

Summer 2019

Deep Reinforcement Learning Approach for Lagrangian Control: Improving Freeway Bottleneck Throughput Via Variable Speed Limit

Reza Vatani Nezafat
Old Dominion University, rvata001@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/cee_etds



Part of the [Civil Engineering Commons](#), and the [Transportation Commons](#)

Recommended Citation

Nezafat, Reza V.. "Deep Reinforcement Learning Approach for Lagrangian Control: Improving Freeway Bottleneck Throughput Via Variable Speed Limit" (2019). Doctor of Philosophy (PhD), dissertation, Civil/Environmental Engineering, Old Dominion University, DOI: 10.25777/dc8z-c232
https://digitalcommons.odu.edu/cee_etds/84

This Dissertation is brought to you for free and open access by the Civil & Environmental Engineering at ODU Digital Commons. It has been accepted for inclusion in Civil & Environmental Engineering Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

**DEEP REINFORCEMENT LEARNING APPROACH FOR LAGRANGIAN CONTROL:
IMPROVING FREEWAY BOTTLENECK THROUGHPUT VIA VARIABLE SPEED LIMIT**

by

Reza Vatani Nezafat
M.S. May 2019, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

CIVIL ENGINEERING

OLD DOMINION UNIVERSITY
2019

Approved by:

Mecit Cetin (Chair)

Sherif S. Ishak (Member)

Hong Yang (Member)

ABSTRACT

DEEP REINFORCEMENT LEARNING APPROACH FOR LAGRANGIAN CONTROL: IMPROVING FREEWAY BOTTLENECK THROUGHPUT VIA VARIABLE SPEED LIMIT

Reza Vatani Nezafat
Old Dominion University, 2019
Director: Dr. Mecit Cetin

Connected vehicles (CVs) will enable new applications to improve traffic flow. The focus of this dissertation is to investigate how reinforcement learning (RL) control for the variable speed limit (VSL) through CVs can be generalized to improve traffic flow at different freeway bottlenecks. Three different bottlenecks are investigated: A sag curve, where the gradient changes from negative to positive values causes a reduction in the roadway capacity and congestion; a lane reduction, where three lanes merge to two lanes and cause congestion, and finally, an on-ramp, where increase in demand on a multilane freeway causes capacity drop. An RL algorithm is developed and implemented in a simulation environment for controlling a VSL in the upstream to manipulate the inflow of vehicles to the bottleneck on a freeway to minimize delays and increase the throughput. CVs are assumed to receive VSL messages through Infrastructure-to-Vehicle (I2V) communications technologies. Asynchronous Advantage Actor-Critic (A3C) algorithms are developed for each bottleneck to determine optimal VSL policies. Through these RL control algorithms, the speed of CVs are manipulated in the upstream of the bottleneck to avoid or minimize congestion. Various market penetration rates for CVs are considered in the simulations. It is demonstrated that the RL algorithm is able to adapt to stochastic arrivals of CVs and achieve significant improvements even at low market penetration rates of CVs, and the RL algorithm is able to find solution for all three bottlenecks. The results also show that the RL-based solutions outperform feedback-control-based solutions.

© 2019 Reza Vatani Nezafat. All Rights Reserved.

Dedicated to
My family

ACKNOWLEDGMENTS

Throughout the writing of this dissertation, I have received a great deal of support and assistance. I would first like to thank my advisor, Dr. Cetin, whose expertise was invaluable in the formulating of the research topic and methodology in particular. I extend my sincere appreciation to my dissertation committee, Dr. Ishak, and Dr. Hong, for their supports and valuable insights.

The research presented in this thesis was funded by The National Center for Strategic Transportation Policies, Investments, and Decisions, which is led by the University of Maryland (UMD). I also want to thank the Civil and Environmental Engineering Visiting Council (CEEVC) for the scholarship they have provided to me.

TABLE OF CONTENT

LIST OF FIGURES.....	IV
LIST OF TABLES.....	VI
CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW	1
1.1) TRAFFIC DYNAMICS, MODELING, AND CONTROL	3
1.2) SIMULATION	5
1.2.1) <i>Longitudinal Driving Behavior</i>	6
1.2.2) <i>Lateral Driving Behavior</i>	6
1.2.3) <i>Driving Behavior at Sags</i>	7
1.3) CONNECTED VEHICLES	8
1.4) DYNAMIC TRAFFIC MANAGEMENT VIA VARIABLE SPEED LIMIT	9
1.5) DEEP REINFORCEMENT LEARNING.....	10
1.6) RESEARCH OBJECTIVE AND RESEARCH QUESTIONS	11
1.7) RESEARCH APPROACH AND RESEARCH SCOPE.....	12
1.8) MAIN CONTRIBUTIONS	14
1.9) OUTLINE OF THE DISSERTATION	14
CHAPTER 2: METHODOLOGY	15
2.1) LONGITUDINAL DRIVING BEHAVIOR.....	15
2.2) LATERAL DRIVING BEHAVIOR	16
2.2.1) <i>Lane Change Incentives</i>	17
2.2.2) <i>Gap Acceptance and Relaxation</i>	20
2.2.3) <i>Synchronization and Gap Creation</i>	21
2.3) CONTROL STRATEGY	21
2.3.1) <i>Deep Reinforcement Learning</i>	22
2.3.2) <i>Feedback Control</i>	27
2.4) ENVIRONMENTS	28
2.4.1) <i>Sag Curve</i>	28
2.4.2) <i>Lane Reduction</i>	31
2.4.3) <i>On-Ramp</i>	34
2.5) RL ARCHITECTURE: ACTION, STATE SPACE, AND REWARD	36
CHAPTER 3: RESULTS	39
3.1) OPTIMUM SOLUTIONS OF THE FEEDBACK CONTROL ALGORITHM	39
3.2) OPTIMUM SOLUTIONS OF RL ALGORITHM.....	41
3.3) PERFORMANCE OF THE RL UNDER NEW DEMAND PATTERNS.....	45
3.4) SENSITIVITY TO THE MARKET PENETRATION RATE OF CONNECTED VEHICLES	51
3.5) ADAPTIVE VSL FOR DIFFERENT MPR OF CVs	56
CHAPTER 4: CONCLUSION	63
REFERENCES	65
VITA	70

LIST OF FIGURES

FIGURE 1 CAPACITY DROP PHENOMENA ILLUSTRATED ON A HYPOTHETICAL FUNDAMENTAL DIAGRAM.....	5
FIGURE 2 IN-RAMP BOTTLENECK.....	12
FIGURE 3 LANE REDUCTION BOTTLENECK.....	13
FIGURE 4 SAG CURVE BOTTLENECK.....	13
FIGURE 5 OVERVIEW OF THE SPECTRUM OF DESIRE AND LANE CHANGING REGIONS.....	17
FIGURE 6 Q-VALUES TABLE EXAMPLE.....	23
FIGURE 7 Q-NETWORK (LEFT) USED TO CALCULATE EQUATION 27, AND THE TARGET-NETWORK (RIGHT) IS USED TO CALCULATE EQUATION 28	24
FIGURE 8 A2C (LEFT) AND A3C (RIGHT) PARALLEL PROCEDURE.....	26
FIGURE 9 UPDATE PROCEDURE OF THE GLOBAL NETWORK.....	27
FIGURE 10 THE GEOMETRY OF THE SAG CURVE BOTTLENECK.....	28
FIGURE 11 DEMAND PROFILE OVER TIME	29
FIGURE 12 SAG CURVE WITH NO CONTROL: INPUT DEMAND AND EXIT FLOWS (TOP RIGHT), THE DENSITY MEASURED BY LOOP DETECTOR (MIDDLE RIGHT), HEAT-MAP (LEFT), AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (BOTTOM RIGHT) WITHOUT A VSL CONTROL SYSTEM.....	30
FIGURE 13 THE GEOMETRY OF THE LANE REDUCTION BOTTLENECK.....	31
FIGURE 14 DEMAND PROFILE OVER TIME	33
FIGURE 15 LANE REDUCTION WITH NO CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND ROW MIDDLE), HEAT-MAP OF SPEED ON TIME –LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	33
FIGURE 16 THE GEOMETRY OF ON-RAMP BOTTLENECK	34
FIGURE 17 DEMAND PROFILE OVER TIME	35
FIGURE 18 ONRAMP WITH NO CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND ROW MIDDLE), HEAT-MAP OF SPEED ON TIME –LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	36
FIGURE 19 ARCHITECTURE OF ACTOR AND CRITIC NETWORK	37
FIGURE 20 SAG CURVE WITH FEEDBACK CONTROL: INPUT DEMAND AND EXIT FLOWS (TOP RIGHT), THE DENSITY MEASURED BY LOOP DETECTOR (MIDDLE RIGHT), HEAT-MAP OF SPEEDS (LEFT), AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (BOTTOM RIGHT)	39
FIGURE 21 LANE REDUCTION WITH FEEDBACK CONTROL: INPUT DEMAND AND EXIT FLOWS (THE SECOND ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND-ROW MIDDLE), HEAT-MAP OF SPEED ON THE TIME-LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL	40
FIGURE 22 ONRAMP WITH FEEDBACK CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND ROW MIDDLE), HEAT-MAP OF SPEED ON TIME-LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	41
FIGURE 23 CONVERGENCE OF THE RL ALGORITHM FOR SAG CURVE BOTTLENECK	42
FIGURE 24 SAG CURVE WITH RL CONTROL: INPUT DEMAND AND EXIT FLOWS (TOP RIGHT), THE DENSITY MEASURED BY LOOP DETECTOR (MIDDLE RIGHT), HEAT-MAP WITH SAMPLE TRAJECTORIES (LEFT), AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (BOTTOM RIGHT) WITHOUT A VSL CONTROL SYSTEM.....	43
FIGURE 25 LANE REDUCTION WITH RL CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND-ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND-ROW MIDDLE), HEAT-MAP OF SPEED ON THE TIME-	

LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	44
FIGURE 26 ONRAMP WITH RL CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND-ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND-ROW MIDDLE), HEAT-MAP OF SPEED ON THE TIME-LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	45
FIGURE 27 SAG CURVE WITH RL CONTROL: INPUT DEMAND AND EXIT FLOWS (TOP RIGHT), THE DENSITY MEASURED BY LOOP DETECTOR (MIDDLE RIGHT), HEAT-MAP WITH SAMPLE TRAJECTORIES (LEFT), AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (BOTTOM RIGHT) WITHOUT A VSL CONTROL SYSTEM.....	46
FIGURE 28 SAG CURVE WITH FEEDBACK CONTROL: INPUT DEMAND AND EXIT FLOWS (TOP RIGHT), THE DENSITY MEASURED BY LOOP DETECTOR (MIDDLE RIGHT), HEAT-MAP WITH SAMPLE TRAJECTORIES (LEFT), AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (BOTTOM RIGHT) WITHOUT A VSL CONTROL SYSTEM.....	47
FIGURE 29 LANE REDUCTION WITH RL CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND ROW MIDDLE), HEAT-MAP OF SPEED ON TIME-LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	48
FIGURE 30 LANE REDUCTION WITH FEEDBACK CONTROL: INPUT DEMAND AND EXIT FLOWS (THE SECOND ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND-ROW MIDDLE), HEAT-MAP OF SPEED ON THE TIME-LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL	49
FIGURE 31 ONRAMP WITH RL CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND-ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND-ROW MIDDLE), HEAT-MAP OF SPEED ON THE TIME-LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	50
FIGURE 32 ONRAMP WITH FEEDBACK CONTROL: INPUT DEMAND AND EXIT FLOWS (SECOND-ROW LEFT), THE DENSITY MEASURED BY LOOP DETECTOR (SECOND-ROW MIDDLE), HEAT-MAP OF SPEED ON THE TIME-LOCATION DIAGRAM (FIRST ROW), AND AVERAGE VEHICLE SPEEDS IN THE CONTROL SECTION (SECOND ROW RIGHT) WITHOUT A VSL CONTROL.....	51
FIGURE 33 MEDIAN PERFORMANCE OF DIFFERENT PENETRATION RATES FOR 50 SIMULATION.....	52
FIGURE 34 BOXPLOTS OF SENSITIVITY ANALYSIS FOR SAG CURVE BOTTLENECK.....	53
FIGURE 35 BOXPLOTS OF SENSITIVITY ANALYSIS FOR LANE REDUCTION BOTTLENECK.....	54
FIGURE 36 EXTREME CASE OF LANE REDUCTION BOTTLENECK AT MPR OF 5% WITH 67% IMPROVEMENT	55
FIGURE 37 BOXPLOTS OF SENSITIVITY ANALYSIS FOR ON-RAMP BOTTLENECK.....	56
FIGURE 38 BOXPLOTS OF SENSITIVITY ANALYSIS FOR SAG CURVE BOTTLENECK.....	57
FIGURE 39 IMPOSED SPEED-LIMIT OF THE ADAPTIVE RL CONTROLLER.....	58
FIGURE 40 BOXPLOTS OF SENSITIVITY ANALYSIS FOR LANE REDUCTION BOTTLENECK.....	59
FIGURE 41 IMPOSED SPEED-LIMIT OF THE ADAPTIVE RL CONTROLLER.....	60
FIGURE 42 BOXPLOTS OF SENSITIVITY ANALYSIS FOR LANE REDUCTION BOTTLENECK.....	61
FIGURE 43 IMPOSED SPEED-LIMIT OF THE ADAPTIVE RL CONTROLLER.....	62

LIST OF TABLES

TABLE 1 CHARACTERISTICS OF THE CAR-FOLLOWING MODEL	29
TABLE 2 CHARACTERISTICS OF THE LMRS MODEL	32

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

All countries allocate millions of dollars every year on transportation systems to mitigate congestion and reduce travel delays. Traditionally, expanding the infrastructure to increase the capacity seems to be a good idea. However, space, environmental, and economic limitations make this solution impossible. Researchers try to come up with more flexible and less costly solutions such as dynamic traffic management (DTM). The main objective of DTM is to optimize the utilization of the available freeway infrastructure using different types of control measures to influence driving behaviors. To do so, they aim to minimize the total travel time of each vehicle by maximizing the exit flow (Papageorgiou *et al.* 2003). Some of the most popular DTM measures are ramp metering, variable speed limit, route guidance, reversible lanes, and lane advice systems (Hegyi *et al.* 2009, Wolshon *et al.* 2006, Papageorgiou *et al.* 2003, Schakel *et al.* 2014).

Recent advancements in the field of artificial intelligence and unprecedented increase in information flow through the internet has resulted in the exponential growth of automation. Similar to other major industries, transportation is also going through massive changes. Connected and automated vehicles (CAV) and connected vehicles (CV) are changing the future of transport with strong promises for improving safety and mobility (Coppola *et al.* 2016). Presence of these technologies in the daily traffic stream enables engineers to come up with innovative solutions to increase the performance of transportation systems in many ways. The first step is to study how existing solutions would operate in the CV environment. For instance, variable message signs (VMS) are one of the most practical infrastructural tools in transportation operation. Many studies have investigated the implementation of variable speed limit (VSL) algorithms in different contexts where speed limits are communicated to the drivers through VMS (Lin *et al.* 2004; Bertini

et al. 2006; *Hegyi et al.* 2008; *Goni Ros et al.* 2014b). Recent studies have reported that infrastructure-to-vehicle (I2V) communication technologies could be a replacement for the VMS, and CVs can serve as the main conduit for information dissemination (*Lee et al.* 2013b; *Vatani Nezafat et al.* 2018). The wealth of data and information generated by CVs can be used to create artificial intelligence solutions too. Therefore, it is essential to study how various types of traditional infrastructure-centric traffic flow control methods would operate in an intelligent environment. Also, the transition from today's traffic to a fully connected environment will be gradual. Therefore, it is important to come up with solutions for partially connected environments, where the traffic is a mixture of regular vehicles and CVs.

The recent revolutionary progress of deep learning has resulted in a great boost for reinforcement learning (RL) as well. RL is one of the subcategories in machine learning, where an agent interacts with an environment and learns how to maximize its reward. Over the past few years, RL agents have reached a state of art performance in different domains such as playing Atari games (*Mnih et al.* 2015), continuous control (*Lillicrap et al.* 2015), and they were able to beat humans in the board game of Go (*Silver et al.* 2017). They are also helping autonomous vehicle industries for different applications such as improving autonomous driving responses to partially observable scenarios (*Shalev-Shwartz et al.* 2016, *Sallab et al.* 2017). In transportation, RL has attracted many researchers too. There are studies on different domains, and researchers are using RL to address DTM in different problems such as signal timing (*Bakker et al.* 2010, *Nagabandi et al.* 2017), ramp metering (*Belletti et al.* 2018), lane change decision making (*Hoel et al.* 2018), lane reduction (*Vinitsky et al.* 2018), and dissipating stop-and-go waves (*Kreidieh et al.* 2018).

Many researchers have tried to overcome the limitations of feedback controllers and online optimization through the RL approaches (Davarynejad *et al.* 2011, El-Tantawy *et al.* 2010, Zhao *et al.* 2011). Some researchers have investigated the implementation of the RL on VSL in a macroscopic simulation environment to increase the exit flow of the system (Zhu *et al.* 2014, Li *et al.* 2017). This research investigates how VSL through RL control in a CV environment can be a generalized solution for different bottlenecks. The controller only manipulates the speed of CVs, which are driving in the control section. This is referred to as Lagrangian control. Furthermore, it shows that RL will learn to adapt itself with the stochastic arrival of CVs, which will make it a perfect choice for partially connected environments.

1.1) Traffic Dynamics, Modeling, and Control

Traffic congestions can have a substantial impact on society in terms of travel delays, travel time unreliability, fuel consumption, and air pollution (van Wee *et al.* 2013). Researchers try to overcome these problems through DTMs. Hence, it is important to understand the fundamentals of traffic flow. Previous studies show a relation between speed and gap distance of vehicles on empirical observations (Treiber *et al.* 2000). Helbing has pointed out that this relation is dependent on driver and vehicle type. Other studies also indicate that an individual driver's behavior can vary based on weather, the geometry of the road, and traffic state (Hoogendoorn *et al.* 2011, Koshi 2003, Helbing *et al.* 2009). These behaviors can be categorized into longitudinal and lateral components. The longitudinal behavior determines the vehicle acceleration, and the lateral behavior pertains to lane selection and change switching decisions.

The fundamental relation between traffic flow, density and speed on a lane depends on the characteristics of the longitudinal driving behavior. Based on empirical evidence, when the density is low, vehicles drive with free-flow speed. They keep up with it until density reaches a certain

point (critical density). After this point, an increase in density would result in lower speed, and consequently, traffic flow would decrease. Capacity is the maximum flow, which can be reached at the critical density. Many studies have described this fundamental relation as a mathematical model (Greenshields *et al.* 1935, Newell 1993).

In terms of lateral behavior, empirical observations show that drivers choose lanes based on preference, traffic condition, and route requirements. The lateral behavior has an important impact on the dynamic of the traffic stream, and yet it has not been studied as much as longitudinal behavior due to the scarcity of reliable data (Hidas *et al.* 2004, Brackstone *et al.* 1996). The progress of video processing technology has helped to collect high-quality trajectory data such as NGSIM, and it has resulted in the development of better and more sophisticated lateral behavior models (Hoogendoorn *et al.* 2003, Kesting *et al.* 2007, Schakel *et al.* 2012).

Traditionally, capacity is defined when the maximum flow reaches the critical density. However, many studies have noted that the capacity of a given location depends on the stochastic nature of bottleneck. For instance, researchers have found that formation of the queue at upstream of a bottleneck decreases discharge flow significantly (Hall *et al.* 1991; Tilch *et al.* 2000). This difference, which is shown in *Figure 1*, is called the capacity drop. Many researchers have investigated different types of bottlenecks to understand the magnitude of the capacity drop and optimal capacity to decrease the probability of breakdown in microscopic level (Treiber *et al.* 2006; Cassidy *et al.* 2005; Sohrabi *et al.* 2017). They have found that bottlenecks emerge not only at sections with abrupt increases in demand, such as on-ramps, or weaving sections where capacity would drop suddenly (Daganzo 1997) but also at basic sections of freeways, including sag curves and tunnels where drivers lose alertness and decrease their speed unconsciously (Koshi *et al.* 1983).

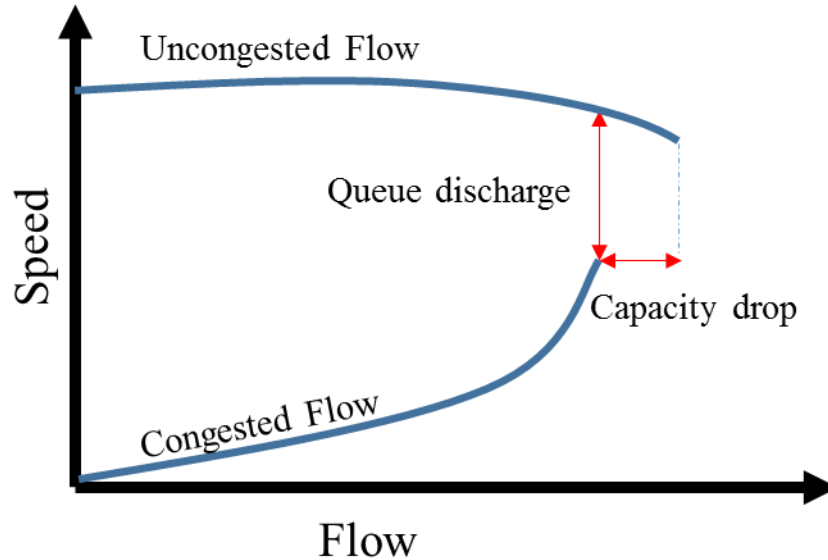


Figure 1 Capacity drop phenomena illustrated on a hypothetical fundamental diagram

Traffic models can be distinguished based on the level of information they provide into different categories such as microscopic, submicroscopic, cellular automata, mesoscopic, and macroscopic ([Ludmann 1998](#), [Nagel *et al.* 1992](#), [Jayakrishnan *et al.* 1994](#)). Microscopic models consider each driver-vehicle as the particle in the traffic stream which is dependent of characteristics of the driver and the vehicle, their interactions to each other, the geometry of the road or environmental conditions such as weather or light. Macroscopic models try to formulate a general relationship between fundamental traffic characteristics (density, flow, speed). They are all based on vehicle conservation law and flow-density-speed relation. Two main types of these models are first-order ([Lighthill *et al.* 1955](#), [Daganzo 1994](#)) and higher-order ([Payne 1971](#), [Spiliopoulou *et al.* 2014](#)).

1.2) Simulation

The main focus of this thesis is to use the abundance of information provided by CVs and create an intelligent environment to control the traffic stream. The stochastic arrival of each CV

has an important role in this environment. Therefore, all of the analyses in this thesis are performed at the microscopic level.

1.2.1) Longitudinal Driving Behavior

There have been many attempts to describe longitudinal driving behavior in a microscopic environment (Chandler *et al.* 1958, Treiber *et al.* 2000, Bando *et al.* 1995, Gipps 1981). They all formulate the model as an ordinary differential equation, which calculates the behavior of any particular vehicle based on the dynamics of the leading vehicle. Models that are more complicated use the information of the vehicle(s) proceeding the leading vehicle to consider the multi-anticipative behavior. In this thesis, one of the most popular and influential microscopic models is adopted, which is called Intelligent Driver Model (IDM) to produce normal longitudinal driving behavior (Treiber *et al.* 2000).

1.2.2) Lateral Driving Behavior

The lane changing behavior of vehicles in a traffic stream can be considered as a multi-step process. The driver knows where her destination in the network is, and characteristics such as on-ramp, off-ramp, or how other mandatory merges would influence the lane choice (Toledo *et al.* 2005). There are many different factors from drivers perspective in lane change decision making, such as maintaining desired speed or adapting to the traffic speed (Keyvan-Ekbatani *et al.* 2016). Drivers would perform an advanced acceleration or deceleration behavior to find a sufficient gap for a desirable and safe lane change (Gipps 1986). If an accepted gap were not available, drivers would indicate their intention of lane changing using turning lights and cooperate with other drivers to create a desirable gap (Schakel *et al.* 2012). In this thesis one of the most recent lane-changing models that consider all of the specifications above is incorporated which is called Lane Changing Model with Relaxation and Synchronization (LMRS) (Schakel *et al.* 2012).

