Statistical Approach for Activity-Based Model Calibration based on Plate Scanning and Traffic Counts Data

Treerapot Siripirote[a], Agachai Sumalee[1,b], H.W. Ho[c], William H.K. Lam[c]

[a] *Department of Civil Engineering, Faculty of Engineering, Srinakharinwirot University, Ongkarak, Nakhonnayok, Thailand*
[b] *Department of Civil Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand*
[c] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China*

Abstract

Traditionally, activity-based models (ABM) are estimated from travel diary survey data. The estimated results can be biased due to low-sampling size and inaccurate travel diary data. For an accurate calibration of ABM parameters, a maximum-likelihood method that uses multiple sources of roadside observations (link counts and/or plate scanning data) is proposed. Plate scanning information (sensor path information) consists of sequences of times and partial paths that the scanned vehicles are observed over the preinstalled plate scanning locations. Statistical performances of the proposed method are evaluated on a test network using Monte Carlo technique for simulating the link flows and sensor path information. Multiday observations are simulated and derived from the true ABM parameters adopted in the choice models of activity pattern, time of the day, destination and mode. By assuming different number of plate scanning locations and identification rates, impacts of data quantity and data quality on ABM calibration are studied. The results illustrate the efficiency of the proposed model in using plate scanning information for ABM calibration and its potential for large and complex network applications.

Keywords: Maximum-likelihood estimation; Plate scanning; Statistical model calibration

1. Introduction

Activity-based models (ABM) have been developed to overcome the limitations from the conventional 4-step models by deriving travel demand from activity participations and activity behavior sequences/patterns (Bhat et al., 1999). ABMs can either be formulated by utility maximization-based (e.g. Bowman et al., 2006; Bifulco et al., 2010) or rule-based approach (e.g. Arentze and Timmermans, 2004). The utility maximization-based ABMs have been widely developed to evaluate traffic policies in many cities such as Portland, Columbus, Sacramento, and Jakarta (Bradley et al., 2010; Vovsha et al., 2004; Yagi and Mohammadian, 2010).

Traditionally, activity-based models are estimated from travel diary survey data (TD) of which estimated results can be biased due to low-sampling size and inaccurate TD. For example of the ABM with complex activity-travel decisions, only 1% of the population was used to estimate the ABM parameters (Bowman et al., 2006). In addition, trip under-reporting, due to response burdens or incomplete/inaccurate trip memory, will also lead to erroneous travel diary data (Bricka and Bhat, 2006). Consequently, the predicted travel demands based on TD may be inconsistent with actual roadside data (e.g. link counts). In order to calibrate ABM parameters

---

[1] Corresponding author. Tel.: +66-89280-1692; fax: +66-2274-0698.
E-mail address: asumalee@gmail.com (A. Sumalee).

that consist with roadside data, Bowman et al. (2006) developed a comprehensive model calibration approach. According to their approach, some ABM parameters in the utility function (e.g. constant terms) are heuristically adjusted to reproduce predicted traffic flows that fit with the collected traffic counts. In recent years, Cools et al. (2010) also conducted ABM calibration by a heuristic method. In their study, the demand of activity-travel patterns was adjusted by randomly weighting their activity-travel pattern choices to reproduce external trip matrix information. The ABM calibration from this method, however, still lacks statistical efficiency to measure how well the model calibration results fit the actual roadside data (e.g. link counts).

Ideally, the data used to calibrate the parameters in ABM should be collected by GPS-based travel surveys such as activity-travel data collected by GPS equipment attached to probe vehicles or carried by travellers. (e.g. Cottrill et al., 2013; Frignani et al., 2010, Axhausen et al., 2003). The GPS-based data can provide a probe of a sequence of activity/time spent/travel decisions, which is an RP-typed data for the activity-based model. To maintain the high accurate TD data and minimize response burden, learning machine technique was conducted to pre-determine the activity-travel choices of the respondents from the historical record of travel diary made (Cottrill et al., 2013). Nevertheless, the sample sizes of the activity-travel data collected by GPS-based travel surveys may be small due to low response rate. As an alternative, it is possible that number plate scanning (PS) could be used to identify travel paths. Information, in the context of vehicle tracking, obtained from plate scanning, which does not require the installation of GPS equipment, is similar to that of the GPS-based data collection method. In PS data collection, vehicle information is obtained at pre-determined locations on the road network.

Furthermore, plate scanning process can identify the same vehicles travelling 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 the methods to collect vehicle identification (VI) data. The data from plate scanning consists of: (i) the vehicle passing time at plate scanning locations, and (ii) the sequence of scanned vehicles along a series of plate scanning locations on road networks. The accuracy of collecting the above data from the plate scanning method is determined from detection rate and identification rate. Detection rate is the proportion of the number of detected vehicles among those vehicles passing sensor locations (i.e. the vehicle is known to pass the sensor and the license plate number of the vehicle can be detected by the sensor). In addition, identification rate is the proportion of detected vehicles that their license plate numbers are correctly identified. To achieve a high identification rate, Ozbay and Ercelebi (2005) proposed a license plate recognition system with more than 90% of license plate samples correctly identified.

Further to the related studies on model calibrations with VI data, Castillo et al. (2008b) developed trip table reconstruction/estimation framework using plate scanning. In recent years, Siripirote et al. (2014) attempted to update travel behavior model parameters and to estimate trip chains by using individual vehicle's timestamps from plate scanning. In this study, a statistical framework for the calibration of activity-based model parameters using sensor path flows from plate scanning information is proposed that is based on the hierarchical activity-travel decision model. This model calibration approach was motivated by similar works done in the area of traffic flow estimations, in which one observes traffic in some links and tries to estimate OD flows or flows on remaining links (e.g. Castillo et al., 2007, 2008a, 2010, 2011, 2013; Maher, 1983; Ng, 2012; He, 2013; Watling et al., 1992, 1994; Yang et al., 1992, 1995; Shen and Wynter, 2012). The ABM calibration proposed in this paper is formulated as a maximum likelihood estimation problem for reproducing the collected link counts and sensor path flow information. The remainder of the paper is organized as follows. Firstly, basic components of the proposed model, including notation and information collected from plate scanning, are described in Section 2. The formulation and solution algorithm of the vehicle reidentification based ABM calibration model are then described in the third section. In

Section 4, numerical example for evaluating the proposed calibration model is setup, solved and discussed. Finally, conclusions are drawn in the last section.

2. Assumptions and Definitions

In this section, the necessary assumptions and definitions for setting up the proposed ABM calibration model (Section 3) are introduced. The definitions in network, activity chain representations, and collected plate scanning data could be respectively found in Section 2.1 and 2.2, while assumptions of the proposed model will be given in Section 2.3.

2.1 Network and activity chain representation

For traffic network $(\mathbf{N}, \mathbf{L})$, $\mathbf{N}$ is the set of nodes and $\mathbf{L}$ is the set of links. Activity location *lo* is located in each traffic zone where $\mathbf{N}_z$ is the set of zone centroids and $\mathbf{L}_z$ is the set of links in traffic zone z ($\mathbf{N}_z \subset \mathbf{N}$ and $\mathbf{L}_z \subset \mathbf{L}$). In this study, the activity location (*lo*) is assumed to be virtually located at the zone centroid, which is one of the nodes in a network, and represents all real activity locations in each traffic zone ($lo \in \mathbf{N}_z$).

As we consider on daily activity-travel participations, user i (i^{th} observed vehicle) makes a plan to perform *activity pattern y*. Let $\mathbf{A}_y$ denotes the daily scheduled activity pattern that consists of an ordered set of the activities:

$$\mathbf{A}_y = \{a_1, ..., a_n, ..., a_{N_y}\}, \quad \forall y \in \{1, ..., Y\} \tag{2.1}$$

where Y is the total daily activity patterns, a_n is an activity performed in sequence $n \in \{1, ..., N_y\}$ of activity pattern y, and N_y is the total number of activities included in activity pattern y. For instance, if the activity pattern 1 (i.e. *y=1*) is Stay-at-home (H)-Working (W)-Stay-at-home (H), then the corresponding variable ($\mathbf{A}_1$) is defined as $\mathbf{A}_1 = \{H, W, H\}$.

To complete daily activity-travel participations, this user also needs to select a *trip chain* that consists of an ordered set of activity location/mode/time of day. Given the list of activities in the specified activity pattern performed by individuals (user i), $\mathbf{A}_y$, the trip chain h (the combined set of activity locations/modes/time of the day daily performed by trip makers starting at origin zone *o*) is denoted as $\mathbf{B}_{y,h}^o$ and defined as follows:

$$\mathbf{B}_{y,h}^o = \{(lo_1, ..., lo_n, ..., lo_{N_{y,h}}), (mo_{1,2}, ..., mo_{n,n+1}, ..., mo_{N_{y,h}-1,N_{y,h}}), (m_{1,2}, ..., m_{n,n+1}, ..., m_{N_{y,h}-1,N_{y,h}})\}$$
$$\forall h \in \{1, ..., H_y\}, \ lo_1 = o, \ lo_n \in \mathbf{N}_z, \ n \in \{1, ..., N_{y,h}\} \tag{2.2}$$

where lo_n is the activity location n where individuals perform an activity, $mo_{n,n+1}$ is the travel mode from activity location lo_n to lo_{n+1}, $m_{n,n+1}$ is the travel period of individuals traveling from lo_n to lo_{n+1}, $N_{y,h}$ is the total number of visits at activity locations of individuals who make trip chain h associated with activity pattern y and, H_y is the total number of trip chains associated with activity pattern y. Note that a trip chain associated with at least one out-of-home activity that begins and ends at the same activity location (i.e. $lo_1 = lo_{N_{y,h}}$) is called a *tour*. A tour that begins at home is called *home-based tour*. In addition, for simple illustrations

throughout the paper, individuals, who make trip chain h (originated at zone o) associated with activity pattern y, perform *activity chain[2] j* denoted as Υ_j and defined as $\Upsilon_j = \mathbf{A}_y \cup \mathbf{B}_{y,h}^o$.

2.2 Data collection from plate scanning

For traffic network $(\mathbf{N}, \mathbf{L})$, let $\bar{l}$ be the link installed with sensor that scans license plate of vehicles (simply called scanned link $\bar{l}$) and $\Im$ be the set of scanned links ($\bar{l} \in \Im$; $\Im \subseteq \mathbf{L}$). When user i completes activity pattern y associated with trip chain h (or activity chain j) travelling on the network, user i has been scanned (identified) by a series of sensors locating on the scanned links. The path travelled by user i can be represented as an ordered set of links passed by the traveller. Sensor path is the path, or partial path, of travelers represented as the ordered set of scanned links identifying the same vehicle using license plate matching technique. The scanned sensor path q of user i, $\mathbf{C}_{q_i}$, can be defined as follows.

$$\mathbf{C}_{q_i} = \{\bar{l}_1^i, ..., \bar{l}_k^i, ..., \bar{l}_{K^q}^i\}, \quad \forall q \in \{1, ..., N_q\} \tag{2.3}$$

where $\bar{l}_k^i$ is the scanned link that user i is identified in order k of the links that he/she is scanned, K^q is the total number of links which are scanned on the sensor path q and, N_q is the total number of sensor paths. Other information that can be collected from plate scanning is the time period of user i passing through scanned link $\bar{l}_k^i$. The ordered set (index e) of travel periods of user i is defined as follows.

$$\mathbf{M}_{e_i} = \{m_1^i, ..., m_k^i, ..., m_{K^q}^i\}, \quad \forall e \in \{1, ..., N_e\} \tag{2.4}$$

where m_k^i is the time period of user i on scanned link $\bar{l}_k^i$, and, N_e is the total number of travel period sets.

To illustrate data collection from plate scanning, a simple network with 3 zone centroids and 4 links is defined and shown in Figure 1(a). For simplicity, it is assumed that there are only 2 specific travel patterns (i.e. H1-W2 and H1-W3) made at the same travel period (say AM period). Also, these travel patterns contain 2 OD pairs and 2 paths (see Table 1). Within this network, user 1 is assumed to travel on pattern H1-W3 (or OD pair 2). If this user can be correctly identified from all the sensors that he or she has passed (full identifiability), the path of user 1 is directly distinguished by scanned sensor path (1,2), or $\mathbf{C}_{q_1} = \{1, 2\}$ (see Figure 1(c)). In this case, the scanned sensor path of full identifiability is called *true sensor path*. Then, the summation of all vehicles observed by true sensor path (1,2) can be reconstructed to be the path flows of OD pair 1. If the sensors on link 1 or 2, however, cannot correctly identify this user, there will be more than one ordered set of scanned sensor paths possibly observed from the path trace of user 1 (i.e. $\mathbf{C}_{q_1} \subset \{(1), (2), (1,2)\}$, see Figure 1(d) and Table 1). In this case, if sensor path (1) is observed, there is more than one OD pair that could possibly be derived (e.g. OD pair 1 and 2 possibly derived from sensor path (1)). Note that ABM parameter calibrations with partial identifiability of plate scanning information is illustrated in section 3. In addition, the solution algorithm for calibrating ABM parameters is presented in appendix A.

