Reinforcement Learning-Based Variable Speed Limit Control Strategy to Reduce Traffic Congestion at Freeway Recurrent Bottlenecks

Zhibin Li, Pan Liu, Chengcheng Xu, Hui Duan, and Wei Wang

Abstract—**The primary objective of this paper was to incorporate the reinforcement learning technique in variable speed limit (VSL) control strategies to reduce system travel time at freeway bottlenecks. A Q-learning (QL)-based VSL control strategy was proposed. The controller included two components: a QL-based offline agent and an online VSL controller. The VSL controller was trained to learn the optimal speed limits for various traffic states to achieve a long-term goal of system optimization. The control effects of the VSL were evaluated using a modified cell transmission model for a freeway recurrent bottleneck. A new parameter was introduced in the cell transmission model to account for the overspeed of drivers in unsaturated traffic conditions. Two scenarios that considered both stable and fluctuating traffic demands were evaluated. The effects of the proposed strategy were compared with those of the feedback-based VSL strategy. The results showed that the proposed QL-based VSL strategy outperformed the feedback-based VSL strategy. More specifically, the proposed VSL control strategy reduced the system travel time by 49.34% in the stable demand scenario and 21.84% in the fluctuating demand scenario.**

Index Terms—**Variable speed limit, reinforcement learning, congestion, freeway, bottleneck.**

I. Introduction

A FREEWAY section becomes a bottleneck when traffic demand exceeds capacity, resulting in capacity drop and increased system travel time [1]–[4]. Variable speed limit control has been introduced as an innovative approach to mitigate congestion and improve traffic operations at freeway bottlenecks. In past, numerous VSL control strategies have been proposed [5]–[27]. Most of the strategies can be classified into two categories: the online optimization approach [7]–[16] and the feedback control approach [17]–[27].

The online optimization approach determines the speed limit in the controlled area by solving optimal control algorithms. Several researchers have considered the freeway traffic control problem involving ramp metering and VSL as a constrained discrete-time optimal control problem that can be solved by an open-loop optimal control tool. The advantage of the online optimization approach is that the control strategy could, theoretically, achieve the optimum system performance. However, the open-loop optimal solution requires accurate models to predict the evolution of freeway traffic flow. Such models may not always be available. In addition, the optimal control algorithm requires large online computing workloads and, as a result, does not allow for large scale applications.

In the feedback-based VSL strategies, the controller automatically adjusts the speed limits to keep the controlled variable, i.e., bottleneck density, to be close to the desired density for flow maximization. Carlson *et al.* [18] proposed a local feedback based VSL control strategy which aimed at reducing delay at a merge bottleneck, and the feedback based controller was upgraded to consider multiple bottlenecks [25]. The strategy relied on readily available real-time measurements of traffic conditions and did not contain online models or demand predictions. As compared to online optimization approach, the feedback VSL strategy achieved similar control effects but was much more efficient and robust to actual traffic conditions.

The feedback-based control strategy also has several limitations. Without online traffic prediction modules, current feedback controllers are unable to take corrective actions until a deviation in the controlled variable has already occurred. More specifically, the feedback-based VSL controller takes actions shortly after the upstream traffic flow has perturbed the bottleneck traffic and generated an error between the measured and desired densities. As a result, there is usually a time lag between the deviation in traffic state and the corrective actions made by the feedback controller. If freeway traffic is largely varying over time, the controller may not be able to achieve the desired state in a timely pattern [28]–[30]. In recent years, the reinforcement learning (RL) approaches have attracted significant attention as a possible solution to address the limitations associated with the online optimization and the feedback controllers [31]–[37]. RL is a type of Machine Learning, and thereby also a branch of Artificial Intelligence. RL allows machines and agents to automatically determine the ideal behavior within a specific context to maximize its performance [38]–[40]. The RL learns directly from the interactions between states and actions through trial and error. The RL agent takes optimal actions under various states according to the long-term accumulation of rewards. As a result, a well-trained RL agent can, theoretically, make predictions on system evolution and achieve a proactive control scheme.

Manuscript received November 30, 2015; revised April 10, 2016, July 26, 2016, December 28, 2016, and March 2, 2017; accepted March 7, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 51508094 and Grant 51322810, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20150612. The Associate Editor for this paper was P. Kachroo. *(Corresponding author: Pan Liu.)*

Z. Li, P. Liu, C. Xu, and W. Wang are with the School of Transportation, Southeast University, Nanjing 210096, China (e-mail: lizhibin@seu.edu.cn; liupan@seu.edu.cn; iamxcc1@gmail.com; wangwei@seu.edu.cn).

H. Duan is with Jiaxing College, Jiangxing 314001, China (e-mail: huiduan@seu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2017.2687620

Several researchers have evaluated the potential of incorporating the RL algorithm with the ramp metering techniques [36], [41]. Previous studies have reported that the RL-based ramp metering algorithm used a more relaxed metering rate than the feedback-based algorithm and resulted in lower queue length. Recently, Zhu and Ukkusuri [42] developed a RL approach for dynamic speed limit control in a large roadway network. In their algorithm the speed limit in each link was allowed to be lower or higher than real-world speed limit to decide the optimal dynamic speed limit scheme. The optimal speed limit control reduced the total travel time in the network by 18%. However, their study focused on large-scale traffic networks and did not pay particular attentions to the specific freeway bottlenecks where the occurrence of capacity drop was the primary reason for the low efficiency of traffic flow. In addition, additional research is also needed to compare the performance of the RL-based VSL control with previous prevailing algorithms.

The primary objective of this study was to incorporate the RL in VSL control strategies to reduce traffic congestion at recurrent merge bottlenecks on freeways. A Q-learning (QL) based VSL strategy was proposed. The effects of the proposed QL-based VSL strategy were then compared to those achieved with the feedback-based VSL strategy.

II. Methodology

RL is inspired by behaviorist psychology considering how agents ought to take actions in an environment in order to maximize the cumulative reward. A RL agent interacts in discrete time steps with its environment which is typically formulated as a Markov decision process (MDP). Optimization of VSL control requires the determination of optimal speed limits. The action of an agent is to activate different speed limits at the decision interval. The transition time from one sate to another state after activating VSL control is unity. Each time the agent takes an action that affects the current state, the state changes. Thus, the VSL control problem can be formulated as a MDP problem and can be processed by RL technique [42].

A. Basic QL Algorithm

The QL is one of the most commonly used RL algorithms [38], [43]. The state set, action set, and reward function are determined for the QL agent. 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. Then the agent receives a reward to evaluate the quality of the transition. The mapping from the environmental state to the selection of action is known as a policy which defines the agent's behavior. By evaluating the rewards of multiple actions, the agent learns how to find a sequence of optimal actions that yields the maximum cumulative rewards over time. A Q-value is assigned to each state-action combination to evaluate the quality of the combination. The set of Q-values can be represented as:

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

where $\mathbf{S}$ is the set of possible states, $\mathbf{A}$ is the set of possible 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 that defines the relative importance of the current rewards and those earned earlier ($0 \leq \gamma \leq 1$). For a non-deterministic environment, the Q-value is updated with every new training sample according to

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

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 which controls how fast the Q-values are altered.

The QL algorithm will converge to the correct Q-values if each action is executed for each state for a plenty number of times and the learning rate is decreased appropriately over time. After the Q-values for various state-action pairs have been estimated during the learning process, the optimal action for a state is determined as the one with the largest Q-value. Then the QL agent can be used for the optimal control according to its knowledge. More details about the QL methodology can be found in Abdulhai *et al.* [31].

B. QL-Based VSL Strategy

A QL-based VSL control strategy was proposed. The flowchart of the proposed strategy is shown in Figure 1. The controller included two components: a QL-based offline agent and an online VSL controller. The central logic of the VSL control was to keep bottleneck traffic operating near its capacity state.

The VSL control within a merge bottleneck area is illustrated in Figure 2 (a). The VSL system included two sections: (1) a controlled section in which the outflow was controlled by adjusting the posted speed limits; and (2) an acceleration section that allowed vehicles to accelerate from low speed within the VSL-induced bottleneck to roughly critical speed as they reached the downstream bottleneck [11], [18].

The VSL control aimed at improving the outflow rate via preventing the capacity drop at active bottlenecks. The mainline traffic flow entering bottlenecks could be controlled by adjusting the speed limits posted in the upstream controlled sections. The most critical issue was to determine the optimal speed limit given the traffic flow states at the bottleneck area. If the speed limit was too low, the VSL control would transfer delay from the target bottleneck to the upstream area. If the speed limit was too high, the VSL control would not be able to fully prevent the capacity drop. In the proposed control strategy, the optimal speed limit for a particular traffic state was determined using the QL agent. The QL perceived a particular traffic state and selected a speed limit. The speed limit posted in the controlled section may lead to the transition

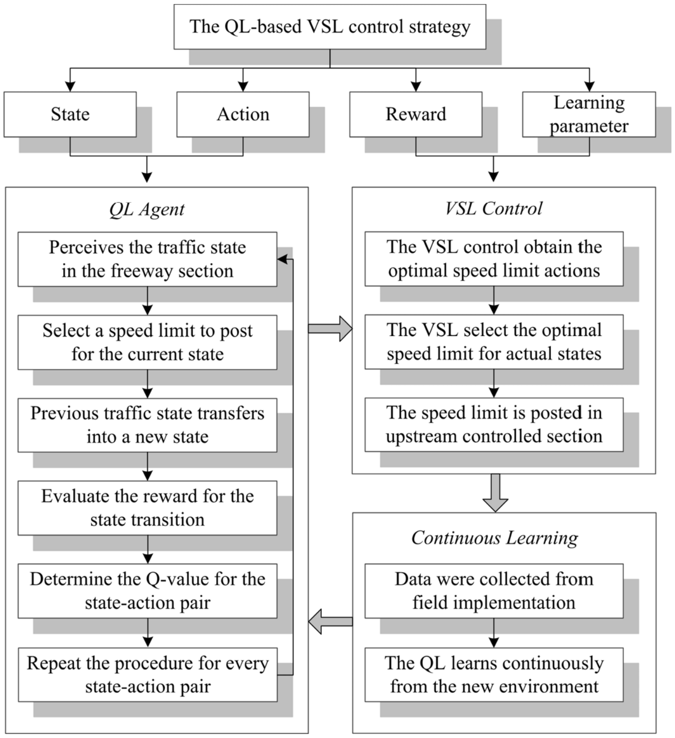

The QL-based VSL control strategy
State
Action
Reward
Learning parameter
QL Agent
Perceives the traffic state in the freeway section
Select a speed limit to post for the current state
Previous traffic state transfers into a new state
Evaluate the reward for the state transition
Determine the Q-value for the state-action pair
Repeat the procedure for every state-action pair
VSL Control
The VSL control obtain the optimal speed limit actions
The VSL select the optimal speed limit for actual states
The speed limit is posted in upstream controlled section
Continuous Learning
Data were collected from field implementation
The QL learns continuously from the new environment

Fig. 1. Flowchart of the QL-based VSL strategy.

