Dynamic activity-travel assignment in multi-state supernetworks under transport and location capacity constraints

Peng Liu, Feixiong Liao, Hai-Jun Huang & Harry Timmermans


To link to this article: http://dx.doi.org/10.1080/23249935.2016.1189739
Dynamic activity-travel assignment in multi-state supernetworks under transport and location capacity constraints

Peng Liu\textsuperscript{a}, Feixiong Liao\textsuperscript{b}, Hai-Jun Huang\textsuperscript{a} and Harry Timmermans\textsuperscript{b}

\textsuperscript{a}School of Economics and Management, Beihang University, Beijing, People’s Republic of China; \textsuperscript{b}Urban Planning Group, Eindhoven University of Technology, Eindhoven, The Netherlands

ABSTRACT

A unified framework coupling activity-based modelling and dynamic traffic assignment has recently been proposed. It formulates dynamic activity-travel assignment (DATA) in multi-state supernetworks as a dynamic user equilibrium. Choices of departure time, route, mode, activity sequence, and activity/parking location are determined endogenously, reflected in time-dependent activity-travel patterns (ATPs). However, capacity constraints associated with activities, parking, and public transit vehicles have not taken into account. This paper extends this approach by formulating these constraints and incorporating them into the DATA framework. Three numerical examples are presented to illustrate the effectiveness of the approach. It is shown that capacity constraints have significant effects on the choice facets.

1. Introduction

Travel demand forecasting has long been a central issue of transportation planning. Over the last decades, the activity-based modelling paradigm of travel demand analysis has gradually replaced the traditional trip-based paradigm (e.g. four-step model) (Henson, Goulias, and Golledge 2009; Shiftan and Ben-Akiva 2011; Pinjari and Bhat 2011; Rasouli and Timmermans 2014). The potential benefits of activity-based modelling have sparked interests in developing operational travel demand forecasting systems (e.g. ALBATROSS (Arentze and Timmermans 2004a), MATSim (Balmer, Axhausen, and Nagel 2006), TASHA (Miller 2009), ADAPTS (Auld and Mohammadian 2009), and SimAgent (Bhat et al. 2012). Activity scheduling and implementation are considered two essential components of such systems. The former primarily focuses on the generation of activity-travel patterns (ATPs), predicting where, when, with whom, with which transport mode and taking which route to conduct activity programmes of a given time frame. Various mechanisms have been applied and developed in the travel behaviour research community (e.g. constraint based, rule based, and utility maximization based). The derived ATPs are often arranged into time-dependent O–D trip matrices fed into the activity implementation stage, where static/dynamic traffic...
assignment models are in play to capture the interaction of travel demand and supply of the transport systems. The outputs are time-dependent travel times between those O–D pairs. Moreover, to accommodate the interdependencies of the complete ATPs, a feedback from activity implementation to activity scheduling has been adopted in a few recent systems (e.g. Auld and Mohammadian 2009; Pendyala et al. 2012).

On the other hand, the shift in the unit of analysis from trips to activities has also been undertaken in the traffic theory community, albeit at a smaller scale. Traditional single route choice models (Wardrop 1952; Merchant and Nemhauser 1978) have been extended via combined route and mode choice (Sheffi 1985) to multi-dimensional choice models pertaining to full activity programmes (Lam and Yin 2001; Lam and Huang 2002, 2003; Ouyang et al. 2011). Along this transition, as far as network-based approaches are concerned, the network supply has been accordingly upgraded from simple one-dimensional traffic networks via multi-modal transport networks (Lozano and Storchi 2002; Carlier et al. 2003) to activity-based supernetworks (Ramadurai and Ukkusuri 2010; Fu and Lam 2014). Such network extensions usually share the same two-tuple (node vs. edge) topology, but are capable of carrying rich information on various choice facets from the demand side and land-use transport system from the supply side. The multi-state supernetwork representation, which was originated to model multi-modal multi-activity scheduling at the individual level (Arentze and Timmermans 2004b; Liao, Arentze, and Timmermans 2010, 2011, 2012, 2013a), has been advanced in terms of the high level of detail captured in the networks.

Recently, Liu et al. (2015) proposed a unified travel demand forecasting framework by formulating dynamic activity-travel assignment (DATA) in multi-state supernetworks as a dynamic user equilibrium (DUE) problem. Choices of departure time, route, mode, activity sequence, activity and parking location, and activity duration are determined endogenously in time-dependent ATPs through the supernetworks. Hence, this framework couples activity scheduling and implementation in the strongest sense of combining both lines of research mentioned above.

However, capacity constraints have not taken into account in the DATA framework, except road capacity for private vehicles (mainly car). Although capacity constraints are ubiquitous at all sorts of facilities, most travel demand forecasting systems have only focused on the capacity of the traffic networks. To increase travel behaviour realism, a few studies also considered capacity constraints at parking locations at the end of the car trips (Lam et al. 2006) and modelled crowdedness in public transit vehicles (Tian, Huang, and Yang 2007), while to the best of our knowledge very few took capacity constraints at activity locations (e.g. supermarkets and restaurants) into consideration. In terms of the choice dimensions captured, these studies more or less fall short of representing the full ATPs. Given the fact that the location/vehicle capacities have significant effects on the choice of activity locations and other choice facets, the absence of capacity constraints in the DATA framework may lead to biased predictions of travel demand.

The aim of this paper, therefore, is to suggest an extended Bureau of Public Road (BPR) function for modelling the effects of capacity constraints of roads for private vehicles, public transit vehicles, and locations for activity and parking. These capacity constraints will be consistently incorporated in the DATA framework, which will not only allow assigning flows to roads and public transit vehicles, but also to activity and parking locations. It is opposed to other assignment models in travel demand analyses that have only focused on the travel component, namely traffic flows. Terminologically, the assignment is about the
determining the flows under the interactions of supply–demand of activity-travel. Previous assignment models do not possess capacity constraints on the supply side for parking and activity locations. In that sense, there is no assignment on the flow to activity and parking locations. Therefore, the DATA framework under capacity constraints involves a more valid analysis of interactions between travel demand and level of service.

