

# Eco-system optimal time-dependent flow assignment in a congested network

Chung-Cheng Lu<sup>a</sup>

Email: [jasonccclu@gmail.com](mailto:jasonccclu@gmail.com)

Jiangtao Liu<sup>b</sup>

Email: [jliu215@asu.edu](mailto:jliu215@asu.edu)

Yunchao Qu<sup>b,c</sup>

Email: [quyunchao0613@gmail.com](mailto:quyunchao0613@gmail.com)

Srinivas Peeta<sup>d</sup>

Email: [peeta@purdue.edu](mailto:peeta@purdue.edu)

Nagui M. Rouphail<sup>e</sup>

Email: [rouphail@ncsu.edu](mailto:rouphail@ncsu.edu)

Xuesong Zhou<sup>b\*</sup>

Email: [xzhou74@asu.edu](mailto:xzhou74@asu.edu)

a. Department of Industrial Engineering and Management, National Taipei University of Technology, Taipei, 106, Taiwan

b. School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, 85287, USA

c. State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, 100044, China

d. School of Civil Engineering, Purdue University, West Lafayette, IN, 47907, USA

e. Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC, 27695-7908, USA

\*Corresponding author. Tel.: 480-965-5827

## Abstract

This research addresses the eco-system optimal dynamic traffic assignment (ESODTA) problem which aims to find system optimal eco-routing or green routing flows that minimize total vehicular emission in a congested network. We propose a generic agent-based ESODTA model and dissect the link travel times and speeds obtained by three mesoscopic traffic simulation methods to highlight the need for a dynamic network loading (DNL) model to effectively generate time-dependent speeds for multi-scale emission analysis. The relationship between link emission and delay is also analyzed. Based on a modified point-queue model, an expanded space-time network is constructed to formulate the ESODTA with constant bottleneck discharge capacities. The resulting linear integer model of the ESODTA is solved by a Lagrangian relaxation-based algorithm. To address the more general ESODTA with time-dependent discharge rates in a congested network, we propose a column generation-based algorithm, which consists of a mesoscopic DNL model and a gradient projection-based descent direction method for updating time-dependent vehicular path assignments. We derive a formula of marginal emission which encompasses the marginal travel time as a special case, and develop an algorithm for evaluating path marginal emissions in a congested network. Numerical experiments are conducted to demonstrate that the proposed algorithm is able to effectively obtain coordinated route flows that minimize the system-wide vehicular emission for medium-scale and large-scale networks.

**Keywords:** green transportation; vehicular emission modeling; eco-routing; marginal emission; multi-scale dynamic network loading.

# 1. Introduction

## 1.1 Motivation and objective

Highway vehicles have been the primary focus of environmental protection and transportation agencies to reduce greenhouse gas emissions, since they account for 72 percent of total transportation emissions (Greene and Schafer, 2003). Among various measures that have been considered for potential vehicular emission reduction (such as energy efficiency improvements, low-carbon alternative fuels, increasing the operating efficiency of the transportation system, and reducing travel), eco-routing or green routing in route guidance provision is receiving increasing attention from the field of green transportation. The idea of green routing is to help drivers make greener choices about their routes by providing the most eco-friendly route in terms of minimum emissions. In a recent laboratory experiment conducted at the University of California at Berkeley, subjects were found to be willing to adjust their route choice behaviour to reduce emissions, exhibiting an average willingness to pay for emissions reduction, or value of green, of 15 cents per pound of CO<sub>2</sub> saved (Gaker et al., 2010; Gaker et al., 2011).

Directing an individual vehicle to a green route can reduce its eco-cost or emission to the environment. However, without an effective system-wide coordination, independent drivers acting non-cooperatively would affect and even worsen traffic conditions and emissions. Instead, system optimal green routing or eco-routing policies that result in minimal total system emission may be of greater interest to public-sector environmental protection and traffic management agencies. The minimal total system emission serves as a benchmark to evaluate the benefits of practical traffic emission reduction measures. Moreover, the resulting green routing policies provide valuable insights for designing those measures. Therefore, this research intends to find time-dependent system optimal green routing policies that minimize the network-wide vehicular emission, which is termed as the Eco-System Optimal DTA (ESODTA) problem.

Classical system optimal dynamic traffic assignment (SODTA) models aim to direct all travelers to paths so as to minimize overall system travel time or cost (e.g., Ghali and Smith, 1995; Peeta and Mahmassani 1995a; Peeta and Ziliaskopoulos, 2001; Peeta and Zhou, 2006). A rich body of literature has been devoted to SODTA models and algorithms, and static traffic assignment (STA) models with environmental considerations have been proposed in a number of past studies (Szeto et al., 2012). Only recently have environment-related objectives and constraints been considered in the dynamic traffic assignment (DTA) context, as a few researchers started to recognize the need for incorporating the effect of traffic dynamics on estimating vehicular emissions so as to thoroughly consider the full set of interacting factors (e.g. Abdul Aziz and Ukkusuri, 2012; Zhou et al., 2015a; Ma et al., 2015).

As the development of ESODTA is still in its infancy, a number of critical but challenging issues need to be addressed to enable ESODTA in congested networks:

(i) Exploring the relationship (or difference) between conventional traffic performance measures (such as travel time, speed, and delay) and emissions is fundamental to the development of ESODTA models, but the properties of the emission generated by vehicles traversing a link were not rigorously analyzed in existing studies.

(ii) It is essential for the dynamic network loading (DNL) model of ESODTA to generate time-dependent speeds so that fine-grained emission estimates can be obtained for multi-scale emission analyses, whereas most of the DNL models underlying classical DTA models focus mainly on link travel times (i.e., the time between the entrance and exit times of a vehicle).

(iii) Congestion effects on vehicular emissions have to be explicitly and effectively taken into account in ESODTA models. Generally, the time in queue is a more appropriate traffic performance measure than the delay when evaluating vehicular emissions and energy consumption in congested networks (Lawson et al., 1997). Thus, it is important for the traffic flow model adopted in the underlying DNL model to explicitly describe vehicles moving in the queue. Although microscopic traffic flow models are desirable for analyzing vehicle trajectories in congested conditions, microscopic traffic simulation can be computationally intensive and typically requires a wide range of detailed geometric data and driving behavior parameters, which can be difficult to calibrate, especially for the purpose of producing high-fidelity emissions estimates.

(iv) Link (or path) marginal emission, the change in emission due to an additional unit of inflow, is critical to the solution algorithms of ESODTA. Although marginal travel time (or delay) has been extensively studied in the literature, none of the existing studies has investigated marginal emission in a congested network. In addition to developing efficient approaches to evaluate link and path marginal emissions, it is important to discuss the relationship between marginal emission and marginal delay.

With the aim of expanding the boundary of SODTA from travel time-based models to emission-based models (i.e., ESODTA), this research addresses the above methodological challenges in enabling SODTA in region-wide emission applications.

## 1.2 Travel time-based and eco-cost-based traffic assignment models

Following the pioneering work of Merchant and Nemhauser (1978), various approaches have been proposed in the past decades to formulate and solve the travel time-oriented system optimal dynamic traffic assignment problem in ideal or general networks, such as mathematical programming (e.g., Merchant and Nemhauser, 1978; Carey, 1987; Carey, 1992; Ziliaskopoulos, 2000), optimal or convex control (e.g., Friesz et al., 1989; LaFortune et al. 1993; Wie et al., 1994), simulation-based optimization (Ghali and Smith, 1995; Peeta and Mahmassani, 1995a; Peeta and Mahmassani, 1995b; Peeta and Zhou, 2006), game theory (Garcia et al., 2000), graphical method (Munoz and Laval, 2005; Shen and Zhang, 2009) and variational inequality (Shen et al., 2007).

As extensions of the user equilibrium and system optimum principles, eco-cost-based, or emission-based, assignment principles have been adopted in a number of STA models. For instance, Benedek and Rilett (1998) presented the emission optimal principle which describes that travellers choose paths so as to minimize the total network emission, rather than total travel time. They also discussed an extension of the user equilibrium principle, the environmental equity principle, in which travellers are assigned in such a way that the amounts of emission on all selected routes are the same. Another line of research was to employ the multi-objective or multi-criterion approaches in traffic assignment models. For example, a multi-criterion system optimum model was proposed by Tzeng and Chen (1993), where the system optimum objective is the sum of total travel time for road users and air pollution for non-users. Nagurney et al. (1998, 2002) presented a multi-class user equilibrium traffic assignment model in which each class of users was assumed to select a route with the least weighted sum of travel time, travel cost and emissions. Zhang et al. (2010) developed a system optimal STA model with the objective being the weighted sum of travel time and emissions. They introduced a cell-based modelling approach for emission concentrations so that either the average or maximum emissions in a network can be considered in the optimization process. A comprehensive review of network equilibrium approaches addressing environmental concerns (e.g., emissions and noise) can be found in Szeto et al. (2012).

Despite the numerous aforementioned studies of STA with environmental considerations, very few DTA models have been developed for environment-related applications. Recently, Abdul Aziz and Ukkusuri (2012) integrated emission-based objective into the traditional travel time-based DTA framework, and developed a SODTA model with dual objectives. They formulated the problem as a nonlinear quadratic program which is readily solved by CPLEX. Zhou et al. (2015a) presented a DTA model and its solution algorithm for a number of emerging emissions and fuel consumption related applications that require both effective microscopic and macroscopic traffic stream representations.

## 1.3 Microscopic and macroscopic traffic flow models for emission estimation

In order to capture the impact of traffic congestion on the energy use and emissions output across different spatial scales (e.g., regions, corridors, segments, and intersections) and various temporal resolutions (e.g., second-by-second, peak hours, and entire day), it is essential for underlying traffic flow models to be able to capture traffic dynamics and describe congestion phenomena (e.g., queue formation, spillback, and dissipation). Microscopic traffic simulation models have been widely used to generate instantaneous speed and acceleration data required by emission models on a vehicle-by-vehicle and second-by-second basis (e.g., Bai et al., 2007; Boriboonsomsin and Barth, 2008; Mandavilli et al., 2008; Panis et al., 2006). However, microscopic traffic simulation is computationally expensive and typically requires a wide range of detailed geometric data and driving behavior parameters, which are difficult to calibrate. Mesoscopic traffic flow modeling may be a more viable approach to strike a balance between the model and computational complexities and the emission resolution.

In their pioneering works, Lighthill and Whitham (1955) and Richards (1956) (LWR) proposed the kinematic wave theory, which rigorously describes traffic flow dynamics by integrating flow conservation constraints, traffic flow models, and partial differential equations (PDEs). Based on a triangular flow density relation, two finite difference-based numerical schemes were proposed to solve the first order kinematic wave problem: (i) by extending deterministic queuing theory, Newell's simplified model (Newell, 1993a, 1993b, 1993c) keeps track of shock waves and queue propagation using cumulative flow counts on links; (ii) Daganzo's Cell Transmission Model (CTM; Daganzo, 1994, 1995a) discretizes a link into many homogenous segments (i.e. cells), and adopts a "supply-demand" or "sending-receiving" framework to model flow dynamics between cells.

Abdul Aziz and Ukkusuri (2012) adopted the CTM as the DNL model underlying their SODTA model and derived link emissions based on the average speed inside a cell in a time interval (e.g. 60 seconds). Note that the average cell-speed approach may not effectively estimate time-dependent emissions, which are highly sensitive to

second-by-second speed variations across different locations. In their work of modeling delay and emission for signalised intersections, [Zhu et al. \(2013\)](#) studied dynamic traffic models that can produce the speed profile, including the car-following model, the point-queue model, the shockwave model, and the CTM, and applied them for both delay and emissions estimation.

Classical point-queue models assume that the link travel time consists of two parts: free-flow travel time and delay. Delay is a typical measure of the impact of congestion on travelers' time. However, for evaluating the congestion effect on vehicular emission and energy consumption, the more appropriate measure is the amount of time actually spent in queue (waiting time or time in queue), which is usually greater than the delay. To effectively measure the time and distance spent by vehicles in a queue, [Lawson et al. \(1997\)](#) proposed using the input-output (or queueing) diagram to determine the spatial and temporal extents of queue upstream of a bottleneck. They derived the relationship between delay and waiting time in queue and constructed the curve depicting the cumulative number of vehicles to have reached the back of the queue as a function of time.

Addressing the need to consistently incorporate different resolutions of traffic descriptions in a traffic flow model, [Leclercq \(2007\)](#) proposed a hybrid LWR model combining both macroscopic and microscopic traffic descriptions and defining simple interfaces to translate the boundary conditions when changing the traffic description. Recently, [Zhou et al. \(2015a\)](#) proposed a mesoscopic DNL model that seamlessly integrates Newell's simplified kinematic wave model (a macroscopic model) and simplified car-following model (a microscopic model) into a unified framework, to evaluate vehicle emission/fuel consumption impact of different traffic management strategies. The advantages of the mesoscopic approach in computational efficiency and in effectively describing free-flow and congested traffic states make it appealing for cross-resolution and multi-scale emission modelling in DTA applications.