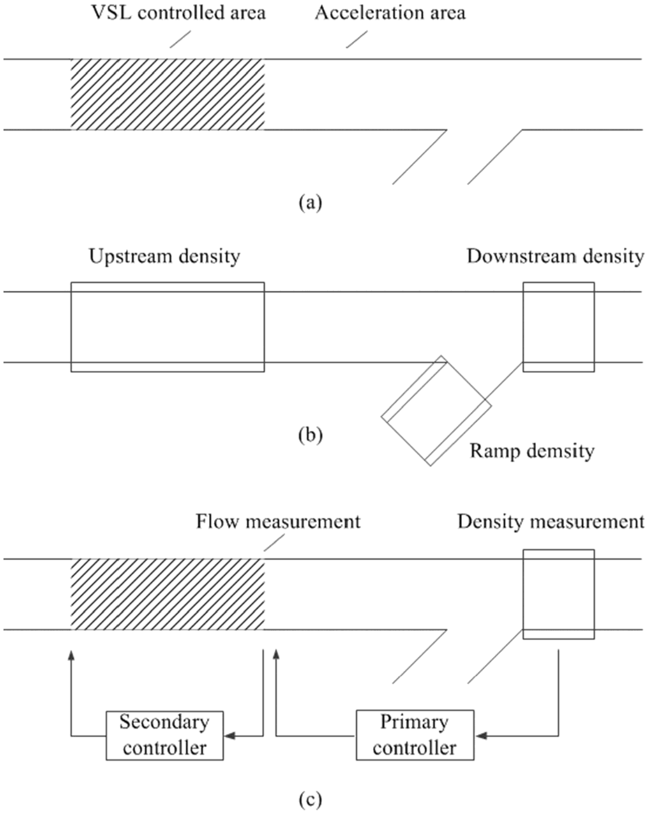


Fig. 2. (a) Freeway section with VSL control; (b) Density measurement in the QL-based VSL control; (c) Algorithm of the feedback-based VSL control.

in freeway traffic sate. The QL then evaluated the reward for the state transition and determined the Q-value for the state-action pair. The procedure was repeated for every state-action pair in the training dataset. The crucial elements in the QL agent included:

1) **State**. The QL agent used a state table as a collection of all the possible states. Traditionally, a state function is needed to describe the operation of traffic flow on freeways. However, due to the complexity of the dynamics of freeway traffic flow, it is quite difficult to obtain a state function that describes precisely how traffic flow may change from one state to another with the VSL control. In the QL agent, continuous state space could be represented by discrete states using a general function approximator to improve learning efficiency. A table with discrete states was used in the present study to depict the traffic flow conditions under the influence of VSL control. The selected state variables should be able to define the traffic flow state of the entire freeway section. Because the learning time increases exponentially as the number of state variables increases, only the most important traffic flow parameters should be considered. In this study, three traffic flow variables were used for defining the traffic state at a freeway merge bottleneck. They were: the density at the immediate downstream of the merge area, the density at the upstream mainline section, and the density on the ramp (see Figure 2 (b)). With the three variables the evolution of traffic operations at a freeway merge bottleneck could be monitored;
2) **Action**. The VSL controlled the mainline traffic flow by adjusting the speed limit posted in the controlled area. Thus, the speed limit was considered the action in the QL-based VSL control strategy. In practice, only discrete values of speed limits are allowed to be posted on VSL signs. In this study, the speed limits that were integer multiples of 5 mph were considered in the action set. To avoid introducing disturbances to traffic flow, the speed limits were set to gradually change over time according to a specific rate;
3) **Reward**. The objective of the QL-based VSL strategy was to reduce the system travel time. Consider a freeway system that consisted of several origins I and destinations I', and a discrete time representation of traffic variables with time index k and time interval η. Assuming that the system received arriving flow $q_i(k)$ at origin i at time k, the total arriving flow equaled $q(k) = q_1(k) + q_2(k) + \ldots + q_I(k)$; and the total exit flow at the destinations equaled $s(k) = s_1(k) + s_2(k) + \ldots + s_{I'}(k)$. The total travel time (TTT) over a time horizon K can be calculated by [40]:

$$TTT = \eta \sum_{k=1}^{K} N(k) \tag{4}$$

where $N(k)$ is the total number of vehicles in the network at time k. $N(k)$ can be calculated as:

$$\begin{aligned} N(k) &= N(k-1) + \eta[q(k-1) - s(k-1)] \\ &= N(0) + \eta \sum_{\kappa=0}^{k-1} [q(\kappa) - s(\kappa)] \end{aligned} \tag{5}$$

where $N(0)$ was the number of vehicles in the network at the initial time of simulation, $q(\kappa)$ was the total arriving flow at time κ, and $s(\kappa)$ was the total exit flow at time κ. Substituting (5) in (4) the following equation is obtained:

$$TTT = \eta \sum_{k=1}^{K} \left[N(0) + \eta \sum_{\kappa=0}^{k-1} q(\kappa) - \eta \sum_{\kappa=0}^{k-1} s(\kappa) \right] \quad (6)$$

In the simulation, the number of vehicles entering the freeway system at each time step from both upstream mainline and on-ramp was recorded. The number of vehicles leaving the freeway system at downstream mainline and off-ramp was also recorded. During the simulation, the backward queue on either mainline or on-ramp did not spread to the origins, and so the entering flow was not affected by the queue on freeways. The *TTT* in our study was calculated according to Eq. (4) to (6). In this way, travel time of the entire traffic system which includes that of the dynamics of merging flows at ramps, the queue, and the backward flow propagation, was considered.

Assuming that the arriving flow q and its spatial and temporal distribution were independent of any control measures, according to Eq. (6) the minimization of the system travel time can be achieved by maximizing the time-weighted exit flows [1]–[3], [44]–[46], which was given by:

$$S = \eta^2 \sum_{k=1}^{K} \sum_{\kappa=0}^{k-1} s(\kappa) \quad (7)$$

Eq. (6) suggested that any control measures that managed to increase the early exit flows of the freeway section would lead to a decrease in the total travel time. For the merge bottleneck in Figure 2 (a), the reduction in total travel time was mainly determined by the bottleneck discharge flow with the VSL.

According to traffic flow theory, flow can be represented by density and each density corresponds to a unique flow [44]–[48]. Density has been considered as an indicator for traffic state pattern [49]–[51]. At the bottleneck area, maximum exit flow is reached when density is equal to its critical value. Exit flow decreases when density deviates from the critical value. Bottleneck density (or occupancy) has been the control objective in many ramp metering algorithms [36], [44], [45] and feedback VSL algorithms [18], [21], [22]. Thus, in our study the reward function in the RL could be determined according to the density measured at the immediate downstream of the bottleneck. The reward function used in our study is discussed in a subsequent section; and

1) **Learning Parameter**. The QL used a larger learning rate in the initial stage of the learning process. The learning rate typically decreased over time to ensure convergence. However, in the QL-based VSL algorithm, different states were explored at different stages of the learning process and thus typical methods for changing learning rate as a function of time were not readily available. In our study, the learning rate for each state-action pair was defined as functions of the number of visits to that pair. As the number of visits to each state-action pair $C(s, a)$ increased, the learning rate was decreased to suppress uncertainties and to converge to the optimal Q-values [36], [41]. The learning rate was estimated by

$$\kappa_{(s,\,a)} = \left[\frac{1}{1 + C(s, a)(1 - \gamma)} \right]^{0.7} \quad (8)$$

Another important consideration in the QL algorithm is to make a balance between exploitation and exploration when selecting actions. The QL should fully learn the information that has been presented in the Q-values. Using pure exploitation may greatly save the learning time, but it may also prohibit the discovery of new potential better actions and lead to local optimization. On the other hand, pure exploration outperforms pure exploitation in the capability of discovering new potential better actions. However, it may result in a random action selection without making use of the learning results and, accordingly, is quite time consuming. Previous studies have used algorithms such as ϵ-greedy and Softmax to balance the exploration and exploitation. One of the limitations of the ϵ-greedy method is that it treats all exploratory actions equally, irrespective of the estimated value of each action. To overcome this limitation, in our study the selection of action followed the Softmax action selection strategy, and the probability of selecting a particular action was determined by a function of the estimated values. Using the Boltzman distribution, the following probability function was obtained [32], [34]:

$$P_s(a) = \frac{e^{Q(s,\,a)/T}}{\sum_{b \in A} e^{Q(s,\,b)/T}} \quad (9)$$

where $P_s(a)$ was the probability of selecting action a at state s, and T was the temperature.