To that effect, the remainder of this paper is organized as follows. Section 2 first introduces the specifications of multi-state supernetwork representation for conducting an activity programme, and then presents the basic assumptions of the approach. Section 3 discusses the mechanisms to calculate link flows, link time expenses, and link disutilities after the incorporation of capacity constraints in a dynamic environment. Subsequently, flow propagation and path (ATP) disutilities are formulated. Section 4 formulates the DATA under capacity constraints based on the user equilibrium model and gives a solution method. Section 5 presents three numerical examples to illustrate the approach. Finally, concluding remarks and a discussion of future research are provided.

2. Representation and assumptions

2.1. Supernetwork representation

An activity-based multi-state supernetwork \( SNK(N, L) \), which is composed of a finite set of nodes, \( N \), and a finite set of directed links, \( L \), is constructed in this study. Let \( l \) denote a link in \( SNK \), belonging to one of four basic link types, that is, \( l \in L_{PVN} \cup L_{PTN} \cup L_T \cup L_A \). \( L_{PVN} \) is the set of physical links in private vehicle networks (PVN), which can only be accessed by private vehicles; \( L_{PTN} \) is the set of physical links in public transit networks (PTN), where travellers can walk and take public transit vehicles; \( L_T \) is the set of transfer links between traffic modes (e.g. parking and picking-up); and \( L_A \) is the set of activity links (e.g. working and shopping). Let \( a, A \) denote an activity and a set of activities, respectively, that is, \( A = \{a: a = 1, \ldots, A\} \). The set of home locations is represented by \( H \), \( H = \{h: h = 1, \ldots, H\}, (H \subset N) \). Let \( p \) denote an ATP, which is simply an ordered set of links, that is, \( \{l_1, \ldots, l_n\} \). Let \( P_h \) denote the set of all feasible ATPs based on home \( h \).

As shown in Figure 1, \( h \) and \( h' \) denote an O–D pair, which refer to the same location, that is, home, but lie at the beginning and end state of conducting an activity programme. The ATP formed by the bold links shows that the traveller leaves home \( h \) by car to conduct an activity at \( a_1 \) with parking at \( r_2 \), then returns home and switches to bike to conduct another activity at \( a_2 \) with parking at \( r_4 \), and finally returns home. Therefore, the links interconnecting PVN and PTN represent parking/picking-up private vehicles, and those interconnecting PVN and PTN represent conducting activities (we refer the readers to Liao, Arentze, and Timmermans 2013a for detailed definitions of link types). Logically, any path from one point to another defines an ATP, expressing a specific combined choice of traffic mode/route, parking/activity location, and activity sequence.

2.2. Model assumptions

In order to facilitate the presentation of the essential ideas, without loss of generality, the following basic assumptions are made in this study.
Assumption 2.1: Travellers are considered as heterogeneous in this study, represented in terms of $I = \{i : i = 1, \ldots, I\}$ traveller classes (Yang and Huang 2004; Han and Yang 2008). Heterogeneous travellers differ in terms of different value of times, different traffic mode or activity choice preference, etc. Furthermore, travellers are assumed to be rational individuals. It means that they have perfect information about SNK and always choose their optimal path from all feasible ATPs to minimize the activity-travel disutility (Lam and Huang 2003; Ouyang et al. 2011).

Assumption 2.2: When link flows are smaller than a specific scale of link capacity, travellers obtain the ordinary (dis)utility; otherwise, the service or resource obtained unit flow (i.e. a traveller) will be reduced, and hence travellers obtain less (more) (dis)utility. More specifically, too many private vehicles on road network lead to congestion and increases travel time (Arnott, De Palma, and Lindsey 1990; Huang and Lam 2002); too many passengers in public transit vehicles lead to crowding and discomfort (Lo, Yip, and Wan 2003; Tian, Huang, and Yang 2007; Fu and Lam 2014); and too many users at activity/parking locations lead to crowding and decline of quality of services. It is noted that this study does not consider spillback effects in private vehicle roads, activity locations and public transit vehicles.

Assumption 2.3: The time period $[0, D]$ of interest is discretized to a finite set of time intervals, $K = \{k : k = 1, \ldots, K\}$. Let $\omega$ be the interval length such that $\omega \cdot K = D$. It is assumed that the inflow to a specific ATP occurs during a time interval, and the model will finally be implemented on computers on the basis of time slices (Lam and Yin 2001; Zhang et al. 2005). Furthermore, the study time period $[0, D]$ is chosen to be large enough such that it can cover all daily activity-travels in SNK.

Assumption 2.4: The travel times of links belonging to PVN or mode transfer are determined by link flows (Lam and Yin 2001; Lam et al. 2006), and those belonging to PTN and activity durations are fixed (Tian, Huang, and Yang 2007; Li et al. 2010). The departure frequency of vehicles
in the PTN is given (Fu and Lam 2014). In addition, let $M_l = \{m_l\}$ be the time table of link $l$ for serving the arrival transit vehicles. Particularly, if the arrival time at a link in PTN is unequal to the entering times of transit vehicles, a waiting time occurs. Other links (e.g. travel links of PVN) are assumed to have no restrictive time table, and travellers can enter these links at any time without waiting. Thus, we suppose the set $K$ defines the time table for these links, that is, $M_l = \{m_l\} = K$.

**Assumption 2.5:** In this paper, the desired entry time window for class $i$ at link $l$ is formulated as $[o_{i,l}, e_{i,l}]$. When a traveller begins to conduct an activity, deviating from the window of desired start times will incur a schedule delay penalty (Small 1982; Arnott, De Palma, and Lindsey 1990; Yang and Meng 1998). It guarantees a widely accepted assumption, namely certain periods of the day are more attractive, resulting in higher utilities for particular activities (Ettema and Timmermans 2003; Ettema et al. 2007). Other links (e.g. travel links of PVN) are assumed to have no restrictive window of desired entry time, and travellers can traverse these links at any time during period $[1, K]$ without penalty. Thus, we suppose $[1, K]$ defines the window of desired entry time for these links.

**Assumption 2.6:** In this study, activity interdependency of household members is not considered (Fu and Lam 2014). We further assume that the disutilities incurred by traversing links are independent and additive. Consequently, time-dependent ATP disutility is calculated by adding up the actual disutilities on those links along that ATP (Chen and Hsueh 1998; Liu et al. 2015).

