Article:

Siripirote, T, Sumalee, A, Watling, DP orcid.org/0000-0002-6193-9121 et al. (1 more author) (2014) Updating of travel behavior parameters and estimation of vehicle trip-chain data based on plate scanning. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 18 (4). pp. 393-409. ISSN 1547-2450

https://doi.org/10.1080/15472450.2013.806761

Updating of travel behavior model parameters and estimation of vehicle trip chain based on plate scanning

Treerapot Siripirote[1], Agachai Sumalee[*,1], David P. Watling[2], Hu Shao[1,3]

[1] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University
[2]Instititute for Transport Studies, University of Leeds
[3]Department of Mathematics, China University of Mining and Technology

ABSTRACT

This paper proposes a maximum-likelihood method to update travel behavior model parameters and estimate vehicle trip chain based on plate scanning. The information from plate scanning consists of the vehicle passing time and sequence of scanned vehicles along a series of plate scanning locations (sensor locations installed on road network). The paper adopts the hierarchical travel behavior decision model, in which the upper tier is an activity pattern generation model, and the lower tier is a destination and route choice model. The activity pattern is individual profile of daily performing activities. To obtain the reliable estimation results, the sensor location schemes for predicting trip chaining are proposed. The maximum-likelihood estimation problem based on plate scanning is formulated to update model parameters. This problem is solved by the Expectation-Maximization (EM) algorithm. The model and algorithm are then tested with simulated plate scanning data in a modified Sioux-Falls network. The results illustrate the efficiency of the model and its potential for an application to large and complex network cases.

Key Words: maximum-likelihood estimation; plate scanning; EM algorithm; trip chaining

1. INTRODUCTION

The underlying assumption of the four-step model ignores the interdependent decision of trips, which may be due to time and/or space constraint, made by travelers over each day. This assumption delinks the relationship between human daily activities and travel pattern. The activity-based model (ABM) has been proposed to overcome this limitation in which the travel demand is derived from activity participation and the sequences or patterns of activity behavior (Bhat et al., 1999). The structure of ABM can be utility maximization-based (e.g. Bowman et al., 2006, 2008) or rule-based approach (e.g. Arentze and Timmermans, 2004). The utility maximization-based ABMs have been widely developed to evaluate the traffic policies in many cities such as Portland, Columbus, Atlanta, and Sacramento (Vovsha et al., 2004).

Traditionally, activity-based models are estimated from travel diary survey data (TD). These estimated results can be biased due to low-sampling size, inaccurate travel diary data, etc. For instance, considering the complex activity-travel decisions in ABM, approximately only 1% of the population was used to estimate the ABM parameters (Bowman and Bradley, 2008). In addition, the under-reporting of trips, due to response burden or uncompleted memorization of trips, leads to inaccurate travel dairy data (Bricka and Bhat, 2006). Consequently, the predicted travel demand from ABM based on TD can be inconsistent with actual roadside data (e.g. link count). In order to calibrate ABM parameters, Bowman et al. (2006) developed a comprehensive model calibration approach to calibrate the ABM parameters with travel survey data. According to his approach, some parameters in the utility of activity-travel decisions are heuristically adjusted to reproduce predicted traffic flows that fit to traffic counts. In recent years, Cools et al. (2010) also conducted ABM calibration by heuristic method. In their study, ABM parameters are adjusted by randomly weighting the chosen activity-travel patterns so as to reproduce the external trip matrix information. The heuristic-based calibrations, however, still lack statistical measures on how well the calibrated ABM parameters can reproduce the collected roadside data (e.g. link counts).

In the other related problems of traffic model calibration, the problem of estimating trip table information from link counts has been a long-running theme of transportation network analysis. The estimation of trip table can be classified by two approaches: (i) top-down approach, and (ii) bottom-up approach. For top-down approach, four-step model is generally used to reproduce internal origin-destination (OD) trips from internal traffic zones. In addition, these OD trips associated with the trips originating from external traffic zones (external OD trips) are used to calculate the trip table from this approach. Furthermore, to calibrate the trip table, trip table adjustment to fit with link counts is heuristically conducted (e.g. applying of origin-destination K-factors (Ortuzar and Willumsen, 2011)). To provide the statistical method for estimating trip tables, the observable data from the road network (e.g. link counts) can directly be utilized in the model calibration from bottom-up approach. By specifying the observable data on road network as the preferred outputs of the model, the parameters of the model (trip table) can be adjusted accordingly. For instance, link counts and prior trip table are directly used in the model to statistically estimate trip table from this approach (Bell, 1991; Cascetta, 1984, 1993; Maher, 1983; Watling, 1994; Yang et al., 1992, 1995). In the context of intelligent transportation systems (ITS), Hu et al. (2001) extended the problem of trip table estimation by using adaptive kalman filtering to estimate dynamic assignment matrices and OD demands. To consider other traffic model estimations, Lao et al. (2012) developed a Gaussian mixture model to estimate travel speeds and classified vehicle volumes using loop detectors. Also, Yuan et al. (2013) adopted the traffic flow model to predict the travel speeds using two traffic data sources (loop detector and floating car data) based on kalman filtering.

Ideally, the data used in the estimation of ABM parameters should be collected by GPS-based travel surveys (activity-travel data collected by the GPS equipment attached to probe vehicles or carried by travelers). For instance, Axhausen et al. (2003) developed an automated process to predict travelers' destinations and trip purposes from vehicle GPS traces. In addition, Frignani et al. (2010) also collected high accurate activity-travel data (e.g. chosen activity type, destination and mode) from internet GPS-based interaction travel feedback system. To identify the travel path using vehicle GPS traces, topological map-matching method is normally used. Due to error in positioning probe vehicles in the digital map, this method sometimes fails to identify the correct travel path. In recent year, Velaga et al. (2012) improved the performance of map-matching method by using error detection and correction technique. Nevertheless, the deployment of the GPS equipment in a large scale to collect activity-travel data for ABM parameter estimation may not be appropriate due to low response rate or high cost of GPS mobile devices. Alternatively, the other method, which can possibly identify activity-travel data, is plate scanning (PS). Compared with the GPS-based data collection method, the information obtained from plate scanning is similar to the GPS-based data in the context of tracking the vehicles. In contrast, plate scanning does not require the installation of the GPS equipment on the tested vehicles. Based on PS data collection, the information of these vehicles is obtained at pre-determined locations on road network. Furthermore, the process of plate scanning can identify the same vehicles traveling along a series of plate scanning locations by matching their license plate numbers. With this method of data collection, plate scanning is considered to be one of methods to collect vehicle re-identification (VI) data. The data from plate scanning consists of: (i) the vehicle passing time at plate scanning locations, and (ii) sequence of scanned vehicles along a series of plate scanning locations on road network. The accuracy of collecting the above data from plate scanning method is determined from detection rate and identification rate. The detection rate is the proportion of the number of vehicles that pass sensor locations and can be detected (i.e. vehicle is known to pass at the sensor, but the license plate number of the vehicle may or may not be identified by sensor). By considering on the detection rate from PS, most of vehicles passing sensors can generally be detected. Consequently, the detection rate is generally higher than other VI methods (e.g. tag reader system). In addition, identification rate is the proportion of detected vehicles, in which their license plate numbers can be correctly identified. Based on intelligent vehicle identification system, Ozbay and Ercelebi (2005) proposed the method of advance license plate recognitions with more than 90% of vehicle samples correctly identified.

In this study, the statistical framework for updating of travel behavior model parameters and estimating vehicle trip chain from plate scanning (PS) is proposed based on the hierarchical travel decision model. This framework was motivated by similar works done in the area of trip table estimation (i.e. particularly based on bottom-up approach). Nevertheless, due to the nature of partial observations, the selection of the route and activity/location, is not fully observed from PS. Thus, in the estimation problem, the “missing variables” are defined to represent the lack of the routes and activity/location data. The paper then defines a maximum likelihood function for this problem based on PS. This proposed model was solved by the Expected-Maximization (EM) algorithm so as to avoid the combinatorial nature of the missing variables. The remainder of the paper is organized as follows. First, some basic components of the proposed model are described in next section, including notation and model formulation. Data collection from PS is also described in the third section. The fourth section then formulates the maximum-likelihood estimation problem (MLP) from plate scanning data and solves this proposed MLP. The optimal location of PS stations (i.e. generally called “sensor locations” throughout the paper) is presented in appendix A. Section 5 then tests the proposed model and algorithm with simulated PS data on a test network. The final section concludes the paper.

2. NOTATION AND MODEL FORMULATION

Considering the approach to update travel behavior model parameters and estimate vehicle trip chain from plate scanning, the following notations are used throughout the paper:

Sets of network components

$\mathbf{O}$ = set of origin zones.

$\mathbf{D}$ = set of destination zones.

$\mathbf{L}$ = set of links, $l \in \mathbf{L}$.

$\mathbf{N}$ = set of nodes.

Measurement variables (of observed user i)

$\bar{l}_k^i$ = the link that user i is identified in order k of links to be scanned (scanned at sensor k).

$y_{k,k+1}^i$ = sensor-to-sensor travel time from sensor k to sensor $k+1$.

t_k^i = time moment of vehicle scanned at sensor k.

m_k^i = travel period of vehicle scanned at sensor k.

Sets of measurements (of observed user i)

$\Upsilon_{x_i} = \{\bar{l}_1^i, ..., \bar{l}_k^i, ..., \bar{l}_{K^x}^i\}$ set of links installed with sensors (scanned links) identifying the same vehicle (or sensor path x).

$\mathbf{Y}_i = \{y_{1,2}^i, ..., y_{k,k+1}^i, ..., y_{K^x-1,K^x}^i\}$ set of sensor-to-sensor travel times between two consecutive sensors.

$\mathbf{T}_i = \{t_1^i, ..., t_k^i, ..., t_{K^x}^i\}$ set of time moments of scanned vehicle though sensors.

$\mathbf{M}_{e'_i} = \{m^i_1, ..., m^i_k, ..., m^i_{K^x}\}$ set of timestamp periods of scanned vehicle though sensors.

Estimation variables

u_{rm} = mean in-vehicle time on route r at time period m.

s_a = mean duration of an activity a.

θ_r, θ_a = coefficient of route choice and activity choice respectively.

d^{xe}_{fh} = travel demand of trip chain h associated with activity pattern f (or activity chain (f,h)) travelling on travel period set $\mathbf{C}^{fh}_e$ scanned by sensor path x.

2.1 Network and activity choice representation

Consider a traffic network $(\mathbf{N}, \mathbf{L})$ with activity location (lo) in each traffic zone[1] where $\mathbf{N}_z$ is the set of zone centroids and $\mathbf{L}_z$ is the set of links in traffic zone z, respectively ($\mathbf{N}_z \subset \mathbf{N}$ and $\mathbf{L}_z \subset \mathbf{L}$). In addition, the activity location (lo) is assumed to virtually locate at the zone centroid where the zone centroid is the node, representative of all real activity locations in that zone ($lo \in \mathbf{N}_z$).

According to daily activity-travel participations, user i (observed vehicle i^{th}) makes a plan to perform activity pattern f. Let $\mathbf{A}_f$ denote the activity pattern consisted of an ordered set of the activities which are daily scheduled to be carried out:

$$\mathbf{A}_f = \{a_1, ..., a_q, ..., a_{Q_f}\} \quad \text{for } f \in \{1, ..., F\} \tag{2.1a}$$

where F = the total daily activity patterns.

a_q = an activity performed in sequence q of activity pattern f, $q \in \{1, ..., Q_f\}$, and Q_f is the total number of activities included in activity pattern f.

For instance, if the activity pattern (f=1) is Stay-at-home(H)-Working(W)-Stay-at-home (H), $\mathbf{A}_1 = \{H, W, H\}$.

