Transportation Research Part C 41 (2014) 30–47

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach

Feng Zhu, Satish V. Ukkusuri *

School of Civil Engineering, Purdue University, West Lafayette, IN 47907-2051, USA

ARTICLE INFO

Article history:
Received 19 November 2013
Received in revised form 24 January 2014
Accepted 24 January 2014

Keywords:
Dynamic speed limit control
Stochastic network
Connected vehicle
Reinforcement learning
Network loading

ABSTRACT

This paper proposes a novel dynamic speed limit control model accounting for uncertain traffic demand and supply in a stochastic traffic network. First, a link based dynamic network loading model is developed to simulate the traffic flow propagation allowing the change of speed limits. Shockwave propagation is well defined and captured by checking the difference between the queue forming end and the dissipation end. Second, the dynamic speed limit problem is formulated as a Markov Decision Process (MDP) problem and solved by a real time control mechanism. The speed limit controller is modeled as an intelligent agent interacting with the stochastic network environment stochastic network environment to assign time dependent link based speed limits. Based on different metrics, e.g. total network throughput, delay time, vehicular emissions are optimized in the modeling framework, the optimal speed limit scheme is obtained by applying the R-Markov Average Reward Technique (R-MART) based reinforcement learning algorithm. A case study of the Sioux Falls network is constructed to test the performance of the model. Results show that the total travel time and emissions (in terms of CO) are reduced by around 18% and 20% compared with the base case of non-speed limit control.

1. Introduction

Speed limits are implemented in the traffic infrastructure system primarily based on the safety consideration. There are many studies which examine the safety benefits of imposing speed limits. To name a few, Kloeden et al. (2001) and Ossiander and Cummings (2002) showed that increasing the speed limit could lead to higher fatal crash rate and more deaths. Lee et al. (2006) proposed a real time crash prediction log-linear model to evaluate the relationship between crash potential (the likelihood of crash) and speed limit. It is demonstrated that 5–17% reduction of crash potential can be achieved by imposing real time variable speed limit control. Nowadays the environmental benefits of speed limits are also noticeable due to the growing traffic pollutions and the awareness of the impact of greenhouse gases, especially in urban areas. (Keller et al., 2008) showed that a 4% reduction of NO_x emissions was obtained when speed limit on Swiss motorways decreased from 120 kph to 80 kph. Madireddy et al. (2011) found that CO_2 and NO_x emissions could be reduced by 25% if speed limits decrease from 50 kph to 30 kph in the study area of Belgium.

1.1. Literature review and motivations

In recent years, the mobility benefits of speed limit control began to attract researchers' interest. Deterioration of traffic condition in urban areas has long been a burden to urban economic development and people's quality of life. From the

* Corresponding author. Address: CIVL G167D, 550 Stadium Mall Drive West Lafayette, IN 47906, USA. Tel.: +1 (765) 494 2296; fax: +1 (765) 496 7996.
E-mail addresses: zhu214@purdue.edu (F. Zhu), sukkusur@purdue.edu (S.V. Ukkusuri).

http://dx.doi.org/10.1016/j.trc.2014.01.014
0968-090X/

demand side, many traffic control management strategies have been proposed to relieve congestion problems. As one of many traffic control strategies, speed limit control is an important strategy to alleviate congestion and emission in traffic networks.

Typically, depending on temporal dimension, the speed limit control problem can be classified to two categories: static and dynamic speed limit control. In the static case, the speed limit control problem is usually considered as part of the network design problem. Yang et al. (2012) revisited the static user equilibrium (UE) problem with speed limits, and investigated the impact of speed limits on total travel time and vehicular emission at a network level. The numerical studies show reduction in both travel time and vehicular emission with an appropriate speed limit design. Based on the notion of user equilibrium, Wang (2013) developed a bi-level programming model to design optimal speed limits considering both network efficiency and equity. No decisive conclusion can be reached on whether the total travel time reduces under the UE principle with speed limits. On the other hand, dynamic speed limit control is implemented in practice especially in highway work zones, known as variable speed limit (VSL) control (NRC, 1998). Empirical studies have shown the effectiveness of VSL in smoothing traffic flow and reducing traffic breakdowns (van den Hoogen, 1994; Raemae, 2002). Kang et al. (2004) developed an on-line algorithm for dynamic VSL control in highway work zone operations. Hegyi et al. (2005) integrated VSL control and ramp metering and solved the coordination control problem by applying the model predictive control (MPC) approach. The model can also be extended to consider main-stream metering. Significant travel time reduction (15%) compared with non-control case is obtained in the numerical case study. Followed the line, Carlson et al. (2010) formulated the integrated VSL control and ramp metering problem as a constrained discrete-time optimal control problem. A universal open-loop optimal control tool (Papageorgiou and Kotsialos, 2002; Kotsialos et al., 2002) by computing suitable feasible directions is applied to solve the problem.

It is worth to note that one limitation of the above studies of speed limits (both static and dynamic) is that the stochastic property of network either from the demand side or supply side is not considered. Another limitation is that most of the above studies utilize macroscopic model to simulate traffic flow propagation. Hence they are not able to capture the more realistic and important characteristics of traffic, e.g. shockwave propagation, spill-back effect of heavy congestions. One motivation of this paper is to fill these research gaps in the literature.

Another motivation of this study is to address the dynamic speed limit control problem under the connected vehicle (CV) environment. CV environment is a recently developed concept owing to the development of wireless communication technology, especially the Dedicated Short Range Communications (DSRC) technology. DSRC has great potential in the area of intelligent transportation system (ITS), as it enables the wireless exchange of information between vehicles, as well as between vehicles and roadside infrastructure. The CV technology is primarily developed to improve the safety of traffic (crash collision avoidance) at intersections. The secondary concern is to alleviate congestion and reduce vehicular emission. Acknowledging the potential, the intelligent transportation system program of the U.S. Department of Transportation (DOT) emphasizes CV research in the ITS Strategic Plan (2010–2014). CV environment facilitates communication platform where vehicles can talk to adjacent vehicles (i.e., Vehicle-to-Vehicle, V2V), to the infrastructure components (i.e., Vehicle-to-Infrastructure, V2I), and also infrastructure to infrastructure communication (i.e., I2I). CV has also received attention in Europe, where it is known as Car to Car (C2C) and Car to X (C2X) technology. Though CV has not been implemented in the real world transportation system yet, many auto companies are expending significant efforts to produce vehicles with communication features. In addition, many test beds are ongoing in US, Europe, and Japan. Recent advances in CV environment offer useful technologies in detection and acquisition of high fidelity data that can be used for more efficient traffic control strategies. In particular, under the CV environment, the speed limit controller has access to the traversing information of the surrounding vehicle, e.g. origin, destination, path taken, speed, distance traveled, etc. Based on the information, the controller produces the time-varied speed limit scheme and sends the speed limit back to the vehicle. The vehicle adjusts its free flow speed accordingly. The interacting process between the controller and the vehicle is going on continuously, hence the realization of dynamic speed limit control.

1.2. Contributions of the paper

Triggered by the above motivations, this paper sets out to propose a novel dynamic speed limit control model based on reinforcement learning approach. In the model, it is required that the traffic flow information of the link is known to the speed limit controller. This setting is technologically possible under the CV environment. Under the CV environment, vehicles are able to communicate with infrastructure, thus the controller is able to inform speed limit to the vehicles. In current practice, it is also plausible to implement the dynamic speed limit by imposing variable speed limit signs and deploying roadside sensors along the links of interest.

The contributions of the paper are mainly twofold. (1) We have developed a link-based dynamic network loading (LDNL) model with the consideration of speed limits under the stochastic network environment. The demand input is generated randomly according to a certain probability distribution. Similarly, the saturation capacities of the specified links are also randomly generated to account for the uncertain supply (e.g., highway crash, lane closure, work zone, etc.). In the proposed LDNL model, every link in the given network is divided into three parts, namely, the beginning part, the main part, and the ending part. The beginning part and the ending part are of fixed size depending on the preset resolution, while the main part allows flexible size depending on the residual length of the link. Flow propagation to and from the main part is strictly analyzed applying the kinematic wave theory. The shockwave propagation is well defined and captured by checking the

difference between the queue forming end and dissipation end of the main part. By making use of the fundamental diagram in the LDNL model, we are able to obtain the link speed profile which can then be used to estimate the vehicular emissions. (2) The dynamic speed limit problem is formulated as a Markov Decision Process (MDP) problem and solved by a real time control mechanism. Different metrics, e.g. total network throughput, delay time, vehicular emissions, can be easily set as the optimized objective in the modeling framework. The speed limit controller is modeled as an intelligent agent interacting with the stochastic network environment by taking actions, which is to determine time-dependent speed limits. Thus, the dynamic speed limit control problem is transferred to finding the optimal policy (mapping between the speed limits and traffic states) that gives the maximum reward measured in terms of total travel time, number of stops, vehicular emission, etc. in the long term. The optimal speed limit scheme is obtained by applying the R-Markov Average Reward Technique (RMART) based reinforcement learning algorithm.

The rest of the paper is structured as follows. Section 2 is devoted to a novel link based dynamic network loading model, including the formulation of the model and link speed profile estimation. Section 3 describes the MDP reformulation of the speed limit assignment problem, and the reinforcement learning approach to solve the problem, including the definition of state, action, reward, and details of the RMART algorithm. Section 4 conducts a numerical case study on the Sioux Falls network. Discussions on the performance of the algorithm and insights for practice are provided. Section 5 concludes the paper and discusses the direction for future research.