1.2.3) Driving Behavior at Sags

Sag curves are one of the main reasons for bottlenecks in hilly regions. They are defined as a transition section in which slope increases gradually from negative to positive values. They can reduce the capacity of the freeway from 10 to 25 percent depending on the magnitude of positive slope and length of transition from downhill to uphill (Okamura et al. 2000). In 2014, Xing et al. noted that up to 60 percent of bottlenecks on Japanese intercity freeways are because of sag curves. Previous studies showed that drivers reduce their desired speeds at sags (Furuichi et al. 2003; Brilon et al. 2004). In 2012, Yoshizawa et al. reported that when drivers reach a sag curve, they cannot fully compensate for the increase in the slope resulting in poor acceleration behaviors. These behaviors are the main reason for speed reduction. However, Laval put forward that when power to mass ratio is considerably large, the reason may be related to insufficient acceleration capability of drivers.

Over the years, researchers have developed microscopic simulation models to imitate the characteristics of traffic flow at sags. To develop these car-following models, some researchers have assumed the negative effect of gradient on vehicle acceleration. However, this assumption is not consistent with empirical data which show that drivers regain their normal driving behavior after passing vertical curves (Koshi et al. 1992; Komada et al. 2009). To overcome this drawback, other studies have used compensation for the limiting effect, which increase in gradient has on vehicle acceleration (Yokota 1998; Oguchi et al. 2009). Researchers can reproduce the longitudinal driving behavior and traffic dynamics at sag curves accurately. The location of the bottleneck in these models is at the bottom of the curve. However, the empirical data shows that the bottleneck should be located at the end of the curve (Brilon et al. 2004; Patire et al. 2011). More recently, in 2012, Goni Ros et al. used IDM as the base car-following model (Treiber et al. 2000). They have introduced another term in the acceleration equation to incorporate the effect of

compensation for driving behavior on sags. The developed model can capture the effect of compensation on a vertical curve and illustrates drivers being able to regain normal behavior once the driver leaves the curve. It also produces the location of the bottleneck at the end of the curve and beginning of uphill which is consistent with empirical observations. They have assumed drivers would compensate the gradient linearly along with the sags in the direction of uphill.

1.3) Connected Vehicles

Connected and autonomous vehicles are reshaping the future of transportation. Sharing data locally with other vehicles or roadside infrastructures helps researchers to come up with new solutions to optimize efficiency (Milanés *et al.* 2014; Goodall *et al.* 2013) and increase the safety of transportation networks (Olia *et al.* 2016). In a study by Talebpour *et al.*, the possibility of shockwave detection through CVs in a microsimulation model have been investigated. Some researchers have used CVs equipped with a speed advisory system to minimize idling at signals (Malenstein 1998; Rakha *et al.* 2011). More recently, Ramezani *et al.* developed an optimization program using CV's environment to determine advisory speeds for connected vehicles in work zones. The CVs program of the USDOT is one of the relatively new technologies that allow vehicles to link directly to its surrounding environment. This technology provides communication between vehicles that are close together, and between vehicles and the nearby infrastructure on the road. The goal of these interactions is to increase the safety, efficiency, and the mobility of the transportation network. Therefore, the Institute of Electrical and Electronics Engineers (IEEE) has proposed a modified version of the Wireless Local Area Network (WLAN) protocol for V2V and I2V communications which is called Dedicated Short-Range Communication (DSRC). The FCC has allocated a dedicated bandwidth of 75MHz in the 5.850 to 5.925GHz band for the DSRC. The maximal communication distance is around 300 meters (Xu *et al.* 2017). Some automakers are already installing DSRC devices in their new vehicles that allow the V2V and I2V communication

to increase safety (Lukin et al. 2006). Despite all of the advantages, it lacks scalability, which means in dense traffic, the protocol is not reliable to provide time-probabilistic characteristics (Lee et al. 2013a). An alternative option to DSRC is the 5G-LTE that is a new, under development, cellular wireless infrastructure. It has the potential to be redesigned as a communication basis for CVs. It offers low latency and high throughputs, but it increases bandwidth demands and needs for real-time critical services (Bailo et al. 2018).

1.4) Dynamic Traffic Management via Variable Speed Limit

The VSL is a flexible speed restriction on a given length of the road. The speed limit changes according to the current traffic state or environmental conditions. Many control strategies have been proposed for the operation of VSL (Chen et al. 2015, Liu et al. 2015, Carlson et al. 2010). Some researchers use online optimization (Hegyi et al. 2005; Kwon et al. 2007; Zegeye et al. 2012; Pasquale et al. 2016). They formulate a VSL on a freeway a constrained discrete-time optimal control problem, which can be solved by open-loop optimal control. This approach can theoretically reach optimum system performance. However, an accurate prediction of traffic flow is necessary for the open-loop optimal solution. Such models need a huge amount of computation. Hence, it is hard to make it practical for large-scale applications. Others use the feedback control approach (Popov et al. 2008; Carlson et al. 2011; Iordanidou et al. 2017). In this approach, the control strategy maximizes flow by automatically adjusting the speed limits to keep the controlled variable, i.e., bottleneck density, to be as close as possible to the desired target value. Since this model relies on real-time measurements of traffic conditions, and it does not need predictions, the VSL strategy is more efficient and robust to actual traffic conditions. However, they are unable to take actions unless an oscillation in the controlled variable is observed. As a consequence of being reactive to observed changes, there would be a delay between the oscillation in the traffic stream

and the response of the controller. This can create drastic changes in VSL when there is large variation in traffic demand (Aström *et al.* 2010).

1.5) Deep Reinforcement Learning

Reinforcement learning (RL) is one of the major sub-disciplines in machine learning alongside with supervised and unsupervised learning. The main aim of RL is to understand how an agent can learn to take actions in an environment to maximize its cumulative reward. The environment can be formulated as a Markov Decision Process (MDP). The agent takes actions based on the current state of the environment. For each action that agent takes, the environment responds back with the new state and a reward signal. The reward signal has a key role in the learning process of the agent. Generally, the reward signal is presented by a simple scalar, which is the only way the agent evaluates its actions. Traditionally, when the environment has a small state space, and its stochastic nature is simple, dynamic programming approaches can be used to solve the problem. However, for more complex environments, it's not feasible.

Therefore, one of the alternatives is to utilize Q-Learning. The main idea is to memorize optimal action in every state through trial and error. This approach also becomes infeasible due to the expansion of state space. To overcome this drawback, instead of storing Q values for all possible states-actions combination, we can build a Q function estimator to approximate the value. One of the best available function estimator approaches is the Artificial Neural Network (ANN) which can build nonlinear models for complex environments (Sutton *et al.* 1998).

The ANN idea has been around since 1960. There have been many studies using ANN as a machine learning approach in all engineering fields, including transportation community (Faghri *et al.* 1992). The advantage of ANN over classical machine learning approaches such as SVM is the ability of this method to perform feature extraction and selection automatically. To get the best

performance out of classical methods, researchers need to select the best features to represent the data, which is time-consuming and involves some heuristic procedures. In a sense, for classical methods, some part of the learning has to be done by the researcher. Up to a few years ago, the performance of ANNs and classical methods were almost the same. With the recent advancements in computational power and an increase in the accumulation of data, researchers have noticed an interesting pattern in the performance of machine learning algorithms. The performance of ANNs increases rapidly, with more data, while the performance of classical methods would not get better after a certain point. This observation has led to a significant increase in new studies about ANNs in the field of computer science. For some tasks, such as image retrieval (Ku et al. 2015), object detection, and tracking (Girshick 2015, Ren *et al.* 2015) ANN has reached a state of the art performance, exceeding the performance of competing methods. The detailed incorporation of RL and ANN is explained later on in the methodology section of this study.

1.6) Research objective and research questions

The main objective of this thesis is to develop new concepts for DTM at freeway bottlenecks and to create more general solutions using artificial intelligence and rich information of the CV environment. To achieve this objective, this thesis proposes to answer the following research questions:

1. Given a CV environment, can we develop VSL control algorithms that are robust and applicable to multiple types of freeway bottlenecks? In particular, how can RL methods be leveraged and designed to improve throughput at these bottlenecks through VSL control?
2. How effective are these VSL algorithms, and to what extent will they mitigate congestion?

3. In what aspects the RL control strategies and outcomes differ from those of feedback control methods?
4. How effective will these algorithms perform under varying market penetration of CVs?
5. Is the RL algorithm able to adapt its controlling strategy to different market penetration of CVs

1.7) Research approach and research scope

Microscopic traffic simulation models have been developed for three types of bottlenecks. These include an on-ramp, lane reduction, and a sag curve, as shown in *Figure 2*, *Figure 3*, and *Figure 4*. The simulation models were integrated with RL algorithms to answer the research questions above. Different approaches to RL has been investigated to find the optimal solution for each bottleneck. Also, a feedback-based controller was developed to be compared with the RL controller and answer the third question. Finally, extensive sensitivity analyses have been conducted to understand the impacts of the penetration rate of CVs on system performance.

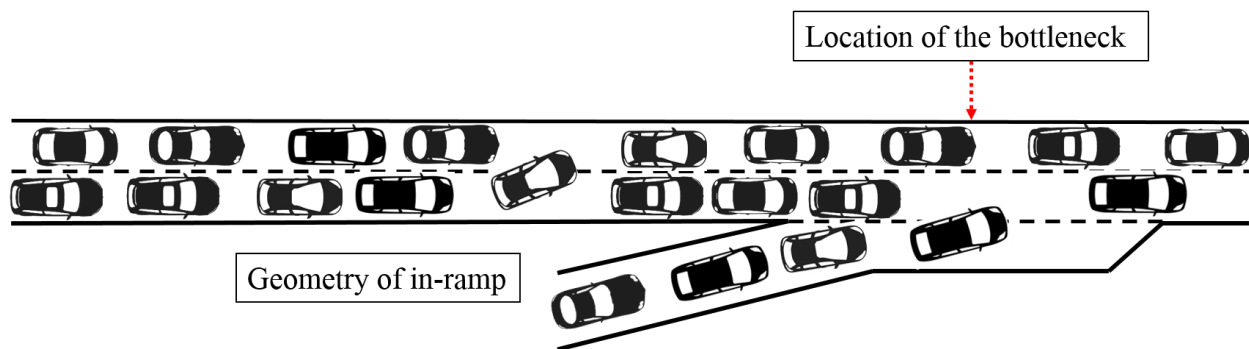


Figure 2 In-ramp bottleneck

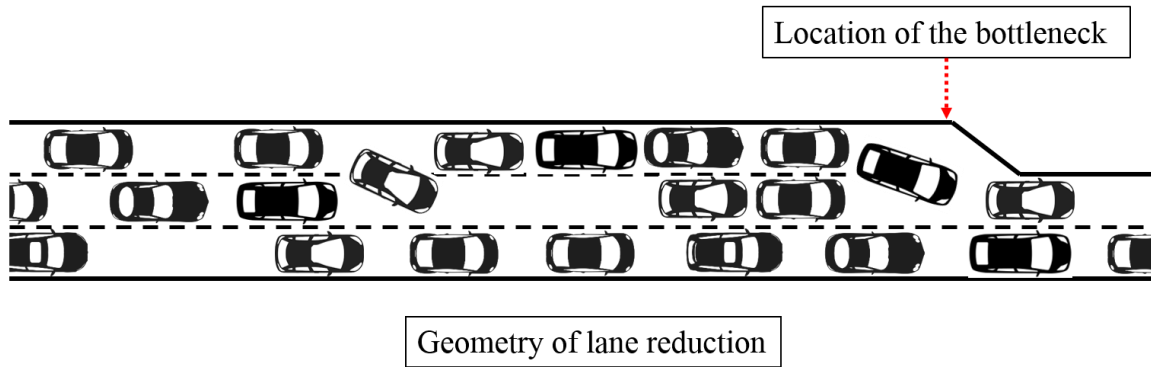


Figure 3 Lane reduction bottleneck

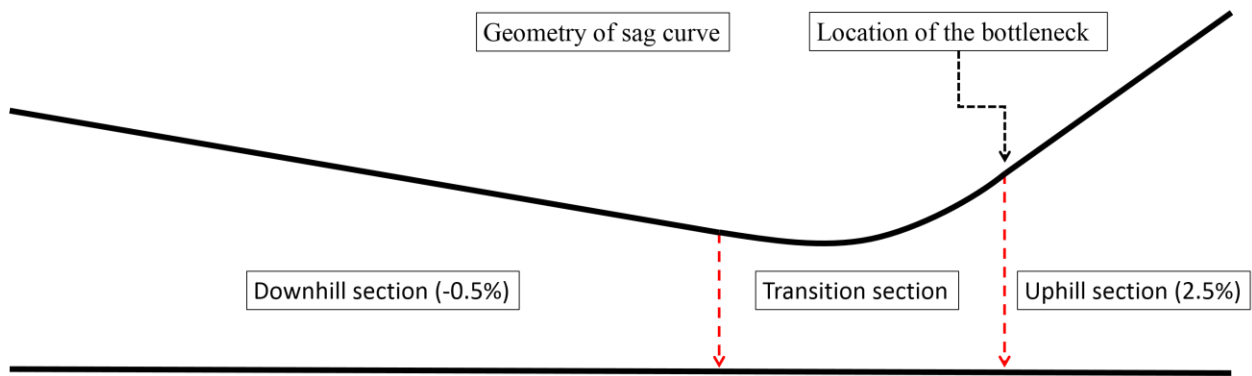


Figure 4 Sag curve bottleneck

The focus of this research is on freeways, where traffic flows in one direction. The reason is to have a complete understanding of the impact of RL control measures on isolated bottlenecks, which are common in freeway networks. In contrary, in urban areas, traffic flow usually breaks down as a result of multiple types of bottlenecks such as intersections. It is crucial noting that the control strategy developed in this thesis could be extended to more complex networks with multiple bottlenecks.

This thesis investigates traffic management strategy using RL in a CV environment for the on-ramp, sag curve, and lane reduction, which are common types of bottlenecks on freeways. The thesis focuses on minimizing the total delay. Moreover, the thesis investigates the impact of the

penetration rate of CVs on the performance of the management strategy. It is considered that the demand profile is given. Other objectives, such as safety, environmental impacts, maintenance, implementation, and road design, are beyond the scope of this study.

1.8) Main contributions

Main contributions of this research are listed as follows:

1. Guidelines for developing DTM measures through the RL algorithm. The thesis develops an RL controller for VSL that is adaptive to the stochastic arrival of CVs.
2. The thesis demonstrates how to generalize the RL controller for VSL to mitigate congestion at different freeway bottlenecks. The method will be generalizable in the sense that one controller can learn to optimize the traffic exit-flow of various bottlenecks. The proposed method can be used to develop new DTM concepts.
3. An extensive sensitivity analysis has been conducted to demonstrate the impact of the penetration rate of CVs on the performance of the RL controller.

1.9) Outline of the dissertation

The remainder of this thesis describes all models and details to demonstrate how the simulation environment and RL controller are developed (Chapter 2). Results and performance of the RL approach for different bottlenecks and different penetration rates are presented in Chapter 3, and finally, Chapter 4 presents the conclusions of the thesis and summarizes the main findings. It discusses the meanings of results and potential future works.

CHAPTER 2: METHODOLOGY

2.1) Longitudinal Driving Behavior

To perform the analyses, a micro-simulation model is developed in Python for a freeway segment which takes into account vertical curves on the road. Therefore, the longitudinal driving behavior is one of the critical components of this simulation model. In this study, the model Goni Ros *et al.* proposed is used. This model can produce normal driving behavior on a basic section of a freeway. It also accounts for the influence of vertical curves on vehicle acceleration. It calculates acceleration from the summation of two terms, as presented in Equation 1. The first component corresponds to car-following behavior, which is a modified version of IDM (Treiber *et al.* 2000) and the second one calculates behavior on uphill.

$$v' = fr(t) + fg(t) \quad (1)$$

The first acceleration term uses speed (v), relative speed (Δv), the desired speed (v_{des}), and spacing to the vehicle ahead (s) to calculate acceleration for the following car.

$$fr(t) = \alpha \times \min \left[1 - \left(\frac{v(t)}{v_{des}} \right)^4, 1 - \left(\frac{s_{des}(v(t), \Delta v(t))}{s(t)} \right)^2 \right] \quad (2)$$

In Equation 2, s_{des} is the desired spacing which is computed using Equation 3. The main influencing factor is the safe gap to the lead vehicle.

$$s_{des}(v(t), \Delta v(t)) = s_0 + v(t) \cdot \tau(v(t)) + \frac{v(t) \cdot \Delta v(t)}{2\sqrt{\alpha \cdot b}} \quad (3)$$

The parameter α is the maximum acceleration, b is the maximum comfortable deceleration, s_0 is the gap in the standstill situation, and τ is the safe time headway as a function of speed. Based on the traffic state, the safe time headway (τ) changes, as shown in Equation 4.

$$\tau(v(t)) = \begin{cases} \tau_f & v(t) \geq v_{crit} \\ \gamma \cdot \tau_f & v(t) < v_{crit} \end{cases} \quad (4)$$

The second term ($fg(t)$) in Equation 1 captures the influence of gradient on vehicle acceleration. This influence is equal to the difference between the gradient at the position of the vehicle ($G(x(t))$) and the compensated gradient by the driver at the time ($G_c(t)$) multiplied by gravity acceleration. This is shown in Equation 5.

$$fg(t) = -g(G(x(t)) - G_c(t)) \quad (5)$$

It is assumed that drivers would compensate linearly for any increase in freeway gradient with maximum gradient compensation rate defined by parameter c .

$$G_c(t) = \begin{cases} G(x(t)) & G(x(t)) \leq G(t_c) + c(t - t_c) \\ G(t_c) + c(t - t_c) & otherwise \end{cases} \quad (6)$$

Where:

$$t_c = \max[t | G_c(t) = G(x(t))] \quad (7)$$

If the increase in grade over time is lower than c , then $G_c(t)$ is equal to $G(x(t))$ and $fg(t)$ is zero. Hence, the acceleration of the vehicle is not affected, and the driver fully compensates for the gradient. On a basic section of the freeway where there is no vertical curve, the second term of Equation 1 would be zero, and the model produces normal driving behavior.

2.2) Lateral Driving Behavior

The LMRS algorithm assumes that the lane-changing behavior depends on the level of desire the driver has. It's a single value derived from a combination of incentive to follow the route (d_r), gain speed (d_s) and keep right (d_b). The desire to change from lane i to lane j is calculated using Equation 8.

$$d^{ij} = d_r^{ij} + \theta_v^{ij} \cdot (d_s^{ij} + d_b^{ij}) \quad (8)$$

The d_r^{ij} is the incentive to follow the road. The d_s^{ij} is the speed incentive. The d_b^{ij} is the incentive to keep right and the θ_v^{ij} is a factor to incorporate discretionary incentives. The desire is an ordinal value between -1 to 1 and depending on the range; it falls the lane changing behavior changes. Negative values of desire mean lane changing is not desired. As it is shown in *Figure 5*, there are four regions for lane changing. The first region is where no lane changing happens. The second region is called free lane changing, which means if its possible, lane changes are initiated without any preparation. The third region is the synchronization area. The vehicle starts synchronizing its speed with the target lane. This goal can be achieved by following the leading vehicle on that lane. If still, the driver was not able to find an applicable gap to perform lane changing the desired value will pass d_{coop} threshold and the vehicle starts cooperative behavior. Using turn indicator, other vehicles will know the lane changing intention, and the potential follower starts creating a gap by following the potential lane changer.

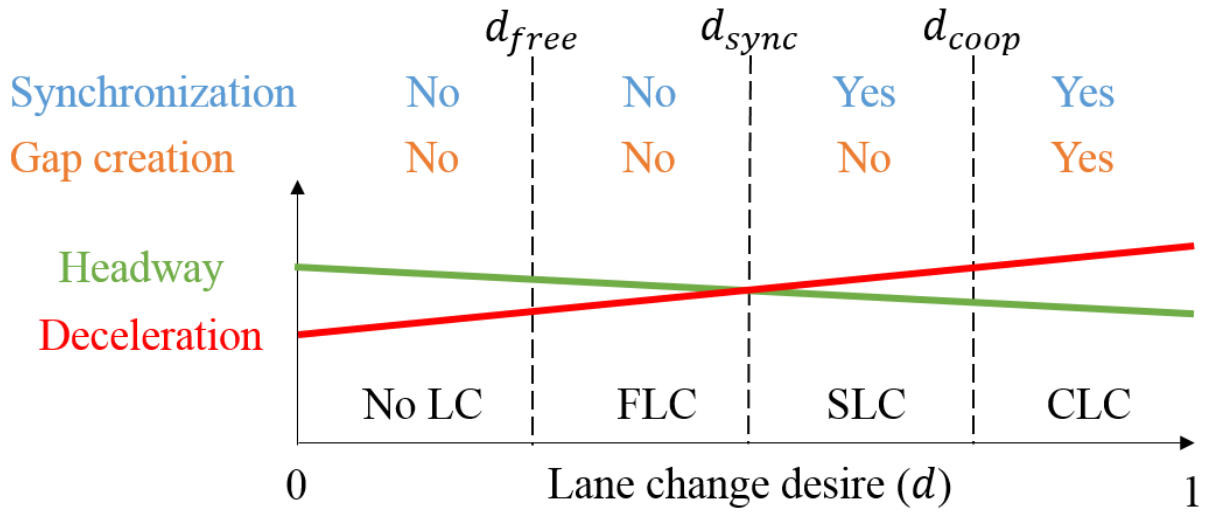


Figure 5 Overview of the spectrum of desire and lane changing regions

2.2.1) Lane Change Incentives

As we discussed before the desire to change lane depend on different incentives. The speed incentive (d_s^{ij}) represent driver's anticipation for speed of vehicles in downstream of the lane. The

anticipation speed (v_{ant}) is a function of speed limit (v_{lim}), maximum vehicle speed (v_{max}) and the speed of potential leading vehicle (v_{lead}) on the target lane. Using v_{lim} and v_{max} the desired speed at lane k is calculated through Equation 9. The δ represent the level of adherence to the speed limit. If it's over one, it results in speeding, and if it's less than one, it results in the opposite.

$$v_{des}^k = \min(\delta \cdot v_{lim}^k, v_{max}) \quad (9)$$

The speed anticipation of leading vehicle (v_{lead}) is calculated as Equation 10. It also depends on how far the leading vehicle is. The s is the current distance headway and x_0 shows to what distance are we considering a leading vehicle in our anticipation.

$$\tilde{v}_{lead} = \left[1 - \frac{s}{x_0}\right] \cdot v_{lead} + \frac{s}{x_0} \cdot v_{des} \quad (10)$$

Finally, the anticipated speed for lane k is calculated using Equation 11. All vehicles on subset m are considered as the leading vehicle. Subset m includes all vehicles on lane k and all vehicles which want to be on lane k ($d^{k-1,k}$ or $d^{k+1,k} \geq d_{coop}$). However, vehicles who want to leave lane k ($d^{k,j} \geq d_{coop}$) are excluded.

$$v_{ant}^k = \min(v_{des}^k, \min(\tilde{v}_{lead}^m)) \quad (11)$$

Using the anticipated speed, the speed incentive can be calculated through Equation 13 & 14, respectively, for left and right lane change. Since we now desire to change lane decrease during acceleration, the α_{gain} is derived from Equation 12 as a reduction factor for the speed incentive. The Δ_s indicates if the lane change is allowed ($\Delta_s = 1$).

$$\alpha_{gain} = \frac{\alpha - \max(\dot{v}, 0)}{\alpha} \quad (12)$$

$$d_s^{i,i-1} = \begin{cases} \alpha_{gain} \frac{v_{ant}^{i-1} - v_{ant}^i}{v_{gain}} & \Delta_s^{i,i-1} = 1 \\ 0 & \Delta_s^{i,i-1} = 0 \end{cases} \quad (13)$$

$$d_s^{i,i+1} = \begin{cases} \alpha_{gain} \frac{\min(v_{ant}^{i-1} - v_{ant}^i, 0)}{v_{gain}} & \Delta_s^{i,i+1} = 1 \text{ and } v_{ant}^i > v_{crit} \\ \alpha_{gain} \frac{v_{ant}^{i-1} - v_{ant}^i}{v_{gain}} & \Delta_s^{i,i+1} = 1 \text{ and } v_{ant}^i \leq v_{crit} \\ 0 & \Delta_s^{i,i+1} = 0 \end{cases} \quad (14)$$