### 3. Link disutility and flow propagation

#### 3.1. Link disutility

During the traverse of a link in SNK, the times of arrival, entry, and departure can be viewed as three different states, arranged in order of time of occurrence. According to Assumption 2.4, the gap between arrival and entry times is defined as waiting time. This is because traveller needs to wait for entering transit vehicle after arrival. According to Assumption 2.5, the gap between actual entry time and desired entry time window $[o_{i,l}, e_{i,l}]$ is defined as schedule delay time. The gap between entry and departure times is described as duration. If the link is for activity, duration represents activity duration; if it is for parking, duration is for searching time; and if it is for travel, it stands for travel time on the link. It should be noted that for activity link, the duration disutility describes the gap between the ideal utility and the utility with the actual duration; while for non-activity link, it describes the gap between the ideal disutility (set as zero) and the disutility with the actual duration. The purpose of this manoeuvre is to model activity and non-activity links in a unified fashion as conducting activity normally generates utility while travel produces disutility.

In summary, link disutility incurred by class $i$ traveller arriving at link $l$ during interval $k$, that is, $\text{dis} U_i^l(k)$ can be expressed by the following generalized function:

$$\text{dis} U_i^l(k) = \text{dis} U_i^{lw} (k) + \text{dis} U_i^{sd} (k) + \text{dis} U_i^{id} (k),$$  

where $\text{dis} U_i^{lw} (k)$, $\text{dis} U_i^{sd} (k)$ and $\text{dis} U_i^{id} (k)$ denote the disutilities of waiting, schedule delay, and duration, incurred by class $i$ traveller arriving at link $l$ during time interval $k$, respectively. With reference to classical bottleneck models (e.g. Arnott, De Palma, and Lindsey 1990; Li, Lam, and Wong 2014), these disutilities are all defined as opportunity costs in economics deducting the utility of in-home.
Without of loss generality, we first define a linear function to calculate the waiting disutility as follows:

\[
\text{dis } U_{i}^{\text{wr}}(k) = \delta_{i}^{l} \cdot w_{i}(k),
\]

where \(\delta_{i}^{l}\) is the disutility of unit waiting time on link \(l\) for class \(i\), \(w_{i}(k)\) is length of waiting time on \(l\) of class \(i\) arriving at \(l\) during \(k\). Obviously, \(\text{dis } U_{i}^{\text{wr}}(k)\) is continuous and increasing with length of waiting time.

According to Assumption 2.4, for public transit link \(l\), there exists a time interval \(m_{l}\) satisfying \(m_{l} \geq k\), which means traveller should be there waiting for the coming transit vehicle. For non-public transit link \(l\), there always exists a time interval \(m_{l}\) satisfying \(m_{l} = k\) as a result of \(M_{l} = K\). Consequently, the waiting time at these non-activity links is always equal to zero. Without of loss generality, waiting time can be expressed as:

\[
w_{i}(k) = \min\{m_{l} : m_{l} \geq k, \forall m_{l} \in M_{l}\} - k.
\]

According to Assumption 2.5, schedule delay disutility captures the penalty of early or late schedule for entering a link. If entry time \(k\) is earlier than \(o_{i}^{l}\), it must be associated with a schedule early time, that is, \(o_{i}^{l} - k\); if entry time \(k\) is later than \(e_{i}^{l}\), it must be associated with a schedule late time, that is, \(k - e_{i}^{l}\); otherwise, no schedule delay involved. It is noted that \(e_{i}^{l} = o_{i}^{l}\) the time window \([o_{i}^{l}, e_{i}^{l}]\) degenerates to a desired entry time. The schedule delay disutility can be expressed as the following function:

\[
\text{dis } U_{i}^{\text{ed}}(k) = \beta_{i}^{l} \cdot \max\{o_{i}^{l} - k, 0\} + \gamma_{i}^{l} \cdot \max\{k - e_{i}^{l}, 0\},
\]

where parameter \(\beta_{i}^{l}\) and \(\gamma_{i}^{l}\) are the unit time penalty of arriving early and late, respectively, on link \(l\) for class \(i\). Obviously, schedule delay disutility \(\text{dis } U_{i}^{\text{ed}}(k)\) is continuous with entry time \(k\). Particularly, entry time \(k\) at non-activity links always satisfies \(k \in [o_{i}^{l}, e_{i}^{l}] = [1, K]\), and thus its schedule delay disutility is always equal to zero.

As mentioned above, for activity link, \(\text{dis } U_{i}^{\text{ad}}(k)\) is defined as the gap between ideal utility \(y_{i}^{l}\) and the utility of the actual duration for \(i\) arriving at \(l\) during \(k\), that is, \(y_{i}^{l}(k)\). For non-activity link, \(\text{dis } U_{i}^{\text{ad}}(k)\) is equal to \(y_{i}^{l}(k)\) directly. Generally, \(\text{dis } U_{i}^{\text{ad}}(k)\) can be calculated by

\[
\text{dis } U_{i}^{\text{ad}}(k) = |y_{i}^{l} - y_{i}^{l}(k)|,
\]

where

\[
y_{i}^{l} = \left\{\begin{array}{ll}
0, & \text{if } l \notin L_{A}, \\
\max\{y_{b}^{l}(k), \forall b \in L_{a(l)}, k \in K\}, & \text{if } l \in L_{A}.
\end{array}\right.
\]

As shown in Equation (6), \(L_{a(l)}\) is the set of links conducting the same activity as that of link \(l\). Obviously, for activity link, \(y_{i}^{l}\) is maximum utility that class \(i\) can gain from conducting the same activity with \(l\), and thus the actual utility is always less than or equal to the ideal utility. For non-activity link, the ideal disutility satisfies \(y_{i}^{l} = 0\), and thus the actual disutility is always more than the ideal disutility.