Notation:

Sets	
C	Set of all the links
C_R	Set of origin links
C_S	Set of destination links
C_O	Set of ordinary links
C_D	Set of diverging links
C_M	Set of merging links
C_{SL}	Set of signalized links (with signalized intersection at the end)
$\Gamma^{-1}(i)$	Set of predecessors of link i
$\Gamma(i)$	Set of successors of link i
Parameters:	
S	Saturation flow rate
d_J	Jam density
N_E	Holding capacity of the beginning or ending part of the link
N_M^i	Holding capacity of the main part of link i
$Q_{B,E}^{i,t}$	Inflow (B) or outflow (E) capacity of link i at time t
δ_i	Ratio of the shockwave speed over the free flow speed at link i, within $[0,1]$
ρ_E^i	Exogenous ratio for the link i (merging or diverging link)
$L_M^{i,t}$	Length of the main part of link i at time t
$L_0^{i,t}$	Length of the beginning or ending part of link i at time t
Variables:	
$x_{B,M,E}^{i,t}$	Occupancy (refers to the number of vehicles) of the beginning, main, or ending part of link i at time t
$y_B M^{i,t}$	Flow from beginning part to main part of link i at time t
$y_{ME}^{i,t}$	Flow from main part to ending part of link i at time t
$y_{EB}^{i,j,t}$	Flow from ending part of link i to the beginning part of downstream link j at time t
$d^{i,t}$	Demand input for origin i at time t
q_i^t	Traffic flow of link i at time t
k_i^t	Density of link i at time t
v_i^t	Speed of link i at time t
V_i^t	Free flow speed (speed limit) of link i at time t
W_i^t	Shockwave speed of link i at time t

2. Link based dynamic network loading model

The cell transmission model (CTM) proposed by Daganzo (1994, 1995) is one of the widely used network loading models. It provides a convergent approximation to a simplified version of the LWR hydrodynamic model (Lighthill and Whitham, 1955; Richards, 1956), whereby the fundamental diagram of traffic flow and density is assumed to be a piecewise linear

function. Thus the model is capable of capturing the traffic propagation phenomena such as spill back, kinematic wave, and physical queue in the network loading process. CTM has been used for various dynamic problems in the last decade. Among the wealth of literature, Lo and Szeto (2002) embedded CTM into the user equilibrium DTA problem using the variational inequality (VI) approach. Szeto and Lo (2004) extended the VI formulation to capture simultaneous route and departure time choice problem with elastic demands. Han et al. (2011) and Ukkusuri et al. (2012) formulated the cell-based dynamic user equilibrium problem using complementarity theory. Besides network analysis, CTM has also been applied in the area of traffic control management (Gomes et al., 2008; Lo, 1999; Lo et al., 2001; Ukkusuri et al., 2010).

One limitation of CTM lays in its incapability of modeling dynamic speed limit. In CTM, a series of homogenous cells are used to represent the road network, and time is discretized into steps. The length of each cell is set to be the distance traveled by the free-flow speed in one time step. Therefore, for the same time step, faster free-flow speed will lead to longer cell length, and vice versa. When there is no congestion, vehicles move from one cell to the next one in one time step. However, in the case of dynamic speed limit control, the free flow speed (assuming equals to the imposed speed limit) is changeable at a time-dependent manner, rendering a different fundamental diagram (flow-density relationship). As in Fig. 1, it shows three types of fundamental diagram under three types of free flow speed (i.e., V_1^t, V_2^t, V_3^t).

In light of the limitation from CTM, we develop a dynamic network loading model that explicitly takes speed limit into consideration. The model is presented in more details in the sections below.

2.1. Link representation for a generalized network

Consider an ordinary link in a generalized network with homogeneous capacity (if the link is with heterogeneous capacities, it can be represented as a combination of several sub-links with homogeneous capacities). Hypothetically, we divide the link into three components, namely, the beginning part, the main part, and the ending part, as shown in Fig. 2. B and E represent the beginning part and the ending part of the ordinary link. These two parts are of identical size. Typically, the beginning part and the ending part are treated as the basic elements of the network representation. Similar to the cell size determination in CTM, the size of the beginning part or the ending part is determined by the multiplication of free-flow speed and time step. On the other hand, M represents the intermediate component of the road segment. The size of M is dependent on the residual length of the link (i.e. the length of the link minus the length of B and E).

As ordinary links are the simplest components in a network, a complex network configuration can be considered as a mixture of ordinary links. To be more specific, a diverging link is treated as an ordinary link (from upstream end) connected with multiple ordinary links (from downstream end), while a merging link is treated as multiple ordinary links (from upstream end) connected with one ordinary link (from downstream end). The link representation for diverging and merging links are shown in Fig. 3. Notice that the connection components of different links in the network are always the basic parts (beginning or the ending part) of links. For the sake of demonstration purpose, we use the shape of a circle to represent the main part of a link. The flow propagation process for all the three parts are covered in the next section.

2.2. Traffic flow propagation in the main part of a link

In order to implement the updating process of flow and occupancies in the link based DNL model, we approximate spatial queue length as proportional to the number of stopped vehicles (represented by occupancies) in the queue. The queue formation end is determined by the accumulative number of vehicles entering the link from the time when the queue begins to form; and the queue dissipation end is determined as the accumulative number of vehicles leaving the link from the time when the queue begins to dissipate. As demonstrated in Fig. 4, point A represents the beginning time of a red phase (queue

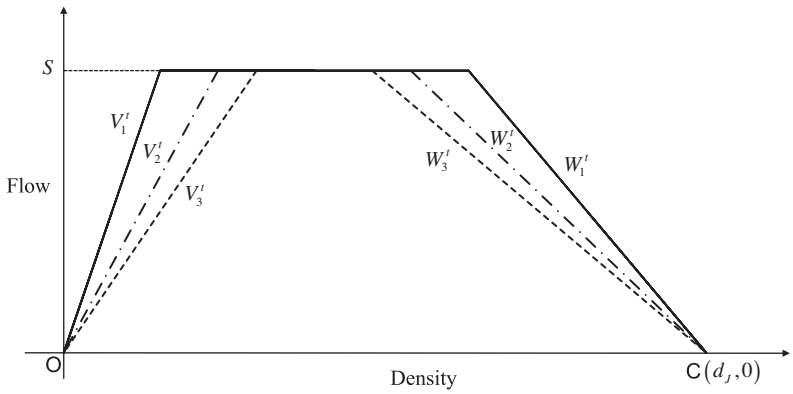

Fig. 1. Effect of speed limits in the fundamental diagram.

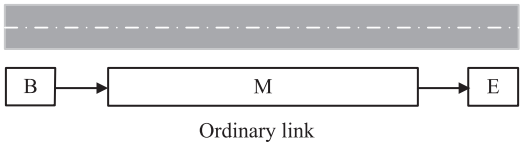

Fig. 2. Link representation of an ordinary link.

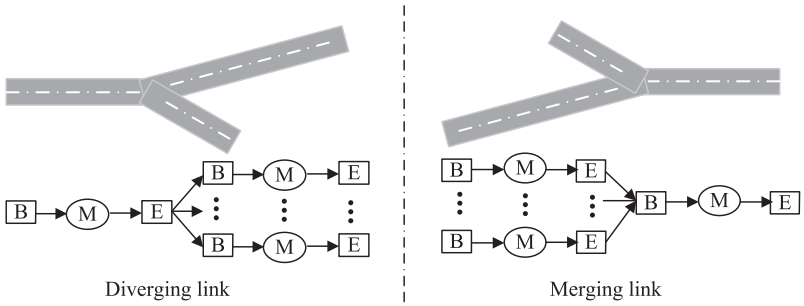

Fig. 3. Link representation of a diverging link and a merging link.

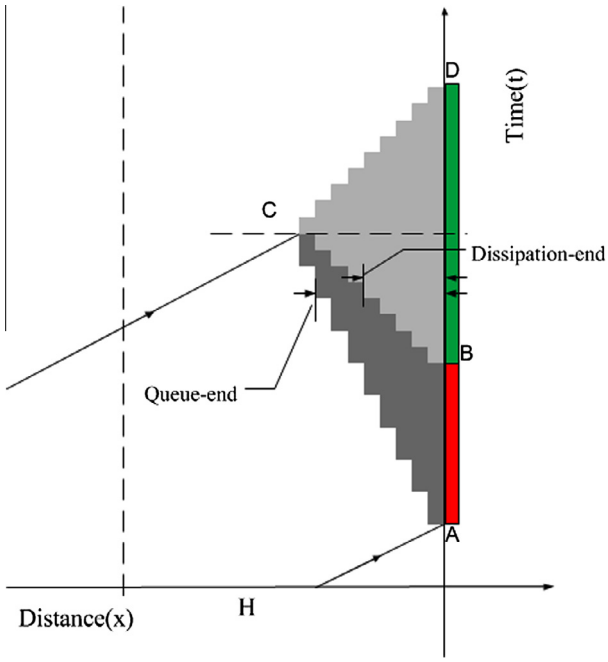

Fig. 4. Demonstration of queue forming end and dissipation end.

begins to form), point B represents the ending time of the red phase (queue begins to dissipate), point C represents the maximum queue length, and point D represents the ending time of green phase (queue stops to dissipate).

Queue forming end determination:

$$x_Q^{i,t} = \sum_{t'}^{t} y_{BM}^{i,t} \tag{1}$$

where $x_Q^{i,t}$ denotes the forming end of the queuing traffic, and t' denotes the beginning time of the queue formation.

Queue dissipation end determination:

$$x_{Q'}^{i,t} = \sum_{t''}^{t} y_{ME}^{i,t} \tag{2}$$

where $x_{Q'}^{i,t}$ denotes the dissipation end of the queuing traffic, and t'' denotes the beginning time of the queue dissipation.

Queue length is determined as the difference between the queue formation end and dissipation end. Thus the queuing state or the free-flow state is inferred by the sign of $L^{i,t} = x_Q^{i,t} - x_{Q'}^{i,t}$.