Individuals can then select trip chain which is consisted of an ordered set of activity locations and an ordered set of paths travelled between any two adjacent activity locations. Given the list of activities in the specified activity pattern performed by individuals, $\mathbf{A}_f$, the trip chain h (the combined set of locations visited and paths travelled by trip makers starting at origin zone o) is denote as $\mathbf{LR}^o_{fh}$, which is expressed as follows.

$$\mathbf{LR}^o_{fh} = \{(lo_1, ..., lo_q, ..., lo_{Q_{fh}}), (r_{1,2}, ..., r_{q,q+1}, ..., r_{Q_{fh}-1, Q_{fh}})\} \tag{2.1b}$$

for $h \in \{1, ..., H_f\}$, $lo_1 = o$, $lo_q \in \mathbf{N}_z$, $q \in \{1, ..., Q_{fh}\}$

where lo_q = activity location q where individual performs an activity, $lo_q \in \mathbf{N}_z$.

Q_{fh} = the total number of visits at activity locations of individual who makes trip chain h of activity locations and paths associated with activity pattern f.

H_f = the total number of trip chains of activity locations and paths associated with activity pattern f.

[1] A traffic zone is a special area designed by state, which usually consists of one or more census blocks.

$r_{q,q+1}$ = path travelled from activity location lo_q to lo_{q+1}.

Note that a trip chain that begins and ends at the same activity location ($lo_1 = lo_Q$) is called a tour. A tour of trip chain begins at home is called home-based tour. In addition, individuals, who make trip chain h associated with activity pattern f, perform activity chain[2] (f,h).

2.2 Assumptions

The general assumptions for this study are:

(i) Travel behaviour model based on utility maximization-based approach (e.g. Bifulco et al., 2010; Bradley et al., 2010; Bowman et al., 2006, 2008; Vovsha et al., 2004) is adopted. In particular, in this study, a group of people who makes at least one out-of-home activity and drives alone without changing to other modes of transport is considered. Consequently, neglecting mode choice consideration, activity-travel participations of this group of people can be distinguished into two tiers (higher tier: activity pattern and lower tier: trip chain (route and activity location).

(ii) The set of feasible activity patterns and trip chains of individual travelers is assumed to be given. In addition, the travel pattern of feasible trip chains is assumed to be home-based tour. These assumptions have also been adopted by the related studies (e.g. Li et al., 2010; Maruyama and Sumalee, 2007).

(iii) The daily activity-travel schedules of trip-makers involve the decisions of activity pattern and trip chain. Trip-makers base their decisions about activity and travel schedules on a tradeoff between the utility or benefits derived from activity participation at different locations and the disutility incurred by travel between activity locations. Here, we assume that all individuals are utility-maximizing decision makers, that is, they schedule their activity patterns/trip chains or activity chains[2] to maximize their perceived trip utility (e.g. Feil et al., 2009; Flotterod et al., 2011; Li et al., 2010).

(iv) The utility gained from activity participation depends on the start time of the activity and its duration. In contrast, the disutility of travel between activity locations depends on the in-vehicle time (e.g. Flotterod et al., 2011; Li et al., 2010; Feil et al., 2009). In this study, the utility of activity and travel is assumed to be a linear function with respect to the in-vehicle time and activity duration.

(v) The in-vehicle time and duration of an activity follow a probability distribution parameterized by its mean and variance (independent normal distribution type).

(vi) For disaggregated travel behavior model (e.g. activity-based model), the utility of activity pattern at upper tier is generally derived from socio-demographics (e.g. household and personal characteristics) and logsum utility of trip chain at lower tier (Bowman et al., 2006 ,2008). However, in this study, we focus on the method to predict the demand of vehicle trip chain from plate scanning at short-term operation analysis, which can be varied from day to day. Socio-demographics may not describe on these variations of activity pattern demand. In contrast, these demand variations are strongly related to the parameters (in utility function at lower tier) predicting trip-chain demand. Consequently, we assume that the utility of activity pattern is solely derived from the logsum utility of trip chain at lower tier (see (2.10)).

2.3 Model formulation

Supply side

---
[2] Activity chain is the combined decision of travelers on activity pattern and trip chain.

Consider a traffic network $(\mathbf{N},\mathbf{L})$ where $\mathbf{N}$ the set of nodes and $\mathbf{L}$ is and the set of links in section 2.1, there are two types of network elements that are defined as follows.

Vehicle path. A vehicle path represents a run of vehicles on the links between two locations. Each vehicle path contains information on in-vehicle time. The in-vehicle time on any path r during time period m, denoted as b_{rm}, is equal to the difference between the departure time of the vehicle from and its arrival time at that path, i.e. $b_{rm} = \vartheta_{rm} - \nu_{rm}$, where ν_{rm} and ϑ_{rm} is the time at which the vehicle begins and ends its journey on path r during time period m, respectively.

Activity link. An activity link represents a place where an individual performs a certain activity. Individuals gain utility or benefit from participation in an activity that is dependent on the activity start time and duration. The duration of activity a, denoted as b_a on activity link, is equal to the difference between the times that travelers enter to and leave from that link, i.e. $b_a = \omega_a - \tau_a$, where τ_a is the enter time of traveler at activity link (or the activity start time) and ω_a is the departure time of traveler from activity link (or the activity end time).

In addition, the time windows of travel, $[\nu_{rm}, \vartheta_{rm}]$, can be defined by travel periods, which can also be determined as a function of departure time and enter time of vehicle path, where time period of ν_{rm} and ϑ_{rm}, $\delta(\nu,\vartheta) = \mathrm{m}$; $\pi_m \le \nu_{rm}, \vartheta_{rm} \le \varpi_m$ for time period $m \in \{1,\dots,M\}$ (M is the number of travel periods), and π_m, ϖ_m is the start and end time of time period m, respectively. Thus, we can also define the set (index e) of travel periods of individuals travelling between any two adjacent activity locations of activity chain (f,h), $\mathbf{C}_e^{fh}$, as follows.

$$\mathbf{C}_e^{fh} = \{m_{1,2},\dots,m_{q,q+1},\dots,m_{Q_{fh}-1,Q_{fh}}\}, \quad e \in \{1,\dots,E\} \tag{2.2}$$

where E is the total number of travel period sets; $m_{q,q+1}$ is travel period of individual traveling on path $r_{q,q+1}$ of activity chain (f,h).

Furthermore, if user i (observed vehicle i) travels on vehicle path (or simply called path) during travel period set $\mathbf{C}_e^{fh}$, vehicle path set including travel period information can be defined as follows.

$$\mathbf{RM}_{fh}^{e} = \{(r_{1,2}, m_{1,2}),\dots,(r_{q,q+1}, m_{q,q+1}),\dots,(r_{Q_{fh}-1,Q_{fh}}, m_{Q_{fh}-1,Q_{fh}})\} \tag{2.3}$$

According to assumption (v), we assume that both durations of performing an activity a at location q and traveling on path $r_{q,q+1}$ during travel period m follow normal distribution, activity duration, b_a, and in-vehicle time, b_{rm}, can then be formulated as follows.

$$b_a \sim N\left(S_a^q, \left(\sigma_a^q\right)^2\right) \qquad \text{for } a \in \mathbf{A}_f \tag{2.4}$$

$$b_{rm} \sim N\left(u_{rm}^q, \left(\sigma_{rm}^q\right)^2\right) \qquad \text{for } rm \in \mathbf{RM}_{fh}^{e} \tag{2.5}$$

where S_a^q, $\left(\sigma_a^q\right)^2$ is the mean and variance of duration of an activity a at location q; u_{rm}^q, $\left(\sigma_{rm}^q\right)^2$ is the mean and variance of in-vehicle time on path $r_{q,q+1}$ during travel period m.

According to assumption (iv), the utility of activity participation and disutility of travel is represented by the utility of trip chain h associated with activity pattern f during travel period set $\mathbf{C}_{e}^{fh}$,V_{fh}^{e} , which can be written as follows.

$$V_{fh}^{e} = \sum_{a \in \mathbf{A}_f} V_{perf}(a) + \sum_{rm \in \mathbf{RM}_{fh}^{e}} V_{travel}(r,m) \tag{2.6}$$

where $V(a)_{perf}$ is measured utility of performing activity a, and $V_{travel}(r,m)$ is measured disutility of traveling on route r during travel period m, which essentially have the following forms:

$$V_{perf}(a) = \theta_a \cdot S_a^q \qquad \text{for } a \in \mathbf{A}_f \tag{2.7}$$

$$V_{travel}(r,m) = -\theta_r \cdot u_{rm}^q \qquad \text{for } rm \in \mathbf{RM}_{fh}^{e} \tag{2.8}$$

where θ_a ,θ_r is the coefficient of an activity choice and route choice, respectively ($\theta_a \geq 0$, $\theta_r \geq 0$, $\theta_a, \theta_r \in \boldsymbol{\theta}$).

In addition, the coefficient of activity choice, θ_a , representing the marginal utility during performing an activity a at its duration depends on two factors: (i) type of activity pattern, f, and (ii) activity start time, τ , (e.g. Bifulco et al., 2010; Bowman et al., 2008). Consequently, θ_a can be classified by the information of τ and f as $\theta_a(\tau, f)$ or a simplified form of $\theta_a(m, f)$ for time period of τ , $\delta(\tau) = m$; $\pi_m \leq \tau \leq \varpi_m$ where m is travel period, and π_m , ϖ_m is the start and end time of time period m, respectively.

Note that the maximum likelihood estimation problem and solution algorithm for updating in-vehicle time/activity duration and coefficients of utility of performing activity and traveling based on utility function (2.6), (2.7), and (2.8) are presented in section 4.

Demand side

The purpose of the demand model is to explain the traveler's choice of activity patterns and trip chains in terms of the attributes of the travelers and of the available activity-travel pattern alternatives. The basic choice alternative which is being modeled is typical daily activity patterns and trip chains (or activity chains) of travelers. It is assumed that the choice sets of activity-travel patterns conducted by travelers are given (assumption (ii)).

According to assumption (i), the structure of the demand model is nested logit. There are, of course, other mathematical forms, which associate choice probabilities with attributes of the alternatives (e.g. structural equation and probit model). The nested logit model, however, has the advantage of representing reasonable hypotheses about choice behavior with correlation in the nest (level of decisions) while remaining tractable for empirical estimation (Ortuzar and Williams, 2011). The model consists of two levels of decisions. The higher tier model is an activity pattern demand generation (2.1a), and the lower tier represents a trip chain (activity location/route choice (2.1b)) model. According to assumption (vi), the utility of activity pattern in this study is assumed to be the logsum utility of trip chains belonging to specified activity pattern at lower tiers. The probability that individuals select trip chain h associated with activity pattern f is then expressed by:

$$P_{fh} = P_f \cdot P_{h/f} = \frac{\exp\left(\psi_1 V_f^*\right)\exp\left(\psi_2 V_{fh}\right)}{\sum_{b=1}^{F}\exp\left(\psi_1 V_b^*\right)\sum_{d=1}^{H_f}\exp\left(\psi_2 V_{fd}\right)} \tag{2.9}$$

where ψ_1 , ψ_2 = scale parameters, b = index of activity patterns.

d = index of trip chain associated with activity pattern f .

V_f^* = logsum utility of trip chains belonging to activity pattern f .