To calculate route incentive, first the desire to leave route k is calculated through Equation 15. The x_r^k is the remaining distance on route k , $t_r^k = x_r^k / v$ is the remaining time with current speed and the n_r^k is several required lane changes. Then the route incentive is calculated using Equation 16. The $\Delta_r^j = 1$ indicates that route j can still be followed.

$$d_r^k = \max\left(1 - \frac{x_r^k}{n_r^k \cdot x_0}, 1 - \frac{t_r^k}{n_r^k \cdot t_0}, 0\right) \quad (15)$$

$$d_r^{ij} = \begin{cases} d_r^i & \Delta_r^j = 1 \text{ and } d_r^i > d_r^j \\ 0 & \Delta_r^j = 1 \text{ and } d_r^i = d_r^j \\ -d_r^j & \Delta_r^j = 1 \text{ and } d_r^i < d_r^j \\ -\infty & \Delta_r^j = 0 \end{cases} \quad (16)$$

The keep right incentive is a constant bias towards the right lane in accordance to “keep right if possible” rule. This incentive is calculated through Equation 17.

$$\begin{aligned}
d_b^{i,i-1} &= 0 \\
d_b^{i,i+1} &= \begin{cases} d_{free} & v_{ant}^{i+1} = v_{des} \text{ and } d_r^{i,i+1} \geq 0 \\ 0 & \text{otherwise} \end{cases}
\end{aligned} \tag{17}$$

If a mandatory lane change is urgent, voluntary lane change incentives will be ignored through θ_v^{ij} which is calculated using Equation 18.

$$\theta_v^{ij} = \begin{cases} 0 & d_r^{ij} \cdot d_v^{ij} < 0 \text{ and } |d_r^{ij}| \geq d_{coop} \\ \frac{d_{coop} - |d_r^{ij}|}{d_{coop} - d_{sync}} & d_r^{ij} \cdot d_v^{ij} < 0 \text{ and } d_{sync} < |d_r^{ij}| < d_{coop} \\ 1 & d_r^{ij} \cdot d_v^{ij} \geq 0 \text{ or } |d_r^{ij}| \leq d_{sync} \end{cases} \tag{18}$$

2.2.2) Gap Acceptance and Relaxation

If the desire to change lane is high enough ($d^{ij} \geq d_{free}$). The vehicle will look for an acceptable gap to perform the lane change. The gap is accepted if both lane changer (c) and a new follower (f) have acceleration higher than a safe deceleration threshold as it is shown in Equation 19.

$$\dot{v}^g \geq -b^c \cdot d^{ij,c} \quad g \in \{c, f\} \tag{19}$$

Since vehicles tend to accept shorter headways for lane changing, Equation 20 updates acceptable headway used in Equation 3 for both lane changer and new follower. The $\langle d^{ij,c} \rangle$ means that desire is bounded between zero and one. Then the acceptable headway of the vehicle will go back to its value over time through Equation 21. This is called relaxation, and it is assumed that the relaxation of headway is exponential with relaxation time τ .

$$T^g(d^{ij,c}) = \min(T^g(d^{ij,c}), \langle d^{ij,c} \rangle \cdot T_{min}^g + (1 - \langle d^{ij,c} \rangle) \cdot T_{max}^g) \quad (20)$$

$$T(t) = T(t - \Delta t) + \{T_{max} - T(t - \Delta t)\} \frac{\Delta t}{\tau} \quad (21)$$

2.2.3) Synchronization and Gap Creation

When the desire to change lane is higher than the synchronization threshold, the vehicle starts synchronizing its speed with the leader on the target lane. Using the car-following model, the new acceleration is calculated, which has to comply with Equation 22. The maximum deceleration of b is the upper bound of synchronization acceleration, which is both comfortable and safe.

$$\dot{v}_{sync}^{ij} > -b \quad (22)$$

If the desire passes the cooperation threshold, the vehicle informs adjacent follower by turn light, and the new follower will create a gap by following the lane changer constrained by Equation 22. Again, the maximum deceleration of b is the upper bound of cooperation acceleration.

2.3) Control Strategy

The objective of the control strategy is to eliminate congestion in a bottleneck and improve the performance of highways. For networks not influenced by other control measures, minimizing the total time that vehicles spend in the system is equivalent to maximizing the exit flow (Papageorgiou *et al.* 2003). As mentioned previously, the capacity of the freeway on a bottleneck ($q_{bottleneck}$) is less than other sections ($q_{Capacity}$). Therefore, the network's exit flow is bound by the capacity of the bottleneck.

$$q_{Exit} \approx q_{bottleneck} < q_{Capacity} \quad (23)$$

One way to maximize the exit flow is to prevent traffic from becoming congested at the bottleneck. Keeping the traffic state uncongested at the bottleneck is possible if the inflow of it gets regulated at a controlled section at the upstream. The inflow of bottleneck is approximately equal to the outflow of the control section, and per the fundamental relation between speed and flow, changing speed on control section changes the inflow of the bottleneck. By dynamically modifying the speed at control section, it is possible to keep the inflow to the bottleneck slightly below its free flow capacity. It will increase the time-weighted sum of the exit flow. When the demand in the upstream is large enough, the congestion would not be prevented entirely. As a result, the control section and upstream would become congested instead of the bottleneck, but the outflow from the controlled part will be higher than the queue discharge capacity of the bottleneck.

2.3.1) Deep Reinforcement Learning

The RL algorithm learns from observations to maximize the reward of an agent by interacting with an environment through possible actions. The action space can be discrete or continuous, and it is taken from a range of actions called A . At time t , the action is represented by a_t , the state as s_t , and the reward as r_t . The main goal of the agent is to learn a policy that maximizes the expectation of future rewards. This policy is represented by $\pi(a|s)$ which calculates the probability of taking action a given the state s . The impact of future rewards is represented by a function called return, which is calculated as in Equation 24. The parameter γ is called discount factor which takes into account the importance of future rewards. It is usually considered to be a constant number between 0.8 to 0.99.

$$R_t = \sum_t^T r_t + \gamma r_{t+1} + \dots + \gamma^{T-t} r_{T-t} \quad (24)$$

One way to estimate our target $\pi(a|s)$ is to estimate the value functions of the policy. There are two different value functions, as they are shown in Equations 25 & 26. One is the state-value function, which calculates the expected return of being in state s , and the other one is action-value function. Similarly, it calculates the expected return of taking action a in state s .

$$V_{\pi}(s_t) = E[R_t|s_t] \quad (25)$$

$$Q_{\pi}(s_t, a_t) = E[R_t|s_t, a_t] \quad (26)$$

The best policy has the optimal action-value function which is represented by $Q^*(S_t, a_t)$. By using this function one can simply choose $a^* = \max_{a \in A} Q^*(s, a)$ at every state and solve the problem. If the environment has a finite discrete state and finite discrete action space, it would be easy to represent Q-values in a table such as the one shown in *Figure 6*. Value-based methods such as tabular Q-learning try to estimate the Q-value function through Equation 27 and 28, which is known as Bellman equation ([Sutton et al. 1998](#)). It is self-evident that an increase in state and action space dimensions makes it impossible to approximate Q-value function through a table.

Q-Learning	a1	a2	...	an
S1				
S2				
S3				
.
Sn				

Figure 6 Q-values table example

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha[y_t - Q_t(s, a)] \quad (27)$$

$$y_t = r_t + \gamma \max_{a_{t+1}} Q_t(s_{t+1}, a_{t+1}) \quad (28)$$

By leveraging deep neural networks, instead of estimating individual Q-values for each state-action pair, we can approximate a function that maps from states to Q-values of different actions. This approach is called Deep Q-Network (DQN) ([Mnih et al. 2015](#)). Since deep neural networks

fluctuate a lot in the process of training, the agent will end up chasing a moving target, and it may never converge. Therefore, Mnih et al. came up with the idea of using two separate neural networks for the Q-value estimation, as illustrated in *Figure 7*. A value network that estimates Q-value of state-action pairs and a target network which is used to estimate $\max_{a_{t+1}} Q_t(s_{t+1}, a_{t+1})$ and help the calculation of y_t in Equation 28. The target network will be frozen and get updated only after a period of time which makes the targets partially stabilized.

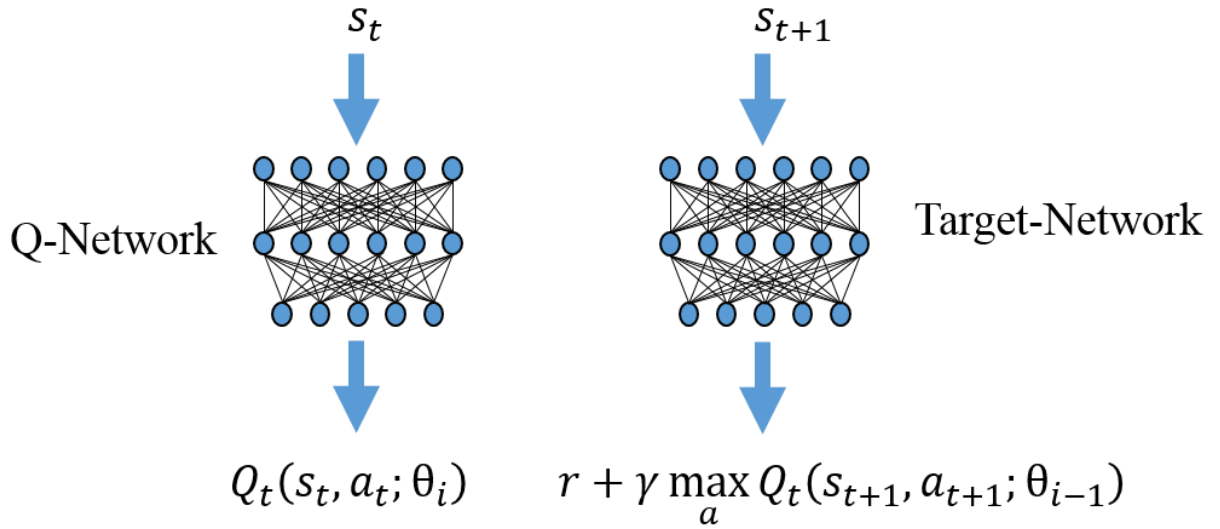


Figure 7 Q-Network (left) used to calculate Equation 27, and the Target-Network (right) is used to calculate Equation 28

An alternative approach to the value-based method is a policy-based method, where an agent will directly find the best policy instead of value functions (Sutton et al. 2018). One of the policy-based methods is called Reinforce (Williams 1992). The algorithm updates parameters in a way that the probability of actions that resulted in higher returns would increase. It approximates the best policy with the function $\pi(a|s; \theta)$ and updates parameters with Equation 29. Since the cumulative reward is needed, this algorithm only updates at the end of each episode.

$$\theta_{t+1} = \theta_t + \alpha \cdot R_t \frac{\partial \log \pi(a|s; \theta)}{\partial \theta} \quad (29)$$

There are hybrid methods combining value-based and policy-based methods. These are called actor-critic methods. These methods optimize the policy (*actor*, $\pi(a|s; \theta^\pi)$) directly but they also estimate the value function (*critic*, $v(s_t; \theta^v)$) to reduce the variance of updates (Sutton et al. 2018). They also allow bootstrapping, which means there is no need to wait for the end of an episode to update parameters. Instead of using return function (R_t) they use Advantage (A_b) as it is shown in Equation 30. The parameter b is called bootstrapping interval.

$$A_b = r_t + \gamma r_{t+1} + \dots + \gamma^{b-t} r_{t+b-1} + v(s_{t+b}; \theta^v) - v(s_t; \theta^v) \quad (30)$$

The parameters of the actor and critic models can be updated by Equation 29. These updates for actor $\pi(a|s; \theta^\pi)$ and critic $v(s_t; \theta^v)$ are shown in Equations 31 & 32.

$$\theta_{t+1}^\pi = \theta_t^\pi + \alpha_\pi \cdot A_b \frac{\partial \log \pi(a|s; \theta^\pi)}{\partial \theta^\pi} \quad (31)$$

$$\theta_{t+1}^v = \theta_t^v + \alpha_v \cdot A_b \frac{\partial v(s; \theta^v)}{\partial \theta^v} \quad (32)$$

Since the process is sequential, the experiences that the agent collects are highly correlated, which is a big issue for a deep neural network to approximate a function from correlated data. One way to break this correlation is experience replay (Mnih et al. 2015). In this approach, the agent will not update the network with immediate experience. Instead, it will store tuples of $[s_t, a_t, r_t, s_{t+1}]$ in a buffer and updates its network using a random sample. This approach costs memory and it forces using old experiences. Another solution is to have asynchronous agents (Mnih et al. 2016), which is adopted in this study.

For the training of the algorithm, the multi-threading ability of Python has been used to compute the interaction of agents and environments in parallel. The parallel computation through multi-threading can be implemented as a synchronous procedure which is called Advantage Actor-Critic (A2C) or as an asynchronous procedure, which then will be called Asynchronous Advantage

Actor-Critic (A3C). For the synchronous implementation (*Figure 8 left*), when all workers (agents) are done, their experiences would be collected and passed to the global network. The global network is a copy of actor and critic which perform optimization of network parameters through gradient descent. After the updates, new network parameters will be passed to all workers, and they start running another episode on the environment. It should be noted that these experiences are highly correlated and to break this correlation experience replay buffer is needed. For the asynchronous implementation (*Figure 8 right*), all agents begin an episode together, but whoever finishes first will go to the global network with a collection of experiences. The global network receives experiences from the worker, updates network parameters, and passes new parameters to the worker to start a new episode. In this approach, workers interact with the environment and the global network asynchronously. Therefore, each one uses the most recent parameters calculated by the global network, and they don't have a correlation.

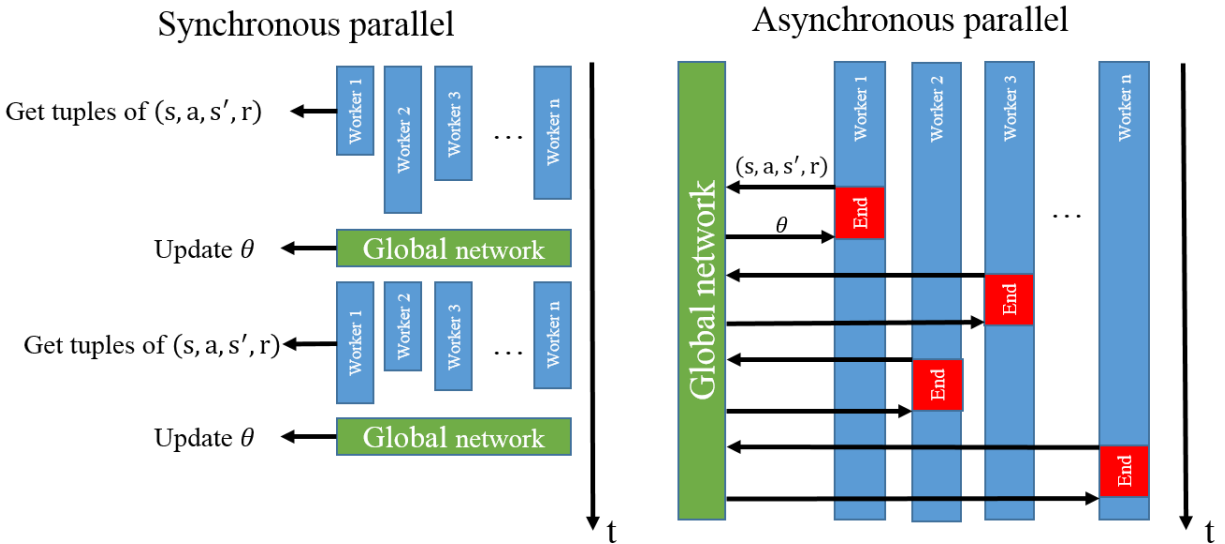


Figure 8 A2C (left) and A3C (right) parallel procedure

The *Figure 9* illustrates the procedure of updates by the global network. First, through the reward and the next state, the critic will be updated. Using the new critic, the advantage function

will be calculated. The advantage value then would be used to calculate the gradient and update the actor.


- 
1. Correspond to action $a \sim \pi_\theta(a|s)$, receive the experience set (s, a, s', r) from the worker
 2. Update v_θ^π using target $r + \gamma v_\theta^\pi(s') \rightarrow$ Update the Critic
 3. Evaluate $A^\pi(s, a) = r(s, a) + \gamma v_\theta^\pi(s') - v_\theta^\pi(s) \rightarrow$ Update the Advantage
 4. $\nabla_\theta J(\theta) \approx \nabla_\theta \log \pi_\theta(a|s) A^\pi(s, a) \rightarrow$ Calculate the gradient
 5. $\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta) \rightarrow$ Update the Actor

Figure 9 Update procedure of the global network

2.3.2) Feedback Control

The controller which calculates a speed limit for the control section is inspired by the ramp-metering control algorithm called ALINEA, which is based on a proportional feedback control law (Papageorgiou *et al.* 1997). It calculates the variable speed limit from Equation 33. The target density (ρ_{Target}) is slightly lower than the critical density of the fundamental diagram, and real-time density (ρ_b) is the estimated density at the bottleneck calculated every T_c seconds. The algorithm would change the speed limit as a proportion (κ) of the difference between target and measured density every time that a new density is calculated.

$$v_{Limit}(t) = v_{Target} + k \times (\rho_{Target} - \rho_b(t - 1)) \quad (33)$$

As evident from Equation 33, in high demand conditions, the controller would keep the density at bottleneck close to target density to prevent breakdown. Whenever demand decreases, the measured density would be significantly less than target density which leads the controller to impose a higher speed limit and, in contrary, if demand increases measured density it would be substantially more than target density which leads the controller to enforce a lower speed limit. The controller always uses the previously estimated density so that drivers would have enough time to cover the distance between the control section and the bottleneck.

2.4) Environments

Three environments have been developed in Python for training the RL algorithm. The control section is 1.0 km long. In this section, only CVs would be informed of the imposed speed limit. Notably, it is assumed that all CVs would comply completely. As soon as connected vehicles leave the control section, their speed will revert to the default desired speed. In other words, the desired speed of CVs is only varied while they are within the control section. The description of the simulation setup for each environment is explained in details as follows.

2.4.1) Sag Curve

For the sag curve bottleneck, the investigated network contains a single-lane freeway with a sag in the middle. The length of the network is 12 km. The road starts with a constant-gradient downhill section followed by a vertical sag curve, and at the end, a constant-gradient uphill section (*Figure 10*). The downhill section has a constant gradient equal to -0.5 percent and the uphill section has a constant slope equal to 2.5 percent. At the vertical sag, the slope increases linearly from -0.5 to +2.5 percent, and the length of the vertical curve is 0.6 km between $x = 10.7$ km and $x = 11.3$ km. The downhill section is long enough to make sure the queue would not reach the entry point of the simulation.

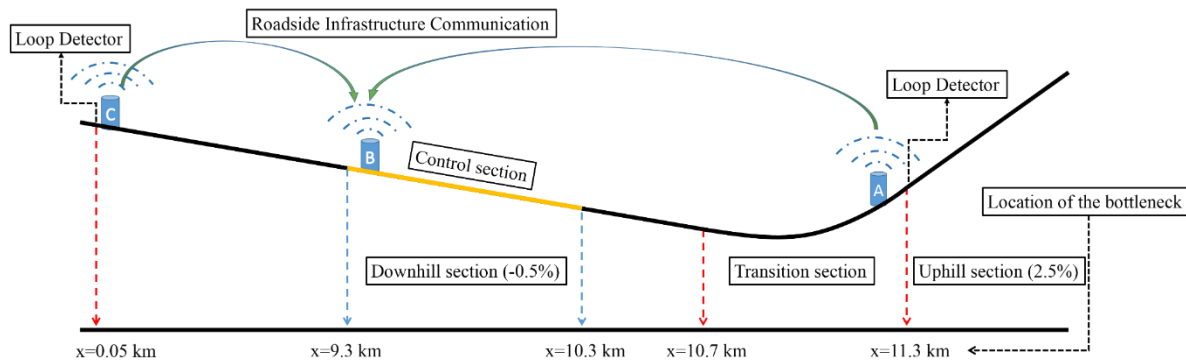


Figure 10 The geometry of the sag curve bottleneck

The speed limit is assumed to be 120 km/h. Characteristics of vehicles and drivers, as defined by the IDM model, are assumed to be homogeneous to prevent the emergence of other types of bottlenecks in the simulation. These model parameters are shown in Table 1.

Table 1 Characteristics of the car-following model

$v_{des} (km/h)$	$a(m/s^2)$	$b(m/s^2)$	$\tau(s)$	$s_0(m)$	$v_{crit} (km/h)$	$\gamma(-)$	$c(s^{-1})$	$\Delta t(s)$
120	1.45	2.1	1.2	3	65	1.15	0.0001	0.5

The control section is between $x = 9.3$ km and $x = 10.3$ km. The downstream end of the controlled section is 0.4 km away from the beginning of the transition section. The distribution of demand over time is illustrated in *Figure 11*. The first 5 minutes is a transition from zero to 2400 veh/h, a capacity higher than the bottleneck capacity. The demand stays at 2400 veh/h for 10 minutes then transitions back to zero across 5 minutes. Beyond this, the demand remains zero until all vehicles have exited the facility.

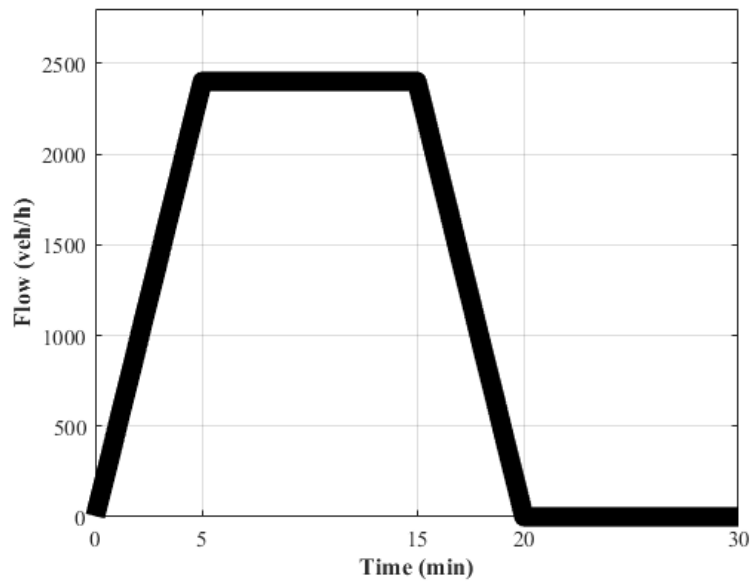


Figure 11 Demand profile over time

With the given parameters above, a microsimulation model was created in Python. In *Figure 12*, a heat map which shows the average speed of vehicles for different times and locations is presented (left chart) for the base case, i.e., when no control strategy is implemented. At the very beginning of the simulation, the effect of the uphill is not significant enough to cause a breakdown at the bottleneck. After a while, a shockwave starts to propagate backward, starting at the bottleneck with constant speed. Since the breakdown is due to the geometry of the road, this shockwave continues to propagate until in-flow decreases. The second shockwave emerges shortly after the first one due to constant over-capacity demand. It shows that the model can reproduce stop-and-go waves at sags, which is consistent with the literature (Goni Ros *et al.* 2014a).

