VARIABLE SPEED LIMITS CONTROL FOR WORK ZONES

VARIABLE SPEED LIMITS CONTROL FOR FREEWAY WORK ZONE WITH SENSOR FAULTS

By SHUMING DU, B.ENG., M.A.Sc

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2020)

Hamilton, Ontario (Civil Engineering)

TITLE: Variable Speed Limits Control for Freeway Work Zone with Sensor Faults

AUTHOR: Shuming Du, B.ENG., M.A.Sc

SUPERVISOR: Dr. Saiedeh Razavi

NUMBER OF PAGES: xvii; 137

Lay Abstract

Freeway work zones can increase congestion with higher travel time, safety risk, emissions and fuel consumption. This research aims to improve traffic conditions near work zones using a variable speed limits control system. By exploiting redundant traffic information, a variable speed limit control system that is insensitive to traffic sensor failures is presented. The proposed system was evaluated under realistic freeway work zone conditions in a simulation environment. The results show that the proposed system can reliably detect sensor failures and consistently provide improvements in mobility, safety and sustainability despite the presence of traffic sensor failures.

Abstract

Freeway work zones with lane closures can adversely affect mobility, safety, and sustainability. Capacity drop phenomena near work zone areas can further decrease work zone capacity and exacerbate traffic congestion. To mitigate the negative impacts caused by freeway work zones, many variable speed limits (VSL) control methods have been proposed to proactively regulate the traffic flow. However, a simple yet robust VSL controller that considers the nonlinearity induced by the associated capacity drop is still needed. Also, most existing studies of VSL control neglected the impacts of traffic sensor failures that commonly occur in transportation systems. Large deviations of traffic measurements caused by sensor faults can greatly affect the reliability of VSL controllers.

To address the aforementioned challenges, this research proposes a fault-tolerant VSL controller for a freeway work zone with consideration of sensor faults. A traffic flow model was developed to understand and describe the traffic dynamics near work zone areas. Then a VSL controller based on sliding mode control was designed to generate dynamic speed limits in real time using traffic measurements. To achieve VSL control fault tolerance, analytical redundancy was exploited to develop an observer-based method and an interacting multiple model with a pseudo-model set (IMMP) based method for permanent and recurrent sensor faults respectively. The proposed system was evaluated under realistic freeway work zone conditions using the traffic simulator SUMO.

This research contributes to the body of knowledge by developing fault-tolerant VSL control for freeway work zones with reliable performance under permanent and recurrent sensor faults. With reliable sensor fault diagnosis, the fault-tolerant VSL controller can

consistently reduce travel time, safety risks, emissions, and fuel consumption. Therefore, with a growing number of work zones due to aging road infrastructure and increasing demand, the proposed system offers broader impacts through congestion mitigation and consistent improvements in mobility, safety, and sustainability near work zones.

Preface

This thesis has been prepared in accordance to the guidelines of the sandwich thesis format from School of Graduate Studies at McMaster University. Three papers are included in this thesis as listed below.

- S. Du and S. Razavi, "Variable speed limit for freeway work zone with capacity drop using discrete-time sliding mode control," *Journal of Computing in Civil Engineering*, vol. 33, no. 2, p. 04019001, 2019.
- S. Du and S. Razavi, "Fault-tolerant control of variable speed limits for freeway work zone using likelihood estimation," *Advanced Engineering Informatics*, vol. 45, p.101133, 2020.
- S. Du and S. Razavi, "Fault-tolerant control of variable speed limits for freeway work zone with recurrent sensor faults," *IEEE Transactions on Intelligent Transportation Systems*, 2020. *Submitted September 8, 2020*, T-ITS-20-09-1842.

The first manuscript is presented in Chapter 2. The work was started in January 2017. The manuscript was submitted in April 2018 and accepted in August 2018. The work was conducted under the supervision of Dr. Saiedeh Razavi. The manuscript is included in this thesis as it provides the design of the VSL controller. My contributions include:

- Development of the VSL controller based on sliding model control
- Implementation and evaluation of the developed system
- Development of the manuscript

The second manuscript is presented in Chapter 3. The work was started in April 2018. The manuscript was submitted in December 2019 and accepted in June 2020. The work was conducted under the supervision of Dr. Saiedeh Razazi. The manuscript is included in this thesis as it provides the design of the VSL controller with fault tolerance to permanent sensor faults. My contributions include:

- Development of the fault-tolerant VSL controller using the observer-based method
- Implementation and evaluation of the developed system
- Development of the manuscript

The third manuscript is presented in Chapter 4. The work was started in April 2018. The manuscript was submitted in September 2020. The work was conducted under the supervision of Dr. Saiedeh Razavi. The manuscript is included in this thesis as it provides the design of the VSL controller with fault tolerance to recurrent sensor faults. My contributions include:

- Development of the fault-tolerant VSL controller using the IMMP-based method
- Implementation and evaluation of the developed system
- Development of the manuscript

The co-author's contributions include:

- Provide supervision and technical advice
- Review and edit the manuscript
- Financially support the research work

Copyright Permission

Acknowledgments

I would like to express my sincere gratitude to my supervisor Dr. Saiedeh Razavi. Your expertise, patience, and generosity have made my studies an enjoyable and memorable experience. You shared your in-depth knowledge with me and guided me on technical level. You encouraged me and gave me advice for my personal and professional development. I am deeply grateful to have your personal and professional guidance during this journey.

I would like to thank my supervisory committee members, Dr. Shahin Sirouspour, Dr. Thia Kirubarajan, and Dr. Wael El-Dakhakhni for your insight comments and advice on my research.

Thanks to my colleagues Dr. Wade Genders, Dr. Jun Wang, and Dr. Nazila Roofigari Esfahan. It has been a pleasure to work with you.

Thanks to Dr. Ratnasingham Tharmarasa for all the advice and discussions. Thanks to Dr. Mark Ferguson and Dr. Hany Hassan for all the intellectually-stimulating discussions. Special thanks to Hui Guo for your continuous encouragement.

To my parents, I am grateful for your encouragement and support throughout my life. Your love has been my greatest motivation.

CONTENTS

Lists of Figures and Tables

List of Figures

List of Tables

List of Abbreviations and Acronyms

IMM	Interacting Multiple Model
IMMP	Interacting Multiple Model with a Pseudo-model Set
KF	Kalman Filter
KWT	Kinematic Wave Theory
ML	Mainline
MPC	Model Predictive Control
PH	Pseudo Health
PI	Proportional-integral
PM	Postmile
PML	Pseudo Mainline
PRP	Pseudo Ramp
PSF	Permanent Sensor Fault
RMSE	Root Mean Square Error
RP	Ramp
RSF	Recurrent Sensor Fault
SCA	State Covariance Adaption
SMC	Sliding Mode Control
SUMO	Simulation of Urban Mobility
TS	Tentative Set
TTC	Time-to-collision
VSL	Variable Speed Limit

Declaration of Academic Achievement

Shuming Du was the main contributor to three manuscripts presented in this thesis. The contributions from the co-author are detailed at the beginning of Chapter 2, 3, and 4.

1 Introduction

1.1 Problem Statement and Motivation

With aging road infrastructure and increasing traffic demand, a growing number of freeway work zones are employed in roadway projects. More than $3.9 billion was spent in road projects for National Highway System in Canada in 2015 [1]. An estimate of $1.3 billion will be invested in Ontario, Canada from 2020 to 2023 for highway restoration and rehabilitation projects [2]. Despite long-term benefits of these maintenance projects, freeway work zones and their associated lane closures can easily lead to congestion and increase travel time, safety risk, emissions, and fuel consumption. In addition, capacity drop phenomena near work zone areas can further reduce work zone capacity and exacerbate traffic congestion [3]. In 2017, approximately 8.8 billion hours delays and additional 3.3 billion gallons of fuel consumption were caused by congestion in the U.S. urban areas [4]. Nearly 10 percent of these congestion costs are attributed to work zones [5]. Around 94 thousand work zone crashes occurred in 2017 in the U.S [6]. Therefore, it is of paramount importance to mitigate the negative impacts caused by freeway work zones. Variable speed limits (VSL) control as one of the intelligent transportation systems has been widely studied to alleviate congestion near freeway work zone areas.

1.1.1 Variable Speed Limits Control

VSL control systems can proactively regulate traffic flow and affect the evolution of the traffic flow near work zone areas. Compared with static speed limits, VSL control can

dynamically adjust speed limits according to real-time traffic conditions or weather conditions. By restricting mainline traffic flow and harmonizing traffic speeds, VSL controllers can mitigate congestion near freeway bottlenecks. Some jurisdictions such as the UK, Australia, and the U.S. have implemented VSL control systems on highways. VSL signs on the M25 highway in the UK from Google street view are shown in Fig. 1.1. VSLs mounted on gantries are displayed to control the traffic flow in response to the traffic conditions.

Fig. 1.1. VSL signs on the M25 highway in the UK (Map data © 2020 Google)

To generate appropriate VSLs, previous studies have developed various VSL control methods. Feedback-based VSL controllers [7], [8] were designed to stabilize traffic states near the critical density via linearization methods such that capacity drop phenomena can be avoided and the maximum work zone capacity can be achieved. However, the discontinuity of the fundamental diagram caused by capacity drop may affect the effectiveness of the linearization methods. In contrast, kinematic wave theory based methods [9], [10] and rule based methods [11], [12] were used to directly derive speed limits for freeway bottlenecks from the fundamental diagram. Since kinematic wave theory based methods and rule based methods heavily relied on the calibrated

fundamental diagram, stochastic characteristics of traffic flow may influence the performance of VSL controllers. In [13], [14], [15], model predictive control methods were developed to solve VSL control as optimization problems. VSLs were generated by maximizing the work zone throughput at the cost of increasing the computational complexity. Methods based on reinforcement learning were also developed to reduce congestion near freeway bottlenecks [16], [17]. However, freeway work zones, especially short-term work zones, may not provide enough traffic data to train and design VSL controllers based on reinforcement learning. Therefore, a VSL controller is needed that can efficiently generate VSLs with consideration of the nonlinearity caused by capacity drop and stochastic characteristics of traffic flow, for freeway work zones.

1.1.2 Sensor Faults

The effectiveness of VSL controllers heavily relies on reliable traffic measurements. With traffic states detected by sensors, VSL control can generate appropriate speed limits in real time. However, stationary traffic sensor faults commonly exist in transportation systems. More than 30 percent of traffic sensors suffer from different types of sensor faults and cannot provide reliable traffic measurements on a typical day in California, U.S. [18]. Due to sensor faults, traffic measurements may have large deviations from the actual traffic states and cause system degradation. However, the previous studies [7]-[17] on VSL control did not consider the impacts of sensor failures. Different types of sensor faults may also have different impacts on VSL controllers.

Sensor faults can be broadly categorized into permanent and recurrent sensor faults. Permanent sensor faults can persist indefinitely until the sensors are repaired whereas

recurrent sensor faults can repeatedly fail and recover without external intervention. Due to uncertain factors such as loose connections and environmental disturbances, the occurrence of recurrent sensor faults can be 10 to 30 times as frequent as permanent sensor faults in electronic systems [19]. Both permanent and recurrent sensor faults also commonly occur in transportation systems [18], [20].

To address the impacts of stationary traffic sensor faults, previous studies have developed various approaches by exploiting the spatial and temporal dependencies of traffic data. Tensor decomposition [21] and generative adversarial networks [22] were proposed to impute faulty traffic measurements offline. However, traffic measurements from a faulty sensor beyond the sensor failure time are incorporated in these offline methods. Since VSL control needs real-time measurements and has limited access to sensor measurements when a sensor fails, offline methods may not be feasible for VSL controllers. In contrast, online methods such as regression models [23], pattern clustering [24], and graph Markov networks [25] were developed to detect sensor faults and provide real-time traffic state estimations when sensor faults occur. However, large historical traffic data were needed to calibrate the estimation models in these online methods. In practice, traffic sensors near short-term freeway work zones may not generate enough data for these online methods. In addition, the previous studies [23]-[25] only considered permanent sensor faults. Accordingly, traffic measurements from a sensor were completely discarded when the sensor fails. In the case of recurrent sensor faults, complete removal of measurements from a faulty sensor may neglect potential faultless data when the sensor restores to working order. Besides, the systems in [23]-[25] may be

vulnerable as fewer traffic sensors were used. Therefore, to ensure the performance of VSL control for freeway work zones when sensor faults occur, it is essential to develop fault diagnosis for permanent and recurrent sensor faults, and provide reliable traffic state estimations for VSL control.

1.2 Research Objectives

The main objective of this research is to develop fault-tolerant VSL control for freeway work zones with consideration of permanent and recurrent sensor faults. To mitigate the negative impacts caused by freeway work zones, VSL control can proactively regulate traffic flow in response to real-time traffic conditions to achieve improvements in mobility, safety, and sustainability. Also, fault tolerance of VSL control to permanent and recurrent sensor faults contributes to practical implementation of the developed VSL controller. To develop the fault-tolerant VSL control system, specific objectives include:

- Design a traffic flow model to understand and describe traffic dynamics near freeway work zone areas.
- Build a nonlinear VSL control with consideration of the discontinuity caused by capacity drop.
- Detect and identify permanent sensor faults as well as recurrent sensor faults.
- Provide reliable traffic state estimations when sensor faults happen.
- Achieve consistent improvements in mobility, safety, and sustainability under fault-free and sensor faults environments.

1.3 Research Methodology

To achieve the aforementioned objectives, the framework of the fault-tolerant VSL control system is shown in Fig. 1.2.

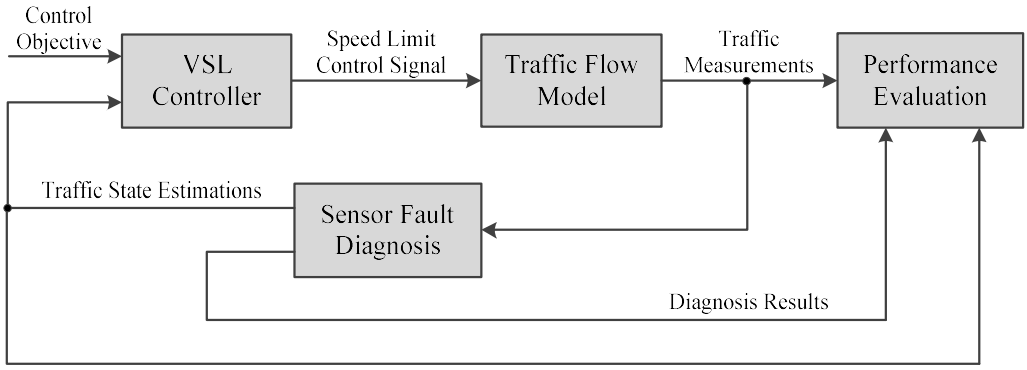


Fig. 1.2. Fault-tolerant VSL control system framework

The VSL controller can generate speed limit control signals to track the control objective which is the critical density using traffic state estimations from the sensor fault diagnosis component. Then the speed limit control signal can affect the evolution of traffic flow that is described by the traffic flow model. Traffic measurements from sensors are sent to the sensor fault diagnosis to detect sensor faults and provide reliable traffic state estimations to the VSL controller. The performance of the fault-tolerant control system is evaluated in the performance evaluation component. The details of each component and the corresponding chapters are discussed as follows.

- VSL controller:

The design of the VSL controller is the focus of Chapter 2. To address the aforementioned limitations of VSL control, a simple yet robust VSL controller that

considers the discontinuity caused by capacity drop is needed. Due to the nonlinearity caused by capacity drop, the VSL controller was designed based on discrete-time sliding mode control. The control objective is to stabilize traffic density near the critical density from the fundamental diagram, such that the maximum work zone throughput can be achieved. With a sliding surface designed, the VSL controller can drive the system state trajectory to the desired equilibrium state using speed limit control signals. The controller with strong robustness is not sensitive to the parameter variations such as noisy traffic demand and traffic disturbance. Due to the stochastic nature of traffic flow, the inherently robustness and variable structure control of the controller make it suitable and efficient to achieve VSL control. To understand the impacts of speed limit control signals on the system state trajectory, the traffic flow model was designed.

- Traffic flow model:

The design of the traffic flow model is another focus of Chapter 2. Capacity drop phenomena and impacts of VSLs are the main considerations in the design of the traffic flow model. Discrete-time traffic flow model was designed using the modified cell transmission model. Capacity drop models were incorporated to consider the nonlinearity of the fundamental diagram near work zone areas. The restricted traffic flow governed by VSLs was used to depict the impacts of VSLs on traffic dynamics.

- Sensor fault diagnosis:

The sensor fault diagnosis component is the focus of Chapter 3 and 4. Specifically, Chapter 3 deals with permanent traffic sensor faults while Chapter 4 focuses on recurrent traffic sensor faults. Analytical redundancy is exploited to achieve fault diagnosis in both

chapters. In Chapter 3, the observer-based method was developed to provide redundant traffic state estimations. By comparing the likelihood estimations from two observers and a Kalman filter, sensor faults with zero flow rates in both stationary mainline traffic sensors and ramp traffic sensors are detected and identified. In Chapter 4, the interacting multiple model with a designed pseudo-model set (IMMP) method was developed to diagnose recurrent sensor faults and provide reliable traffic state estimations to the VSL controller. Different types of recurrent sensor faults such as faults with zero flow rate, partial mainline sensor faults, and faults with abnormally high flow rates are considered. To reduce the computational complexity caused by different types of sensor faults, an adaptive model set was designed to reduce the number of models running in parallel. State covariance adaption and the pseudo-model set were introduced to compensate for the discrepancies between the model parameters and the extent of corresponding sensor faults. Then reliable traffic state estimations are provided using the probabilistically weighted sum of state estimations from models in effect.

- Performance evaluation:

Traffic simulation was used to evaluate the performance of the proposed system. For a safety-critical system like VSL control, real-world development and implementation is highly risky and may have serious consequences if the development fails. In contrast, traffic simulation, which is widely used in transportation planning, offers a safe development environment before implementation in real world. Various models and methods can be extensively studied and evaluated using traffic simulation. In this research, the traffic simulator Simulation of Urban Mobility (SUMO) [26] was used to

evaluate the proposed system.

It is essential to reproduce real-world traffic dynamics in traffic simulation with good accuracy. Realistic work zone conditions are considered in Chapter 2, 3, and 4. Before implementation of the proposed system, the microscopic models such as car-following and lane changing models in simulation were calibrated and validated using historical traffic data in [18] to replicate traffic conditions in real world. The calibrated models are to minimize the discrepancies of the system performance between the simulation and the reality and ensure the effectiveness of the proposed system.

Performance evaluation is included in Chapter 2, 3, and 4 to provide analytical results of the VSL control system. Since the focus of Chapter 2 is to develop a VSL controller, the performance evaluation focuses mainly on the controller performance. In Chapter 2, performance measures include work zone throughput, traffic densities, VSL commands, speed evolution, travel time, time-to-collision, emissions, and fuel consumption near work zone areas. Chapter 3 and 4 mainly focus on sensor fault diagnosis; the reliability of sensor faults detection and identification was added to performance analysis. Specifically, permanent sensor faults were detected using likelihood estimations in Chapter 3 while mode probabilities were used to diagnose recurrent sensor faults in Chapter 4. The impacts of sensor faults on state estimations and VSL control were considered. Estimations of density and ramp flow were evaluated in Chapters 3 and 4. Other performance measures used in Chapter 2 for mobility, safety, and sustainability were also studied in Chapter 3 and 4 to analyze the fault tolerance of the VSL controller.

1.4 Thesis Organization

The framework of the research and the thesis organization is shown in Fig. 1.3.

Chapter 1 presents an overview of the problems along with research motivations, research objectives, and research methodology to achieve the specific research objectives.

Chapter 2 introduces the VSL controller for a freeway work zone without consideration of sensor faults. The focus of Chapter 2 is to develop the traffic flow model and the VSL controller. The performance of the proposed VSL controller was evaluated with analytical results. This chapter includes a published journal article.

Chapter 3 extends the VSL control with fault tolerance to permanent sensor faults. The focus of Chapter 3 is to achieve sensor fault diagnosis and provide reliable traffic state estimations to VSL control. Evaluation under realistic freeway work zone conditions demonstrates the developed system can produce consistent VSL performance despite permanent sensor faults. This chapter includes a published journal article.

Chapter 4 presents fault-tolerant VSL control with the occurrence of recurrent sensor faults. The focus of Chapter 4 is to achieve fault tolerance to recurrent sensor faults for VSL control. Results show the proposed system can accurately estimate traffic states, reliably diagnose recurrent sensor faults, and consistently improve mobility, safety, and sustainability near work zone areas. This chapter includes a submitted manuscript.

Chapter 5 summarizes the work presented in this thesis and discusses contributions, limitations and future work.

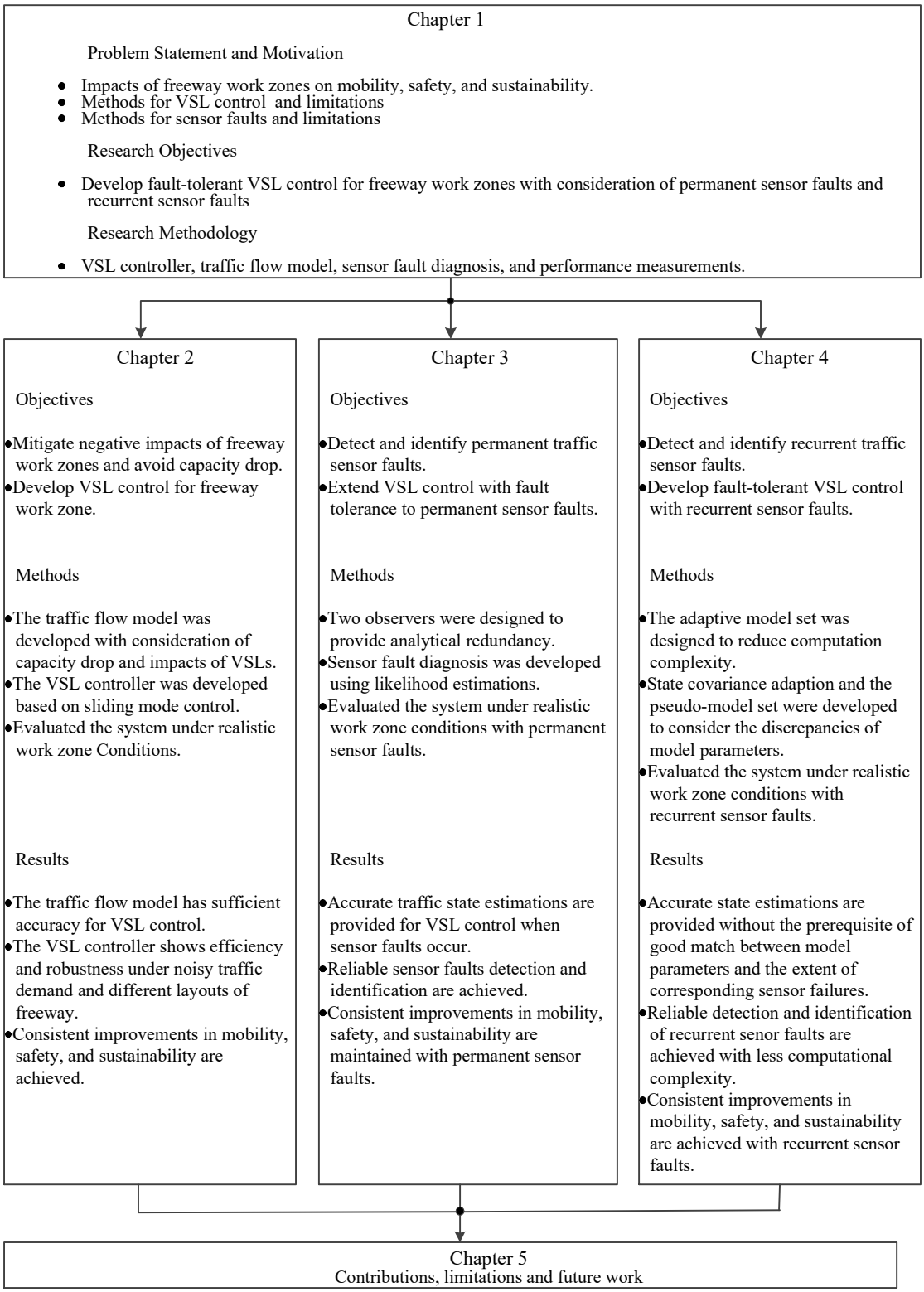


Fig. 1.3. Thesis organization framework

2 Variable Speed Limit for Freeway Work Zone with Capacity Drop Using Discrete-Time Sliding Mode Control

To address the negative impacts of freeway work zones, various VSL control methods have been proposed. However, a simple yet robust VSL controller is still needed. This chapter presents a VSL controller based on sliding mode control with consideration of the discontinuity caused by capacity drop. The proposed system is evaluated under realistic freeway work zone conditions using the traffic simulator SUMO. In comparison to a kinematic wave theory based method, the developed system shows robustness under noisy traffic demand and different freeway layouts. The following journal article is included in this chapter.

- S. Du and S. Razavi, "Variable speed limit for freeway work zone with capacity drop using discrete-time sliding mode control," *Journal of Computing in Civil Engineering*, vol. 33, no. 2, p. 04019001, 2019.

The co-author's contributions include:

- Provide supervision and technical advice
- Review and edit the manuscript
- Financially support the research work

2.1 Abstract

Freeway work zones with lane closures have direct negative impacts on travel time, safety, and environmental sustainability. Capacity drop at the onset of congestion can further reduce the discharging rate at work zone areas and worsen traffic conditions. Existing studies have developed various variable speed limit (VSL) control methods to mitigate the congestion; however, a simple yet robust VSL control strategy that considers the nonlinearity induced by capacity drop is still lacking. To address the above-mentioned issues, this study proposes a VSL strategy using a nonlinear traffic flow model and discrete-time sliding mode control for a freeway work zone. The developed traffic flow model incorporates the nonlinearity caused by the capacity drop at the work zone using the cell transmission model. The sliding mode controller is designed to drive the traffic state, which is acquired from the traffic flow model, to the desired equilibrium state with different convergence rates. Under speed limit constraints, the VSL scheme is generated to regulate the traffic flow and mitigate the congestion near the work zone area. The proposed system is implemented and evaluated using the traffic microscopic simulator SUMO. The results indicate that the proposed VSL control can consistently improve traffic mobility, safety and environmental sustainability under noisy traffic demand and different control scenarios. Compared with the uncontrolled scenario, the developed system shows improvements by approximately reducing 17% of average travel time, 90% of safety risks, and 6% of NOx, CO2 emissions and fuel consumption.

2.2 Introduction

Freeway traffic congestion is a critical issue in modern societies as it leads to higher travel time, safety risks, environmental emissions and fuel consumption. Estimated 888 million hours delays, 90 thousand crashes, and 310 million gallons of fuel waste have resulted from work zones in the U.S. [27]. To improve traffic conditions in Canada, $3.9 billion was invested into the National Highway System for the expansion of road networks, rehabilitation of freeway assets, deployment of intelligent transportation systems, etc. in 2015 [1]. However, despite long-term benefits from these freeway projects such as road expansion and maintenance, the associated work zones create traffic bottlenecks which result in substantial traffic delays and economic losses. Therefore effective control is needed to regulate the traffic flow at freeway work zone areas to improve traffic mobility, safety and environmental sustainability.

The maximum flow rate at a work zone bottleneck can further decrease when a queue forms upstream of the bottleneck [28], [29]. The capacity decreases due to a work zone lane closure. Therefore, when demand exceeds the reduced capacity, traffic can easily become congested, forming a queue upstream of the work zone. Such congestion can further diminish the work zone capacity. This capacity drop phenomenon not only worsens traffic conditions but also leads to a discontinuous and nonlinear fundamental diagram around the work zone area [30], [31]. Due to the nonlinearity, it becomes a challenging problem to control the traffic flow around the work zone bottleneck [32]. To address this problem and to prevent the occurrence of capacity drop, variable speed limit (VSL) control, one of the intelligent transportation system technologies, has been widely

studied. A VSL strategy can dynamically change the speed limit upstream of the work zone bottleneck, whereby the traffic mainline inflow is restricted [33] or the traffic homogenization is improved [34]. For the last two decades, many different approaches have been studied to develop the VSL scheme.

To avoid capacity drop and improve traffic conditions, a number of studies have developed various local feedback VSL strategies with the creation of some distance as an acceleration zone upstream of the bottleneck. In [7], a local proportional-integral (PI) feedback VSL controller was proposed to regulate traffic flow. The acceleration area was created for vehicles to accelerate from the congestion area thereby avoiding capacity drop. Another local feedback controller whose second loop was replaced by a lookup table was developed in [35]. The VSL feedback controller integrated with a delay balancing algorithm was extended to the multiple bottlenecks scenario [8]. This PI local feedback controller was also studied in the microscopic simulation [36]. From previous studies, the local feedback controllers were designed via the linearization around the critical density in the fundamental diagram. However, the effectiveness of this linearization could be affected because of the discontinuity of the fundamental diagram caused by capacity drop. This discontinuity around the critical density can easily occur when the capacity drop leads to flow disruption [37]. Thus without consideration of the nonlinearity induced by capacity drop, the performance of the local feedback VSL controllers may be affected.

Since potential improvements could be made by exploiting abundant information from traffic flow theory, many researchers have sought to utilize kinematic wave theory to design VSL controllers. A theoretical framework based on kinematic wave theory for a

VSL strategy was developed in [38], where the maximum flow rate was sustained by transferring the capacity flow from congested states to uncongested states via an acceleration zone immediately upstream of the bottleneck. In [39], a VSL controller using kinematic wave theory was also developed, where an acceleration zone was incorporated to improve traffic stability. A link queue model with a PI controller was developed to take capacity drop phenomena into account [40]. A bang-bang controller was introduced to calculate speed limits using calibrated relations between traffic speed, density, and flow [41]. However, the aforementioned studies on kinematic wave theory to develop VSL control heavily relied on deterministic characteristics of traffic flow. Since critical parameters of the fundamental diagram could vary frequently [42], it may influence the effectiveness of VSL control due to the stochastic nature of traffic flow. Furthermore, the impact of an acceleration zone was not fully investigated. Our study shows that inappropriate settings of the length of an acceleration zone still lead to capacity drop. This will be further discussed in the following sections of this paper.

Traffic prediction methods were also studied to design a VSL strategy to mitigate traffic congestion. Model predictive control (MPC) was used to design a VSL controller to suppress the shockwave effects [43]. After traffic states were predicted, the objective function was optimized to generate the VSL scheme. The fundamental diagram was used to modify traffic prediction models in favor of online computation [44]. Later, the discrete first-order MPC and mode dependent cell transmission model on MPC were developed in [14] and [45] respectively to further improve the performance of VSL controllers. A Kalman filter was also adopted into the MPC method to improve state

prediction [46]. Instead of using macroscopic models, microscopic models were utilized to optimize speed limits using MPC by predicting the states of each vehicle [15], [47]. However, using MPC methods to develop VSL control requires sufficient accuracy of traffic state prediction which makes it a challenging problem when the nonlinearity induced by capacity drop is encountered. In addition, the complexity has greatly increased in the optimization process to derive the speed limits solution. This may make it difficult to implement MPC-based VSL control in a real-world environment.

To reduce the complexity and improve traffic conditions, rule-based methods and lane change control were studied to develop a speed limits strategy. The rule-based decision tree in [11] was used to design speed limits to improve safety. The estimations of delays and crash costs were utilized as rules to calculate speed limits through minimizing the cost estimation. On the other hand, lane change control combined with speed limits was adopted in [48] to eliminate capacity drop at a bottleneck. Although the rule-based VSL control largely reduces computational complexity, the stochastic nature of traffic flow may influence the controller performance. With respect to lane change control, extensive experiments are needed to develop an appropriate lane change scheme under different speed limits.