---

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

Table 1
Travel pattern, OD pair, and sensor path information

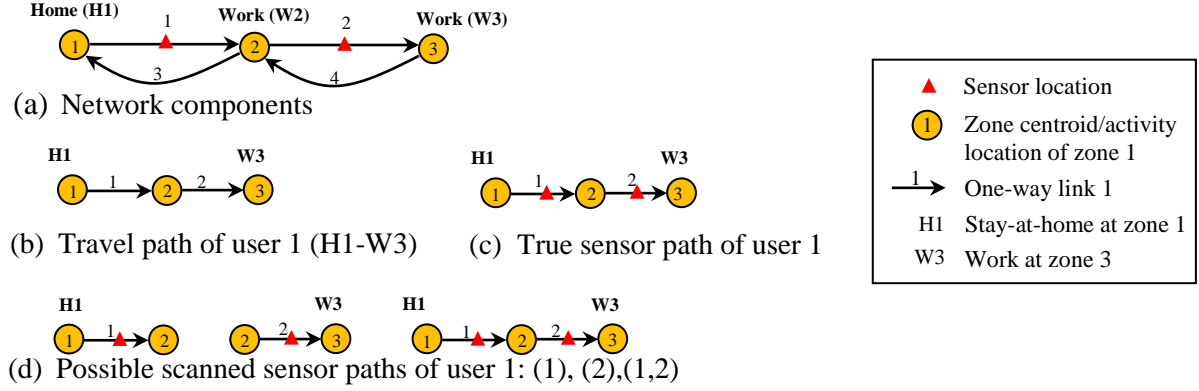| No.	Travel pattern	OD pair	Path	True sensor path s	Possible scanned sensor path q
1	H1-W2	1	(1)	(1)	(1)
2	H1-W3	2	(1,2)	(1,2)	(1), (2), (1,2)

Fig. 1. Network description and sensor path information

2.3 Assumptions

In this paper, the following assumptions are generally taken in model formulation and numerical example:

(i)   An activity-travel behaviour model that is based on random utility maximization-based approach is adopted for this study. The utility of performing an activity pattern/trip chain is based on socio-demographics, land-use data, and transport system characteristics. (e.g. Bifulco et al., 2010; Bowman et al., 1998, 2000, 2006, 2008; Bradley et al., 2010).

(ii) The choice set of activity pattern/trip chain (activity chains) of travellers is assumed to be given. In addition, the feasible trip chain (travel pattern) is assumed to be a home-based tour. These assumptions have also been adopted in the related studies (e.g. Cools et al., 2010; Li et al., 2010; Maruyama and Sumalee, 2007; Siripirote et al., 2014).

(iii) All users passing any sensors are assumed to be fully detected (i.e. detection rate = 100%). In addition, traffic counts at these sensor locations are accordingly taken as error-free (e.g. Castillo et al., 2008b; Mínguez et al., 2010). However, some observed travelers might not be correctly identified (i.e. identification rate ≤ 100%).

3.   Formulation for vehicle reidentification based ABM calibration

In this study, statistical method to calibrate an activity-based model using plate scanning (PS) data is proposed. Plate scanning (PS) data are much more informative than link counts data as PS data can generate more known equations for calibrating ABM parameters (e.g. Castillo et al., 2008b for path reconstruction problem). The structured relations between demands of activity chains and observed traffic flows (especially PS flows) are described as follows:

3.1 Demand side

Based on assumption (i), demand function in the activity-based model is assumed to be explicitly related to the exogenous variables (e.g. Li et al., 2012) via a K-dimensional parameter vector $\boldsymbol{\alpha}$, with $\alpha_k \in \boldsymbol{\alpha}$, the kth coefficient of ABM. Let $\Pr_j^{o,p}(\boldsymbol{\alpha})$ be the proportion that population type p (classified by person and household demographic (e.g. gender and household income)) in origin zone o select a specific activity chain j. Let $\boldsymbol{\Gamma}$ be the choice set of activity chains and let $J=|\boldsymbol{\Gamma}|$ denote the number of activity chains, $j \in \{1,...,J\}$. Let $\mathbf{N}^o$ be the random vector of the realized number of population in traffic zone o (i.e. $\mathbf{N}^o = (n_1^o,...,n_p^o,...,n_{|p|}^o)^{\mathrm{T}}$) where p is the type of population and $|p|$ is the number of population types. To specify the distribution of $\mathbf{N}^o$, the poisson distribution of $\mathbf{N}^o$ is assumed, where $n_p^o \sim \mathrm{Poisson}(\mathrm{E}[n_p^o])$; $p \in \{1,...,|p|\}$, $o \in \mathbf{N}_z$ and $\mathrm{E}[\ ]$ indicates an expected value of variables in the bracket. Regarding the thinning property of poisson distributions, the realized demands of activity chain j, u_j (fractions of number of population, $\mathbf{N}^o$; $\forall\ o$) can be expressed by:

$$u_j \sim \mathrm{Poisson}(\mathrm{E}[u_j(\boldsymbol{\alpha})]) \quad \forall j \in \{1,...,J\} \tag{3.1a}$$

where $\mathrm{E}[u_j(\boldsymbol{\alpha})] = \sum_{o,p} \mathrm{E}[n_o^p]\Pr_j^{o,p}(\boldsymbol{\alpha})$ \hfill (3.1b)

Note that the expected number of population, $\mathrm{E}[n_o^p]$, is normally obtained from population synthesis based on Monte Carlo simulation in ABM and assumed to be given for this study.

Let $\mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})] = (\mathrm{E}[u_1(\boldsymbol{\alpha})],...,\mathrm{E}[u_j(\boldsymbol{\alpha})],...,\mathrm{E}[u_J(\boldsymbol{\alpha})])^{\mathrm{T}}$ be the vector of expected activity chain demands. According to the central limit theorem, if the number of population is large enough, a normal approximation of activity chain demands can be used as follows:

$$\mathbf{u} \sim \mathrm{N}(\mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]\ ,\ \Sigma_u) \tag{3.2}$$

where $\Sigma_u = \mathrm{diag}(\mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})])$ is the diagonal matrix in which diagonal is $\mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]$ and off-diagonal is zero. Also, N() indicates a dimensioned multivariate normal distribution with arguments representing its mean vector and covariance matrix respectively. The major benefit of working this normal approximation is the mathematical tractability of the corresponding likelihood theory. With actual dispersions of $\mathbf{u}$ which can vary from space and time, the normal model can simply be modified to account for different levels of dispersions (coefficient of variations) ϕ, i.e. $\mathbf{u} \sim \mathrm{N}(\mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})], \phi\Sigma)$ which can account for underdispersion ($\phi < 1$) or overdispersion ($\phi > 1$) in comparison of the poisson case ($\phi = 1$).

To consider the relations between the activity chain demands and the standard forms of OD flows, let OD pair w representing a trip (by specific mode mo and travel period m) from origin zone o to specific destination zone. Based on our normal approximation, the vector of OD flows, $\mathbf{t}$, can then be mapped by the activity chain demand matrix, $\mathbf{u}$, as the following equation.

$$\mathbf{t} = \boldsymbol{\delta} \cdot \mathbf{u} \sim \mathrm{N}(\boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]\ ,\ \boldsymbol{\delta} \cdot \Sigma_u \cdot \boldsymbol{\delta}^{\mathrm{T}}) \tag{3.3a}$$

where $\boldsymbol{\delta} = (\delta_w^j)$ is the OD demand-activity chain demand proportion matrix. We have mean OD flows: $\mathrm{E}[\mathbf{t}(\boldsymbol{\alpha})] = \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]$ by definition, and OD demand pair w, t_w, is defined as:

$$t_w = \sum_j \delta_w^j \cdot u_j \quad \forall\, w \tag{3.3b}$$

where δ_w^j is the number of OD pair w that contains in activity chain j.

3.2 Supply side

To consider *the stochastic supply side* of the activity-based model, the basic characteristics of measurements on links and paths (i.e. link flows and path flows) are considered. Let $\mathbf{R}_w$ be the non-empty path set of the OD pair w, define $\mathbf{R} = \cup_w \mathbf{R}_w$ be the path set of all OD pairs. In addition, let $m = |\mathbf{R}|$ denote the total number of paths. Let $\boldsymbol{\lambda} = (\lambda_1,...,\lambda_r)^{\mathrm{T}}$ be the vector of mean path flows, which the mean OD flows are the summation of corresponding mean path flows: $\mathrm{E}[t_w(\boldsymbol{\alpha})] = \sum_{r \in \mathbf{R}_w} \lambda_r$. Let $p_{r,w}$ be the proportion of traffic flows on path r in the demand of OD pair w, and let $\mathbf{P}$ be the path-OD proportion matrix (i.e. $\mathbf{P} = (p_{r,w})$). The proportion $p_{r,w} = \lambda_r / \mathrm{E}[t_w(\boldsymbol{\alpha})]$ for $r \in \mathbf{R}_w$ is normally obtained from traffic assignment model such as stochastic traffic assignment (Clark and Watling, 2002; Watling, 2006).

Let $\mathbf{f} = (f_1,...,f_m)^{\mathrm{T}}$ be the random vector of realized path flows. After assigning the proportion of OD demand on the paths, path flows are the deterministic function of OD demand. Also, the distribution of path flows based on normal approximation of travel demands can be formulated by the follows:

$$\mathbf{f} = \mathbf{P} \cdot \mathbf{t} \sim \mathrm{N}(\mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})] \,,\, \mathbf{P} \cdot \boldsymbol{\delta} \cdot \Sigma_u \cdot \boldsymbol{\delta}^{\mathrm{T}} \cdot \mathbf{P}^{\mathrm{T}}) \tag{3.4}$$

where the mean path flows: $\boldsymbol{\lambda} = \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})] = \mathbf{P} \cdot \mathrm{E}[\mathbf{t}(\boldsymbol{\alpha})]$ by the definition. (3.4a)

In this study, two sources of traffic measurements are used to calibrate ABM parameters. The first is traffic counts of which link counting locations are located on every plate scanning locations to calculate the identifiability of each sensor. If there are n scanned links in the network, a vector of link counts $\mathbf{v} = (v_1,...,v_n)^{\mathrm{T}}$ can be observed and related to all path flows by:

$$\mathbf{v} = \boldsymbol{\Delta} \cdot \mathbf{f} \tag{3.5}$$

where $\boldsymbol{\Delta} = (\Delta_{i,j})$ is the link-path incidence matrix for the scanned links only, defined by:

$$\Delta_{i,j} = \begin{cases} 1 & \text{path } j \text{ contains scanned link } i \\ 0 & \text{otherwise.} \end{cases}$$

Based on the normal approximation of the demand, the distribution of link counts is given by:

$$\mathbf{v} \sim \mathrm{N}(\boldsymbol{\Delta} \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})] \,,\, \boldsymbol{\Delta} \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \Sigma_u \cdot \boldsymbol{\delta}^{\mathrm{T}} \cdot \mathbf{P}^{\mathrm{T}} \cdot \boldsymbol{\Delta}^{\mathrm{T}}) \tag{3.6}$$

where the mean link counts: $\mathrm{E}[\mathbf{v}] = \boldsymbol{\Delta} \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})] = \boldsymbol{\Delta} \cdot \boldsymbol{\lambda}$ by the definition. (3.6a)

The second source of traffic observations is traffic plate scanning information. Let Q and S denote the random scanned sensor path set for a given user (i.e. set of scanned links in (2.3)) and R be the random path set, which the user made. Also, let q and r be the realizations of Q and R, respectively. Note that the (true) sensor path set s represents the scanned sensor path in which all the sensors along the given path set r can identify that user.

