	Truck tractor	wet		64	40	Ford WinStar Van	wet
45	20	Truck tractor	wet		65	10	Audi Car	wet
46	20	Truck tractor	wet		66	10	Audi Car	wet
47	25	Truck tractor	wet		67	15	Audi Car	wet
48	25	Truck tractor	wet		68	15	Audi Car	wet
49	30	Truck tractor	wet		69	20	Audi Car	wet
50	30	Ford WinStar Van	wet		70	20	Audi Car	wet
51	10	Ford WinStar Van	wet		71	25	Audi Car	wet
52	10	Ford WinStar Van	wet		72	25	Audi Car	wet
53	15	Ford WinStar Van	wet		73	30	Audi Car	wet
54	15	Ford WinStar Van	wet		74	30	Audi Car	wet
55	20	Ford WinStar Van	wet		75	35	Audi Car	wet
56	20	Ford WinStar Van	wet		76	35	Audi Car	wet
57	25	Ford WinStar Van	wet		77	40	Audi Car	wet
58	25	Ford WinStar Van	wet		78	40	Audi Car	wet
59	30	Ford WinStar Van	wet		79	45	Audi Car	wet
60	30	Ford WinStar Van	wet		80	45	Audi Car	wet

Appendix: B

Traffic Speed Estimation Using Single Loop Event Data Compared with Those from Dual Loop Event Data

The following are more plots for speed estimation using single loop event data compared with those estimated from dual loops as described in Chapter 3. The purpose is to show the effectiveness and reliability of the algorithm. Data Source: BHL (Figure 2.2) on 03/01/2011, for Station 1, 2, 3, 6, and 7 in both East Bound and West Bound. For Station 1 EB, the speed comparison and the raw and mode duration for single loop and raw and moving-window (MW) Averaged dual-loop duration are plotted. Only the speeds are plotted for all the other stations and directions.

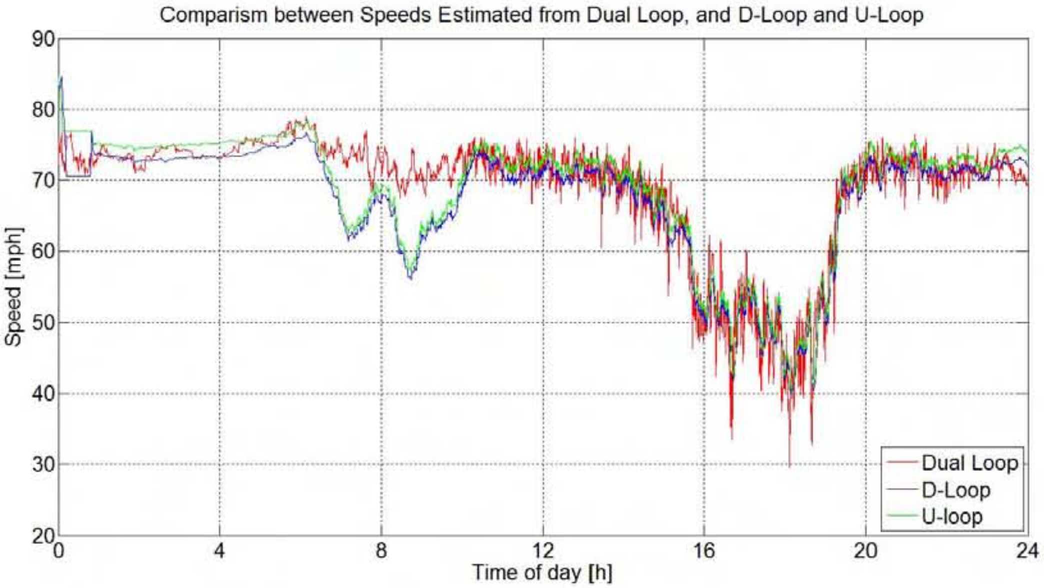


Figure B. 1 BHL EB, Station 1, Lane 1, Speed estimation comparison

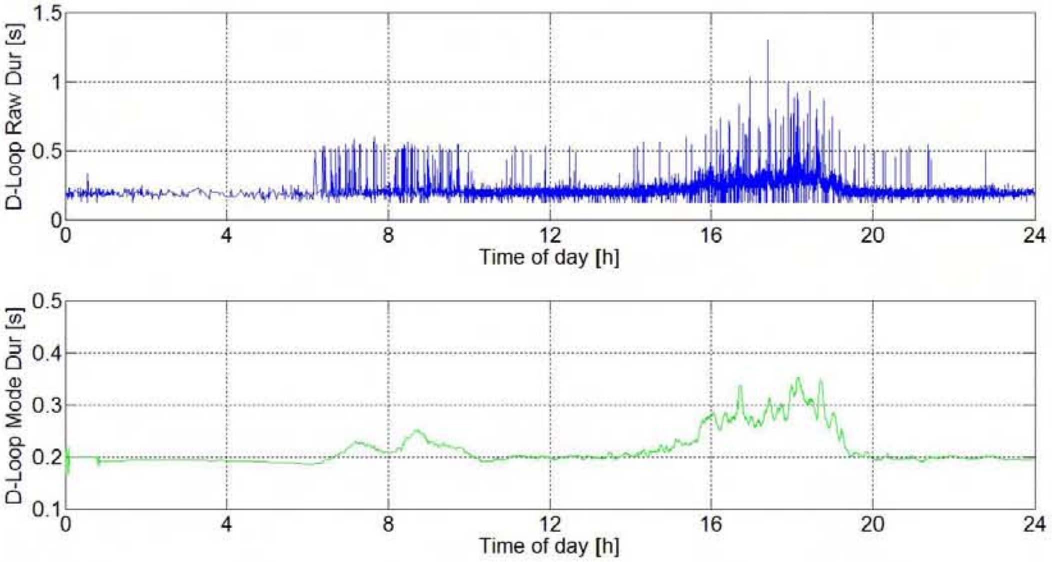


Figure B. 2 BHL EB, Station 1, Lane 1, Raw and Mode Durations: D-loop

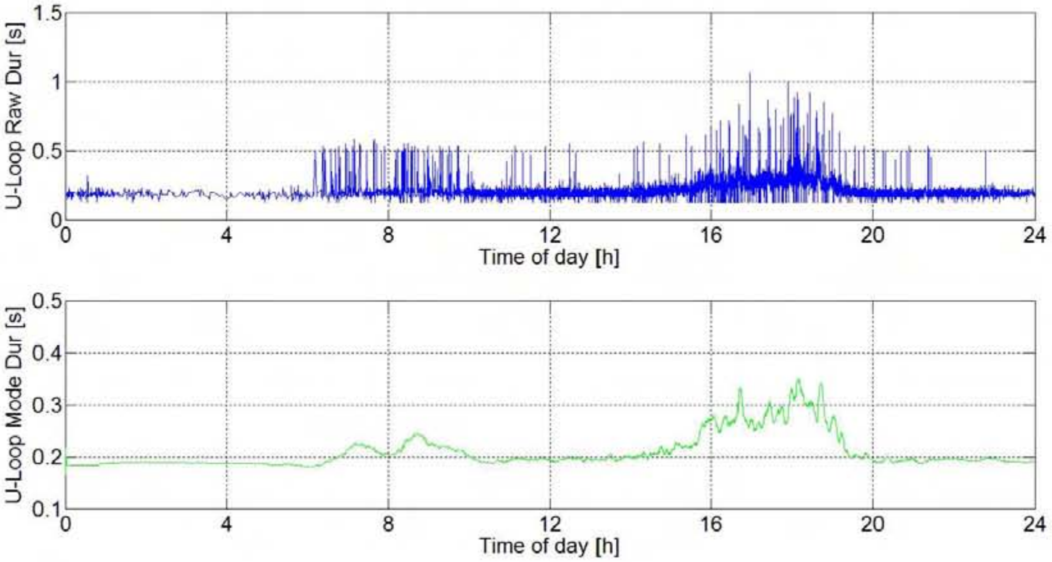


Figure B. 3 BHL EB, Station 1, Lane 1, Raw and Mode Durations: U-loop

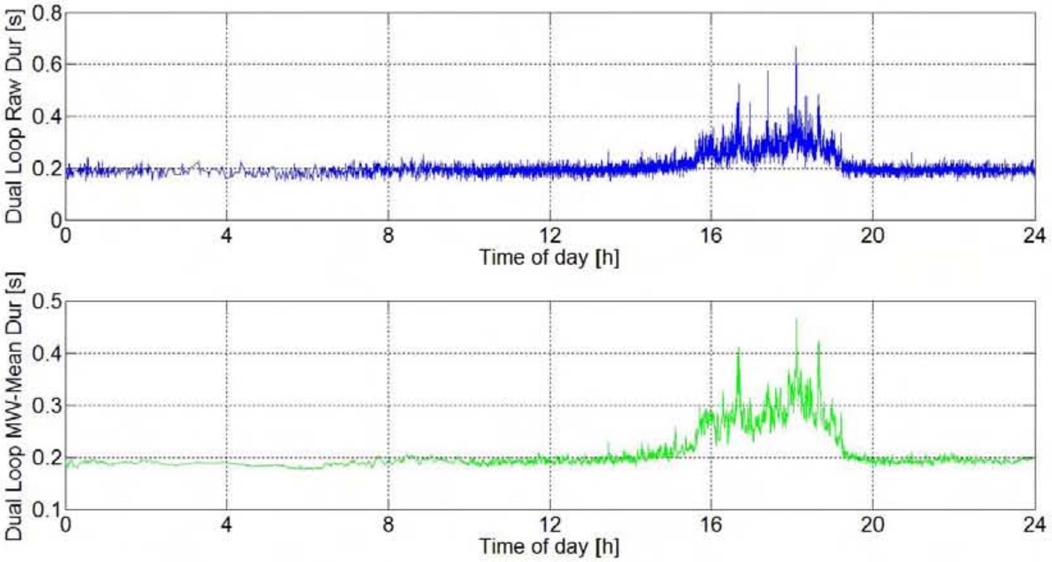


Figure B. 4 BHL EB, Station 1, Lane 1, Raw and MW Averaged Duration of Dual-Loops

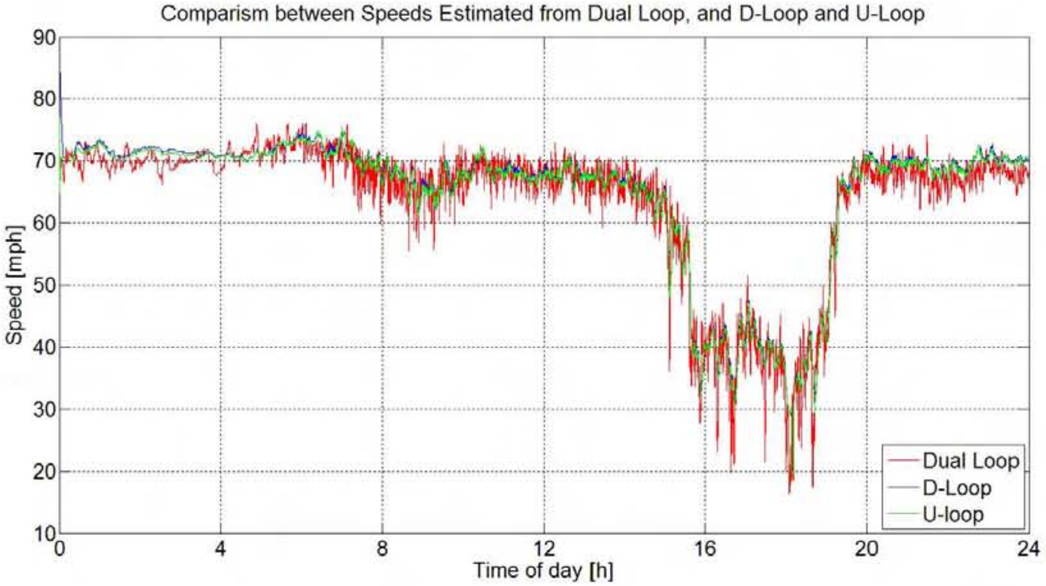


Figure B. 5 BHL EB, Station 1, Lane 2, Speed estimation comparison

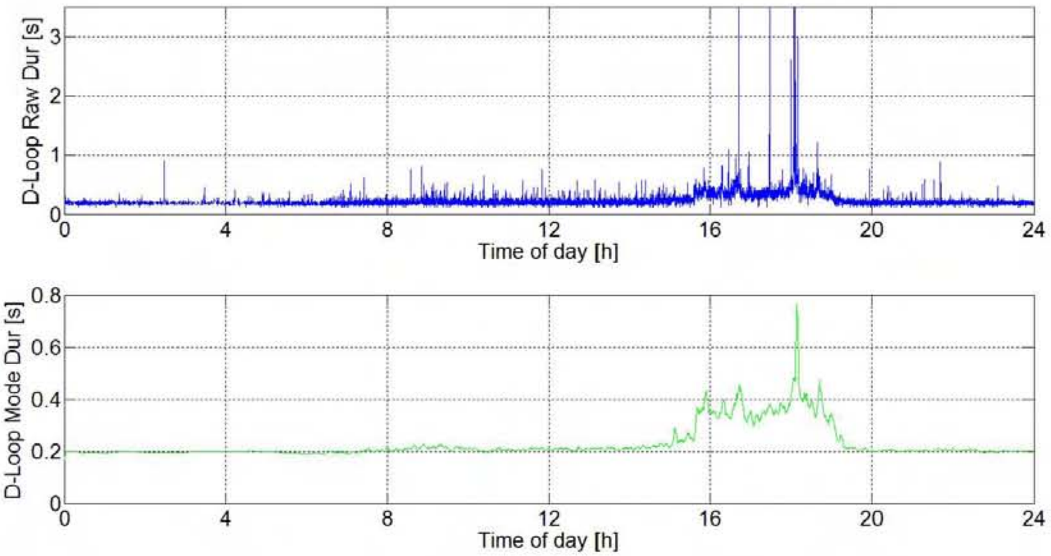


Figure B. 6 BHL EB, Station 1, Lane 2, Raw and Mode Durations: D-loop

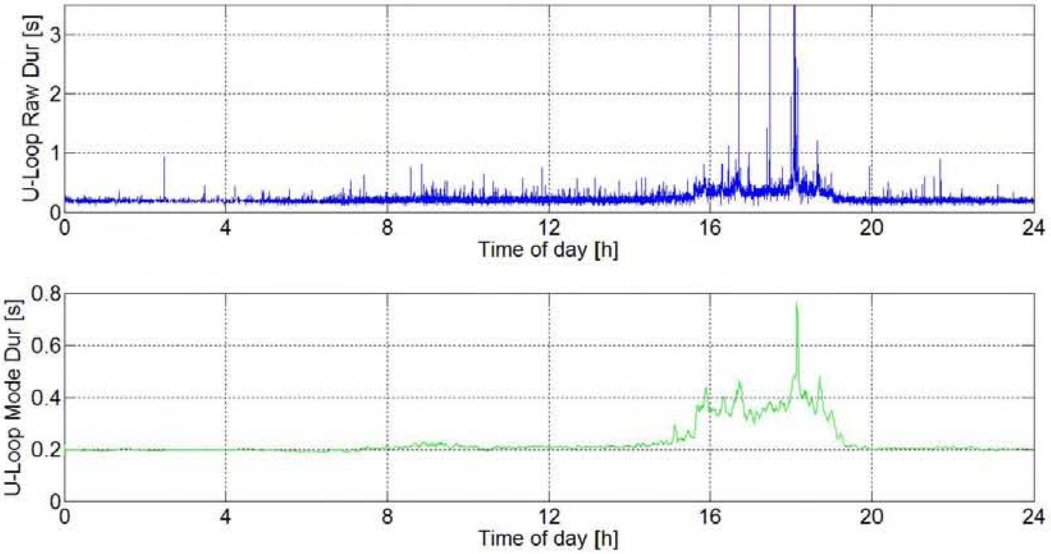


Figure B. 7 BHL EB, Station 1, Lane 2, Raw and Mode Durations: U-loop

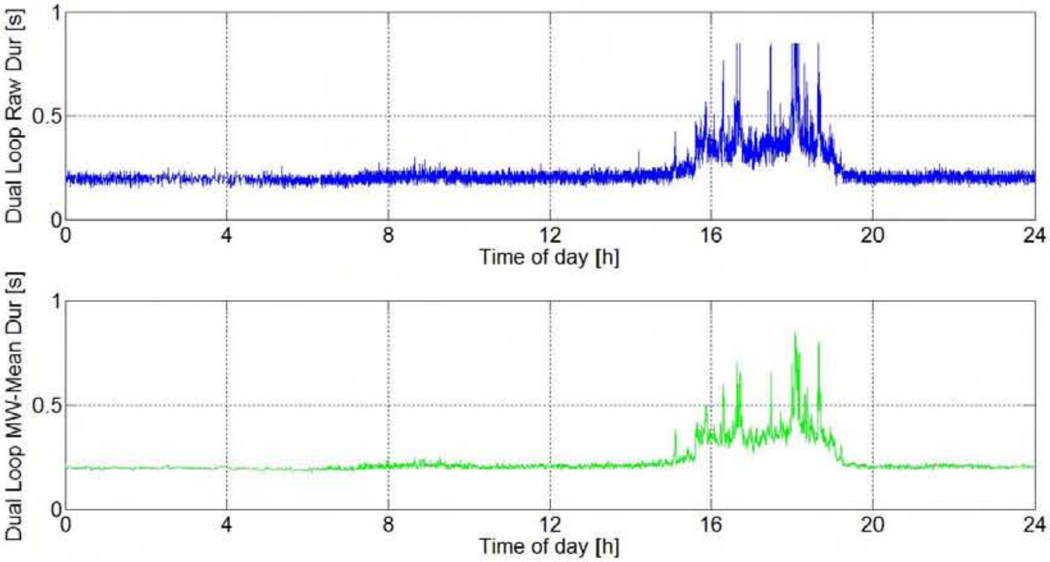


Figure B. 8 BHL EB, Station 1, Lane 2, Raw and MW Averaged Duration of Dual-Loops

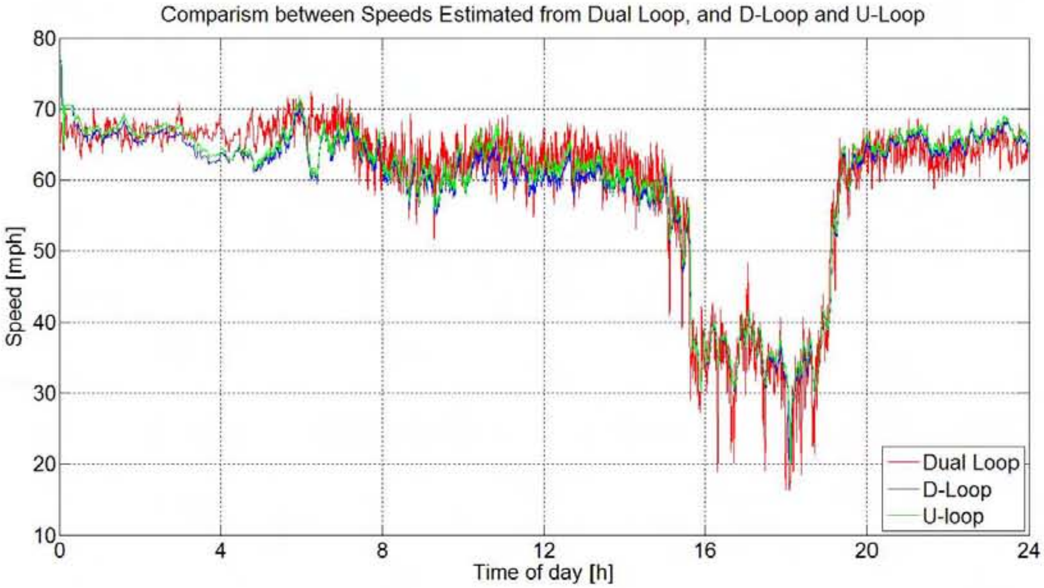


Figure B. 9 BHL EB, Station 1, Lane 3, Speed estimation comparison

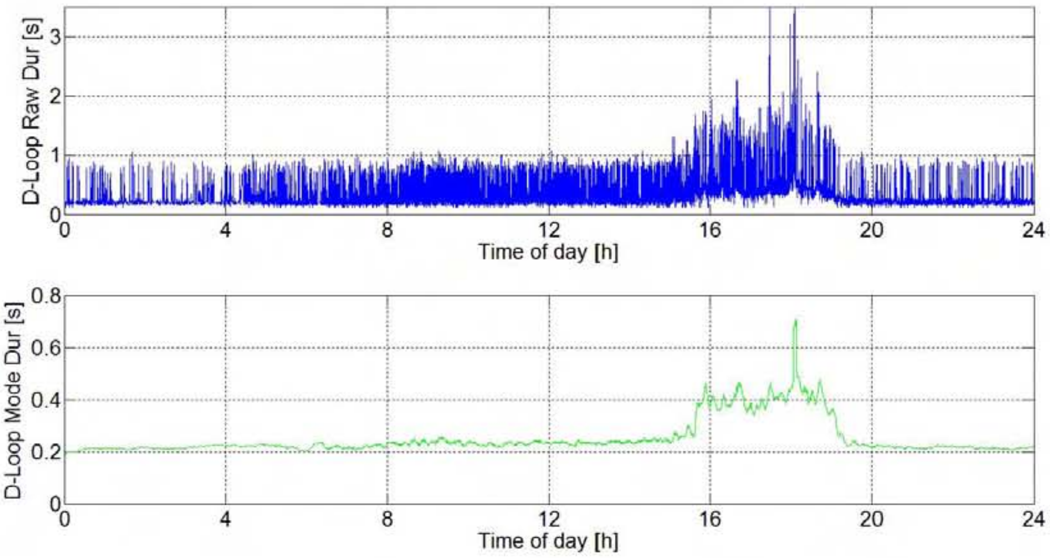


Figure B. 10 BHL EB, Station 1, Lane 3, Raw and Mode Durations: D-loop

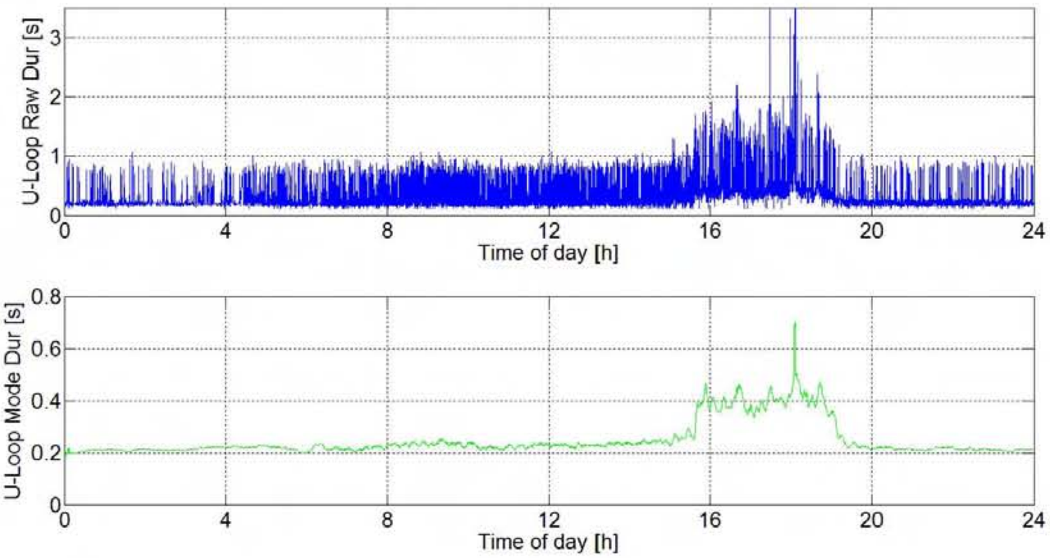


Figure B. 11 BHL EB, Station 1, Lane 3, Raw and Mode Durations: U-loop

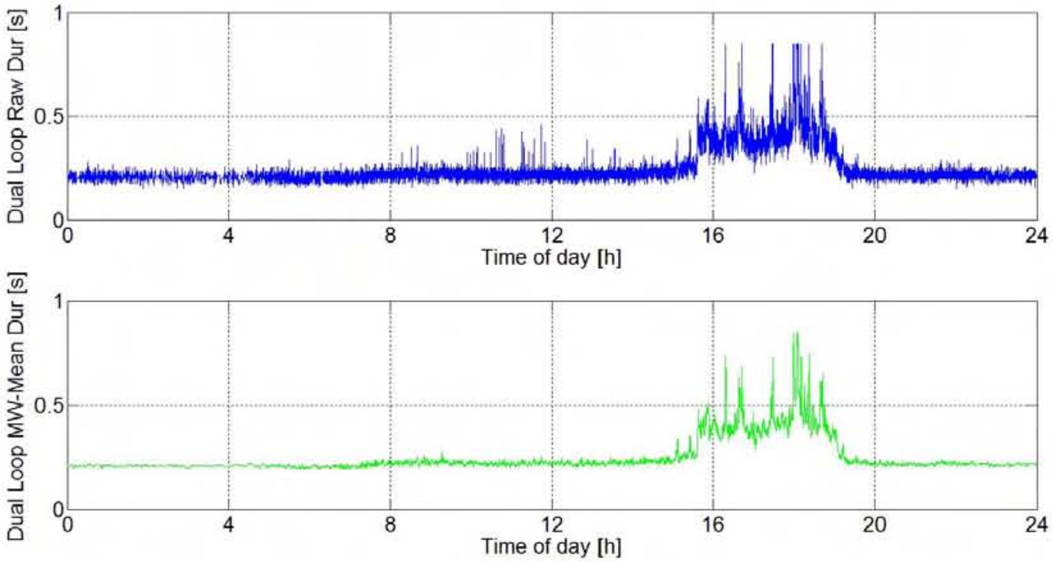


Figure B. 12 BHL EB, Station 1, Lane 3, Raw and MW Averaged Duration of Dual-Loops