Without of loss generality, we define a linear function to formulate the actual (dis)utility \(y_{i}^{l}(k)\) as follows:

\[
y_{i}^{l}(k) = \alpha_{i}^{l}(k) \cdot t_{i}^{l}(k),
\]

where \(t_{i}^{l}(k)\) and \(\alpha_{i}^{l}(k)\) are actual duration and the (dis)utility of unit time on \(l\), respectively, for class \(i\) arriving at \(l\) during \(k\). Obviously, \(y_{i}^{l}(k)\) is continuous and increasing with \(\alpha_{i}^{l}(k)\) and \(t_{i}^{l}(k)\).
As known, BPR function was originally proposed to capture the relationship between traffic flows and travel times in the road network from the viewpoint of statistic physics. However, it should be noted as well that the disutility of conducting activities is context dependent and not readily captured only by the queuing mechanism. Despite the shortcomings, BPR represents the reality to a large extent. Thus, it is suitable for capturing search time for parking space and crowding because of congestion in the public transit vehicles. Studies on these issues can be found in separate papers on parking (e.g. Lam et al. 2006), public transit (e.g. Lo, Yip, and Wan 2003; Fu and Lam 2014) and general traffic networks (e.g. Liu et al. 2015). According to Assumptions 2.2 and 2.4, the length of actual duration \( t_i^l(k) \) can be calculated by:

\[
t_i^l(k) = \begin{cases} 
  t_i^l \cdot \left[ 1 + \eta_i \cdot \left( \frac{\max(u_i(k) - \lambda_i c_i, 0)}{c_i} \right)^{\theta_i} \right], & \text{if } l \in L_{PVN} \cup L_T, \\
  t_i^l, & \text{if } l \in L_{PTN} \cup L_A,
\end{cases}
\]  

where \( t_i^l \) is the free flow travel time or desired duration on link \( l \) for class \( i \), \( c_i \) is capacity, \( u_i(k) \) is inflow on \( l \) during \( k \), and \( \eta_i, \theta_i \) are the corresponding parameters resembling BPR function. Parameter \( \lambda_i \) is a scaling coefficient of the capacity, satisfying \( \lambda_i \in [0, 1] \). More specifically, for private vehicle or mode transfer link \( l \), if link inflow \( u_i(k) \) is more than \( \lambda_i c_i \), \( t_i^l(k) \) will be increasing with \( u_i(k) \) due to congestion; otherwise, \( t_i^l(k) \) is equal to free flow travel time \( t_i^l \). If \( \lambda_i = 0 \), Equation (8) is equivalent to the common BPR function adopted by Lam and Yin (2001). It should be noted that Equation (8) cannot guarantee the principle of First-In-First-Out on private vehicle links (suggested by Carey 1992; Carey, Ge, and McCartney 2003; Szeto and Lo 2005 and so on).

According to Assumptions 2.2 and 2.4, the (dis)utility of unit actual duration, that is, \( \alpha_i^l(k) \) can be calculated by:

\[
\alpha_i^l(k) = \begin{cases} 
  \alpha_i^l, & \text{if } l \in L_{PVN} \cup L_T, \\
  \alpha_i^l \cdot \left[ 1 + \eta_i \cdot \left( \frac{\max(u_i(k) - \lambda_i c_i, 0)}{c_i} \right)^{\theta_i} \right], & \text{if } l \in L_{PTN}, \\
  \alpha_i^l \cdot \left[ 1 - \eta_i \cdot \left( \frac{\max(q_i(k) - \lambda_i c_i, 0)}{c_i} \right)^{\theta_i} \right], & \text{if } l \in L_A,
\end{cases}
\]

where \( \alpha_i^l \) is the free flow (dis)utility of unit travel time on link \( l \), \( c_i \) is the capacity of public transit vehicle or capability of activity service, \( q_i(k) \) is link occupancy (i.e. the number of travellers on \( l \) during \( k \)), and \( \eta_i, \theta_i \) are the corresponding parameters ‘resembling’ BPR function. If flows in a transit vehicle (or at an activity location) are more than \( \lambda_i c_i \), unit time (dis)utility \( \alpha_i^l(k) \) will be decreasing (increasing) due to crowding (or decline of quality of services); otherwise, \( \alpha_i^l(k) \) is equal to \( \alpha_i^l \). When \( \lambda_i = 0 \), Equation (9) degenerates into the function adopted by Nielsen (2000), Lo, Yip, and Wan (2003) as well as Fu and Lam (2014).

As shown, Equations (8) and (9) are highly general functions; therefore, the parameters with different values may represent different attributes of links. For example, when \( \eta_i \) is close to 0, it implies that delay time (or disutility) is not sensitive to congestion (or crowding); when \( \theta_i \) is equal to 1, it implies that delay time (or disutility of unit time) is linear. It should be noted that Equations (8) and (9) cannot enable upper bounds on the capacities. However,
BPR function is the starting point universally adopted to model the effects of congestion mentioned above. It can also be extended to represent point-queue system (e.g. Liu et al. 2015). Similar to traffic flow propagation on roads, BPR function can be also applied for that regard as long as they can be represented as links of the ATPs, which is the fact in the supernetwork.

Substituting Equations (6)–(9) into Equation (5), we finally have a unified function to model duration-based disutilities for different type links as follows:

$$\text{dis} U_{id}^j(k) = |y_i^j - \alpha_i^j \cdot t_i^j| + \alpha_i^j \cdot t_i^j \cdot \eta_i \cdot \left( \max\{x_i(k) - c_i, 0\} \right)^{\theta_i},$$  

(10)

where

$$x_i(k) = \begin{cases} u_i(k), & \text{if } l \notin L_A, \\ q_i(k), & \text{if } l \in L_A. \end{cases}$$  

(11)

As shown in Equation (10), $|y_i^j - \alpha_i^j \cdot t_i^j|$ can be considered as free flow travel disutility, which is independent on link flows, and $\alpha_i^j \cdot t_i^j \cdot \eta_i \cdot \left( \max\{x_i(k) - c_i, 0\} / c_i \right)^{\theta_i}$ can be considered as overload disutility (e.g. congestion, crowding and decline of quality of service), which is dependent on link flows.

