SPEED DISTRIBUTION BASED APPROACH FOR SHOCKWAVE DETECTION IN A CONNECTED DRIVING ENVIRONMENT

A Thesis

by

CONNIE BETH XAVIER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Alireza Talebpour
Committee Members,	Yunlong Zhang
	Natarajan Gautam
Head of Department,	Robin Autenrieth

December 2017

Major Subject: Civil Engineering

Copyright 2017 Connie Beth Xavier

ABSTRACT

One of the key advantages of connectivity in highway environments is the possibility of shockwave detection at the onset of formation, which can provide more flexibility in mitigating congestion. Past research used data from radar guns and loop detectors to show that a rapid increase in speed variance could be an indicator of shockwave formation. This paper investigates the possibility of utilizing connected vehicles data and vehicle trajectory data to determine if any increase in speed variance over distances could be an indicator of shockwave formation. Moreover, the effects of limited information in a connected driving environment on shockwave detection based on speed variance were explored.

Two datasets were evaluated: I-5 Connected Vehicles dataset and NGSIM US 101 dataset. Six segments analyzed in the I-5 dataset showed that a jump in speed variance could detect congestion earlier than looking at average speed alone. The NGSIM US 101 scenarios of 100, 50 and 10 percent market penetration rates (MPRs) were analyzed assuming 100, 80, and 50 percent of speed data were received at each time step. For MPRs of 100 and 50 percent, speed variance was able to identify the six shockwaves in the dataset. The RMSE, calculated for various MPRs, showed an inverse relationship to MPR.

The impact of misinformation from potential cyberattacks or equipment malfunctions was also tested on the US 101 dataset. Speed variance was more robust than average speed when speeds were reported either higher or lower than actual speeds.

When speeds were falsely reported as a combination of higher and lower than actual speeds, variance continually increased, though a jump in variance was still an indication of shockwave formation. When incorrect speeds were reported for only a high variance interval by 1-5 mph and 5-10 mph, speed variance remained a strong indicator of congestion formation. Analyzing the US 101 dataset with larger distance intervals, by individual lanes, and by different lane aggregations improved variance based shockwave detection reliability. Shockwaves detected earlier and more reliably can delay shockwave propagation and further reduce negative impacts on safety, performance, and emissions.

DEDICATION

This thesis is dedicated to my family for their constant support, love, and friendship.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Talebpour, for supporting me and guiding me throughout the preparation of this research. I would also like to thank my committee members, Dr. Zhang and Dr, Gautam, for their willingness to provide feedback on this work. Thanks to Amr Elfar of Northwestern University as well for the collaboration efforts and research exchanges that helped form this research.

I also want to express my gratitude to Texas A&M University and the Texas A&M Engineering Experiment Station for the financial support of my graduate studies.

Thanks also goes to the Institute of Transportation Engineers student chapter at Texas A&M University for making my graduate study an enjoyable experience. Thanks to my mother and father for always encouraging and supporting me. Last but not least, I credit and thank God for every blessing and opportunity.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Professor Alireza Talebpour [advisor], Professor Yunlong Zhang of the Department of Civil Engineering and Professor Natarajan Gautam of the Department of Industrial and Systems Engineering.

All work for the thesis was completed by the student, under the advisement of Professor Alireza Talebpour of the Department of Civil Engineering.

Funding Sources

Graduate study was supported by the Diversity Fellowship from Texas A&M University and by a Research Assistantship with the Texas A&M Engineering Experiment Station.

NOMENCLATURE

BSM	Basic Safety Message
CV	Connected Vehicle
ITS	Intelligent Transportation Systems
MPR	Market Penetration Rate
PDR	Packet Delivery Rate
RMSE	Root-Mean-Square-Error

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The development of intelligent transportation systems (ITS) including wireless communication technologies like dedicated short-range and cellular communications has sparked research into the benefits of connected vehicle (CV) applications. One major focus of the U.S. Department of Transportation's ITS Strategic Plan 2015-2019 is the adoption and deployment of CV systems (*1*). Past research has indicated the benefits of connected vehicles to reduce energy consumption, decrease emissions, and improve safety and mobility (*2-5*). CV technology provides the opportunity to increase throughput and improve the stability of traffic flow on highways (*6*). In a CV network, individual vehicles can communicate with each other and with the infrastructure through vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications respectively. CVs send out basic safety messages (BSM) that include real-time updates of a vehicle's status including its location, speed, and acceleration from its on-board unit to other vehicles' on-board units or to the infrastructure's roadside unit and receive the same types of information in return. These technologies allow in-depth knowledge of the microscopic traffic flow state within a segment. The vehicle trajectory data gained from having a connected vehicle system can be used to detect, and therefore delay shockwave propagation in advance of an upcoming bottleneck or traffic jam.

Inductive loop detectors that are sawed into the pavement are one of the most common types of vehicle detection methods. However, with construction

implementation costs and the ability to only relay macroscopic traffic characteristics of speed and occupancy at specific locations, connected vehicle data is preferred as it allows for greater accuracy. Stephens et al. infused data from a connected vehicle platoon updated every 0.1 miles with data from in-pavement traffic sensors placed every half-mile (*7*). The average speeds calculated from the connected vehicle platoon showed the back of queue could be detected at an earlier time and milepost than using infrastructure-based data alone (*7*). From the speed distribution of individual vehicles captured with CV technologies, the speed variance of vehicles can also be calculated to detect the starting point of congestion formation possibly even more accurately and earlier than average speed alone.

Currently, several vehicle manufacturers have deployed connected vehicle technologies into their newest models (*8*). In addition, the National Highway Traffic Safety Administration has proposed a new safety standard to require all new light vehicles to have V2V capabilities (*9*). However, the current market penetration rate (MPR) of vehicles equipped with V2X communications is very low. Even with low market penetration rates, the traffic flow state of a system can still be assessed, though the limitations in accuracy must be explored.

With increasing market penetration of these connected vehicle technologies, there are also cybersecurity concerns. Lead cybersecurity experts and automotive stakeholders see cybersecurity as one of the biggest threats to vehicle manufacturers and find the current state of the industry to be unprepared for these potential attacks (*10; 11*). Cybersecurity has become a major issue due to increasing complexity from

approximately 100 million lines of code in a vehicle, to increasing connectivity causing vehicles to become more accessible, and to more personal data being available in car networks which can lead to identity theft (*12*). The on-board unit and roadside unit can also malfunction and send false data. Therefore, this research explores the effect of different MPRs and misinformation on speed variance used to detect shockwaves.

MOTIVATION AND CONTRIBUTION

A shockwave occurs when traffic transitions between a free-flow state and a congested state (*13*). It involves a discontinuity of flow or density which causes vehicles to change speeds quickly in a short amount of time (*14*). Vehicles approaching the shockwave have to all pass through the jammed area. This causes many negative impacts on overall system performance by increasing travel times through a segment, on safety from constant acceleration/deceleration behavior of vehicles, and on emissions from increased fuel consumption (*15*). The queue discharge rate from a shockwave is also lower than the free flow capacity of a freeway which affects performance (*15;16*). Overall, the impacts on safety, performance, and emissions show that detecting shockwaves at the onset of formation is an important step to mitigate its consequences which is the focus of this research. Early detection allows drivers to be more aware of the situation ahead, reduces speed differentials, slows down congestion formation, and smooths overall traffic flow. It allows for control strategies like variable speed limit systems to be initiated earlier. The shockwave detection method used in this research is also tested against different factors to test and improve its reliability.

PROBLEM STATEMENT

Previous studies used radar guns and loop detectors to show that a rapid increase in speed variance could be an indicator of shockwave formation. This study revisits the idea and utilizes connected vehicles data and vehicle trajectory data to determine if any increase in speed variance over distances could be an indicator of shockwave formation.

RESEARCH OBJECTIVES

This study has three major objectives that support the overall problem statement, as outlined below:

1. To determine whether speed variance calculated from connected vehicle data can be used to detect congestion formation,
2. To examine the effects of partial connectivity through various MPR and packet delivery rate (PDR) scenarios and of misinformation on speed variance and its ability to detect shockwave formation. PDR is defined as the percentage of speed data received at each time step. MPR is defined as the percentage of connected vehicles over the 15-minute period, and
3. To examine the effect of individual lanes and lane aggregations on speed variance used to detect shockwaves.

RESEARCH APPROACH

How objectives 1-3 above will be achieved in this study are outlined in steps 1-3 below, respectively:

1. Average speeds and speed variances calculated from a CV demonstration along I-5 in Seattle with 19 equipped vehicles will be analyzed for several segments that show shockwave formation to determine if there is a connection between speed variance and average speed. Average speed and speed variance will also be analyzed for a vehicle trajectory dataset on a segment of US 101 over a 15-minute period considering 100% MPR and 100% PDR.
2. The following additional percentages of MPR will be analyzed for the US 101 vehicle trajectory dataset: 50 and 10. For each MPR case, the following PDR percentages will be analyzed: 100, 80, and 50. Speed variance will be calculated for each MPR and PDR case. For every scenario besides a 100% MPR and 100% PDR, ten iterations will be completed. From these iterations, the lowest MPR and PDR case where high speed variance can be used to detect shockwave formation can be found. Speed data will also be manipulated and speed variances will be recalculated to examine the effects of misinformation.
3. The US 101 dataset will be analyzed by individual lanes and different lane aggregations to improve the use of speed variance as a shockwave detection method.

The main programs that will be used are Matlab, Microsoft Excel, and Google Earth. Speed variance will be calculated using (1) which is the standard sample variance equation. The v represents the average speed and the v_i represents ith speed in a particular time and distance interval. The n is the number of speeds recorded in a

particular time and distance interval. Average speeds will also be calculated for each interval.

$$s^2 = \frac{\Sigma (v_i - \overline{v})^2}{n - 1} \quad (1)$$

THESIS ORGANIZATION

This thesis consists of seven chapters. Chapter I provides an introduction to the research, the objectives, and the approach. Chapter II presents background information and relevant literature on the major research topics. Chapter III presents the two major datasets used in the study: I-5 and US 101. Chapter IV analyzes the data and presents the results of using speed variance as an indicator of congestion formation. Chapter V analyses the effects of partial connectivity and misinformation on using speed variance to detect shockwaves in the US 101 dataset. Chapter VI attempts to improve the shockwave detection method by performing different lane analyses. Chapter VII provides a summary of the work, lists major findings and limitations of this study and proposes future work.

CHAPTER II

LITERATURE REVIEW

The following literature review covers background information on the traffic flow theory of shockwave formation, on current shockwave detection methods and control strategies, and on cybersecurity problems associated with connected vehicle technology.

TRAFFIC FLOW THEORY

In the presence of a traffic jam due to an incident or bottleneck, traffic moves from a free flow state to a congested state to create a shockwave. Three different traffic flow states have been defined in literature: free flow, synchronized flow, and traffic jams (*17*). The synchronized flow state is characterized by the inability to pass other vehicles due to higher densities than in the free flow state (*17*). Observations of traffic flow on German highways have shown that there are common characteristics of phase transitions between the three traffic flow states (*18*). Flow and speed were calculated using loop detectors placed along a segment that had two off-ramps and one on-ramp (*18*). Traffic can transition from a free-flow state to either a synchronized or jam state depending on whether there is a development of a local perturbation, as caused by inflow at ramps, or random perturbation (*18; 19*). Phase transitions in traffic flow cause a sharp decrease in vehicle speeds (*18*). The speed variance in the synchronized flow state was observed to be lower than in the free flow state due to the higher density which restricts movement

(*18*). In addition, speed variance for different lanes in synchronized flow is similar unlike in free flow (*18; 20*). Stop-and-go phenomena are also created from phase transitions from free to synchronized to the jammed state (*21*). From the synchronized flow state, flow can transition into a higher density state where random local perturbations can expand and form a jam (*21*).

Previous studies have suggested using both macroscopic models (*22; 23*) and microscopic models (*24*) that the broadening of the speed distribution can indicate traffic breakdown. There is an increase in speed variance within the free flow state immediately before it turns into the jammed state (*25; 26*). There is a rapid increase in speed variance near the jamming point (*22*). Similar to fluid dynamics, the transition between free flow and higher density traffic states is marked by large fluctuations (*22*). R.D. Kuhne (*22*) plotted standard variation of speed against the equilibrium density and found that a sharp increase in standard variation was associated with the instabilities formed by jams and stop-start waves. Therefore, the speed variance of vehicles can be tracked across distance and time, and any sharp increases may suggest an upcoming jamming point. The main limitations of previous studies were their reliance on radar data and the ability to only detect breakdown at the location of the radar. CV technology provides the opportunity to track changes in speed across a segment for greater shockwave detection accuracy.

Past research had also found that shockwaves were initiated by and grow in amplitude with lane changing (*27; 28*). Shockwaves may be attributed not only to car-following behavior but also to lane changing (*27*). This research also explores the idea

by analyzing the multi-lane US 101 dataset for the connection between speed variance and the number of lane changes.

SHOCKWAVE DETECTION METHODS

The traditional method for shockwave detection is to track speed and density changes using a time-space graph that shows traffic states and a density-flow graph, or fundamental diagram, that shows the state's density and flow values (*13; 15; 29*). A shockwave is detected by checking whether the flow and speed values are below certain thresholds (*29*). Lighthill and Whitham (*30*) developed one of the original shockwave theories by relating traffic flow to kinematic waves. Richards (*31*) related traffic flow with a continuous "fluid" density and the speed-density diagram and used a shearing process for following shockwaves. One of the main limitations to traditional approaches is that the density values needed are hard to measure along freeways. Density is often estimated from loop detectors that provide occupancy values. Loop detectors provide aggregate information and are installed at set distance intervals. Therefore, detecting the start of a shockwave is often dependent on the number and spacing of the detectors.

Another approach to detecting bottlenecks and oscillations on freeways is the wavelet transformation method (*32*). Wavelet transform is a decomposition tool that extracts information from stationary time-series data (*32; 33*). Wavelet-based energy is used to identify the location of a bottleneck and queue formation by tracking changes in average speeds from loop detectors (*32*). Wavelet transformation was also used to analyze individual vehicle trajectories from NGSIM data to detect oscillations, or

shockwaves, by tracking deceleration behavior (*32*). Wavelet-based energy peaks are tracked as the shockwave propagates over time and space (*32*). Talebpour et al. (*34*) modeled acceleration behavior using a similar approach assuming vehicle trajectory data will be available with connected vehicles. One study applied a numerical algorithm to the NGSIM dataset and estimated the propagation speed to be about 11.4 mph for all shockwaves regardless of the speed before the segment becomes congested but did not look specifically into the start of the shockwave (*35*). This paper focuses on shockwave detection and presents a different approach to the wavelet transformation based method. Once a shockwave is detected, different control strategies can be put into place including a speed harmonization or variable speed limit system (*36-38*). The earlier a shockwave can be detected, the faster these systems can be initiated. Benefits of a variable speed limit system include improvements in safety (*39*), delaying or preventing traffic breakdown (*40*), and environmental benefits (*41*).

CYBERSECURITY

Connected vehicle systems are susceptible to cyberattacks, and robust security systems are required in their deployment (*42*). Applications on vehicle carry-in devices with possible virus and malware increase the chances of an attack to vehicle electronics (*43*). Cyberattacks can originate from the infrastructure including roadside units, security systems, and other vehicles (*44*). The V2V wireless interface is susceptible to spreading malware quickly if compromised (*45*). Attacks could potentially happen from anywhere which was made evident when researchers in Pittsburgh were able to take control over a

vehicle traveling in St. Louis by accessing the short-range wireless connections to Bluetooth units inside the vehicle to take over in-vehicle networks (*46*). Wireless attacks can pose major threats to passenger safety (*46*). Limited connectivity is one of the main difficulties of vehicle cybersecurity (*43*). Another high threat for connected vehicles are fake BSMs which can generate wrong driver reactions (*44*). Manipulating BSMs can include falsifying a vehicle's global position and speed. Therefore, this project looks at the cybersecurity issues of partial connectivity with each market penetration scenario and of misinformation. This study is not concerned with the origin or cause of these cyberattacks, but rather how the manipulation of messages sent between the infrastructure and the vehicle can affect the reliability of a speed distribution based approach for shockwave detection. Incorrect speed messages can be sent due to equipment malfunction as well. These issues will be analyzed by looking at different packet delivery rates (PDR) to simulate message delivery failures and by introducing false messages into the system to examine the effects of misinformation on speed variance.

CHAPTER III

DATASETS

This chapter introduces the two main datasets used for the analysis: the I-5 CV small-scale demonstration dataset and the US 101 vehicle trajectory dataset.

I-5 INFLO DATASET

In order to determine whether there was a connection between shockwave formation and high speed variance, data was analyzed from the Intelligent Network Flow Optimization (INFLO) prototype small-scale demonstration. In the INFLO demonstration, vehicles were equipped with CV systems to send BSMs (*7*). These vehicles drove the I-5 corridor in Washington in small platoons from Monday, January 12, 2015 to Friday, January 16, 2015 (*7*). BSMs were sent through on-board units in the vehicle using dedicated short-range communications when passing by roadside safety units (RSU) along the highway and using cellular radio otherwise (*7*). BSMs contained information such as a vehicle's location and speed (*7*). Vehicles drove in a loop on I-5, starting nine miles south of downtown and ending fourteen miles north of downtown before exiting and turning around (*7*). Points of high congestion surround the downtown area (*7*). Figure 1 shows a Google Earth overview of the I-5 study corridor in Seattle (*7*). Light green pins indicate the north (Edmond, WA) and south (Tukwila, WA) entrance and exit points of the demonstration route.