V_{fh} = utility of trip chain h associated with activity pattern f at lower tier (see (2.6), (2.7), and (2.8) for utility formulation).

$$V_f^* = \left(\frac{1}{\psi_2}\right)\ln\sum_{d=1}^{H_f}\exp\left(\psi_2 V_{fd}\right) \tag{2.10}$$

Limitations of demand model

The demand model as developed here incorporates several behavioral assumptions. The most fundamental assumption is that travel arises from a traveler choice process in which the alternatives that are considered are complete daily activity-travel participations. These activity-travel participations are constructed by two hierarchical levels of decisions (activity pattern and trip chain). In addition, the utility of these participations is simply derived from related activity-travel attributes (in-vehicle times and activity durations). There are, however, some limitations of such a demand model:

- As the activity-travel interactions between members in household are not captured in this model, the trips in trip chains made by individuals are independent to other travelers.
- Regarding to assumption (i), a single group of trip makers is considered. In practice, the preferences of activity pattern or trip chain (or perceived utility) can be different from one group to others (taste variations). These preferences are normally dependent on individual or household demographics, which are not included in the utility function. This utility function is simply assumed to be identically perceived by the members of the same behavioral homogeneous group (e.g. Li et al., 2010). Consequently, model coefficients (θ) are derived from the utility function only responding to the particular group.
- Regarding to assumption (ii), the trip makers have the fixed plan on conducting all activities in the daily activity patterns without rescheduling such a plan. This assumption would be unsuitable when non-scheduled activities are the major concerns of the analysis. Nevertheless, the advantage of this approach is that the model estimation results are consistent and efficiently used to test the new transportation policy applications.

The demand model, of which its limitations are described above, is used for a general framework to update the demand model parameters and estimate vehicle trip chains from plate scanning. The extensions of this framework such as the complex demand model representing activity-travel interaction of members in household or other groups of travelers can be adopted for further studies.

Data needs for model estimation

In general, the model estimation is constructed by using travelers' attributes (in-vehicle times and activity durations) and chosen activity patterns and trip chains of travelers. However, these attributes and chosen activity patterns/trip chains are not directly observed by plate scanning. To obtain such information, the method is presented in section 4 and 5.

3. DATA COLLECTION FROM PLATE SCANNING

For network $(\mathbf{N},\mathbf{L})$, let $\bar{l}$ denote the link installed by a sensor scanning a vehicle license plate (simply called scanned link $\bar{l}$). When user i makes an activity pattern f associated with trip chain h traveling on the network, user i has been scanned by a series of sensors locating on the scanned links. Path traveled by user i can be represented by an ordered set of links that this user orderly passes. Compared to travel path, sensor path is the path or partial path of travelers represented by the ordered set of scanned links identifying the same vehicle in the sequence order using license plate matching technique. Observed sensor path x of user i, Υ_{x_i}, can be defined as follows.

$$\Upsilon_{x_i} = \{\bar{l}_1^{\,i},...,\bar{l}_k^{\,i},...,\bar{l}_{K^x}^{\,i}\} \text{ for } x \in \{1,..., X\} \tag{3.1}$$

where $\bar{l}_k^{\,i}$ = the scanned link that user i is identified in order k of links to be scanned.

K^x = the total number of times of links to be scanned in sensor path x.

X = the total number of sensor paths.

Other information that we can collect from plate scanning is the time moment of user i pass through scanned link $\bar{l}_k^{\,i}$. The ordered set of observed time moment of user i is defined by:

$$\mathbf{T}_i = \{t_1^i,...,t_k^i,...,t_{K^x}^i\} \quad \text{for } k = 1,2,..., K^x. \tag{3.2}$$

where t_k^i = the time moment that user i passes the sensor k on scanned link $\bar{l}_k^{\,i}$.

In addition, the different of time moments between any two adjacent scanned links, which are registered by user i, can be represented by the observed sensor-to-sensor travel time. The ordered set of observed sensor-to-sensor travel time , $\mathbf{Y}_i$, of user i can be described as follows.

$$\mathbf{Y}_i = \{y_{1,2}^i,...,y_{k,k+1}^i,...,y_{K^x-1,K^x}^i\},\ y_{k,k+1}^i > 0 \text{ for } k = 1,2,..., K^x\text{-1}. \tag{3.3}$$

where $y_{k,k+1}^i$ = the sensor-to-sensor travel time from scanned link $\bar{l}_k^{\,i}$ (sensor k) to $\bar{l}_{k+1}^{\,i}$ (sensor k+1); $y_{k,k+1}^i = t_{k+1}^i - t_k^i$.

According to the model resolution, which is usually represented in the unit of time period of a day, we can directly convert the ordered the set of observed time moment of user i, $\mathbf{T}_i$, to the ordered set (index e') of timestamp periods (time periods of a day) as follows.

$$\mathbf{M}_{e_i'} = \delta(T_i) = \{\delta(t_1^i),...,\delta(t_k^i),...,\delta(t_{K^x}^i)\} \text{ for } k = 1,2,..., K^x; \delta(t_k^i) = m_k^i\ ;\pi_m \le t_k^i \le \varpi_m \tag{3.4}$$

where m_k^i = timestamp period of user i passing sensor k on scanned link $\bar{l}_k^{\,i}$.

π_m, ϖ_m = the start and end time of time period m, respectively.

To illustrate data collection from plate scanning, a simple network with 9 nodes (including two zone centroids) and 12 links is defined in Figure 1(a). User 1 and user 2 perform trip chain H1-W2-H1 on path 1 and path 2, respectively (Figure 1(b)). According to data observed from plate scanning, the partial paths of user 1 and user 2 are observed by sensor path 1, ($\Upsilon_1 = \{1,2,5,12\}$), and sensor path 2, ($\Upsilon_2 = \{1,7,12,6\}$), respectively. Furthermore, as user 1 and user 2 passed these scanned links as Υ_1 and Υ_2 (Figure 1(c)), the ordered set of observed time moments at scanned links, $\mathbf{T}_{i=1}$ = {8:00,8:15,8:28,17:30} and $\mathbf{T}_{i=2}$ = {8:03,8:20, 17:30,17:47}, could be recorded. Given two possible travel periods, period 1 = [6:30-9:30] and period 2 = [16:30-19:30], the ordered set of timestamp period of user 1 and user 2 ($\mathbf{M}_{i=1}$ = {1,1,1,2} and $\mathbf{M}_{i=2}$ = {1,1,2,2}) could be calculated from ordered set of time moment, $\mathbf{T}$. In addition, the observed sensor-to-sensor travel times of both users (in min.), which are calculated from the difference of time moments at two consecutive sensors, are equal to $\mathbf{Y}_{i=1}$ = {15,13,542} and $\mathbf{Y}_{i=2}$ = {17,550,17}. Note that the values of the sensor-to-sensor travel times from scanned link 5 to scanned link 12 of user 1 (542 min.) and from scanned link 7 to scanned link 12 of user 2 (550 min.) are high, because these values include duration of work at zone 2.

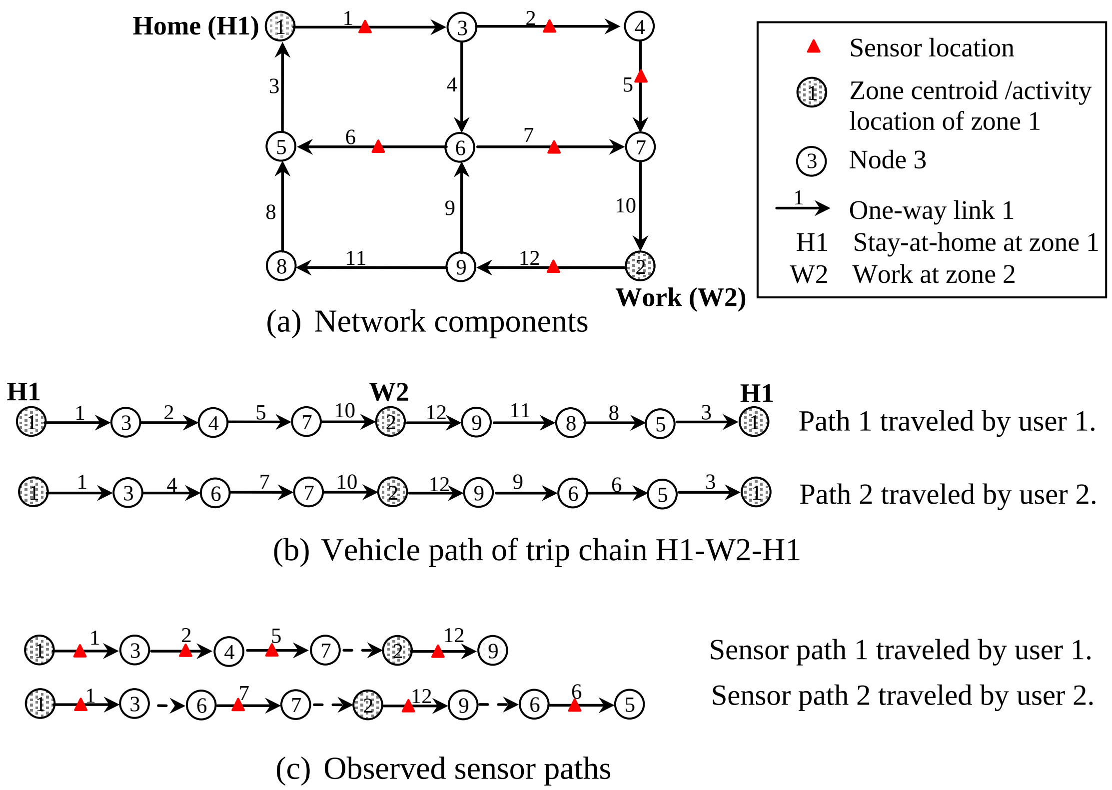


Figure 1 Network description and sensor path.

4. UPDATING OF TRAVEL BEHAVIOR MODEL PARAMETERS AND ESTIMATION OF VEHICLE TRIP CHAIN BASED ON PLATE SCANNING

4.1 Maximum likelihood estimation problem

According to assumption (ii), the choice sets of activity chain (f,h) of travelers are given (i.e. $f \in \{1,\ldots,F^x\}$ and $h \in \{1,\ldots,H_f^x\}$), where F^x is the total number of feasible activity patterns (2.1a) derived from sensor path x, and H_f^x is the total number of feasible trip chains (2.1b) associated with activity pattern f derived from sensor path x.

For each feasible choice of activity chains, time of the day of travelers can be represented by the set of travel periods daily travelling between any two adjacent activity locations. In addition, travel period of travelers ($\mathbf{C}_e^{fh}$ in (2.2)), can be directly distinguished from observed

$\mathbf{M}_{e'}$ in (3.4), where the sensors are specifically located on the links as the proposed sensor location schemes in appendix A.

Consider user i traveling on sensor path x during timestamp period set $\mathbf{M}_{e'}$ (3.4), if the chosen activity chain (f,h) during travel period set $\mathbf{C}_e^{fh}$ (2.2) can directly be observed from plate scanning, then the indicator variables can be observed as follows:

$$z_{iefh}^{x} = \begin{cases} 1, & \text{if user i, is sequentially detected by sensor } 1,..,k,...,K^{x} \text{ and matched with} \\ & \text{sensor path x, and selects activity chain } (f,h) \text{ during travel period set } C_e^{fh}. \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

In order to obtain the updated model parameters, rewritten in a vector form, $\Omega = \{\boldsymbol{\theta}, \psi, \mathbf{u}, \mathbf{S}\}$, the conventional maximum-likelihood of all users for nested logit choice model (2.9) is formulated as follows:

$$q\ (\Omega|\mathbf{z}) = \prod_{x=1}^{X}\prod_{e=1}^{E^x}\prod_{i=1}^{n_{xe}}\prod_{f=1}^{F^x}\prod_{h=1}^{H_f^x}\left(P_{fh}^{xe}(\Omega)\right)^{z_{iefh}^{x}} \tag{4.2}$$

where E^x = the total number of travel period sets (2.2) derived from sensor path x.

n_{xe} = the number of vehicles observed by sensor path x during travel period set $\mathbf{C}_e^{fh}$.