#### 1.4 Numerical solution methods for system optimal traffic assignment models

Both exact and heuristic methods have been developed in the literature to solve the SODTA problem on ideal networks or general networks with multiple origin-destination (O-D) pairs. Exact methods were mainly applied to solve the link-based SODTA problem formulated as mathematical programming or optimal control problems. For instance, [Merchant and Nemhauser \(1978\)](#) solved a piecewise linear version of their model by a one-pass simplex method. [Wie et al. \(1994\)](#) developed an augmented Lagrangian method in conjunction with the conjugate gradient method to solve the discrete time optimal control formulation of the problem. On the other hand, heuristics based on some predefined averaging schemes (e.g., [Magnanti and Perakis, 1997](#)), such as the method of successive averages (MSA), have been used for solving path-based SODTA problems ([Peeta and Mahmassani, 1995a; Peeta and Mahmassani, 1995b; Peeta and Zhou, 2006; Shen et al., 2007](#)). The drawback of using MSA is that it uses an across-the-board step size for updating path assignments, so the degree to which the path flows deviate from optimality conditions is not taken into account for different O-D pairs and departure intervals. This may lead to a slow convergence or even failure to converge for some problem instances. (e.g., [Mounce and Smith, 2007](#)). To improve the convergence and the solution quality, [Sbayti et al. \(2007\)](#) proposed an efficient MSA-based implementation technique that uses a sorting technique in updating vehicle assignments based on a selected path travel attribute (e.g. travel time). [Lu et al. \(2009\)](#) developed a path-swapping method which was shown to outperform the MSA on several large network tests.

The link marginal delay (or travel time), which represents the change in delay due to an additional unit of link inflow, is critical to the solution algorithms of SODTA. [Ghali and Smith \(1995\)](#) presented an analytical approach to evaluate link marginal delays on a congested link, based on link cumulative flow curves. [Peeta and Mahmassani, \(1995a\)](#) developed a numerical method based on mesoscopic traffic simulation to evaluate link and path marginal travel times. [Peeta and Zhou \(2006\)](#) extended these concepts to the stochastic case under O-D randomness for multiple user classes. [Shen et al. \(2007\)](#) showed that it is necessary to explicitly trace the propagation of path flow perturbation in evaluating path marginal travel times and proposed an evaluation method of path marginal delays ([Qian and Zhang, 2011; Qian et al. 2012](#)). Focusing on the connection from dynamic system optimum to dynamic user equilibrium conditions, [Carey and Watling \(2012\)](#) developed an analytical model for calculating the marginal costs and externalities based on simplified kinematic wave model. [Lu et al. \(2013\)](#) further examined the partial derivatives of link flow and density and path travel time with respect to an additional unit of perturbation flow. While link and path marginal delays were investigated in the context of SODTA, to the authors' current knowledge, there is no study that addresses link and path marginal emissions.

## 1.5 Overview of the paper

In Section 2, we present a generic agent-based mathematical formulation to model the ESODTA problem, which aims to minimize total network-wide vehicular emission. By dissecting the link travel times and speeds obtained by three mesoscopic traffic flow models: point-queue model, average cell-speed model, and modified point-queue model, we highlight the importance of a dynamic network loading (DNL) model to effectively generate time-dependent speeds for multi-scale emission analysis.

In Section 3, for the case of constant bottleneck discharge rates and without queue-spillback effect (i.e., the queueing speed is constant), we derive an important property of link emission describing the relationship between emission and delay. Based on the modified point-queue model, an expanded space-time network is constructed to formulate the ESODTA with constant discharge rates as a linear integer model, which is solved by a Lagrangian relaxation-based algorithm.

In Section 4, for the case of time-dependent bottleneck discharge rates and queueing speed (e.g., at signalized intersections), we propose a column generation-based algorithm, which consists of a mesoscopic DNL model and a gradient projection-based descent direction method for updating time-dependent path assignments. The mesoscopic DNL model tightly integrates Newell’s simplified kinematic wave and car-following models to generate time-dependent and location-dependent vehicular speed profile and evaluate vehicle emissions from cumulative vehicle counts (Zhou et al., 2015a). We derive a formula of link marginal emission which encompasses the link marginal travel time as a special case, and develop an algorithm for evaluating path marginal emissions in a congested network.

Section 5 presents two numerical examples: one for the ESODTA with constant bottleneck capacities while the other for the ESODTA with time-dependent bottleneck capacities. We compare the results of the ESODTA with those of the travel time-based SODTA and of the user equilibrium DTA (or UEDTA). Concluding remarks are given in Section 6.

## 2. Agent-based ESODTA model

The notations used to present the generic agent-based ESODTA model are defined as follows.

### Indices:

$\tau$	Index of departure time interval
$w$	Index of OD pair
$p$	Index of path for time-dependent OD pair $(w, \tau)$
$i, j, k$	Node index in physical network
$l, (i, j)$	Index of link $l = (i, j)$ , $l = 1, 2, \dots, L$
$t$	Index of simulation time interval
$f$	Index of agent (vehicle) with its departure time $\tau$ , OD pair $w$ , and path $p$ ; $f = f(w, \tau, p)$

### Sets:

$N$	Set of nodes in the physical network
$E$	Set of road links in the physical network
$F$	Set of agents (vehicles)
$P(w, \tau)$	Path set of OD pair $w$ and departure time interval $\tau$

### Parameters:

$d(w, \tau)$	Number of vehicles for OD pair $w$ and departure time interval $\tau$
$\mathbf{d}$	Vector of time-dependent demands of all OD pairs
$v_F(l)$	Free-flow speed on link $l$ , which is also the speed limit on link $l$
$v_Q(l, t)$	Queueing speed (or speed in queue) on link $l$ at time interval $t$

### Variables:

$r(w, \tau, p)$	Number of vehicles on path $p$ of OD pair $w$ and departure time interval $\tau$
$q(l, t)$	Time-dependent link flow of link $l$ in time interval $t$
$\mathbf{q}$	Vector of link flows
$TT(f, l, t)$	Agent $f$ ’s time-dependent travel time on link $l$ in time interval $t$



<b>TT</b>	Vector of agents' path travel times
$EC(f)$	Total path emission or eco-cost of agent $f$
$EC(f, l)$	Emission or eco-cost of agent $f$ on link $l$
$EC(f, l, v(t))$	Instantaneous emission of agent $f$ with (time-dependent) speed $v$ on link $l$ in time interval $t$
$EC(f, l, t', t'')$	Emission of agent $f$ on link $l$ with entering time $t'$ and leaving time $t''$
<b>EC</b>	Vector of path emission cost of all agents
$TE$	Total system emission cost

## 2.1 Model formulation

Consider a road network  $G = (N, A)$  with a set of nodes  $N$  and a set of links  $E$ . Each link is denoted as a directed link  $l = (i, j)$  from upstream node  $i$  to downstream node  $j$ . In the ESODTA, all (green) travellers are assumed to behave cooperatively in their route choices to minimize the total emission. For simplicity and with no loss of generality, three basic assumptions are made for this model: (i) departure time choices are not considered and time-dependent O-D travel desires are assumed to be given; (ii) link attributes, such as free-flow travel time, bottleneck discharge rate, density-flow relationship, are given to perform a dynamic network loading process and generate time-dependent vehicular trajectories; (iii) the emission cost function is given to estimate vehicular emissions. With these assumptions, the proposed ESODTA aims to determine the time-varying agent (path) flows which minimize the total network emission.

The agent-based ESODTA is presented as follows

### Agent-based ESODTA

$$\text{Min}_{\mathbf{r} \in \Omega} TE = \sum_f EC(f) \quad (1)$$

$$\text{Subject to } [\mathbf{q}, \mathbf{TT}, \mathbf{EC}] = \text{DNLE}(\mathbf{d}) \quad (2)$$

$$EC(f) = \sum_{l \in p_f} EC(f, l) \quad (3)$$

$$\sum_{p \in P(w, \tau)} r(w, \tau, p) = d(w, \tau), \quad \forall w, \tau \quad (4)$$

$$r(w, \tau, p) \geq 0, \quad \forall w, \tau, p \in P(w, \tau) \quad (5)$$

The objective function, Eq.(1), minimizes the network-wide vehicular emission. Constraint (2) states that, given the time-dependent OD demand vector  $\mathbf{d}$ , the estimated path emissions of all agents,  $\mathbf{EC}$ , is obtained by a dynamic network loading (DNL) model for emission estimation,  $\text{DNLE}(\mathbf{d})$ . Constraint (3) describes that the time-varying path emission  $EC(f)$  of agent/vehicle  $f$  is assumed to be the sum of the emissions on its constituent links. Constraint (4) is the demand flow balance constraint for each OD pair  $w$  and each departure time interval  $\tau$ . Constraint (5) requires non-negative path flows. Note that the above formulation is a generic ESODTA model in which different emission cost functions,  $EC(f, l)$ , and DNL models,  $\text{DNLE}(\mathbf{d})$ , can be embedded for multi-scale emission optimization applications.

## 2.2 Dynamic network loading model for emission estimation

To enable multi-scale emission optimization applications, the ESODTA model relies on an effective DNL model ( $\text{DNLE}(\mathbf{d})$ ) for obtaining high-resolution emission estimations. Essentially, given the network (link and node) data and time-dependent OD demands, the DNL model generates a set of vehicles along with their attributes (e.g., OD, departure time, type and age) that are loaded to the network, and performs traffic simulation to evaluate time-dependent link speed and acceleration/deceleration profiles, which are used in the emission estimation function for computing emission costs. The major difference between the  $\text{DNLE}(\mathbf{d})$  model for the ESODTA and classical DNL models for general DTA applications is that the former is used to generate detailed vehicle trajectories and time-dependent link speeds, whereas the latter's major outputs are link travel times and (aggregated) flows. As the context of DTA is extended from travel time-based models to emission or eco-cost based applications, it is necessary for the  $\text{DNLE}(\mathbf{d})$  model to output time-dependent speeds for estimating vehicular emissions.

Technically, both microscopic and mesoscopic traffic flow models can be embedded in  $\text{DNLE}(\mathbf{d})$  to construct detailed time-dependent speed profiles for emission estimation. Microscopic models have been widely used to generate vehicle emissions estimates by evaluating driving speed and acceleration characteristics/profiles on a vehicle-by-vehicle and second-by-second basis (e.g., [Ahn et al., 2002](#); [Nam et al., 2002](#); [Stathopoulos and Noland, 2003](#)). Although microscopic traffic flow models, such as cellular automatic (CA) and car following models, are desirable for analyzing vehicular delays in congested conditions, microscopic simulation can be computationally intensive and typically requires a wide range of detailed geometric data and driving behavior parameters, which can

be difficult to calibrate, especially for the purpose of producing high-fidelity emissions estimates. To strike a balance between the model and computational complexities and the emission resolution, this research adopts the mesoscopic modeling approach.

Fig. 1 depicts the trajectories of a vehicle on a link generated by three different methods based on mesoscopic traffic simulation that can be applied to obtain time-dependent travel times. The link index  $l$  is omitted for clarity. Assume that there is an agent entering the link at time  $t_1$  and leaving the link at time  $t_4$ . The agent's trajectories described by the three methods are as follows. With Method 1 (point-queue model), the agent moves at free-flow speed  $v_F$  until it reaches the downstream node at time  $t_3$ , and then stops at this node until the capacity is available for discharging this agent at time  $t_4$  (e.g., point-queue model). With Method 2 (average cell-speed model), the agent moves at an average speed  $\bar{v}$  through the link (e.g., Abdul Aziz and Ukkusuri, 2012). With Method 3 (modified point-queue model), the agent moves at a free-flow speed  $v_F$  until reaches the back of the queue at time  $t_2$ , and then at a slower speed  $v_Q$  before the bottleneck (e.g., Lawson et al., 1997).

Table 1 summaries the travel times and travel speeds obtained by the three methods under free-flow and congested conditions. Although the three methods output the same link travel time for the agent, the other four measures and the vehicle trajectory differ significantly in the three methods. Of particular concern are the waiting time (or time in queue) and speed in queue of the vehicle, which are critical for estimating emissions. The average speed obtained using the first method cannot truly reflect the speed changes of vehicles, leading to inaccurate estimates of emission in general. The second method based on the point-queue model is able to approximate the delay of the vehicle at the bottleneck, but the time and speed in queue are not effectively represented by this method, which will underestimate the agent's emission on the link. On the other hand, Method 3 explicitly takes into account the physical queue in describing the vehicle trajectory under congested condition, resulting in a more reasonable estimate of emission for the vehicle.

Typically, the queueing speed  $v_Q$  is related to the bottleneck discharge capacity. Assume that the discharge capacity is constant and there is no queue-spillback effect, then the queueing speed is constant, which can be obtained by the flow-speed relationship curve. Consequently, the vehicular emission on this link is a linear function of the delay. This important property of link emission will be rigorously derived in Section 3. Moreover, based on Method 3, this study constructs the expanded space-time network to model the ESODTA.

If there are signal control strategies or incidents on the link, the link capacity and the queueing speed will be time-dependent. Spillback occurs when the street block is so congested that it has no space for any entering traffic. Under these two conditions, the link or bottleneck discharge capacity and queueing speed will be time-dependent, and the vehicular emission estimation will be much more complex, compared to the case with a constant discharge capacity. To deal with the more general but complex ESODTA problem with time-dependent link capacities, this study adopts the DNLE(d) model, which was developed by Zhou and Taylor (2014), in a column generation-based solution algorithm presented in Section 4. Zhang et al. (2013), Ma et al. (2014), and Doan and Ukkusuri (2015) offered in-depth discussions on the use of point queue, spatial queue, cell transmission and continuous-time double queue models in SODTA. Recently, Ma et al. (2015) further developed an innovative traffic emission pricing model for dynamic traffic networks with single destinations, in which a first-best dynamic emission pricing scheme is examined to systematically consider both portions of free-flow travel time and shockwave travel times.