One of the advantages of the proposed strategy was that the QL agent would continuously optimize the speed limits after the proposed control strategy was implemented in the field. With the QL agent, the actual rewards for various state-action pairs were obtained from the field data. After a certain period of time, the new rewards for different state-action pairs would be updated and the QL agent would learn continuously the optimal speed limits in an offline pattern from the new environment. The frequency for updating optimal speed limits in the online controller depended on several factors, including the variability of freeway traffic flow, the time that was required to collect new data, and the computational time it took for training the control strategy. An iterative procedure was followed to learn continuously from the field applications of VSL control and to keep optimizing the control strategy. As such, the proposed RL-based VSL strategy actually combined the advantages of both online optimization and feedback control strategies.

C. Feedback-Based VSL Strategy

The local feedback-based VSL strategy developed by Carlson *et al.* [18] is briefly introduced in this section. The implementation of the feedback controller is shown in Figure 2 (c). The freeway segment was divided into two sections: (1) a VSL controlled section in which the outflow from the upstream section to the downstream bottleneck was controlled by dynamically adjusting the speed limits posted on VSL signs. A flow measurement was located immediately downstream of the VSL controlled area; and (2) an acceleration section which allowed vehicles to accelerate from low speed within the VSL-induced bottleneck to roughly critical speed as they reached the downstream bottleneck. The objective of the algorithm was to keep the traffic at bottleneck operating near its capacity by controlling the occupancy. When the current occupancy exceeded a pre-determined threshold, the VSL control reduced the speed limits to reduce the outflow of the upstream controlled area, and accordingly, to starve the inflow to the downstream bottleneck. The traffic parameters required by the VSL control and traffic dynamics were simulated using a modified cell transmission model (CTM).

A simple dead-beat feedback controller could be too slow, oscillating or unstable for the control case [18]. Thus, the proposed VSL controller used a less direct controller design which had a two-loop feedback cascade control structure (see Figure 2 (c)). The procedure for designing and turning in the cascade controllers started from the internal loop and moved to the external loop. The internal loop contained a secondary controller that regulated the outflow q_{VSL} by adjusting the VSL rate b that was measured by the speed limit divided by free flow speed. The outflow was fed back and compared with the desired flow given by the primary controller in the external loop. The primary controller compared the measured density at the bottleneck area with the setpoint density defined by the operator for throughput maximization.

The secondary controller was an integral I controller with the transfer function given by

$$b(k) = b(k-1) + K_I e_q(k) \tag{10}$$

where K_I was the controller parameter, and $e_q(k) = \hat{y}(k)\text{-}q_{\mathrm{VSL}}(k)$ was the flow control error.

The primary controller in the external loop compared the measured density d at the bottleneck area with the setpoint density χ, which was set to be equal to the critical density for throughput maximization. The primary controller updated its output (the desired flow in the secondary controller) to drive the downstream occupancy closer to the reference value. The primary loop controller was specified to be a proportional-integral (PI) controller which provided for a desired zero steady-state error, while keeping a satisfactory transient response and disturbance rejection. The PI-type controller was given as

$$\hat{y}(k) = \hat{y}(k-1) + (K'_P + K'_I)e_d(k) - K'_P e_d(k-1) \tag{11}$$

where K'_I and K'_P are the controller parameters, and $e_d(k) = \chi - d(k)$ is the density control error. Whenever the secondary controller furnishes a VSL rate $b(k)$ that exceeds one of its bounds $b(k) \in [b_{\min}, 1]$, the value of $b(k)$ must be truncated to the respective bound and used as $b(k-1)$ for the next control period to avoid the wind-up effect [17]. The same applies for $\hat{y}_{VSL}(k)$ with $\hat{y}_{VSL}(k) \in [q_{\min}, q_{\max}]$ with appropriately fixed bounds. For practical consideration, the calculated speed limit value is rounded to the nearest 5 mph during the application.

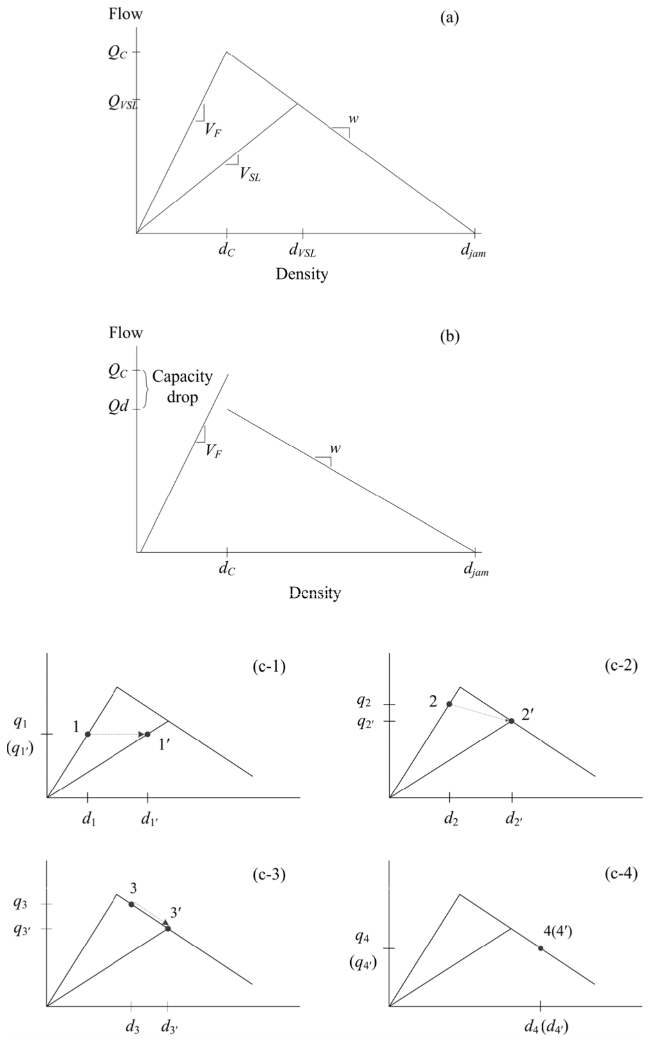


Fig. 3. (a) Fundamental diagram of traffic flow in CTM; (b) Fundamental diagram of traffic flow with capacity drop; and (c) change of traffic states with VSL control.

III. Simulation Network

A. Development of Simulation Model

In our study, the decision of VSL control strategy was made on the basis of aggregated traffic data from inductive loop detectors. For example, freeway loop detectors usually report traffic data in 30-s time intervals. Detailed information about individual vehicles was actually not available when developing the VSL control strategy. Thus, instead of using microscopic simulation models, a modified CTM was used for modeling the traffic flow at freeway bottleneck areas where VSL control was applied. By dividing the corridor into sub-sections, i.e., cells, the CTM predicted the macroscopic traffic flow characteristics by evaluating the flow and density at a finite number of intermediate points at different time steps [48]. The length of the cell was chosen such that it was equal to the distance traveled by free flow traffic in each evaluation time step, and was set to be fixed throughout the simulation. Traffic in each cell operated according to the fundamental diagram which was approximated by a triangular shape.

To more accurately reproduce the traffic flow affected by the VSL control, several modifications were made to the

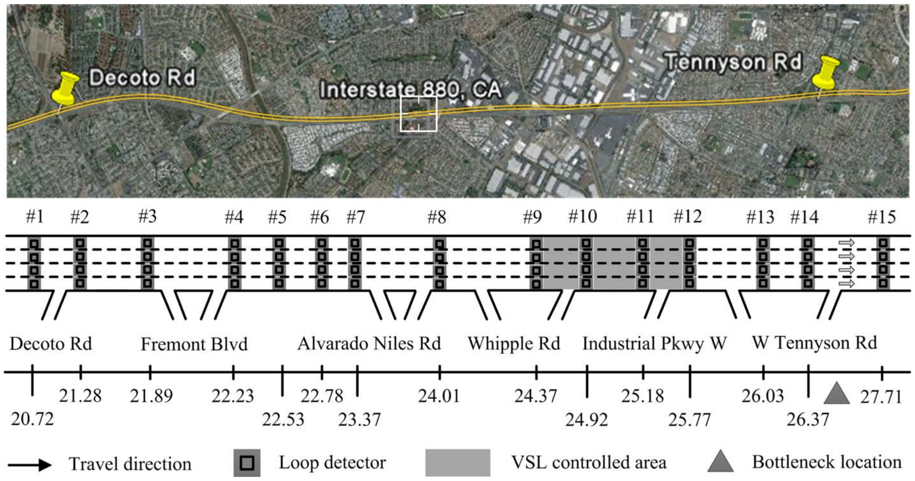


Fig. 4. Illustration of study freeway section.

