An integrated Bayesian Approach for passenger flow assignment in metro network

Journal Article

Author(s):
Sun, Lijun; Lu, Yang; Jin, Jian Gang; Lee, Der-Horng; Axhausen, Kay W.

Publication date:
2015-03

Permanent link:
https://doi.org/10.3929/ethz-b-000089672

Originally published in:
Transportation Research Part C: Emerging Technologies 52, https://doi.org/10.1016/j.trc.2015.01.001

An Integrated Bayesian Approach for Passenger Flow Assignment in Metro Networks

Lijun Sun[a,b], Yang Lu[c,*], Jian Gang Jin[d], Der-Horng Lee[b], Kay W. Axhausen[a,e]

[a] *Future Cities Laboratory, Singapore-ETH Centre, Singapore, 138602, Singapore*
[b] *Department of Civil & Environmental Engineering, National University of Singapore, Singapore 117576, Singapore*
[c] *Singapore-MIT Alliance for Research and Technology (SMART), Singapore, 138602, Singapore*
[d] *School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*
[e] *Institute for Transport Planning and Systems (IVT), ETH Zürich, Zürich CH-8093, Switzerland*

Abstract

This paper proposes an integrated Bayesian statistical inference framework to characterize passenger flow assignment model in a complex metro network. In doing so, we combine network cost attribute estimation and passenger route choice modeling using Bayesian inference. We build the posterior density by taking the likelihood of observing passenger travel times provided by smart card data and our prior knowledge about the studied metro network. Given the high-dimensional nature of parameters in this framework, we apply the variable-at-a-time Metropolis sampling algorithm to estimate the mean and Bayesian confidence interval for each parameter in turn. As a numerical example, this integrated approach is applied on the metro network in Singapore. Our result shows that link travel time exhibits a considerable coefficient of variation about 0.17, suggesting that travel time reliability is of high importance to metro operation. The estimation of route choice parameters conforms with previous survey-based study, showing that showing that the disutility of transfer time is about twice of that of in-vehicle travel time in Singapore metro system.

Keywords: Bayesian inference, Metro network, Travel time, Smart card, Route choice

1. Introduction

With the increasing demand and range of urban mobility, metro systems are playing more and more important roles in urban transportation, particularly in high-density mega-cities. Taking Singapore's Mass Rapid Transit (MRT) system as an example, around two million metro trips were made daily in the year 2012. Compared to other transport modes, metro systems have dedicated and exclusive rail-based infrastructures, making it possible to provide superior service with higher speeds and larger capacity. Due to their superiority, metro systems not only attract but also suffer from high passenger demand — especially during rush hours when passenger demand exceeds its designed capacity for not only trains, but also platforms — experiencing over-crowdedness, disturbances and disruptions time and again. These adverse impacts can jeopardize passenger's traveling experience and therefore should be minimized. From operators' point of view, understanding passenger demand and flow assignment patterns in a complex metro network becomes crucial to maintaining service reliability and developing efficient failure response strategies.

To characterize a passenger flow assignment model for metro network, two factors are of the most importance: O-D demand matrix and route choice behavior. Because of the widely adopted tap-in-tap-out fare collection system, the station-to-station OD matrix in a metro network is known; however the route choice decisions are usually not directly observable, therefore a widely used approach is to first develop a route choice model — characterized by some critical cost attributes influencing passenger perception, such as in-vehicle time, number of transfers and fare paid — and then employ observed

*Corresponding author. Address: Singapore-MIT Alliance for Research and Technology (SMART), 1 Create Way, #09-01, CREATE Tower, Singapore, 138602, Singapore

Email addresses: `sunlijun@nus.edu.sg` (Lijun Sun), `luyang@smart.mit.edu` (Yang Lu), `jiangang.jin@sjtu.edu.cn` (Jian Gang Jin), `dhl@nus.edu.sg` (Der-Horng Lee), `axhausen@ivt.baug.ethz.ch` (Kay W. Axhausen)

preference data to calibrate model parameters. Despite the mathematical modeling, in principle there are two crucial issues to be solved in this approach before applying it on metro networks. The first is to accurately measure each attribute in the model, such as different stages of travel time and transit fares mentioned above. These values are used as input and assumed to be known in advance. In practice, experimenters need to determine such network properties by using train operation data and field surveys. However, accurate evaluation of route attributes, such as in-vehicle time, waiting time and transfer time, could be challenging considering possible congestion or interruption scenarios. The second issue is to obtain enough field observations, which register individual route choice preferences to support parameter estimation. However, in practice one may encounter many difficulties. On one hand, in the absence of detailed train operation logs recording train departure/arrival time and trajectories, it is difficult to measure exact network attributes, such as in-vehicle time, waiting time and transfer time. On the other hand, as most metro networks are designed as closed systems where passengers only leave traces at boarding/alighting stations for the purpose of fare collection, operators have limited knowledge on passenger route choice and trajectory within the system. In other words, we know little about which train or which transfer station an individual passenger has taken during his/her trip in the case where multiple alternative routes exist. In order to obtain passenger route choice preference data, a conventional approach is to conduct field surveys in train stations, asking people the exact route they will take to reach their destinations. However, some shortcomings of these methods have been identified, such as being subject to bias and errors and being both time-consuming and labor-intensive in conducting surveys and processing the data. In addition, since most surveys are conducted with focus on particular location and time, the results are often limited in scale and diversity. As a result, developing alternative methods to reveal individual route choice preference in large-scale networks remains challenging.