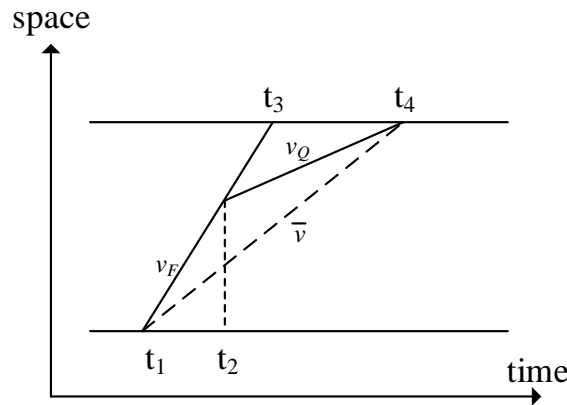


Fig. 1 Three mesoscopic methods for obtaining an agent's trajectory



Table 1 Travel speeds and travel times obtained by the three methods under free-flow and congestion

	Free-flow speed	Free-flow travel time	Speed in queue	Waiting time (time in queue)	Total travel time
Method 1	$v_F$	$t_3 - t_1$	0	$t_4 - t_3$	$t_4 - t_1$
Method 2	$\bar{v}$	$t_4 - t_1$	$\bar{v}$	0	$t_4 - t_1$
Method 3	$v_F$	$t_2 - t_1$	$v_Q$	$t_4 - t_2$	$t_4 - t_1$

### 2.3 Emission estimation and emission cost function

The link emission cost represents the impacts of a vector of vehicular emissions (such as air pollutants: CO, NO and HC and greenhouse gases CO<sub>2</sub>) on the environment and depends typically on the speed-based emission profile in a time-space cell. The total emission of agent  $f$  through link  $l$ ,  $EC(f, l, t', t'')$ , is an integration of the instantaneous emission cost  $EC(f, l, v(t))$  during the entrance time  $t'$  and the exit time  $t''$ .

$$EC(f, l, t', t'') = \int_{t'}^{t''} EC(f, l, v(t)) dt. \quad (6)$$

Typically, vehicular emission is a function of vehicle speed and acceleration, while correction factors can also be applied to the function to take into account different vehicle types, roadway characteristics, driving patterns and weather conditions.

The emission costs per unit time of vehicles traveling in free-flow speed and congested speed can be determined using a speed-based emission model (e.g., Szeto et al., 2012). For instance, the MOVES model (U.S. EPA, 2009) calculates the vehicle emission rates based on two factors: emission source (vehicle characteristics) and vehicle operating mode, where the latter is represented as vehicle specific power (VSP). The VSP is a function of vehicle speed, acceleration, and road grade, which account for kinetic energy, rolling resistance, aerodynamic drag, and gravity. An example of the VSP, given by U.S. EPA (2009), is as follows.

$$VSP(v, a) = \frac{A \cdot v + B \cdot v^2 + C \cdot v^3 + m \cdot v \cdot a}{m} \quad (7)$$

where  $v, a, m$  are the speed, acceleration and weight of vehicle  $f$ , respectively, and  $A$  is the rolling term,  $B$  the rotating term, and  $C$  the drag term. Then,  $VSP$  and  $v$  jointly determine the operating mode through a two-dimensional lookup function. The emission rate of a vehicle can be obtained as a function of the operating mode, vehicle type, link characteristics (e.g., grade) and other relevant factors. For example, Stein and Walker (2003) followed similar methodology to estimate regression models that correlate speed and CO emission from MOBILE 6.2. The CO emission rate (gm. of CO per vehicle per second) at the speed  $v$  (miles per hour) can be expressed as,

$$EC(v) = -0.064 + 0.0056v + 0.00026(v - 50)^2 \quad (8)$$

According to U.S. EPA, among the primary air pollutants, the top contributor that requires an air-quality standard is carbon monoxide (CO) (EPA, 2000). Further, air quality data for 2008 reported by the EPA identifies metropolitan areas exceeding the CO emission thresholds set by National Ambient Air Quality Standards (NAAQS) (Zhang et al., 2010; Hallmark et al., 2000). Thus, CO emission from on-road vehicles is a major issue that requires attention in transportation planning.

## 3. A linear integer programming model for the ESODTA

This section presents a linear integer programming model for the ESODTA with constant bottleneck discharge rates. We first discuss an important property of link emission for this case and describe the expanded space-time network, based on which the linear integer programming model for the ESODTA is developed. Then, we present the Lagrangian relaxation-based solution algorithm for solving the model.

### 3.1 Property of link emission

According to Newell's kinetic wave and car-following theory (Newell, 1993, 2002), a vehicle drives at its free-flow speed  $v_F$  under uncongested conditions. When a bottleneck forms at the end of the link, for keeping safe, the

vehicle will decelerate to a slower speed  $v_Q$  at the location  $d_Q$  (the back of the queue) and follow with the leading car. The time-space diagram, shown in Fig. 2, depicts a typical vehicle trajectory on a link with a single bottleneck at the downstream end. The solid line depicts the actual trajectory of vehicle  $f$  on link  $l$ , while the dotted line represents the desired trajectory of the vehicle through the bottleneck.

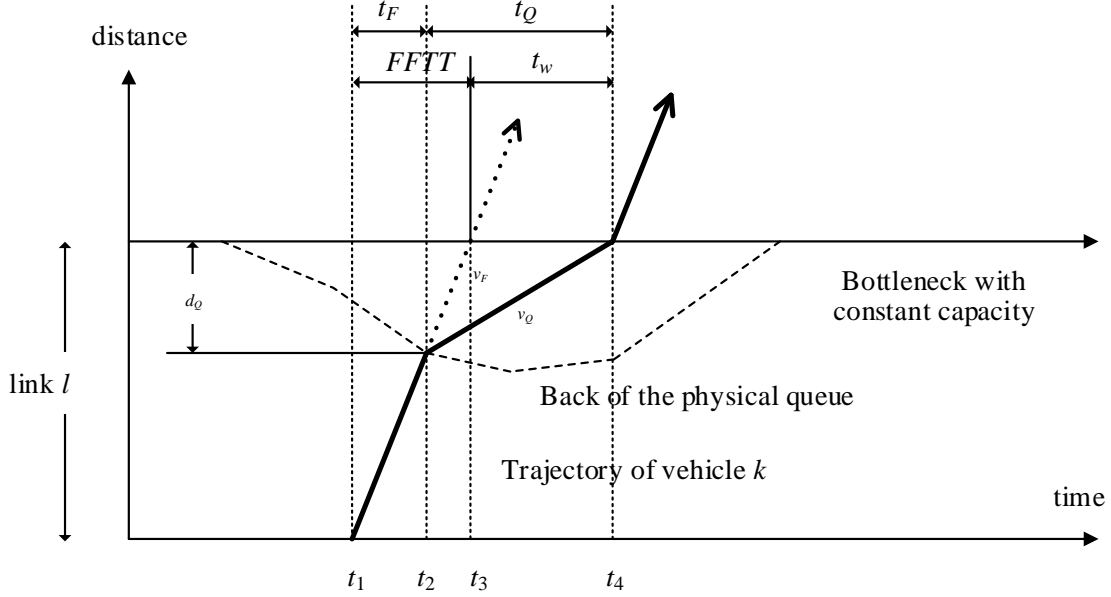


Fig. 2 Vehicle trajectory on a link with a bottleneck (adopted from Lawson et al., 1997)

The vehicle enters the link at time  $t_1$ , reaches the back of the queue (represented by the dashed curve) at time  $t_2$ , desires to leave the link at time  $t_3$ , and eventually leaves the link at time  $t_4$  due to capacity restriction. Here, we denote the travel time under free-flow speed as  $t_F = t_2 - t_1$ , the travel time under queueing speed (i.e., time in queue) as  $t_Q = t_4 - t_2$ ; the free-flow travel time  $FFTT = t_3 - t_1$ , and delay  $t_w = t_4 - t_3$ . The total travel time  $TT$  on link  $l$  can be described as follows:

$$TT = FF TT + t_w = t_F + t_Q. \quad (9)$$

It is obvious that the time in queue  $t_Q$  is greater than the delay  $t_w$ , because vehicles traveling at free-flow speed would naturally reach the back of the queue (which has physical length) before they would have reached the bottleneck with free-flow travel time  $FFTT$  (without any obstruction). Fig. 2 shows that the delay varies with the distance traveled in the vertical queue  $d_Q$ , as  $d_Q = t_Q \cdot v_Q = v_F \cdot (t_Q - t_w)$ , and the time in queue  $t_Q$  can be derived as follows:

$$t_Q = \frac{t_w \cdot v_F}{v_F - v_Q}. \quad (10)$$

Eq.(10) indicates that the time in the (physical) queue  $t_Q$  is a fixed multiple of the delay  $t_w$  (Lawson et al., 1997). According to Eq.(9) and Eq.(10), the free-flow travel time can also be derived as follows:

$$t_F = FF TT - \frac{t_w \cdot v_Q}{v_F - v_Q}. \quad (11)$$

With the assumption of constant discharge capacity, both  $v_F$  and  $v_Q$  are constant. The link emission of agent  $f$  entering and exiting link  $l$  at times  $t_1$  and  $t_4$ , respectively, consists of two parts: the emission when traveling at the free-flow speed  $v_F$  ( $EC(f, l, v_F)$ ) and the emission when traveling at the queueing speed  $v_Q$  ( $EC(f, l, v_Q)$ ), as follows.

$$EC(f, l, t_1, t_4) = EC(f, l, v_F) \cdot t_F + EC(f, l, v_Q) \cdot t_Q. \quad (12)$$

For clarity, the agent, link and time indices are omitted from the notations here. That is,  $EC_F = EC(f, l, v_F)$ , and  $EC_Q = EC(f, l, v_Q)$ , then Eq.(12) can be re-written as Eq.(13).

$$EC = EC_F \cdot t_F + EC_Q \cdot t_Q. \quad (13)$$

where  $EC_F$  and  $EC_Q$  can be determined using a function that describes the relationship between emission and speed, such as the emission function, Eq.(8).

**Proposition 1:** With a constant discharge rate, the vehicular emission can be expressed as a linear combination of the free-flow travel time  $FFTT$ , the emission at free-flow speed  $EC_F$  and the delay  $t_w$ .

**Proof:**

The total link emission  $EC = EC_F \cdot t_F + EC_Q \cdot t_Q$ , and according to Eqs.(10)-(13), we can obtain the total link emission as follows:

$$EC = \left( FFTT - \frac{t_w \cdot v_Q}{v_F - v_Q} \right) \cdot EC_F + \left( \frac{t_w \cdot v_F}{v_F - v_Q} \right) \cdot EC_Q. \quad (14)$$

Assume that  $\lambda = \frac{EC_Q}{EC_F} > 1$ , then total link emission can be re-written as a linear function of the delay  $t_w$ .

$$EC = EC_F \cdot \left( FFTT + t_w \cdot \frac{\lambda \cdot v_F - v_Q}{v_F - v_Q} \right) \quad (15)$$

This completes the proof.

It is also important to note that if  $EC_F = 1$  and  $\lambda = 1$  (i.e.,  $EC_Q = EC_F$ ), then  $\frac{\lambda \cdot v_F - v_Q}{v_F - v_Q} = 1$ , and the link travel time is a special case of the link emission.

$$EC = EC_F \cdot \left( FFTT + t_w \cdot \frac{v_F - v_Q}{v_F - v_Q} \right) = FFTT + t_w = TT. \quad (16)$$

### 3.2 Expanded space-time network

Additional notations used in presenting the expanded space-time network and the ESODTA model with constant bottleneck discharge rates are defined as follows.

**Sets:**

- $V$  Set of vertices in the expanded space-time network
- $A$  Set of arcs in expanded space-time network

**Parameters:**

- $O(f)$  Origin node of agent  $f$
- $D(f)$  Destination node of agent  $f$
- $DT(f)$  Departure time of agent  $f$
- $AT(f)$  Given assumed arrival time of agent  $f$  at the destination, and it is a given large value. The waiting cost on the destination node is 0.
- $s_{i,j}$  Free-flow travel time of link  $(i,j)$ , which is an integer multiple of one time interval
- $t_{i,j}$  Travel time (free-flow or congested travel time) of arc  $(i,j)$
- $EC_F(l)$  Unit emission cost under constant free-flow speed  $v_F(l)$  on link  $l$
- $EC_Q(l)$  Unit emission cost under constant queueing speed  $v_Q(l)$  on link  $l$
- $EC_{i,j,t,t'}$  Emission cost of an agent traveling on arc  $(i,j)$  with entrance time  $t$  and exit time  $t'$
- $Cap_{i,j}$  Outflow capacity of arc  $(i,j)$

**Variables:**

- $x_{i,j,t,t'}^f$  Binary decision variable indicating whether agent  $f$  travels on arc  $(i,j)$  with entrance time  $t$  and exit time  $t'$  ( $x_{i,j,t,t'}^f = 1$ ), or not ( $x_{i,j,t,t'}^f = 0$ ).

In order to obtain the  $v_Q$  on the congested links, the physical network needs to be modified to capture the queue based on the assigned link capacity. Three cases are considered in this study: (i) at the merge point, the assigned outflow capacities of upstream links are proportional to the downstream link's inflow capacity based on the number of lanes or the given upstream link's capacity values; (ii) at the diverge point, the assigned inflow capacities of downstream links is assumed to be the given outflow capacity of the upstream link; (iii) otherwise, the outflow capacity of the upstream link is equal to the inflow capacity of the downstream link. In case (i) and case (ii), virtual waiting nodes need to be created for the queue. One simple example is illustrated in Fig. 3 for the capacity allocation and virtual waiting nodes building.