If $L^{i,t} > 0$, it indicates the link is in queuing state. thus we have:

$$y_{ME}^{i,t} = \min\left\{Q_E^{i,t}, \delta_{ME}(N_E - x_E^{i,t})\right\} \tag{3}$$

According to shockwave theory, the backward shockwave of the ending part reaches the beginning part of the link after $\frac{L_M^{i,t}}{W_i^t}$ time steps, hence we get:

$$y_{BM}^{i,t} = \min\left\{\begin{matrix} {}_{y_{ME}}^{i,t-\frac{L_M^{i,t}}{W_i^t}}, x_B^{i,t}, Q_B^{i,t}, \delta_{BM}(N_M^i - x_M^{i,t}) \end{matrix}\right\} \tag{4}$$

If $L^{i,t} \leqslant 0$, it indicates the link is in free-flow state. According to shockwave theory, the forward shockwave of the beginning part reaches the ending part of the link after $\frac{L_M^{i,t}}{V_i^t}$ time steps, hence we have:

$$y_{ME}^{i,t} = \min\left\{\begin{matrix} {}_{y_{BM}}^{i,t-\frac{L_M^{i,t}}{V_i^t}}, Q_E^{i,t}, \delta_{ME}(N_E - x_E^{i,t}) \end{matrix}\right\} \tag{5}$$

$$y_{BM}^{i,t} = \min\left\{x_B^{i,t}, Q_B^{i,t}, \delta_{BM}(N_M^i - x_M^{i,t})\right\} \tag{6}$$

To sum up, we have:

$$y_{ME}^{i,t} = \begin{cases} \min\{Q_E^{i,t}, \delta_{ME}(N_E - x_E^{i,t})\}, & \text{if } L^{i,t} > 0 \\ \min\left\{\begin{matrix} {}_{y_{BM}}^{i,t-\frac{L_M^{i,t}}{V_i^t}}, Q_E^{i,t}, \delta_{ME}(N_E - x_E^{i,t}) \end{matrix}\right\}, & \text{else} \end{cases} \tag{7}$$

$$y_{BM}^{i,t} = \begin{cases} \min\left\{\begin{matrix} {}_{y_{ME}}^{i,t-\frac{L_M^{i,t}}{W_i^t}}, x_B^{i,t}, Q_B^{i,t}, \delta_{BM}(N_M^i - x_M^{i,t}) \end{matrix}\right\}, & \text{if } L^{i,t} > 0 \\ \min\{x_B^{i,t}, Q_B^{i,t}, \delta_{BM}(N_M^i - x_M^{i,t})\}, & \text{else} \end{cases} \tag{8}$$

2.3. The whole formulation of the link based DNL model

Occupancy conservation recursive equations:

For the beginning part:

$$x_B^{i,t+1} = \begin{cases} x_B^{i,t+1} = x_B^{i,t} + d^{i,t} - y_{BM}^{i,t}, & \forall i \in C_R \\ x_B^{i,t} + \sum_{i' \in \Gamma^{-1}(i)} y_{EB}^{i',i,t} - y_{BM}^{i,t}, & \forall i \in C/C_R \end{cases} \tag{9}$$

In order to capture the uncertainty from the traffic demand side, we have:

$$d^{i,t} = p_i(t) D_i(t) \tag{10}$$

where $D_i(t)$ is a fixed value, and $p_i(t)$ denotes a random value within $(0,1)$ generated by a certain probability distribution. Based on empirical data, the typical distribution assumptions for demand input include multivariate normal distribution, lognormal distribution, and multivariate lognormal distribution (Zhao and Kockelman, 2002; Krishnamurthy and Kockelman, 2003; Siu and Lo, 2008; Ng et al., 2010).

For the main part:

$$x_M^{i,t+1} = x_M^{i,t} + y_{BM}^{i,t} - y_{ME}^{i,t} \tag{11}$$

For the ending part:

$$x_E^{i,t+1} = x_E^{i,t} + y_{ME}^{i,t} - \sum_{j \in \Gamma(i)} y_{EB}^{i,j,t} \tag{12}$$

Accounting for the change of speed limits:

Note that the size of the basic part (the beginning part or the ending part) is determined by the multiplication of the free flow speed and time step. As the speed limit changes, the size of the basic part changes accordingly. Thus we need to adjust the occupancies of the beginning and ending part proportionally to the change of size (which is proportionally to the change of speed limits). To maintain the link occupancy conservation, the occupancy of the main part needs to be adjusted as well. If the speed limit of link i is changed for the time $t+1$, before applying (9), (11), and (12), for the beginning part and ending part, we have:

$$x_B^{i,t+1} = x_B^{i,t}\frac{V_i^{t+1}}{V_i^t}, \quad x_E^{i,t+1} = x_E^{i,t}\frac{V_i^{t+1}}{V_i^t} \tag{13}$$

For the main part, we have:

$$x_M^{i,t+1} = x_B^{i,t} + x_M^{i,t} + x_E^{i,t} - x_B^{i,t+1} - x_E^{i,t+1} \tag{14}$$

Flow propagation recursive equations:
For connectors pertaining beginning part:

$$y_{BM}^{i,t} = \begin{cases} \min\left\{ y_{ME}^{i,t-\frac{L^{i,t}}{W_i^t}}, x_B^{i,t}, Q_B^{i,t}, \delta_{BM}(N_M^i - x_M^{i,t}) \right\}, & \text{if } L^{i,t} > 0 \\ \min\left\{ x_B^{i,t}, Q_B^{i,t}, \delta_{BM}(N_M^i - x_M^{i,t}) \right\}, & \text{else} \end{cases} \tag{15}$$

For connectors pertaining main part:

$$y_{ME}^{i,t} = \begin{cases} \min\left\{ Q_M, Q_E^{i,t}, \delta_{ME}(N_E - x_E^{i,t}) \right\}, & \text{if } L^{i,t} > 0 \\ \min\left\{ y_{BM}^{i,t-\frac{L^{i,t}}{V_i^t}}, Q_E^{i,t}, \delta_{ME}(N_E - x_E^{i,t}) \right\}, & \text{else} \end{cases} \tag{16}$$

For connectors pertaining ending part of ordinary links:

$$y_{EB}^{i,j,t} = \min\{x_E^{i,t}, Q_E^{i,t}, Q_B^{j,t}, \delta_{EB}(N_B - x_B^{j,t})\}, \quad j \in \Gamma(i), \quad \forall i \in C_O \tag{17}$$

For connectors pertaining ending part of signalized links:

$$y_{EB}^{i,j,t} = \min\{x_E^{i,t}, Q_{signal}^{i,t}, Q_B^{j,t}, \delta_{EB}(N_B - x_B^{j,t})\}, \quad j \in \Gamma(i), \quad \forall i \in C_{SL} \tag{18}$$

Where $Q_{signal}^{i,t} = \begin{cases} Q_E^{i,t}, \text{ if phase is activated} \\ 0, \text{else} \end{cases}$

For connectors pertaining ending part of merging links:

$$y_{EB}^{i,j,t} = \min\left\{ x_E^{i,t}, Q_E^{i,t}, Q_B^{j,t}, \rho_E^i \delta_{EB}(N_B - x_B^{j,t}) \right\}, \quad j \in \Gamma(i), \quad \forall i \in C_M \tag{19}$$

For connectors pertaining ending part of diverging links:

$$y_{EB}^{i,j,t} = \min\{\rho_E^i x_E^{i,t}, Q_E^{i,t}, Q_B^{j,t}, \delta_{EB}(N_B - x_B^{j,t})\}, \quad j \in \Gamma(i), \quad \forall i \in C_D \tag{20}$$

In order to capture the uncertainty from the infrastructure supply side, we have:

$$Q_E^{i,t} = \begin{cases} p_i(t) \cdot S, \text{if link } i \text{ subjects to uncertainty} \\ S, \text{else} \end{cases} \tag{21}$$

Note that S is a fixed value, representing the saturation flow; and $p_i(t)$ denotes a random value generated by certain probability distribution dependent on the type of uncertain event (e.g. a highway crash, lane closure, work zone etc.). Typical probability distributions include the lognormal distribution (Jia et al., 2010) and Weibull distribution (Brilon et al., 2005, 2007). Note that the similar idea to describe a stochastic traffic network environment is also discussed in (Sumalee et al., 2011).

Remark 1: The ratio of the shockwave speed over the free flow speed, i.e., δ_{EB} or δ_{BM}, is a flexible parameter. It is needed calibration from the empirical data before putting into practical use. For a given section of road in the real world, shockwave speed may not change with the change of free flow speed. This setting can be easily achieved by allowing a flexible ratio of the two types of speeds.
Remark 2: The LDNL model differentiates from the link transmission model (Yperman, 2007) in the following ways. (1) Links are hypothetically divided into three parts. This setting makes it possible and easy to accommodate the dynamic change of speed limits. (2) Occupancies are updated directly in the three parts, and flows are propagated between parts.

We do not apply the node model or the concept of accumulated number of vehicles as in (Yperman, 2007). (3) The traffic flow propagation is more accurate, as the backward wave propagation of congestion is well defined and captured by checking the difference between the queue forming end and dissipation end of the main part.

2.4. Link average speed estimation

The advantage of the link based DNL lies in that it covers the whole range of traffic dynamics including queue formation, dissipation and kinematic wave. With the density or cell occupancy determined, the mean speed at the link level can be derived, making it applicable for more accurate emission estimation.

Based on the fundamental diagram (Fig. 5) of traffic flow, the traffic speed of each link can be calculated as:

$$v_i^t = \frac{q_i^t}{k_i^t} \tag{22}$$

Moreover, from the fundamental diagram of the link based DNL, we get:

$$q_i^t = \min(V_i^t \cdot k_i^t, Q_E^{i,t}, W_i^t \cdot (d_J - k_i^t)) \tag{23}$$

Substituting (23) into (22), we obtain:

$$v_i^t = \min\left(V_i^t, \frac{Q_E^{i,t}}{k_i^t}, \frac{W_i^t \cdot (d_J - k_i^t)}{k_i^t}\right) \tag{24}$$

Note that the density of the link i is calculated as $k_i^t = \frac{x_B^{i,t} + x_M^{i,t} + x_E^{i,t}}{L_M^{i,t} + 2L_0^t}$. Hence (24) is written as:

$$v_i^t = \min\left(V_i^t, \frac{Q_E^{i,t}(L_M^{i,t} + 2L_0^t)}{x_B^{i,t} + x_M^{i,t} + x_E^{i,t}}, \frac{W_i^t \cdot (d_J(L_M^{i,t} + 2L_0^t) - x_B^{i,t} - x_M^{i,t} - x_E^{i,t})}{x_B^{i,t} + x_M^{i,t} + x_E^{i,t}}\right) \tag{25}$$

