IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 20, 2019, accepted April 1, 2019, date of publication April 9, 2019, date of current version April 19, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2909524

# Spatial and Temporal Analyses for Estimation of Origin-Destination Demands by Time of Day Over Year

LIANG SHEN[1], HU SHAO[1,2], TING WU[3], AND WILLIAM H. K. LAM[2]
[1]School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China
[2]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong
[3]Department of Mathematics, Nanjing University, Nanjing 210093, China

Corresponding author: Ting Wu (tingwu@nju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Project 71671184 and Project 71571096, in part by the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project PolyU 152628/16E, and in part by the Social Science Fund of Jiangsu Province under Project 14GLC001.

**ABSTRACT** This paper proposes a two-stage model for the estimation of origin–destination (OD) demands by the time of day over the year with the use of offline traffic data from the real-time travel information system. In the first stage, a travel time recursive function is proposed to use the offline travel speed data for the investigation of the spatial and temporal relationships between time-dependent OD demands and traffic counts. As such, it is not required to carry out the time-consuming dynamic traffic assignment (DTA) process which is frequently used in the conventional time-dependent OD estimation models. Using the results in the first stage together with the available traffic count data, a least-squares method is adopted to formulate the time-dependent OD demand estimation problem as a quadratic programming model in the second stage. A solution algorithm is adapted for solving the proposed model. Then, the proposed method is easy for implementation in practice. Particularly, when the traffic accident occurs in the network, the estimated time-dependent OD demands can be helpful for understanding the complex travel behavior (e.g., departure time choice) under uncertainty condition. The numerical examples are presented to illustrate the applications of the proposed model.

**INDEX TERMS** OD demand estimation, dynamic traffic assignment, real-time traffic information, least squares, quadratic programming.

## I. INTRODUCTION

Time-dependent origin-destination (OD) traffic demand is one of the necessary and basic data for traffic monitoring and management. Particularly, time-dependent OD demand can be used as an input for dynamic traffic assignment (DTA) systems to capture the spatial and temporal distributions of traffic flows in transportation network. Existing methods for time-dependent OD estimation can be classified into two types as follows.

- The first type of method refers to the conventional survey-based method for gathering OD demand information. Such method is difficult to be applied in a large-scale network due to its high price in terms of monetary, time and labor costs [1].
- The second type of method is the traffic surveillance data-based method. In the past decades, substantial research efforts using optimization methods have been invested in estimating the time-dependent OD demand from a variety of traffic surveillance data, such as traffic counts, vehicle trajectory, turning movements [1]–[10]. The traffic surveillance data can be collected automatically by the sensor systems at low price in transportation networks. Thus, such methods for time-dependent OD estimation are very popular.

It should be pointed out that the key idea of surveillance data-based method is to make the estimated time-dependent OD demand approximate the observed data (e.g. traffic counts) from the viewpoint of statistics. Mathematically, the estimated time-dependent OD demand is kept being adjusted until it matches the observed data (e.g. traffic counts) as much as possible. For such purpose, the estimated

The associate editor coordinating the review of this manuscript and approving it for publication was Najah Abu Ali.

time-dependent OD demand is assigned to the network links by DTA process [11]–[15] so as to acquire the time-dependent link choice proportion and then the estimated traffic counts. Using the estimated traffic counts and the surveillance data, the statistical tools can be used for the time-dependent OD estimation problem. Moreover, the time-dependent OD can be estimated by minimizing the distance between the observed and estimated traffic counts (least squares method), or maximizing the likelihood of the observed traffic counts (maximum likelihood method). For instance, Cascetta *et al.* [16] extended the concepts of static OD estimation problem and formulated a generalized least square (GLS) based framework for estimating dynamic OD demands. Lu *et al.* [17] proposed a single-level dynamic OD demand estimation model, without using link proportions, based on the mathematical program with complementary (or equilibrium) constraints approach. Lu *et al.* [18] presented an enhanced Simultaneous perturbation stochastic approximation (SPSA) algorithm, called Weighted SPSA (W-SPSA), which incorporates the information of spatial and temporal correlation in a traffic network to limit the impact of noise and improve convergence and robustness. Cipriani *et al.* [19] presented a new method to solve the simultaneous adjustment of a dynamic traffic demand matrix, searching for a reliable solution with acceptable computational times for off-line applications and using as an input traffic counts and speeds, prior OD matrices and other aggregate demand data (traffic demand productions by zone). However, the main difficulty of these methods comes from the implementation of DTA. DTA itself is difficult and challenging in its own right [1]. DTA model needs extensive calibration [7] and is very time-consuming for implementation.

To avoid the DTA process, some researchers have studied the problem of travel demand estimation using mobile phone data [20]–[24]. For example, Iqbal *et al.* [20] proposed a methodology to develop OD matrices using mobile phone Call Detail Records (CDR) and limited traffic counts. Toole *et al.* [21] presented a full implantation of a travel demand model that uses new, big data resources as input. Ge and Fukuda [22] proposed an approach to estimating trip flows with aggregated mobile phone datasets and low-cost surveys. Huang *et al.* [23] proposed a human mobility model that combines mobile phone signaling data and urban transportation data. The model provides a way to predict real time urban travel demand at high temporal and spatial resolution. Ni *et al.* [24] proposed a multi-layered Hierarchical Flow Network (HFN) to formulate the simultaneous estimation problem of different levels of traffic demand and behavioral parameters. Meanwhile, the authors maped different layers of HFN to different big data sources, including household travel surveys, mobile phone data, floating car data, and sensor data. Although the mobile phone data is very helpful for time-dependent OD demand estimation, such kind of data is not easy to access due to privacy issue. Alternatively, this paper proposes a new method for time-dependent OD demand estimation. Such method utilizes the data from automatic vehicle identification (AVI) sensor system. And the corresponding method is different from that using mobile phone data.

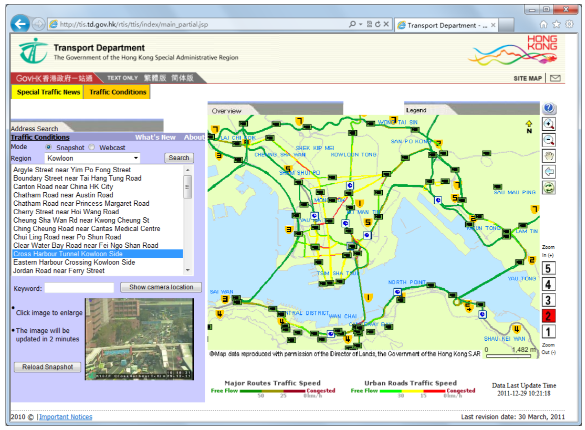

**FIGURE 1.** Real-time travel speed map in Hong Kong.

In Hong Kong road network, time-dependent link choice proportion can be obtained by the time-dependent traffic speed information, which is available in the journey time indication system (JTIS) of Hong Kong Transport Department. Specifically, the JTIS can provide the mean traffic speeds every two minutes on the major road links of Hong Kong over the whole year. Such traffic speed data shown in Figure 1, is also available to public by "Traffic Speed Map (TSM)" on Hong Kong Transport Department's homepage (http://tis.td.gov.hk/rtis/ttis/index/main_partial.jsp). Different from conventional approaches, JTIS utilizes the data from automatic vehicle identification (AVI) sensor system to estimate the time-dependent mean travel speed on road links instead of carrying out DTA. Details on JTIS can be referred to Tam and Lam [25]. It is noted that the time-dependent traffic speed provided by JTIS is a direct result of assigning time-dependent OD demand to the links (say DTA). Thus, the time-dependent speed can in turn be used to deduce the time-dependent link choice proportion. To do so, a travel time recursive formula, on the basis of the time-dependent flow speeds shortest path method by Sung *et al.* [26] is proposed. Using this formula the completed (actual) travel time of each path starting from any time interval can be calculated. Then, under any of the dynamic traffic assignment equilibrium assumptions [27], [28], time-dependent link choice proportion can be easily calculated. The flowchart for calculating time-dependent link choice proportion is shown in Figure 2.

An illustrative example is proposed to further demonstrate how to calculate the time-dependent link choice proportion so as to capture the spatial and temporal relationships between the time-dependent OD demands and the traffic accounts. The illustrative example network is shown in Figure 3. The observed travel speed information is presented in Table 1. It can be seen that there is only one path between each OD pair. Thus, the path flow is equal to the OD demand. And there is no need of the DTA equilibrium assumption in this

**TABLE 1.** Spatial and temporal relationship between traffic counts and time-depend OD demands.

| Time interval | Mean speed (km/minute) | | | Spatial and temporal relationship | Equation no. |
|---|---|---|---|---|---|
| | Link 1 | Link 2 | Link 3 | Link 3 | |
| $t_1$ [8:00, 8:02) | 0.5 | 1 | 1 | $v_3^{t_1} = 0\times q_{14}^{t_1} + \frac{3}{4}\times q_{24}^{t_1}$ | (1) |
| $t_2$ [8:02-8:04) | 1 | 0.5 | 1 | $v_3^{t_2} = 1\times q_{14}^{t_1} + \frac{1}{2}\times q_{14}^{t_2} + \frac{1}{4}\times q_{24}^{t_1} + \frac{1}{2}\times q_{24}^{t_2}$ | (2) |
| $t_3$ [8:04-8:06) | 1 | 1 | 1 | $v_3^{t_3} = \frac{1}{2}\times q_{14}^{t_2} + \frac{1}{2}\times q_{14}^{t_3} + \frac{1}{2}\times q_{24}^{t_2} + \frac{1}{2}\times q_{24}^{t_3}$ | (3) |
| … | … | … | … | … | … |