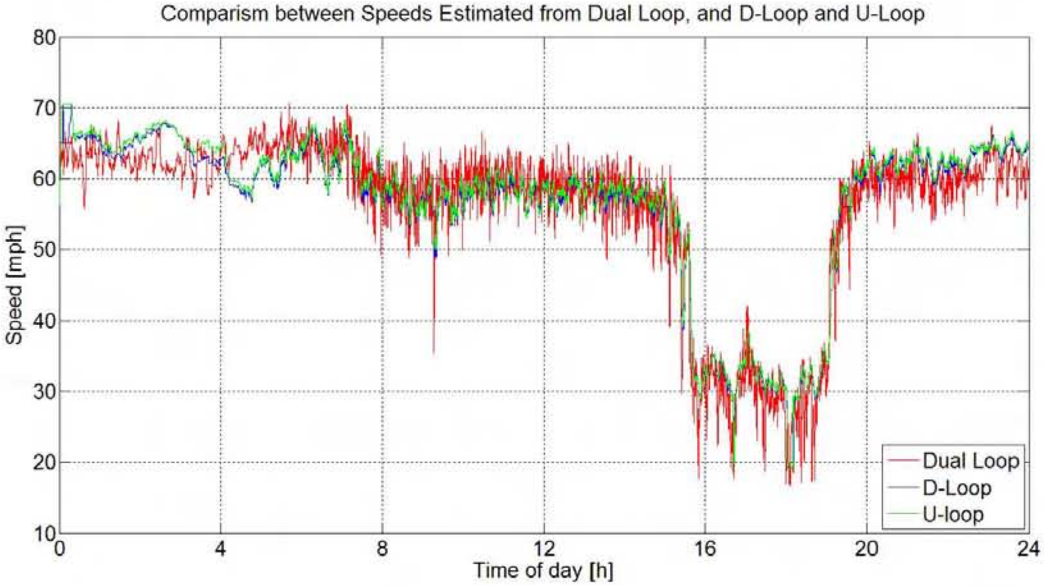


Figure B. 13 BHL EB, Station 1, Lane 4, Speed estimation comparison

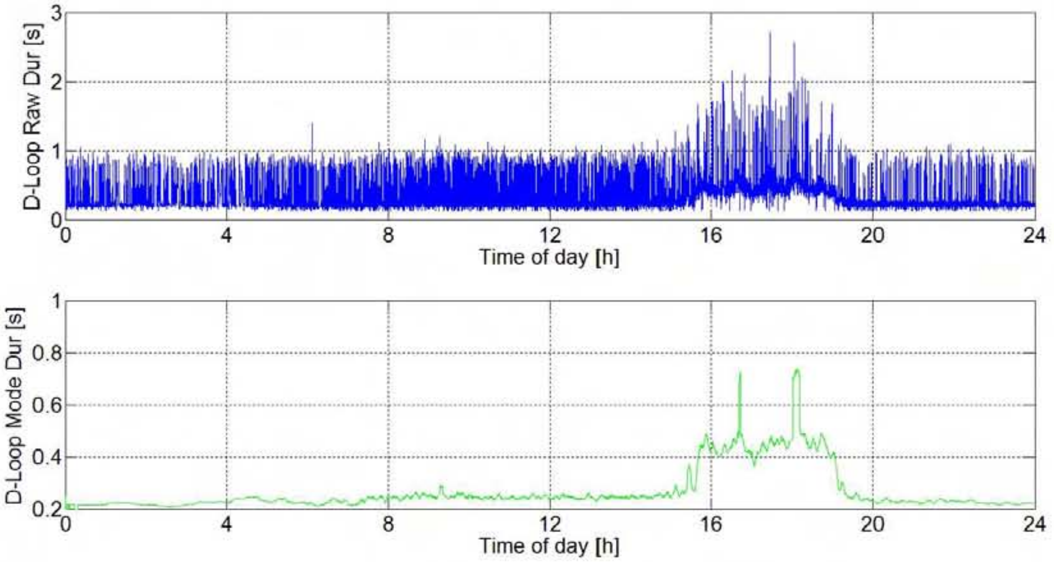


Figure B. 14 BHL EB, Station 1, Lane 4, Raw and Mode Durations: D-loop

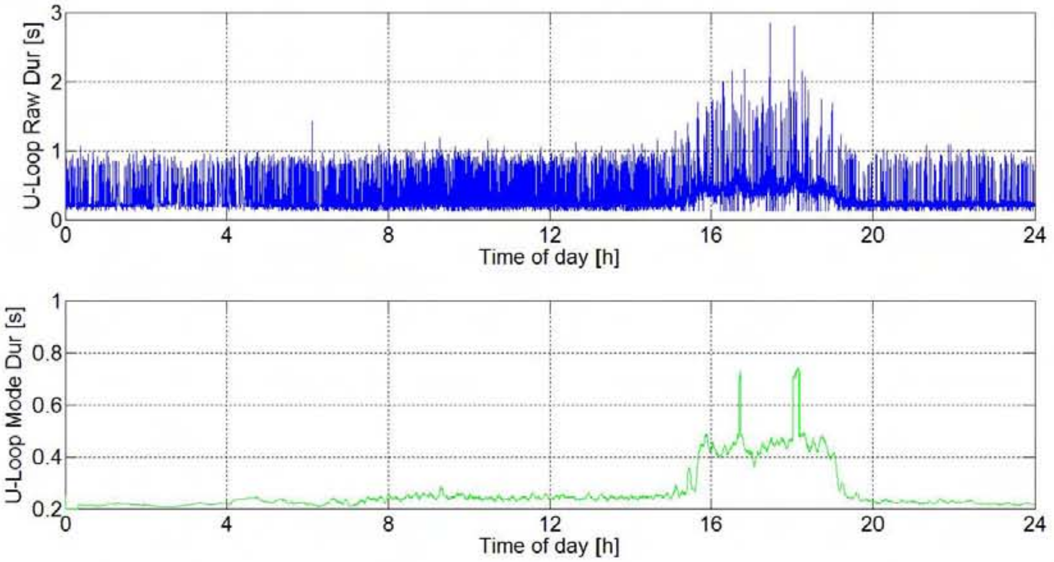


Figure B. 15 BHL EB, Station 1, Lane 4, Raw and Mode Durations: U-loop

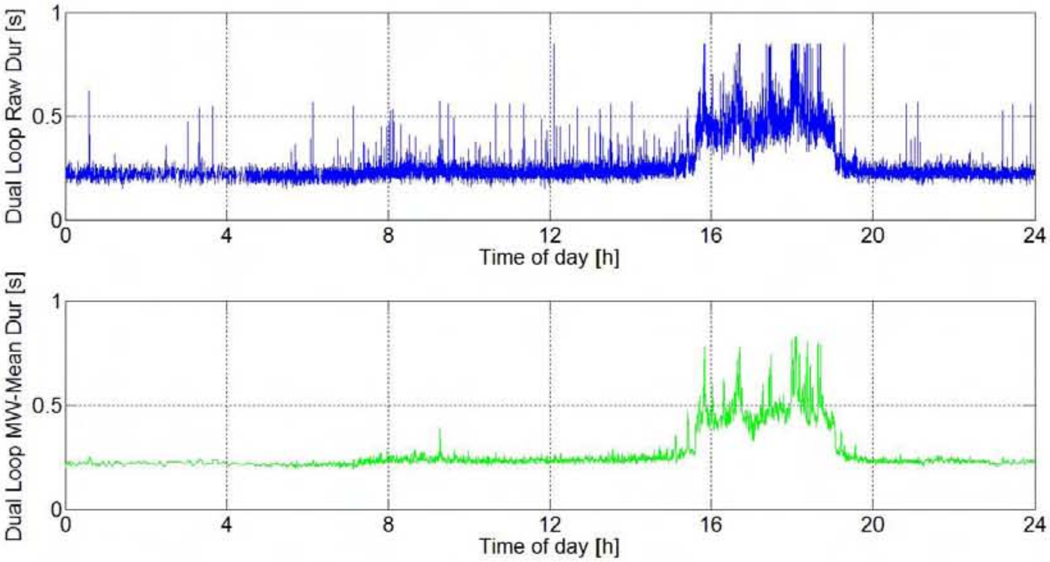


Figure B. 16 BHL EB, Station 1, Lane 4, Raw and MW Averaged Duration of Dual-Loops

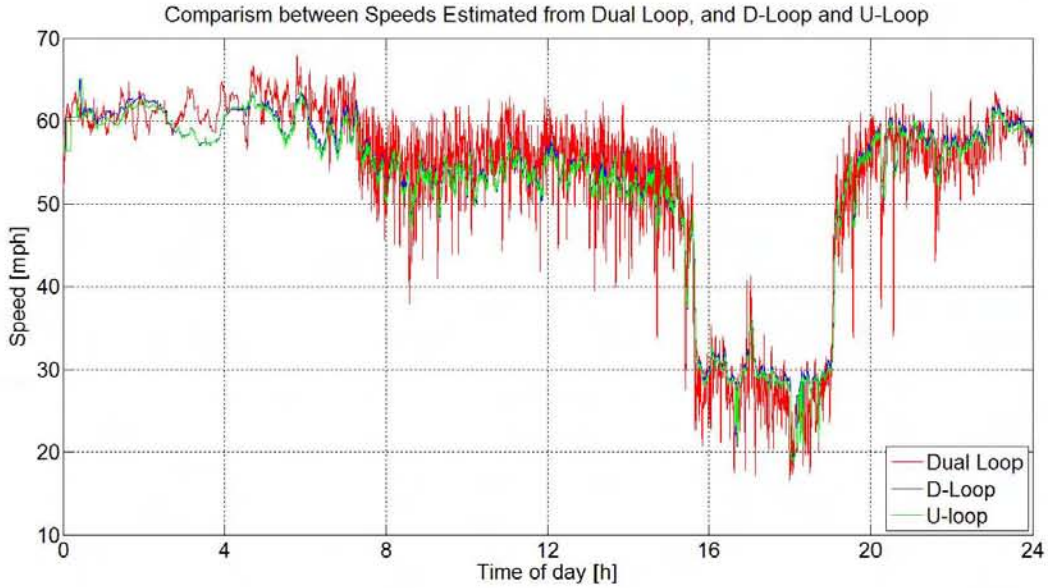


Figure B. 17 BHL EB, Station 1, Lane 5, Speed estimation comparison

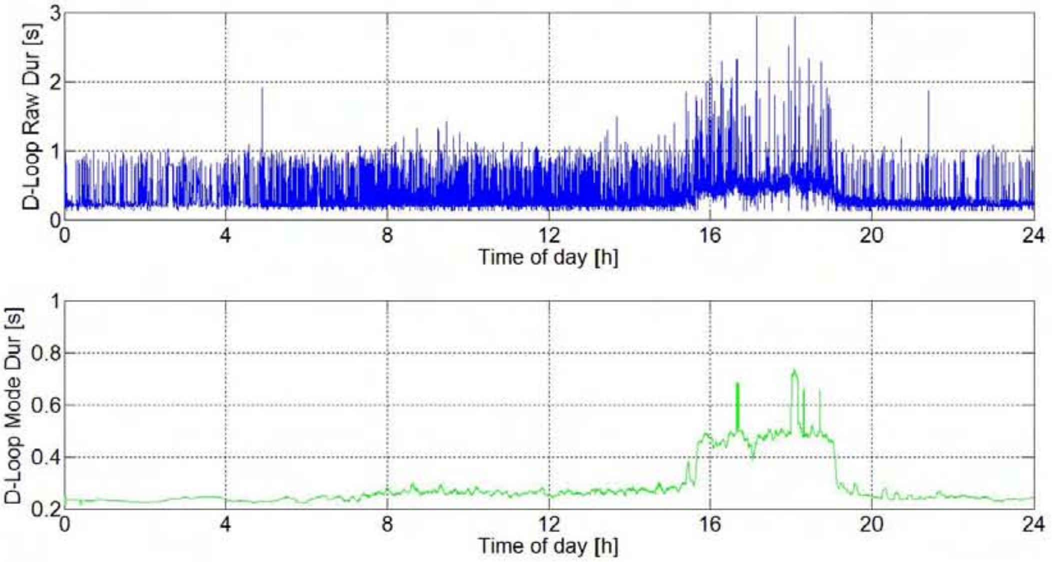


Figure B. 18 BHL EB, Station 1, Lane 5, Raw and Mode Durations: D-loop

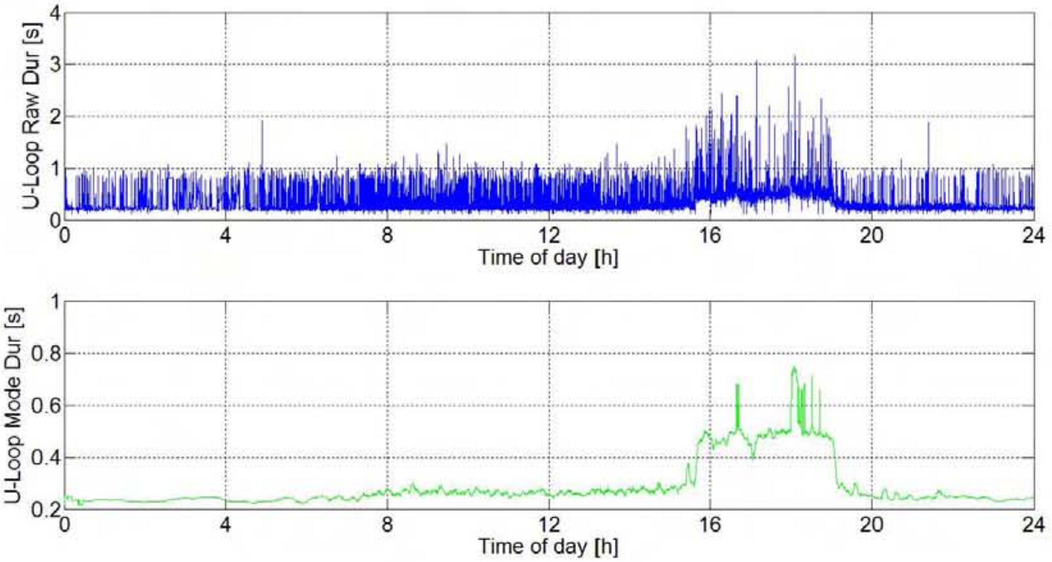


Figure B. 19 BHL EB, Station 1, Lane 5, Raw and Mode Durations: U-loop

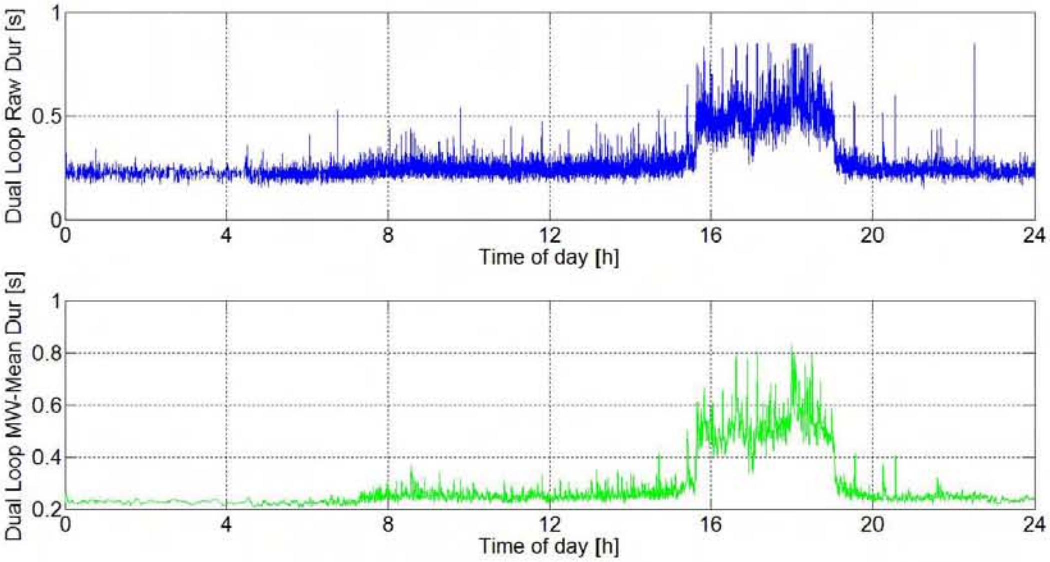


Figure B. 20 BHL EB, Station 1, Lane 5, Raw and MW Averaged Duration of Dual-Loops

