A Hybrid Route Choice Model for Dynamic Traffic Assignment

Zhen (Sean) Qian · H. Michael Zhang

Abstract Network user equilibrium or user optimum is an ideal state that can hardly be achieved in real traffic. More often than not, every day traffic tends to be in disequilibrium rather than equilibrium, thanks to uncertainties in demand and supply of the network. In this paper we propose a hybrid route choice model for studying non-equilibrium traffic. It combines pre-trip route choice and en-route route choice to solve dynamic traffic assignment (DTA) in large-scale networks. Travelers are divided into two groups, habitual travelers and adaptive travelers. Habitual travelers strictly follow their pre-trip routes which can be generated in the way that major links, such as freeways or major arterial streets, are favored over minor links, while taking into account historical traffic information. Adaptive travelers are responsive to real-time information and willing to explore new routes from time to time. We apply the hybrid route choice model in a synthetic medium-scale network and a large-scale real network to assess its effect on the flow patterns and network performances, and compare them with those obtained from Predictive User Equilibrium (PUE) DTA. The results show that PUE-DTA usually produces considerably less congestion and less frequent queue spillback than the hybrid route choice model. The ratio between habitual and adaptive travelers is crucial in determining realistic flow and queuing patterns. Consistent with previous studies, we found that, in non-PUE DTA, supplying a

Z. (Sean) Qian · H. M. Zhang (✉)
Department of Civil and Environmental Engineering,
University of California Davis, Davis CA, USA
e-mail: hmzhang@ucdavis.edu

Z. (Sean) Qian
e-mail: zqian@ucdavis.edu

medium sized group (usually less than 50%) of travelers real-time information is more beneficial to network performance than supplying the majority of travelers with real-time information. Finally, some suggestions are given on how to calibrate the hybrid route choice model in practice to produce realistic results.

Keywords Dynamic traffic assignment · Hybrid route choice · Traveler heterogeneity

1 Introduction

Dynamic traffic assignment (DTA) aims at determining the time-dependent link flow pattern given a set of Origin-Destination (O-D) traffic demand for a given time period, by following certain behavioral assumptions on the travelers and physical models that govern the flow of traffic on links and through junctions. The study of DTA comprises three main elements: travelers' route choices, departure time choices and the flow dynamics of traffic. This paper is concerned with travelers' route choices.

Route choice is naturally made by individuals (Bovy and Stern 1990), therefore lends itself to discrete choice models. Various forms of probabilistic route choice models, such as Multinomial Logit (Dia et al. 2001), Cross Nested Logit (Vovsha and Bekhor 1998), Path Size Logit (Ramming 2002), C-Logit (Cascetta et al. 1996) and so forth have been proposed. They assume the probability of choosing each route from a route set for each traveler is related to his utility that consists of various attributes, such as gender, age, perceived travel time, pre-trip information, etc. Some of these models have recently been extended to investigate how real-time information can affect travelers' route choices (Hawas 2004; Lee et al. 2010; Mahmassani and Liu 1999; Peeta and Yu 2005; Reddy et al. 1995; Xu et al. 2011) and the stochasticity of route choices (Nie et al. 2011). The real-time information was also considered as part of the cognitive cost of individual travelers in route choice modeling (Gao et al. 2011).

Out of computational considerations, route choices in dynamic traffic assignment are modeled in a much simpler way than the previously mentioned discrete choice models: travelers simply choose the cheapest or shortest route that are present to them. The resulting flow pattern is usually referred to as a user-equilibrium (UE) flow pattern. Some studies consider the stochastic route choices in steady-state networks by assuming certain traffic information are provided to travelers (Bovy and Fiorenzo-Catalano 2007; Ukkusuri and Patil 2007). In the dynamic context, there are generally two types of UE in the literature. One is the so-called Boston User Equilibrium (BUE) (Friesz et al. 1993), which is an adaption of the static Wardroppian UE. It assumes a traveler chooses the shortest route only based on the prevailing traffic condition at the time of his choice decision (Carey and Ge 2011; Kuwahara and Akamatsu 2001; Ran et al. 1993), also known as the minimum instantaneous

cost path (Ghali 1995). The other UE type is the so-called Predictive User Equilibrium (PUE). Under this behavioral assumption, travelers choose the shortest route based on "anticipated" travel times, or travel times that they actually experienced from previous days. The result is a UE in which the actual travel times/costs for travelers from any O-D pair are minimal and identical (Friesz et al. 1993), regardless of the routes they take.

In real life, travelers' route choice behavior is likely to be more complex than what was assumed in both BUE and PUE. For example, travelers may not consider all the possible routes but have several pre-trip routes in mind prior to their departure, which are selected from their day-to-day traveling experiences. Moreover, these pre-selected routes may not be user-optimal ones. Although travel time and schedule delay costs are dominant factors in travelers' route choice decisions, several other factors, such as road accessibility, pavement conditions, and so on, may influence their decisions as well. Besides these factors, a traveler's personality should also play an important role in his or her route choice. For example, a conservative traveler may stick to his chosen route from day to day while an adventurous traveler may be more willing to explore new routes based on his actual travel experiences. Thus real traffic is more likely to be the product of various types of choice decisions rather than cost-minimizing BUE or PUE applied uniformly across the entire traveling population. It is therefore of particular interest to develop a route choice model that combines various types of traffic information and considers various kinds of travelers. However, such route choice models, particularly in the context of traffic disequilibria in large-scale dynamic networks, are rare in the literature. Peeta and Mahmassani (1995) are among the few who analyzed the combination of BUE and PUE. They distinguished travelers by their route choices following either of system optimal, UE, historical and real-time information, in order to predict the future O-D demands and to optimize the provision of real-time information. Pel et al. (2009) also adopted route choices other than BUE/PUE. They introduced a hybrid route choice model where all travelers have a pre-trip route, but they all consider real-time traffic conditions in seeking the new routes. This model, however, requires path enumeration and its application to large-scale networks is very limited. In addition, Kant (2008) conducted an interesting study that combines BUE and PUE to form a new type of UE. He assumes travelers' travel cost consists of the actual travel costs and the expected travel costs. Then the solution framework of a PUE is directly applied to solve the new "combined" UE. However, this approach requires a pre-determined route set with very limited number of routes for each O-D pair, and is thus also problematic in handling large-scale networks.