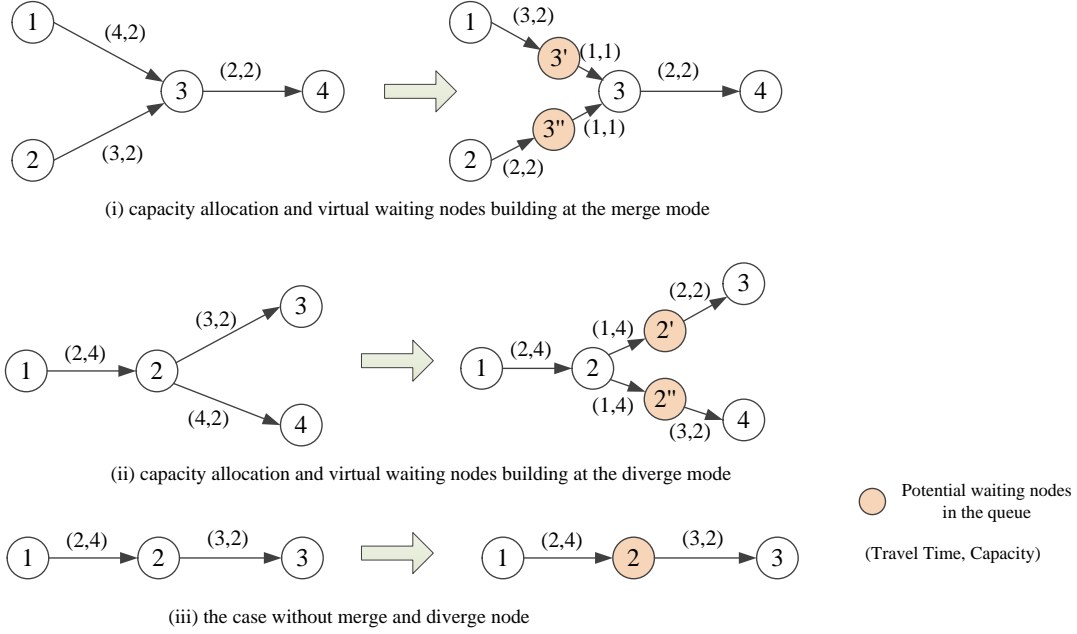


Fig. 3 Network modification for the capacity allocation and potential waiting nodes building

Based on the modified physical traffic network, the expanded space-time network is constructed for the ESODTA with constant bottleneck discharge rates. Consider a space-time network, where  $V$  is the set of vertices and  $A$  is the set of arcs. Note that the physical network is represented by nodes and links, while vertices and arcs are defined to illustrate the expanded space-time network in this paper. A virtual waiting node  $i$  (including all destination nodes) is extended to a set of vertices  $(i, t)$  at each time interval  $t$  in the optimization horizon,  $t = 1, 2, \dots, T$ . In addition, the origin nodes are not allowed for waiting. In the proposed space-time network representation, we consider two types of arcs as follows.

(1) **Traveling arcs:** A traveling arc represents the vehicle moves on a link  $(i, j)$  from time  $t$  to time  $t'$  in the physical network or from vertex  $(i, t)$  to vertex  $(j, t')$  in the space-time network, where  $t' - t = s_{i,j}$ . It is noted that the link free-flow travel time should be integer multipliers of one time interval.

(2) **Waiting arcs:** A waiting arc represents the vehicle waits at a virtual waiting node  $j$  (i.e., the queue) for a time interval; that is, this arc is incident from vertex  $(j, t)$  and incident to vertex  $(j, t + 1)$ . The arc travel time is 1 and arc capacity is set as a large constant,  $M$ . In addition, a large arrive time  $AT(f)$  at the destination  $D(f)$  for agent  $f$  will be assumed and the waiting cost is 0 on the destination nodes.

Fig. 4(a) depicts a simple physical network with three nodes, two links, one O-D pair (1, 3), and time-dependent travel demand input. By using dynamic network modeling processed described in Tong et al. (2015) and Li et al. (2015), the corresponding expanded space-time network is displayed in Fig. 4(b).

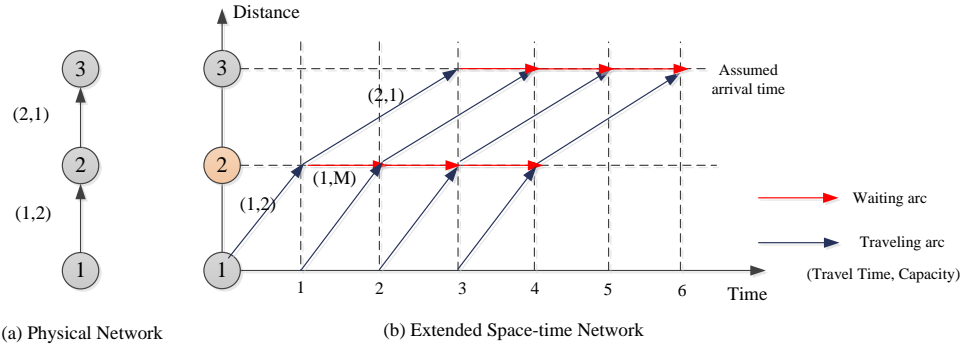


Fig. 4 A simple network and its corresponding space-time network

### 3.3 ESODTA mathematical formulation

Based on the expanded space-time network, the ESODTA problem with constant bottleneck discharge rates (ESODTA-constant discharge rates) can be formulated as follows.

$$\text{Minimize } Z = \sum_f \sum_{(i,j,t,t') \in A} EC_{i,j,t,t'} \times x_{i,j,t,t'}^f \quad (17)$$

Subject to

$$\sum_{(i,j,t-t_{i,j,t}) \in A} x_{i,j,t-t_{i,j,t}}^f - \sum_{(j,k,t,t+t_{j,k}) \in A} x_{j,k,t,t+t_{j,k}}^f = \begin{cases} -1 & \forall f \in F, j = O(f) \text{ and } t = DT(f) \\ 1 & \forall f \in F, j = D(f) \text{ and } t = T \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

$$\sum_{f \in F} x_{i,j,t,t+t_{i,j}}^f \leq Cap_{i,j}, \quad \forall (i,j,t,t+t_{i,j}) \in A, t = 1, 2, \dots, T \quad (19)$$

$$x_{i,j,t,t'}^f = \{0, 1\} \quad (20)$$

In the above formulation, the objective is to minimize the total vehicular emission shown as Eq.(17). The flow balance constraints are presented as Eq.(18). Eq.(19) describes the outflow capacity constraints on traveling arcs. The constraints shown as Eq.(20) requires all decision variables to be binary.

The emission costs  $EC_{i,j,t,t'}$  in the objective function are discussed as follows. According to Proposition 1 in Section 3.1, the link emission cost is a function of the delay  $t_w(i,j)$  on link  $(i,j)$ :

$$EC(i,j) = EC_F(i,j) \times \left( s_{i,j} + t_w(i,j) \times \frac{\lambda \times v_F - v_Q}{v_F - v_Q} \right) = \alpha(i,j) + \beta(i,j) \times t_w(i,j) \quad (21)$$

where  $\alpha(i,j) = EC_F(i,j) \times s_{i,j}$  is the free-flow emission cost on link  $(i,j)$ , and  $\beta(i,j) = EC_F(i,j) \times \frac{\lambda \times v_F - v_Q}{v_F - v_Q}$  is the queueing emission cost on the waiting arc of link  $(i,j)$  for one time interval. Thus, according to the design of the expanded space-time network, the link emission of one vehicle at time  $t$  consists of two parts: (i) the emission at free-flow speed on the traveling arc before reaching the virtual waiting node or bottleneck, where the emission cost is  $EC_{i,j,t,t+s_{i,j}} = EC_F(i,j) \times s_{i,j}$ ; (ii) the emission at queueing speed on waiting arcs (waiting at the virtual node) for  $t_w(i,j)$  time intervals, where the emission cost of one waiting arc is  $EC_{j,j,t,t+1} = \beta(i,j)$ .

### 3.4. Lagrangian relaxation-based solution algorithm

This subsection presents the Lagrangian relaxation-based algorithm for solving the ESODTA model with constant bottleneck discharge rates. In this algorithm, the capacity constraints, shown in Eq.(19), are dualized to the objective function with the multipliers  $\mu_{i,j,t,t'}$ ,  $\forall (i,j,t,t')$ , as follows.

$$\text{Min } Z^{LR} = \sum_f \sum_{(i,j,t,t')} (EC_{i,j,t,t'} + \mu_{i,j,t,t'}) \times x_{i,j,t,t'}^f - \sum_{(i,j,t,t')} \mu_{i,j,t,t'} \times Cap_{i,j} \quad (22)$$

Then, the relaxation problem is to find the time-dependent least-cost path problem for each agent in the expanded space-time network. The Lagrangian relaxation-based algorithm is described as follows:

Step 1: Initialization. Let iteration  $n = 1$ , initialize the multipliers  $\mu_{i,j,t,t'}^n = 0$ .

Step 2: Solve the relaxation problem for each agent as the time-dependent least-cost path problem.

Step 3: Update the Lagrangian multipliers as follows:

The step-size updating is based on the method of successive averages (MSA):  $\gamma^n = \frac{1}{n+1}$

$$\mu_{i,j}^{n+1}(t) = \max\{0, \mu_{i,j}^n(t) + \gamma^n \times (\sum_f x_{i,j,t,t+s_{i,j}}^f - Cap_{i,j})\} \quad (23)$$

Step 4: Check termination condition: if  $n < N_{max}$ , then  $n = n + 1$  and return to step 2. Otherwise, stop the algorithm.

## 4. ESODTA with time-dependent bottleneck capacities

The ESODTA model, presented in Section 3.2, assumes constant bottleneck discharge rates, but bottleneck capacities are generally time-dependent in the presence of signal controls at intersections or incidents on links, which result in time-dependent queueing speeds. Moreover, if a downstream link is so congested that the queue spills back to the current link, then there is no space for entering traffic. When link or bottleneck capacities and queueing speeds are time-dependent, the vehicular emission estimation will be much more complex, compared to the

case with constant discharge capacities. To tackle this difficulty, this research adopts the dynamic network loading model, DNLE(**d**) developed by Zhou et al. (2015a), to estimate link and path emissions when bottleneck capacities are time-dependent. The DNLE(**d**) combines both macroscopic and microscopic traffic descriptions based on Newell's simplified kinematic wave model and simplified car-following model. Specifically, Newell's simplified kinematic wave model is employed in the dynamic mesoscopic traffic simulation package, DTALite, which outputs link arrival and departure times for each vehicle. Given the link arrival and departure times of individual vehicles, we adopt Newell's simplified linear car following model (Newell, 2002) to reconstruct the detailed vehicle trajectories which can be used to derive second-by-second vehicle speeds and accelerations. Then, the time-dependent speeds and accelerations are input to the MOVES model to generate vehicular emissions (see the description in Section 2.3).

The column generation-based ESODTA algorithm embeds: (i) the above-mentioned DNLE(**d**) to evaluate link and path emissions for the set of agents with assigned paths, and (ii) the gradient projection-based method to update the path assignment of the agents. Additional notations used in this section are defined as follows.

#### Sets or vectors:

$P_m(w, \tau)$  Set of paths in iteration  $m$  for the agents of OD pair  $w$  and departure time  $\tau$   
 $P_m$  Set of paths in iteration  $m$  for all of the agents in the network;  $P_m = \{P_m(w, \tau), \forall w, \tau\}$

#### Indices:

$n$  index of inner loop iterations in the column generation-based algorithm  
 $m$  index of outer loop iterations in the column generation-based algorithm

#### Parameters:

$M_{max}$  Maximum number of outer loop iterations  
 $N_{max}$  Maximum number of inner loop iterations

#### Variables:

$r(w, \tau, p)^n$  Number of agents on path  $p$  of OD pair  $w$  and departure time  $\tau$  in iteration  $n$   
 $\mathbf{r}^n$  Path flow vectors (a feasible solution) in iteration  $n$ ;  $\mathbf{r}^n = \{r(w, \tau, p)^n, \forall w, \tau, p\}$   
 $EC(w, \tau, p)^n$  Total emission of the agents on path  $p$  of OD pair  $w$  and departure time  $\tau$  in iteration  $n$   
 $TE_w^r(\mathbf{r}^n)$  Total emission of the agents of OD pair  $w$  and departure time  $\tau$ , evaluated at the feasible solution  $\mathbf{r}^n$   
 $TE(\mathbf{r}^n)$  Total system emission, evaluated at the feasible solution  $\mathbf{r}^n$

## 4.1 Solution algorithm

### 4.1.1 Column generation-based algorithmic framework

The column generation-based algorithm generates time-dependent least marginal emission paths as needed in the outer loop and solves a reduced (or restricted) ESODTA problem in the inner loop (e.g., Lu et al., 2009). The column generation-based approach operates as follows (see Fig. 5). In each outer loop iteration  $m$ , the time-dependent least-cost path algorithm, developed by Ziliaskopoulos and Mahmassani (1993), is applied to find the time-dependent least marginal emission path for each O-D pair and each departure time interval. New paths, if any, are added to augment the current subset of feasible paths,  $P_m$ , in iteration  $m$ . A gradient projection-based descent direction method, presented in Section 4.1.2, is then used to solve the reduced ESODTA problem defined on  $P_m$ . The algorithm terminates and outputs time-varying path flows obtained in the current iteration, if no new path is found or a preset convergence criterion is satisfied.

The gradient projection-based descent direction method proceeds iteratively and forms the inner loop in the column generation-based algorithmic framework. In each inner loop iteration  $n$ , the updated path flows  $\mathbf{r}^n$  and the corresponding total system emission  $TE(\mathbf{r}^n)$  and link marginal emissions are evaluated by the DNLE(**d**) model. If the difference between the objective values in two successive iterations (i.e.,  $TE(\mathbf{r}^n) - TE(\mathbf{r}^{n-1})$ ) is less than a preset threshold or a preset convergence criterion (e.g.,  $n = N_{max}$ ) is satisfied, the inner loop terminates and the algorithm returns to the outer loop.