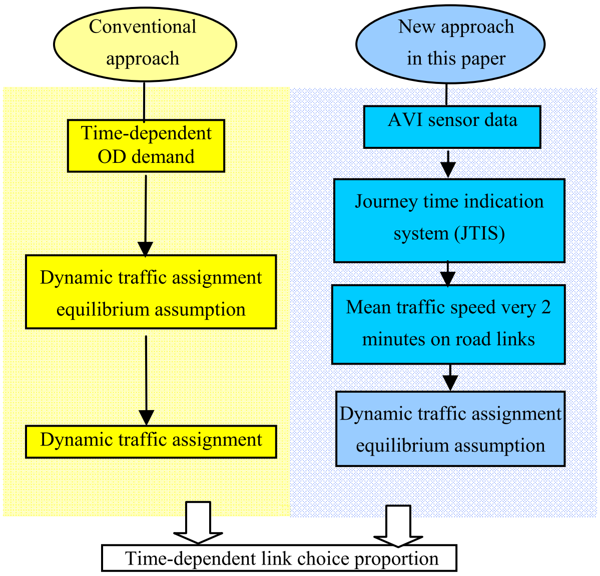


**FIGURE 2.** Flowchart for calculating time-dependent link choice proportion.

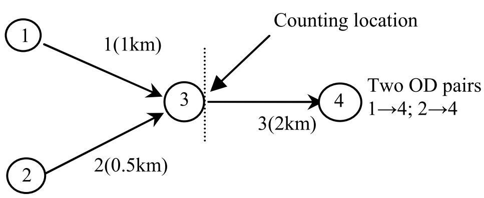


**FIGURE 3.** An illustrative network.

example. The length of each link is denoted in the parentheses of the link number near each link. The length of the time interval is 2 minutes. During the first time interval $i$, traffic flows between OD pair 1→4, $q_{14}^{t_1}$, does not reach the counting location. Thus, the traffic flow on link 3 during $t_1$, $v_3^{t_1}$, does not conclude the information of $q_{14}^{t_1}$. Similarly, only $\frac{3}{4}$ of $q_{24}^{t_1}$ can pass through the counting location. Thus, we can get the flow conservation condition by a linear equation: $v_3^{t_1} = 0 \times q_{14}^{t_1} + \frac{3}{4} \times q_{24}^{t_1}$, which describes the spatial and temporal relationship between traffic counts and time-dependent OD demands. Likewise, we can get Equations (2) and (3) for time intervals $t_2$ and $t_3$. If there is another counting location in the same network, we can also get the system of linear equations between time-dependent OD demands and traffic counts. These systems of linear equations include the same decision variables, i.e., the time-dependent OD demands of all OD pairs. Thus, they can be combined as a new system of linear equations. The solution of such system equations is the concerned time-dependent OD demands. It is known that the system of linear equations can be easily solved for large-scale problem. Therefore, the proposed method is applicable for large-scale network.

The proposed time-dependent OD estimation model is different from the existing ones. Existing models on time-dependent OD estimation using traffic surveillance data can be further divided into several groups with respect to three criteria (1) whether or not a DTA component is incorporated into the estimation process [29]–[31]; (2) whether or not the network is general [1]; (3) whether or not multiple sources of observed data are used. Table 2 summarizes some of the existing models using the above classification criteria. It can be seen from this table that most existing models on OD estimation problems either include the DTA process or link-path indicator except the works by Ma and Zhen [32]. In Ma and Zhen [32], the authors presented a data-driven framework that estimates day-to-day dynamic OD using high-granular traffic counts and speed data collected over many years based on the statistical equilibrium assumption [33], [34]. The proposed framework statistically clusters daily traffic data into typical traffic patterns using t-Distributed Stochastic Neighbor Embedding (t-SNE) and k-means methods. Inspired by the idea of Ma and Zhen [32], this paper proposes a two-stage model for estimation of time-dependent origin-destination demands using off-line traffic data from real-time travel information system. The time-dependent path/link indicator in Ma and Zhen [32] is extended from a Boolean parameter to a non-Boolean parameter using a newly proposed travel time recursive function in this paper. Based on the above discussions, the method proposed in our paper extends most of the existing studies for the time-dependent OD estimation.

The proposed model does not include the DTA process, which needs extensive calibration and is very time-consuming. Alternatively, a system of linear equations is adopted to figure out the spatial and temporal relationship between time-dependent OD demand and traffic counts. As a result, the proposed method is applicable to general network. The estimated off-line time-dependent OD demands have many applications in traffic monitoring and management.

**TABLE 2. Classification of time-dependent OD estimation models.**

| Models | DTA | | Networks | | Sources of data | | Link-path indicator | |
|---|---|---|---|---|---|---|---|---|
| | DTA | Non-DTA | Simple networks | General networks | Single source | Multiple sources | Logical variable | Nonlogical variable |
| Cremer and Keller [36,37] | | √ | √ | | √ | | / | / |
| Bell [38], Chang and Wu [39], Wu and Chang [40] | √ | | √ | | √ | | / | / |
| Lu [17], Kang [29], Osorio [35], Tavana [41], Zhou [42], Zhou and Mahmassani [43] | √ | | | √ | √ | | / | / |
| Ma and Qian [32] | | √ | | √ | | √ | √ | |
| Proposed model in this paper | | √ | | √ | | √ | | √ |

Moreover, it is noted that the travel speed information of the whole year is available in JTIS of Hong Kong. As a result, the time-dependent OD demand can be obtained by using the proposed model. Particularly, when the traffic accident occurs in the network, the estimated time-dependent OD demands can be helpful for understanding the complex travel behavior (e.g. departure time choice) under uncertainty condition.

The rest of the paper is organized as follows. In the next section, a two-stage model is proposed for the estimation of time-dependent OD demand together with mathematical proofs for the solution existence and uniqueness. Then, a heuristic solution algorithm is proposed in Section 3. Numerical examples and results are discussed in Section 4. Finally, conclusions and further studies are given in Section 5.

## II. MODEL FORMULATION

### A. LIST OF SYMBOLS

| | |
|---|---|
| $\mathbf{G} = (\mathbf{N}, \mathbf{A})$ | A road network, with $\mathbf{N}$ being the set of nodes and $\mathbf{A}$ being the set of links, respectively. |
| $\mathbf{R}$ | Set of OD pairs, $\mathbf{R}$ is a subset of $\mathbf{N} \times \mathbf{N}$. |
| $r$ | Origin node. |
| $s$ | Destination node. |
| $\mathbf{K}_{rs}$ | Non-empty path set between OD pair $rs$. |
| $t_i$ | Time interval $i$, $t_i = [t_i^-, t_i^+)$ |
| $t_w$ | The length of each time interval, i.e., $t_w = t_i^+ - t_i^-$ |
| $q_{rs}^{t_i}$ | Traffic flow between OD pair $rs$ during time interval $t_i$. |
| $f_{rs}^{t_i}$ | Traffic flow on path $k \in \mathbf{K}_{rs}$ during time interval $t_i$. |
| $p_{rs}^{k,t_i}$ | Choice proportion of path $k \in \mathbf{K}_{rs}$ during time interval $t_i$, $f_{rs}^{k,t_i} = p_{rs}^{k,t_i} q_{rs}^{t_i}$ |
| $v_a^{t_i}$ | Estimated traffic counts on link $a$ during time interval $t_i$ |
| $\tilde{v}_a^{t_i}$ | Observed traffic counts on link $a$ during time interval $t_i$ |
| $s_a^{t_i}$ | Mean travel speed on link $a$ during time interval $t_i$ |
| $T_{rs}^{k,t_i}$ | Travel time of $f_{rs}^{t_i}$ between OD pair $rs$ |
| $\bar{l}_{a_i}$ | Distance of link $a_i$ |
| $n_1$ | Number of time intervals of estimated OD demand |
| $n_2$ | Number of time intervals of observed link flows |
| $\delta_{rs}^{k,a}(t_j, t_i)$ | The time-dependent link-path indicator |

### B. ASSUMPTIONS

To facilitate the presentation of the essential ideas without loss of generality, the following basic assumptions are made in this paper:

A1. It is assumed that the length of all the time intervals is equal to 2 minutes, which is consistent with the JTIS system in Hong Kong.

A2. As the length of the time interval in the JTIS is short (say 2 minutes), the departure flow rate during the period of each time interval is assumed to be uniform. That is to say the fluctuation of departure flow within a time interval is ignored in this paper.

A3. The length of each link is sufficiently long so that the mean travel time on each link is no less than 1 minute.

A4. It is assumed that the path set is given and fixed for each OD pair. In reality, it may be very difficult to identify the used path set. This set of path can be limited to around six to eight paths for a realistic network and can be based on the actually chosen paths obtained by observation and/or interview surveys [44]–[46].

A5. Travelers are assumed to make their path choice decisions according to the stochastic dynamic user-optimal (SDUO) path choice model [28].

A6. It is assumed that the link travel speeds are uniformly distributed. And the temporospatial variation of travel speeds due to presence of shock waves (e.g. signalized arterial) is omitted.

A7. For simplicity, it is assumed that the path set is predetermined and fixed for all time intervals. That is to say if a particular path flow of the first time interval is assigned to

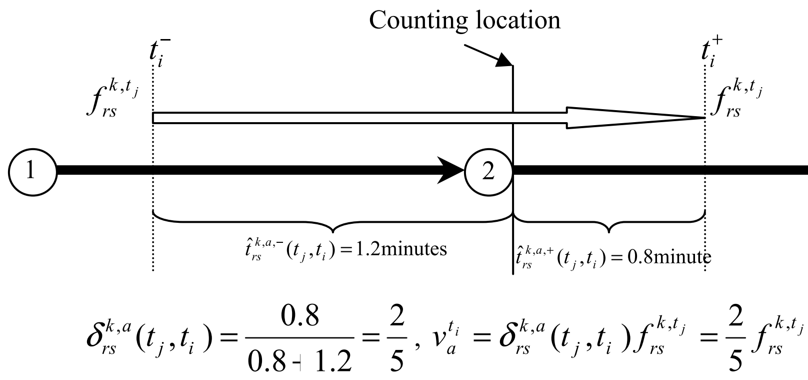


**FIGURE 4.** **Illustration on the time-dependent traffic counts and path flows.**

a link a, it will always be assigned to link a for all the time intervals.