From extensive previous studies, it can be seen that VSL control can effectively mitigate traffic congestion. However, the following limitations exist:

1) Neglect of the discontinuity and nonlinearity caused by capacity drop phenomena. The discontinuity of the fundamental diagram occurs at work zone areas when traffic is under congested states. However, this discontinuity and nonlinearity were not considered

in the design of local PI feedback VSL controllers.

2) Absence of consideration of stochastic characteristics and disturbances inherent in traffic flow. The VSL strategy derived based on the kinematic wave theory or rule-based method overlooks the actual traffic disturbance, particularly at an acceleration zone.

3) High complexity in predictive control. The complexity is attributed to the prediction and optimization process. In the prediction, multiple time steps of traffic states are calculated and predicted. Meanwhile, high computation load is involved in the optimization process.

To overcome the aforementioned issues, this study proposes the VSL controller for a freeway work zone area using the discrete-time sliding mode control. Given the capacity drop phenomenon, the traffic flow model is first developed at the work zone bottleneck based on the cell transmission model. The discontinuity induced by the capacity drop is incorporated into the nonlinear discrete-time traffic flow model. To take the stochastic nature of traffic into account, the discrete-time sliding mode controller with different convergence rates, which has strong robustness and simple design, is developed to generate the VSL scheme using the built traffic flow model. Thus the sophisticated prediction and optimization are avoided in the controller. The effectiveness of the proposed VSL control system, with respect to improvements on traffic mobility, safety and environmental sustainability, was assessed in the traffic microscopic simulator SUMO [26].

The remainder of this paper is organized as follows: the nonlinear traffic flow model is built considering the capacity drop at the work zone area; following that, the discrete-

time sliding model control is developed using the discrete-time traffic flow model; with the performance indicators introduced, the proposed system is implemented and evaluated with analytical results; finally limitations, future work and conclusions of this paper are presented.

2.3 System Model

Freeway work zones are set up due to lane expansion, pavement repair or other construction operations. Consequently, partial lane(s) will be closed and lead to capacity loss. As shown in Fig. 2.1, only 2 lanes are open at the work zone area with construction warning signs placed along the 3-lane freeway.

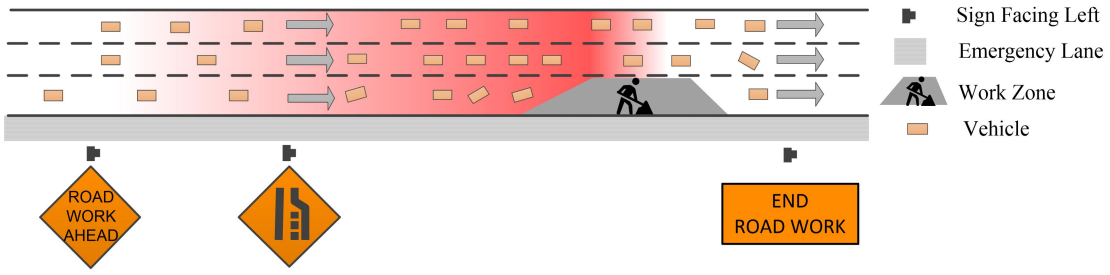


Fig. 2.1. Traffic network with a work zone

When traffic demand exceeds the work zone capacity, a queue can form at the beginning of the lane-dropped area and propagate upstream at a certain speed. Three aspects of detrimental effects emerge when the congestion happens: 1) traffic throughput decreases and more delays occur due to capacity drop; 2) it aggravates the potential crash risks as there are more sudden brakes when vehicles with high speeds encounter the back-propagation queue wave; 3) environmental emissions and fuel consumption increase due

to more frequent accelerations and decelerations upstream of the work zone area.

To develop effective control to alleviate the congestion near work zone areas, it is essential to first understand the traffic dynamics at the work zone area. Therefore, in this section, the general traffic flow model will be discussed, followed by the model for the work zone bottleneck.

2.3.1 General Traffic Flow Model

In the modified cell transmission model, traffic density as the state variable is used to reproduce traffic dynamics [49]. As shown in Fig. 2.2, the freeway is partitioned into n+3 segments including the VSL control zone, the acceleration zone and the work zone.

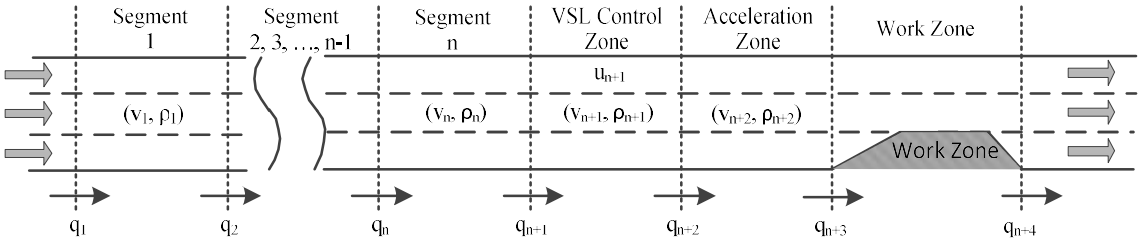


Fig. 2.2. Freeway layout with multiple segments (not to scale)

The length, average speed and density in each segment are represented as L_i, v_i and ρ_i respectively. VSLs are imposed at the VSL control zone with a speed limit sign installed at the beginning of the VSL control zone and another speed limit sign installed at the beginning of the acceleration zone, which allows vehicles to accelerate or keep maximum free speed to get through the work zone area. This study considers only one freeway work zone, for which the effects of traffic states inside and downstream of the

work zone segment are not considered. Multiple work zones and the effects of other downstream bottlenecks will be studied in the future work.

According to the conservation law [50], traffic density can be updated as:

$$\rho_i(k+1) = \rho_i(k) + \frac{\Delta T_s}{L_i}[q_i(k) - q_{i+1}(k)], i = 1, 2, \cdots, n+2 \tag{2.1}$$

where $q_i(k)$ is the total traffic flow from segment i to segment i+1 during the sample time interval $\Delta T_s[k, k+1)$. The length L_i satisfies the condition that $L_i \geq v_{free,i}\Delta T_s$ where $v_{free,i}$ is the free flow speed at the segment i. To calculate the traffic inflow $q_i(k)$ in Eq. (2.1), we assume the triangular fundamental diagram [51], then

$$\begin{cases} q_i(k) = \min\{S_{i-1}(k), R_i(k)\} \\ S_{i-1}(k) = \min\{v_{i-1}(k)\rho_{i-1}(k), q_{\max,i-1}\} \\ R_i(k) = \min\{q_{\max,i}, \omega(\rho_j - \rho_i)\} \\ \quad i = 1, 2, \cdots, n+2 \end{cases} \tag{2.2}$$

where $S_{i-1}(k)$ is the maximum traffic flow capable of supplying from segment i-1 to segment i, and $R_i(k)$ is the maximum traffic flow capable of receiving by segment i. ω and ρ_j are the backward propagating wave speed and jam density respectively in the fundamental diagram. $q_{\max,i}$ is the maximum flow, i.e. the capacity at segment i.

It can be seen that traffic dynamics along the freeway can be derived using the Eq. (2.1) and Eq. (2.2). However, imposed speed limits and capacity drop phenomena at the work zone have a great impact on the traffic flow $q_{n+2}(k)$ and $q_{n+3}(k)$ which needs further discussions on the density evolution at the acceleration zone in the following subsection.

2.3.2 Model for Work Zone Bottleneck

Under the assumption of triangular fundamental diagram, the model parameters are depicted in Fig. 2.3.

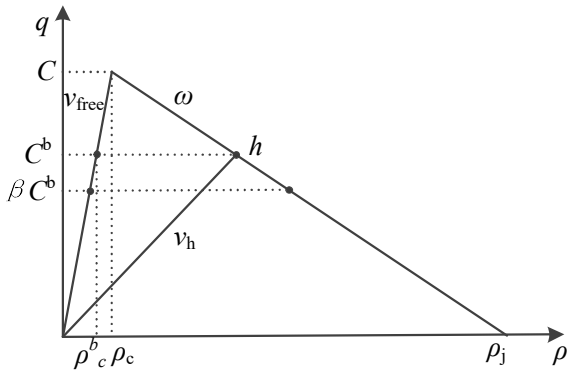


Fig. 2.3. Fundamental diagram with capacity drop

The capacity C of the 3-lane freeway can be calculated as $C = v_{free}\rho_c$ where v_{free} and ρ_c are the free flow speed and critical density respectively. Accordingly, the capacity C^b of the work zone bottleneck is derived as $C^b = v_{free}\rho_c^b = 2/3C$ in the case of the 3-lane freeway with one lane closure at the work zone. ρ_c^b is the critical density corresponding to the capacity C^b.

The impact of capacity drop can be described in the fundamental diagram, as it makes the bottleneck capacity C^b decrease to βC^b where β (0,1) is the drop factor [32], [37], [52]. When traffic demand increases and the density $\rho_{n+2}(k)$ is larger than the critical density ρ_c^b at the bottleneck, the traffic flow $q_{n+3}(k)$ will exceed the work zone capacity C^b. This results in a queue forming upstream of the work zone and moving backward

with the wave speed ω. Then the work zone capacity C^b drops to be the capacity βC^b. Therefore the traffic flow $q_{n+3}(k)$ can be expressed as:

$$q_{n+3} = \begin{cases} v_{n+2}(k)\rho_{n+2}(k) & , \quad \rho_{n+2}(k) \le \rho_c^b \\ \beta C^b & , \quad \rho_{n+2}(k) > \rho_c^b \end{cases} \tag{2.3}$$

In the VSL control zone, the speed limit $u_{n+1}(k)$ is imposed to restrict the traffic flow. Therefore the impact of the speed limit can be reflected through the variation of $q_{\max,n+2}$ which is calculated as $q_{\max,n+2} = \omega\rho_j u_{n+1}(k) / [u_{n+1}(k) + \omega]$. Thus the traffic flow $q_{n+2}(k)$ can be derived as

$$q_{n+2} = \begin{cases} \min\{v_{n+1}(k)\rho_{n+1}(k), \dfrac{\omega\rho_j u_{n+1}(k)}{u_{n+1}(k)+\omega}\} & , \quad \rho_{n+2}(k) \le \rho_c \\ \min\{v_{n+1}(k)\rho_{n+1}(k), \dfrac{\omega\rho_j u_{n+1}(k)}{u_{n+1}(k)+\omega}, \omega(\rho_j - \rho_{n+2}(k))\} & , \quad \rho_{n+2}(k) > \rho_c \end{cases} \tag{2.4}$$

By substituting the Eq. (2.3) and Eq. (2.4) into the general traffic flow model Eq. (2.1), the evolution of the density $\rho_{n+2}(k)$ at the acceleration zone can be derived and formulated as the Eq. (2.5) which shows the dynamics of traffic state $\rho_{n+2}(k)$ given the speed limit control at the *k*th time step.

$$\rho_{n+2}(k+1) = f(\rho_{n+2}(k), u_{n+1}(k)) \tag{2.5}$$

The objective of the VSL scheme is to stabilize the density $\rho_{n+2}(k)$ at the critical density ρ_c^b of the bottleneck such that the capacity drop could be avoided and the maximum flow rate is achieved. Accordingly, the travel delay, the sudden brakes, and extra emissions will reduce. However, the Eq. (2.5) is a nonlinear state equation. Thus the

discrete-time sliding mode control, which is capable of stabilizing the nonlinear state equation Eq. (2.5) by driving the density state towards the equilibrium point with different convergence rates, will be discussed in the next section.

2.4 Sliding Mode Control Design

Sliding mode control is a nonlinear control method which can achieve the global stability compared with the local feedback controller. By creating the equilibrium state on a designed sliding surface, the controller can generate the switching control signal thereby driving the system to reach the sliding surface with finite time [53]. Since the switching control signal can be simple and not precise, sliding model control with strong robustness is not sensitive to parameter variation such as noisy traffic demand and traffic disturbance. Therefore sliding mode control has the strength of simple design, global stability and good robustness.

In this section, the traffic flow model is further simplified and discussed given the existence of speed limits; and the controller is proposed with its stability discussed afterwards. In practice, the measurement of traffic states and control of speed limits are conducted in discrete time steps. Therefore, the discretization of both work zone model and sliding mode control are studied. Meanwhile, the speed limit constraints are also considered. As shown in Fig. 2.4, by using the detected traffic data, the discrete-time work zone model can provide traffic states evolution to the controller. Then, the discrete-time sliding mode controller with the saturation function will generate the speed limits

schemes to regulate the traffic flow upstream of the work zone. Under the speed limit constraints, the VSLs will be imposed in the calibrated traffic network.

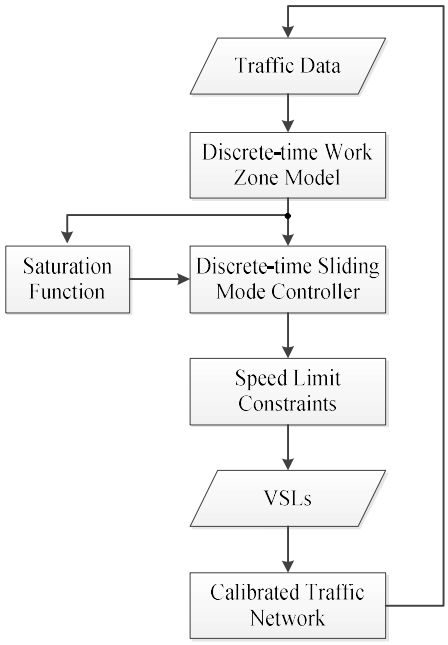


Fig. 2.4. VSL control framework

2.4.1 Discrete-time Work Zone Model

In the Section 2.3, the work zone model was formulated as Eq. (2.5). This model is simplified in this subsection so as to develop the sliding mode controller. In Eq. (2.4), the calculation of the traffic flow $q_{n+2}(k)$ also relies on the density $\rho_{n+2}(k)$ at the acceleration zone. Thus traffic demand which is correlated with the traffic density is discussed.

When traffic demand is light and lower than the work zone capacity C^b, there is no need to restrict the traffic flow with speed limits. On the other hand, when the traffic demand exceeds the work zone capacity C^b, the VSL is employed and results in the following relation:

$$v_{n+1}(k)\rho_{n+1}(k) = \frac{\omega\rho_j u_{n+1}(k)}{u_{n+1}(k)+\omega} \tag{2.6}$$

Consequently, the traffic flow $q_{n+2}(k)$ can be rearranged as Eq. (2.7) with the speed limit.

$$q_{n+2} = \begin{cases} \dfrac{\omega\rho_j u_{n+1}(k)}{u_{n+1}(k)+\omega} & , \quad \rho_{n+2}(k) \le \rho_c \\ \min\{\dfrac{\omega\rho_j u_{n+1}(k)}{u_{n+1}(k)+\omega}, \omega(\rho_j - \rho_{n+2}(k))\} & , \quad \rho_{n+2}(k) > \rho_c \end{cases} \tag{2.7}$$

With regard to the condition $\rho_{n+2}(k) > \rho_c$ in Eq. (2.7), the relation:

$$\frac{\omega\rho_j u_{n+1}(k)}{u_{n+1}(k)+\omega} \le \omega(\rho_j - \rho_{n+2}(k)) \tag{2.8}$$

is satisfied and constrained by the stability of the discrete-time sliding mode control discussed in the following stability theorem. Therefore, by combining the Eq. (2.1), Eq. (2.3), Eq. (2.7) and Eq. (2.8), the simplification of discrete-time the work zone model can be expressed as:

$$\rho_{n+2}(k+1) = \rho_{n+2}(k) + \frac{\Delta T_s}{L_{n+2}}[\frac{\omega\rho_j u_{n+1}(k)}{u_{n+1}(k)+\omega} - \psi(k)] \tag{2.9}$$

where $\psi(k)$ can be computed by

$$\psi(k) = \begin{cases} v_{n+2}(k)\rho_{n+2}(k) & , \quad \rho_{n+2}(k) \le \rho_c^b \\ \beta C^b & , \quad \rho_{n+2}(k) > \rho_c^b \end{cases} \tag{2.10}$$

The Eq. (2.9) is the simplified form of the state equation Eq. (2.5) for the controller design. However, the nonlinearity and discontinuity of the work zone model still exist.

2.4.2 Discrete-time Sliding Mode Controller Design

To avoid the capacity drop and achieve the maximum flow rate, the objective of the discrete-time sliding mode controller is to generate the speed limits scheme to stabilize the traffic density $\rho_{n+2}(k)$ at the critical density ρ_c^b. On the other hand, the VSL controller based on kinematic wave theory controls the speed limit at the speed v_h, which is the speed corresponding to the congested state *h* at capacity flow shown in Fig. 2.3, to avoid the capacity drop. However, traffic disturbances may affect the effectiveness of the VSL controller based on kinematic wave theory. The impacts will be further discussed in the simulation and results section.

To drive the system to the equilibrium state, the sliding surface is designed as:

$$s(k) = c[\rho_{n+2}^e - \rho_{n+2}(k)] \tag{2.11}$$

where c is a constant non-zero parameter and ρ_{n+2}^e represents the time invariant equilibrium state, i.e. the critical density ρ_c^b. From the sliding surface, it can be seen that when the system trajectory has reached the sliding surface and stayed on the surface, the control objective is achieved.

To force the system trajectory to reach the surface, the discrete-time sliding mode control is developed based on the reaching process [53]. The reaching law is designed as Eq. (2.12) to drive the system from the initial state to reach and stay confined to the surface Eq. (2.11):

$$s(k+1) = s(k) - \Delta T_s \varepsilon \,\mathrm{sgn}(s(k)) - \Delta T_s q s(k) \tag{2.12}$$

where ε and q are positive constant parameters and sgn($s(k)$) is the signum function. In Eq. (2.12), the component $-\Delta T_s q s(k)$ makes the traffic density trajectory move asymptotically towards the surface at an exponential convergence rate. However, it cannot guarantee the finite-time reaching process. Hence, the component $-\Delta T_s \varepsilon \operatorname{sgn}(s(k))$ is added to avoid the zero-reaching rate and force the traffic density to converge to the equilibrium state within finite time.

Under the sliding mode control, the traffic density $\rho_{n+2}(k+1)$ can be derived by combining the reach law Eq. (2.12) with the sliding surface $s(k+1)$, as

$$\rho_{n+2}(k+1) = \rho_{n+2}^{e} - \frac{1}{c}[s(k) - \Delta T_s \varepsilon \operatorname{sgn}(s(k)) - \Delta T_s q s(k)] \tag{2.13}$$

Using the above Eq. (2.13), the evolution of traffic density $\rho_{n+2}(k+1)$ is derived from the aspect of sliding mode control design. Meanwhile, the dynamics of the density $\rho_{n+2}(k+1)$ can also be obtained using the work zone model Eq. (2.9) driven by the speed limit $u_{n+1}(k)$. Therefore, by substituting Eq. (2.13) into Eq. (2.9), the simplification of speed limit $u_{n+1}(k)$ is derived as:

$$u_{n+1}(k) = g(\rho_{n+2}(k), s(k), \psi(k)) \tag{2.14}$$

The stability of the VSL control system can be analyzed using the following theorem.

Theorem 1: For the discrete-time system of Eq. (2.9), if the sliding surface is designed as Eq. (2.12) with $\varepsilon > 0$, $q > 0$, $2 - \Delta T_s q \gg 0$ *and small sample time interval* ΔT_s, *the discrete-time control law of Eq. (2.14) can guarantee the asymptotical stability of sliding surface Eq. (2.11) such that the system of Eq.(2.9) can reach the equilibrium state and*

achieve desired density.

Proof: The Lyapunov analysis method is utilized to prove theorem 1. The Lyapunov function is constructed as

$$V(k) = s^2(k) \tag{2.15}$$

To achieve the system stability, it should satisfy the following equation:

$$\Delta V(k) = V(k+1) - V(k) = s^2(k+1) - s^2(k) < 0 \tag{2.16}$$

Then the equivalent equation of Eq. (2.16) can be formulated as [54]:

$$\begin{cases} [s(k+1) - s(k)]\,\mathrm{sgn}(s(k)) < 0 \\ [s(k+1) + s(k)]\,\mathrm{sgn}(s(k)) > 0 \end{cases} \tag{2.17}$$

By substituting the sliding surface Eq. (2.12) into Eq. (2.17),

$$\begin{cases} -\Delta T_s \varepsilon - \Delta T_s q \left| s(k) \right| < 0 \\ 2\left| s(k) \right| - \Delta T_s \varepsilon - \Delta T_s q \left| s(k) \right| > 0 \end{cases} \tag{2.18}$$

The Eq. (2.18) is satisfied when the conditions that $\varepsilon > 0$, $q > 0$, $2 - \Delta T_s q \gg 0$ and small sample time interval are met. Thus the asymptotical stability is guaranteed via the Eq. (2.19):

$$\begin{cases} V(k) = s^2(k) \geq 0 \\ \Delta V(k) < 0 \end{cases} \tag{2.19}$$

Though the stability of designed controller is illustrated as Theorem 1, the signum function in Eq. (2.13) can easily lead to the chattering phenomenon due to the hard switching control signal. To eliminate the adverse effects of the chatting on speed limits, the linear saturation function Eq. (2.20) where $c\Delta > 0$ is utilized to replace the signum function such that smooth control can be achieved.

$$Sat(s(k))\begin{cases} 1 & , \quad s(k) > c\Delta \\ s(k)/(c\Delta) & , \quad |s(k)| \le c\Delta \\ -1 & , \quad s(k) < -c\Delta \end{cases} \tag{2.20}$$

In Section 2.4.1, Eq. (2.8) is satisfied because of the stability of the discrete-time sliding mode control. It can be seen that the traffic flow $q_{n+2}(k)$ cannot be restricted by the speed limit at VSL control zone if Eq. (2.8) is not satisfied. Then the density $\rho_{n+2}(k)$ would exceed the critical density ρ_c^b and cause congestion at the work zone area. However, this phenomenon conflicts with the stability of the designed control shown in Theorem 1. Therefore Eq. (2.8) is satisfied.

Apart from the control stability, the appropriate parameters ε, q and c are needed to develop an efficient controller. ε is correlated with the system robustness. Higher value of ε can help to reject the external disturbance. The parameter q has the impact on the reaching rate. The system trajectory reaching the sliding surface can be faster when q is a larger value. Also, a larger c can improve the system transient response. However, too large values of ε, q and c will generate a large control signal and lead to the chattering problem. Hence, the trade-off among the transient response, robustness and overshoot issue should be considered to design these three parameters.

In this study, two sets of parameters are selected. When the traffic is in congested states, too low speed limits should be avoided to prevent the prolonged delays [42]. However, in the transition phase from free flow states to congested states, a quick response is needed as a queue can form fast and easily under high traffic demand [52]. Therefore, the overshoot issue is the main concern under congested states while the

transient response is the primary consideration with uncongested states. To satisfy different levels of demand, two sets of parameters, one for uncongested states and the other for congested states, are designed. Thus the designed control can have different convergence rates to the sliding surface.

2.4.3 Speed Limit Constraints

The following speed limits constraints should be considered to impose the speed limits control.

1) *Speed limit range*: In practical implementation, only discrete speed limits are displayed as it is hard for human drivers to follow continuous speed limits. Therefore, the possible VSLs range belongs to the set that has speed limits from minimum speed limit 16 km/h (10 mi/h) to maximum legitimate limit 113 km/h (70 mi/h) with the increment speed of 8 km/h (5 mi/h).
2) *Discretization*: The continuous speed limit is rounded up to be its closest integral multiple of the increment speed which also satisfies the speed limit range constraint. Meanwhile, the VSL control time interval is selected to be integral multiple of the sample time interval to maintain the system performance and avoid too frequent speed limit change.
3) *Speed limit variation*: To avoid sudden speed changes and reduce potential safety risks, the speed limit difference between two consecutive control time steps satisfies variation threshold v_{Δ} .

2.5 Simulation and Results

2.5.1 Simulation Setup

The proposed VSL control system was evaluated on a 9 km (5.6 mi) southbound segment of Interstate 15, which has the maximum legitimate 113 km/h (70 mi/h) speed limit, in San Bernardino County, California, United States. Construction work zones for freeway barriers were set up along this segment on July 22 and August 2, 2016. Both work zones started at the same position and resulted in one of the three lanes closure. As shown in Fig. 2.5, the blue (thick) line indicates the freeway network while the red (thin) line shows a work zone starts from the position State PM (postmile) 40.2 and extends 4.2 km (2.6 mi) along the I-15S freeway.

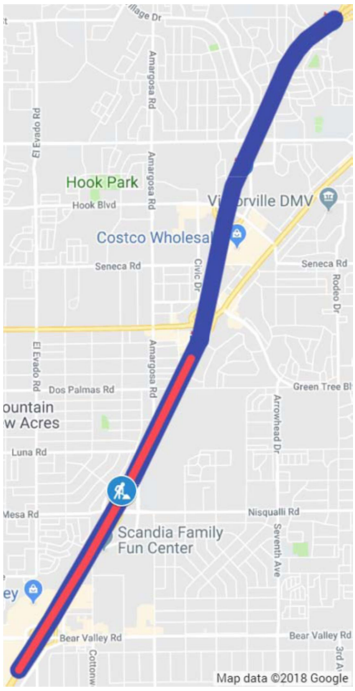


Fig. 2.5. Traffic simulation network with work zone (Map data © 2018 Google)

To evaluate the proposed VSL control, the traffic simulation network with the work zone was calibrated and validated using the real traffic measurement provided by California PeMS (Freeway Performance Measurement System) database [18]. PeMS database archives freeway traffic data such as flow rate, speed, occupancy from detectors over the traffic network in California. In this study, the traffic network in Fig. 2.5 was first built in the traffic simulator SUMO. Then one loop detector located upstream of the work zone and the other at the merge point to the work zone (i.e. the beginning of the work zone) were utilized to calibrate and validate the microscopic model. The work zone traffic data were selected during the congested states between 13:30 and 14:00 on August 2, 2016 to calibrate the car following model and lane changing model. Then the calibrated microscopic model was validated using the traffic data between 12:00 and 12:30 on July 22, 2016. The calibrated and validated results are shown in Table 2.1.

Table 2.1 Comparison Between PeMS Data and Simulation Data

Traffic Condition	August 2, 2016			July 22, 2016		
	PeMS	Simulation	Percentage Error %	PeMS	Simulation	Percentage Error %
Upstream Flow Rate (veh/h)	3082	3001	2.6	2917	3065	5.1
Upstream Speed (km/h)	47.1	46.0	2.3	46.5	44.6	4.1
Work Zone Throughput (veh/h)	2707	2901	7.2	2652	2702	1.9
Merge Speed (km/h)	19.0	16.7	12.1	21.2	18.5	12.7

It can be seen that the errors of calibrated and validated results are less than 10% in the upstream flow rate, upstream speed and work zone throughput. The errors of speeds which are around 20 km/h at the merge point are slightly larger than 10% compared with other measurements. Overall, the calibrated and validated results show that the calibrated microscopic model can provide sufficient accuracy to replicate the realistic work zone traffic conditions.

With the calibrated and validated traffic network, the configuration of the upstream of work zone was designed. For the upstream of work zone, it was divided into n+2 segments including the VSL control zone and acceleration zone. Two types of acceleration zones with the length of 0.5 km (0.3 mi) and 0.8 km (0.5 mi) were designed respectively to assess the impacts of the acceleration zone and robustness of the proposed system. Without the loss of generality, for the other n+1 upstream segments, their lengths were set to be the same as the acceleration zone. Therefore, ten segments and six segments upstream of the work zone were designed for 0.5 km (0.3 mi) and 0.8 km (0.5 mi) acceleration zone respectively. Two speed limit signs were installed at the two ends of the VSL control zone. The loop detectors were installed at the start of freeway, the boundaries of acceleration zone and the end of work zone.

For the calibration, the triangular fundamental diagram was calibrated using the traffic flow, speed, density generated from SUMO simulation. Specifically, vehicles were generated and travelled along a closed circular route. Then traffic density increased gradually to cover the desired range with both free flow states and congested states. The flow, speed and density were aggregated during 30 s time interval. The estimation for a

3-lane freeway with respect to the capacity C, critical density ρ_c, jam density ρ_j, free flow speed v_{free} were 4800 veh/h, 44 veh/km, 270 veh/km and 108 km/h respectively. The details about the calibration are explained in the SUMO documentation [55].

The Krauss car-following model and LC2013 lane-changing model with vehicle parameters were utilized [55]. For the calibrated vehicle parameters, the length of vehicles and the minimum gap were set to be 4.3 m and 4.0 m respectively. The maximum acceleration and deceleration ability of vehicles were set as 2.7 m/s^2 and 4.5 m/s^2 respectively. The factor of the eagerness for performing lane changing to gain speed was set as 50. The speed deviation was set as 0.1 to allow for 10% speed variation of the posted speed limit. The driver imperfection factor was 0.5 to represent the imperfection to adapt to the posted speed limit.

For the sliding mode controller, the parameters $c_1 = 2$, $\varepsilon_1 = 6$, $q_1 = 10$ and $c_2 = 10$, $\varepsilon_2 = 50$, $q_2 = 120$ were selected for uncongested states and congested states respectively. Because one lane is closed at the work zone area, the work zone capacity C^b can be calculated as $2/3\,C$ around 3200 veh/h. The capacity drop factor β was 0.94 with the dropped capacity 3000 veh/h and Δ in the saturation function was 15 veh/km. Ideally, the critical density ρ_c^b at work zone would be 30 veh/km calculated by C^b / v_{free}. However, the average speed inside acceleration zone normally drops by 10 km/h to 15 km/h with the present of speed limit upstream [56]. Therefore the critical density ρ_c^b was calibrated to be 35 veh/km. The speed v_h was calculated as 32 km/h (20 mi/h). The 15 s sample and 30 s control time interval were selected to maintain the controller performance as well as

avoid frequent speed limit changes. The speed limit variation 8 km/h (5 mi/h) was selected as the threshold v_{Δ}.

To evaluate the effectiveness and robustness of the proposed system, different scenarios are considered. Noisy traffic demand and different lengths of acceleration zone are used to demonstrate the robustness of the developed control system. Since KWT based control (kinematic wave theory based control) heavily relies on the deterministic characteristics of traffic flow, KWT based control is selected for comparison. The summary of simulated scenarios is shown in Table 2.2.

Table 2.2 Scenario Summary

Scenario	Acceleration Zone		No-Control	Discrete-time SMC	KWT Based Control
	0.8 km	0.5 km			
1	X		X		
2	X			X	
3	X				X
4		X	X		
5		X		X	
6		X			X

The 0.8 km acceleration zone with no-control, discrete-time SMC (sliding mode control) and KWT based control is evaluated in scenario 1, 2 and 3 respectively. The similar comparisons of the methods are also assessed in scenario 4, 5 and 6 with the 0.5 km acceleration zone configuration. The KWT based control has the same speed limit constraints as the proposed control. Also, the time, when the speed v_h in KWT based control is active and in control, is the same as the time when the restricted speed limits

are active under the proposed control. The restricted speed limits are speed limits lower than the maximum legitimate speed limit. The short length 0.5 km was selected due to the minimum length constraint of the cell transmission model while the long length 0.8 km was selected to be long enough to provide consistent and reliable results.

The simulation was run for 10 replications with different random seeds. As the warm-up period, low traffic demand of 1000 veh/h was generated for 10 minutes. This 10 minute simulation data was discarded. Then noisy demand, which has random fluctuation added on the mean value of traffic flow, was generated for 1 h shown in Fig. 2.6.