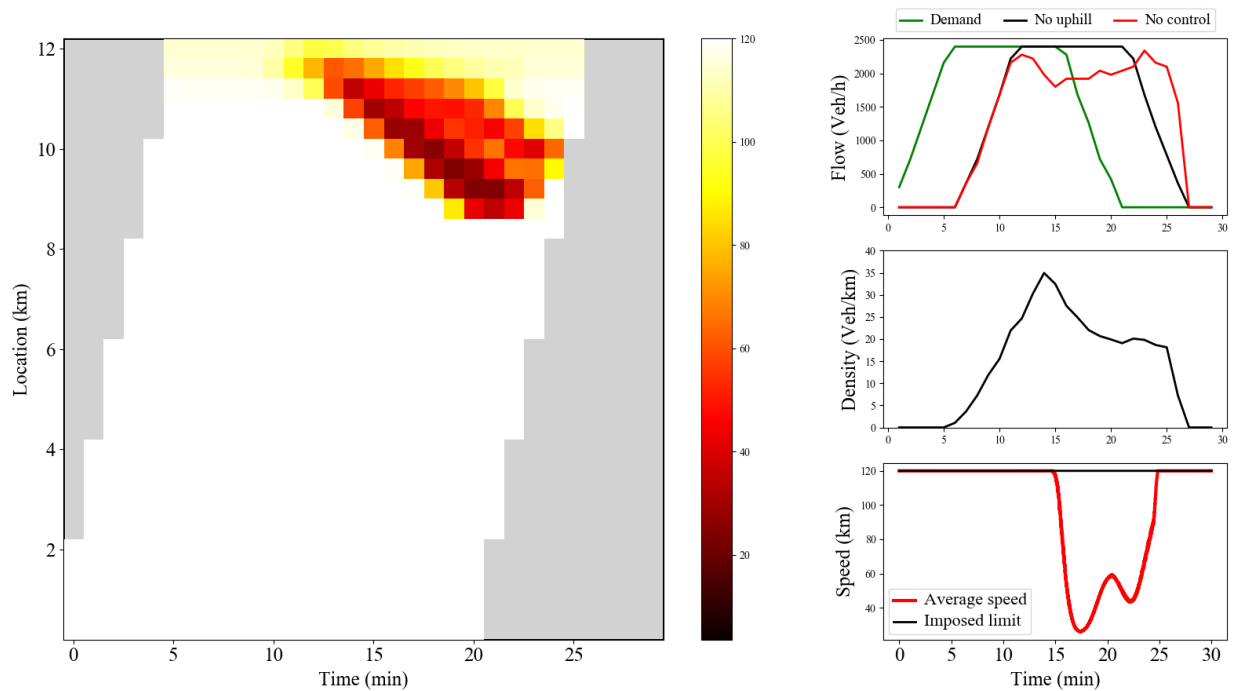


Figure 12 Sag curve with no control: Input demand and exit flows (top right), the density measured by loop detector (middle right), heat-map (left), average vehicle speeds in the control section (bottom right) without a VSL control system

Three other charts are included to provide additional performance measures for analysis. At the top right, the input demand over time (green line), as well as exit flow rates, are depicted

for the base case (red line). The second chart, density versus time plot at the middle right, shows the measured density by the loop detector at the uphill (see *Figure 10*). The chart on the bottom right shows the observed speed (red line) at the control section as well as the imposed VSL (It is assumed that when the control section is empty, the average speed is 120 km/h).

2.4.2) Lane Reduction

For the lane reduction bottleneck, the investigated network contains a three-lane freeway where one lane will be dropped towards the end of the network. The length of the network is 6 km. The road has a constant-gradient of zero. The first lane drops at $x=5.7$ km. The control section starts at $x=1.7$ km and ends at $x=2.7$ km. The control section only controls the speed-limit of lane two and lane three. The speed limit of the dropped lane will not be affected. Two loop detectors are placed to read the density on the road. The first one is located at the bottleneck ($x=5.7$ km), and the second one at the entrance of the network ($x=0.05$ km). There are three roadside units (RSU) to communicate the information with the RL algorithm and CVs. The first RSU (A) is located at the bottleneck to inform the RL algorithm with the density information collected by the loop detector. The second RSU (B) is located at the entrance of the control section, and the third RSU (C) is located at the entrance of the network to inform the RL algorithm about the density at the entrance of network. The geometry of this bottleneck is presented in *Figure 13*.

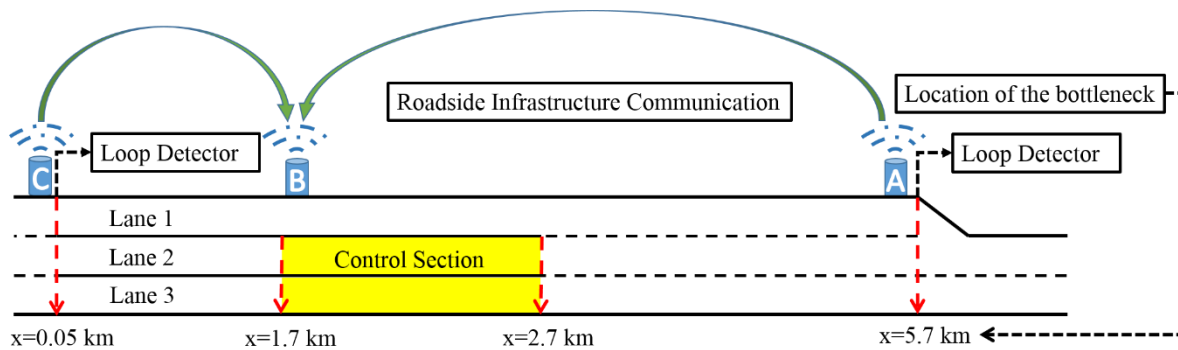


Figure 13 The geometry of the lane reduction bottleneck

The speed limit is 120 km/h. Characteristics of vehicles and drivers, as defined by the IDM model, are assumed to be same as sag curve problem (Table 1). The LMRS lane-changing model was developed to perform lane changing behavior. Characteristics of the LMRS model is assumed to be homogenous. These hypothetical model parameters are shown in Table 2. Since the VSL only control lanes two and three, the desire to move to lane one would increase before entering the control section. Therefore, the lane-changing is forbidden for all lanes starting from the first loop detector (C) until the end of the control section to prevent other types of bottlenecks at the control section.

Table 2 Characteristics of the LMRS model

$T_{min}(s)$	$\tau(s)$	$x_0(m)$	$t_0(s)$	d_{free}	d_{sync}	d_{coop}	$v_{gain}(\frac{km}{h})$
0.56	25	295	43	0.365	0.577	0.78	69.6

The distribution of demand over time is illustrated in *Figure 14*. The first 5 minutes is a transition from zero to 4200 veh/h, a capacity higher than the bottleneck capacity. The demand stays at 4200 veh/h for 10 minutes then transitions back to zero across 5 minutes. Beyond this, the demand remains zero until all vehicles have exited the facility.

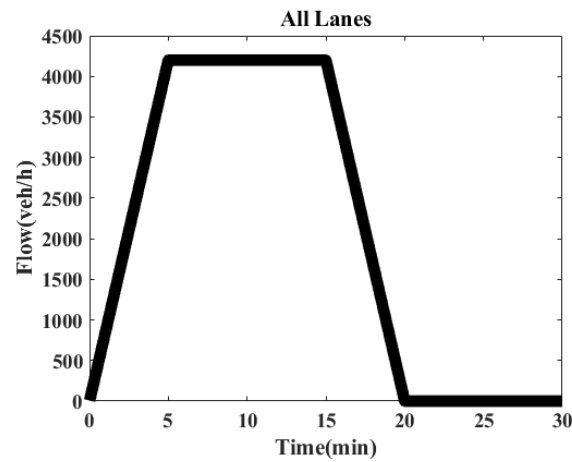


Figure 14 Demand profile over time

With the given parameters above, a microsimulation model was developed in Python. In *Figure 15*, a heat map which shows the average speed of vehicles for different times and locations are shown (left chart) for the base case, i.e., when no control strategy is implemented.

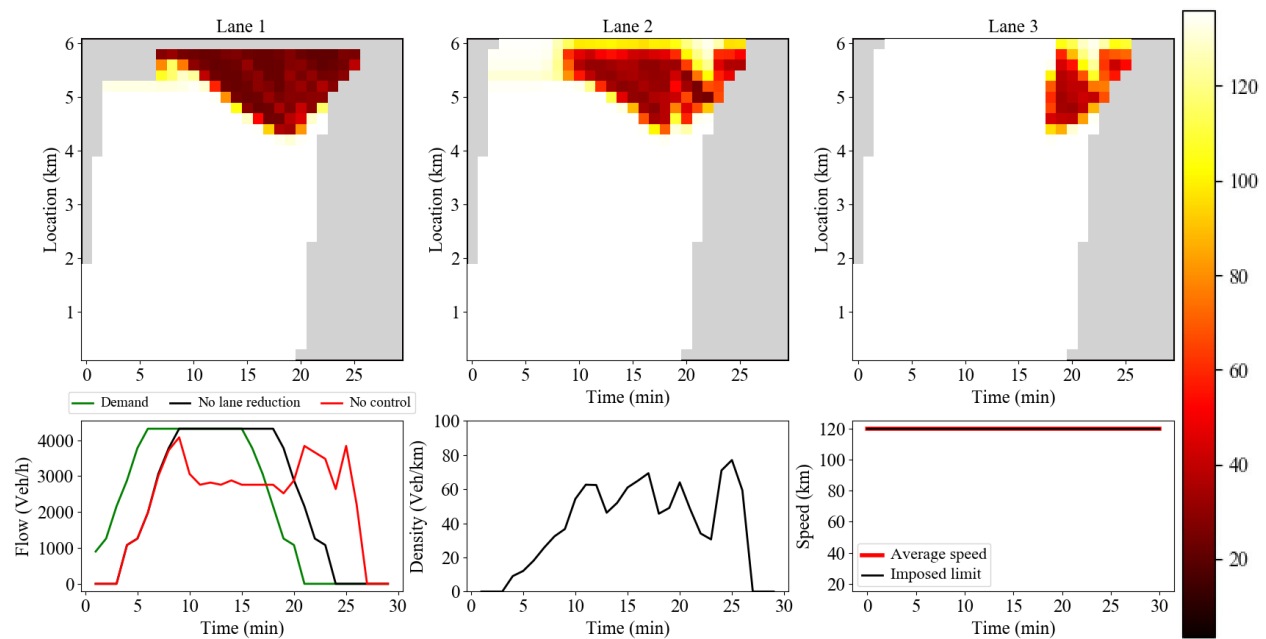


Figure 15 Lane reduction with no control: Input demand and exit flows (second row left), the density measured by loop detector (second row middle), heat-map of speed on time –location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

2.4.3) On-Ramp

For the on-ramp bottleneck, the investigated network contains a two-lane freeway with an on-ramp which has a parallel acceleration lane of 150 meters. The length of the network is 6 km. The road has a constant-gradient of zero, and the parallel acceleration lane ends at $x=5.7$ km. The control section starts at $x=1.7$ km and ends at $x=2.7$ km. The control section only controls speed-limit of main lanes. The speed limit of onramp will not be affected. Two loop detectors have been implemented to read the density in the environment. The first one is located at the bottleneck ($x=5.7$ km), and the second one is located at the entrance of the network ($x=0.05$ km). There are three roadside units (RSU) to communicate the information with the RL algorithm and CVs. The first RSU (A) is located at the bottleneck to inform the RL algorithm with the density information collected by the loop detector. The second RSU (B) is located at the entrance of the control section, and the third RSU (C) is located at the entrance of the network to inform the RL algorithm about the density at the entrance of network.

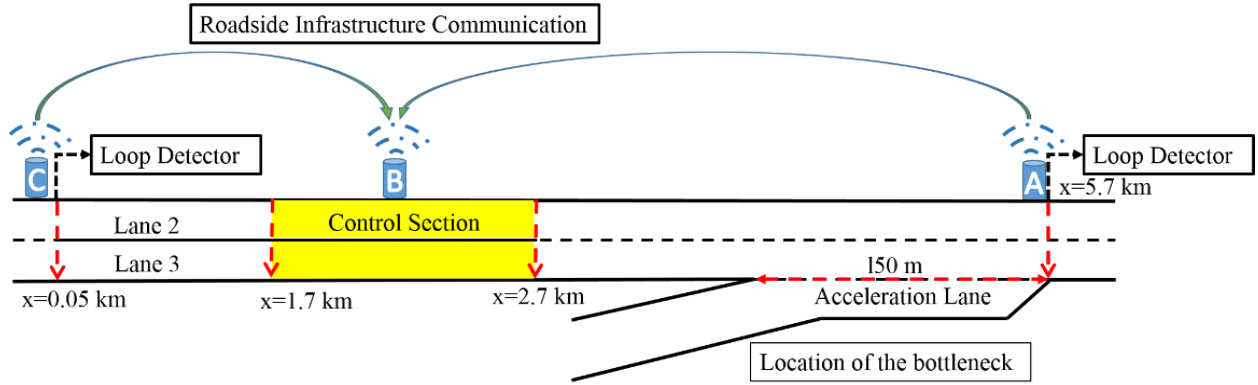


Figure 16 The geometry of on-ramp bottleneck

The speed limit is 120 km/h. Characteristics of vehicles and drivers, as defined by the IDM model, are assumed to be same as sag curve problem (Table 1). Characteristics of the LMRS model in this bottleneck are assumed to be the same as the lane reduction bottleneck (Table 2). The lane-

changing behavior is forbidden from the first loop detector until the end of the control section to prevent other types of bottlenecks at the control section. The *Figure 16*, illustrates the geometry of on-ramp bottleneck.

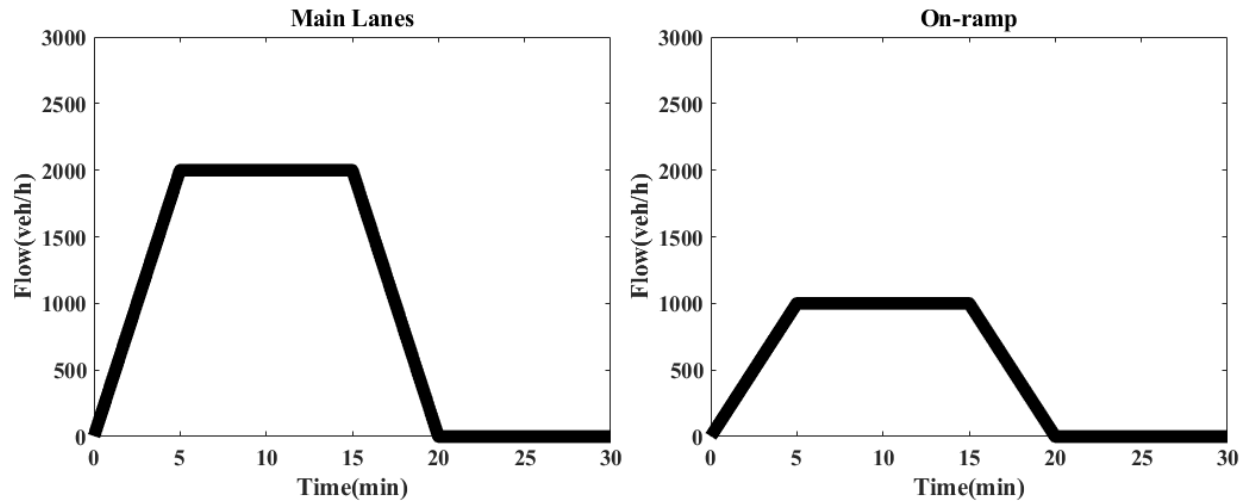


Figure 17 Demand profile over time

The demand profile over time is illustrated in Figure 17. The first 5 minutes is a transition from zero to 2000 veh/h, for the main road and to 1000 veh/h for the ramp. The demand stays overcapacity for 10 minutes, then transitions back to zero across 5 minutes. Beyond this, the demand remains zero until all vehicles have exited the facility. A microsimulation model was created in Python. In Figure 18, a heat map which shows the average speed of vehicles for different times and locations are shown (left chart) for the base case, i.e.

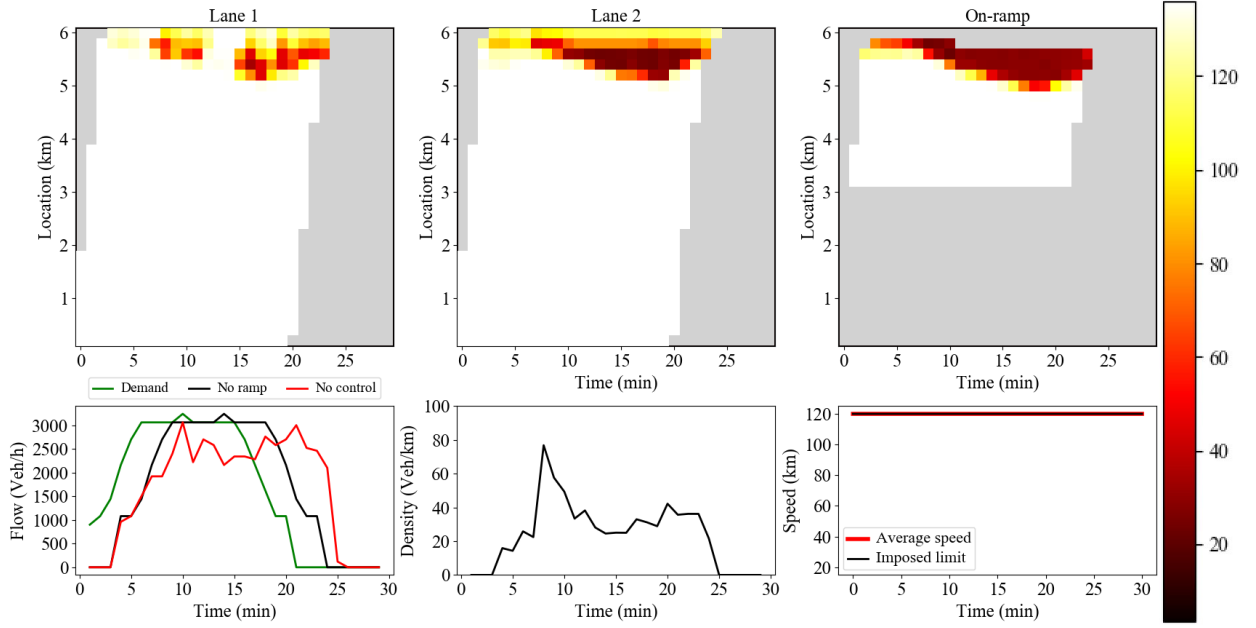


Figure 18 Onramp with no control: Input demand and exit flows (second row left), the density measured by loop detector (second row middle), heat-map of speed on time –location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

2.5) RL Architecture: Action, State Space, and Reward

This research, has been conducted on an Intel(R) Xeon(R) CPU with eight core. Each core has two threads which can be utilized for multi-threading processing. Therefore, an A3C algorithm with 16 agents was trained to optimally control a variable speed limit with five possible actions (possible actions= $[-10, -5, 0, 5, 10]$). In every minute t , the agent must take one of the five possible actions which will lead to an increase, decrease, or no change to the speed limit at the control section at time $t-1$. To represent the state space, two key measurements from the sensors are considered: The density at the bottleneck ($\rho_b(t)$), and the density at the network entrance ($\rho_e(t)$). To capture the temporal variations and features of states, ten consecutive states are stacked together as shown in Equation 34 and then fed to the neural networks.

$$States = \{[\rho_b(t-9), \rho_e(t-9)], \dots, [\rho_b(t), \rho_e(t)]\} \quad (34)$$

The actor and critic are both neural networks with the input shape of 10×2 and four fully connected hidden layers with 64, 128, 128, and 128 units as it is shown in *Figure 19*. All units in hidden layers have rectified linear unit (RELU) as their activation function. The output layer of the Actor has five units which is equal to number of actions with a Softmax as the activation function. The output layer of the Critic has only one unit which calculates the value of being in a state with a linear activation function.

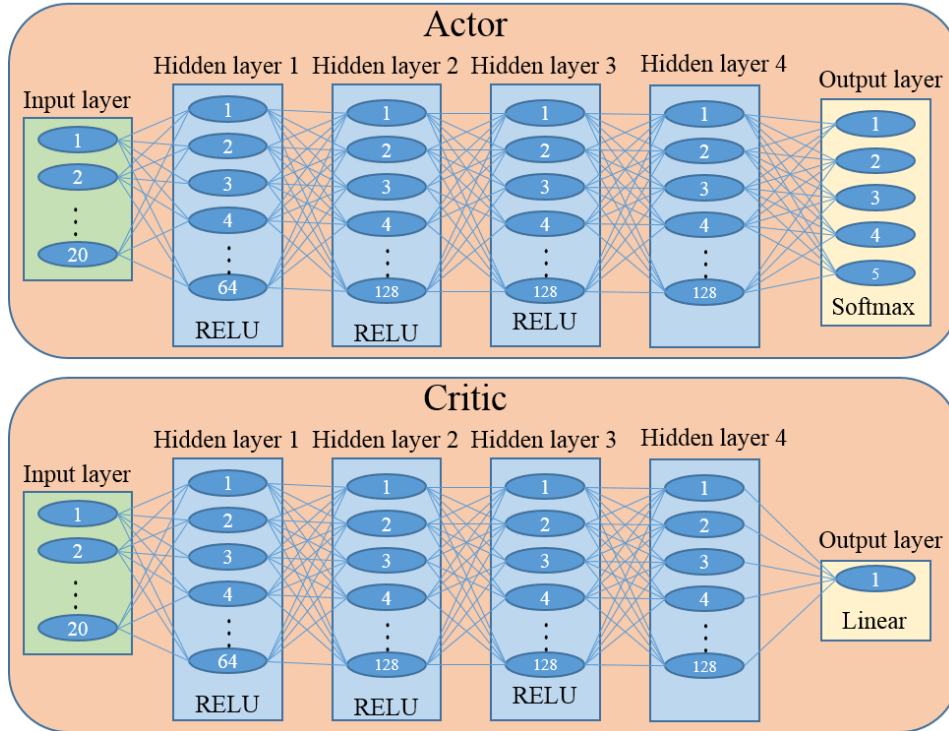


Figure 19 Architecture of Actor and Critic network

The main objective of the VSL is to decrease the total travel time of all vehicles in the system. The delay for a vehicle is computed relative to the hypothetical scenarios mentioned above, where the bottleneck is assumed not to influence the traffic flow. For the simulated demand,

vehicles travel through the corridor at free-flow speeds. For a given scenario, the total delay is the sum of the individual vehicle delays (Equation 35). For example, when there is no VSL control system, the system performs, as shown in *Figure 12*, *Figure 15* and *Figure 18*. This total delay (TD) is then taken as a reference and compared to the total delay under the VSL control strategies. Since the main objective is to decrease total delay, it is not possible to calculate total delay unless the episode ends. Therefore, the reward of all non-terminal states is considered to be zero. Consequently, the reward for the terminal state is defined as shown in Equation 36. The terminal reward is discounted and distributed for all action state pairs taken in an episode with discount factor (γ) equals to 0.99.

$$TD = TTT_{Bottleneck} - TTT_{NoBottleneck} \quad (35)$$

$$Reward(t) = \begin{cases} 0 & t \neq T \\ \frac{TD_{NoControl} - TD_{Controlled}}{TD_{NoControl}} & t = T \end{cases} \quad (36)$$

CHAPTER 3: RESULTS

The resulting system performance for different bottlenecks is explained in this chapter. Moreover, an extensive sensitivity analysis has been investigated for each bottleneck to understand the impact of different level of connectivity on the performance of models.

3.1) Optimum Solutions of the feedback control algorithm

To compare the RL model with an existing model in the literature, a feedback control was developed in Python based on Equation 33. Optimal parameters of this controller were found through a grid search. *Figure 20* shows the optimal solutions of the feedback controller for the sag curve bottleneck. The feedback controller was able to eliminate 53% of the delay caused by sag curve when the penetration rate of CVs is 100%.

