Project ID: NTC2014-SU-R-08

Evaluating and Calibrating Emission Impacts of Traffic Management Strategies through Simplified Emission Estimation Model and Mesoscopic Dynamic Traffic Simulators

Final Report

by

Xuesong Zhou
Associate Professor, School of Sustainable Engineering and the Built Environment
Arizona State University, Tempe, AZ, xzhou74@asu.edu

Jiangtao Liu
Graduate Student, School of Sustainable Engineering and the Built Environment
Arizona State University, Tempe, AZ, jliu215@asu.edu

Yunchao Qu
Post-doc Scholar, School of Sustainable Engineering and the Built Environment
Arizona State University, Tempe, quyunchao0613@gmail.com

for

National Transportation Center at Maryland (NTC@Maryland)
1124 Glenn Martin Hall
University of Maryland
College Park, MD 20742

February, 2016

ACKNOWLEDGEMENTS

This project was funded by the National Transportation Center @ Maryland (NTC@Maryland), one of the five National Centers that were selected in this nationwide competition, by the Office of the Assistant Secretary for Research and Technology (OST-R), U.S. Department of Transportation (US DOT).

DISCLAIMER

Table of Contents

Exclusive Summary

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 importance for a dynamic network loading (DNL) model to effectively generate high-fidelity vehicle trajectories and 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. 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. We derive a formula of marginal emission which encompasses the marginal delay as a special case, and develop an algorithm for evaluating path marginal emissions in a congested network.

1. Motivation and Objective

Highway vehicles have been the primary focus of environment protection and transportation agencies to reduce greenhouse gas emissions, since they account for 72 percent of total transportation emission (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 CO2 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 a minimal total system emission may be of a 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 aims 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 1995; Peeta and Ziliaskopoulos, 2001). A rich body of literature has devoted to the models and algorithms of the SODTA, and static traffic assignment (STA) models with environmental considerations have been found in a number of past studies (Szeto et al., 2012). Only until recently have the 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., 2015).

As the development of ESODTA is still in the infant stage, 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 emission 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 high-fidelity vehicular trajectories and time-dependent speeds so that fine-grained emission estimations can be obtained for multi-scale emission analyses, whereas most of the DNL models underlying classical DTA models concern mainly on the 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.

2. Literature Review

2.1 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; Ziliaskopoulos, 2000), optimal or convex control (e.g., Friesz et al., 1989; Lafortune et al. 1993; Wie et al., 1994), simulation model (Ghali and Smith, 1995; Peeta and Mahmassani, 1995), game theory (Garcia et al., 2000), graphic 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 chose 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 of the aforementioned numerous studies of STA with environmental considerations, very few DTA models have been developed for environment-related applications. Recently, 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. (2015) 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.

2.2 Microscopic and Macroscopic Traffic Flow Models for Emission Estimation

In order to capture the impact of traffic congestions to 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 be calibrated. Mesoscopic traffic flow modeling approach may be a more viable alternative, in order 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 deference-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) descritizes 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 at 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 incorporating 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. (2015) 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.

2.3 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 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, 1995; 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 OD 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, (1995) developed a numerical method based on mesoscopic traffic simulation to evaluate link and path marginal travel times. 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). 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, the study on the link and path marginal emissions were non-existent.

3. Agent-based ESODTA Model

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

Indices:

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

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 τ

Parameters:

$d(w, \tau)$	Number of vehicle of OD pair w and departure time interval τ
d	Vector of time-dependent demands of all OD pairs
vFl	Free flow speed on link l, which is also the speed limit on link l
vQl,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 τ
ql,t	Time-dependent link flow of link l at time interval t
q	Vector of link flows
TTf,l,t	Agent f's time-dependent travel time on link l at time interval t
TT	Vector of agents' path travel times
ECf	Total path emission or eco-cost of agent f
ECf,l	Emission or eco-cost of agent f on link l

$EC_{f,l,vt}$ Instantaneous emission of agent f with (time-dependent) speed v on link l at time interval t

$EC_{f,l,t',t''}$ Emission of agent f on link l with entering time t' and leaving time t''

EC Vector of path emission cost of all agents

TE Total system emission cost

3.1 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 origin-destination (OD) 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

$$\text{Min}_{\mathbf{r}\in\Omega}\ TE=\sum_f EC_f \tag{1}$$

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

$$EC_f=\sum_{l\in p} f_{EC_{f,l}} \tag{3}$$

$$\sum_{p\in P_{(w,\tau)}} r_{(w,\tau,p)}=d_{w,\tau},\quad \forall w,\tau \tag{4}$$

$$r_{w,\tau,p}\geq 0,\quad \forall w,\tau,p\in P_{(w,\tau)} \tag{5}$$

The objective function, Eq.(1), minimizes the network-wide vehicular emission. Constrain (2) states that, given the time-dependent OD demand vector **d**, the estimated path emissions of all agents, **EC**, is obtained by a dynamic network loading (DNL) model for emission estimation, DNLE(**d**). Constrain (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 τ. 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, DNLE(**d**), can be embedded for multi-scale emission optimization applications.

3.2 Dynamic Network Loading Model for Emission Estimation

To enable multi-scale emission optimization applications, the ESODTA model relies on an effective DNL model (DNLE(**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 DNLE(**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 DNLE(**d**) model to output high-fidelity vehicle trajectories and time-dependent speeds for estimating vehicular emissions.

Technically, both microscopic and mesoscopic traffic flow models can be embedded in DNLE(**d**) to construct detailed vehicle trajectories and 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 $t1$ and leaving the link at time $t4$. The agent's trajectories described by the three methods are as follows. With Method 1 (point-queue model), the agent moves at a free flow speed vF until reaching the downstream node at time $t3$, and then stops at this node until the capacity is available for discharging this agent at time $t4$ (e.g., point-queue model). With Method 2 (average cell-speed model), the agent moves at an average speed v through the link (e.g., Aziz and Ukkusuri, 2012). With Method 3 (modified point-queue model), the agent moves at a free flow speed vF until reaches the back of the queue at time $t2$, and then at a slower speed vQ 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, hence resulting in a more reasonable estimate of emission for the vehicle.

Typically, the queueing speed vQ is related to the bottleneck discharge capacity. Assume that the discharge capacity is constant and there is no queue-spillback effect, then the queuing 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 4. 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 queuing 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 queuing 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 et al. (2015), in a column generation-based solution algorithm presented in Section 5.

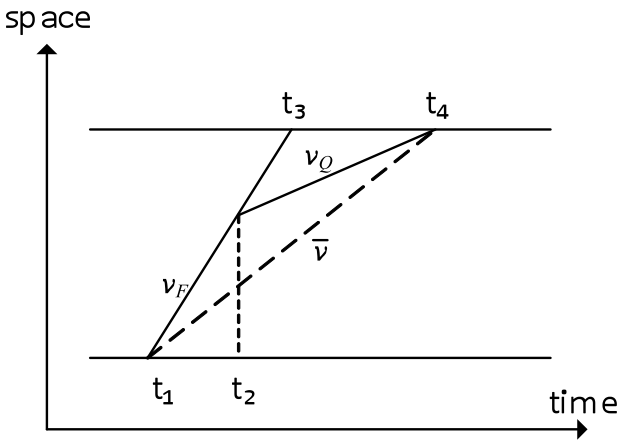

Fig.1 Three mesoscopic methods for obtaining an agent's trajectory

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	vF	$t3-t1$	0	$t4-t3$	$t4-t1$
Method 2	v	$t4-t1$	v	0	$t4-t1$
Method 3	vF	$t2-t1$	vQ	$t4-t2$	$t4-t1$

3.3 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_2) 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, ECf,l,t',t'', is an integration of the instantaneous emission cost $ECf,l,v(t)$ during the entrance time t' and the exit time t''.

$$ECf,l,t',t''=t't''ECf,l,v(t)dt. \quad (6)$$

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.

$$VSPv,a=A\cdot v+B\cdot v2+C\cdot v3+m\cdot v\cdot am \quad (7)$$

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,

$$E(v)= -0.064 + 0.0056v + 0.00026\, v - 502 \quad (8)$$

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.

4. 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.

4.1 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 vF under uncongested conditions. When a bottleneck forms at the end of the link, for keeping safe, the vehicle will decelerate to a slower speed vQ at the location dQ (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.