****

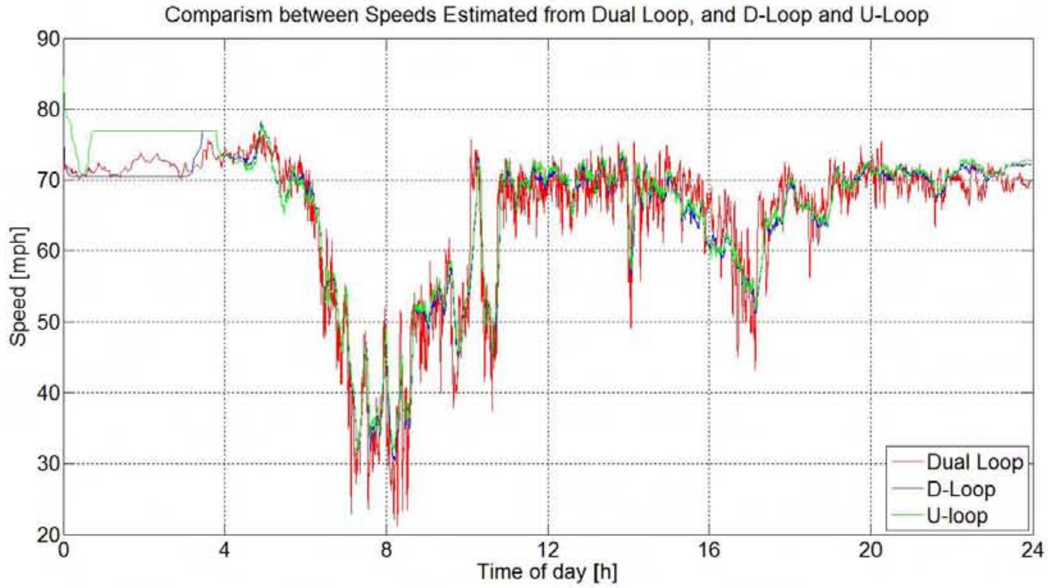


Figure B. 21 BHL WB, Station 1, Lane 1, Speed estimation comparison

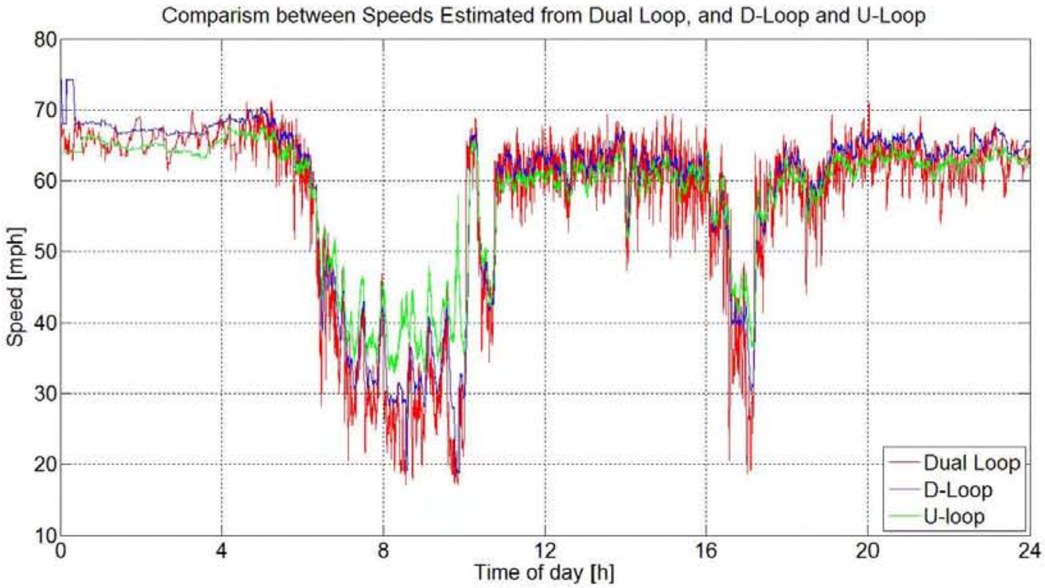


Figure B. 22 BHL WB, Station 1, Lane 2, Speed estimation comparison

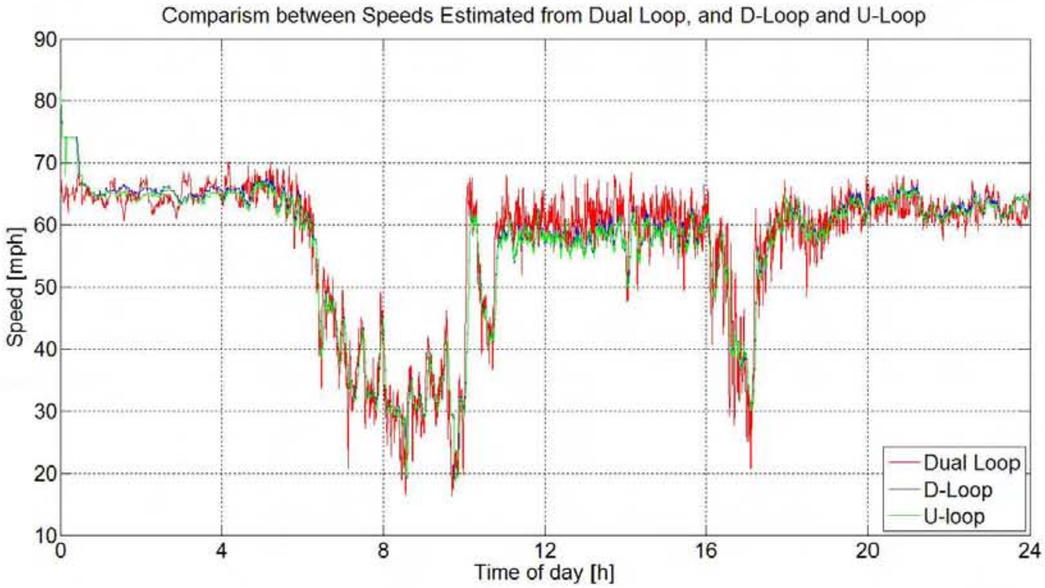


Figure B. 23 BHL WB, Station 1, Lane 3, Speed estimation comparison

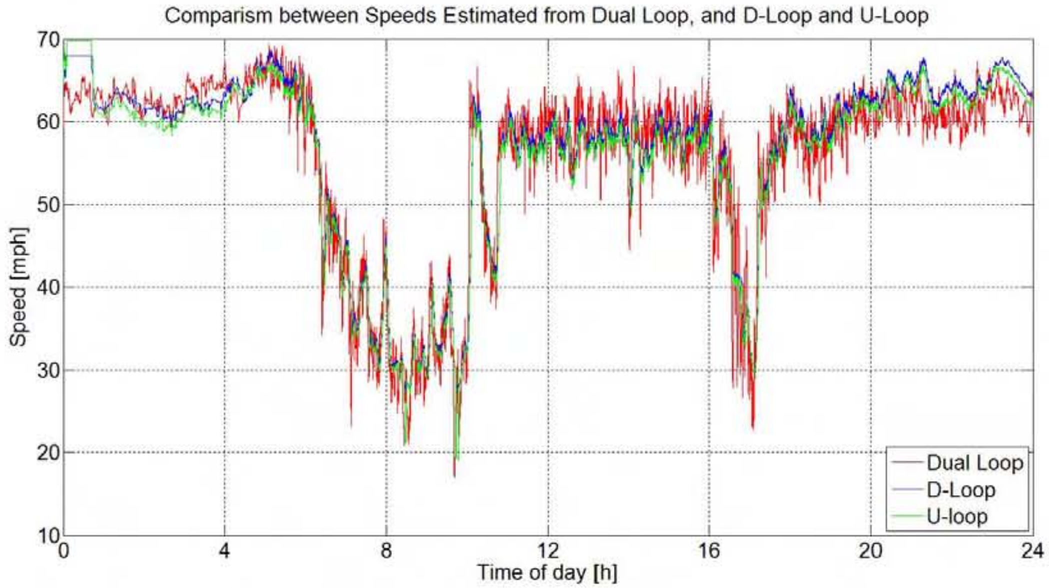


Figure B. 24 BHL WB, Station 1, Lane 4, Speed estimation comparison

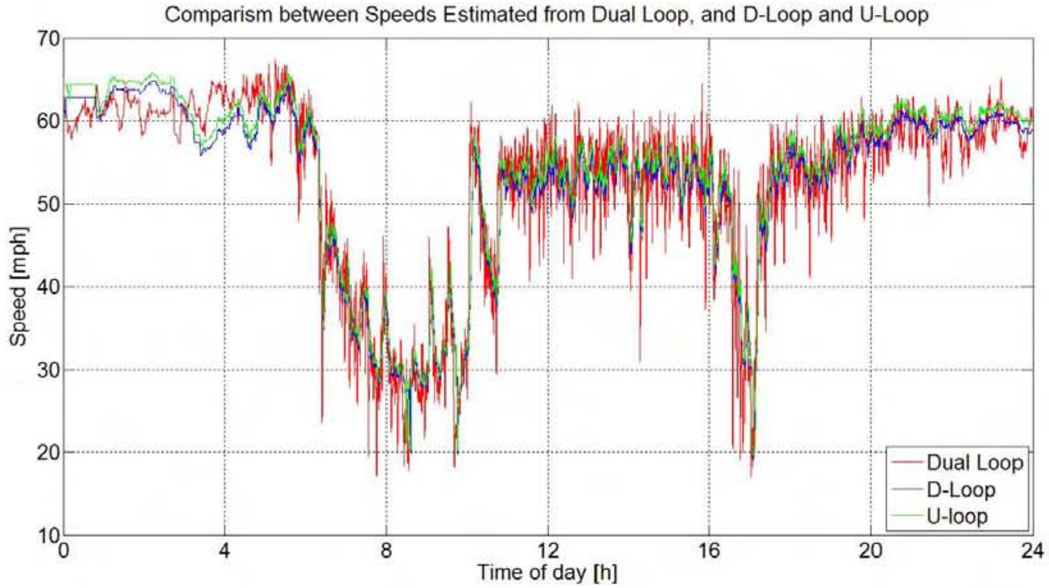


Figure B. 25 BHL WB, Station 1, Lane 5, Speed estimation comparison

****
****

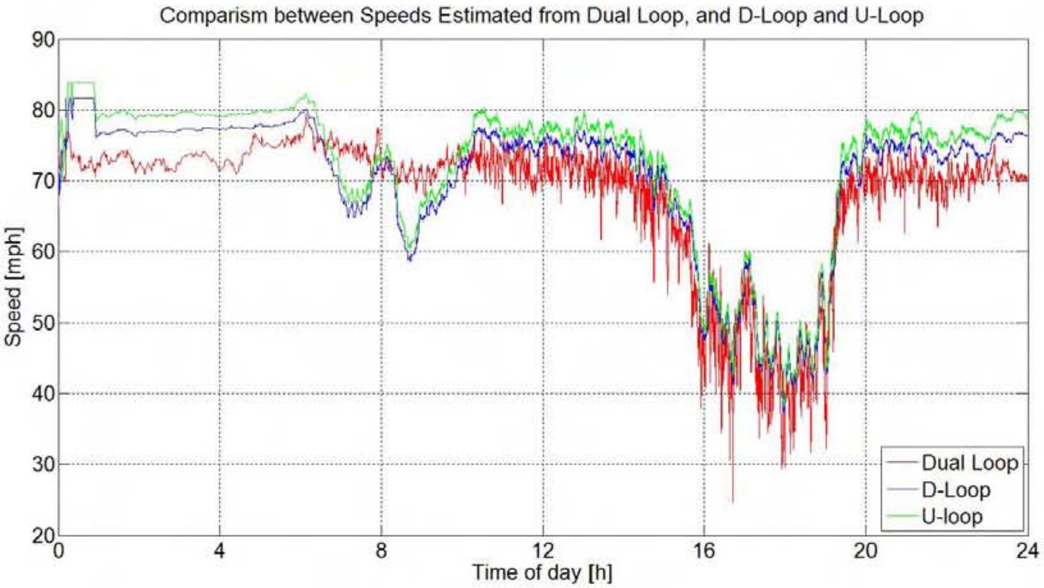


Figure B. 26 BHL EB, Station 2, Lane 1, Speed estimation comparison

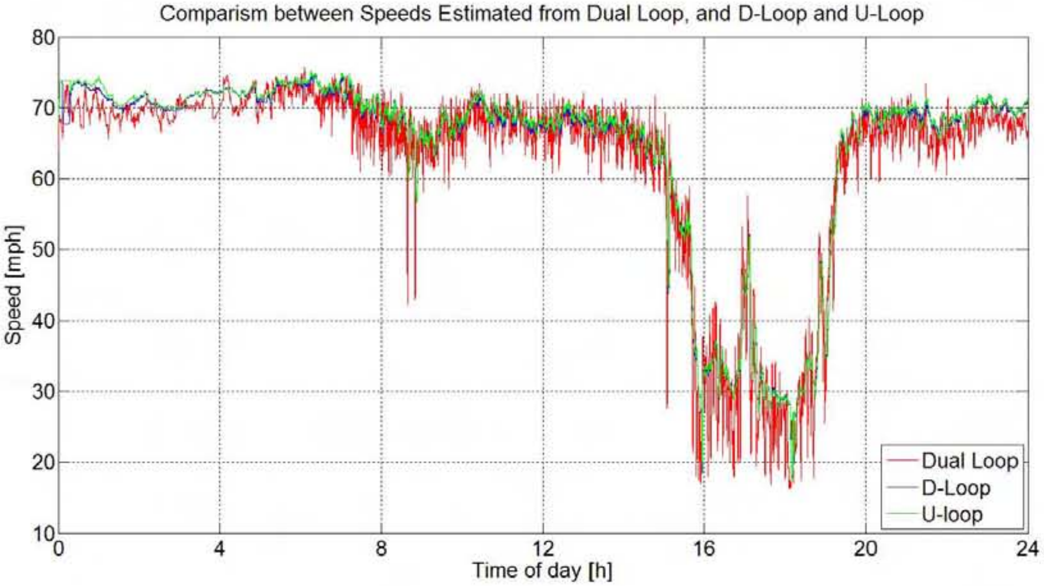


Figure B. 27 BHL EB, Station 2, Lane 2, Speed estimation comparison

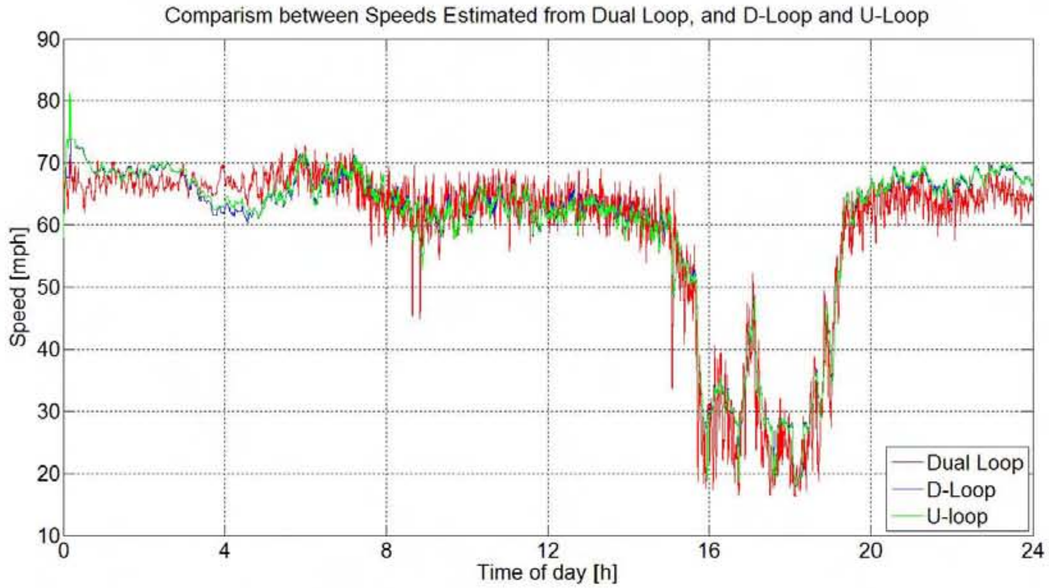


Figure B. 28 BHL EB, Station 2, Lane 3, Speed estimation comparison

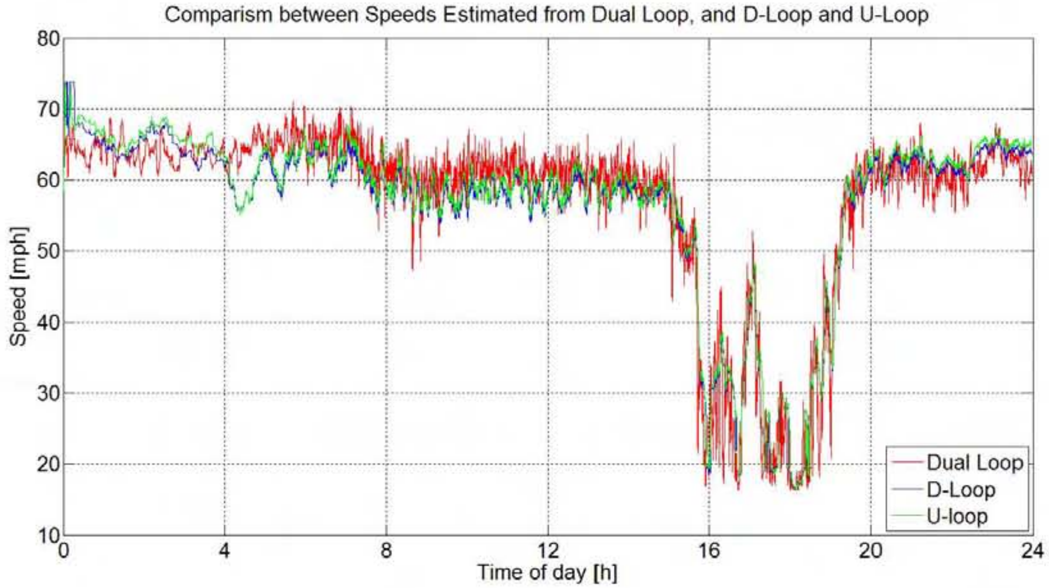


Figure B. 29 BHL EB, Station 2, Lane 4, Speed estimation comparison

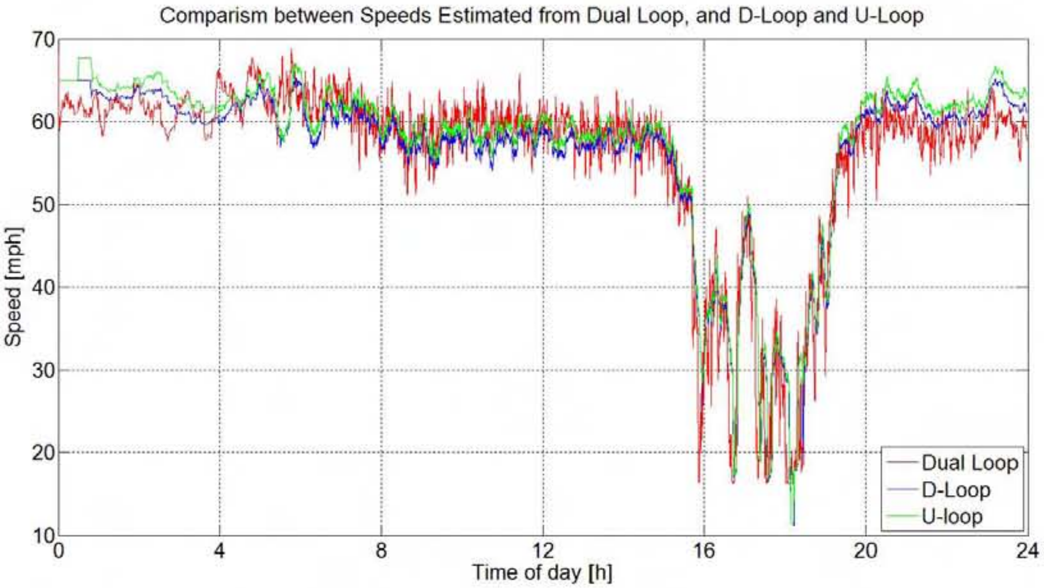


Figure B. 30 BHL EB, Station 2, Lane 5, Speed estimation comparison

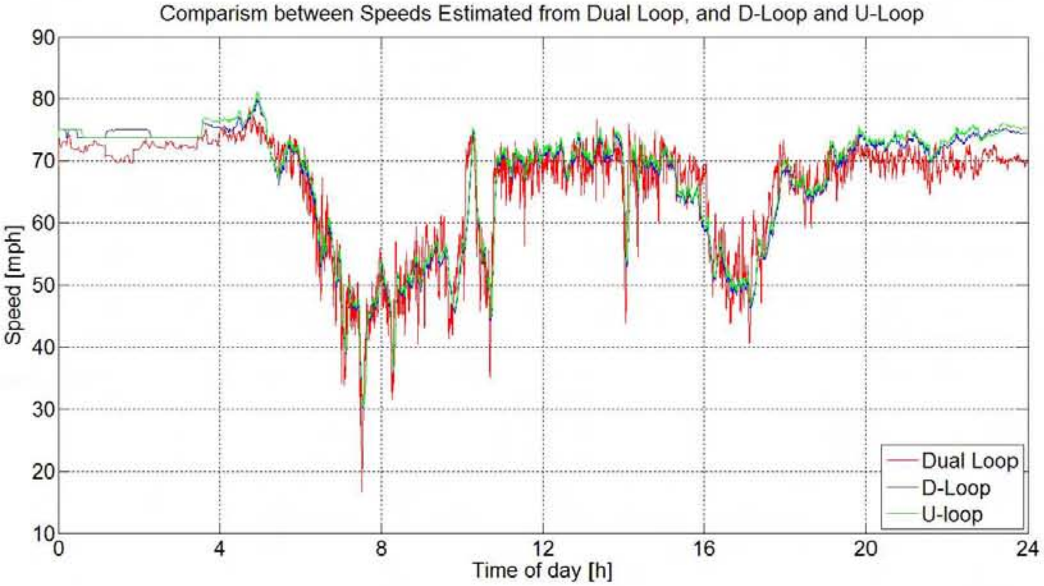


Figure B. 31 BHL WB, Station 2, Lane 1, Speed estimation comparison

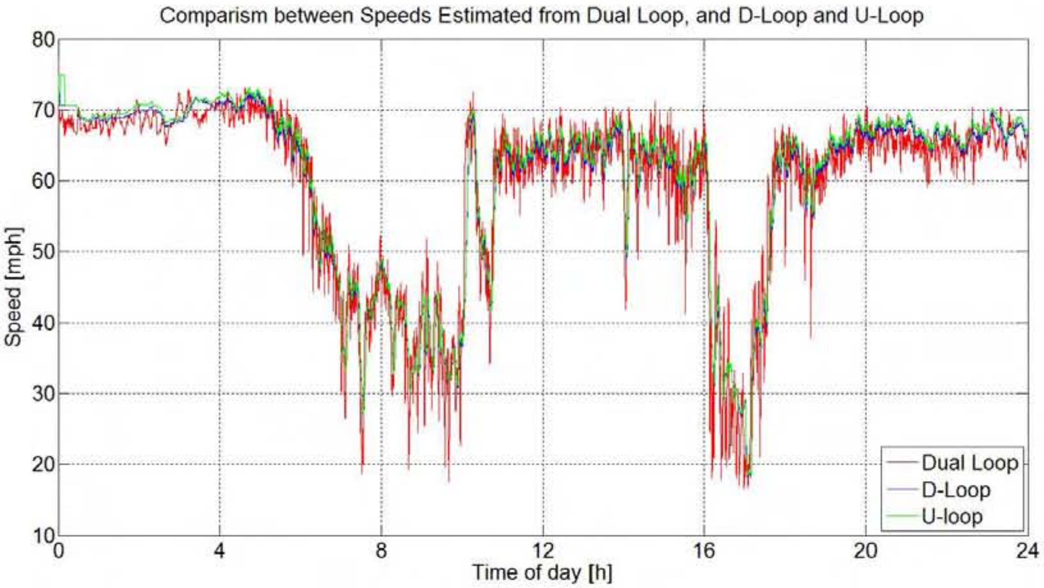


Figure B. 32 BHL WB, Station 2, Lane 2, Speed estimation comparison

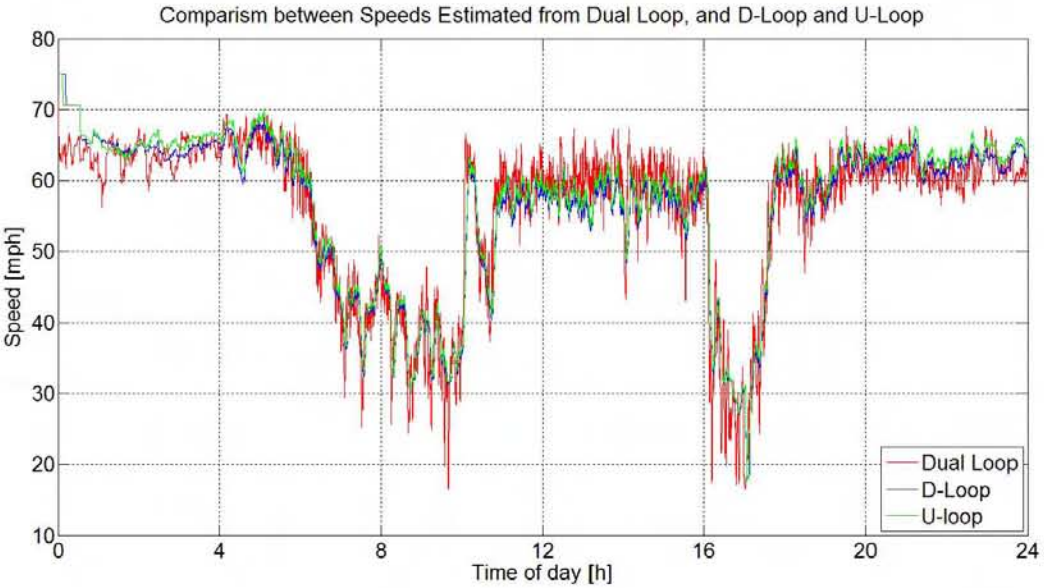


Figure B. 33 BHL WB, Station 2, Lane 3, Speed estimation comparison

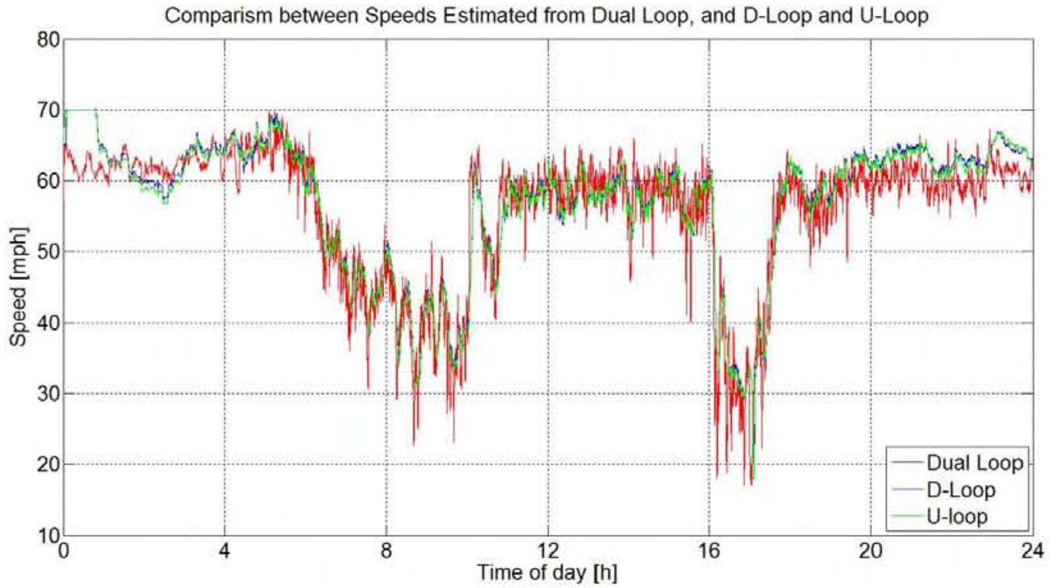


Figure B. 34 BHL WB, Station 2, Lane 4, Speed estimation comparison