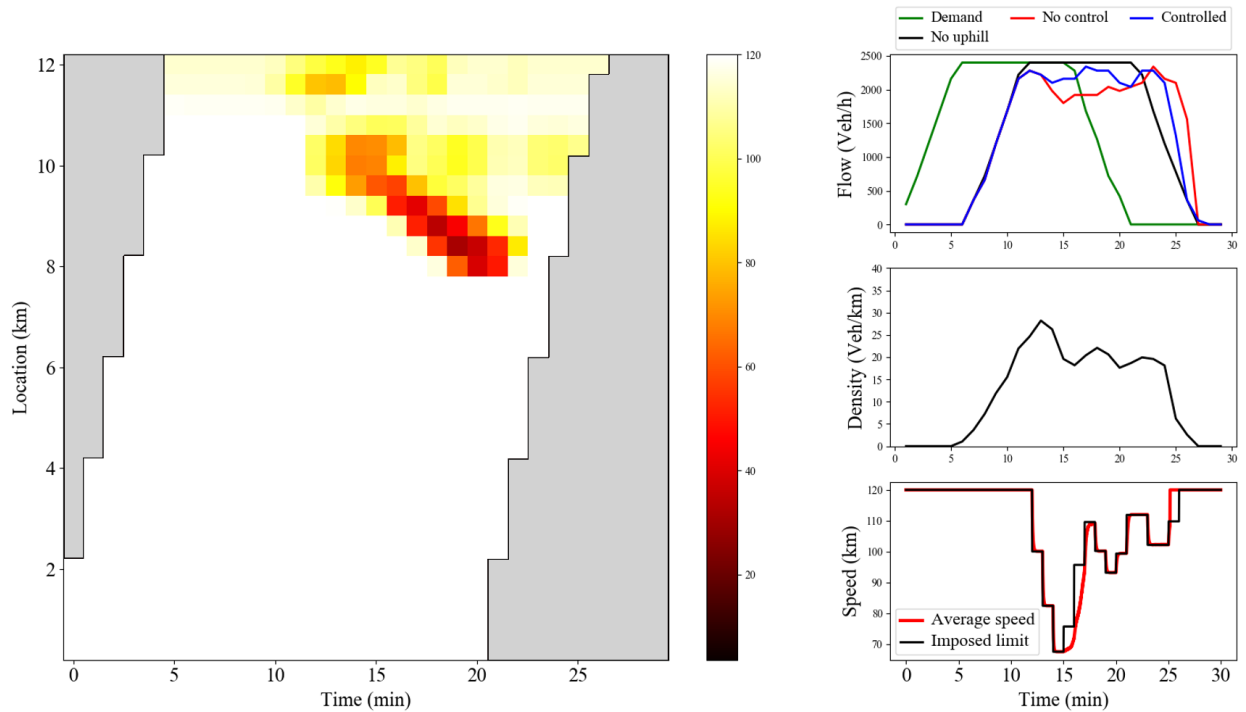


Figure 20 Sag curve with feedback control: Input demand and exit flows (top right), the density measured by loop detector (middle right), heat-map of speeds (left), average vehicle speeds in the control section (bottom right)

Figure 21 shows the optimal solutions of the feedback controller for the lane reduction bottleneck.

The feedback controller was able to eliminate 21% of the delay caused by on-ramp when market penetration rate of CVs was 100%. The VSL has reached to speed as low as 20 km/h.

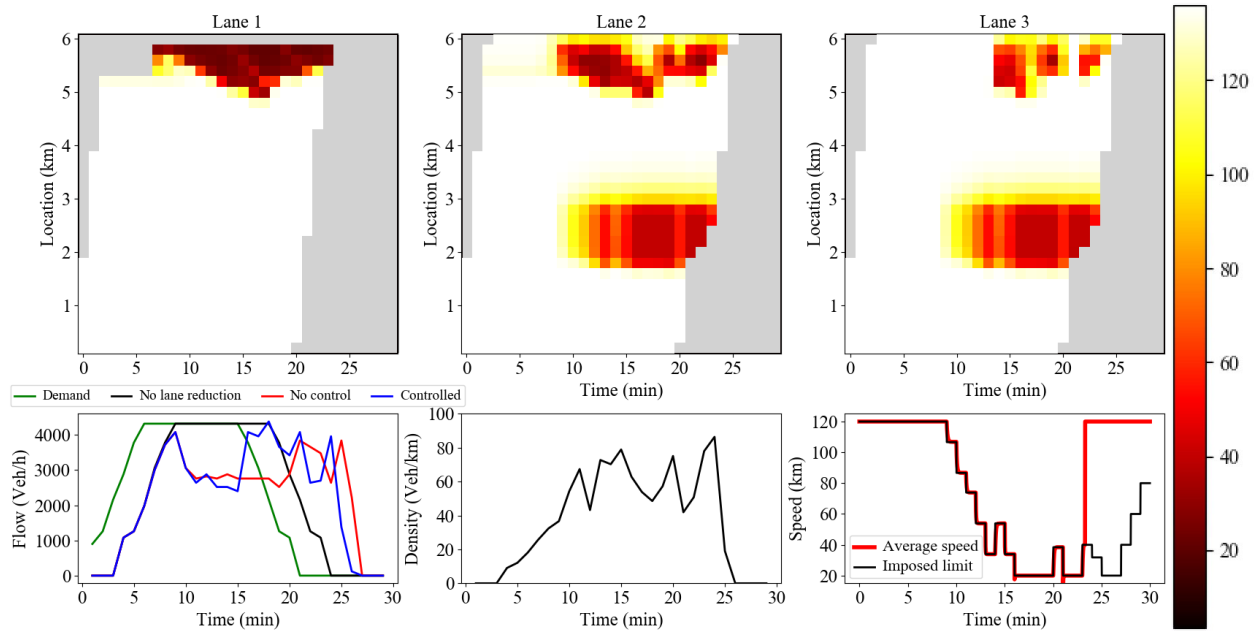


Figure 21 Lane reduction with feedback control: Input demand and exit flows (the second row left), the density measured by loop detector (second-row middle), heat-map of speed on the time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

The Figure 22, shows the optimal solutions of the feedback controller for the on-ramp bottleneck.

The feedback controller was able to eliminate 13% of the delay caused by on-ramp. When the market penetration rate of CVs was 100%. The VSL has reached to speed as low as 20 km/h.

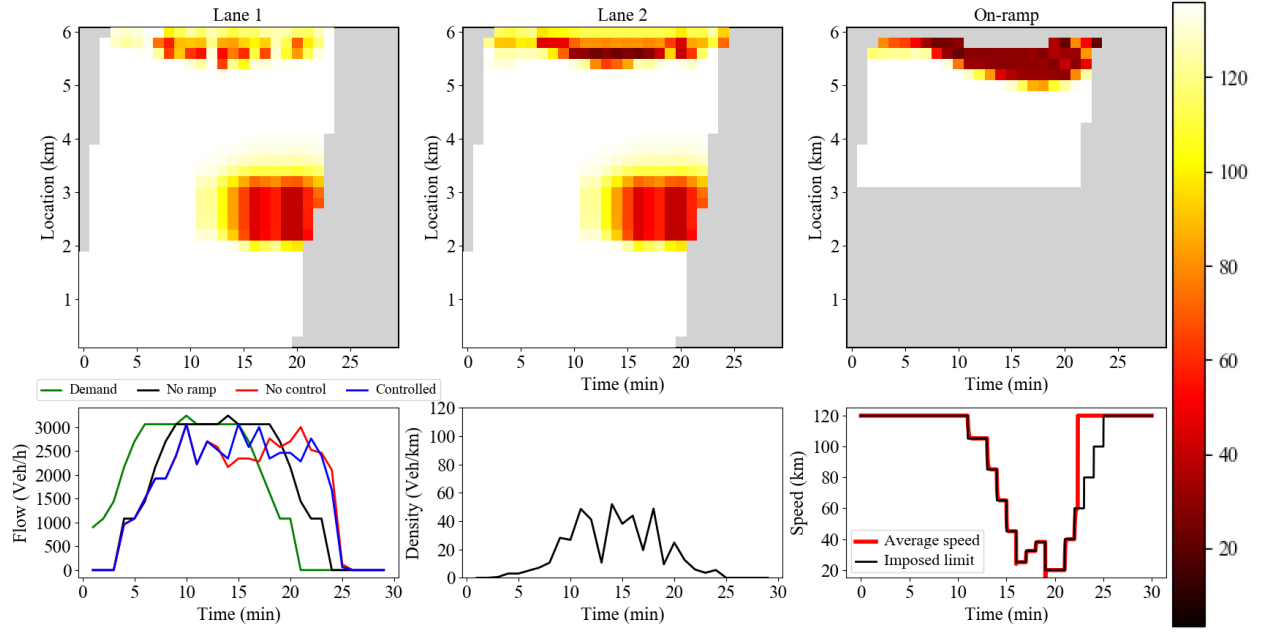


Figure 22 Onramp with feedback control: Input demand and exit flows (second row left), the density measured by loop detector (second row middle), heat-map of speed on time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

3.2) Optimum Solutions of RL algorithm

For each bottleneck, the RL algorithm has been trained. An episode is a simulation with the given demand profile. It ends when all vehicles exit the network. An agent will go through the simulation and take actions based on its Actor and Critic (the given ANNs from global network). When an episode ends the worker give its collected experiences to the global network. The global network will update the Actor and Critic's networks and and give back the new parameters to the worker. This process is done asynchronously in parallel with 16 agents. *Figure 23* shows 50-points moving average for total delay improvement of the RL algorithm for the sag curve bottleneck. The algorithm convergence to the optimal solution after around ten thousand episodes. Since the action space of the RL algorithm is discrete, the Q-values have discontinuity, and the divergence from the optimal solution can be seen between 12000 and 14000 episodes.

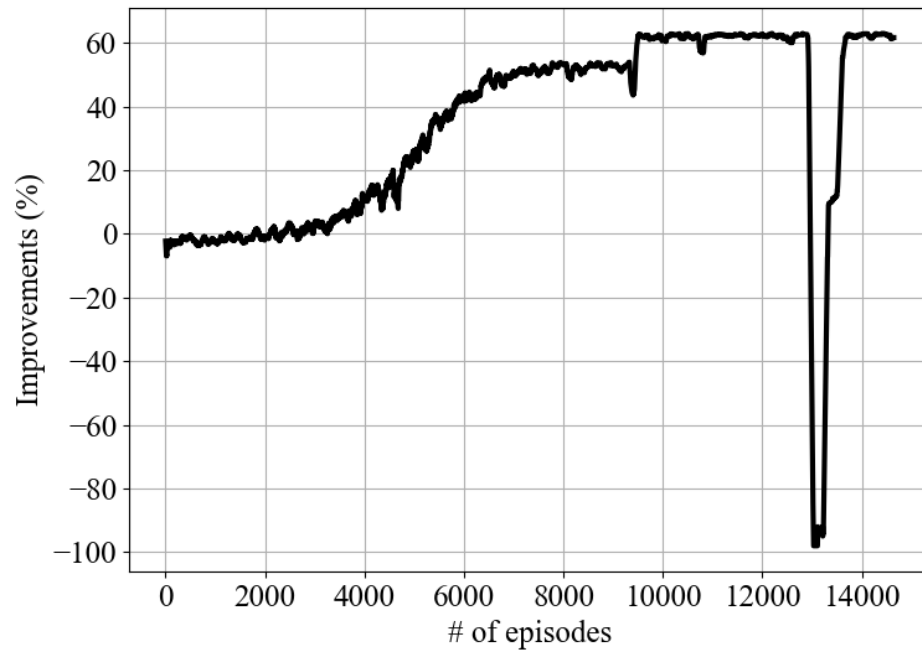


Figure 23 Convergence of the RL algorithm for sag curve bottleneck

The performance of the controller is shown in *Figure 24* for the sag curve bottleneck. The market penetration rate (MPR) is 100%. The VSL reduced the total delay by 63% compared to the no control scenario. The shockwave is moved upstream of the control section. The density at the bottleneck stays below 20 veh/km at the onset of congestion which is the critical density. In the bottom-right chart, it is clear that vehicles adhere to the imposed VSLs since their average speeds follow the VSLs.

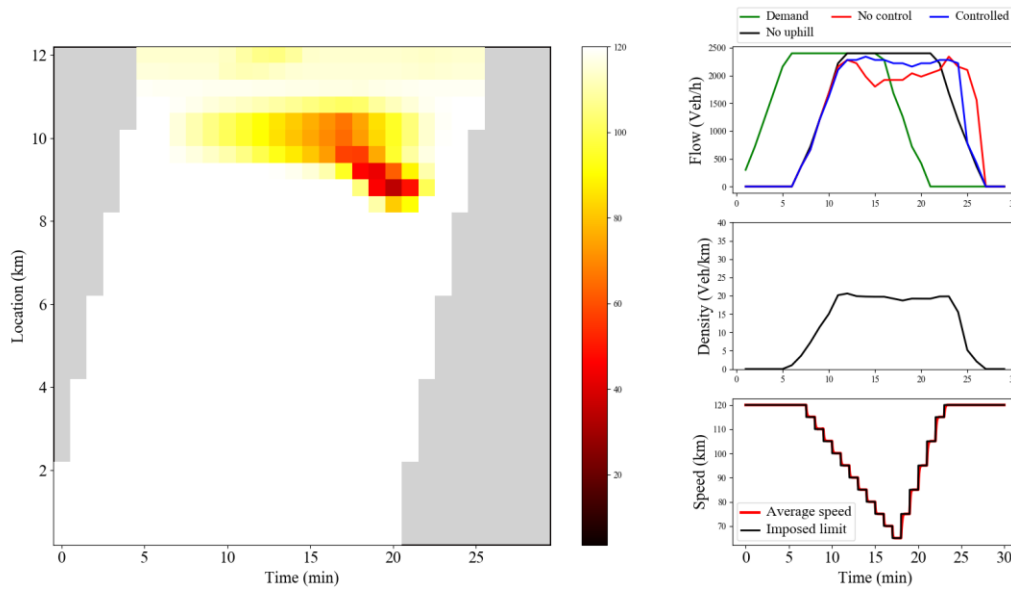


Figure 24 Sag curve with RL control: Input demand and exit flows (top right), the density measured by loop detector (middle right), heat-map with sample trajectories (left), average vehicle speeds in the control section (bottom right) without a VSL control system

The *Figure 25*, illustrates the performance of the RL algorithm for lane reduction bottleneck. The MPR is 100%. Even though the algorithm was not able to clear the congestion at the bottleneck, it was able to reduce the impact of the capacity drop. The VSL reduced the total delay by 64% compared to the no control scenario.

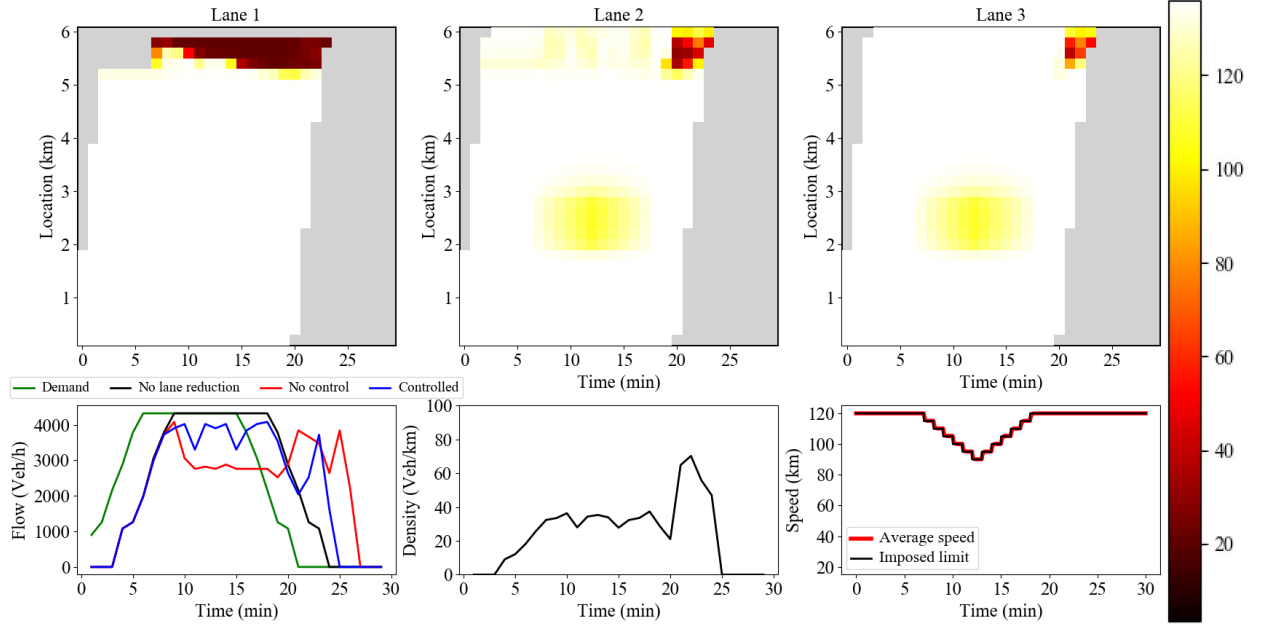


Figure 25 Lane reduction with RL control: Input demand and exit flows (second-row left), the density measured by loop detector (second-row middle), heat-map of speed on the time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

Figure 26 illustrates the performance of the RL algorithm for the on-ramp bottleneck. The MPR is 100%. Same as the lane reduction problem, the algorithm was not able to clear the congestion at the bottleneck, but it was able to reduce the impact of the capacity drop. The VSL reduced the total delay by 19% compared to the no control scenario. This reduction is 13% with the feedback control algorithm as discussed before. Performance of RL controller for on-ramp bottleneck is significantly lower than the other two bottlenecks. The reason should be due to the limited availability of lane changing for on-ramp. Vehicles only have the acceleration lane to perform the lane changing and enter the main road. Therefore, it's harder to dissipate the congestion. It is notable that all RL solutions outperform the feedback controller. They also make a smooth patterns compared to feedback control (Figure 20, Figure 21, and Figure 22).

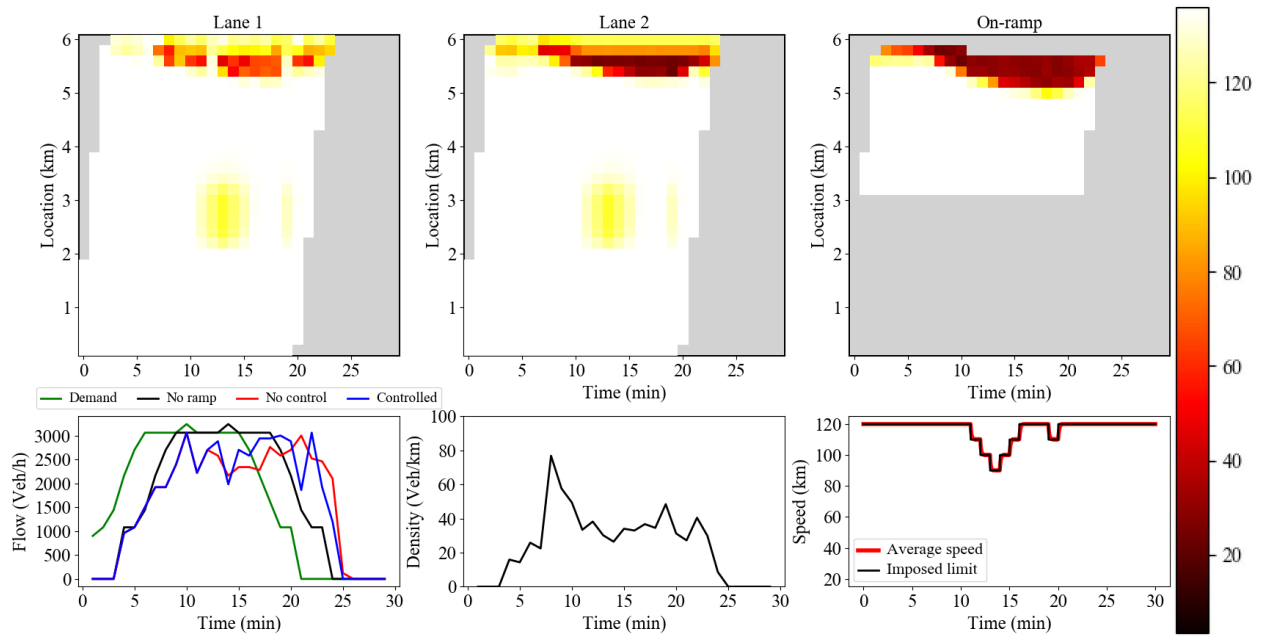


Figure 26 Onramp with RL control: Input demand and exit flows (second-row left), the density measured by loop detector (second-row middle), heat-map of speed on the time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

3.3) Performance of the RL under new demand patterns

As it is shown in the previous section, the RL algorithm was able to solve all three bottlenecks through VSL. It outperforms the feedback control and produces a much smoother pattern for VSL. It is also important to see how the RL algorithm trained for these short demands would perform on new demand patterns. Therefore, a scenario where a short period peak demand followed by longer peak demand was investigated for all three bottlenecks to see how the RL algorithm would react to the fluctuation in demand. The *Figure 27*, illustrates the performance of the RL algorithm on sag curve bottleneck for the longer demand. The RL was able to eliminate 48% of delay caused by the sag curve.

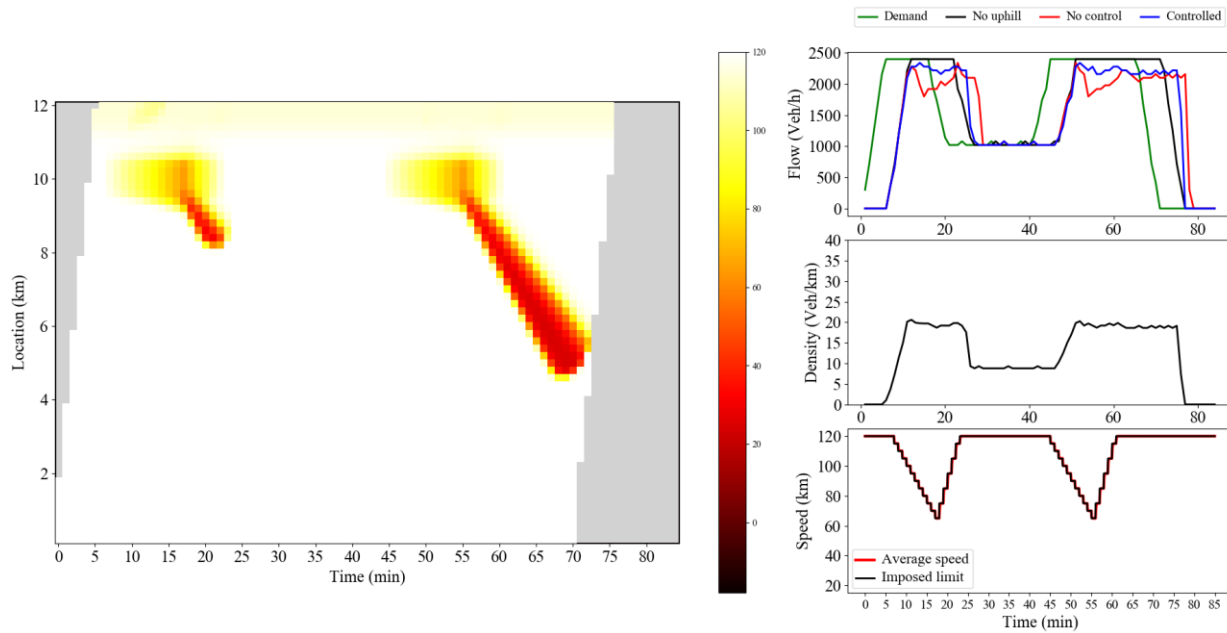


Figure 27 Sag curve with RL control: Input demand and exit flows (top right), the density measured by loop detector (middle right), heat-map with sample trajectories (left), average vehicle speeds in the control section (bottom right) without a VSL control system

The *Figure 28*, shows the result of the feedback control on the same demand pattern. The feedback controller is also able to respond effectively to fluctuation in demand, but it was able to only eliminate 33% of the delay caused by the sag curve. It should be noted all three algorithms were trained only on short demand with one peak. These results show that the RL algorithm is robust to the length of demand.