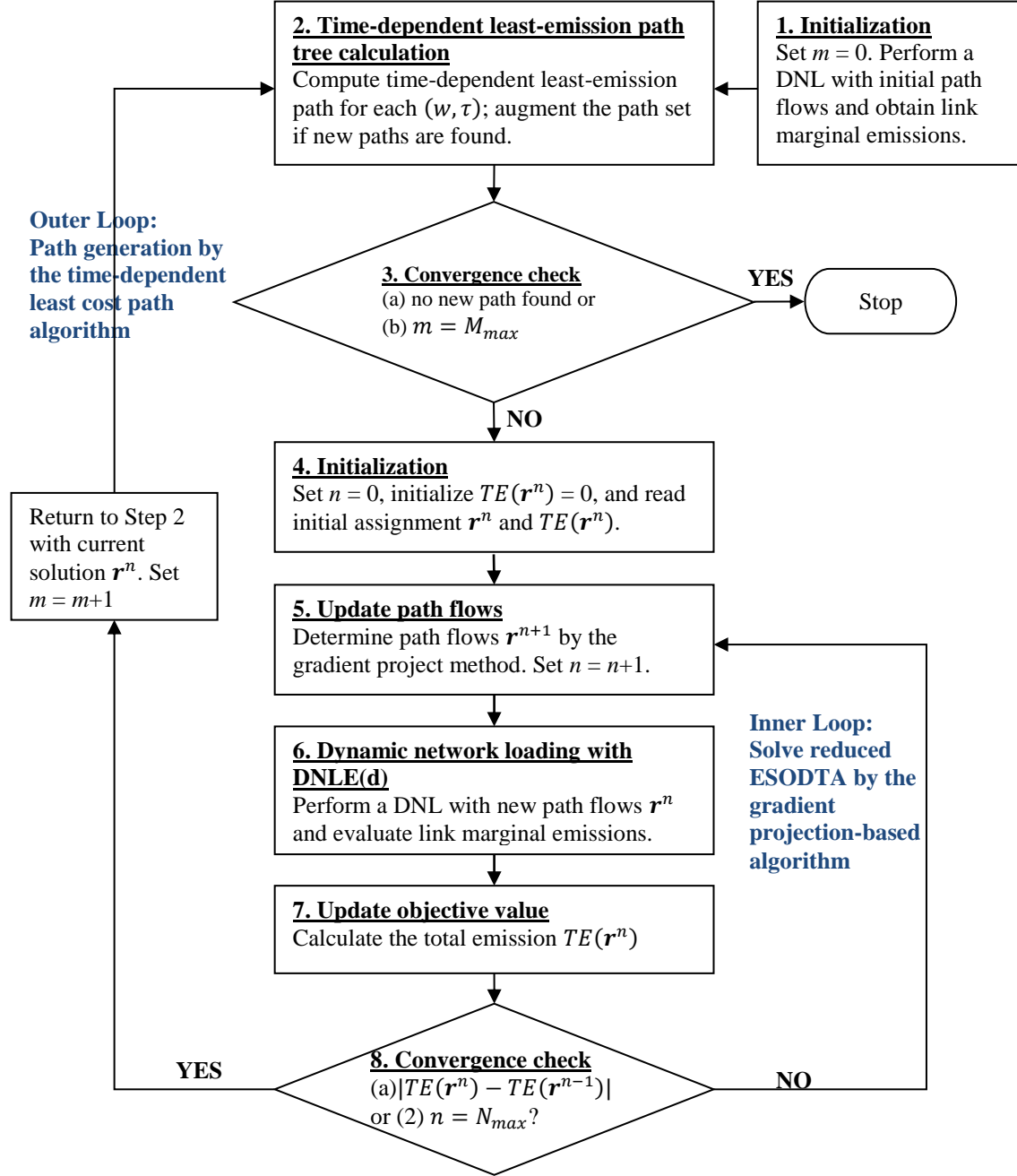


Fig. 5 Flow chart of the column generation-based ESODTA algorithm

#### 4.1.2 Gradient projection-based descent direction method

The reduced ESODTA problem, defined by a subset of feasible paths  $P_m$  in outer loop iteration  $m$ , is solved by the gradient projection-based descent direction method in the inner loop to obtain least-emission path flows on the existing paths. With a feasible solution  $\mathbf{r}^n$  in inner loop iteration  $n$ , the method adopts a search direction along the feasible descent direction based on the gradient,  $\nabla TE(\mathbf{r}^n)$ :

$$\mathbf{r}^{n+1} = Proj_{\Omega}[\mathbf{r}^n - s^n \times \nabla TE(\mathbf{r}^n)], \quad (24)$$

where  $s^n \in (0, 1)$  is the step size in inner loop iteration  $n$ .  $Proj_{\Omega}[u]$  denotes the unique projection of path flow vector  $u$  onto the feasible space  $\Omega$  and is defined as the unique solution of the problem:  $\text{Min}_{y \in \Omega} \|u - v\|$ . Accordingly, the new iterate  $\mathbf{r}^{n+1}$  is obtained by updating the current iterate  $\mathbf{r}^n$  along the direction  $-\nabla TE(\mathbf{r}^n)$  with a move size  $s^n$ .

To facilitate solving the ESODTA problem on large networks with multiple O-D pairs, the proposed method decomposes the original problem into many sub-problems, each of which corresponds to a  $(w, \tau)$  pair, by assuming that cross-network marginal effects are negligible (e.g., [Zhang et al., 2009](#)). Let  $p^*$  be the referenced least marginal emission path of a  $(w, \tau)$  pair. Given a feasible solution  $\mathbf{r}^n$ , the flow balance conservation constraints [Eq.\(4\)](#) can be rearranged as follows:

$$r(w, \tau, p^*)^n = d(w, \tau) - \sum_{p \in P(w, \tau) \setminus p^*} r(w, \tau, p)^n, \forall w, \tau. \quad (25)$$

Then, with this rearrangement, the objective function corresponding to a pair  $(w, \tau)$  can be written as follows:

$$\begin{aligned} TE_w^\tau(\mathbf{r}^n) &= \sum_{p \in P(w, \tau) \setminus p^*} r(w, \tau, p)^n \times EC(w, \tau, p)^n + r(w, \tau, p^*)^n \times EC(w, \tau, p^*)^n \\ &= \sum_{p \in P(w, \tau) \setminus p^*} r(w, \tau, p)^n [EC(w, \tau, p)^n - EC(w, \tau, p^*)^n] + d(w, \tau) \times EC(w, \tau, p^*)^n. \end{aligned} \quad (26)$$

The first-order partial derivative of  $TE_w^\tau(\mathbf{r}^n)$  with respect to a particular path flow  $r(w, \tau, p)$  is

$$\begin{aligned} \nabla TE_{wp}^\tau(\mathbf{r}^n) &= \partial TE_w^\tau(\mathbf{r}^n) / \partial r(w, \tau, p) \\ &= EC(w, \tau, p)^n - EC(w, \tau, p^*)^n + \sum_{p' \in P(w, \tau) \setminus p^*} \left[ r(w, \tau, p')^n (\eta_{wp'}^{\tau, n} - \eta_{wp^*}^{\tau, n}) \right] + d(w, \tau) \times \eta_{wp^*}^{\tau, n}. \end{aligned} \quad (27)$$

where  $\eta_{wp}^{\tau, n} = \partial EC(w, \tau, p)^n / \partial r(w, \tau, p)$  denotes the path marginal emission which represents the change in path emission due to an additional unit of the path inflow,  $r(w, \tau, p)$ . Note that if within-path-set marginal effects are also ignored, then

$$\nabla TE_{wp}^\tau(\mathbf{r}^n) = EC(w, \tau, p)^n - EC(w, \tau, p^*)^n + r(w, \tau, p)^n [\eta_{wp}^{\tau, n} - \eta_{wp^*}^{\tau, n}] + d(w, \tau) \times \eta_{wp^*}^{\tau, n}. \quad (28)$$

Regarding the step size  $s^n$ , this study adopts a scheme of mixed step sizes as follows.

$$s^n = 1/m, \text{ if } n = 0; s^n = 1, \text{ otherwise.} \quad (29)$$

Recall that  $m$  is the (outer loop) iteration counter. The decreasing step size  $s^n = 1/m$  follows the MSA. Based on [Eqs.\(24\), \(27\), and \(29\)](#), the gradient projection-based descent direction method derives the following path flow updating scheme in the inner loop of the algorithmic framework.

$$\begin{aligned} r(w, \tau, p)^{n+1} &= \text{Max}\{0, r(w, \tau, p)^n - s^n \times \nabla TE_{wp}^\tau(\mathbf{r}^n)\}, \forall p \in P(w, \tau) \setminus p^*, \\ r(w, \tau, p^*)^{n+1} &= d(w, \tau) - \sum_{p \in P(w, \tau) \setminus p^*} r(w, \tau, p)^{n+1}. \end{aligned} \quad (30)$$

## 4.2 Marginal emission evaluation

The gradient projection-based method requires evaluation of path marginal emissions,  $\eta_{wp}^\tau, \forall w, \tau, p$ , and their constituent link marginal emissions  $\eta_l^t, \forall l, t$ . This subsection presents the evaluation of link and path marginal emissions. We define the additional notations used in this section as follows.

$w$	backward wave speed
$k_{jam}$	jam density
$length(l)$	length of link $l$
$nlanes(l)$	number of lanes on link $l$
$A(l, t)$	cumulative number of vehicles that have arrived at link $l$ at time $t$
$V(l, t)$	cumulative number of vehicles that have waited at the vertical queue of link $l$ at time $t$
$D(l, t)$	cumulative number of vehicles that have departed from link $l$ at time $t$
$q^{max}(l, t)$	maximum flow rate on link $l$ at time $t$
$cap^{in}(l, t)$	inflow capacity of link $l$ at time $t$
$cap^{out}(l, t)$	outflow capacity of link $l$ at time $t$
$FFTT(l)$	free-flow travel time on link $l$ ; i.e., $length(l)/v_f$
$BWTT(l)$	backward wave travel time on link $l$ ; i.e., $length(l)/w$

#### 4.2.1 Link marginal emissions

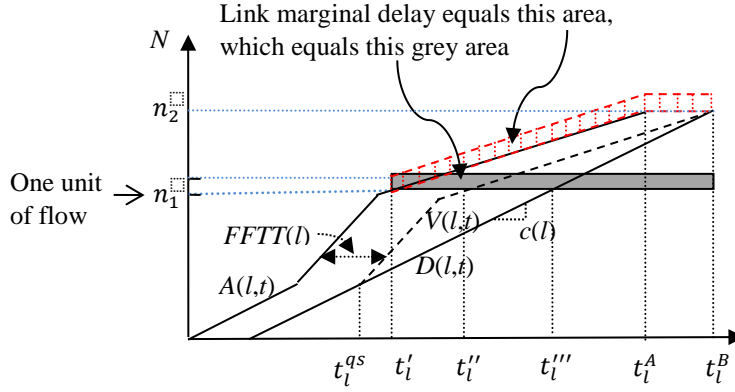


Fig. 6 Illustration of link marginal travel time on a congested link

Fig. 6 depicts the cumulative arrival  $A(l, t)$ , virtual arrival  $V(l, t)$ , and departure curves  $D(l, t)$  for a congested link  $l$ . The (outflow) capacity of the link is  $c(l)$ . The queue starts at  $t_l^{qs}$  and dissipates at  $t_l^B$  on the link. Let  $t_l'$ ,  $t_l''$ , and  $t_l'''$  be the times when an additional vehicle ( $n_1$ ) arrives at link  $l$ , joins the queue of the bottleneck, and leaves the link, respectively. According to Ghali and Smith (1995), the link marginal travel time due to the additional vehicle is equal to the gray area (i.e.,  $mTT(l, t_l') = t_l^B - t_l'$ ), which includes the travel time of the additional vehicle ( $t_l''' - t_l'$ ) and its impact on the travel times of the vehicles behind the unit of flow ( $t_l^B - t_l'''$ ). Note that if this additional vehicle does not encounter a queue, the link marginal travel time equals  $FFTT(l) = t_l'' - t_l'$ . The vehicles arriving between  $t_l'$  and  $t_l^A$  experience the additional travel time  $1/c(l)$ , because it takes  $1/c(l)$  to discharge this perturbation vehicle. In our notation, this means that the change in the delay  $\Delta t_w = 1/c(l)$ . The impact on the travel times of the impacted vehicles arriving between  $t_l'$  and  $t_l^A$  can be derived from Fig. 6 as follows.

$$mTT(l, t_l') = t_l^B - t_l''' = (n_2 - n_1) \frac{1}{c(l)}. \quad (31)$$

Similarly, the link marginal emission can be divided into two parts: the emission of the additional vehicle and its impact on the emissions of the following vehicles (or the change in emission for the impacted vehicles). Let  $\Delta EC(l, t)$  be the change in emission for one vehicle entering the link behind the unit perturbation vehicle. The following proposition derives the change in emission for the impacted vehicles,  $mEC(l, t)$ , and its relationship with the change in travel time for the impacted vehicles,  $mTT(l, t)$ .

**Proposition 2:**  $mEC(l, t) = (n_2 - n_1)\Delta EC(l, t)$ , which is a multiple of the change in travel time for the impacted vehicles:  $mEC(l, t) = \gamma(l) \times mTT(l, t)$ .

