Reinforcement Learning-Based Variable Speed Limits Control to Reduce Crash Risks Near Traffic Oscillations on Freeways

Zhibin Li,* Chengcheng Xu, Yanyong Guo, and Pan Liu
Are with the School of Transportation, Southeast University, China.
E-mail: lizhibin@seu.edu.cn

Ziyuan Pu
Is within the Department of Civil and Environmental Engineering,
University of Washington, United States.

Digital Object Identifier 10.1109/MITS.2019.2907631
Date of current version: 9 January 2020

*Corresponding author

©SHUTTERSTOCK.COM/NUT SUPHAWONG

1939-1390/21©2021IEEE

***Abstract*—Traffic oscillation on freeways results in abrupt and frequent vehicle deceleration, which remarkably increases crash risks. Artificial Intelligence based traffic control provides a new opportunity for addressing the safety issues. This study aims at proposing a Reinforcement Learning (RL)-based variable speed limits (VSL) control algorithm to reduce crash risks associated with oscillations. The state, action and reward, which are the critical components in the RL, were designed carefully for safety improvement. The RL was trained to learn the optimal speed limit for various traffic states to achieve a goal of safety optimization. A rear-end crash risk model was applied to assess crash risks associated with oscillations near freeway bottlenecks. The cell transmission model was modified as the simulation platform for evaluating the control effects. The results showed that after the training process, the proposed RL-based VSL control successfully reduced the crash risks by 19.4%. A continuous online learning function was developed in RL to enhance the robustness of our strategy. The results showed that with continuous learning, the RL-based VSL control performed reasonably well under lower driver compliance situations.

I. Introduction

Traffic oscillations on freeways refer to the stop-and-go or slow-and-go driving conditions which propagate upstream against the direction of travel in congested traffic [1]. Oscillation forces approaching vehicles adjust their traveling speeds frequently and abruptly. Previous studies have reported that most of collision risks near freeway bottlenecks are highly associated with traffic oscillations [2]–[4].

Variable speed limit (VSL) has been used as an innovative approach to improve traffic safety on freeways. Control strategies heavily affect safety effects of VSL. Early VSL control algorithms were open loop which used deterministic speed limit values for particular traffic conditions [5]–[7]. Later, studies proposed close loop-based algorithms which adjusted speed limits based on control feedbacks [8]–[10]. However, the aforementioned VSL strategies are reactive in nature and cannot respond proactively to traffic changes. The feedback features may lose the best control opportunity and cause some oscillations during the settling time, which in consequence decrease control effects. Some studies proposed Model Predictive Control (MPC)-based VSL strategies which could respond proactively to traffic changes [11]–[12]. However, such strategies require accurate models to predict evolution of freeway traffic flow which may not always be available. Some recent studies developed integrated VSL and automated vehicle driving techniques in a connected vehicle environment to reduce crash risks [13]–[15]. The drawback of such strategies is that they required known information of all vehicles which is not applicable currently.

In recent years, Artificial Intelligence (AI) has attracted an increasing attention in many areas. AI is usually defined as the intelligent agent that perceives its environment and takes actions that maximize its chance of success at some goal [16]. The AI technique has a potential for application in traffic control tasks. Some recent studies have applied the reinforcement learning (RL) method for traffic control purposes [17]–[20]. A well-trained RL agent can make predictions on system evolution and achieve a proactive control sc`heme [21]–[22]. Previously, some studies have incorporated the RL with the ramp metering control [23]–[24]. Recently, Zhu and Ukkusuri [25] developed a RL for dynamic speed limit control to reduce travel time in a large roadway network. In our previous study [26], we proposed a RL-based VSL control to reduce total travel time near a freeway recurrent bottleneck.

The literature review suggested that none of previous studies has applied the RL in VSL control to reduce crash risks. Thus, this study aims at evaluating the potential of incorporating the RL in the VSL control to reduce crash risks associated with traffic oscillations near freeway recurrent bottlenecks. The key factors such as state, action and reward for safety optimization is designed to achieve such goal.