### C. STAGE I: TIME-DEPENDENT LINK CHOICE PROPORTION

The proposed method for time-dependent OD estimation adopts a two-stage modeling framework. In the first stage, the time-dependent link choice proportion is calculated by the travel speed information. Then, in the second stage, the time-dependent OD demand is estimated by a modified least squares model. In this paper, the study time period $[0, t)$ is divided into a sequence of time intervals:

$$t_1 = [t_1^-, t_1^+), \quad t_2 = [t_2^-, t_2^+), \ldots, t_n = [t_n^-, t_n^+), \ldots \quad (1)$$

where $t_1^- = 0, t_i^+ = t_{i+1}^-, t_n^+ = t$ and $[0, t) = \bigcup_{i=1}^{n} t_i$. Denote $t_w$ as the length of time interval:

$$t_w = t_i^+ - t_i^- i = 1, 2, \ldots \quad (2)$$

According to assumption A1, it is assumed $t_w = 2$ in this paper. The time-dependent OD demand $q_{rs}^{t_i}$ can be expressed as the summation of time-dependent path flows as follows:

$$q_{rs}^{t_i} = \sum_{k \in \mathbf{K}_{rs}}^{f_{rs}^{k,t_i}} \quad (3)$$

Denote $p_{rs}^{k,t_i}$ as the time-dependent path choice proportion. Then, it follows from Equation (3) that

$$f_{rs}^{k,t_i} = p_{rs}^{k,t_i} q_{rs}^{t_i} \quad (4)$$

As shown in Figure 4, the counting location for collecting link traffic counts are assumed at the upstream node (entrance) of the link, the estimated traffic counts on link $a$ during time interval $t_i$ are assumed as, $v_a^{t_i}$, which is the total number of vehicles passing through the counting location of link $a$ during the time interval $t_i$.

As shown in Figure 4, the time-dependent link-path indicator $\delta_{rs}^{k,a-}(t_j, t_i) = 0.4$, that is to say in time interval $t_i = [t_i^-, t_i^+)$, 40% traffic flow of time-dependent path flow $f_{rs}^{k,t_j}$ have passed the counting location. In this figure, $\hat{t}_{rs}^{k,a,-}(t_j, t_i) = 1.2$minutes and $\hat{t}_{rs}^{k,a,+}(t_j, t_i) = 0.8$minute are given and fixed for illustration purpose. In our proposed model, the $\hat{t}_{rs}^{k,a,-}(t_j, t_i)$ and $\hat{t}_{rs}^{k,a,+}(t_j, t_i)$ can be calculated by the speed information according to equation (10), which is shown in Figure 6. Mathematically, $v_a^{t_i}$ can be expressed by the summation of the all path flows passing through the counting location during time interval $t_i$. Then, $v_a^{t_i}$ can be expressed as follows.

$$v_a^{t_i} = \sum_{j=1}^{i} \sum_{rs \in \mathbf{R}} \sum_{k \in \mathbf{K}_{rs}} \delta_{rs}^{k,a}\left(t_j, t_i\right) f_{rs}^{k,t_j} \quad (5)$$

where $\delta_{rs}^{k,a}\left(t_j, t_i\right)$ is the time-dependent link-path indicator defined as follows.

$$\delta_{rs}^{k,a}(t_j, t_i) = \frac{\hat{t}_{rs}^{k,a,+}(t_j, t_i)}{\hat{t}_{rs}^{k,a,-}(t_j, t_i) + \hat{t}_{rs}^{k,a,+}(t_j, t_i)}, \quad j \leq i \quad (6)$$

In the above equation, $\hat{t}_{rs}^{k,a,-}(t_j, t_i)$ is the travel time of $f_{rs}^{k,t_j}$ before the counting location during time interval $t_i$; $\hat{t}_{rs}^{k,a,+}(t_j, t_i)$ is the travel time of $f_{rs}^{k,t_j}$ after the counting location during time interval $t_i$ as shown in Figure 4. It is obvious that the travel time interval $t_i = [t_i^-, t_i^+)$ is 2 minutes as shown in Figure 4. For simplicity, we assumed that travelers depart at 8:00 ($t_i^- = 8:00$). Then, the travel time of $f_{rs}^{k,t_j}$ before the counting location during time interval $t_i$ can be calculated according to the speed data and the travel time is 1.2 minutes. Meanwhile, the travel time of $f_{rs}^{k,t_j}$ after the counting location during time interval $t_i$ can also be obtained according to the speed data and the travel time is 0.8 minutes. Thus, the traffic flow on link $a$(between nodes 2 and 3) is $v_a^{t_i} = \delta_{rs}^{k,a}(t_j, t_i) f_{rs}^{k,t_j} = \frac{2}{5} f_{rs}^{k,t_j}$. That is to say, only 0.4 of $f_{rs}^{k,t_j}$ can pass through the counting location during time interval $t_i$. It should be pointed out that $\delta_{rs}^{k,a}(t_j, t_i)$ can be any number in the interval [0,1]. Particularly, $\delta_{rs}^{k,a}(t_i, t_i) = 1$, if the counting location is located at the origin node. Under this condition, $\hat{t}_{rs}^{k,a,-}(t_i, t_i) = 0$ because all the traffic flow of time interval $t_i$ is observed by the counting location. Then, it follows that $\delta_{rs}^{k,a}(t_j, t_i) = 1$ according to Equation (6).

Substitute Equation (6) into Equation (5) we can get that

$$v_a^{t_i} = \sum_{j=1}^{i} \sum_{rs\in\mathbf{R}} \sum_{k\in\mathbf{K}_{rs}} \delta_{rs}^{k,a}\left(t_j, t_i\right) p_{rs}^{k,t_i} q_{rs}^{t_i} \tag{7}$$

where $\delta_{rs}^{k,a}\left(t_j, t_i\right) p_{rs}^{k,t_i}$ is the time-dependent link choice proportion, which indicates the spatial and temporal relationship between the time-dependent OD demands and link traffic counts. According to assumption A5, the path choice proportion, $p_{rs}^{k,t_i}$, can be calculated as follows.

$$p_{rs}^{k,t_i} = \frac{\exp(-\theta T_{rs}^{k,t_i})}{\sum_{k'\in\mathbf{K}_{rs}} \exp(-\theta T_{rs}^{k',t_i})} \tag{8}$$

where $\theta$ is the dispersion parameter. It can be seen that the $p_{rs}^{k,t_i}$ is a function with respect to the completed (or actual) travel times $T_{rs}^{k,t_i}$ of all paths between OD pair $rs$ during time interval $t_i$. A new method, named spatial-temporal scan method, is proposed for calculating $T_{rs}^{k,t_i}$ using the time-dependent travel speed data from JTIS. The mean travel speed on link $a$ during time interval $t_i$ is denoted as $s_a^{t_i}$, which is available in JTIS of Hong Kong. Denote $\hat{T}_{rs}^{k}(t_j^-, l)$ as the mean travel time starting from the location $l$ to the destination node ($l$ is the distance from the origin node on path $k$) at time $t_j^-$ as shown in Figure 5.

For convenience, it is assumed that path $k$ consists of link sequence $\{a_1, a_2, \ldots, a_n\}$. Denote

$$l_{a_i} = \sum_{j=1}^{i} \bar{l}_{a_j} \tag{9}$$

where $\bar{l}_{a_j}$ is the length of link $a_j$. Then, $\hat{T}_{rs}^{k}(t_i^-, l)$ can be expressed by the following travel time recursive function (10), where $l_{a_0} = 0$.

There are five cases for the calculation of $\hat{T}_{rs}^{k}(t_i^-, l)$. Cases 1, 2 and 3 represent that the concerned traffic flow does not reach the destination node in the next time interval $t_i^+$ as shown in Figure 6(a). Case 1 represents that the concerned traffic flow still remains at the same link ($a_j$) in next time interval. Case 2 represents that the concerned traffic flow passes through link $a_j$ and remains at the next link $a_{j+1}$ in next time interval. Case 3 represents that the concerned traffic flow passes through links $a_j$ and $a_{j+1}$ remains at link $a_{j+2}$ in next time interval. It should be noted that the concerned traffic flow can at most pass through two links in one time interval. Thus, there are only three cases if the concerned traffic flow does not reach the destination node in next time interval. Cases 4 and 5 represent that the concerned traffic flow reaches the destination node in the next time interval $t_i^+$. Case 4 represents that the concerned traffic flow successively passes two links and reaches the destination in next time interval as shown in Figure 6(b). Case 5 represents that the concerned traffic flow starts at the last link of the path and reaches the destination in next time interval as shown in Figure 6(c).

$$\hat{T}_{rs}^{k}(t_i^-, l) = \begin{cases} t_w + \hat{T}_{rs}^{k}\left(t_{i+1}^-, t_w s_{a_j}^{t_i} + l\right) \quad \text{if} \begin{cases} l_{a_{j-1}} \le l < l_{a_j} \\ t_w s_{a_j}^{t_i} + l \le l_{a_j} \end{cases} \quad \text{(case 1)} \\ t_w + \hat{T}_{rs}^{k,t_i}\left(t_{i+1}^-, \left(t_w - \frac{l_{a_j} - l}{s_{a_j}^{t_i}}\right) s_{a_{j+1}}^{t_i} + l_{a_j}\right) \\ \quad \text{if} \begin{cases} l_{a_{j-1}} \le l < l_{a_j}, \quad t_w s_{a_j}^{t_i} + l > l_{a_j} \\ \left(t_w - \frac{l_{a_j} - l}{s_{a_j}^{t_i}}\right) s_{a_{j+1}}^{t_i} + l_{a_j} \le l_{a_{j+1}} \end{cases} \quad \text{(case 2)} \\ t_w + \hat{T}_{rs}^{k,t_i}\left(t_{i+1}^-, \left(t_w - \frac{l_{a_j} - l}{s_{a_j}^{t_i}} - \frac{\bar{l}_{a_{j+1}}}{s_{a_{j+1}}^{t_i}}\right) s_{a_{j+2}}^{t_i} + l_{a_{j+1}}\right) \\ \quad \text{if} \begin{cases} l_{a_{j-1}} \le l < l_{a_j}, \quad t_w s_{a_j}^{t_i} + l > l_{a_j} \\ l_{a_n} > \left(t_w - \frac{l_{a_j} - l}{s_{a_j}^{t_i}}\right) s_{a_{j+1}}^{t_i} + l_{a_j} > l_{a_{j+1}} \end{cases} \quad \text{(case 3)} \\ \frac{l_{a_j} - l}{s_{a_j}^{t_i}} + \frac{\bar{l}_{a_n}}{s_{a_n}^{t_i}} + \hat{T}_{rs}^{k}(t_i^-, l) \\ \quad \text{if} \begin{cases} l_{a_{j-1}} \le l < l_{a_j}, \quad t_w s_{a_j}^{t_i} + l > l_{a_j} \\ \left(t_w - \frac{l_{a_j} - l}{s_{a_j}^{t_i}}\right) s_{a_{j+1}}^{t_i} + l_{a_j} \ge l_{a_n} \end{cases} \quad \text{(case 4)} \\ \frac{l_{a_n} - l}{s_{a_n}^{t_i}} + \hat{T}_{rs}^{k}(t_i^-, l) \text{if} \begin{cases} l_{a_{n-1}} \le l < l_{a_n} \\ t_w s_{a_n}^{t_i} + l > l_{a_n} \end{cases} \quad \text{(case 5)} \end{cases}$$
$$\forall k \in \mathbf{K}_{rs}, \quad rs \in \mathbf{R}, \quad i = 1, 2, 3, \ldots \tag{10}$$