The emergence and wide deployment of automated fare collection (AFC) systems open a new data-driven approach for metro network analysis. Taking advantage of smart card-based fare collection systems, in which individual passenger's tapping-in/out transactions are recorded, researchers are now able to better understand metro operation with large quantities of real-world observations (Pelletier et al., 2011). Such data set also provides us with a good opportunity to study passenger behavior in a data-driven approach. In doing so, researchers have tried to combine passenger travel time information with train operation logs (Kusakabe et al., 2010; Sun and Xu, 2012; Zhou and Xu, 2012). However, without further investigating travel time variability, these approaches essentially assume that train services are always punctual to timetables and hence network cost attributes are assumed to be deterministic, even though there is clear evidence showing that train services can be delayed/disrupted by excessive passenger demand. On the other hand, owing to the uncertainty in travel time, the difficulties in revealing individual trajectory from tap-in/tap-out information still remain, preventing us from collecting accurate preference data. In view of these unsolved issues, this paper presents the development and empirical verification of a new integrated metro assignment framework using Bayesian inference approach. Taking advantage of large quantities of real-world observations provided by smart card data, the suggested model simultaneously estimate network attributes and passenger route choice preference. Consequently, the proposed framework utilizes only travel time observations along with static network data to construct the passenger flow assignment model in a closed metro network. With low social-economic cost and implementation convenience, such approach is appealing for metro operations and maintenance.

Bayesian inference method is a well established statistical model which has been applied to various transportation applications, including O-D estimation, route choice modeling and flow assignment inference (Hazelton, 2008, 2010; Wei and Asakura, 2013). It enables us to find a posterior distribution which integrates all our prior knowledge with the available observations. Although in this sense it is a powerful tool for our inference problem, in practice it is difficult to implement such models owing to the difficulty in computing the Bayesian posterior analytically. However, thanks to the rapid increase of computational power, nowadays we can characterize properties of the Bayesian posterior using computational approaches, of which the most notable one is Markov Chain Monte Carlo (MCMC) methods (Robert and Casella, 2004; Robert, 2014). The proposed framework in this paper is also based on solution algorithms provided by MCMC methods.

The contribution of this paper is threefold. First, we construct an integrated network characterization and flow assignment framework through a data-driven approach, allowing us to better understand passenger route choice behavior from large quantities of smart card observations. Second, by taking travel time variability caused by possible interruption during metro operation into consideration, our model can better characterize network travel time and its uncertainty given any origin-destination (O-D) pairs, providing better travel information to metro users. Finally, as will be shown in the following, the Bayesian formulation has the capacity to estimate network cost attributes and characterize passenger route choice model in a simultaneous manner, showing great potential in practice, in particular in cities with large/complex metro networks such as Beijing, London, New York, Seoul and Tokyo.

This remainder of this paper is organized as follows: in Section 2, we review previous studies on several related topics, including travel time reliability, passenger route choice behaviour, the use of smart card data in understanding metro operation and flow assignment, and in particular the application of Bayesian inference in transport network modeling; in Section 3, we propose the modelling framework, which contains several components including reconstructing network, identifying choice set and building the Bayesian inference model. In Section 4, we present the variable-at-a-time solution algorithm to characterize Bayesian posterior distribution. As an illustration, we apply the proposed framework on the simplified Singapore MRT network as a case study in Section 5. Finally, we conclude our study, summarize our main findings and discuss future research directions in Section 6.

2. Literature Review

Travel time reliability on urban road networks has been documented extensively in the literature, for both buses and private vehicles (Li et al., 2012; Strathman and Hopper, 1993; van Nes and van Oort, 2009). However, as mentioned, metro systems have long been considered as punctual to timetables (except during service interruptions/disruptions) and metro travel time reliability has attracted little attention in the literature. This is likely due to the lack of empirical observations regarding metro travel time reliability, which has now become available with the emergence of smart card data.

The large quantities of smart card transactions offer us a great opportunity to investigate passenger transit behavior and demand patterns (Bagchi and White, 2005). For example, Barry et al. (2002) first used smart card data to estimate metro O-D demand. By analyzing transit smart card data in Seoul, Park et al. (2008) suggested that smart card holders exhibit no difference from other users in terms of travel behavior, and travel patterns can be analyses in an aggregated manner. Using the same data set, Jang (2010) presented an empirical study on identifying transfer patterns of inter-modal transportation. Apart from understanding travel behavior, smart card data have been used to improve public transit services at strategic, tactical and operational levels as well. A comprehensive review of using smart card data at different levels of management can be found in (Pelletier et al., 2011). Using passenger demand extracted from smart card data in Singapore, Jin et al. (2014) studied a practical problem about integrating localized bus service with metro network in order to enhance the resilience to service disruptions of metro systems, offering new design principles of multi-modal transit networks. Using the same data set, Sun et al. (2014) proposed three optimization models to design demand-driven timetables for a single-track metro service. The results show that demand-sensitive timetables have great potential in reducing passenger waiting time and crowdedness on trains.