**Proof:**

Since the change in travel time is the sum of the change in the time spent in queue and the change in the actual free-flow travel time (i.e.,  $\Delta t_w(l, t) = \Delta t_Q(l, t) + \Delta t_F(l, t)$ ) and  $\Delta t_w(l, t) = 1/c(l)$ ,

$$\Delta t_F(l, t) = \frac{1}{c(l)} - \Delta t_Q(l, t), \quad (32)$$

According to Eq.(10), the change in the time spent in queue,  $\Delta t_Q(l, t)$ , is a fixed multiple of the change in the travel time,  $\Delta t_w(l, t)$ :

$$\Delta t_Q(l, t) = \frac{\Delta t_w(l, t)}{1 - \frac{v_Q(l)}{v_F(l)}} = \frac{1/c(l)}{1 - \frac{v_Q(l)}{v_F(l)}}. \quad (33)$$

Then, according to Eq.(12), the change in emission for one vehicle entering the link behind the unit perturbation vehicle is as follows.

$$\begin{aligned} \Delta EC(l, t) &= \Delta t_Q(l, t)EC(v_Q(l)) + \Delta t_F(l, t)EC(v_F(l)) \\ &= \frac{1}{c(l)} \left[ \frac{1}{1 - \frac{v_Q(l)}{v_F(l)}} EC(v_Q(l)) + \left(1 - \frac{1}{1 - \frac{v_Q(l)}{v_F(l)}}\right) EC(v_F(l)) \right]. \end{aligned} \quad (34)$$

Based on Eq.(31), the change in emission for the impacted vehicles can be derived as follows.

$$mEC(l, t) = (n_2 - n_1)\Delta EC(l, t). \quad (35)$$

Let  $\gamma(l) = \left[ \frac{1}{1 - \frac{v_Q(l)}{v_F(l)}} EC(v_Q(l)) + (1 - \frac{1}{1 - \frac{v_Q(l)}{v_F(l)}}) EC(v_F(l)) \right]$ . Then,  $mEC(l, t) = (n_2 - n_1) \frac{\gamma(l)}{c(l)} = \gamma(l) \times mTT(l, t)$ .

This completes the proof.

The link marginal emission  $\eta_l^t$  is the sum of the emission of the additional vehicle and  $mEC(l, t)$ . Note that in Eq.(35)  $\gamma(l)$  is a constant multiplier, because  $v_F(l)$  and  $v_Q(l)$  are given constants, and  $EC(v_F(l))$  and  $EC(v_Q(l))$  can be determined according to the emission function, such as Eq.(8). The queueing speed  $v_Q(l)$  of link  $l$  is determined mainly based on the bottleneck discharge capacity. For instance, Lu et al. (2013) discussed a method for the outflow capacity at merge or diverge junctions.

#### 4.2.2 Evaluation of path marginal emissions

Evaluating path marginal emissions in dynamic and congested traffic networks requires explicitly tracing the perturbation propagation of an additional unit of inflow along a path. This issue has also been recognized by Shen et al. (2007), Qian and Zhang (2011) and Lu et al. (2013) on evaluating path marginal delays (or travel times).

Consider a freeway or an arterial segment with two sequential links without merges and diverges, link  $l$  and link  $l+1$ . Under congested conditions, there are three basic cases of interest, when the additional unit of vehicle arrives at this segment at time  $t'_l (= \tau, \text{the departure time})$ .

(i) There is a bottleneck on the downstream link  $l+1$  and the queue on link  $l+1$  does not spill back to link  $l$ ; link  $l$  is in free-flow condition while link  $l+1$  is partially congested (Fig. 7(a)).

(ii) There is a bottleneck on each of the two links, and the two bottlenecks are independent (e.g., link  $l+1$  is sufficiently long so that the queue in the downstream does not spill back to the upstream). This is in fact the case in which both links are partially congested (Fig. 7(b)).

(iii) There is a bottleneck on the downstream link  $l+1$  and the queue on link  $l+1$  spills back to link  $l$ ; that is, link  $l$  is partially congested while link  $l$  is fully congested (Fig. 8).

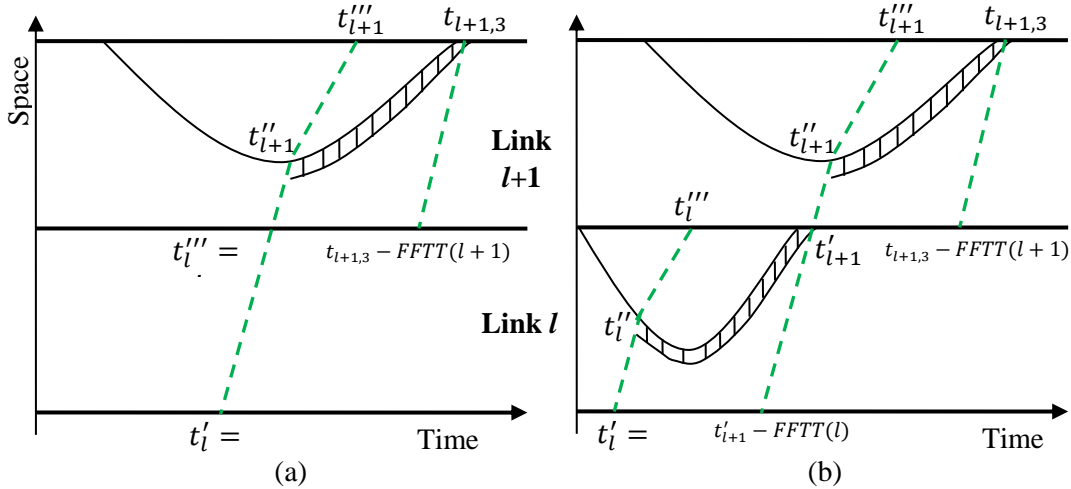


Fig. 7 Path marginal emission analysis without queue spillback

In case (i), the first link is not impacted by that additional vehicle, while the second link's impacted regime is from  $t'_{l+1}$  to  $t_{l+1,3} - FFTT(l+1)$ . In case (ii), the first link has an impacted regime spanning a time period from  $t'_l$  to  $t'_{l+1} - FFTT(l)$ , and the second link's impacted regime is from  $t'_{l+1}$  to  $t_{l+1,3} - FFTT(l+1)$ . Note that in case (ii),  $t'_{l+1} = t_{l,3}$ ; that is, the perturbation propagates to link  $l+1$  when the queue on link  $l$  vanishes at time  $t_{l,3}$ .

In case (iii), if the additional vehicle does not encounter the queue on link  $l$  (Fig. 8(a)), then the link marginal emission on this link is zero. The perturbation then moves to link  $l+1$  with an impacted period from  $t'_{l+1}$  to  $t_{l+1,1}$ . After detecting the next nearest event timestamp,  $t_{ne} = t_{l+1,1}$ , which corresponds to a queue spillback event, we need to trace back to link  $l$  to take time period  $[t_{l+1,1} - FFTT(l), t_{l+1,2} - FFTT(l)]$  into consideration. Finally, we move to link  $l+1$  to cover the last impacted regime from  $t_{l+1,2}$  to  $t_{l+1,3} - FFTT(l+1)$ . Thus, the proposed

algorithm for evaluating path marginal emissions needs to incorporate a backtracking mechanism, in order to explicitly consider the difference pieces of the impacted regime over multiple links.

If the additional vehicle encounters the queue on link  $l$  (Fig. 8(b)), the first link's impacted regime spans a time period from  $t'_l$  to  $t_{l,3} - FFTT(l)$ , and the second link's impacted regime is from  $t_{l,3}$  to  $t_{l+1,3} - FFTT(l+1)$ . Again, the perturbation can enter link  $l+1$  at time  $t_{l+1,2}$  only when the queue on link  $l$  vanishes at time  $t_{l,3}$ .

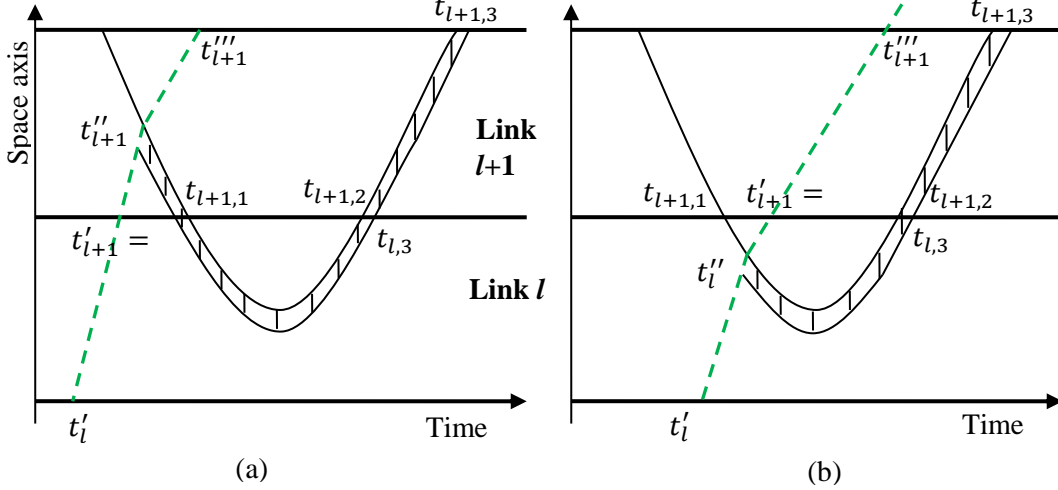


Fig. 8 Path marginal emission analysis with queue spillback from downstream link

Based on the above analysis for the impacted regimes on two consecutive links due to an additional vehicle, this research proposes a method for evaluating the path marginal emission of a path  $p \in P(w, \tau)$  with multiple links  $l = 1, \dots, L$ . The path marginal emission due to an additional vehicle consists of the path emission of that vehicle and additional emissions generated by impacted vehicles. Starting from the first link  $l=1$  and given departure time  $\tau$ , the proposed algorithm, presented in Algorithm 1, keeps accumulating path marginal emission by adding the link marginal emission obtained using Proposition 2 for each impacted period and traces the perturbation propagation based on the next nearest event timestamp which is used to guide the evaluation procedure advancing to next link (or impacted period) or returning to last link in the queue spillback case.

**Algorithm 1:** Evaluation of the path marginal emission for a triplet  $(p, w, \tau)$

Initialize  $t' = \tau$ ,  $l = 1$ , and  $\eta_{wp}^\tau = ec_{wp}^\tau$ , where  $ec_{wp}^\tau$  denotes the emission of path  $p \in P(w, \tau)$ .

**Do while** link index  $l \leq L$  ( $L$  is the number of links)

**Step 1:** Obtain the next (nearest) event timestamp,  $t_{ne}$ . The next event may correspond to the end time of congestion (i.e.,  $t_3$ ), the beginning of queue spillback (i.e.,  $t_1$ ) or  $t_{ne} = t' + FFTT(l)$  (under uncongested conditions).

**Step 2:** Evaluate link marginal emission,  $\eta_l^{t'}$ , using Proposition 2.

**Step 3:** Accumulate the path marginal emission  $\eta_{wp}^\tau = \eta_{wp}^\tau + \eta_l^{t'}$ .

**Step 4:** Move to the next link, and update the starting time,  $t'$ , of next analysis period. If next event timestamp  $t_{ne}$  corresponds to the end time of congestion or is in a uncongested time period, then  $t' = t_{ne} + FFTT(l)$  and  $l = l + 1$ ; otherwise ( $t_{ne}$  corresponds to a queue spillback case)  $l = l - 1$  and  $t' = t_{ne} - FFTT(l)$ .

**End**

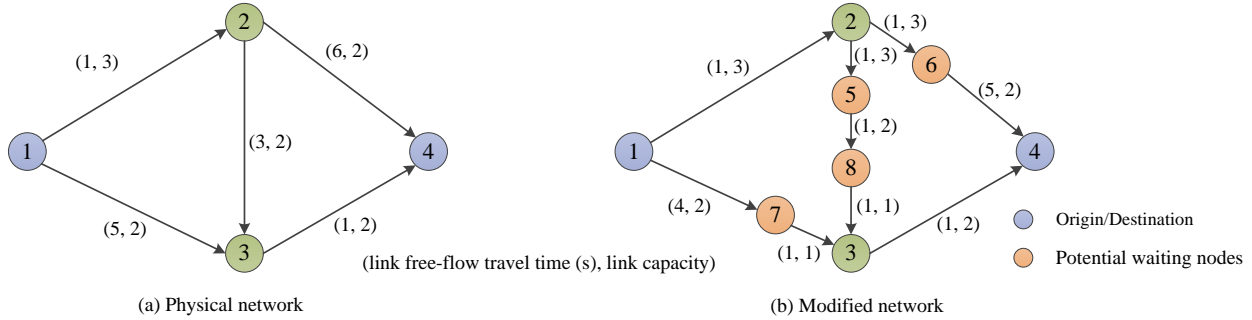
## 5. Numerical Experiments

This section presents two numerical examples. The first one demonstrates the ESODTA model with constant bottleneck capacities, presented in Section 3, and the second one illustrates the solution algorithm for the ESODTA with time-dependent bottleneck capacities, presented in Section 4. All of the numerical tests are conducted on a Dell Precision T7610 Workstation with 20 CPU cores (40 computational threads) and 192G RAM.

## 5.1 Numerical example for the ESODTA with constant bottleneck capacities

For illustrative purpose, one simple network is created with 4 nodes, 5 links, and one OD pair (1,4), as shown in Fig. 9(a). The link free-flow travel time and link capacity of each link are displayed in the parenthesis near that link. The time-dependent travel demands from node 1 to node 4 are 5 agents, 4 agents, and 1 agent departing at time 0, 1, and 2, respectively. Based on the description in Section 3.2, the traffic network is modified and shown in Fig. 9(b) by considering the merge and diverge junctions to better represent travel delay for emission calculation. The corresponding expanded space-time network is depicted in Fig. 9 (c). All of the links are assumed to have the same free-flow speed value of 60 mph. The fundamental q-k diagrams under different link capacities are shown in Fig.10. Based on the  $v_f$  and  $v_q$  on each link and the emission cost function in Eq.(8), we can calculate the link emission cost at free-flow and delay conditions using Eq.(21).