Then, the concerned path travel time $T_{rs}^{k,t_i}$ can be calculated by equation (10) as follows.

$$T_{rs}^{k,t_i} = \hat{T}_{rs}^{k}(t_i^-, 0) \tag{11}$$

### D. STAGE II: TIME-DEPENDENT OD ESTIMATION MODEL

Denote the observed time-dependent traffic counts on link $a$ during time interval $t_i$ as $\tilde{v}_a^{t_i}$. Then, the corresponding time-dependent OD demand estimation model can be formulated as the following least squares model.

$$\min_{q_{rs}^{t_i}} \left\{ \gamma \sum_{i=1}^{n_2} \sum_{a\in\tilde{\mathbf{A}}} \left(v_a^{t_i} - \tilde{v}_a^{t_i}\right)^2 + (1-\gamma) \sum_{i=1}^{n_1} \sum_{rs\in\mathbf{R}} \left(q_{rs}^{t_i} - \bar{q}_{rs}^{t_i}\right)^2 \right\} \tag{12a}$$

$$\text{s.t. } q_{rs}^{t_i} \ge 0 i = 1, 2, \ldots n_1 \tag{12b}$$

where $\gamma$ ($0 \le \gamma \le 1$) is a non-negative weighting parameter, which determines the relative importance of the first term in Equation (12a); $\bar{q}_{rs}^{t_i}$ is the historical time-dependent OD demand; $n_1$ is the number of the time intervals of estimated OD demand; $n_2$ is the number of the time intervals of observed link flows. Generally, it is required that $n_1 \le n_2$ so that the observed link flows can carry the information of the estimated OD demands.

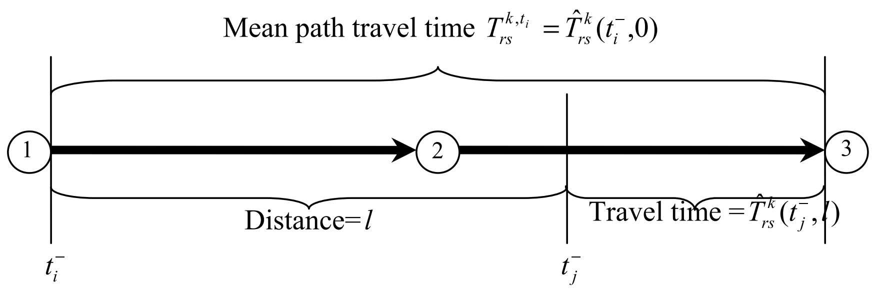


**FIGURE 5.** Illustration on the path travel time calculation.

Substitute Equation (7) into optimization problem (12) follows that

$$\min_{q_{rs}^{t_i}} \left\{ \gamma \sum_{i=1}^{n_2} \sum_{a \in \tilde{\mathbf{A}}} \left( \sum_{j=1}^{i} \sum_{rs \in \mathbf{R}} \sum_{k \in \mathbf{K}_{rs}} \delta_{rs}^{k,a}(t_j, t_i)\, p_{rs}^{k,t_i} q_{rs}^{t_i} - \tilde{v}_a^{t_i} \right)^2 + (1-\gamma) \sum_{i=1}^{n_1} \sum_{rs \in \mathbf{R}} \left( q_{rs}^{t_i} - \bar{q}_{rs}^{t_i} \right)^2 \right\} \tag{12a}$$

$$\text{s.t. } q_{rs}^{t_i} \geq 0 i = 1, 2, \ldots n_1 \tag{12b}$$

*Proposition 1:* There exists a unique solution for the optimization problem (12) if $0 \leq \gamma < 1$.

*Proof:* The Hessian matrix of the objective function is positive definite. And the feasible set of (12) is a closed and convex set. Thus, there exists a unique solution for the optimization problem (12).

It should be noted that if the historical time-dependent OD demand $\bar{q}_{rs}^{t_i}$ is unavailable (or inaccurate), the weighting parameter $\gamma$ should be set as $\gamma = 1$. Then, solution of minimization problem (12a) is equivalent to solving a system of linear equations and multiple solutions may exist for Problem (12).

## III. SOLUTION ALGORITHM

The two-stage problems can be solved separately as there is no feedback between the two stages, which is different from the conventional bi-level programming. The solution algorithm is proposed with the following steps.

*Step 1:* Initialization.

*Step 2:* Stage I: Calculate the completed travel time for each path at each time interval according to Equation (10). Then, calculate the path and link choice proportions according to Equations (8) and (7).

*Step 3:* Stage II: According to the link choice proportion calculated in Stage I, solve the constrained least squares problem (12) to obtain the time-dependent OD demands.

*Remark:* In Step 3, the constrained least squares problem belongs to Quadratic Programming model, which can be solved by many software, such as LINGO and MATLAB.

## IV. NUMERICAL EXAMPLES

The purposes of the numerical examples are to illustrate: (a) spatial effects of the accuracy of time-dependent OD demand estimation; (b) temporal effects of the accuracy of time-dependent OD demand estimation; (c) effects of accuracy of observed data. (d) efficiency of the proposed solution algorithm. A number of sensitivity tests and investigations will be carried out as follows for these purposes.

### A. EXAMPLE 1: SMALL EXAMPLE

To demonstrate the properties of the model formulation, a small tractable network is chosen for the tests. A simple network with six nodes, seven links, two O-D pairs ($1 \to 3$ and $2 \to 4$) and four paths, as shown in Figure 7, is used to illustrate the applications of the proposed model.The relevant input data for this example network are given in Tables 3, 4, 5, 6 and 7. The dispersion parameter is assumed to be $\theta = 1$. The estimated path choice proportions are shown in Table 7. In this example, the Percentage Root Mean Square Error (PRMSE) is used to demonstrate the accuracy of the estimation.

$$\text{PRMSE} = \sqrt{\frac{1}{n_1 |\mathbf{R}|} \sum_{rs \in \mathbf{R}} \sum_{i=1}^{n_1} \left( \frac{q_{rs}^{t_i} - \tilde{q}_{rs}^{t_i}}{\tilde{q}_{rs}^{t_i}} \right)^2} \tag{13}$$

It is known that the higher value the PRMSE, the lower value the accuracy of estimation. To clearly show the effects of the observed data, it is assumed that $\gamma = 1$ in the examples of ***A.1***, ***A.2*** and ***A.3*** so that the effects of historical OD demand information can be excluded.

#### 1) SPATIAL EFFECTS OF THE ACCURACY OF TIME-DEPENDENT OD DEMAND ESTIMATION

As aforementioned, there are spatial and temporal relationships between time-dependent OD demands and the traffic counts. The two kinds of relationships could affect the accuracy of the OD demand estimation. In practice, the spatial relationship leads to a class of optimization problem named counting location (or sensor location) problem. This kind of problem explicitly addresses the issue of where to put the counting sensor so as to achieve sufficient OD demand information with minimal number of sensors.