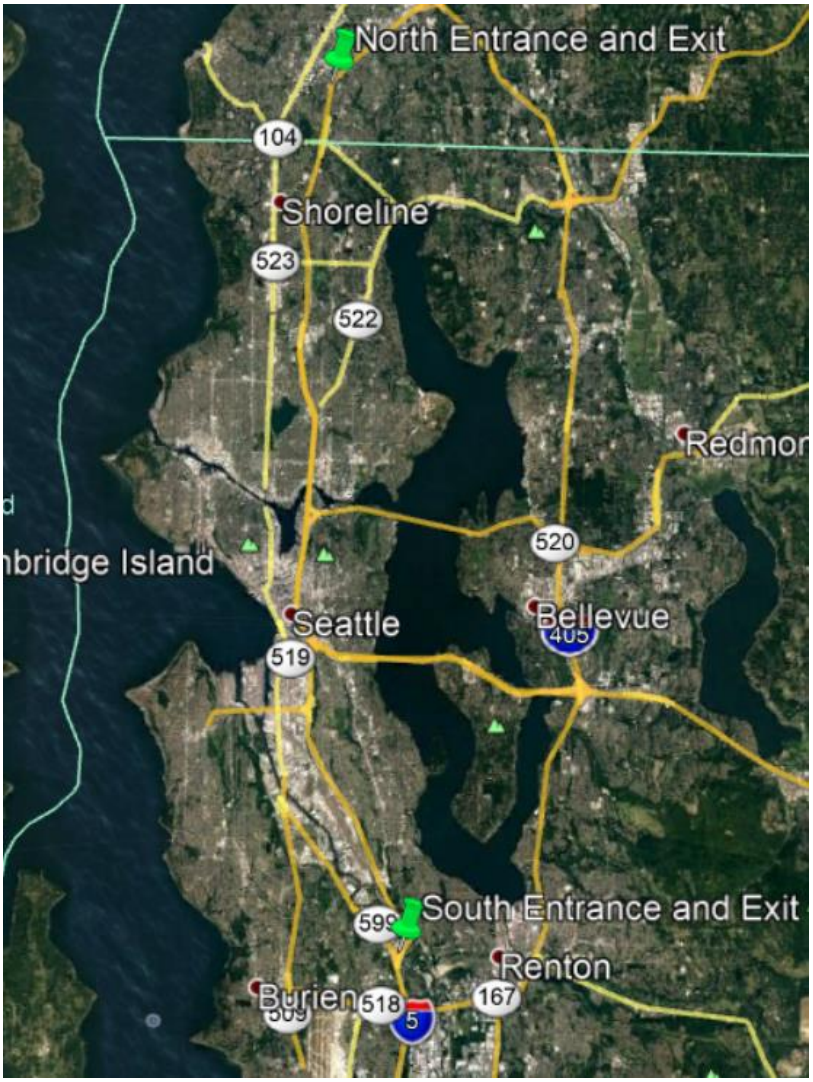


Figure 1: I-5 INFLO small-scale corridor overview (*7*)

Data from the INFLO demonstration collected on January 16, 2015 were analyzed for this study (*47*). There were 19 equipped vehicles that completed the demonstration on Friday morning (*7*). Data were recorded for both the northbound (NB) and southbound (SB) directions of I-5. NB vehicle data were recorded along I-5 from South 129th Street bridge to the 220th Street Southwest exit which is a segment length of about 22.5 miles. SB vehicle data were recorded along I-5 from about 212th Street

Southwest to the Interurban Avenue exit which is a segment length of about 23.5 miles. Over 30,000 data points were recorded for both directions combined. On Friday morning, vehicles were released around 30 seconds apart from each other (*7*). Drivers were told to drive normally, passing other vehicles and changing lanes according to conditions, and were not required to stay in the same order (*7*). The participants completed two loops along the test route (*7*).

US 101 DATASET

Vehicle trajectory data was analyzed from the Next Generation SIMulation (NGSIM) program (*48*). In the NGSIM program, vehicle trajectory data was collected on the southbound direction of US 101/Hollywood Freeway in Los Angeles, CA on Wednesday, June 15th, 2005, from 7:50 a.m. to 8:35 a.m. The study segment was between the on-ramp at Ventura Boulevard and the off-ramp at Cahuenga Boulevard, contained five main lanes, and was approximately 2,100 feet long. Data was collected through eight video cameras mounted on top of a nearby building and a computer program was used to convert the images into vehicle trajectory data. The final dataset includes each vehicle's geographical coordinates, speeds, lane positions, and accelerations that passed through the section at every one-tenth of a second. This project looks at the data collected from 7:50-8:05 a.m.

CHAPTER IV

SPEED VARIANCE AS AN INDICATOR OF CONGESTION FORMATION

This chapter analyzes shockwave formation in both the I-5 and US 101 datasets and explores differences in using average speeds versus speed variances to detect shockwaves. Matlab was used to create all average speed and speed variance plots.

I-5 DATA ANALYSIS AND RESULTS

Data Preprocessing

The dataset came with the timestamp of the BSM in one-minute intervals in Universal Coordinated Time from 1:47 p.m. to 5:33 p.m. The speed of the vehicle at the time of the BSM was reported in mph with at least four decimal places. The vehicle's location at the time of the BSM was reported by latitude and longitude in degrees. The mile marker (MM) location of the vehicle on the corridor at the time of the BSM was rounded to two decimal places. MM information was used to group the data into intervals. Data points with negative mile marker locations were filtered out. Vehicles that reported speeds of exactly 0 mph were also filtered from the dataset, so vehicles that may have stalled or stopped after taking an exit were excluded from the analysis. Heading was given on a range of 0 – 360. Data points were plotted on Google Earth with a heading label. From the plot, headings that fell within the SB and NB range were used to sort the data by direction. After plotting the dataset on Google Earth for the NB and SB directions separately, it was found that certain data points were scattered especially

at locations when the highway went through a tunnel. Sometimes the BSMs were from vehicles located on exit ramps.

After cleaning up the data, BSMs from vehicles at MM 156.5-164 and MM 169-179 were analyzed for the NB and SB directions respectively. The highest speed limit along the corridor is 60 mph, and there are segments with active variable speed limit signs. Average speeds above 45 mph were determined to be in an uncongested state. Average speeds between 25 and 45 mph were determined to be in a transition state, and average speeds below 25 mph were determined to be in a congested state.

Distance Interval Selection

CVs transmit messages every 0.1 seconds, but the data came with a timestamp in one-minute intervals, so average speed and speed variance were calculated every minute. Intervals of different distances were tested to determine which interval length would give the best results. Based on the table of average speeds, three segments that made a transition from uncongested to congested flow states were selected for the NB and SB directions respectively. The 0.2-mile segments did not show a high speed variance to indicate the start of congestion for all the NB cases. Similarly, the 0.25-mile segments did not show a high speed variance for all the NB cases. This could be due to the low number of BSMs reported in each interval. As market penetration rises, the number of BSMs in each time and MM interval will increase. Then smaller intervals can be used to locate the start of congestion even more accurately.

The 0.5-mile interval showed a jump in variance for all of the NB and SB cases. The detection of congestion with speed variances in 0.5-mile segments is similar to what

is currently used with loop detectors. In order to shorten the interval length to detect congestion at earlier mile markers, a rolling horizon approach updated every 0.25 miles was used. This way, the speed variances of 0.5-mile segments would detect the start of congestion 0.25 miles earlier. The optimal segment length will most likely vary depending on the number of CVs and accuracy of the data. For this study, the rolling horizon approach using 0.5-mile segments updated every 0.25 miles showed the jump in speed variance values needed to detect the start of congestion.

Results

Figure 2 shows the contour plot of average speed and speed variance for each time and mile marker interval of a segment in the NB direction. This segment took a long time to converge to the congested state compared to other segments. All intervals that did not have data were given an average speed of 60 mph and speed variance of 0. The 158.5-159 MM interval at 15:12 where the average speed first dipped below 45 mph also had a speed variance of over 100. This speed variance is 23 times the amount of the speed variance in the previous interval. The average speed variance for the free flow state of the segment was 3.4. The sharp increase in speed variance from interval to interval is an indication of congestion. After the two intervals of high speed variance over 100, the speed variance significantly drops. This is most likely due to the higher densities of vehicles in the congested state restricting speed differences. The segment reaches the congestion state (under 25 mph) starting at 15:16 in MM 160.5-161. Therefore, the high speed variance was able to detect congestion formation two miles before and four minutes earlier than using the average speed alone.

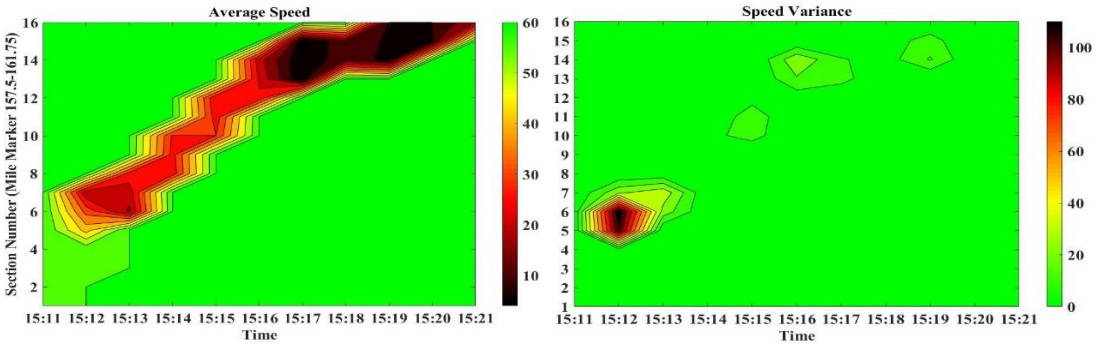


Figure 2: Example of long-time convergence to congested state (northbound)

Figure 3 shows the contour plot of average speed and speed variance for each time and mile marker interval of a segment in the SB direction with a short-time convergence to the congested state compared to other segments. Average speeds under 45 mph are reported early on in the segment. However, this does not necessarily mean congestion will form. A very high speed variance of 126 was reported at 15:43 in the MM 171.75-172.25 interval with an average speed of 45 mph. The segment reached the congestion state at 15:44 in MM 171.25-171.75. Therefore, a high speed variance was able to detect congestion 0.5 miles before and one minute earlier than using the average speed alone, although the time to congestion was shorter.

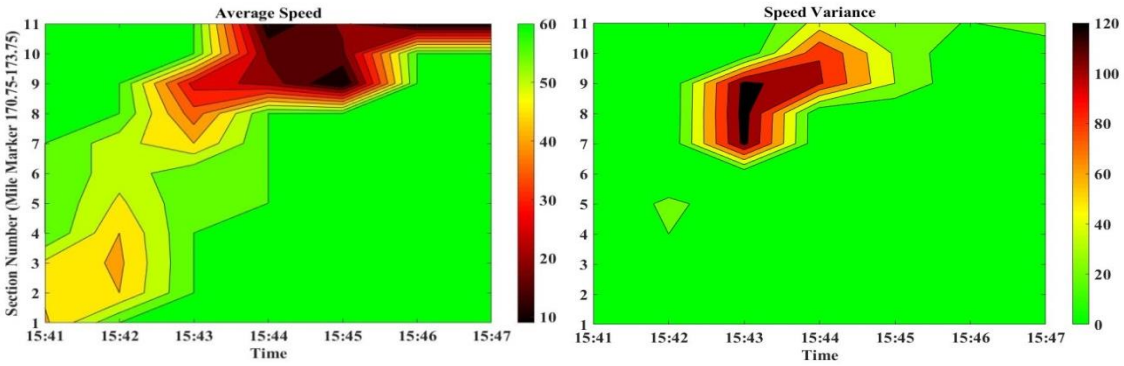
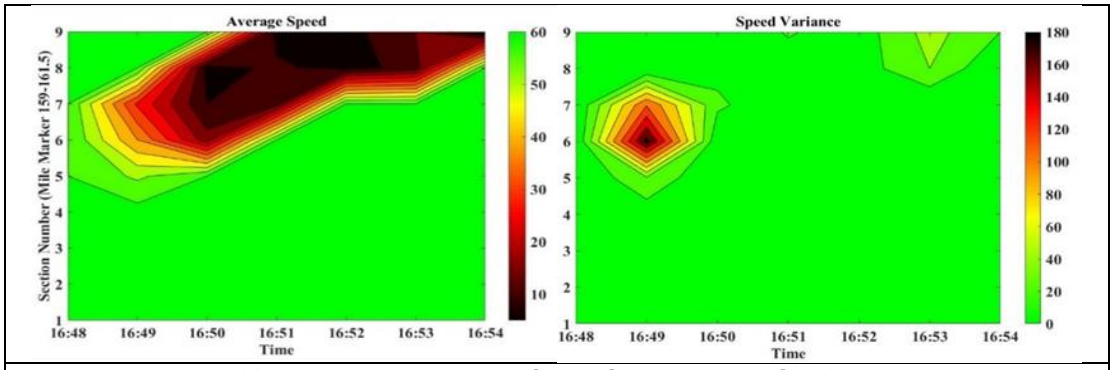
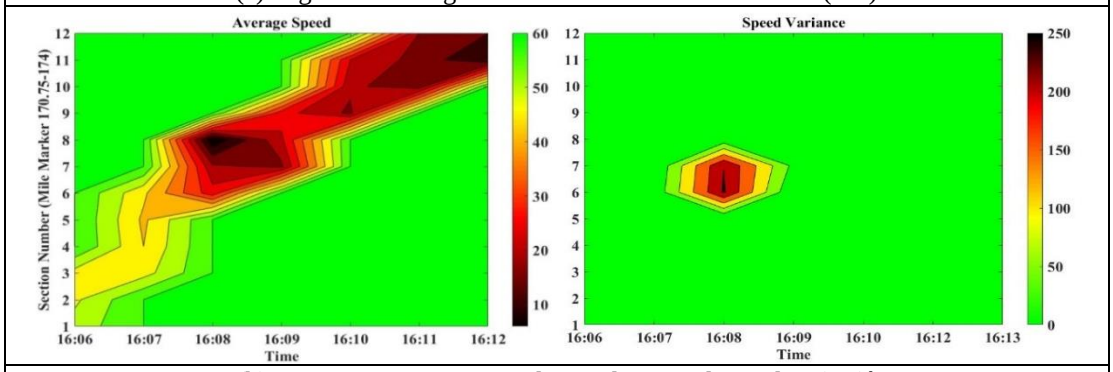


Figure 3: Example of short-time convergence to congested state (southbound)

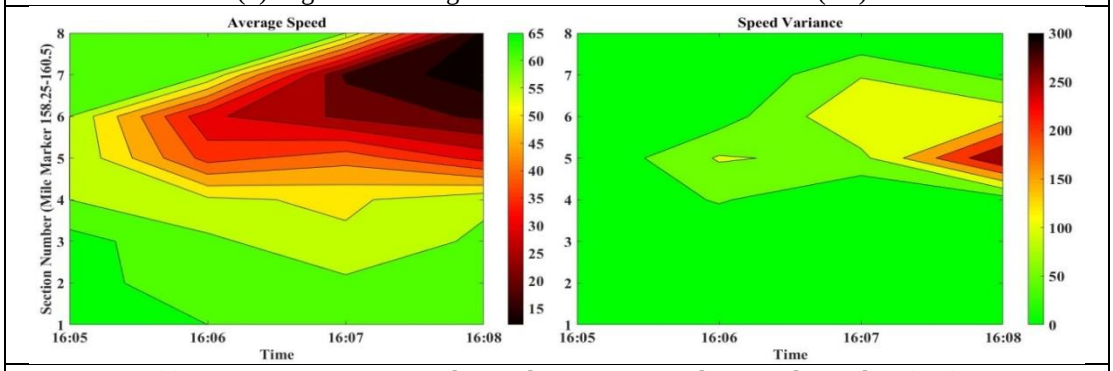
The four other segments analyzed can be seen in Figure 4. All six segments had a time and mile marker interval with a speed variance of over 100. For the dataset analyzed, a speed variance of over 100 indicates a high probability of leading to congestion. A high speed variance was able to detect congestion at least one minute and/or 0.25 miles earlier than when the average speed dropped below 25 mph indicating the congestion state. The exact value may differ depending on the market penetration rate. However, if the speed variance is tracked over time and distance, sharp increases in speed variance with an average speed in the transition range have shown to be strong indicators of congestion formation. An average speed that is in the transition range does not necessarily lead to the congested state every time. Speed variance calculated along with the average speed can be a more reliable indicator of the possibility of congestion formation.