3. Reinforcement learning for dynamic speed limit control

Notation:

s_i^t	Current state of link i at time t
$\tilde{s}_i^t$	Next state of link i at time t
a_i^t	Action(speed limit) of link i at time t
$r_i^t(s_i^t, a_i^t, \tilde{s}_i^t)$	Observed reward when the agent takes action a_i^t in state s_i^t, and moves to state $\tilde{s}_i^t$ at link i
I	Speed limit rate, e.g. 5 mph
V_0	Base speed limit, e.g. 20 mph
$\rho(s_i^t, a_i^t)$	Average reward of state-action pair (s_i^t, a_i^t) at link i
$\bar{Q}(s_i^t, a_i^t)$	Q-value of state-action pair (s_i^t, a_i^t) at link i
$\alpha^{(k)}$	Learning rate for the Q-values at kth iteration
$\beta^{(k)}$	Learning rate for the average reward at kth iteration
γ	Discount factor for reward value
ε	Greedy value

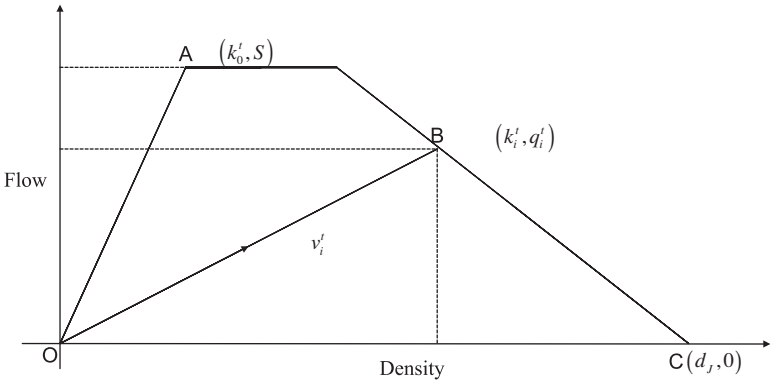

Fig. 5. Flow-density fundamental diagram for the link based DNL model.

3.1. Dynamic speed limit problem as Markov Decision Process (MDP)

Optimization of traffic control (different objectives can be filled in here, e.g. total travel time, number of stops, vehicular emission) requires the determination of optimal speed limits. To attain the optimal objective, the speed limit controller has to allocate different speed limits to vehicles traveling on the link depending on the congestion level and different time period of the day. The controller takes decision at specific intervals that is determined beforehand by the speed limit planner. The traffic network is the environment and the traffic controllers act as agents in this context. Due to the uncertain traffic demand and supply, the network environment is stochastic. The action of an agent is defined as to activate different speed limits at the decision interval. Note that, the transition time from one state to another state after activating any of the speed limits is unity (or same for all cases). Thus, the dynamic speed limit problem has all the elements of MDP. Each time the agent takes an action that impacts the current environment, the state of the environment changes. The problem is to find the optimal policy (mapping between the speed limit activations and traffic states) that gives the maximum reward measured in terms of total travel time, number of stops, vehicular emission, etc. in the long term. Let set T, S_t, A_t, P_t, R_t denote the set of time steps T, state s_i^t, action a_i^t, transition probability, and cost $r_i^t()$ respectively, then the dynamic speed limit problem is characterized by $\{T, S_t, A_t, P_t, R_t\}$, and can be formulated into the infinite time horizon MDP as:

$$\bar{Q}^*(s_i^t, a_i^t) = \max_{a_i(\cdot)\in A_t} \bar{Q}(s_i^t, a_i^t), \forall i \tag{26}$$

Reinforcement learning techniques have been effectively applied to solve practical problems involving optimal control and optimization in different disciplines of science and engineering. In general, any method applying the sampling based techniques to solve the optimal control problems or its variants can be defined as reinforcement learning (RL). The agent (the speed limit controller) interacts with the environment (the system or any representative model) by taking some action and the environment reacts to that action by changing its state. In addition, the environment also interacts with the agent to determine how much reward it gains by performing that action. The reward gives a measure of the effectiveness of the action taken by the agent to reach its optimization goals. In the context of dynamic speed limit problem the speed limit controller is the agent and the traffic network (which is dynamic and random) is the environment. The reinforcement learning (RL) system mainly consists of three components: the state, action, and reward, as defined in the following sections.

3.2. State of the speed limit controller

State of the speed limit controller is a measure of the real time traffic environment, or the evolution of traffic flow. It is obvious that the evolution of traffic flow is a continuous process. In other words, the state of traffic environment has infinite dimensions. Setting the state with high dimensions may increase the difficulty of solving the problem ("curse of dimension"). Depending on the density, here we characterize the state of the speed limit controller into four congestion levels: free flow state (stated value as 1), slight congestion state (stated value as 2), moderate congestion state (stated value as 3), heavy congestion state (stated value as 4). Specifically, the discretization of state for a controlled link is as below.

$$s_i^t = \begin{cases} 1, & \text{if } k_i^t \leqslant 0.25 \cdot d_J \\ 2, & \text{else if } k_i^t \leqslant 0.50 \cdot d_J \\ 3, & \text{else if } k_i^t \leqslant 0.75 \cdot d_J \\ 4, & \text{else if } k_i^t \leqslant d_J \end{cases} \tag{27}$$

3.3. Actions of the speed limit controller

The action taken by the speed limit controller is defined as to assign speed limits to the traffic on the controlled link. Similar to the definition of state, the speed limit V_i^t is a continuous function. However, considering the real world implementation and the difficulty of solving high dimension problems, here we characterize the possible speed limits as below.

$$V_i^t = V_0 + a_i^t \cdot I, \quad a_i^t = \{1, 2, 3, \cdots, A\} \tag{28}$$

where A is a positive integer, and $V_0 + A \cdot I$ denotes the maximum speed limit.

Reinforcement learning algorithms in general require a balance between exploitation and exploration in the strategies for selecting optimal action. The simplest action rule is to select the action (or one of the actions) with the highest estimated state-action value (complete greedy behavior). In other words, the agent always tries to maximize the immediate reward using the immediate knowledge without any attempt to explore other possible actions. To balance between exploitation and exploration we apply the ε – greedy method (Sutton and Barto, 1998). In this method, the agent chooses the action that result in the maximum state-action value in most cases except in a few cases where a random action is chosen to explore other possible actions. The probability of this random behavior is ε and the probability of selecting the optimal action converges to greater than $1 - \varepsilon$. One should note that, the advantage of ε – greedy methods over the greedy methods is highly dependent on the type of problem. For instance, with higher variance in the reward values the ε – greedy methods might perform better.

3.4. Reward function

The reward function definition is closely related the optimized objective of the whole model. Depending on the metrics of interest, we can define reward as total travel time, queue length, waiting time, delay time, vehicular emission, etc. In this study, we adopt the total travel time as the intended objective. The reward function is defined as below.

$$r_i^t(s_i^t, a_i^t, \tilde{s}_i^t) = -\left(x_B^{i,t} + x_M^{i,t} + x_E^{i,t}\right) \tag{29}$$

Note that though we specify reward as the total travel time in this study, other types of reward are also readily fit into the modeling framework.

3.5. RMART algorithm description

Due to the uncertain traffic demand and supply, traffic volume of a link is a stochastic process and the state in the reinforcement learning system is highly dependent on that. Two distinct properties of traffic dynamics are: the similarity of traffic pattern (e.g., the traffic pattern at a particular link on each Sunday during 11 am-noon) and heterogeneity in the network congestion. To account for these attributes, this research deploys an average reward technique which is also known as advanced off-policy R-Markov Average Reward Technique(R-MART) (Gosavi, 1997; Sutton and Barto, 1998). Like most RL based schemes, the proposed algorithm has two phases: learning phase and implementation phase. The implementation phase takes place after the learning phase. The key difference in the techniques stated above is the process of updating the state-value function (Aziz et al., 2013). During the learning phase the agents update the state-action value by interacting with the environment. Balancing the exploration and exploitation is important at this phase. Initially, the algorithm starts with ε – greedy using higher ε value. Then the ε value gradually decreases towards the end of the learning phase. During the implementation period, the algorithm emphasizes on exploitation with very small ε value. Since the only change from the learning to implementation phase is the action selection strategy, only the learning phase algorithm is described below.

R-MART is an advanced off-policy temporal difference (TD) control algorithm. This method does not divide the experience into separate episodes with finite returns. The value function is a measurement of how good the agent performs under the state by taking the action, i.e., the value function measures the expected return by considering future reward that can be

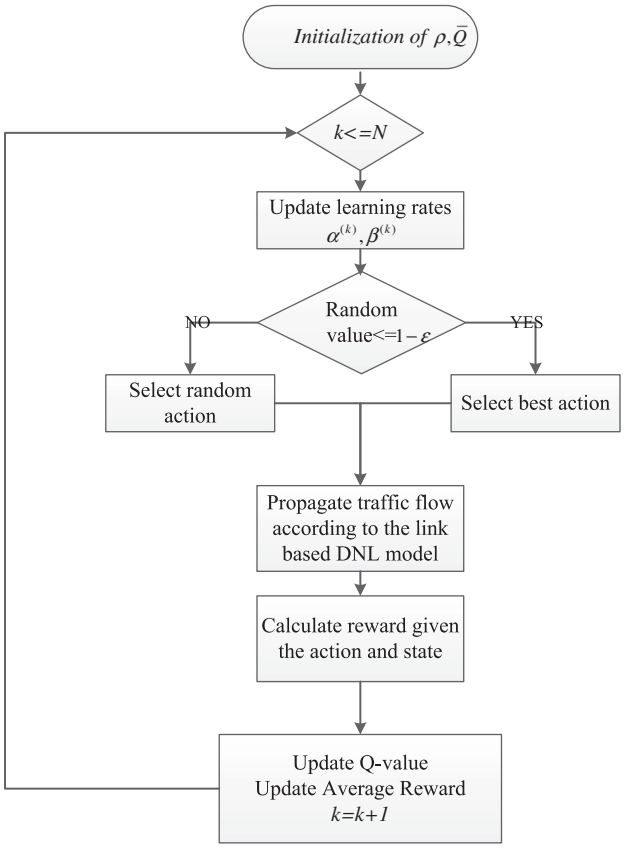

Fig. 6. Flow chart of the R-MART algorithm.

expected. Precisely, the value function is defined with respect to the average expected reward per time step and is defined as:

$$\rho(s_i^t, a_i^t) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(r_i^t(\cdot)) \tag{30}$$