The probability of observing on scanned sensor path set q, given that true sensor path set made by the user is s, contributes to the possibility that the identification of that user only occurs at the sensors in sensor path set q and that user can not be identified from the

remaining sensors. If the identification rate on any sensor/scanned link i (of sensor path set q ($\mathbf{C}_q$)) is observed to be equal to θ_i, the proportion (that the user observed by scanned sensor path set q travels on true sensor path s ($\mathbf{C}_s$)) is generally defined by:

$$\Pr(Q = q | S = s) = \prod_{i \in \mathbf{C}_q} \theta_i \cdot \prod_{j \in \mathbf{C}_s \cap j \notin \mathbf{C}_q} (1 - \theta_j) \tag{3.7a}$$

Then, let N_q and N_s be, respectively, the number of sensors in scanned sensor path set q and true sensor path set s. If the identification rate on all scanned links is assumed to be equal to θ for convenient illustrations (e.g. Castillo et al., 2008b), the proportion of the user, travelling on true sensor path s and observed by scanned sensor path set q, is defined by:

$$\Pr(Q = q | S = s) = \theta^{N_q} (1 - \theta)^{N_s - N_q} \tag{3.7b}$$

In addition, let $\omega_{s,r}$ be the true sensor path–path incident representing that a mapping relation between true sensor path (s) and travel path (r) where:

$$\omega_{s,r} = \begin{cases} 1 & \text{true sensor path } s \text{ contains all sensors located along path } r \\ 0 & \text{otherwise.} \end{cases} \tag{3.7c}$$

Then, the proportion that the user, travelling on path r and observed by scanned sensor path set q, can accordingly be defined by:

$$\Pr(Q = q | R = r) = \Pr(Q = q | S = s) \cdot \omega_{s,r} \tag{3.7d}$$

It is worth nothing that if sensors are located on a network where any true sensor path is uniquely mapped to travel path, a path is uniquely mapped to a true sensor path (e.g. illustrative network in Figure1). Regarding a sensor location scheme for reconstructing travel paths from sensor paths (sensor location model, M_1, in Mínguez et al. (2010)), the value of $\Pr(Q = q | R = r)$ is equal to one while path r is uniquely mapped to scanned sensor path q under the condition that each sensor has full identifiability ($\theta = 100\%$). Due to such full identifiability condition, any scanned sensor path is identical to true sensor path ($q=s$).

Let $\Psi = (\psi_{q,r} = \Pr(Q = q | R = r))$ be the scanned sensor path-path choice proportion matrix, where the elements ($\psi_{q,r}$) in the matrix Ψ are obtained from (3.7d) above. It follows that we can obtain the scanned sensor path flows, $\mathbf{g} = (g_1, ..., g_z)^{\mathrm{T}}$, and let $z = |\mathbf{g}|$ denote the total numbers of scanned sensor paths. Regarding the plate scanning observations, the condition distributions of scanned sensor path flows are multinomial as follows:

$$\mathbf{g} | \mathbf{f} \sim \text{Multinomial}(\mathbf{f}, \Psi) \tag{3.8a}$$

where the proportion of each sensor paths (Ψ) is defined in (3.7d) (e.g. Kwon and Varaiya, 2005; Asakura et al., 2000; Van der Zijpp, 1997). The mean of scanned sensor path flows is then defined by:

Mean of $\mathbf{g}$: $\text{E}[\mathbf{g}] = \Psi \cdot \text{E}[\mathbf{f}] = \Psi \cdot \boldsymbol{\lambda}$ (3.8b)

In addition, the variance-covariance matrix of sensor path flows (Σ_g) is expressed in (3.8c)-(3.8d) as follows:

Variance of $\mathbf{g}$: $\text{Var}(g_q, g_q) = \sum_r \psi_{q,r} \cdot f_r \cdot (1 - \psi_{q,r})$ for $q = 1, ..., |\mathbf{g}|$ (3.8c)

Covariance of $\mathbf{g}$: $\mathrm{Cov}(g_q, g_{q'}) = -\sum_r \psi_{q,r} \cdot f_r \cdot \psi_{q',r}$ for $q \neq q'$ (3.8d)

Following a normal approximation in the demand side and sensor path flows based on the central limit theorem for large population, the approximated distribution of sensor path flows can also be expressed by:

$$\mathbf{g} = \Psi \cdot \mathbf{f} \sim \mathrm{N}(\mathrm{E}[\mathbf{g}], \Sigma_g)$$ (3.9)

where the mean sensor path flows: $\mathrm{E}[\mathbf{g}] = \Psi \cdot \mathbf{P} \cdot \delta \cdot \mathrm{E}[\mathbf{u}(\alpha)] = \Psi \cdot \mathrm{E}[\mathbf{f}] = \Psi \cdot \lambda$ by the definition, and variance-covariance of scanned sensor path flows, Σ_g, is described in (3.8c)-(3.8d).

The link flows vector ($\mathbf{v}$) and sensor path flows vector ($\mathbf{g}$) are not statistically independent, since they are observed contemporaneously. However, these two types of traffic observations can be separated by decomposing the link counts as $\mathbf{v} = \mathbf{v}_{trk} + \mathbf{v}_{not}$ where $\mathbf{v}_{trk} = \mathbf{H} \cdot \mathbf{g}$ is the contribution to the link counts from scanned sensor path flows and $\mathbf{H} = (h_{i,j})$ is the link-sensor path incidence matrix for the scanned links only, defined by:

$$h_{i,j} = \begin{cases} 1 & \text{sensor path } j \text{ contains scanned link } i \\ 0 & \text{otherwise.} \end{cases}$$

, and hence $\mathbf{v}_{not}$ is the contribution from those vehicles, which are not scanned:

$$\mathbf{v}_{not} = \mathbf{v} - \mathbf{v}_{trk}$$ (3.10a)

After replacing the definition of $\mathbf{v}$ and $\mathbf{v}_{trk}$ into (3.10a), it yields:

$$\mathbf{v}_{not} = \mathbf{v} - \mathbf{H} \cdot \mathbf{g} = \Delta \cdot \mathbf{f} - \mathbf{H} \cdot \Psi \cdot \mathbf{f} = (\Delta - \mathbf{H} \cdot \Psi) \cdot \mathbf{f}$$ (3.10b)

The vector $\mathbf{g}$ and $\mathbf{v}_{not}$ are independent under standard assumptions (e.g. Parry and Hazelton, 2013). The distribution of $\mathbf{g}$ is given by (3.9), while $\mathbf{v}_{not}$ corresponding non-scanned link flows is distributed by:

$$\mathbf{v}_{not} \sim \mathrm{N}((\Delta - \mathbf{H} \cdot \Psi) \cdot \lambda , (\Delta - \mathbf{H} \cdot \Psi)\mathbf{P}\delta \cdot \Sigma_u \cdot \delta^{\mathrm{T}}\mathbf{P}^{\mathrm{T}}(\Delta - \mathbf{H} \cdot \Psi)^{\mathrm{T}})$$ (3.11)

Then, with observed traffic flows described above ($\mathbf{v}_{not}$ and $\mathbf{g}$), the ABM parameters (α) are calibrated by using maximum-likelihood estimation approach (see Appendix A for more details).

4. Numerical example

This numerical example adopts a simplified version of the activity-based model framework originally proposed by Bowman et al. (1998, 2001) based on a random utility maximization. In this study, travellers are facing two major choices: i) Daily activity pattern choice (e.g. H-W-H), and ii) Tour choice. The travellers' tour choices consist of a mandatory first tour and an optional second tour of which each of these tours contains the time-of-day, destination and mode choices. The choice dimensions for this application are:

- daily activity pattern choice;
- tour choice(trip chain model), consisting of:
  (a) first tour (mandatory):

(i) time-of-day choice[3];
(ii) destination and mode choice
  (b) second tour (optional)
(i) time-of-day choice;
(ii) destination and mode choice.

Based on this framework, demand model for generating link counts and plate scanning information in the test network is described (see Section 4.1). With the generated information from Section 4.1, the proposed ABM calibration model is adopted to estimate the required parameters (Section 4.2) and the results will be analysed in Section 4.3.

4.1 Model specification

For illustration purpose, a single category of population (i.e. workers) with *home-based tour* travel pattern and no activity-travel interactions of members in household are considered in this numerical example. Note that the extensions of the demand model (e.g. representing activity-travel interaction of members in household (Bradley and Vovsha, 2005)) can also be adopted in this illustrative demand model framework for future studies. The choices for out-of-home activities are work (W) or maintenance (O) (e.g. shopping), while stay-at-home (H) is the only at-home activities considered. In this numerical example, the following four choice models are considered.

Daily activity pattern choice model (Ap): This model reproduces the choice of daily activity pattern, y, for each origin zone o. In this numerical example, four different activity patterns (i.e. $y \in \{1, 2, 3, 4\}$) are considered: H-W-H ($y = 1$), H-O-W-H ($y = 2$), H-W-O-H ($y = 3$), and H-W-H-O-H, ($y = 4$).

First tour time-of-day choice model (Ftod): This model reproduces the choice of the time-of-day for the first tour (t_1). In this numerical example, three different choices of t_1 are considered ($t_1 \in \{1, 2, 3\}$) and the details could be found in Table 2. For example of a traveler performing an activity pattern of H-W-H associated with a time-of-day choice (t_1=1), this traveler starts the trip from home to work during AM peak and returns home during MD period.

Second tour time-of-day choice model (Stod): This model reproduces the choice of the time-of-day t_2 in the second tour, which is only applicable to the activity pattern H-W-H-O-H. The choice set of this second tour dimension is considered as a function of the time constraints of the first tour (e.g. the second tour cannot start until the first tour has ended). For instance, if travellers perform time-of-day choice in first tour (t_1=3), there is only one time of day in second tour (t_2=3) possibly scheduled.

Table 2
Time-of-day alternatives (first and second tour)

t_1	First tour		t_2	Second tour	
	Start	End		Start	End
1	AM[a]	MD[a]	1	PM[a]	PM
2	AM	PM	2	PM	OP
3	AM	OP[a]	3	OP	OP

[a] AM = 7:00-9:00, MD = 12:30-14:30, PM = 17:30-19:30, and OP = 20:00 – 22:00.

---

[3] Time-of-day choice is a combined choice of start tour and end tour travel period.

Destination and mode choice model for the first tour (Fdm) *and second tour* (Sdm): These models reproduce the followings: i) Choice of the destination zone, *lo*, for work purpose in the first tour and maintenance purpose in second tour, and ii) Travel mode, *mo*. In this numerical example, the three nearest zones to the origin (but not including it) are considered to be the choices of *lo* (e.g. $lo = 1$ for the nearest zone), while car and bus are taken as the choices for the travel model (*mo*). Consequently, the combined choice of mode and destination for first tour (b_1) or second tour (b_2) can be considered with 6 alternatives. Note that the destination for maintenance purposes before/after work in activity pattern (H-O-W-H and H-W-O-H) is assumed to be located at the same work zone centroid for illustration purpose.

Test demand model has 2 levels of decisions (activity pattern choice at higher tier and trip chain/tour choice at lower tier). Activity pattern is normally dependent on time of day of performing activities at lower tier (Bowman and Ben-Akiva, 2001). In addition, time of the day in second tour (t_2) is temporally constrained by first tour (t_1). Based on these relations, two levels of nests under activity pattern (nests of t_1, t_2) are constructed. Without loss of generality, the probability that person type p originating from zone o performs activity chain j is expressed by:

$$\Pr_j^{o,p}(\boldsymbol{\alpha}) = \Pr(b_2 \mid y, t_1, b_1, t_2)\Pr(t_2 \mid y, t_1, b_1)\Pr(b_1 \mid y, t_1)\Pr(t_1 \mid y)\Pr(y) \tag{4.1}$$

where the probability of selecting daily activity pattern y, $\Pr(y)$, and the corresponding utilities are defined as follows:

$$\Pr(y) = \frac{\exp(V_y + V_y')}{\sum_{y'} \exp(V_{y'} + V_{y'}')} \tag{4.2a}$$

$$V_y' = \frac{1}{\mu^{t_1}} \ln \sum_{t_1'} \exp[(V_{t_1'} + V_{y,t_1'}')\mu^{t_1}] \tag{4.2b}$$

$$V_y = \mathrm{Asc}_1(y) + \alpha_6 X_6 + \alpha_7 X_7 + \alpha_8 X_8 \tag{4.2c}$$

where μ^{t_1} is the scaling parameter at time of day t_1, X_k and Asc_1 are respectively the kth model's attribute and alternative specific constant of the choice model (see Table 4).