### 3.2. Flow propagation

In order to formulate how path inflows spread through the dynamic supernetwork, we formulate the essential flow constraints in this subsection. Corresponding to a traveller’s arrival, entry and departure of a link, there are three basic variables, that is, arrival-flow, inflow and out-flow to describe the states of link flow. To start with, let $r_{hip}^{hi}(j)$ represent the arrival-flow of class $i$ traveller on link $l$ during interval $j$ which departs from home $h$ entering ATP $p$ during interval $k$. It can be expressed as

$$r_{hip}^{hi}(j) = \delta_{hip}^{hi}(j) \cdot f_{hi}^p(k),$$  

(12)

where $f_{hi}^p(k)$ is the inflow of $i$ on $p$ from $h$ during $k$; indicator variables $\delta_{hip}^{hi}(j)$ is equal to 1, if existing $i$ entering $p$ from $h$ during $k$ and arriving at $l$ during $j$; otherwise, 0. For any $l_b \in p$, this is detailed as:

$$\delta_{lb}^{hi}(j) = \begin{cases} 1, & \text{if } \text{Int}(k + t_{iw}^1 + t_{id}^1 + t_{iw}^2 + t_{id}^2 + \cdots + t_{iw}^{l_b-1} + t_{id}^{l_b-1} + 0.5) = j, \\ 0, & \text{otherwise}, \end{cases}$$  

(13)

where the nested waiting times and durations satisfy $t_{iw}^1 = w_{hi}^1(k), t_{id}^1 = \text{Int}(t_{i}^1(k) + 0.5)$, $t_{lw}^1 = w_{hi}^2(k + t_{iw}^1 + t_{id}^1), t_{id}^2 = \text{Int}(t_{i}^2(k + t_{iw}^1 + t_{id}^1) + 0.5), \ldots$, for short. Similar requirement was also adopted by Chen and Hsueh (1998), Lam and Yin (2001).

Furthermore, according to the principle of flow propagation, for any arrival time $j$, the in-flow and out-flow of $i$ on $l$ which departs from $h$ entering $p$ during $k$ can be expressed as
the following functions, respectively:

\[ u_{lk}^{\text{hip}} (j + w_l(j)) = r_{lk}^{\text{hip}} (j), \] (14)

\[ v_{lk}^{\text{hip}} (j + w_l(j) + \ln(t_l(j) + 0.5)) = r_{lk}^{\text{hip}} (j). \] (15)

Equations (14) and (15) depict the relationships among link-specific arrival-flow, in-flow, out-flow, and duration. Given these link-specific flows, the arrival-flow, in-flow, and out-flow on \( l \) during \( j \), that is, \( r_l(j) \), \( u_l(j) \) and \( v_l(j) \), can be expressed as follows:

\[ r_l(j) = \sum_{h \in H} \sum_{i \in I} \sum_{p \in P_h} \sum_{k \in K} r_{lk}^{\text{hip}} (j), \] (16a)

\[ u_l(j) = \sum_{h \in H} \sum_{i \in I} \sum_{p \in P_h} \sum_{k \in K} u_{lk}^{\text{hip}} (j), \] (16b)

\[ v_l(j) = \sum_{h \in H} \sum_{i \in I} \sum_{p \in P_h} \sum_{k \in K} v_{lk}^{\text{hip}} (j). \] (16c)

Besides the link arrival-flow, in-flow and out-flow, link occupancy can be deemed as the fourth significant variable for describing the state of link flow. It can be calculated by:

\[ q_l(j) = \sum_{j' \leq j} r_l(j') - \sum_{j' \leq j} v_l(j'). \] (17)

In order to understand how travellers make their decisions, we need more detail of activity-travel path disutility. According to Assumption 2.6, time-dependent ATP disutility can be calculated by the following function:

\[ \text{dis} U_{hi}^{p} (k, f) = \sum_{l \in p} \sum_{j \in K} \delta_{lk}^{\text{hip}} (j) \cdot \text{dis} U_{ji}^{l} (j), \] (18)

where \( \text{dis} U_{hi}^{p} (k) \) represents time-dependent ATP disutility incurred by \( i \) departing from \( h \) and entering \( p \) during \( k \).

4. Data model and solution method

4.1. Data model

According to Assumption 2.1, travellers are assumed to be rational individuals who always choose their optimal path from all feasible ATPs to minimize the activity-travel disutility. Thus, the dynamic ATPs will converge to an equilibrium state in which no traveller wants to change his/her route unilaterally nor change his/her activity-travel schedule. The DUE model for the DATA problem can be expressed by finding \( f \) (i.e. vector of all time-dependent ATP inflows) such that the following conditions hold:

\[ \text{dis} U_{hi}^{p} (k, f) \begin{cases} = \text{dis} U_{\min}^{hi} (f), & \text{if } f_{hi}^{p} (k) > 0, \forall h, i, p, k, \\ \geq \text{dis} U_{\min}^{hi} (f), & \text{if } f_{hi}^{p} (k) = 0, \forall h, i, p, k, \end{cases} \] (19a)

\[ \sum_{p \in P_h} \sum_{k \in K} f_{hi}^{p} (k) = f_{hi}^{P}, \forall h, i, \] (19b)

\[ f_{hi}^{p} (k) \geq 0, \forall h, i, p, k, \] (19c)
where \( \text{dis} U^{hi}_{p}(k,f) \) is the equilibrium activity-travel disutility incurred by \( i \) departing from \( h \) entering \( p \) during \( k \) under flow pattern \( f \), \( \text{dis} U^{hi}_{\text{min}}(f) \) is the corresponding minimum disutility for \( i \) from \( h \), and \( F^{hi} \) is the total number of \( i \) at \( h \).

For Equation (19a), it is well known that, at an equilibrium state, for each \( i \) at \( h \) only those time-dependent ATPs (i.e. activity-travel path with specific departure time) that have the minimal disutility are used, and those time-dependent ATPs that are not used should have a disutility higher than or equal to the minimal disutility. Equations (19b) and (19c) represent the flow conservation and non-negativity constraints, respectively.