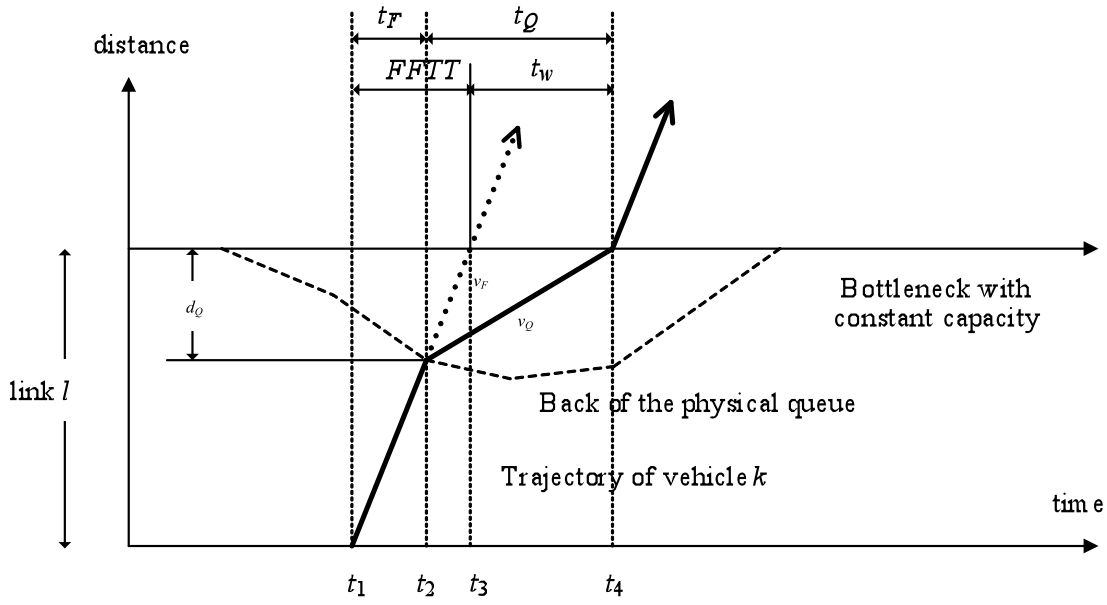

Fig. 2 Vehicle trajectory on a link with a bottleneck (adopted from Lawson et al., 1997)

The vehicle enters the link at time *t1*, reaches the back of the queue (represented by the dashed curve) at time *t2*, desires to leave the link at time *t3*, and eventually leaves the link at time *t4* due to capacity restriction. Here, we denote the travel time under free flow speed as

$tF=t2-t1$, the travel time under queueing speed (i.e., time in queue) as $tQ=t4-t2$; the free flow travel time $FFTT=t3-t1$, and delay $tw=t4-t3$. The total travel time TT on link l can be described as follows:

$$TT=FFTT+tw=tF+tQ. \quad (9)$$

It is obvious that the time in queue tQ is greater than the delay tw, 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 dQ, as $dQ=tQ \cdot vQ=vF \cdot (tQ-tw)$, and the time in queue tQ can be derived as follows:

$$tQ=tw \cdot vFvF-vQ. \quad (10)$$

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

$$tF=FFTT-tw \cdot vQvF-vQ. \quad (11)$$

With the assumption of constant discharge capacity, both vF and vQ are constant. The link emission of agent f entering and exiting link l at times $t1$ and $t4$, respectively, consists of two parts: the emission when traveling at the free-flow speed vF and the emission when traveling at the queueing speed vQ, as follows.

$$ECf,l,t1,t4=ECf,l,vF \cdot tF+ECf,l,vQ \cdot tQ. \quad (12)$$

For clarity, the agent, link and time indices are omitted from the notations here. That is, $ECF=ECf,l,vF$, and $ECQ=ECf,l,vQ$, then Eq.(12) can be re-written as Eq.(13).

$$EC=ECF \cdot tF+ECQ \cdot tQ. \quad (13)$$

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 ECF and the delay tw.

Proof:

The total link emission $EC=ECF \cdot tF+ECQ \cdot tQ$, and according to Eqs.(10)-(13), we can obtain the total link emission as follows:

$$EC=FFTT-tw \cdot vQvF-vQ \cdot ECF+tw \cdot vFvF-vQ \cdot ECQ. \quad (14)$$

Assume that $\lambda=ECQ ECF>1$, then total link emission can be re-written as a linear function of the delay tw.

$$EC= ECF \cdot FFTT+tw \cdot \lambda \cdot vF-vQvF-vQ \quad (15)$$

This completes the proof.

It is also important to note that if $ECF=1$ and $\lambda=1$ (i.e., $ECQ=ECF$), then $\lambda \cdot \upsilon F-\upsilon Q \upsilon F-\upsilon Q=1$, and we can obtain that the link travel time is a special case of the link emission.

$$EC= ECF \cdot FFTT+tw \cdot \upsilon F-\upsilon Q \upsilon F-\upsilon Q=FFTT+tw=TT. \quad (16)$$

4.2 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:

Of 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, and it is a given large value

si,j Free-flow travel time of link (i,j), which is an integer multiplier of one time interval

ti,j Travel time (free flow or congested travel time) of arc (i,j)

$ECFl$ Unit emission cost under constant free flow speed υFl on link l

$ECQl$ Unit emission cost under constant queueing speed υQl on link l

ECi,j,t,t' Emission cost of an agent traveling on arc (i, j) with entrance time t and exit time t'

$Capi,j$ Outflow capacity of arc (i,j)

$Leni,j$ Length of arc (i,j)

$ki,jjam$ Jam density of arc (i,j)

Variables:

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

The expanded space-time network is constructed based on the modified point-queue model (i.e., Method 3 in Section 3.2) 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

present the expanded space-time network in this paper. A vertex in the expanded network represents a physical node or virtual waiting node i at time t (or a node-time pair (i,t)). The space-time network is expanded by adding a virtual waiting node j' on physical link (i,j). In addition, the corresponding node-time pairs (j',t) for that virtual waiting node j' at all time intervals $t=1,2,...,T$ are added to the expanded network. Note that if j is a destination node, it is not necessary to add the virtual waiting node j', because the destination node j can be used as a waiting node. Moreover, if a node is the origin for one agent and the destination for other agents, we need to divide this node to two nodes to ensure that a destination node is not an origin node and vice versa. As a result, the problem becomes a single origin to single destination problem for each agent. With the addition of a virtual waiting node j', link (i,j) is divided into two sections: (i,j') and (j',j). The first section represents a vehicle moves at the free flow speed, while the second section is used to discharge that vehicle from the bottleneck after waiting at the virtual node j' for the available capacity. There are two types of arc connecting the vertices in the expanded space-time network.

Traveling arcs: A traveling arc represents the vehicle moves on a link (i,j) from time t to time t'. There are two cases. In the first case, a traveling arc connects vertex (i,t) (i.e., physical node i at time t) to vertex $(j',t+s_{i,j}-1)$ (i.e., virtual node j' at time $t+s_{i,j}-1$), representing a vehicle moves at the free flow speed on link (i,j). The arc travel time is $s_{i,j}-1$ and arc capacity is $Cap_{i,j}$. In the second case, a traveling arc is incident from vertex (j',t) (i.e., virtual node j' at time t) and incident to vertex $(j,t+1)$ (i.e., physical node j at time $t+1$), representing a vehicle discharges from the bottleneck on link (i,j). The arc travel time is 1 and arc capacity is $Cap_{i,j}$.

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.

For each agent f, we can find its source vertex ($O(f)$, $DT(f)$) corresponding to the origin node $O(f)$ and the departure time $DT(f)$ in the extended space-time network, and its sink vertex can be obtained by assuming a large arrive time $AT(f)$ at the destination node $D(f)$. Moreover, the costs on the waiting arcs of destination nodes are set as 0.

Fig. 3(a) depicts a simple physical network with three nodes, two links, and two OD pairs: (1, 2) and (2, 3). Note that since node 2 represents simultaneously the origin of the first OD pair and the destination of the second OD pair, this node is divided into node 2 and node 4, as shown in Fig. 3(b), and hence the first OD pair becomes (1, 4). The travel time and capacity of link (2, 4) are 1 and a large constant M, respectively. Then we add one virtual node $2'$ on link (1, 2) as a

waiting node, as shown in Fig. 3(c). Fig. 4 displays the corresponding expanded space-time network.