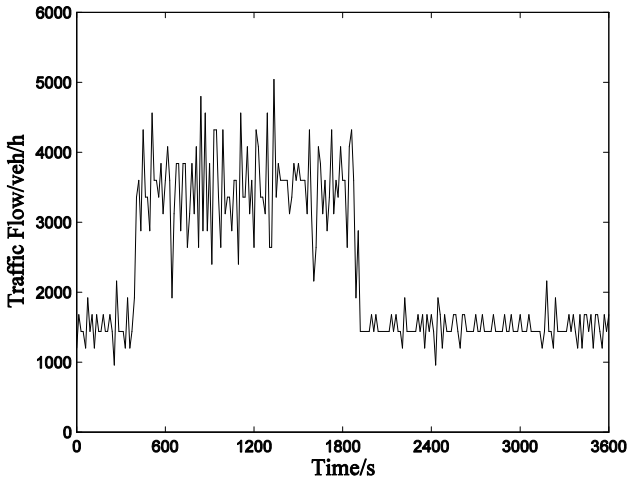


Fig. 2.6. Noisy traffic demand for 1 h

The average traffic demand is 1500 veh/h for the first 300 s then the average high traffic demand of 3600 veh/h is introduced which lasts for 1500 s, followed by the average low demand of 1500 veh/h from the time 1800 s to 3600 s. The noisy traffic demand is utilized for all the scenarios.

2.5.2 Simulation of the 0.8 km Acceleration Zone

For the configuration of 0.8 km acceleration zone, the no-control, discrete-time sliding mode control and KWT based control are implemented respectively.

The work zone throughputs under different control scenarios are illustrated in Fig. 2.7.

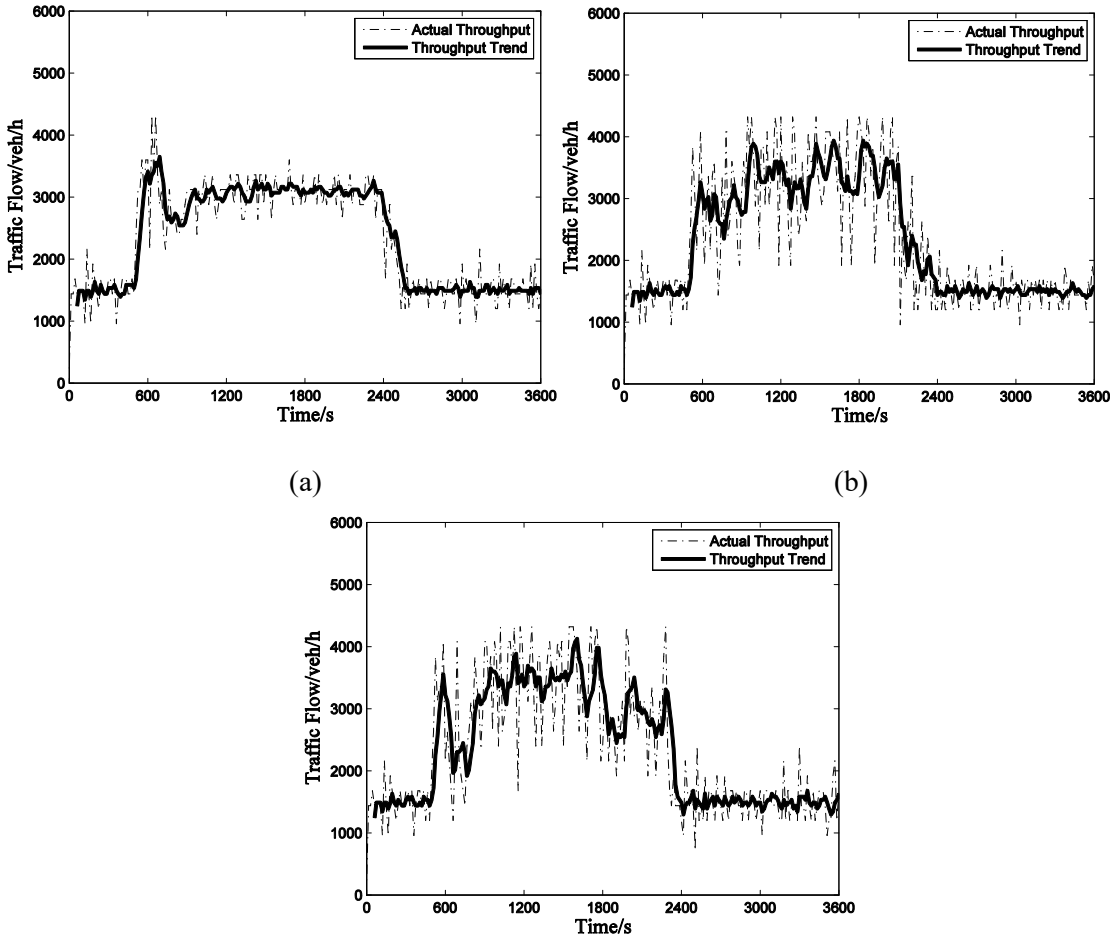


Fig. 2.7. Work zone throughput in (a) scenario 1; (b) scenario 2; and (c) scenario 3

The dashed line shows the actual throughput sampled every 15 s while the solid line shows the throughput trend using the moving average technique. It can be seen from Fig.

2.7(b) that the discrete-time sliding mode control maintains higher traffic throughput around 3200 veh/h during the high demand time and no significant drop of the throughput is observed. During the transition phase 600 s to 900 s when the low demand increases to the high demand, there is a consistent high throughput under the discrete-time sliding model control while the throughputs decrease in the no-control and KWT based control scenarios. Meanwhile, the capacity drop phenomenon occurs in scenario 1. In Fig. 2.7(a), the throughput is stabilized around 3000 veh/h due to the capacity drop from the time 900 s to 2300 s. Under the KWT based control, although the traffic throughput is kept around the work zone capacity between the time 900 s and 1800 s as shown in Fig. 2.7(c), the capacity drop phenomenon still arises at the time 1800 s and lasts for about 9 minutes. When the accumulated queue is fully dissolved around the time 2300 s, the traffic throughput in scenario 3 reduces to around 1500 veh/h with the low traffic demand.

The effectiveness and robustness of the proposed control system are illustrated from Fig. 2.7. The proposed control has the ability to avoid the capacity drop phenomenon. Compared with no-control scenario, the traffic throughput in scenario 2 is maintained around the work zone capacity during the high demand period. Meanwhile, the proposed control is not sensitive to the noisy demand. In both scenario 1 and 3, the traffic throughput drops at the transition phase. Since the noisy demand is highly unstable, particularly at the transition phase, the frequent lane changing behaviour causes this throughput drop phenomenon. Also, the static equilibrium speed limit v_h derived from KWT based control method may still restrict the traffic flow when a relatively low demand appears. However, the similar throughput decrease is not shown under the

proposed control. In contrast, the proposed system can maintain the high throughput and have a good robustness to the noisy demand. Furthermore, the proposed control is robust to the traffic disturbances, particularly at the acceleration zone. Though the traffic throughput maintains around the work zone capacity for about 15 minutes under the KWT based control, the capacity drop still happens and leads to a lower traffic throughput. This capacity drop occurs when a queue forms due to the potential traffic disturbances such as inappropriate accelerations, decelerations. However, the work zone throughput in scenario 2 is kept at the work zone capacity which shows the robustness of

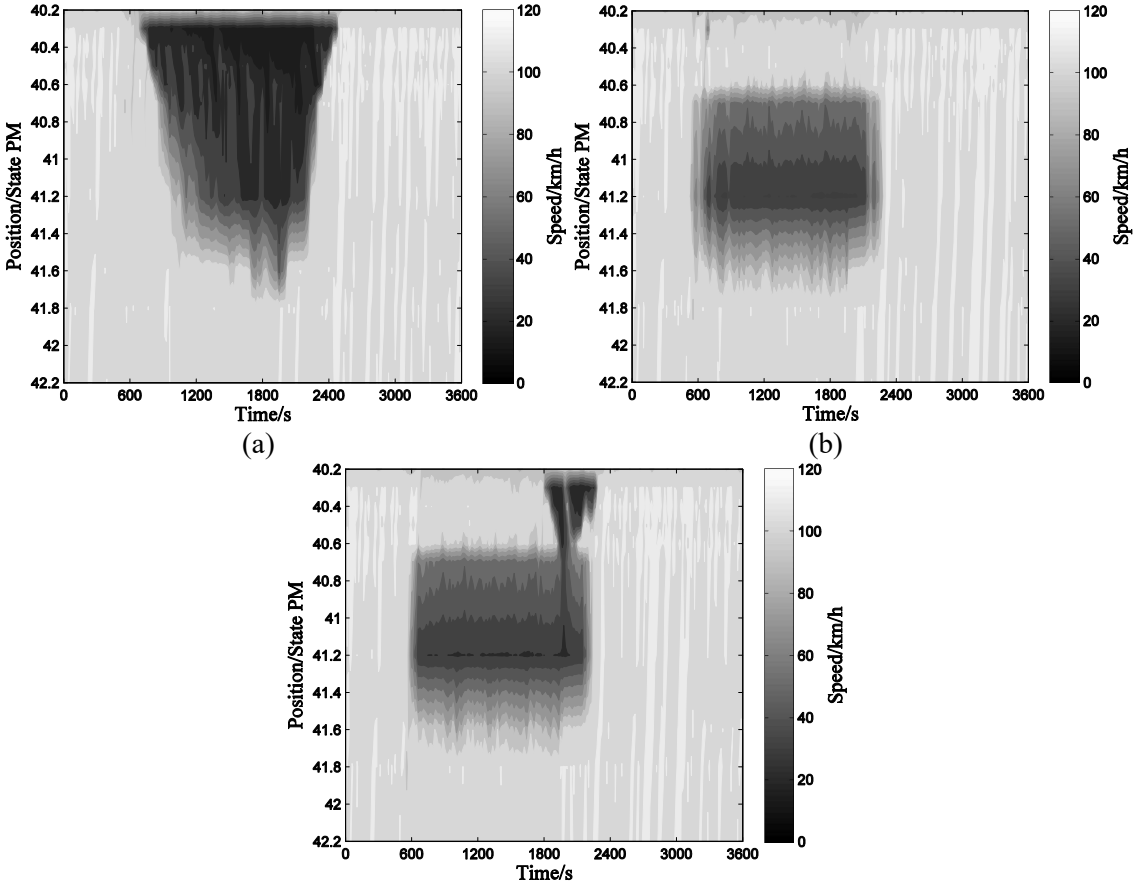
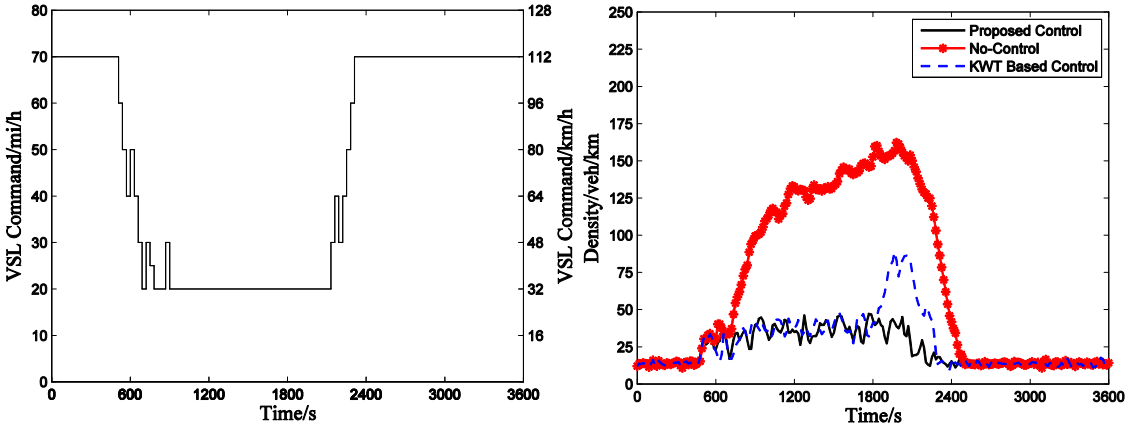


Fig. 2.8. Traffic speed profile in (a) scenario 1; (b) scenario 2; and (c) scenario 3

proposed system to the potential traffic disturbances.

Fig. 2.8 describes the spatio-temporal evolution of the speed 3.2 km (2 mi) upstream of the work zone area. As presented, vehicles travel with much lower speed in scenario 1 than the speeds in scenario 2 and 3. Consequently, more travel delay is involved in the scenario 1. The capacity drop effects under KWT based control discussed in Fig. 2.7(c) are also shown in Fig. 2.8(c). It can be seen that vehicles have slower speed than the speed under discrete-time sliding mode control from the time step 1800 s to 2300 s at the acceleration zone. This slower speed indicates the queue forms at the work zone area and more vehicles join the tail of queue with slow speeds. Compared with the no-control scenario, the proposed control solves the congestion about 5 minutes earlier. Also, under the discrete-time sliding mode control, the length of the freeway affected by the congestion is much shorter than the affected freeway without any control.

The VSL command which is the speed limits generated by the proposed system is shown in Fig. 2.9(a) and the density at the acceleration zone is illustrated in Fig. 2.9(b). As we can see, the VSL command is generated under the speed limit constraints and has restricted speed limits during the high traffic demand period. With respect to the density at the acceleration zone, it is controlled around the 35 veh/km under the high traffic demand using the proposed control. However, in no-control scenario, the density increases rapidly and reaches the maximum density around 160 veh/km. Though the density under KWT based control is similar to the proposed control at first, the high density is still created between the time 1800 s and 2300 s as the queue forms at the work zone area and more vehicles are fed into the acceleration zone.



(a) (b)

Fig. 2.9. (a) VSL command; and (b) density at the acceleration zone

The performance measurements shown in Table 2.3 are also utilized to demonstrate the effectiveness of the proposed system with improvements in traffic mobility, safety and environmental sustainability.

Table 2.3. Performance Measurement for 0.8 km Acceleration Zone

Parameter	No-Control	Discrete-time SMC	Improvement %	KWT Based Control	Improvement %
Average Travel Time T_{total} (min)	2.82	2.38	-15.6	2.63	-6.74
Time-to-collision p (%)	7.81	0.231	-97.0	2.41	-69.1
NOx (kg)	0.614	0.567	-7.65	0.586	-4.56
CO2 (t)	1.51	1.41	-6.62	1.46	-3.31
Fuel Consumption (L)	651	609	-6.45	628	-3.53

The average travel time is computed using the link-based method [57]:

$$T_{total} = \frac{\sum_{j=1}^{N}\sum_{i=1}^{n+3} T_{L_i}^{j}}{N} \tag{2.21}$$

where $T_{L_i}^{j}$ is the average travel time for segment L_i; N is the total number of the sampled travel time, i.e. the travel time is estimated every $3600/N$ time interval.

The probability of time-to-collision (TTC) indicator is utilized as the surrogate safety measure. To calculate the probability, the TTC is first calculated by:

$$TTC_i = \frac{x_{l,i} - x_{f,i}}{v_{f,i} - v_{l,i}}, v_{f,i} > v_{l,i}, x_{l,i} > x_{f,i} \tag{2.22}$$

where $x_{l,i}$ and $v_{l,i}$ are the position and velocity of the leading vehicle at time step *i*; $x_{f,i}$ and $v_{f,i}$ are the position and velocity of the following vehicle at time step *i*.

3 seconds (approximate 100 m with free flow speed) are considered as safe following distance in the freeway [58]. Hence, only when the distance between the leading vehicle and following vehicle is less than 100 m, the TTC will be considered and calculated using Eq. (2.22). The threshold 1.5 second is selected as minimum perception and reaction time to avoid the crash [59]. Therefore, the probability of TTC is computed by:

$$p = \frac{\text{Total Number of } TTC_{crash}}{\text{Total Number of } TTC \text{ Considered}} = \frac{\text{Number of } (TTC_i < 1.5\text{s})}{\text{Total Number of } TTC \text{ Considered}} \tag{2.23}$$

The default emission model based on HBEFA3 model is utilized to calculate the NOx, CO2 emissions and fuel consumption. The calculation mainly considers the acceleration and speed of each vehicle. Details about the HBEFA3 model for the different emission classes and parameter fitting process can be found in [60].

It can be seen that both the discrete-time sliding mode control and KWT based control have improvements from the aspects of traffic mobility, safety and environmental emissions compared with the no-control scenario. However, the proposed control presents more improvements. The largest improvement lies in safety measurement. The probability of time-to-collision is decreased by over 90%. It greatly improves the freeway driving safety. This improvement attributes to the reduction of sudden accelerations and decelerations using the proposed VSL control. Vehicles are able to gradually reduce their speeds and therefore the overall speed variation is reduced. Furthermore, the results show about 15% of saving in the travel time. Because of the avoidance of capacity drop, the work zone throughput is maintained at the bottleneck capacity which improves freeway mobility. Since the travel time is reduced and the throughput is maintained at a high value, fewer vehicles are in congested states and less accelerations and decelerations are generated. This helps to reduce the emissions and fuel consumption by 7% on average and improve environmental sustainability.

2.5.3 Simulation of the 0.5 km Acceleration Zone

The no-control, discrete-time sliding mode control and KWT based control are implemented for the configuration of 0.5 km acceleration zone through scenarios 4, 5 and 6 respectively.

The work zone throughput and spatiotemporal evolution of the speed in these three scenarios are presented in Fig. 2.10 and Fig. 2.11 respectively.

In Fig. 2.10(a), the capacity drop phenomenon occurs and the throughput fluctuates around 3000 veh/h between 1000 s and 2400 s. Under the KWT based control, due to the

noisy demand and the static speed limit, the throughput decreases around the time 800 s. Neither Fig. 2.10(b) nor Fig. 2.10(c) shows apparent capacity drop. Also, the queue is dissipated around the same time step 2300 s in scenario 5 and 6. The similar performance

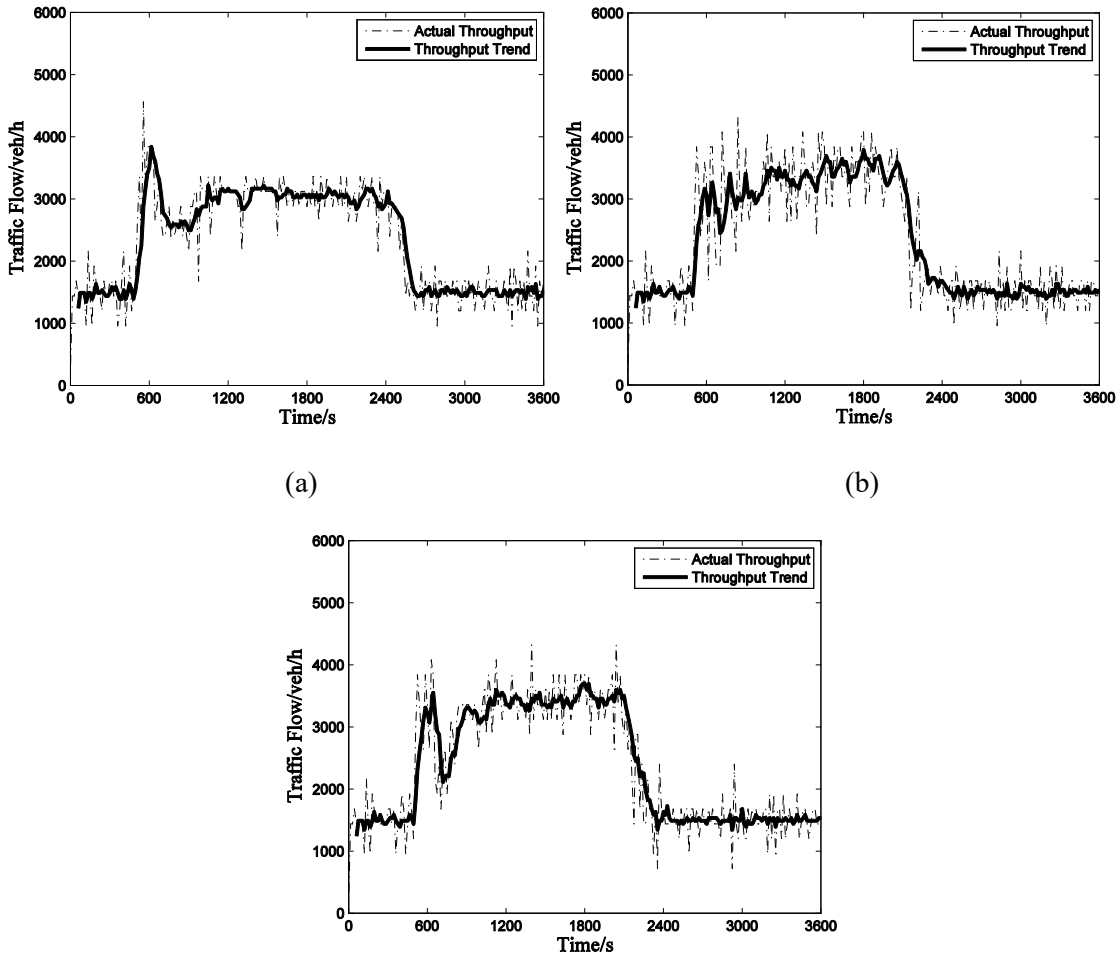


Fig. 2.10. Work zone throughput in (a) scenario 4; (b) scenario 5; and (c) scenario 6

in work zone throughput between the proposed control and KWT based control can be attributed to: (1) 0.5 km is not long enough for vehicles to fully accelerate and the traffic flow rate cannot recover from the VSL control zone. Therefore, the limited length of the

acceleration zone will naturally constrain the traffic flow rate. As a result, a queue will not form and capacity drop phenomena will not occur; (2) intuitively the shorter length of acceleration zone with smaller number of vehicles will have less traffic disturbances.

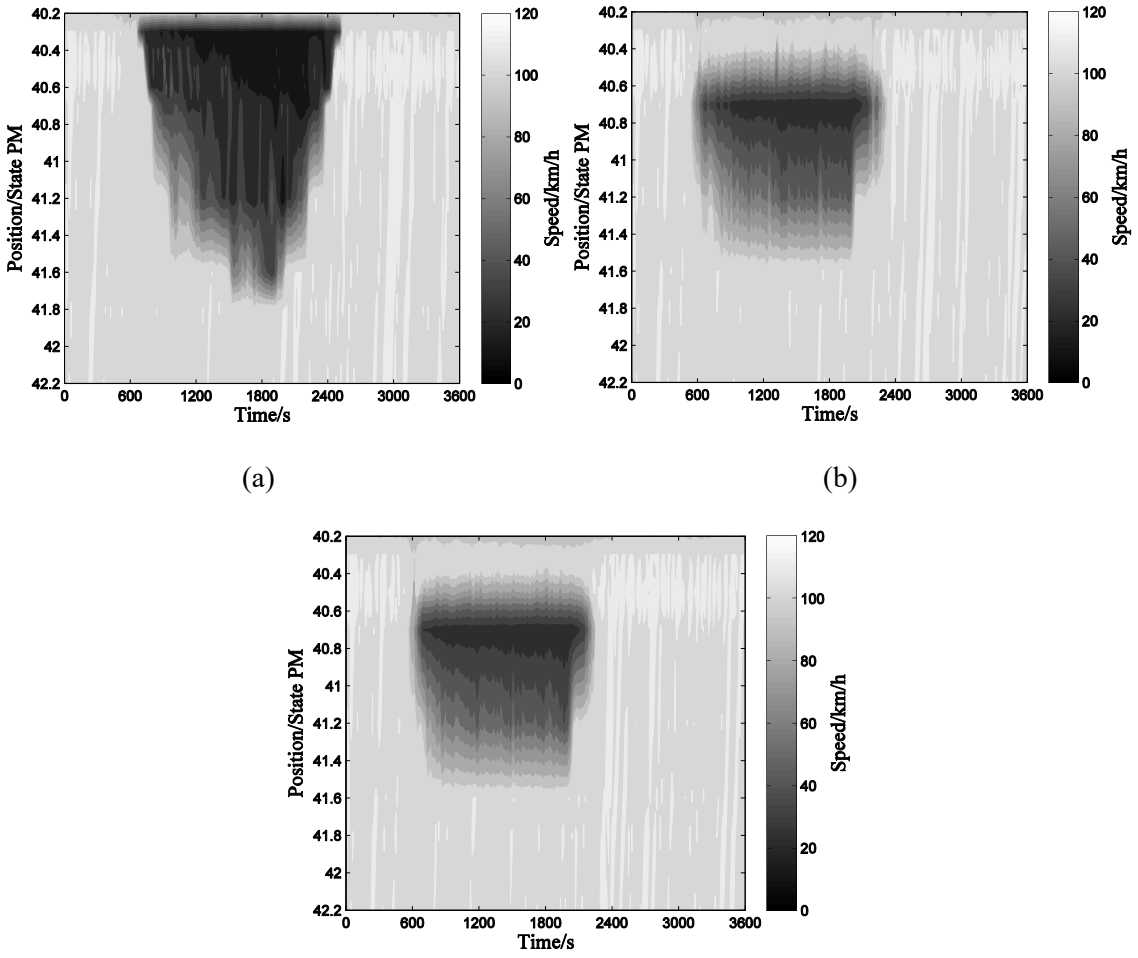
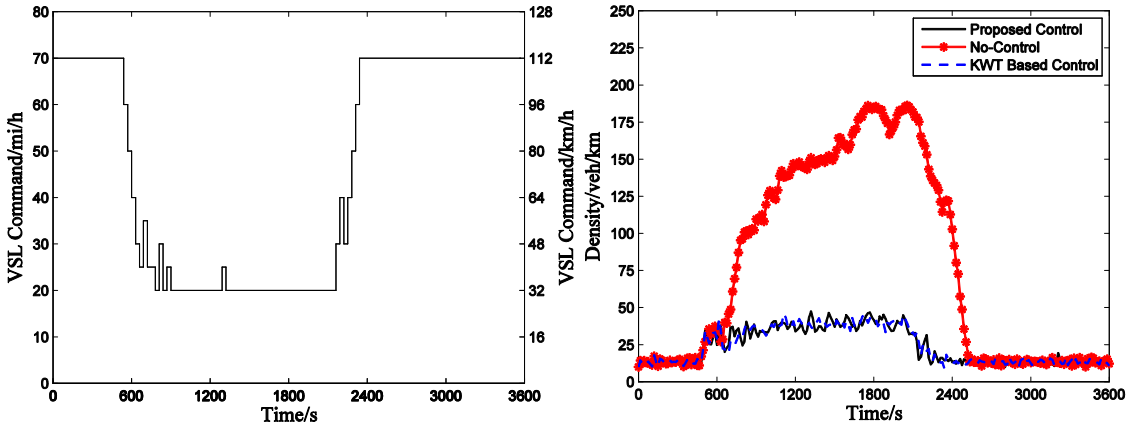


Fig. 2.11. Traffic speed profile in (a) scenario 4; (b) scenario 5; and (c) scenario 6

The similarity of the performance between the proposed control and KWT based control is also shown in Fig. 2.11(b) and Fig. 2.11(c). Both have the similar shape of slow

speed area. For the no-control scenario, Fig. 2.11(a) illustrates a longer queue forms and longer recovery time is required.

Fig. 2.12(a) and Fig. 2.12(b) show the VSL command generated by discrete-time sliding mode control and the density at the acceleration zone respectively.



(a) (b)

Fig. 2.12. (a) VSL command; and (b) density at the acceleration zone

The VSL command follows the speed limit constraints and has the restricted speed limits during the high demand period. Accordingly, densities in the discrete-time sliding mode control and KWT based control are stabilized around 35 veh/km under the high demand. Thus there is no congestion occurring immediate upstream of the work zone area. However, in the no-control scenario, the density exceeds the critical density and increases up to about 180 veh/km. Since the flow exceeds the work zone capacity, the queue forms at the beginning of work zone, followed by the capacity drop phenomenon.

The following performance measurements in Table 2.4 are utilized to evaluate the effectiveness of the proposed control compared with the no-control and KWT based

control scenarios. As presented in Table 2.4, both discrete-time sliding mode control and KWT based control outperform the no-control scenario from the aspects of traffic mobility, safety and environmental sustainability. In addition to the similar work zone throughput, speed profiles and density trends between the proposed control and KWT based control, the similar improvements in Table 2.4 are achieved between scenario 5 and 6. Still, the largest improvement lies in safety followed by improvements of average travel time, environmental emissions and fuel consumption.

Table 2.4. Performance Measurement for 0.5 km Acceleration Zone

Parameter	No-Control	Discrete-time SMC	Improvement %	KWT Based Control	Improvement %
Average Travel Time T_{total} (min)	2.85	2.33	-18.2	2.32	-18.6
Time-to-collision p (%)	7.77	0.625	-92.0	0.781	-89.9
NOx (kg)	0.613	0.581	-5.22	0.580	-5.38
CO2 (t)	1.51	1.44	-4.64	1.44	-4.64
Fuel Consumption (L)	651	620	-4.76	620	-4.76

Through the analysis of the performance for no-control, discrete-time sliding mode control and KWT based control under different freeway layouts, the proposed control system shows the following strengths:

1) The developed discrete-time traffic flow model has the ability to model the capacity drop and provide sufficient accuracy for the designed controller. The nonlinear traffic flow model is able to take the discontinuity and nonlinearity induced by capacity drop into consideration.

2) The proposed discrete-time sliding mode control is not sensitive to different lengths of acceleration zones and has a good robustness to traffic disturbances and the noisy demand. Compared with the KWT based control, the proposed control is able to avoid the capacity drop and show consistent improvements under different lengths of acceleration zones. The traffic flow has inherent stochastic characteristics and potential disturbances. Nevertheless, consistent improvements are achieved in both scenario 2 and 5. Compared with the KWT based control with 0.8 km acceleration zone, the proposed system is robust to the potential disturbance, particularly at the acceleration zone, such that queue formation is prevented and the capacity drop is avoided. Meanwhile, the developed discrete-time sliding mode control can adapt to the fluctuant demand and generate effective speed limits. Consistent improvements of the proposed system show the developed controller is not sensitive to the noisy demand.

3) The proposed system has simple design and efficient calculation. Compared with MPC methods, the prediction of multiple time steps and multiple traffic states are avoided and only one step of the density state at the acceleration zone is calculated and predicted. The sophisticated optimization process is also not needed in the proposed control system and the speed limits scheme is generated only by using the actual detected traffic states. This improves the computation efficiency and makes it easy to implement.

4) The features of discrete time and discontinuous control inherent in discrete-time sliding mode control enable the natural implementation in VSL control.

2.6 Limitations and Future Work

At the current stage, the developed VSL control system is evaluated using the calibrated and validated freeway with work zone area. The factors of work zone configuration such as lateral clearance and presence of activity are not taken into account in this paper. The more detailed work zone configuration will be considered. Also, other control methods such as model predictive control, PI feedback control will be investigated as comparison strategies to assess the proposed control system.

In this study, the speed of vehicles is controlled in one VSL control zone. The VSL strategy for multiple control zones upstream of the bottleneck needs to be studied in the future. Multiple control zones will have longer control areas. Thus the deployment of multiple control zones can potentially improve system performance. The coordination for multiple control zones will be studied as future work. Meanwhile, the identification of optimal speed variation threshold, sample interval and control interval will be integrated into the control design in the future.

Two different lengths of control zone and acceleration zone are considered and assessed in this study. By incorporating the impacts of different lengths on the dynamics of traffic flow, the identification of the optimal setting of the length will be studies to further alleviate the congestion happened around work zone areas.