It should be noted that if interval length \( \omega \) is infinite close to zero (i.e. continuous-time), the link flows will be continuous with path flows, and then the existence of the DUE can be guaranteed (Nagurney 1999). Unfortunately, the DUE model is formulated and computed on the basis of time slices. According to Equations (12)–(15), link arrival-flows, in-flows and out-flows maybe not continuous with ATP flows \( f \) owing to half adjust of link travel times. Consequently, the value of \( \omega \) should be small enough so that it can ensure the existence of the proposed discrete- time DUE; or at least ensure that the computing result by a solution method is very close to the corresponding continuous-time DUE solution (i.e. within an acceptable tolerance). In addition, due to the inclusion of 0–1 integer indicator variables \( \delta_{hk}^{hp}(j) \), the ATP disutilities calculated by Equation (18) are non-convex (Janson 1991; Lam and Yin 2001). Consequently, the DUE conditions (19a)–(19c) may have multiple local solutions.

### 4.2. Solution method

In this paper, we adopt a route-swapping solution method (Huang and Lam 2002; Szeto and Lo 2006; Liu et al. 2015) to solve the DATA problem Equations (19a)–(19c). The details are presented as follows:

**Step 0:** Initialization. Set parameters \( \rho > 0, \mu > 0 \), and convergence tolerance \( \varepsilon > 0 \); choose scale parameters sequence \( \{\rho_{\tau}\} \); set iteration index \( \tau = 0 \) and let \( f_{0} \) be initial inflow vector.

**Step 1:** Compute ATP disutility. Compute waiting disutility \( \text{dis} U^{w}_{hi}(k)_{\tau} \) by Equations (2) and (3); compute schedule delay disutility \( \text{dis} U^{s}_{hi}(k)_{\tau} \) by Equation (4); compute duration disutility \( \text{dis} U^{d}_{hi}(k)_{\tau} \) by Equation (10); compute link travel disutility \( \text{dis} U^{l}_{hi}(k)_{\tau} \) by Equation (1) and ATP disutility \( U^{hi}_{p}(k)_{\tau} \) by Equation (18).

**Step 2:** Check the stopping criterion. The iteration terminates if

\[
\frac{\sum_{h \in H} \sum_{i \in I} \sum_{p \in P_{h}} \sum_{k \in K} f^{hi}_{p}(k)_{\tau} \cdot [\text{dis} U^{hi}_{p}(k)_{\tau} - \text{dis} U^{hi}_{\text{min}}(f_{\tau})]}{\sum_{h \in H} \sum_{i \in I} \sum_{p \in P_{h}} \sum_{k \in K} f^{hi}_{p}(k)_{\tau} \cdot \text{dis} U^{hi}_{\text{min}}(f_{\tau})} < \varepsilon. \tag{20}
\]

**Step 3:** Update ATP inflow. For \( h \in H, i \in I \), compute

\[
f^{hi}_{p}(k)_{\tau+1} = \begin{cases} \max\{0, f^{hi}_{p}(k)_{\tau} - \rho_{\tau} f^{hi}_{p}(k)_{\tau} [\text{dis} U^{hi}_{p}(k)_{\tau} - \text{dis} U^{hi}_{\text{min}}(f_{\tau})]\}, & \text{if} \ (p,k) \notin PK^{hi}_{\tau}, \\ f^{hi}_{p}(k)_{\tau} + \sum_{(p,k) \in PK^{hi}_{\tau}} \frac{[f^{hi}_{p}(k)_{\tau} - f^{hi}_{p}(k)_{\tau+1}]}{|PK^{hi}_{\tau}|}, & \text{if} \ (p,k) \in PK^{hi}_{\tau}, \end{cases} \tag{21}
\]
where

\[ PK^hi = \{(p, k) : \text{dis} U^hi_p(k) = \text{dis} U^hi_{\text{min}}(f_t), \forall p, k\}, \]  

(22)

let \( \tau = \tau + 1 \), and return to step 1.

In Step 0, the initial inflow vector \( f_0 \) is simply averaged across all time-dependent ATP. Set \( \rho(\tau) = \rho / \text{Int}(\tau / \mu + 1) \), which satisfies \( \lim_{\tau \to \infty} \rho_\tau = 0 \), \( \rho_\tau f^hi_p(k) = 0 \) and \( \sum_\tau \rho_\tau f^hi_p(k) = \infty \). Therefore, sequence \( \{f_\tau : \tau = 0, 1, 2, \ldots\} \) generated by this solution method will converge to a flow pattern even if the ATP disutilities are not monotonic (Huang and Lam 2002). In Step 2, the left-hand side of the stopping criterion is a gap function measuring how close a \( \tau \)th solution is to equilibrium Equations (19a)–(19c). If Equation (20) is met, the procedure is terminated and the final solution is obtained. In Step 3, Equations (21) and (22) generate a new flow pattern \( f_{\tau+1} \) which deducts flows from the time-dependent paths with higher disutility than the minimum disutility and adds these flows onto the paths with minimum disutility. The sum of the deductions is equal to the sum of the additions so as to ensure the validity of the solution from the swapping procedure. Meanwhile, this shift guarantees that inflows are non-negative.

5. Numerical examples

In this section, we present the numerical results of three examples to illustrate the proposed DATA model with capacity constraints. Example 1, containing a many-to-one public transit line depicted in Figure 2 (first used by Tian, Huang, and Yang 2007), is meant to illustrate the commuting properties of in-vehicle crowding and departure time choice in PTN. Example 2 is tested on a network depicted in Figure 4 (first used by Chen and Hsueh 1998) for illustrating the effects of capacity constraints on quality of service and activity location choice. Example 3 is tested on a multi-mode network depicted in Figure 6 for illustrating multi-source capacity constraints and multi-dimensional choices.

5.1. Example 1 (departure time choice with in-vehicle crowding)

As shown in Figure 2, both home locations \( h_1 \) and \( h_2 \) have 1000 residents (only one class), who need go to \( w \) for work in the morning. The work place \( w \) is connected with \( h_1 \) and \( h_2 \) by walking links 1, 2 and 3 and transit links 4 and 5. The travel time and disutility of unit walk time on links 1, 2 and 3 are assumed to be \( t_1^1 = t_2^1 = t_3^1 = 10 \text{ min} \) and \( \alpha_1^1 = \alpha_2^1 = \alpha_3^1 = 0.1/\text{min} \), respectively. The travel time, unit waiting time disutility and unit duration disutility of transit links 4 and 5 are assumed to be \( t_4^1 = t_5^1 = 20 \text{ min} \) and \( \alpha_4^1 = \alpha_5^1 = 0.15/\text{min} \), respectively. In addition, a linear in-vehicle crowding disutility function is adopted (i.e. \( \theta_4 = \ldots \)
θ_5 = 1). Set capacities \( c_4 = c_5 = 50 \) persons per transit run, parameters \( \eta_4 = \eta_5 = 0.15 \) and \( \lambda_4 = \lambda_5 = 0 \).