As mentioned earlier, route choice, departure time choice and traffic flow evolution are three essential elements of a DTA problem. Even with the same route/departure time choice model, various flow models could result in considerably different flow patterns. In the literature, several flow models were introduced to formulate the DTA problem, which include exit-flow function based traffic model (Merchant and Nemhauser 1978), the delay-function based model (Friesz et al. 1993), the point queue (PQ) model (Smith 1993), the

spatial queue (SQ) model (Kuwahara and Akamatsu 2001) and kinematic wave (LWR) model (Daganzo 1994). Nie and Zhang (2005) compared these various kinds of so-called link models and found that the kinematic wave model, which models queue spillback in the form of shock waves, provides a realistic description of traffic flow propagation. However, queue spillback in large-scale networks, though exists in the real world, may propagate in an unexpected way that could lead to serious network gridlock, which in most cases is caused by unrealistic route choices. Daganzo (1998) explored the case of queue spillover with a PUE route choice under the spatial-queue model, and showed that the results could be naturally chaotic in the sense that flow patterns may be fairly sensitive to small changes in the network settings. This indicates that the stability of the PUE state is closely related to route choice and queue spillback. Therefore, for large-scale networks, it is crucial to analyze how a certain route choice model interacts with the queue spillback mechanism such that it can be appropriately calibrated in real world applications.

Rather than treating all travelers identically, this paper speculates that some travelers are likely to follow their pre-determined routes while others update their routes en-route in response to real-time information. Following this, we propose a hybrid route choice model coupled with the CTM implementation of the kinematic wave traffic flow model (Lighthill and Whitham 1955; Richards 1956), and show how route choices and queue spillback affect the resulting flow patterns. Computational procedures to realize the models and numerical experiments will be provided, and issues related to model calibration are also discussed.

The rest of this paper is organized as follows. We first briefly review the route choice models embedded in dynamic user-equilibrium (DUE) in Section 2, then propose a hybrid route choice model in Section 3. Section 4 shows, in a small example, the interaction between route choice and queue spillback phenomena. In Section 5 we analyze the flow patterns and network performance with respect to parameters in the hybrid route choice model in both a synthetic and a large-scale real network. Finally, we conclude the paper in Section 6, with a discussion about the calibration of route choice parameters.

2 Route choice models embedded in DUE

Because the DUE route choice is central to our paper, in this section we recap both the PUE route choice and BUE route choice models used in dynamic traffic assignment.

2.1 PUE (predictive) route choice

PUE assumes travelers choose their routes prior to their departure, and then strictly follow them till they reach their destinations. Their choices are based on the shortest paths with respect to actual travel time/cost they experience.

On one hand, a dynamic network loading (DNL) procedure is used to determine the time-varying link flow and the actual travel time/cost, and then to determine the routes traveled; On the other hand, the DNL procedure requires knowing the route choices of all travelers prior to their departure. Therefore, PUE usually requires an iterative procedure in which DNL and shortest path calculation alternate until both the route choice and flow patterns converge.

According to PUE, all travelers can perfectly predict the traffic conditions of the entire network at any time based on their day-to-day experience. They are aware of their actual travel time/cost prior to their departure which is used to determine their shortest routes, regardless of real-time traffic information. This type of route choice is more suitable to model regular commuting traffic without disruptions (such as incidents) for a prolonged period of time during which the total traffic demand remains relatively constant.

2.2 BUE (en-route) route choice

The DTA in Boston User Equilibrium can be decomposed to all-or-nothing static traffic assignment for each time interval (Kuwahara and Akamatsu 2001). In other words, travelers will first determine their routes based on free-flow travel time of all the links in the network, and then the network is loaded by the O-D demand according to the route choices till the first assignment time interval ends. At the beginning of the next time interval, travelers update their route choices by taking the shortest paths with respect to the prevailing (instantaneous) travel times/costs at that time. Flow and travel cost of each link will then be updated based on the new route choice decisions. Repeat this process where the update of route choices and network loading alternate till all the travelers end their trips. Thus, only one network loading is needed and each traveler acts as if he makes en-route decisions of route choice to be taken at each assignment time interval according to the instantaneous travel times/costs. Therefore, we also call the route choice model embedded in the BUE en-route route choice.

The en-route route choice assumes that travelers only have real-time information about the current network conditions, and since they do not predict traffic conditions in the future, they always make the route choice decision that is most appealing at the current time interval, although the chosen route can be far from the best route when the actual travel time/cost is considered. This type of en-route route choice may describe the travelers' behavior in response to a sudden change of demand or roadway capacity where travelers are unable to obtain historical day-to-day information and have to deviate from their pre-trip routes. With the adoption of variable message boards, highway advisory radio and mobile internet devices such as smart phones, travelers nowadays can get up-to-the-minute information in some travel market such that they can make en-route route adjustments in response to changes in network traffic conditions rather than sticking to their pre-determined routes.

3 A hybrid route choice model

In this section, we propose a route choice model which combines pre-trip route choice and en-route route choice and incorporates several factors other than travel time/cost.

3.1 The general model

Suppose the network is represented by a directed graph that includes a set of nodes, N, and a set of links, A. Let a denote the link index, $a \in A$. Let R and S denote the set of origin nodes and destination nodes, respectively. $r - s$ represents an O-D pair, where $r \in R$ and $s \in S$. K_t^{rs} and q_t^{rs} is the set of paths and O-D demand for an O-D pair $r - s$ departing at time t, respectively. The generalized travel cost of commuters departing at time t on path p of O-D pair rs, $c_p^{rs}(t)$, consists of I number of terms $(w_{1,pt}^{rs}, w_{2,pt}^{rs}..., w_{I,pt}^{rs})$ which represent those factors that travelers perceive on path p of O-D pair rs departing at time t (including travel time, schedule delay cost, toll, etc.) and are weighted by scalers λ_i,

$$c_p^{rs}(t) = \sum_{i=1}^{I} \lambda_i w_{i,pt}^{rs} \tag{1}$$

Let $[0, T]$ be an assignment horizon (i.e., the analysis period). The network is assumed to be empty at $t = 0$. Corresponding to the assignment period, we define a loading horizon $[0, T']$, where T' marks the time when the network is empty. Furthermore, let ϕ_a denote an assignment interval, a discrete duration during which the departure flow rate for any O-D pair is assumed to be constant (m_a is the number of assignment intervals, i.e., $T = m_a \phi_a$). ϕ_l is the loading interval, a discrete duration during which network conditions are assumed to be stationary (a loading horizon consists of m_l loading intervals of uniform length, i.e., $T' = m_l \phi_l$). ϕ_a must be a multiple of ϕ_l.