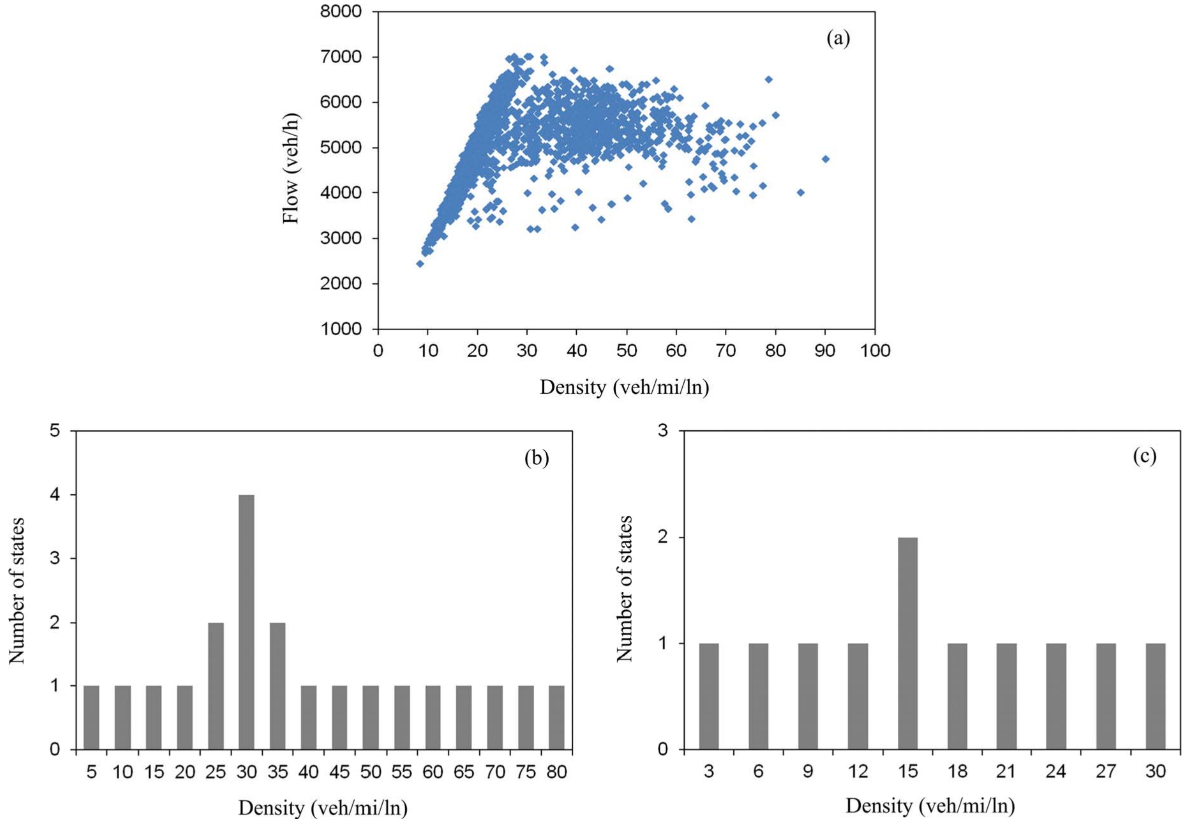


Fig. 5. (a) Actual flow and density data on the study site; (b) Selection of states on freeway mainlines in the QL; and (c) Selection of states on ramps in the QL.

traditional CTM [52]. Assuming that cell i was characterized by its triangular shaped fundamental diagram, the left limb of the triangle in Figure 3 (a) represented the sending function and the right limb represents the receiving function. The sending function represented the vehicles that could supply to the downstream cell $i+1$ with a flow rate of $\sigma_i(k)$, where k was the time step. The receiving function represented the available space in cell i which determined how many vehicles could enter cell i from the upstream cell $i-1$ with a flow rate of $\delta_i(k)$.

With the VSL control, the sending and receiving functions were determined by the minimum value between the speed limit V_{SL} and the free flow speed V_f:

$$\sigma_i(k) = \min\{\min\{V_{SL}(k), V_f\} \cdot d_i(k) \cdot n_i, Q_{VSL}\} \quad (12)$$

$$\delta_i(k) = \min\{w_i \cdot (d_{i,jam} - d_i(k)) \cdot n_i, Q_{VSL}\} \quad (13)$$

where $d_i(k)$ was the density at cell i at time k, n_i was the number of lanes, Q_{VSL} was the maximum flow under current speed limit, w_i was the kinematic wave speed, and $d_{i,jam}$ was the jam density.

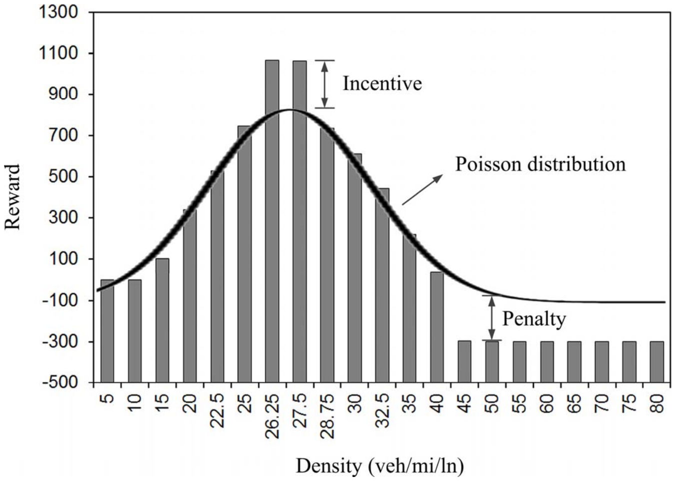


Fig. 6. Reward for different states in the QL.

In the real world drivers may not fully comply with the posted speed limits. A new parameter called overspeed was introduced here to take into account the situation that traffic speed is higher than the posted speed limit in unsaturated traffic conditions. The actual vehicle speed was given by

$$V'_{SL}(k) = \min(V_f, V_{SL}(k) + V_O) \tag{14}$$

where $V'_{SL}(k)$ was the actual traffic speed when the speed limit was $V_{SL}(k)$ at time k, V_f was the free flow speed, and V_O was the magnitude of overspeed. By replacing the speed limit $V_{SL}(k)$ in Eq. (12) and (13) with the actual speed $V'_{SL}(k)$, the impact of overspeed on traffic flow was reflected in the CTM.

The discharge flow drops below the bottleneck capacity after congestion forms [1]–[4]. To model the capacity drop, it was assumed that the bottleneck cell was characterized by an inverse λ-shaped fundamental diagram (see Figure 3 (b)). The flow was calculated by the left limb if capacity drop did not occur, and was calculated by the right limb after capacity drop occurred. Note that because capacity drop did not affect free flow speed, the size of bottleneck cell was not influenced by capacity drop, and remained constant during simulation. The sending function with capacity drop was determined by

$$\sigma_i(k) = \begin{cases} V_F \cdot d_i(k) \cdot n_i, & if\ d_i(k) \le d_C \\ Q_d, & if\ d_i(k) > d_C \end{cases} \tag{15}$$

where Q_C was the capacity of the bottleneck (veh/h), Q_d was the maximum discharge flow rate (veh/h) after capacity drop, and d_C was the critical density.

In the CTM, the flow rate was determined as the minimum value of the sending and receiving functions. The evolution of density and speed within each cell were calculated according to the flow rate. The use of VSL control resulted in changes in traffic states, as shown in Figure 3 (c). Assuming that traffic was in free flow state 1, if the flow rate q_1 was smaller than the Q_{VSL} with the speed limit V_{SL}, traffic state 1 would transfer to state 1′ with higher density. If the flow rate in the current state was greater than Q_{VSL}, such as the state 2 or 3, traffic state would transfer to state 2′ or 3′ with lower flow rate and higher density. For the heavily congested state 4 in which flow rate was smaller than Q_{VSL}, the VSL control would not change the traffic state (4=4′). To simulate the effect of acceleration section in the CTM, the maximum allowed speed within the acceleration section was set up to be the critical speed. In such way, the transition of traffic state from VSL controlled area to bottleneck area can be simulated.

The parameters in the CTM were calibrated using field data. Four traffic flow parameters were considered for the calibration of the fundamental diagram, including the free flow speed, the capacity flow, the discharge flow after capacity drop, and the speed of kinematic wave [48]. The capacity flow and discharge flow after capacity drop were calculated using the cumulative vehicle count curves at the bottleneck location [53]. The speed of kinematic wave was calculated by monitoring the traffic states at the detector stations located upstream of the active bottleneck [54].

B. Freeway Network

The research team selected a 6-mile long freeway section on the Interstate 880 freeway in Oakland, United States. The study site was located on the northbound freeway between mileposts 21 to 27. The freeway section was plagued by a merge recurrent bottleneck at its downstream end (see Figure 4). The bottleneck activated during both morning and afternoon peak periods on weekdays. Real-time traffic data were obtained from the Highway Performance Measurement System (PeMS). The PeMS database provided 30-s raw loop detector data, including vehicle count, vehicle speed, and detector occupancy. To reproduce the actual traffic flow features near the bottleneck area, the parameters in the CTM were calibrated using the field data collected from the study site. The free flow speed was found to be 65 mph. The capacity of the freeway mainline before capacity drop was found to be 1750 veh/h/ln. The magnitude of capacity drop was found to be 8.1%. The speed of the kinematic wave was estimated to be 9.2 mph. The simulation results were compared with the field data for model validation. The results showed that the calibrated CTM generated reasonable estimates for the traffic flow characteristics at the bottleneck. The average difference between the simulated and field measured flow rate was 12.6%, and the average difference between the simulated and field measured speed was 10.8%.