Without loss of generality, we define link 6 as working in the corresponding SNK. The working hour and unit working time utility are assumed to be \( t_6^1 = 480 \) min and \( \alpha_6^1 = 0.2/\text{min} \), respectively. The desired start time window of working is set as \( [o_6^1, e_6^1] = [9:00, 9:00] \), which means 9:00 a.m. is the only one desired start time. The parameter of unit time penalty of arriving early and late are assumed to be \( \beta_6^1 = 0.05 \) and \( \gamma_6^1 = 0.2/\text{min} \), respectively. Transit time table is assumed to be \( \{6:00, 6:10, 6:20 \ldots 22:00\} \). Interval length is set as 1 min. Finally, we set parameters \( \mu = 500 \), \( \rho = 0.01 \) and \( \varepsilon = 10^{-4} \).

Figure 3 shows the solution results of time-dependent ATP inflows and disutilities departing from \( h_1 \) and \( h_2 \).

In fact, travellers may change their departure time choice spontaneously during morning commuting peak in order to avoid excessive in-vehicle crowding. It is worth mentioning that if there is no in-vehicle crowding effect (caused by capacity constraint in PTN), travellers will only choose the transit run with respect to on-time arrival. The set of departure times consists of discrete time intervals, that is, 6:30, 6:40 \ldots 8:20 that are chosen by travellers from \( h_1 \) during the morning peak period. Thereinto, departure times 7:10, 7:20 \ldots 8:10 are associated with an identical and maximum number of traveller, that is, \( f_1^{1,1}(k) = 109 \). If travellers from \( h_2 \) board one transit run, a certain number of travellers from \( h_1 \) are already in the vehicle. In addition, the on-time arrival transit run carries a maximum number of travellers from both \( h_1 \) and \( h_2 \). The time-increasing rate of the travellers arriving early at working place is less than the time-declining rate of the late-arrival travellers, because the unit time penalty of arriving early is smaller than that of arriving late. These flow patterns are identical to that proved and shown in Tian, Huang, and Yang (2007). Furthermore, travellers only choose those departure times without waiting for transit runs.

5.2. Example 2 (activity location choice with quality of service)

As shown in Figure 4, there are in total 2000 travellers work at, which is connected with restaurants \( r_1 \) and \( r_2 \) (can be chosen for having lunch) by directed walking links 1–6.
travel times of links 1–6 are fixed and assumed to be 16, 8, 10, 10, 5, and 5 min, and unit walking time disutility is set as 0.1/min. Without loss of generality, we define links 7 and 8 as having lunch at \( r_1 \) and \( r_2 \) in the corresponding SNK, respectively. The desired start time window of having lunch is set as \([12:00, 12:00]\). There are two feasible paths, that is, \( p_1 = \{1, 7, 3, 4\} \) and \( p_2 = \{2, 8, 5, 6\} \).

In addition, the length of lunch time, unit time penalty of early and late are assumed to be 30 min, 0.1 and 0.15/min, respectively. A quadric in-location crowding disutility function is adopted (i.e. \( \theta_7 = \theta_8 = 2 \)). Set capacities \( c_7 = c_8 = 100 \), parameters \( \eta_7 = \eta_8 = 0.5 \) and \( \lambda_7 = \lambda_8 = 0.5 \). Set 1000 travellers have a preference for restaurants \( r_1 \) and the others prefer \( r_2 \), specifically \( \alpha_{17} = \alpha_{28} = 0.4 \) and \( \alpha_{18} = \alpha_{27} = 0.35/\text{min} \). Interval length is set as 1 min. Finally, we set parameters \( \mu = 400, \rho = 0.02 \) and \( \varepsilon = 10^{-4} \).

Figure 5 shows the solution results of the time-dependent ATP inflows and disutilities departing for restaurants \( r_1 \) and \( r_2 \). It demonstrates the equilibrium state that the used time-dependent paths have the least disutility and unused time-dependent paths have higher disutilities, which illustrates that the computing results are in accord with DUE model. If considering capacity constraints at the locations, travellers’ trade-off between the decline of quality of services and schedule delay disutilities will lead to the peak period for having lunch. As shown in Figure 5(a) and 5(b), the departure peak period for having lunch at \( r_1 \) is from 11:37 to 11:47, and it is shorter that for \( r_2 \) (from 11:41 to 11:57), because the travel cost between \( w \) and \( r_2 \) is smaller. It is noteworthy that \( r_1 \) and \( r_2 \) are both chosen by class 1 traveller, although they have a higher preference for \( r_1 \). In addition, there is no class 2 traveller choosing \( r_1 \), because the class 2 traveller prefers to \( r_2 \) while \( r_2 \) is closer to
Furthermore, without capacity constraints at the locations, the departure time interval at equilibrium state will degenerate into a point-in-time. As the travel times of links 1–6 are assumed to be fixed, it possibly results in the phenomenon that all travellers choose the same location or the same time for having lunch without considering the decline of quality of services. It means that travellers will only choose the desired start time for having lunch, which is unrealistic.

### 5.3. Example 3 (multiform capacity constraints and choices)

As shown in Figure 6, there are in total 2000 travellers who live at $h$. Of these, 1000 have private cars, while the others do not. Work place $w$ is connected with $h$ by private vehicle links (i.e. links 1 and 2) and public transit links (i.e. links 3 and 4). Transit time table is assumed to be $\{6:00, 6:10, 6:20, \ldots, 22:00\}$. Shopping location $s_1$ is connected with $w$ via walking links 5 and 6; $s_2$ is connected with $h$ via private vehicle links 7 and 8. In addition, three parking places $r_h$, $r_w$, and $r_s$ are located at $h$, $w$ and $s_2$, respectively. We obtain the corresponding activity-based multi-state supernetwork, as shown in Figure 7. In this SNK, besides traffic links 1–8, other links numbered 9–17 are defined. Links 9, 10, and 11 represent working, shopping at $s_1$ and shopping at $s_2$, respectively; links 12–17 represent picking up or parking at $r_h$, $r_w$ or $r_s$, respectively. Detailed parameter settings for all activity-travel links are given in Table 1.