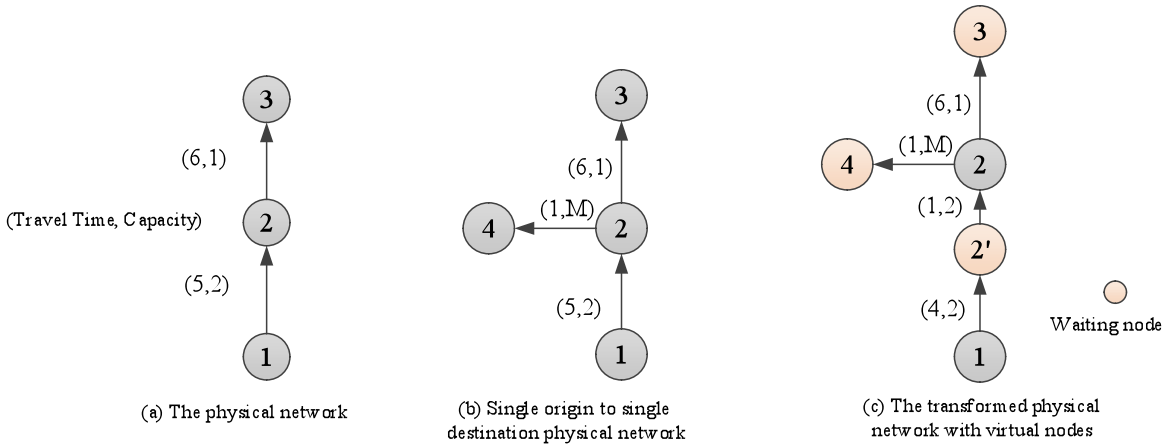

Fig. 3 Original and expanded physical network

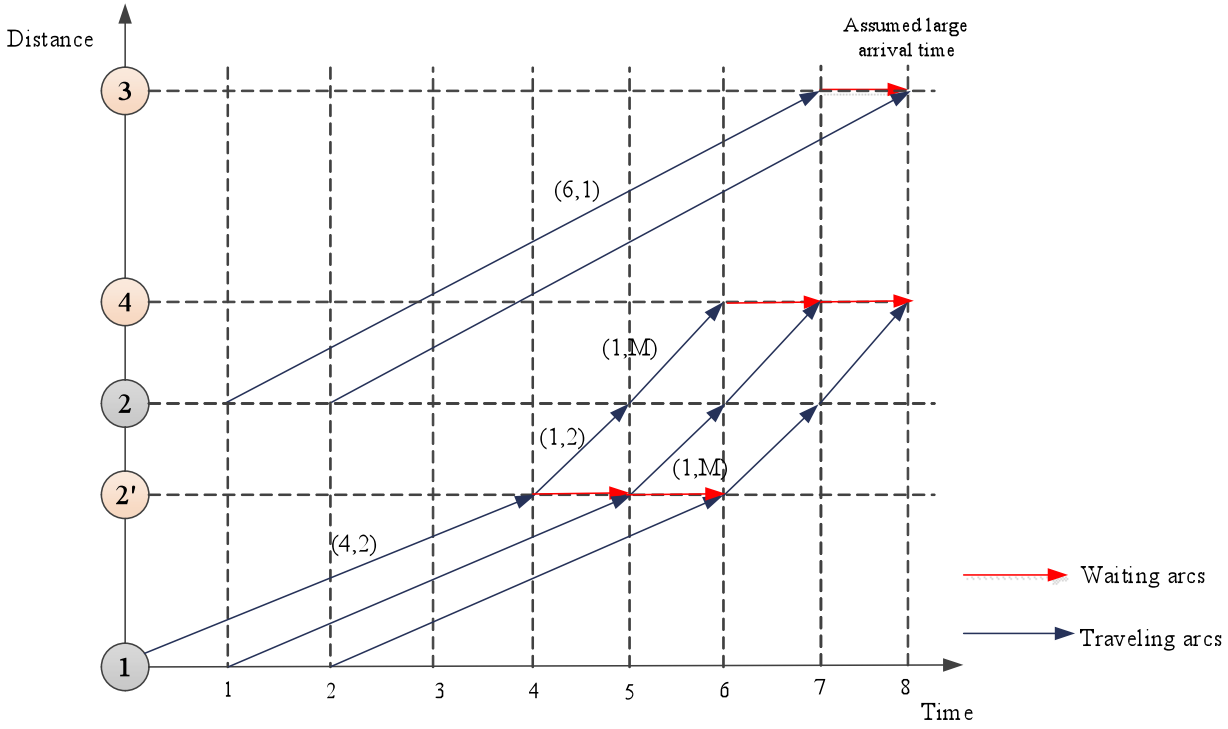

Fig. 4 The expanded space-time network corresponding to the network depicted in Fig. 3(b)

4.3 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.

Minimize $Z=f(i,j,t,t')ECi,j,t,t'\times xi,j,t,t'f$ (17)

Subject to

$$\sum_{i,j,t-ti,j,t\in A} x_{i,j,t-ti,j,t}(f) - \sum_{j,k,t,t+tj,k\in A} x_{j,k,t,t+sj,k}(f) = -1 \quad \forall f\in F,\ j=O_f \text{ and } t=DT(f) \quad 1 \quad \forall f\in F, j=D_f \text{ and } t=T \quad 0 \quad otherwise \tag{18}$$

$$\sum_{f\in F} x_{i,j,t,t+ti,j}(f) \le Cap_{i,j},\ \forall (i,j,t,t+ti,j)\in A, t=1,2,\ldots,T \tag{19}$$

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

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 of the decision variables are binary.

The emission costs $EC_{i,j,t,t'}$ in the objective function are discussed as follows. According to Proposition 1 in Section 4.1, the link emission cost is a function of the delay $tw_{i,j}$ on link i,j:

$$EC_{i,j} = ECF_{i,j} \times s_{i,j} + tw_{i,j} \times \lambda \times \upsilon_F - \upsilon_Q \upsilon_F - \upsilon_Q = a_{i,j} + \beta_{(i,j)} \times tw_{i,j} \tag{21}$$

where $a_{i,j} = ECF_{i,j} \times s_{i,j}$ is the free-flow emission cost on link (i,j), and $\beta_{i,j} = ECF_{i,j} \times \lambda \times \upsilon_{Fi,j} - \upsilon_{Qi,j} \upsilon_{Fi,j} - \upsilon_{Qi,j}$ 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 three parts: (i) the emissions at free-flow speed on the traveling arc before reaching the virtual waiting node or bottleneck, where the emission cost is $EC_{i,j',t,t+s_{i,j}-1} = ECF_{i,j} \times (s_{i,j} - 1)$; (ii) the emission at queueing speed on waiting arcs (waiting at the virtual node) for $tw_{i,j}$ time intervals, where the emission cost of one waiting arc is $EC_{j',j',t,t+1} = \beta_{(i,j)}$; (iii) the emission at free-flow speed on the traveling arc after discharging from the virtual waiting node, where the emission cost is $EC_{j',j,t,t+1} = ECF_{i,j}$.

4.4 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.

$$\text{Min} ZLR = f(i,j,t,t')(EC_{i,j,t,t'} + \mu_{i,j,t,t'}) \times x_{i,j,t,t'}(f) - (i,j,t,t')\mu_{i,j,t,t'} \times Cap_{i,j} \tag{21}$$

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): $\gamma_{n=1}{}^{n+1}$

$$\mu_{i,jn+1t} = \max\{0, \mu_{i,jnt} + \gamma_n \times (fx_{i,j,t,t+s_{i,jf}} - Cap_{i,jOut})\} \quad (22)$$

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

5. ESODTA with Time-dependent Bottleneck Capacities

The ESODTA model, presented in Section 4.2, 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 queuing 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., 2015, 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 3.3).

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:

$Pm(w,\tau)$ Set of paths in iteration m for the agents of OD pair w and departure time τ

Pm Set of paths in iteration m for all of the agents in the network; $Pm=\{Pmw,\tau, \forall w,\tau\}$

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:

$Mmax$ Maximum number of outer loop iterations

$Nmax$ Maximum number of inner loop iterations

Variables:

$r_{w,\tau,pn}$ Number of agents on path p of OD pair w and departure time τ in iteration n

$\boldsymbol{r}n$ The path flow vectors (a feasible solution) in iteration n; $\boldsymbol{r}n=\{r_{w,\tau,pn},\forall w,\tau,p\}$

$EC_{w,\tau,pn}$ Total emission of the agents on path p of OD pair w and departure time τ in iteration n

$TEw\tau\boldsymbol{r}n$ Total emission of the agents of OD pair w and departure time τ, evaluated at the feasible solution $\boldsymbol{r}n$

$TE\boldsymbol{r}n$ Total system emission, evaluated at the feasible solution $\boldsymbol{r}n$

5.1 Solution Algorithm

5.1.1 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, Pm, in iteration m. A gradient projection-based descent direction method, presented in Section 5.1.2, is then used to solve the reduced ESODTA problem defined on Pm. The algorithm terminates and outputs time-varying path flows obtained in the current iteration, if there is not any new path 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 $\boldsymbol{r}n$ and the corresponding total system emission $TE\boldsymbol{r}n$ and link marginal emissions are evaluated by the DNLE($\mathbf{d}$) model. If the difference between the objective values in two successive iterations (i.e., $TE\boldsymbol{r}n-TE\boldsymbol{r}n-1$) is less than a preset threshold or a preset convergence criterion (e.g., $n=Nmax$) is satisfied, the inner loop terminates and the algorithm returns to the outer loop.