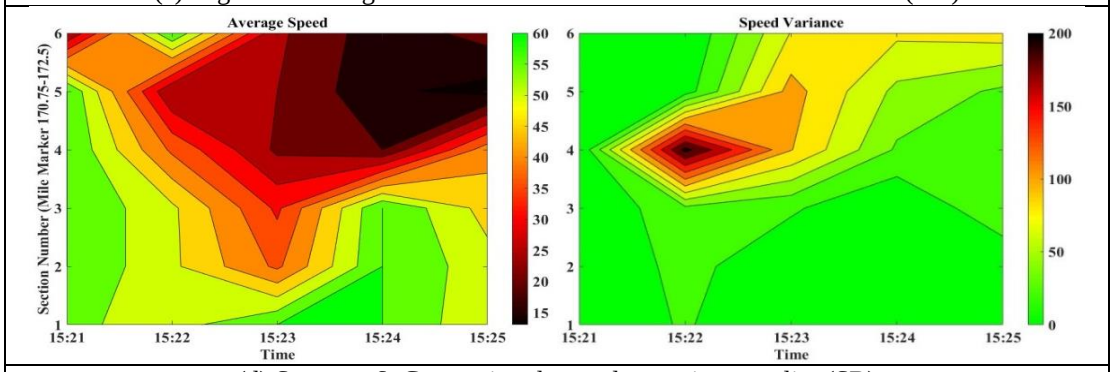
(a) Segment 3: Congestion detected one minute earlier (NB)



(b) Segment 4: Congestion detected 0.25 miles earlier (SB)`



(c) Segment 5: Congestion detected one minute and 0.25 miles earlier (NB)



(d) Segment 6: Congestion detected one minute earlier (SB)

Figure 4: Average speed vs. speed variance for segments along I-5

Since only a small number of the total flow of vehicles along I-5 were connected, the average speed calculated may not accurately reflect the actual state of the system. However, a connection was still found between the average speed and speed variance during shockwave formation. The following section looks at trajectory data from all vehicles within a segment, simulating a fully connected environment, to examine how speed variance changes in relation to average speed.

US 101 DATA ANALYSIS AND RESULTS

Data Preprocessing

Data was analyzed from 7:50 a.m. – 8:05 a.m. To preprocess the data, Montanino and Punzo's (*49*) multistep corrections were applied to the dataset but no outliers were found according to the acceleration threshold given in the corrections. In the first 100 ft. of the 2100 ft. segment, there was a lack of data most likely due to an error in recording all the vehicles. Intervals with an insufficient number of data points were plotted but not included in the analysis.

Speed Variance over a Fixed Point vs. a Distance

All previous research that had found a connection between speed variance and a transition into the jammed state had used aggregated data from fixed-location detectors to make their conclusions. In this study, connected vehicle data was aggregated from several locations over a distance. Aggregating speed data over one point versus a distance can produce different results. To ensure that the connection still holds between high speed variance and congestion formation, data from US 101 was used to find

average speed and speed variance as vehicles passed close (≤16 feet since this was the average length of a vehicle) to the midpoint of a distance interval and for all vehicles passing over the distance interval. Both versions show similar results, especially in average speeds. However, the speed variances calculated over a distance are higher than when calculated at a fixed point. This is expected, since speeds are more likely to vary over a segment than at a fixed point. Figure 5 shows a sample of the results of average speed and speed variance calculated over a distance versus a fixed point for US 101 over a 5-minute period from 7:50:20 a.m. – 7:55:20 a.m.

Distance and Time Interval Selection

As mentioned previously, there is an optimal time/distance interval for each dataset. Distance and time intervals were chosen empirically and logically as a way to break up the segment and track shockwave propagation. Lu et al. (*50*) suggested using 500 ft. sections for modeling fundamental diagrams using NGSIM vehicle trajectories. However, since the segment is only 2100 ft. long, this would be too long to be able to track the shockwave. Using a time interval of 20 seconds, average speeds and speed variances were calculated for the 15-minute period with distance intervals of 100, 150, and 200 feet. 150 feet showed the clearest results. Then, using a distance interval of 150 feet, average speed and speed variance were calculated for 10, 20, 30 and 60 second time intervals. Generally, the shockwaves in the data became clearer as the time interval shortened. Time and distance intervals of 20 seconds and 150 feet were chosen for further analyses. The plots of the other distance and time intervals can be found in the Appendix.

Time Interval (sec)	Midpoint of Distance Interval (ft)								
	100	300	500	700	900	1100	1300	1500	1700
0-20	23	26	27	31	31	32	33	35	40
20-40	23	27	30	32	32	31	31	34	37
40-60	24	26	30	32	29	27	27	35	37
60-80	20	23	27	23	20	24	30	35	38
80-100	19	22	17	18	26	30	32	34	36
100-120	17	17	21	27	31	33	35	36	37
120-140	17	21	26	31	32	35	36	38	39
140-160	23	25	28	32	33	35	38	40	42
160-180	25	29	28	30	32	34	37	39	40
180-200	26	29	29	29	30	30	34	39	41
200-220	28	28	28	30	24	24	31	36	39
220-240	27	25	25	20	19	24	28	36	38
240-260	21	20	17	15	19	28	30	35	38
260-280	14	16	17	22	24	27	31	35	38
280-300	15	23	24	27	27	27	30	36	38

(a) Average Speed Over Fixed Point

Time Interval (sec)	Distance Interval (ft)								
	0-200	200-400	400-600	600-800	800-1000	1000-1200	1200-1400	1400-1600	1600-1800
0-20	23	25	28	31	32	33	35	36	39
20-40	24	26	30	33	33	32	32	34	37
40-60	25	26	30	32	30	28	28	35	37
60-80	20	22	27	25	22	25	30	35	38
80-100	20	20	17	19	26	31	33	35	36
100-120	17	17	21	28	32	33	35	37	37
120-140	19	21	26	31	33	36	37	38	39
140-160	23	25	28	32	34	35	38	40	42
160-180	26	28	29	30	33	34	38	39	40
180-200	27	28	30	30	31	31	35	39	41
200-220	28	28	29	30	25	25	31	37	39
220-240	28	25	25	21	19	25	30	36	38
240-260	21	19	17	16	21	28	31	35	39
260-280	14	15	18	23	25	28	32	36	38
280-300	17	22	24	28	28	27	30	36	39

(b) Average Speed Over Distance

Time Interval (sec)	Midpoint of Distance Interval (ft)								
	100	300	500	700	900	1100	1300	1500	1700
0-20	6	13	13	10	19	24	20	16	8
20-40	8	7	14	11	5	24	48	15	13
40-60	10	17	8	14	44	68	83	15	7
60-80	37	21	15	78	137	97	61	14	8
80-100	18	20	120	147	60	26	19	10	10
100-120	58	93	72	22	7	9	9	13	16
120-140	55	13	20	19	7	8	17	9	9
140-160	7	6	7	9	13	12	14	11	10
160-180	6	4	8	24	10	11	10	8	5
180-200	13	5	7	4	8	16	9	6	7
200-220	21	5	6	10	20	40	14	9	18
220-240	18	7	11	60	51	31	64	11	11
240-260	27	47	70	69	48	20	34	20	11
260-280	33	42	35	16	8	10	15	18	24
280-300	39	8	7	10	9	12	15	7	7

(c) Speed Variance Over Fixed Point

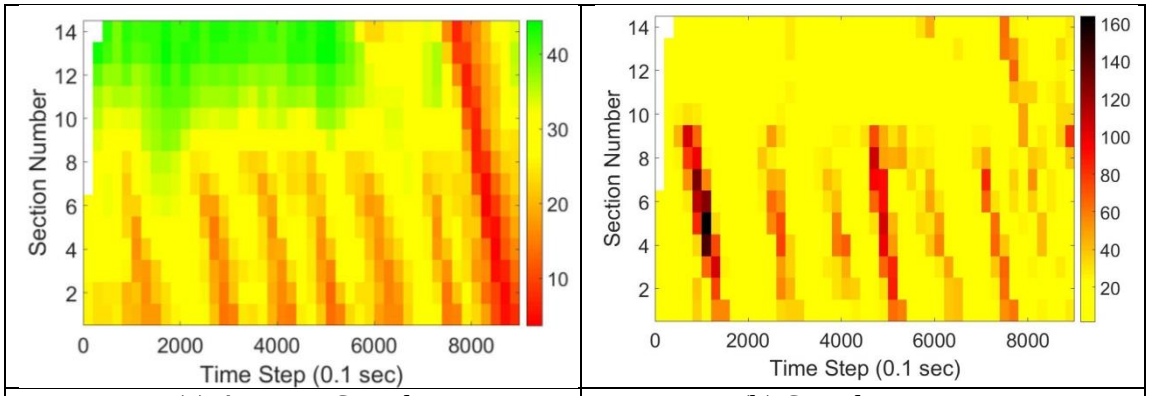
Time Interval (sec)	Distance Interval (ft)								
	0-200	200-400	400-600	600-800	800-1000	1000-1200	1200-1400	1400-1600	1600-1800
0-20	7	12	14	14	25	36	38	16	13
20-40	9	11	17	13	9	26	44	20	16
40-60	10	17	11	15	47	78	96	19	9
60-80	39	30	35	97	145	119	77	20	10
80-100	16	36	134	172	104	40	27	11	11
100-120	64	90	84	33	9	10	11	15	18
120-140	50	19	27	20	10	11	19	10	10
140-160	10	9	8	11	14	12	15	13	12
160-180	7	7	8	20	13	12	12	8	6
180-200	12	6	6	5	9	16	14	7	8
200-220	21	7	10	19	32	39	24	11	16
220-240	19	9	14	67	65	40	57	14	13
240-260	40	51	81	89	63	29	38	23	13
260-280	38	40	41	25	22	23	30	22	24
280-300	37	9	12	14	14	17	22	10	7

(d) Speed Variance Over Distance

Figure 5: Average speeds and speed variances calculated over a distance vs. a fixed point for US 101 from 7:50:20-7:55:20 a.m.

Average Speed and Speed Variance

Average speed and speed variance were calculated for the five main lanes of US 101 using time and distance intervals of 20 seconds and 150 feet from 7:50 a.m. – 8:05 a.m. Figure 6 shows the results. The section number is the number of each 150 feet section from 0 – 2100 feet and the time step is in units of 0.1 seconds from 7:50 a.m. The time steps plotted are the times at the midpoint of each 20 second interval from 7:50 a.m. – 8:05 a.m.



(a) Average Speed (b) Speed Variance

Figure 6: US 101 from 7:50 – 8:05 a.m. in 150 ft. and 20 sec. intervals

From the figure, seven distinct segments with congestion (average speeds < 20 mph) can be seen within the 15-minute period. These are most likely caused by the merging maneuvers of vehicles from the on- and off-ramps. By color-coding the intervals with high speed variance, a similar pattern is shown in the speed variance plot. The highest average speed calculated was 44 mph, so the transition range between uncongested and congested conditions was determined to be between 20 and 30 mph.

When looking at the values, the shockwaves are more distinct in the speed variance data than in the average speed data. Each shockwave is separated by intervals of speed variance values less than 12. After a few intervals of high speed variance when vehicles are transitioning to the congested or uncongested state, the speeds stabilize so the variance decreases. This is a contrast to the average speed data which shows a gradual decrease in speed values, so there is no clear line of transition between the states. The difference in ranges (uncongested, transition, and congested) is also much higher in the speed variance data than in the average speed data. Therefore, jumps in the speed variance are a more reliable indicator of a shockwave formation than looking solely for drops in average speed. Typically, sharp increases in speed variance were able to detect the shockwave in the same or one time/distance interval before average speeds were in the congested range.

SUMMARY

This chapter analyzed CV data collected from a demonstration along I-5 in Washington in both the NB and SB directions. A rolling horizon approach with 0.5-mile intervals updated every 0.25 miles was used to plot the data. A total of six segments analyzed showed that a speed variance of over 100 with an average speed in the transition range could accurately detect congestion formation at least one minute and/or 0.25 miles earlier than using the average speed alone. The exact value for high speed variance may differ depending on the market penetration rate and interval size but a sharp increase in speed variance is a strong indicator of congestion formation. A different interval size and

higher market penetration rate may detect shockwaves even earlier and more accurately. Average speeds and speed variances were calculated for a vehicle trajectory dataset of US 101 over a 15-minute period using 150 feet and 20 second intervals. From the 100% connectivity case, it was observed that an increase in speed variance was an indicator of congestion formation. This finding can be used to initiate speed harmonization applications including variable speed limit systems to increase throughput and delay shockwave propagation with a more uniform traffic flow. The earlier a shockwave can be detected, the faster the variable speed limit system can be implemented and the more effective it will be on resulting traffic operations.

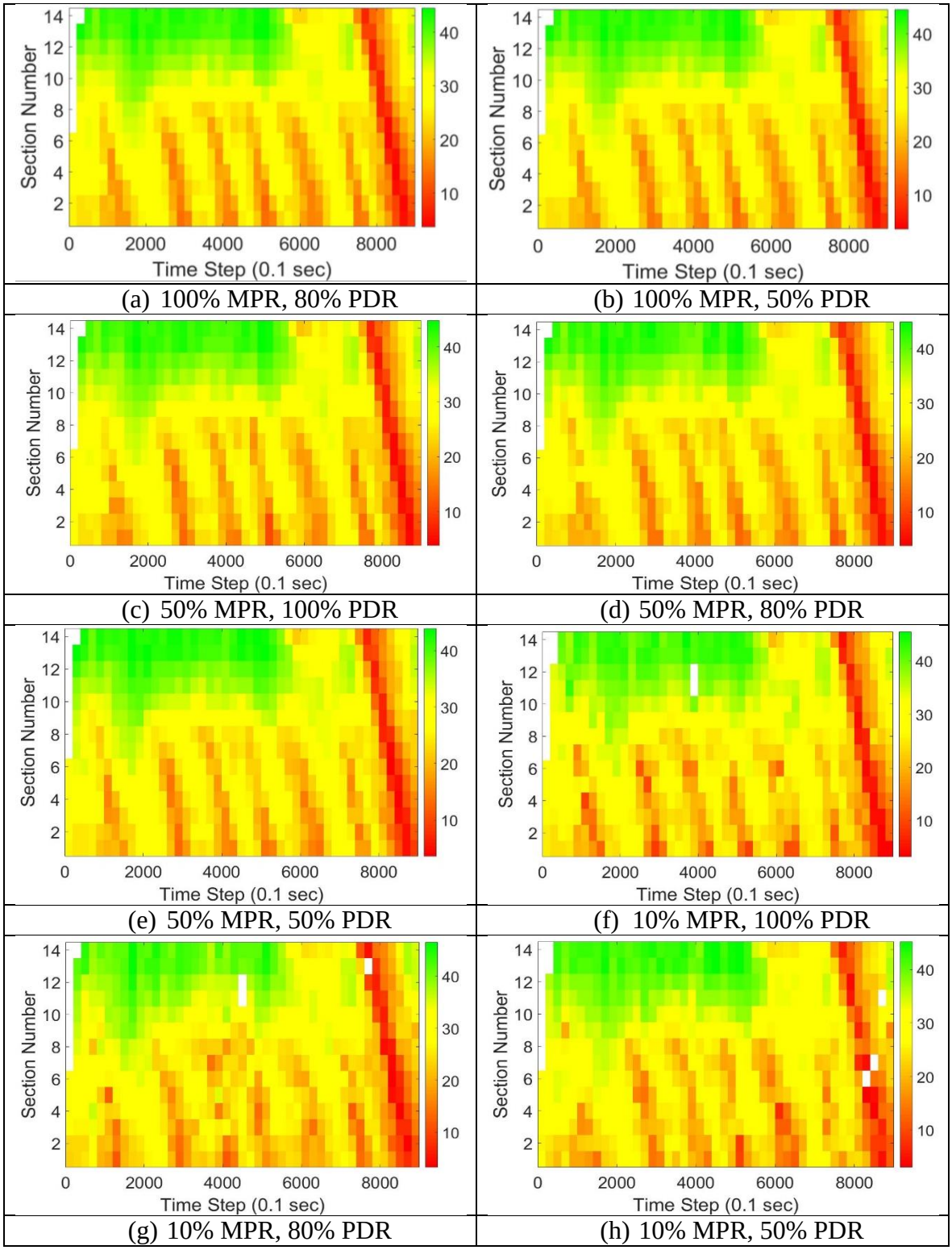
CHAPTER V

EFFECTS OF PARTIAL CONNECTIVITY AND MISINFORMATION ON SHOCKWAVE DETECTION ACCURACY

This chapter analyzes the effects of partial connectivity and misinformation on using speed variance for shockwave detection. This chapter also provides a comparison between the results of the I-5 segments and the lower connectivity cases of US 101.

PARTIAL CONNECTIVITY

For every MPR, the BSMs of only a percentage of the total number of vehicles that had passed through the segment over the 15-minute period were considered. The dataset came with a unique vehicle ID number for every vehicle that passed through the segment. Vehicles that were "connected" were chosen randomly. For every PDR, only a percentage of the BSMs of the "connected" vehicles were received at each time interval over the whole segment. This was done to simulate message delivery failures due to connectivity issues or cyberattacks. The BSMs that were received from the vehicles were randomly chosen and updated at each time interval. MPR rates of 100%, 50%, and 10% were analyzed along with PDR rates of 100%, 80%, and 50%. Figure 7 and Figure 8 show how average speed and speed variance respectively change as the connectivity decreases with each case for one iteration. Intervals with insufficient data points were color-coded white.