II. RL-Based VSL Control Strategy

A RL interacts in discrete time steps with its environment which is typically formulated as a Markov decision process (MDP). The VSL control problem can be formulated as a MDP problem and can be processed by RL technique [25]–[26]. The overall methodological framework includes four parts. A RL agent first needs to decide the state set, action set, reward function and learning parameters which affect the learning performance. Then The VSL control strategy is proposed based on the characteristics of study site, VSL sign location, and upper and lower speed limit values. The simulation platform which considers vehicle movement, road geometry and crash risk prediction model should be developed for training and testing purposes. The continuous learning function is enabled in our proposed RL-based VSL control agent in order to account for uncertainties in real environment such as severe congestion, bad weather and local driver behaviors.

A. Basic RL Algorithm

Q-Learning is one of the most commonly used RL algorithms [27]. At each time step, the agent perceives the state of the environment and takes an action to transfer the current state to a new state. The agent receives a reward to evaluate the quality of the transition. By evaluating the rewards of multiple actions, the agent learns how to find a sequence of optimal actions that yields a maximum cumulative reward. A Q-value is assigned to each state-action combination:

$$\mathbf{Q}: \mathbf{S} \times \mathbf{A} \rightarrow \mathbf{R} \tag{1}$$

where $\mathbf{S}$ is the set of states, $\mathbf{A}$ is the set of actions, and $\mathbf{R}$ is the set of rewards. In an infinite horizon discounted reward problem, the agent's goal is to maximize

$$\sum_{t=0}^{\infty} \gamma^t R_t \tag{2}$$

where R_t is the reward at time step t, and γ^t is the discount factor $(0 \leqslant \gamma \leqslant 1)$. The Q-value is updated with new training samples:

$$Q^{t+1}(s_t, a_t) = Q^t(s_t, a_t) + \kappa_{(s,a)}[R_{t+1} + \gamma \cdot \max Q^t(s_{t+1}, a_{t+1}) - Q^t(s_t, a_t)] \tag{3}$$

where $Q^{t+1}(s_t, a_t)$ is the Q-value for the state-action pair (s_t, a_t) at time step $t+1$, R_{t+1} is the reward received after performing action a_t at state s_t and then moves to the new state s_{t+1}, and $\kappa(s, a)$ is the learning rate.

B. RL-Based VSL Control Strategy

The critical elements in the RL-based VSL control include the state, action, reward, and learning parameters:

1) *State*. The QL agent uses a state table as a collection of possible states. Continuous state space was represented by discrete values. As learning time increases exponentially as the number of states increases, only traffic parameters that have a dominant effect on traffic dynamics should be considered for defining the states. The reduced speed limit not only affects the crash risk in the VSL controlled section, but also affects the risk in the immediate upstream section and downstream section. Thus, density within the three sections are used to be the state set (see Fig. 1). Upper and lower boundaries for density were determined via multiple simulation runs. As traffic flow changes sensitively near the critical density, a small interval of 2 veh/mi/ln was used to divide states near critical density. A larger interval of 8 veh/mi/ln was used to divide other traffic conditions. We found that using such design could achieve good results while keep a reasonable learning time. The final state set was {8, 16, 24, 26, 28, 30, 38, 46, 54, 62, 70, 78, 86, 94, 102 veh/mi/ln}. The total state space was 15 × 15 × 15 = 3375.
2) *Action*. The speed limit was designed as the action in the QL-based VSL control. The speed limit range is set to be from 20 mph to 65 mph with an increase of 5 mph. That is, the set of action used in the RL algorithm is {20, 25, 30, 35, 40, 45, 50, 55, 60, 65}. To avoid disturbances to traffic, the maximum speed limit change rate was set to be 10 mph per minute.
3) *Reward*. The objective of the QL-based VSL control is to minimize the total risks of crashes. The total crash risk (*TCR*) over a time horizon K is given by:

$$TCR = \sum_{i=1}^{I} \sum_{k=1}^{K} R(i, k) \tag{4}$$

where $R(i, k)$ is the crash risk in section i and time k, and I is the number of sections. The crash risk was calculated using a crash risk prediction model as introduced in the next section. The reward function was calculated as the decrease of crash risk with the VSL control as compared to the no control case. An additional penalty was added to the reward value if the VSL increased crash risks. The reward is given as:

$$\text{Reward} = -\frac{TCR_{\text{VSL}} - TCR_{\text{NO}}}{TCT_{\text{NO}}} \times 100 \tag{5}$$