The probability of selecting time-of-day t_1 in the first tour, $\Pr(t_1 \mid y)$, and the corresponding utility functions are defined as follows:

$$\Pr(t_1 \mid y) = \frac{\exp(V_{t_1} + V_{y,t_1}')}{\sum_{t_1'} \exp(V_{t_1'} + V_{y,t_1'}')} \tag{4.3a}$$

$$V_{y,t_1'}' = \frac{1}{\mu^{t_2}} \ln \sum_{t_2'} \exp[(V_{t_2'})\mu^{t_2}] \tag{4.3b}$$

$$V_{t_1} = \mathrm{Asc}_2(t_1) + \alpha_{11} X_{11} + \alpha_{12} X_{12} \tag{4.3c}$$

where μ^{t_2} is the scaling parameter at time of day in second tour, t_2. The probability of selecting destination/mode choice b_1 in the first tour, $\Pr(b_1 \mid y, t_1)$, and the corresponding utility function are defined as follows:

$$\Pr(b_1|y,t_1) = \frac{\exp(V_{b_1}+V'_{b_1})}{\sum_{b'_1}\exp(V_{b'_1}+V'_{b'_1})} \tag{4.4a}$$

$$V'_{b_1} = \frac{1}{\mu^{b_1}}\ln\sum_{b'_2}\exp[(V_{b'_2})\mu^{b_1}] \tag{4.4b}$$

$$V_{b_1} = \mathrm{Asc}_3(b_1)+\alpha_{14}TT_{14}+\alpha_{15}Emp+\alpha_{16}X_{16} \tag{4.4c}$$

where μ^{b_1} is the scaling parameter at destination and mode in first tour, b_1, and *Emp* is the number of employment in log scale. The probability of selecting time-of-day t_2 in the second tour, $\Pr(t_2|y,t_1,b_1)$, and the corresponding utility function are defined as follows:

$$\Pr(t_2|y,t_1,b_1) = \frac{\exp(V_{t_2})}{\sum_{t'_2}\exp(V_{t'_2})} \tag{4.5a}$$

$$V_{t_2} = \mathrm{Asc}_4(t_2)+\alpha_{19}X_{19}+\alpha_{20}X_{20} \tag{4.5b}$$

The probability of selecting destination/mode choice b_2 in the second tour, $\Pr(b_2|y,t_1,b_1,t_2)$, and the corresponding utility are defined as follows:

$$\Pr(b_2|y,t_1,b_1,t_2) = \frac{\exp(V_{b_2})}{\sum_{b'_2}\exp(V_{b'_2})} \tag{4.6a}$$

$$V_{b_2} = \mathrm{Asc}_5(b_2)+\alpha_{22}TT+\alpha_{23}Retail+\alpha_{24}X_{24} \tag{4.6b}$$

where *Retail* is the number of retails in log scale.

In this study, a test network with 5 traffic zones (including activity H, W, and O in each zone), 38 links, and 13 nodes (Figure 2) is adopted. The true model parameters (α^{true}) are defined Table 3. The given number of population in each traffic zones, which is categorized by personal and household characteristics, is presented in Table 4 and Figure 3.

As mentioned in the beginning of this section, car and bus are the 2 travel modes considered in this network. Feasible paths travelled by both car and bus are assumed to be generated by the k-shortest path method (Yen, 1971). Travels demands (OD flows) of each travel periods (AM, MD, PM, or OP) are obtained through travel demand model (3.1b) derived from the true model parameters (see Table 3) and assigned to the network in a logit-based stochastic user equilibrium (SUE) manner. The path choice proportion, $p_{r,w}$, is obtained from the traffic assignment based on the given route choice parameters and the OD flows derived from true ABM parameter α^{true} in (3.3b) (e.g. Cascetta and Russo, 1997). In addition, the mean path flows expressed in (3.4a) were simultaneously calculated by multiplying the OD flows to the network with known path choice proportions as mentioned above.

To define the location of traffic observations for ABM parameter calibrations, optimal 15 sensor locations are obtained from design of optimal plate scanning location (i.e. model M_1 in Mínguez et al., 2010). To determine the identification rate (θ) at each sensors, link counting locations are also located on plate scanning locations. With these 15 sensor locations (Figure 2), all 54 paths derived from 18 OD pairs can be uniquely mapped to the true sensor paths (see Table 5). This implies that, under the full identifiability condition ($\theta = 100\%$), the summation

of all flows following any true sensor path is equal to the true flows of the path uniquely mapped. For instance, true sensor path (2, 20) is uniquely mapped path no.1. In other words, path no.1 can be reconstructed from true sensor path (2, 20). As a result, the flows on path no.1 are equal to the observed flows on this true sensor path.

Table 3
Setting of true model parameters (α^{true}) and model's attributes (X)

No.	Model	Variable name	Type of variable	Coeff. (α^{true})
1		H-O-W-H specific const. (Asc$_1$ of H-W-H = 0)	Asc$_1$	$\alpha_1^{true} = -2.73$
2		H-W-O-H specific const.	Asc$_1$	$\alpha_2^{true} = -0.73$
3		H-W-H-O-H specific const.	Asc$_1$	$\alpha_3^{true} = 0.91$
4	Activity pattern (Ap)	Scaled parameter (μ^{t1})	μ^{t1}	$\alpha_4^{true} = 2.12$
5		Scaled parameter (μ^{t2})	μ^{t2}	$\alpha_5^{true} = 1.00$
6		Dummy[a]: Male + H-W-H	$X_6 = 1$ or 0	$\alpha_6^{true} = 1.04$
7		Dummy: Family with at least one child+ H-O-W-H	$X_7 = 1$ or 0	$\alpha_7^{true} = 3.05$
8		Dummy: Family without child + H-W-O-H	$X_8 = 1$ or 0	$\alpha_8^{true} = 0.78$
9	Time of day in first tour (Ftod)	AM to PM specific const. (Asc$_2$ of AM-MD = 0)	Asc$_2$	$\alpha_9^{true} = -0.27$
10		AM to OP specific const.	Asc$_2$	$\alpha_{10}^{true} = -0.87$
11		Dummy: Full time worker + tour time: AM to MD	$X_{11} = 1$ or 0	$\alpha_{11}^{true} = -1.99$
12		Dummy: Part-time worker + tour time: AM to OP	$X_{12} = 1$ or 0	$\alpha_{12}^{true} = 0.94$
13	Destination and mode in first tour (Fdm)	Bus specific const. (Asc$_3$ of car = 0)	Asc$_3$	$\alpha_{13}^{true} = 1.54$
14		Generalised travel time	TT	$\alpha_{14}^{true} = -0.04$
15		Number of employments (log scale)	Emp	$\alpha_{15}^{true} = 0.39$
16		Dummy: High household income + car	$X_{16} = 1$ or 0	$\alpha_{16}^{true} = 2.68$
17	Time of day in second tour (Stod)	PM to OP specific const. (Asc$_4$ of PM-PM = 0)	Asc$_4$	$\alpha_{17}^{true} = -0.53$
18		OP to OP specific const.	Asc$_4$	$\alpha_{18}^{true} = -1.79$
19		Dummy: Full time worker + tour time: PM to PM	$X_{19} = 1$ or 0	$\alpha_{19}^{true} = -2.14$
20		Dummy: Part-time worker + tour time: OP to OP	$X_{20} = 1$ or 0	$\alpha_{20}^{true} = 1.25$
21	Destination and mode in second tour (Sdm)	Bus specific const. (Asc$_5$ of car = 0)	Asc$_5$	$\alpha_{21}^{true} = -0.34$
22		Generalised travel time	TT	$\alpha_{22}^{true} = -0.10$
23		Number of retails (log scale)	Retail	$\alpha_{23}^{true} = 0.49$
24		Dummy: High household income + car	$X_{24} = 1$ or 0	$\alpha_{24}^{true} = 3.37$
25		Scaled parameter (μ^{b1})	μ^{b1}	$\alpha_{25}^{true} = 0.50$

[a] Dummy variable (X) = 1 if specific demographic/choice is selected, 0 otherwise.
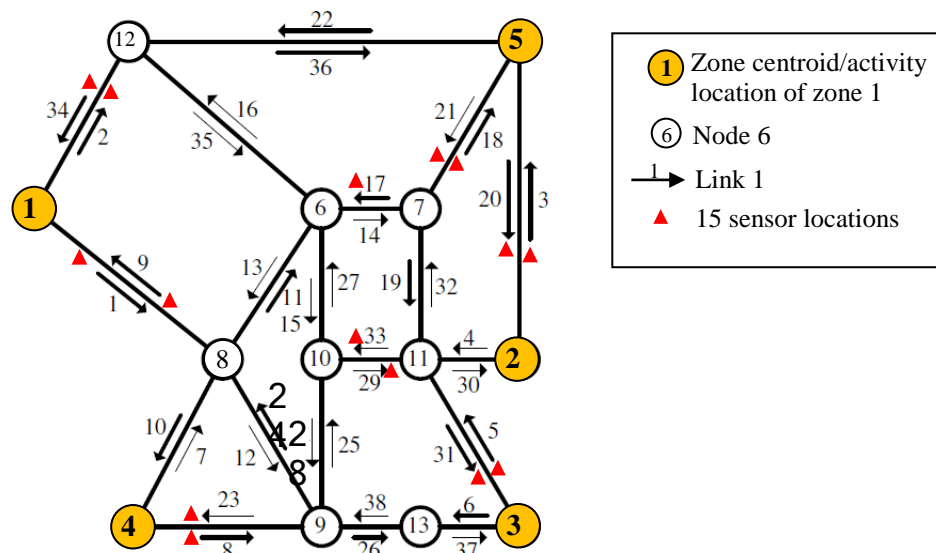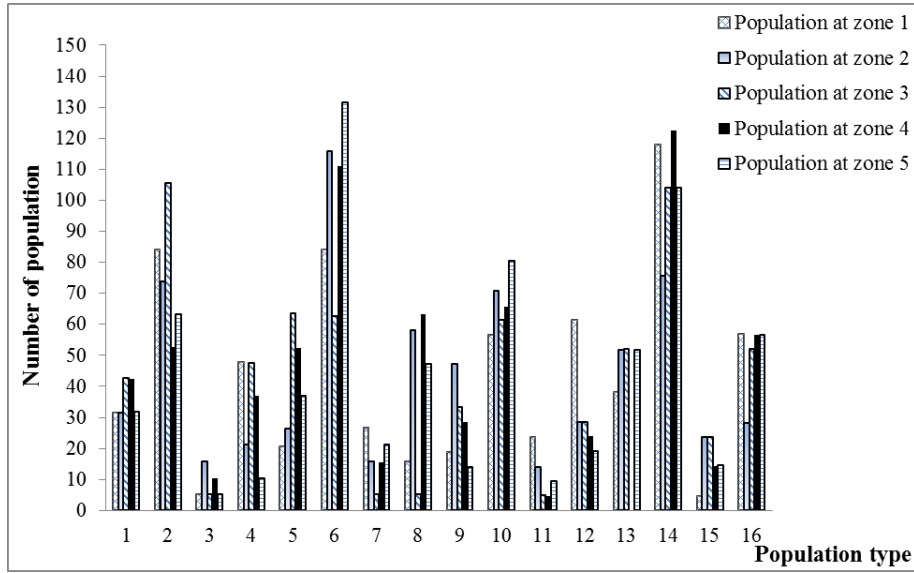
Fig. 2. Test network

Table 4
16 population types

Type (p)	Gender	Family with child	Household income	Career		Type (p)	Gender	Family with child	Household income	Career
1	Female	No	Low	Part-time[a]		9	Male	No	Low	Part-time
2	Female	No	Low	Full-time[b]		10	Male	No	Low	Full-time
3	Female	No	High	Part-time		11	Male	No	High	Part-time
4	Female	No	High	Full-time		12	Male	No	High	Full-time
5	Female	Yes	Low	Part-time		13	Male	Yes	Low	Part-time
6	Female	Yes	Low	Full-time		14	Male	Yes	Low	Full-time
7	Female	Yes	High	Part-time		15	Male	Yes	High	Part-time
8	Female	Yes	High	Full-time		16	Male	Yes	High	Full-time

[a] part-time worker

[b] full-time worker

Fig. 3. Number of population, $E[n_o^p]$ (by population type, p, and origin zone, o)

4.2 Evaluation method and test case

To evaluate the performance of the proposed model calibration, prior model parameters, $\boldsymbol{\alpha}^0$, were set to deviate from true model parameters, $\boldsymbol{\alpha}^{true}$, for representing a real application. The three settings of prior model parameters, $\boldsymbol{\alpha}^0$, were considered as follows:

- *Amplification*: all true model parameters are multiplied by an amplified factor (β_1), to represent that the amplitude of (prior) ABM parameters is over-estimated from travel diary samples, i.e. $\boldsymbol{\alpha}^0 = \beta_1 \boldsymbol{\alpha}^{true}; \beta_1 > 1$.