(a) 100% MPR, 80% PDR (b) 100% MPR, 50% PDR

(c) 50% MPR, 100% PDR (d) 50% MPR, 80% PDR

(e) 50% MPR, 50% PDR (f) 10% MPR, 100% PDR

(g) 10% MPR, 80% PDR (h) 10% MPR, 50% PDR

Figure 7: Average speed profiles from 7:50 – 8:05 a.m. on US 101

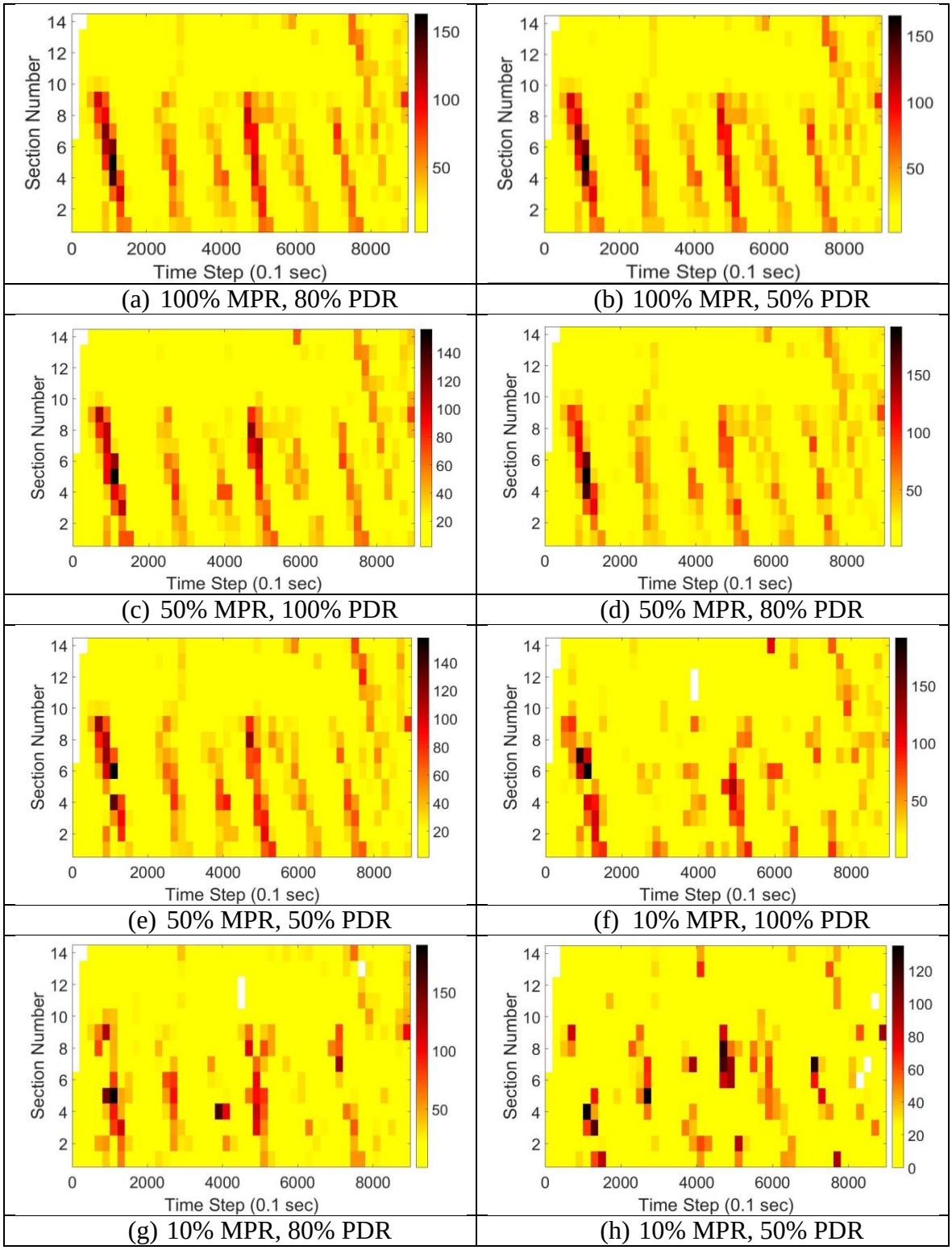


Figure 8: Speed variance profiles from 7:50 – 8:05 a.m. on US 101

The results of the lower connectivity cases depend on which BSMs are sent. Therefore, more iterations will be completed for each case. As shown by the figures, all cases of 100% MPR and 50% MPR in the speed variance data showed the six shockwaves clearly. In particular, the first shockwave is more clearly pronounced in the speed variance data than in the average speed data. However, for the lower connectivity cases, jumps in speed variance may not be a good indicator for shockwave formation than looking at average speed alone.

Table 1 shows the percent difference in the highest speed variance found in the range of each shockwave for lower connectivity cases from the base 100% MPR and 100% PDR case. Shockwaves with a maximum speed variance value that differs greater than 15% are in red. The number of red values increases as the connectivity level (found by multiplying percentages of MPR and PDR) decreases.

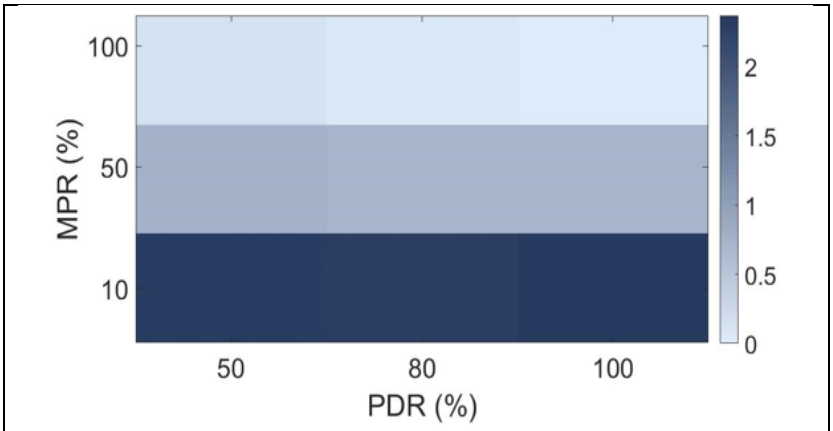
Table 1: Percent Difference from Highest Speed Variance found in 100% MPR/100% PDR Case for Each Shockwave

			Shockwave Number					
MPR (%)	PDR (%)	Connectivity Level (%)	1	2	3	4	5	6
100	80	80	-0.7%	0.3%	-0.1%	0.4%	-1.3%	-0.2%
100	50	50	0.8%	0.5%	0.9%	-0.8%	-2.6%	1.5%
50	100	50	-4.4%	4.6%	1.7%	11.6%	-10.1%	-14.6%
50	80	40	-3.9%	1.3%	**21.3%**	10.4%	-3.2%	-12.2%
50	50	25	**16.9%**	-13.1%	9.0%	-9.6%	**21.7%**	3.9%
10	100	10	**16.9%**	-10.5%	-3.1%	**21.0%**	**38.6%**	-3.9%
10	80	8	**16.2%**	**26.0%**	**135.9%**	10.4%	5.2%	**55.8%**
10	50	5	**-19.1%**	**77.0%**	**36.4%**	**16.6%**	8.2%	**49.9%**

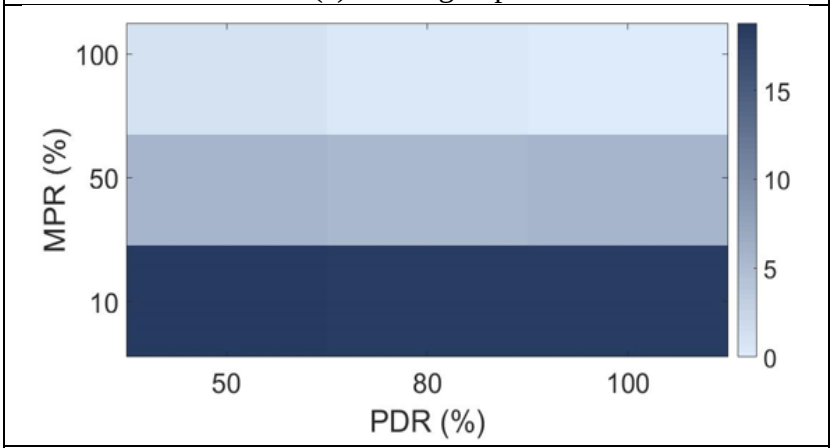
ROOT-MEAN-SQUARE-ERROR

The root-mean-square-errors (RMSE) of average speed and speed variance were found for each scenario with respect to the original 100% MPR and 100% PDR case. The average RMSE values found from ten iterations is shown in Figure 9. Distance/time intervals that did not record a value were ignored in the calculation. Both RMSE values of average speed and speed variance show similar patterns. In general, RMSE increases as connectivity decreases. However, the RMSE values for each MPR case are similar across PDR values and do not strictly increase or decrease with PDR. This is likely due to the randomness in the PDR calculation that is updated at every time interval. The difference in RMSE values between MPR cases is larger for the speed variance data than the average speed data. This could be the result of greater variability associated with the speed variance calculation.

To examine the relationship between RMSE and MPR more closely, the MPR was varied at 5% intervals and the RMSE value was calculated with respect to the 100% MPR/100% PDR case each time. The results are shown in Figure 10 for one iteration at each MPR interval. As the percent of connected vehicles decreases, the RMSE value increases. In particular, the RMSE value appears to rapidly increase below around 35% MPR. This could explain why speed variance was an unreliable indicator for shockwave formation in the 10% MPR scenarios.

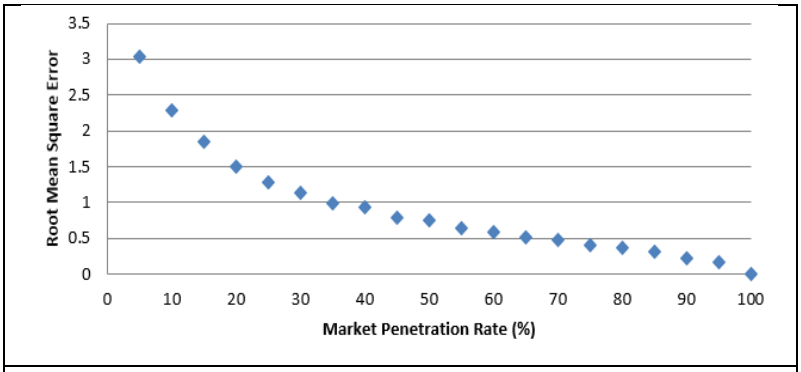


(a) Average Speeds

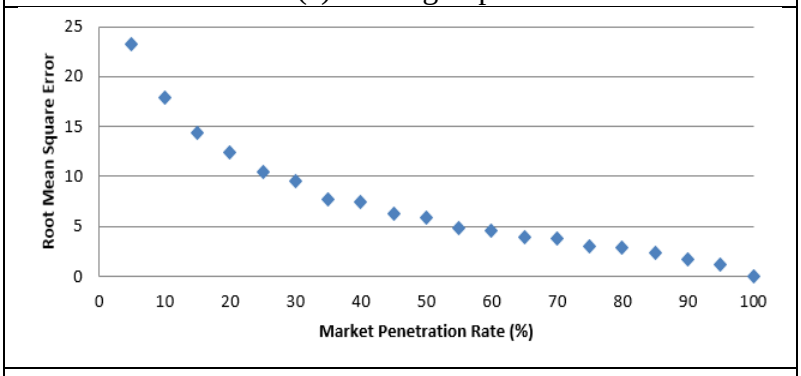


(b) Speed Variances

Figure 9: Average RMSE values from 10 iterations w.r.t 100% MPR/100% PDR case



(c) Average Speeds



(d) Speed Variances

Figure 10: RMSE vs. MPR from one iteration w.r.t 100% MPR/100% PDR case

DIFFERENCES BETWEEN I-5 AND US 101 RESULTS

The analysis of connected vehicle data from the I-5 small-scale INFLO demonstration hinted that speed variance could be an indicator of shockwave formation even at low connectivity levels. However, the US 101 data results, both empirically and from the RMSE values, show that this may not be true. In both datasets, vehicles were allowed to change lanes and move freely. One explanation could be due to the difference in distance intervals used in calculating the average speed and speed variance values. The speed variance values were calculated for the I-5 demonstration in half-mile long intervals. Speed variance was not a strong indicator of shockwave formation at smaller

distance intervals. Since the US 101 vehicle trajectory data section was only 2100 feet long, shorter intervals had to be used. Figure 11 shows a comparison between the speed variance profiles of US 101 using 150 ft. and 500 ft. intervals with 20 second time intervals at low connectivity levels. Overall, the shockwaves are clearer with larger distance intervals. The 500 ft. distance intervals indicate that speed variance can be used to detect shockwaves with at least 20% MPR while the 150 ft. intervals suggest the breakdown point to be between 30% and 40% MPR. While longer intervals may improve the reliability of shockwave detection, smaller intervals allow for more precision in detecting the starting point of congestion formation. Aggregating the data reduces noise in the data and makes it appear cleaner.

The RMSE value for the 10% MPR case with respect to the 100% MPR case was calculated using different distance intervals. Table 2 shows the average RMSE values of US 101 over ten iterations for each distance interval. The results clearly show that as the distance interval increases, the error decreases. This supports the empirical results shown in Figure 11. There is about a 27% decrease in error when using 500 ft. intervals than 100 ft. intervals. The difference in results between US 101 and I-5 could also be attributed to focusing on specific segments in the I-5 dataset that fit well for shockwave analysis.

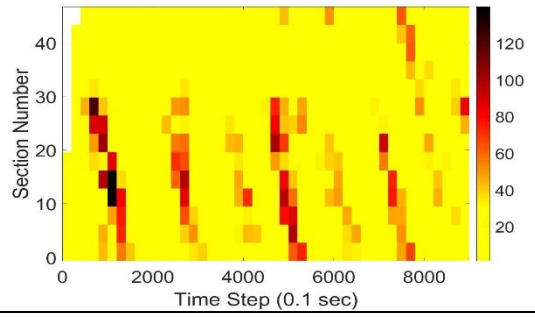
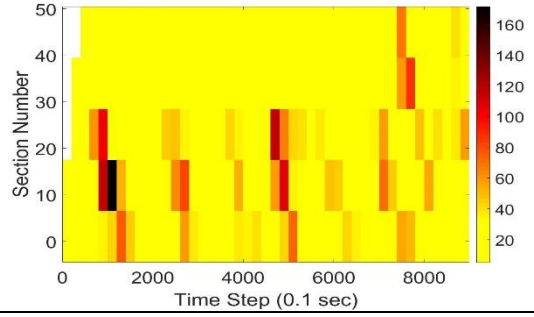
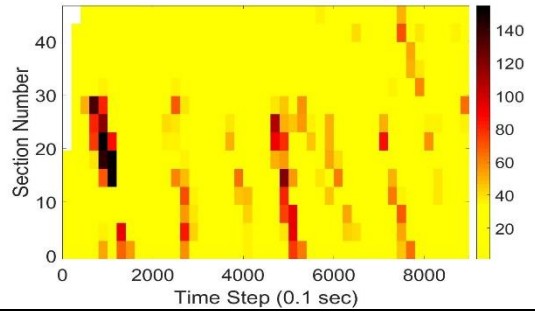
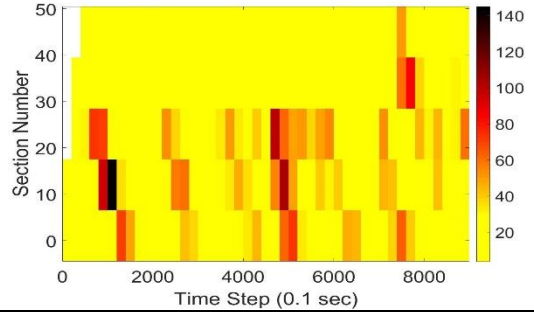
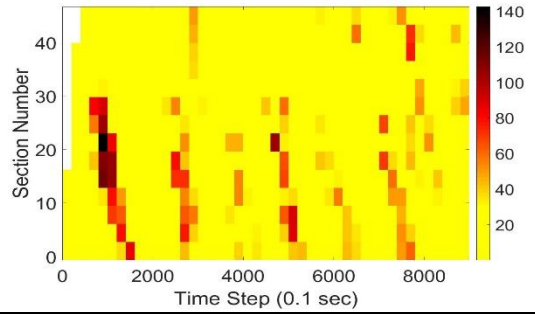
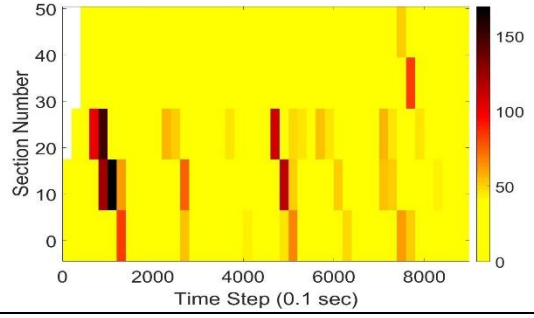
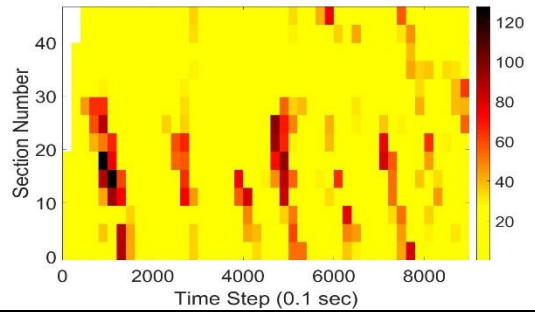
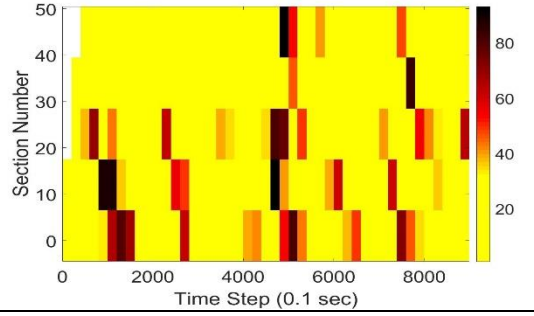
150-foot intervals	500-foot intervals
	
(a) 40% MPR	(b) 40% MPR
	
(c) 30% MPR	(d) 30% MPR
	
(e) 20% MPR	(f) 20% MPR
	
(g) 10% MPR	(h) 10% MPR

Figure 11: Speed variance profiles from 7:50 – 8:05 a.m. of US 101 in 150 ft. and 500 ft. intervals

Table 2: Average RMSE Values of 10% MPR Case from 100% MPR Case for Each Scenario over 10 Iterations of US 101 Data

Distance Intervals	Average Speed Variance RMSE
100 ft.	19.3
200 ft.	17.3
300 ft.	15.6
400 ft.	15.4
500 ft.	14.1

MISINFORMATION IMPACTS

In order to test the impacts of potential cyberattacks or equipment malfunctions in a connected driving environment, several possible scenarios were simulated and tested on the US 101 dataset. In the case of cyberattacks, there are several ways a system can be hacked into, so assumptions were made. The focus of this research is to examine the effect of falsely reported speeds on shockwave detection reliability. It was assumed that small changes in speed (<5 mph) would have minimal impact on the average speed and speed variance of a particular time and distance interval. Similarly, very large differences (>10 mph) could be picked out as outliers. Therefore, the effects of intermediate differences in individual speeds (5-10 mph) were examined.