P_{fh}^{xe} = probability of user i selecting activity chain (f,h) on travel period set $\mathbf{C}_e^{fh}$ by (2.9), which the feasible choices of activity chains are derived from sensor path x.

However, the chosen activity chain (f,h) of user i and the model's attributes (i.e. mean in-vehicle time, $u_{rm}^{q} \in \mathbf{u}$ in (2.8) and mean activity duration, $S_a^q \in \mathbf{S}$ in (2.7)) are not directly observed from plate scanning (i.e. $z_{iefh}^{x}, u_{rm}^{q},$ and S_a^q are not directly observed).

According to assumption (v), we assume that both in-vehicle time $(u_{rm}^{q}, (\sigma_{rm}^{q})^2)$ and duration of activity $(S_a^q, (\sigma_a^q)^2)$ are distributed normally and independently. As a result, for any travelers possibly performing activity chain (f,h) during travel period set $\mathbf{C}_e^{fh}$ and identified by sensor k and k+1 in the network, the sensor-to-sensor travel time between sensor k and k+1, which is an aggregated form of u_{rm}^{q} and S_a^q, also follows the normal distribution shown in (4.3), (4.4), and (4.5) below.

$$y_{fh}^{ke} \sim N\left(E(y_{fh}^{ke}), (\sigma_{fh}^{ke})^2\right) \tag{4.3}$$

$$E(y_{fh}^{ke}) = u_{rm'}^{k} + S_{a'}^{k} \qquad \text{for } a' \in \mathbf{A}_f,\ rm' \in \mathbf{RM}_{fh}^{e} \tag{4.4}$$

$$(\sigma_{fh}^{ke})^2 = (\sigma_{rm'}^{k})^2 + (\sigma_{a'}^{k})^2 \quad \text{for } a' \in \mathbf{A}_f,\ rm' \in \mathbf{RM}_{fh}^{e} \tag{4.5}$$

where y_{fh}^{ke} = random sensor-to-sensor travel time between sensor k to k+1 of travelers performing activity chain (f,h) during travel period set $\mathbf{C}_e^{fh}$.

$E(y_{fh}^{ke}), \sigma_{fh}^{ke}$ = mean and standard deviation of y_{fh}^{xke}, $y_{fh}^{xke} \in \mathbf{Y}$ and $\sigma_{fh}^{ke} \in \boldsymbol{\sigma}$.

$u^k_{rm'}, \sigma^k_{rm'}$ = mean and standard deviation of in-vehicle time between sensor k to k+1 of travelers performing activity chain (f,h).

$S^k_{a'}, \sigma^k_{a'}$ = mean and standard deviation of total activity durations of travelers performing activity chain (f,h) between sensor k to k+1.

According to travelers performing activity chain (f,h), if there is no activity performed between sensor k to k+1, sensor-to-sensor travel time will be equal to the in-vehicle time. Thus, $S^k_{a'}$ and $\sigma^k_{a'}$ will be excluded from (4.4) and (4.5).

To estimate zone-to-zone in-vehicle time, network example in Figure 2 is used to illustrate the relations between observed sensor-to-sensor in-vehicle time and modeled zone-to-zone in-vehicle time. The expected travel times from activity location (zone centroid) q to q+1 can then be presented in a form of mean sensor-to-sensor travel time from any sensor k to k+1, $u^k_{rm'}$, as follows.

$$u^k_{rm'} = \gamma^{k,q} u^q_{rm} + (1-\gamma^{k,q}) u^{q+1}_{rm} \quad \text{for } k \in \{1,..,K^x - 1\},\ q \in \{1,...,Q_{fh}\} \tag{4.6}$$

where u^q_{rm} = mean in-vehicle time of travelers from location (zone centroid) q to q+1, $u^q_{rm} \in \mathbf{u}$.

$\gamma^{k,q}$ = ratio of in-vehicle time (from location (or zone centroid) q to sensor k+1) to in-vehicle time (from sensor k to sensor k+1) (i.e. $0 \le \gamma \le 1$).

After replacing $u^k_{rm'}$ in (4.4) by the definition in (4.6), the mean sensor-to-sensor travel time in (4.4) can then be formulated in a form of zone-to-zone in-vehicle time as follows.

$$E(y^{ke}_{fh}) = \gamma^{k,q} u^q_{rm} + (1-\gamma^{k,q}) u^{q+1}_{rm} + S^k_{a'} \tag{4.7}$$

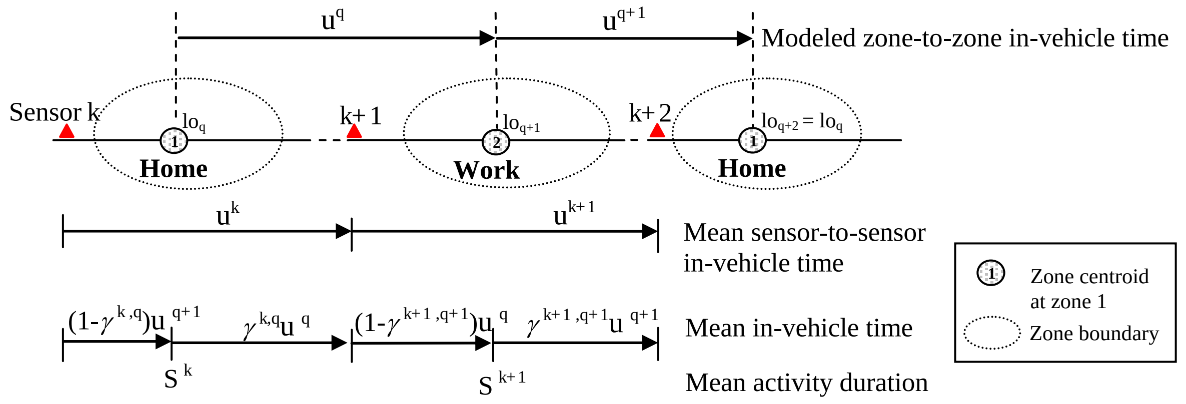


FIGURE 2 Representation of in-vehicle time and activity duration (chain H1-W2-H1).

Since trips are assumed to be started or ended at zone centroids, the distribution of actual activity locations in traffic zone also leads to significant travel time variation. For instance, if activity locations are dispersed in large traffic zone, in-vehicle time variation from location q to q+1, $(\sigma^q_{rm})^2$ tends to be high. However, the information of actual locations, where the trips start and end, is not directly observed from plate scanning. Consequently, in-vehicle time variation between two actual activity locations is not explicitly known. Nevertheless, the variance of zone-to-zone in-vehicle times can be presented in a form of variance of in-vehicle time from any sensor k to k+1, $(\sigma^{k}_{rm'})^2$ shown in (4.8a).

$$(\sigma^{k}_{rm'})^2 = \gamma^{k,q}(\sigma^{q}_{rm})^2 + (1-\gamma^{k,q})\cdot(\sigma^{q+1}_{rm})^2 \tag{4.8a}$$

where $(\sigma^{q}_{rm'})^2$ is variance of in-vehicle time of travelers between two centroids (from lo_q to lo_{q+1}).

After replacing $(\sigma^{k}_{rm'})^2$ in (4.5) by the definition in (4.8a), the sensor-to-sensor travel time variation (4.5) can then be rearranged as follows.

$$(\sigma^{ke}_{fh})^2 = \gamma^{k,q}(\sigma^{q}_{rm})^2 + (1-\gamma^{k,q})\cdot(\sigma^{q+1}_{rm})^2 + (\sigma^{k}_{a'})^2 \tag{4.8b}$$

If the actual decisions of any travelers (z^{x}_{iefh} =1 or 0 for $i = 1,\ldots,n_{xe}$) can be known, the mean and variance of sensor-to-sensor travel time (sensor k to k+1) of activity chain (f,h) can be estimated by solving the first order derivation from maximum likelihood problem (4.2) as follows (Ben-Akiva and Lerman, 1985).

$$E(y^{ke}_{fh}) = \frac{\sum_{x=1}^{X}\sum_{i=1}^{n_{xe}} y^{xie}_{k,k+1}\cdot z^{x}_{iefh}}{\sum_{x=1}^{X}\sum_{i=1}^{n_{xe}} z^{x}_{iefh}} \quad \text{and} \quad \left(\sigma^{ke}_{fh}\right)^2 = \frac{\sum_{x=1}^{X}\sum_{i=1}^{n_{xe}} z^{x}_{iefh}\left(y^{xie}_{k,k+1} - E(y^{ke}_{fh})\right)^2}{\sum_{x=1}^{X}\sum_{i=1}^{n_{xe}} z^{x}_{iefh}} \tag{4.9}$$

where $y^{xie}_{k,k+1}$ is sensor-to-sensor travel time from sensor k to k+1on sensor path x observed from user i traveling during travel period set $\mathbf{C}^{fh}_{e}$. In addition, the mean in-vehicle time, $u^{k}_{rm'}$ is equal to $E(y^{ke}_{fh})$ for the case that travelers does not perform any activity. On the other hand, the mean activity duration, $S^{k}_{a'}$, is equal to $E(y^{ke}_{fh})-u^{k}_{rm'}$ for the case that travelers perform only an activity in between sensor k and k+1.

In practice, there could be more than one feasible activity chain of individuals derived. Consequently, the value of z^{x}_{iefh} can not be directly observed from plate scanning. To deal with this problem, indicator variable z is basically treated as a missing variable and solved by the method in section 4.2.

To update model parameter, rewritten in a vector form $\Lambda = \{\boldsymbol{\theta}, \boldsymbol{u}, \mathrm{S}, \boldsymbol{\sigma}, \psi\}$, based on maximum-likelihood estimation approach, a joint probability mass function (2.9) and probability density function (normal distribution) for the random variables (y^{ke}_{fh}) in (4.3) assumed to underlying our observations (**y**) in (3.3), under the case that **z** can be estimated, can be written down as follows:

$$q\ (\Lambda\,|\,\mathbf{y},\mathbf{z}) = \prod_{x=1}^{X}\prod_{e=1}^{E^x}\prod_{i=1}^{n_{xe}}\prod_{f=1}^{F^x}\prod_{h=1}^{H^x_f}\left\{P^{xe}_{fh}(\Omega)\prod_{k=1}^{K^x-1}\left(\frac{1}{\sqrt{2\pi}\sigma^{ke}_{fh}}\exp\left(-\frac{1}{2}\left(\frac{y^{xie}_{k,k+1}-E(y^{ke}_{fh})}{\sigma^{ke}_{fh}}\right)^2\right)\right)\right\}^{z^{x}_{iefh}} \tag{4.10}$$

4.2 Solution algorithm

As mentioned in section 4.1, the variable **z**, which represents the selection of activity pattern/path/activity location from travelers, is unobserved and cannot be directly identified by plate scanning (due to the nature of partial observation). Thus, in the estimation problem, one option is to treat **z** as “missing variables” (Watling et al., 1992, 1994) and apply the EM (Expectation-Maximization) algorithm to solve the estimation problem with missing **z**. By taking log of the function (4.10), we can obtain the “complete data log-likelihood function.” The EM algorithm will then iterate between the two steps: (i) maximizing this likelihood function with respect to the parameters, $\Lambda = \{\boldsymbol{\theta}, \boldsymbol{u}, \mathrm{s}, \boldsymbol{\sigma}, \psi\}$ (the M-step) given the expected value of **z**, and (ii) evaluating the expectations E(**z**) over the missing variables **z** given the parameter values (the E-step). Note that the reader can skip from the details of the EM algorithm as follows to the proposed algorithm without loss of generality of this paper.