- *Random without bias*: all true model parameters are independently drawn from normal distribution with mean $\boldsymbol{\alpha}^{true}$ and variance $\beta_2 \boldsymbol{\alpha}^{true}$, i.e. $\boldsymbol{\alpha}^0 \sim N(\boldsymbol{\alpha}^{true}, \beta_2 \boldsymbol{\alpha}^{true})$.

- *Random with bias*: all true model parameters are independently drawn from normal distribution with mean $\beta_3\alpha^{true}$ and variance $\beta_4\alpha^{true}$, i.e. $\alpha^0 \sim N(\beta_3\alpha^{true}, \beta_4\alpha^{true})$; $\beta_3 \neq 1$.

According to the simulation of traffic, a Monte-Carlo method was used to draw traffic flows from travel demand $E[\mathbf{u}(\alpha)]$, a function of true model parameters, α^{true} shown in (3.1b) assigning to test network with SUE manner. A simulation with 10 simulations (replications)

Table 5

The set of scanned links (from 15 sensors) and its combination in order to identify all the paths[a]

od pair	o	d	Path (r)	Scanned link no.[b]														
				1	2	3	5	8	9	17	18	20	21	23	29	31	33	34
1	1	2	1		X							X						
1	1	2	2		X										X			
1	1	2	3	X											X			
2	1	3	4		X										X	X		
2	1	3	5	X											X	X		
2	1	3	6		X											X		
3	1	4	7	X														
3	1	4	8	X										X				
3	1	4	9		X									X				
4	1	5	10		X													
4	1	5	11		X						X							
4	1	5	12	X							X							
5	2	1	13			X												X
5	2	1	14														X	X
5	2	1	15						X									X
6	2	3	16													X		
6	2	3	17			X							X			X		
6	2	3	18														X	
7	2	5	19			X												
7	2	5	20								X							
7	2	5	21								X					X		
8	3	1	22					X									X	X
8	3	1	23					X		X								X
8	3	1	24					X	X							X		
9	3	2	25					X										
9	3	2	26					X			X	X						
9	3	2	27											X				
10	3	4	28					X						X		X		
10	3	4	29											X				
10	3	4	30					X		X				X				
11	3	5	31					X			X							
11	3	5	32			X		X										
11	3	5	33					X			X					X		
12	4	1	34						X									
12	4	1	35					X	X									
12	4	1	36					X										X
13	4	3	37					X							X	X		
13	4	3	38					X										
13	4	3	39					X								X		
14	4	5	40					X				X			X			
14	4	5	41					X				X						
14	4	5	42				X	X							X			
15	5	1	43															X
15	5	1	44							X			X					X
15	5	1	45					X	X				X					
16	5	2	46									X						
16	5	2	47										X					
16	5	2	48						X				X		X			
17	5	3	49										X			X		

17	5	3	50					X			X	
17	5	3	51		X			X		X	X	
18	5	4	52					X	X			X
18	5	4	53		X			X	X			
18	5	4	54				X	X				X

[a] Identification rate of each sensor is assumed to be 100% (θ=100%).
[b] "X" represents the existence of sensors on the link.

were made (N=10). In each simulation (replication), scanned sensor path flows ($\hat{\mathbf{g}}$) are drawn from a multinomial variate in (3.8a) and link counts ($\hat{\mathbf{v}}_{not}$) are drawn from a normal variate in (3.11). Similarly, prior $\boldsymbol{\alpha}^0$ is simulated from the three settings above.

To simulate the variations of observations in multiple days, link flows and sensor path flows are drawn 10 times to represent the observations of link counts and sensor path flows in 10 different days. With such multiday link flows and sensor path flows, the value of $\bar{a}$ in (3.22) was set as 10 (days). In order to illustrate the efficiency of the proposed method (Section 3) in calibrating the ABM parameters, the following three scenarios of collected data are setup and the corresponding ABMs are calibrated:

Scenario I – Use both link count and sensor path flow information (Lc+Trk)
Scenario II – Use only link count information (Lc only)
Scenario III – Use only sensor path flow information (Trk only)

In general, the accuracy of model calibration depends on the quality (identification rate) and quantity (number of sensors) of observations. For instance, low identification rate will lead to error in ABM calibrations. On the other hand, the lack of sensors to uniquely identify the chosen paths can induce errors in calibrating the ABM parameters. In order to study the aforementioned impacts, the sensitivity analysis of the calibration accuracy with respect to the identification rate (i.e. by taking θ = 100%, 90%, 70%, and 40%) and the number of sensors is carried out based on the settings in Scenario I~III.

Let $\boldsymbol{\alpha}^0$ be the vector of the initial (prior) values of the model parameters, $\boldsymbol{\alpha}^*$ be the vector of the calibrated values of the model parameters in scenario I, II, or III, and $\boldsymbol{\alpha}^{true}$ be the vector of the true model parameters. The statistical performance of the estimated k^{th} parameter in the activity-based model (section 4.1), α_k, can be measured by the mean squared errors (MSE) and percentage error reduction (MSE%) as follows:

$$\text{MSE}(\alpha^0) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{(\alpha_k^{0,(i)} - \alpha_k^{true})^2}{K} \qquad (4.7)$$

$$\text{MSE}(\alpha^*) = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} \frac{(\alpha_k^{*,(i)} - \alpha_k^{true})^2}{K} \qquad (4.8)$$

where N is the number of datasets (replications) of traffic observations and prior $\boldsymbol{\alpha}^0$, K is the number of parameters considered in sub ABM model or whole ABM model, α_k^{true} is the true value of kth parameter in $\boldsymbol{\alpha}^{true}$, $\alpha_k^{0,(i)}$ is initial value of kth prior parameter in i^{th} dataset, and $\alpha_k^{*,(i)}$ is the calibrated value of kth parameter in i^{th} dataset. To measure the amounts of error reduction from model calibrations (Scenario I ~ III) compared with error of initial values, the percentage error reduction of MSE, MSE%, is formulated as follows:

$$\text{MSE\%}(\alpha) = \frac{(\text{MSE}(\alpha^0) - \text{MSE}(\alpha^*))}{\text{MSE}(\alpha^0)} \times 100\% \qquad (4.9)$$

4.3 Analysis of the results

In the numerical example, the test network with 15 installed sensors (Figure 2) is adopted. The three scenarios (described in Section 4.2) of which the demand is derived from the true model parameters (Table 3) are setup and the corresponding ABM parameters are calibrated. Table 6 shows the calibrated parameters and the corresponding statistical performance of the proposed scenarios. Note that with 100% identification rate, the likelihood estimation of model calibrations from sensor path flows (scenario III) is identical to the likelihood estimation of model calibrations from sensor path flows and link counts (scenario I).

Table 6
Percentage error reduction (MSE%(α)) from model calibrations with simulated traffic data[a]

No.	Setting of prior ABM parameters	Choice model in ABM	Prior parameters	Scenario II: Link counts only		Scenario I[b]: Link counts and sensor path flows	
			MSE	MSE	MSE%	MSE	MSE%
1	Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)	Ap	2.017	1.444	28%	0.125	94%
2	Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)	Ftod	0.463	1.968	-325%	0.035	92%
3	Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)	Fdm	1.290	0.323	75%	0.015	99%
4	Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)	Stod	1.132	1.908	-69%	0.100	91%
5	Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)	Sdm	0.755	0.497	34%	0.251	67%
1	Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)	Ap	4.082	1.518	63%	0.009	100%
2	Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)	Ftod	2.288	1.009	56%	0.035	98%
3	Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)	Fdm	3.606	0.215	94%	0.015	100%
4	Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)	Stod	3.117	1.277	59%	0.100	97%
5	Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)	Sdm	3.461	0.900	74%	0.251	93%
1	Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)	Ap	2.786	1.712	39%	0.135	95%
2	Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)	Ftod	1.678	1.841	-10%	0.035	98%
3	Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)	Fdm	0.965	0.227	77%	0.015	98%
4	Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)	Stod	1.674	2.085	-25%	0.100	94%
5	Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)	Sdm	3.370	1.025	70%	1.001	70%
1	Set4:($\alpha^0 = 2.0\alpha^{true}$)	Ap	3.173	1.516	52%	0.007	100%
2	Set4:($\alpha^0 = 2.0\alpha^{true}$)	Ftod	1.431	0.983	31%	0.035	98%
3	Set4:($\alpha^0 = 2.0\alpha^{true}$)	Fdm	2.435	0.227	91%	0.015	99%
4	Set4:($\alpha^0 = 2.0\alpha^{true}$)	Stod	2.410	1.200	50%	0.100	96%
5	Set4:($\alpha^0 = 2.0\alpha^{true}$)	Sdm	2.935	1.017	65%	1.001	66%

[a] the number of sensors = 15.   [b] the identification rate (θ) = 100%.

The proposed ABM calibrations with link counts and sensor path flows (scenario I) reproduce the calibrated ABM parameters close to the true values for all settings of prior ABM parameters (Table 6). It could be seen that scenario I shows the smallest level of error (MSE) for all choice models in ABM. Also, the combined use of link counts and sensor path flows (scenario I) outperforms the sole use of link counts (scenario II) in all choice models or any settings of prior parameters (α^0). This can be described by the different number of known equations (traffic observations) used to calibrate ABM parameters in scenario I~III. In scenario II, the known equations are created from the link counts classified by travel periods and travel modes (15 sensors x 4 periods x 2 modes = 120 equations). In scenario III, the known equations are created from 87 scanned sensor paths derived from 15 sensors (87 sensor paths x 4 periods x 2 modes = 696 equations). Consequently, the 816 equations in scenario III, which consists 120

equations of link counts and 696 equations of scanned sensor path flows, are 6.8 times of number of equations in scenario II.

To compare the errors between calibrated parameters ($\boldsymbol{\alpha}^*$) and prior parameters ($\boldsymbol{\alpha}^0$), statistical performances of the model calibration in the three scenarios are studied by using percentage error reduction (MSE%). A 100% of MSE% indicates that the calibrated parameters are totally free from the errors incurred in the prior parameters (i.e. the true value of the parameters, $\boldsymbol{\alpha}^{true}$, is obtained after calibration). The results from Table 6 also show that high error reductions (>50%) could be obtained for estimating parameters of all choice models (scenario I).

Impact of identification rate

In following experiments, it is focused on the quality of plate scanning observations, which is determined by the identification rate of the sensor. Given 40% of links in the test network (Figure 2) observed (i.e. 15 sensors) with four different levels of identification rates (i.e. 100%, 90%, 70% and 40%), the errors of the calibrated parameters are shown in Table 7. The results in Table 7 show that the parameter calibration error, MSE, (percentage error reduction, MSE%) increases (decreases) with the decreasing identification rate in scenario I and III. With at least 40% of identification rate ($\theta \geq 40\%$), the errors of calibrated ABM parameters in scenario I are smaller than the errors of calibrated ABM parameters in scenario II and III for most settings of prior, $\boldsymbol{\alpha}^0$. In addition, with acceptable identification rate (e.g. $\theta \geq 70\%$), ABM calibration results in scenario I are more reliable and lead to a significant improvement in the model calibrations (MSE% $\geq 40\%$).