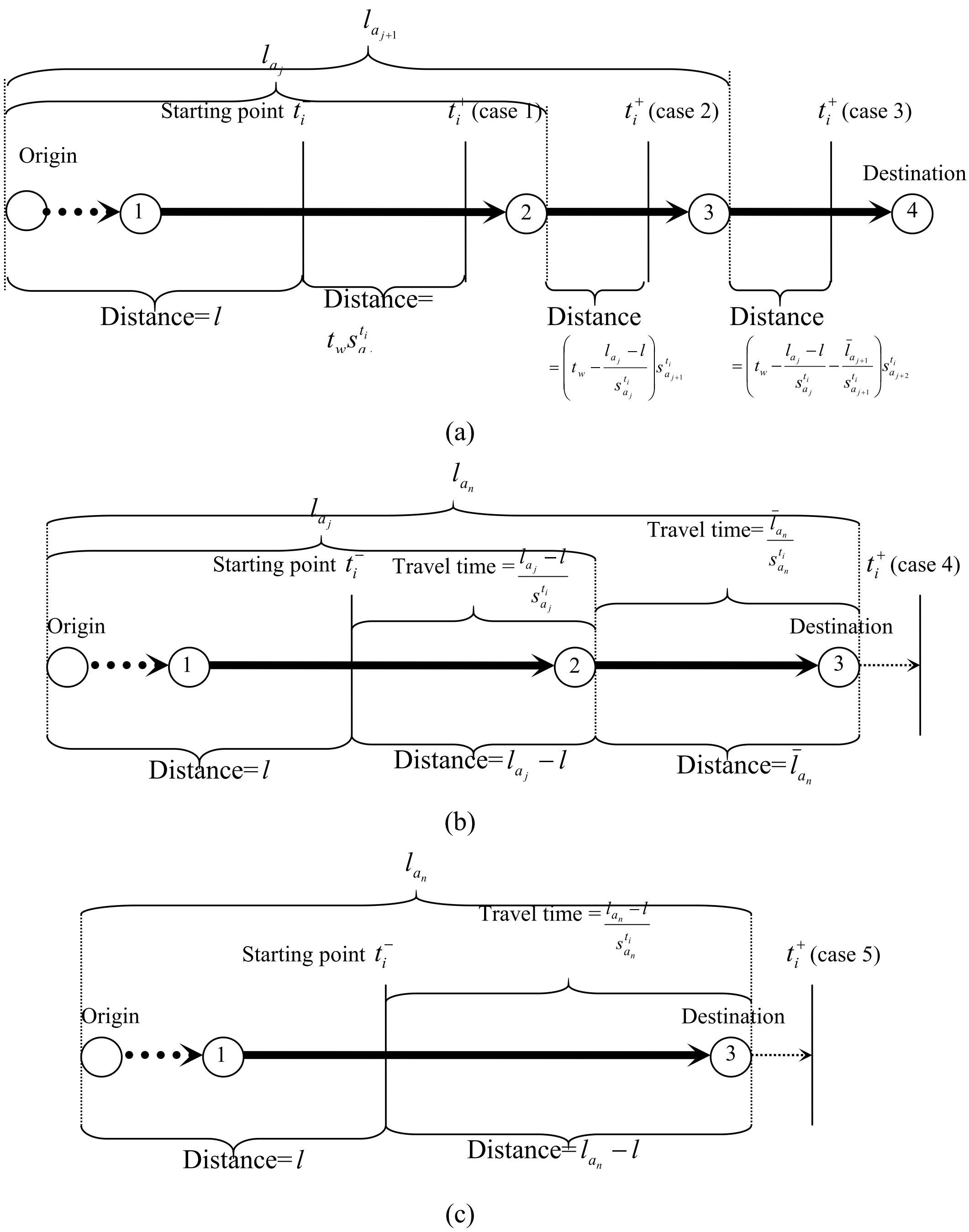


**FIGURE 6.** Illustration on the five cases for the calculation of the travel time recursive function (10).

**TABLE 3.** Network characteristics.

| | Link no. | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Distance (km) | 1.2 | 4 | 9 | 1.2 | 1.8 | 7.5 | 1.8 |
| Free flow travel time (min/km) | 1.2 | 5.0 | 6.0 | 1.2 | 1.5 | 5.0 | 1.5 |
| Capacity (pcu/min) | 25 | 50 | 20 | 25 | 15 | 20 | 15 |

Sensor location problem has achieved lots of attention in the literature. In this example, several sensitivity tests are carried out to shown the spatial effects of the accuracy of time-dependent OD estimation. To do so, 4 scenarios with different counting locations (observed links) are set as shown in Figure 8. It can be seen from Figure 8 that different

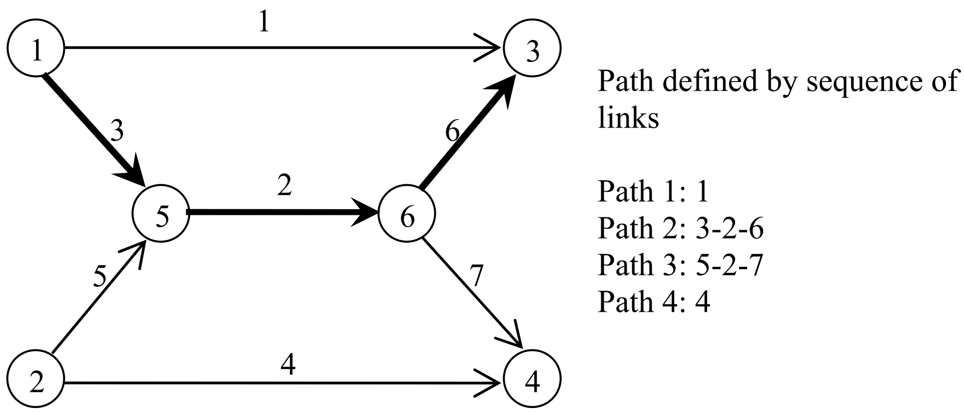


**FIGURE 7.** Illustrative example.

**TABLE 4.** Observed time-dependent link flow.

| No. | Time interval $t_i$ | Link flow (pcu /min) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $\tilde{v}_{a_1}^{t_i}$ | $\tilde{v}_{a_2}^{t_i}$ | $\tilde{v}_{a_3}^{t_i}$ | $\tilde{v}_{a_4}^{t_i}$ | $\tilde{v}_{a_5}^{t_i}$ | $\tilde{v}_{a_6}^{t_i}$ | $\tilde{v}_{a_7}^{t_i}$ |
| 1 | 8:00-8:02 | 22.8 | 1.2 | 37.2 | 23.9 | 36.1 | 0.0 | 0.0 |
| 2 | 8:02-8:04 | 29.8 | 8.4 | 35.2 | 28.5 | 36.5 | 0.0 | 0.0 |
| 3 | 8:04-8:06 | 31.8 | 9.3 | 38.2 | 28.3 | 41.7 | 0.0 | 0.0 |
| 4 | 8:06-8:08 | 35.6 | 42.6 | 39.4 | 26.0 | 49.0 | 0.0 | 19.2 |
| 5 | 8:08-8:10 | 37.0 | 47.3 | 43.0 | 31.1 | 48.9 | 0.0 | 26.6 |
| 6 | 8:10-8:12 | 0.0 | 37.1 | 0.0 | 0.0 | 0.0 | 18.4 | 30.0 |
| 7 | 8:12-8:14 | 0.0 | 38.3 | 0.0 | 0.0 | 0.0 | 17.3 | 29.9 |
| 8 | 8:14-8:16 | 0.0 | 41.8 | 0.0 | 0.0 | 0.0 | 17.7 | 36.4 |
| 9 | 8:16-8:18 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 18.2 | 0.0 |
| 10 | 8:18-8:20 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 20.1 | 0.0 |
| 11 | 8:20-8:22 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

**TABLE 5.** Time-dependent travel speed from JTIS.

| No. | Time interval $t_i$ | Mean traffic speed on links (km /min) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $s_{a_1}^{t_i}$ | $s_{a_2}^{t_i}$ | $s_{a_3}^{t_i}$ | $s_{a_4}^{t_i}$ | $s_{a_5}^{t_i}$ | $s_{a_6}^{t_i}$ | $s_{a_7}^{t_i}$ |
| 1 | 8:00-8:02 | 0.25 | 0.80 | 1.50 | 0.16 | 0.93 | 1.50 | 0.93 |
| 2 | 8:02-8:04 | 0.08 | 0.80 | 1.49 | 0.19 | 1.17 | 1.49 | 1.17 |
| 3 | 8:04-8:06 | 0.15 | 0.80 | 1.50 | 0.19 | 1.16 | 1.50 | 1.16 |
| 4 | 8:06-8:08 | 0.09 | 0.80 | 1.50 | 0.17 | 1.01 | 1.50 | 1.01 |
| 5 | 8:08-8:10 | 0.08 | 0.80 | 1.46 | 0.30 | 1.20 | 1.45 | 1.20 |
| 6 | 8:10-8:12 | 0.11 | 0.80 | 1.50 | 0.21 | 1.19 | 1.50 | 1.19 |
| 7 | 8:12-8:14 | 0.09 | 0.80 | 1.50 | 0.14 | 0.79 | 1.50 | 0.79 |
| 8 | 8:14-8:16 | 0.08 | 0.80 | 1.50 | 0.15 | 0.90 | 1.49 | 0.90 |
| 9 | 8:16-8:18 | 0.11 | 0.80 | 1.50 | 0.26 | 1.20 | 1.50 | 1.20 |
| 10 | 8:18-8:20 | 0.09 | 0.80 | 1.50 | 0.26 | 1.20 | 1.50 | 1.20 |
| 11 | 8:20-8:22 | 0.25 | 0.80 | 1.50 | 0.28 | 1.20 | 1.50 | 1.20 |

counting locations resulted in different estimation errors. For example, when two counting locations are used (say observed links 2 and 3), the estimated time-dependent OD demands are almost equal to the actual ones with PRMSE = 0%. Even if for scenarios of one counting location, the PRMSEs are different. For example, when traffic counts of link 1 is used, the PRMSE = 21.5%, whereas PRMSE = 6.4% if traffic counts of link 2 is used. This example evidenced that the counting locations have significant effects of the estimation error of time-dependent OD demands.

### 2) TEMPORAL EFFECTS OF THE ACCURACY OF TIME-DEPENDENT OD DEMAND ESTIMATION

The temporal relationship between the time-dependent OD demands and the traffic counts should also be considered when using the proposed model in practice. As discussed in

**TABLE 6.** Actual and prior time-dependent OD demands.

| No. | Time interval $t_i$ | Actual OD demand (pcu/min) | | Prior OD demand (pcu/min) | |
|---|---|---|---|---|---|
| | | $\tilde{q}_{13}^{t_i}$ | $\tilde{q}_{24}^{t_i}$ | $\bar{q}_{13}^{t_i}$ | $\bar{q}_{24}^{t_i}$ |
| 1 | 8:00-8:02 | 60 | 60 | 36 | 36 |
| 2 | 8:02-8:04 | 65 | 65 | 39 | 39 |
| 3 | 8:04-8:06 | 70 | 70 | 42 | 42 |
| 4 | 8:06-8:08 | 75 | 75 | 45 | 45 |
| 5 | 8:08-8:10 | 80 | 80 | 48 | 48 |

**TABLE 7.** Estimated time-dependent path choice proportion.

| No. | Time interval $t_i$ | Real-time mean link travel time (min) | | | |
|---|---|---|---|---|---|
| | | $p_{rs}^{1,t_i}$ | $p_{rs}^{2,t_i}$ | $p_{rs}^{3,t_i}$ | $p_{rs}^{4,t_i}$ |
| 1 | 8:00-8:02 | 0.38 | 0.62 | 0.60 | 0.40 |
| 2 | 8:02-8:04 | 0.46 | 0.54 | 0.56 | 0.44 |
| 3 | 8:04-8:06 | 0.45 | 0.55 | 0.60 | 0.40 |
| 4 | 8:06-8:08 | 0.47 | 0.53 | 0.65 | 0.35 |
| 5 | 8:08-8:10 | 0.46 | 0.54 | 0.61 | 0.39 |