4) *Learning Parameter*. The learning rate is determined as a function of visits to that state-action combination [23]–[25]:

$$\kappa_{(s,a)} = \left[\frac{1}{1 + C(s, a)(1-\gamma)}\right]^{0.7} \tag{6}$$

The QL algorithm needs to balance the exploitation and exploration when selecting actions. In our study, the selection of action followed the Softmax strategy [18], [23].

III. Model Development

A. Simulation Model for VSL Control

The decision of VSL control is made based on aggregated traffic variables. The RL-based VSL determines speed limits according to traffic densities. Besides, the crash risk prediction model for evaluating safety effects of VSL was also proposed based on aggregated loop detector data. Detailed information from individual vehicles is actually not considered. Thus, in this study, the cell transmission model (CTM) was used for the simulation of traffic flow on freeway mainlines. The CTM is a macroscopic simulation approach proposed by Daganzo based on kinematic wave theory [28]. The popularity of CTM is due to its low computation requirements and the ease with which it can be calibrated [29]–[30]. The CTM predicts macroscopic traffic flow characteristics at different time steps according to the fundamental diagram.

To build the simulation platform for RL-based VSL control tests, several modifications were further made to the

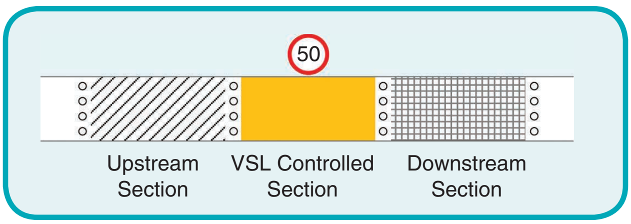


FIG 1 Determination of state set in the RL agent.

traditional CTM. Details of the simulation model development can be found in our previous study [26], [31]. Major modifications are: 1) The impact of VSL control on the fundamental diagram was specified. In such way, the CTM can simulate traffic operation under the control of VSL; 2) By specifying the inverse λ-shaped fundamental diagram for the bottleneck cell, the CTM can generate the phenomenon of capacity drop at a proper magnitude; 3) A stochastic component was introduced in the sending flow function for the bottleneck cell to generate oscillation waves repeatedly in the congested traffic conditions. Once generated, those waves propagate toward upstream sections; 4) The crash risk prediction model was integrated into the simulation platform; and 5) The RL-based VSL control strategy was integrated into the CTM.

B. Crash Risk Prediction Model

To evaluate the safety effects of VSL control, a relationship must be established between the risks of crashes and freeway traffic variables. In this study, the crash risk prediction model developed in the authors' previous study was used to assess crash risks [32]. The model was developed for rear-end crashes near freeway bottleneck areas with oscillation waves, which is consistent with the VSL control environment in the study.

A rear-end collision risk index that took into consideration the characteristics of deceleration trajectories of vehicles was proposed. A case-control strategy design was used to identify the factors that affected the risks of crashes. A logistic regression model was then developed to relate the rear-end collision likelihood to some traffic variables. The model was calibrated based on the 3-year crash and traffic data on Interstate 880 in California, US. The output of the model is the crash risk $R(i, k)$ in each section i at each time step k which are calculated according to the traffic speed and occupancy data at the upstream and downstream loop detector stations. For more information regarding the crash risk prediction model please refer to [32]. Note that if other collision models are used, the optimal RL control strategy may be changed. But researchers can follow our RL-based VSL control framework to obtain the optimal strategy for their models.