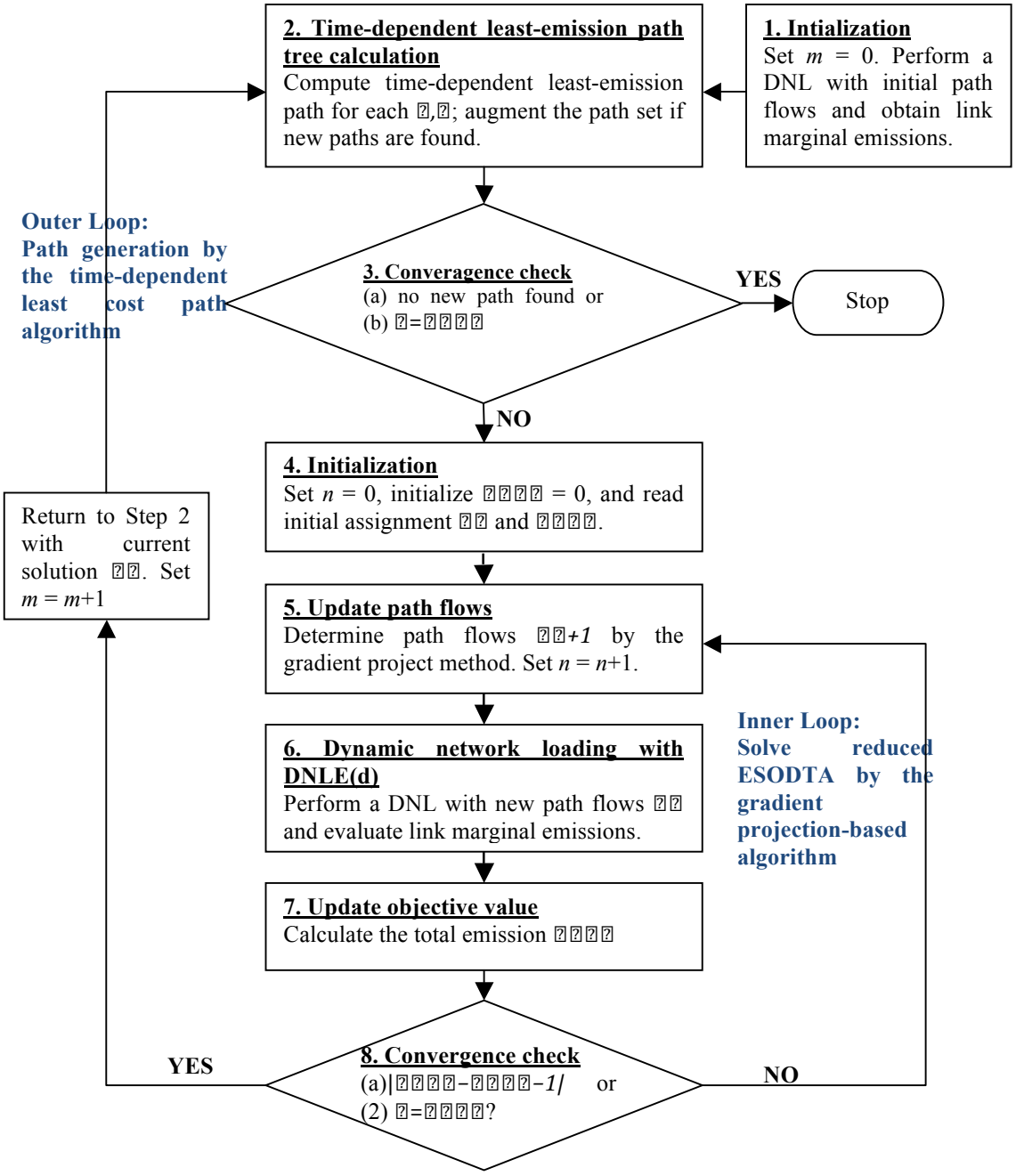

Fig. 5 Flow chart of the column generation-based ESODTA algorithm

5.1.2 Gradient Projection-based Descent Direction Method

The reduced ESODTA problem, defined by a subset of feasible paths *Pm* 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 ***rn*** in inner loop iteration

n, the method adopts a search direction along the feasible descent direction based on the gradient, $\nabla TE\boldsymbol{r}n$:

$$\boldsymbol{r}n+1=Proj\Omega[\boldsymbol{r}n-sn\times\nabla TE\boldsymbol{r}n], \tag{23}$$

where $sn\in(0, 1)$ 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 Ω and is defined as the unique solution of the problem: $\mathrm{Min}y\in\Omega||u-v||$. Accordingly, the new iterate $\boldsymbol{r}n+1$ is obtained by updating the current iterate $\boldsymbol{r}n$ along the direction $-\nabla TE\boldsymbol{r}n$ with a move size sn.

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,τ 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,τ pair. Given a feasible solution $\boldsymbol{r}n$, the flow balance conservation constraints Eq.(4) can be rearranged as follows:

$$r(w,\tau,p*)n=dw,\tau-p\in Pw,\tau\backslash p*rw,\tau,pn,\ \forall w,\tau. \tag{24}$$

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

$$\begin{aligned}TEw\tau\boldsymbol{r}n&=p\in Pw,\tau\backslash p*rw,\tau,pn\times ECw,\tau,pn+r(w,\tau,p*)n\times EC(w,\tau,p*)n\\&=p\in Pw,\tau\backslash p*rw,\tau,pnECw,\tau,pn-EC(w,\tau,p*)n+dw,\tau\times EC(w,\tau,p*)n.\end{aligned} \tag{25}$$

The first-order partial derivative of $TEw\tau\boldsymbol{r}n$ with respect to a particular path flow rw,τ,p is

$$\begin{aligned}\nabla TEw\tau\boldsymbol{r}n&=\partial TEw\tau\boldsymbol{r}n\partial r(w,\tau,p)\\&=ECw,\tau,pn-EC(w,\tau,p*)n+p'\in Pw,\tau\backslash p*r(w,\tau,p')n\eta wp'\tau,n-\eta wp*\tau,n+dw,\tau\times\eta wp*\tau,n.\end{aligned} \tag{26}$$

where $\eta wp\tau,n=\partial ECw,\tau,pn\partial r(w,\tau,p)$ 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

$$\nabla TEw\tau\boldsymbol{r}n=ECw,\tau,pn-EC(w,\tau,p*)n+rw,\tau,pn\eta wp\tau,n-\eta wp*\tau,n+dw,\tau\times\eta wp*\tau,n. \tag{27}$$

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

$$sn=1/m, \text{ if } n=0;\ sn=1, \text{ otherwise.} \tag{28}$$

Recall that m is the (outer loop) iteration counter. The decreasing step size $sn=1/m$ follows the MSA. Based on Eqs.(23), (26), and (28), the gradient projection-based descent direction method derives the following path flow updating scheme in the inner loop of the algorithmic framework.

$$\begin{aligned}&rw,\tau,pn+1=\mathrm{Max}0,\ rw,\tau,pn-sn\times\nabla TEw\tau\boldsymbol{r}n,\ \forall p\in P(w,\tau)\backslash p^*,\\&r(w,\tau,p*)n+1=d(w,\tau)-p\in Pw,\tau\backslash p*rw,\tau,pn+1.\end{aligned} \tag{29}$$

5.2 Marginal Emission Evaluation

The gradient projection-based method requires evaluation of path marginal emissions, $\eta_{w,p,\tau}, \forall w,\tau,p$, and their constituent link emissions $mEC_{l,t}, \forall l,t$. 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

$qmax(l,t)$ Maximum flow rate on link l at time t

$capin(l,t)$ Inflow capacity of link l at time t

$capout(l,t)$ 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., $length(l)/w$

5.2.1 Link Marginal Emissions for One Vehicle

This research derives the link marginal emission due to an additional vehicle arriving at link l at time t, based on the cumulative arrival and departure curves. The link marginal emission is referred to as the change in the emission on the link at time t due to an additional unit of link inflow. We also illustrate the relationship between link marginal travel time (or delay) $mTT(l,t)$ and link marginal emission $mEC(l,t)$.

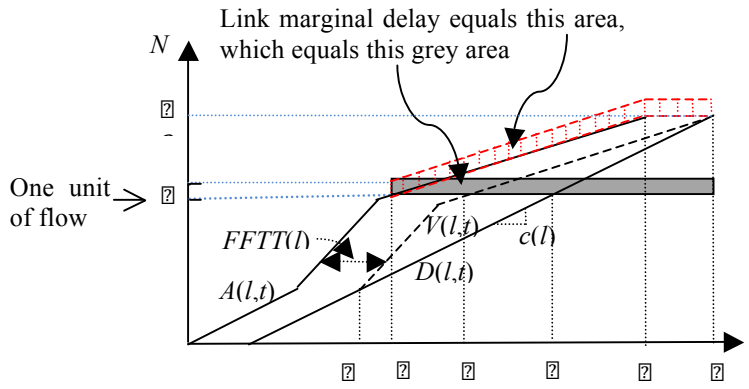

Fig. 6 Illustration of link marginal travel time on a congested link

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 $tlqs$ and dissipates at tlB on the link. Let tl', tl'', and tl''' be the times when an additional vehicle ($n1$) arriving at link l, joining in the queue of the bottleneck, and leaving the link, respectively. According to Ghali and Smith (1995), the link marginal delay (or travel time) due to the additional vehicle is equal to the gray area (i.e., $mTTl,tl'=tlB-tl'$). On the other hand, if this additional vehicle does not encounter a queue, the link marginal delay equals $FFTT(l)$. They also showed that the vehicles arriving between tl' and tlA experience the additional delay $1/c(l)$, because it takes $1/c(l)$ to discharge this perturbation vehicle. In our notation, this means that the change in the delay $\Delta tw=1/c(l)$. The link marginal delay at time tl' can be derived from **Fig. 6** as follows.

$$mTTl,tl'=tlB-tl'=n2-n11c(l). \tag{30}$$

Let $\Delta ECl,t$ be the change in emission for one vehicle entering the link behind the unit perturbation vehicle. The following proposition derives the link marginal emission and states its relationship with link marginal travel time.

Proposition 2: The link marginal emission is $mECl,t=n2-n1\Delta ECl,t$ and a multiple of the link marginal travel time (or marginal delay): $mECl,t=\gamma(l)\times mTTl,t$.