**TABLE 8.** Resultant time-dependent OD demands under different values of $n_1$.

| No. | Time interval $t_i$ | Actual OD demand (pcu/min) | | Estimated OD demand (pcu/min) | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $n_2=3$ | | $n_2=5$ | | $n_2=7$ | |
| | | $q_{13}^{t_i}$ | $q_{24}^{t_i}$ | $q_{13}^{t_i}$ | $q_{24}^{t_i}$ | $q_{13}^{t_i}$ | $q_{24}^{t_i}$ | $q_{13}^{t_i}$ | $q_{24}^{t_i}$ |
| 1 | 8:00-8:02 | 60.0 | 60.0 | 86.9 | 60.0 | 62.6 | 60.0 | 60.0 | 59.4 |
| 2 | 8:02-8:04 | 65.0 | 65.0 | 53.2 | 65.0 | 67.3 | 65.0 | 60.0 | 66.6 |
| 3 | 8:04-8:06 | 70.0 | 70.0 | 93.0 | 70.0 | 52.8 | 70.0 | 70.4 | 70.0 |
| 4 | 8:06-8:08 | 75.0 | 75.0 | 96.7 | 72.3 | 76.1 | 53.1 | 70.6 | 82.4 |
| 5 | 8:08-8:10 | 80.0 | 80.0 | 99.2 | 52.7 | 66.8 | 71.9 | 80.0 | 68.4 |
| PRMSE | | | | **24.5%** | | **13.6%** | | **6.4%** | |

the model formulation, the number of time intervals for the observed link flows, $n_2$, is usually greater than $n_1$, which is the number of time intervals for estimated OD demands. Otherwise, the observed data is not sufficient enough to capture the information of requested time-dependent OD demand. This can be evidenced by the results shown in

Table 8. In this example, it is assumed that $n_1 = 5$, i.e., 5 time intervals OD demands are requested to be estimated. Traffic counts of link 2 are used in this example. It can be seen that the PRMSE of the estimated time-dependent OD demand decreases as the $n_2$ increases. That is to say that the more the temporal data used, the higher the accuracy of estimation.

### 3) EFFECTS OF ACCURACY OF OBSERVED DATA

1) In the above two examples, it is assumed that the network is under very ideal condition. Specifically, it is assumed that after 8:10 there is no traffic flow departure from the two origin nodes as shown in Table 4. Thus, the observed counts in Table 4 only carry the information of the time-dependent OD demands from 8:00 to 8:10. However, such ideal condition may not hold in reality. For example, there exist traffic flows starting from the origin nodes from 8:10 to 8:14 as shown in Table 9. In this table, the link flow without disturbance is shown in parentheses if this link flow suffers from disturbance. It can be seen that the observed traffic counts under the two conditions, i.e., with and without disturbances, are different. For instance, the disturbed traffic counts for on link 2 during time interval 8:10-8:12 is 47.9 pcu/min, while the one without disturbance is 37.1 pcu/min. In this example, it is assumed that the traffic counts in link 2 are observed with $n_1 = 5$ and $n_2 = 7$. The corresponding results are shown in Table 10. It can be seen that estimation error increases from 6.4% to 17.7% when there is disturbance in the observed data. And the resultant OD demand in the last time interval (8:00-8:10) is the most inaccurate one. This phenomenon also indicates what how to deal with the disturbances in observations in

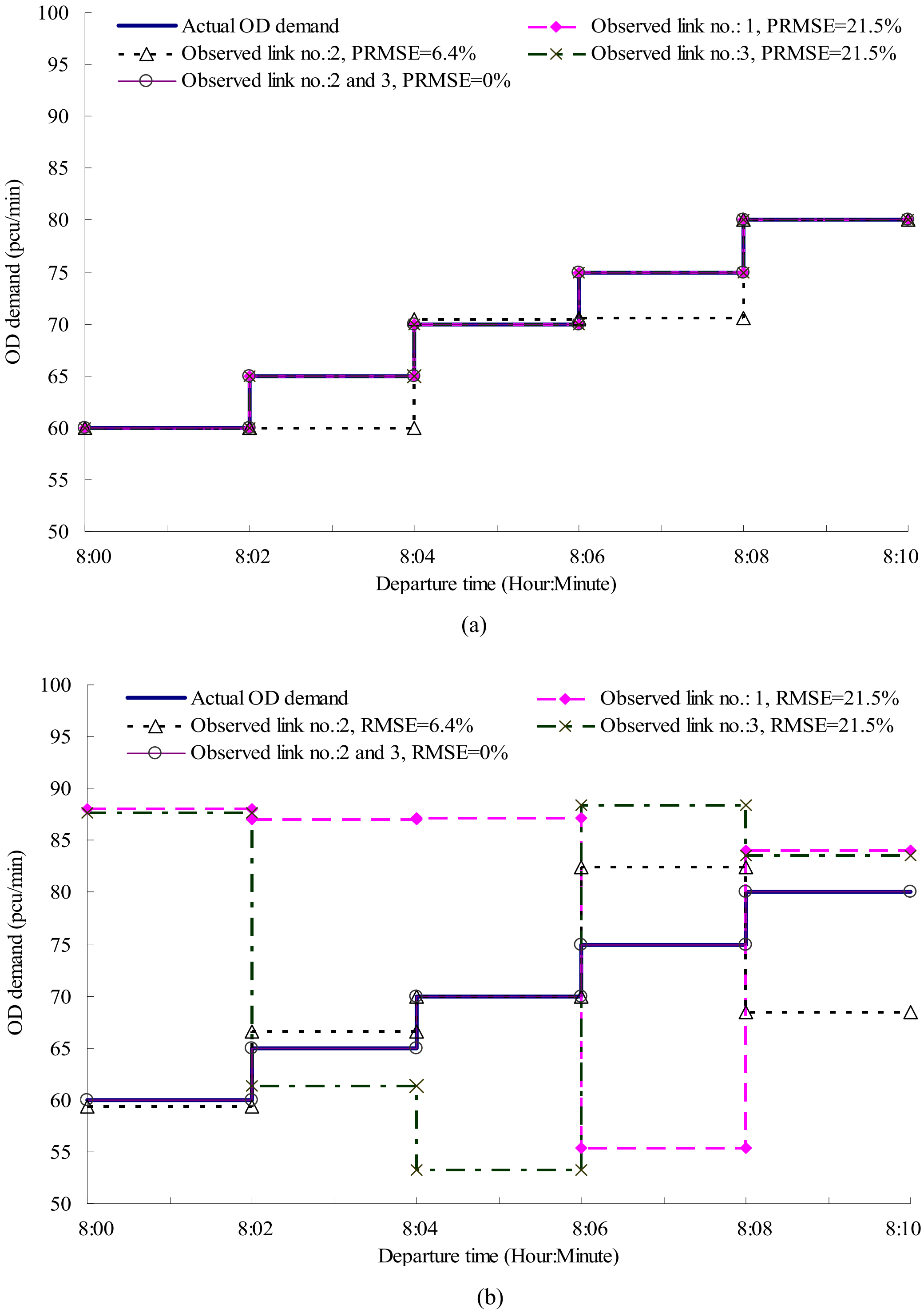


**FIGURE 8.** Resultant time-dependent OD demands of different scenarios on observed link.

an important issue for the application of the proposed model.

### B. EXAMPLE 2: MEDIUM-SIZED EXAMPLE

The well-known medium-sized Sioux Falls network as shown in Figure 9 is used to illustrate the efficiency the proposed solution algorithm. This example network consists of 24 nodes and 76 links. The dispersion parameter in SDUO is set as $\theta = 1$. The travel speed information (every two minutes) for this example network is randomly generated, which is omitted in this paper due to the length limitation. 152 paths are initially generated for the 96 OD pairs (Origin nodes: 1, 2, 4 and 5, Destination nodes: 13, 20, 21 and 24; Origin nodes: 13, 20, 21 and 24,

**TABLE 9. Observed time-dependent link flow with disturbance.**

| No. | Time interval $t_i$ | Link flow (pcu /min) | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $\tilde{v}_{a_1}^{t_i}$ | $\tilde{v}_{a_2}^{t_i}$ | $\tilde{v}_{a_3}^{t_i}$ | $\tilde{v}_{a_4}^{t_i}$ | $\tilde{v}_{a_5}^{t_i}$ | $\tilde{v}_{a_6}^{t_i}$ | $\tilde{v}_{a_7}^{t_i}$ |
| 1 | 8:00-8:02 | 22.8 | 1.2 | 37.2 | 23.9 | 36.1 | 0.0 | 0.0 |
| 2 | 8:02-8:04 | 29.8 | 8.4 | 35.2 | 28.5 | 36.5 | 0.0 | 0.0 |
| 3 | 8:04-8:06 | 31.8 | 9.3 | 38.2 | 28.3 | 41.7 | 0.0 | 0.0 |
| 4 | 8:06-8:08 | 35.6 | 42.6 | 39.4 | 26.0 | 49.0 | 0.0 | 19.2 |
| 5 | 8:08-8:10 | 37.0 | 47.3 | 43.0 | 31.1 | 48.9 | 0.0 | 26.6 |
| 6 | 8:10-8:12 | 35(0) | **47.9(37.1)** | 45(0) | 36.5(0) | 43.5(0) | 18.4 | 30.0 |
| 7 | 8:12-8:14 | 33(0) | 38.3 | 47(0) | 36(0) | 44(0) | 17.3 | 29.9 |
| 8 | 8:14-8:16 | 0.0 | 80.4(41.8) | 0.0 | 0.0 | 0.0 | 17.7 | 36.4 |
| 9 | 8:16-8:18 | 0.0 | 44.8(0) | 0.0 | 0.0 | 0.0 | 18.2 | 32.1(0) |
| 10 | 8:18-8:20 | 0.0 | 46.8(0) | 0.0 | 0.0 | 0.0 | 20.1 | 16.5(0) |