IV. Experiment Design

The freeway segment selected for the evaluation of the RL-based VSL control was on Interstate 880 in California (see Fig. 2). The section contains a recurrent bottleneck at its downstream end (see the dark triangle in Fig. 2). The resulting queues from the bottleneck typically increase in length and propagate towards upstream sections, resulting in large traffic disturbances and high collision potentials within the area.

The CTM parameters were calibrated using the empirical loop detector data collected on the study site. There are fifteen loop detectors within the study area and one-month traffic data were collected from the Freeway Performance Measurement System (PeMS) database which reported average speed, occupancy and flow every 30 s. The procedure presented in our previous studies was followed to calibrate the parameters [31], [33]. The free flow speed was found to be 65 mph. The capacity was 1900 veh/h/ln. The magnitude of capacity drop was 6.7%. The speed of the kinematic wave was estimated to be 12 mph. Each simulation run lasted for 100 min after the initial 10 min warm-up period. VSL does not work in the warm-up period. The peak period started from 20 min and lasted to 60 min with a mainline demand of 7120 veh/h. The off-peak mainline demand was set to be 4000 veh/h.

The VSL signs were assumed to be installed on each detector location in Fig. 2. The speed limit in each road segment was calculated independently based on the VSL control strategy. The RL-based VSL algorithm was trained in the following steps. In each cycle, the algorithm received the system density states, and selected a speed limit action that was to be posted in the next cycle. After the action was implemented, the freeway traffic evolved to a new state. Then the algorithm received a reward for the selected action and specified the Q-value for that state-action pair. The Softmax method was followed to select actions. The learning process continued until the Q-values for all the state-action pairs reached convergence. For more details regarding the learning parameter calibration and the Softmax strategy in the RL, please refer to our previous study [26]. The RL learning time in our experiment is about 20 min.

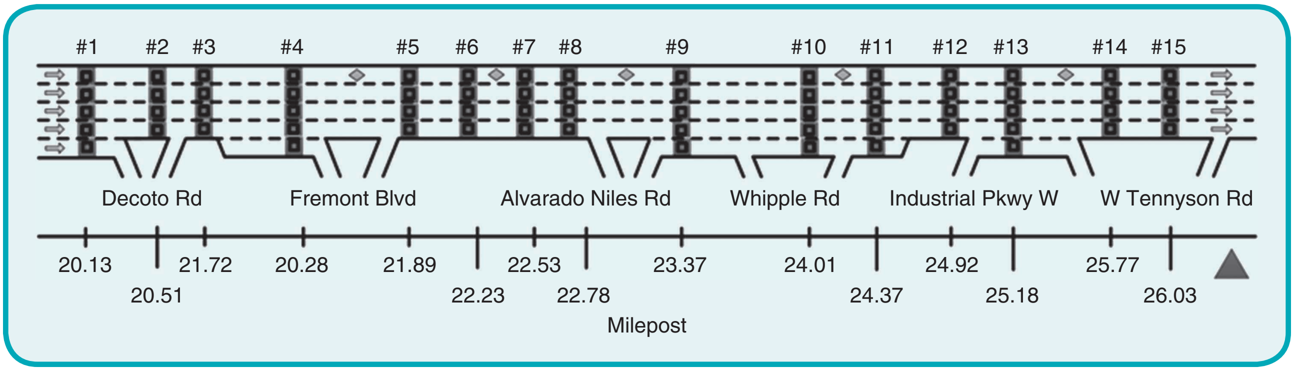


FIG 2 Study site for RL-based VSL control.

V. Evaluation of Control Strategy

A. Safety Effects of RL-based VSL Control

The trained RL-based VSL control was applied for the study segment and the effects on traffic operation and crash risk were estimated in the simulation model. For comparison purpose, the measurements without control were also estimated. The results are shown in Table I. They suggested that the RL-based VSL control strategy effectively reduced the crash risks within the study segment. More specifically, in the no control case, the total crash risk associated with traffic oscillation was 1075.9%/mi/h. With the RL-based VSL control, the total crash risk was 867.6%/mi/h which implied a 19.4% reduction as compared to the no control. The total travel time with the VSL control was slightly increased by 1.5%. This is because our VSL control reduced vehicle speeds earlier before reaching downstream congestion. The speed reduction period with the VSL control is larger than the no control case due to the gradual spatial and temporal change of traffic speed (see Fig. 3).