Table 7
Errors (MSE%(α)) of ABM calibration for different identification rates (θ)

Setting of prior ABM parameters (α^0)	Prior parameters	Scenario II: Link counts only		θ	Scenario I: Link counts and sensor path flows		Scenario III: sensor path flows only	
	MSE	MSE	MSE%		MSE	MSE%	MSE	MSE%
Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)				100%	0.234	82%	0.234	82%
Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)				90%	0.606	53%	0.764	40%
Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)	1.279	1.264	1%	70%	0.765	40%	0.872	32%
Set1:($\alpha^0 \sim N(\alpha^{true}, 0.8\alpha^{true})$)				40%	0.810	37%	0.949	26%
Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)				100%	0.195	94%	0.195	94%
Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)				90%	0.716	79%	0.757	78%
Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)	3.439	1.073	69%	70%	0.922	73%	0.943	73%
Set2:($\alpha^0 \sim N(2.0\alpha^{true}, 0.7\alpha^{true})$)				40%	1.164	66%	1.319	62%
Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)				100%	0.237	89%	0.237	89%
Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)				90%	0.940	57%	0.975	56%
Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)	2.210	1.369	38%	70%	1.303	41%	1.352	39%
Set3:($\alpha^0 \sim N(0.4\alpha^{true}, 0.7\alpha^{true})$)				40%	1.316	40%	1.424	36%
Set4:($\alpha^0 = 2.0\alpha^{true}$)				100%	0.194	93%	0.194	93%
Set4:($\alpha^0 = 2.0\alpha^{true}$)				90%	0.649	75%	0.655	75%
Set4:($\alpha^0 = 2.0\alpha^{true}$)	2.593	1.076	58%	70%	0.844	67%	0.847	67%
Set4:($\alpha^0 = 2.0\alpha^{true}$)				40%	0.965	63%	1.237	52%

Another improvement in model calibrations with traffic observations can be seen by calculating the mean path flows derived from the calibrated, $\boldsymbol{\alpha}^*$. The statistical performance of the estimated mean path flows on any path r can be measured by the MSE as follows:

$$\text{MSE}(\lambda_r^*) = \sum_{i=1}^{N} \frac{(\lambda_r^{*,(i)} - \lambda_r^{true})^2}{N} \tag{4.10a}$$

$$\text{MSE}(\lambda_r^0) = \sum_{i=1}^{N} \frac{(\lambda_r^{0,(i)} - \lambda_r^{true})^2}{N} \tag{4.10b}$$

where the mean path flows: $\boldsymbol{\lambda}^* = \mathbf{P} \cdot \boldsymbol{\delta} \cdot \text{E}[\mathbf{u}(\boldsymbol{\alpha}^*)]$, $\boldsymbol{\lambda}^0 = \mathbf{P} \cdot \boldsymbol{\delta} \cdot \text{E}[\mathbf{u}(\boldsymbol{\alpha}^0)]$, $\boldsymbol{\lambda}^{true} = \mathbf{P} \cdot \boldsymbol{\delta} \cdot \text{E}[\mathbf{u}(\boldsymbol{\alpha}^{true})]$.

As one of main aims for calibrating travel demand model is to validate the modeled traffic flows (derived from ABM) with observed traffic flows (traffic counts on screenline/cordonline) for obtaining a reliable vehicle distance traveled (VKT), the calibrated path flows (scenario I~II) are calculated to evaluate the performance of calibration results shown in Table 8.

The results (Table 8) show that the error (MSE) of mean path flows derived from calibrated $\boldsymbol{\lambda}^*$ from 15 sensors in scenario I~II with identification rate ($\theta = 100\%$ or 40%) is significantly smaller than the error of mean path flows derived from prior model parameters, $\boldsymbol{\alpha}^0$ (set1 in Table 6 or Table 7). Since we simultaneously calibrate all model parameters in the maximum likelihood problem (MLP in (A.7), (A.9), or (A.11)), there can possibly be at least one combination of calibration results that still fit to traffic flow observations (non-unique solution). For instance, with set of prior parameters no.1 (randomly deviated from true parameters with normal distribution assumed), percentage error reductions of calibrated parameters of time of day model in first tour (ftod) and second tour (stod) using link counts only are -325% and -69% (see Table 6). These errors would be occurred due to more than one combination of calibration results may possibly fit link counts and prior model parameters. To deal with this problem, all model parameters calibrated with sufficient plate scanning information (scenario I or III) are close to the true values (MSE% > 0) and the calibration results seem to be unique solution of the MLP.

For instance, the calibrated path flows with link count only (scenario II) can outperform the prior path flows (derived from prior model parameters, $\boldsymbol{\alpha}^0$), even though the calibrated model parameters (scenario II) is not much better than prior model parameters as compared to true model parameters (Table 7). In addition, the mean path flows of the calibrated $\boldsymbol{\lambda}^*$ under full identifiability ($\theta = 100\%$) tend to be closer to mean (true) path flows, $\boldsymbol{\lambda}^{true}$ (derived from $\boldsymbol{\alpha}^{true}$) than those path flows under partial identifiability ($\theta = 40\%$).

Impact of the number of sensors

In this experiment, a different number of sensors with a fixed identification rate (90%) are installed into the test network for collecting the link counts and sensor path flows information. With this setup, the ABM for Scenario I~III is calibrated in three levels of number of population (i.e. medium level: n_o^p; $\forall o, p$ shown in Figure 3, low level: $0.05n_o^p$, or high level: $50n_o^p$) and the errors of the calibrated parameters are shown in Table 9. In Table 9, it could be seen that the MSE (MSE%) increases (decreases) as the number of sensors is decreased from 15 (or 39% of links) to 5 (or 13% of links) for all levels of the number of populations. These trends indicate that the error of parameter calibrations increases while the number of sensors decreases.

This observation could be explained by the fact that the decreasing number of sensors will result in fewer number of traffic observations (number of known equations in ABM calibrations). With a low number of sensors, the ABM parameter calibrations are less accurate.

Table 8
The comparison of the mean path flows derived from the calibrated ABM parameters using 15 sensors (Figure 2) in scenario I~II with different of identification rates (θ=100% or 40%)

r	o	d	Sensor path topology	Path topology	True path flows	MSE of prior path flows	MSE of mean path flows[a]		
							Scenario II: Lc only	Scenario I: Lc+Trk θ=100%	θ=40%
1	1	2	(2,20)	(2,36,20)	140	3,010	2.78	0.002	1.59
2	1	2	(2,29)	(2,35,15,29,30)	100	1,600	1.33	0.003	0.86
3	1	2	(1,29)	(1,12,25,29,30)	53	441	0.40	0.0005	0.24
4	1	3	(2,29,31)	(2,35,15,29,31)	128	1,938	2.86	0.25	4.77
5	1	3	(1,29,31)	(1,12,25,29,31)	53	608	0.47	0.04	0.73
6	1	3	(2,31)	(2,35,14,19,31)	77	699	1.03	0.09	1.72
7	1	4	(1)	(1,10)	718	85,060	105.67	4.44	43.57
8	1	4	(1,23)	(1,12,23)	219	7,726	9.74	0.40	4.23
9	1	4	(2,23)	(2,35,15,28,23)	156	3,636	4.99	0.19	2.72
10	1	5	(2)	(2,36)	647	43,303	31.47	8.31	34.49
11	1	5	(2,18)	(2,35,14,18)	102	1,195	0.92	0.19	0.90
12	1	5	(1,18)	(1,11,14,18)	102	1,184	0.84	0.19	0.88
13	2	1	(3,34)	(3,22,34)	127	1,714	3.64	0.03	1.65
14	2	1	(33,34)	(4,33,27,16,34)	105	2,246	1.23	0.02	0.85
15	2	1	(17,34)	(4,32,17,16,34)	61	850	0.41	0.01	0.29
16	2	3	(31)	(4,31)	406	18,959	13.19	1.67	12.87
17	2	3	(3,21,31)	(3,21,19,31)	125	1,872	1.21	0.15	1.19
18	2	3	(33)	(4,33,28,26,37)	95	1,258	0.67	0.08	0.63
19	2	5	(3)	(3)	957	164,219	80.47	1.58	36.23
20	2	5	(18)	(4,32,18)	302	16,512	8.34	0.16	3.61
21	2	5	(33,18)	(4,33,27,14,18)	252	11,831	6.74	0.11	2.56
22	3	1	(5,33,34)	(5,33,27,16,34)	127	2,193	2.67	0.24	4.46
23	3	1	(5,17,34)	(5,32,17,16,34)	77	814	0.99	0.09	1.65
24	3	1	(5,33,9)	(5,33,28,24,9)	53	355	0.51	0.04	0.82
25	3	2	(5)	(5,30)	354	14,643	9.85	1.32	9.68
26	3	2	(5,18,20)	(5,32,18,20)	125	1,872	1.21	0.15	1.19
27	3	2	(29)	(6,38,25,29,30)	147	2,715	1.59	0.18	1.59
28	3	4	(5,33,23)	(5,33,28,23)	379	19,033	8.99	1.25	14.84
29	3	4	(23)	(6,38,23)	325	10,374	5.76	4.45	9.11
30	3	4	(5,17,23)	(5,32,17,15,28,23)	156	2,693	1.40	0.66	2.20
31	3	5	(5,18)	(5,32,18)	727	62,938	23.56	4.65	28.03
32	3	5	(5,3)	(5,30,3)	96	1,110	0.41	0.08	0.49
33	3	5	(5,33,18)	(5,33,27,14,18)	138	2,406	0.90	0.13	0.97
34	4	1	(9)	(7,9)	477	35,974	45.02	1.76	20.32
35	4	1	(8,9)	(8,24,9)	219	7,726	9.74	0.40	4.23
36	4	1	(8,34)	(8,25,27,16,34)	397	26,230	33.40	1.47	13.86
37	4	3	(8,29,31)	(8,25,29,31)	555	35,233	15.31	6.71	27.53
38	4	3	(8)	(8,26,37)	172	3,391	1.48	0.65	2.66
39	4	3	(8,31)	(8,25,27,14,19,31)	134	2,046	0.91	0.41	1.62
40	4	5	(8,29,18)	(8,25,29,32,18)	266	9,977	15.63	0.03	3.35
41	4	5	(8,18)	(8,25,27,14,18)	140	2,735	4.25	0.01	0.94
42	4	5	(8,29,3)	(8,25,29,30,3)	62	496	0.75	0.003	0.20
43	5	1	(34)	(22,34)	507	30,028	20.20	4.56	22.03
44	5	1	(21,17,34)	(21,17,16,34)	131	1,580	1.57	0.32	1.67
45	5	1	(21,17,9)	(21,17,13,9)	213	4,115	3.16	1.12	3.19
46	5	2	(20)	(20)	704	89,391	48.26	1.01	20.32
47	5	2	(21)	(21,19,30)	302	16,512	8.34	0.16	3.61
48	5	2	(21,17,29)	(21,17,15,29,30)	505	46,267	21.23	0.35	9.66
49	5	3	(21,31)	(21,19,31)	623	45,715	18.31	3.36	20.48
50	5	3	(20,31)	(20,4,31)	144	2,498	0.93	0.18	1.09
51	5	3	(21,17,29,31)	(21,17,15,29,31)	193	4,832	1.42	0.30	1.96
52	5	4	(21,33,23)	(21,19,33,28,23)	203	5,843	8.78	0.02	1.96
53	5	4	(21,17,23)	(21,17,15,28,23)	93	1,216	1.89	0.004	0.42

| 54 | 5 | 4 | (20,33,23) | (20,4,33,28,23) | 171 | 3,981 | 6.44 | 0.01 | 1.43 |
| **Total MSE** | | | | | | **866,791** | **603** | **54** | **394** |

[a] mean path flows are the summation of expected path flows of all travel patterns, modes, and travel periods.

This will result in, for example, an estimation of ABM parameters of scenario III (medium level of population) that causes MSE to increase from 0.764 for 15 sensors to 1.355 for 5 sensors. Similar interpretation could also be applied to the results of MSE%.

Table 9
Percentage error reductions (MSE%(α)) of ABM calibration for different number of sensors[a]

Number of sensors	Number of additional link counting stations	% of links to be observed	Scenario II: Link counts only		Scenario I: Link counts and sensor path flows		Scenario III: sensor path flows only	
			MSE	MSE%	MSE	MSE%	MSE	MSE%
Case I: Low number of population (total population = 175)								
5	-	13%	1.913	-26.5%	1.581	-4.5%	1.681	-11.1%
10	-	26%	1.866	-23.4%	1.520	-0.5%	1.613	-6.6%
15	-	39%	1.706	-12.7%	1.478	2.3%	1.571	-3.8%
15	5	53%	1.677	-10.8%				
15	10	66%	1.631	-7.8%				
15	15	79%	1.588	-5.0%				
15	20	92%	1.563	-3.3%				
15	23	100%	1.531	-1.2%				
Case II: Medium number of population (total population = 3,500)								
5	-	13%	1.465	-14.5%	1.294	-1.1%	1.355	-5.9%
10	-	26%	1.423	-11.2%	1.093	14.5%	1.150	10.1%
15	-	39%	1.264	1.2%	0.606	52.6%	0.764	40.3%
15	5	53%	1.228	4.0%				
15	10	66%	1.218	4.7%				
15	15	79%	1.166	8.9%				
15	20	92%	1.144	10.5%				
15	23	100%	1.133	11.4%				
Case III: High number of population (total population = 175,000)								
5	-	13%	1.162	23.5%	0.776	48.9%	0.801	47.3%
10	-	26%	0.987	35.0%	0.723	52.4%	0.784	48.4%
15	-	39%	0.894	41.2%	0.602	60.4%	0.640	57.9%
15	5	53%	0.815	46.4%				
15	10	66%	0.809	46.8%				
15	15	79%	0.808	46.8%				
15	20	92%	0.787	48.2%				
15	23	100%	0.779	48.7%				

[a] the identification rate (θ) = 90%.