We introduce two groups of travelers: travelers who are willing to deviate from their pre-determined routes and those who are not. The reason is simple. Some conservative travelers, once they determine which routes to take and get familiar with those particular routes, would rather stick to them than risking on finding new (or unknown) routes that may actually turn out to be worse than their previous routes, unless the congestion they experience in their current routes becomes unacceptable to them. Those travelers are normally reluctant to deviate from their prescribed routes. We call this group of travelers **habitual travelers** (the proportion of this type of travelers is $1 - \theta$). On the other hand, some adventurous travelers are more willing to explore new routes in response to their travel experience and/or up-to-date traffic information. They may be equipped with devices that offer real-time navigation, or they may be familiar with the entire network and are able to change their routes to avoid the congestion. We call this group of travelers **adaptive travelers** (the proportion of this type of travelers is θ). In a real network, the proportion θ, also referred

to as the **Diversion Ratio**, may not be a constant and may change with respect to network conditions. For example, the diversion ratio can increase in the event of a major accident or highway reconstruction project. Nevertheless, we expect the diversion ratio to be relatively stable for a network at least in the short run barring the occurrences of various major incidents.

For the habitual travelers, their routes are determined based on a number of factors, such as travel distance, historical travel times, and personal preference for major streets and freeways. These routes, however, may not be the same as the DUE routes when everyone is a habitual traveler, since now some of the travelers are adaptive travelers, PUE is no longer achievable. Let P_t^{rs} denote the set of those routes that habitual travelers departing at time t between O-D pair rs strictly follow (The generation of P_t^{rs} will be discussed in Section 3.2). The proportion of travelers who use a path $p \in P_t^{rs}$ in the group of habitual travelers, also known as the prescribed route rate (Pel et al. 2009), is assumed to follow the Logit model with respect to the generalized travel cost,

$$\pi_p(t) = \frac{\exp(-c_p(t))}{\sum_{p' \in P_t^{rs}} \exp(-c_{p'}(t))} \tag{2}$$

Therefore, the number of habitual travelers who depart at time t between O-D pair rs and use path $p \in P_t^{rs}$ is,

$$q_p(t) = (1 - \theta)q_t^{rs}\pi_p(t) \tag{3}$$

For adaptive travelers, we assume that they always take their respective shortest path with respect to the instantaneous travel cost at each time interval. Adaptive travelers behave in a similar way as in the en-route route choice embedded in BUE, but the time period at which travelers update their shortest paths using the instantaneous travel cost can be relaxed from the assignment interval, ϕ_a in the BUE, to an arbitrary time interval in multiples of the loading time interval, i.e. $\gamma\phi_l$ (where γ is an integer). $\gamma\phi_l$ indicates how frequent the adaptive travelers are able to obtain up-to-date traffic information and choose an alternative route if necessary. It is easy to see that if all the travelers are adaptive travelers (i.e. $\theta = 1$) and let $\lambda = \phi_a/\phi_l$, then the hybrid route choice model essentially solves BUE.

It is crucial to define the instantaneous travel time of link a at entry time t, $\tau_a(t)$, which equals $l_a/s_a(t)$ where $s_a(t)$ is the instantaneous travel speed of link a at entry time t and l_a the length of link a. Given the density of link a at time t, $k_a(t)$, $s_a(t)$ is estimated by,

$$s_a(t) = \begin{cases} u_a & \text{if } k_a(t) \leq k_{a,m} \\ \dfrac{k_{a,j} - k_a(t)}{k_{a,j} - k_{a,m}} \dfrac{q_m}{k_a(t)} & \text{if } k_a(t) > k_{a,m} \end{cases} \tag{4}$$

where $k_{a,j}$ is the jam density of link a, $k_{a,m}$ the critical density and q_m the maximum flux of link a (also known as the capacity). If the density of link a at

time t is smaller than $k_{a,m}$, then the instantaneous travel speed is the free-flow speed of link a, i.e. u_a; otherwise, it equals to the division of the flux of link a at time t by its density at time t where the flux can be solved by the triangular fundamental diagram of link a.

A DTA with this hybrid route choice model no longer requires an itera-tive solution procedure and the resultant flow pattern does not satisfy user equilibrium conditions in any sense. Instead, a one-shot DNL is applied to perform the assignment. During the DNL process, a shortest path calculation is needed in every certain time interval to obtain the new routes for those adaptive travelers, and the DNL is finished when every traveler reaches her destination.

3.2 Generating pre-trip route sets

We discuss in this subsection how those habitual travelers choose their pre-trip routes prior to the execution of the DTA.

An easy and reasonable way of generating pre-trip routes is to include the K shortest paths (with respect to free-flow travel time) in the set of pre-trip routes for each O-D pair *rs*, and then apply Eq. (3) to stochastically assign some habitual travelers with one of those routes. This method is referred as "ordinary K shortest path (KSP) generation" in the rest of the paper.

Though the ordinary KSP generation provides several alternative routes for habitual travelers, it usually does not result in a realistic route set because (1) those travelers may have a special affinity to the freeway and major arterial streets, and (2) they may use their travel experiences in addition to the free-flow travel times to determine their routes. To address these problems, we proposed two modifications to the the ordinary KSP generation method.

3.2.1 Road-hierarchy-based route set generation

Habitual travelers, being conservative, may prefer the freeway and major arterials over minor streets even if the latter may have shorter travel times. Some empirical observations seem to support this statement. For example, when a freeway is highly congested but both the on-ramp and off-ramp adjacent to the congested section are not used, one can bypass the congestion by first taking the off-ramp then use the on-ramp to get back to the freeway to save travel time. However, few travelers actually do this in reality, because most travelers would rather not bother to make extra efforts (such as changing lanes, switching to unfamiliar links, etc.) to reduce their delay unless this reduction is significant. Furthermore, when travelers select their own set of possible routes, they may mainly look at those major roadways and those minor streets are negligible. For instance, the least costly (first best) route choice of a traveler is a freeway connecting his house and office. Although a route consisting of the same freeway and with, however, other minor links from house/office to the freeway is theoretically his second least costly choice

in terms of travel time/cost, he may take the route with the major arterial connecting his house and office as the second best choice. When the first best route is highly congested, one may consider the major arterial rather than the theoretical second best route. Now, various navigation tools also provide alternative routes that differ in terms of freeways and major arterials, rather than those that only differ in minor streets.

The above empirical evidence suggests that habitual travelers' route choice may be hierarchical, that is, they divide the network into multi-level networks, with freeways and major arterials constitute the upper-level network and minor streets the lower-level network. They'll favor links in the upper-level network over links in the lower-level network when they choose their routes.