Table I. Effects of the RL-based VSL control.

Measurement	No Control	RL-Based VSL	VSL in [29]
Total travel time (h)	197.3	200.3	200.4
Percent of increase of total travel time	/	1.5	1.6
Total crash risk (%/mi/h)	1075.9	867.6	825.2
Percent of reduction of total crash risk	/	19.4	23.3

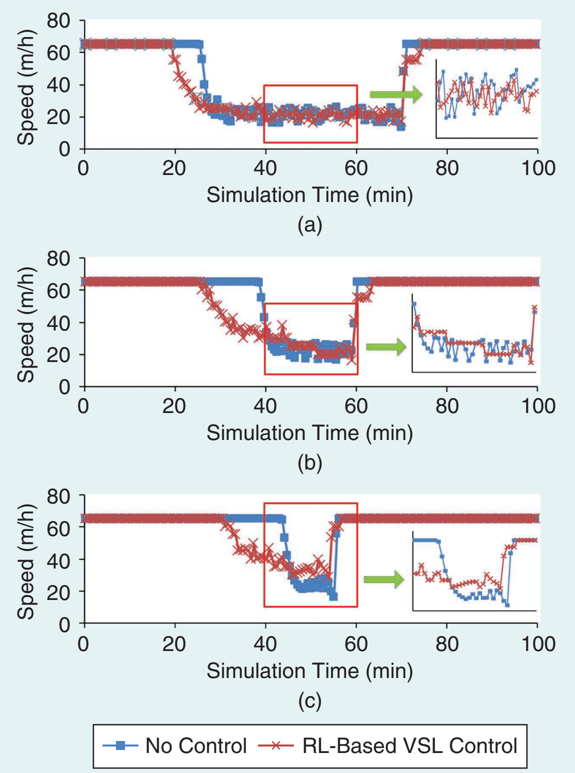


FIG 3 Speed curve with and without VSL control at different locations. (a) Location #13, (b) Location #10, and (c) Location #7.

The research team further compared the speed profiles with and without the VSL control. The speed data measured at three loop detector locations are shown in Fig. 3. The comparison suggested that two major differences existed between the two speed profiles: (1) In the no control case, traffic speed dropped suddenly from the free flow speed (65 mph) to the speed in congestion (around 20 mph). With the RL-based VSL control, a smoother reduction of speed was achieved as the VSL reduced speed gradually before oscillation arrived; and (2) The traffic variation within the congestion region was smoother with the RL-based VSL control than the no control (see Fig. 3), which indicated a reduced collision risks [32], [34]. The reason for the better control effects is that the trained RL agent has the capability of predicting traffic evolution and state transition. Thus, the controller can make a proactive action in response to traffic changes in order to reduce crash risks.

We further compared the RL-based VSL with the optimal VSL strategy proposed in our previous study [31]. The optimal strategy applied a genetic algorithm to find the best values for the control factors in VSL strategies for safety maximization. The optimal strategy reduced total crash risk by 23.3% which is higher than the RL-based VSL. The result is reasonable since the optimal strategy can be considered as the mathematical optimization to the control task. As our RL agent used limited and discontinued state representatives, its performance is greatly affected. Complex algorithms such as Deep Q-Learning with continuous states could be explored in our future studies.