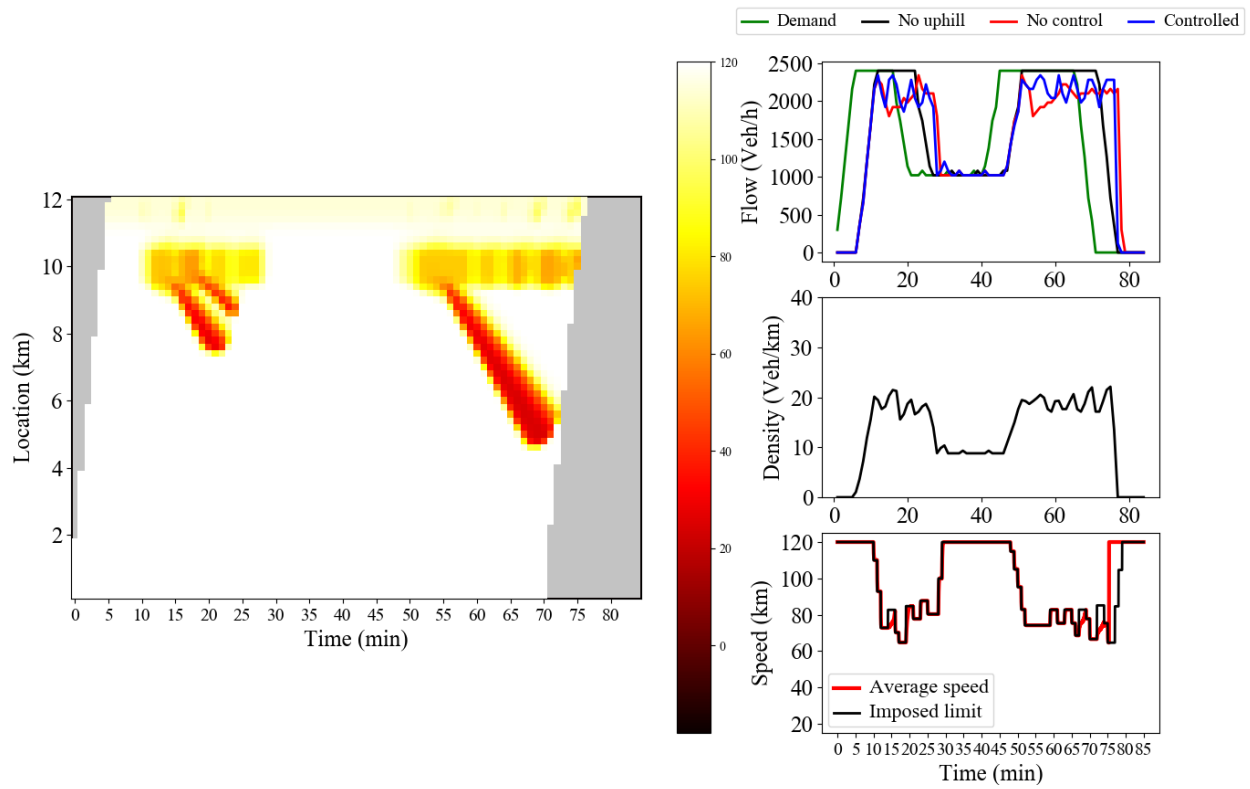


Figure 28 Sag curve with feedback control: Input demand and exit flows (top right), the density measured by loop detector (middle right), heat-map with sample trajectories (left), average vehicle speeds in the control section (bottom right) without a VSL control system

The *Figure 29*, illustrates the performance of the RL algorithm on lane reduction bottleneck for the new demand. The RL was able to eliminate 36% of delay caused by the lane reduction.

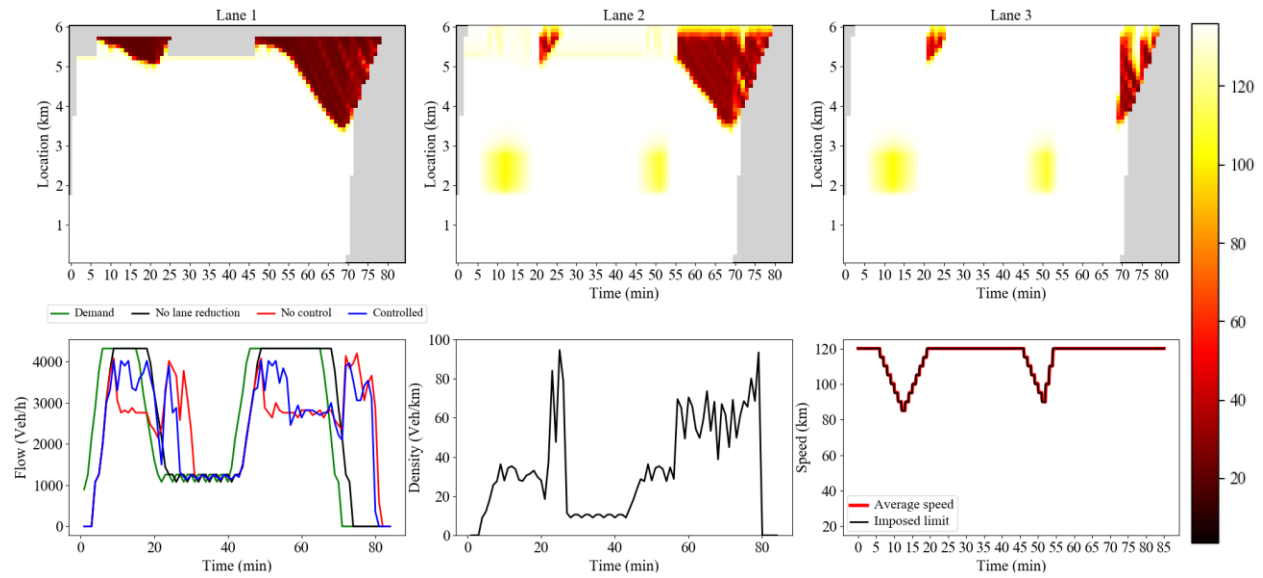


Figure 29 Lane reduction with RL control: Input demand and exit flows (second row left), the density measured by loop detector (second row middle), heat-map of speed on time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

The *Figure 30*, shows result of the feedback control on the same demand pattern. The feedback controller is also able to respond effectively to fluctuation in demand, but it was able to only eliminate 10% of the delay caused by the sag curve.

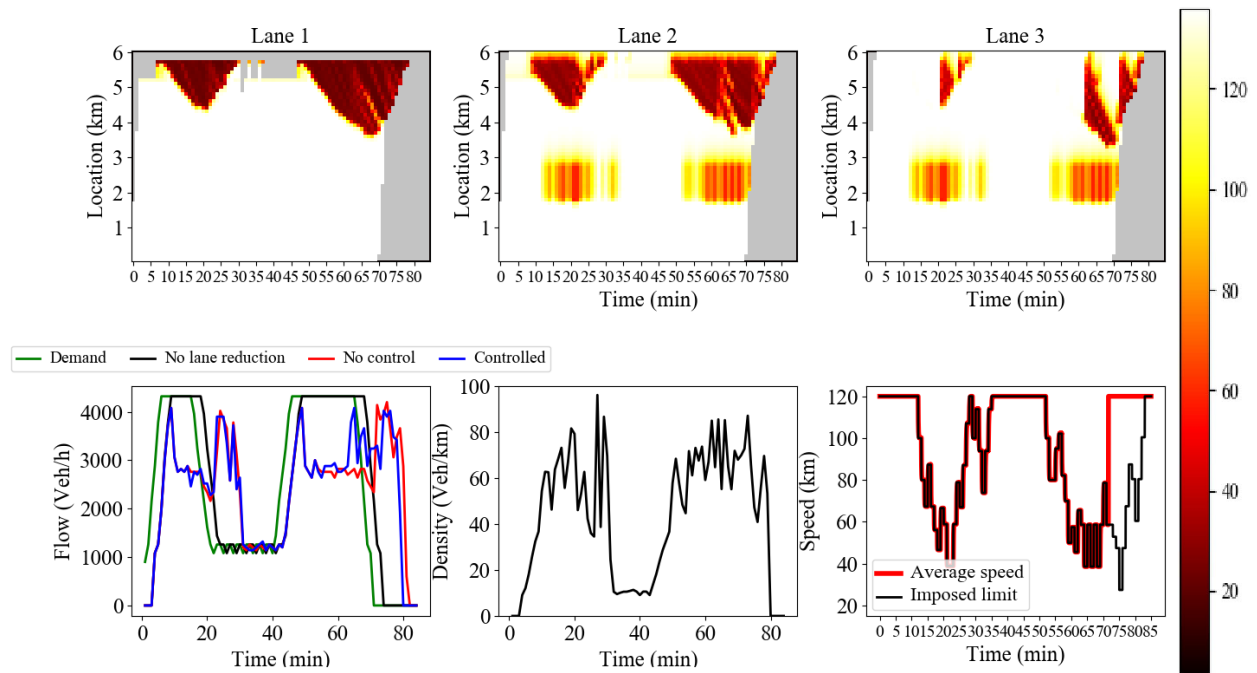


Figure 30 Lane reduction with feedback control: Input demand and exit flows (the second row left), the density measured by loop detector (second-row middle), heat-map of speed on the time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

The *Figure 31*, illustrates the performance of the RL algorithm on lane reduction bottleneck for the longer demand. The RL was able to eliminate 11% of delay caused by the lane reduction.

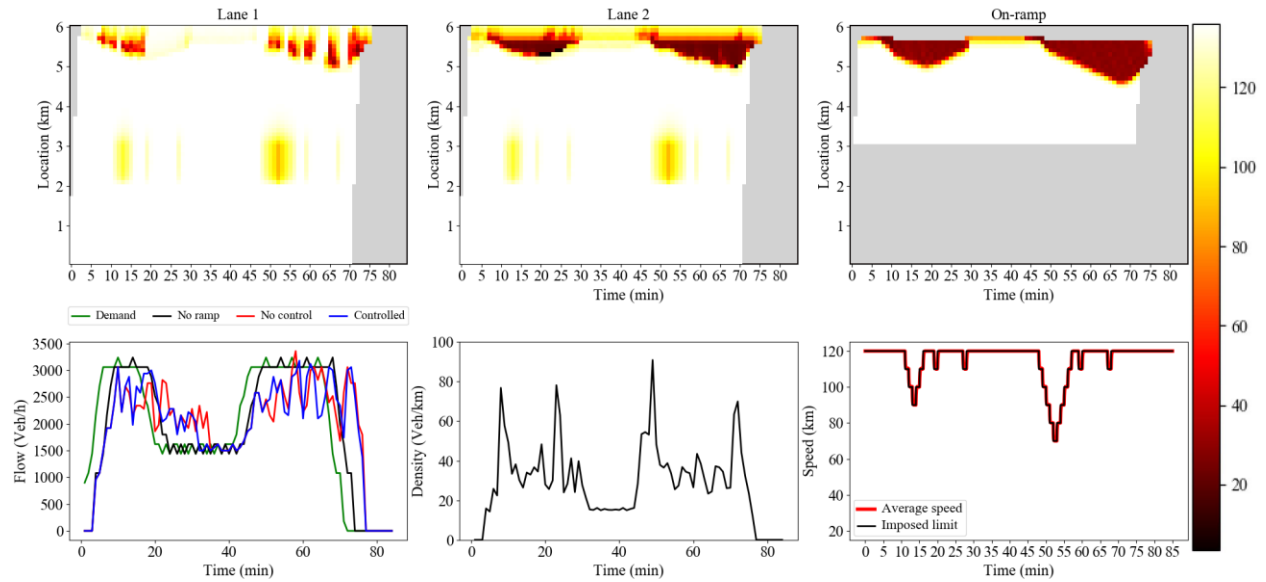


Figure 31 Onramp with RL control: Input demand and exit flows (second-row left), the density measured by loop detector (second-row middle), heat-map of speed on the time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

The *Figure 32*, shows the result of the feedback control on the same demand pattern. The feedback controller is also able to respond effectively to fluctuation in demand, but it was able to only eliminate 5% of the delay caused by the sag curve.

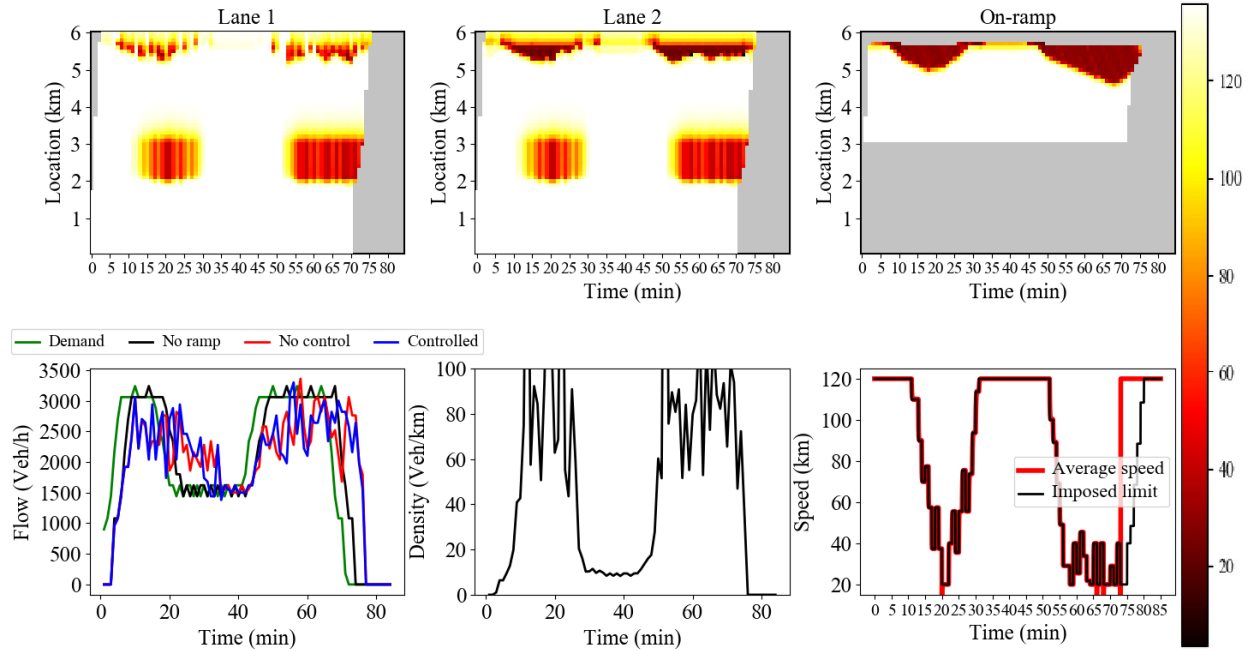


Figure 32 Onramp with feedback control: Input demand and exit flows (second-row left), the density measured by loop detector (second-row middle), heat-map of speed on the time-location diagram (first row), and average vehicle speeds in the control section (second row right) without a VSL control

3.4) Sensitivity to the Market Penetration Rate of Connected Vehicles

In this section, the MPR level of the CVs is varied to understand the impacts of different MPR on system performance. As arrivals of CVs are assumed to be random, each generated vehicle is predicted to be a CV based on the set MPR. Consequently, the total number of CVs in the system (n) has a Binomial distribution. Each market penetration level is simulated 50 times to account for the variability in n as well as their arrival times. Median performance of the three scenarios is illustrated in Figure 33. Since bottlenecks are different, the magnitude of improvements for these three models are not comparable, but the trend shows reducing MPR will result in lower improvement. For the lane reduction and the on-ramp bottlenecks, as the percentage of CVs increases, the median improvement also increases at a constant rate. But the sag curve problem has a sudden jump when the percentage of CVs goes beyond 30%. It should be noted this models are only trained 100% MPR.

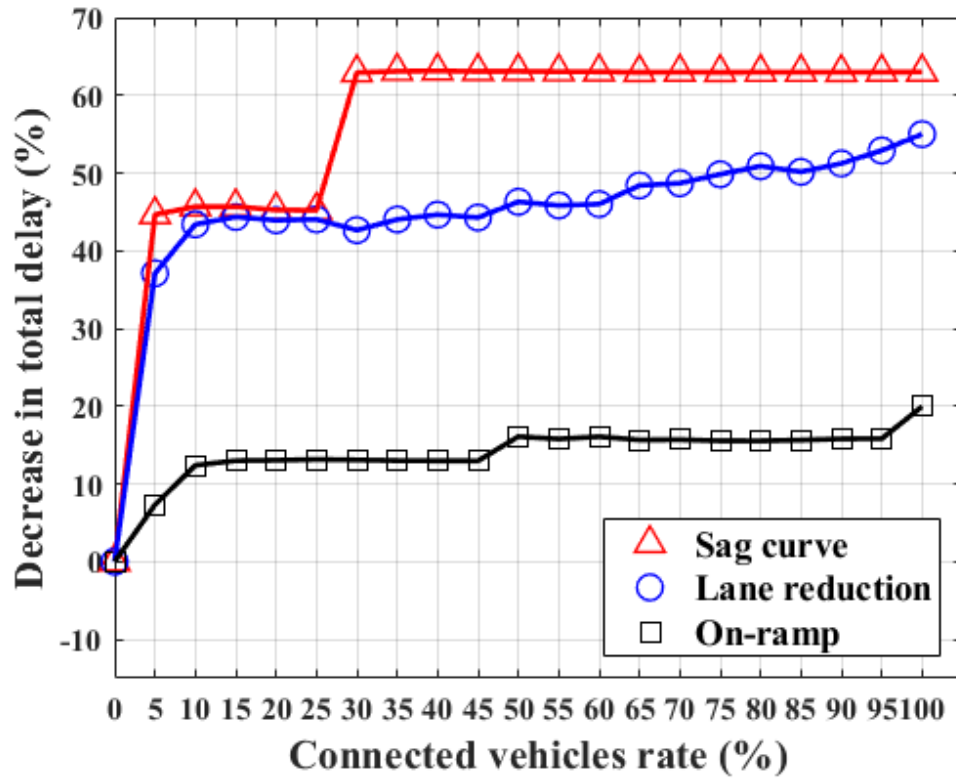


Figure 33 Median performance of different penetration rates for 50 simulation

The results of 50 runs for the sag curve bottleneck are presented as boxplots in Figure 34. When the MPR is less than 35%, the performance has a wide range of variation. It is possible to get a “negative” performance for MPRs less than 15%. The density diagram shown previously in *Figure 12* (middle right) shows how density at the bottleneck would change if there were no control strategy. The most critical period for control strategy is between 5 to 10 minutes from the beginning of the simulation, which is the transition period from the uncongested to the congested traffic. After this period, congestion starts to propagate and grows backward. If the density of connected vehicles over time within this period is not enough to mitigate the initiation of the breakdown, the queue will extend from the bottleneck location (i.e., uphill) to the control sections. It will make a recovery to normal operations almost impossible.

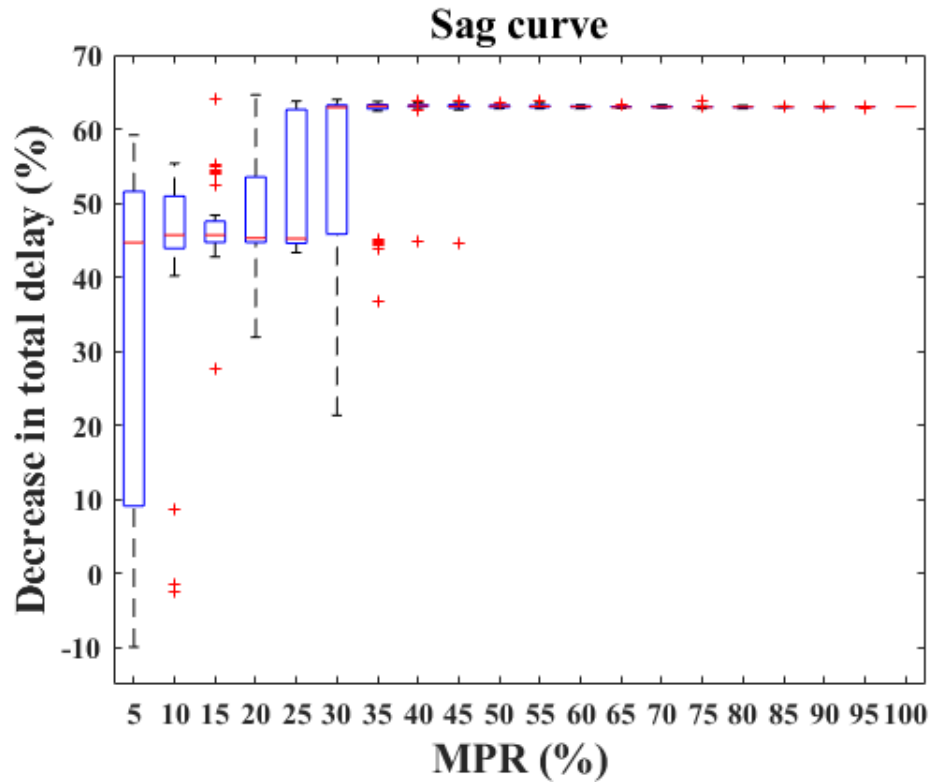


Figure 34 Boxplots of sensitivity analysis for sag curve bottleneck

The results of 50 runs for the lane reduction bottleneck is presented as boxplots in *Figure 35*. Performance of the RL algorithm for this bottleneck has a wide range of variation. The magnitude of variation tends to increase when the percentage of CVs decrease. The model can get a “negative” performance for MPRs less than 15%. It is notable that for the MPR of 5%, the model has widest variation range. It can have negative performance as well as a positive performance. The positive performance can even get higher than 100% connectivity performance. To understand how this is possible, the extreme case of this boxplot where the model was able to have 67% improvement was investigated in details.

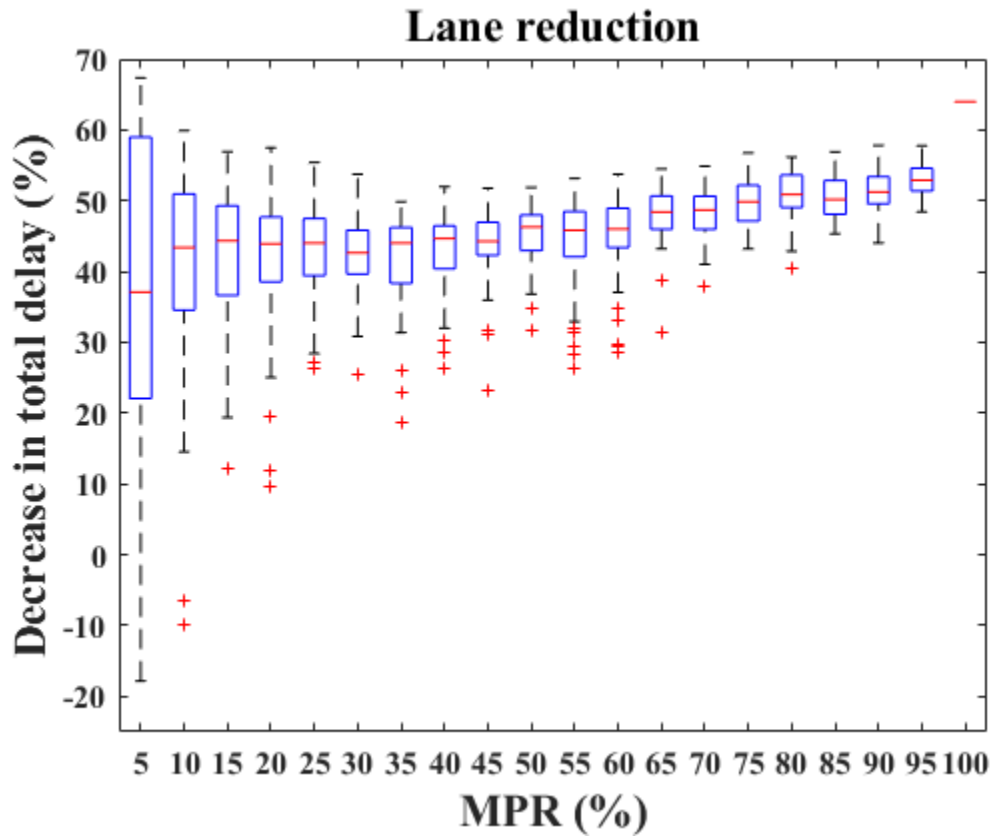


Figure 35 Boxplots of sensitivity analysis for lane reduction bottleneck

As it is shown in *Figure 36*, the RL algorithm performs its normal behavior, but CVs in the control section are randomly distributed in a way that creates a distinct wave-shaped average speed behavior. The speed diagram is zoomed in for better observation (Second row right). Since the model has a discrete action space and boundaries for changing the speed limit, it is not able to reproduce this behavior. Therefore, the RL algorithm with the current action space is not able to reach this level of performance.