To implement this route choice hierarchy, one can label links by different levels according to the free-flow speed and/or lane capacity (for example, 45 miles per hour and 3 lanes each way). An easy way of implementing such a two-level hierarchical route choice model is to weight significantly less on those major links than those minor links, so that the major links can attract more flow as compared to the case where links in both levels are weighted equally. Therefore, in a network with hierarchical roadway facilities (such as freeway, major arterials, minor arterials, collectors and so forth), we construct the following static travel cost for habitual travelers (on path p) when using KSP to generate their pre-trip route sets,

$$c_p^{rs} = \sum_{a \in A} \delta_{ap} \lambda_a c_a \tag{5}$$

where δ_{ap} is the link-route incidence indicator, equal to 1 if link a is on path p (equal to 0 otherwise). c_a is the generalized travel cost on link a. In particular, if only travel time is concerned, then c_a is the free-flow travel time of link a at time t. Let ϵ represent a small positive number. $\lambda_a = \epsilon < 1$ if link a is a major link and $\lambda_a = 1$ if link a is a minor link. This method is referred as "hierarchical KSP generation". The choice of the parameter ϵ is by trial-and-error so as to produce the DNL results that best match the reality.

3.2.2 PUE route set generation

The KSP generation methods discussed above use free-flow travel times to generate the route set for habitual travelers. This may be reasonable if the habitual travelers have no knowledge of actual travel times on those routes. In reality, however, those travelers often experience longer travel times on free-flow travel times due to traffic congestion, and will not determine their routes solely on free-flow travel times. In order to generate such a route set that incorporates historical traffic information, we can first perform a PUE-DTA and use the resultant routes as the pre-trip route set for habitual travelers. We assume the proportions of flow that are assigned to each route under the PUE will be applied to the group of habitual travelers. This method is referred as "PUE route set generation".

4 Queue spillback and the hybrid route choice

In this section, we use an example to show that under the CTM implementation of the LWR kinematic wave model, queue spillback is closely related to the route choice and the diversion ratio is crucial in determining queuing patterns. For simplicity, we assume the travel time is the sole factor in constructing the generalized travel cost.

The LWR model considers the physical space a vehicle takes, so queues in this model take up space and can block traffic from entering a downstream link if space runs out on that link. Moreover, the LWR model describes queue growth in a more realistic way through its shock wave mechanism. Here we adopt the CTM implementation of the LWR model for traffic movement on links (Daganzo 1994). For traffic flow through junctions, we use a general node traffic model given by Nie and Zhang (2010). It should be pointed out that node models play an important role in determining queuing spillback.

Suppose there are M approaches at an intersection and those approaches are marked as $1, 2, 3, ...M$. Flows leaving any exit will diverge to any other approaches and meanwhile, any approach also serves as a merge point of flow from any other approaches. Therefore, merges and diverges of all the approaches occur simultaneously. Let v_{ij} be the number of vehicles moving from approach i to j during a loading time interval. In order to describe traffic movements through intersections, we first define the demand (supply) of a link at time t, $D(t)$ ($S(t)$), as the maximum number of vehicles that are allowed to leave (enter) the link at t. Assume the vehicle proportion at upstream link i heading for downstream link j, a_{ij}, is known according to the vehicles in the last cell of link i under the LWR model, and demand and supply for each cell is computed based on the fundamental diagram. v_{ij} reads (Nie and Zhang 2010),

$$\text{virtual demand } \overline{vd}_i(t) = \min\left(D_i(t), \min_j \left\{ \frac{S_j(t)}{a_{ij}(t)} \right\} \right) \tag{6}$$

$$\text{virtual supply } \overline{vs}_i(t) = \min\left(S_i(t), \sum_j a_{ij}(t) D_j t \right)$$

$$v_{ij}(t) = \min\left(\overline{vd}_i(t) a_{ij}(t), \overline{vs}_j(t) \frac{\overline{vd}_i(t) a_{ij}(t)}{\sum_k \overline{vd}_k(t) a_{kj}(t)} \right)$$

which is a generalization or streamlined version of several node models (Daganzo 1994, 1995; Jin and Zhang 2003, 2004; Lebacque 1996). As a matter of fact, Eq. (6) does not strictly enforce FIFO at diverge nodes (Nie 2006), but since we do not pursue PUE in this paper, the FIFO condition can be relaxed.

Now consider the network of Fig. 1(b), where travelers from link 0 heading for the destination can choose either link 1 or link 3 to go through an intermediate node. Link 1 has a lower free-flow travel time than link 3 and it is preferred by travelers under slight congestion. Link 2 serves as a bottleneck due to a lane drop and a queue may grow and back up to both link 1 and

link 3. Figure 1(a) gives the fundamental diagram of link 1 with the capacity C_m, free-flow speed u_f and jam density k_j.

Due to the shorter free-flow travel time on link 1 than link 3, habitual travelers will take link 1 rather than link 3. There could be a case that all adaptive travelers also use link 1 regardless of the diversion ratio. To see this, suppose a worst case on link 1 where the outflow rates of link 1 achieves minimum when it equals to the capacity of link 2 (i.e. $C_i = C_{m,2}$) and link 4 is not used, so that the travel speed on link 1 achieves the minimum. Therefore, flow on link 1 is at density k_i and the traversal speed is u_i. Link 3 will never be used if its free-flow travel time is larger than the travel time on link 1, i.e.

$$\frac{l_1}{u_i} < \frac{l_3}{u_{f,3}} \tag{7}$$

Therefore, link 1 is over-saturated (i.e. the queue regulates its inflow) and a queue could spillover to upstream links. In this case, the queue may spill back upstream, and this can be unrealistic when link 3 is not that much longer than link 1 and in practice it may be an acceptable choice to some travelers. No diversion ratio can reduce such oversaturation under this setting.

Even though Eq. (7) does not hold (i.e. link 3 offers relatively competitive free-flow travel time), there is still a case that the queue on link 1 spills over and link 3 is not used, if the diversion ratio is not properly set. If the diversion ratio is low, then most of demands will be loaded on the pre-trip route, i.e. link 0 to link 1 to link 2. It is easy to see that the queue spills back to link 0

Fig. 1 (**a**) The fundamental diagram of link 1; (**b**) a sample network
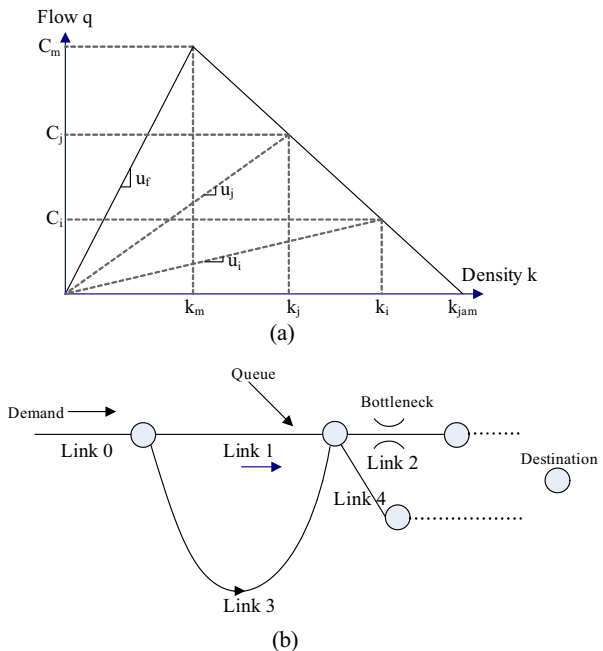
due to the low diversion ratio. On the other hand, if the diversion ratio is high, most travelers are willing to switch to a route with shortest instantaneous travel time. Suppose both links 2 and 4 are used and link 2 is still a bottleneck link. In this case, the outflow rates of link 1 becomes C_j and $C_j > C_i$. Therefore, the instantaneous travel speed on link 1 is now u_j. Link 3 will never be used if its free-flow travel time is larger than the travel time on link 1, i.e.