B. Continuous Learning Under Uncertainties

In the RL-based VSL control, an online continuous learning agent was proposed to keep optimizing the actions after the VSL control is implemented on local freeways. The RL agent collects the actual rewards for various traffic state-action pairs from new environments. The new rewards are updated in the Q-table, which determines the optimal control actions, after a certain time period. Then the QL agent can learn the new optimal actions with the updated table. As such, the proposed RL-based VSL control strategy will be able to accommodate to the uncertainties in the practice. One of the critical concerns with the VSL control in practice is related to the ways drivers respond to the posted speed limits. Such issue can be addressed by the continuous learning of the RL-based VSL strategy. An example is given to help illustrate how continuous learning helps address the overspeed issue. In the experiment,

Table II. safety effects of continuous learning in VSL control.			
Overspeed Magnitude	5 mph	10 mph	15 mph
Reduction of total crash risk (%)			
RL-based VSL without continuous learning	15.5	11.0	6.8
RL-based VSL with continuous learning	18.0	16.1	12.8
VSL strategy in [29]	18.62	13.3	8.1

the magnitude of overspeed was set to be 5 to 15 mph. The control effects with and without the continuous learning were estimated and the results are shown in Table II.

The results suggested that in general, the VSL control effects in reducing crash risks decreased as the magnitude of overspeed increased. For example, the reduction of total crash risk was 15.5% when the overspeed was 5 mph. With the continuous learning, the safety effects of the VSL control were obviously improved. For example, the reduction in crash risk increased to 18.0% at the same level of overspeed. The superiority of the continuous learning module was even more obvious when the overspeed became larger. Such results suggest that the continuous learning enhances the robustness of our strategy. We further evaluated the performance of the optimal VSL strategy proposed in [31]. The results suggested that the performance of the optimal strategy was greatly affected by the overspeed factor. It indicates the robustness of the optimal VSL strategy is not very good as compared to the RL-based VSL control with continuous learning function.

VI. Conclusions

This study proposed an optimal RL-based VSL control strategy to reduce crash risks associated with traffic oscillations near freeway bottlenecks. The state, action, and reward were carefully designed in the RL agent for the control purpose. The RL agent was then trained to learn the optimal speed limit for various traffic states. With a rear-end crash risk prediction model, the safety effects of the RL-based VSL were evaluated in the simulation platform developed based on the CTM. The results showed that the proposed algorithm reduced total crash risks by 19.4% while only increased total travel time by 1.5%. An online continuous learning function was developed in RL to enhance the robustness of our strategy. The results showed that with continuous learning, the RL-based VSL control increased the safety effects under lower driver compliance situations.

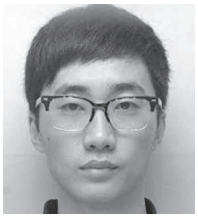
The RL-based VSL agent can automatically determine the ideal control action in environments for safety improvement. It can also mimic the cognitive function of human minds such as continuous learning ability and new environment adaptiveness, which is consistent with the nature of AI. Note that our strategy works well for fixed bottlenecks such as those caused by ramps, lane reduction, and work zone that have recurrent congestion patterns. For bottlenecks that are caused by incidents or do not create traffic oscillations, our strategy may not well in reducing crash risks. In addition, a RL-based VSL strategy that jointly maximizing safety and reducing travel time on freeways will be encouraging. Future studies may focus on those scenarios.

Acknowledgment

This work was jointly supported by the National Natural Science Foundation of China (71871057, 51678132), and the Fundamental Research Funds for the Central Universities (2242018K41009).

About the Authors

Zhibin Li received his Ph.D. degree from the School of Transportation at the Southeast University, China in 2014. He is currently a professor in Southeast University. From 2015 to 2017, he worked as a postdoctoral researcher in the University of Washington and the Hong Kong Polytech University. From 2010 to 2012, he was a visiting student in the University of California, Berkeley. His research interests include Intelligent Transportation, Traffic Safety, Data Mining, Traffic Control, etc. He received China National Scholarship twice, and won the Best Doctoral Dissertation Award by China Intelligent Transportation Systems Association in 2015.