Four different scenarios were tested on the first shockwave detected within the US 101 dataset plotted with 20 sec. and 150 ft. intervals. There was a pronounced increase in variance indicating shockwave formation for the first shockwave. Incorrect speeds were reported for a low variance interval in the uncongested range and the first high variance interval indicating shockwave formation. The low variance interval was chosen a time and distance step preceding the high variance interval since it was a backward propagating shockwave. For the first shockwave, the low variance interval

was from 7:50:40-7:51:00 a.m. over section 10 and had a variance value of 26.25 and an average speed value of 32.94 mph. The next interval with a sharp increase in variance was from 7:51:00-7:51:20 a.m. over section 9 and had a variance value of 103.26 and an average speed value of 26.51 mph. The first two scenarios tested assumed that a certain percentage of the speeds reported a 5-10 mph difference in actual speed. The falsely reported speeds were chosen randomly and the actual difference in speed was also chosen randomly. As mentioned previously, cyberattacks can generate from anywhere, and there can be an equipment malfunction on the receiving and sending end of a BSM. The overall trend in results is similar regardless of whether there is a malfunction with the vehicle's onboard unit affecting the sending of messages or if there is a malfunction with the roadside unit affecting the receiving of messages.

Figure 12 shows the effect of increasing a certain percentage of the speeds by 5-10 mph on the low variance and high variance (sharp increase in variance) interval's average speed and speed variance. Figure 13 shows the result of randomly decreasing actual speeds by 5-10 mph. When a certain percentage of speeds are falsely reported higher, the average speed increases. With high percentages of higher reported speeds, the high variance interval is no longer reporting a low average speed, so evidence of shockwave formation disappears. The opposite is true when the speeds are consistently reported lower than actual speeds causing the low variance interval to appear congested. However, the variance is not linearly increasing or decreasing with an increase or decrease in speeds but shows an inverted U relationship to the percentage of falsely reported speeds. The difference in variance between the low variance and high variance

interval remains large throughout. Therefore, a large difference in variance is still a good indicator of shockwave formation. This shows that speed variance is a more reliable measure for shockwave detection than tracking average speed alone.

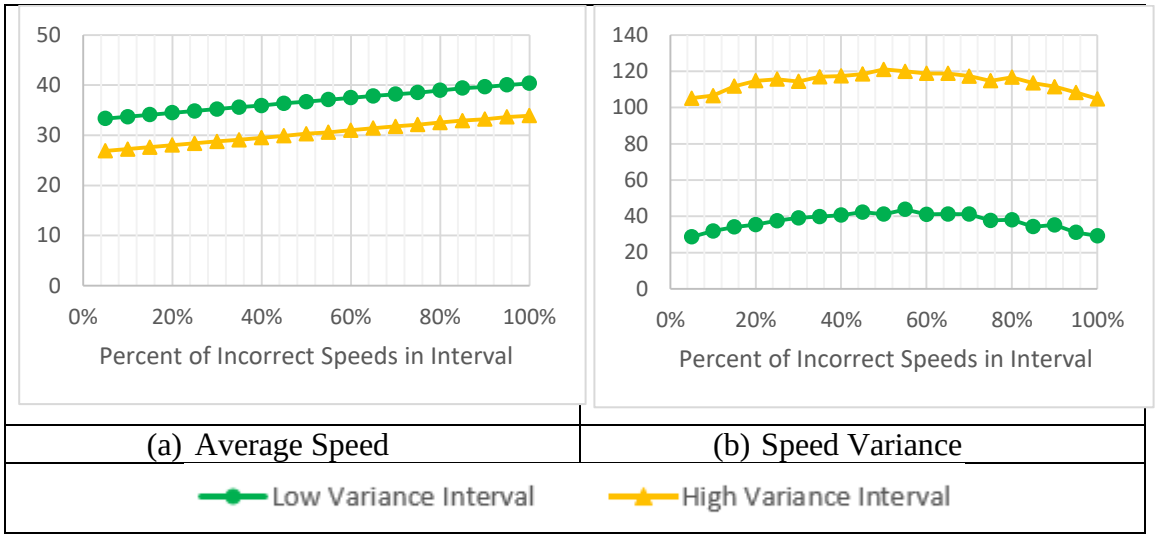


Figure 12: Effect of falsely reported higher speeds by 5-10 mph

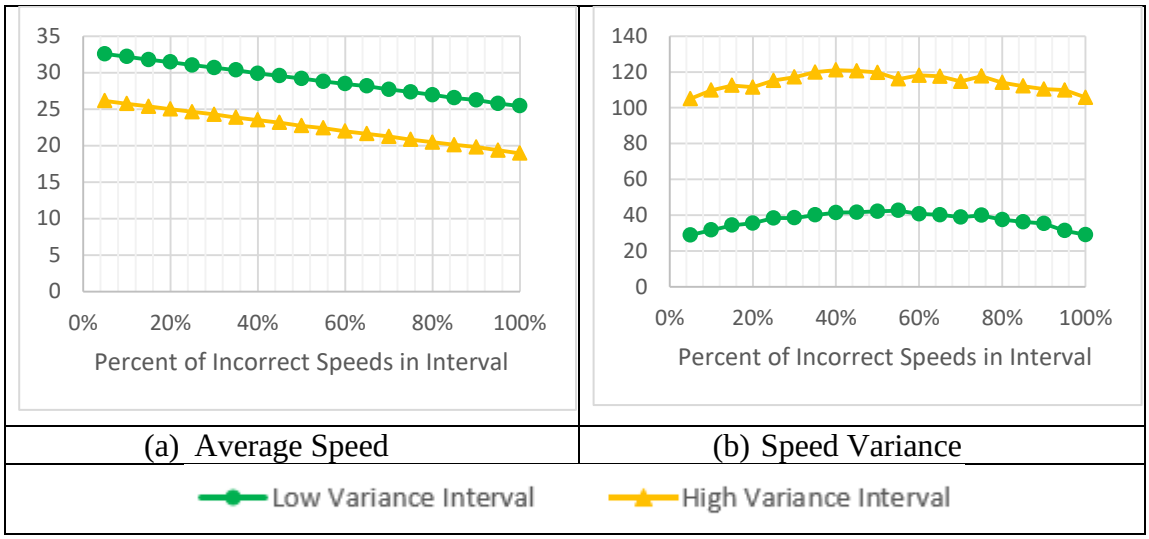


Figure 13: Effect of falsely reported lower speeds by 5-10 mph

Next, percentages of speeds were randomly either increased or decreased by 5-10 mph. Figure 14 shows the results of increasing or decreasing a certain percentage of the speeds by 5-10 mph on the low variance and high variance interval's average speed and speed variance. Due to the randomly chosen increase and decrease in speeds, the differences in speeds balanced out, so the average speed did not vary much with a higher number of falsely reported speeds. The speed variance results show a linearly increasing trend to the percentage of falsely reported speeds. Although the variance continually increases, the difference in speed variance between the low variance and high variance interval remains fairly constant. Therefore, the variance can be tracked across time and space, and a rapid increase in variance between two intervals could still be used as an indicator of shockwave formation.

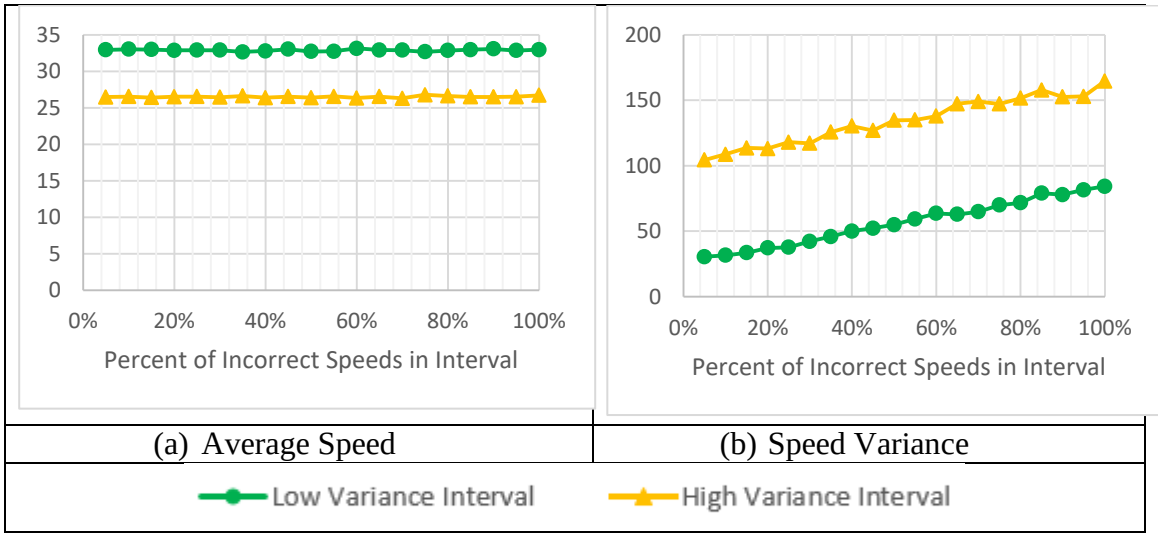


Figure 14: Effect of falsely reported higher and lower speeds by 5-10 mph

Lastly, speeds were manipulated for only the low variance interval and its recalculated speed variance was compared to the speed variance of the high variance interval. The difference was found by subtracting the recalculated low variance interval's speed variance from the unchanged high variance interval's speed variance value of 103.26. Figure 15 plots these differences for various speed manipulation ranges. Speeds were randomly increased and decreased within the specified speed ranges. The proposed shockwave detection method relies on a jump in speed variance values across time and space. If the difference remains significant, shockwaves can still be detected although speeds may have been falsely reported. The results show that when speeds are incorrectly reported within 1-5 mph, there is little effect on the speed variance difference between the two intervals. When speeds are incorrectly reported within 5-10 mph, the difference only significantly drops with very high percentages of incorrect speeds. When speeds are incorrectly reported within 10-15 mph, there reaches a point at 50% incorrect speeds where there is no difference between the speed variances of the two intervals. The same breakdown point drops to 25% when speeds are incorrectly reported within 15-20 mph. Negative differences indicate that the speed variance calculated for the low variance interval was higher than the speed variance for the high variance interval, making it appear as though the shockwave occurred earlier.

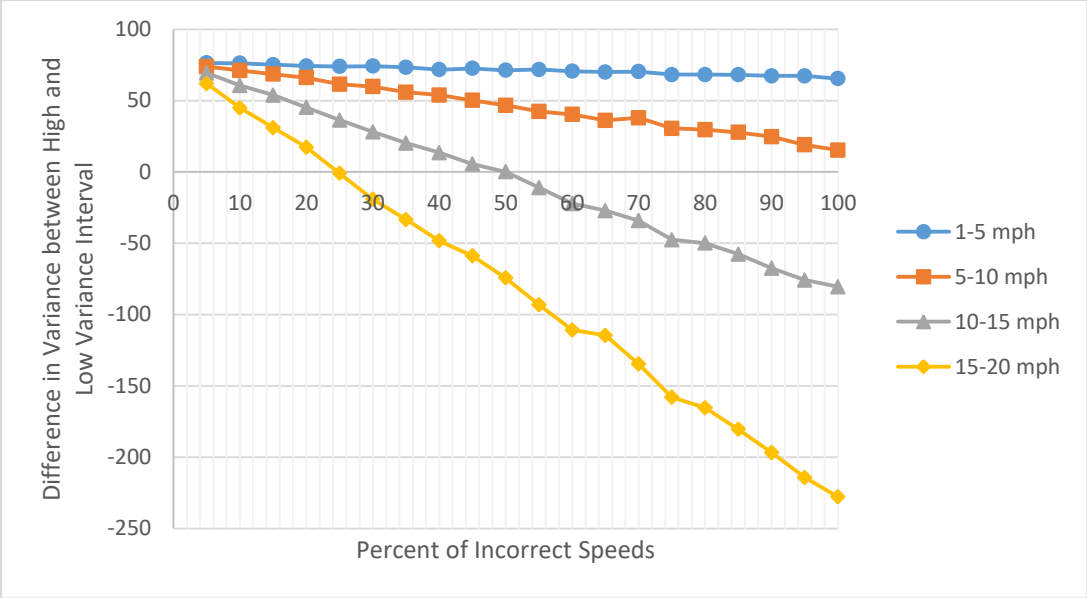


Figure 15: Speed variance differences between a low variance and high variance interval across various speed manipulation ranges

SUMMARY

Average speeds and speed variances were calculated for a vehicle trajectory dataset of US 101 over a 15-minute period for different MPR and PDR scenarios using 150 feet and 20 second intervals. All cases of 100% MPR and 50% MPR showed the six shockwaves in the dataset clearly. RMSE values with respect to the 100% connectivity case were averaged over ten iterations for the lower connectivity cases. In general, RMSE increases as connectivity decreases. However, the RMSE values for each MPR case are similar across PDR values and do not strictly increase or decrease with PDR. MPR was varied at 5% intervals. RMSE values were calculated from one iteration and showed that the error increases greatly below around 35% MPR. The difference in results between the I-5 and US 101 dataset may be attributed to the difference in length

of the segments. Plots of US 101 data using 500 ft. intervals showed the six shockwaves clearer in lower connectivity cases than in the plots made with 150 ft. intervals. The RMSE error decreased as the distance interval size increased from 100 ft. to 500 ft.

The effect of different potential cyberattacks and connectivity failures were tested on one shockwave in the US 101 dataset. The scenarios of falsely increasing, decreasing, and both increasing and decreasing actual speeds by 5-10 mph were tested for an interval with low variance and an interval with high variance. Speed variance was more robust than average speed to changes in speeds when speeds were reported to be either strictly increasing or decreasing. When speeds were both increasing and decreasing, average speeds remained steady while variance continually increased. However, the difference in variance between the two intervals remained steady with an increasing percentage of incorrect speeds which supports the idea of tracking changes in variance to detect shockwave formation. When speeds were incorrectly reported for only one of the intervals, speed variance remained a strong indicator of congestion formation when speeds were incorrectly reported by 1-5 mph and 5-10 mph for all percentages of incorrect speeds.

CHAPTER VI

LANE-BY-LANE AND LANE CHANGE ANALYSIS

This chapter further analyzes the shockwaves in the US 101 dataset and attempts to improve the speed distribution based shockwave detection method by analyzing individual lanes, different lane aggregations, and the relationship to the number of lane changes.

LANE-BY-LANE ANALYSIS

To improve the accuracy of using speed variance for shockwave detection, speed variances were calculated for each lane separately. Figure 16 shows the speed variance profiles by lane for US 101. Lane 5 is the leftmost lane and lane 1 is the rightmost lane by the auxiliary lane. The results show that the shockwaves are only clearly seen in the plot of lane 1. This is to be expected since the shockwaves are being caused by lane-changing maneuvers from the on-ramp and off-ramp. Analyzing each lane individually can narrow down which lanes are most affected and further pinpoint the cause of the shockwave. This is not as clear in the average speed profiles by lane, which show shockwave formation for several lanes. These speed profiles are located in the Appendix.

Figure 17 shows the speed variance profiles by different lane aggregations. Shockwaves 1 and 4 can be clearly seen in all graphs. Figure 17(c) appears to be the clearest and to have the least amount of noise among the four figures.