2.7 Conclusions

In this paper, a VSL control strategy for freeway with the work zone area is proposed using the developed discrete-time nonlinear work zone model and discrete-time sliding

mode control. The work zone model incorporates the capacity drop and impacts of speed limits to provide the sufficient accuracy to the controller. The discrete-time sliding mode control drives the traffic states to achieve the desired equilibrium state with different convergence rates. The inherently discrete time and discontinuous control of the controller makes it simple and efficient to achieve VSL control. The effectiveness of the proposed system is evaluated under a calibrated simulation environment. Compared with the KWT based control, the strong robustness of the designed controller is demonstrated by the performance under noisy traffic demand and different layouts of the freeway. Meanwhile, the proposed system shows the ability to consistently improve traffic mobility, safety and environmental sustainability near work zone areas.

3 Fault-Tolerant Control of Variable Speed Limits for Freeway Work Zone Using Likelihood Estimation

Traffic sensor faults can greatly affect the effectiveness of VSL control. However, most existing studies did not consider sensor faults in the design of VSL control. This chapter presents a fault-tolerant VSL controller for a freeway work zone with permanent sensor faults. Likelihood estimations are used to detect and identify faults of stationary mainline and ramp sensors. An observer-based method is developed to provide accurate traffic state estimations when sensor faults occur by exploiting analytical redundancy. The developed system demonstrates reliable sensor fault diagnosis and consistent improvements of traffic conditions near work zone areas. The following journal article is included in this chapter.

- S. Du and S. Razavi, "Fault-tolerant control of variable speed limits for freeway work zone using likelihood estimation," *Advanced Engineering Informatics*, vol. 45, p.101133, 2020.

The co-author's contributions include:

- Provide supervision and technical advice
- Review and edit the manuscript
- Financially support the research work

3.1 Abstract

Freeway work zones with lane closures can lead to disruption to local traffic and cause significant impacts on mobility, safety, and environmental sustainability. To mitigate traffic congestion near work zone areas, many variable speed limits (VSL) control approaches have been developed. However, VSL control systems, as safety-critical transportation management systems, are prone to the occurrence of traffic sensor faults. Faulty sensors can cause great deviations of traffic measurements and system degradation. Therefore, this study aims to develop a fault-tolerant VSL control strategy for freeway work zones with consideration of mainline and ramp sensor faults. To analyze the traffic dynamics near work zone areas, a traffic flow model is built. Then a sliding mode controller in the previous study is utilized for VSL control. In addition to the traffic states estimated by a Kalman filter, two observers are developed to provide analytical redundancy of traffic state estimations. By comparing the logarithm of the likelihood estimations from the Kalman filter and two observers, a fault diagnosis scheme is designed to detect and identify the faults of mainline and ramp sensors. Then the VSL controller can be reconfigured accordingly in the case of sensor faults. The proposed system is implemented and evaluated under realistic freeway work zone conditions using the traffic simulator SUMO. The results show that the developed system can accurately detect and identify sensor faults in real time. Consistent improvements in mobility, safety and sustainability are also achieved under fault-free and sensor faults scenarios.

3.2 Introduction

Freeway work zones play an important role in maintaining and improving road infrastructure. With the increasing demand for highway expansion and maintenance, a growing number of work zones are set up to conduct road construction activities such as road pavement, resurfacing and rehabilitation. However, traffic congestion can easily form near work zone areas and cause longer travel time, higher safety risk, more emissions and fuel consumption. According to the urban mobility report [4], additional 8.8 billion hours and 3.3 billion gallons of fuel were wasted by urban Americans in 2017 due to traffic congestion. Nearly 10 percent of the overall congestion is caused by work zones [5]. More than 800 work zone fatalities occurred in 2017 [6]. To mitigate the impacts caused by work zones, variable speed limits (VSL) control systems have been widely studied. With the knowledge of traffic states near a work zone, VSL control can impose dynamic speed limits upstream of the work zone so as to restrict or homogenize traffic flow before entering the work zone [33], [34]. Accordingly, traffic conditions near the work zone area can be improved.

Many methods have been proposed in previous studies to design VSL control systems. A rule-based controller was developed to impose dynamic speed limits based on the densities near a traffic bottleneck [61]. The model predictive control was employed to optimize the VSL scheme to improve the bottleneck throughput [13], [15]. VSL control was also proposed to regulate the traffic flow in a network with multiple bottlenecks using a proportional-integral feedback controller [8]. With the development of connected vehicle technology, the effectiveness of dynamic speed limits control was evaluated

under connected environments [62]. Consistent improvements in traffic mobility and safety were shown from the aforementioned studies using different types of controllers and the controller with the presence of connected vehicles. However, the effectiveness of these developed VSL controllers relied on the assumption that reliable traffic state estimations from stationary traffic sensors are available. In practice, stationary traffic sensors such as radar sensors, loop detectors, and cameras are prone to various sensor faults including sensor malfunction and transmission failure. In 2018, more than 30 percent of stationary traffic sensors cannot provide reliable traffic measurements on a typical day in California due to sensor faults [18]. Since a large amount of missing traffic data can easily lead to VSL controller degradation, it is essential to address the adverse effects caused by stationary traffic sensor faults.

To deal with traffic sensor faults, system redundancy is exploited to generate redundant traffic state estimations to replace faulty measurements in the case of sensor faults. Two types of system redundancies, physical redundancy and analytical redundancy, are mainly adopted [63]. Physical redundancy requires an extra set of traffic sensors to be installed, which becomes impractical to implement in a large scale system like freeway network due to hardly affordable cost. In contrast, cost-effective analytical redundancy generates redundant traffic state estimation from a mathematical system model. Therefore, analytical redundancy is mainly introduced and discussed in this study.

A variety of approaches have taken advantage of the spatiotemporal feature of traffic flow to produce analytical redundancy and recover missing traffic data. A Bayesian tensor decomposition model was proposed to learn the spatiotemporal patterns in traffic

flow to impute the missing traffic data [21]. Two cokriging methods were developed to study the spatiotemporal dependency of traffic data in order to recover corrupted traffic data [64]. A multi-view learning model was trained to estimate the missing data with good accuracy from a time series traffic dataset [65]. However, these learning models exploited the temporal information from measurements after the missing data point to achieve accurate traffic state estimation. The data recovering process was conducted offline. Since VSL control systems demand real-time traffic state estimation and online data imputation when a sensor fault occurs, it may be difficult to obtain accurate estimations with offline learning models and maintain the performance of VSL control systems. To address the demand of online imputation for traffic control systems, many model-based approaches have been developed. A calibrated regression model was developed in [66] to predict the missing traffic data. In [67], a probabilistic principal component analysis was utilized to develop an imputation optimization algorithm to provide accurate state estimation with missing data. Multiple linear regression models were proposed to detect the missing data samples and impute them accordingly [23]. Nevertheless, these model-based approaches needed historical traffic measurement data near the faulty sensors area to calibrate the imputation models. However, with respect to freeway work zones, particularly short-term work zones, historical traffic data may not be available. The lack of historical traffic data makes it difficult to calibrate the imputation models so as to provide accurate traffic state estimation for a VSL controller. Therefore, a sensor fault diagnosis scheme, which can achieve online traffic data imputation without the demand of historical traffic measurement data, is needed for a VSL controller.

Likelihood estimation based methods have been mainly studied to perform online sensor fault diagnosis in engine control area. A likelihood ratio based method was proposed to detect the bleed valve fault in gas turbine engine [68]. The residuals generated by two virtual sensors were analyzed using maximum-likelihood voting algorithm to detect the faulty speed sensor of electric vehicle powertrains [69]. The likelihood estimation was used to assess the sensor fault severity of in railway traction drive [70]. However, these likelihood estimation methods developed in engine control applications can hardly be applied to sensor fault diagnosis for VSL controllers due to significant differences of system models. In addition, the likelihood estimation methods in [68], [69] have difficulties in addressing faults of multiple sensors like mainline sensors and ramp sensors considered in this study.

To address the impacts of sensor faults, this paper aims to develop a fault-tolerant VSL control system for freeway work zones using the likelihood estimation. Three main contributions are made including:

1) Consistently maintain the performance of VSL controller with the presence of sensor faults. The VSL controller can be reconfigured according to different types of sensor faults. The system performance is evaluated from the aspects of mobility, safety and sustainability.

2) Reliably detect the stationary traffic sensor faults without the prerequisite of historical traffic data. The designed fault diagnosis scheme can detect sensor faults in real time when a fault occurs.

3) Accurately identify the faults of stationary traffic mainline sensors and ramp

sensors. Faults of different sensors can be identified via the comparison of the logarithm of the likelihood estimations generated from the Kalman filter and two observers.

The rest of this paper is organized as follows. The system framework is first presented. Then the traffic flow model with augmented traffic states is introduced. After the discussion of the sliding mode controller for VSL control, the design of traffic state estimation based on Kalman filter is presented. Two observers are developed for stationary traffic mainline and ramp sensors. Then the fault diagnosis scheme is designed. Afterwards, the effectiveness of the developed system is evaluated with analytical results. Finally, conclusions and future work are discussed.

3.3 Methodology

The framework of the fault-tolerant VSL control system for freeway work zones is shown in Figure 3.1.

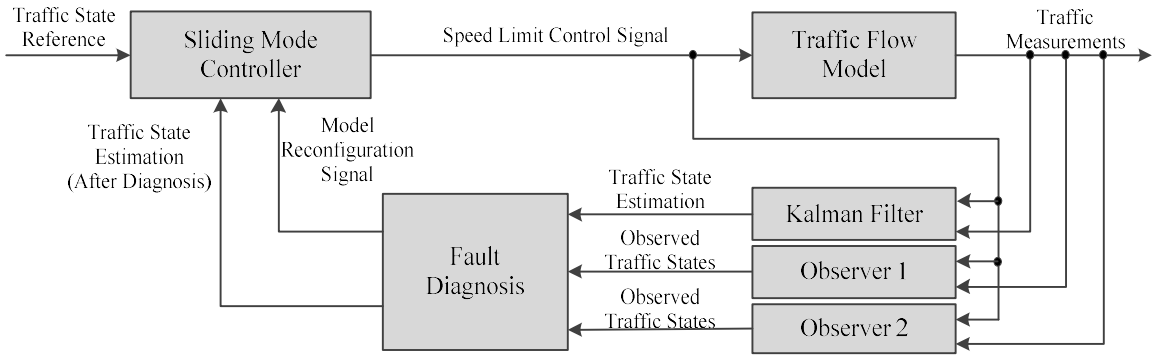


Fig. 3.1. System framework

The sliding mode controller, which is developed in the previous study of the authors

[71], can read the traffic state estimation from the fault diagnosis component to derive the speed limit control signal so as to track the traffic state reference. Then the speed limit signal can affect the evolution of the traffic flow model to improve traffic conditions near work zone areas. To consider the capacity drop phenomenon around work zone areas, the capacity drop model in [3], [14] is incorporated in the traffic flow model. Both the control signal and traffic measurements are fed to the Kalman filter, observer 1 and observer 2 to obtain traffic state estimation and observed traffic states respectively. These two observers are designed to provide redundant traffic states in the case of stationary traffic sensor faults. After the process of fault diagnosis, the reliable traffic state estimation is sent to the controller and reconfiguration signal is generated as well when a sensor fault occurs. More details of each component in the system framework can be found in the following sections.

3.3.1 Traffic Flow Model

The layout of a freeway work zone with a lane closure is presented in Figure 3.2.

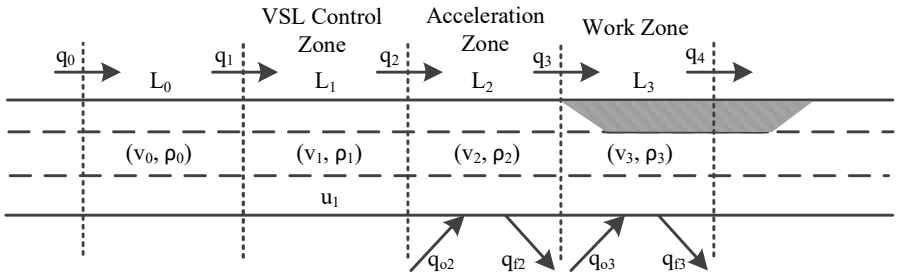


Fig. 3.2. Layout of freeway work zone area (not to scale)

The freeway network near the work zone area is partitioned into multiple segments: the VSL control zone, acceleration zone and work zone. The length of each segment *i* is expressed as L_i. The traffic inflow q_i, outflow q_{i+1}, average speed v_i and density ρ_i are denoted for each segment *i*. The design of the fault-tolerant VSL controller can be affected by ramp flows at the acceleration zone and work zone. Therefore, the on-ramp flow q_{o2}, off-ramp flow q_{f2}, on-ramp flow q_{o3} and off-ramp flow q_{f3} are considered in Figure 3.2. It should be noted that in practice there may be no on-ramps or off-ramps at some freeway segments.

To detect the traffic states near work zone areas, probe sensors and stationary sensors are adopted in this study. Specifically, connected vehicles with the ability to transmit their speeds and locations are utilized as the probe sensors. These connected vehicles are assumed to have no autonomous driving capability and share the same driver behavior with conventional vehicles. With the moving of connected vehicles which are mixed with the conventional vehicles, the probe sensors can detect the traffic states at different segments during the traveling while the stationary sensors can only monitor the traffic states at a fix location or segment. In this study, two stationary mainline sensors are installed within the acceleration zone and work zone respectively. Also, stationary ramp sensors are installed to monitor the ramp flows at the acceleration zone and work zone.

Since one lane is closed due to the presence of the work zone in Figure 3.2, traffic congestion can easily happen near work zone areas. We assume no traffic congestion occurs downstream of the work zone. Two speed limit signs are installed to improve traffic conditions. One is installed at the beginning of VSL control zone to control the

speed of vehicles inside the VSL control zone. The other is installed at the beginning of the acceleration zone to allow vehicles to accelerate to free flow speed to travel through the work zone.

The evolution of traffic density can be derived using the conservation law [50] as:

$$\rho_i(k+1) = \rho_i(k) + \frac{\Delta T_s}{L_i}[q_i(k) - q_{i+1}(k) + q_{ri}(k)], i = 2,3 \tag{3.1}$$

where *k* stands for the *k*th time step. ΔT_s is the sample time interval and $q_{ri}(k)$,which is calculated as Eq. (3.2), represents the ramp flows at segment *i*.

$$q_{ri}(k) = q_{oi}(k) - q_{fi}(k) \tag{3.2}$$

The density $\rho_2(k)$ at the acceleration zone is mainly considered because the objective of the VSL controller is to stabilize the density $\rho_2(k)$ at the critical density [71]. Also, the density $\rho_3(k)$ needs to be obtained to provide redundant traffic states which will be discussed in the Section 3.3.4. The traffic flow q_2 , q_3 and q_4 in Eq. (3.1) can be derived under the assumption of triangular fundamental diagram [51] as:

$$\begin{cases} q_2(k) = \min\{v_1(k)\rho_1(k), \omega\rho_j u_1(k)/[u_1(k)+\omega], \omega[\rho_j - \rho_2(k)]\} \\ q_3(k) = \min\{v_2(k)\rho_2(k), a\rho_2(k) + b\} \\ q_4(k) = v_3(k)\rho_3(k) \end{cases} \tag{3.3}$$

where ω, ρ_j and $u_1(k)$ are the backward wave speed, jam density and speed limit control signal respectively. The constant parameter *a* and *b* are obtained from the capacity drop model in [3], [14]. The critical density, jam density and maximum dropped capacity of the freeway network are used to calculate the parameter *a* and *b*. Capacity drop occurs when a queue forms upstream of a bottleneck [28], [30], [72]. In Fig. 3.2, the

work zone capacity C^b ideally equals to 2/3 the full road capacity C due to one closed lane. However, when a queue forms upstream of the work zone, the work zone capacity C^b can be further decreased because of the capacity drop phenomenon. Therefore, the constant parameter *a* and *b* are introduced to consider the dropped work zone capacity.

It can be seen that the speed limit $u_1(k)$ can affect the evolution of traffic dynamics using the traffic flow $q_2^u(k) = \omega\rho_j u_1(k)/[u_1(k)+\omega]$. When traffic demand is lower than the work zone capacity, $v_1(k)\rho_1(k)$ is less than $q_2^u(k)$ and VSL control is not needed to restrict the traffic flow because vehicles can travel through the work zone area without congestion. On the other hand, when the traffic demand is higher than the work zone capacity, the flow $q_2^u(k) < \omega[\rho_j - \rho_2(k)]$ is constrained by the stability of the sliding mode controller which is discussed in the following section. Intuitively, it can be seen that when $q_2^u(k) > \omega[\rho_j - \rho_2(k)]$, the speed limits cannot restrict the traffic flow and traffic congestion is unavoidable at the acceleration zone.

Since the ramp flow $q_{r2}(k)$ can also affect the density $\rho_2(k)$ which is the control objective, augmented traffic states as $\boldsymbol{x}(k) = [\rho_2(k) \quad \rho_3(k) \quad q_{r2}(k)]^{\mathrm{T}}$ is established. Thus, the evolution of the augmented traffic states as the traffic flow model can be derived as:

$$\boldsymbol{x}(k+1) = \boldsymbol{A}(k)\boldsymbol{x}(k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{\varepsilon}(k) \tag{3.4}$$

where $\boldsymbol{u}(k)$ is the system input and $\boldsymbol{\varepsilon}(k)$ is assumed to be white process noise with the covariance $\boldsymbol{Q}(k)$. $\boldsymbol{A}(k)$ and $\boldsymbol{B}(k)$ are system matrices. Specifically, the traffic flow model

Eq. (3.4) is obtained under without capacity drop and with capacity drop scenarios as:

$$\begin{bmatrix} \rho_2(k+1) \\ \rho_3(k+1) \\ q_{r2}(k+1) \end{bmatrix} = \begin{bmatrix} 1-\frac{\Delta T_s}{L_2}v_2(k) & 0 & \frac{\Delta T_s}{L_2} \\ \frac{\Delta T_s}{L_3}v_2(k) & 1-\frac{\Delta T_s}{L_3}v_3(k) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho_2(k) \\ \rho_3(k) \\ q_{r2}(k) \end{bmatrix} + \begin{bmatrix} \frac{\Delta T_s}{L_2}q_2^u(k) \\ \frac{\Delta T_s}{L_3}q_{r3}(k) \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1(k) \\ \varepsilon_2(k) \\ \varepsilon_3(k) \end{bmatrix} \quad (3.5)$$

$$\begin{bmatrix} \rho_2(k+1) \\ \rho_3(k+1) \\ q_{r2}(k+1) \end{bmatrix} = \begin{bmatrix} 1-\frac{\Delta T_s}{L_2}a & 0 & \frac{\Delta T_s}{L_2} \\ \frac{\Delta T_s}{L_3}a & 1-\frac{\Delta T_s}{L_3}v_3(k) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \rho_2(k) \\ \rho_3(k) \\ q_{r2}(k) \end{bmatrix} + \begin{bmatrix} \frac{\Delta T_s}{L_2}[q_2^u(k)-b] \\ \frac{\Delta T_s}{L_3}[b+q_{r3}(k)] \\ 0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1(k) \\ \varepsilon_2(k) \\ \varepsilon_3(k) \end{bmatrix} \quad (3.6)$$

where a random walk equation of the ramp flow $q_{r2}(k+1) = q_{r2}(k) + \varepsilon_3(k)$ is treated as an augmented traffic state. $\varepsilon_3(k)$ is selected to accommodate the ramp flow variation. The average speed $v_2(k)$ and $v_3(k)$ in Eq. (3.5) and (3.6) are detected by connected vehicles at the acceleration zone and work zone [73]. Thus the traffic flow model considering the evolution of the augmented traffic states is established.

3.3.2 Sliding Mode Controller

Sliding mode control is a variable structure control method for nonlinear systems [74]. With multiple control structures, a sliding mode controller can generate switching control signals to change the trajectory of a system. To drive the system state to the desired equilibrium state, a sliding surface is designed such that the trajectory of the system can converge and stay confined to the sliding surface. The main feature of sliding mode control is its strong robustness to system uncertainties and external disturbances. Because

of the nonlinearity caused by capacity drop phenomena and traffic disturbances near work zone areas [37], sliding mode control is utilized to design the VSL controller.

The objective of the VSL controller is to stabilize the traffic density $\rho_2(k)$ at the work zone critical density ρ_c^b so that the capacity drop phenomenon can be avoided. The VSL controller based on the sliding mode control method in [71] is employed in this study. To develop the VSL controller, a sliding surface is first designed as:

$$s(k) = c[\rho_c^b - \rho_2(k)] \tag{3.7}$$

where the nonzero constant parameter c is selected to have a proper convergence rate. Then a reaching law, which can drive the states to the designed surface and make the states confined on the surface, is designed as [53]:

$$s(k+1) = s(k) - \Delta T_s \eta \,\mathrm{sgn}(s(k)) - \Delta T_s q s(k) \tag{3.8}$$

where η and q are positive constant parameters and sgn represents a signum function. Thanks to the $-\Delta T_s q s(k)$ part, the trajectory of system states is forced to move to the sliding surface Eq. (3.7) with an exponential rate, while the switching signal $-\Delta T_s \eta \,\mathrm{sgn}(s(k))$ can help to avoid the zero reaching rate and make the states constrained to the sliding surface.

By combining the Eq. (3.7) and (3.8), the evolution of traffic density $\rho_2(k)$ can be derived from the perspective of sliding mode control. Meanwhile, the evolution of traffic density $\rho_2(k)$ is also analyzed in Eq. (3.4) from the perspective of traffic flow model. The comparison of density $\rho_2(k)$ from these two sources can be used to derive the speed

limit control signal as a function *g*:

$$u_1(k) = g(\rho_2(k), v_2(k), s(k)) \tag{3.9}$$

which is correlated with the traffic density $\rho_2(k)$, average speed $v_2(k)$ at the acceleration zone, and the difference between $\rho_2(k)$ and the work zone critical density ρ_c^b.

To guarantee the stability of the sliding mode controller, the condition that $\eta > 0$, $q > 0$ and $2 - \Delta T_s q \gg 0$ should be satisfied as proved in the previous study [71].

The practical speed limit constraints should also be considered: 1) the speed limits should fall in the range of the minimum speed limit $v_{\min}$ and the maximum speed limit $v_{\max}$ with the incremental speed Δv; 2) the continuous speed limit $u_1(k)$ is rounded up to the closest integer speed limit which is an integral multiple of the incremental speed Δv; 3) the maximum speed limit difference between two consecutive control time steps is constrained by $\Delta v_{\max}$;4) the speed limit control time interval ΔT_c should be an integral multiple of the sample time interval ΔT_s to avoid frequent change of speed limits.

3.3.3 Kalman Filter for Traffic State Estimation

Kalman filters are linear state estimators for stochastic systems. By using noisy sensor measurements in a stochastic system, the Kalman filter can recursively estimate system states. The recursive filter process consists of the time update and measurement update [75]. In the time update step, the Kalman filter can predict the current system state. When a new measurement is available, the state prediction can be updated using a weighted average in the measurement update step.

To estimate the traffic state $\boldsymbol{x}(k)$ in Eq. (3.4), the measurement equation should be derived. In this study, two stationary mainline sensors are installed to detect the traffic outflow $q_3(k)$ and $q_4(k)$ at the acceleration zone and work zone respectively. By utilizing the detected average speed $v_2(k)$ and $v_3(k)$ at the acceleration zone and work zone, the traffic density $\rho_i(k)$ can be measured using $\rho_i(k) = q_{i+1}(k) / v_i(k)$. Since the ramp flow is detected by the installed stationary ramp sensors, the measurement equation is derived as:

$$\boldsymbol{y}^{(0)}(k) = \boldsymbol{C}^{(0)}(k)\boldsymbol{x}^{(0)}(k) + \boldsymbol{\zeta}^{(0)}(k) \tag{3.10}$$

where $\boldsymbol{y}^{(0)}(k) = [\rho_2(k) \quad \rho_3(k) \quad q_{r2}(k)]^{\mathrm{T}}$ and $\boldsymbol{C}^{(0)}(k) = \boldsymbol{I}_{3*3}$. $\boldsymbol{\zeta}^{(0)}(k)$ is the measurement noise assumed to be Gaussian noise of zero mean with covariance $\boldsymbol{R}^{(0)}(k)$.

By combining Eq. (3.4) with Equation (3.10), the Kalman filter for traffic state estimation can be obtained as:

$$\begin{cases} \hat{\boldsymbol{x}}^{(0)}(k+1 \mid k+1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}^{(0)}(k \mid k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{K}^{(0)}(k+1)\overline{\boldsymbol{y}}^{(0)}(k+1) \\ \boldsymbol{y}^{(0)}(k+1) = \boldsymbol{C}^{(0)}(k+1)\boldsymbol{x}^{(0)}(k+1 \mid k+1) + \boldsymbol{\zeta}^{(0)}(k+1) \end{cases} \tag{3.11}$$

where $\boldsymbol{K}^{(0)}(k+1)$ is the gain matrix and $\overline{\boldsymbol{y}}^{(0)}(k+1)$ is the innovation vector calculated in Eq. (3.12).

$$\overline{\boldsymbol{y}}^{(0)}(k+1) = \boldsymbol{y}^{(0)}(k+1) - \boldsymbol{C}^{(0)}(k+1)[\boldsymbol{A}(k)\hat{\boldsymbol{x}}^{(0)}(k \mid k) + \boldsymbol{B}(k)\boldsymbol{u}(k)] \tag{3.12}$$

Moreover, the innovation covariance matrix $\boldsymbol{S}^{(0)}(k+1)$ is derived as

$$\boldsymbol{S}^{(0)}(k+1) = \boldsymbol{C}^{(0)}(k+1)[\boldsymbol{A}(k)\boldsymbol{P}^{(0)}(k \mid k)\boldsymbol{A}^{\mathrm{T}}(k) + \boldsymbol{Q}(k)]\boldsymbol{C}^{(0)^{\mathrm{T}}}(k+1) + \boldsymbol{R}^{(0)}(k) \tag{3.13}$$

where $\boldsymbol{P}^{(0)}(k \mid k)$ is the state covariance matrix.

With the initial traffic states $\hat{\boldsymbol{x}}^{(0)}(k \mid k)$ and state covariance matrix $\boldsymbol{P}^{(0)}(k \mid k)$, the gain matrix $\boldsymbol{K}^{(0)}(k+1)$ can be updated accordingly. When a new measurement is available, the innovation vector is calculated to achieve the measurement update and the traffic states $\hat{\boldsymbol{x}}^{(0)}(k+1 \mid k+1)$ estimation at the next time step. This recursive process is performed to estimate the traffic states given a new measurement.

3.3.4 Observer

In a control system, an observer can provide the internal system state which can be used to develop the system controller. In the cases where the system state cannot be determined directly from the sensor measurement, an observer is needed. By using the sensor measurement, an observer can estimate the system state based on the evolution of the dynamic system. In this study, observers are designed to provide analytical redundancy for VSL control systems. It can be seen that in Eq. (3.9), the control strategy relies on the accurate traffic estimation of $\rho_2(k)$ to generate the reasonable control signal $u_1(k)$. However, the accuracy of the density estimation $\rho_2(k)$ can be greatly affected if the measurement Eq. (3.10) fails to provide the reliable mainline traffic measurement $q_3(k)$ and the ramp flow measurement $q_{r2}(k)$. Therefore, it is critical to provide redundant state estimation of $\rho_2(k)$ and $q_{r2}(k)$ to avoid VSL control failures and ensure the system performance even with sensor faults.

Two observers are designed for the state estimation of traffic density $\rho_2(k)$ and ramp flow $q_{r2}(k)$ respectively. To avoid using faulty sensors, the observer below is designed in

the case of mainline sensor faults:

$$\begin{cases} \hat{\boldsymbol{x}}^{(1)}(k+1\,|\,k+1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}^{(1)}(k\,|\,k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{K}^{(1)}(k+1)\overline{\boldsymbol{y}}^{(1)}(k+1) \\ \boldsymbol{y}^{(1)}(k+1) = \boldsymbol{C}^{(1)}(k+1)\boldsymbol{x}^{(1)}(k+1\,|\,k+1) + \boldsymbol{\zeta}^{(1)}(k+1) \end{cases} \tag{3.14}$$

where $\boldsymbol{y}^{(1)}(k) = [\rho_3(k) \quad q_{r2}(k)]^{\mathrm{T}}$ and $\boldsymbol{C}^{(1)}(k) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. $\boldsymbol{\zeta}^{(1)}(k) = [\zeta_2 \quad \zeta_3]^{\mathrm{T}}$ is the measurement noise assumed to be Gaussian noise of zero mean with covariance $\boldsymbol{R}^{(1)}(k)$.

It can be seen that Eq. (3.14) does not rely on the measurement $q_3(k)$ to estimate the traffic states. Therefore, if the stationary mainline sensor fault occurs at the acceleration zone, the observer Eq. (3.14) can still provide reliable state estimations.

Similarly, in the case of the ramp flow sensor fault, the observer is designed with the similar structure

$$\begin{cases} \hat{\boldsymbol{x}}^{(2)}(k+1\,|\,k+1) = \boldsymbol{A}(k)\hat{\boldsymbol{x}}^{(2)}(k\,|\,k) + \boldsymbol{B}(k)\boldsymbol{u}(k) + \boldsymbol{K}^{(2)}(k+1)\overline{\boldsymbol{y}}^{(2)}(k+1) \\ \boldsymbol{y}^{(2)}(k+1) = \boldsymbol{C}^{(2)}(k+1)\boldsymbol{x}^{(2)}(k+1\,|\,k+1) + \boldsymbol{\zeta}^{(2)}(k+1) \end{cases} \tag{3.15}$$

but with $\boldsymbol{y}^{(2)}(k) = [\rho_2(k) \quad \rho_3(k)]^{\mathrm{T}}$, $\boldsymbol{C}^{(2)}(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $\boldsymbol{\zeta}^{(2)}(k) = [\zeta_1 \quad \zeta_2]^{\mathrm{T}}$ and $\boldsymbol{R}^{(2)}(k)$.

Then the recursive process of traffic state estimation in the Section 3.3.3 can be applied to these two observers shown in Eq. (3.14) and Eq. (3.15).

To make sure the observers can work properly, the observability of these two observers is presented. Since the observability matrix can be calculated using Eq. (3.16):

$$\boldsymbol{O}(k, k+3) = \begin{bmatrix} \boldsymbol{C}^{(i)} \\ \boldsymbol{C}^{(i)}\boldsymbol{A}(k) \\ \boldsymbol{C}^{(i)}\boldsymbol{A}(k+1)\boldsymbol{A}(k) \end{bmatrix} \tag{3.16}$$

the observability matrices for Eq. (3.4) under with capacity drop and without capacity

drop scenarios are derived as follows:

$$rank\left[\boldsymbol{O}_{no-drop}\right] = rank\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{\Delta T_s}{L_2} v_2(k) & 1-\frac{\Delta T_s}{L_3} v_3(k) & 0 \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \tag{3.17}$$

$$rank\left[\boldsymbol{O}_{drop}\right] = rank\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{\Delta T_s}{L_2} a & 1-\frac{\Delta T_s}{L_3} v_3(k) & 0 \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \tag{3.18}$$

Because *a* is a nonzero constant parameter and the average speed $v_2(k)$ is bounded to positive speed during the experiment, the rank of both observability matrices equals to 3. Therefore, the observer Eq. (3.14) is observable. Similarly, the observer Eq. (3.15) can be proved to be observable under with capacity drop and without capacity drop scenarios.

3.3.5 Fault Diagnosis

The fault diagnosis scheme can diagnose sensor faults and reconfigure the VSL controller when a sensor fault occurs. The probability based method using the likelihood estimation is developed to detect and identify the faults of stationary mainline and ramp sensors. Likelihood estimation is utilized to calculate the probability of the sensor measurement conditioned on a system mode. In this study, the system modes include

health sensor mode and sensor fault modes. To determine the system mode, likelihood estimation can measure the plausibility provided by the sensor measurement for each possible system mode. Detecting zero flow rates with the presence of vehicles is the leading sensor fault among loop detectors [76]. This type of fault can be caused by the failure of sensor control units, communication lines, etc. Therefore, the stationary sensor faults of zero flow rates are mainly considered. In this study, the concurrent sensor faults and probe sensors faults are not considered.