As shown in Figure 7, a path denoted by the bold links shows that the traveller first picks up the car at home $h$ and chooses link 1 to $w$ for working, parks the car at $r_w$, after work, walks to $s_1$ for shopping, and finally goes back to pick up the car at $r_w$ and drives home. This path can be represented as an ordered set $\{12, 1, 15, 9, 5, 10, 6, 14, 2, 13\}$. In fact, any path is a one-to-one mapping with a combined choice of route, mode, destination, and sequence.

### Table 1. Parameter settings of Example 3 (minutes or per minute).

<table>
<thead>
<tr>
<th>Number of link</th>
<th>Link attribute</th>
<th>Traveller attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_l$</td>
<td>$\eta_l$</td>
</tr>
<tr>
<td>Links 1 and 2</td>
<td>80</td>
<td>0.15</td>
</tr>
<tr>
<td>Links 3 and 4</td>
<td>50</td>
<td>0.15</td>
</tr>
<tr>
<td>Links 5 and 6</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Links 7 and 8</td>
<td>100</td>
<td>0.15</td>
</tr>
<tr>
<td>Link 9</td>
<td>2000</td>
<td>0.5</td>
</tr>
<tr>
<td>Links 10 and 11</td>
<td>100</td>
<td>0.5</td>
</tr>
<tr>
<td>Links 12–17</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 7. Multi-state supernetwork of Example 3.

Table 2. Feasible paths of Example 3.

<table>
<thead>
<tr>
<th>Link order</th>
<th>Number of path</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
<th>7th</th>
<th>8th</th>
<th>9th</th>
<th>10th</th>
<th>11th</th>
<th>12th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path 1</td>
<td>12</td>
<td>1</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>14</td>
<td>2</td>
<td>13</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Path 2</td>
<td>12</td>
<td>1</td>
<td>15</td>
<td>9</td>
<td>14</td>
<td>7</td>
<td>17</td>
<td>11</td>
<td>16</td>
<td>8</td>
<td>2</td>
<td>13</td>
<td>/</td>
</tr>
<tr>
<td>Path 3</td>
<td>12</td>
<td>1</td>
<td>15</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>2</td>
<td>13</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Path 4</td>
<td>12</td>
<td>1</td>
<td>7</td>
<td>17</td>
<td>11</td>
<td>8</td>
<td>15</td>
<td>9</td>
<td>14</td>
<td>2</td>
<td>13</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Path 5</td>
<td>3</td>
<td>9</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Path 6</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>6</td>
<td>9</td>
<td>4</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
</tbody>
</table>

The list of all feasible ATPs is presented in Table 2. It should be noted that class 2 traveller can only choose paths 5 and 6. The time horizon is taken from 6:00 to 22:00. Interval length was set as 2 min. Finally, we set parameters $\mu = 1000$, $\rho = 0.002$ and $\epsilon = 10^{-4}$.

Figure 8 displays the solution results of time-dependent ATP inflows and disutilities for class 1 and 2 travellers. All owning private vehicle travellers (i.e. class 1) choose path 2, because path 2 is provided with the lowest disutility in six feasible ATPs; similarly, all class 2 travellers choose path 5, because it is provided with the lower disutility in two feasible ATPs for no car travellers. In fact, the disutilities on paths 2 and 5 are increasing with the inflow of travellers. The travellers will still choose paths 2 and 5 until the number is increased beyond a point that paths 2 and 5 for each class have no longer the least disutilities. Then, other paths will be used and the path flow will be redistributed. Furthermore, as shown in Figure 8(a), class 1 traveller is concentrated around the continuous period of time, that is, 8:24–8:42 for departure. However, as shown in Figure 8(b), class 2 traveller is concentrated in the set
of discrete time points, that is, 7:44, 7:54, . . . , 9:04 for departure. The cause of this obvious distinction is different choices on traffic mode. It is worth mentioning that if there are no capacity constraints taking effects and exerting influences on congestion, crowding and decline of quality of services, the inflow peak of travellers will become very steep or even all-together in one moment, which is, apparently unrealistic.

6. Conclusion

The integration of activity-based modelling and dynamic traffic assignment has gained ever-increasing popularity for the travel demand analysis in transportation research. A unified framework closely coupling them has recently been proposed by formulating the DATA in multi-state supernetworks. Multi-dimensional choice facets can be determined endogenously in the time-dependent full daily ATPs. In addition to that, this paper proposes an approach to formulate capacity constraints of road usage for private vehicles and public transits, as well as parking and activity locations. The formulations have been consistently incorporated into the DATA framework, which enhances the realism of DATA to capture the choices of ATPs and traffic flows. Three numerical examples are presented to illustrate the effectiveness and feasibility of the approach. It is shown that capacity constraints have significant effects on the choice of ATPs and the ensuing traffic flows.

Nevertheless, several issues are worth considering. First, budget constraints of time and money have not been taken into account. According to previous empirical studies, budget constraints like those formulated in this paper may have also strong effects on the choice of ATPs and hence traffic flows. Second, nearly all travel demand forecasting systems have still focused on individual ATPs. Whereas, the multi-state supernetwork representations are also powerful to represent joint ATPs (Liao, Arentze, and Timmermans 2013b), which allows the possibility of modelling joint ATPs at the traffic assignment stage in the DATA framework. Third, the present study adopts the choice mechanism of utility maximization under the deterministic representation of activity-travel time. The extensions of DATA under uncertainty and other choice mechanisms are also of great interest. These issues will be addressed in the future research.
Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This research is jointly supported by the National Natural Science Foundation of China [grant number 71361130011], Netherlands Organisation for Scientific Research, and grants from the National Basic Research Program of China [grant number 2012CB725401] and the Innovation Foundation of BUAA for Ph.D. graduates. These supports are gratefully acknowledged.

References