In addition, the MSE% tends to be decreased while the number of population is decreased due to the normal approximation of traffic flows distributions (i.e. see (3.6), (3.9), and (3.11)) associated with case I: low number of population (the central limit theorem). Regarding the choice dimension in the ABM, it consists 306 choices of activity chains including 3 patterns with single tour (y =1,2,3) (3 ap x 3 tod x 6 dm) and the pattern with two tours (y=4) (1 ap x [2 tod (t_1=1,2) x 6 dm x 3 stod(t_2=1,2,3) x 6 sdm + 1 tod(t_1=3) x 6 dm x 1 stod(t_2=3) x 6 sdm]). Since there is sufficient number of population defined in case II~III to make a full representative image of the underlying activity chains (306 choices), the ABM calibrations with traffic data show a significant improvement (MSE%>0). As the cost of link counts observations per station is normally less than the cost of plate scanning observations, the ABM

calibrations with the additional sensors for link counts only (i.e. 5,10,15,20,or 23 additional link counting stations, see Table 9) are also conducted. However, it is clearly seen that with acceptable number of sensors (39% of links installed with 15 sensors), the MSE of ABM calibrations in scenario I for all three levels of populations is significantly less than the MSE in the ABM calibrations with link counts only.

5. Conclusions

A statistical method for calibrating parameters in activity-based models with link counts and number plate scanning information was presented in this paper. The proposed calibration approach was formulated as a maximum likelihood estimation method and solved by using sequential quadratic programming. Statistical performances (MSE and MSE%) of this method were evaluated on a test network with different settings in qualities and quantities of roadside observations simulated from travel demand and supply model as mentioned.

In addition, ideally, the household and travel survey (HTS) data would be collected for at least a week to represent the weekly variations of activity-travel patterns made by travellers. However, in practices, the household and travel survey data is repeatedly collected for only 2-3 consecutive days (including working and non-working day) due to budget constraints. To calibrate the ABM parameters by using traffic flows data (i.e. link counts and sensor path flows), traffic flows data for at least one week should be collected to deal with the day-to-day variations of activity-travel patterns. In this research, multiday traffic flow observations are assumed to follow normal distribution for illustration purpose (see section 3.2). To cover HTS survey durations (2-3 days) and day-to-day variations in a week, traffic flows were simulated for 10 days in this study. To consider the more variations in ABM calibrations, the ABM calibrations associated with week-to-week or month-to-month variations of activity-travel patterns would be conducted in the future study (e.g. Wolf et al., 2004).

The accuracy of the calibrated parameter is, in general, satisfactory, showing the capability of the proposed method to significantly reduce biases in the ABM parameters that are normally estimated from household and travel diary surveys (considered as the prior parameters in this study). Based on the synthetic data sets of the test network, parameters calibrated from number plate scanning information, with a sufficient identification rate and number of sensors, are found to be more accurate than the traditional calibration approaches that use only link counts. In particular, when there is only a small number of sensors (or observations), the calibrations based on plate scanning observations outperform its link counts counterparts with the same number of observation stations (sensors). This is described by larger number of known equations (traffic observations) in the likelihood estimation problems generated from plate scanning observations. In the numerical example, it is found that reduction in the identification rates or the number of observation stations (sensors) will decrease the accuracy of the ABM parameter calibrations. However, with acceptable number of sensors and identification rates, the model calibration results with plate scanning clearly show a good fit to the true model parameters and true path flows.

Appendix A

A.1 Maximum-likelihood estimation problem (MLP)

The ABM parameters (α) can be calibrated by maximizing the likelihood to reproduce observed traffic flows (i.e. link counts vector, $\hat{\mathbf{v}}_{not}$, and scanned sensor path flows vector, $\hat{\mathbf{g}}$). In addition, another objective of the MLP is to calibrate the ABM parameters in which the calibration results are improved from prior ABM parameters (α^0). Note that the prior parameters, α^0, which are assumed to be given for this study, can be obtained by another conventional maximum-likelihood estimation problem using household and travel diary survey data (e.g. Bowman et al., 2006). Since observed traffic data ($\hat{\mathbf{v}}_{not}$ and $\hat{\mathbf{g}}$) and prior ABM parameters (α^0) can generally be considered to be statistically independent (e.g. Cascetta and Russo, 1997) and the observed flows in vector $\hat{\mathbf{g}}$ and vector $\hat{\mathbf{v}}_{not}$ are independent based on the standard assumption (e.g. Parry and Hazelton, 2013), the likelihood estimation problem can be formulated as follows:

$$L(\hat{\mathbf{g}}, \hat{\mathbf{v}}_{not}, \alpha^0 | \alpha) = L(\alpha^0 | \alpha) \cdot L(\hat{\mathbf{v}}_{not} | \alpha) \cdot L(\hat{\mathbf{g}} | \alpha) \tag{A.1}$$

The maximum likelihood estimators of the parameter vector (α) can then be obtained by maximizing (A.1) or their natural logarithms:

$$\alpha^* = \underset{\alpha \in \Omega}{\text{argmax}} \ [\ln \ L(\alpha^0 | \alpha) + \ln \ L(\hat{\mathbf{v}}_{not} | \alpha) + \ln \ L(\hat{\mathbf{g}} | \alpha)] \tag{A.2}$$

where Ω is the feasible set (region) of α, $\alpha \in \Omega$. Since errors in traffic counts have a joint multivariate normal (MVN) distribution (shown in (3.11)), it follows that:

$$\ln L(\hat{\mathbf{v}}_{not} | \alpha) = \ln[\exp \ -\frac{1}{2}[\hat{\mathbf{v}}_{not} - (\Delta - \mathbf{H} \cdot \Psi) \cdot \lambda]^T \Sigma_{v_{not}}^{-1} [\hat{\mathbf{v}}_{not} - (\Delta - \mathbf{H} \cdot \Psi) \cdot \lambda]] + const. \tag{A.3}$$

After replacing the mean path flows: $\lambda = \mathbf{P} \cdot \delta \cdot \text{E}[\mathbf{u}(\alpha)]$ into (A.3), it yields:

$$\ln L(\hat{\mathbf{v}}_{not} | \alpha) = -\frac{1}{2}[\mathbf{v}_{not} - (\Delta - \mathbf{H} \cdot \Psi)\mathbf{P} \cdot \delta \cdot \text{E}[\mathbf{u}(\alpha)]]^T \Sigma_{v_{not}}^{-1} [\hat{\mathbf{v}}_{not} - (\Delta - \mathbf{H} \cdot \Psi)\mathbf{P} \cdot \delta \cdot \text{E}[\mathbf{u}(\alpha)]]$$
$$+ \ const. \tag{A.4}$$

Also, errors in scanned sensor path flows have a joint multivariate normal (MVN) distribution shown in (3.9), it follows that:

$$\ln L(\hat{\mathbf{g}} | \alpha) = \ln[\exp -\frac{1}{2}[\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \delta \cdot \text{E}[\mathbf{u}(\alpha)]]^T \Sigma_g^{-1} [\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \delta \cdot \text{E}[\mathbf{u}(\alpha)]]] + const.$$
$$= -\frac{1}{2}[\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \delta \cdot \text{E}[\mathbf{u}(\alpha)]]^T \Sigma_g^{-1} [\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \delta \cdot \text{E}[\mathbf{u}(\alpha)]] + const. \tag{A.5}$$

In general, values of α^0 in (A.1) are obtained through another MLP[4], they are asymptotically and normally distributed around the "true" vector α with dispersion matrix Σ_{α^0} (e.g. Ben-

---

[4] Estimation of ABM parameters (α) from travel diary sample (which the estimated parameters are considered to be the prior parameters (α^0) in this study) is shown as follows:

Akiva and Lerman, 1985; Cascetta and Russo, 1997). In this scenario, the log likelihood function for the prior parameters ($\boldsymbol{\alpha}^0$) can be expressed by:

$$
\begin{aligned}
\ln L(\boldsymbol{\alpha}^0 | \boldsymbol{\alpha}) &= \ln[\ \exp\ -\frac{1}{2}[\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}]^{\mathrm{T}} \Sigma_{\alpha^0}^{-1} [\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}]] + const. \\
&= -\frac{1}{2}[\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}]^{\mathrm{T}} \Sigma_{\alpha^0}^{-1} [\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}] + const.
\end{aligned}
\tag{A.6}
$$

After putting (A.4), (A.5), and (A.6) into (A.2), maximum-likelihood estimation problem (MLP), given $\boldsymbol{\alpha}^0$, $\hat{\mathbf{v}}_{not}$, and $\hat{\mathbf{g}}$, can then be formulated by:

$$
\begin{aligned}
\boldsymbol{\alpha}^*_{senarioI} = \underset{\boldsymbol{\alpha} \in \Omega}{\mathrm{argmax}}[&-\frac{1}{2}[\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}]^{\mathrm{T}} \Sigma_{\alpha^0}^{-1} [\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}] \\
&-\frac{1}{2}[\hat{\mathbf{v}}_{not} - (\boldsymbol{\Delta} - \mathbf{H} \cdot \Psi) \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]]^{\mathrm{T}} \Sigma_v^{-1} [\hat{\mathbf{v}}_{not} - (\boldsymbol{\Delta} - \mathbf{H} \cdot \Psi) \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]] \\
&-\frac{1}{2}[\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]]^{\mathrm{T}} \Sigma_g^{-1} [\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]]]
\end{aligned}
\tag{A.7}
$$

To evaluate the performance of the ABM calibration with two traffic sources (plate scanning flows and link counts), the conventional ABM calibration with single traffic source (link counts vector, $\hat{\mathbf{v}}$) is also conducted. The likelihood estimation problem with observed link flows only (scenario II) can be formulated as follows:

$$
L(\hat{\mathbf{v}}, \boldsymbol{\alpha}^0 | \boldsymbol{\alpha}) = L(\boldsymbol{\alpha}^0 | \boldsymbol{\alpha}) \cdot L(\mathbf{v} | \boldsymbol{\alpha})
\tag{A.8}
$$

Similar to the ABM calibration with two traffic sources as mentioned ((A.1) – (A.7)), the maximum likelihood estimators of the parameter vector ($\boldsymbol{\alpha}$) under the scenario II, that only link counts are used to calibrate, can be obtained by:

$$
\begin{aligned}
\boldsymbol{\alpha}^*_{senarioII} = \underset{\boldsymbol{\alpha} \in \Omega}{\mathrm{argmax}}[&-\frac{1}{2}[\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}]^{\mathrm{T}} \Sigma_{\alpha^0}^{-1} [\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}] \\
&-\frac{1}{2}[\hat{\mathbf{v}} - \boldsymbol{\Delta} \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]]^{\mathrm{T}} \Sigma_v^{-1} [\hat{\mathbf{v}} - \boldsymbol{\Delta} \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]]]
\end{aligned}
\tag{A.9}
$$

Let us assume a sample of M for which we observe their choices of activity chains ($j = 1,\ldots, J$) and the model attributes of each available alternative (e.g. individual 1 selects activity chain alternative 1, individual 2 selects activity chain alternative 3,…, individual m selects activity chain alternative j,…, etc.).

Then, the likelihood function of all individual selections of activity chains is given by the product of logit choice model probabilities (e.g. $\mathrm{Pr}_j(\boldsymbol{\alpha})$ defined in (4.1)) as follows:

$$
L(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \prod_{j=1}^{J} \left(\mathrm{Pr}_{m,j}(\boldsymbol{\alpha}) \right)^{z_{m,j}}
\tag{i}
$$

where z is the indicator variable:
$$
z_{m,j} = \begin{cases} 1 & \text{if activity chain } j \text{ was chosen by individual } m \\ 0 & \text{otherwise} \end{cases}
$$

We normally maximise $L(\boldsymbol{\alpha})$ or their natural logarithms ($l(\boldsymbol{\alpha})$) where:

$$
l(\boldsymbol{\alpha}) = \sum_{m=1}^{M} \sum_{j=1}^{J} z_{m,j} \mathrm{Pr}_{m,j}(\boldsymbol{\alpha})
\tag{ii}
$$

When the likelihood function $L(\boldsymbol{\alpha})$ is maximised, a vector of estimated parameter ($\boldsymbol{\alpha}^*$) is obtained. The estimator of this MLP ($\boldsymbol{\alpha}^*$) has asymptotically normal distribution ($\mathrm{N}(\boldsymbol{\alpha}, \mathrm{S}^2)$ where S is standard deviation, e.g. see Ben-Akiva and Lerman, 1985; Ortuzar and Willumsen, 2011).