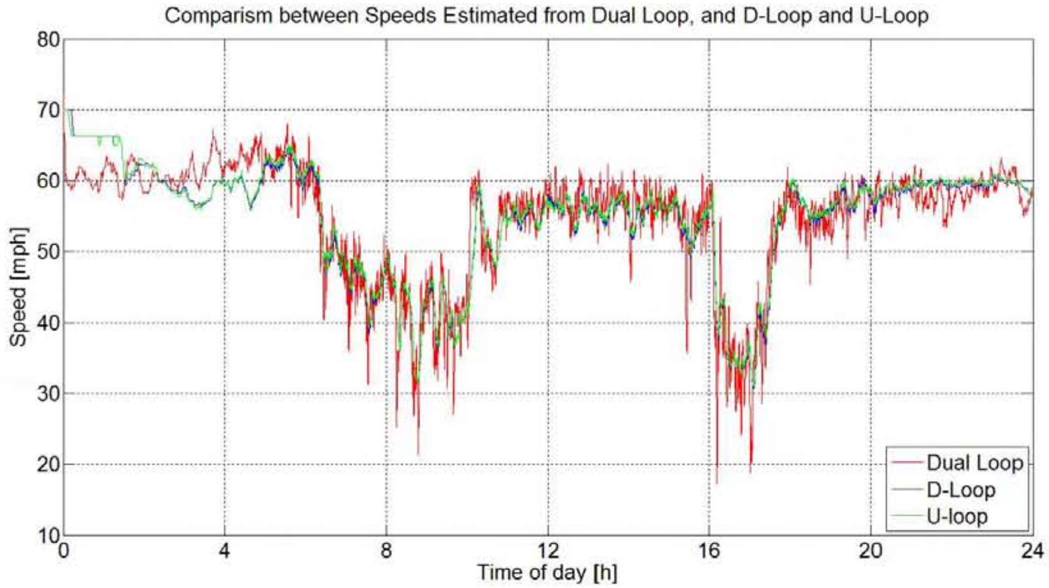


Figure B. 35 BHL WB, Station 2, Lane 5, Speed estimation comparison

****

****

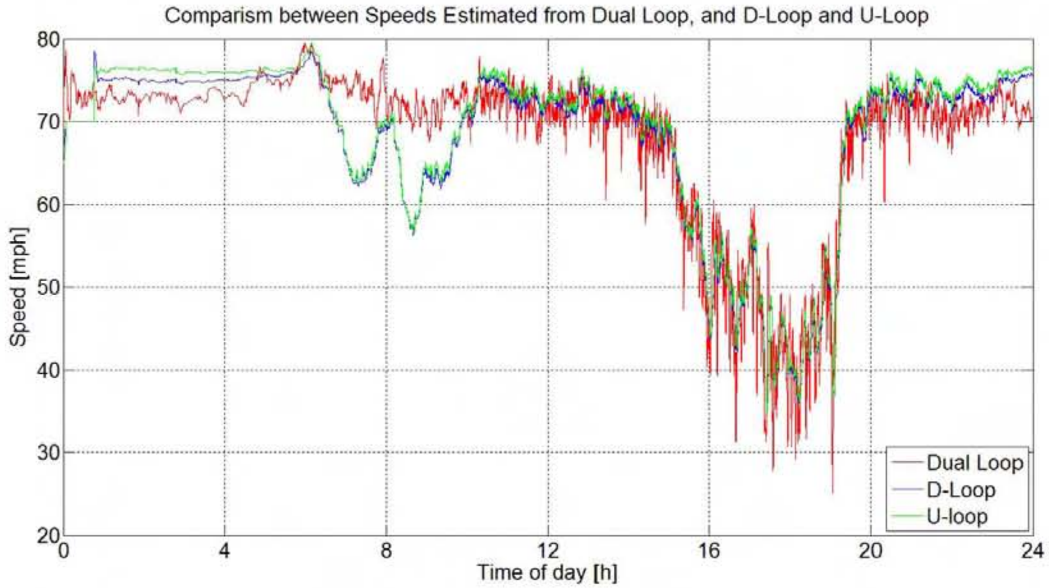


Figure B. 36 BHL EB, Station 3, Lane 1, Speed estimation comparison

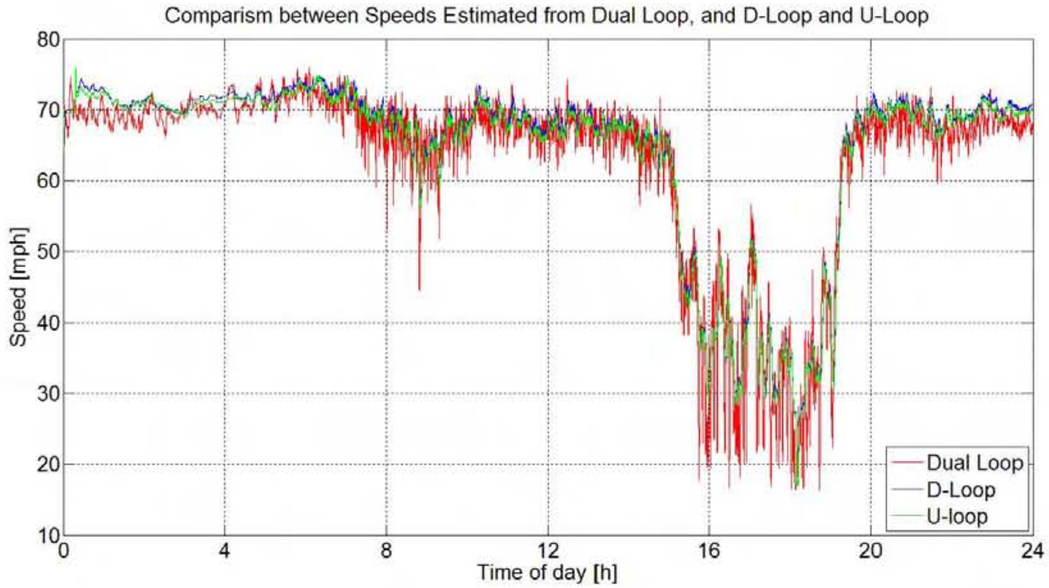


Figure B. 37 BHL EB, Station 3, Lane 2, Speed estimation comparison

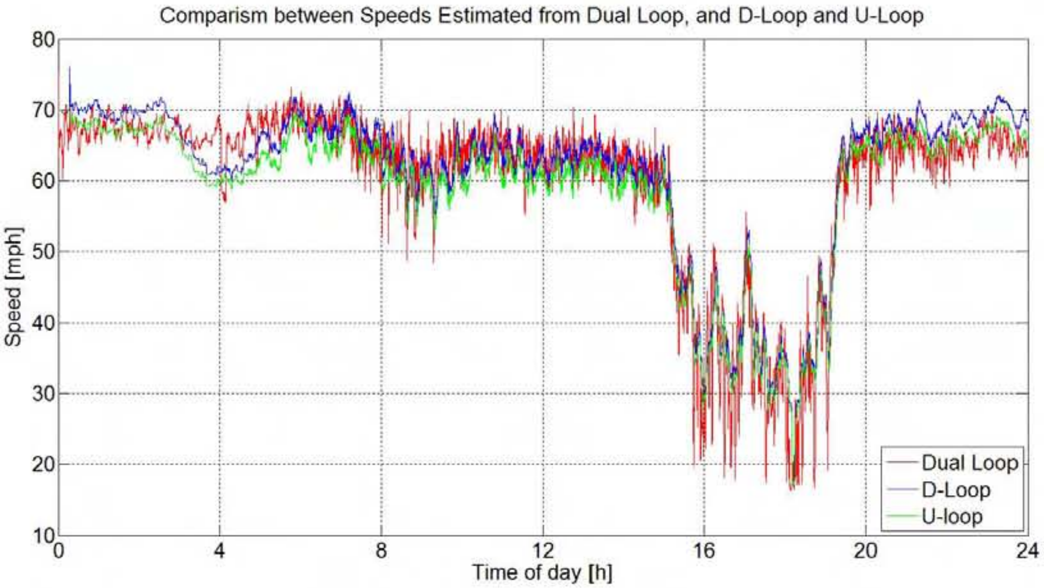


Figure B. 38 BHL EB, Station 3, Lane 3, Speed estimation comparison

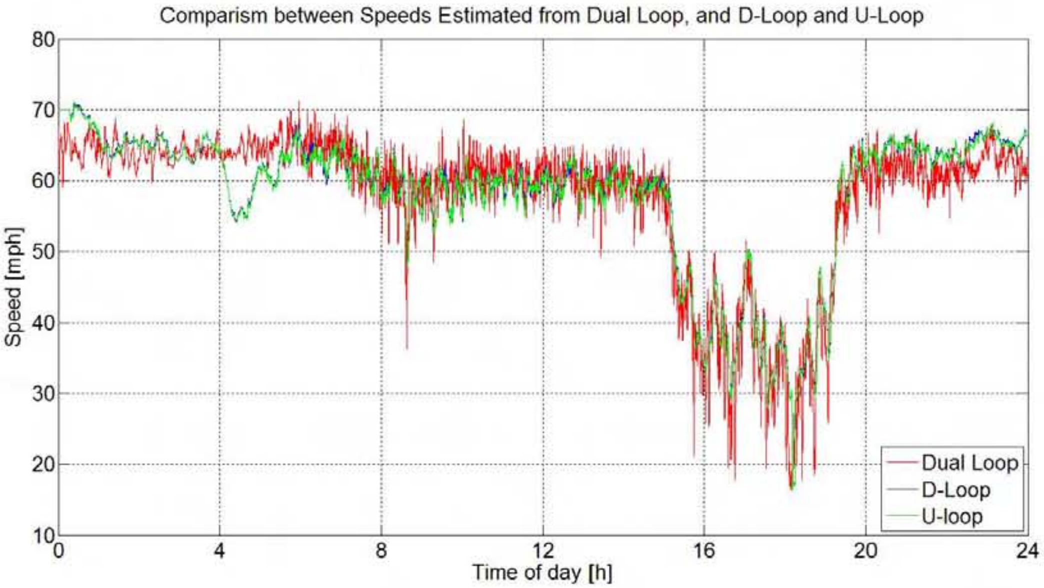


Figure B. 39 BHL EB, Station 3, Lane 4, Speed estimation comparison

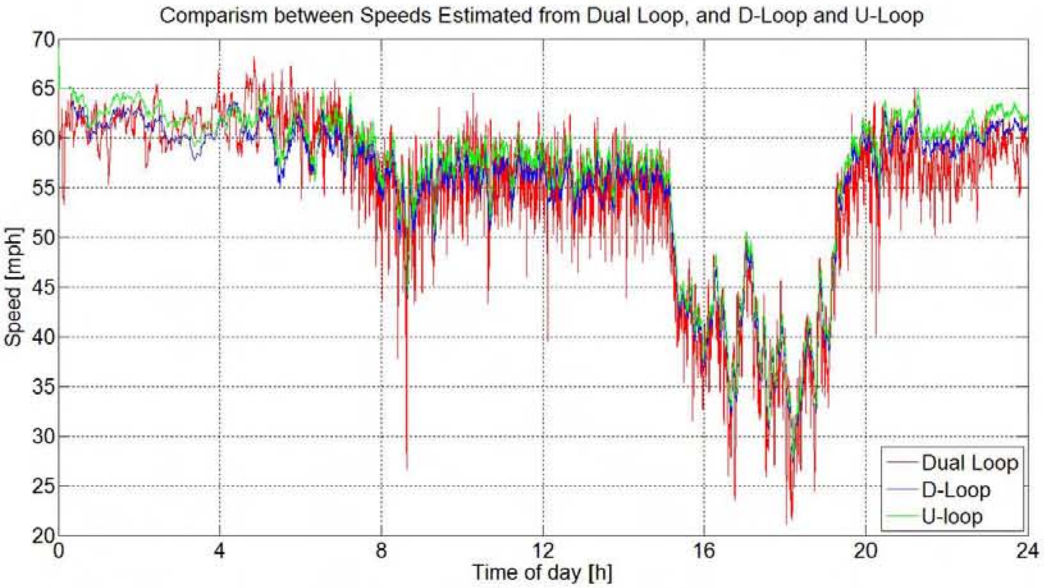


Figure B. 40 BHL EB, Station 3, Lane 5, Speed estimation comparison

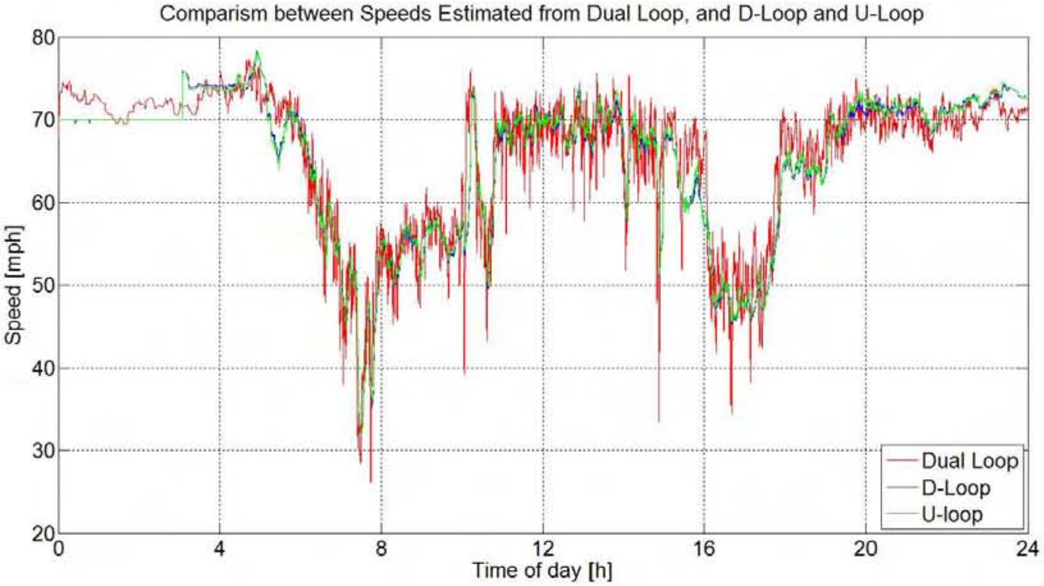


Figure B. 41 BHL WB, Station 3, Lane 1, Speed estimation comparison

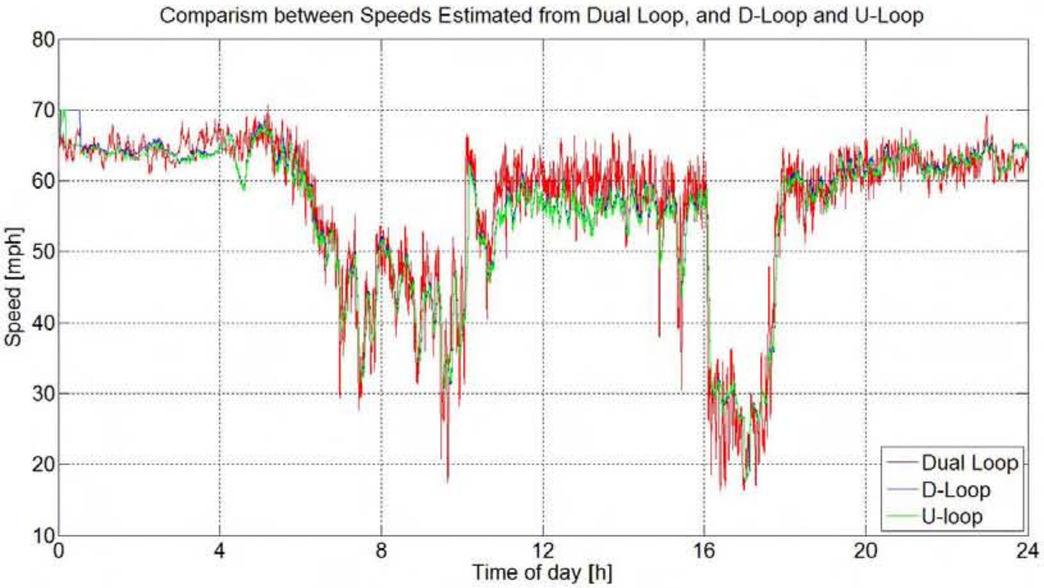


Figure B. 43 BHL WB, Station 3, Lane 3, Speed estimation comparison

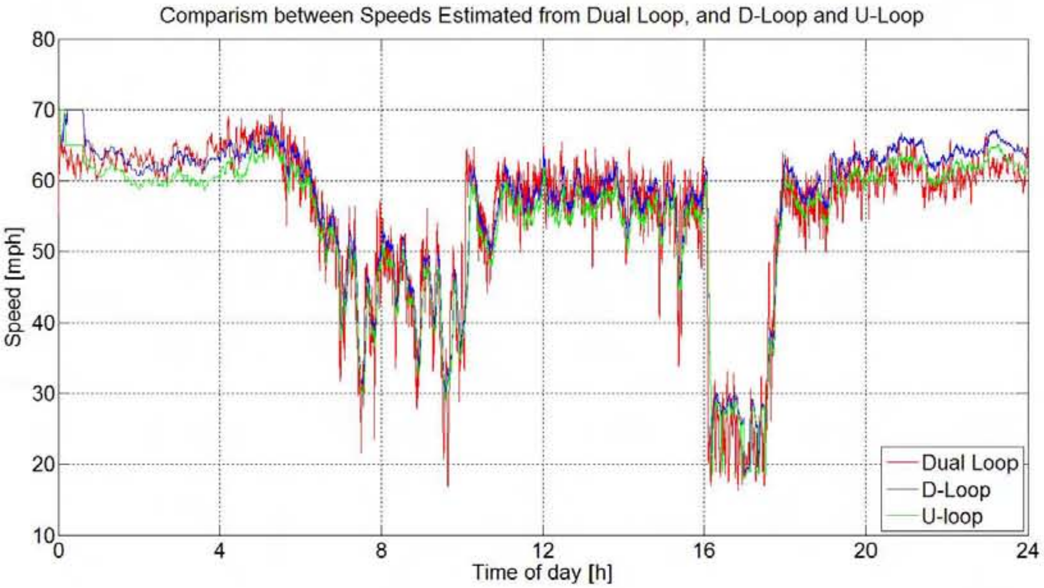


Figure B. 44 BHL WB, Station 3, Lane 4, Speed estimation comparison

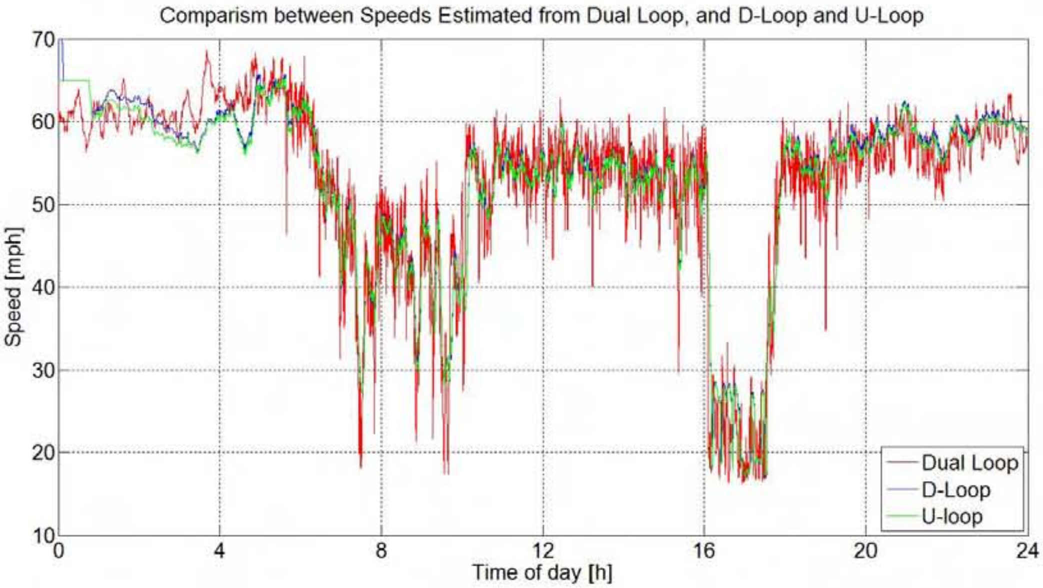


Figure B. 45 BHL WB, Station 3, Lane 5, Speed estimation comparison

****

****

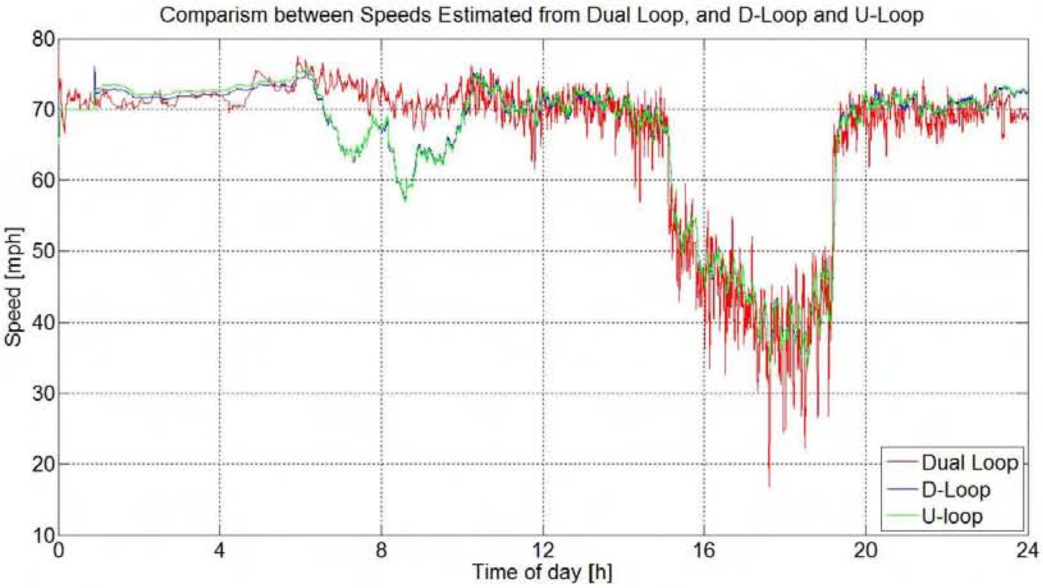


Figure B. 46 BHL EB, Station 6, Lane 1, Speed estimation comparison

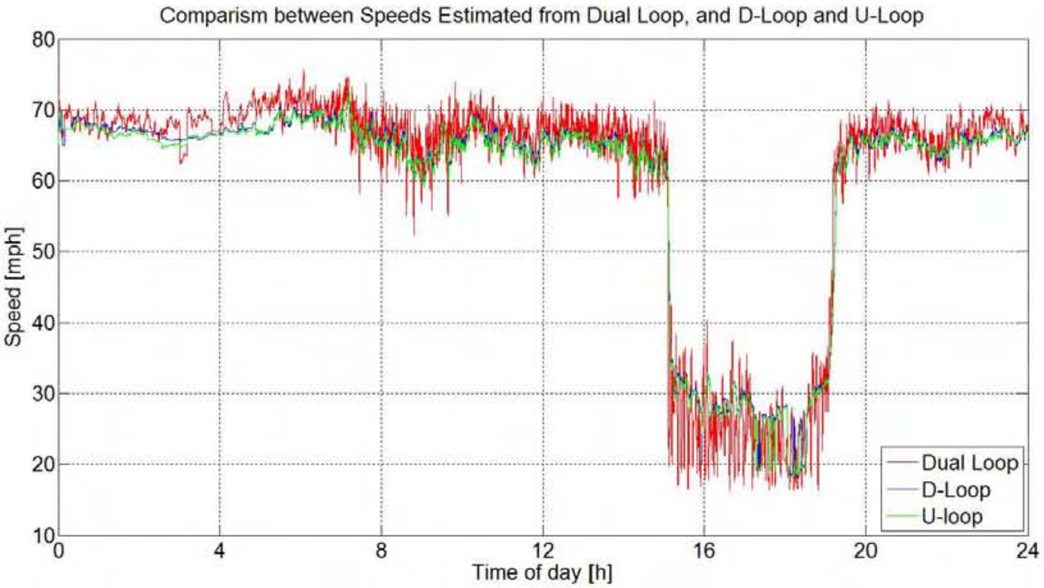


Figure B. 47 BHL EB, Station 6, Lane 2, Speed estimation comparison

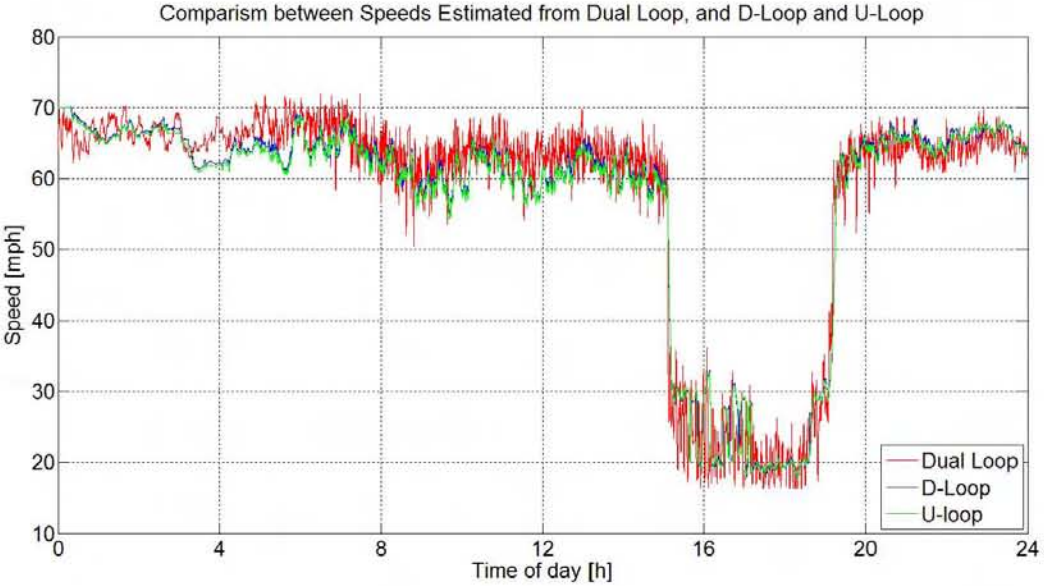


Figure B. 48 BHL EB, Station 6, Lane 3, Speed estimation comparison

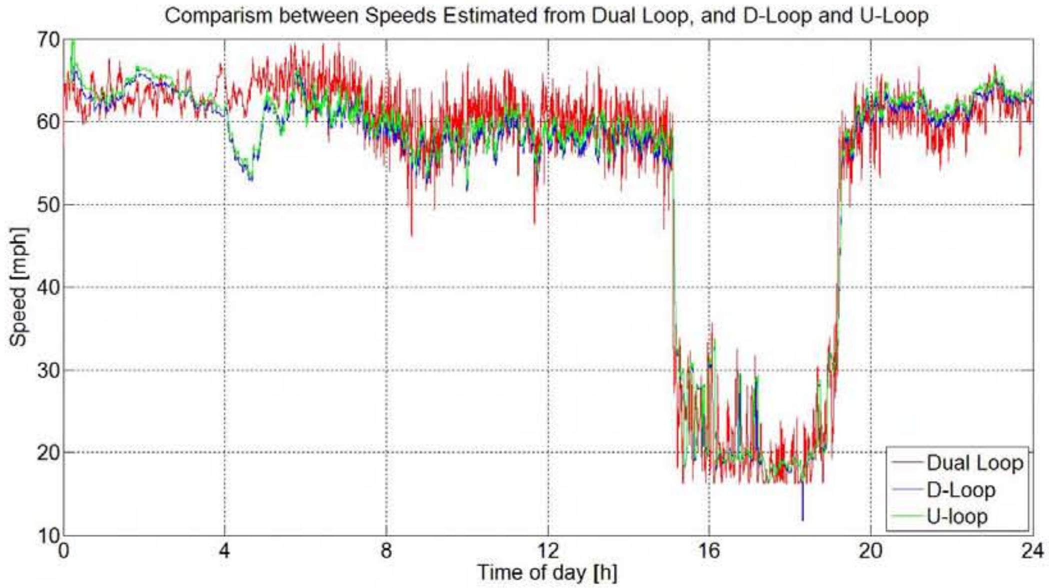