Proof:

Since the change in delay is the sum of the change in the time spent in queue and the change in the actual free flow travel time (i.e., $\Delta twl,t=\Delta tQl,t+\Delta tF(l,t)$) and $\Delta twl,t=1/c(l)$,

$$\Delta tFl,t=1c(l)-\Delta tQ(l,t), \tag{31}$$

According to Eq.(10), the change in the time spent in queue, $\Delta tQl,t$, is a fixed multiple of the change in the delay, $\Delta twl,t$:

$$\Delta tQl,t=\Delta twl,t1-\upsilon Q(l)\upsilon F(l)=1/c(l)1-\upsilon Q(l)\upsilon F(l). \quad (32)$$

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

$$\Delta ECl,t=\Delta tQl,tECvQ(l)+\Delta tFl,tECvF(l)$$
$$=1c(l)11-\upsilon Q(l)\upsilon F(l)ECvQ(l)+(1-11-\upsilon Q(l)\upsilon F(l))ECvF(l). \quad (33)$$

Based on Eq.(30), the link marginal emission can be derived as follows.

$$mECl,t=n2-n1\Delta ECl,t. \quad (34)$$

Let $\gamma(l)=11-\upsilon Q(l)\upsilon F(l)ECvQ(l)+(1-11-\upsilon Q(l)\upsilon F(l))ECvF(l)$. Then, $mECl,t=n2-n1\gamma(l)c(l)=\gamma(l)\times mTTl,t$.

This completes the proof.

5.2.2 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 tl' (= τ, 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).

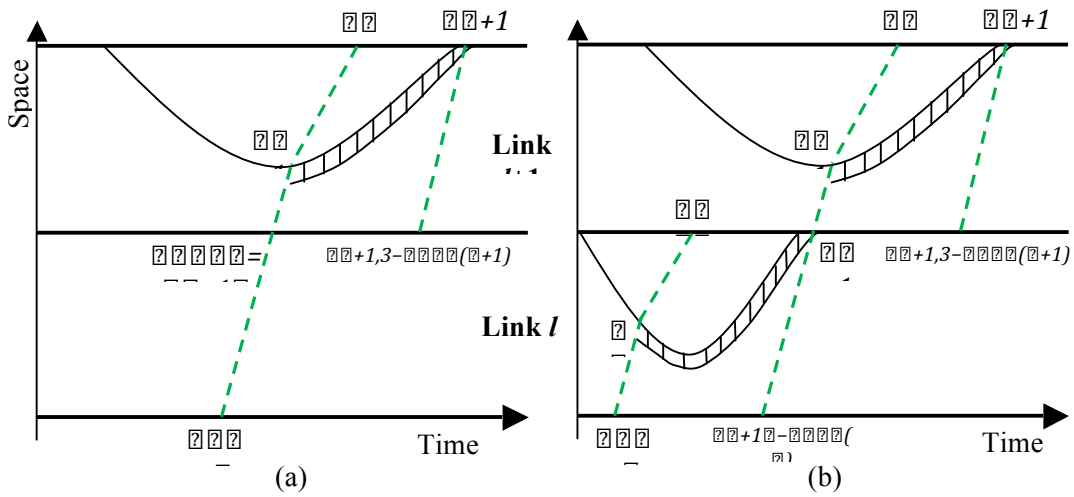

Fig. 7 Path marginal emission analysis without queue spillback

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

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 $tl+1'$ to $tl+1,1$. After detecting the next nearest event timestamp, $tne=tl+1,1$, which corresponds to a queue spillback event, we need to trace back to link l to take time period $[tl+1,1-FFTT(l), tl+1,2-FFTT(l)]$ into consideration. Finally, we move to link $l+1$ to cover the last impacted regime from $tl+1,2$ to $tl+1,3-FFTT(l+1)$. 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 tl' to $tl,3$, and the second link's impacted regime is from $tl,3$ to $tl+1,3$. Again, the perturbation can enter link $l+1$ at time $tl+1,2$ only when the queue on link l vanishes at time $tl,3$.

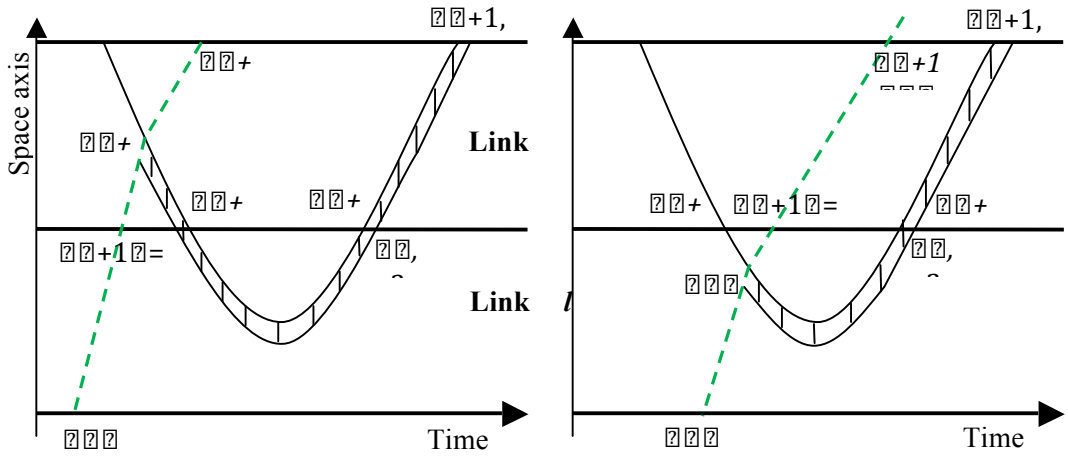

Fig. 8 Path marginal emission analysis with queue spillback from downstream link

Based on the above analysis for the impacted regimes on two consecutive links due to an additional vehicle, this research proposes the method for evaluating the path marginal emission of a path $p \in P(w,\tau)$ with multiple links $l=1,...,L$. 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 τ, the proposed algorithm, presented in Algorithm 1, keeps accumulating path marginal emission by adding the link marginal emission obtained using Algorithm 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,τ)

Initialize $t'=\tau$, $l=1$, and $\eta wp\tau=ecwp\tau$, where $ecwp\tau$ denotes the emission of path $p \in P(w,\tau)$.

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

Step 1: Obtain the next (nearest) event timestamp, *tne*. The next event may correspond to the end time of congestion (i.e., *t3*), the beginning of queue spillback (i.e., *t1*) or $tne=t'+FFTT(l)$ (under uncongested conditions).

Step 2: Evaluate link marginal emission, $\eta l(t',tne)$, using Algorithm 2 (presented in Section 5.2.3).

Step 3: Accumulate the path marginal emission $\eta wp\tau=\eta wp\tau+\eta l(t',tne)$.

Step 4: Move to the next link, and update the starting time, t', of next analysis period. If next event timestamp *tne* corresponds to the end time of congestion or is in a

uncongested time period, then $t'=tne+FFTT(l)$ and $l=l+1$; otherwise (tne corresponds to a queue spillback case) $l=l-1$ and $t'=tne-FFTT(l)$.

End

Evaluation of link marginal emission needs to determine the change in link emission due to an additional vehicle. Fig. 9 depicts the space-time cumulative flow count diagram, with given cumulative arrival and departure count profiles, $A(l,t)$ and $D(l,t)$, as the boundary conditions. Assume that an additional vehicle ja arrives at the upstream end of link l at time t', enters the congested regime at t'', and finally departs from the downstream end of the link at time t'''. Vehicle ja should reach the stop bar at time $t_*=t'+FFTT(l)$, if it does not encounter a queue. Let j' be the vehicle in the existing vehicle set that enters the link right before time t'. Based on the FIFO principle, the impact of the additional vehicle occurs on vehicles $j'+1$, $j'+2$, …, $j3$, arriving behind j', where $j3$ is the vehicle leaving the link right before time $t3$ (when the queue vanishes) or the last impacted vehicle. Also, there is no impact on those vehicles reaching the stop bar after time $t3$ as they do not encounter a queue. Thus, evaluating link marginal emissions focuses on traffic state changes inside the polyhedron t'', t''', $t3$, $t2$ and $t1$, as shown in Fig. 9. Superscripts – and + are used in the figure and the following discussion to represent prior and posterior traffic states for the perturbation. For example, $N-(x,t)$ and $N+(x,t)$ correspond to the cumulative flow counts at position x and time t before and after the perturbation flow arrives, respectively.