When a new measurement is available, the measurement likelihood is estimated by the Kalman filter and two observers. The likelihood can be estimated as:

$$\begin{aligned} & p(\boldsymbol{y}^{(j)}(k+1)) = N(\bar{\boldsymbol{y}}^{(j)}(k+1); \mathbf{0}, \boldsymbol{S}^{(j)}(k+1)) \\ & = \frac{1}{\sqrt{(2\pi)^n \left|\boldsymbol{S}^{(j)}(k+1)\right|}} \exp[-\frac{1}{2}(\bar{\boldsymbol{y}}^{(j)}(k+1))^{\mathrm{T}}(\boldsymbol{S}^{(j)}(k+1))^{-1}\bar{\boldsymbol{y}}^{(j)}(k+1)] \end{aligned} \tag{3.19}$$

where j=0,1,2. By using Eq. (3.19), the Kalman filter and two observers can assess the likelihood of a new measurement. When low anomaly likelihood $p(\boldsymbol{y}^{(0)}(k+1))$ is calculated from the Kalman filter, there is a high probability that a sensor fault occurs. To detect if there is a sensor fault, a threshold λ_1 is defined and utilized in Eq. (3.20).

$$-\log(p(\boldsymbol{y}^{(0)}(k+1))) < \lambda_1 \tag{3.20}$$

If Eq. (3.20) is satisfied, then there is no sensor fault. However, when the Eq. (3.20) is violated, a sensor fault is detected. The negative logarithm of the likelihood is utilized in Eq. (3.20) instead of the likelihood to better illustrate the likelihood estimation in the Section 3.4.

After a sensor fault is detected, identifying the sensor fault caused by either stationary

mainline sensors or stationary ramp sensors is necessary. A threshold λ_2 for the likelihood of the two observers is defined and utilized in Eq. (3.21):

$$-\log(p(\boldsymbol{y}^{(l)}(k+1))) < \lambda_2, l = 1,2 \tag{3.21}$$

When Eq. (3.21) is violated, the sensor faults can be identified. For example, when $-\log(p(\boldsymbol{y}^{(1)}(k+1))) > \lambda_2$, meaning a large negative logarithm of the likelihood is estimated by observer 1, this can be caused by abnormal measurements from the ramp flow sensors. Thus, the stationary ramp sensor fault is identified. Both threshold λ_1 and λ_2 are selected by considering the model uncertainty and traffic measurement noises.

Once a sensor fault is detected and identified, the fault diagnosis scheme can reconfigure the VSL controller accordingly to provide reliable traffic state estimation and ensure the system performance. When there is no sensor fault detected, meaning the system operates under health mode, the estimation from the Kalman filter is fed to the sliding mode controller to maintain the traffic density $\rho_2(k)$ at the critical density. However, when a stationary mainline sensor fault is detected and identified, the traffic state estimation from the Kalman filter may greatly deviate from the actual traffic condition. The fault diagnosis scheme can replace the traffic state estimation from the Kalman filter with the observed states from the observer in Eq. (14). Similarly, when a stationary ramp sensor fault occurs, the fault diagnosis can reconfigure the controller to use the observed traffic states from the observer in Eq. (15). By detecting, identifying the sensor faults, and reconfiguring the controller, the fault diagnosis can achieve fault-tolerant VSL control for freeway work zones.

3.4 Experiments and Results

The fault-tolerant VSL control system was evaluated on a freeway segment with the length of 3.2 km (2 mi) on SR99N in California, US as shown in Fig. 3.3.

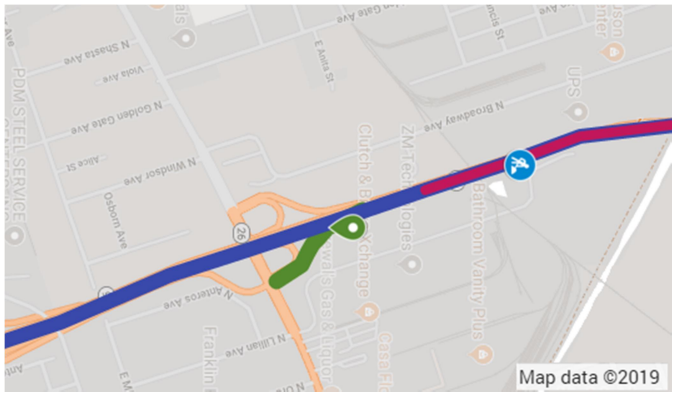


Fig. 3.3. Freeway network with a work zone (Map data © 2019 Google)

The blue line in Fig. 3.3 represents the selected freeway segment. On May 3rd, 2018, a 1 km (0.6 mi) construction work zone was established starting from the State Postmile (PM) 19.7. The work zone area is shown in the red line with a construction icon in Fig. 3.3. The green line with a marker which locates upstream of the work zone represents the on-ramp from Fremont Street. One of the three lanes was closed at the construction work zone, which resulted in severe traffic congestion upstream of the work zone. Two stationary mainline sensors installed at State PM 19.6 and State PM 20.1 were utilized to detect the traffic outflows of the acceleration zone and work zone. The ramp sensor at State PM 19.4 was installed at the on-ramp from Fremont Street. No other ramps were available in the acceleration zone and work zone area. Therefore, the stationary mainline

sensor at State PM 19.6 and on-ramp senor at State PM 19.4 were utilized to simulate the stationary sensor faults. The selected freeway segment in Fig. 3.3 was built in traffic simulator SUMO. The microscopic model was calibrated and validated using the realistic traffic data in California freeway management database [18].

The fundamental diagram was calibrated. The free flow speed, road capacity, work zone critical density ρ_c^b, backward wave speed ω, capacity drop parameter *a* and *b* were 104 km/h (65 mi/h), 5500 veh/h, 42 veh/km, 21 km/h, -0.95 km/h and 3730 veh/h respectively. The lengths of the acceleration zone and work zone were set as 500 m and 650 m respectively according to the realistic sensor location. Based on the realistic traffic data statistics analysis [73], [77], the standard deviations of the measurements from traffic sensors were selected as 25 veh/h and 3 veh/km respectively. 10 % of the connected vehicle market penetration rate was chosen to simulate the traffic flow. The controller parameter *c*, η and *q* were chosen as 2, 15 and 40 based on the previous study [71]. The speed limits $v_{\max}$, $v_{\min}$, Δv and $\Delta v_{\max}$ were selected as 104 km/h (65 mi/h), 16 km/h (10 mi/h), 8 km/h (5 mi/h) and 8 km/h (5 mi/h) respectively. The sample time interval ΔT_s and control time interval ΔT_c were chosen as 15 s and 60 s respectively. Different values of the threshold λ_1 and λ_2 were examined and tested through the simulation. To accurately detect the senor faults and avoid false alarms, λ_1 and λ_2 were selected as 20 and 15 respectively.

To demonstrate the effectiveness of the fault-tolerant VSL control system, three scenarios under fault-tolerant control were considered: 1) health mode without sensor

faults; 2) with the stationary mainline sensor fault occurring at 2835 s; 3) with the stationary ramp sensor fault occurring at 2835 s. To ensure consistent simulation results, 10 repeated simulations were conducted with different random seeds. As a warm-up period, the first 5 minute simulation data were discarded. Then the simulation was run for 6000 s. Low mainline traffic demand of 2200 veh/h lasted for the first 1200 s. Then the mainline demand increased to 3500 veh/h at 1800 s and maintained at high demand for 1500 s, followed by the dropped demand of 2500 veh/h at 3900 s. This low demand lasted until the end of the simulation. In terms of the on-ramp traffic demand, the low on-ramp demand of 200 veh/h lasted until 850 s. Then the on-ramp demand increased to 300 veh/h and lasted for 2200 s. Eventually the on-ramp demand decreased to 200 veh/h at 4260 s and maintained at that demand until the end of the simulation.

The traffic density estimation at the acceleration zone, on-ramp flow estimation and negative loglikelihood estimation are shown in Fig. 3.4. The health estimations in Fig. 3.4 are from the Kalman filter while the reference measurements are obtained using the built-in functionality in the traffic simulator SUMO. Through comparison with the reference measurements, accurate density estimation and on-ramp flow estimation are achieved under the health mode in Fig. 3.4 (a) and (b). Moreover, Fig. 3.4(a) demonstrates that the fault-tolerant VSL controller can stabilize the traffic density at the acceleration zone around the critical density 42 veh/km. In addition to the accruate density estimation and on-ramp flow estimation, the estimations of negative loglikelihood in Figure 3.4(c), from the Kalman filter and two observers, are all below the defined threshold and no false alarms are generated.

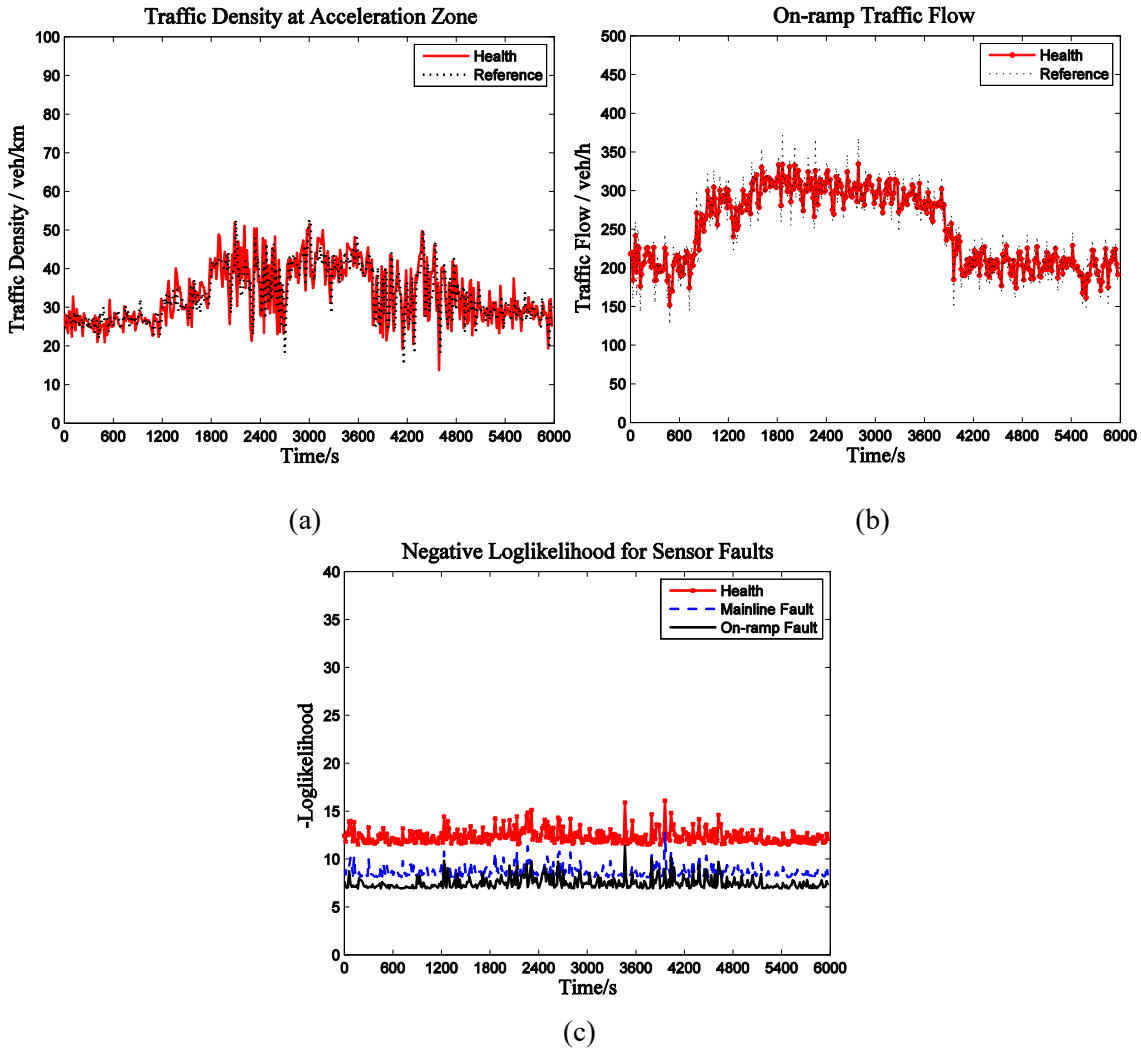
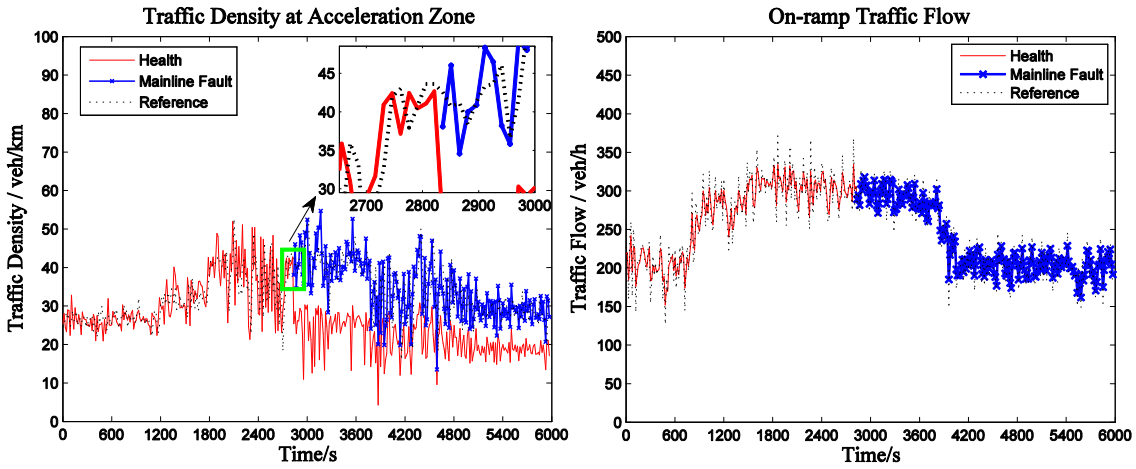


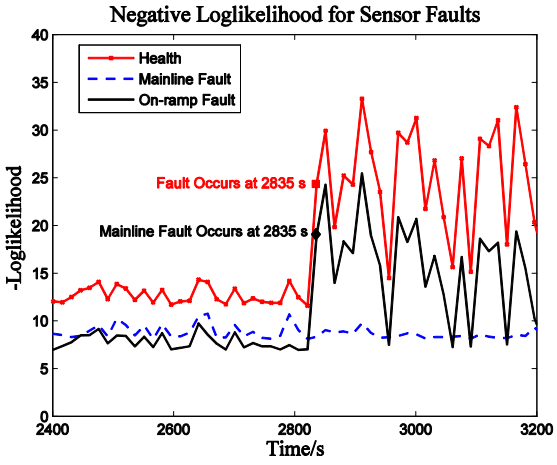
Fig. 3.4. Under scenario 1: (a) density estimation at the acceleration zone; (b) on-ramp flow estimation; (c) negative loglikelihood estimation

In Fig. 3.5, the estimations of traffic density at the acceleration zone, on-ramp flow and likelihood are presented under the stationary mainline sensor fault scenario. Due to the mainline sensor fault, the density estimation from the Kalman filter, which is used under health mode, deviates significantly from the reference measurement in Fig. 3.5(a). The unreliable density estimation may cause great impacts on VSL controller. However, the observed traffic states from the observer, which is designed for the mainline sensor

fault, can immediately come into effect and replace the corrupted state estimation caused by the mainline sensor fault at 2835 s. With reliable observed states, the VSL controller manages to stabilize the density at the acceleration zone near critical density even after the occurrence of the mainline sensor fault as shown in Fig. 3.5(a).



(a) (b)



(c)

Fig. 3.5. Under scenario 2: (a) density estimation at the acceleration zone; (b) on-ramp flow estimation; (c) negative loglikelihood estimation

Since the on-ramp flow sensor can work properly under scenario 2, both the Kalman filter and the observer can accurately estimate the on-ramp flow rate. However, the estimation of on-ramp flow is still being replaced by the observed traffic states in Fig. 3.5(b) to avoid any potential impacts from the Kalman filter. Fig. 3.5(c) demonstrates the process of mainline sensor fault detection and identification. When the negative loglikelihood estimation from the Kalman filter for the health mode exceeds the predefined threshold at 2835 s, a sensor fault is detected. Then the estimation from the observer, which is designed for the on-ramp sensor fault, exceeds the threshold as shown in Fig. 3.5(c). Thus the mainline sensor fault is identified. Meanwhile the VSL controller is reconfigured accordingly to maintain the density in Fig. 3.5(a) near the critical density. From Fig. 3.5, it can be seen that accurate sensor fault detection and identification can ensure the VSL system performance in the case of sensor faults.

Fig. 3.6 illustrates the impacts of stationary on-ramp sensor fault on estimations of density at the acceleration zone, on-ramp flow and the likelihood. Since the stationary mainline sensor is still functional, the density estimations, from the Kalman filter for health mode and two observers, are close to the reference measurements in Fig. 3.6(a). However, the estimation from the Kalman filter is still replaced by the observed states in case of potential estimation deviation. In Fig. 3.6(b), the on-ramp flow estimation from the Kalman filter immediately collapses when the on-ramp sensor fault occurs. However, the observer, which is designed for the on-ramp sensor fault, can still estimate accurate on-ramp flow rate shown in blue cross line.

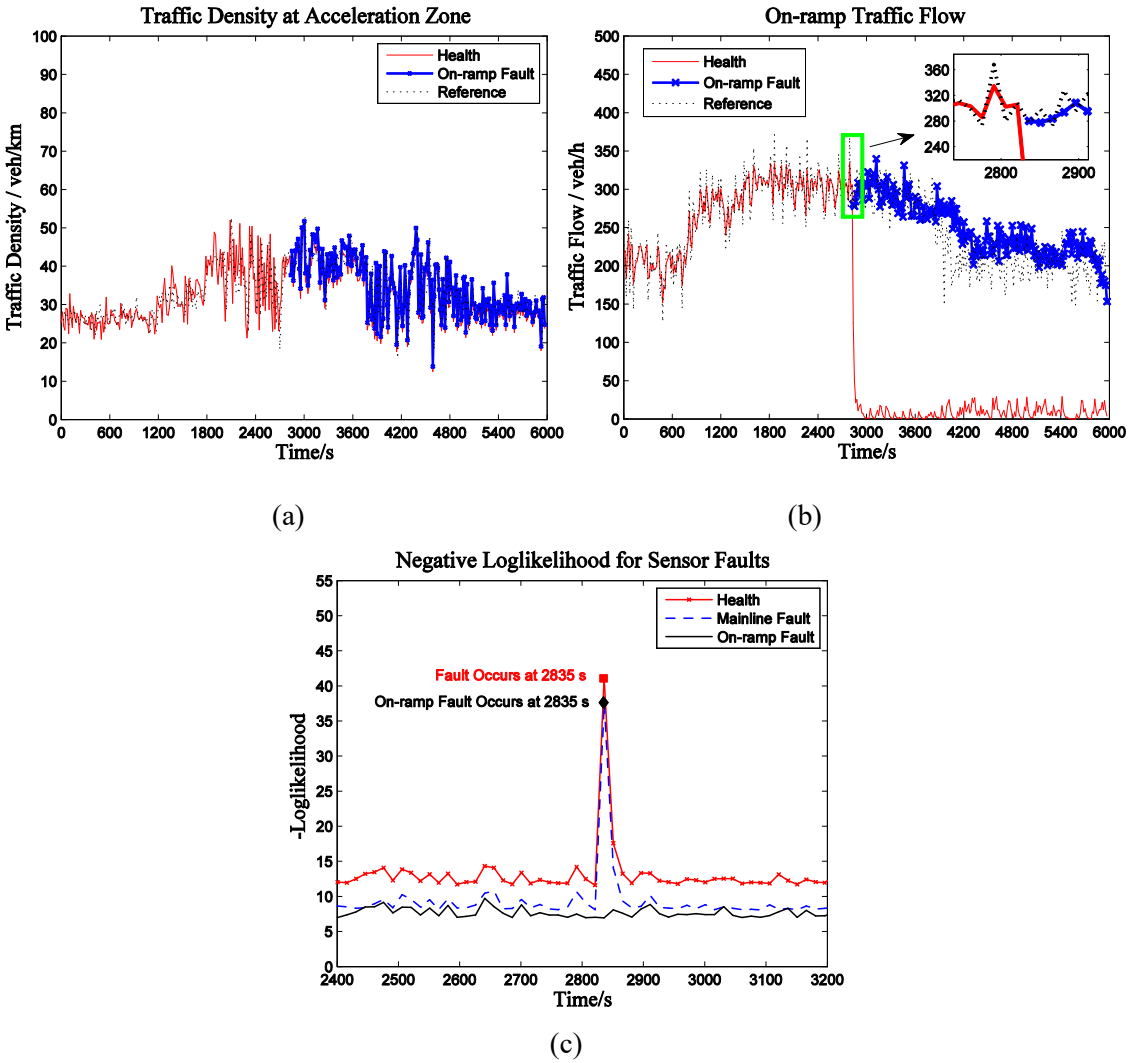


Fig. 3.6. Under scenario 3: (a) density estimation at the acceleration zone; (b) on-ramp flow estimation; (c) negative loglikelihood estimation

With the successful sensor fault diagnosis shown in Fig. 3.6(c), accurate on-ramp flow estimation is achieved in Fig. 3.6(b) by using the observed traffic states after the sensor fault. Fig. 3.6(c) demonstrates that the on-ramp sensor fault can be accurately detected and identified. A sensor fault is detected when the negative loglikelihood estimation from the Kalman filter exceeds the threshold. Then the abnomal negative loglikelihood

estimation from the observer, which is designed in the case of the mainline sensor fault, confirms that on-ramp sensor fault occurs.

From Fig. 3.4, 3.5 and 3.6, it can be seen that fault-tolerant VSL control can accurately estimate the traffic density and ramp flow under with and without sensor faults scenarios. Moreover, sensor faults can be reliably detected and identified online using the real time traffic data.

To quantitatively assess the accuracy of the estimation of the fault-tolerant VSL control system, the RMSE (root mean square error) analysis of the traffic density at the acceleration zone and on-ramp flow rate is presented in Table 3.1.

Table 3.1. RMSE for Estimation of Density and On-ramp Flow

	Density Estimation (veh/km)	On-ramp Flow Estimation (veh/h)
Scenario 1	6.36	27.4
Scenario 2	7.33	27.4
Scenario 3	6.38	38.0

It can be seen that small estimation error (around 6 veh/km in density and 30 veh/h in on-ramp flow) is achieved under all three scenarios. The mainline sensor fault mainly affects the density estimation resulting in a slight RMSE increase and has little impacts on the on-ramp flow estimation. In contrast, the on-ramp sensor fault has more impacts on the on-ramp flow estimation than the density estimation.

The effectiveness of the fault-tolerant VSL control is also evaluated through the performance measurements in Table 2. The performance without VSL control is utilized

as a comparison. The probability of time-to-collision is calculated based on that minimum 1.5 s is needed to avoid a collision [78]. The travel time upstream of the work zone is considered. The consistent improvements are shown among all three scenarios. The largest improvement lies in safety as 98% of the probability of collisions is avoided using the fault-tolerant VSL control. Around 10% of travel delays and fuel consumption can be saved and less CO2 emission is generated using the VSL control system.

Table 3.2. Performance Measurement under Scenario 1, 2 and 3

	No-Control	1	Improve-ment (%)	2	Improve-ment (%)	3	Improve-ment (%)
Time-to-collision p (%)	8.06	0.150	98.1	0.157	98.1	0.150	98.1
Average Travel Time T (min)	3.45	2.94	14.8	2.99	13.3	2.94	14.8
Fuel Consumption (l)	1004	913	9.06	932	7.17	914	8.96
CO2 (t)	2.33	2.12	9.01	2.16	7.30	2.12	9.01

Through the analysis of the performance of the developed system under three different scenarios, the fault-tolerant VSL control demonstrates the ability to accurately estimate traffic states under the health mode without false alarms. The stationary mainline and on-ramp sensor fault are detected and identified in real time. With the successful fault diagnosis using the likelihood estimation method, the accuracy of the estimations for the traffic density and the on-ramp flow is guaranteed. This ensures reliable VSL performance with the presence of sensor faults. Consistent performance

improvements with respect to mobility, safety and sustainability among all three scenarios demonstrate the fault tolerance of the developed system.

3.5 Conclusions and Future Work

A fault-tolerant VSL control system for freeway work zone using the likelihood estimation method is developed in this study. The Kalman filter using augmented traffic states can provide accurate estimations of the traffic density and ramp flow for the VSL controller under the health mode. Meanwhile, the observers have the ability to produce analytical redundancy for traffic state estimation near work zone areas. By exploiting the analytical redundancy, the likelihood estimation method successfully detects and identifies the stationary sensor faults online using real time traffic data without the requirement for historical traffic data. With the sensor fault detection and identification, the fault diagnosis can reconfigure the VSL controller to maintain the system performance. The fault-tolerant VSL control system demonstrates the ability to consistently improve mobility, safety and sustainability near freeway work zone areas even with the occurrence of sensor faults.

In this study, only stationary sensor faults are considered. The impacts of concurrent sensor faults and probe sensor faults will be studied. Meanwhile, more types of sensor faults such as abnormal noises and lag transmission may have different sensor fault phenomena. The solution to address other types of sensor faults will be considered.

4 Fault-Tolerant Control of Variable Speed Limits for Freeway Work Zone with Recurrent Sensor Faults

Recurrent sensor faults exhibit different features compared with permanent sensor faults and commonly occur in transportation systems. This chapter presents fault-tolerant VSL control for freeway work zones with consideration of recurrent sensor faults. An IMMP-based method is proposed to address different types of faults in stationary mainline and ramp sensors. An adaptive model set is designed to reduce the computational complexity. State covariance adaption and a pseudo-model set are developed to compensate for the discrepancies between the model parameters and the extent of corresponding sensor failures. The proposed system demonstrates reliable sensor fault diagnosis without the prerequisite of a good match between model parameters and sensor faults. Consistent improvements in mobility, safety, and sustainability are achieved under fault-free and recurrent sensor fault scenarios. The following submitted manuscript is included in this chapter.

- S. Du and S. Razavi, "Fault-tolerant control of variable speed limits for freeway work zone with recurrent sensor faults," *IEEE Transactions on Intelligent Transportation Systems*, 2020. *Submitted September 8, 2020*, T-ITS-20-09-1842.

The co-author's contributions include:

- Provide supervision and technical advice
- Review and edit the manuscript
- Financially support the research work

4.1 Abstract

Congestion associated with freeway work zones can adversely affect mobility, safety, and sustainability. Variable speed limits (VSL) control has been widely studied to mitigate the congestion caused by lane closures at work zones. However, most VSL controllers are designed without consideration of traffic sensor faults, especially recurrent sensor faults (RSFs) that commonly exist in freeway transportation systems. Therefore, this study proposes an interacting multiple model approach with a pseudo-model set (IMMP) to achieve VSL control with fault tolerance to different types of RSFs. With the design of a traffic flow model, an adaptive model set is developed using likelihood estimation to reduce the associated computational complexity. To ensure reliable RSF diagnosis, state covariance adaption is proposed to compensate for potential discrepancies caused by improper model parameters. A pseudo-mode set is designed to provide accurate traffic state estimations for VSL control without the prerequisite of a good match between the model parameters and the extent of corresponding sensor failures. The proposed system is evaluated under a realistic work zone environment using the traffic simulator SUMO. The results demonstrate that the system can achieve reliable RSF diagnosis and consistent improvements in mobility, safety, and sustainability near a freeway work zone area despite RSFs.

4.2 Introduction

Freeway work zones with lane closures are increasingly employed in roadway projects due to aging infrastructure and increasing traffic demand. Despite long-term

benefits of such infrastructure projects, temporary freeway work zones can create traffic disruptions and cause detrimental effects on mobility, safety, and the environment. In 2017, congestion caused approximately 8.8 billion hours of traffic delays and an extra 3.3 billion gallons of fuel in the U.S. urban areas [4]. Nearly 10 percent of U.S. traffic congestion expenses are attributed to work zones [5]. During the past decade, more than five thousand fatal crashes happened in the U.S. in relation to work zones [6]. With the significant impacts of freeway work zones, it is critical to ensure traffic efficiency and safety near work zone areas. Variable speed limit (VSL) control has been widely studied to mitigate the impacts of work zones. By imposing dynamic speed limits upstream of a work zone, VSL control can proactively regulate traffic flow, avoid capacity drop, and improve traffic conditions around work zone areas.

Many VSL control methods have been proposed by previous studies. Feedback-based methods [7], [8], [48], [71], [79] and kinematic wave theory-based approaches [9], [10] were developed for VSL controllers to stabilize traffic states around the critical density such that capacity drop phenomena can be avoided. VSL control was also solved as an optimization problem to reduce travel delays using optimal control [13], [14] and reinforcement learning methods [16], [17]. With emerging automated vehicle technologies, VSL control was also designed under a connected vehicle environment [80], [81]. However, the aforementioned work assumed reliable traffic measurements were available in the design of VSL controllers. Few studies considered the impacts of faulty sensors. According to [18], approximately 30 percent of stationary traffic sensors suffered from different types of sensor faults on a typical day in California. The large

volume of faulty data caused by sensor faults may affect the performance of a safety-critical system like VSL control and cause system degradation. Therefore, it is essential to consistently improve traffic conditions near work zones using VSL control despite the probable traffic sensor faults.

To mitigate the negative impacts of sensor faults, the spatial and temporal dependencies of traffic sensor data have been exploited by researchers to address faulty data. Previous studies can be broadly categorized into offline and online approaches. Offline methods such as tensor decomposition [21], multimodal deep learning [82], and generative adversarial networks [83] have been demonstrated to effectively recover missing traffic data. Nonetheless, after a faulty sensor returned to health status, measurements from this sensor were also incorporated to impute faulty data in [21], [82], [83]. Since VSL control needs real-time traffic state estimations, limited access to data beyond the sensor failure time makes offline methods infeasible for VSL controller design. In contrast, online methods can provide real-time estimations with a sensor malfunction. In [84] and [85], a pattern clustering tool and a fuzzy C-mean algorithm were proposed respectively to learn traffic data patterns so as to estimate faulty data. In [23], multiple linear regressions were used to detect abnormal sensor measurements and impute data in real time. In another study [86], a graph Markov network was built to predict missing data using neighboring road links. However, large historical traffic data were demanded by the previous studies [23], [84], [85], [86] to train and calibrate their proposed models. Sensors near freeway work zone areas, especially short-term work zones, may not generate enough data to ensure that models are well calibrated to provide

accurate estimations for the faulty traffic data. To overcome the limitation of large data demand, the previous study of the authors [87] used observer-based likelihood estimations for sensor fault diagnosis. However, recurrent sensor faults (RSFs) were not considered in any of the above-mentioned studies [23], [84]-[87], which can pose challenges to accurate estimation of traffic states.