RMART assumes ergodic process, i.e., it does not depend on the initial state. For any initialized state the long-term average should yield the same value. One clear distinction from other temporal difference technique is the use of relative value functions. The values are relative to average reward under the active policy (Gosavi, 1997; Sutton and Barto, 1998). RMART uses the concept of average reward over long term instead of discounted reward used in *Q*-learning and SARSA. (Tsitsiklis and Van Roy, 2002) made an analytical comparison between the discounted (*Q*-learning) and average reward techniques (RMART) and showed that as the discount factor approaches one, the value function by discounted technique approaches the

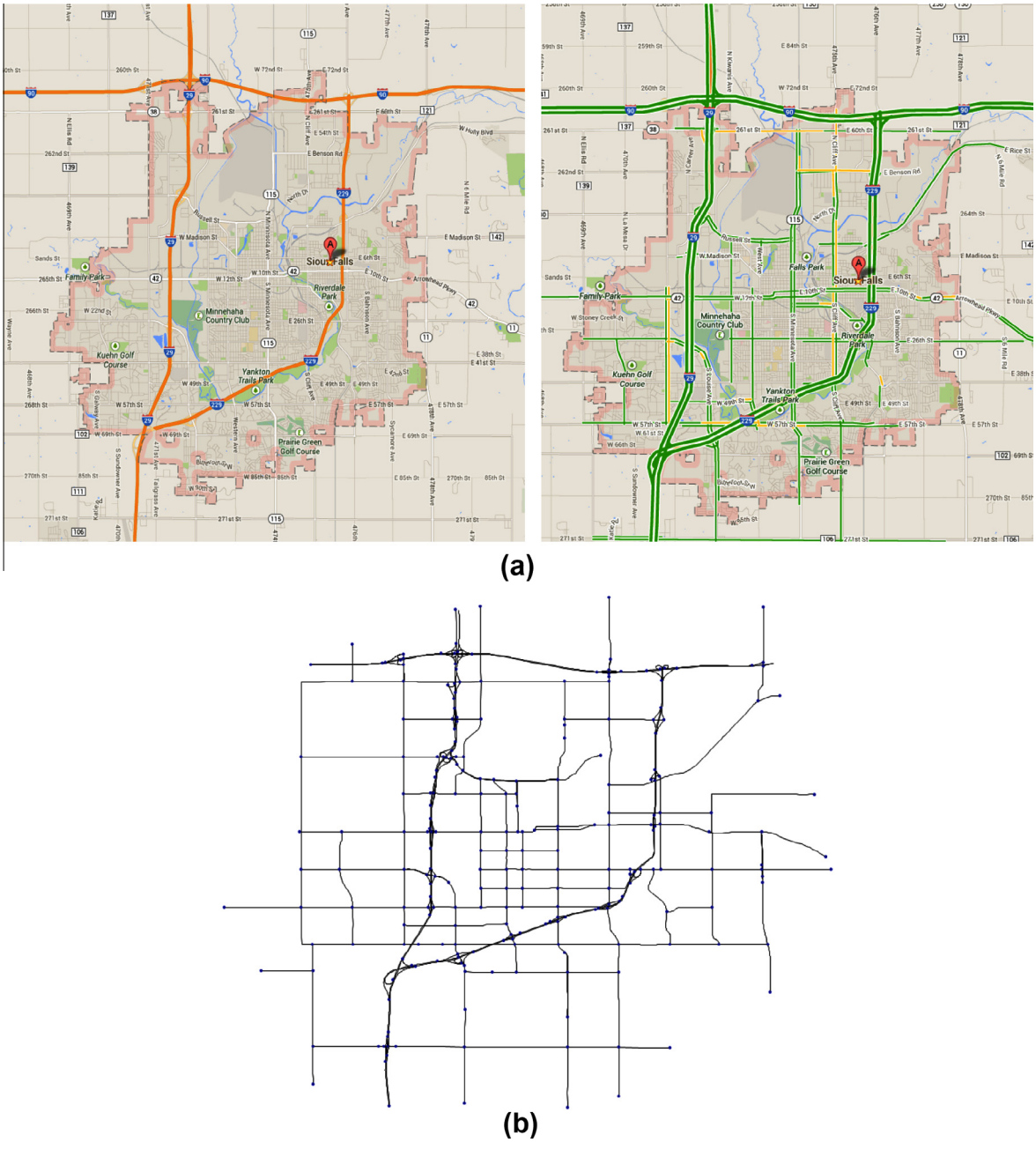

Fig. 7. (a) Google map of the Sioux Falls network (b) link representation of the Sioux Falls network.

differential value function by average reward technique. In addition, average reward methods offer computational advantages (Tadepalli and Ok, 1998). Fig. 6 presents the flow chart of the RMART algorithm to solve the dynamic speed limit problem. The Pseudo code for the RMART algorithm is also presented at Appendix A.

Table 1
Parameter settings for the test network.

Jam density	200 veh/km
Time step	2 s
Demand variation interval	60 s
Capacity variation interval	300 s
Speed limit rate	7.5 mph
V_0	20 mph
Duration	3600 s
Shockwave speed over free flow speed ratio	0.4
Saturation flow rate	1 veh/step

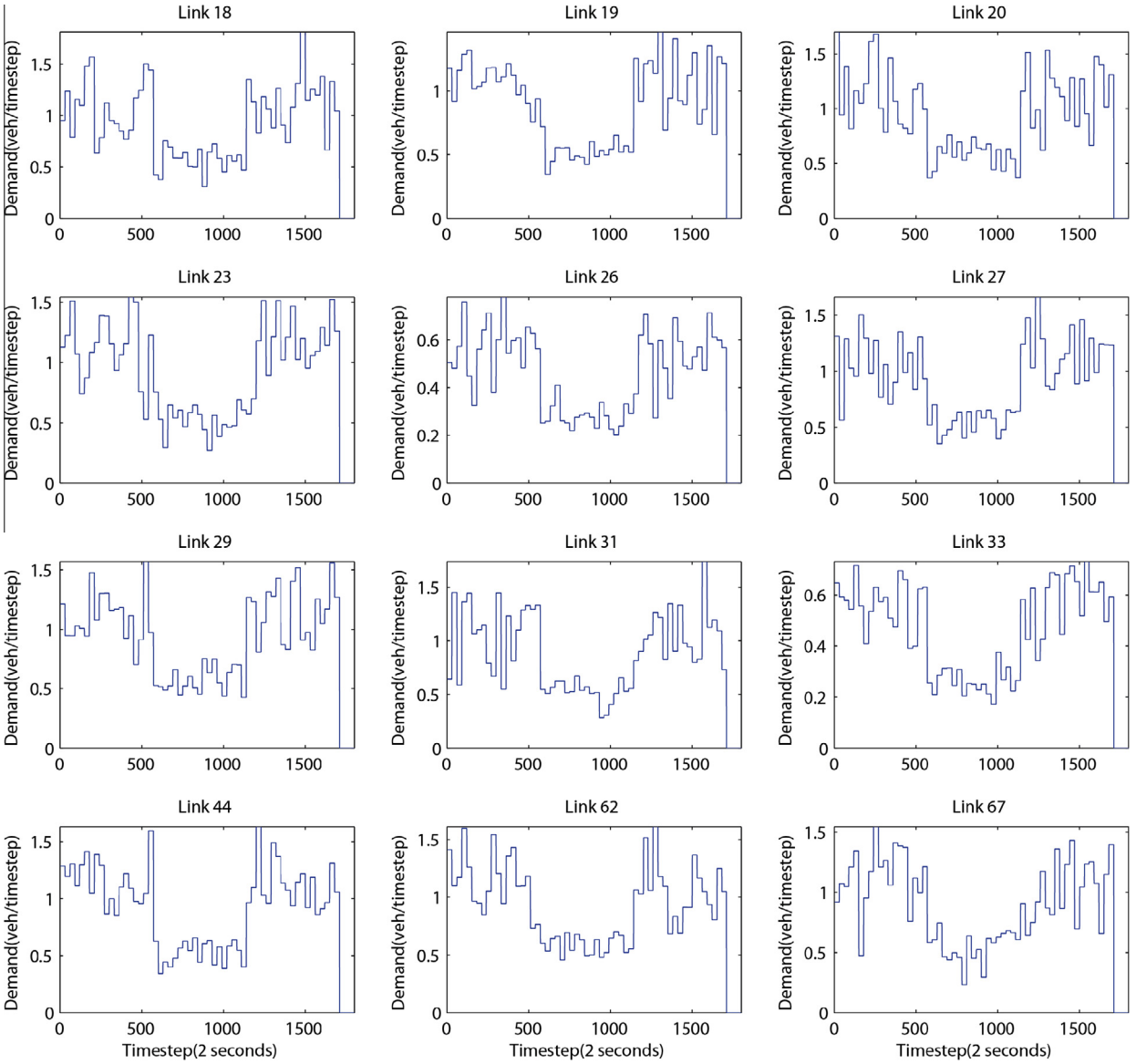

Fig. 8. Demand input variation of origin link.

4. Test case study

4.1. Experiment design

The test case study is conducted on the Sioux Falls network as shown in Fig. 7. Fig. 7a shows the Sioux Falls network from Google map. Fig. 7b shows the link representation of the network. The network configuration (e.g. length of links, connectivity between links) in Fig. 7b is consistent with the actual network (Fig. 7a). There are 547 links, 346 nodes in Fig. 7b. In this study we do not consider all the links under dynamic speed limit control. The selection criteria of links for dynamic speed limit control are: (1) the base speed limit (the actual speed limit in real world) is greater than 25 mph (links with speed limit of 25 mph or less in the real world are probably located in residential areas, hence may not subject to dynamic changes); (2) the length of the link is greater than 150 m (150 m is a tentative threshold parameter to sort out short links in the network); (3) the links do not connect to origin nodes or destination nodes. Under such criteria, 371 links are finally selected as links for dynamic speed limit control.