Ziyuan Pu received a B.S. degree in transportation engineering from Southeast University, Nanjing, China (2010) and a M.S. degree in civil and environmental engineering from University of Washington, Seattle, US (2015). He is currently working toward the Ph.D. degree in civil and environmental engineering at University of Washington, Seattle, WA, US. Since 2015, he has been a Research Assistant with the Smart Transportation Applications and Research Laboratory (STAR Lab), University of Washington. His research interests include remote sensing technology, deep learning, machine learning, traffic data mining and intelligent transportation systems.

Chengcheng Xu received the Ph.D. (2014) degree from the School of Transportation at the Southeast University in Nanjing, China. He is currently an associate professor with the Key Laboratory of Traffic Planning and Management at the Southeast University. He won the Best Doctoral Dissertation Award by China Intelligent Transportation Systems Association in 2016. His research interest focuses on traffic safety, data mining, traffic operation, machine learning, etc.

Yanyong Guo received the M.S. degree in transportation engineering from Chang'an University, Xi'an, China, in 2012, and the Ph.D. degree in transportation engineering from Southeast University, Nanjing, China, in 2016. From 2014 to 2015, he was a Visiting Ph.D. Student with The University of British Columbia, where he is currently a Postdoctoral Fellow. His research interests include road safety evaluations, traffic conflicts techniques, advanced statistical techniques for safety evaluation, and pedestrian and cyclist behaviors. He received the China National Scholarship, in 2014, and the Best Doctoral Dissertation Award from the China Intelligent Transportation Systems Association, in 2017.

Pan Liu received his Ph.D. degree in Civil Engineering from the University of South Florida in Tampa, US in 2006. He is currently a professor at the School of Transportation at the Southeast University in Nanjing, China. His research interests include traffic operations and safety, and intelligent transportation systems. He is the recipient of the Distinguished Young Scientist Foundation of NSFC in 2013, and the award of the Program for New Century Excellent Talents funded by the Ministry of Education, China.

References

[1] M Mauch and M. J. Cassidy, "Freeway traffic oscillations: Observations and predictions," in *Proc. Transportation and Traffic Theory in Century Int. Symp. Transportation and Traffic Theory*, Transportation Center, University of California, 2004, pp. 653–674.

[2] D. Chen, J. A. Laval, S. Ahn, and Z. Zheng, "Microscopic traffic hysteresis in traffic oscillations: A behavioral perspective," *Transp. Res. B, Methodol.*, vol. 46, no. 10, pp. 1440–1453, 2012.

[3] Z. Li, K. Chung, and M. J. Cassidy, "Collisions in freeway traffic: Influence of downstream queues and interim means to address them," *Transp. Res. Rec.*, vol. 2396, no. 1, pp. 1–9, 2013.

[4] J. A. Laval and L. Leclercq, "A mechanism to describe the formation and propagation of stop-and-go waves in congested freeway traffic," *Philosoph. Trans. A, Math. Phys. Eng. Sci.*, vol. 368, no. 1928, pp. 4519–4541, 2010.

[5] C. Lee, B. Hellinga, and F. Saccomanno, "Evaluation of variable speed limits to improve traffic safety," *Transp. Res. C, Emerg. Technol.*, vol. 14, no. 3, pp. 213–228, June 2006.

[6] M. Abdel-Aty, J. Dilmore, and A. Dhindsa, "Evaluation of variable speed limits for real-time freeway safety improvement," *Accid. Anal. Prev.*, vol. 38, no. 2, pp. 335–345, Mar. 2006.

[7] P. Allaby, B. Hellinga, and M. Bullock, "Variable speed limits: Safety and operational impact of a candidate control strategy for freeway applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 671–680, Dec. 2007.

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

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

[10] G. R. Iordanidou, I. Papamichail, C. Roncoli, and M. Papageorgiou, "Feedback-based integrated motorway traffic flow control with delay balancing," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2319–2329, 2017.

[11] G. S. V. D. Weg, A. Hegyi, S. P. Hoogendoorn, and B. De Schutter, "Efficient model predictive control for variable speed limits by optimizing parameterized control schemes," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 2015, pp. 1137–1142.