RSFs commonly exist in transportation systems [18], [20]. Compared with permanent sensor faults (PSFs) which persist indefinitely until the sensors are repaired, RSFs can repeatedly appear and disappear without external intervention due to loose connection, surrounding interference or aging components. In electronic systems, the occurrence of RSFs can be 10 to 30 times more frequent than PSFs [88]. Uncertainties associated with faulty time intervals of RSFs may considerably increase maintenance costs caused by multiple maintenance site visits, and reduce system reliability. Thus it is of great importance to achieve RSF tolerance in traffic state estimations. Estimations of abnormal traffic data in the previous studies [23], [84]-[87] were limited to PSFs, which results in neglecting potential reliable measurements when a faulty traffic sensor restores to working order in RSFs. Besides, estimations are more susceptible to errors as fewer sensors are used. Another important consideration is the type of sensor malfunctions. In [23], [84]-[87], only one type of sensor faults was investigated. Different types of sensor malfunctions can also affect the estimation accuracy as shown in the results of this study. Therefore, to ensure the effectiveness of VSL control, a sensor fault diagnosis scheme is needed that can reliably detect RSFs with different types of malfunctions, and accurately estimate traffic states online.

To achieve fault diagnosis of different types of sensor faults, interacting multiple model (IMM) methods have been studied in motor control related fields. IMM methods use a set of models depicting a sensor health mode and all possible sensor faults. By obtaining mode probabilities and estimations of multiple models, IMM approaches can identify the type of sensor faults and improve state estimation. In [89], a distributed IMM system was designed to cope with different types of sensor faults in a robotic system. Variances were adjusted using a compensation parameter to improve IMM performance [90]. However, the fixed IMM structure used in [89], [90] may cause high computational costs as more models will be added and processed when more types of sensor faults are considered. Variable IMM structure [91] and an event-triggered fault diagnosis method [92] were proposed to avoid adding new models by estimating the extent of sensor failures. Nevertheless, the accuracy of state estimations in [89]-[92] highly relied on the degree of a match between the model and the corresponding sensor fault mode. Despite accurately identifying the type of sensor faults, a poor state estimation may still happen when preselected or adaptive model parameters in IMM cannot represent the extent of sensor failures with high accuracy. Given different types of RSFs, an IMM approach is needed to adaptively adjust the model set and provide accurate traffic state estimations even with an imperfect match between models and the corresponding sensor faults.

In view of the above-mentioned analysis, this study presents a fault-tolerant VSL control system for a freeway work zone with different types of RSFs. The control system proposed in this study offers the following three distinct contributions:

1) It consistently produces good VSL control performance with the occurrence of

RSFs. The VSL control performance is evaluated in terms of mobility, safety, emissions, and fuel consumption.

2) It reliably detects and identifies different types of RSFs in stationary mainline and ramp traffic sensors. An adaptive model set is designed to process RSFs with low computational complexity. State covariance adaption is built to compensate for the discrepancies caused by improper model parameters.

3) It accurately estimates traffic states without the prerequisite of a good match between the model parameters and the extent of corresponding sensor failures. An IMM approach is augmented by a pseudo-model set (IMMP) to provide traffic state estimations for VSL control.

The remainder of this paper is structured as follows. The traffic flow model is proposed in Section 4.3. Section 4.4 presents the IMMP-based fault-tolerant VSL control system with consideration of different types of RSFs in stationary mainline and ramp sensors. Evaluation of the effectiveness of the proposed system with analytical results is provided in Section 4.5. Conclusion is presented in Section 4.6.

4.3 Traffic Flow Model

A freeway work zone with lane closures can cause road capacity loss and easily lead to congestion near the work zone area. Additionally, a queue forming upstream of the work zone can result in more capacity loss due to capacity drop phenomena [3]. To reduce congestion near a freeway work zone, a VSL controller can impose dynamic speed limits so as to affect the traffic dynamics around the work zone area. Thus it is

essential to build a traffic flow model to understand and describe the evolution of traffic flow near the work zone.

The layout of a freeway work zone area is presented in Fig. 4.1. The congestion is shown in gradient color. The mixed traffic flow with conventional vehicles and connected vehicles traverses the work zone area where one lane is closed. We assume these connected vehicles have no self-driving functionalities and only transmit their speeds and positions.

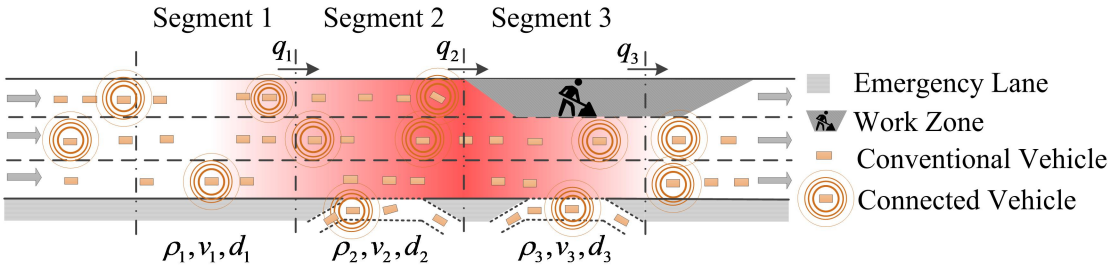


Fig. 4.1. Freeway work zone area (not to scale)

As shown in Fig. 4.1, the freeway is partitioned into multiple segments. For each segment n, the density ρ_n, the average speed v_n, the segment length d_n, the ramp flow q_{rn}, and the traffic outflow q_n are denoted. The lengths of segment 2 and 3 are selected such that each segment only has on-ramps or off-ramps. It should be noted there may be no on-ramps or off-ramps at some freeway segments in the real world. Segment 1, 2, and 3 are the VSL zone, the acceleration zone, and the work zone respectively. Dynamic speed limits are imposed at the VSL zone to control the speed of vehicles. The acceleration zone allows vehicles to accelerate to free speed to traverse the work zone area. The work

zone segment is the segment in which a lane closure exists due to construction work.

Probe sensors and stationary sensors are used in this study to measure traffic states. Connected vehicles as probe sensors can monitor the traffic states at different segments when they traverse the work zone area. Two stationary mainline traffic sensors are located at segment 2 and segment 3 to monitor the traffic states at these segments respectively. Stationary ramp sensors are installed to measure ramp flows.

The traffic state vector is defined as $\boldsymbol{x}_k = [\rho_{2,k} \quad \rho_{3,k} \quad q_{r2,k}]^{\mathrm{T}}$ where k represents the kth time step. The density $\rho_{2,k}$ is selected since the objective of VSL control [71] is to stabilize $\rho_{2,k}$ at the critical density ρ_c which is determined by the fundamental diagram. The ramp flow $q_{r2,k}$ is also included as it can directly affect the density $\rho_{2,k}$. The density $\rho_{3,k}$ is selected to provide the redundant traffic state which will be discussed in Section 4.4.1. By using the conservation law [50] and the sensor measurement vector $\boldsymbol{z}_k$, the traffic flow model is designed as

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}\boldsymbol{u}_{k-1} + \boldsymbol{\delta}_{k-1} \\ \boldsymbol{z}_k = \boldsymbol{C}_k\boldsymbol{x}_k + \boldsymbol{\omega}_k \end{cases} \tag{4.1}$$

Specifically,

$$\boldsymbol{A}_{k-1} = \begin{bmatrix} 1 - v_{2,k-1}\Delta T / d_2 & 0 & \Delta T / d_2 \\ v_{2,k-1}\Delta T / d_3 & 1 - v_{3,k-1}\Delta T / d_3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.2}$$

$$\boldsymbol{B}_{k-1} = \begin{bmatrix} \Delta T / d_2 & 0 & 0 \\ 0 & \Delta T / d_3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.3}$$

$$\boldsymbol{u}_{k-1} = [q_{1,k-1} \quad q_{r3,k-1} \quad 0]^{\mathrm{T}} \tag{4.4}$$

$$\boldsymbol{C}_k = \boldsymbol{I}_{3*3} \tag{4.5}$$

$$\boldsymbol{\delta}_{k-1} = \begin{bmatrix} \delta_{1,k-1} & \delta_{2,k-1} & \delta_{3,k-1} \end{bmatrix}^{\mathrm{T}} \tag{4.6}$$

$$\boldsymbol{\omega}_k = \begin{bmatrix} \omega_{1,k} & \omega_{2,k} & \omega_{3,k} \end{bmatrix}^{\mathrm{T}} \tag{4.7}$$

$\boldsymbol{\delta}_k$ and $\boldsymbol{\omega}_k$ are the process and measurement noises assumed to be white noises with the covariance $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ respectively. ΔT is the sample time interval. The average speed $v_{n,k-1}$ can be detected using probe sensors [73]. The ramp flow $q_{rn,k}$ is measured by the stationary ramp sensor at the segment *n*. By using the known relation $\rho = q / v$, the traffic density $\rho_{n,k}$ can be obtained using the average speed $v_{n,k}$ and the flow measurement from the stationary mainline sensor within the segment *n*. The traffic flow $q_{1,k-1}$ is governed by the dynamic speed limit *u* imposed at the VSL zone. By imposing speed limit *u* and regulating the flow $q_{1,k-1}$, a VSL controller in [71] is used to stabilize the density $\rho_{2,k}$ at the critical density ρ_c when traffic demand exceeds the work zone capacity. Readers are referred to [71] for the derivation of the VSL controller based on sliding mode control.

Since the objective of VSL control is to stabilize the density $\rho_{2,k}$, the estimation accuracy of $\rho_{2,k}$ greatly affects VSL control performance. However, the estimation accuracy of $\rho_{2,k}$ depends on the reliability of model Eq. (4.1). When a stationary mainline or ramp senor fault occurs, the measurement matrix $\boldsymbol{C}_k$ is no longer the identity matrix and will be determined by the extent of sensor failures. The discrepancies caused by sensor faults in model Eq. (4.1) can cause a poor estimation of the density $\rho_{2,k}$, resulting

in a potential VSL control failure. Thus different types of RSFs are considered in the fault-tolerant control design as explained in the following section.

4.4 IMMP-Based Fault-Tolerant Control Design

An IMM method is a multiple model approach for hybrid systems. A bank of models, where each model represents a fault mode, is processed in parallel using an IMM method. With the occurrence of RSFs, the fault mode with the highest possibility is selected at each time step as the current system mode and the combined estimation is calculated accordingly [75]. The jump-linear hybrid system can be established as

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{A}_{k-1}(s_k)\boldsymbol{x}_{k-1} + \boldsymbol{B}_{k-1}(s_k)\boldsymbol{u}_{k-1} + \boldsymbol{\delta}_{k-1}(s_k) \\ \boldsymbol{z}_k = \boldsymbol{C}_k(s_k)\boldsymbol{x}_k + \boldsymbol{\omega}_k(s_k) \end{cases} \tag{4.8}$$

where s_k refers to the sensor fault mode at the *k*th time step. $\boldsymbol{A}_{k-1}(s_k)$, $\boldsymbol{B}_{k-1}(s_k)$, $\boldsymbol{C}_{k-1}(s_k)$, $\boldsymbol{\delta}_{k-1}(s_k)$, and $\boldsymbol{\omega}_{k-1}(s_k)$ are mode dependent matrices under the mode s_k which is modeled as a Markov chain with the transition probability

$$\sigma_{ij} = P\{s_k^j \mid s_{k-1}^i\} \tag{4.9}$$

where *i* and *j* are the mode indexes. The fault mode s_{k-1}^i and s_k^j are selected from a set *S* which consists of all possible sensor fault modes.

4.4.1 Sensor Faults

Ensuring reliability of traffic sensors that are used by VSL control is of critical importance. Since the objective of VSL controllers [71] is to stabilize the density $\rho_{2,k}$,

this study focuses on the faults of the stationary mainline sensors and ramp sensors at segment 2 in Fig. 4.1. The faults of probe sensors will be studied in our future work.

To provide reliable measurements of the density $\rho_{2,k}$, redundant traffic data are needed. One solution is to provide physical redundancy by installing an extra set of stationary traffic sensors at the same location at segment 2. However, the cost of physical redundancy will be prohibitive in a large-scale implementation. Thus, this study exploits cost-effective analytical redundancy using the density $\rho_{3,k}$ in state vector $\boldsymbol{x}_k$.

Stationary traffic sensors suffer from different types of faults. According to [18], [20], detecting zero flow rates with the presence of vehicles and outputting abnormally high flow rates are the leading stationary sensor faults in transportation systems; these two types of faults are mainly considered in this study. Besides, a mainline freeway generally has multiple lanes. Given the possibility of lane-based sensors, the partial mainline sensors with zero flow rates are also considered. Considering the RSFs, we assume there are no transitions from fault to fault. Thus the set of the total sensor fault modes is defined as

$$S = S^H \cup S^{ML} \cup S^{RP} \tag{4.10}$$

Specifically,

$$\begin{cases} S^H = \{s^h\} \\ S^{ML} = \{s^{m1}, s^{m2}, s^{m3}\} \\ S^{RP} = \{s^{r1}, s^{r2}\} \end{cases} \tag{4.11}$$

where S^H, S^{ML}, and S^{RP} represent the set of the sensor health mode, the mainline sensor fault modes, and the ramp sensor fault modes respectively. s^h is denoted as the sensor

health mode when no faults occur. s^{m1}, s^{m2}, and s^{m3} stand for three different mainline sensor faults, i.e. zero flow rates, abnormally high flow rates, and partial mainline sensors with zero flow rates. s^{r1} and s^{r2} are denoted for ramp sensor faults of zero flow rates and abnormally high flow rates respectively.

Different sensor fault modes can be depicted by modifying traffic flow models. As discussed in Section 4.3, the measurement matrix $\boldsymbol{C}_k^j$, where j refers to any mode s^j in the set S, may not be the identity matrix. For example, when a mainline sensor fault occurs, the first row of the measurement matrix $\boldsymbol{C}_k^j$ will be multiplied by a factor α which is determined by the extent of the sensor failure. The factor α is larger than 1 when the fault is related to abnormally high flow rates. Accordingly, α will be 0-1 and 0 respectively for a partial mainline sensor failure and a failure with zero flow rates. Similarly, when a ramp sensor fault occurs, the third row of the matrix $\boldsymbol{C}_k^j$ will be multiplied by a factor β which is selected based on the extent of the ramp sensor failure.

4.4.2 Adaptive Model Set

An adaptive model set can reduce the computation complexity and improve the efficiency of an IMM method. A fixed structure of an IMM method will consume more computational resources when more models are added to the model set. In addition, an improvement in estimations may not be achieved with the use of more models [91]. In Section 4.4.1, a total of six fault modes are considered in Eq. (4.10). However, it is very unlikely that a traffic sensor has different types of faults at the same time. Extra

computation burdens can be caused when six models are processed all the time. To improve the cost effectiveness, an adaptive model set S_k is developed as

$$S_k = \{s_k^h, s_k^m, s_k^r\} \tag{4.12}$$

where s_k^m and s_k^r can be any mode in the set S^{ML} and S^{RP} respectively at the *k*th time step. s_k^h is denoted as the health mode at the *k*th time step.

To determine s_k^m and s_k^r, a predictive model set $S_{k|k-1}$ is obtained as

$$S_{k|k-1} = \{s_k^j \mid P\{s_k^j \mid s_{k-1}^i\} > 0, s_{k-1}^i \in S_{k-1}, \varphi_{k-1}^i > \varepsilon\} \tag{4.13}$$

where φ_{k-1}^i is the mode probability of the mode s_{k-1}^i and ε is selected to avoid using a mode with a very low probability for the prediction of possible modes at the *k*th time step.

Using the predictive model set $S_{k|k-1}$ and the predetermined model sets S^{ML} and S^{RP}, we can obtain tentative model sets for the mainline and ramp sensor faults

$$\begin{cases} S_k^{ML,t} = S_{k|k-1} \cap S^{ML} \\ S_k^{RP,t} = S_{k|k-1} \cap S^{RP} \end{cases} \tag{4.14}$$

Then the mode s_k^m and s_k^r can be selected from the tentative model sets $S_k^{ML,t}$ and $S_k^{RP,t}$ using the likelihood estimation as

$$\begin{cases} s_k^m = \underset{s_k^{l_1,t} \in S_k^{ML,t}}{\arg\max}\, p(\boldsymbol{z}_k \mid \bar{\boldsymbol{x}}_{k-1|k-1}^{l_1,t}, \bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_1,t}) \\ s_k^r = \underset{s_k^{l_2,t} \in S_k^{RP,t}}{\arg\max}\, p(\boldsymbol{z}_k \mid \bar{\boldsymbol{x}}_{k-1|k-1}^{l_2,t}, \bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_2,t}) \end{cases} \tag{4.15}$$

where the likelihood estimation is calculated using the mixed state estimation $\bar{\boldsymbol{x}}_{k-1|k-1}^{l_1,t}$, $\bar{\boldsymbol{x}}_{k-1|k-1}^{l_2,t}$ and state covariance $\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_1,t}, \bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_2,t}$. The derivation of the mixed state estimation

and state covariance is discussed in Section 4.4.5.

By using Eq. (4.15), the adaptive model set Eq. (4.12) can be obtained. Note that Eq. (4.15) is used when there are multiple modes in the tentative model set $S_k^{ML,t}$ or $S_k^{RP,t}$. When there is only one mode in $S_k^{ML,t}$ or $S_k^{RP,t}$, the mode s_k^m or s_k^r can be determined directly without using Eq. (4.15). In the cases where there are no modes in $S_k^{ML,t}$ or $S_k^{RP,t}$ meaning the probability of mainline sensor faults or ramp sensor faults at the *k*th time step is close to zero or negligible, the mode s_{k-1}^m or s_{k-1}^r at the last time step is selected. With the negligible mode probability, impacts of mode s_{k-1}^m or s_{k-1}^r on the IMM estimations are negligible as well. Thus the adaptive model set is designed for a recursive IMM process.

4.4.3 State Covariance Adaption

The model parameter α and β related to the extent of sensor failures can affect the likelihood estimations in Eq. (4.15). With the selection of α and β, the sensor fault modes and the corresponding traffic flow models are determined. Then the likelihood estimations in Eq. (4.15) are calculated using the traffic flow models. However, α and β may not depict the extent of sensor failure accurately, resulting in insignificant likelihood estimations of all modes in the tentative model sets. This may cause difficulties in selecting the adaptive model set S_k.

To compensate for the discrepancies between the model parameters and the extent of corresponding sensor failures, a mixed state covariance adaption is designed. The compensation is activated when the likelihood estimations in Eq. (4.15) satisfy

$$\max_{s_k^{l,t} \in (S_k^{ML,t} \cup S_k^{RP,t})} p(\boldsymbol{z}_k \mid \bar{\boldsymbol{x}}_{k-1|k-1}^{l,t}, \bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l,t}) < \gamma \tag{4.16}$$

where the mode $s_k^{l,t}$ belongs to the union of $S_k^{ML,t}$ and $S_k^{RP,t}$. The mixed traffic state estimation $\bar{\boldsymbol{x}}_{k-1|k-1}^{l,t}$ and state covariance $\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l,t}$ are derived under the mode $s_k^{l,t}$. γ is selected as the threshold of the minimum likelihood estimation.

With the condition in Eq. (4.16) satisfied, the mixed state covariance $\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l,t}$ can be compensated to account for the discrepancies caused by inappropriate model parameters. A compensation factor λ_k is introduced to compensate for $\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l,t}$ as

$$\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l,t} := \lambda_k \bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l,t} \tag{4.17}$$

By comparing the residual vectors of the models in tentative model sets $S_k^{ML,t}$ and $S_k^{RP,t}$ with a base residual vector, the factor λ_k is calculated as

$$\lambda_k = \left(\min_{s_k^{l,t} \in (S_k^{ML,t} \cup S_k^{RP,t})} \left\| \bar{\boldsymbol{y}}_k^{l,t} \right\| \right) / \left\| \bar{\boldsymbol{y}}_k^{base} \right\| \tag{4.18}$$

Specifically,

$$\left\| \bar{\boldsymbol{y}}_k^{base} \right\| = \frac{1}{N} \sum_{m=1}^{N} \min_{s_{k-m}^j \in S_{k-m}} \left\| \boldsymbol{y}_{k-m}^j \right\| \tag{4.19}$$

where $\|\bullet\|$ denotes the Euclidean norm. $\bar{\boldsymbol{y}}_k^{l,t}$ and $\boldsymbol{y}_{k-m}^j$ are the residual vectors under the mode $s_k^{l,t}$ and s_{k-m}^j respectively. The derivation of $\bar{\boldsymbol{y}}_k^{l,t}$ and $\boldsymbol{y}_{k-m}^j$ is discussed in Section 4.4.5. $\left\| \bar{\boldsymbol{y}}_k^{base} \right\|$ is calculated using a moving window with the size of *N*.

The minimum Euclidean norm is used in Eq. (4.18) and Eq. (4.19) to avoid overcompensation. The model matching the actual extent of sensor failure should have a

smaller Euclidean norm of the residual vector than other models'. Therefore, the minimum Euclidean norms of $\bar{\boldsymbol{y}}_k^{l,t}$ and $\boldsymbol{y}_{k-m}^{j}$ are used to ensure a smaller compensation factor is obtained. Thus the mixed state covariance adaption is built to compensate for the discrepancies between the predetermined model parameters and the actual extent of sensor failures.

4.4.4 Pseudo-mode Set Design

The accuracy of IMM estimations relies on the selection of model parameters. In Section 4.4.3, the mixed state covariance adaption is developed in Eq. (4.17) given the discrepancies caused by model parameters. The reliability of traffic sensor fault detection and identification can be improved by using Eq. (4.17). However, an IMM method will still use these model parameters to estimate traffic states. Without a good match between the model parameters and the extent of sensor failures, the estimation accuracy of the traffic state vector $\boldsymbol{x}_k$ can be affected. To overcome the dependence of the estimation accuracy on the selection of model parameters α and β, the IMMP approach is developed using a pseudo-model set S^p as

$$S^p = \{s^{h,p}, s^{m,p}, s^{r,p}\} \tag{4.20}$$

where $s^{h,p}$, $s^{m,p}$, and $s^{r,p}$ are the pseudo health mode, the pseudo mainline sensor fault mode, and the pseudo ramp sensor fault mode respectively.

Compared with the adaptive model set S_k in Eq. (4.12), the pseudo-model set S^p is time-invariant whereas the model set S_k needs to be calculated at each time step.

Specifically, the traffic flow model under any mode $s^{j,p}$ in the model set S^p can be expressed as

$$\begin{cases} \boldsymbol{x}_k^{j,p} = \boldsymbol{A}_{k-1}^{j,p}\boldsymbol{x}_{k-1}^{j,p} + \boldsymbol{B}_{k-1}^{j,p}\boldsymbol{u}_{k-1} + \boldsymbol{\delta}_{k-1}^{j,p} \\ \boldsymbol{z}_k^{j,p} = \boldsymbol{C}_k^{j,p}\boldsymbol{x}_k^{j,p} + \boldsymbol{\omega}_k^{j,p} \end{cases} \tag{4.21}$$

The matrix $\boldsymbol{A}_{k-1}^{j,p}$, $\boldsymbol{B}_{k-1}^{j,p}$, $\boldsymbol{u}_{k-1}$, and $\boldsymbol{\delta}_{k-1}^{j,p}$ in Eq. (4.21) can be obtained by using Eq. (4.2)-(4.4), (4.6) under the respective modes. With respect to $\boldsymbol{C}_k^{j,p}$ and $\boldsymbol{\omega}_k^{j,p}$, Eq. (4.5) and Eq. (4.7) can be used under the pseudo health mode $s^{h,p}$ whereas Eq. (4.22) and Eq. (4.23) are used under the pseudo mainline sensor fault mode $s^{m,p}$ and the pseudo ramp sensor fault mode $s^{r,p}$ respectively.

$$\boldsymbol{C}_k^{m,p} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \boldsymbol{\omega}_k^{m,p} = [\omega_{2,k} \quad \omega_{3,k}]^{\mathrm{T}} \tag{4.22}$$

$$\boldsymbol{C}_k^{r,p} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \boldsymbol{\omega}_k^{r,p} = [\omega_{1,k} \quad \omega_{2,k}]^{\mathrm{T}} \tag{4.23}$$

Accordingly, $\boldsymbol{z}_k^{m,p} = [\rho_{3,k} \;\; q_{r2,k}]^{\mathrm{T}}$ and $\boldsymbol{z}_k^{r,p} = [\rho_{2,k} \;\; \rho_{3,k}]^{\mathrm{T}}$ can be obtained. The observabilities of the model Eq. (4.21) under respective modes are proved in [87].

By taking advantage of the mode probabilities of the adaptive model set S_k, the IMMP method can estimate traffic state estimations without the impacts of the model parameters. For example, the traffic flow model under the pseudo mainline sensor fault mode in Eq. (4.21), (4.22) does not rely on the measurements from mainline sensors. Thus the model parameters, which depict the extent of mainline sensor failures, will not affect the estimations of traffic states. Nevertheless, when we only use the pseudo model

set S^p in a traditional IMM method to provide state estimations, the mode probabilities cannot be calculated correctly as different measurements are used by the models in S^p. Since both S_k and S^p have three modes including the health mode, the mainline sensor fault mode, and the ramp sensor fault mode, the mode probabilities calculated for S_k can be shared with the mode probabilities for S^p. Thus the strength of estimating traffic states with the pseudo-mode set can be kept with the shared mode probabilities from the adaptive model set.

4.4.5 IMMP-Based Fault-tolerant Control Scheme

The IMMP-based fault-tolerant control system consists of the VSL controller, the traffic flow model, and the RSF diagnosis as shown in Fig. 4.2. The reliable density estimation from the RSF diagnosis component is fed to the VSL controller. Then the speed limit control signal is generated to affect the evolution of traffic states near a work zone area. The RSF diagnosis is designed by using the IMM algorithm with the pseudo-mode set and the adaptive model set. The steps of RSF diagnosis are presented as follows. Note that $i, j = 0,1,2$ with 0, 1, and 2 referring to the health mode, the mainline sensor fault mode, and the ramp sensor fault mode respectively.

1) *Tentative model set selection (TS).* The tentative model set $S_k^{ML,t}$ and $S_k^{RP,t}$ are selected in Eq. (4.14) using the mode probability φ_{k-1}^i, the mainline sensor fault model set S^{ML}, and the ramp sensor fault model set S^{RP}.

2) *Mixing*. The mixed state estimation and covariance in Eq. (4.15) and Eq. (4.16) can

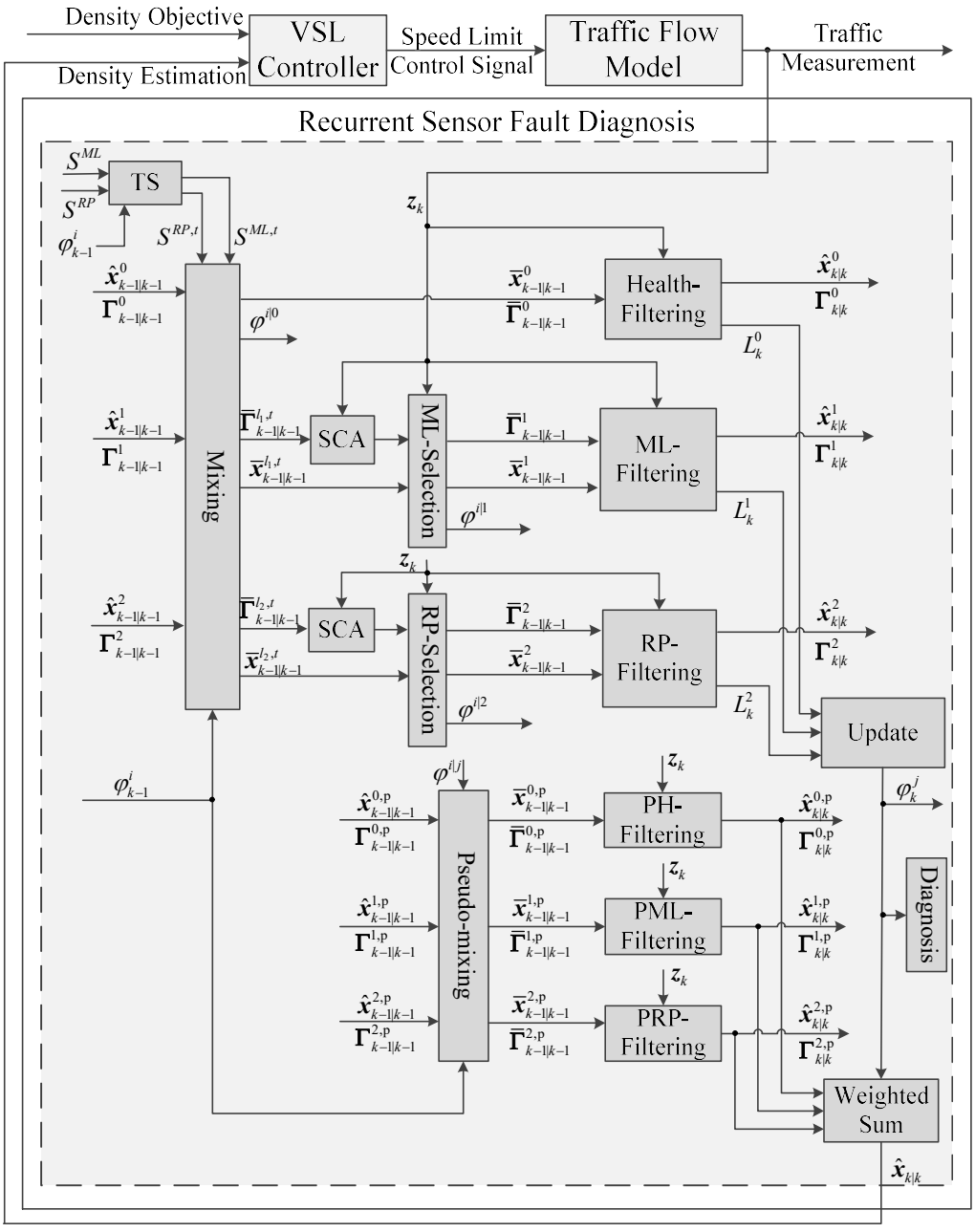


Fig. 4.2. Diagram of the IMMP-based fault-tolerant control. TS: tentative set. SCA: state covariance adaption. ML: mainline. RP: ramp. PH: pseudo health. PML: pseudo mainline. PRP: pseudo ramp.

be calculated by fusing the state estimation $\hat{\boldsymbol{x}}^{i}_{k-1|k-1}$ and covariance $\boldsymbol{\Gamma}^{i}_{k-1|k-1}$ at the last time

step under the mode s_{k-1}^{i} as

$$\bar{\boldsymbol{x}}_{k-1|k-1}^{l_j,t} = \sum_{s_{k-1}^{i} \in S_{k-1}} \varphi^{i|l_j,t} \hat{\boldsymbol{x}}_{k-1|k-1}^{i} \tag{4.24}$$

$$\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_j,t} = \sum_{s_{k-1}^{i} \in S_{k-1}} \varphi^{i|l_j,t} [\boldsymbol{\Gamma}_{k-1|k-1}^{i} + (\hat{\boldsymbol{x}}_{k-1|k-1}^{i} - \bar{\boldsymbol{x}}_{k-1|k-1}^{l_j,t})(\hat{\boldsymbol{x}}_{k-1|k-1}^{i} - \bar{\boldsymbol{x}}_{k-1|k-1}^{l_j,t})^{\mathrm{T}}] \tag{4.25}$$

where the conditional probability $\varphi^{i|l_j,t}$ is calculated as

$$\varphi^{i|l_j,t} = \varphi_{k-1}^{i} \sigma_{i(l_j,t)} / \varphi_{k-1}^{l_j,t} \tag{4.26}$$

Specifically,

$$\varphi_{k-1}^{l_j,t} = \sum_{s_{k-1}^{i} \in S_{k-1}} \varphi_{k-1}^{i} \sigma_{i(l_j,t)} \tag{4.27}$$

where $\sigma_{i(l_j,t)}$ is the mode transition probability from s_{k-1}^{i} to $s_{k}^{l_j,t}$. φ_{k-1}^{i} is the mode probability of the mode s_{k-1}^{i}.