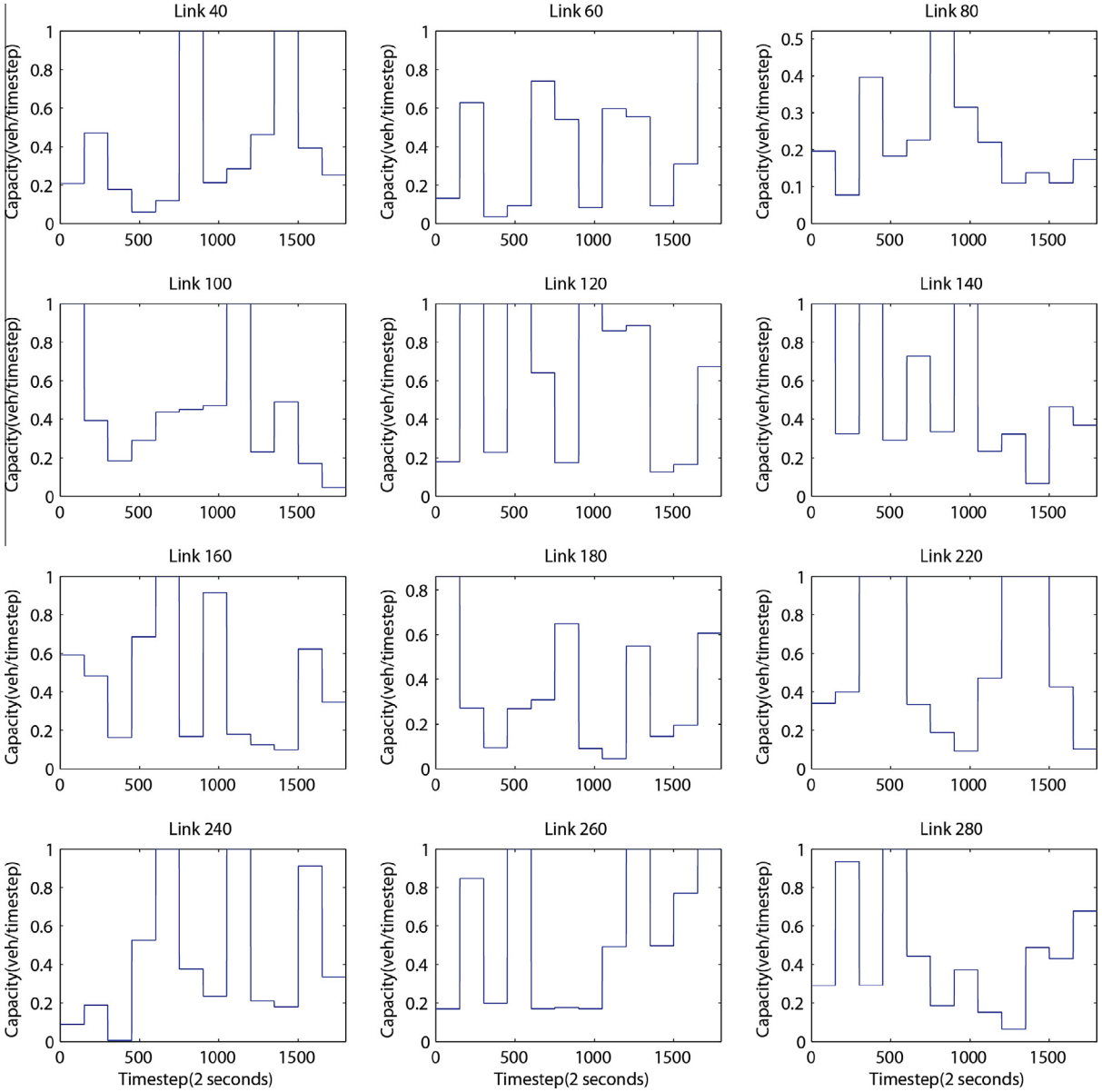

Fig. 9. Exit capacity variation of links under speed limit control.

For the action space of the controller, we set $V_0 = 20, I = 7.5, A = 8$, hence the range of the speed limit is between 20 mph and 80 mph. The saturation flow is set as 1800 (vph). The duration of the simulation is 3600 s with a time step size of 2 s. Hence there are 1800 time steps. For the input demand, we randomly generate demand values according to the lognormal probability distribution, with two peaks (mean value of 750 vph), and the deviation percentage value is 0.20. Furthermore, the demand input is generated every 30 time steps (i.e. every 60 s) from 0 to 1700 time steps. In such setting, there is no demand input for the last 100 time steps. Hence the traffic in the network can be cleared at the end of the simulation. Moreover, to model the random disturbance of traffic (e.g. a highway crash, lane closure, work zone etc.), the capacity of the exit part in some link is generated randomly according to the lognormal probability distribution. Mean value of the capacity is 1080vph, and deviation percentage is 0.20. We select links whose index number is divisible by 20 (20 is just a tentative threshold parameter) for such uncertain capacity setting. Similar to the demand input generation, exit capacity of the link is changing every 150 time steps (300 s) in the simulation. More details on the preset parameters can be found in Table 1.

In the experiment design, we only set the total travel time as the optimized objective. Meanwhile, we are also interested in measuring the environmental benefits of implementing the dynamic speed limit control. Hence we compute the emissions in terms of CO by applying the follow relationship between CO and speed (Aziz and Ukkusuri, 2012):

$$E_{CO} = -0.064 + 0.0056 \cdot v + 0.00026(v - 50)^2 \tag{31}$$

where E_{CO} is the emission rate in terms of CO (g/s), and v is the speed of the vehicle (mph).

4.2. Result analysis

A sample of the demand inputs of the origin links are presented in Fig. 8. It is shown that the traffic demand is generated randomly every 30 time steps, with two peaks at a hight demand level of 0.8 veh/step (note that the saturation flow rate is 1 veh/step). Moreover, a sample of the capacity of the specified links are presented in Fig. 9. Due to the setting in the experiment design, the capacity is fluctuating every 150 time steps. Still, we see that the capacity is not a fixed value but generated randomly. The uncertainty from both the traffic demandside and the infrastructure supply side defines the stochastic traffic network environment.

Note that in this study, we set total travel time as the optimized objective. The results of the total travel time, and emission (in terms of CO) for different simulation runs are shown in Fig. 10. In this study, we set the maximum iteration as 100. It

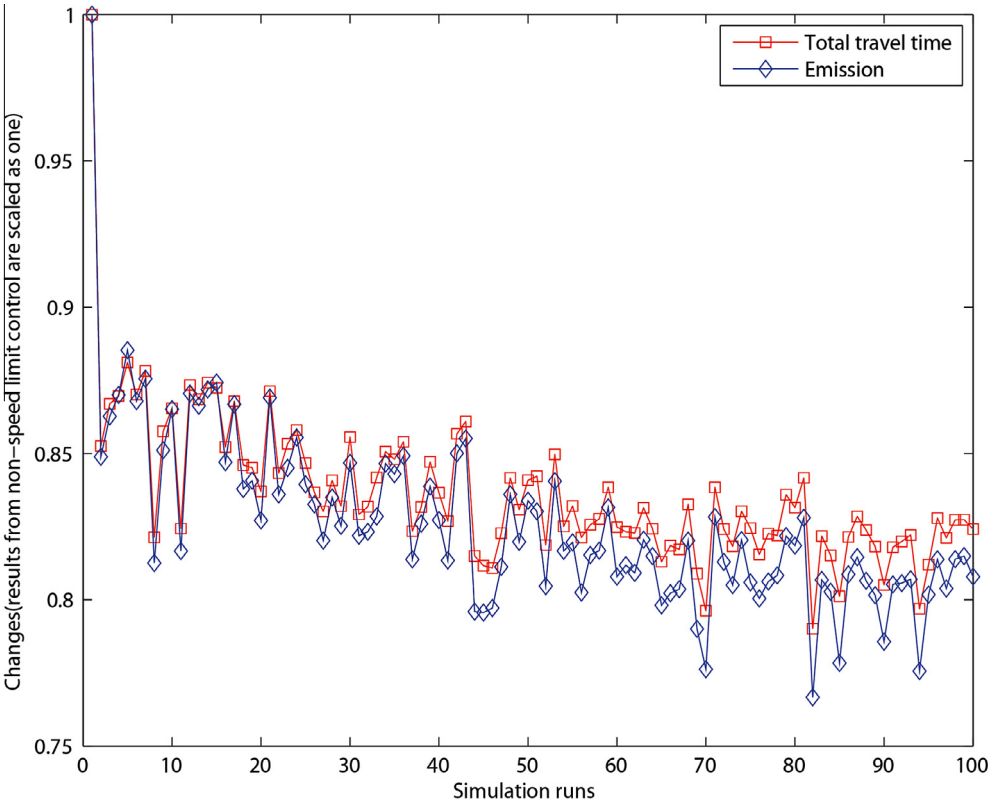

Fig. 10. Variation of total travel time and emission (in terms of CO) of different simulation runs.

is clear to see that, as iteration goes on, the total travel time and emission are getting lower. From around the 80th iteration, the total travel time is fluctuating within a small range of 0.82 (18% reduction), and the total emission is fluctuating within the range of 0.80 (20% reduction). Note that the results of the non-speed limit control case are set as the base for comparison. The trend of the total travel time confirms the R-MART algorithm's performance in minimizing the total travel time. On the other hand, the result is not converging to a stable point. This is also expected. One reason is that the R-MART algorithm takes action according ε – greedy method. It does not take the action acoording to R-MART algorithm every time, but also explores the possibility of other actions by a small probabilty $(1-\varepsilon)$. Another reason is due to the uncertainty from both the traffic demand side and supply side.