$$\frac{l_1}{u_j} < \frac{l_3}{u_{f,3}} \qquad (8)$$

Only if the diversion ratio is in a reasonable range, not too low and not too high, then both links 1 and 3 will be used and the queue on link 1 will not spill back. One can also find other examples to illustrate a similar phenomenon. Therefore, an appropriate diversion ratio is crucial in determining queuing patterns, and should be carefully calibrated in practice.

5 Numerical experiments

In this section, we perform several numerical experiments to evaluate the proposed hybrid route choice model with different diversion ratios and with three methods of generating pre-trip route sets for the habitual travelers: ordinary KSP, hierarchical KSP and PUE route set generation. As a benchmark case, the PUE route choice will also be implemented and its results are used to assess the performance of the hybrid route choice model. The numerical experiments are carried out in two networks, one medium-size synthetic network and another large-size real network. We assume the travel time is the sole factor in constructing the generalized travel cost.

5.1 A synthetic network

We synthesize a corridor network which consists of three residential areas and a Central Business District (CBD) connected by a three-lane freeway and two-lane major arterial road, as shown in Fig. 2.

We assume in the morning peak hour (1 h assignment horizon), travelers commute, either from the residential area or the external area (represented
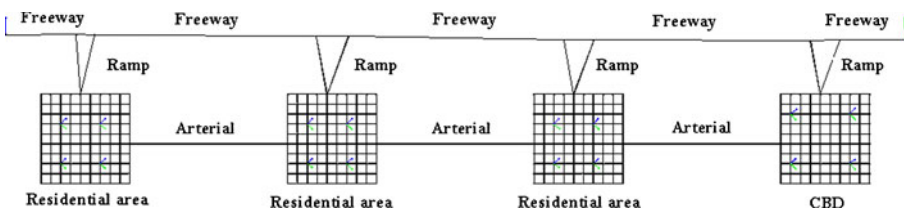

Fig. 2 A synthetic corridor network

Table 1 The total travel time and total delay of selected scenarios

	TTT(h)	TD(h)	Diversion ratio=0.2		Diversion ratio=0.5		Diversion ratio=0.8	
			TTT(h)	TD(h)	TTT(h)	TD(h)	TTT(h)	TD(h)
PUE	812	166						
PUE generation			749	104	842	191	935	261
Ordinary KSP generation			1,719	1,070	961	302	992	312
Hierarchical KSP generation			1,514	843	980	302	1,007	321

TTT Total travel time; *TD* Total delay

by the origin at the beginning of the freeway), to the CBD area. The external origin has a total demand of 400 vehicles to each of the residential areas, 1,200 vehicles to the CBD and 100 vehicles to the external destination (at the end of the freeway). Each residential area also has a total demand of 400 vehicles to other residential areas, 1,280 vehicles to the CBD and 200 vehicles to the external destination. Further, we allocate the total demands to 12 five minute assignment intervals following a trapezoidal flow pattern to mimic the flow profile in the peak hour.

Each residential area and the CBD area are represented by 9×9 grid subnetworks and four origins and four destinations. Each link of the grid subnetworks stands for a minor local street and is set to have an identical capacity of 1,500 vehicles per hour and a speed limit of 25 mile per hour, but its length is randomly chosen from a uniform distribution within $[0.08, 0.12]$ (miles). The grid subnetwork connects the origins and destinations to the freeway ramps and the major arterial road, so that travelers have choices of choosing either the freeway or the arterial to go to the CBD in the morning commute. There are in all 478 nodes, 1,528 links and 129 O-D pairs. Each segment of the freeway and the arterial has the same length, but the freeway has a total capacity of 6,000 vph and 65 mph speed limit, while the arterial road has a total capacity of 3,600 vph and a speed limit of 40 mph. For each segment, the free-flow travel time via the ramps and the freeway and via the arterial road is set to be 3 min and 5 min, respectively, so that the freeway is a preferred route and the arterials may also be used if the freeway is congested.

For all the scenarios studied, the loading time interval is 5 s, and adaptive travelers update their travel information/routes every 90 s. The diversion ratio is set to be 0.2, 0.5 or 0.8 to represent a low, medium or high proportion of adaptive travelers. For those habitual travelers, we let $K = 2$ in the KSP computation. The results in total travel cost and total delay[1] for all scenarios are shown in Table 1.

---

[1]Total delay equals the total vehicle-hour-traveled (VHT), subtracted by the total free-flow travel time of all vehicle trips taking the same routes as where VHT is computed.

5.1.1 The effects of diversion ratio

As seen from Table 1, under the KSP route set generation, a 0.2 diversion ratio leads to significantly larger total travel time and total delay than a 0.5 or 0.8 diversion ratio. On the other hand, a high diversion ratio yields slightly higher total free-flow travel time (which equals $TTT - TD$) than a low diversion ratio, because more travelers may deviate to a longer route with less congestion under a higher diversion ratio. Interestingly, a 0.5 diversion ratio yields less queuing delay than a 0.2 or 0.8 diversion ratio, which may be explained by the following: (1) as explained in Section 4, a medium diversion ratio can prevent queuing spillback under certain conditions; and (2) intuitively, a diversion ratio, if too small, can lead to severe queuing on the freeway due to the high demand of habitual travelers. If the diversion ratio is set to be too high, then compared to a medium diversion ratio, some bottleneck links are used more intensively by adaptive travelers after they update their routes. This explains why in practice, a moderate diversion ratio may produce a flow pattern with less queuing delay than a high or a low one.