Figure B. 49 BHL EB, Station 6, Lane 4, Speed estimation comparison

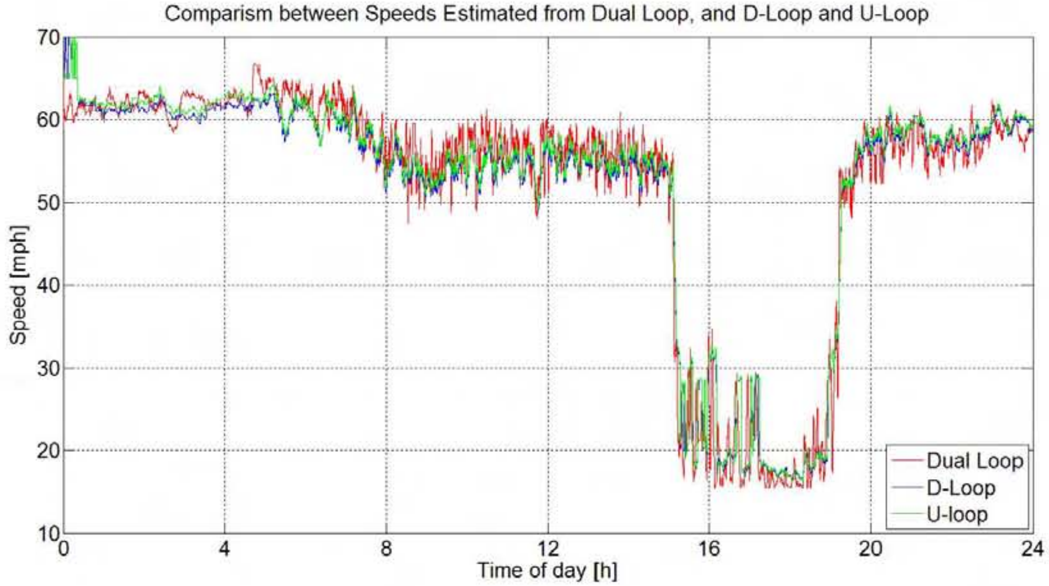


Figure B. 50 BHL EB, Station 6, Lane 5, Speed estimation comparison

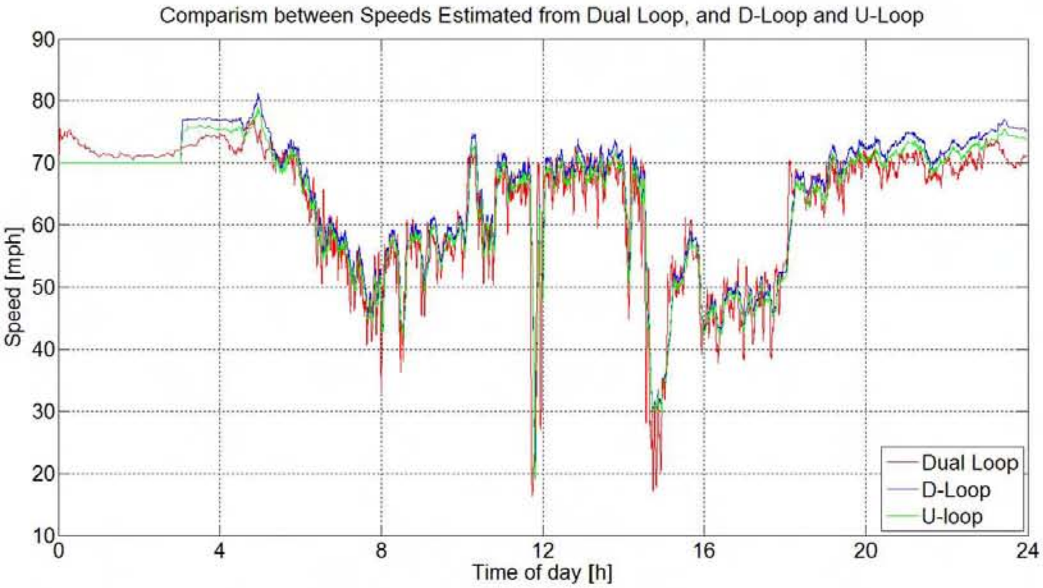


Figure B. 51 BHL WB, Station 6, Lane 1, Speed estimation comparison

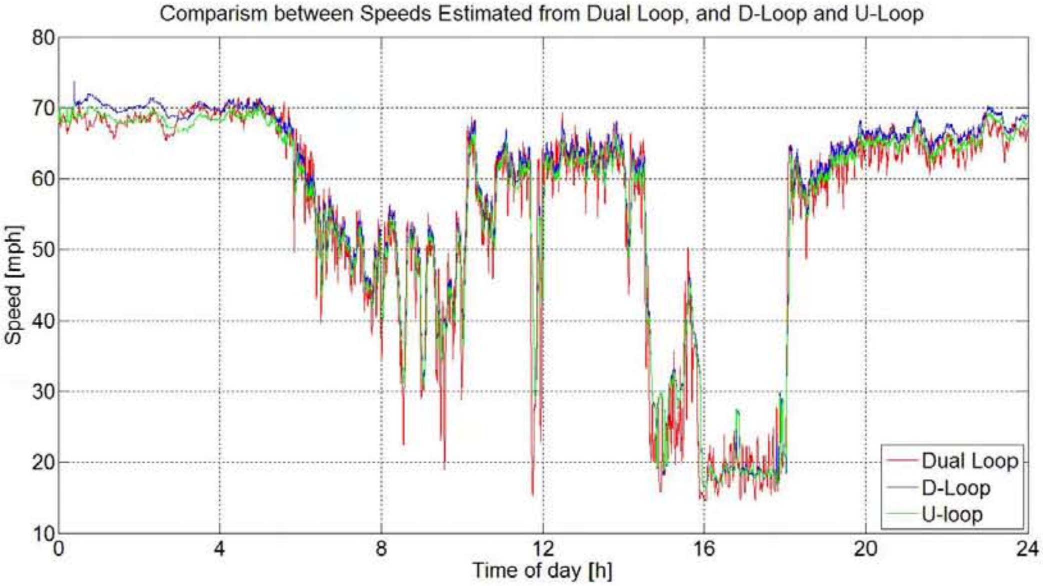


Figure B. 52 BHL WB, Station 6, Lane 2, Speed estimation comparison

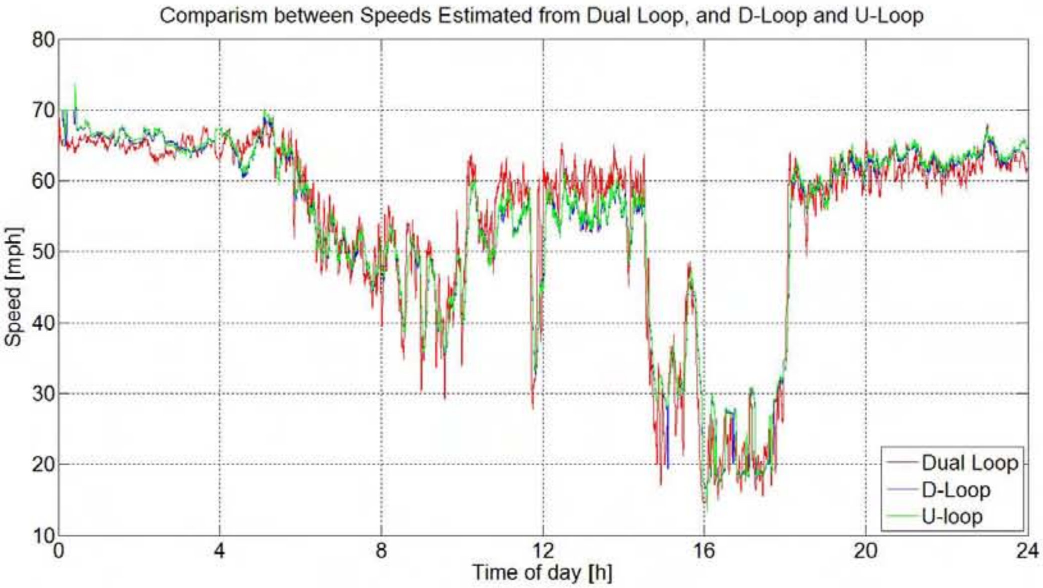


Figure B. 53 BHL WB, Station 6, Lane 3, Speed estimation comparison

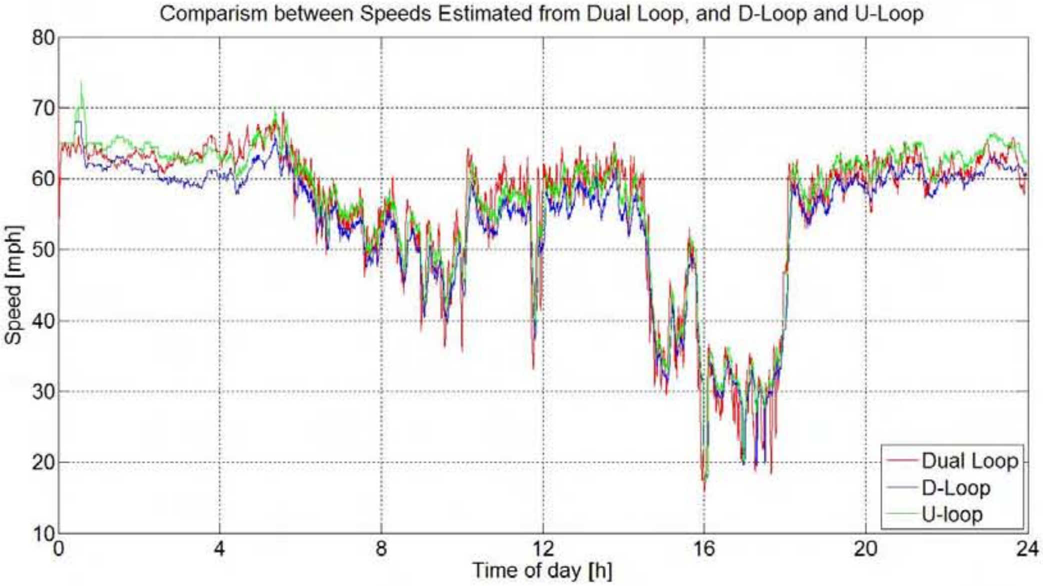


Figure B. 54 BHL WB, Station 6, Lane 4, Speed estimation comparison

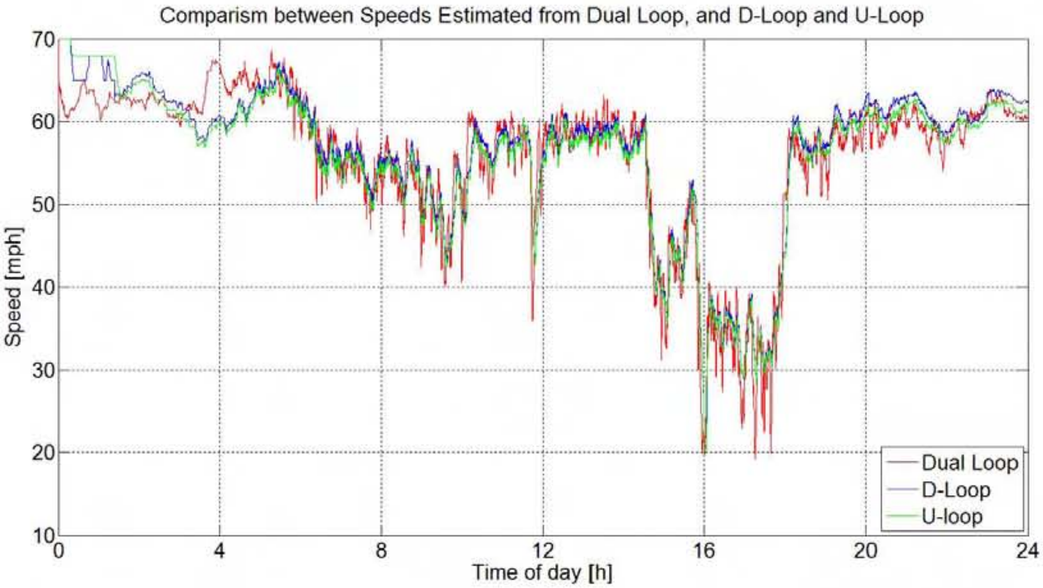


Figure B. 55 BHL WB, Station 6, Lane 5, Speed estimation comparison

****

****

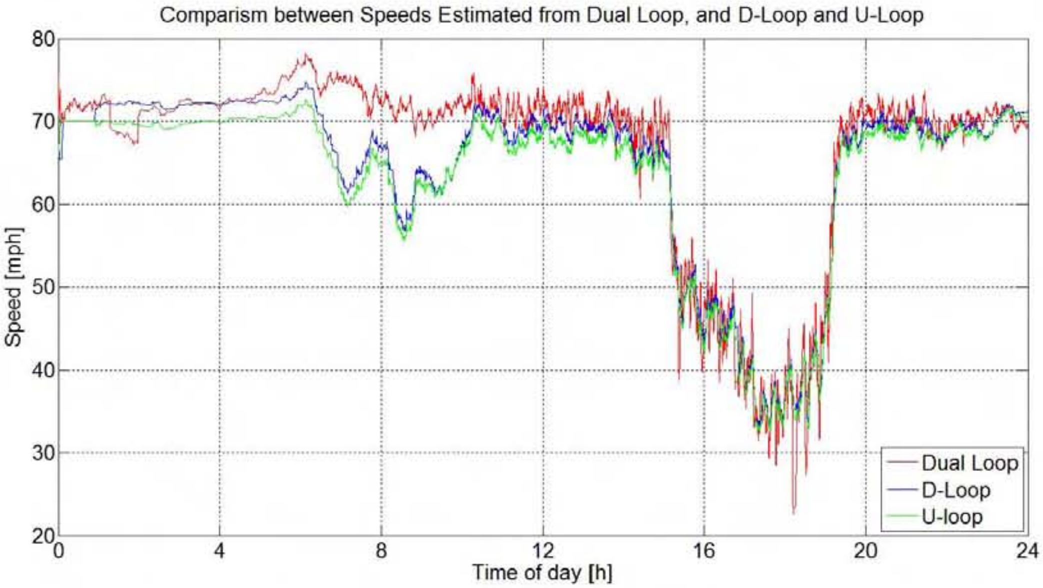


Figure B. 56 BHL EB, Station 7, Lane 1, Speed estimation comparison

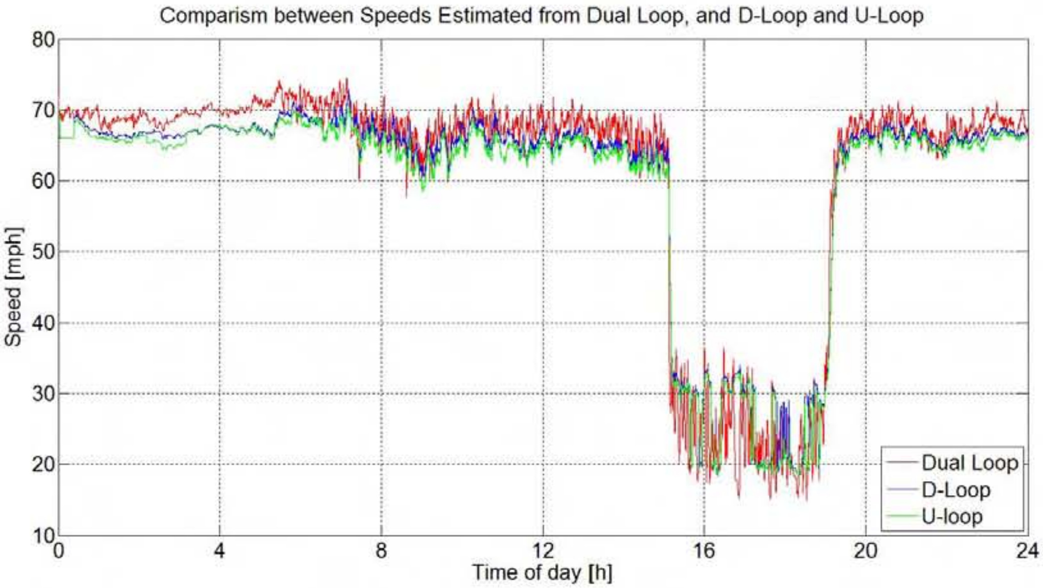


Figure B. 57 BHL EB, Station 7, Lane 2, Speed estimation comparison

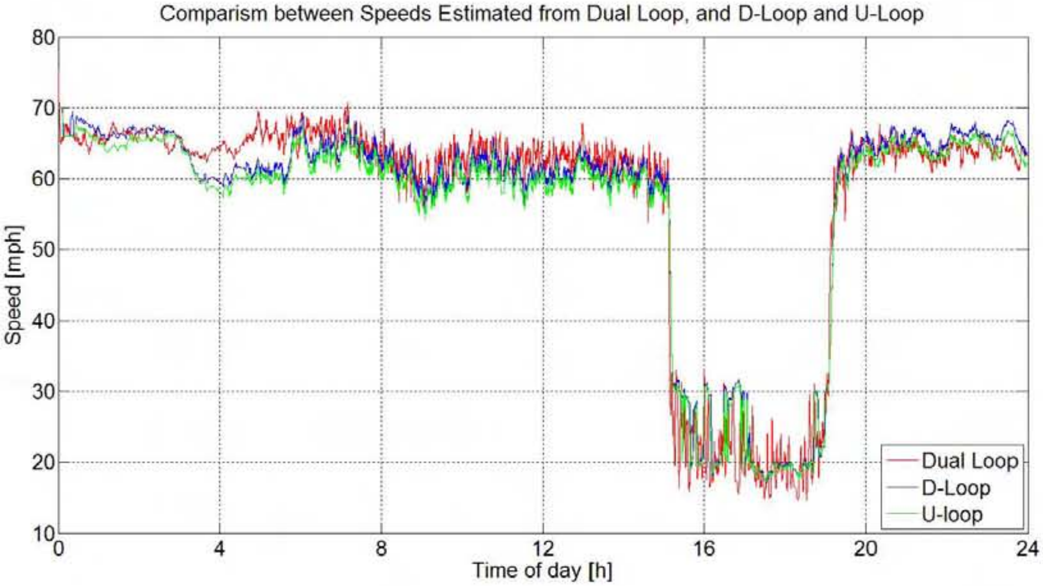


Figure B. 58 BHL EB, Station 7, Lane 3, Speed estimation comparison

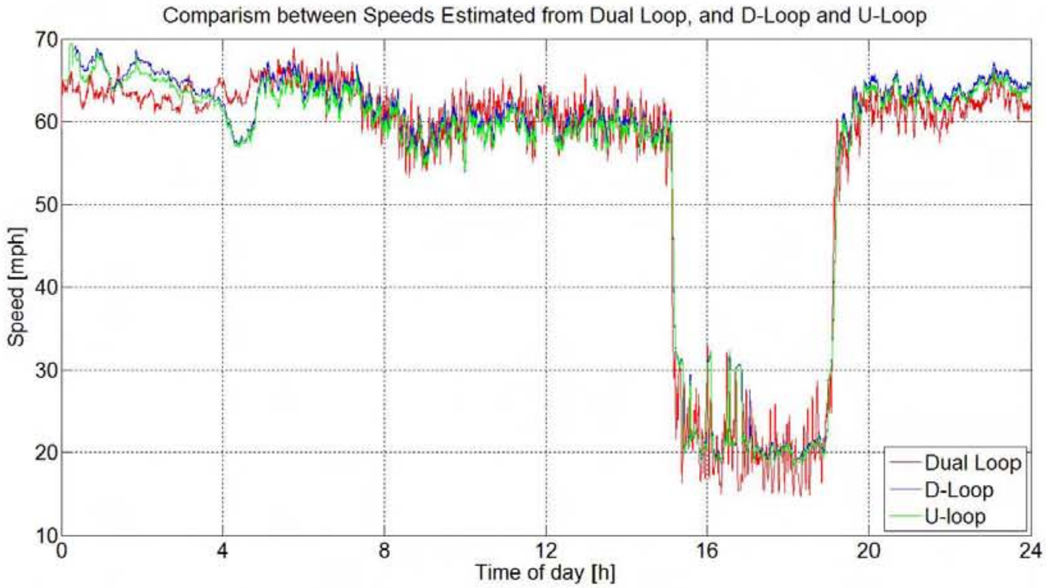


Figure B. 59 BHL EB, Station 7, Lane 4, Speed estimation comparison

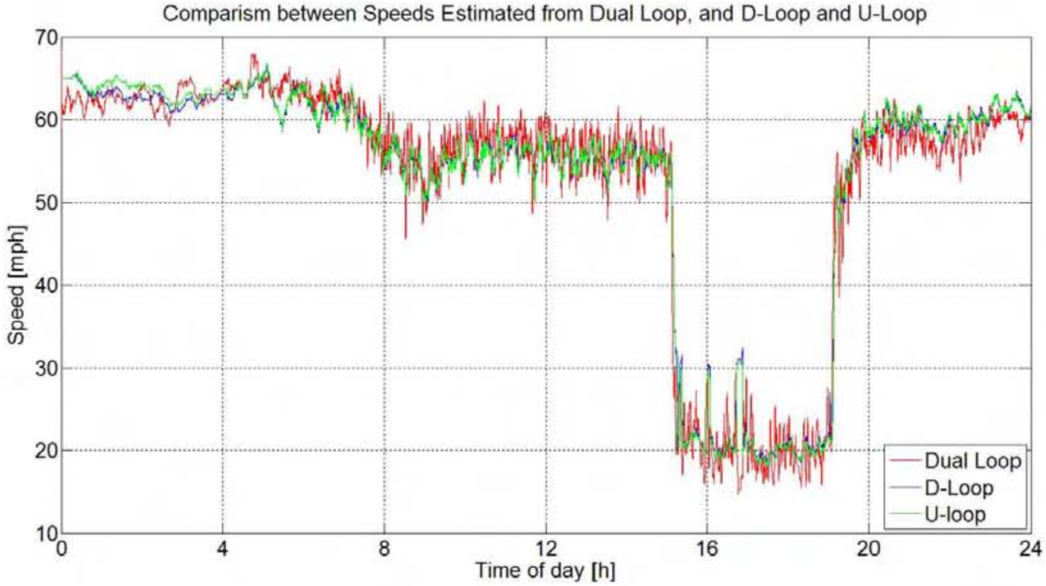


Figure B. 60 BHL EB, Station 7, Lane 5, Speed estimation comparison

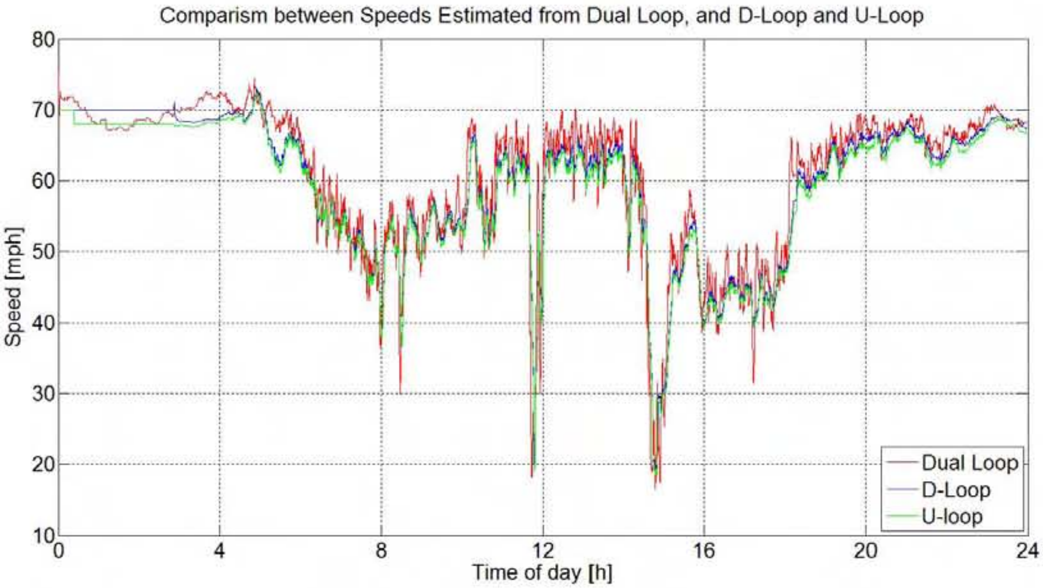


Figure B. 61 BHL WB, Station 7, Lane 1, Speed estimation comparison

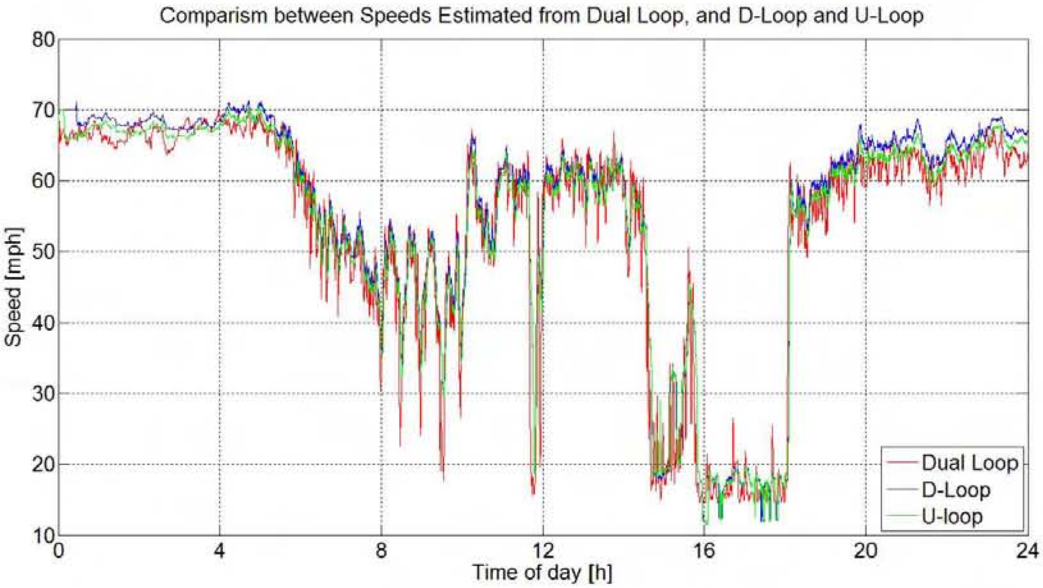


Figure B. 62 BHL WB, Station 7, Lane 2, Speed estimation comparison

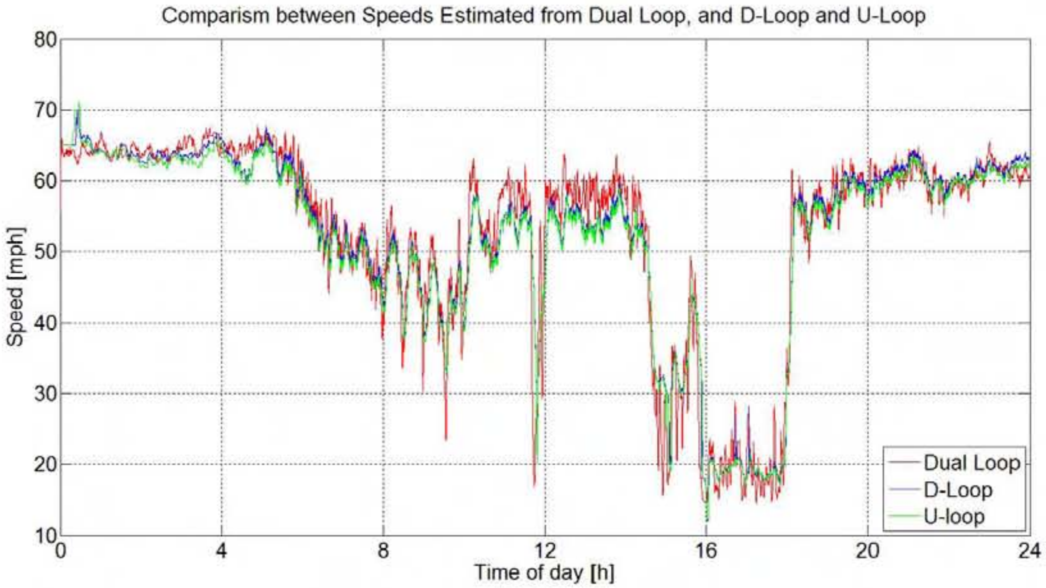