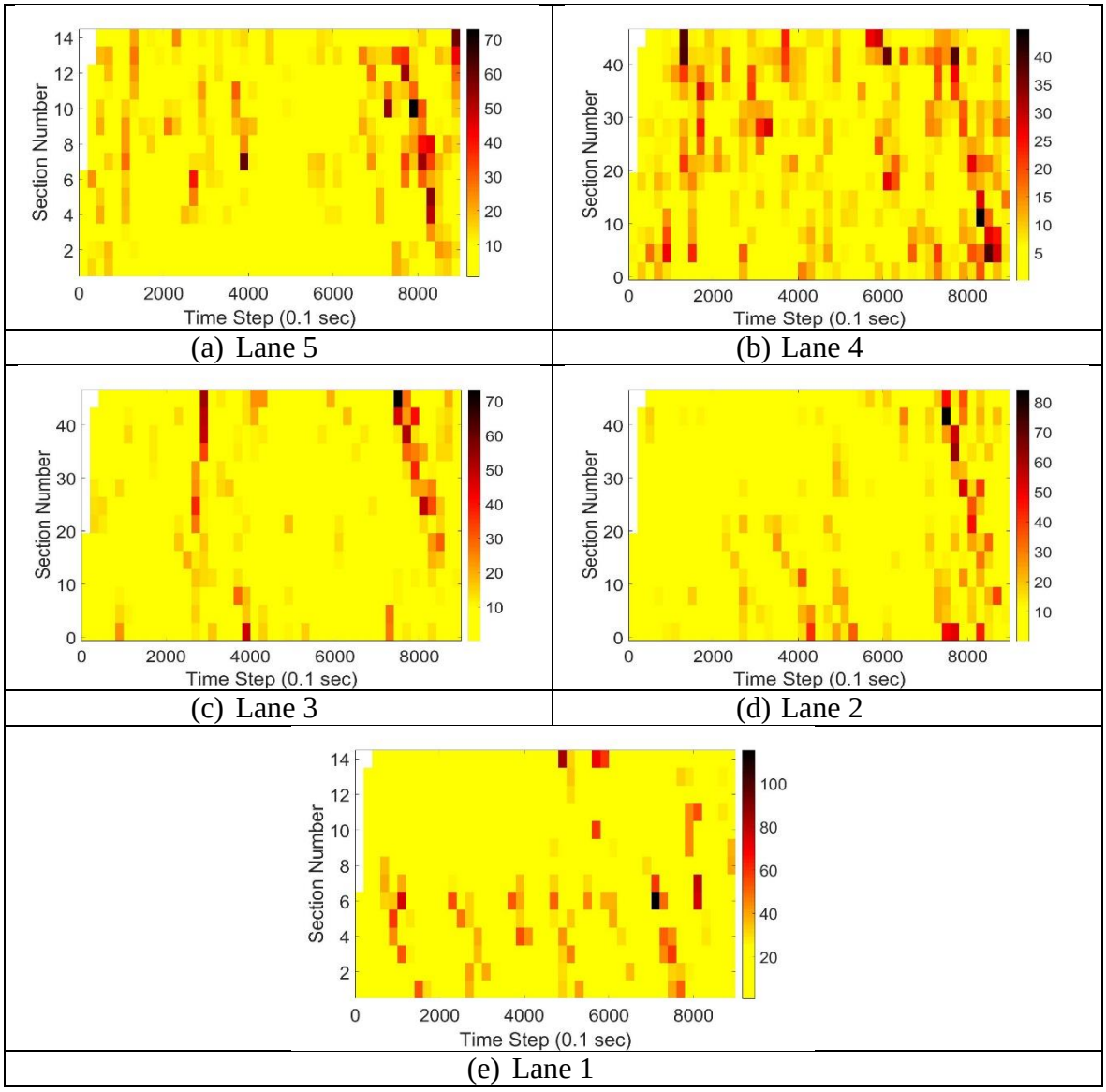
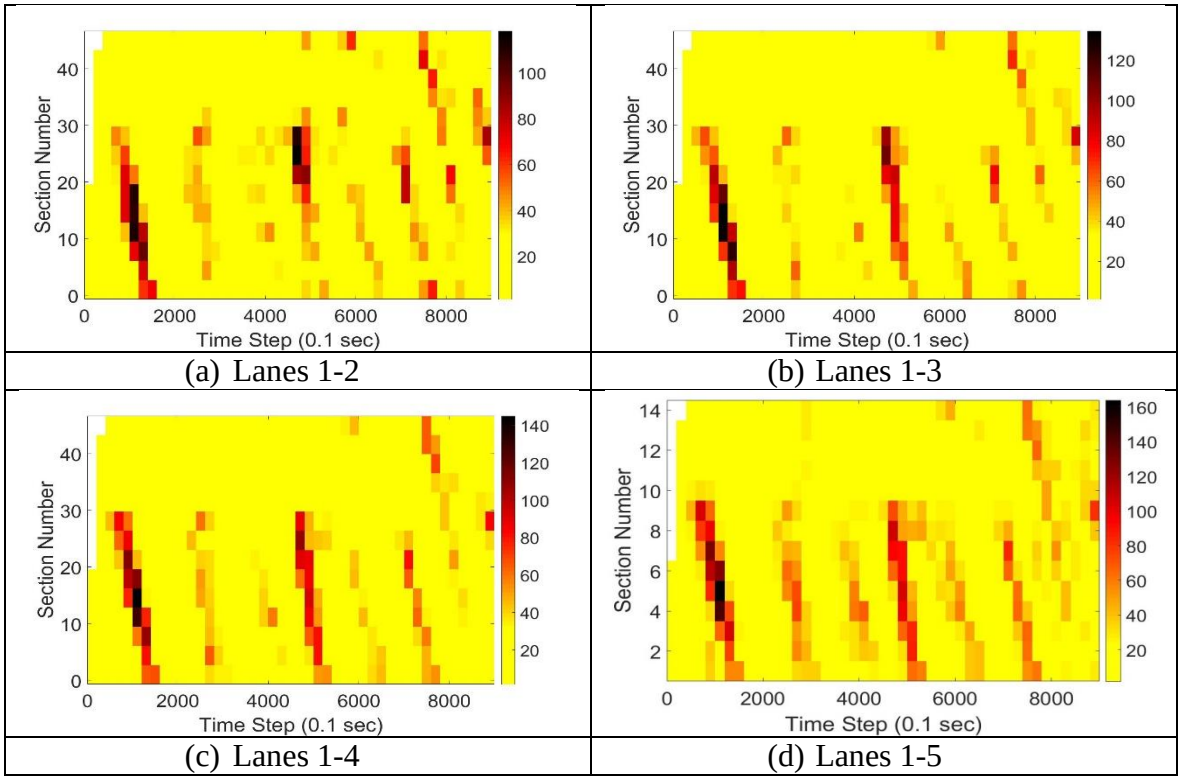


Figure 16: Speed variance profiles from 7:50 – 8:05 a.m. on US 101 by lane



(a) Lanes 1-2 (b) Lanes 1-3

(c) Lanes 1-4 (d) Lanes 1-5

Figure 17: Speed variance profiles from 7:50 – 8:05 a.m. on US 101 by different lane aggregations

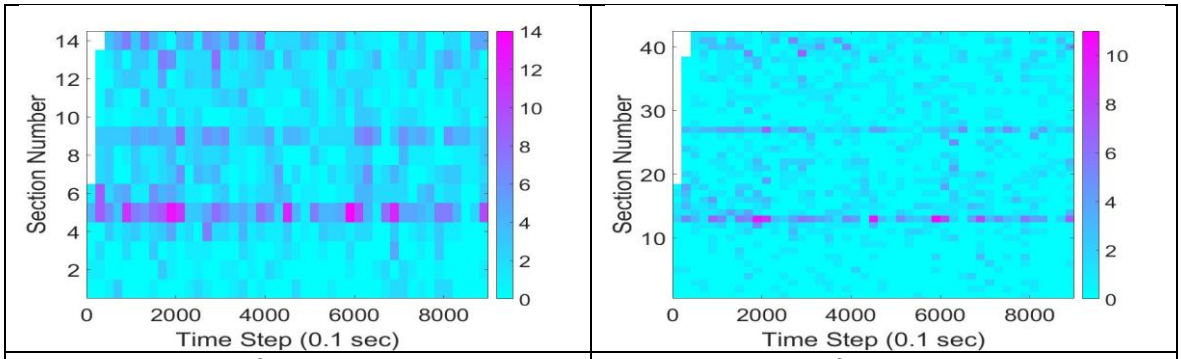
Table 3 shows the average RMSE results for the 10% MPR case with respect to the 100% MPR case for each lane aggregation over ten iterations. Lane 1 has the lowest RMSE value while aggregating data from all the lanes has the highest RMSE value. Looking at a single lane reduces the amount of noise in the data. Aggregating lanes 1 through 4 has the second lowest error which is supported empirically by Figure 17. It is interesting to note that the RMSE values do not strictly increase or decrease as more lanes are analyzed, suggesting that this relationship may be unique to each dataset.

Table 3: Average RMSE Values for 10% MPR Case w.r.t 100% MPR Case for Each Scenario over 10 Iterations for US 101

	Average Speed Variance RMSE
Lane 1	12.8
Lanes 1-2	18.4
Lanes 1-3	18.0
Lanes 1-4	16.9
Lanes 1-5	18.7

LANE CHANGE ANALYSIS

Since Ahn and Cassidy (*27*) had found that shockwaves could be initiated by lane changes, it was hypothesized that the number of lane changes would increase as vehicles approached the point of shockwave formation, which is the merge point for the US 101 segment. The number of lane changes was first calculated for 150 ft. sections of US 101. The results showed a jump in the number of lane changes in section 5, which is where the merge point is located, but not a gradual increase in the number of lane changes over distances. The distance interval was then further broken up, and the number of lane changes was calculated for 50 ft. intervals. However, the results again showed only a high jump in the number of lane changes near the merge point. Therefore, no relationship was found between shockwave formation and number of lane changes for the US 101 dataset. Figure 18 shows the results.



(a) 150 ft. intervals (b) 50 ft. intervals

Figure 18: Lane change profiles from 7:50 – 8:05 a.m. on US 101 in 20 sec. intervals

The correlation coefficient was calculated between the speed variance and lane change profiles of 150 ft. and 20 sec. intervals. The correlation coefficient was 0.006 indicating that no relationship exists between the number of lane changes and speed variance for the US 101 dataset. This supports the empirical observation found from the plots of the data. This result may be attributed to the short length of the segment.

SUMMARY

Speed variances were calculated and plotted by lane for the US 101 dataset. The results show that the shockwaves are only clearly seen in lane 1, the lane closest to the auxiliary lane, due to lane changes from the on- and off-ramp. Analyzing the speed variances of each lane individually can narrow down which lanes are most affected and further pinpoint the cause of the shockwave. Speed variance profiles were plotted by different lane aggregations and the RMSE was found for 10% MPR with respect to 100% MPR case for each lane aggregation over ten iterations. Lane 1 had the lowest RMSE value while aggregating data from all the lanes has the highest RMSE error. The

number of lane changes was also plotted over 150 ft. and 50 ft. intervals. The correlation coefficient between speed variance and number of lane changes was 0.006 indicating that no relationship exists between the number of lane changes and speed variance for the US 101 dataset.

CHAPTER VII

SUMMARY AND FUTURE RESEARCH

SUMMARY

The purpose of this study was to utilize connected vehicles data and vehicle trajectory data to determine if any increase in speed variance over distances could be an indicator of shockwave formation. Earlier and more reliable shockwave detection can delay shockwave propagation in advance of a traffic jam and further reduce the negative impacts on safety, performance, and emissions. The shockwave detection method used in this research was tested against different factors including individual lanes, lane aggregations, and differently sized intervals to test and improve its reliability. Since this detection method relies on speed and location data from vehicles, one problem it faces is the currently low market penetration rates of CVs. In addition, there is an increased threat of cyberattacks and equipment malfunctions due to a CV's complexity and connectivity capabilities. Therefore, this study also explored the impacts of partial connectivity and misinformation on speed variances calculated to detect shockwaves. The idea for this study came from past literature on traffic flow theory that found that a rapid increase in speed variance could be an indicator of shockwave formation. However, the previous studies had used data from loop detectors and radar guns to show this and was therefore limited to detecting shockwaves at specific locations. CV technology allows for more accurate detection across time and space.

The I-5 INFLO demonstration and the NGSIM US 101 datasets were analyzed in this research. Matlab, Google Earth, and Excel were the primary programs used to analyze the data. The I-5 dataset only had data from 19 connected vehicles. Average speeds and speed variances were plotted in one minute and 0.5-mile intervals using a rolling horizon approach updated every 0.25 miles. Average speeds and speed variances were also plotted for the US 101 dataset from 7:50 a.m. – 8:05 a.m. in 150 ft. and 20 sec. intervals. Six shockwaves most likely caused by lane-changing maneuvers from the on- and off-ramp could be seen in the plots of the data.

The effects of partial connectivity and misinformation were tested on the US 101 dataset. MPRs of 100%, 50%, and 10% and PDRs of 100%, 80%, and 50% were tested. MPR was defined as the percent of connected vehicles over the 15-minute period. PDR was defined as the percentage of speed data received and was updated at each time step. RMSE values with respect to the 100% MPR/100% PDR case were found for the lower connectivity cases.

Four different misinformation scenarios were simulated and tested on one shockwave in the US 101 dataset. Specifically, an interval with low speed variance and the next interval with a jump in speed variance were analyzed. The first scenario assumed a percentage of the speeds were reported 5-10 mph higher than actual speeds. The second scenario assumed a percentage of the speeds were reported 5-10 mph lower than actual speeds, and the third scenario assumed a combination of lower and higher reported speeds by 5-10 mph. The effect of different percentages of incorrect speeds on the intervals' average speed and speed variance were analyzed. The fourth scenario

manipulated the speeds for only the low speed variance interval and plotted the differences between the variances of the two intervals for different falsely reported speed ranges.

Speed variance profiles were plotted by individual lanes and different lane aggregations for the US 101 dataset. Average RMSE values from ten iterations were also found for the 10% MPR case with respect to the 100% MPR case for the different lane aggregations. The number of lane changes was calculated using both 150 ft. and 50 ft. distance intervals. The correlation coefficient was also calculated between the speed variances and number of lane changes in 150 ft. intervals.

This chapter summarizes the major findings of the study. The limitations are also discussed after the findings. Finally, recommendations for future research are given.

FINDINGS

The major findings from the I-5 data analysis and the 100% connectivity case of the US 101 dataset are as follows:

- An analysis of six segments from the I-5 dataset showed that speed variances calculated over distances could be used to identify shockwaves. The six segments showed that a jump in speed variance could detect congestion 0.25 miles and/or one minute earlier than when the average speed dropped to the congested range.
- The 100% connectivity case of the US 101 dataset showed that an increase in speed variance is an indicator of congestion formation.

The major findings from the partial connectivity cases of US 101 and causes for the differences in results between I-5 and US 101 are as follows:

- All cases of 100% MPR and 50% MPR showed the six shockwaves in the US 101 dataset.
- In general, RMSE increases as connectivity decreases especially below around 35% MPR. No relationship was found across PDR.
- The difference in results between I-5 and US 101 at low connectivity intervals could be attributed to the length of the intervals used in calculating the speed variance. 500 ft. intervals were tested on the US 101 dataset. The 500 ft. distance intervals indicated that speed variance could be used to detect shockwaves with at least 20% MPR while the 150 ft. intervals suggested the breakdown point to be between 30% and 40% MPR. RMSE also decreases as the distance interval increases. Aggregating data reduces not only noise but also the precision in locating the start of a shockwave.

The major findings on the impacts of misinformation on the speed distribution based shockwave detection method are as follows:

- As the percentage of falsely reported speeds increased, the average speeds increased or decreased depending on whether speeds were reported higher or lower than their actual values. Speed variance showed an inverted U relationship to the percentage of incorrect speeds.
- When actual speeds were randomly increased and decreased, average speeds remained steady while variance continually increased. The difference between

the low and high variance intervals remained fairly constant indicating that speed variance can still be tracked along time and space and a huge jump in interval variance values could indicate shockwave formation.

- When speeds were manipulated for only one of the intervals by 1-5 mph and 5-10 mph, the difference in speed variance values between the two intervals remained significant for all percentages of incorrect speeds. The difference in speed variance values dropped to zero when speeds were manipulated within 10-15 mph and 15-20 mph with 50% and 25% incorrect speeds respectively.

The major findings from the individual lane analyses, lane aggregations, and number of lane changes are as follows:

- Speed variance profiles by lane showed that most of the shockwaves could only be clearly seen in lane 1, the lane closest to the on- and off-ramp. Analyses by lane can further point out the cause of the shockwave and show which lanes are most affected. This was not as clear in the average speed profiles.
- Lane 1 had the lowest RMSE value while aggregating data from all the lanes had the highest RMSE value. RMSE values did not strictly increase or decrease as more lanes were analyzed together.
- No relationship was observed between shockwave formation and the number of lane changes in the 150 ft. and 50 ft. lane change profiles for the US 101 dataset.
- The correlation coefficient was calculated to be 0.006 indicating that no relationship exists between the number of lane changes and speed variance for the US 101 dataset.

LIMITATIONS

The major limitations of this research are summarized as follows:

- The shockwave analysis in the US 101 dataset was limited by the size of the 2100-foot long segment. Although the I-5 INFLO demonstration only had a few connected vehicles, data was collected over several miles. This could be one reason why speed variance was observed to detect congestion much earlier than average speed for the I-5 dataset than the US 101 dataset.
- The shockwaves present in the US 101 dataset were all most likely caused by lane-changing maneuvers from the on- and off-ramp. Other types of shockwaves were not analyzed using the speed distribution based shockwave detection method.
- Misinformation in CV data can occur in a variety of different ways. This study focused on only four possible scenarios and only manipulated speed data.