IV. Training of QL-Based VSL Strategy

The state table was determined using one-week traffic flow data measured at the study site. Note that longer study period could be used in practice to more accurately take into account possible variations in traffic states. The continuous flow-density space was then disaggregated to generate the states that would be included in the QL agent. The flow and density data from the freeway stretch were plotted in Figure 5 (a). Most of the observations had a density between 5 to 80 veh/mi. Because freeway traffic flow changed sensitively near the critical density, small intervals were used to divide traffic states near the critical point, while relatively larger intervals were used in free flow and congested flow conditions. As shown

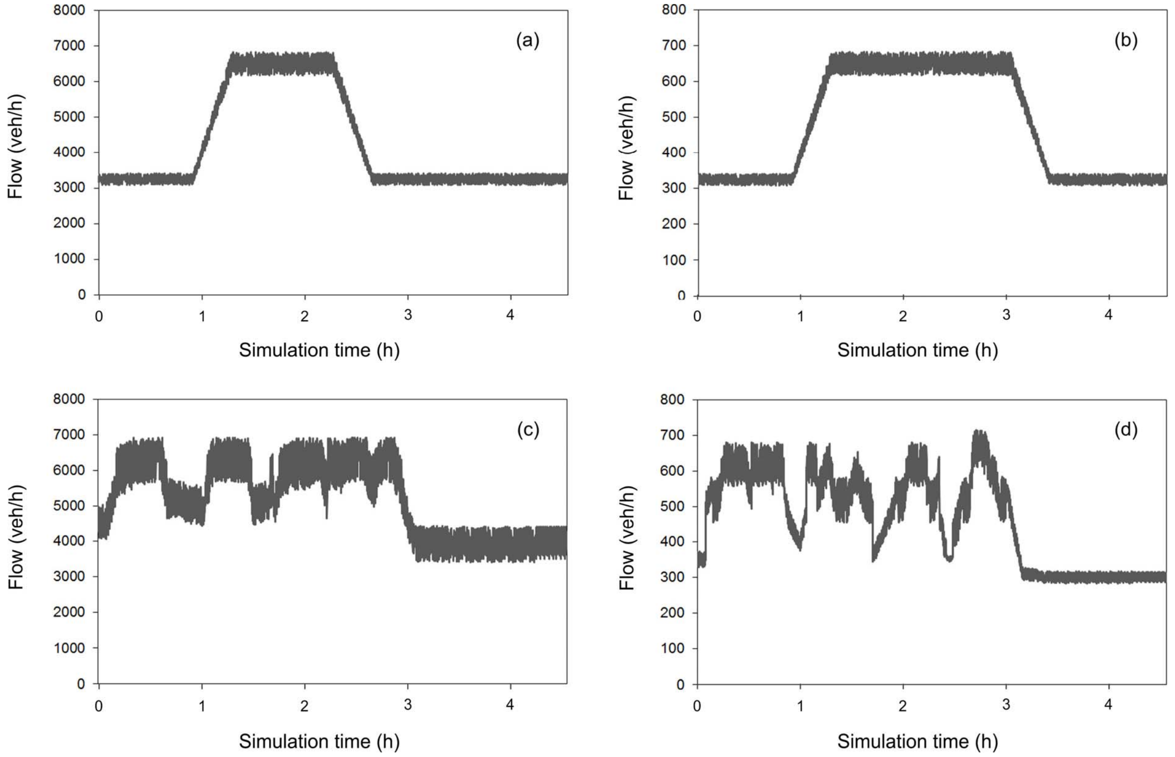


Fig. 7. (a) Traffic demand on mainline in the stable demand scenario; (b) traffic demand on W Tennyson Rd ramp in the stable demand scenario; (c) Traffic demand on mainline in the fluctuating demand scenario; and (d) Traffic demand on W Tennyson Rd ramp in the fluctuating demand scenario.

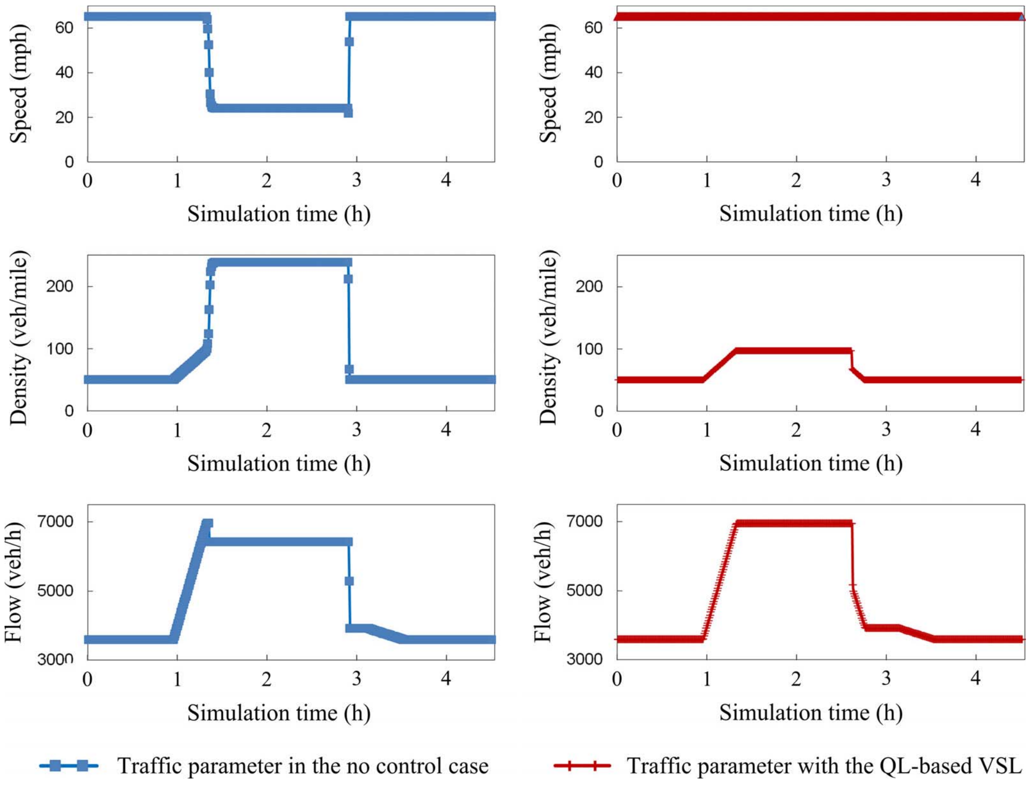


Fig. 8. Traffic operations at the bottleneck in the stable demand scenario.

in Figure 5 (b), the mainline traffic states, including the downstream and the upstream sections, were divided with an interval of 1.25 to 2.5 veh/mi near the critical density, and with an interval of 5 veh/mi in free flow and congested conditions. The continuous traffic state was approximated to the nearest center of discrete states. The same procedure was

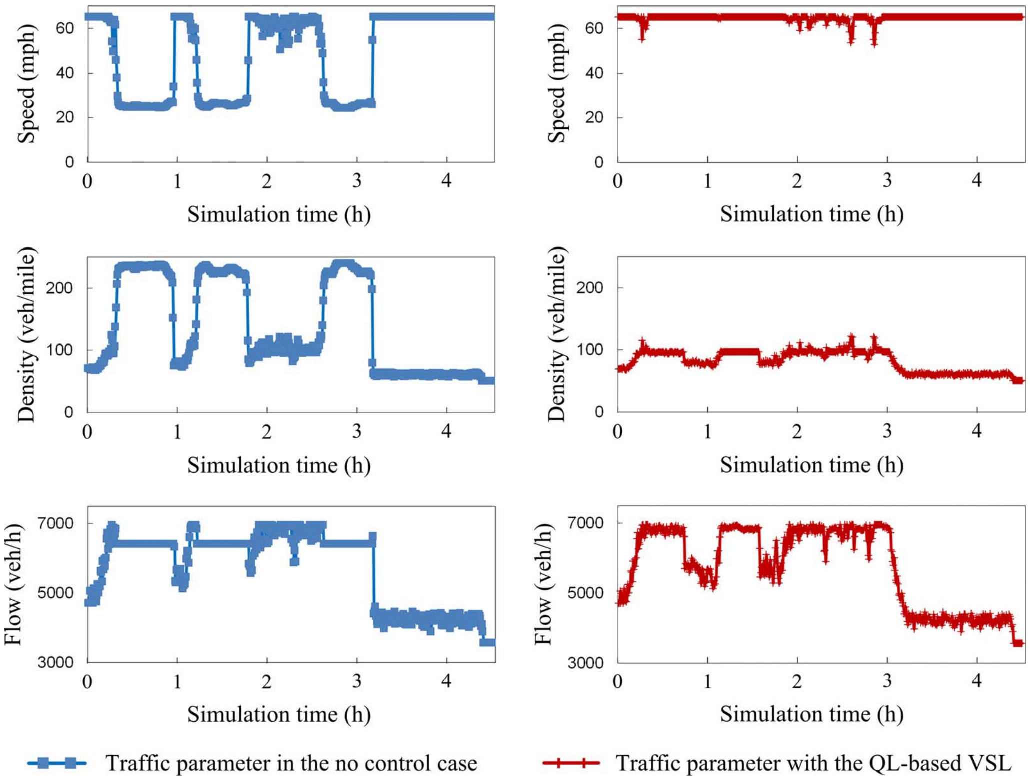


Fig. 9. Traffic operations at the bottleneck in the fluctuating demand scenario.

followed for the state selection for the ramp traffic flow. A total of 4851 traffic states were identified initially and included in the state table.

The proposed VSL control strategy adjusted speed limit from 20 to 65 mph with an increment of 5 mph. The action set was given by {20, 25, 30, 35, 40, 45, 50, 55, 60, 65}. The simulation period for each state-action pair was set to be five minutes to ensure the stability of the traffic state after an action was implemented. The effect of a specific action was evaluated by a reward that was defined on the basis of the downstream density after the section was taken. The largest reward was received if the downstream density equaled critical density. The reward decreased as the traffic state deviated from the capacity state. In this study, the reward function was determined according to the Poisson distribution function given by:

$$R(s) = \mu \cdot \Pr(X = s) = \mu \cdot \frac{\lambda^s e^{-\lambda}}{s!} \tag{16}$$

where $R(s)$ was the reward for density state s, μ was the parameter which determined the magnitude of reward, $\Pr(X = s)$ was the probability function for the state s, and λ was the Poisson parameter.