Figure B. 63 BHL WB,   Station 7, Lane 3, Speed estimation comparison

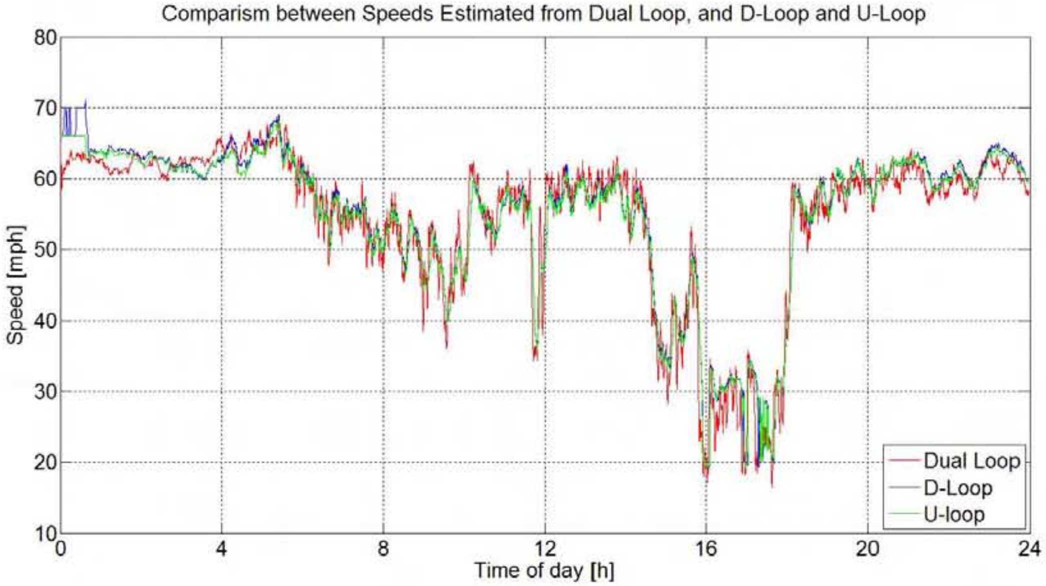


Figure B. 64 BHL WB,   Station 7, Lane 4, Speed estimation comparison

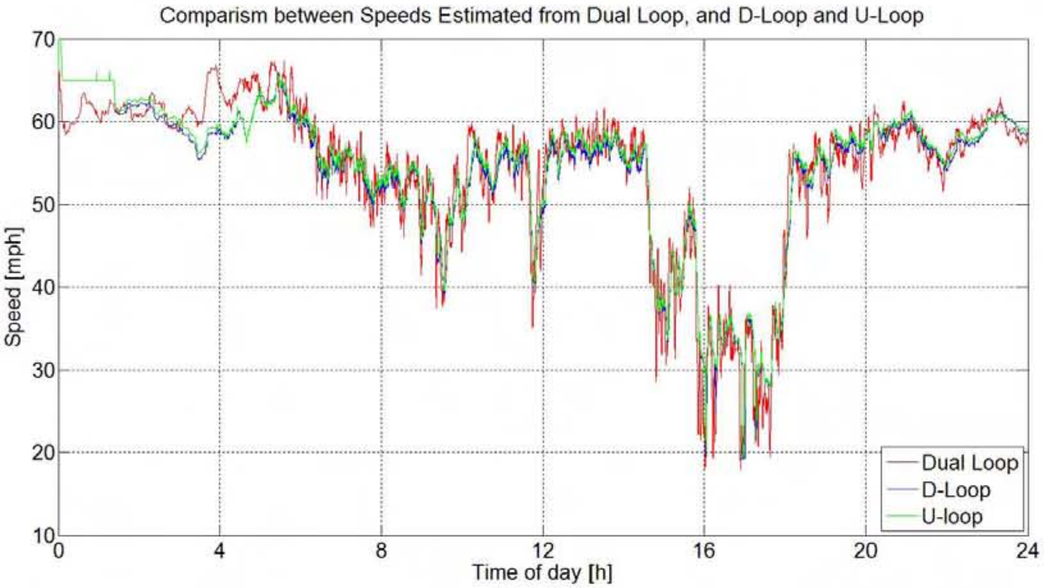


Figure B. 65 BHL WB, Station 7, Lane 5, Speed estimation comparison

Appendix: C

Vehicle Length-Based Classification from Single Loop Event Data

The following are more plots for vehicle length estimation and classification using single loop data compared with those from dual loop speed as described in Chapter 3. The purpose is to show the effectiveness and reliability of the algorithm. Data Source: BHL (Figure 2.2) on 03/01/2011, for Station 7 in both East Bound and West Bound of all 5 lanes. The plots include the number of vehicles in 6 classes every 30min.

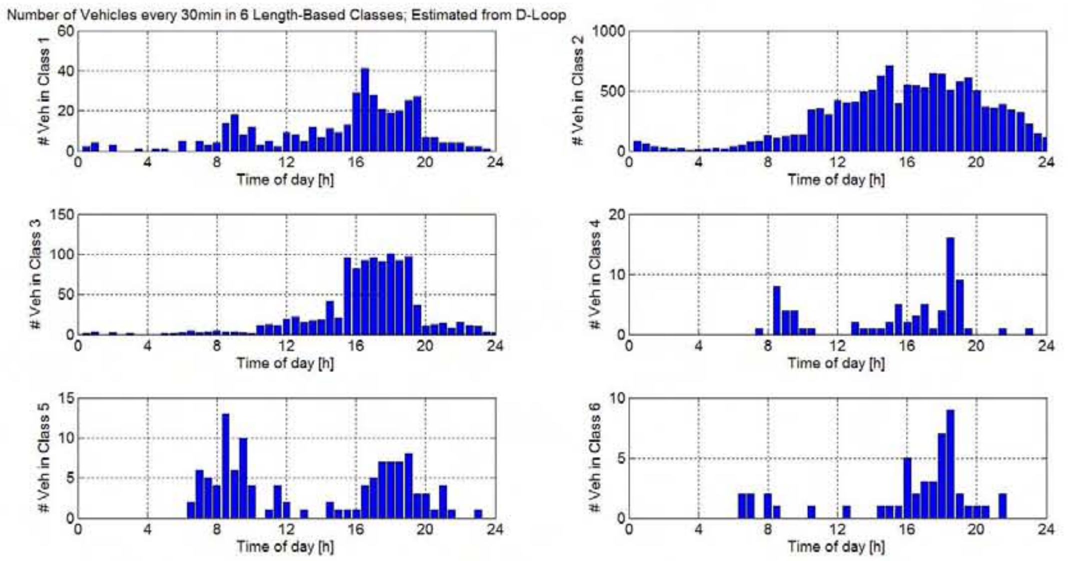


Figure C. 1 BHL EB, Station 7, Lane 1, Vehicle Classification Every 30min; D-loop

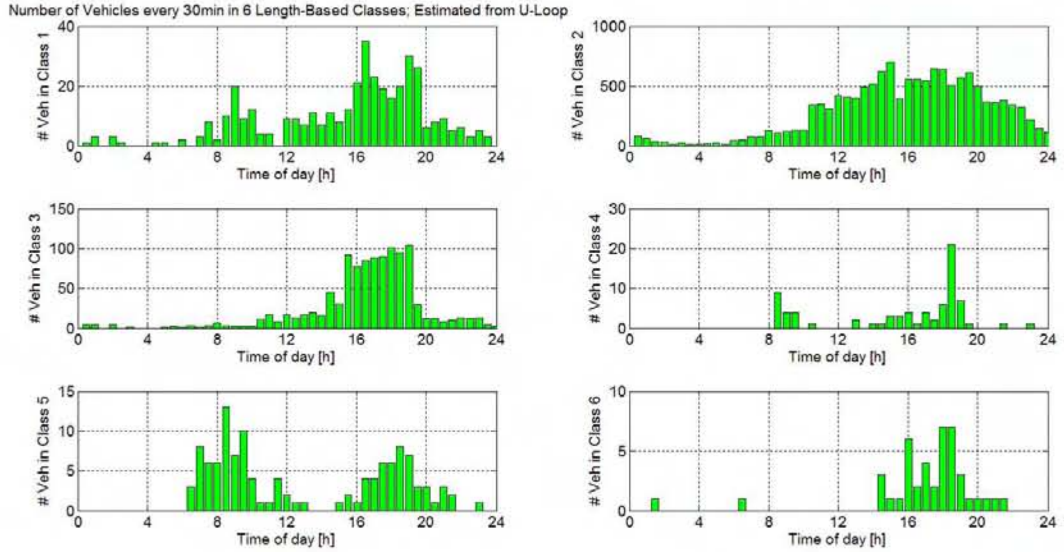


Figure C. 2 BHL EB, Station 7, Lane 1, Vehicle Classification Every 30min; U-loop

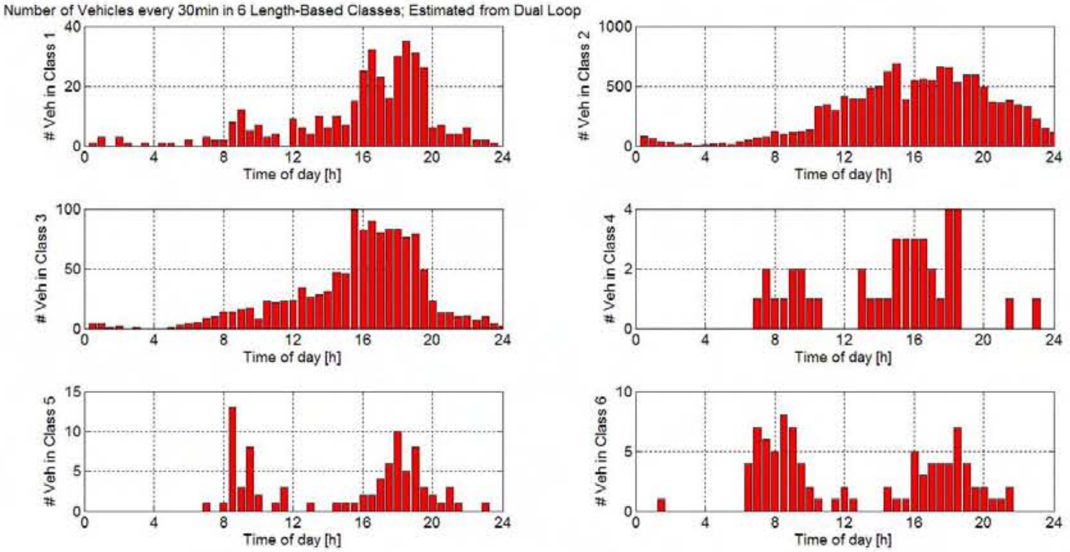


Figure C. 3 BHL EB, Station 7, Lane 1, Vehicle Classification Every 30min; Dual-loop

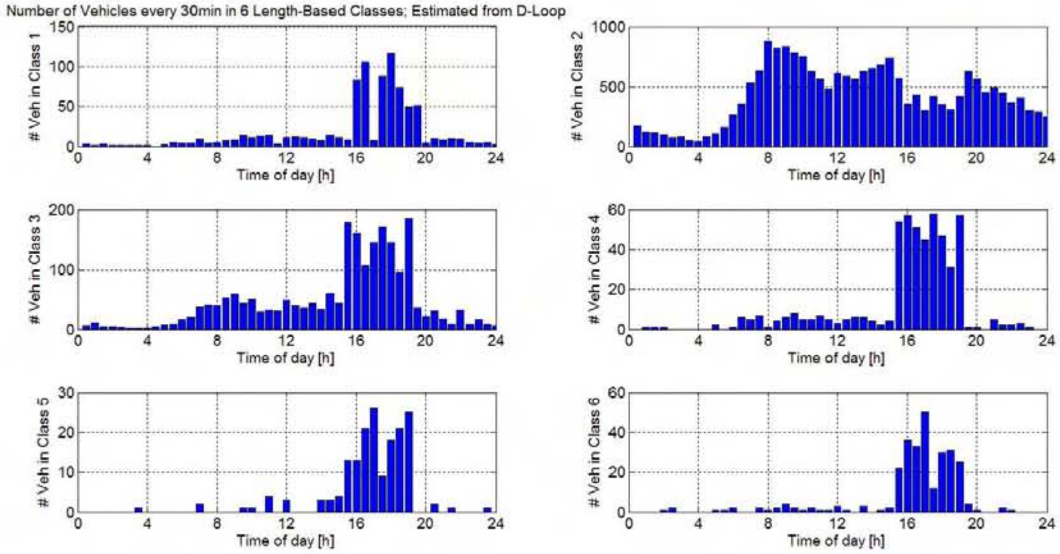


Figure C. 4 BHL EB, Station 7, Lane 2, Vehicle Classification Every 30min; D-loop

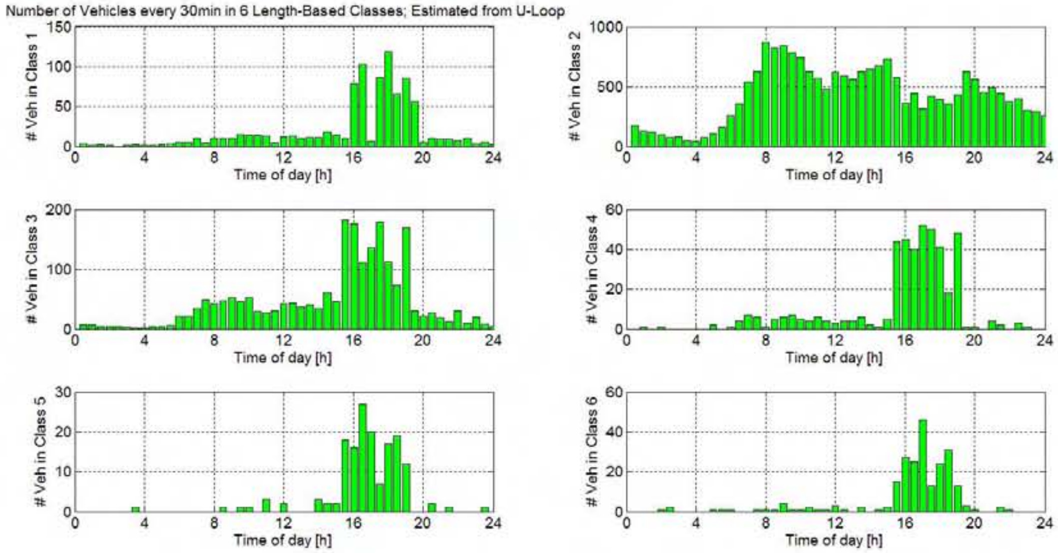


Figure C. 5 BHL EB, Station 7, Lane 2, Vehicle Classification Every 30min; U-loop

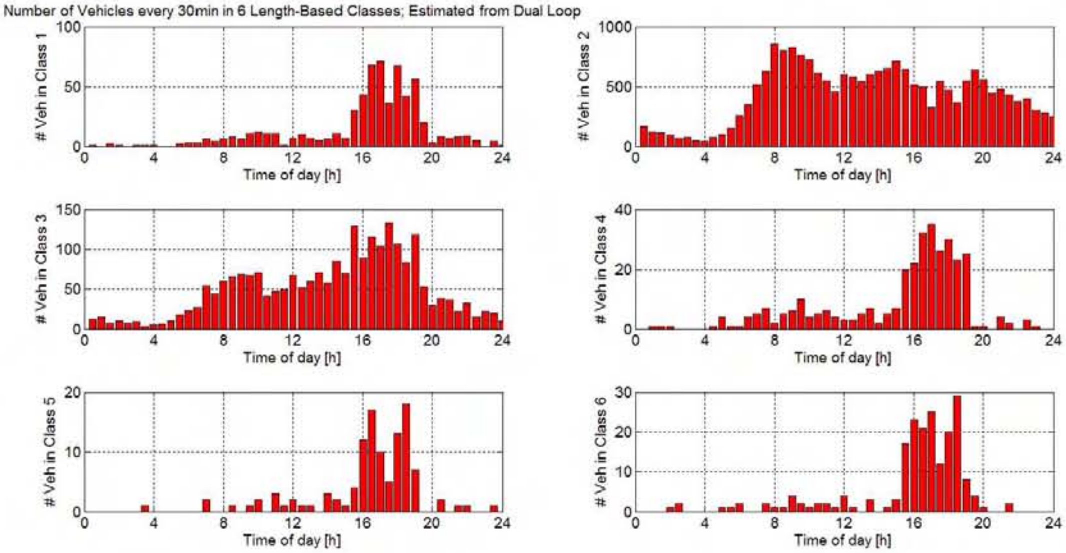


Figure C. 6 BHL EB, Station 7, Lane 2, Vehicle Classification Every 30min; Dual-loop