The ESODTA model formulated based on the expanded space-time network (Section 3.3) is solved using the commercial solver, GAMS. The Lagrangian relaxation approach, presented in Section 3.4, is also implemented to obtain a lower bound. The comparison between the optimal solution and the lower bound is shown in Fig. 11. The final gap after 50 iteration is about 8%. In addition, we solve the travel time-based SODTA problem, in which the link emission cost in the objective function Eq.(17) is replaced by the link travel time. Table 2 lists the time-dependent arc flows in the ESODTA and the travel time-based SODTA models. The optimal total travel time and emission of the ESODTA are 67 seconds and 1095.72gm, respectively, while they are 63 seconds and 1104.84 gm, respectively, in the travel time-based SODTA model. The emission reduction in ESODTA is gained at the expense of an increased total travel time. In practice, the decision maker chooses between the ESODTA and the travel time-based SODTA depending on his or her goal. The related source code can be downloaded at the ResearchGate website:[https://www.researchgate.net/publication/295856604\\_Eco-system\\_optimal\\_time-dependent\\_flow\\_assignment\\_in\\_a\\_congested\\_network#share](https://www.researchgate.net/publication/295856604_Eco-system_optimal_time-dependent_flow_assignment_in_a_congested_network#share)





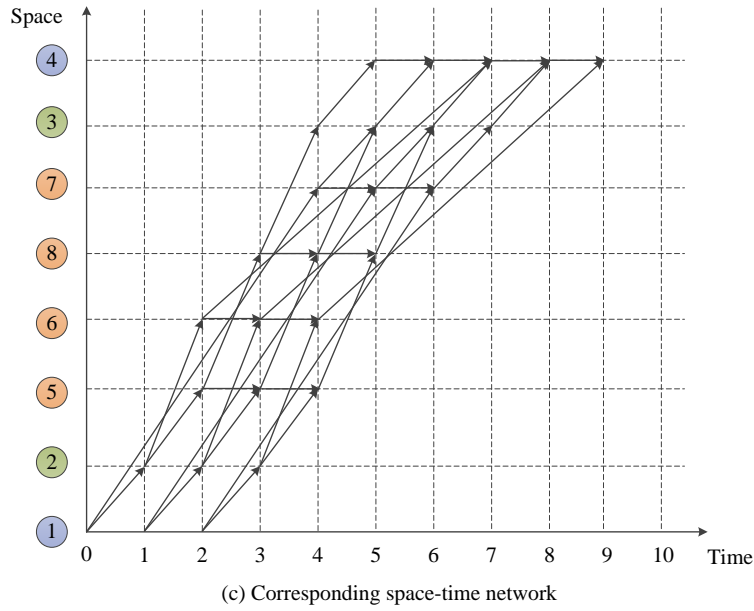
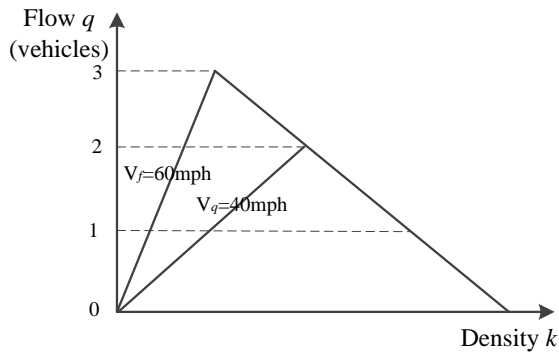
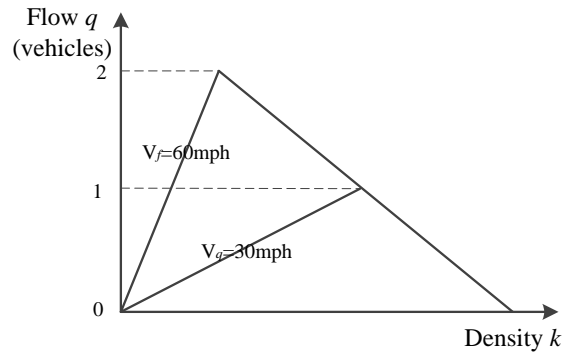


Fig. 9 Physical traffic network, modified traffic network, and its corresponding space-time network



(a) the fundamental  $q$ - $k$  diagram for link capacity = 3



(b) the fundamental  $q$ - $k$  diagram for link capacity = 2

Fig.10 Fundamental  $q$ - $k$  diagrams under different link capacities

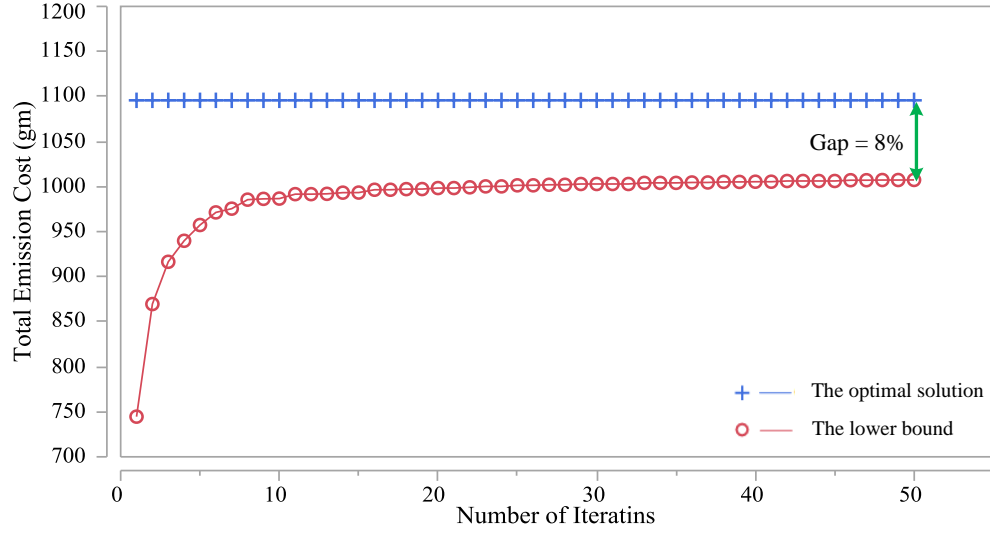


Fig. 11 Result comparison between the optimal solution and the lower bound in 50 iterations

Table 2 Time-dependent arc flows in the ESODTA and the travel time-based SODTA models

Arc( $i, j, t, t'$ )	Travel time-based SODTA	ESODTA	Arc( $i, j, t, t'$ )	Travel time-based SODTA	ESODTA
(1,2,0,1)	3	3	(3,4,6,7)	2	2
(1,2,1,2)	3	3	(3,4,7,8)	1	1
(1,2,2,3)	1	1	(2,6,1,2)	2	1
(2,5,1,2)	1	2	(2,6,2,3)	2	2
(2,5,2,3)	1	1	(2,6,3,4)	0	1
(2,5,3,4)	1	0	(6,4,2,7)	2	0
(5,8,2,3)	1	1	(6,4,3,8)	2	2
(5,8,3,4)	1	1	(6,4,4,9)	0	2
(5,8,4,5)	1	1	(5,5,2,3)	0	1
(8,3,3,4)	1	0	(5,5,3,4)	0	1
(8,3,4,5)	1	0	(6,6,2,3)	0	1
(8,3,5,6)	1	1	(6,6,3,4)	0	1
(1,7,0,4)	2	2	(8,8,3,4)	0	1
(1,7,1,5)	1	1	(8,8,4,5)	0	1
(1,7,2,6)	0	0	(7,7,4,5)	1	1
(7,3,4,5)	1	1	(7,7,5,6)	1	1
(7,3,5,6)	1	1	(4,4,5,6)	1	1
(7,3,6,7)	1	1	(4,4,6,7)	3	3
(3,4,4,5)	1	1	(4,4,7,8)	4	5
(3,4,5,6)	2	2	(4,4,8,9)	10	8

## 5.2 Numerical example for the ESODTA with time-dependent bottleneck capacities

This numerical experiment is conducted on the Salt Lake City regional traffic network with 13,923 nodes, 26,768 links and 2,302 zones shown in Fig. 12 (Zhou, et al., 2015b). The total number of vehicles is about 1.35 million for a 3-hour demand loading period, 15:00-18:00.

Although the primary focus of this experiment is on the ESODTA model with time-dependent bottleneck capacities (Section 4), we also solve the travel time-based user equilibrium DTA (UEDTA) problem (Lu et al., 2009) and compare the results of the two models. Average trip time and average CO emission of the two models for 40 iterations are depicted in Fig. 13. We can observe that: (i) in Fig. 13(a) although the objective of the ESODTA is to minimize the total emission, its final average trip time is also less than that of the UEDTA model; (ii) in Fig. 13(b) the final average CO emission of the ESODTA model is less than that of the UEDTA model, because travelers choose the least marginal emission routes in the ESODTA model.

According to the final average trip time and average CO emission of the two models, one vehicle can save 1300 grams of CO emissions and 3 minutes of travel time if the routing policy of the ESODTA model is adopted. Considering the total demand of 1.35 million vehicles, significant savings in total emission and travel time can be achieved by adopting the routing policy of the ESODTA model. Note that, in order to improve the computation efficiency, the two solution algorithms were implemented using a parallel computing technique. The computational time for one iteration of the solution algorithms is about 1 minute and 10 seconds in the computing environment used.



Fig. 12 Salt Lake City regional traffic network

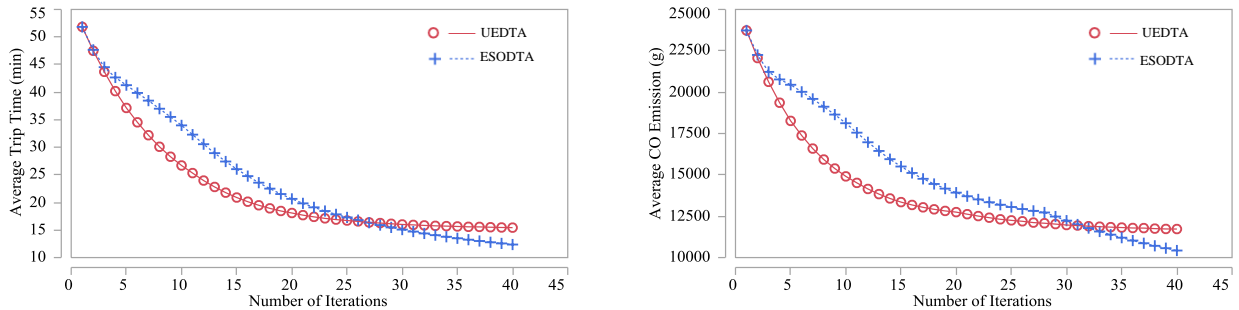


Fig. 13 Comparisons of average trip time and average CO emission of ESODTA and UEDTA models

## 6. Conclusions

This research generalizes the normative capability of DTA to sustainable transportation network modelling in terms of emission. The ESODTA model is proposed to obtain time-varying path flows in a congested network where

travellers behave cooperatively in selecting paths to minimize total network emissions. For the case of constant bottleneck capacities, the model is formulated on the expanded space-time network constructed using the modified point-queue model (Section 3). For the case of time-dependent bottleneck capacities, the model is solved by the column generation-based algorithm, which consists of a mesoscopic DNL model and a gradient projection-based descent direction method for updating time-dependent path assignments. Furthermore, link marginal emission is derived and its relationship with link marginal travel time is discussed. We also propose a numerical approach for evaluating path marginal emissions. The numerical examples demonstrate the effectiveness of the model and the algorithm for obtaining the routing policy which minimizes total network emissions, and the emission reduction in ESODTA is gained at the expense of an increased total travel time. In addition, compared to the UEDTA, significant savings in total emission and travel time can be achieved by adopting the routing policy of the ESODTA model.

The solution to the ESODTA model provides a benchmark (or lower bound) in total system emission when drivers are unanimously guided by a central controller against other less environmentally-efficient routing policies, such as green user equilibrium (GUE) flow patterns. Moreover, the Eco SO path flows provide a basis of generating green routing policies for advanced eco-friendly route guidance provision systems that consider different market penetration rates. Link and path marginal emissions, which are essential input to the ESODTA algorithm, are also valuable for deriving emission charges on road users.

## Acknowledgements

The research reported here was partially sponsored by U.S. Department of Transportation (DOT) University Transportation Centers, “Development and Demonstration of Advanced Methods for Quantifying Freight Truck Activity, Energy Use, and Emissions”, partially funded by U.S. Department of Energy’s (DOE) Advanced Research Projects Agency – Energy (ARPA-E), “Traveler Response Architecture using Advanced Novel Signaling for Network Efficiency in Transportation (TRANSNET)”, partially funded by National Science Foundation – United States under Grant No. CMMI 1538105 “Collaborative Research: Improving Spatial Observability of Dynamic Traffic Systems through Active Mobile Sensor Networks and Crowdsourced Data”, and partially funded by Utah Department of Transportation, “Simplified web-based decision support method for traffic management and work zone analysis”. The work presented in this paper remains the sole responsibility of the authors.

## References

- Abdul Aziz, H.M., Ukkusuri, S.V., 2012. Integration of environmental objectives in a system optimal dynamic traffic assignment model. *Computer-Aided Civil & Infrastructure Engineering*, 27(7), 494–511.
- Bai, S., Nie, Y., Niemeier, D. A., 2007. The impact of speed post-processing methods on regional mobile emissions estimation. *Transportation Research Part D* 12(5), 307-324.
- Benedek, C.M. and Rilett, L.R., 1998. Equitable traffic assignment with environmental cost function. *Journal of Transportation Engineering*, 124, 16-22.
- Boriboonsomsin, K., Barth, M., 2008. Impacts of freeway high-occupancy vehicle lane configuration on vehicle emissions. *Transportation Research Part D* 13(2), 112-125.
- Brockfeld, E., Kühne, R., Wagner, P., 2004. Calibration and validation of microscopic traffic flow models. *Transportation Research Record* 1876, 62–70.
- Carey, M., 1987. Optimal time-varying flows on congested networks. *Operations Research* 35(1), pp. 58-69.
- Carey, M., 1992. Nonconvexity of the dynamic traffic assignment problem. *Transportation Research Part B* 26(2), pp. 127-133.
- Carey, M., and Watling, D. 2012. Dynamic traffic assignment approximating the kinematic wave model: System optimum, marginal costs, externalities and tolls. *Transportation Research Part B: Methodological*, 46(5), 634-648.
- Daganzo, C. F., 1994. The cell transmission model: a simple dynamic representation of highway traffic. *Transportation Research Part B* 28(4), pp. 269-287.
- Daganzo, C. F., 1995a. The cell transmission model, part II: network traffic. *Transportation Research Part B* 29(2), pp. 79-93.
- Daganzo, C. F., 2006. In traffic flow, cellular automata = kinematic waves. *Transportation Research Part B* 40(5), 396-403.