In the M-step, taking log of function (4.10) yields the complete-data log-likelihood of:

$$L(\Lambda \mid \mathbf{y}, \mathbf{z}) = \sum_{x=1}^{X} \sum_{e=1}^{E^x} \sum_{i=1}^{n_{xe}} \sum_{f=1}^{F^x} \sum_{h=1}^{H_f^x} z_{iefh}^{x} \left(\ln(P_{fh}^{xe}(\Omega)) - \sum_{k=1}^{K^x-1} \ln(\sqrt{2\pi}\sigma_{fh}^{ke}) - \sum_{k=1}^{K^x-1} \tfrac{1}{2} \left(\frac{y_{k,k+1}^{xie} - E(y_{fh}^{ke})}{\sigma_{fh}^{ke}} \right)^2 \right) \quad (4.11)$$

The complete-data log-likelihood is linear in **z**. Thus, in the context of the EM algorithm, the expected complete-data log-likelihood is just (4.10) with each z replaced by its conditional expectation $E\left[\mathbf{z} \middle| \mathbf{Y}, \Lambda\right]$, where **Y** represents the vector of random variables in (4.3), $\mathbf{Y} = \{..., y_{fh}^{ke}, ...\}$, corresponding to particular values of observed sensor-to-sensor travel time **y**, and **z** is vector form of scalar z. Then, given conditional expectation E(**z**) (from E-step), model parameter vector, Λ, is updated by maximizing the completed-data log-likelihood (4.11). However, simultaneous updating of model’s attributes (or vector from $\mathbf{s}$ and $\mathbf{u}$) and coefficients of nested choice model ($\boldsymbol{\theta}$) in vector Λ can lead to the combinatorial problem. To avoid the combinatorial problem of updating Λ from the observed sensor-to-sensor travel time, **y**, the process to update Λ is divided into two steps:

First, given any scalar z (from E-step) and observed **y**, the mean activity duration ($\mathbf{s}$) and mean in-vehicle time ($\mathbf{u}$) are updated by maximizing log-likelihood of probability density function of observed sensor-to-sensor travel time (**y**) as follows.

$$L(\mathbf{s}, \mathbf{u}, \boldsymbol{\sigma} \mid \mathbf{y}, \mathbf{z}) = \sum_{x=1}^{X} \sum_{e=1}^{E^x} \sum_{i=1}^{n_{xe}} \sum_{f=1}^{F^x} \sum_{h=1}^{H_f^x} z_{iefh}^{x} \left(-\sum_{k=1}^{K^x-1} \ln(\sqrt{2\pi}\sigma_{fh}^{ke}) - \sum_{k=1}^{K^x-1} \tfrac{1}{2} \left(\frac{y_{k,k+1}^{xie} - E(y_{fh}^{ke})}{\sigma_{fh}^{ke}} \right)^2 \right) \quad (4.12)$$

where the random variable y_{fh}^{ke} can be formulated on the basis of the mean sensor-to-sensor travel time $E(y_{fh}^{ke})$ (or in a form of $\mathbf{s}$ and $\mathbf{u}$) and variance σ_{fh}^{ke}, which are expressed in (4.7) and (4.8b) of section 4.1, respectively.

Second, given **z** (from E-step) and model’s attributes ($\mathbf{s}$ and $\mathbf{u}$) updated from problem (4.12), route and activity choice coefficients ($\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_a$) and scale parameter ψ are simultaneously updated by maximizing log-likelihood of nested logit choice model (2.9) as follows.

$$L(\boldsymbol{\theta}, \psi \mid \mathbf{s}, \mathbf{u}, \mathbf{z}) = \sum_{x=1}^{X} \sum_{e=1}^{E^x} \sum_{i=1}^{n_{xe}} \sum_{f=1}^{F^x} \sum_{h=1}^{H_f^x} z_{iefh}^{x} \ln(P_{fh}^{xe}(\Omega)) \quad (4.13)$$

For the E-step, the uncompleted-data **z** is estimated by all updated values of model parameters (Λ) from the M-step as follows:

$$E\left[z^{x}_{iefh}\,|\mathbf{Y},\Lambda\right]=0.\Pr\left(z^{x}_{iefh}=0\,|\mathbf{Y},\Lambda\right)+1.\Pr\left(z^{x}_{iefh}=1|\mathbf{Y},\Lambda\right)=\Pr\left(z^{x}_{iefh}=1|\mathbf{Y},\Lambda\right) \tag{4.14}$$

Then, by standard laws of conditional probabilities:

$$\Pr\left(z^{x}_{iefh}=1|\mathbf{Y},\Lambda\right)=\frac{\Pr\left(z^{x}_{iefh}=1,\mathbf{Y}|\Lambda\right)}{\Pr\left(\mathbf{Y}|\Lambda\right)} \tag{4.15}$$

By the Bayes' rule (and the fact that z^{x}_{iefh} can only be 0 or 1):

$$\Pr\left(\mathbf{Y}|\Lambda\right)=\sum_{j=0}^{1}\Pr\left(z^{x}_{iefh}=j,\mathbf{Y}|\Lambda\right) \tag{4.16}$$

Putting (4.15) and (4.16) into (4.14), it gives:

$$E\left[z^{x}_{iefh}\,|\mathbf{Y},\Lambda\right]=\frac{\Pr\left(z^{x}_{iefh}=1,\mathbf{Y}\,|\Lambda\right)}{\sum_{j=0}^{1}\Pr\left(z^{x}_{iefh}=j,\mathbf{Y}|\Lambda\right)} \tag{4.17}$$

The probability distributions of selecting activity chain (f,h) associated with observed **y** required for the numerator and denominator of (4.17) can be defined by (4.10) and another application of Bayes' rule, yielding:

$$\Pr\left(z^{x}_{iefh}=j,\mathbf{y}|\Lambda\right)=\sum_{\substack{\text{all combinations } \mathbf{z} \\ \text{with } z_{iefh}=j}} q(\mathbf{y},\mathbf{z}|\Lambda) \tag{4.18}$$

Since the decision on activity chain (f,h) of user i is independent from other users, the combinations **z** in (4.18) consist of $z^{x}_{iefh}=1$ when user i selects activity chain (f,h) and $z^{x}_{iefh}=0$ for other cases that activity chain (f,h) is not selected from user i. Consequently, we can compute the expected z-value for user i selecting activity chain (f,h) as follows:

$$E\left[z^{x}_{iefh}\,|\mathbf{Y},\Lambda\right]=\frac{w^{x}_{iefh}}{\sum_{b=1}^{F_x}\sum_{c=1}^{H^x_f} w^{x}_{iebc}} \tag{4.19}$$

where $$w^{x}_{iebc}=P^{xe}_{bc}(\Omega)\prod_{k=1}^{K^x-1}\left(\frac{1}{\sqrt{2\pi}\sigma^{ke}_{bc}}\exp\left(-\frac{1}{2}\left(\frac{y^{xie}_{k,k+1}-E(y^{ke}_{bc})}{\sigma^{ke}_{bc}}\right)^2\right)\right) \tag{4.20}$$

Regarding to the procedure of the two tiers mentioned above, the proposed algorithm we can be summarized as below.

Proposed algorithm

Step 1: Initialize parameters $(\Lambda^0=\{\boldsymbol{\theta}^0,\boldsymbol{u}^0,S^0,\boldsymbol{\sigma}^0,\psi^0\})$; Set iteration: n = 0.

Step 2: From E-step, given Λ^n, find expected traveler's chosen activity chain, E(**z**), by solving (4.19) and (4.20). Then, set n=n+1.

Step 3: From M-step, find updated parameters $(\hat{\Lambda}^n=\{\hat{\boldsymbol{\theta}}^n,\hat{\boldsymbol{u}}^n,\hat{S}^n,\hat{\boldsymbol{\sigma}}^n,\hat{\psi}^n\})$ as follows.

Step 3.1: find $\hat{\mathbf{u}}^n$, $\hat{\mathbf{s}}^n$, and $\hat{\boldsymbol{\sigma}}^n$ by solving problem (4.12), given E(**z**) from step 2 and observed sensor-to-sensor travel time, **y**.

Step 3.2: find $\hat{\boldsymbol{\theta}}^n$ and $\hat{\psi}^n$ by solving problem (4.13), given E(**z**) from step 2 and updated mean in-vehicle time and activity duration ($\hat{\mathbf{u}}^n$, $\hat{\mathbf{s}}^n$) from step 3.1.

Step 4: If the convergence criterion is met, stop and final updated $\Lambda = \hat{\Lambda}^n$; otherwise go to step 2.

The convergence criterion bases on the maximum relative gap of all parameters in Λ at successive iterations: $\max \left| \dfrac{\hat{\Lambda}^n - \hat{\Lambda}^{n-1}}{\hat{\Lambda}^{n-1}} \right| \le \varepsilon$ where ε is the error tolerance.

Step 5: Estimate activity-chain demand, $d_{fh}^{xe} = \sum_{i=1}^{n_{xe}} z_{iefh}^{x}$.

Note that the initial Λ^0 can usually be obtained from prior data, which assumed to be known for this study. Furthermore, the accuracy of updating of model parameter (Λ) by the proposed algorithm with plate scanning data also depends on the sensor location. More details on the design of sensor locations can be found in appendix A.

5. NUMERICAL EXAMPLE

In this section, the proposed algorithm is tested with a modified Sioux Falls network. This network consists of 35 nodes, 98 links, and 10 traffic zones shown in Figure 3(a). In addition, Figure 3(b) presents the available activity types in each zone (i.e. zone 1-3 including stay-at-home (H) and transition (T), zone 4-7 including work (W), and zone 8-10 including maintenance (O)). The proposed sets of sensors of cordon-line-based scheme (see appendix A) in Figure 3(b) can collect 2,000 registered vehicles (i.e. the vehicle population travelling in the network), of which 440 vehicles of vehicle population have their trips originated from zone 1. The remains have their trips originated from zone 2 and 3.

5.1 Setting of the test case

Some network conditions are defined for the tests:

1) The choice alternation of out-of-home activities includes work (W) and maintenance (O) purposes. In addition, at-home activities include transition (T) and stay-at-home (H) purposes.
2) Four possible types of activity patterns have been explicitly observed.
   - Home-based work tour (H-W-H)
   - Home-based work tour including maintenance before work (H-O-W-H).
   - Home-based work tour including maintenance after work (H-W-O-H).
   - Home-based work tour with one secondary tour for maintenance (H-W-T-O-H).
3) The model resolution is divided into four periods (period 1: AM [6:30–9:29], period 2: Midday [9:30–15:59], period 3: PM [16:00–18:59], and Period 4: Other [19:00–6:29]).

   For instance, when the users perform trip chain H1-W5-T1-O8-H1 during travel period set {1,3,4,4}, these users start their first trip from home at zone 1 on period 1 and travel after work at zone 5 on period 3 to home for a transition purpose (e.g. taking a short break after work at home). Secondly, they leave home on period 4 for maintenance at zone 8 and return home within the same period.