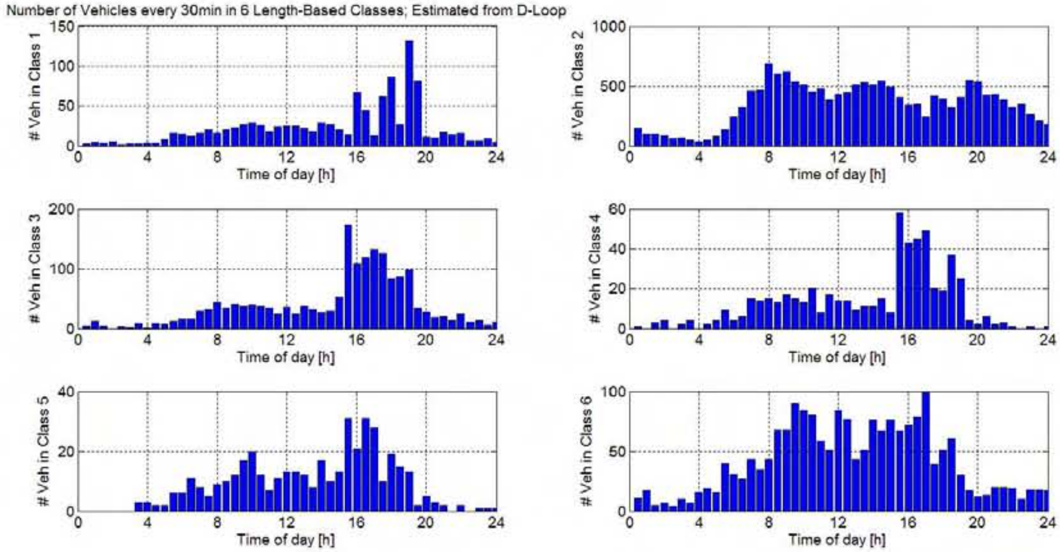


Figure C. 7 BHL EB, Station 7, Lane 3, Vehicle Classification Every 30min; D-loop

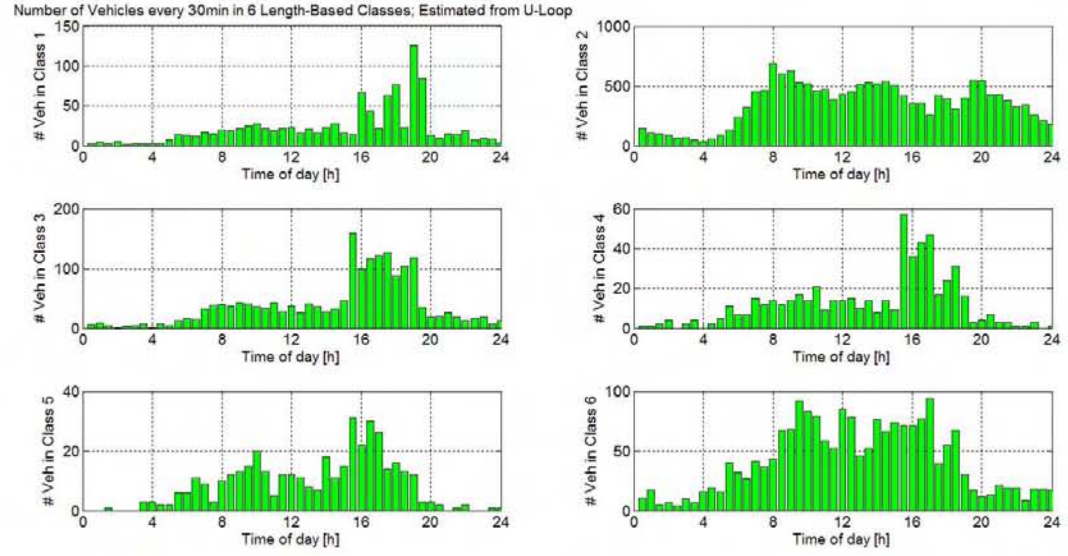


Figure C. 8 BHL EB, Station 7, Lane 3, Vehicle Classification Every 30min; U-loop

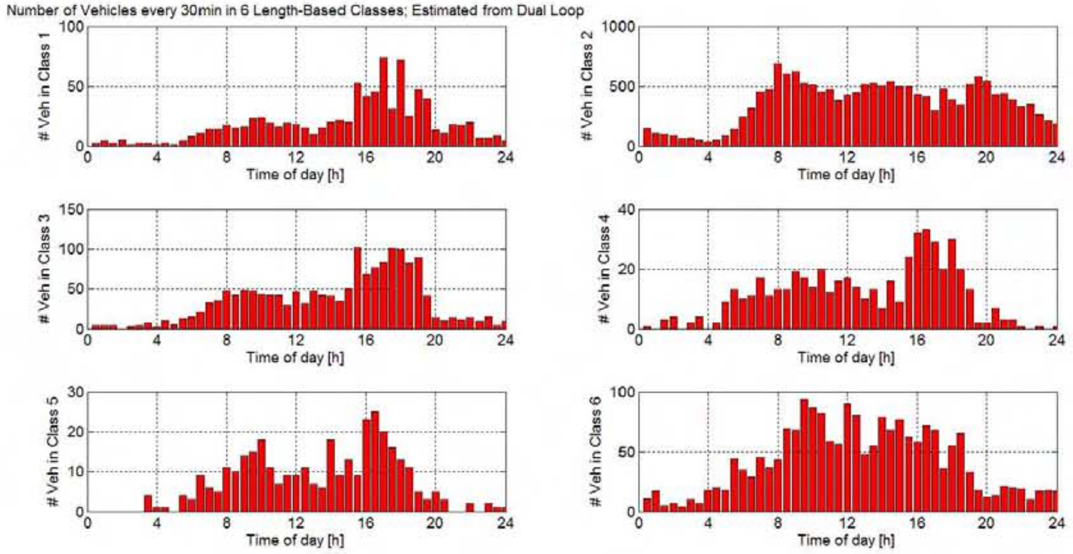


Figure C. 9 BHL EB, Station 7, Lane 3, Vehicle Classification Every 30min; Dual-loop

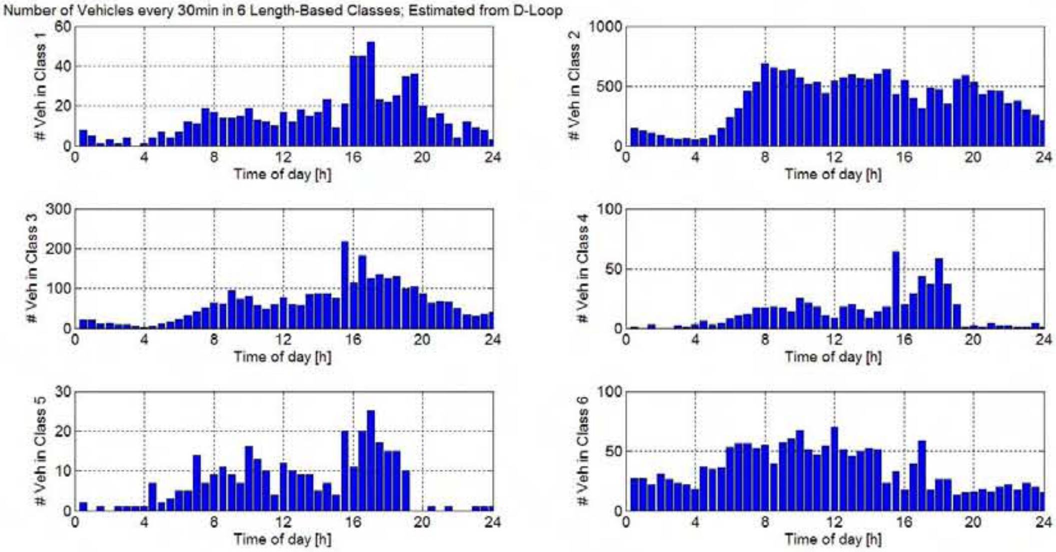


Figure C. 10 BHL EB, Station 7, Lane 4, Vehicle Classification Every 30min; D-loop

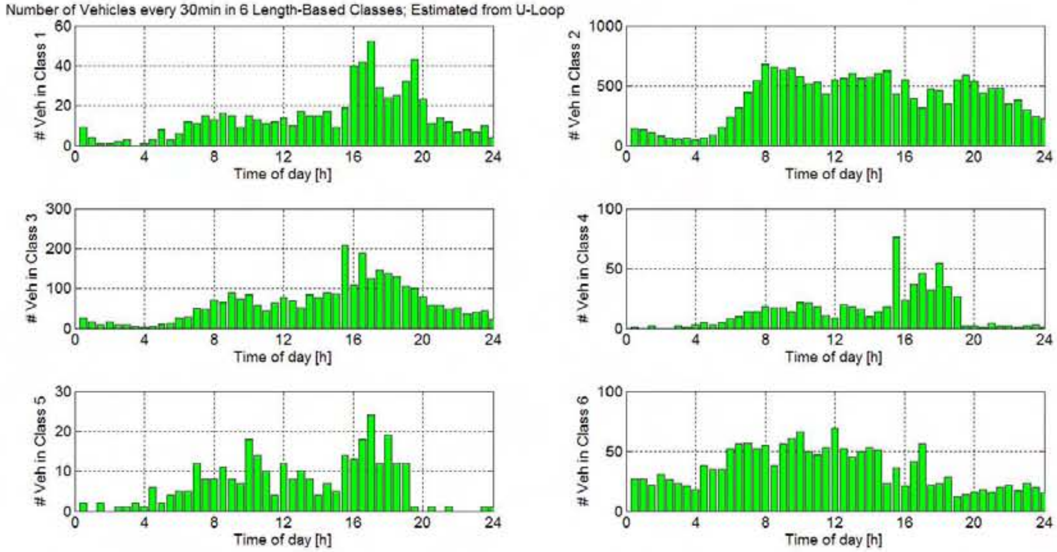


Figure C. 11 BHL EB, Station 7, Lane 4, Vehicle Classification Every 30min; U-loop

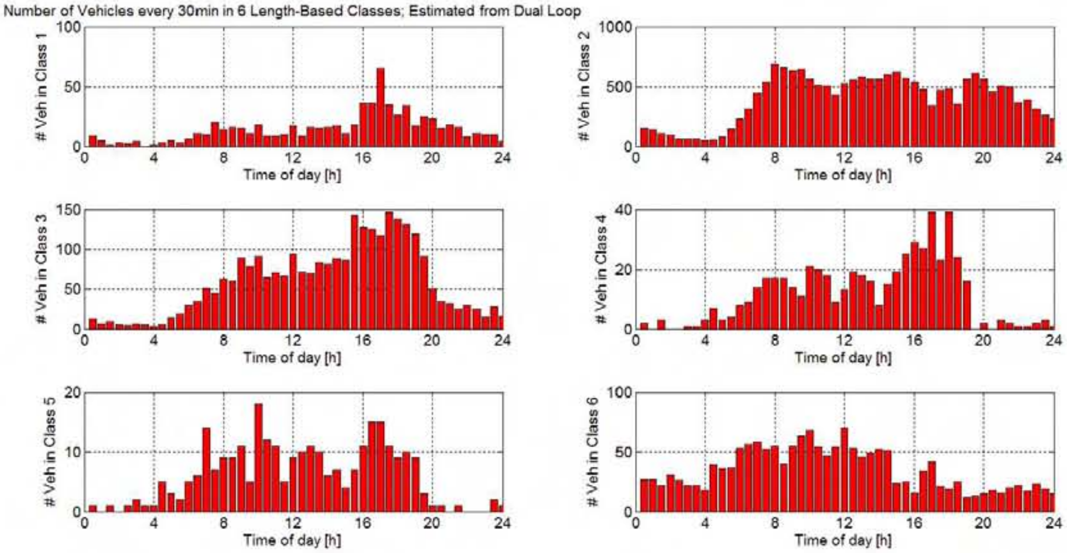


Figure C. 12 BHL EB, Station 7, Lane 4, Vehicle Classification Every 30min; Dual-loop

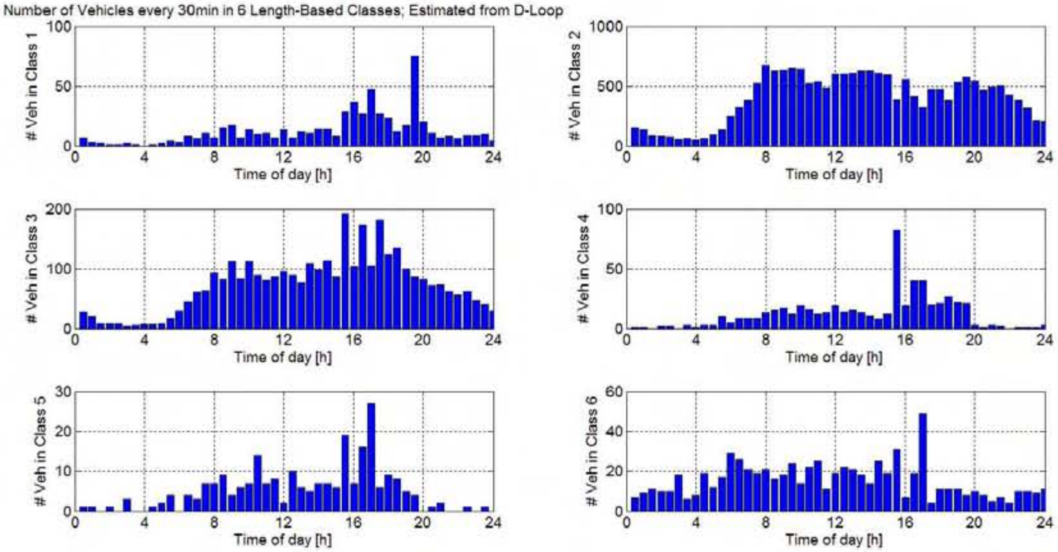


Figure C. 13 BHL EB, Station 7, Lane 5, Vehicle Classification Every 30min; D-loop

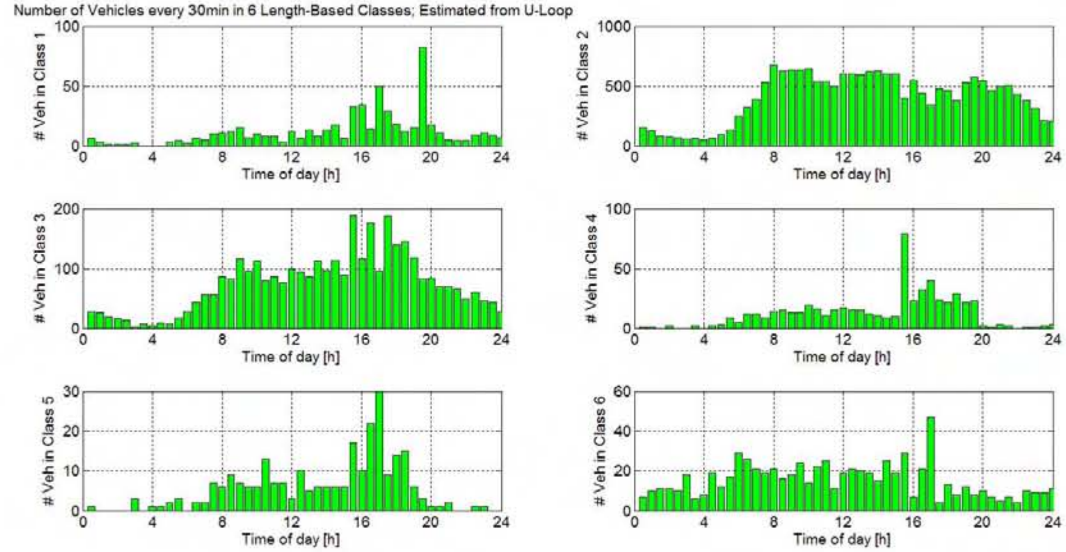


Figure C. 14 BHL EB, Station 7, Lane 5, Vehicle Classification Every 30min; U-loop

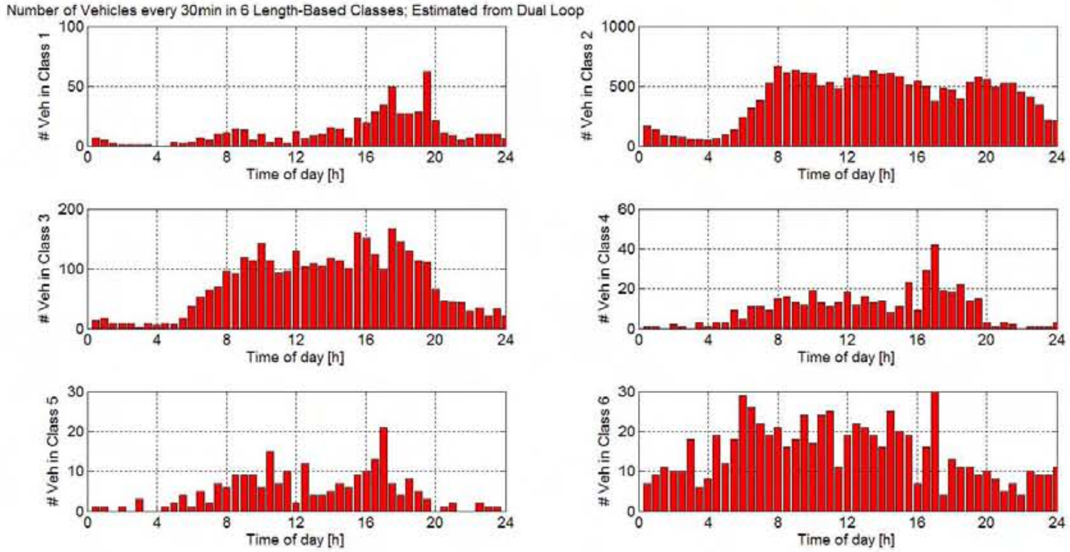


Figure C. 15 BHL EB, Station 7, Lane 5, Vehicle Classification Every 30min; Dual-loop

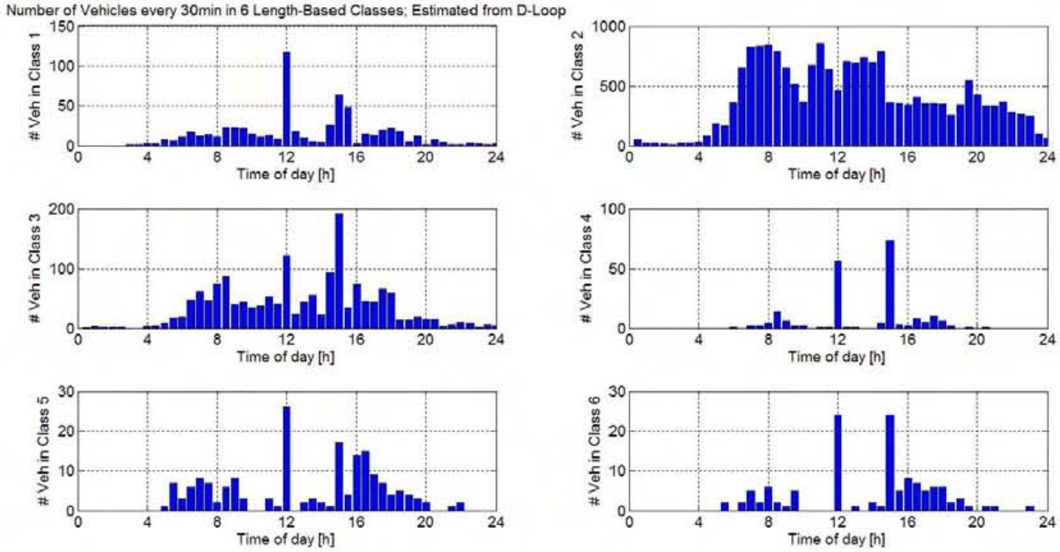


Figure C. 16 BHL WB, Station 7, Lane 1, Vehicle Classification Every 30min; D-loop

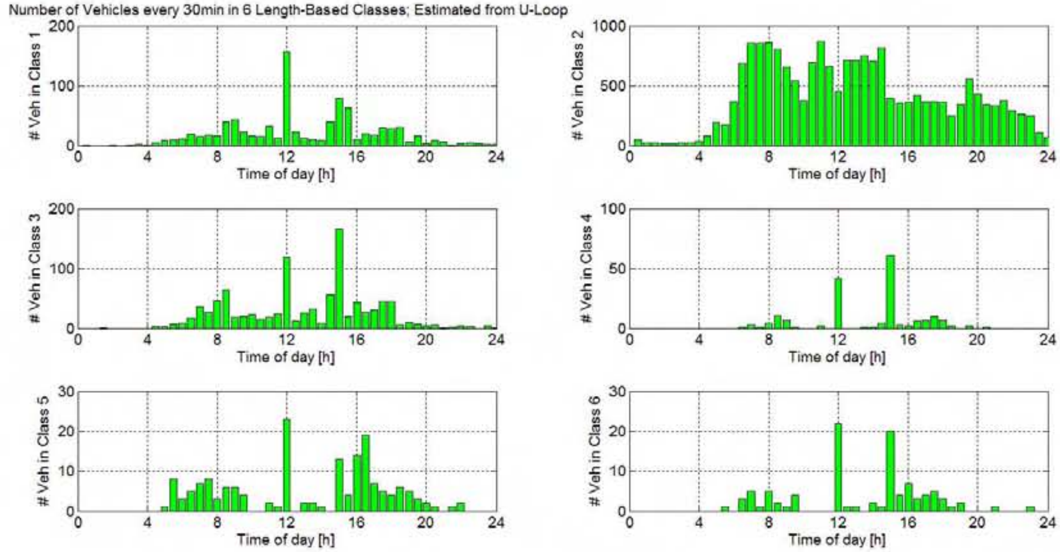


Figure C. 17 BHL WB, Station 7, Lane 1, Vehicle Classification Every 30min; U-loop

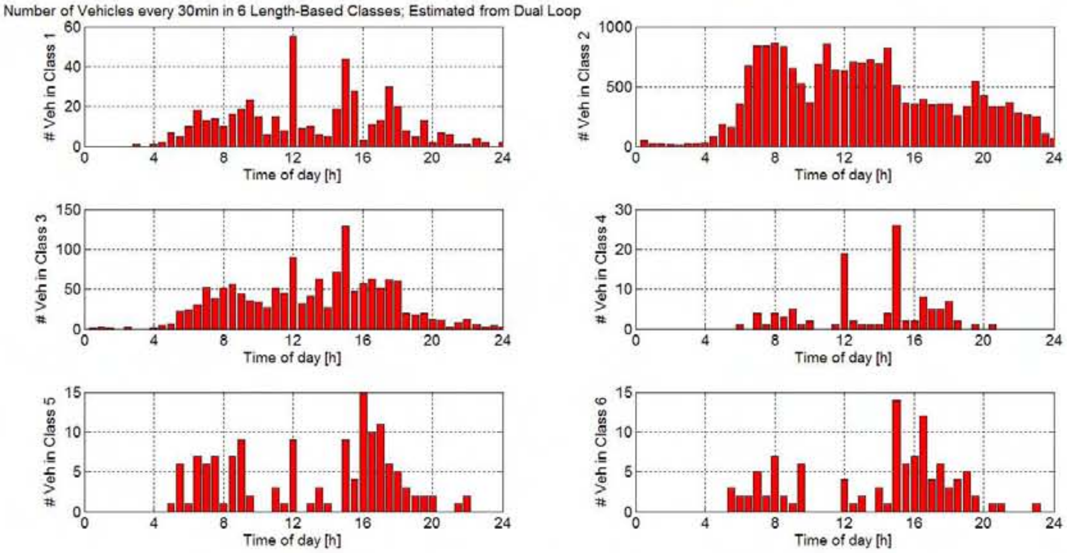


Figure C. 18 BHL WB, Station 7, Lane 1, Vehicle Classification Every 30min; Dual-loop