We plot in Fig. 3(a) and (b) the time-varying volumes on a freeway link and an arterial link in the middle section against three diversion ratios under the ordinary KSP route set generation. All four scenarios yield approximately the same time-varying volumes in the first 200 loading intervals (around 20 min). This is because when the network is not congested or mildly congested, travelers within the same O-D pair, regardless of their willingness to switch routes, will use the same routes that are preferred in terms of free-flow travel time. However, the volumes of four scenarios start to differ significantly after the network is loaded with high demands. Under the low diversion ratio, a large percentage of travelers will still follow the freeway, which leads to much more severe congestion on the freeway than the case of a high diversion ratio where most travelers will switch routes and the case of PUE. The freeway link cumulates up to 600 vehicles when the diversion ratio is 0.2, as compared to 400 vehicles when the diversion ratio 0.5, and 200 vehicle when it is 0.8 or in PUE. The low diversion ratio also results in a long loading tail after the assignment horizon ends, i.e. the network does not clear up until 25 min after the assignment horizon ends, while the network clears up within 6 min in all other scenarios.

Generally, of all four scenarios, PUE yields the highest share of arterial road usage, while the low diversion ratio has the lowest share. By checking the cumulative number of vehicles on the arterial link, PUE almost triples the number of travelers who use the arterial as compared to the case of 0.2 diversion ratio. Because PUE assumes all travelers can predict the network condition after day-to-day experience, such a user optimum leads to less congestion and higher usage of the arterial than the hybrid route choice model. For the hybrid route choice model, a medium diversion ratio has substantially higher arterial share than a low one, slightly higher arterial share than a high diversion ratio. This is consistent with the result that a medium diversion ratio produce less overall congestion than a high or low one.
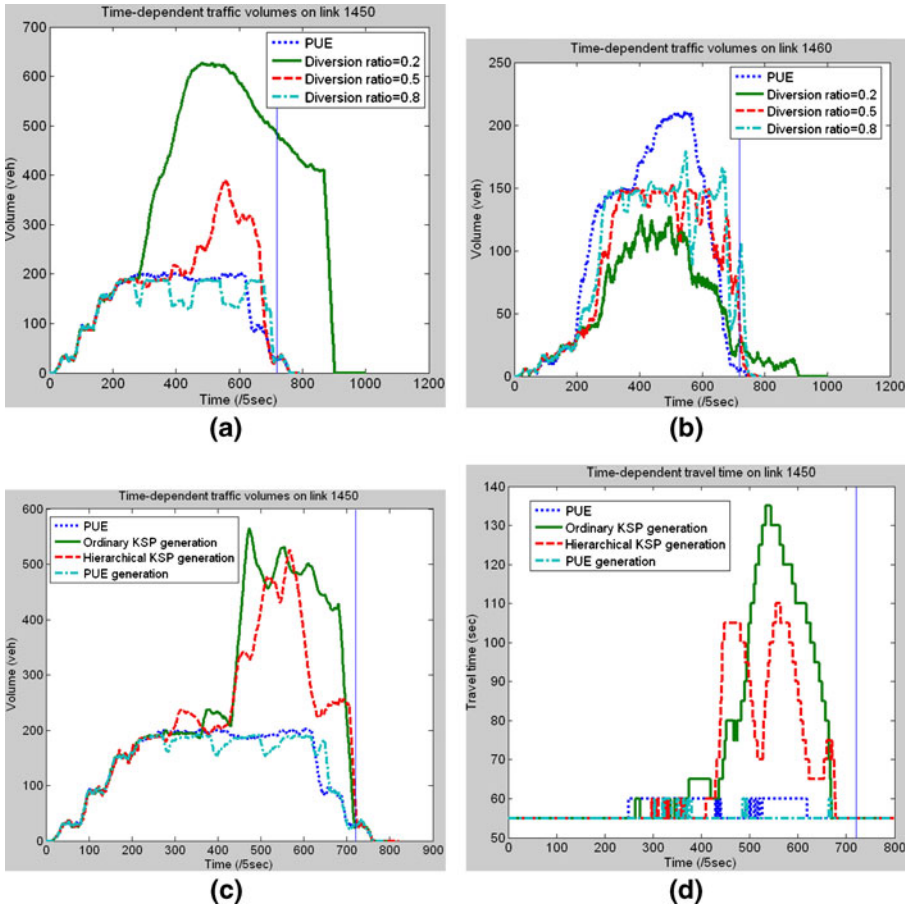
Fig. 3 (**a**) Time-varying volumes on a freeway link w.r.t. different diversion ratio; (**b**) Time-varying volumes on an arterial link w.r.t. different diversion ratios; (**c**) Time-varying volumes on a freeway link w.r.t. different methods of generation pre-trip route sets; (**d**) Time-varying travel times on a freeway link w.r.t. different methods of generation pre-trip route sets. The *vertical line* indicates the time when the assignment horizon ends

As can be seen from Fig. 3(b), compared to the PUE, the hybrid route choice model yields less volumes on the arterial road during the time (from 200th interval to 240th interval) when the congestion on the freeway starts to propagate quickly, and it also yields more volumes during the time (the 610th interval to the end) when the queue on the freeway starts to dissipate. This is because in PUE, travelers can predict the actual traffic congestion on all links and can therefore better respond to the congestion propagation. As for the hybrid route choice model, because adaptive travelers make en-route route choices and their prediction is based on instantaneous travel time, they may underestimate the congestion on the freeway when the queue builds up and overestimate the congestion when the queue starts to dissipate. Thus,

compared to PUE, the route diversion in the hybrid route choice model always lags behind the time they would divert to optimize their travel cost in response to the actual congestion propagation, and obviously such a route choice may be far from user optimal (but still could be realistic, dependent on the real travel behavior).

We also observe a large fluctuation of volumes on both the freeway and the arterial road when the diversion ratio is set to be high. Because travelers will update their current traffic information every 18 loading time intervals, if the diversion ratio is high, then a large number of adaptive travelers will deviate to a certain route (e.g. the arterial road) at the beginning of certain time intervals and result in unexpected congestion of the new route. This congestion may spread quickly and makes the freeway route advantageous again, which leads a large number of adaptive travelers to again switch back to the freeway route. Thus, under a high diversion ratio, a large number of travelers intermittently choose either the freeway or the arterial, and that is why we see a large fluctuation of volumes on both links.

5.1.2 Effects of pre-trip routes

We plot in Fig. 3(c) and (d) the time-varying volumes and travel times on a freeway link in the middle section against the three methods of generating pre-trip route sets for habitual travelers where the diversion ratio is set to be 0.5. The weighting factor on the major links (freeway links and arterial links) is set to be $\lambda_a = 0.01$ in the hierarchical KSP calculation.

Generally, if half of the travelers are aware of the historical information and follow their pre-trip routes, then time-varying flows on this freeway link is close to the results obtained from PUE. However, if those travelers only use the free-flow travel time to determine their pre-trip routes so that the freeway is overwhelmingly used, then the freeway could be far more congested than the case where the historical information is used to determine the pre-trip routes, particularly after the 400th time interval where the congestion becomes severe on the freeway. Compared to the maximum 60 s in the case of PUE or the case where the historical information is used by habitual travelers, using KSP to generate pre-trip route sets yields up to 135 s travel time on the freeway link, and thus the TTT and TD of the network are significantly larger as well.