3) *State covariance adaption (SCA)*. The measurement $\boldsymbol{z}_k$ is used for the mixed state covariance adaption. When Eq. (4.16) is not satisfied, $\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_j,t}$ will not be adjusted. However, when Eq. (4.16) is satisfied, the compensation factor λ_k is obtained in Eq. (4.18) using Eq. (4.28)

$$\boldsymbol{y}_{k}^{l_j,t} = \boldsymbol{z}_k - \boldsymbol{C}_{k}^{l_j,t} (\boldsymbol{A}_{k-1}^{l_j,t} \bar{\boldsymbol{x}}_{k-1|k-1}^{l_j,t} + \boldsymbol{B}_{k-1}^{l_j,t} \boldsymbol{u}_{k-1}) \tag{4.28}$$

Then we can adjust the mixed state covariance $\bar{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_j,t}$ in Eq. (4.17) accordingly. Note that the residual vector $\boldsymbol{y}_{k-m}^{j}$ in Eq. (4.19) is used to update the $\left\| \bar{\boldsymbol{y}}_{k}^{base} \right\|$ and will be discussed in the step 6 *filtering*.

4) *Mainline and ramp sensor fault mode selection (ML-, RP-Selection)*. With the

calculation of the mixed state estimation $\overline{\boldsymbol{x}}_{k-1|k-1}^{l_j,t}$ and state covariance $\overline{\boldsymbol{\Gamma}}_{k-1|k-1}^{l_j,t}$, Eq. (4.15) can be used to select the mainline and ramp sensor fault modes when there are multiple modes in the model set $S_k^{ML,t}$ or $S_k^{RP,t}$. More details can be found in Section 4.4.2 for the cases where the number of modes in $S_k^{ML,t}$ or $S_k^{RP,t}$ equals to one or zero.

It should be noted that when the mainline and ramp sensor fault modes are determined, the mode $s_k^{l_j,t}$ will be modified as the mode s_k^j in the following steps. All the mode-related matrices will be modified accordingly as well.

5) *Pseudo-mixing (PH-, PML-, PRP-, filtering)*. With the selection of the mainline and ramp sensor fault modes, the conditional mode probability $\varphi^{i|j}$ can be determined and shared with the pseudo-model set. Thus the mixed state estimation $\overline{\boldsymbol{x}}_{k-1|k-1}^{j,p}$ and state covariance $\overline{\boldsymbol{\Gamma}}_{k-1|k-1}^{j,p}$ can be obtained as

$$\overline{\boldsymbol{x}}_{k-1|k-1}^{j,p} = \sum_{s_{k-1}^{i,p} \in S^p} \varphi^{i|j} \hat{\boldsymbol{x}}_{k-1|k-1}^{i,p} \tag{4.29}$$

$$\overline{\boldsymbol{\Gamma}}_{k-1|k-1}^{j,p} = \sum_{s_{k-1}^{i,p} \in S^p} \varphi^{i|j} [\boldsymbol{\Gamma}_{k-1|k-1}^{i,p} + (\hat{\boldsymbol{x}}_{k-1|k-1}^{i,p} - \overline{\boldsymbol{x}}_{k-1|k-1}^{j,p})(\hat{\boldsymbol{x}}_{k-1|k-1}^{i,p} - \overline{\boldsymbol{x}}_{k-1|k-1}^{j,p})^{\mathrm{T}}] \tag{4.30}$$

where the mode $s_{k-1}^{i,p}$ refers to the *i*th mode in the pseudo-model set at the last time step.

6) *Filtering (Health-, ML-, RP-, PH-, PML-, PRP-, filtering)*. With the mixed state estimation $\overline{\boldsymbol{x}}_{k-1|k-1}^{j}$ and covariance $\overline{\boldsymbol{\Gamma}}_{k-1|k-1}^{j}$ at the last time step, three Kalman filters (KFs) for the models in the adaptive model set can be employed to obtain the state estimation $\hat{\boldsymbol{x}}_{k|k}^{j}$ and covariance $\boldsymbol{\Gamma}_{k|k}^{j}$ at the current *k*th time step. Readers can refer to [75] for the recursive KF algorithm. Similarly, three KFs for the models in the pseudo-model set can

be used to obtain the state estimation $\hat{\boldsymbol{x}}_{k|k}^{j,p}$ and covariance $\boldsymbol{\Gamma}_{k|k}^{j,p}$ at the *k*th time step using $\overline{\boldsymbol{x}}_{k-1|k-1}^{j,p}$ and $\overline{\boldsymbol{\Gamma}}_{k-1|k-1}^{j,p}$.

The likelihood estimation L_k^j in Fig. 4.2 can be obtained as

$$L_k^j = p(\boldsymbol{z}_k \mid \overline{\boldsymbol{x}}_{k-1|k-1}^j, \overline{\boldsymbol{\Gamma}}_{k-1|k-1}^j) \tag{4.31}$$

The residual vector $\boldsymbol{y}_k^j$ in Eq. (4.32) can be added to update $\left\| \overline{\boldsymbol{y}}_k^{base} \right\|$ in Eq. (4.19) for the usage at the next time step.

$$\boldsymbol{y}_k^j = \boldsymbol{z}_k - \boldsymbol{C}_k^j (\boldsymbol{A}_{k-1}^j \overline{\boldsymbol{x}}_{k-1|k-1}^j + \boldsymbol{B}_{k-1}^j \boldsymbol{u}_{k-1}) \tag{4.32}$$

7) *Update*. The posterior mode probability φ_k^j at current *k*th time step can be calculated using the prior probability in Eq. (4.27) and the likelihood estimation in Eq. (4.31) as

$$\varphi_k^j = L_k^j \varphi_{k-1}^j \tag{4.33}$$

8) *Diagnosis*. Different types of RSFs are detected and identified in the *diagnosis* step. Specifically, the mode probability φ_k^0, φ_k^1, and φ_k^2 of the health mode, the mainline sensor fault mode, and the ramp sensor fault mode respectively are used. With $f = 1, 2$, a sensor fault can be detected using

$$\max_f \varphi_k^f > \varphi^{th} \tag{4.34}$$

where φ^{th} is the predetermined threshold. When Eq. (4.34) is satisfied, a sensor fault is detected. A mainline sensor fault occurs when $\varphi_k^1 > \varphi_k^2$ whereas a ramp sensor fault occurs when $\varphi_k^1 < \varphi_k^2$. The specific type of the mainline or ramp sensor fault is determined by the adaptive model set S_k in *mainline and ramp sensor fault mode selection* step. In

contrast, when Eq. (4.34) is not satisfied, no sensor faults occur.

9) *Weighted sum*. The combined estimation from the pseudo-model set is calculated using the mode probabilities of the models in the adaptive mode set as

$$\hat{\boldsymbol{x}}_{k|k} = \sum_{j=0}^{2} \varphi_k^j \hat{\boldsymbol{x}}_{k|k}^{j,p} \tag{4.35}$$

Then the density estimation $\hat{\rho}_{2,k}$ in the probabilistically weighted sum $\hat{\boldsymbol{x}}_{k|k}$ is fed back for VSL control.

With traffic sensor measurements, the RSF diagnosis can perform these nine steps recursively to achieve sensor fault diagnosis and provide reliable traffic state estimations for the VSL controller.

4.5 Experiment and Results

To demonstrate the fault-tolerance of the VSL control system to different types of RSFs, we evaluated the system on a 4.2 km (2.6 mi) section of the northbound SR99 in California, U.S. as shown in Fig. 4.3. The blue line, which starts at A and ends at B, shows the mainline freeway of the section. The position A and B correspond to the State Postmile (PM) 17.5 and 20.1 respectively. The thick red line with a worker icon stands for a construction work zone with one of three lanes closed in May 2018. This work zone with the length of 1 km started at the State PM 19.7. The on-ramp at East Fremont Street is shown in the yellow line with an arrow upstream of the work zone in Fig. 4.3.

Two stationary mainline sensors are installed at the State PM 19.6 and 20.1 respectively. The ramp sensor is installed at the State PM 19.4. With consideration of the realistic locations of the sensors and the work zone, the lengths of the VSL zone, the

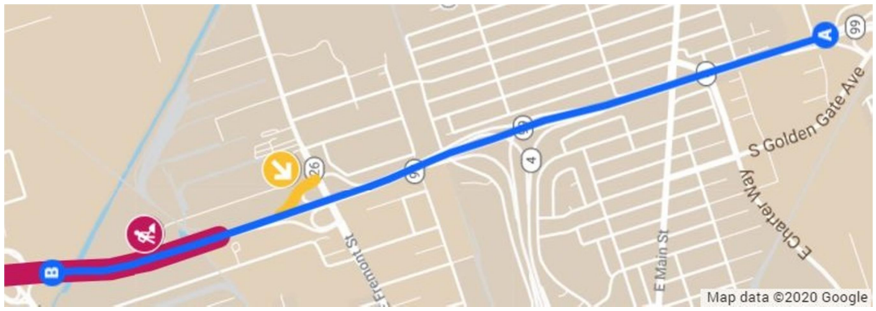


Fig. 4.3. Section of the freeway SR99 with a work zone (Map data © 2020 Google)

acceleration zone and the work zone were selected as 800 m, 500 m, and 650 m respectively. It should be noted that the VSL controller is not sensitive to the lengths of the divided zones [71]; various lengths can be selected if mobile sensors such as wheeled sensor stations are available. Thus, the stationary mainline sensors at the State PM 19.6 and 20.1 can measure the traffic outflows of the acceleration zone and the work zone respectively. The ramp sensor at the State PM 19.4 can detect the on-ramp flow to the acceleration zone. There are no other ramps in the acceleration zone and the work zone. Accordingly, different types of RSFs of the mainline sensor at the State PM 19.6 and the ramp sensor at the State PM 19.4 are considered.

The section of the freeway SR99 was built in the traffic simulator SUMO. The microscopic models such as car- following and lane changing models were calibrated and validated using the realistic traffic measurements from [18].

Dynamic traffic demand is used to assess the effectiveness of the IMMP-based fault diagnosis with different types of RSFs. During the warm-up for the simulation, low traffic demand was first generated for 5 minutes in the freeway section. With the warm-up simulation data discarded, the simulation was run for 7200 s with varying traffic

demand. For the mainline freeway, the low traffic demand of 2200 veh/h was generated for the first 700 s. Then the demand started increasing and reached the high demand of 3400 veh/h at 1300 s. This high demand of 3400 veh/h lasted until 4700 s. Afterwards, the demand started decreasing and reduced to 2200 veh/h at 5300 s. Then the demand maintained at 2200 veh/h for the rest of the simulation. For the ramp, the traffic demand of 150 veh/h was introduced for the first 300 s. Then the demand rose to 270 veh/h at 1100 s. The demand of 270 veh/h lasted for 4100 s, followed by the demand reducing to 150 veh/h at 5610 s. The demand maintained at 150 veh/h until the end of the simulation.

Five different types of RSFs are considered. The partial mainline sensor fault occurs when the sensor of the rightmost lane of the mainline freeway fails. Given the sensor faults analysis in [18], [20], 3000 veh/h and 9000 veh/h were selected as the abnormally high flow rates detected by the sensors for the ramp flow and the mainline flow respectively. The sequence of sensor fault modes will be discussed in the following figures. In the following analysis, H, F1, F2, F3, F4, and F5 represent the sensor health mode, the mainline sensor fault with zero flow rates, the ramp sensor fault with zero flow rates, the mainline sensor fault with abnormally high flow rates, the ramp sensor fault with abnormally high flow rates, and the partial mainline sensor fault respectively.

The model parameter α was selected as 0, 0.5, and 2 respectively for the fault F1, F5, and F3 whereas the model parameter β was chosen as 0 and 2 respectively for the fault F2 and F4. The impacts of the predetermined α and β will be discussed in the following figures. The following parameters were used: $\varepsilon = 0.01, \gamma = 10^{-50}, \varphi^{th} = 0.5$. ε was selected to avoid using the modes with negligible mode probabilities for the model set prediction.

γ was chosen based on the range of the likelihood estimations in [87] to detect faults timely and avoid too many false alarms. φ^{th} was used to ensure the dominant mode was selected as the system mode. The probability of staying at the same mode was selected as 0.6. The probability from the health mode to any fault was 0.08 whereas the probability from any fault to the health mode was 0.4. We also tested various probabilities of staying at the same mode. Varying the probability from 0.6 to 0.8 has limited impacts on the results in this study. With the realistic traffic measurements analysis in [73], the standard deviations of the process noise for the density and the ramp flow were selected as 3 veh/km and 25 veh/h respectively whereas the standard deviations of the measurement noise for the density and the ramp flow were chosen as 10 veh/km and 25 veh/h. The sample time interval was 15 s and the control time interval was 60 s. The market penetration rate of connected vehicles was chosen as 10 %. The window size N was selected as 3 via empirical testing.

Three scenarios are considered to assess the effectiveness of the components in Section 4.4.2-4.4.4 in the fault-tolerant VSL control: 1) only the adaptive model set is used; 2) both the adaptive model set and the state covariance adaption are used; 3) all three components including the adaptive model set, the state covariance adaption, and the pseudo-model set are used. To ensure consistent results, the simulation was run for 10 times with different random seeds. The same RSFs sequence in Fig. 4.4, 4.5, and 4.6 was used for all three scenarios.

The system modes which are the RSF diagnosis results are shown in Fig. 4.4(a), 4.5(a), and 4.6(a) whereas the mode probabilities for each mode are shown in Fig. 4.4(b),

4.5(b), and 4.6(b). It can be seen that different types of RSFs occur throughout the simulation. It should be noted that faults of zero flow rates and partial sensor failures are only considered under high demand as these faults have limited impacts on VSL control when traffic demand is low.

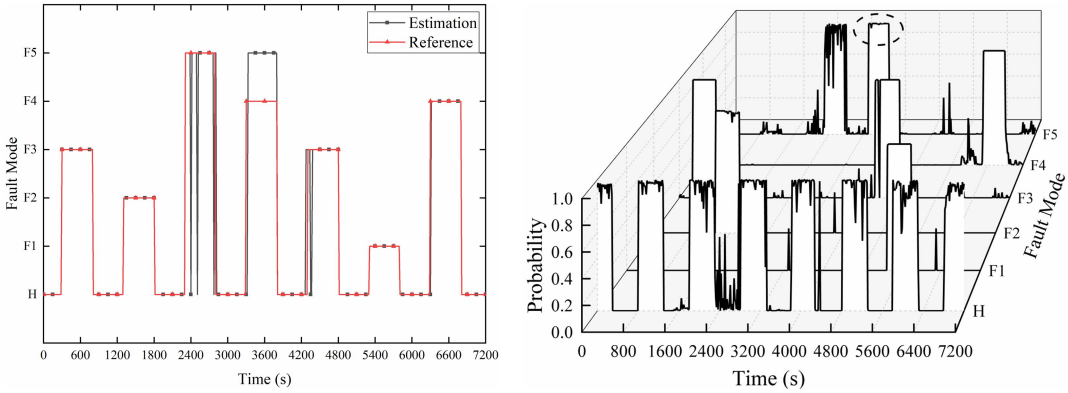


Fig. 4.4. RSF diagnosis in scenario 1: (a) system modes; (b) mode probabilities

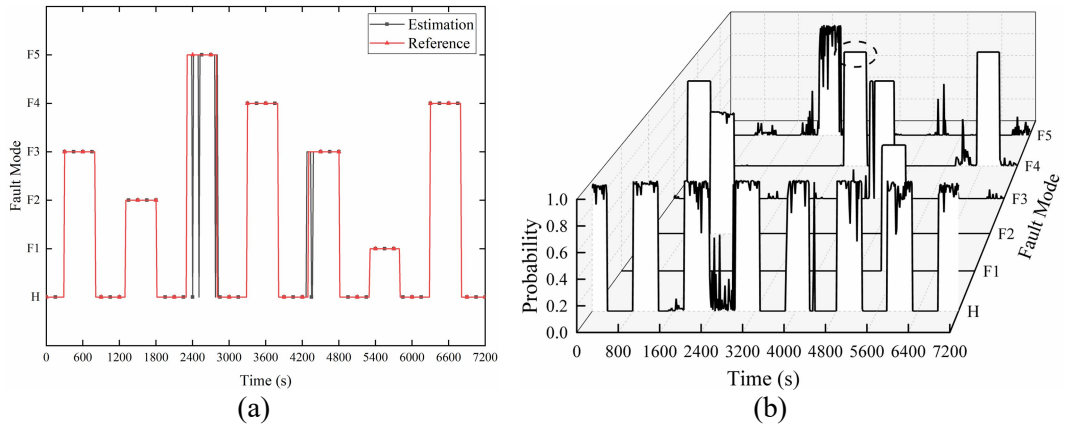


Fig. 4.5. RSF diagnosis in scenario 2: (a) system modes; (b) mode probabilities

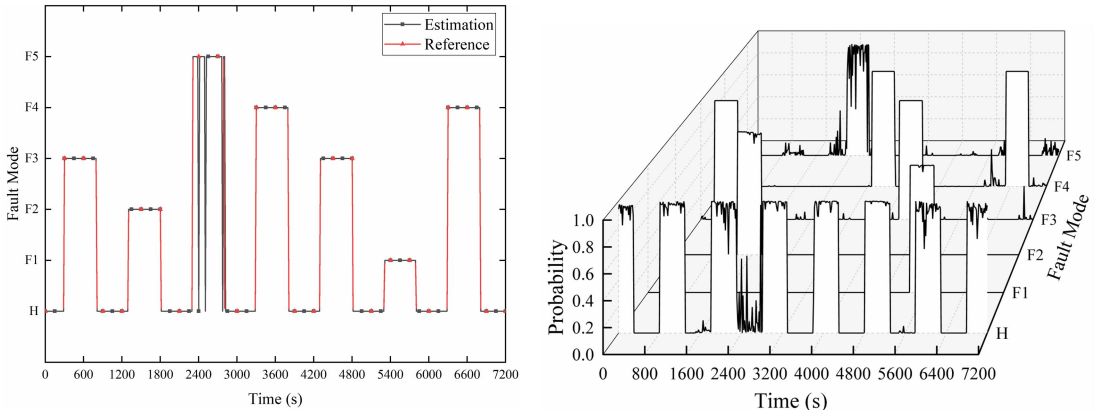
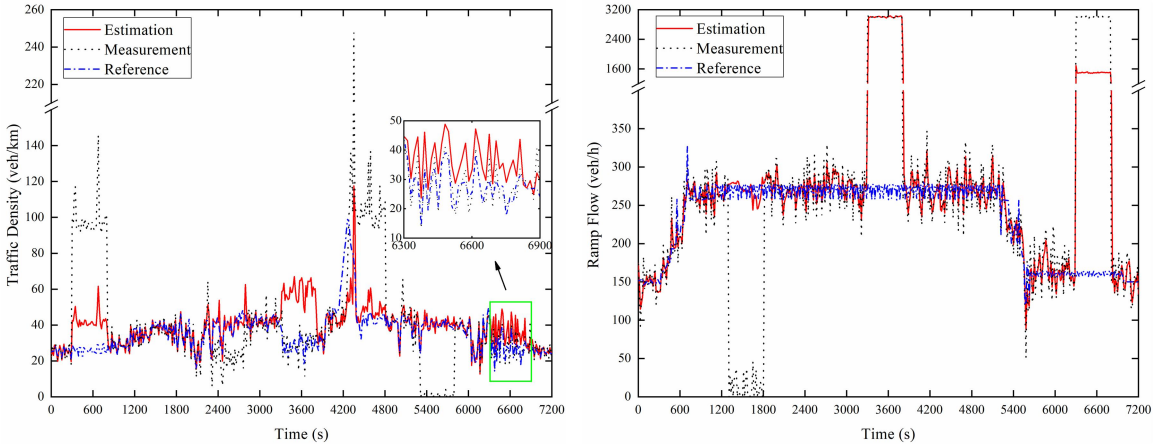


Fig. 4.6. RSF diagnosis in scenario 3: (a) system modes; (b) mode probabilities

Comparison with the actual sensor fault modes shows that the RSF diagnosis generally detects different types of RSFs accurately for all three scenarios. However, the actual fault F4 is identified as F5 around the time step 3600 s in scenario 1 in Fig. 4.4(a). The reason is that the improper selection of the model parameter for the F4 model leads to too small likelihood estimations for all models when F4 occurs. Consequently, no faults are detected at first and faulty measurements are used to estimate traffic states. This causes a large error in state estimations and leads to incorrect fault identification afterwards. In contrast, the reliable identification of F4 is shown in scenario 2 in Fig. 4.5(a) with the state covariance adaption. By adjusting the state covariance, the likelihood estimations are increased to allow for improper model parameters. Thus F4 can be reliably detected. The major difference of mode probabilities between scenario 1 and 2 is also illustrated with dash circles in Fig. 4.4(b) and 4.5(b). The mode probabilities of F5 are close to one in scenario 1 around 3600 s whereas the mode probabilities of F4 are close to one with the state covariance adaption in scenario 2. Compared with scenario 1

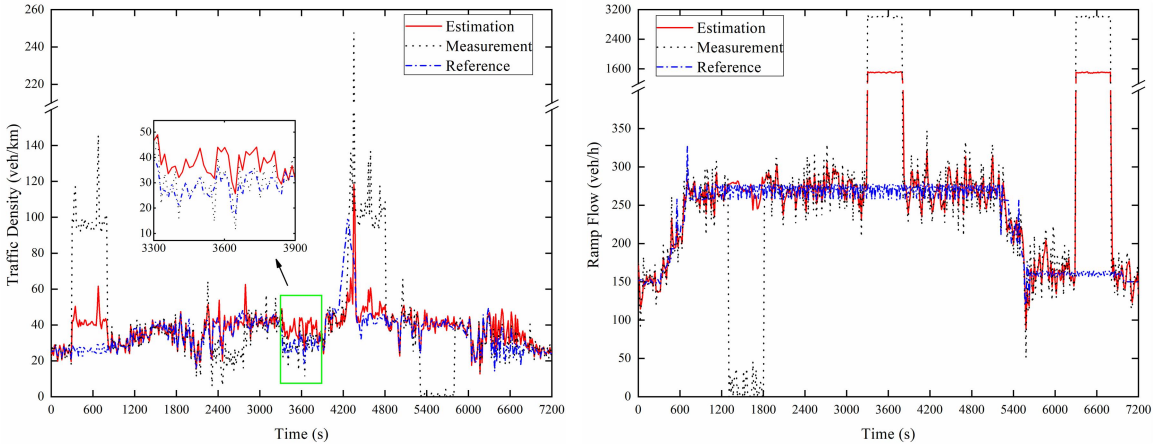
and 2, Fig. 4.6 shows improvements in both system modes and mode probabilities with the pseudo-mode set. The modes are detected more reliably in Fig. 4.6(a) and the probabilities of actual modes dominate most of the simulation time in Fig. 4.6(b). The performance sensor fault diagnosis is improved as more accurate traffic states are estimated using the pseudo-model set.

The estimations of density and ramp flow are shown in Fig. 4.7, 4.8, and 4.9 for the respective scenarios.



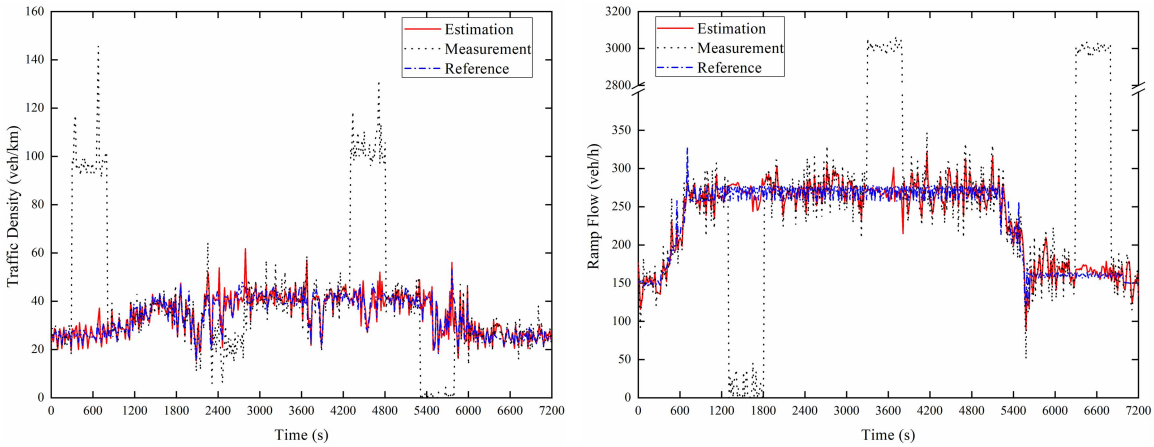
(a) (b)

Fig. 4.7. State estimations in scenario 1: (a) density; (b) ramp flow



(a) (b)

Fig. 4.8. State estimations in scenario 2: (a) density; (b) ramp flow



(a) (b)

Fig. 4.9. State estimations in scenario 3: (a) density; (b) ramp flow

In Fig. 4.7, 4.8, and 4.9, the measurements are generated from traffic sensors whereas the built-in functions in SUMO are used to generate the true density and ramp flow for reference. Fig. 4.7 (a) and (b) illustrate that the estimations of the density and ramp flow deviate greatly from the true density and ramp flow during the fault F3 and F4 due to the improper model parameters. As shown in Fig. 4.8 (a) and (b), with the correct identification of the fault F4 in scenario 2 using the state covariance adaption, less deviation in the estimations is achieved around 3600 s. Although reliable sensor fault diagnosis is shown in scenario 2, the improper model parameters still cause a large error in the estimations during the fault F3 and F4. In contrast, the density and ramp flow are consistently estimated with good accuracy under different types of sensor faults in scenario 3 shown in Fig. 4.9 (a) and (b). Since the estimations with the pseudo-model set do not rely on the model parameters, accurate state estimation can still be provided for VSL control without a good match between the model parameters and the extent of sensor failures.

The accuracy of estimations of the density and ramp flow is assessed quantitatively using RMSE (root mean square error) analysis in Table 4.1 for all three scenarios.

The estimations are analyzed for each fault mode. Due to the incorrect identification of the fault F4 in scenario 1, a large error of estimations of the density and ramp flow is caused during the fault F4. Moreover, the estimation of the ramp flow under the health mode in scenario 1 is also affected as faulty sensor data are used to estimate traffic states. Slightly better estimations are achieved in scenario 2 with the state covariance adaption when the fault diagnosis can reliably detect RSFs. Meanwhile, scenario 3 shows more accurate estimations with the pseudo-model set under different fault modes.

Table 4.1. RMSE for Estimation of the Density and Ramp Flow

Fault Mode	Density (veh/km)			Ramp Flow (veh/km)		
	1	2	3	1	2	3
H	7.9	7.5	4.6	134.4	34.4	16.4
F1	8.4	8.4	6.9	21.5	21.5	21.5
F2	5.7	5.7	5.7	48.1	48.1	48.0
F3	13.4	13.3	3.5	16.7	16.6	16.5
F4	26.8	10.1	3.2	2444.3	1287.6	13.1
F5	4.7	4.7	4.8	11.2	11.2	11.1

The performance of sensor fault diagnosis in all three scenarios is also analyzed quantitatively in Table 4.2. Precision, recall, and F_1-score are used as the quantitative measures of effectiveness. For a specific fault mode such as F1, precision is calculated by dividing the true positive cases (the identified F1 mode matches the true F1 mode) by the

sum of true positive cases and false positive cases (the identified F1 mode does not match the true F1 mode) whereas recall is calculated by dividing the true positive cases by the sum of true positive cases and false negative cases (the true F1 mode is identified as other modes rather than F1). F_1-score provides the combined measure of precision and recall. Scenarios 2 and 3 achieve high precision scores, recall scores, and F_1-scores for different types of sensor faults. It can be seen that the state covariance adaption and the pseudo-model set can greatly improve the reliability of RSF diagnosis. In contrast, in scenario 1, a low recall score is obtained for F4 as scenario 1 cannot reliably identify the fault F4. Meanwhile, incorrectly identifying the actual fault F4 as F5 causes the low precision for the fault F5. Thus, low F_1-scores are obtained for both F4 and F5 in scenario 1.

Table 4.2. RMSE for Estimation of the Density and Ramp Flow

Fault Mode	1			2			3		
	Precision	Recall	F_1-score	Precision	Recall	F_1-score	Precision	Recall	F_1-score
H	97.1	98.9	98.0	98.6	98.6	98.6	98.5	99.1	98.8
F1	95.9	100	97.8	100	100	100	100	99.2	99.6
F2	100	98.5	99.2	100	98.5	99.2	100	98.5	99.2
F3	96.5	99.2	97.8	96.5	99.2	97.8	97.5	99.2	98.4
F4	100	24.2	32.7	100	100	100	100	100	100
F5	41.0	92.4	56.4	96.9	92.4	94.6	98.4	92.4	95.3

[a]Unit:%

To evaluate the VSL control performance, the spatial and temporal evolution of the speed in the freeway section is presented in Fig. 4.10. Compared with the no control

scenario in Fig. 4.10 (a), scenario 1 and 2 show no significant improvements in terms of the congestion mitigation. In Fig. 4.10 (b) and (c), light congestion is created around 4300 s at the beginning of the work zone. This congestion is caused by inappropriate VSL control signals due to inaccurate density estimations in scenario 1 and 2. Similar lengths and duration of congestion are created in Fig. 4.10 (a), (b), and (c). However, the congestion is created with a shorter length and resolved earlier in Fig. 4.10 (d) under the IMMP-based fault-tolerant VSL control in scenario 3. It can be seen that accurate traffic state estimations are essential to ensure the effectiveness of VSL control.

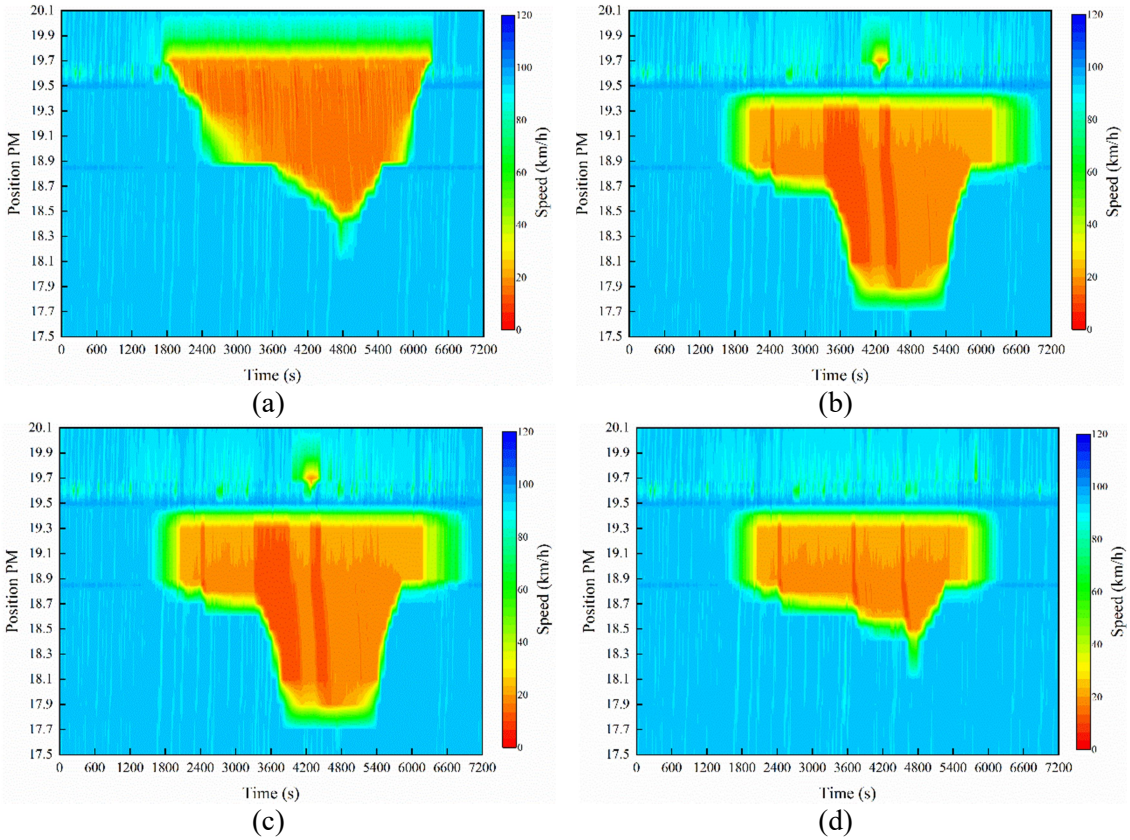


Fig. 4.10. Speed profile in the SR99 freeway section under (a) no control; (b) scenario 1; (c) scenario 2; (d) scenario 3

The VSL performance is also analyzed quantitatively in Table 4.3. Average travel time, time-to-collision [71], CO2 emission, and fuel consumption are selected as the measures of effectiveness.