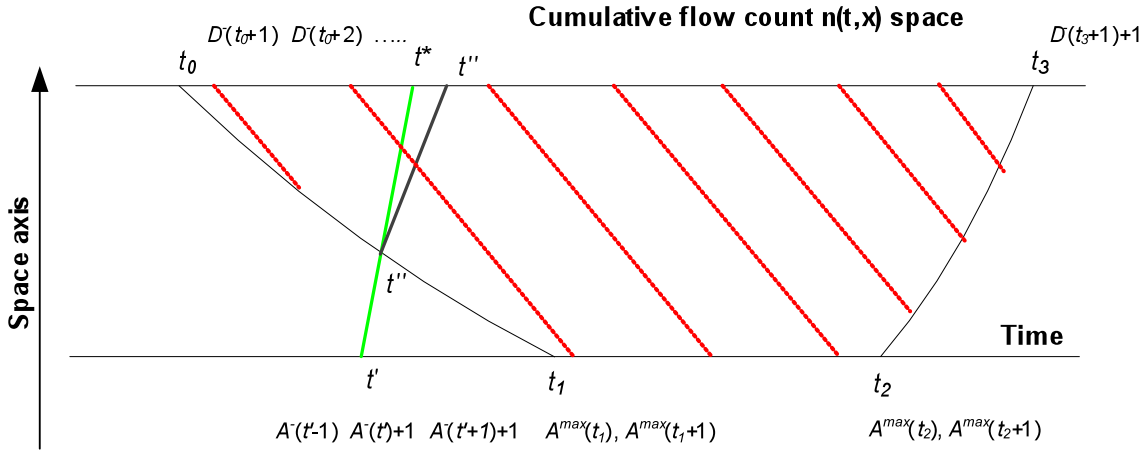

Fig. 9 Illustration of the polygon in which traffic state changes need to be analyzed for evaluating link marginal emission due to an additional vehicle arriving at time t'

The cumulative arrival counts after the arrival of an additional vehicle at time t' are determined as follows:

$$A+l,t=\min Amaxl,t,A-l,t+1,\ \forall t\geq t', \tag{35}$$

where $Amaxl,t$ is defined by the following queue spillback constraint (Zhou et al., 2015):

$$Amax_{l,t} = D_{l,t-BWTT} + kjam_{(l)} \times length_{(l)} \times nlanes_{(l)}. \quad (36)$$

Due to the outflow capacity constraint, the cumulative departure counts at the downstream end of the link remains unchanged before the end of congestion, $t3$.

$$D^{+}_{l,t} = D^{-}_{l,t}, \ \forall t0 \le t \le t3. \quad (37)$$

Right after $t3$, the last vehicle in the vertical queue will be discharged. Hence,

$$D^{+}_{l,t} = D^{-}_{l,t+1}, \ \forall t \ge t3. \quad (38)$$

Eq.(37) implies that the additional vehicle ja needs to take the original departure time slot of vehicle $j'+1$, due to the FIFO constraint. Then, a vehicle j entering the link after time t' (i.e., an impacted vehicle) will take the original departure time slot of its follower $j+1$; that is, the departure time of vehicle j (denoted as Dep_j) is pushed forward as follows:

$$Dep_j = Dep_{j+1}, \ j = j'+1, j'+2, \ldots, j3. \quad (39)$$

The left portion of Figure 10 depicts a sample vehicle trajectory diagram before adding an additional vehicle ja at time t'. In the right portion of the figure, vehicle ja first drives at free-flow speed on the link and then its position is controlled by the position of lead vehicle j=1 due to the backward wave propagation. Finally, it has to follow vehicle j=2's original trajectory, which is also controlled by vehicle j=1's positions through backward wave, and then ja takes the departure time slot of vehicle j=2.

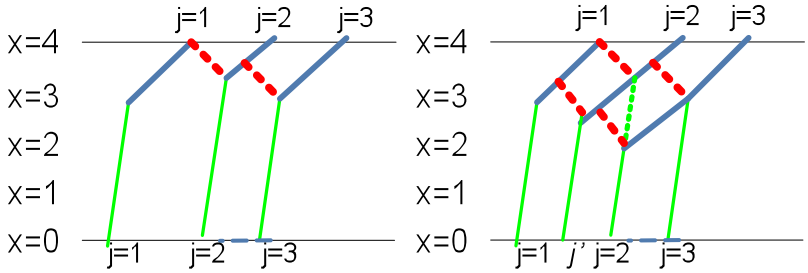

Fig. 10 Illustration of vehicle trajectory changes after the arrival of vehicle ja at time t'

(The dash red lines represent the backward wave controlling a follower vehicle's position.)

To numerically evaluate link marginal emission due to an additional vehicle, one can use Newell's simplified car-following model to reconstruct vehicle trajectories with a perturbation vehicle arriving at time t', and then compare the total link emission before and after the perturbation. However, link marginal emission values are required at numerous perturbation arrival times, repeating this process for each different perturbation arrival time t' ($t0 - FFTT_l \le t' \le t3 - FFTT_{(l)}$) could be extremely computationally expensive.

5.2.3 Traffic State Changes in Space-time Cells

To avoid evaluating link marginal emission for each possible perturbation arrival time on a congested link, this research proposes an approximation method that is able to efficiently determine traffic state changes due to an additional vehicle. The idea begins with analyzing how traffic states in a space-time cell are affected by the perturbation at different vehicle arrival times, where a space-time cell *(x,t)* is defined over space between position *x* and $x+\Delta X$ and over time between time *t* and $t+\Delta T$, as illustrated in Fig. 11. The discussion below focuses on the traffic state change in *N(x,t)* in each space-time cell of a link.

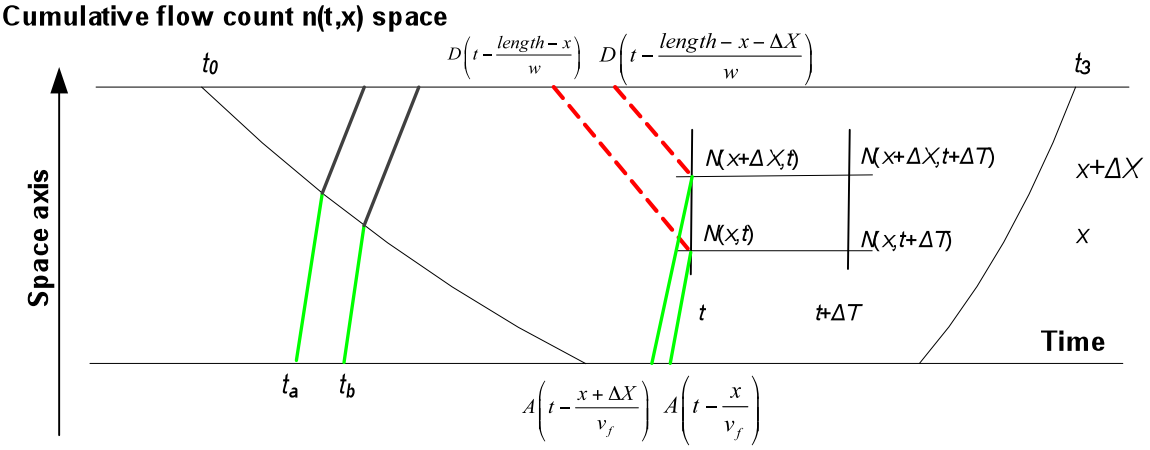

Fig. 11 Illustration of traffic state analysis in a space-time cell

To determine the cumulative flow count in any space-time cell *(x,t)* of a link, the proposed approximation method for determining link marginal emission adopts Newell's three-detector principle (Newell, 1993a):

$$Nx,t=minAl,t-xVf,Dl,t-lengthl-xw+kjaml\times lengthl-x\times nlanesl, \quad (40)$$

where *x* is the distance from the upstream end of link *l* under consideration.

Proposition 3: Consider two different perturbation arrival times *ta* and *tb* at link *l* for the unit perturbation flow, where $t0\le ta<tb<t-x/vf$ (see Fig. 11). The changes in traffic density and volume in a cell *(x,t)* corresponding to the two different perturbation arrival times are the same.

Proof:

Denote *Na+x,t* and *Nb+x,t* as the posterior cumulative counts at space-time position *(x,t)* due to an additional vehicle entering link *l* at *ta* and *tb*, respectively. According to Eq.(35), the posterior cumulative arrival counts at time $t-x/vf$ are the same for both arrival times *ta* and *tb*.

$$Aa+l,t-xVf=Ab+l,t-xVf=A-l,t-xVf+1. \quad (41)$$

According to Eq.(37), the posterior cumulative departure counts at time $t-lengthl-xw$ are also the same for both perturbation arrival times.

$$Da+l,t-lengthl-xw=Db+l,t-lengthl-xw=D-l,t-lengthl-xw \quad (42)$$

Thus, we can obtain the same cumulative flow count at (x, t) under different perturbation arrival times, according to Newell's three detector principle in Eq.(40).

$$Na+x,t=Nb+x,t=$$

$$minA-l,t-xVf+1,D-l,t-lengthl-xw+kjaml\times lengthl-x\times nlanesl. \quad (43)$$

We can further derive that $Na+x+\Delta X,t=Nb+x+\Delta X,t$, so the density of cell x,t is independent of different perturbation arrival times.

$$kx,t=1\Delta X\times nlanes(l)Nx+\Delta X,t-Nx,t. \quad (44)$$

We can also obtain the same flow rate between time t and $t+\Delta T$ under different perturbation arrival times.

$$qx,t=1\Delta TNx+\Delta T,t-Nx,t. \quad (45)$$

This completes the proof.