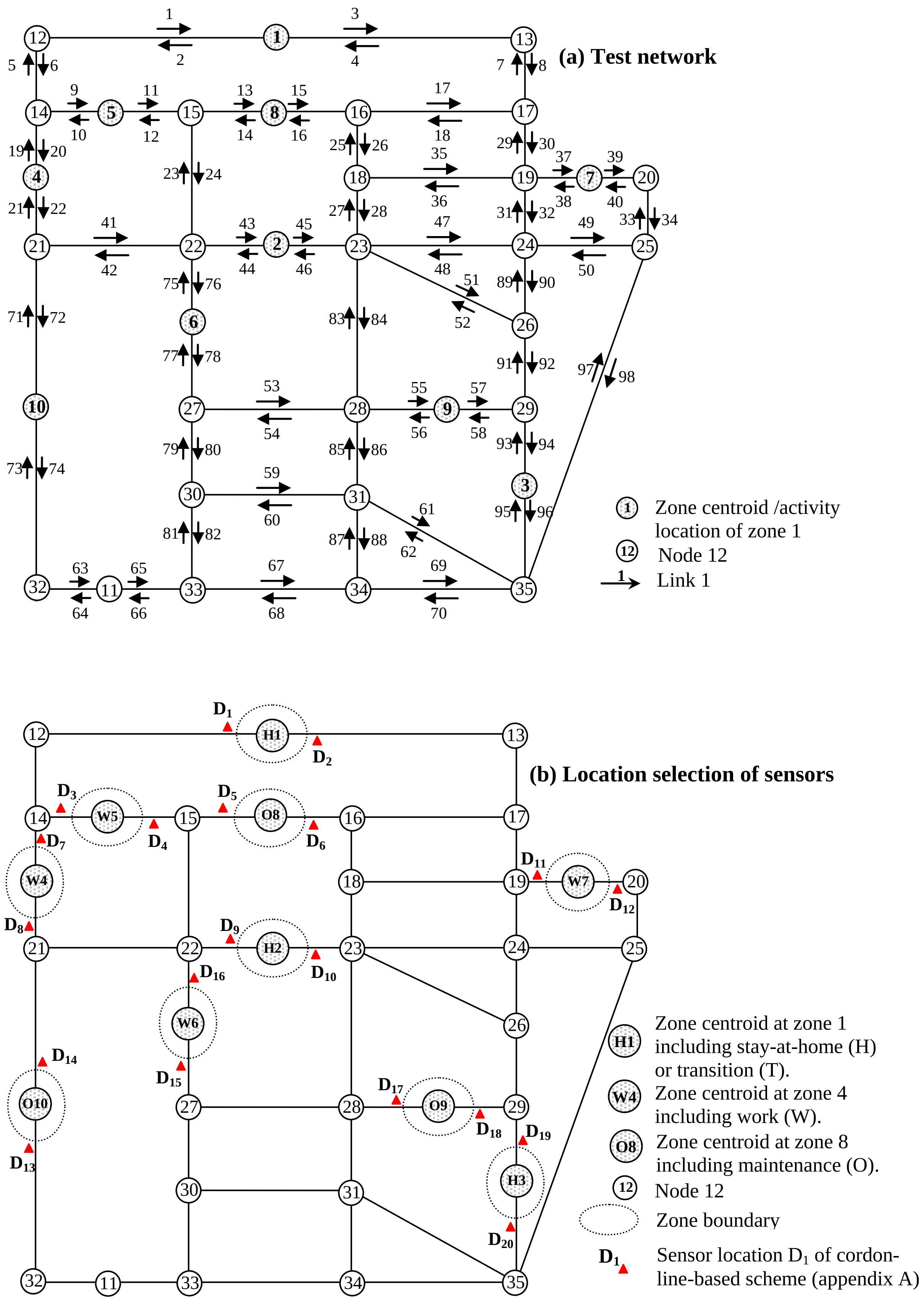


FIGURE 3 Test network and sensor location of cordon-line base scheme.

4) Travel time variation between two adjacent sensors during travel period set $\mathbf{C}_{\mathrm{e}}^{\mathrm{fh}}$, $(\sigma_{\mathrm{fh}}^{\mathrm{ke}})^2$, in (4.8b) can be estimated as a linear function of mean sensor-to-sensor travel time, it yields:

$$(\sigma_{\mathrm{fh}}^{\mathrm{ke}})^2 = \alpha \mathrm{E}(\mathrm{y}_{\mathrm{fh}}^{\mathrm{ke}}) \tag{5.1}$$

where α is a constant representing level of travel time variations.

5.2 Simulation and evaluation method

According to the simulation method in this study, given pre-specified true model parameters $\Lambda^{\mathrm{true}} = \{\boldsymbol{\theta}^{\mathrm{true}}, \boldsymbol{\mu}^{\mathrm{true}}, \mathbf{S}^{\mathrm{true}}, \boldsymbol{\sigma}^{\mathrm{true}}, \psi^{\mathrm{true}}\}$, the demand of activity pattern/location/path choice or activity chain can be calculated from nested choice logit model (2.9).

Based on such the activity-chain demand, the simulated sensor-to-senor travel time (**y**) of each traveler observed from plate scanning is obtained from a Monte-Carlo method by sampling **y** from normal variates in (4.3). For instance, given a true demand of vehicles passing on the sensor path (D2,D3,D5,D2) to perform trip chain H1-W5-H1 on vehicle path (2,6,9,11,13,15,17,7,4), these vehicles will stop at zone 1 (between D2 and D3) for stay-at-home purpose (H) and stop at zone 5 (between D3 and D5) for work purpose (W). Consequently, the simulated sensor-to-sensor time (from sensor D3 to sensor D5) based on this demand is drawn from the mean of in-vehicle time plus duration of work and its variance ($\boldsymbol{\mu}^{\mathrm{true}}$, $\mathbf{S}^{\mathrm{true}}$, $\boldsymbol{\sigma}^{\mathrm{true}}$). After the simulation, the initial model parameters in Λ^0 are set as the initial values for finding final updated parameters in $\hat{\Lambda}$ from solution algorithm (section 4.2).

In this study, evaluation of the statistical performance of updating of model parameters, $\hat{\Lambda}$, is also carried out through a Monte-Carlo method by randomly sampling initial estimates of true parameter, Λ^{true}. In particular, initial Λ^0 is generated by drawing from a normal variate $\Lambda^0 \sim (1+\eta)\cdot\Lambda^{\mathrm{true}}$ where η is a random error term of initial Λ^0 with a mean equal to $\hat{\eta}$ and a variance equal to σ_η^2. The statistical performance of final updated $\hat{\Lambda}$ can be measured by the percentage reduction of mean square error from initial Λ^0, defined as follows (Cascetta and Russo, 1997).

$$\mathrm{MSE}\%(\beta_{\mathrm{j}}) = [\mathrm{MSE}(\beta_{\mathrm{j}}^0) - \mathrm{MSE}(\hat{\beta}_{\mathrm{j}})] / \mathrm{MSE}(\beta_{\mathrm{j}}^0)\cdot 100\% \quad \text{for } \beta_{\mathrm{j}}^0 \in \Lambda^0 \text{and } \hat{\beta}_{\mathrm{j}} \in \hat{\Lambda} \tag{5.2}$$

where N = total number of trials of a dataset (Λ^0 and $\hat{\Lambda}$) with the same $\hat{\eta}$ and σ_η^2.

$\mathrm{MSE}\%(\beta_{\mathrm{j}})$ = percentage reduction of the mean square error of parameter in vector Λ.

$$\mathrm{MSE}(\beta_{\mathrm{j}}^0) = \sum_{\mathrm{n}=1}^{\mathrm{N}} (\beta_{\mathrm{n,j}}^0 - \beta_{\mathrm{j}}^{\mathrm{true}})^2 / \mathrm{N} \tag{5.3}$$

$$\mathrm{MSE}(\hat{\beta}_{\mathrm{j}}) = \sum_{\mathrm{n}=1}^{\mathrm{N}} (\hat{\beta}_{\mathrm{n,j}} - \beta_{\mathrm{j}}^{\mathrm{true}})^2 / \mathrm{N} \tag{5.4}$$

where $\beta_{\mathrm{j}}^{\mathrm{true}}$ = $\mathrm{j^{th}}$ parameter in true parameter vector Λ^{true}.

$\mathrm{MSE}(\beta_{\mathrm{j}}^0)$ = mean square error of initial value of $\mathrm{j^{th}}$ parameter in Λ^0 from N datasets; $\beta_{\mathrm{n,j}}^0$ is initial value of $\mathrm{j^{th}}$ parameter in Λ^0 of $\mathrm{n^{th}}$ trial.

$\mathrm{MSE}(\hat{\beta}_{\mathrm{j}})$ = mean square error of updated value of $\mathrm{j^{th}}$ parameter in $\hat{\Lambda}$ from N datasets; $\hat{\beta}_{\mathrm{n,j}}$ is updated value of $\mathrm{j^{th}}$ parameter in $\hat{\Lambda}$ of $\mathrm{n^{th}}$ trial.

5.3 Computational results

To examine the performance of updating of model parameters based on plate scanning, two tests were conducted as follows:

5.3.1 Random initialization test

To investigate initialization of in-vehicle time and activity duration (presenting the qualities of prior data), initial values of in-vehicle times and activity durations are set randomly to deviate from the true values set from the base case (Table 2). Given other model parameters, the results of updated in-vehicle time and activity duration (in Table 1) show that:

- The performance of the algorithm in updating the mean in-vehicle times/activity durations and variance of sensor-to-sensor travel time is generally satisfactory due to high percentage error reduction (MSE% of u^q, S_a, $(\sigma_{fh}^{ke})^2$ > 50%). This implies that the MSE of the updated model is substantially improved from the initial model, even though in the tests the initial parameters in the calibration process of the updated model were highly deviated from the true values ($\hat{\eta}$ =2.5).
- When the initial parameters of the variance of sensor-to-sensor time were set to be substantially different from the true value ($\hat{\eta} \geq 1.7$), the improvement of the updated model (parameters) in terms of MSE% can be limited. This may be due to the non-convex nature of the maximum likelihood estimation problem where the solution may be at the local optimal when the start point of the estimation problem is far from the global optimal solution (true parameters).
- From other results, the proposed algorithm generally performed satisfactorily in updating the activity duration parameters (mean of work, transition and maintenance durations). However, the algorithm performed not as well in updating the mean travel time parameters. This may be due to the issue of non-identification of the travel route based on the sensor-to-sensor time data. Nevertheless, the improvement of the MSE% for the case of updating the mean travel time is still acceptable (in the range of 58%-94% improvement compared to non-updated models).

5.3.2 The sensitivity test

With true parameters of mean in-vehicle time, u^q, (from zone to zone) and mean activity duration, S_a, (base case in Table 2), the updating of the parameters from the initial parameters in random initialization test has a good result (MSE% > 50%). In this initialization test, there is a significant difference between S_a and u^q (e.g. $S_{a=W}$ = 480 min. and u^q =18.8 min.). To further test on the robustness of the proposed algorithm, the sensitivity test with the reduction of S_a from base case was conducted (Table 2). Also, there are two levels of travel time variations set in this test (low (α=0.5) or high (α=1.5), see (5.1)). Value of α in base case and case 3 is equal to 0.5, and value of α in the other cases (case 2 and 4) is equal to 1.5. The initial values of u^q, S_a, $(\sigma_{fh}^{ke})^2$ from setting no.3 in Table 1 are adopted to represent the normal applications of these values. The numerical results are described as follows:

Updating of in-vehicle time

- The results (Table 2) show that, for the cases that a gap between S_a and u^q is similar (e.g. base case and case 2), the absolute percentage error (APE) in u^q increases when level of travel time variation, α , increases from 0.5 to 1.5. For instance, the APE of u^q in case 2 is larger than the APE of u^q in base case and the APE of u^q in case 4 is larger than the APE of u^q in case 3.

- With the same level of travel variation, α, the error (APE) in updating of u^q increase when activity duration, S_a, decreases (e.g. APE of u^q (case 2) < APE of u^q (case 4)).

Updating of activity duration

- For the cases with the same value of α, the error of updating of mean activity durations (work, maintenance and transition) tends to increase when the difference between S_a and u^q decreases. For instance, with the same value of α (α=1.50), the error of updating of $S_{a=W}$, $S_{a=O}$, or $S_{a=T}$ in case 2 is much less than the error in case 4 (Table 2).
- With the smallest values of S_a associated with high value of α (in case 4), the updating of mean activity duration, S_a has the worst result (Table 2).