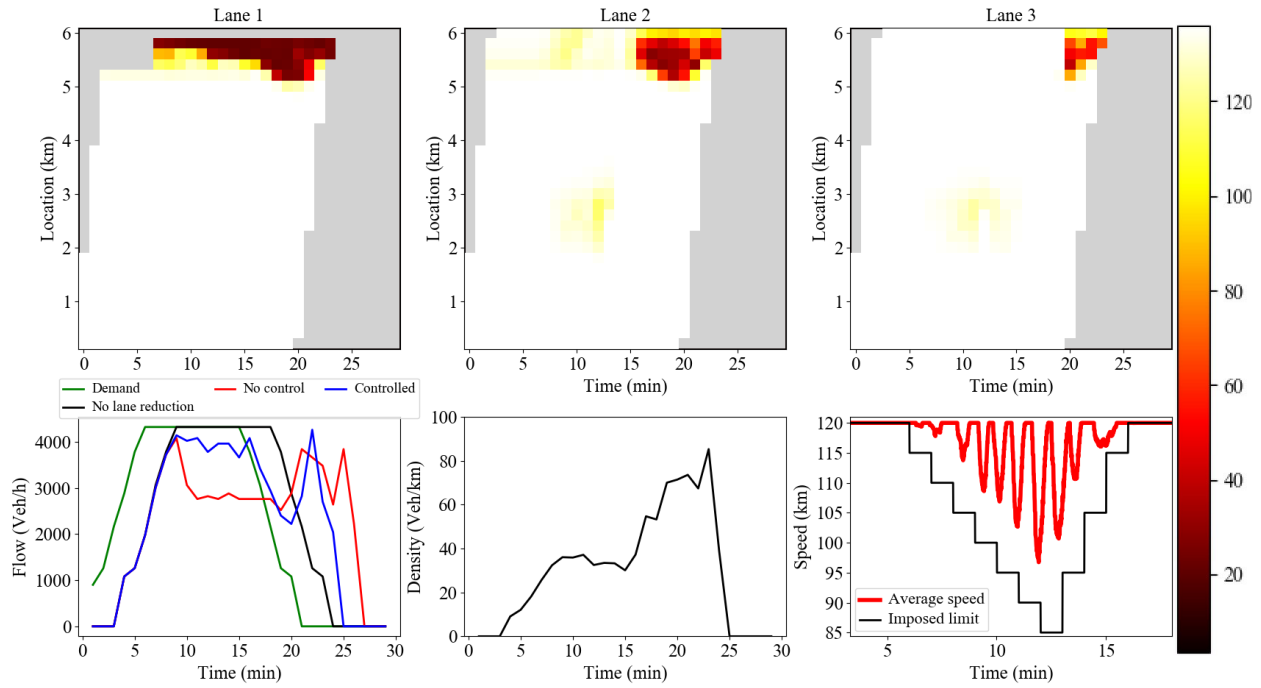


Figure 36 extreme case of lane reduction bottleneck at MPR of 5% with 67% improvement

The results of 50 runs for the on-ramp bottleneck is presented as boxplots in *Figure 37*. Performance of RL algorithm for this bottleneck has a wide range of variation too. The magnitude of variation tends to increase when the percentage of CVs decrease. The model has “negative” performances for MPRs less than 25%. The variation performances for MPRs higher than 50% are in the same range.

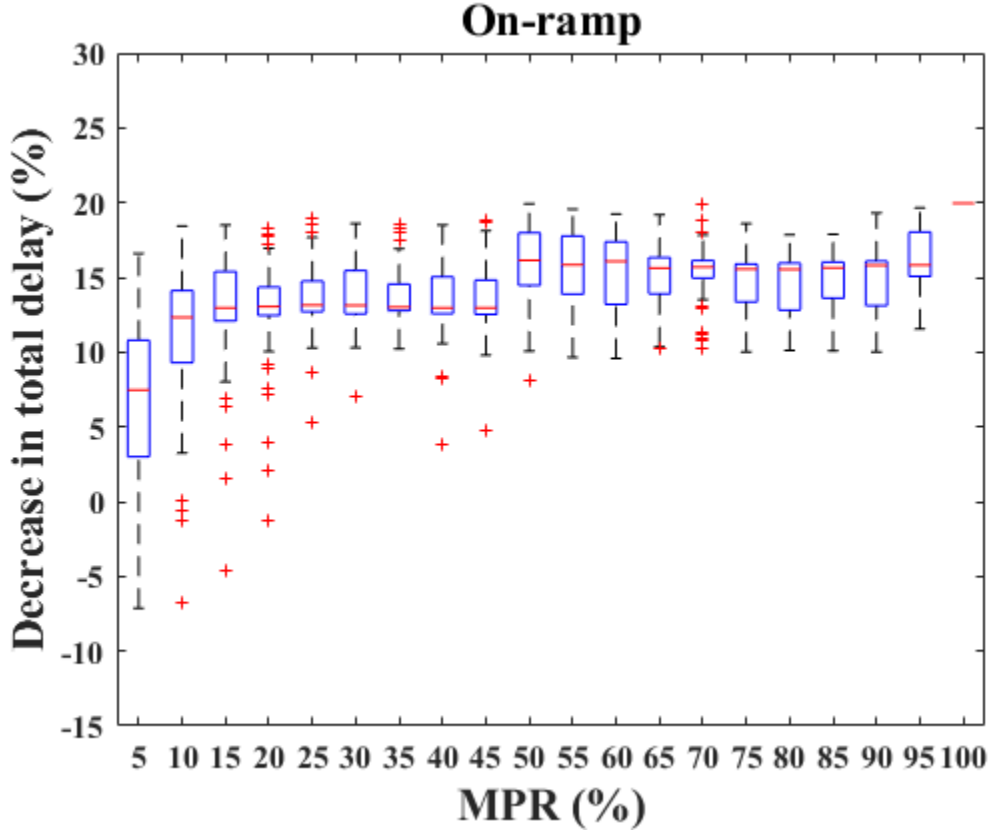


Figure 37 Boxplots of sensitivity analysis for on-ramp bottleneck

3.5) Adaptive VSL for different MPR of CVs

One of the features of the RL algorithm is the ability to add new inputs to the model with a minor modification. In our setup, we have an RSU and a loop detector at the entrance of the network. Since CVs can communicate with the RSU and the loop detector can count the number of vehicles passed, the percentage of CVs can be calculated at each interval of simulation. Percentage of CVs at the previous interval was added to the neural network of the model as an input (Equation 37). Thus, the model can adapt itself and choose different solutions for different MPRs. We use the previous interval to take in to account the time that it takes for each CV to reach the control section. The architecture of the neural network remains the same except the input vector has 10×3 elements.

$$States = \{[\rho_b(t-9), \rho_e(t-9), \%CV(t-10)], \dots, [\rho_b(t), \rho_e(t), \%CV(t-1)]\} \quad (37)$$

The model has been trained on the sag curve bottleneck environment. At the beginning of each episode, a random MPR level would be chosen from a uniform distribution (5% to 100% with an increment of 5). Once chosen, the MPR level will stay constant for the whole simulation.

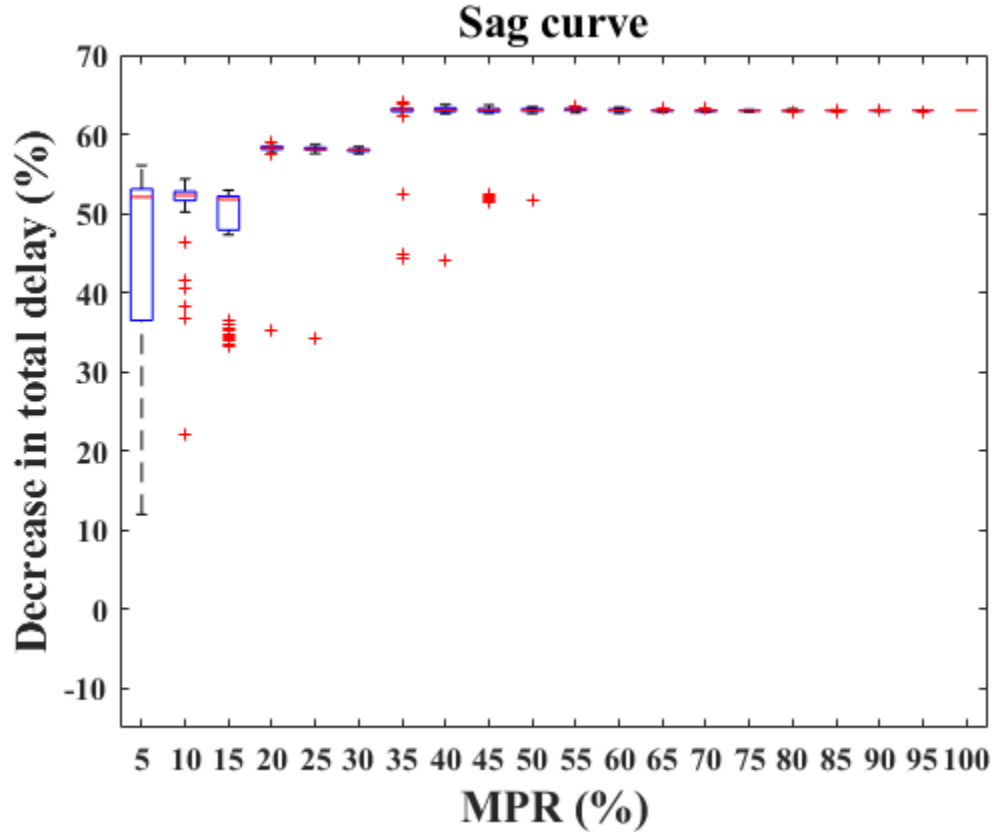


Figure 38 Boxplots of sensitivity analysis for sag curve bottleneck

The results of 50 runs of adaptive RL algorithm for the sag curve bottleneck are presented as boxplots in *Figure 38*. The RL algorithm was able to reduce the variation of performances significantly compared to the nonadaptive version (*Figure 34*). There are no negative results, and the model has significant improvement even for lower MPRs. There can be seen three different levels of performances which correspond to different solutions that the algorithm selects to solve the problem. *Figure 39*, illustrate these three patterns. Based on the MPR levels, the RL controller

observes the system state at the entrance of the network; it chooses which pattern has a higher expected value of an improvement.

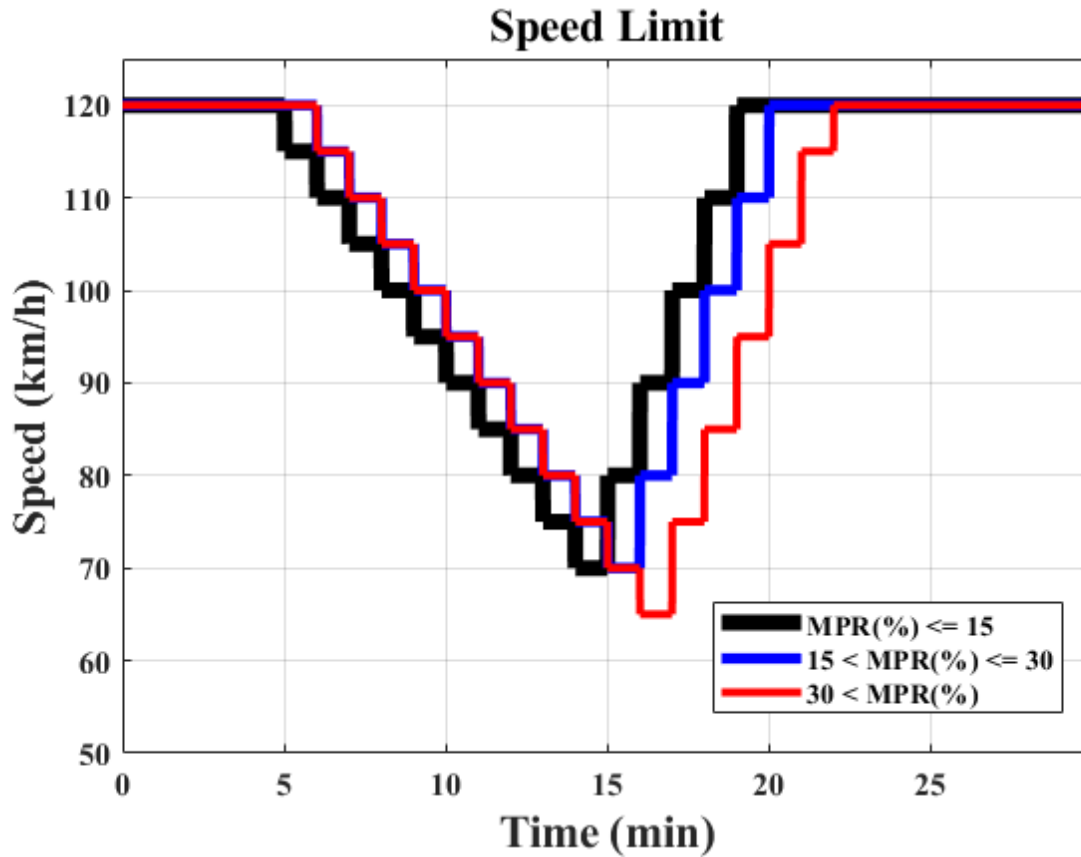


Figure 39 Imposed speed-limit of the adaptive RL controller

The same approach has been investigated for the other two bottlenecks. The results of 50 runs of adaptive RL algorithm for the lane reduction bottleneck are presented as boxplots in *Figure 40*. The adaptive RL algorithm was able to show better performance in comparison to the non-adaptive version (*Figure 35*). The negative performances are eliminated, and overall median performances are shifted up.

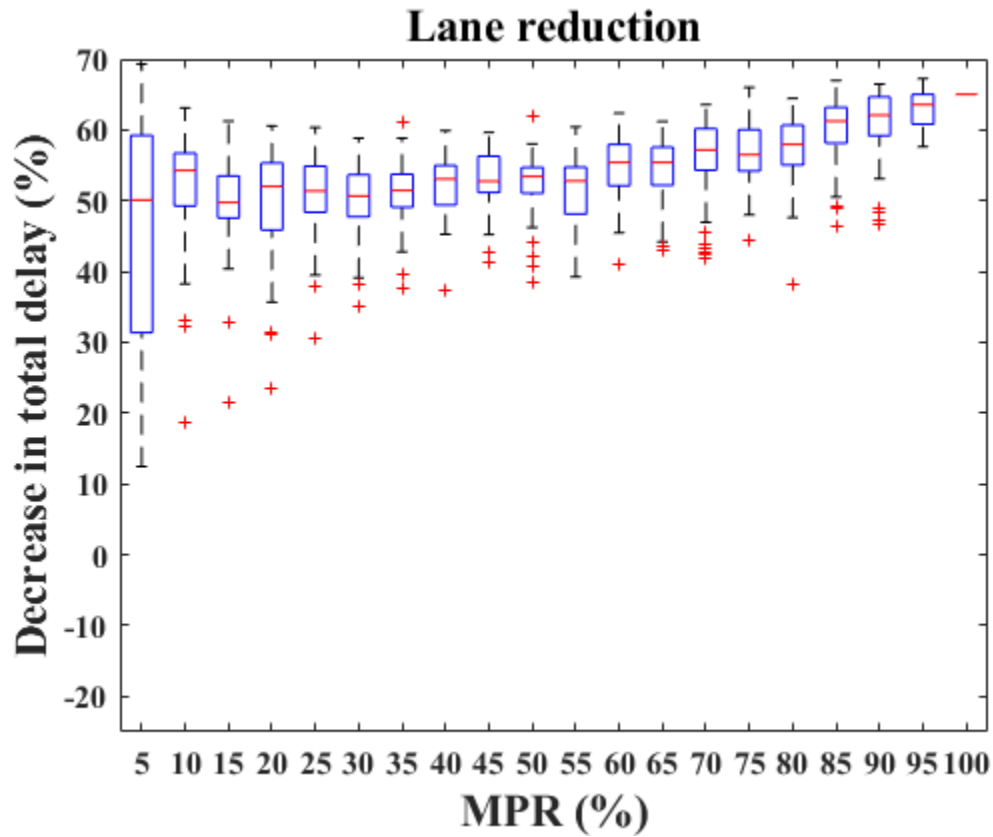


Figure 40 Boxplots of sensitivity analysis for lane reduction bottleneck

The adaptive RL solutions for different MPR levels are illustrated in *Figure 41*. The controller chooses between seven different patterns to perform under different MPRs. The CVs are the means by which the controller regulates the traffic flow in the stream. When the MPR decreases the controller gradually stretches the length and increases the magnitude of its speed reduction to interact with more CVs.

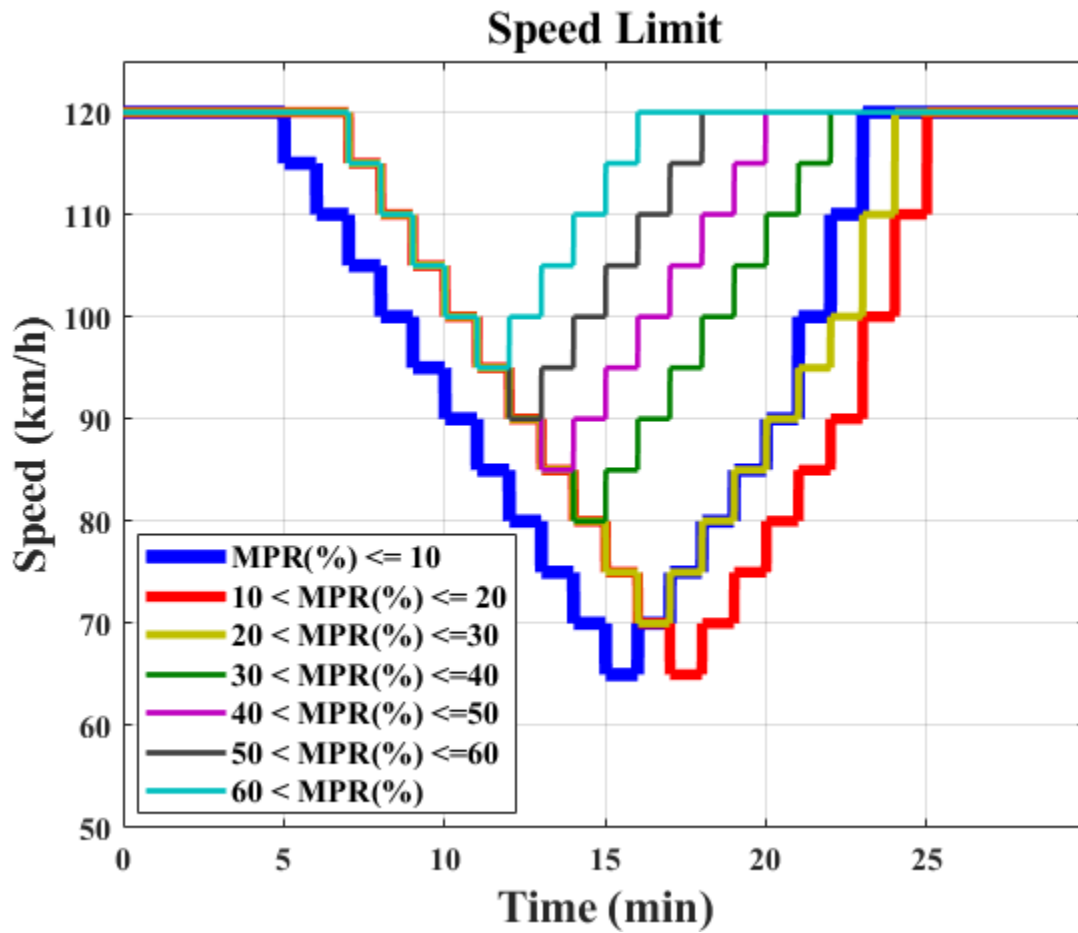


Figure 41 Imposed speed-limit of the adaptive RL controller

The results of 50 runs of adaptive RL algorithm for the on-ramp bottleneck are presented as boxplots in *Figure 42*. Similar to the other bottlenecks, as compared to the non-adaptive version, the adaptive controller was able to reduce the variation in performance significantly. The controller shows stability for majority of MPRs but less than 20% the variation starts to increase and there are still negative performances at 5% MPR.

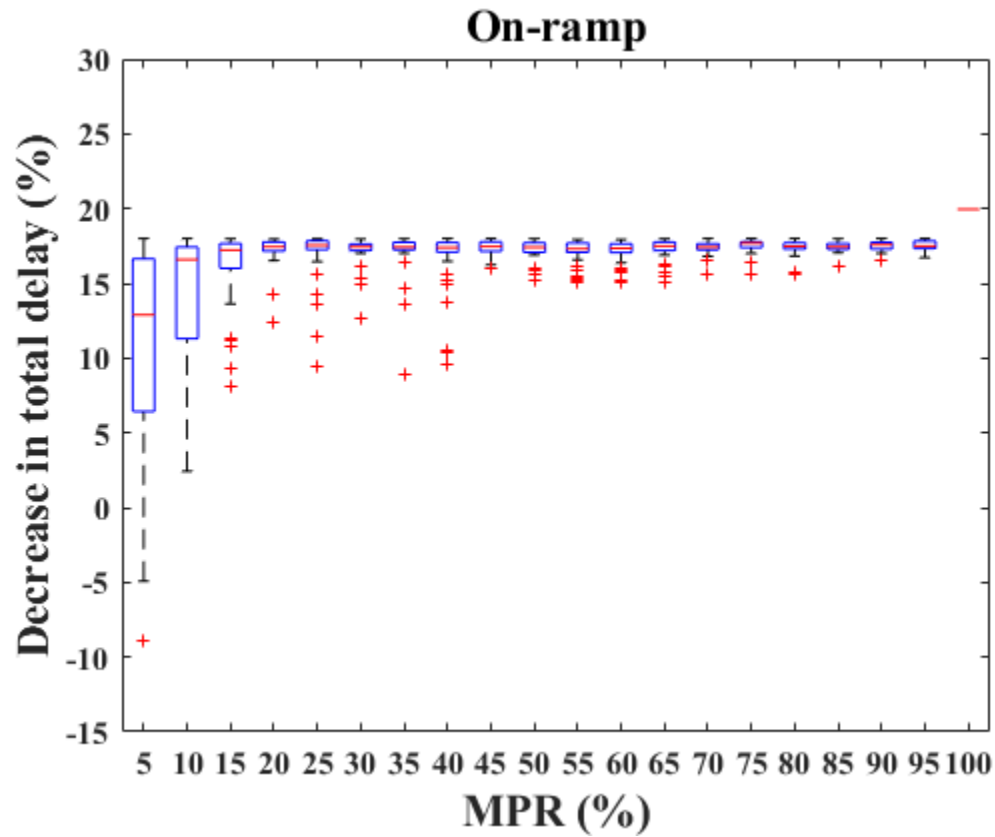


Figure 42 Boxplots of sensitivity analysis for lane reduction bottleneck

Figure 43, illustrates different patterns the controller chooses to interact with the environment in response to the different MPRs. Through five different patterns the controller tries to interact with the environment in the most efficient way and maximize the reward expectation which is equal to reduction in the total delay.