The parameter μ was set to be 1×10^4. The parameter λ was set to be equal to the critical density (26.9 veh/mi). To increase the convergence speed of the QL algorithm, an extra incentive of 200 was added to the rewards of the two states near critical density. A penalty of 400 was added to the rewards for severely congested traffic states. The incentive and penalty values were selected from multiple tests. The final rewards considered in the QL algorithm are shown in Figure 6.

The following parameters in the RL agent should be carefully determined: the learning rate κ, the discount factor γ, and the temperature T. As shown in section II, the learning rate was determined by both the number of the visits to the pair and the discount factor. Note that a discount factor was used here to compensate for the decrease in value of rewards over time. In general, a smaller discount factor indicated that the agent would put more importance to the current rewards. Thus, with the increase in the value of γ, the agent would account for future rewards more strongly. The closer the value of γ was to 1, the longer time the system would take to reach convergence. Following the suggestions provided by previous studies, the discount factor was set to be 0.8 in the present study [37], [41], [55], [56].

For the temperature parameter, when T was large, each action would have approximately the same probability of being selected (more exploration). When T was small, actions would be selected proportionally to their estimated value (more exploitation). In the present study, the QL was forced to explore each action until all actions had been tried at least once for every state. Then the Softmax method was followed to choose better actions. The temperature was defined separately for each state as a function of the number of visits to that state. This dependence would ensure that the Softmax policy would become greedy policy after all state-action pairs in a state had been visited enough [41]. At the initial stage a high temperature ($T_0 = 1000$) was set to make all the actions to be selected with equal chances. With the increase of iterations, a decreased temperature (to $T \approx 1$) was set to generate large difference in selection probability for the actions with different Q-values. The actions with higher estimated

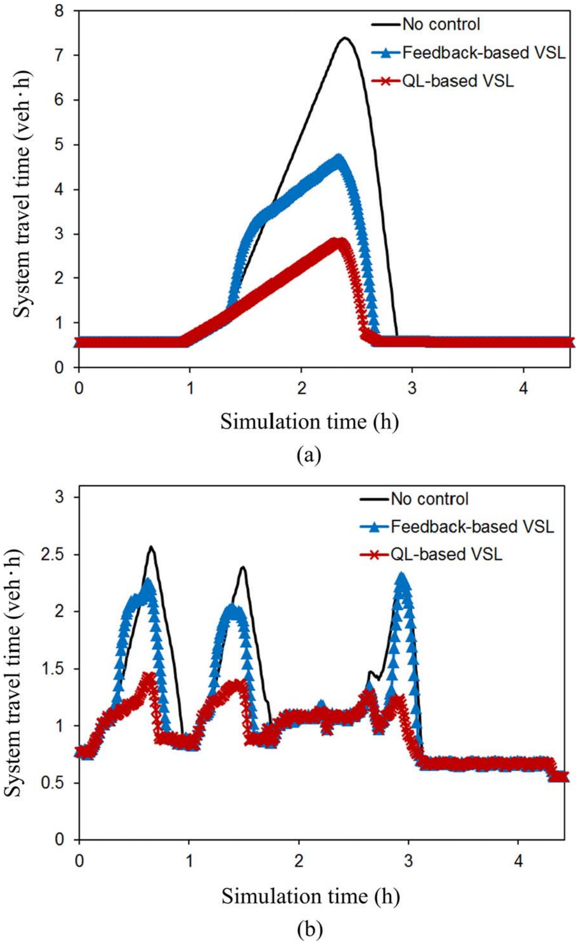


Fig. 10. (a) System travel in the no control case and with different VSL strategies in the stable demand scenario; and (b) System travel in the no control case and with different VSL strategies in the fluctuating demand scenario.

rewards would be chosen with higher priorities. The learning process continued until the Q-values for all the state-action pairs reached convergence. Several control periods from 30 s to 5 min were tested. Results suggested that the QL-based VSL with a control period of 30 s had the best control effects. In each cycle, the algorithm received the system states and selected a speed limit action. Then the algorithm received a reward for the selected action and updated the Q-value for that state-action pair. The off-line learning time of the QL-based agent for the 4851 states was about 30 min.

V. Evaluation of Proposed Control Strategy

The effects of the QL-based VSL control strategy were evaluated using the modified CTM. The duration of simulation was set to be 4 hours with a 30-min warm-up period. Two traffic demand scenarios were considered, including a stable demand scenario and a fluctuating demand scenario. As shown in Figure 7, traffic demand increased after one hour and the peak period lasted for two hours in the stable demand scenario. In the fluctuating demand scenario, the traffic demand on both mainline and ramp was subject to strong variations over time. The control period was initially set to be 30 s to identify the maximum achievable control effects.

The changes in traffic flow features at the freeway bottleneck with the influence of VSL control are depicted in Figure 8 and 9. For comparison we also considered the situation in which none of the control strategies was used.

In the stable demand scenario, traffic congestion formed after 1.5 h, resulting in decreased speed and increased density at the bottleneck area. The capacity drop occurred shortly after the formation of congestion, resulting in increased system travel time. With the VSL control, the capacity drop was prevented and the congestion duration was reduced. In the fluctuating demand scenario, traffic congestion formed repeatedly within the section, resulting in large variation in traffic flow variables and decreased bottleneck discharge flow rate (see Figure 9). With the VSL control the bottleneck congestion was nearly eliminated and the bottleneck flow was increased.

The system travel time with and without the VSL control are compared in Figure 10. In the stable demand scenario, the total travel time was reduced from 1046 to 530 veh·h with the proposed QL-based VSL control strategy, indicating a 49.34% reduction in the system travel time. In the fluctuating demand scenario, the proposed VSL control strategy reduced the system total system travel time by 21.84%. The results suggested that the proposed QL-based VSL strategy significantly reduced traffic congestion at freeway recurrent bottlenecks.

The proposed QL-based VSL strategy was compared to the feedback-based VSL strategy. The parameters in the feedback-based VSL algorithm were calibrated to optimize the control effects [18], [57]. In the stable demand scenario, the feedback-based VSL strategy reduced the total travel time by 24.96%. In the fluctuating demand scenario, the feedback-based based VSL strategy reduced the total travel time by 11.02%. The results suggested that the QL-based VSL control strategy outperformed the feedback control strategy in reducing system travel time at freeway recurrent bottlenecks.

The research team further compared the speed limits that were determined using these two strategies (see Figure 11). Two findings were obtained: (1) when traffic congestion was going to form at the bottleneck, the QL-based VSL reduced the upstream speed limit quickly to 35 mph to prevent capacity drop at the downstream bottleneck. While the feedback-based VSL control reduced speed limit after the congestion has formed and capacity drop has occurred. The speed limit was further reduced to 15 mph to eliminate the bottleneck congestion[1]; (2) during peak periods the speed limit with the QL-based VSL was quite stable while the speed limit under the feedback-based VSL oscillated over time, posting disturbances to mainline traffic.

The differences in the control effects can be attributed to the fact that the feedback-based VSL control adjusts the speed limit after traffic congestion has formed and generates a system error, while the QL-based VSL control adjusts the speed limits prior to the formation of traffic congestion. As mentioned before, the trained QL agent had the capability of predicting

[1] Note that the overshooting effect usually can be reduced by adjusting the parameters in the feedback-based algorithm. In this study multiple parameters were tested but the overshooting cannot be fully prevented. The reason would be that the due to the hysteretic nature, the feedback algorithm cannot prevent the capacity drop that occurs quickly after the form of congestion.

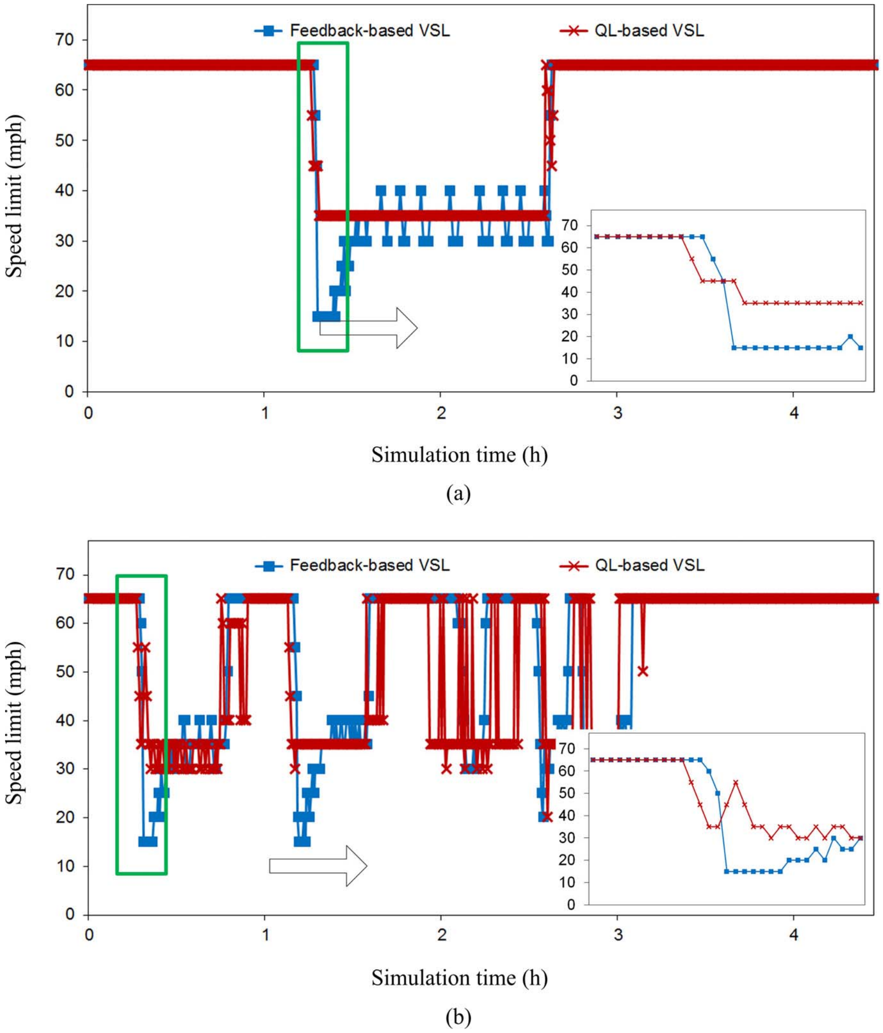


Fig. 11. (a) Comparison of speed limits with different VSL strategies in the stable demand scenario; and (b) Comparison of speed limits with different VSL strategies in the fluctuating demand scenario.

state transitions. As shown in Figure 11, the QL-based VSL controller started to reduce speed limits 60 s earlier than the feedback controller did. In addition, the feedback algorithm was found to be sensitive to the deviation in downstream density. Because only discrete speed limits were posted, the VSL controller may not control the downstream density exactly at its critical value. The accumulation of small control errors would finally cause a change in speed limit, resulting in the oscillation of speed limit. While in the QL-based VSL controller, the speed limit did not change until the deviation transferred downstream density from one discrete state to another.