Let $Arrj$ and $Dep(j)$ be the arrival time and departure time of vehicle j on link l. In light of Proposition 3, the posterior trajectories of an impacted vehicle j (i.e., $t'\le Arrj\le t3-FFTT(l)$) on link l due to an additional vehicle arriving at different times ta and tb are the same, which is denoted as $xa+j,t=xb+j,t,\ Arrj\le t\le Dep(j)$. Because the emission of an impacted vehicle j on link l is determined based on the second-by-second vehicle trajectory or position $x+(j,t)$ on link l, its posterior link emissions due to the unit perturbation flow arriving at time ta and tb are the same, which is denoted as $ECa+j=ECb+j$.

This proposition greatly facilitates evaluating link marginal emission for different perturbation arrival times on a link. The proposed method simply reconstructs trajectories of all impacted vehicles due to an additional vehicle entering a link right after the time $t'=t0-FFTT(l)$ (where $t0$ is the time when the queue starts on that link) and computes corresponding vehicle emissions on that link. These reconstructed trajectories and vehicle emissions can be re-used for evaluating posterior link emissions of impacted vehicles (i.e., $t'\le Arrj\le t3-FFTT(l)$) at different link arrival times t' between $t0-FFTT(l)$ and $t3-FFTT(l)$.

The proposed method is presented in Algorithm 2, which consists of two stages. The first stage, including Steps 1 to 3, is to reconstruct impacted vehicle trajectories and obtain the emission changes due the unit perturbation flow arriving at time $t'=t0-FFTT(l)$. The second stage, Step 4, determines the link marginal emissions for different perturbation arrival times in a specific time period between t' and tne, where tne is defined as the next (nearest) event

timestamp which corresponds to the beginning of queue spillback (i.e., $t1$) or the end time of congestion (i.e., $t3-FFTT(l)$), as shown in Fig. 9.

Algorithm 2: Evaluation of link marginal emission on link l between time t' and tne

Step 1: Reconstruct trajectory according to Newell's simplified car-following model (Zhou et al., 2015) and calculate prior emission for each impacted vehicle, $EC-j,\ j=j0,j0+1,...,j3$, where $j0$ denotes the vehicle arriving right after time $t0-FFTT(l)$ and $j3$ is the vehicle arriving link l right before time $t3-FFTT(l)$.

Step 2: Insert an additional vehicle right after $t0-FFTT(l)$. This vehicle takes the original departure time slot of vehicle $j0$, Dep_{j_0}, and pushes forward the departure times of all impacted vehicles; that is $Dep+j=Dep-j+1$ for $j=j0,j0+1,...,j3$ (i.e., Eq.(39)). Then, adopt Newell's simplified car-following model (Zhou et al., 2015) to reconstruct vehicle trajectory and calculate the posterior emission $EC+j$ for each impacted vehicle $j=j0,j0+1,...,j3$.

Step 3: Calculate the emission change for each impacted vehicle j as $\Delta ECj=EC+j-EC-j$, for $j=j0,j0+1,...,j3$.

Step 4: For the given perturbation arrival time t' and next event time stamp tne, find the vehicle index $j'+1$ with arrival time right after t' and jne with arrival time right before tne. The link marginal emission of impacted vehicles can be approximated as

$$mEC(t',tne)=j=j'+1jne\Delta ECj \tag{46}$$

Note that Steps 1 to 3 of Algorithm 2 are pre-processing steps that are called only once for evaluating link marginal emission of a link, while Step 4 is executed for each different t' and tne.

Recall that, in Ghali and Smith's (1995) approach (see section 5.2.1), the link marginal delay $mTT=t3-t'$. Considering link travel delay as the counterpart of link emission in our research, the link marginal delay can be derived using Eq.(46) as well. Specifically, for each impacted vehicle, its additional travel delay is $\Delta twj=1/capout$, as it has to wait one more headway to be discharged from the vertical queue. In the case of constant $capout$, it is easy to obtain that the number of vehicles arriving at the vertical queue after time t' and before $t3$ is $t3-t'\times capout$. Thus, the corresponding link marginal delay turns out to be

$$\eta l(t',t3)=j=j'+1j3\Delta twj=t3-t'\times capout\times 1capout=t3-t', \tag{47}$$

which is consistent with the result obtained by Algorithm 2.

5.2.4 Before and After Impact Analysis

The subsection presents a thorough examination on how the traffic state in terms of cumulative counts in different space-time regimes is impacted due to an additional vehicle. In the following analysis, the space-time diagram is decomposed into three regimes (see Fig. 12). For narrative convenience, let $D(x,t)=Dl,t-lengthl-xw+kjaml\times lengthl-x\times nlanesl$.

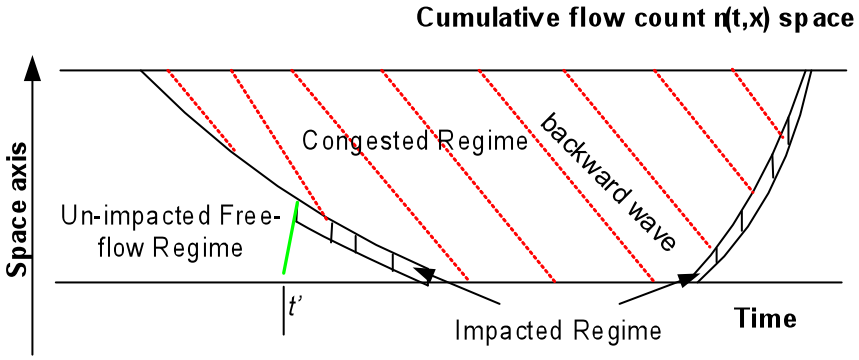

Fig. 12 Three different regimes regarding traffic state change due to an additional vehicle

(1) Congestion regime where $N-x,t=D-x,t$

In the congestion regime, according to Eq.(35),

$$A+l,t-xvf=\min Amaxl,t-xvf,A-l,t-xvf+1> A-(l,t-xvf)\geq D-(x,t)=D+(x,t) \quad (48)$$

Then, based on Newell's three-detector principle in Eq.(40), we can derive that $N+x,t=\min A+l,t-xvf,D+(x,t)=N-x,t$. Thus, for a space-time position x,t in the congestion regime where $N-x,t=D-x,t$, the cumulative count is the same before and after the arrival of an additional vehicle; that is, $N+x,t=N-x,t$. In other words, the traffic state in this congested regime is controlled by the unchanged backward wave, so the increase in the cumulative arrival count does not make the forward wave become the "winner" in the minimization relationship in Eq.(40); the arrival of an additional vehicle does not change the congested state. Accordingly, if all the four corner points of a time-space cell belong to the congestion regime, the density and flow in that cell do not change due to an additional vehicle.

(2) Unaffected free-flow regime where $N-x,t=A-(l,t-xvf)$ and $A-(l,t-xvf)+1<D-(x,t)$

The condition, $A-(l,t-xvf)+1<D-(x,t)$, is used to distinguish the unaffected free-flow regime from the impacted free-flow regime. In the formal case,

$$A+l,t-xvf=A-(l,t-xvf)+1<D-x,t=D+x,t. \quad (49)$$

Thus, the original cumulative arrival count is low enough so that an increase by one vehicle does not change the free-flow traffic state at position (x, t), where the forward wave still dominates.

$$N+x,t=\min A+l,t-xvf,D+(x,t)=A-l,t-xvf+1. \quad (50)$$

In the unaffected free-flow regime, it is easy to show that the corresponding flow and density do not change due to an additional vehicle.

(3) Impacted free-flow regime, where $N^-(x,t)=A^-(l,t-x/v_f)$ and $A^-(l,t-x/v_f)+1\geq D^-(x,t)$

It is easy to derive that

$$N^+(x,t)=\min\{A^+(l,t-x/v_f),D^+(x,t)\}=\min\{A^-(l,t-x/v_f)+1,\ D^-(x,t)\}=D^-(x,t). \quad (51)$$

That is, the traffic state in terms of cumulative count of a time-space position in the impacted free-flow regime will change from uncongested $A^-(l,t-x/v_f)$ to congested $D^-(x,t)$, due to an additional vehicle.

The above analysis results are summarized in Table 2.

Table 2 Three different regimes in terms of traffic state change after adding a vehicle

Regimes	Prior traffic state	Posterior traffic state	Traffic state change
Congested	$N^-(x,t)=D^-(x,t)$	$N^+(x,t)=N^-(x,t)$	Remain unchanged
Unaffected free-flow	$N^-(x,t)=A^-(l,t-x/v_f)$, and $A^-(l,t-x/v_f)+1<D^-(x,t)$	$N^+(x,t)=N^-(x,t)+1$	Remain unchanged, though cumulative arrival count increased by 1
Impacted free-flow	$N^-(x,t)=A^-(l,t-x/v_f)$, and $A^-(l,t-x/v_f)+1>D^-(x,t)$	$N^+(x,t)=D^-(x,t)$	Both traffic density and flow will change due to an additional inflow.

6. Conclusions

This research generalizes the normative capability of DTA to sustainable transportation network modeling. The ESODTA model is proposed to obtain path flows in a congested network where travelers behave cooperatively in selecting paths to minimize total network emission. The mesoscopic DNL model which seamlessly integrates Newell's simplified kinematic wave model and simplified car-following model into a unified framework was developed to enable the internally consistent DTA for temporally cross-resolution and spatially multi-scale emission modeling. Moreover, we presented, for the first time in literature, the computational algorithm of link and path marginal emissions based on Newell's three-detector principle and thorough examination on traffic state changes due to an additional vehicle. These path marginal emissions are critical input to the gradient projection-based descent direction method for solving the ESODTA problem.