An interesting result is that using the hierarchical KSP generation indeed significantly reduce the queuing as compared to the case with only the ordinary KSP generation. When such a hierarchical way of generating pre-trip routes is considered, habitual travelers commuting from the residential area to the CBD have choices of either the freeway or the arterial road. However, if no hierarchical road preference is incorporated, then the first two shortest paths for most O-D pairs will always be such that the freeway is chosen in both routes with distinctions on some local streets. Therefore, by using the hierarchical KSP to generate the route sets, more habitual travelers will use the arterial road instead of the freeway than the ordinary KSP generation.

Table 2 The total travel time and total delay of the Sacramento network w.r.t DR

	DR	0.3	0.4	0.45	0.5	0.55	0.6	0.7
Ordinary KSP generation	ATT(min)	G	41.53	38.87	37.74	35.61	G	G
	AD(min)	G	15.65	12.87	11.58	9.36	G	G
	ADS (mile)	G	25.49	25.49	25.46	25.51	G	G
Hierarchical KSP generation	ATT(min)	63.93	47.13	42.77	38.18	37.16	37.58	G
	AD(min)	37.67	20.92	16.49	12.01	10.84	10.96	G
	ADS (mile)	25.72	25.67	25.70	25.60	25.61	25.67	G

DR Diversion ratio; *ATT* Average travel time; *AD* Average travel delay; *ADS* Average travel distance; *G* Gridlock

As seen from Table 1, the PUE route generation with a 0.2 diversion ratio has less TTT and TD than PUE. This indicates that if we only provide real-time information to a small proportion of travelers and meanwhile assume that most of travelers use the historical information to travel, then this may considerably benefit the network. However, if the real-time information is offered to a large number of the travelers and they all use it, then the network performance may be worse off. This is because the optimal route at the current time interval, if used by an excessive number of travelers, may build up an unexpected queue and sometimes this is not desirable. It should be noted that the range of diversion ratio found here is consistent with the market penetration found in previous traveler information studies (e.g., Mahmassani and Jayakrishnan 1991; Mahmassani and Liu 1999), and this consistency can be viewed as an initial validation of the proposed hybrid route choice model.

5.2 The Sacramento metropolitan area network

We also applied the hybrid route choice model to a large network which covers the Sacramento metropolitan area. The network consists of 2,556 nodes, 7,221 links and 729 O-D pairs. We collected 5 min traffic counts for a period of 24 h on 25 segments of major freeways and 32 major arterials on the periphery of downtown Sacramento. The time-varying counts are then used to estimate a morning peak 6-h (6:00am–12:00pm) time-dependent O-D demands by a logit path flow estimator algorithm (Bell et al. 1997). Other parameters are: loading time interval 10 s; $K = 3$ in the KSP calculation; and the traffic information will be updated every 5 min. The weighting scaler on the major links is set to be 0.2 in the hierarchical KSP route generation. Table 2 shows the average travel time, average travel delay and average travel distance using different diversion ratios and pre-trip route generation methods.[2]

As discussed before, the diversion ratio, if set too high or too low, may cause serious queuing. In this case, the DTA procedure terminates without

---

[2]We do not include the results from PUE because a gridlock occurs during the iterative procedure of DNL, and solving the issue of gridlock in the DTA algorithm is beyond the scope of the paper.

gridlock only when the diversion ratio is within a certain range, i.e., from $0.35 \sim 0.55$ if the ordinary KSP generation is used, or from $0.25 \sim 0.65$ if the hierarchical KSP generation is used. The wider range of the acceptable diversion ration provided by the hierarchical KSP generation indicates that it can better prevent excessive concentration of queuing than the ordinary KSP generation. Overall a 0.55 diversion ratio looks reasonable in the sense that the resultant ATT and AD and time-varying link flow on designated links (25 segments of freeways and 32 major arterials) approximately match the actual observation. When the diversion ratio is less than 0.55, less travelers are willing to switch routes and travelers are subject to more queuing delay on average. In contrast, when the diversion ratio is larger than 0.55, more travelers respond to real-time information and switch routes, but this may cause excessive queuing on certain routes or links due to herding, which may even produce a gridlock.

If hierarchical KSP is used to generate the route sets, then for most O-D pairs, the second or the third shortest path is more likely to include the arterial links as an alternative to the preferred freeway route, as compared to the ordinary KSP generation. It assigns more travelers on the major arterials, which tends to distribute the flow more evenly for those major links (i.e. those weighed less in computing shortest paths, freeway or major arterials) and prevent excessive queuing that leads to gridlock. Given the same diversion ratio, using hierarchical KSP always yields longer average distance and larger average delay and cost than the ordinary KSP. This is because hierarchical KSP induces travelers to travel more on major links rather than the minor links (e.g. local streets), and they generally take slightly longer distances. In addition, under hierarchical KSP, it is more likely that flows are concentrated on those major links with fewer minor links assigned with flows, which results in a larger queuing delay, as compared to the ordinary KSP.

6 Conclusions

Travelers' route choice behavior in real life follows neither Boston UE or Predictive UE. Rather, in choosing their routes travelers may have both historical experience and real-time information. This paper proposes a hybrid route choice model where travelers are divided into two groups, habitual travelers and adaptive travelers. Habitual travelers strictly follow their pre-trip routes. We speculate that route choice of habitual travelers may be such that the major roads, such as the freeway or major arterials, are favored over minor road links and they also consider historical traffic information based on their day-to-day driving experiences. Therefore, we propose two new methods of generating their pre-trip route sets, hierarchical K shortest path and Predictive-UE generation. On the other hand, adaptive travelers are responsive to real-time information and willing to change routes. The hybrid model may be more realistic in describing travelers' route choices in regards to travel costs. Unlike

the PUE, the new hybrid model requires only one shot dynamic network loading.

We study how the choice of diversion ratio and generation methods for habitual travelers' route set affect the resulting flow and queuing patterns and found that,

– The hybrid model is easy to calibrate and can work efficiently for large-scale networks. It is likely to produce realistic results as indicated by the large-scale numerical example.
– A medium level of diversion ratio can reduce excessive concentration of queuing compared to a high or low level of diversion ratio. If it is set too high or too low, then queues tend to concentrate on certain links, which can eventually lead to gridlock.
– In most cases, PUE yields significantly less congestion and queue spillovers than the hybrid route choice model, because travelers in the former can anticipate what would happen for their entire trip and hence can make better choices, while in the latter the adaptive travelers make myopic choice decisions based on the prevailing traffic conditions that may prove to be a bad choice as traffic conditions change.
– Pre-trip routes generated by PUE (i.e. using historical information) for habitual travelers tend to yield considerably less delay and queues in the network than by ordinary KSP which only uses free-flow travel time information. Pre-trip routes generated by hierarchical KSP may reduce queuing compared to those generated by ordinary KSP, thanks to the former's ability to spread demand among freeways and major arterials.
– When habitual travelers use historical information to determine their pre-trip routes, offering real-time information to a small portion of travelers can actually achieve better network performance than PUE where all travelers are assumed to have perfect information all the time. However, if the majority of travelers are offered with real-time information, network performance may actually become much worse than offering no one real-time information.