- 1 Doan, K., Ukkusuri, S. V. 2015. Dynamic system optimal model for multi-OD traffic networks with an advanced  
2 spatial queuing model. *Transportation Research Part C: Emerging Technologies*, 51, 41-65.
- 3 Dowling, R., Skabardonis, A., 1993. Improving average travel speeds estimated by planning models. *Transportation*  
4 *Research Record* 1366, 68–74.
- 5 Friesz, T. L., Luque, J., Tobin, R. L., Wie, B.-Y., 1989. Dynamic network traffic assignment considered as a  
6 continuous time optimal control problem. *Operations Research* 37(6), pp. 893-901.
- 7 Gaker, D., Zheng, Y., Walker, J., 2010. Experimental economics in transportation: focus on social influences and  
8 provision of information. *Transportation Research Record* 2156, 47-55.
- 9 Gaker, D., Vautin, D., Vij, A., Walker, J. L., 2011. The power and value of green in promoting sustainable transport  
10 behavior. *Environmental Research Letters* 6(3), 1-11.
- 11 Garcia, A., Reaume, D., Smith, R. L., 2000. Fictitious play for finding system optimal routings in dynamic traffic  
12 networks. *Transportation Research Part B* 34(2), 147-156.
- 13 Ghali, M.O. and Smith, M.J., 1995. A model for the dynamic system optimum traffic assignment problem.  
14 *Transportation Research Part B* 29(3), 155-170.
- 15 Greene, D., Schafer, A., 2003. Reducing greenhouse gas emissions from U.S. transportation. Center for Climate and  
16 Energy Solutions (<http://www.c2es.org/docUploads/ustransp.pdf>).
- 17 Helali, K., Hutchinson, B., 1994. Improving road link speed estimates for air quality models. *Transportation*  
18 *Research Record* 1444, 71–77.
- 19 Hurdle, V. F., Son, B., 2000. Road test of a freeway model. *Transportation Research Part A* 34(7), 537-564.
- 20 Lafortune, S., Sengupta, R., Kaufman, D. E., Smith, R., 1993. Dynamic system-optimal traffic assignment using a  
21 state space model. *Transportation Research Part B* 27(6), 451-472.
- 22 Lawson, T. W., Lovell, D. J., and Daganzo, C. F., 1996. Using Input-Output Diagram to Determine Spatial and  
23 Temporal Extents of a Queue Upstream of a Bottleneck. *Transportation Research Record*, No. 1572, pp. 140-147
- 24 Leclercq, L., 2007. Hybrid approaches to the solutions of the “Lighthill-Whitham-Richards” model. *Transportation*  
25 *Research Part B* 41(7), 701-709.
- 26 Li, P., Mirchandani, P., Zhou, X. 2015. Solving simultaneous route guidance and traffic signal optimization problem  
27 using space-phase-time hypernetwork. *Transportation Research Part B: Methodological* 81: 103-130.
- 28 Lighthill, M., Whitham, G., 1955. On kinematic waves II: a theory of traffic flow on long crowded roads. *Proc.*  
29 *Royal Society of London, Part A* 229 (1178), 317-345.
- 30 Lo, H. K., Szeto, W. Y., 2002. A cell-based variational inequality formulation of the dynamic user optimal  
31 assignment problem. *Transportation Research Part B* 36(5), 421-443.
- 32 Lu, C.-C., Mahmassani, H. S., Zhou, X., 2009. Equivalent gap function-based reformulation and solution algorithm  
33 for the dynamic user equilibrium problem. *Transportation Research Part B* 43(3), 345-364
- 34 Ma, R., Ban, X. J., Pang, J. S. 2014. Continuous-time dynamic system optimum for single-destination traffic  
35 networks with queue spillbacks. *Transportation Research Part B: Methodological*, 68, 98-122.
- 36 Ma, R., Ban, X. J., & Szeto, W. Y. 2015. Emission Modeling and Pricing in Dynamic Traffic  
37 Networks. *Transportation Research Procedia*, 9, 106-129.
- 38 Magnanti, T., Perakis, G., 1997. Averaging schemes for variational inequalities and systems of equations.  
39 *Mathematics of Operations Research* 22(3), 568-587.
- 40 Mandavilli, S., Rys, M. J., Russell, E. R., 2008. Environmental impact of modern roundabouts. *International Journal*  
41 *of Industrial Ergonomics* 38(2), 135-142.
- 42 Merchant, D. K., Nemhauser, G. L., 1978. A model and an algorithm for the dynamic traffic assignment problems.  
43 *Transportation Science* 12(3), 183-199.
- 44 Mounce, R., Smith, M., 2007. Uniqueness of Equilibrium in steady state and dynamic traffic networks. In: Allsop,  
45 R.E., Bell, M.G.H., and Heydecker, B.G. (Eds), *Transportation and Traffic Theory*, Elsevier, 281-299.
- 46 Munoz, J. C., Laval, J. A., 2006. System optimum dynamic traffic assignment graphical solution method for a  
47 congested freeway and one destination. *Transportation Research Part B* 40(1), 1-15.
- 48 Nagurney, A., Ramanujam, P. and Dhanda, K.K., 1998. Multimodal traffic network equilibrium model with  
49 emission pollution permits: Compliance vs noncompliance. *Transportation Research, Part D*, 35, 349–374.
- 50 Nagurney, A., Dong, J. and Mokhtarian, P.L., 2002. Traffic network equilibrium and the environment: A  
51 multicriteria decision-making perspective. In E. Kontogiorgos, B. Rustem, & S. Siokos (Eds.), *Computational*  
52 *methods in decision-making, economics and finance* (pp. 501–523). Dordrecht: Kluwer Academic Publishers.
- 53 Newell, G. F., 1993a. A simplified theory on kinematic waves in highway traffic, part I: general theory.  
54 *Transportation Research Part B* 27(4), 281-287.

- 1 Newell, G. F., 1993b. A simplified theory on kinematic waves in highway traffic, part II: queueing at freeway  
2 bottlenecks. *Transportation Research Part B* 27(4), 289-303.
- 3 Newell, G. F., 1993c. A simplified theory on kinematic waves in highway traffic, part III: multi-destination flows.  
4 *Transportation Research Part B* 27(4), 305-313.
- 5 Newell, G., 2002. A simplified car-following theory: a lower order model, *Transportation Research B* 36(3), 195–  
6 205.
- 7 Ni, D., Leonard, J. D., Williams, B. M., 2006. The network kinematic waves model: a simplified approach to  
8 network traffic. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 10(1), 1-14.
- 9 Panis, L., Broekx, S., Ronghui, L., 2006. Modeling instantaneous traffic emission and the influence of traffic speed  
10 limits. *Science of the Total Environment* 371(1-3), 270–285.
- 11 Peeta, S., Mahmassani, H. S., 1995a. System optimal and user equilibrium time-dependent traffic assignment in  
12 congested networks. *Annals of Operation Research* 60(1), 81-113.
- 13 Peeta, S., Mahmassani, H.S., 1995b. Multiple user classes real-time traffic assignment for on-line operations: A  
14 rolling horizon solution framework. *Transportation Research, Part C: Emerging Technologies* 3(2), 83-98.
- 15 Peeta, S., Zhou, C., 2006. Stochastic quasi-gradient algorithm for the off-line stochastic dynamic traffic assignment  
16 problem. *Transportation Research, Part B: Methodological* 40(3), 179-206.
- 17 Peeta, S., Ziliaskopoulos, A.K., 2001. Foundations of dynamic traffic assignment: The past, the present and the  
18 future. *Networks and Spatial Economics* 1(3/4), 233-266.
- 19 Qian, Z. and Zhang, H.M., 2011. Computing individual path marginal cost in networks with queue spillbacks.  
20 *Transportation Research Record* 2263, 9-18.
- 21 Qian, Z., Shen, W., Zhang, H. M.. 2012. System-optimal dynamic traffic assignment with and without queue  
22 spillback: Its path-based formulation and solution via approximate path marginal cost." *Transportation research*  
23 *part B: methodological* 46(7) 874-893.
- 24 Richards, P. I., 1956. Shock waves on the highway. *Operations Research* 4(1), pp.42-51.
- 25 Sbayti, H., Lu, C.-C., Mahmassani, H. S., 2007. Efficient implementations of the method of successive averages in  
26 simulation-based DTA models for large-scale network applications. *Transportation Research Record* 2029, 22-  
27 30.
- 28 Shen, W., Nie, Y., Zhang, H. M., 2007. On path marginal cost analysis and its relation to dynamic system-optimal  
29 traffic assignment. In: Allsop, R.E., Bell, M.G.H., and Heydecker, B.G. (Eds), *Transportation and Traffic Theory*,  
30 Elsevier, 319-352.
- 31 Shen, W., Zhang, H. M., 2009. On the morning commute problem in a corridor network with multiple bottlenecks:  
32 Its system-optimal traffic flow patterns and the realizing tolling scheme. *Transportation Research Part B* 43(3),  
33 267-284.
- 34 Tong, L., Zhou, X., Miller, H. J. 2015. Transportation network design for maximizing space–time  
35 accessibility. *Transportation Research Part B: Methodological*, 81, 555-576.
- 36 Tzeng, G.H., and Chen, C.H., 1993. Multiobjective decision making for traffic assignment. *IEEE Transactions on*  
37 *Engineering Management*, 40, 180–187.
- 38 U.S. EPA, 2009. Draft Moto Vehicle Emission Simulator (MOVES) – Software Design and Reference Manual.  
39 Technical report EPA-420-B-09-007, U. S. Environmental Protection Agency.
- 40 Wie, B.-W., Tobin, R. L., Friesz, T. L., 1994. The augmented Lagrangian method for solving dynamic network  
41 traffic assignment models in discrete time. *Transportation Science* 28(3), 204-220.
- 42 Zhang, H. M., Nie, Y., Qian, Z. 2013. Modelling network flow with and without link interactions: the cases of point  
43 queue, spatial queue and cell transmission model. *Transportmetrica B: Transport Dynamics* 1.1, 33-51.
- 44
- 45 Zhang, K., Mahmassani, H. S., Lu, C.-C., 2009. Probit-based time-dependent stochastic user equilibrium:  
46 reformulation and solution algorithm. Paper presented at the 88th Annual Meeting of Transportation Research  
47 Board, January 11-15, 2009, Washington, D.C., USA.
- 48 Zhang, Y., Lv, J. and Ying, Q., 2010. Traffic assignment considering air quality. *Transportation Research Part D*, 15,  
49 497-502.
- 50 Zhou, X., Taylor, J., 2014. DTALite: A queue-based mesoscopic traffic simulator for fast model evaluation and  
51 calibration. *Cogent Engineering* 1 (1), 961345.
- 52 Zhou, X., Tanvir, S., Lei, H., Taylor, J., Liu, B., Roupail, N.M. and Frey, H.C., 2015a. Integrating a simplified  
53 emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of  
54 traffic management strategies. *Transportation Research Part D*, 37, 123-136.



- 1 Zhou, X., Zlatkovic, M., and Farhan, M., 2015b. Simplified web-based decision support method for traffic  
2 management and work zone analysis. For Utah Department of Transportation Research Division. Report No.  
3 UT-15.09.
- 4 Zhu, F., Lo, H. K. and Lin, H.-Z., 2013. Delay and emissions modelling for signalized intersections.  
5 *Transportmetrica B: transport dynamics* 1.2: 111-135.
- 6 Ziliaskopoulos, A. K., 2000. A linear programming model for the single destination system optimum dynamic traffic  
7 assignment problem. *Transportation Science* 34(1), 1-12.
- 8 Ziliaskopoulos, A. K., Mahmassani, H. S., 1993. Time dependent shortest-path algorithm for real-time Intelligent  
9 Vehicle Highway System applications. *Transportation Research Record* 1408, 94-100.
- 10
- 11 Yang, L., Zhang, Y., Li, S., Gao, Y., 2016. A two-stage stochastic optimization model for the transfer activity choice  
12 in metro networks. *Transportation Research Part B* 83, 271-297.
- 13 Yin, J., Tang, T., Yang, L., Gao, Z., Ran, B., 2016. Energy-efficient metro train rescheduling with uncertain time-  
14 variant passenger demands: An approximate dynamic programming approach. *Transportation Research Part B*  
15 91, 178-210.
- 16
- 17 Nguyen and Pallottino, 1988; Spiess and Florian, 1989; De Cea and Fernandez, 1993; Nuzzolo et al., 2001; Gao et al., 2004
- 18
- 19 Lam et al., 1999; Nguyen et al., 2001; Cepeda et al., 2006; Szeto and Jiang, 2014a;
- 20
- 21
- 22 Poon et al., 2004; Hamdouch et al., 2004; Hamdouch and Lawphongpanich, 2008; Nuzzolo et al., 2012; Hamdouch et al., 2014; Verbas and  
23 Mahmassani, 2015;
- 24
- 25 Niu and Zhou, 2013; Szeto and Jiang, 2014b; Niu et al., 2015;
- 26