FUTURE RESEARCH

Recommendations for future research are listed as follows:

- Analyze data from other CV deployments over longer segments to observe whether the speed distribution based method can detect shockwaves earlier than solely looking at average speeds. The relationship between number of lane changes and speed variance and the effect of differently sized intervals can be reanalyzed with a longer segment analysis.

- Analyze different types of shockwaves to determine whether the speed variance values depend on the cause of the shockwave.
- Manipulate location data as well as speed data and test the resulting impact on average speed and speed variance. Consider partial connectivity cases as well.

REFERENCES

1. Barbaresso, J., Cordahi, G., Garcia, D., Hill, C., Jendzejec, A., and K. Wright. *USDOT's Intelligent Transportation Systems (ITS) Strategic Plan 2015-2019*. Publication FHWA-JPO-14-145. FHWA, U.S. Department of Transportation, Dec. 2014.
2. Zeng, X., Balke, K., and P. Songchitruksa. *Potential Connected Vehicle Applications to Enhance Mobility, Safety, and Environmental Security*. No. SWUTC/12/161103-1. Southwest Region University Transportation Center, Texas Transportation Institute, Texas A&M University System, 2012.
3. M. McGurrin, M. Vasudevan, and P. Tarnoff. *Benefits of Dynamic Mobility Applications Preliminary Estimates from the Literature*. Publication FHWA-JPO-13-004. United States Department of Transportation RITA JPO, Dec. 2012.
4. Hendrickson, C., and C. Harper. *Safety and Cost Assessment of Connected and Autonomous Vehicles*. The National USDOT University Transportation Center for Safety, Project Report, 2015.
5. Guler, S. I., Menendez, M., and L. Meier. Using Connected Vehicle Technology to Improve the Efficiency of Intersections. In *Transportation Research Part C: Emerging Technologies*, Vol. 46, Sept 2014, pp. 121-131.
6. Mahmassani, H. S. 50th Anniversary Invited Article—Autonomous Vehicles and Connected Vehicle Systems: Flow and Operations Considerations. In *Transportation Science,* Vol. 50, No. 4, 2016, pp. 1140-1162.

7. Stephens, D., Timcho, T., Schroeder, J., Brown, J., Bacon, P., Smith, T., Balke, K., Charara, H., and S. Sunkari. *Intelligent Network Flow Optimization (INFLO) Prototype Seattle Small-Scale Demonstration Report*. Publication FHWA-JPO-15-223. United States Department of Transportation RITA JPO, May 2015.
8. Visiongain Ltd. *Top 20 Connected Car Companies 2016*. PRNewswire, www.prnewswire.com/news-releases/top-20-connected-car-companies-2016-584101521. Accessed March 14, 2017.
9. *Federal Motor Vehicle Safety Standards: V2V Communications.* National Highway Traffic Safety Administration, Vol. 82, No. 8, Jan. 2017.
10. Ashley, S. *Cybersecurity for robot cars.* SAE International. Accessed March 14, 2017.
11. Bigelow, P. *How the Connected Car Will Defend Against Hackers*. http://blog.caranddriver.com/how-the-connected-car-will-defend-against-hackers/. Accessed March 14, 2017.
12. *Cyber Security in the Connected Vehicle Report 2016 Brochure*. TU Automotive, 2016, http://www.tu-auto.com/cybersecurity-report/. Accessed March 14, 2017.
13. Treiber, M., and A. Kesting. *Traffic Flow Dynamics: Data, Models and Simulation.* Springer-Verlag Berlin Heidelberg, 2013.
14. Kuhne, R., and P. Michalopoulos. Continuum flow models. In *A State of the Art Report Revised Monograph on Traffic Flow Theory*, FHWA, Washington, DC, 1997, pp. (5)1-51.

15. Hegyi, A., S. Hoogendoorn, M. Schreuder, H. Stoelhorst, and F. Viti. SPECIALIST: A dynamic speed limit control algorithm based on shock wave theory. In *11th International IEEE Conference on Intelligent Transportation Systems,* IEEE, 2008. pp. 827-832.
16. Kerner, B. S., and H. Rehborn. Experimental features and characteristics of traffic jams. In *Physical Review E,* Vol. 53, No. 2, 1996, pp. R1297-R1300.
17. Kerner, B.S., and H. Rehborn. Experimental Properties of Complexity in Traffic Flow. In *Physical Review E,* Vol. 53, May 1996, pp. R4275-R4278.
18. Kerner, B.S., and H. Rehborn, Experimental Properties of Phase Transitions in Traffic Flow. In *Physical Review Lett,* Vol. 79, Nov. 1997, pp. 4030-4033.
19. Helbing, D., M. Treiber, A. Kesting, and M. Schönhof. Theoretical vs. Empirical Classification and Prediction of Congested Traffic States. In *The European Physical Journal B-Condensed Matter and Complex Systems,* Vol. 69, No. 4, 2009, pp. 583-598.
20. Kerner, B.S. Phase Transitions in Traffic Flow. In *Traffic and Granular Flow '99,* Springer Berlin Heidelberg, 2000, pp. 253-283.
21. Kerner, B.S. Experimental Features of Self-Organization in Traffic Flow. In *Physical Review Lett,* Vol. 81, Oct. 1998, pp. 3797-3800.
22. Kuhne, R.D. Macroscopic Freeway Model for Dense Traffic – Stop-Start Waves and Incident Detection. In *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory,* 1984, pp. 21-42.

23. Kuhne, R.D. Freeway Speed Distribution and Acceleration Noise—Calculations from a Stochastic Continuum Theory and Comparison with Measurements. In *Proceedings of the 10th International Symposium on Transportation and Traffic Theory*, 1987. pp. 119-137.
24. Krauss, S., P. Wagner, and C. Gawron. Metastable States in a Microscopic Model of Traffic Flow. In *Physical Review E,* Vol. 55, May 1997, pp. 5597--5602.
25. Helbing, D. Improved Fluid-dynamic Model for Vehicular Traffic. In *Physical Review E,* Vol. 51, April 1995, pp. 3164--3169.
26. Kuhne, R.D. Freeway Traffic Control Using a Dynamic Traffic Flow Model and Vehicle Reidentification Techniques. In *Freeway Operations, Highway Capacity, and Traffic Flow,* 1991, pp. 251-259.
27. Ahn, S., and M. J. Cassidy. Freeway traffic oscillations and vehicle lane-change maneuvers. In *17th International Symposium on Traffic and Transportation Theory*, 2007. pp. 691-710.
28. Schakel, W. J., B. Van Arem, and B. D. Netten. Effects of Cooperative Adaptive Cruise Control on Traffic Flow Stability. In *13th International IEEE Conference on Intelligent Transportation Systems*, IEEE, 2010. pp. 759-764.
29. Hegyi, A., and S. Hoogendoorn. Dynamic Speed Limit Control to Resolve Shock Waves on Freeways - Field Test Results of the SPECIALIST Algorithm. In *13th International IEEE Conference on Intelligent Transportation Systems*, IEEE, 2010. pp. 519-524.

30. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, No. 229, The Royal Society, 1955. pp. 317-345.
31. Richards, P. I. Shock Waves on the Highway. In *Operations research*, Vol. 4, No. 1, 1956, pp. 42-51.
32. Zheng, Z., S. Ahn, D. Chen, and J. Laval. Applications of Wavelet Transform for Analysis of Freeway Traffic: Bottlenecks, Transient Traffic, and Traffic Oscillations. In *Transportation Research Part B: Methodological*, Vol. 45, No. 2, 2011, pp. 372-384.
33. Daubechies, I. *Ten Lectures on Wavelets.* Vol. 61 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.
34. Talebpour, A., and H. S. Mahmassani. Modeling Acceleration Behavior in a Connected Environment. In *Midyear Meetings and Symposium Celebrating 50 Years of Traffic Flow Theory*, Portland, OR, 2014, pp. 87-91.
35. Lu, X.-Y., and A. Skabardonis. Freeway Traffic Shockwave Analysis: Exploring the NGSIM Trajectory Data. Presented at the *86th Annual Meeting of the Transportation Research Board*, Washington, DC, 2007.
36. Khondaker, B., and L. Kattan. Variable Speed Limit: A Microscopic Analysis in a Connected Vehicle Environment. In *Transportation Research Part C: Emerging Technologies*, Vol. 58, 2015, pp. 146-159.

37. Lu, X.-Y., S. E. Shladover, I. Jawad, R. Jagannathan, and T. Phillips. A Novel Speed-Measurement Based Variable Speed Limit/Advisory Algorithm for a Freeway Corridor with Multiple Bottlenecks. Presented at the *Transportation Research Board 94th Annual Meeting*, 2015.
38. Ramezani, H., and R. Benekohal. Optimized Speed Harmonization with Connected Vehicles for Work Zones. In *18th International Conference on Intelligent Transportation Systems*, IEEE, 2015. pp. 1081-1086.
39. Abdel-Aty, M., J. Dilmore, and A. Dhindsa. Evaluation of Variable Speed Limits for Real-time Freeway Safety Improvement. In *Accident Analysis & Prevention*, Vol. 38, No. 2, 2006, pp. 335-345.
40. Hegyi, A. *Model Predictive Control for Integrating Traffic Control Measures*. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2004.
41. Zegeye, S. K., B. De Schutter, H. Hellendoorn, and E. Breunesse. Reduction of Travel Times and Traffic Emissions Using Model Predictive Control. In *American Control Conference*, IEEE, 2009. pp. 5392-5397.
42. Garcia, D., C. Hill, and J. Wagner. *Revolutionizing our Roadways: Cybersecurity Considerations for Connected and Automated Vehicle Policy*. Texas A&M Transportation Institute, Transportation Policy Research Center, 2015.
43. Onishi, H. Paradigm Change of Vehicle Cyber Security. In *4th International Conference on Cyber Conflict*, IEEE, 2012. pp. 1-11.

44. Petit, J., and S. E. Shladover. Potential Cyberattacks on Automated Vehicles. In *IEEE Transactions on Intelligent Transportation Systems,* Vol. 16, No. 2, 2015, pp. 546-556.

45. Weimerskirch, A., and R. Gaynier. An Overview of Automotive Cybersecurity: Challenges and Solution Approaches. In *Proceedings of the 5th International Workshop on Trustworthy Embedded Devices*, Denver, CO, 2015. pp. 53.

46. *Vehicle Cybersecurity: DOT and Industry Have Efforts Under Way, but DOT Needs to Define its Role in Responding to a Real-world Attack*. United States Government Accountability Office, 2016.

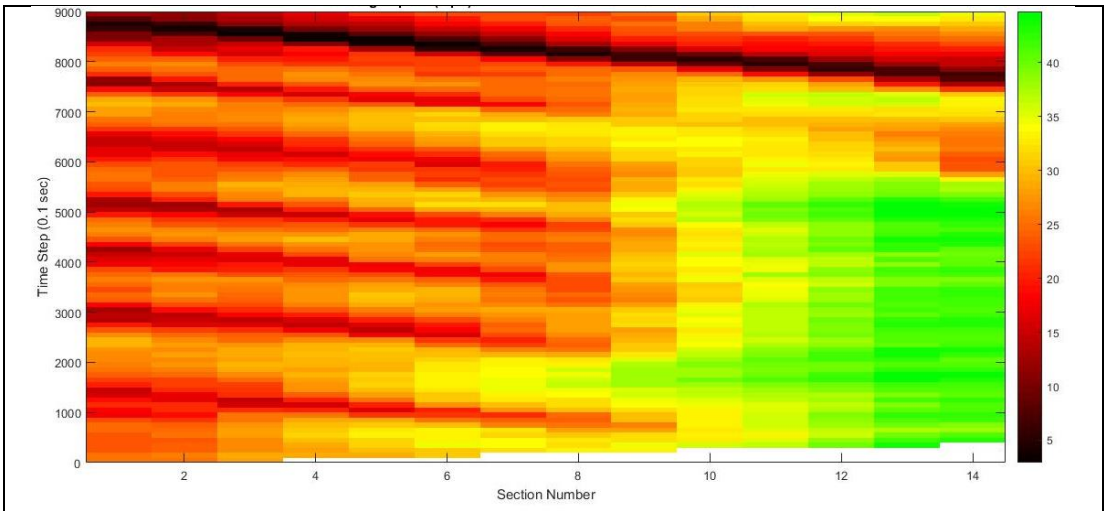
47. *INFLO Prototype Seattle Small-Scale Demonstration Data.* Battelle, Columbus, OH., 2016.

48. *Next Generation Simulation (NGSIM) US Route 101 Dataset*. FHWA USDOT ITS JPO, Washington DC, 2016.

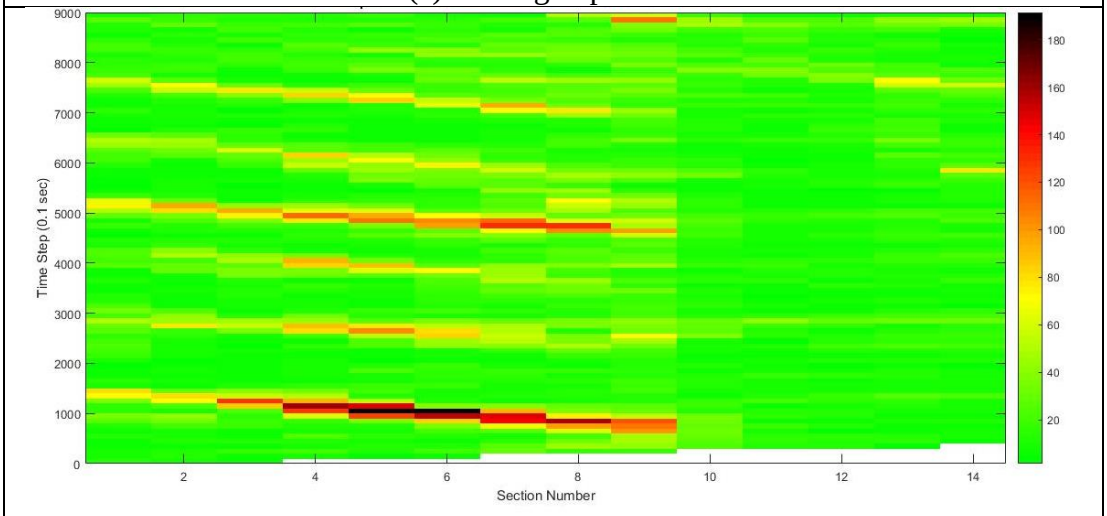
49. Montanino, M., and V. Punzo. Making NGSIM Data Usable for Studies on Traffic Flow Theory: Multistep Method for Vehicle Trajectory Reconstruction. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2390, 2013, pp. 99-111.

50. Lu, X.-Y., P. Varaiya, and R. Horowitz. Fundamental Diagram Modelling and Analysis Based NGSIM Data. In *Proceedings of the 12th IFAC Symposium on Control in Transportation Systems,* Vol. 42, No. 15, 2009, pp. 367-374.