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 GSO 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. On the other hand, the proposed DNL model can be encapsulated in traffic emission and energy evaluation and optimization models that require both fine-grained and coarse-grained traffic flow representations. Particularly, the high-resolution, second-by-second vehicle trajectories generated by the DNL model can be used to derive speed and acceleration estimates that are essential input to state-of-the-art emission modeling systems (e.g., MOVES).

7. Acknowledgements

Dr. Jason Lu from the Department of Transportation and Logistics Management, National Chiao Tung University, Taiwan has contributed significantly to this report and related journal papers. Dr. Srinivas Peeta from Purdue University and Dr. Nagui Rouphail from North Carolina State University also offered very constructive comments to the research framework and analysis methods.

8. References

Aziz, H.M.A., 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.

Benedek, C.M. and Rilett, L.R., (1998). Equitable traffic assignment with environmental cost function. Journal of Transportation Engineering, 124, 16-22.

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.

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.

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.

Dowling, R., Skabardonis, A., 1993. Improving average travel speeds estimated by planning models. Transportation Research Record 1366, 68–74.

Friesz, T. L., Luque, J., Tobin, R. L., Wie, B.-Y., 1989. Dynamic network traffic assignment considered as a continuous time optimal control problem. Operations Research 37(6), pp. 893-901.

Gaker, D., Zheng, Y., Walker, J., 2010. Experimental economics in transportation: focus on social influences and provision of information. Transportation Research Record 2156, 47-55.

Gaker, D., Vautin, D., Vij, A., Walker, J. L., 2011. The power and value of green in promoting sustainable transport behavior. Environmental Research Letters 6(3), 1-11.

Garcia, A., Reaume, D., Smith, R. L., 2000. Fictitious play for finding system optimal routings in dynamic traffic networks. Transportation Research Part B 34(2), 147-156.

Ghali, M.O. and Smith, M.J., 1995. A model for the dynamic system optimum traffic assignment problem. Transportation Research Part B 29(3), 155-170.

Greene, D., Schafer, A., 2003. Reducing greenhouse gas emissions from U.S. transportation. Center for Climate and Energy Solutions (http://www.c2es.org/docUploads/ustransp.pdf).

Helali, K., Hutchinson, B., 1994. Improving road link speed estimates for air quality models. Transportation Research Record 1444, 71–77.

Hurdle, V. F., Son, B., 2000. Road test of a freeway model. Transportation Research Part A 34(7), 537-564.

Lafortune, S., Sengupta, R., Kaufman, D. E., Smith, R., 1993. Dynamic system-optimal traffic assignment using a state space model. Transportation Research Part B 27(6), 451-472.

Lawson, T. W., D. J. Lovell and C. F. Daganzo. Using Input-Output Diagram to Determine Spatial and Temporal Extents of a Queue Upstream of a Bottleneck. Transportation Research Record, No. 1572, 1996, pp. 140-147

Leclercq, L., 2007. Hybrid approaches to the solutions of the “Lighthill-Whitham-Richards” model. Transportation Research Part B 41(7), 701-709.

Lighthill, M., Whitham, G.., 1955. On kinematic waves II: a theory of traffic flow on long crowded roads. Proc. Royal Society of London, Part A 229 (1178), 317-345.

Lo, H. K., Szeto, W. Y., 2002. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. Transportation Research Part B 36(5), 421-443.

Lu, C.-C., Mahmassani, H. S., Zhou, X., 2009. Equivalent gap function-based reformulation and solution algorithm for the dynamic user equilibrium problem. Transportation Research Part B 43(3), 345-364

Magnanti, T., Perakis, G., 1997. Averaging schemes for variational inequalities and systems of equations. Mathematics of Operations Research 22(3), 568-587.

Mandavilli, S., Rys, M. J., Russell, E. R., 2008. Environmental impace of modern roundabouts. International Journal of Industrial Ergonomics 38(2), 135-142.

Merchant, D. K., Nemhauser, G. L., 1978. A model and an algorithm for the dynamic traffic assignment problems. Transportation Science 12(3), 183-199.

Mounce, R., Smith, M., 2007. Uniqueness of Equilirbium in steady state and dynamic traffic networks. In: Allsop, R.E., Bell, M.G.H., and Heydecker, B.G. (Eds), Transportation and Traffic Theory, Elsevier, 281-299.

Munoz, J. C., Laval, J. A., 2006. System optimum dynamic traffic assignment graphical solution method for a congested freeway and one destination. Transportation Research Part B 40(1), 1-15.

Nagurney, A., Ramanujam, P. and Dhanda, K.K., 1998. Multimodal traffic network equilibrium model with emission pollution permits: Compliance vs noncompliance. Transportation Research, Part D, 35, 349–374.

Nagurney, A., Dong, J. and Mokhtarian, P.L., 2002. Traffic network equilibrium and the environment: A multicriteria decision-making perspective. In E. Kontoghiorges, B. Rustem, & S. Siokos (Eds.), Computational methods in decision-making, economics and finance (pp. 501–523). Dordrecht: Kluwer Academic Publishers.

Newell, G. F., 1993a. A simplified theory on kinematic waves in highway traffic, part I: general theory. Transportation Research Part B 27(4), 281-287.

Newell, G. F., 1993b. A simplified theory on kinematic waves in highway traffic, part II: queueing at freeway bottlenecks. Transportation Research Part B 27(4), 289-303.

Newell, G. F., 1993c. A simplified theory on kinematic waves in highway traffic, part III: multi-destination flows. Transportation Research Part B 27(4), 305-313.

Newell, G., 2002. A simplified car-following theory: a lower order model, Transportation Research B 36(3), 195–205.

Ni, D., Leonard, J. D., Williams, B. M., 2006. The network kinematic waves model: a simplified approach to network traffic. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 10(1), 1-14.

Panis, L., Broekx, S., Ronghui, L., 2006. Modeling instantaneous traffic emission and the influence of traffic speed limits. Science of the Total Environment 371(1-3), 270–285.

Peeta, S., Mahmassani, H. S., 1995. System optimal and user equilibrium time-dependent traffic assignment in congested networks. Annals of Operation Research 60(1), 81-113.

Qian, Z. and Zhang, H.M., 2011. Computing individual path marginal cost in networks with queue spillbacks. Transportation Research Record 2263, 9-18.

Richards, P. I., 1956. Shock waves on the highway. Operations Research 4(1), pp.42-51.

Sbayti, H., Lu, C.-C., Mahmassani, H. S., 2007. Efficient implementations of the method of successive averages in simulation-based DTA models for large-scale network applications. Transportation Research Record 2029, 22-30.

Shen, W., Nie, Y., Zhang, H. M., 2007. On path marginal cost analysis and its relation to dynamic system-optimal traffic assignment. In: Allsop, R.E., Bell, M.G.H., and Heydecker, B.G. (Eds), Transportation and Traffic Theory, Elsevier, 319-352.

Shen, W., Zhang, H. M., 2009. On the morning commute problem in a corridor network with multiple bottlenecks: Its system-optimal traffic flow patterns and the realizing tolling scheme. Transportation Research Part B 43(3), 267-284.

Tzeng, G.H., and Chen, C.H., 1993. Multiobjective decision making for traffic assignment. IEEE Transactions on Engineering Management, 40, 180–187.

U.S. EPA, (2009). Draft Moto Vehicle Emission Simulator (MOVES) – Software Design and Reference Manual. Technical report EPA-420-B-09-007, U. S. Environmental Protection Agency.

Wie, B.-W., Tobin, R. L., Friesz, T. L., 1994. The augmented Lagrangian method for solving dynamic network traffic assignment models in discrete time. Transportation Science 28(3), 204-220.

Zhang, K., Mahmassani, H. S., Lu, C.-C., 2009. Probit-based time-dependent stochastic user equilibrium: reformulation and solution algorithm. Paper presented at the 88th Annual Meeting of Transportation Research Board, January 11-15, 2009, Washington, D.C., USA.

Zhang, Y., Lv, J. and Ying, Q., 2010. Traffic assignment considering air quality. Transportation Research Part D, 15, 497-502.

Zhou, X., Tanvir, S., Lei, H., Taylor, J., Liu, B., Rouphail, N.M. and Frey, H.C. (2015) Integrating a simplified emission estimation model and mesoscopic dynamic traffic simulator to efficiently evaluate emission impacts of traffic management strategies. Transportation Research Part D, 37, 123-136.

Zhu, F., Lo, H. K. and Lin, H.-Z., 2013. Delay and emissions modelling for signalised intersections. Transportmetrica B: transport dynamics 1.2: 111-135.

Ziliaskopoulos, A. K., Mahmassani, H. S., 1993. Time dependent shortest-path algorithm for real-time Intelligent Vehicle Highway System applications. Transportation Research Record 1408, 94-100.

Ziliaskopoulos, A. K., 2000. A linear programming model for the single destination system optimum dynamic traffic assignment problem. Transportation Science 34(1), 1-12.