Updating of coefficients of demand model

- The updated model coefficients θ_r or θ_a in all test cases are different from true values (Table 2). This implies that, due to non-convex maximization problem (solved by the EM algorithm), the updated model coefficients (θ_r,θ_a) are non-unique solutions. Consequently, there can be more than one solution of updating of (θ_r,θ_a) that can possibly reproduce the observations (sensor-to-sensor travel times).

Trip chain estimation

To exemplify trip chain estimation based on observed sensor-to-sensor travel time (**y**), true trip-chain demands derived from true Λ^{true} in case 3 (Table 2) are set to represent the case that activity durations of any purposes are smallest (or smallest values of (S_a - u^q)). The estimated trip chains are illustrated with 6 sensor paths of total simulated 50 sensor paths on test network. Path topology and path travel time of feasible trip chains of each sensor paths are described in Table 4. The results of trip chain estimations are (table 3):

- Due to the settings of smallest activity duration (S_a) in case 3, some travelers, performing the maintenance ($S_{a=O}$ = 6 min.) in the activity pattern (H-O-W-H and H-W-O-H), cannot be predicted in the E-step of the EM algorithm. As a result, trip chains of H-W-H are over-estimated and trip chains of H-O-W-H or H-W-O-H are under-estimated (Table 3).
- In contrast, the results from Figure 4 imply that, with largest difference between S_a and u^q, the smallest errors (MMSEs) of estimated trip chains in base case and case 2 are obtained. In addition, the MMSE in case 2 higher than the MMSE in base case, due to higher value of α. For the remaining cases (case 3-4), the performance of trip chain estimations tends to decrease when level of travel time variations, α, increases.

Discussions of results

The proposed algorithm performs well in updating the model attributes and predicting trip chains in some conditions (high activity durations and low travel time variations). This implies that, due to the large values of (S_a - u^q), the sensor-to-sensor travel time distributions (y) of the alternative (y_1) that travelers do at least one activity in a single traffic zone of two adjacent sensors are distinguished from another alternative (y_2) that travelers just drive through without doing any activity. Consequently, the chosen activities with two different distributions (y_1, y_2) are effectively predicted. By using proposed sensor locations (appendix A), the choices of travelling between two adjacent sensors (sensor k to k+1) are delimited to such the two alternatives. As a result, the chosen activity patterns made by most travelers are correctly predicted by a series of sensor-to-sensor travel times, **y**, (E-step). In accordance with accurate expected choice of activity patterns, E(**z**), model attributes are well updated.

Table 1 Updated results from random initial values after 15 trials (N=15).

Variable	Name	Index	Setting of initial variables 1	2	3	4	5	6	7
u^q	Mean in-vehicle time from zone to zone (min.)	Mean $\hat{\eta}$	0.20	0.50	0.70	1.00	1.50	2.00	2.50
		Var. σ_η^2	0.02	0.05	0.07	0.10	0.15	0.20	0.25
		MMSE%[a]	68	87	58	94	85	68	68
$S_{a=W}$	Mean work duration (min.)	Mean $\hat{\eta}$	0.20	0.50	0.70	1.00	1.50	2.00	2.50
		Var. σ_η^2	0.02	0.05	0.07	0.10	0.15	0.20	0.25
		MSE%	~100[b]	~100	~100	~100	~100	~100	~100
$S_{a=O}$	Mean maintenance duration (min.)	Mean $\hat{\eta}$	0.20	0.50	0.70	1.00	1.50	2.00	2.50
		Var. σ_η^2	0.02	0.05	0.07	0.10	0.15	0.20	0.25
		MSE%	~100	~100	~100	~100	~100	~100	~100
$S_{a=T}$	Mean transition duration (min.)	Mean $\hat{\eta}$	0.20	0.50	0.70	1.00	1.50	2.00	2.50
		Var. σ_η^2	0.02	0.05	0.07	0.10	0.15	0.20	0.25
		MSE%	97	97	~100	~100	99	99	~100
$(\sigma_{fh}^{ke})^2$	Variance of sensor-to-sensor time (5.1)	Mean $\hat{\eta}$	0.30	0.70	1.00	1.50	1.70	2.50	3.00
		Var. σ_η^2	0.03	0.07	0.10	0.15	0.17	0.25	0.30
		MSE%	~100	~100	~100	~100	71	60	52
Computational time[c] (min.)			48.8	52.2	53.8	53.9	53.1	54.4	54.2

[a]MMSE% of u^q is the average of MSE% of in-vehicle times from all OD pairs. [b]Value is close to 100%.
[c]All settings of initial variables are processed by Quad CPU i5-2400 @ 3.10 GHz and Ram 4 GB.

Table 2 Updated results from various settings of mean activity duration (S_a) and level of travel time variations (α).

Variable Name	Index	Case Base	2	3	4
Average of mean in-vehicle time, u^q, of all OD pairs (min.)	**True**	**18.8**			
	Est.	18.6	17.9	19.6	20.6
	APE	1.06	4.79	4.26	9.57
$S_{a=W}$, Mean work duration (min.)	**True** (S_a- u^q)	**480.0** (461.2)		**24.0** (5.2)	
	Est.	479.8	479.9	23.7	23.6
	APE	0.05	0.03	1.18	1.48
$S_{a=O}$, Mean maintenance duration (min.)	**True** (S_a- u^q)	**120.0** (101.2)		**6.0** (-12.8)	
	Est.	119.9	120.0	4.4	4.0
	APE	0.09	0.001	27.20	33.59
$S_{a=T}$, Mean transition duration (min.)	**True** (S_a- u^q)	**45.0** (26.2)		**2.3** (-16.5)	
	Est.	45.3	43.6	4.2	4.9
	APE	0.58	3.10	86.11	119.26

Table 2 Updated results from various settings of mean activity duration (S_a) and level of travel time variations (α) (Continue.).

Variable Name	Index	Case: Base	2	3	4
α level of travel time variation (5.1)	**True**	**0.50**	**1.50**	**0.50**	**1.50**
	Est.	0.49	1.49	0.92	3.55
	APE	0.55	0.52	83.73	136.46
θ_r, Route choice coefficient	**True**	**0.80**			
	Est.	0.82	0.81	0.60	0.62
	APE	2.55	0.97	25.01	22.76
θ_a , Activity choice coefficient of pattern H-W-H at $\mathbf{C}_e$ ={1,3}	**True**	**0.91**			
	Est.	0.36	0.35	0.02	0.00
	APE	60.28	61.92	97.58	100.00
θ_a, Activity choice coefficient of pattern H-W-O-H at $\mathbf{C}_e$ ={1,3,4}	**True**	**0.71**			
	Est.	0.28	0.27	0.00	0.00
	APE	73.58	60.77	100.00	100.00
θ_a, Activity choice coefficient of pattern H-W-T-O-H at $\mathbf{C}_e$ ={1,3,4,4}	**True**	**0.81**			
	Est.	0.40	0.39	2.00	2.00
	APE	50.37	51.46	146.91	146.91
θ_a, Activity choice coefficient of pattern H-O-W-H at $\mathbf{C}_e$ ={1,1,3}	**True**	**0.69**			
	Est.	0.27	0.26	0.00	0.00
	APE	60.82	62.90	100.00	100.00
Computational time[a] (min.)		55.9	55.0	54.1	41.0

[a]All settings of initial variables are processed by Quad CPU i5-2400 @ 3.10 GHz and Ram 4 GB.

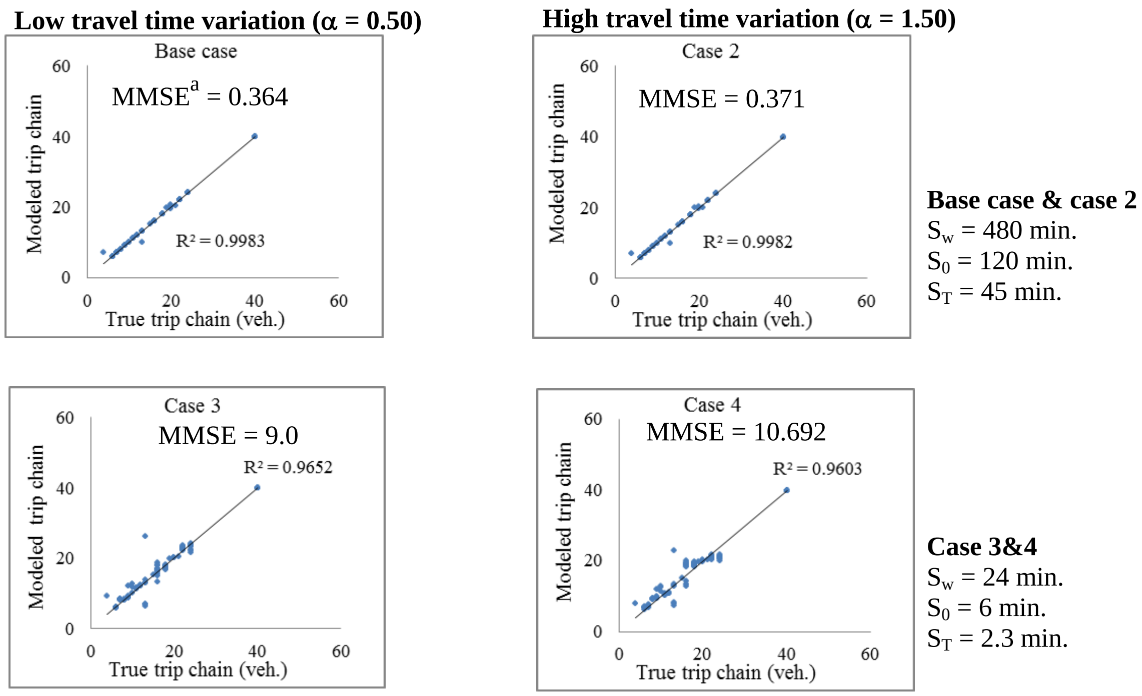
Table 3 Example of estimated results of trip chains originating from zone 1. [a]

x	Sensor path	Feasible trip chain	Path[b]	Time stamp period set	True flows (veh.)	Estimated flows (veh.)	MSE
1	D1 D18 D15 D4 D1	H1-O9-W6-H1	1	{1,1,3,3,1}	20.0	18.4	3.41
		H1-W6-H1	1	{1,1,3,3,1}	20.0	21.6	3.41
2	D2 D3 D5 D2	H1-W5-O8-H1	2	{1,3,4,1}	20.0	16.9	9.93
		H1-W5-H1	2	{1,3,4,1}	20.0	23.1	9.93
3	D1 D3 D16 D5 D4 D1	H1-W6-O8-H1	3	{1,1,3,3,4,1}	10.0	8.7	2.61
		H1-W5-O8-H1	3	{1,1,3,3,4,1}	10.0	9.0	1.51
		H1-W5-H1	3	{1,1,3,3,4,1}	10.0	11.0	1.51
		H1-W6-H1	3	{1,1,3,3,4,1}	10.0	11.3	2.61
4	D2 D3 D16 D5 D2	H1-W6-O8-H1	4	{1,1,3,4,1}	10.0	8.5	2.70
		H1-W5-O8-H1	4	{1,1,3,4,1}	10.0	8.8	1.79
		H1-W5-H1	4	{1,1,3,4,1}	10.0	11.2	1.78
		H1-W6-H1	4	{1,1,3,4,1}	10.0	11.5	2.71
5	D1 D11 D6 D4 D1	H1-W7-O8-H1	5	{1,1,3,4,1}	20.0	17.5	7.84
		H1-W7-H1	5	{1,1,3,4,1}	20.0	22.5	7.84
6	D2 D11 D6 D2	H1-W7-O8-H1	6	{1,3,4,1}	20.0	17.0	9.56
		H1-W7-H1	6	{1,3,4,1}	20.0	23.0	9.56

[a]True model parameters are set from case 3 in sensitivity test shown in Table 2.
[b]Given path topology and path travel time are shown as table 4.