APPENDIX

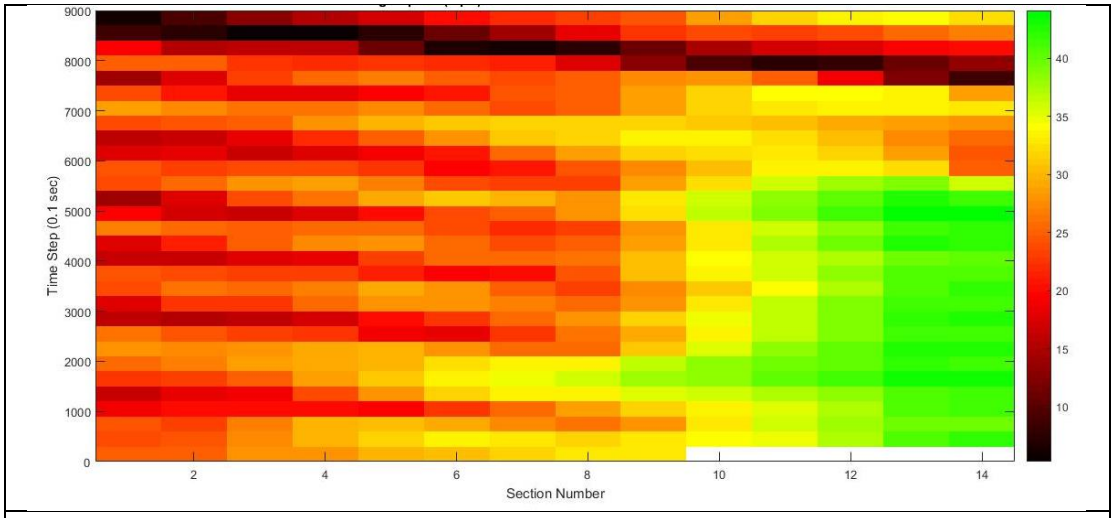


(a) Average Speed

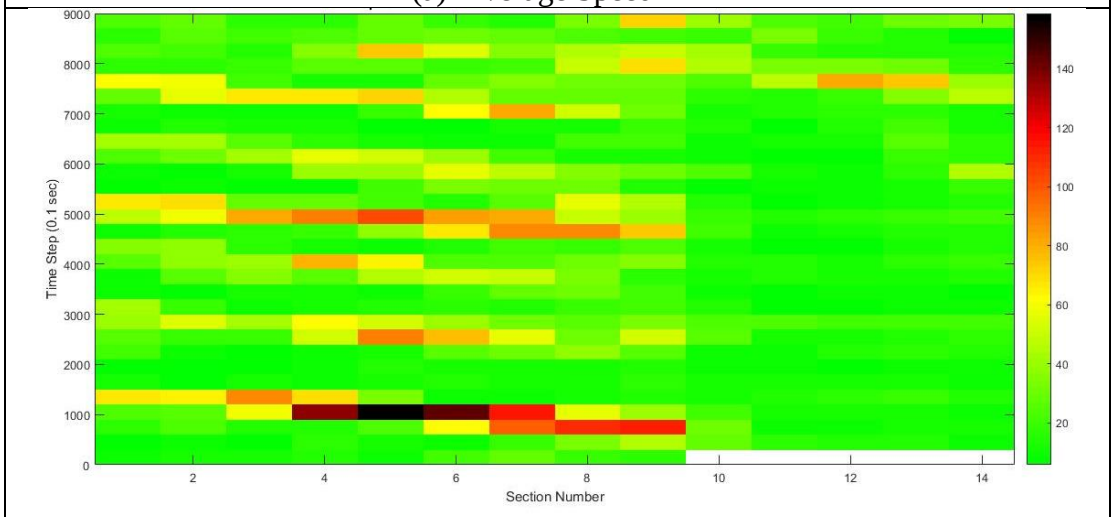


(b) Speed Variance

Figure A-1: US 101 from 7:50 – 8:05 a.m. in 150 ft. and 10 sec. intervals

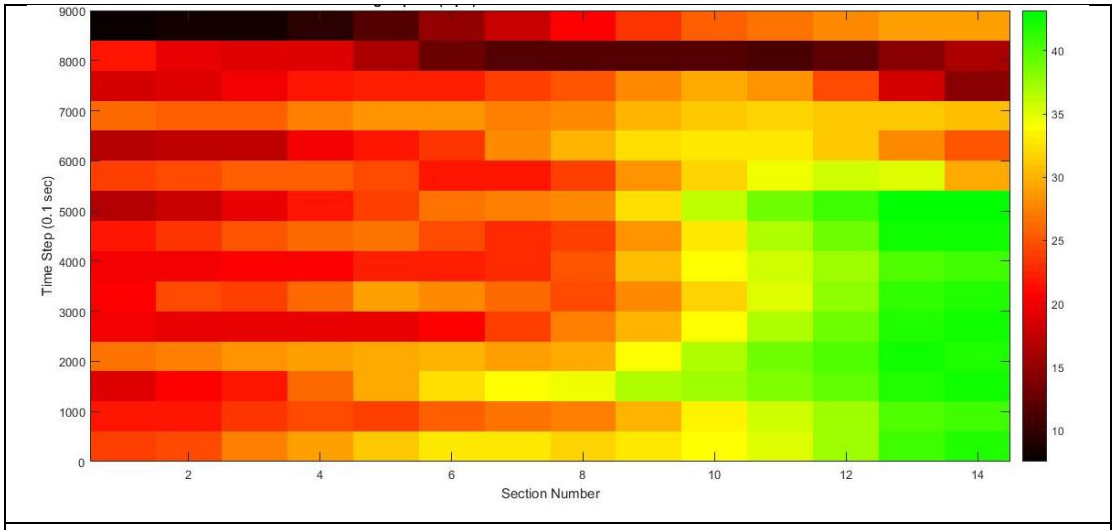


(a) Average Speed

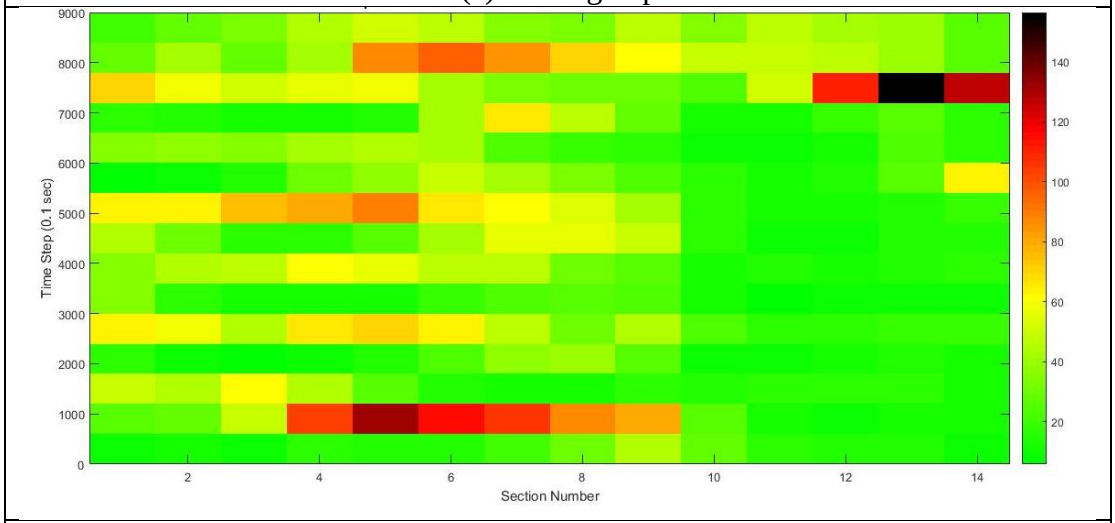


(b) Speed Variance

Figure A-2: US 101 from 7:50 – 8:05 a.m. in 150 ft. and 30 sec. intervals

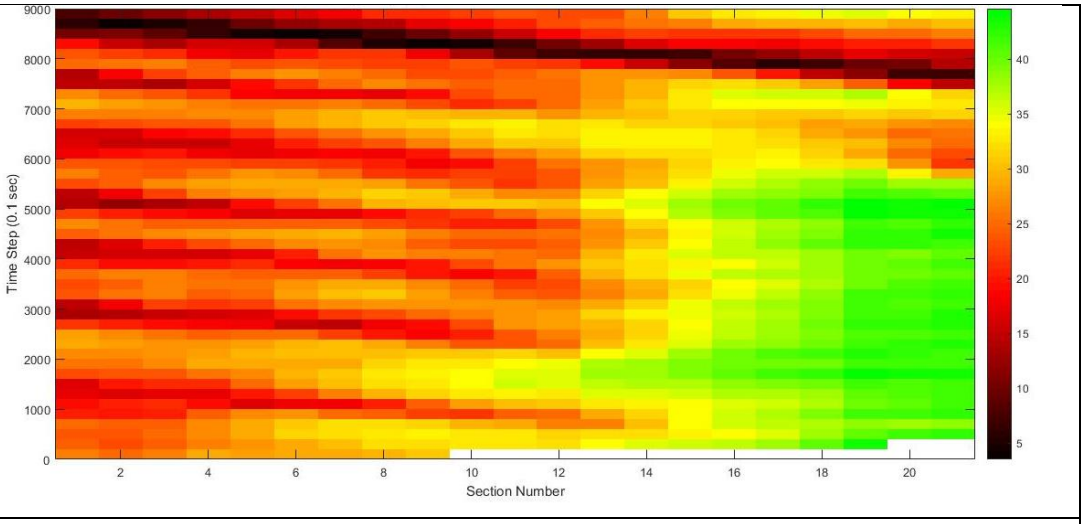


(a) Average Speed

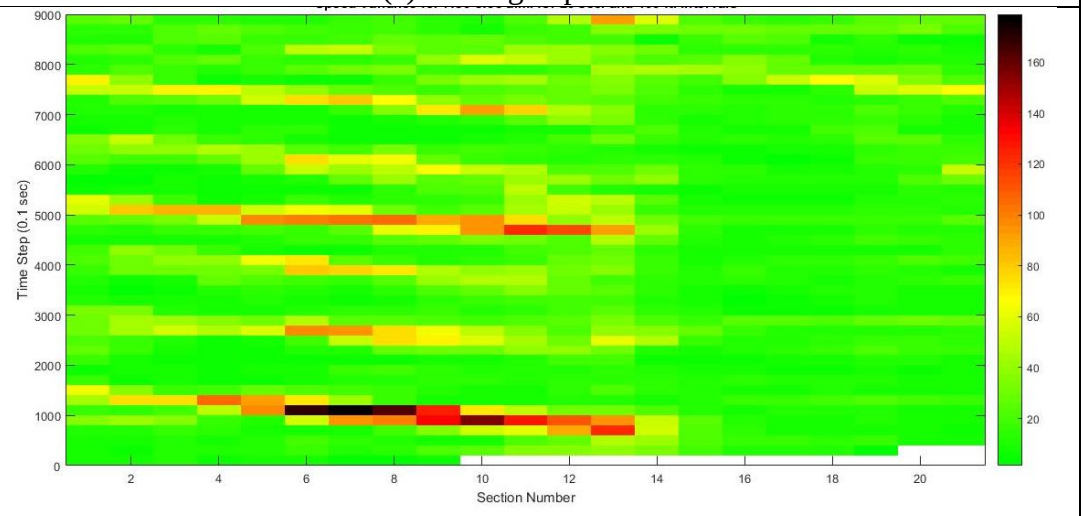


(b) Speed Variance

Figure A-3: US 101 from 7:50 – 8:05 a.m. in 150 ft. and 60 sec. intervals

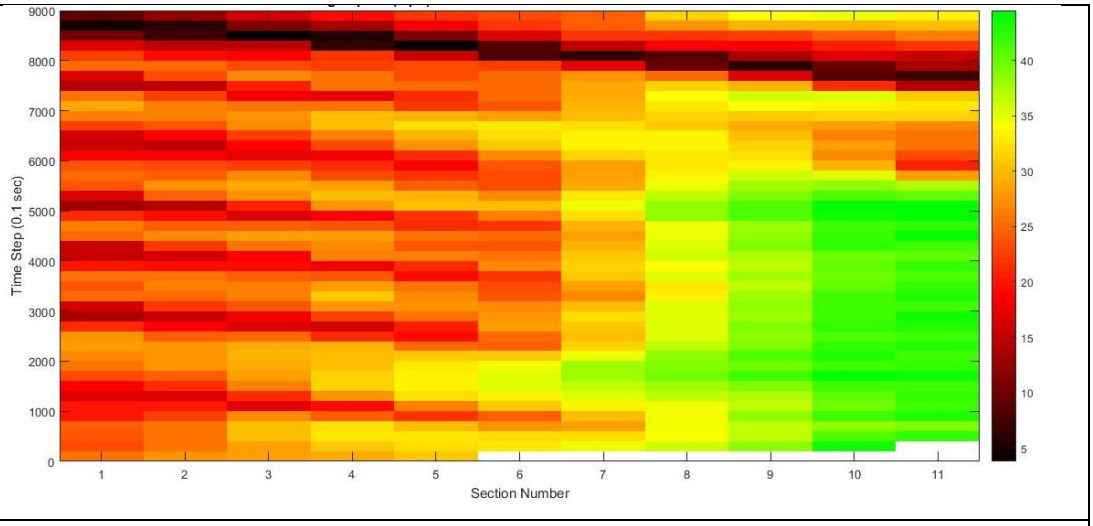


(a) Average Speed

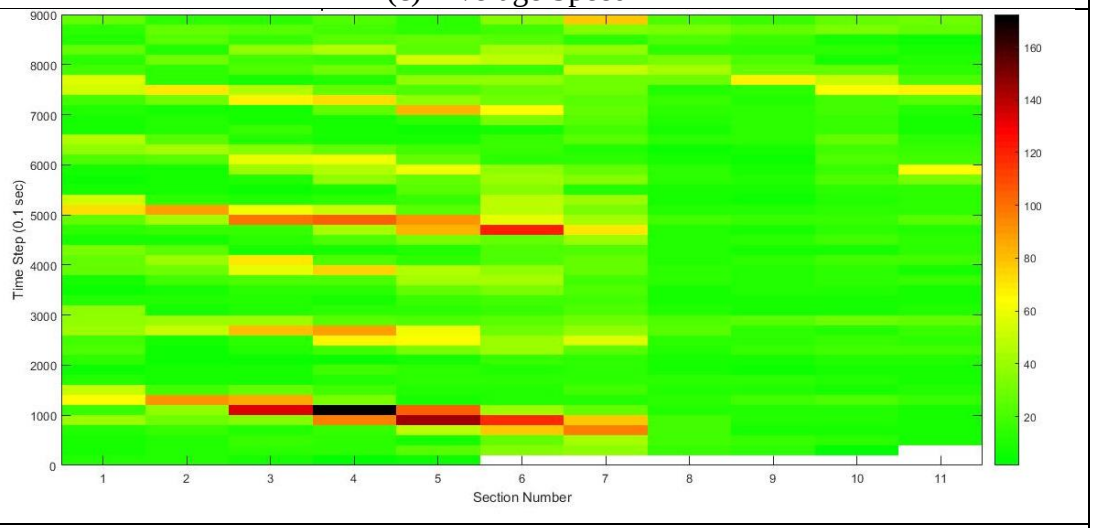


(b) Speed Variance

Figure A-4: US 101 from 7:50 – 8:05 a.m. in 100 ft. and 20 sec. intervals



(c) Average Speed



(d) Speed Variance

Figure A-5: US 101 from 7:50 – 8:05 a.m. in 200 ft. and 20 sec. intervals

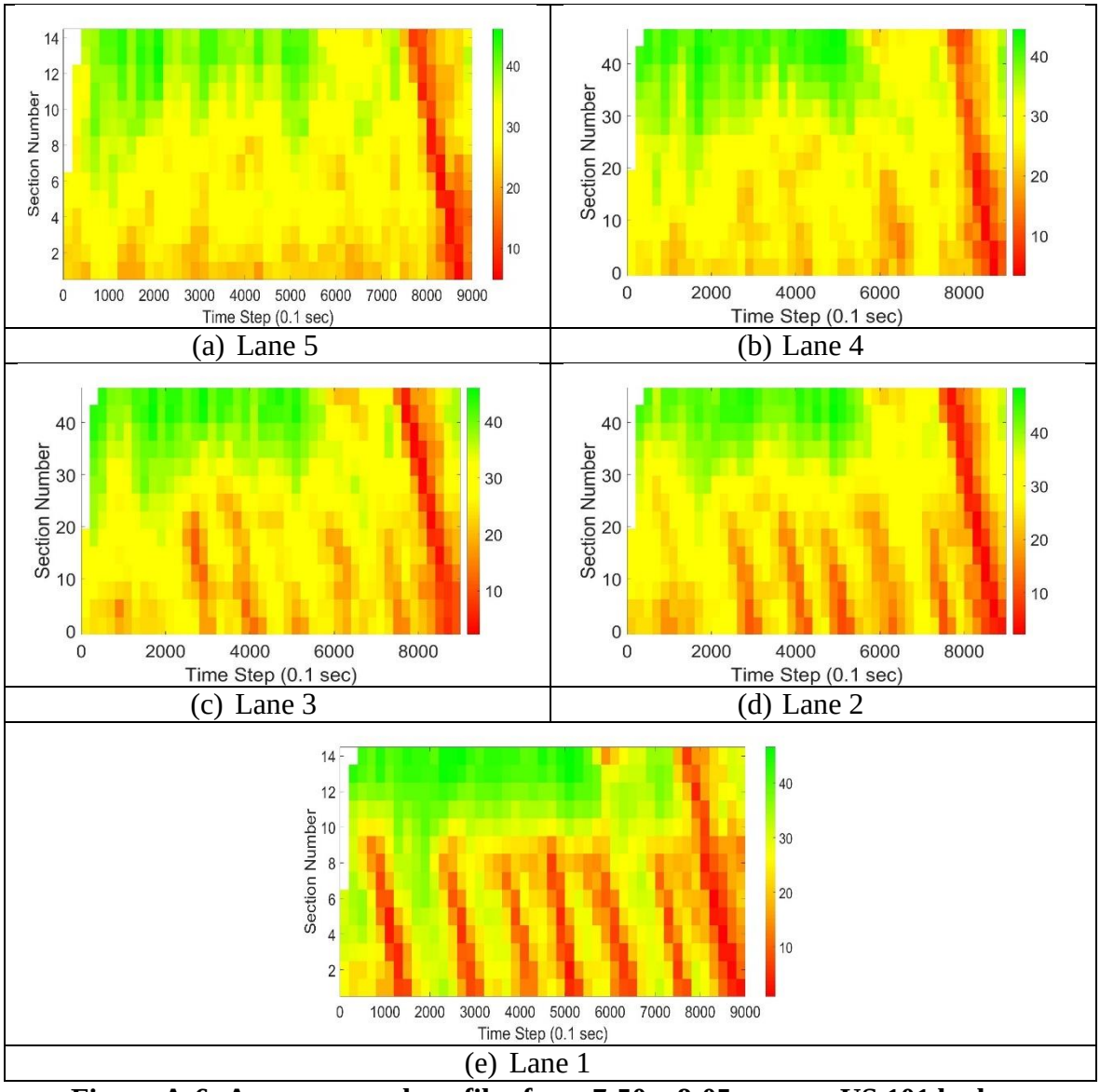


Figure A-6: Average speed profiles from 7:50 – 8:05 a.m. on US 101 by lane