Then constrains for speed change rate considering practical applications were added to test for the control effects that could be achieved in more realistic conditions. The maximum change rate of speed limit was set to be 10 mph per 1 min. It was found that the inclusion of the maximum speed change rate slightly reduced the control effects. More specifically, the QL-based and the feedback-based VSL reduced the total travel time by 43.25% and 16.24% in the stable demand scenario, and by 14.46% and 6.91% in the fluctuating demand scenario. The QL-based VSL control strategy outperformed the feedback control strategy in reducing system travel time.

VI. Continuous Learning

Traditionally, the online optimization based VSL strategy does not contain a feedback component. The proposed QL-based VSL strategy combined the advantages of both the online optimization and the feedback-based control strategies. An offline continuous learning agent was used to keep optimizing the actions after the VSL control was implemented. The actual rewards for various traffic state-action pairs can be obtained from field applications. The new rewards can be updated in the state table after a certain time period. Then the QL agent can learn the new optimal actions with the updated state table. As such, the proposed VSL control strategy will be able to accommodate the uncertainties in practical applications. Theoretically, the longer the QL-based algorithm is used, the better the control effects the system will be able to achieve. As a result, the QL-based VSL strategy may potentially have high transferability and adaptability for practical applications. One of the critical concerns with VSL control is related to the way drivers respond to the posted speed limits. Such issue

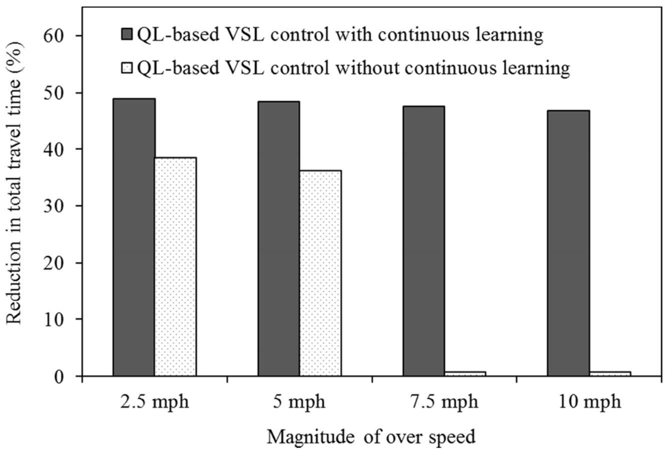


Fig. 12. Effects of QL-based VSL strategies with continuous learning.

can easily be addressed by the continuous learning function of the proposed QL-based strategy. An example is given here to help illustrate how continuous learning helps to address the overspeed issue with VSL control. The magnitude of overspeed in Eq. (14) was set to be 2.5 to 10 mph with an increment of 2.5 mph. The VSL controller started using the optimal actions that did not consider overspeed cases. The actual rewards associated with various state-action pairs were collected for 10 hours in the simulation environment and then updated in the state table. The QL agent learned the optimal actions with the new state table, and then updated the optimal actions in the online VSL controller. The procedure repeated several times until the optimal actions did not change. The simulation results illustrated in Figure 12 showed that without continuous learning, the VSL control had no effects when the magnitude of overspeed was larger than 5 mph, while with the continuous learning, the VSL control effectively reduced the system travel time given large magnitude of overspeed.

VII. Conclusions and Discussion

A QL-based VSL control strategy that aimed at reducing system travel time at freeway recurrent bottlenecks was proposed. A modified CTM was used to model the traffic flow under the influence of VSL control. The simulation results suggested that the proposed strategy effectively reduced traffic congestion at freeway recurrent bottlenecks. The total travel time was reduced by 49.34% and 21.84% in the stable and the fluctuating demand scenarios. The RL agent was trained in an off-line scheme. After the optimal control strategy is obtained, the RL-based VSL does not contain heavy online computing workload which enables it for real-time traffic control. In addition, through the continuous learning function, the RL-based VSL can learn from the differences between model predictions and observation in real world, and update the optimal control strategy. Thus, the RL-based VSL can be considered more robust to uncertainties in traffic flow. The well-trained RL-based VSL has the capability of predicting traffic state transitions and acts in a proactive control scheme. In our simulation tests, improved performances in reducing system travel time were observed as compared to the feedback-based VSL control. But RL method also has some limitations. For example, the RL agent requires sufficient training to learn the optimal control strategies before it can be applied. If the agent is not well trained, the control effects could be greatly reduced. Also the RL agent performs like a black box which may reduce its transferability to other traffic scenarios or road segments. The feedback-based VSL control is simple and straightforward for local practitioners to use, and is expected to have good transferability to different scenarios.

The proposed QL-based VSL may have the potential to be implemented in practical engineering applications to improve traffic operations on freeways. To implement the proposed strategy, a VSL controlled area needs to be designed upstream of the bottleneck and loop detectors should be placed according to Figure 2. The RL-based VSL algorithm needs to be trained in the simulation testing environment to obtain the optimal control strategies. With the continuous learning function, the proposed VSL strategy will be able to accommodate the uncertainties in the field applications and lead to the optimal control effects. This study showed that through continuous learning, the control effects of VSL were improved under large overspeed situations.

In the real world, an accurate estimation of traffic state with loop detector data becomes more challenging. Thus, the RL-based VSL control could be in conjunction with some advanced traffic state estimation methods [58]–[64] to strengthen the offline modeling component of the overall process. Note that the performance of feedback controllers could also be further enhanced using online models or traffic prediction models. With such models the feedback controllers will be able to take actions based on the predicted traffic states. In this way, the feedback controller could achieve a proactive control scheme. However, online traffic prediction models are still underdevelopment and no sophisticated models that can be incorporated into the VSL control system are currently available. Thus, the development of feedback controller with online traffic prediction models could be considered by future studies. It would be interesting to compare the effects of such feedback control method with the RL based VSL control.

In the present study, a small freeway network with a merge bottleneck was used to test the effect of the proposed VSL control. Other types of bottlenecks caused by lane reduction, traffic incident or work zone also need to be investigated. A larger traffic network with multiple links or closely spaced bottlenecks could also be considered for validating the performance of the RL-based VSL control. Besides, this study used bottleneck density control as the reward function in the RL method, and successful reduction of travel time was observed. Future study could consider other reward measurements and objective functions for system optimization. In addition, the coordination of multiple reinforcement learning agents with VSL and ramp metering control algorithms can be evaluated to improve the overall performance in reducing traffic congestions. Furthermore, some advanced machine learning methodologies, such as deep learning [65], [66], could be applied when developing the VSL control strategies. Authors recommend that future studies could focus on those issues.

REFERENCES

[1] M. J. Cassidy and J. Rudjanakanoknad, "Increasing the capacity of an isolated merge by metering its on-ramp," *Transp. Res. B, Methodol.*, vol. 39, no. 10, pp. 896–913, 2005.

[2] K. Chung, J. Rudjanakanoknad, and M. J. Cassidy, "Relation between traffic density and capacity drop at three freeway bottlenecks," *Transp. Res. B, Methodol.*, vol. 41, no. 1, pp. 82–95, 2007.

[3] L. Zhang and D. Levinson, "Ramp metering and freeway bottleneck capacity," *Transp. Res. A, Policy Practice*, vol. 44, no. 4, pp. 218–235, 2011.

[4] S. Oh and H. Yeo, "Microscopic analysis on the causal factors of capacity drop in highway merging sections," in *Proc. 91th Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2012, pp. 1–23.

[5] X. Y. Lu, S. E. Shladover, I. Jawad, R. Jagannathan, and T. Phillips, "Novel algorithm for variable speed limits and advisories for a freeway corridor with multiple bottlenecks," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2489, pp. 86–96, Dec. 2015.

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

[7] H. Liu, L. Zhang, D. Sun, and D. Wang, "Optimize the settings of variable speed limit system to improve the performance of freeway traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3249–3257, Jun. 2015.

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

[9] E. Kwon, D. Brannan, E. Kwon, K. Shouman, C. Isacason, and B. Arseneau, "Development and field evaluation of variable advisory speed limit system for work zone," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2015, pp. 12–18, Dec. 2007.

[10] R. C. Carlson, I. P. M. Papamichail, and A. Messmer, "Optimal mainstream traffic flow control of large-scale motorway networks," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 2, pp. 193–212, 2010.