Table 4 Example of path topology and path travel time of trip chains.

Path	Path topology	Path travel time (min.)
1	3 8 30 32 90 92 58 56 54 77 75 23 12 10 5 1	95
2	2 6 9 11 13 15 17 7 4	80
3	2 6 9 11 24 76 75 23 13 14 12 10 5 1	92
4	2 6 9 11 24 76 75 23 13 15 17 7 4	92
5	3 8 30 37 38 29 18 16 14 12 10 5 1	61
6	3 8 30 37 38 29 18 16 15 17 7 4	61



[a]MMSE of trip chain is the average of MSE of all trip chains.

Figure 4 Comparisons between true and modeled trip-chain demand from various setting of mean activity duration (S_a) and level of travel time variation (α) in sensitivity test (Table 2).

6. CONCLUSIONS

In this paper, a method based on the maximum-likelihood technique for updating travel behavior models parameters and estimation of vehicle trip chain by using sensor-to-sensor travel time of each sensor path collected from plated scanning is proposed. The solution algorithm, based on EM approach, uses the travel time of any user observed from any adjacent sensors making the specified sensor path at a certain time period. The structure of demand model is a nested logit. Based on plate scanning data collection, the chosen activity patterns/routes/locations are forecasted by maximizing the joint function between probability density function of observed sensor-to-sensor travel time and probability of nested logit choice model of activity-travel pattern. The proposed model and algorithm were tested with a modified Siouxfalls network, where the optimal set of sensors is installed. In addition, the relative gap of solution algorithm is less than the error tolerance (ε = 0.0001). The test results, then, showed that solution algorithm can produce a good updating of model parameters, Λ, when level of travel time variation, α, is low or mean activity duration is high. However, smaller activity duration and higher travel time variations leading to high overlap of sensor-to-sensor travel time distributions among activity-chain alternations (shown in sensitivity test) tend to update model parameters less accurate. The possible applications of the proposed method are:

- To update the importance attributes of travel demand model from the prior data (e.g. activity duration, travel time) without conducting the new household and travel survey (HTS).
- To predict trip chain demand based on plate scanning (using sensor-to-sensor travel time).

APPENDIX A

The design on sensor location scheme

A cordon-line-based sensor location scheme is proposed for this study, where any activity location, lo_q, in traffic zone is contained in the cordon line of sensors. Then, this scheme can fully detect vehicles before or after making an activity in a single traffic zone. The observed travel time (**y**) between two sensors of cordon line is used to predict the possibility (or the indicator variables **z** in (4.1)) of making this activity between them. As a result, large possible activity location choices can be diminished to a binary choice of an activity location visited by any user whether he/she passes through or stops to make the activity in a specified traffic zone.

The characteristics of the algorithm for cordon-line-based scheme are described as follows.

- The objective (A.1) is to find minimum numbers of links installed with sensors that can provide the conditions of cordon-line-based scheme.
- To satisfy such a scheme, there must be at least one sensor contained on each feasible path between any two adjacent activity locations (lo_q, lo_{q+1}) of any activity chain (f,h). This requirement (A.2) for collecting the set of observed timestamp period, **M**, can be used to distinguish the set of travel period, **C**. Consequently, for any path of activity chain (f,h), the path consists at least the link l located with a sensor. This constraint is also known as sensor location scheme of path coverage (see Yang et al. (1998) and Castillo et al. (2008, 2011) for more applications of sensor location schemes).
- In addition, the constraint (A.3) is another specific characteristic of the cordon-line-based scheme in that any activity location in specified traffic zone can be contained in a cordon line of sensors. In other words, the travelers can pass on the link installed with sensors at least one time before arriving to and after departing from a single traffic zone where they perform an activity. Consequently, for each sensor paths of cordon-line-based-scheme, feasible trip chains are delimited to only choices of performing an activity in single traffic zone located in between two adjacent sensors.

The binary linear programming problem of this scheme can then be formulated as follows:

$$\min \sum_{l \in \mathbf{L}} u_l , \tag{A.1}$$

subject to

$$\sum_{l \in \mathbf{L}'} u_l b_l^{p^{O,D}} \geq 1; \ \forall O, D, p, \ b_l^{p^{O,D}} = \begin{cases} 1 & \text{if link } l \text{ is in path } p \text{ from O to D.} \\ 0 & \text{otherwise.} \end{cases} \tag{A.2}$$

$$\sum_{l \in \mathbf{L}'} u_l d(p_1^{O,D1}, p_2^{O,D2}, l) \geq 1; \;\; \forall O, D1, D2, \quad p_1^{O,D1} \neq p_2^{O,D2} \tag{A.3}$$

$$, \; d(p_1^{O,D1}, p_2^{O,D2}, l) = \begin{cases} 1 & \text{if } b_l^{p_1^{O,D1}} \neq b_l^{p_2^{O,D2}}; \; D1 \neq D2 \; . \\ 0 & \text{otherwise.} \end{cases}$$

where $\mathbf{L}'$ = the set of all links excluding any link in origin zone O, destination zone D1 and zone D2, $\mathbf{L}' \subset$ set of link, $\mathbf{L}$.

$p_1^{O,D1}$ = path 1 (from origin O to destination D1).

$b_l^{p_1^{O,D1}}$ = link-path incident value of path 1 containing with link l.

$d(p_1^{O,D1}, p_2^{O,D2}, l)$ = the indicator variable [$d(p_1^{O,D1}, p_2^{O,D2}, l)$ = 1: link l is in path 1 and not in path 2 or vice versa; and $d(p_1^{O,D1}, p_2^{O,D2}, l)$ = 0: otherwise].

u_l = a binary value such that it takes one when the link l is located by a sensor, and 0 otherwise.

ACKNOWLEDGEMENTS

The authors are indebted to the General Research Fund of the Hong Kong Research Council (RGC) (Project: PolyU 5290/09E) for support of this research.

REFERENCES

Arentze, T.A., & Timmermans, H.J.P. (2004). A learning-based transportation oriented simulation system. Transportation Research Part B, **38**(7), 613-633.

Axhausen, K.W., Schönfelder, S., et al. (2003). 80 weeks of GPS-traces: Approaches to enriching the trip information, Arbeitsbericht Verkehrs- und Raumplanung, **178**, Institut für Verkehrsplanung und Transportsysteme, ETH Zürich, Zürich.

Bell, M.G.H. (1991). The estimation of origin-destination matrices by constrained generalised least squares. Transportation Research Part B, **25**(1), 13-22.

Ben-Akiva, M., & Lerman, S.R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge.

Bhat, C.R. & Koppelmen, F.S. (1999). Handbook of Transportation Science. New York, Kluwer academic publishers.

Bifulco, G., Cartenì, A., & Papola, A. (2010). An activity-based approach for complex travel behaviour modelling. European Transport Research Review, **2**, 209-221.

Bradley, M., Bowman, J.L., & Griesenbeck, B., (2010). SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution. Journal of Choice Modeling, **3**(1), 5-31.

Bricka, S., & Bhat, C.R. (2006). A Comparative analysis of GPS-based and travel survey-based data. Transportation Research Record, **1972**, 9-20.

Bowman, J.L., & Bradley, M.A. (2008, October). Activity-based model: approaches used to achieve integration among trips and tours thoughtout the day. Paper presented at the European Transport Conference, Leeuwenhorst, The Netherlands.

Bowman, J.L., Bradley, M.A., & Gibb, J. (2006, September). The Sacramento activity-based travel demand model: estimation and validation results. Paper presented at the European Transport Conference, Strasbourg, France.

Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. Transportation Research Part B, **18**, 289-299.

Cascetta, E., Inaudi, D., & Marquis, G. (1993). Dynamic estimators of origin-destination matrices using traffic counts. Transportation Science, **27**(4), 363–373.

Cascetta, E., & Russo, F. (1997). Calibrating aggregate travel demand models with traffic counts: Estimators and statistical performance. Transportation, **24**(3), 271-293.

Castillo, E., Menéndez, J.M., & Jiménez, P. (2008). Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. Transportation Research Part B, **42**(5), 455-481.

Castillo, E., Gallego, I., Menéndez, J.M. & Jiménez, P. (2011): Link flow estimation in traffic networks on the basis of link flow observations, Journal of Intelligent Transportation Systems: Technology,Planning, and Operations, **15**(4), 205-222.

Cools, M., Moons, E., & Wets, G. (2010). Calibrating activity-based models with external origin–destination information: overview of possibilities. Transportation Research Record, **2175**, 98-110.

Feil, M., Balmer, M., & Axhausen, K.W. (2009). Generating comprehensive all-day schedules: expanding activity-based travel demand modelling. European Transport Planning Colloquium.

Flötteröd, G., Chen, U., & Nagel, K. (2011). Behavioral calibration and analysis of a large-scale travel microsimulation. Networks and Spatial Economics, A journal of Infrastructure Modeling and Computation, 1-22.

Frignani, M.Z., Auld, J., Mohammadian, A.K., & Williams, C. (2010). Urban travel route and activity choice survey: Internet-based prompted-recall activity travel survey using global positioning system data. Transportation Research Record, **2183**, 19-28.

Hu, S.R., Madanat, S., Krogmeier, J., & Peeta, S. (2001). Estimation of dynamic assignment matrices and OD demands using adaptive kalman filtering, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, **6**(3), 281-300.

Li, Z.C., Lam, W.H.K, Wong S.C., & Sumalee A. (2010). An activity-based approach for scheduling multimodal transit services. Transportation, **37**(5), 751-774.

Lao, Y., Zhang, G., Corey, J., & Wang, Y. (2012) Gaussian mixture model-based speed estimation and vehicle classification using single-loop measurements, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, **16**(4), 184-196.

Maruyama, T., & Sumalee A. (2007). Efficiency and equity comparison of cordon- and area-based road pricing schemes using a trip-chain equilibrium model. Transportation Research Part A, **41**(7), 655-671.

Maher, M.J. (1983). Inferences on trip matrices from observations on link volumes: A bayesian statistical approach. Transportation Research Part B, **17**(6), 435-447.

Ortuzar, J.D. & Willumsen, L.G. (2011). Modelling Transport (fourth Edition). Chichester, Wiley.

Ozbay, S., & Ercelebi, E. (2005). Automatic vehicle identification by plate recognition. World Academy of Science, Engineering and Technology, **9**, 222-225.

Velaga, N.R., Quddus, M.A., & Bristow, A.L. (2012). Improving the performance of a topological map-matching algorithm through error detection and correction, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, **16**(3), 147-158.

Vovsha, P., Bradley, M. & Bowman, J.L. (2004, May). Activity-based travel forecasting models in the United States: Progress since 1995 and Prospects for the Future. Paper presented at the EIRASS Conference on Progress in Activity-Based Analysis, Maastricht.

Watling, D.P. (1994). Maximum likelihood Estimation of an origin-destination matrix from a partial registration plate survey. Transportation Research Part B, **28**(4), 289-314.

Watling, D.P., & Maher, M.J. (1992). A statistical procedure for estimating a mean origin-destination matrix from a partial registration plate survey. Transportation Research Part B, **26**(3), 171-193.

Yang, H., Sasaki, T., et al. (1992). Estimation of origin-destination matrices from link traffic counts on congested networks. Transportation Research Part B, **26**(6), 417-434.

Yang, H., & Zhou, J. (1998). Optimal traffic counting locations for origin-destination matrix estimation. Transportation Research Part B, **32**(2), 109–126.

Yang, H. (1995). Heuristic algorithms for the bilevel origin-destination matrix estimation problem. Transportation Research Part B, **29**, 231–242.

Yuan, Y., Lint, H.V., Wageningen-Kessels, F.V., & Hoogendoorn, S. (2013) Network-wide traffic state estimation using loop detector and floating car data, Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, Accepted for publication.