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

[13] Y. Han, D. Chen, and S. Ahn, "Variable speed limit control at fixed freeway bottlenecks using connected vehicles," *Transp. Res. B, Methodol.*, vol. 98, pp. 113–134, 2017.

[14] D. Chen, S. Ahn, and A. Hegyi, "Variable speed limit control for steady and oscillatory queues at fixed freeway bottlenecks," *Transp. Res. B, Methodol.*, vol. 70, pp. 340–358, 2014.

[15] Z. He, L. Zheng, L. Song, and N. Zhu, "A jam-absorption driving strategy for mitigating traffic oscillations," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 802–813, 2017.

[16] D. Poole, A. Mackworth, and R. Goebel, *Computational Intelligence: A Logical Approach*. New York: Oxford Univ. Press, 1998.

[17] B. Abdulhai, R. Pringle, and G. J. Karakoulas, "Reinforcement learning for true adaptive traffic signal control," *J. Transp. Eng.*, vol. 29, no. 3, pp. 278–285, 2003.

[18] S. El-Tantawy and B. Abdulhai, "Towards multi-agent reinforcement learning for integrated network of optimal traffic controllers," *Transp. Lett.*, vol. 2, no. 2, pp. 89–110, 2010.

[19] T. Liu and X. Hu, "A bi-level control for energy efficiency improvement of a hybrid tracked vehicle," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1616–1625, 2018.

[20] T. Liu, B. Wang, and C. Yang, "Online Markov Chain-based energy management for a hybrid tracked vehicle with speedy Q-learning," *Energy*, vol. 160, pp. 544–555, 2018.

[21] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press, 1998.

[22] S. El-Tantawy and B. Abdulhai, "Comparative analysis for temporal difference learning methods in adaptive traffic signal control," presented at the 12th World Conf. Transport Research, Lisbon, Portugal, 2010.

[23] D. Zhao, X. Bai, F. Wang, and J. Xu, "DHP method for ramp metering of freeway traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 990–999, 2011.

[24] K. Rezaee, B. Abdulhai, and H. Abdelgawad, "Self-Learning adaptive ramp metering: Analysis of design parameters on a test case in Toronto," presented at 92nd Annu. Meeting of the Transportation Research Board, Washington, D.C., 2013.

[25] F. Zhu and S. V. Ukkusuri, "Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach," *Transp. Res. C, Emerg. Technol.*, vol. 41, pp. 30–47, 2014.

[26] Z. Li, P. Lin, C. Xu, H. Duan, and W. Wang, "Reinforcement learning-based variable speed limit control strategy to reduce traffic congestion at freeway recurrent bottlenecks," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3204–3217, 2017.

[27] C. Watkins and P. Dayan, "Q-Learning," *Mach. Learn.*, vol. 8, no. 3, pp. 279–292, 1992.

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

[29] H. B. Celikoglu, "Dynamic classification of traffic flow patterns simulated by a switching multimode discrete cell transmission model," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2539–2550, 2014.

[30] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "Piecewise-linearized cell transmission model and parameter calibration methodology," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1965, pp. 183–191, 2006.

[31] Z. Li, P. Liu, W. Wang, and C. Xu, "Development of a control strategy of variable speed limits to reduce rear-end collision risks near freeway recurrent bottlenecks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 866–877, 2014.

[32] Z. Li, S. Ahn, K. Chung, D. R. Ragland, W. Wang, and J. W. Yu, "Surrogate safety measure for evaluating rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks," *Accid. Anal. Prev.*, vol. 64, pp. 52–61, 2014.

[33] Z. Li, P. Liu, C. Xu, and W. Wang, "Optimal mainline variable speed limit control to improve safety on large-scale freeway segments," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 31, no. 5, pp. 366–380, 2016.

[34] Z. Li, W. Wang, R. Chen, and C. Xu, "Evaluation of the impacts of speed variation on freeway traffic collisions in various traffic states," *Traffic Inj. Prev.*, vol. 14, no. 8, pp. 861–866, 2013.

ITS