**TABLE 10. Resultant time-dependent OD demand of different scenarios with respect to the accuracy of traffic flow data (observed link no.: 2).**

| No. | Time interval $t_i$ | Actual OD demand (pcu/min) | | Estimated OD demand (pcu/min) | | | |
|---|---|---|---|---|---|---|---|
| | | | | Without disturbance | | With disturbance | |
| | | $q_{13}^{t_i}$ | $q_{24}^{t_i}$ | $q_{13}^{t_i}$ | $q_{24}^{t_i}$ | $q_{13}^{t_i}$ | $q_{24}^{t_i}$ |
| 1 | 8:00-8:02 | 60.0 | 60.0 | 60.0 | 59.4 | 59.3 | 60.0 |
| 2 | 8:02-8:04 | 65.0 | 65.0 | 60.0 | 66.6 | 73.9 | 65.0 |
| 3 | 8:04-8:06 | 70.0 | 70.0 | 70.4 | 70.0 | 90.3 | 70.0 |
| 4 | 8:06-8:08 | 75.0 | 75.0 | 70.6 | 82.4 | 75.0 | 80.9 |
| 5 | 8:08-8:10 | 80.0 | 80.0 | 80.0 | 68.4 | ***98.2*** | ***48.6*** |
| PRMSE | | | | **6.4%** | | **17.7%** | |

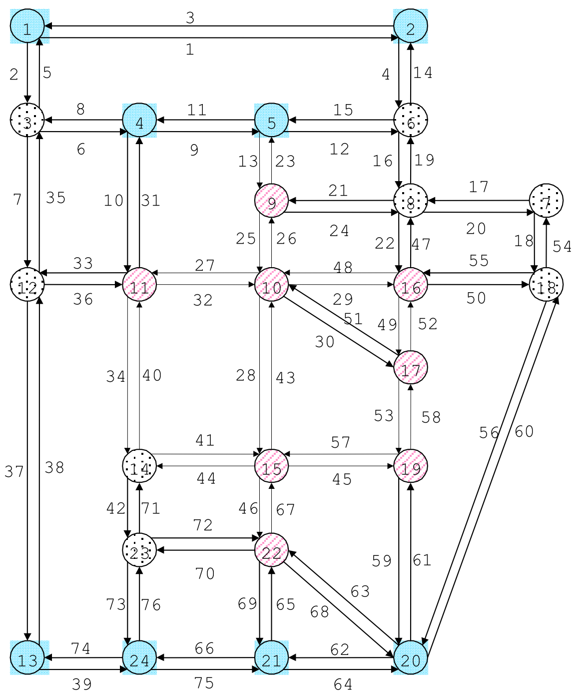


**FIGURE 9. Sioux Falls network.**

Destination nodes: 1, 2, 4 and 5; Origin nodes: 6, 7, 8 and 18, Destination nodes: 3, 12, 14 and 23; Origin nodes: 3, 12, 14 and 23, Destination nodes: 6, 7, 8 and 18;Origin nodes: 9, 10, 11 and 16, Destination nodes: 15, 17, 19 and 22; Origin nodes: 15, 17, 19 and 22, Destination nodes: 9, 10, 11 and 16). The path set is assumed to be fixed. It is assumed that the traffic counts on all 76 links can be observed.

As the number of OD pairs is greater than the number of links, the historical time-dependent OD demand information is used in this paper. Specifically, it is assumed that $\gamma = 0.5$. The historical time-dependent OD demand is set as identical to the actual time-dependent OD demand. The stopping tolerance criteria $\tau$ is set to be 0.001. The estimation errors are PRMSE $=$ 5.6%. Details of the results are omitted. The proposed model was solved using Matlab code on a laptop (Intel(R) Core(TM) i3-3120M CPU@2.50GHz $\times 2$, RAM 8G). The CPU time is about 977 seconds. It was also found that if the historical data ($\gamma = 1$) is not used for this network, multiple solutions existed. This phenomenon is reasonable. When number of OD pairs is greater than the number links, it is impossible to uniquely identify the optimal OD demand matrix. That is to say, this problem of identifiability, which is regarded as a common difficulty of OD estimation, still exists in our model. How to overcome this difficulty reveals significant investigation in the further studies. It is evidenced from this example that proposed algorithm can be applied to a medium-sized network.

## V. CONCLUSIONS AND FURTHER STUDIES

A two-stage model has been proposed in this paper for estimation of day-to-day OD demands using off-line traffic data from real-time travel information system. In stage I, a travel time recursive function has been proposed to use the travel speed data in Hong Kong JTIS for the investigation of the spatial and temporal relationships between time-dependent OD demands and traffic counts. As such, the time-consuming DTA process is not required. Thus, it has the potential for implementation in large-scale road networks. In stage II, a least-squares method is adopted to formulate the time-dependent OD demand estimation problem as a quadratic programming model. The solution existence and uniqueness conditions of the proposed model have been proved. The two-stage model can be easily solved as there is no feed back between the two stages. Numerical examples have demonstrated the spatial and temporal effects of the accuracy of time-dependent OD demand estimation. It was also found that the accuracy of observed data can influence the estimation errors of the proposed model.

On the basis of the model proposed in this paper, the following extensions can be envisaged.

As indicated in the numerical examples, spatial and temporal distributions of observed traffic counts can influence results of the model. In view of this, it could be interesting to further investigate how to optimize the traffic counting locations and how to determine observed time intervals for data collection.

- How to use the real data in a large-scale network to validate the model's results as well as conduct the compassion [47] with traditional DTA-based models reveals further investigations. Some newly developed solution algorithms for VI could be adopted to solve the DTA for high computational efficiency [48]–[50].
- In our model the accuracy of the speed information directly influences the accuracy of the time-dependent OD demand estimation. To collect accurate speed information, many detectors are required to detect various kinds of travel speeds, particularly in a congested link. This could raise another research topic on how to locate the detector so as to achieve sufficient speed information. Such topic would be an interesting extension of this paper.
- Assumption A3 may limit the application of the proposed method to certain network. In this paper, this assumption is made for mathematical convenience. Further studies could be carried out to relax this assumption so as to make the proposed model more closed to the real situation.
- In this paper, SDUO assumption for path choice is adopted. However, there are several other assumptions for dynamic path choice behaviors, such as dynamic user-optimal principle and reliability-based stochastic-dynamic-user-equilibrium [51], [52]. All the assumptions can also be applied in the proposed as an extension of the proposed model by modifying Equation (8).
- The proposed time-dependent OD demand estimation model has the potential for implementation in large-scale road networks. Then, it can be applied to path finding problem under network uncertainty and network design [53]–[63].
- It is assumed that the path set is pre-determined and fixed for all time intervals. In reality, the path set between an OD pair may be changed for different time intervals. The proposed model should be extended to relax this assumption for accounting the time-dependent path set for OD estimation problem.

## REFERENCES


[1] Y. Nie and H. M. Zhang, "A variational inequality formulation for inferring dynamic origin-destination travel demands," *Transp. Res. B, Methodol.*, vol. 42, nos. 7–8, pp. 635–662, 2008.

[2] M. J. Maher, "Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach," *Transp. Res. B, Methodol.*, vol. 17, no. 6, pp. 435–447, 1983.

[3] B. N. Janson, "Most likely origin-destination link uses from equilibrium assignment," *Transp. Res. B, Methodol.*, vol. 27, no. 5, pp. 333–350, 1993.

[4] D. P. Watling, "Maximum likelihood estimation of OD matrices from a partial registration plate number," *Transp. Res. B, Methodol.*, vol. 28, no. 4, pp. 289–314, 1994.

[5] H. P. Lo, N. Zhang, and W. H. K. Lam, "Estimation of an origin-destination matrix with random link choice proportions: A statistical approach," *Transp. Res. B, Methodol.*, vol. 30, no. 4, pp. 309–324, 1996.

[6] H. Kim, S. Baek, and Y. Lim, "Origin-destination matrices estimated with a genetic algorithm from link traffic counts," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1771, no. 1, pp. 156–163, 2001.

[7] H. Kim and R. Jayakrishnan, "The estimation of a time-dependent OD trip table with vehicle trajectory samples," *Transp. Planning Technol.*, vol. 33, no. 8, pp. 747–768, 2010.

[8] Z. Y. Liu, Q. Meng, and G. Gomes, "Estimating link travel time functions for heterogeneous traffic flows on freeways," *J. Adv. Transp.*, vol. 50, no. 8, pp. 1683–1698, 2016.

[9] X. Li, J. Kurths, C. Gao, J. Zhang, Z. Wang, and Z. Zhang, "A hybrid algorithm for estimating origin-destination flows," *IEEE Access*, vol. 6, pp. 677–687, 2018.

[10] L. Nie, Y. Li, and X. Kong, "Spatio-temporal network traffic estimation and anomaly detection based on convolutional neural network in vehicular ad-hoc Networks," *IEEE Access*, vol. 6, pp. 40168–40176, 2018.

[11] Q. Meng and H. L. Khoo, "A computational model for the probit-based dynamic stochastic user optimal traffic assignment problem," *J. Adv. Transp.*, vol. 46, no. 1, pp. 80–94, 2012.

[12] Q. Meng, W. H. K. Lam, and L. Yang, "General stochastic user equilibrium traffic assignment problem with link capacity constraints," *J. Adv. Transp.*, vol. 42, no. 4, pp. 429–465, 2008.

[13] W. Y. Szeto and H. K. Lo, "Dynamic traffic assignment: Properties and extensions," *Transportmetrica*, vol. 2, no. 1, pp. 31–52, Jan. 2006.

[14] W. Y. Szeto, "Enhanced lagged cell-transmission model for dynamic traffic assignment," *Transp. Res. Rec.*, vol. 2085, pp. 76–85, Dec. 2008.

[15] Y. Jiang, W. Y. Szeto, J. Long, and K. Han, "Multi-class dynamic traffic assignment with physical queues: Intersection-movement-based formulation and paradox," *Transportmetrica A, Transp. Sci.*, vol. 12, no. 10, pp. 878–908, 2016.

[16] E. Cascetta, D. Inaudi, and G. Marquis, "Dynamic estimators of origin-destination matrices using traffic counts," *Transp. Sci.*, vol. 27, no. 4, pp. 363–373, 1993.

[17] C.-C. Lu, X. Zhou, and K. Zhang, "Dynamic origin–destination demand flow estimation under congested traffic conditions," *Transp. Res. C, Emerg. Technol.*, vol. 34, pp. 16–37, Sep. 2013.