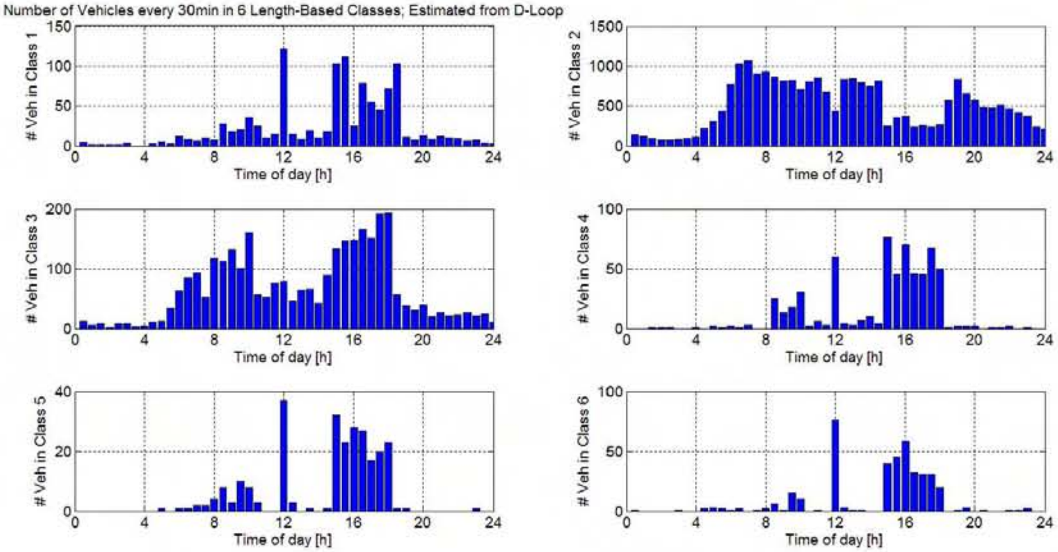


Figure C. 19 BHL WB, Station 7, Lane 2, Vehicle Classification Every 30min; D-loop

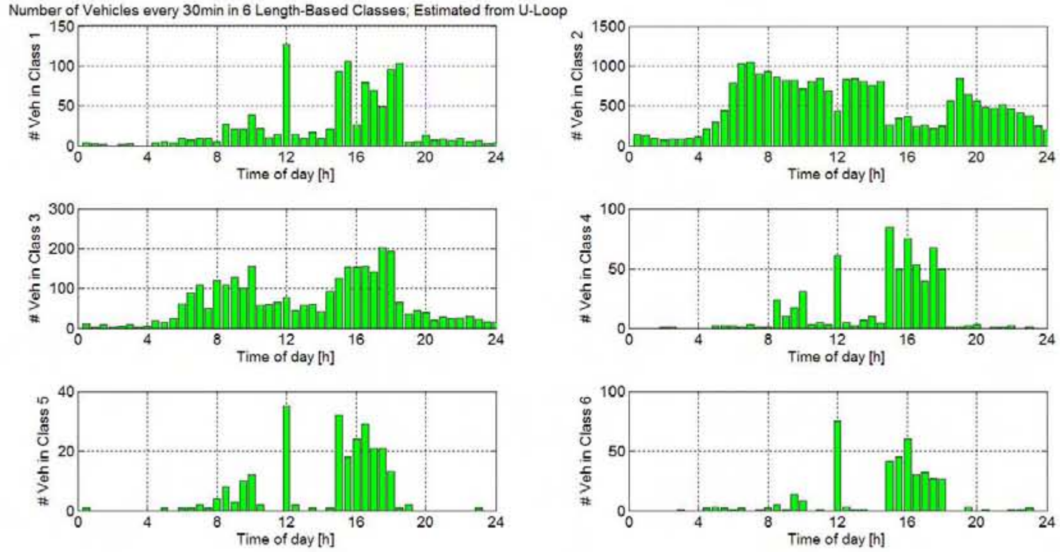


Figure C. 20 BHL WB, Station 7, Lane 2, Vehicle Classification Every 30min; U-loop

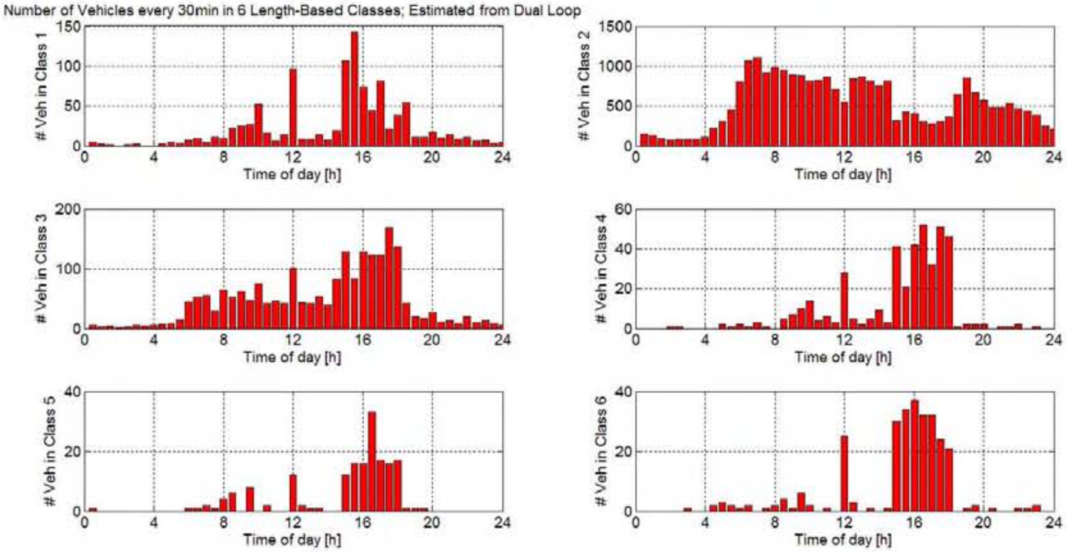


Figure C. 21 BHL WB, Station 7, Lane 2, Vehicle Classification Every 30min; Dual-loop

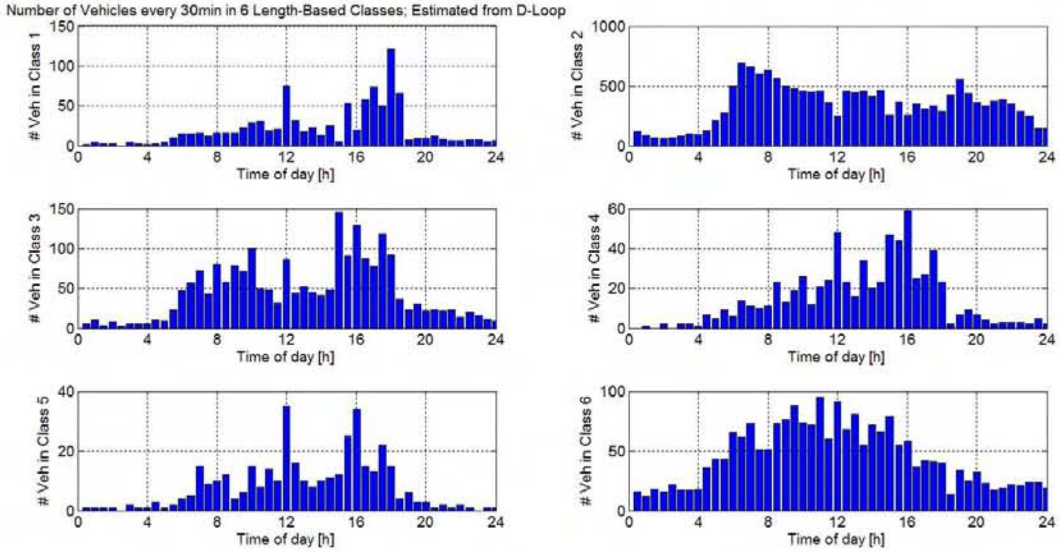


Figure C. 22 BHL WB, Station 7, Lane 3, Vehicle Classification Every 30min; D-loop

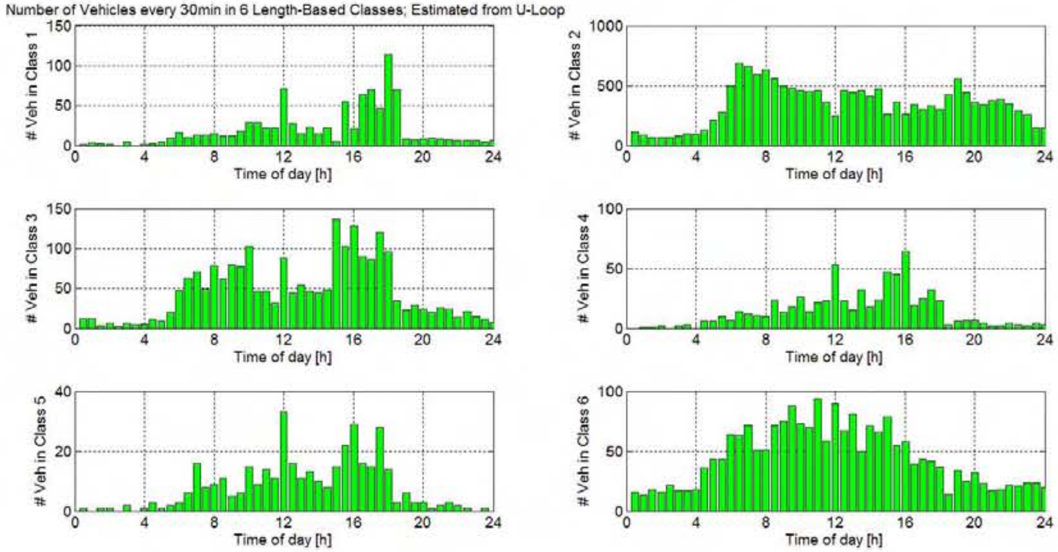


Figure C. 23 BHL WB, Station 7, Lane 3, Vehicle Classification Every 30min; U-loop

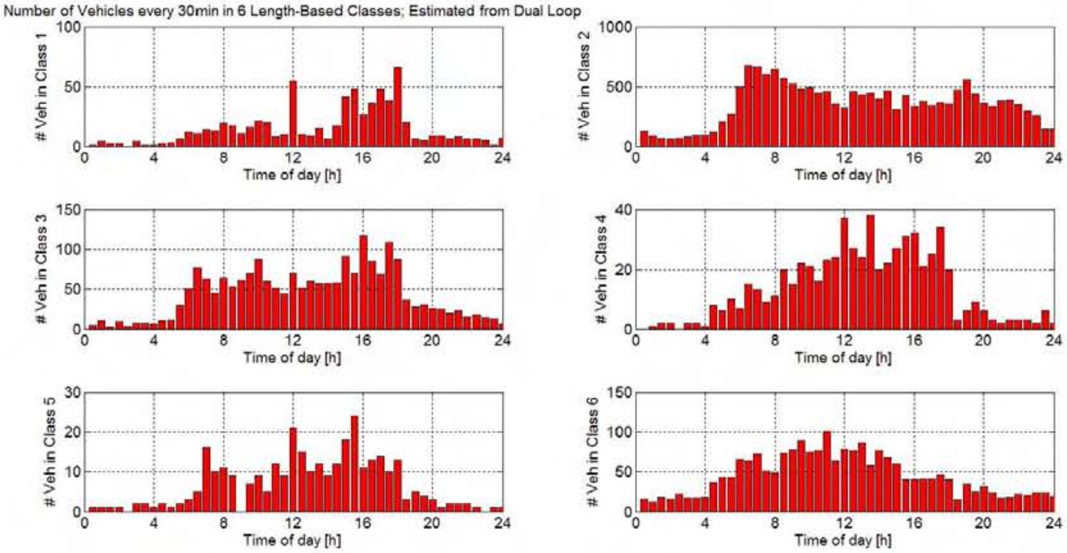


Figure C. 24 BHL WB, Station 7, Lane 3, Vehicle Classification Every 30min; Dual-loop

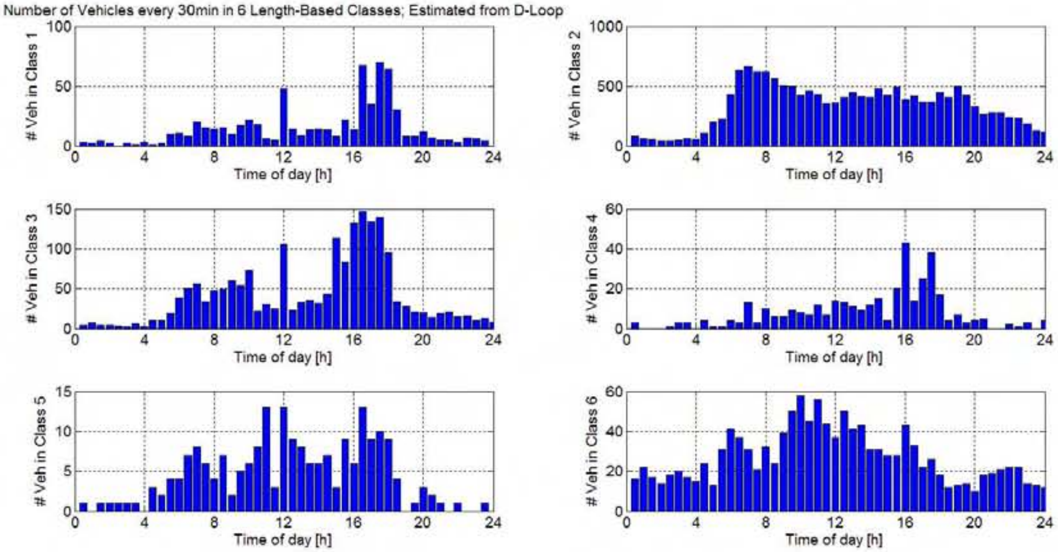


Figure C. 25 BHL WB, Station 7, Lane 4, Vehicle Classification Every 30min; D-loop

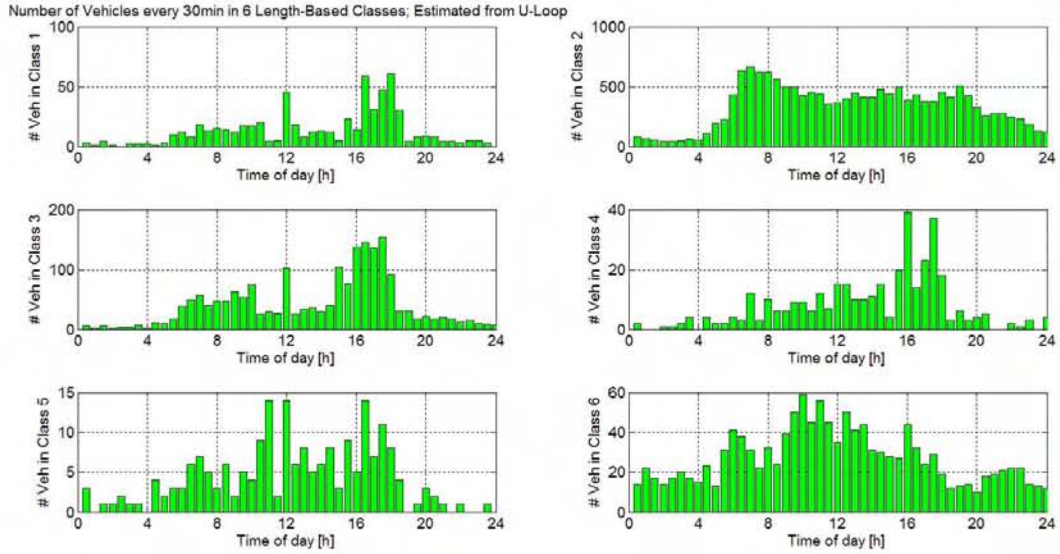


Figure C. 26 BHL WB, Station 7, Lane 4, Vehicle Classification Every 30min; U-loop

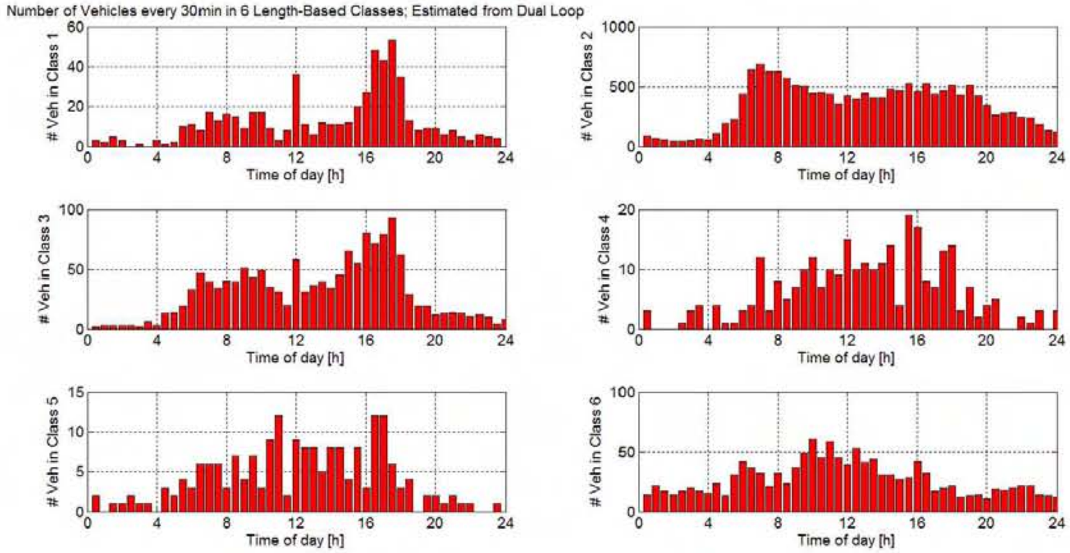


Figure C. 27 BHL WB, Station 7, Lane 4, Vehicle Classification Every 30min; Dual-loop

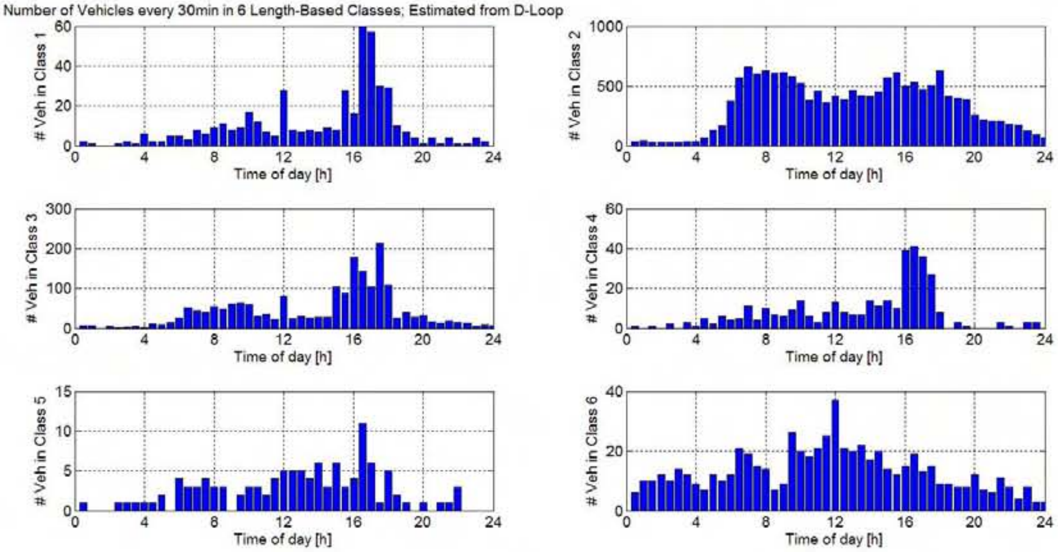


Figure C. 28 BHL WB, Station 7, Lane 5, Vehicle Classification Every 30min; D-loop

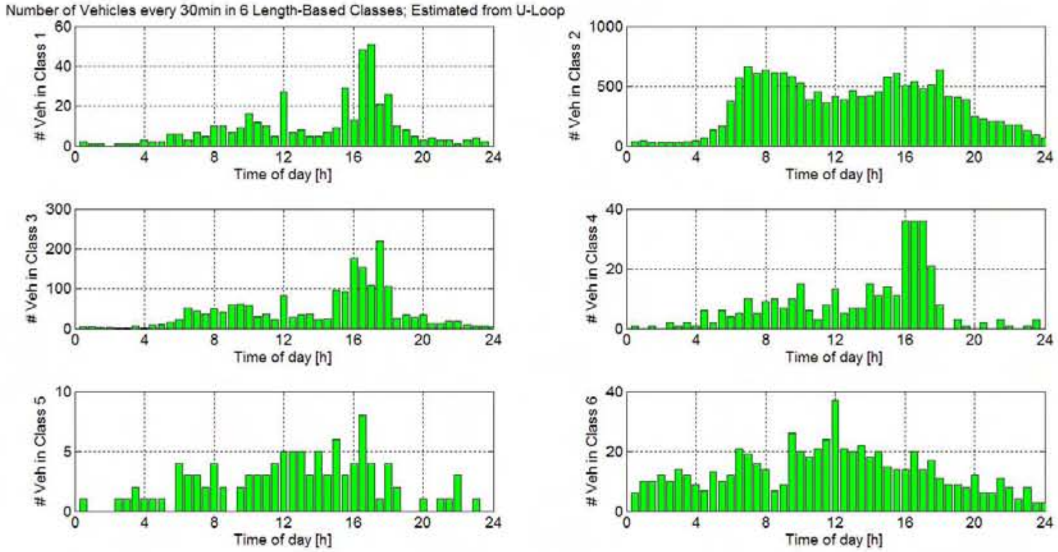


Figure C. 29 BHL WB, Station 7, Lane 5, Vehicle Classification Every 30min; U-loop

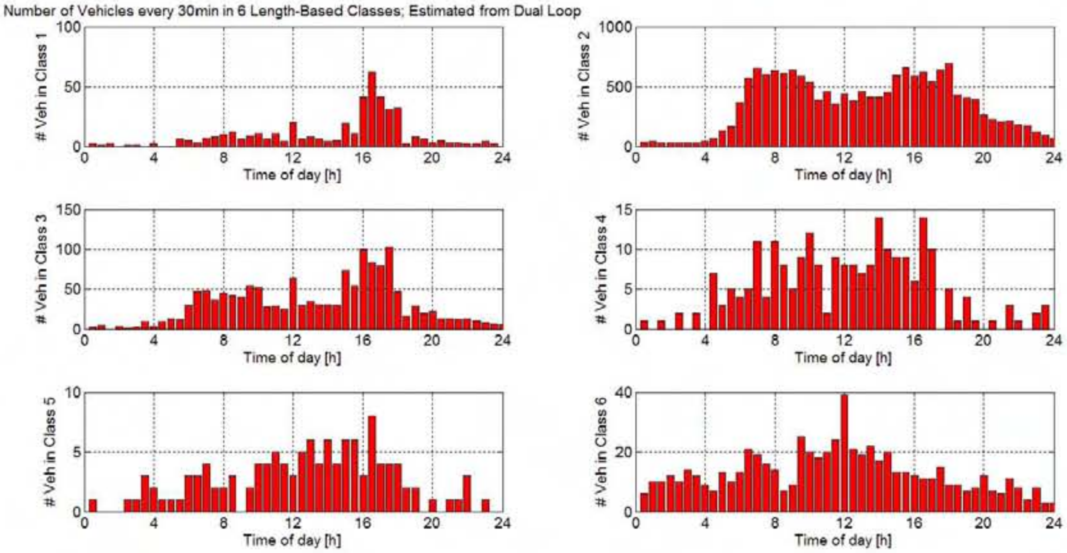


Figure C. 30 BHL WB, Station 7, Lane 5, Vehicle Classification Every 30min; Dual-loop