Table 4.3. Performance Measurement of Fault-tolerant VSL Control

	No-Control	1	Improve-ment (%)	2	Improve-ment (%)	3	Improve-ment (%)
Average Travel Time (min)	4.7	5.4	-14.9%	5.3	-12.8%	4.2	+10.6%
Time-to-collision (%)	8.2	2.6	+68.3%	2.5	+69.5%	0.65	+92.1%
CO2 Emission(t)	5.1	5.5	-7.8%	5.4	-5.9%	4.8	+5.9%
Fuel Consumption (l)	2184.9	2332.0	-6.7%	2321.3	-6.2%	2086.2	+4.5%

Compared with the no control scenario, scenario 1 and 2 make the congestion worse in terms of travel time, CO2 emission, and fuel consumption. However, large improvements in safety are achieved under scenario 1 and 2. In contrast, IMMP-based fault-tolerant VSL control in scenario 3 shows consistent improvements from the aspects of mobility, safety, emissions, and fuel consumption. Thus with reliable RSF diagnosis using the state covariance adaption and accurate state estimations using the pseudo-model set, the VSL control performance can be consistently improved.

4.6 Conclusion

An IMMP-based fault-tolerant VSL control approach is proposed to achieve fault tolerance to different types of RSFs in stationary mainline and ramp traffic sensors. With

the design of the traffic flow model, the adaptive model set can reduce the computational complexity with fewer models and improve the efficiency of the IMM recursive algorithm. The state covariance adaption can adaptively compensate for the discrepancies caused by the improper model parameters. Using the adaptive model set and the state covariance adaption, the RSF diagnosis can reliably detect and identify different types of RSFs. Moreover, the design of the pseudo-model set can provide accurate traffic state estimations for VSL control without the prerequisite of a good match between the model parameters and the extent of the corresponding sensor failures. The IMMP-based fault-tolerant VSL control shows consistent improvements in mobility, safety and sustainability for freeway work zones with the occurrence of different types of RSFs.

5 Conclusion

Mitigation of the negative impacts of freeway work zones is of great importance in modern societies with aging infrastructure and increasing traffic demand. This research successfully developed fault-tolerant VSL control for freeway work zones with consistent improvements in mobility, safety, and sustainability.

The VSL controller presented in this research can generate VSLs to proactively regulate traffic flow in response to real-time traffic measurements from sensors. Since sensor faults commonly occur in transportation system, sensor fault diagnosis of this system detects and identifies different types of sensor faults. In addition, reliable traffic states are estimated by the sensor fault diagnosis component to ensure the effectiveness of VSL control. Thus, fault tolerance of VSL control to permanent and recurrent sensor faults is achieved with consistent improvements in mobility, safety, and sustainability.

5.1 Contributions

In view of negative impacts of freeway work zones, limitations of VSL control, and limited VSL control studies considering sensor faults, the main academic contributions of this thesis are presented as follows.

- Simple yet robust VSL Controller:

The efficient and robust VSL controller was developed in this research. Existing VSL control methods demand high computational costs such as multiple steps of traffic prediction in the optimization process [13], [14] and extensive model parameters calibrations [16], [17]. Thus the simple VSL control was developed based on sliding

mode control to efficiently generate VSLs for traffic flow management. Compared with the kinematic wave theory based methods [9], [10], the VSL controller also shows the strength of insensitivity to traffic disturbance owing to variable control structures. To avoid the potential impacts of linearization near the critical density in [7], [8], the nonlinear traffic flow model was developed to explicitly incorporate the discontinuity of the fundamental diagram caused by capacity drop. Under realistic freeway work zone conditions, the VSL controller demonstrates strong robustness under noisy traffic demand and different layouts of freeway. Improvements in travel time, time-to-collision, emissions and fuel consumptions are accomplished using the VSL controller.

- Observer-based fault tolerance for VSL control:

The observer-based method was developed for sensor fault diagnosis in the VSL control system. Most existing studies on VSL control have a major limitation that is the assumption of reliable traffic measurements. Therefore, two observers were designed to provide reliable traffic state estimations in the case of permanent sensor faults. Moreover, the demand of large historical data in the traffic data imputation methods [23], [24], [25] is avoided using these two observers. Likelihood estimations were obtained to diagnose senor faults and reconfigure the VSL controller accordingly. Results show that consistent performance of VSL control is achieved despite permanent sensor faults.

- IMMP-based fault tolerance for VSL control:

The VSL control is further extended with fault tolerance to recurrent sensor faults using the IMMP-based method. The main contributions of the developed IMMP-based method for VSL control are made with respect to computational complexity and model

parameters compensation. With the fixed structure of IMM methods [89], [90], more models will be added and processed when more types of faults are considered. To reduce computational complexity, the adaptive model set was developed with variable structure of model set such that fewer models were running in parallel with consideration of different types of recurrent sensor faults. Another limitation from the existing studies [89]-[92] is that the accuracy of state estimations depends on the degree of a match between the model parameters and the extent of corresponding sensor failures. Therefore, the state covariance adaption and the pseudo-model set were developed to fairly compensate the discrepancies caused by improper model parameters. The fault-tolerant VSL system demonstrates consistent improvements of traffic conditions with VSL control near work zone areas in the case of different types of recurrent sensor faults.

This research also offers broader impacts and contributes to practical traffic systems and other congestion-related transportation problems. Since sensor faults commonly occur in transportation system in real world, the fault-tolerant VSL control in this thesis guides traffic engineers with practical VSL implementations. In addition, the framework of the fault-tolerant VSL control can easily be extended to other freeway traffic management systems such as ramp metering. Although this study mainly focuses on bottlenecks caused by freeway work zones, fault-tolerant VSL control can also be applied to mitigate the negative impacts of recurrent bottlenecks caused by other sources of congestion, such as congestion formation during rush hours. With aging infrastructure, increasing demand, and common sensors failures, the fault-tolerant VSL control in this

research contributes to an effective traffic management system with consistent improvements in mobility, safety, and sustainability.

5.2 Limitations and Future Work

The fault-tolerant VSL control was investigated in this thesis to mitigate the negative impacts of freeway work zones and address sensor faults. VSL control for work zones and fault tolerance are the main focuses of this research. Therefore, future work can be conducted to address limitations associated with VSL control and faults tolerance.

With respect to VSL control, speed limit schemes can be introduced by developing VSL control at network level. In this research, only one freeway work zone is considered. When multiple work zones exist, a queue forming near a work zone may propagate to another work zone upstream. Multiple work zones may have different traffic patterns than one isolated work zone. Centralized or decentralized VSL control should be considered to improve traffic conditions for multiple work zones. In addition, drivers may choose a different route when longer travel time is needed to traverse work zone area. Combining a route guidance algorithm with VSL control may contribute to traffic management of not only freeways but also arterial roads and local roads.

Impacts of configurations of work zones needs to be further studied in the design of VSL control. In this thesis, the number of lane closure is considered to obtain work zone capacity. However, work zone capacity can also be affected by factors such as activities inside work zones, locations of work zones, widths of work zones and lengths of work zones [93], [94]. The previous study of the authors [15] demonstrates less travel time can

be achieved using VSL control when connected work zones share road work information with travellers before they reach work zone bottlenecks. Potential improvements of VSL control may be obtained by taking advantage of abundant information about configurations of work zones.

The optimal locations of VSL signs remains an open problem, especially with the emerging of automated vehicles. Although the VSL control in this research shows robustness under different freeway layouts, the optimal VSL signs deployment is not studied. Conventional vehicles follow speed limits posted at fixed locations such as gantries or roadside equipment. Placing speed limit signs optimally can potentially make VSL control more effective. Besides, automated vehicles can change their speeds anywhere as long as they receive speed limit command. The flexibility of VSL deployment offered by automated vehicles may potentially alleviate congestion as automated vehicles can adjust their speeds timely without encountering the physical speed limit signs. However, it can be challenging to control the speeds of automated vehicles in mixed traffic flow. For example, when automated vehicles want to decelerate and follow a new lower speed limit command, the deceleration can be highly dangerous when conventional vehicles are still travelling according to a higher speed limit. Developing a VSL controller with consideration of VSL sign deployment in mixed traffic flow can be studied to achieve more effective traffic management.

The impacts of compliance rates, trucks and rubbernecking need to be further studied for practical implementations. This study assumes full compliance, for which strict and perhaps automated enforcement is required. Without a high compliance rate, the

performance of VSL control will likely be affected. Therefore, an efficient VSL controller under stochastic compliance needs to be studied in the future.

Another area for advancing this study is adding other types of vehicles such as heavy duty trucks, light duty trucks and motorcycles to the simulation environment of the study. In its current form, only passenger cars are used in the simulations. . With the presence of different types of microscopic models for different types of vehicles, a traffic flow model needs to be developed for VSL control that can accurately reproduce the traffic dynamics near work zone areas. The rubbernecking near work zone areas is also one of the factors causing traffic congestion. An effective control strategy needs to be designed to alleviate the impacts of work zone areas on the traffic flow in the opposite direction.

For future real-world implementation of the developed VSL control system, it is essential to ensure a safe and efficient VSL control strategy in practical applications. Thus, the pre-selected parameters associated with posted speed limits such as the minimum and maximum speed limits, speed limit increments, and minimum duration of the speed limit need to be further investigated in collaboration with traffic control departments. Since factors such as the severity of congestions, safety risks, and the attributes of work zones are highly related to the effectiveness of implementing VSL control, a decision support tool with consideration of these factors can be developed to guide decision makers in the deployment of VSL Control systems.

To advance the proposed fault tolerance method, actuator faults and probe sensor faults can also be investigated in future work. Specifically, actuator faults refers to faulty VSL signs whereas probe sensor faults refers to failures related to connected vehicles or

automated vehicles. In the case of faulty VSL signs caused by factors such as a power outage, the restriction of traffic flow would not be effective. Consequently, a queue can still form upstream a work zone and lead to capacity drop. Maintaining the performance of VSL control with the occurrence of faulty VSL signs needs to be studied. In addition, probe sensor faults occur when connected vehicles cannot send their speeds and/or locations accurately. Inaccurate measurements can be attributed to various factors such as large communication delay in probe sensors that can lead to a mismatch between measurements and actual traffic conditions. Faulty measurements from probe sensors can cause great deviations in traffic state estimations and potentially lead to VSL control failure. Addressing probe sensor faults also needs to be considered in future studies.

References

[1] Transport Canada, “Transportation in Canada,” Ottawa, ON, Canada. Rep. 15357E, 2016.

[2] Ontario Ministry of Transportation. “Published Plans and Annual Reports 2019-2020,” ON, Canada. Rep.2369-1972, Nov. 2019.

[3] M. Kontorinaki, A. Spiliopoulou, C. Roncoli, M. Papageorgiou, “First-order traffic flow models incorporating capacity drop: Overview and real-data validation,” *Transp. Res. Part B Methodol.*, vol.106, pp. 52-75, Dec. 2017.

[4] D. Schrank, B. Eisele, and T. Lomax, “2019 Urban Mobility Report,” The Texas A&M Transportation Institute. Aug. 2019.

[5] S. Chien, L. Zhao, and J. Singh, “Evaluating the Effectiveness of Traffic Diversion and Managed Lanes on Highway Work Zones,” CAIT, Piscataway, NJ, USA, Rep. CAIT-UTC-051, Apr. 2016.

[6] American Road&Transportation Builders Association, “Work Zone Fatal Crashes and Fatalities,” 2018. [Online]. Available: https://www.workzonesafety.org/crash-information/work-zone-fatal-crashes-fatalities.

[7] R. C. Carlson, I. Papamichail, and M. Papageorgiou, “Local feedback-based mainstream traffic flow control on motorways using variable speed limits,” *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1261–1276, Dec. 2011.

[8] G. R. Iordanidou, I. Papamichail, C. Roncoli, and M. Papageorgiou, "Feedback-Based Integrated Motorway Traffic Flow Control with Delay Balancing," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2319–2329, Sep. 2017.

[9] A. Hegyi and S. P. Hoogendoorn, "Dynamic speed limit control to resolve shock waves on freeways - Field test results of the SPECIALIST algorithm," in *Proc. IEEE ITSC*, Funchal, Portugal, 2010, pp. 519-524.

[10] Y. Han and S. Ahn, "Variable Speed Release (VSR): Speed Control to Increase Bottleneck Capacity," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 298–307, Jan. 2020.

[11] A. Talebpour, H. S. Mahmassani, and S. H. Hamdar, "Speed harmonization: Effectiveness evaluation under congested conditions," *Transp. Res. Rec.*, vol. 2391, no. 1, pp. 69-79, Jan. 2013.

[12] P. Lyu, Y. Lin, L. Wang, and X. Yang, "Variable speed limit control for delay and crash reductions at freeway work zone area," *J. Transp. Eng. Part A Syst.*, vol.143, no. 12, p. 04017062, Dec. 2017.

[13] G. S. Van De Weg, A. Hegyi, S. P. Hoogendoorn, and B. De Schutter, "Efficient Freeway MPC by Parameterization of ALINEA and a Speed-Limited Area," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 16–29, Jan. 2019.

[14] Y. Han, A. Hegyi, Y. Yuan, S. Hoogendoorn, M. Papageorgiou, and C. Roncoli, "Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits," *Transp. Res. Part C Emerg. Technol.*, vol. 77, pp. 405–420, Apr. 2017.

[15] S. Du, S. Razavi, and W. Genders, “Optimal variable speed limit control under connected work zone and connected vehicle environment, ” In *Proc., 34th Int. Symp. on Automation and Robotics in Construction*, Taibei, Taiwan, 2017, pp.997-1004.

[16] Y. Wu, H. Tan, L. Qin, and B. Ran, “Differential variable speed limits control for freeway recurrent bottlenecks via deep actor-critic algorithm,” *Transp. Res. Part C Emerg. Technol.*, vol. 117, p. 102649, Aug. 2020.

[17] Z. Li, P. Liu, C. Xu, H. Duan, and W. Wang, “Reinforcement Learning-Based Variable Speed Limit Control Strategy to Reduce Traffic Congestion at Freeway Recurrent Bottlenecks,” *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3204–3217, Nov. 2017.

[18] C. D. of Transportation, (2018) *Caltrans Performance Measurement System (PeMS).* [Online]. Available: http://pems.dot.ca.gov/.

[19] D. Zhou, Y. Zhao, Z. Wang, X. He, and M. Gao, “Review on Diagnosis Techniques for Intermittent Faults in Dynamic Systems,” *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 2337–2347, Mar. 2020.

[20] R. Rajagopal and P. Varaiya, “Health of California’s Loop Detector System,” Univ. of California, Berkeley, CA, USA, Rep. UCB-ITS-PRR-2007-13, Aug. 2007.

[21] X. Chen, Z. He, and L. Sun, “A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation,” *Transp. Res. Part C Emerg. Technol.*, vol. 98, pp. 73–84, Jan. 2019.

[22] Y. Chen, Y. Lv, and F. Y. Wang, "Traffic Flow Imputation Using Parallel Data and Generative Adversarial Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1624–1630, Apr. 2020.

[23] Y. Bie, X. Wang, and T. Z. Qiu, "Online method to impute missing loop detector data for urban freeway traffic control," *Transp. Res. Rec.*, vol. 2593, no. 2593, pp. 37–46, Jan. 2016.

[24] I. Laña, I. (Iñaki) Olabarrieta, M. Vélez, and J. Del Ser, "On the imputation of missing data for road traffic forecasting: New insights and novel techniques," *Transp. Res. Part C Emerg. Technol.*, vol. 90, pp. 18–33, May 2018.

[25] Z. Cui, L. Lin, Z. Pu, and Y. Wang, "Graph Markov network for traffic forecasting with missing data," *Transp. Res. Part C Emerg. Technol.*, vol. 117, p.102671, Aug. 2020.

[26] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker. "Recent development and applications of SUMO—Simulation of urban mobility," *Int. J. Adv. Syst. Meas*, vol.5 no. (3–4), pp. 128–138, 2012.

[27] FHWA. "Work zone management program." fhwa.com https://ops.fhwa.dot.gov/wz/resources/facts_stats.htm (accessed Feb. 13, 2018).

[28] J. H. Banks, "The two-capacity phenomenon: Some theoretical issues," *Transp. Res. Rec.*, no.1320, pp.234–241, 1991.

[29] K. Chung, J. Rudjanakanoknad, and M.J. Cassidy, "Relation between traffic density and capacity drop at three freeway bottlenecks," *Transp. Res. Part B*, vol. 41, no. 1, pp. 82–95, Jan. 2017.

[30] F. L. Hall, V. F. Hurdle, and J. H. Banks, "Synthesis of recent work on the nature of speed-flow and flow-occupancy (or density) relationships on freeways," *Transp. Res. Rec.*, no. 1365, pp.12–18, 1992.

[31] J. Li and H. M. Zhang, "Modeling space–time inhomogeneities with the kinematic wave theory." *Transp. Res. Part B*, vol.54, pp.113–125, Aug. 2013.

[32] W. L. Jin, Q. J. Gan, and J. P. Lebacque, "A kinematic wave theory of capacity drop," *Transp. Res. Part B*, vol. 81, pp. 316–329, Nov. 2015.

[33] F. Soriguera, I. Martínez, M. Sala, and M. Menéndez, "Effects of low speed limits on freeway traffic flow," *Transp. Res. Part C*, vol. 77, pp. 257–274, Apr. 2017.

[34] A. Garcia-Castro and A. Monzon, "Homogenization effects of variable speed limits," *Transport and Telecommun. J.*, vol.15, no.2, pp. 130–143, Apr. 2014.

[35] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Comparison of local feedback controllers for the mainstream traffic flow on freeways using variable speed limits," *J. Intell. Transp. Syst.,* vol.17, no.4, pp.268–281, Nov. 2013.

[36] E. R. Müller, R. C. Carlson, W. Kraus, and M. Papageorgiou, "Microsimulation analysis of practical aspects of traffic control with variable speed limits." *IEEE Trans. Intell. Transp. Syst.*, vol.16, no.1, pp.512–523, Feb. 2015.

[37] K. Kim and M. J. Cassidy, "A capacity-increasing mechanism in freeway traffic," *Transp. Res. Part B*, vol.46, pp.1260–1272, Nov. 2012.

[38] D. Chen and S. Ahn, "Variable speed limit control for severe non-recurrent freeway bottlenecks," *Transp. Res. Part C*, vol.51, pp.210–230, Feb. 2015.

[39] Y. Han, D. Chen, and S. Ahn, "Variable speed limit control at fixed freeway

bottlenecks using connected vehicles," *Transp. Res. Part B*, vol.98, pp.113–134, Apr. 2017.

[40] H. Y. Jin, and W. L. Jin, "Control of a lane-drop bottleneck through variable speed limits," *Transp. Res. Part C*, vol.58, pp.568–584, Sep. 2015.

[41] H. Yang and H. Rakha, "Feedback control speed harmonization algorithm: Methodology and preliminary testing," *Transp. Res. Part C*, vol.81, pp.209–226, Aug. 2017.

[42] M. Papageorgiou, E. Kosmatopoulos, and I. Papamichail, "Effects of variable speed limits on motorway traffic flow," *Transp. Res. Rec.*, no.2047, pp.37–48. Jan. 2008.

[43] A. Hegyi, B. De Schutter, and J. Hellendoorn, "Optimal coordination of variable speed limits to suppress shock waves," *IEEE Trans. Intell. Transp. Syst.*, vol.6, no.1, pp.102–112, Mar. 2005.

[44] Md. Hadiuzzaman, T.Z. Qiu, and X.Y. Lu, "Variable speed limit control design for relieving congestion Caused by active bottlenecks," *J. Transp. Eng.*, vol.139, no.4, pp.358–370, Apr. 2013.

[45] A. Csikós, and B. Kulcsár, "Variable speed limit design based on mode dependent cell transmission model," *Transp. Res. Part C*, vol.85, pp.429–450, Dec. 2017.

[46] X. Yang, Y. Lu, and Y. Lin, "Optimal variable speed limit control system for freeway work zone operations." *J. Comput. Civ. Eng.*, vol.31, no.1, p.04016044, Jan. 2017.

[47] B. Khondaker, and L. Kattan, "Variable speed limit: A microscopic analysis in a connected vehicle environment," *Transp. Res. Part C*, vol.58, pp.146–159, Sep. 2015.

[48] Y. Zhang, and P. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Trans. Intell. Transp. Syst.*, vol.18, no.7, pp.1812–1823, Jul. 2017.

[49] L. Munoz, X. Sun, R. Horowitz, and L. Alvarez, "Traffic density estimation with the cell transmission model," In *Proc. Amer. Contr. Conf.*, Denver, USA, 2003, pp.3750-3755.

[50] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transp. Res. Part B*, vol.28, no. 4, pp.269-287, Aug. 1994.

[51] G. F. Newell, "A simplified theory of kinematic waves in highway traffic, part II: Queueing at freeway bottlenecks," *Transp. Res. Part B*, vol.27, no.4, pp.289–303, Aug. 1993.

[52] R. L. Bertini and M. T. Leal, "Empirical study of traffic features at a freeway lane drop," *J. Transp. Eng.*, vol.131, no.6, pp.397–407, Jun. 2005.

[53] W. Gao, Y. Wang, and A. Homaifa, "Discrete-time variable structure control systems," *IEEE Trans. Ind. Electron.*, vol.42, no.2, pp.117–122, Apr. 1995.

[54] S. Z. Sarpturk, Y. Istefanopulos, and O. Kaynak, "On the stability of discrete-time sliding mode control systems," *IEEE Trans. Automat. Control.*, vol.32, no.10, pp.930–932, Oct. 1987.

[55] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz, "SUMO-Simulation of urban MObility: An overview," In *Proc. 3rd Int. Conf. Adv. Syst. SIMUL*, IARIA, Barcelona, 2011.

[56] Minnesota Department of Transportation, "Speed limits in work zone guidelines," Saint Paul, MN, USA, Oct. 2014.

[57] S.I. Chien, and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *J. Transp. Eng.*, vol.129, no.6, pp.608–616, Nov. 2003.

[58] Government of Ontario. (1990). *Highway Traffic Act*, Toronto, Canada.

[59] A. R. A. Van der Horst, "Time-to-collision as a cue for decision making in braking," In *Proc. 6th ICTCT*, Salzburg, Austria, 1994, pp.1-12.

[60] D. Krajzewicz, M. Behrisch, P. Wagner, R. Luz, and M. Krumnow, "Second generation of pollutant emission models for SUMO." *Behrisch M., Weber M. (eds) Modeling Mobility with Open Data*, Springer, Cham, 2015, pp.203–221.

[61] J.R.D. Frejo, B. De Schutter, "SPERT: a speed limit strategy for recurrent traffic jams," *IEEE Trans. Intell. Transp. Syst.*, vol.20, no.2, pp.692–703, Feb. 2019.

[62] A. Mittal, E. Kim, H.S. Mahmassani, Z. Hong, "Predictive dynamic speed limit in a connected environment for a weather affected traffic network: A case study of Chicago," *Transp. Res. Rec.*,vol.2672, no.19, pp.13–24, Aug. 2018.

[63] M. Blanke, M. Kinnaert, J. Lunze, M. Staroswiecki, *Diagnosis and fault-tolerant control, 2nd ed.*, Springer, Berlin, 2006.

[64] B. Bae, H. Kim, H. Lim, Y. Liu, L.D. Han, and P.B. Freeze, "Missing data imputation for traffic flow speed using spatio-temporal cokriging," *Transp. Res. Part C Emerg. Technol.*, vol.88, pp.124–139, Mar. 2018.

[65] L. Li, J. Zhang, Y. Wang, and B. Ran, "Missing value imputation for traffic-related time series data based on a multi-view learning method," *IEEE Trans. Intell. Transp. Syst.*, vol.20, no.8, pp.2933–2943, Aug. 2019.

[66] K. Henrickson, Y. Zou, Y. Wang, "Flexible and robust method for missing loop detector data imputation," *Transp. Res. Rec.* vol.2527, no.1, pp.29–36, Jan. 2015.

[67] A. Liu, C. Li, W. Yue, and X. Zhou, "Real-time traffic prediction: a novel imputation optimization algorithm with missing data," In *Proc. IEEE Glob. Commun. Conf.* ,Abu Dhabi, United Arab Emirates, 2018, pp.1–7.

[68] P. Li, J. Jayasuriya, A. Hills, A.R. Mills, and V. Kadirkamanathan, "A likelihood ratio-based approach to bleed valve event detection in gas turbine engines," In *Proc. UKACC 12th Int. Conf. Control.*, Sheffield, UK, 2018, pp.408–413.

[69] B. Tabbache, M.E.H. Benbouzid, A. Kheloui, and J.M. Bourgeot, "Virtual-sensor-based maximum-likelihood voting approach for fault-tolerant control of electric vehicle powertrains," *IEEE Trans. Veh. Technol.*, vol.62, no.3, pp.1075–1083, Mar. 2013.

[70] F. Garramiola, J. del Olmo, J. Poza, P. Madina, and G. Almandoz, "Integral sensor fault detection and isolation for railway traction drive," *Sensors*., vol.18, no.5, p.1543, May 2018.

[71] S. Du and S. Razavi, “Variable speed limit for freeway work zone with capacity drop using discrete-time sliding mode control,” *J. Comput. Civ. Eng.*, vol.33, no.2, p.04019001, Mar. 2019.

[72] B.L. Smith, L. Qin, and R. Venkatanarayana, “Characterization of freeway capacity reduction resulting from traffic accidents,” *J. Transp. Eng.*, vol.129, no.4, pp.362–368, Jul. 2003.

[73] N. Bekiaris-liberis, C. Roncoli and M. Papageorgiou, “Highway traffic state estimation with mixed connected and conventional vehicles,” *IEEE Trans. Intell. Transp. Syst.*, vol.17, no.12, pp.3484–3497, Dec. 2016.

[74] W. Perruquetti and J.P. Barbot, *Sliding mode control in engineering*, Taylor & Francis, 2002.

[75] Y. Bar-Shalom, X.R. Li, T. Kirubarajan, *Estimation with applications to tracking and navigation: theory, algorithms, and software*. New York: Wiley, 2001.

[76] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya, “Detecting errors and imputing missing data for single-loop surveillance systems,” *Transp. Res. Rec.* vol.1855, no.1, pp.160–167, Jan. 2003.

[77] T.Z. Qiu, X.Y. Lu, A.H.F. Chow, and S.E. Shladover, “Estimation of freeway traffic density with loop detector and probe vehicle data,” *Transp. Res. Rec.*, vol.2178, no.1, pp.21–29, Jan. 2010.

[78] W. Genders and S. Razavi, “Impact of connected vehicle on work zone network safety through dynamic route guidance”, *J. Comput. Civ. Eng.*, vol.30, no.2, p.04015020, Mar. 2016.

[79] J. R. José, I. Papamichail, M. Papageorgiou, and B. De Schutter, "Macroscopic modeling of variable speed limits on freeways," *Transp. Res. Part C Emerg. Technol.*, vol. 100, pp. 15–33, Jan. 2019.

[80] A. Ghiasi, X. Li, and J. Ma, "A mixed traffic speed harmonization model with connected autonomous vehicles," *Transp. Res. Part C Emerg. Technol.*, vol. 104, pp. 210–233, Jul. 2019.

[81] A. A. Malikopoulos, S. Hong, B. B. Park, J. Lee, and S. Ryu, "Optimal Control for Speed Harmonization of Automated Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2405–2417, Jul. 2019.

[82] L. Li, B. Du, Y. Wang, L. Qin, and H. Tan, "Estimation of missing values in heterogeneous traffic data: Application of multimodal deep learning model," *Knowledge-Based Syst.*, vol. 194, p. 105592, Apr. 2020.

[83] Y. Chen, Y. Lv, and F. Y. Wang, "Traffic Flow Imputation Using Parallel Data and Generative Adversarial Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1624–1630, Apr. 2020.

[84] I. Laña, I. (Iñaki) Olabarrieta, M. Vélez, and J. Del Ser, "On the imputation of missing data for road traffic forecasting: New insights and novel techniques," *Transp. Res. Part C Emerg. Technol.*, vol. 90, pp. 18–33, May 2018.

[85] J. Tang, G. Zhang, Y. Wang, H. Wang, and F. Liu, "A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation," *Transp. Res. Part C Emerg. Technol.*, vol. 51, pp. 29–40, Feb. 2015.

[86] Z. Cui, L. Lin, Z. Pu, and Y. Wang, "Graph Markov network for traffic forecasting with missing data," *Transp. Res. Part C Emerg. Technol.*, vol. 117, p. 102671, Aug. 2020.

[87] S. Du and S. Razavi, "Fault-tolerant control of variable speed limits for freeway work zone using likelihood estimation," *Adv. Eng. Informatics*, vol. 45, p. 101133, Aug. 2020.

[88] D. Zhou, Y. Zhao, Z. Wang, X. He, and M. Gao, "Review on Diagnosis Techniques for Intermittent Faults in Dynamic Systems," *IEEE Trans. Ind. Electron.*, vol. 67, no. 3, pp. 2337–2347, Mar. 2020.

[89] N. Sadeghzadeh-Nokhodberiz and J. Poshtan, "Distributed Interacting Multiple Filters for Fault Diagnosis of Navigation Sensors in a Robotic System," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 47, no. 7, pp. 1383–1393, Jul. 2017.

[90] S. Zhao, B. Huang, and F. Liu, "Detection and Diagnosis of Multiple Faults with Uncertain Modeling Parameters," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 5, pp. 1873–1881, Sep. 2017.

[91] J. Ru and X. R. Li, "Variable-structure multiple-model approach to fault detection, identification, and estimation," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 5, pp. 1029–1038, Sep. 2008.

[92] Z. Jin, Y. Hu, C. Li, and C. Sun, "Event-Triggered Fault Detection and Diagnosis for Networked Systems with Sensor and Actuator Faults," *IEEE Access*, vol. 7, pp. 95857–95866, Jul. 2019.

[93] H. Kevin, A. Kondyli, D. Arguea, L. Elefteriadou, and F. Sullivan, "Estimation Of Freeway Work Zone Capacity Through Simulation And Field Data," *Transp. Res. Rec.*, vol.2130, no.1, pp.16–24, Jan. 2009.

[94] J. Wang and Q. Meng, "Estimating capacity and traffic delay in work zones: An overview," *Transp. Res. Part C Emerg. Technol.*,vol.35, pp.34-45, Oct. 2013.