[18] L. Lu, Y. Xu, and C. Antoniou, "An enhanced SPSA algorithm for the calibration of dynamic traffic assignment models," *Transp. Res. C, Emerg. Technol.*, vol. 51, pp. 149–166, Feb. 2015.

[19] E. Cipriani, M. Florian, and M. Mahut, "A gradient approximation approach for adjusting temporal origin–destination matrices," *Transp. Res. C, Emerg. Technol.*, vol. 19, pp. 270–282, Apr. 2011.

[20] M. S. Iqbal, C. F. Choudhury, P. Wang, and C. M. González, "Development of origin-destination matrices using mobile phone call data," *Transp. Res. C, Emerg. Technol.*, vol. 40, no. 1, pp. 63–74, Mar. 2014.

[21] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 162–177, Sep. 2015.

[22] Q. Ge and D. Fukuda, "Updating origin–destination matrices with aggregated data of GPS traces," *Transp. Res. C, Emerg. Technol.*, vol. 59, pp. 291–312, Aug. 2016.
[23] Z. R. Huang *et al.*, "Modeling real-time human mobility based on mobile phone and transportation data fusion," *Transp. Res. C, Emerg. Technol.*, vol. 96, pp. 251–269, Nov. 2018.
[24] L. Ni, X. Wang, and X. Chen, "A spatial econometric model for travel flow analysis and real-world applications with massive mobile phone data," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 510–526, Jan. 2018.
[25] M. L. Tam and W. H. K. Lam, "Using automatic vehicle idenification data for travel time estimation in hong kong," *Transportmetrica*, vol. 4, no. 3, pp. 179–194, 2008.
[26] K. Sung, M. G. H. Bell, M. Seong, and S. Park, "Shortest paths in a network with time-dependent flow speeds," *Eur. J. Oper. Res.*, vol. 121, no. 1, pp. 32–39, 2000.
[27] B. Ran and D. Boyce, *Modeling Dynamic Transportation Networks: An Intelligent Transportation System Oriented Approach*. Berlin, Germany: Springer-Verlag, 1996.
[28] J. Liu, Z. He, and S. Ma, "An $N$-path logit-based stochastic user equilibrium model," *IEEE Access*, vol. 6, pp. 20971–20986, 2018.
[29] Y. Kang, "Estimation and prediction of dynamic origin-destination (O-D) demand and system consistency control for real-time dynamic traffic assignment operation," Ph.D. dissertation, Dept. Civil Environ. Eng., Univ. Texas Austin, Austin, TX, USA, 1999.
[30] G.-L. Chang and X. Tao, "An integrated model for estimating time-varying network origin–destination distributions," *Transp. Res. A, Policy Pract.*, vol. 33, no. 5, pp. 381–399, 1999.
[31] S. Peeta and A. K. Ziliaskopoulos, "Foundations of dynamic traffic assignment: The past, the present and the future," *Netw. Spatial Econ.*, vol. 1, nos. 3–4, pp. 233–266, 2001.
[32] W. Ma and Z. Qian, "Estimating multi-year 24/7 origin-destination demand using high-granular multi-source traffic data," *Transp. Res. C, Emerg. Technol.*, vol. 96, pp. 96–121, Nov. 2018.
[33] W. Ma and Z. Qian, "On the variance of recurrent traffic flow for statistical traffic assignment," *Transp. Res. C, Emerg. Technol.*, vol. 81, pp. 57–82, Aug. 2017.
[34] W. Ma and Z. Qian, "Statistical inference of probabilistic origin-destination demand using day-to-day traffic data," *Transp. Res. C, Emerg. Technol.*, vol. 88, pp. 227–256, Mar. 2018.
[35] C. Osorio, "Dynamic origin-destination matrix calibration for large-scale network simulators," *Transp. Res. C, Emerg. Technol.*, vol. 98, pp. 186–206, Jan. 2019.
[36] M. Cremer and H. Keller, "Dynamic identification of O-D flow from traffic counts at complex intersections," presented at the 8th Int. Symp. Transp. Traffic Theory, Toronto, ON, Canada, Jun. 1981.
[37] M. Cremer and H. Keller, "A systems dynamics approach to the estimation of entry and exit O-D flows," presented at the 9th Int. Symp. Transp. Traffic Theory, Delft, The Netherlands, Jul. 1984.
[38] M. G. H. Bell, "The real time estimation of origin-destination flows in the presence of platoon dispersion," *Transp. Res. B, Methodol.*, vol. 25, nos. 2–3, pp. 115–125, 1991.
[39] G.-L. Chang and J. Wu, "Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors," *Transp. Res. B, Methodol.*, vol. 28, no. 2, pp. 141–160, 1994.
[40] J. Wu and G.-L. Chang, "Estimation of time-varying origin-destination distributions with dynamic screenline flows," *Transp. Res. B, Methodol.*, vol. 30, no. 4, pp. 277–290, 1996.
[41] H. Tavana, "Internally-consistent estimation of dynamic network origin-destination flows from intelligent transportation systems data using bi-level optimization," Ph.D. dissertation, Dept. Civil Environ. Eng., Univ. Texas Austin, Austin, TX, USA, 2001.
[42] X. Zhou, "Dynamic origin-destination demand estimation and prediction for off-line and on-line dynamic traffic assignment operation," Ph.D. dissertation, Dept. Civil Environ. Eng., Univ. Maryland, College Park, MD, USA, 2004.
[43] X. Zhou and H. S. Mahmassani, "A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework," *Transp. Res. B, Methodol.*, vol. 41, no. 8, pp. 823–840, 2007.
[44] E. Cascetta, A. Nuzzolo, F. Russo, and A. Vitetta, "A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks," in *Proc. 13th Conf. Int. Symp. Transp. Traffic Theory*. Lyon, France: Elsevier Science, 1996, pp. 697–711.
[45] E. Cascetta, F. Russo, and A. Vitetta, "Stochastic user equilibrium assignment with explicit path enumeration: Comparison of models and algorithms," presented at the Int. Fed. Autom. Control, Transp. Syst., Chania, Greece, Jun. 1997.
[46] H. K. Lo and A. Chen, "Traffic equilibrium problem with route-specific costs: Formulation and algorithms," *Transp. Res. B, Methodol.*, vol. 34, no. 6, pp. 493–513, 2000.
[47] Z. He, L. Zheng, P. Chen, and W. Guan, "Mapping to cells: A simple method to extract traffic dynamics from probe vehicle data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 3, pp. 252–267, 2017.
[48] Q. Meng, Z. Liu, and S. Wang, "Asymmetric stochastic user equilibrium problem with elastic demand and link capacity," *Transportmetrica A, Transp. Sci.*, vol. 10, no. 4, pp. 304–326, 2014.
[49] Z. Liu, Q. Meng, and S. Wang, "Variational inequality model for cordon-based congestion pricing under side constrained stochastic user equilibrium conditions," *Transportmetrica A, Transp. Sci.*, vol. 10, no. 8, pp. 693–704, 2014.
[50] J. Long, W. Szeto, Z. Gao, H. Huang, and Q. Shi, "The nonlinear equation system approach to solving dynamic user optimal simultaneous route and departure time choice problems," *Transp. Res. B, Methodol.*, vol. 83, pp. 179–206, Jan. 2016.
[51] W. Y. Szeto, Y. Jiang, and A. Sumalee, "A cell-based model for multi-class doubly stochastic dynamic traffic assignment," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 26, no. 8, pp. 595–611, 2011.
[52] W. Y. Szeto, Y. Jiang, K. I. Wong, and M. Solayappan, "Reliability-based stochastic transit assignment with capacity constraints: Formulation and solution method," *Transp. Res. C, Emerg. Technol.*, vol. 35, pp. 286–304, Oct. 2013.
[53] Y. Jiang and W. Y. Szeto, "Time-dependent transportation network design that considers health cost," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 1, pp. 74–101, 2015.
[54] W. Y. Szeto and A. B. Wang, "Reliable network design under supply uncertainty with probabilistic guarantees," *Transportmetrica A, Transp. Sci.*, vol. 12, no. 6, pp. 504–532, 2016.
[55] W. Zhang, Z. He, W. Guan, and G. Qi, "Day-to-day rerouting modeling and analysis with absolute and relative bounded rationalities," *Transportmetrica A, Transp. Sci.*, vol. 14, no. 3, pp. 256–273, 2018.
[56] H. Shao, W. H. K. Lam, and K. S. Chan, "The problem of searching the reliable path for transportation network with uncertainty," in *Proc. 9th Conf. Hong Kong Soc. Transp. Stud.*, 2004, pp. 226–234.
[57] S. Wang, Q. Meng, and H. Yang, "Global optimization methods for the discrete network design problem," *Transp. Res. B, Methodol.*, vol. 50, pp. 42–60, Apr. 2013.
[58] Z. Y. Liu, S. Wang, B. Zhou, and Q. Cheng, "Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 58–72, Jun. 2017.
[59] Z. Zhang, Y. Wang, P. Chen, Z. He, and G. Yu, "Probe data-driven travel time forecasting for urban expressways by matching similar spatiotemporal traffic patterns," *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 476–493, Dec. 2017.
[60] N. Zhu, Y. Liu, S. Ma, and Z. He. "Mobile traffic sensor routing in dynamic transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2273–2285, Oct. 2014.
[61] Z. B. He, W. Guan, and S. F. Ma, "A traffic-condition-based route guidance strategy for a single destination road network," *Transp. Res. C, Emerg. Technol.*, vol. 32, pp. 89–102, Jul. 2013.
[62] B. Y. Chen, W. H. K. Lam, A. Sumalee, Q. Q. Li, H. Shao, and Z. Fang, "Finding reliable shortest paths in road networks under uncertainty," *Netw. Spatial Econ.*, vol. 13, no. 2, pp. 123–148, 2013.
[63] M. Xu, Q. Meng, and K. Liu, "Network user equilibrium problems for the mixed battery electric vehicles and gasoline vehicles subject to battery swapping stations and road grade constraints," *Transp. Res. B, Methodol.*, vol. 99, pp. 138–166, May 2017.


Authors' photographs and biographies not available at the time of publication.