It is hoped that these findings can help practitioners choose a route choice model and calibrate its parameters against real data. While more experiments need to be carried out to confirm these initial findings, it does seem that the use of historical information, hierarchial network, and moderate level of diversion can help spread congestion and avoid excessive concentration of queuing and gridlock. Our future work will incorporate learning in the route choice decision process and study the resulting day-to-day dynamics and its asymptotic patterns.

References

Bell MGH, Shield CM, Busch F, Kruse G (1997) A stochastic user equilibrium path flow estimator. Transp Res, Part C Emerg Technol 5(3–4):197–210

Bovy P, Stern E (1990) Wayfinding in transport networks. Kluwer Academic Publishers

Bovy PH, Fiorenzo-Catalano S (2007) Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica 3(3):173–189

Carey M, Ge YE (2011) Comparison of methods for path flow reassignment for dynamic user equilibrium. Netw Spat Econ. doi:10.1007/s11067-011-9159-6

Cascetta E, Nuzzolo A, Russo F, Vitetta A (1996) A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Proceedings from the thirteenth international symposium on transportation and traffic theory

Daganzo C (1998) Queue spillovers in transportation networks with a route choice. Transp Sci 32(1):3–11

Daganzo CF (1994) The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transp Res, Part B 28:269–287

Daganzo CF (1995) The cell transmission model, part II: network traffic. Transp Res, Part B 29:79–93

Dia H, Harney D, Boyle A (2001) Dynamics of drivers' route choice decisions under advanced traveler information systems. Roads and Trasp Res 10:2–12

Friesz TL, Bernstein D, Smith TE, Tobin RL, Wei BW (1993) A variational inequality formulation of the dynamic network equilibrium problem. Oper Res 41:179–191

Gao S, Frejinger E, Ben-Akiva M (2011) Cognitive cost in route choice with real-time information: an exploratory analysis. Transp Res, Part A Policy Pract 45(9):916–926

Ghali M (1995) A note on the minimum instantaneous cost path of the dynamic traffic assignment problem. Ann Oper Res 60:115–120

Hawas YE (2004) Development and calibration of route choice utility models: neuro-fuzzy approach. J Transp Eng 130(2):171–182

Jin W, Zhang HM (2003) On the distribution schemes for determining flows through a merge. Transp Res, Part B 37:521–540

Jin W, Zhang HM (2004) Multicommodity kinematic wave simulation model for network traffic flow. Transp Res Rec 1883:59–67

Kant P (2008) Route choice modelling in dynamic traffic assignment. Master thesis, University of Twente. http://essay.utwente.nl/58303/

Kuwahara M, Akamatsu T (2001) Dynamic user optimal assignment with physical queues for many-to-many od pattern. Transp Res, Part B 35:461–479

Lebacque J (1996) The godunov scheme and what it means for first order traffic flow models. In: Procedings of international symposium of transport and traffic theory, pp 79–102

Lee C, Ran B, Yang F, Loh WY (2010) A hybrid tree approach to modeling alternate route choice behavior with online information. J of Intell Transp Sys: Technology, Planning, and Operations 14(4):209–219

Lighthill MJ, Whitham GB (1955) On kinematic waves. ii. a theory of traffic flow on long crowded roads. Proc Royal Soc 229:317–345

Mahmassani HS, Jayakrishnan R (1991) System performance and user response under real-time information in a congested traffic corridor. Transp Res, Part A 25(5):293–307

Mahmassani HS, Liu YH (1999) Dynamics of commuting decision behaviour under advanced traveller information systems. Transp Res, Part C 7(2–3):91–107

Merchant D, Nemhauser G (1978) A model and an algorithm for the dynamic traffic assignment problem. Transp Sci 12:183–199

Nie X, Zhang HM (2005) A comparative study of some macroscopic link models used in dynamic traffic assignment. Netw Spat Econ 5:89–115

Nie Y (2006) A variational inequality approach for inferring dynamic origin-destination travel demands. PhD thesis, University of California at Davis

Nie Y, Zhang HM (2010) Solving the dynamic user optimal assignment problem considering queue spillback. Netw Spat Econ 10(1):49–71

Nie YM, Wu X, de Mello TH (2011) Optimal path problems with second-order stochastic dominance constraints. Netw Spat Econ. doi:10.1007/s11067-011-9167-6

Peeta S, Mahmassani HS (1995) Multiple user classes real-time traffic assignment for online operations: a rolling horizon solution framework. Transp Res, Part C Emerg Technol 3(2):83–98

Peeta S, Yu JW (2005) A hybrid model for driver route choice incorporating en-route attributes and real-time information effects. Netw Spat Econ 5:21–40

Pel AJ, Bliemer MC, Hoogendoorn SP (2009) Hybrid routing choice modeling in dynamic traffic assignment. Transp Res Rec 2091:100–107

Ramming S (2002) Network knowledge and route choice. PhD thesis, Massachusetts Institute of Technology

Ran B, Boyce DE, Leblanc LJ (1993) A new class of instantaneous dynamic user-optimal traffic assignment models. Oper Res 41(1):192–202

Reddy PDVG, Yang H, Vaughn KM, Abdel-Aty MA, Kitamura R, Jovanis PP (1995) Design of an artificial simulator for analyzing route choice behavior in the presence of information system. Math Comput Model 22(4–7):119–147

Richards PI (1956) Shock waves on the highway. Proc R Soc 4:42–51

Smith M (1993) A new dynamic traffic model and the existence and calculation of dynamic user equilibria on congested capacity-constrained road networks. Transp Res, Part B 26:49–36

Ukkusuri S, Patil G (2007) Exploring user behavior in online network equilibrium problems. Transp Res Rec: Journal of the Transportation Research Board 2029:31–38

Vovsha P, Bekhor S (1998) The link-nested logit model of route choice: overcoming the route overlapping problem. Transp Res Rec 1645:133–142

Xu H, Zhou J, Xu W (2011) A decision-making rule for modeling travelers route choice behavior based on cumulative prospect theory. Transp Res, Part C Emerg Technol 19(2):218–228