[11] R. C. Carlson, I. P. M. Papamichail, and A. Messmer, "Optimal motorway traffic flow control involving variable speed limits and ramp metering," *Transp. Sci.*, vol. 44, no. 22, pp. 238–253, 2010.

[12] X. Y. Lu, P. Varaiya, R. Horowitz, D. Su, and S. E. Shladover, "Novel freeway traffic control with variable speed limit and coordinated ramp metering," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2229, pp. 55–65, Sep. 2011.

[13] M. Hadiuzzaman and T. Z. Qiu, "Cell transmission model based variable speed limit control for freeways," *Can. J. Civil Eng.*, vol. 40, no. 1, pp. 46–56, 2013.

[14] X. Yang, Y. Lu, and G. L. Chang, "Proactive optimal variable speed limit control for recurrently congested freeway bottlenecks," in *Proc. 92rd Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2013, pp. 1–29.

[15] M. Hadiuzzaman, J. Fang, C. Lan, and T. Z. Qiu, "Impact of mainline demand levels and control parameters on multiobjective optimization involving proactive optimal variable speed limit control," in *Proc. 93rd Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2014, pp. 1–19.

[16] C. Pasquale, D. Anghinolfi, S. Sacone, S. Siri, and M. Papageorgiou, "A comparative analysis of solution algorithms for nonlinear freeway traffic control problems," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Jun. 2016, pp. 1773–1778.

[17] A. Popov, A. Hegyi, R. Babuška, and H. Werner, "Distributed controller design approach to dynamic speed limit control against shockwaves on freeways," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2086, pp. 93–99, Dec. 2008.

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

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

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

[21] R. C. Carlson, I. Papamichail, and M. Papageorgiou, "Integrated feedback ramp metering and mainstream traffic flow control on motorways using variable speed limits," *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 209–221, Sep. 2014.

[22] 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, Apr. 2013.

[23] J. Lee, J. D. Daniel, G. B. Joe, and B. Park, "Simulation-based evaluations of real-time variable speed limit for freeway recurring traffic congestion," in *Proc. 92rd Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2013, pp. 1–19.

[24] Z. Li, P. Liu, W. Wang, and C. Xu, "Development of control strategy of variable speed limits for improving traffic operations at freeway bottlenecks," *J. Central South Univ.*, vol. 21, no. 6, pp. 2526–2538, 2014.

[25] G. R. Iordanidou, C. Roncoli, and I. P. M. Papamichail, "Feedback-based mainstream traffic flow control for multiple bottlenecks on motorways," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 610–621, Feb. 2015.

[26] Y. Zhang and P. A. Ioannou, "Combined variable speed limit and lane change control for highway traffic," *IEEE Trans. Intell. Transp. Syst.* [Online]. Available: http://ieeexplore.ieee.org/document/7728060/

[27] G. R. Iordanidou, I. R. C. Papamichail, and M. Papageorgiou, "Feedback-based integrated motorway traffic flow control with delay balancing," *IEEE Trans. Intell. Transp. Syst* [Online]. Available: http://ieeexplore.ieee.org/document/7811261/

[28] K. J. Åström and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[29] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, 6th ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2009.

[30] W. S. Levine, *The Control Handbook*, 2th ed. New York, NY, USA: CRC, 2010.

[31] 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.

[32] 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.

[33] C. Jacob and B. Abdulhai, "Machine learning for multi jurisdictional optimal traffic corridor control," *Transp. Res. A, Policy Practice*, vol. 44, no. 22, pp. 53–64. 2010.

[34] 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, Dec. 2011.

[35] M. Davarynejad, A. Hegyi, J. Vrancken, and J. van den Berg, "Motorway ramp-metering control with queuing consideration using Q-learning," in *Proc. 14th IEEE Int. Conf. Intell. Transp. Syst.*, Washington, DC, USA, Oct. 2011, pp. 1652–1658.

[36] K. Rezaee, B. Abdulhai, and H. Abdelgawad, "Self-learning adaptive ramp metering: Analysis of design parameters on a test case in Toronto," in *Proc. 92rd Annu. Meeting Transp. Res. Board*, Washington, DC, USA, 2013, pp. 10–18.

[37] S. El-Tantawy and B. Abdulhai, "Comparative analysis for temporal difference learning methods in adaptive traffic signal control," in *Proc. 12th World Conf. Transp. Res.*, Lisbon, Portugal, 2010, pp. 1–21.

[38] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.

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

[40] A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2003.

[41] K. Rezaee, B. Abdulhai, and H. Abdelgawad, "Application of reinforcement learning with continuous state space to ramp metering in real-world conditions," in *Proc. 15th IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2012, pp. 1590–1595.

[42] 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, Apr. 2014.

[43] S. Mahadevan, "Average reward reinforcement learning: Foundations, algorithms, and empirical results," *Mach. Learn.*, vol. 22, nos. 1–3, pp. 159–195, 1996.

[44] M. Papageorgiou and A. Kotsialos, "Freeway ramp metering: An overview," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 271–281, Apr. 2002.

[45] M. J. Cassidy, "Freeway on-ramp metering, delay savings, and diverge bottleneck," *Transp. Res. Rec.*, vol. 1856, pp. 1–5, Jan. 2003.

[46] C. F. Daganzo, *Fundamentals of Transportation and Traffic Operations*. Oxford, U.K.: Pergamon, 1997.

[47] P. Kachroo and S. Sastry, "Traffic assignment using a density-based travel-time function for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 5, pp. 1438–1447, May 2016.

[48] 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.

[49] 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, Jun. 2014.

[50] H. B. Celikoglu, "An approach to dynamic classification of traffic flow patterns," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 28, no. 4, pp. 273–288, 2013.

[51] H. B. Celikoglu and M. A. Silgu, "Extension of traffic flow pattern dynamic classification by a macroscopic model using multivariate clustering," *Transp. Sci.*, vol. 50, no. 3, pp. 966–981, 2016.

[52] 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, Apr. 2014.

[53] 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, Jan. 2006.

[54] M. Mauch and M. J. Cassidy, "Freeway traffic oscillations: Observations and predictions," in *Proc. 15th Int. Symp. Transp. Traffic Theory*, 2002, pp. 653–674.

[55] A. Gozavi, *Simulation-Based Optimization: Paramitric Optimization Techniques and Reinforcement Learning*. Norwell, MA, USA: Kluwer, 2003, ch. 9.

[56] T. S. Mostafa and H. Talaat, "Intelligent geographical information system for vehicle routing (IGIS-VR): A modeling framework," in *Proc. 13th IEEE Int. Annu. Conf. Intell. Transp. Syst.*, Madeira Island, Portugal, Sep. 2010, pp. 801–805.

[57] A. Kouvelas, K. Aboudolas, E. Kosmatopoulos, and M. Papageorgiou, "Adaptive performance optimization for large-scale traffic control systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1434–1445, Apr. 2008.

[58] M. Treiber and D. Helbing, "Reconstructing the Spatio-temporal traffic dynamics from stationary detector data," *Cooperat. Transp. Dyn.*, vol. 1, no. 3, pp. 3-1–3-24, 2002.

[59] Y. Wang and M. Papageorgiou, "Real-time freeway traffic pattern estimation based on extended Kalman filter: A general approach," *Transp. Res. B, Methodol.*, vol. 39, no. 2, pp. 141–167, 2005.

[60] Y. Wang, M. Papageorgiou, and A. Messmer, "Real-time freeway traffic state estimation based on extended Kalman filter: A case study," *Transp. Sci.*, vol. 41, no. 2, pp. 167–181, 2007.

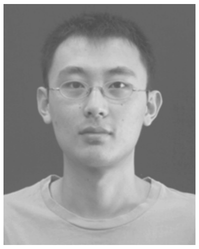
[61] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transp. Res. B, Methodol.*, vol. 39, no. 2, pp. 141–167, 2005.

[62] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[63] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PLoS One* vol. 10, no. 3, p. e0119044, 2015.

Zhibin Li received the Ph.D. degree from the School of Transportation, Southeast University, Nanjing, China, in 2014. From 2010 to 2011, he was a Visiting Scholar with University of California at Berkeley, Berkeley. He was with University of Washington as a Research Associate from 2015 to 2016. He is currently a researcher with Southeast University. He received the China National Scholarship in 2012 and 2013 and the Best Doctoral Dissertation Award from the China Intelligent Transportation Systems Association in 2015.

Pan Liu received the Ph.D. degree in civil engineering from University of South Florida, Tampa, USA, in 2006. He is a Professor with the School of Transportation, Southeast University, Nanjing, China. His research interests include traffic operations and safety, and intelligent transportation systems. He was a recipient of the Distinguished Young Scientist Foundation of NSFC in 2013.



Chengcheng Xu received the Ph.D. degree from the School of Transportation, Southeast University, Nanjing, China, in 2014. He is an Assistant Professor with the Key Laboratory of Traffic Planning and Management, Southeast University. He received the Best Doctoral Dissertation Award from the China Intelligent Transportation Systems Association in 2016.

Hui Duan received the B.S. degree from China University of Mining and Technology in 2011 and the M.S. degree from the School of Transportation, Southeast University, Nanjing, China, in 2014. She is an Assistant Professor with Jiaxing College, Jiangxing, China. Her research interests include traffic safety, traffic control, and intelligent transportation systems.

Wei Wang received the M.S. and Ph.D. degrees in civil engineering from Southeast University, Nanjing, China, in 1985 and 1989, respectively. He is a Professor with the School of Transportation, Southeast University. His research interests include urban transportation planning and intelligent transportation systems. He is a member of the Model Traffic Technology Panel of the National High-Tech Research and Development Program of China (863 Program). He received the National Distinguished Teacher Award of China in 2007.