To further understand how the speed limits under DSL are different from the speed limits in the real world, we plot the speed limit variations from the last simulation run of the proposed speed limit controller, as well as the base speed limit with no control (actual speed limit obtained from the real world). Figs. 11 and 12 show the speed limit and incoming flow variation for the links with real world speed limit of 45 mph and 65 mph. Each figure contains 16 sample links of the category in the network. Based on Figs. 11 and 12, we see that the speed limits under DSL are significantly different from that of the real world. For some link, it is better to increase the speed limit for some period of time. However, for some other link, it is more beneficial to decrease the speed limit. A static scheme of speed limits may impede the traffic network's efficiency in terms of total travel time and vehicular emission (as shown in Fig. 10). This observation confirms the need for dynamic speed limit control. Moreover, the variation of speed limits under DSL seems highly correlated with the incoming flow of the corresponding link. As shown in Figs. 11 and 12, the speed limits under DSL for link 325, 364, 390, 391 are higher than that under non-DSL. The incoming flow for these links are quite low (less than 0.2 veh/time step), suggesting the congestion level is low. On the other hand, for link 323, 354, 385, the speed limits under DSL are lower than that under non-DSL. The incoming flow for these links are quite high (close to or higher than 0.5 veh/time step), suggesting a higher congestion level. Though such

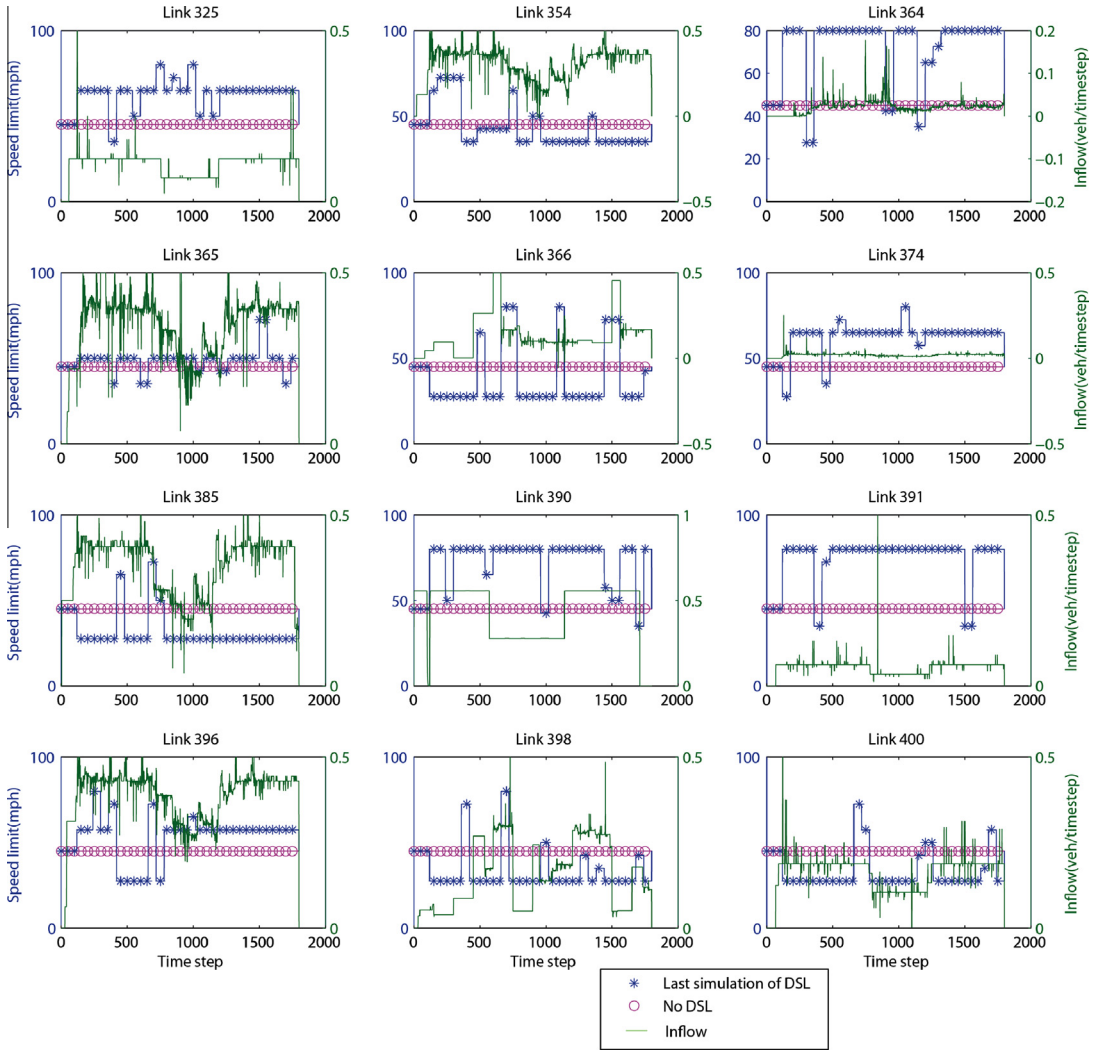

Fig. 11. Speed limit and incoming flow variation of links (real world speed limit 45 mph).

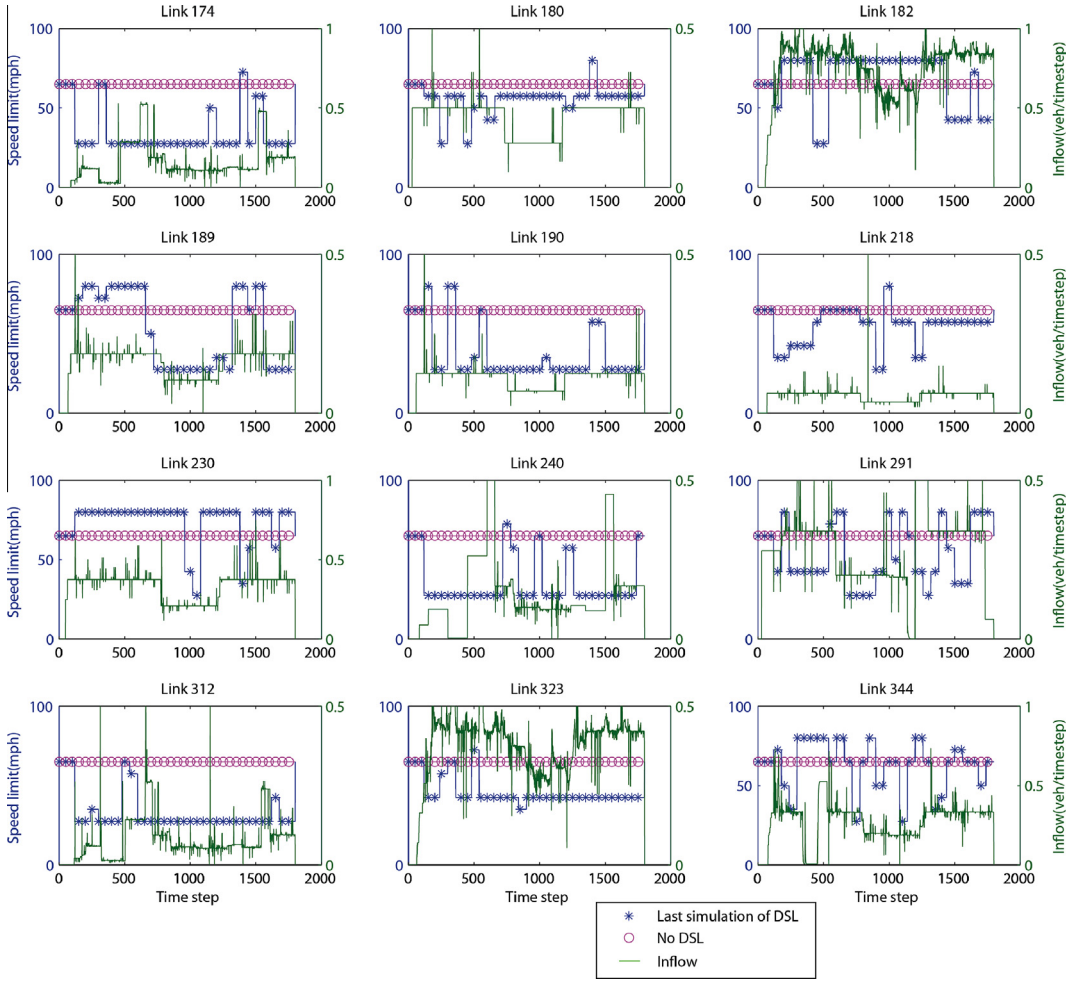

Fig. 12. Speed limit and incoming flow variation of links (real world speed limit 65 mph).

findings cannot be generalized to all the links in the network and are limited by the sample size, it provides useful insights for urban design.

5. Conclusions

This paper proposes a novel dynamic speed limit control model based on reinforcement learning approach. In the model, it is required that the traffic flow information of the link is known to the speed limit controller. This setting is technologically possible under the CV environment. Further, the model is built upon a stochastic network environment. The traffic demand input of origin links and exit capacity of some selected links generated randomly according to certain probability distribution to account for the uncertainty from both demand and supply sides. For the underlying traffic flow model, we develop a link-based dynamic network loading (LDNL) model accounting for speed limits. In the proposed LDNL model, every link of the given network is divided into three components, namely, the beginning part, the main part, and the ending part. The beginning part and the ending part are of basic size depending on the preset resolution, while the main part allows flexible size depending on the residual length of the link. The shockwave propagation is well defined and captured by checking the difference between the queue forming end and dissipation end of the main part. By making use of the fundamental diagram in the LDNL model, we are able to obtain the speed profile which can then be used to estimate the vehicular emission. Moreover, the dynamic speed limit problem is modeled as a Markov Decision Process (MDP) problem. Different metrics, e.g. total network throughput, delay time, vehicular emissions, can be easily set as the optimized objectives in the modeling framework. The speed limit controller is modeled as an intelligent agent interacting with the stochastic network environment by taking actions, which is to determine time-dependent speed limits. The optimal speed limit scheme is obtained by applying the R-Markov Average Reward Technique (R-MART) based reinforcement learning algorithm.

In the numerical case study, we reconstructed the Sioux Falls network to demonstrate the performance of the model. In the case study, we set total travel time as the optimized objective. The results of the total travel time, and emission (in terms

of CO) are improving over simulation runs. From around the 80th iteration, the total travel time is fluctuating within a samll range of 0.82, and the total emission is flucturating within the range of 0.80, compared with the result of non speed limit control. The trend of the total travel time confirms the R-MART algorithm's performance in minimizing the total travel time. Moreover, the comparison between the resulted speed limit (DSL case) and the actual speed limit (non-DSL case) in the real world shows that, for some link, it is better to increase the speed limit for some period of time; for some other link, it is more beneficial to decrease the speed limit. We find that dynamic speed limit is highly correlated to the incoming flow of the link. Such insightful findings can be useful for urban planners or policy makers in the decision making process of transportation planning.