To compare the performance of ABM calibrations with observed plate scanning flows and link counts, ABM calibration with single traffic source (observed plate scanning flows, $\hat{\mathbf{g}}$) is also conducted. The likelihood estimation problem with only using plate scanning flows to calibrate ABM parameters (scenario III) can then be formulated as follows:

$$L(\hat{\mathbf{g}}, \boldsymbol{\alpha}^0 | \boldsymbol{\alpha}) = L(\boldsymbol{\alpha}^0 | \boldsymbol{\alpha}) \cdot L(\hat{\mathbf{g}} | \boldsymbol{\alpha}) \tag{A.10}$$

Also, the maximum likelihood estimators of the parameter vector ($\boldsymbol{\alpha}$) under this scenario III, that only plate scanning flows are used to calibrate, can be obtained by:

$$\boldsymbol{\alpha}^*_{senarioIII} = \underset{\boldsymbol{\alpha} \in \Omega}{\operatorname{argmax}} [-\frac{1}{2}[\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}]^{\mathrm{T}} \Sigma_{\alpha^0}^{-1} [\boldsymbol{\alpha}^0 - \boldsymbol{\alpha}]$$
$$-\frac{1}{2}[\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\alpha)]]^{\mathrm{T}} \Sigma_g^{-1} [\hat{\mathbf{g}} - \Psi \cdot \mathbf{P} \cdot \boldsymbol{\delta} \cdot \mathrm{E}[\mathbf{u}(\alpha)]]] \tag{A.11}$$

A.2 Multiday observations

The optimization problem for calibrating the AMB parameters ($\boldsymbol{\alpha}^*$) based on the observed link flows (on link set, $\mathfrak{I}$) and scanned sensor path flows (on sensor path set, $\mathfrak{R}$) from plate scanning, which are specified by mode and travel period, from $\bar{a}$ days is rewritten as:

$$Z(\boldsymbol{\alpha}^* | \boldsymbol{\alpha}^0, \hat{\mathbf{v}}_{not}, \hat{\mathbf{g}}) = \ln L(\boldsymbol{\alpha}^0 | \boldsymbol{\alpha}) + \ln L(\hat{\mathbf{v}}_{not}^1, ..., \hat{\mathbf{v}}_{not}^{\bar{a}} | \boldsymbol{\alpha}) + \ln L(\hat{\mathbf{g}}^1, ..., \hat{\mathbf{g}}^{\bar{a}} | \boldsymbol{\alpha}) \tag{A.12}$$

In this study, the sequential quadratic programming (SQP) in MATLAB software was adopted for solving this optimization problem (A.12). In addition, based on the properties of MLP, the variance-covariance matrix of α^* is approximated by $-\nabla_\alpha^2 Z^{-1}|_{\alpha=\alpha^*}$ (e.g. Nakayama et al., 2009).

The t-value of $\alpha_k, \alpha_k \in \boldsymbol{\alpha}$, for representing the confidence level of calibration, is defined as:

$$\text{t-value}(\alpha_k) = \frac{\alpha_k^*}{\sqrt{[-\nabla_\alpha^2 Z^{-1}|_{\alpha=\alpha^*}]_{kk}}} \tag{A.13}$$

A.3 Example of Maximum-likelihood estimation

Following the network specification described in section 2.2 (see Figure 1), only destination choice model in ABM is calibrated for illustration (e.g. destination choices from home at zone 1). Given two specific travel patterns (i.e. H1-W2 and H1-W3), the probability of selecting destination zone d originated from zone o, $\Pr(d|o)$, and the corresponding utility are defined as follows:

$$\Pr(d|o) = \frac{\exp(V_d)}{\sum_{d'} \exp(V_{d'})} \quad ; V_d = \alpha \cdot TT_{(o,d)} \tag{A.14}$$

where $TT_{(o,d)}$ is travel cost from origin o to destination d (given $TT_{(\mathrm{H1,W2})} = 1$ and $TT_{(\mathrm{H1,W3})} = 3$).

In addition, given expected number of population at zone 1 (i.e. $\mathrm{E}[N_1] = 1,000$), the expected travel demand ($\mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})]$) is expressed by the follows:

$$\mathrm{E}[\mathbf{u}(\boldsymbol{\alpha})] = \begin{bmatrix} u_{\mathrm{H1-W2}}(\alpha) \\ u_{\mathrm{H1-W3}}(\alpha) \end{bmatrix} = \begin{bmatrix} N_1 \cdot \Pr(d=2|o=1) \\ N_1 \cdot \Pr(d=3|o=1) \end{bmatrix} = \begin{bmatrix} 881 \\ 119 \end{bmatrix} \tag{A.15}$$

,and the variance-covariance, $\Sigma_u = \begin{bmatrix} u_{\text{H1-W2}}(\alpha) & 0 \\ 0 & u_{\text{H1-W3}}(\alpha) \end{bmatrix} = \begin{bmatrix} 881 & 0 \\ 0 & 119 \end{bmatrix}$ (A.16)

According to the network specification mentioned, we can obtain the following matrices:

$$\boldsymbol{\delta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \boldsymbol{\Lambda} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \ \mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$ (A.17)

Since there are three scanned sensor paths ($\mathbf{C}_1 = \{1\}$, $\mathbf{C}_2 = \{2\}$, $\mathbf{C}_3 = \{1,2\}$ see (2.3) for the definitions) which can be observed from plate scanning technique (see Table 1), the link-sensor path incidence matrix (**H**) is:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$ (A.18)

Regarding the plate scanning observations under illustrative settings of identification rate on link 1 and 2 equal to θ, the scanned sensor path-path choice proportion matrix is formulated by:

$$\Psi = \begin{bmatrix} \theta & \theta(1-\theta) \\ 0 & \theta(1-\theta) \\ 0 & \theta^2 \end{bmatrix}$$ (A.19)

To illustrate the model calibrations with traffic data, prior parameter (α^0) is set to deviate from true value ($\alpha^0 = 0.50\alpha^{\text{true}}$ where given $\alpha^{\text{true}} = -1.0$, $\Sigma_{\alpha^0} = 1$). To simulate traffic observations ($\hat{\mathbf{g}}$ and $\hat{\mathbf{v}}_{not}$), a simulation with 10 replications has been done. In each simulation (replication), scanned sensor path flows ($\hat{\mathbf{g}}$) are drawn from (3.8a) and simulated link flows ($\hat{\mathbf{v}}_{not}$) are drawn from (3.11). Then, three scenarios of likelihood estimation problems ((A.7), (A.9), or (A.11)) are solved to compare the performance of calibration results. The errors of calibration results are shown as follows:

Table A1
Errors (MSE) of model parameter calibrations in the illustrative network (Figure 1)

Prior parameters MSE	Scenario II: Link counts only MSE	θ	Scenario I: Link counts and sensor path flows MSE	Scenario III: sensor path flows only MSE
		95%	0.000134	0.000145
0.25000	0.00249	90%	0.000279	0.000319
		80%	0.000550	0.000707
		70%	0.000769	0.001195

The results from Table A1 show that the calibration error (MSE in (4.8)) from traffic simulations (scenario I ~ III) is significantly smaller than the error (MSE in (4.7)) of prior value, α^0. The calibration error using plate scanning information (scenario I or III) tends to increase while identification rate (θ) decreases due to normal approximation of the distribution of scanned sensor path flows in the likelihood estimation problems.

Acknowledgements

References

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

Asakura, Y., Hato, E. and Kashiwadani, M., 2000. OD matrices estimation model using AVI data and its application to the Han-Shin expressway network. *Transportation* 27(4), 419–438.

Axhausen, K.W., Schönfelder, S., Wolf, J., Oliveira, M., Samaga, U., 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*. Kluwer academic publishers, New York.

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

Bowman, J. L., Bradley, M. A., Shiftan, Y., Lawton, K. and Ben-Akiva, M. E., 1998. Demonstration of an activity based model system for Portland. In Selected Proceedings from *the 8th World Conference on Transport Research*, Transport Modelling/Assessment.

Bowman, J.L. and Ben-Akiva, M.E., 2001. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A* 35(1), 1-28.

Bowman, J.L., Bradley, M.A., Gibb, J., 2006. *The Sacramento activity-based travel demand model: estimation and validation results*. Paper presented at the European Transport Conference, Strasbourg, France, September 18-20.

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.

Bradley, M., Vovsha, P., 2005. A model for joint choice of daily activity pattern types of household members. *Transportation* 32(5), 545-571.

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

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

Castillo, E., Conejo, A.J., Pruneda, R.E., Solares, C., 2007. Observability in linear systems of equations and inequalities Applications, *Computers and Operation Research* 34 (6), 1708–1720.

Castillo, E., Mene´ndez, J.M., Sa´nchez-Cambronero, S., 2008a. Predicting traffic flow using Bayesian networks. *Transportation Research Part B* 42(5), 482-509.

Castillo, E., Menéndez, J.M., Jiménez, P., 2008b. 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., Sa´nchez-Cambronero, S., Rivas, A., 2010. Matrix tools for general observability analysis in traffic networks. IEEE Intelligent Transportation Systems 11(4), 799–813.

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.

Castillo, E., Nogal, M. Rivas, A., Sa´nchez-Cambronero, S., 2013. Observability of traffic networks. Optimal location of counting and scanning devices. *Transportmetrica B* 1(1), 68-102.

Clark, S.D., Watling, D.P., 2002. Sensitivity analysis of the prohibit-based stochastic user equilibrium assignment model. *Transportation Research Part B* 36 (7), 617–635.

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.

Cottrill, C.D., Pereira, F.C., Zhao, F., Dias, I.F., Lim, H.B., Ben-Akiva, M., Zegras, P.C., 2013. Future mobility survey: experience in developing a smartphone-based travel survey in Singapore. *Transportation Research Record* 2354, 59-67.

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.

He, S.X., 2013. A graphical approach to identify sensor locations for link flow inference. *Transportation Research Part B* 51, 65-76.

Kwon, J. and Varaiya, P., 2005. Real-time estimation of origin-destination matrices with partial trajectories from electronic toll collection tag data. *Transportation Research Record* 1923, 119-126.

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.

Li, Z.C., Lam, W.H.K, Wong S.C., 2012. Modeling intermodal equilibrium for bimodal transportation system design problems in a linear monocentric city. *Transportation Research Part B* 46(1), 30-49.

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.

Mínguez, R., Sánchez-Cambronero, S., Castillo, E., Jiménez, P., 2010. Optimal traffic plate scanning location for OD trip matrix and route estimation in road networks. *Transportation Research Part B* 44(2), 282-298.

Nakayama, S., Conners, R.D., Watling, D.P., 2009. Estimation of Parameters of Network Equilibrium Models: A Maximum Likelihood Method and Statistical Properties of Network Flow. *Transportation and Traffic Theory 2009: Golden Jubilee*, 39-56.

Ng, M.W., 2012. Corrigendum to: Synergistic Sensor Location for Link Flow Inference without Path Enumeration: A node-based Approach. *Transportation Research Part B* 46 (6), 781–788.

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.

Parry, K., Hazelton, M.L., 2012. Estimation of Origin-Destination Matrices from Link Counts and Sporadic Routing Data. *Transportation Research Part B* 46(1), 175-188.

Shen, W., Wynter, L., 2012. A new one-level convex optimization approach for estimating origin–destination demand. *Transportation Research Part B* 46(10), 1535-1555.

Siripirote, T., Sumalee, A., Watling, D.P., Shao, H., 2014. Updating of travel behavior model parameters and estimation of vehicle trip chain based on plate scanning. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 18(4), 393-409.

Van der Zijpp, N.J., 1997. Dynamic OD-matrix estimation from traffic counts and automated vehicle identification data. *Transportation Research Record* 1607, 87–94.

Vovsha, P., Bradley, M. and Bowman, J.L., 2004. *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, The Netherlands, May 28-31.

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.

Watling, D.P., 2006. User equilibrium traffic network assignment with stochastic travel times and late arrival penalty. *European Journal of Operational Research*, 175 (3), 1539-1556.

Wolf, J., Schönfelder, S., Samaga, Oliveira, U.M., and Axhausen, K.W., 2004. 80 weeks of GPS-traces: Approaches to enriching the trip information, *Transportation Research Record*, 1870, 46-54.

Yagi, S., Mohammadian, A., 2010. An activity-based microsimulation model of travel demand in the Jakarta metropolitan area. *Journal of Choice Modelling* 3(1), 32-57.

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

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

Yen, J.Y., 1971. Finding the *K* shortest loopless paths in a network. *Management Science* 17, 712–716.