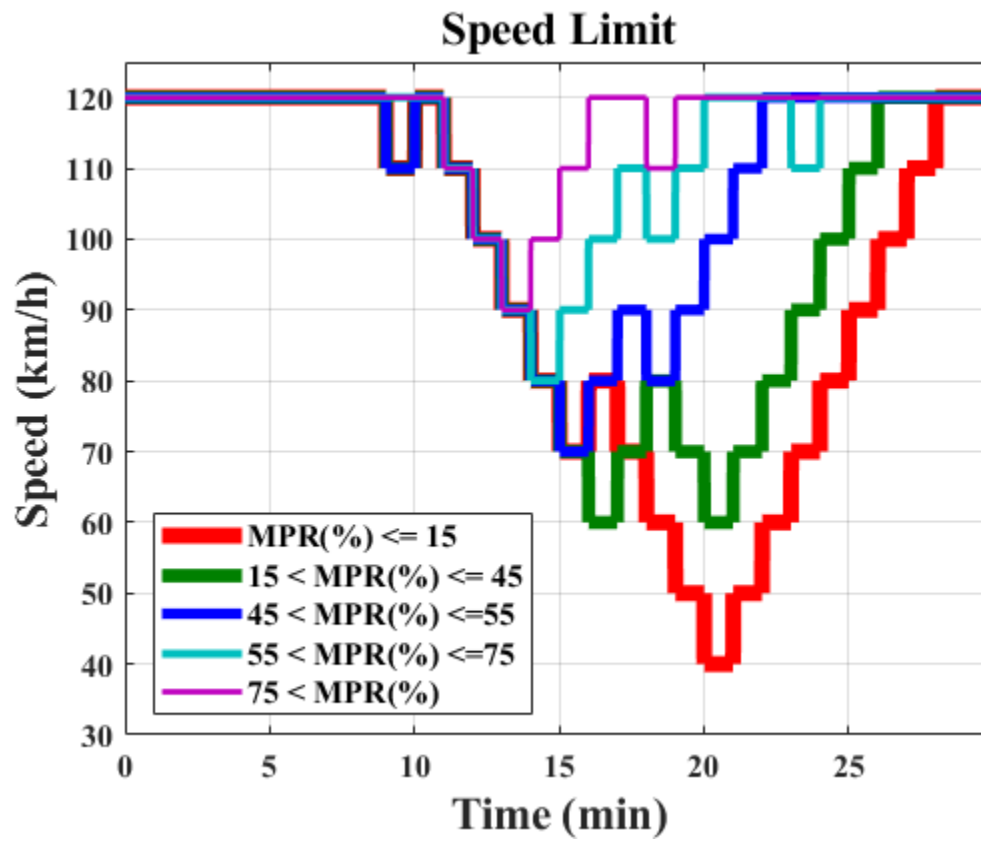


Figure 43 Imposed speed-limit of the adaptive RL controller

CHAPTER 4: CONCLUSION

In this study, an RL algorithm is developed and implemented in a simulation environment for controlling VSLs to manipulate the inflow of vehicles to the bottleneck on a freeway to minimize delays and increase throughput. Infrastructure-to-vehicle (I2V) communication for CVs is assumed to be available for implementing the VSL strategy. Asynchronous Advantage Actor-Critic (A3C) has been developed to control the VSL. Through the A3C control algorithm, the speed of CVs is manipulated in the control section at the upstream of the bottleneck to minimize congestion. Three different bottlenecks are investigated to understand the effectiveness of the RL algorithm on controlling VSL. These bottlenecks include a sag curve, lane reduction, and an on-ramp. Various market penetration rates for CVs are considered in simulations. It is demonstrated that the RL algorithm can be a general solution for all three bottlenecks in the sense that the same modeling structure can effectively improve system performance.

An RL agent is trained for each bottleneck. The optimal solution for different bottlenecks is not the same, but the results show that the RL outperforms feedback controller in all bottlenecks. If MPR is high enough, the RL controller is able to reduce nearly 63% of the total delay for the sag curve bottleneck, 64% of the total delay for lane reduction, and 19% of total the delay for on-ramp bottleneck. A sensitivity analysis shows that even with low MPR (e.g., 15%) the system can reduce delays significantly but the variation in performance can be high. The RL controller might induce a negative performance for lower MPRs (e.g., at less than 15% MPR for sag curve or lane reduction, and at less than 25% for on-ramp). The performance of the RL controller is bounded by its action space. The results demonstrate that not only the MPR of CVs but also how CVs are distributed in the traffic stream is critical for system performance. While MPR could be high, uneven distribution of CVs and lack of CVs at the critical periods as congestion is building up may

cause a deterioration in system performance. Hence, such systems should be designed with care, and the temporal and spatial distribution of CVs should also be accounted for while evaluating system performance.

It is shown that by adding CVs arrival rate as an input, the RL controller can learn to adapt its solution to the MPR in the traffic stream. The controller uses different patterns to change the VSL depending on what rate of CVs it senses at the upstream. This approach increased the performance of the controller significantly, even for low MPRs. It also prevents any negative performance for sag curve and lane reduction bottleneck. For the on-ramp bottleneck there are still some negative performances when the MPR is low.

A similar control strategy can be applied to networks with multiple bottlenecks, where capacity drop may occur. Further evaluation of the strategy will require investigating heterogeneous traffic. More complex VSL solutions such as separate VSL for each lane could also be considered to explore the performance of the proposed control strategy.

REFERENCES

- Aström & Murray 2010. *Feedback systems: an introduction for scientists and engineers*, Princeton university press.
- Bailo, Dzielicki, Smith, Spulber, Chen & Schultz 2018. *The Great Divide: What Consumers Are Buying vs. The Investments Automakers & Suppliers Are Making in Future Technologies, Products & Business Models*. Center for Automotive Research.
- Bakker, Whiteson, Kester & Groen 2010. *Traffic light control by multiagent reinforcement learning systems*. Interactive Collaborative Information Systems. Springer.
- Bando, Hasebe, Nakayama, Shibata & Sugiyama 1995. *Dynamical model of traffic congestion and numerical simulation*. Physical review E, 51, 1035.
- Belletti, Haziza, Gomes & Bayen 2018. *Expert level control of ramp metering based on multi-task deep reinforcement learning*. IEEE Transactions on Intelligent Transportation Systems, 19, 1198-1207.
- Bertini, Boice & Bogenberger 2006. *Dynamics of variable speed limit system surrounding bottleneck on german autobahn*. Transportation Research Record: Journal of the Transportation Research Board, 149-159. <https://doi.org/10.3141/1978-20>.
- Brackstone & McDonald 1996. *The microscopic modelling of traffic flow: Weaknesses and potential developments*.
- Brilon & Bressler 2004. *Traffic flow on freeway upgrades*. Transportation Research Record: Journal of the Transportation Research Board, 112-121. <https://doi.org/10.3141/1883-13>.
- Carlson, Papamichail & Papageorgiou 2011. *Local feedback-based mainstream traffic flow control on motorways using variable speed limits*. IEEE Transactions on Intelligent Transportation Systems, 12, 1261-1276.
- Carlson, Papamichail, Papageorgiou & Messmer 2010. *Optimal mainstream traffic flow control of large-scale motorway networks*. Transportation Research Part C: Emerging Technologies, 18, 193-212.
- Cassidy & Ahn 2005. *Driver turn-taking behavior in congested freeway merges*. Transportation Research Record: Journal of the Transportation Research Board, 140-147. <https://doi.org/10.3141/1934-15>.
- Chandler, Herman & Montroll 1958. *Traffic dynamics: studies in car following*. Operations research, 6, 165-184.
- Chen & Ahn 2015. *Variable speed limit control for severe non-recurrent freeway bottlenecks*. Transportation Research Part C: Emerging Technologies, 51, 210-230.
- Coppola & Morisio 2016. *Connected car: technologies, issues, future trends*. ACM Computing Surveys (CSUR), 49, 46. <https://doi.org/10.1145/2971482>.
- Daganzo 1994. *The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory*. Transportation Research Part B: Methodological, 28, 269-287.
- Daganzo 1997. *Fundamentals of transportation and traffic operations*, Pergamon Oxford.
- Davarynejad, Hegyi, Vrancken & Van Den Berg. *Motorway ramp-metering control with queuing consideration using Q-learning*. 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), 2011. IEEE, 1652-1658.
- El-Tantawy & Abdulhai. *Temporal Difference Learning-Based Adaptive Traffic Signal Control*. Proc., 12th World Conference on Transport Research, Lisbon, Portugal, 2010.
- Faghri & Hua 1992. *Evaluation of artificial neural network applications in transportation engineering*. Transportation Research Record, 1358, 71.
- Furuichi, Yamamoto, Kotani & Iwasaki 2003. *Characteristics of spatial speed change at motorway sag sections and capacity bottlenecks*. 82nd Annual Meeting of the Transportation Research Board, Washington, DC.
- Gipps 1981. *A behavioural car-following model for computer simulation*. Transportation Research Part B: Methodological, 15, 105-111.
- Gipps 1986. *A model for the structure of lane-changing decisions*. Transportation Research Part B: Methodological, 20, 403-414.
- Girshick 2015. *Fast r-cnn*. arXiv preprint arXiv:1504.08083.
- Goni Ros, Knoop, Van Arem & Hoogendoorn 2012. *Car-following behavior at sags and its impacts on traffic flow*. 91st Annual Meeting Transportation Research Board, Washington, USA, 22-26 January 2012; Authors version. TRB. <http://resolver.tudelft.nl/uuid:483db79a-6028-423b-932e-29c52bf77ba3>.
- Goni Ros, Knoop, Van Arem & Hoogendoorn 2014a. *Empirical analysis of the causes of stop-and-go waves at sags*. IET Intelligent Transport Systems, 8, 499-506. <http://dx.doi.org/10.1049/iet-its.2013.0102>.
- Goni Ros, Knoop, Van Arem & Hoogendoorn 2014b. *Mainstream traffic flow control at sags*. Transportation Research Record: Journal of the Transportation Research Board, 57-64. <https://doi.org/10.3141/2470-06>.

- Goodall, Smith & Park 2013. *Traffic signal control with connected vehicles*. Transportation Research Record: Journal of the Transportation Research Board, 65-72. <https://doi.org/10.3141/2381-08>.
- Greenshields, Channing & Miller. *A study of traffic capacity*. Highway research board proceedings, 1935. National Research Council (USA), Highway Research Board.
- Hall & Agyemang-Duah 1991. *Freeway capacity drop and the definition of capacity*. Transportation research record.
- Hegyi, Bellemans & De Schutter 2009. *Freeway traffic management and control*. Encyclopedia of Complexity and Systems Science, 3943-3964.
- Hegyi, De Schutter & Hellendoorn 2005. *Optimal coordination of variable speed limits to suppress shock waves*. IEEE Transactions on intelligent transportation systems, 6, 102-112.
- Hegyi, Hoogendoorn, Schreuder, Stoelhorst & Viti. *SPECIALIST: A dynamic speed limit control algorithm based on shock wave theory*. Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on, 2008. IEEE, 827-832. <https://doi.org/10.1109/ITSC.2008.4732611>.
- Helbing 2001. *Traffic and related self-driven many-particle systems*. Reviews of modern physics, 73, 1067.
- Helbing, Treiber, Kesting & Schönhof 2009. *Theoretical vs. empirical classification and prediction of congested traffic states*. The European Physical Journal B, 69, 583-598.
- Hidas & Wagner. *Review of data collection methods for microscopic traffic simulation*. World Conference on Transport Research Istanbul, Turkey, 2004.
- Hoel, Wolff & Laine. *Automated speed and lane change decision making using deep reinforcement learning*. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018. IEEE, 2148-2155.
- Hoogendoorn, Hoogendoorn, Brookhuis & Daamen 2011. *Adaptation longitudinal driving behavior, mental workload, and psycho-spacing models in fog*. Transportation Research Record: Journal of the Transportation Research Board, 20-28.
- Hoogendoorn, Van Zuylen, Schreuder, Gorte & Vosselman 2003. *Microscopic traffic data collection by remote sensing*. Transportation Research Record, 1855, 121-128.
- Iordanidou, Papamichail, Roncoli & Papageorgiou 2017. *Feedback-based integrated motorway traffic flow control with delay balancing*. IEEE Transactions on Intelligent Transportation Systems, 18, 2319-2329.
- Jayakrishnan, Mahmassani & Hu 1994. *An evaluation tool for advanced traffic information and management systems in urban networks*. Transportation Research Part C: Emerging Technologies, 2, 129-147.
- Kesting, Treiber & Helbing 2007. *General lane-changing model MOBIL for car-following models*. Transportation Research Record, 1999, 86-94.
- Keyvan-Ekbatani, Knoop & Daamen 2016. *Categorization of the lane change decision process on freeways*. Transportation research part C: emerging technologies, 69, 515-526.
- Komada, Masukura & Nagatani 2009. *Effect of gravitational force upon traffic flow with gradients*. Physica A: Statistical Mechanics and its Applications, 388, 2880-2894. <https://doi.org/10.1016/j.physa.2009.03.029>.
- Koshi 2003. *An interpretation of a traffic engineer on vehicular traffic flow*. Traffic and Granular Flow'01. Springer.
- Koshi, Iwasaki & Ohkura. *Some findings and an overview on vehicular flow characteristics*. Proceedings of the 8th International Symposium on Transportation and Traffic Flow Theory, 1983. University of Toronto: Toronto, Ontario, 403-426.
- Koshi, Kuwahara & Akahane 1992. *Capacity of sags and tunnels on Japanese motorways*. ite Journal, 62, 17-22. <http://www.plan.civil.tohoku.ac.jp/kuwahara/publications/1992-002.pdf>.
- Kreidieh, Wu & Bayen. *Dissipating stop-and-go waves in closed and open networks via deep reinforcement learning*. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018. IEEE, 1475-1480.
- Ku, Chou & Peng. *Discriminatively-learned global image representation using CNN as a local feature extractor for image retrieval*. Visual Communications and Image Processing (VCIP), 2015, 2015. IEEE, 1-4.
- Kwon, Brannan, Shouman, Isackson & Arseneau 2007. *Development and field evaluation of variable advisory speed limit system for work zones*. Transportation Research Record: Journal of the Transportation Research Board, 12-18.

- Laval 2009. *Effects of geometric design on freeway capacity: Impacts of truck lane restrictions*. Transportation research part B: methodological, 43, 720-728. <https://doi.org/10.1016/j.trb.2009.01.003>.
- Lee & Lim 2013a. *An empirical study on ad hoc performance of DSRC and Wi-Fi vehicular communications*. International Journal of Distributed Sensor Networks, 9, 482695.
- Lee & Park. *Evaluation of variable speed limit under connected vehicle environment*. Connected Vehicles and Expo (ICCVE), 2013 International Conference on, 2013b. IEEE, 966-967.
- Li, Liu, Xu, Duan & Wang 2017. *Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks*. IEEE transactions on intelligent transportation systems, 18, 3204-3217.
- Lighthill & Whitham 1955. *On kinematic waves II. A theory of traffic flow on long crowded roads*. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 229, 317-345.
- Lillicrap, Hunt, Pritzel, Heess, Erez, Tassa, Silver & Wierstra 2015. *Continuous control with deep reinforcement learning*. arXiv preprint arXiv:1509.02971.
- Lin, Kang & Chang. *Exploring the effectiveness of variable speed limit controls on highway work-zone operations*. Intelligent transportation systems, 2004. Taylor & Francis, 155-168. <https://doi.org/10.1080/15472450490492851>.
- Liu, Zhang, Sun & Wang 2015. *Optimize the settings of variable speed limit system to improve the performance of freeway traffic*. IEEE Transactions on Intelligent Transportation Systems, 16, 3249-3257.
- Ludmann 1998. *Beeinflussung des Verkehrsablaufs auf Strassen: Analyse mit dem fahrzeugorientierten Verkehrssimulationsprogramm PELOPS*, Forschungsges. Kraftfahrwesen.
- Lukin, Konovalov, Scherbakov & Breed. *Dedicated Short-Range Communication System for Vehicle-to-Vehicle Data Transmission on the Basis of Chaotic Waveform codes (DSRC-VVDT)*. Microwaves, Radar & Wireless Communications, 2006. MIKON 2006. International Conference on, 2006. IEEE, 442-445. <https://doi.org/10.1109/MIKON.2006.4345212>.
- Malenstein. *Automated Video Speed Enforcement and Trajectory Control Combined with Fully Automated Processing*. TOWARDS THE NEW HORIZON TOGETHER. PROCEEDINGS OF THE 5TH WORLD CONGRESS ON INTELLIGENT TRANSPORT SYSTEMS, HELD 12-16 OCTOBER 1998, SEOUL, KOREA. PAPER NO. 2022, 1998.
- Milanés, Shladover, Spring, Nowakowski, Kawazoe & Nakamura 2014. *Cooperative adaptive cruise control in real traffic situations*. IEEE Transactions on Intelligent Transportation Systems, 15, 296-305. <https://doi.org/10.1109/TITS.2013.2278494>.
- Mnih, Badia, Mirza, Graves, Lillicrap, Harley, Silver & Kavukcuoglu. *Asynchronous methods for deep reinforcement learning*. International conference on machine learning, 2016. 1928-1937.
- Mnih, Kavukcuoglu, Silver, Rusu, Veness, Bellemare, Graves, Riedmiller, Fiedelnd & Ostrovski 2015. *Human-level control through deep reinforcement learning*. Nature, 518, 529.
- Nagabandi, Yang, Asmar, Kahn, Levine & Fearing 2017. *Neural network dynamics models for control of under-actuated legged millirobots*. arXiv preprint arXiv:1711.05253.
- Nagel & Schreckenberg 1992. *A cellular automaton model for freeway traffic*. Journal de physique I, 2, 2221-2229.
- Newell 1993. *A simplified theory of kinematic waves in highway traffic, part I: General theory*. Transportation Research Part B: Methodological, 27, 281-287.
- Oguchi & Konuma 2009. *Comparative study of car-following models for describing breakdown phenomena at sags*. Proceedings of the 19th ITS World Congress. <https://ci.nii.ac.jp/naid/10026130670/en/>.
- Okamura, Watanabe & Watanabe 2000. *An empirical study on the capacity of bottlenecks on the basic suburban expressway sections in Japan*. Proceedings of the 4th International Symposium on Highway Capacity. https://ntl.bts.gov/lib/8000/8600/8612/11_65.pdf.
- Olia, Abdelgawad, Abdulhai & Razavi 2016. *Assessing the potential impacts of connected vehicles: mobility, environmental, and safety perspectives*. Journal of Intelligent Transportation Systems, 20, 229-243. <https://doi.org/10.1080/15472450.2015.1062728>.
- Papageorgiou, Diakaki, Dinopoulou, Kotsialos & Wang 2003. *Review of road traffic control strategies*. Proceedings of the IEEE, 91, 2043-2067. <https://doi.org/10.1109/JPROC.2003.819610>.
- Papageorgiou, Hadj-Salem & Middelham 1997. *ALINEA local ramp metering: Summary of field results*. Transportation Research Record: Journal of the Transportation Research Board, 90-98. <https://doi.org/10.3141/1603-12>.

- Pasquale, Anghinolfi, Sacone, Siri & Papageorgiou. *A comparative analysis of solution algorithms for nonlinear freeway traffic control problems*. Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on, 2016. IEEE, 1773-1778.
- Patire & Cassidy 2011. *Lane changing patterns of bane and benefit: Observations of an uphill expressway*. Transportation research part B: methodological, 45, 656-666. <https://doi.org/10.1016/j.trb.2011.01.003>.
- Payne 1971. *Model of freeway traffic and control*. Mathematical Model of Public System, 51-61.
- Popov, Hegyi, Babuška & Werner 2008. *Distributed controller design approach to dynamic speed limit control against shockwaves on freeways*. Transportation Research Record: Journal of the Transportation Research Board, 93-99.
- Rakha & Kamalanathsharma. *Eco-driving at signalized intersections using V2I communication*. Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on, 2011. IEEE, 341-346.
- Ramezani & Benekohal. *Optimized speed harmonization with connected vehicles for work zones*. Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on, 2015. IEEE, 1081-1086.
- Ren, He, Girshick & Sun. *Faster r-cnn: Towards real-time object detection with region proposal networks*. Advances in neural information processing systems, 2015. 91-99.
- Sallab, Abdou, Perot & Yogamani 2017. *Deep reinforcement learning framework for autonomous driving*. Electronic Imaging, 2017, 70-76.
- Schakel, Knoop & Van Arem 2012. *Integrated lane change model with relaxation and synchronization*. Transportation Research Record, 2316, 47-57.
- Schakel & Van Arem 2014. *Improving traffic flow efficiency by in-car advice on lane, speed, and headway*. IEEE Transactions on Intelligent Transportation Systems, 15, 1597-1606.
- Shalev-Shwartz, Shammah & Shashua 2016. *Safe, multi-agent, reinforcement learning for autonomous driving*. arXiv preprint arXiv:1610.03295.
- Silver, Schrittwieser, Simonyan, Antonoglou, Huang, Guez, Hubert, Baker, Lai & Bolton 2017. *Mastering the game of go without human knowledge*. Nature, 550, 354.
- Sohrabi, Ermagun & Ovaici 2017. *Finding Optimum Capacity of Freeways Considering Stochastic Capacity Concept*. 96th Annual Meeting Transportation Research Board, Washington DC, USA, 8-12 January TRB.
- Spiliopoulou, Kontorinaki, Papageorgiou & Kopelias 2014. *Macroscopic traffic flow model validation at congested freeway off-ramp areas*. Transportation Research Part C: Emerging Technologies, 41, 18-29.
- Sutton & Barto 1998. *Reinforcement learning: an introduction* MIT Press. Cambridge, MA.
- Sutton & Barto 2018. *Reinforcement learning: An introduction*, MIT press.
- Talebpoor, Mahmassani & Hamdar 2013. *Speed harmonization: evaluation of effectiveness under congested conditions*. Transportation Research Record: Journal of the Transportation Research Board, 69-79.
- Tilch & Helbing 2000. *Evaluation of single vehicle data in dependence of the vehicle-type, lane, and site*. Traffic and Granular Flow'99. Springer. https://doi.org/10.1007/978-3-642-59751-0_31.
- Toledo, Choudhury & Ben-Akiva 2005. *Lane-changing model with explicit target lane choice*. Transportation Research Record, 1934, 157-165.
- Treiber, Hennecke & Helbing 2000. *Congested traffic states in empirical observations and microscopic simulations*. Physical review E, 62, 1805. <https://doi.org/10.1103/PhysRevE.62.1805>.
- Treiber, Kesting & Helbing 2006. *Understanding widely scattered traffic flows, the capacity drop, and platoons as effects of variance-driven time gaps*. Physical Review E, 74, 016123. <https://doi.org/10.1103/PhysRevE.74.016123>.
- Van Wee, Annema & Banister 2013. *The transport system and transport policy: an introduction*, Edward Elgar Publishing.
- Vatani Nezafat, Beheshtitabar, Cetin, Williams & List 2018. *Modeling and Evaluating Traffic Flow at Sag Curves When Imposing Variable Speed Limits on Connected Vehicles*. Transportation Research Record, 0361198118784169.
- Vinitsky, Parvate, Kreidieh, Wu & Bayen. *Lagrangian control through deep-rl: Applications to bottleneck decongestion*. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018. IEEE, 759-765.
- Williams 1992. *Simple statistical gradient-following algorithms for connectionist reinforcement learning*. Machine learning, 8, 229-256.

- Wolshon & Lambert 2006. *Reversible lane systems: Synthesis of practice*. Journal of Transportation Engineering, 132, 933-944.
- Xing, Muramatsu & Harayama 2014. *Balance lane use with VMS to mitigate motorway traffic congestion*. International journal of intelligent transportation systems research, 12, 26-35.
<https://doi.org/10.1007/s13177-013-0067-7>.
- Xu, Li, Zhao, Zhang & Wang 2017. *DSRC versus 4G-LTE for connected vehicle applications: a study on field experiments of vehicular communication performance*. Journal of Advanced Transportation, 2017.
- Yokota 1998. *A study of AHS effects on traffic flow at bottlenecks*. Towards the new horizon together. Proceedings of the 5th world congress on intelligent transport systems, Seoul, Korea. 12-16 October 1998, Paper No. 3200.
- Yoshizawa, Shiomi, Uno, Iida & Yamaguchi 2012. *Analysis of car-following behavior on sag and curve sections at intercity expressways with driving simulator*. International Journal of Intelligent Transportation Systems Research, 10, 56-65. <https://doi.org/10.1007/s13177-011-0043-z>.
- Zegeye, De Schutter, Hellendoorn, Breunese & Hegyi 2012. *A predictive traffic controller for sustainable mobility using parameterized control policies*. IEEE Transactions on Intelligent Transportation Systems, 13, 1420-1429.
- Zhao, Bai, Wang, Xu & Yu 2011. *DHP method for ramp metering of freeway traffic*. IEEE Transactions on Intelligent Transportation Systems, 12, 990-999.
- Zhu & Ukkusuri 2014. *Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach*. Transportation research part C: emerging technologies, 41, 30-47.

VITA

Since graduating with a BSc from the University of Guilan, Iran in 2014, Reza has been a Ph.D. student at Old Dominion University studying in the Civil and Environmental Engineering Department. In the course of his graduate studies, he has focused on development, and evaluation of artificial intelligence and machine learning approaches for transportation applications.