There are multiple research directions along the stream of this paper. (1) This study only simulates traffic flow propagation on determined routes. The driver behavior (system optimal or user equilibrium) is not involved. It may be interesting to revisit the problem in the context of dynamic traffic assignment. (2) We only consider single objective (i.e. total travel time) in the proposed reinforcement learning framework. We have not addressed the multi-objective problem. How to balance the tradeoff between different metrics (e.g. total throughput, delay, emission) is also an interesting research direction. (3) The proposed RL framework is a localized optimization framework. Exploring a cooperative RL framework to optimize the system wide performance of dynamic speed limit is also worthwhile for future research.

Appendix A

Pseudo code for the R-MART based Algorithm:

Initialization:

Set initial values for $\rho(s_i^t, a_i^t)$, *and* $\bar{Q}(s_i^t, a_i^t)$ *for all state-action pairs* (s_i^t, a_i^t).

k=1.

While $k \leqslant N$ *(the maximum number of simulation runs)* **Do**

Update *learning rates for Q values and average reward:*

$$\alpha^{(k)} = 10\frac{\log(k+2)}{k+2}$$

$$\beta^{(k)} = \frac{A}{B+k}, \quad A \text{ and } B \text{ are scalars}$$

Action determination *according to* ε*-greedy method*

Traffic flow propagation *according to the LDNL model*

Update *reward:*

Calculate reward $r_i^t(s_i^t, a_i^t, \tilde{s}_i^t)$ *for choosing action* a_i^t *and next state,* $\tilde{s}_i^t$.

Update *Q-values:*

$$\bar{Q}(s_i^t, a_i^t) \leftarrow \bar{Q}(s_i^t, a_i^t) + \alpha^{(k)}\left[r_i^t(\cdot) - \rho(s_i^t, a_i^t) + \max_{\tilde{a}_i}\bar{Q}(\tilde{s}_i^t, \tilde{a}_i^t) - \bar{Q}(s_i^t, a_i^t)\right]$$

Update *average reward:*

If $\bar{Q}(s_i^t, a_i^t) = \max_{a_i}\bar{Q}(s_i^t, a_i^t)$

Then

$$\rho(s_i^t, a_i^t) \leftarrow \rho(s_i^t, a_i^t) + \beta^{(k)}\left[r_i^t(\cdot) - \rho(s_i^t, a_i^t) + \max_{\tilde{a}_i}\bar{Q}(\tilde{s}_i^t, \tilde{a}_i^t) - \max_{a}\bar{Q}(s_i^t, a_i^t)\right]$$

Update $s_i^t \leftarrow \tilde{s}_i^t$; $k \leftarrow k+1$.

End

References

Aziz, H.M.A., Ukkusuri, S.V., 2012. Integration of environmental objectives in a system optimal dynamic traffic assignment model. Comput.-Aided Civil Infrastruct. Eng. 27 (7), 494–511.

Aziz, H., Zhu, F., Ukkusuri, S.V., 2013. Reinforcement learning based signal control using R-Markov Average Reward Technique (RMART) accounting for neighborhood congestion information sharing. In: Proceedings of 92nd Transportation Research Board Meeting, National Academies (Washington, D.C., January 2013).

Brilon, W., Geistefeldt, J., Regler, M., 2005. Reliability of freeway traffic flow: a stochastic concept of capacity. In: Proceedings of the 16th International Symposium on Transportation and Traffic Theory, College Park, Maryland, pp. 125–144.

Brilon, W., Geistefeldt, J., Zurlinden, H., 2007. Implementing the concept of reliability for highway capacity analysis. Trans. Res. Rec.: J. Transport. Res. Board 2027, 1–8.

Carlson, R.C., Papamichail, I., Papageorgiou, M., Messmer, A., 2010. Optimal motorway traffic flow control involving variable speed limits and ramp metering. Transport. Sci. 44 (2), 238–253.

Daganzo, C.F., 1995. The cell transmission model. Part II: Network traffic. Transport. Res. Part B: Methodol. 29, 79–93.

Daganzo, C.F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transport. Res. Part B: Methodol. 28, 269–287.
Gomes, G., Horowitz, R., Kurzhanskiy, A.A., Varaiya, P., Kwon, J., 2008. Behavior of the cell transmission model and effectiveness of ramp metering. Transport. Res. Part C: Emerg. Technol. 16 (4), 485–513.
Gosavi, A., 1997. Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning. Springer.
Han, L.S., Ukkusuri, S., Doan, K., 2011. Complementarity formulations for the cell transmission model based dynamic user equilibrium with departure time choice, elastic demand and user heterogeneity. Transport. Res. Part B – Methodol. 45, 1749–1767.
Hegyi, A., De Schutter, B., Hellendoorn, H., 2005. Model predictive control for optimal coordination of ramp metering and variable speed limits. Transport. Res. Part C: Emerg. Technol. 13 (3), 185–209.
Van den Hoogen, E., 1994. Control by Variable Speed Signs: Results of the Dutch Experiment in Seventh International Conference on 'Road Traffic Monitoring and Control'. IEE, pp. 145–149.
Jia, A., Williams, B.M., Rouphail, N.M., 2010. Identification and calibration of site specific stochastic freeway breakdown and queue discharge. Transport. Res. Rec.: J. Transport. Res. Board 2188, 148–155.
Kang, K., Chang, G.L., Zou, N., 2004. Optimal dynamic speed-limit control for highway work zone operations. Transp. Res. Rec. 1877, 77–84.
Keller, J., Andreani-Aksoyoglu, S., Tinguely, M., Flemming, J., Heldstab, J., Keller, M., Zbinden, R., Prevot, A., 2008. The impact of reducing the maximum speed limit on motorways in Switzerland to 80 km h^{-1} on emissions and peak ozone. Environ. Modell. Software 23 (3), 322–332.
Kloeden, C.N., Ponte, G., McLean, A.J., 2001. Traveling Speed and the Risk of Crash Involvement on Rural Roads.
Krishnamurthy, S., Kockelman, K., 2003. Propagation of Uncertainty in Transportation Land Use Models: Investigation of DRAM-EMPAL and UTTP Predictions in Austin, Texas. In Transport. Res. Rec.: J. Transport. Res. Board, No. 1831, TRB, National Research Council, Washington, D.C., pp. 24, 219–229.
Kotsialos, A., Papageorgiou, M., Mangeas, M., Haj-Salem, H., 2002. Coordinated and integrated control of motorway networks via non-linear optimal control. Transport. Res. Part C: Emerg. Technol. 10 (1), 65–84.
Lee, C., Hellinga, B., Saccomanno, F., 2006. Evaluation of variable speed limits to improve traffic safety. Transport. Res. Part C: Emerg. Technol. 14 (3), 213–228.
Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves. II. A theory of traffic flow on long crowded roads. Proc. R. Soc. Lond. Ser. A Math. Phys. Sci. 229, 317.
Lo, H.K., Szeto, W.Y., 2002. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. Transport. Res. Part B: Methodol. 36 (5), 421–443.
Lo, H.K., 1999. A novel traffic signal control formulation. Transport. Res. Part A: Policy Practice 33, 433–448.
Lo, H.K., Chang, E., Chan, Y.C., 2001. Dynamic network traffic control. Transport. Res. Part A: Policy Practice 35, 721–744.
Madireddy, M., De Coensel, B., Can, A., Degraeuwe, B., Beusen, B., De Vlieger, I., Botteldooren, D., 2011. Assessment of the impact of speed limit reduction and traffic signal coordination on vehicle emissions using an integrated approach. Transport. Res. Part D: Transport Environ. 16 (7), 504–508.
Ossiander, E.M., Cummings, P., 2002. Freeway speed limits and traffic fatalities in Washington State. Accid. Anal. Prev. 34 (1), 13–18.
Papageorgiou, M., Kotsialos, A., 2002. Freeway ramp metering: an overview. IEEE Trans. Intell. Transp. Syst. 3 (4), 271–281.
Ng, M.W., Kockelman, K., Waller, S.T., 2010. Relaxing the multivariate normality assumption in the simulation of transportation system dependencies. Transport. Lett.: Int. J. Transport. Res. 2 (2), 63–74.
Raemae, P., 2002. Effects of weather-controlled variable message signing on driver behaviour. Nordic Road Transport Res. 1, 17–19.
Richards, P.I., 1956. Shock waves on the highway. Operat. Res. 4, 42.
Siu, B.W.Y., Lo, H.K., 2008. Doubly uncertain transportation network: degradable capacity and stochastic demand. Eur. J. Oper. Res. 191, 166–181.
Sumalee, A., Zhong, R.X., Pan, T.L., Szeto, W.Y., 2011. Stochastic cell transmission model (SCTM): a stochastic dynamic traffic model for traffic state surveillance and assignment. Transport. Res. Part B: Methodol. 45 (3), 507–533.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. A Bradford Book.
Szeto, W.Y., Lo, H.K., 2004. A cell-based simultaneous route and departure time choice model with elastic demand. Transport. Res. Part B – Methodol. 38, 593–612.
Tadepalli, P., Ok, D., 1998. Model-based average reward reinforcement learning. Artif. Intell. 100 (1–2), 177–224.
National Research Council (U.S.). Transportation Research Board. Committee for Guidance on Setting and Enforcing Speed Limits 1998. Managing speed: review of current practice for setting and enforcing speed limits. Transportation Research Board, National Research Council. National Academy Press.
Tsitsiklis, J.N., Van Roy, B., 2002. On average versus discounted reward temporal-difference learning. Mach. Learn. 49 (2–3), 179–191.
Ukkusuri, S.V., Ramadurai, G., Patil, G., 2010. A robust transportation signal control problem accounting for traffic dynamics. Comput. Oper. Res. 37, 869–879.
Ukkusuri, S.V., Han, L., Doan, K., 2012. Dynamic user equilibrium with a path based cell transmission model for general traffic networks. Transport. Res. Part B: Methodol. 46 (10), 1657–1684.
Wang, S., 2013. Efficiency and equity of speed limits in transportation networks. Transport. Res. Part C: Emerg. Technol. 32, 61–75.
Yang, H., Wang, X., Yin, Y., 2012. The impact of speed limits on traffic equilibrium and system performance in networks. Transport. Res. Part B: Methodol. 46 (10), 1295–1307.
Yperman, I., 2007. The Link Transmission Model for dynamic network loading. PhD Dissertation.
Zhao, Y., Kockelman, K., 2002. The propagation of uncertainty through travel demand models: an exploratory analysis. Ann. Regional Sci. 36, 145–163.