Bus smart card systems record not only boarding/alighting stop/time, but also the ID of the vehicle. Thus, it may play the same role as data collected from automated vehicle location (AVL) and automated passenger counting (APC) systems (Lee et al., 2012). However, for metro systems, in which smart card readers are not deployed on trains but at stations, we cannot identify the particular train that an individual passenger takes from the transactions directly. Thus, it remains a challenge to understand metro trips at a microscopic level, in particular when travel time variability is taken into account. Besides, without an in-depth understanding of passenger route choice behavior, the flow assignment problem still need to be studied carefully.

In terms of calibrating flow assignment models, the field has long been relying on collecting preference data (e.g. stated preference and revealed preference) from field surveys and analyses. Thanks to the emergence of smart card data, the challenge now may shift to reveal passenger route choice using historical transactions rather than collecting route choice data with physical surveys. In doing so, Kusakabe et al. (2010) developed a methodology to estimate the exact train which an individual passenger occupies during his/her journey. This method could be used to accurately estimate train occupancy, which is an important factor influencing passenger's perception on service quality. Zhou and Xu (2012) proposed a maximum likelihood estimation method of individual passenger route choice given his/her entry and exit times. Based on the individual estimation, a flow assignment model was proposed to map the macroscopic passenger flow in reality for comparison. Given that the model relies on service timetable, it cannot characterize special events such as passengers being left-behind by a full train. Using the same data set in Beijing, Sun and Xu (2012) introduced the stochastic cost nature of different segments of a metro trip — walking-in, waiting, transfer and walking-out. The method first characterizes the distribution of travel time on each alternative and then uses the mean and variance (moments) to estimate the weight parameter of each component. This approach also requires accurate train operation timetables/logs as input, which may not available for other cases. However, these studies essentially ignore the stochastic nature of train travel time between successive stations, assuming that trains are always punctual to the scheduled timetable and requiring scheduled timetable data as input. By analyzing real-world passenger travel in Singapore, we found clearly that there is an increasing trend of standard deviation of travel time against mean travel time, suggesting that variability increases with travel time. In order to infer the exact train that one passenger took in the absence of operation logs, **?** proposed a linear regression model to estimate train operation properties on a single-track and used the results to compute individual trajectory during a metro trip. By aggregating trajectories for all passengers by time, this method can help to identify train/service trajectory and estimate spatial-temporal occupancy of trains. However, the approach is only applied on a single-track, while at a network level the transfers and synchronization between different services need to be further investigated. In a recent paper, Zhu et al. (2014) presented an framework to calibrate passenger flow assignment model in metro networks based on genetic algorithm. The core of this framework is to first generate candidate set by using statistically-based criteria and then use genetic algorithm to find optimal solution.

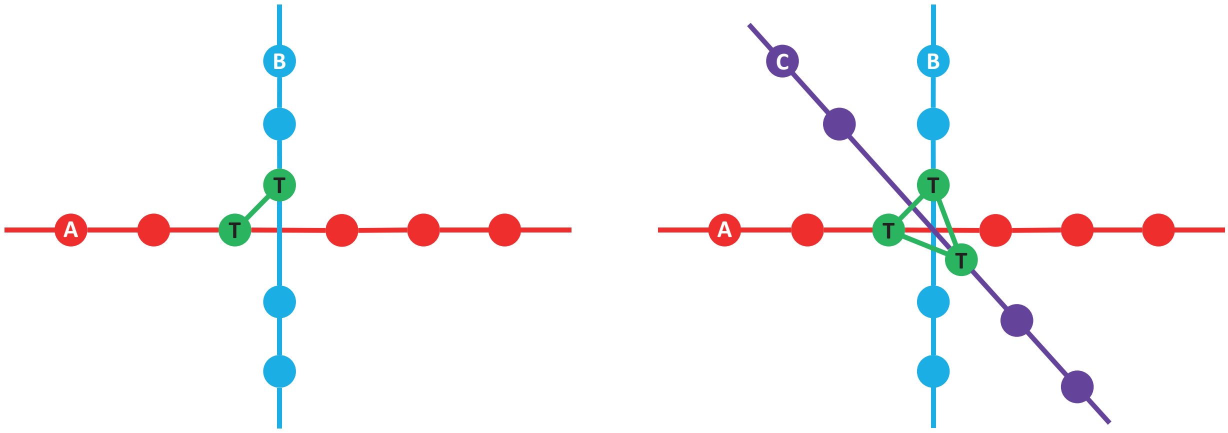
All previous studies focus on one particular part of the overall problem. To our knowledge, in the literature little attention was paid to deal with the case where both network characteristics and passenger choice behavior are unavailable/unknown, and few researchers characterized route choice behavior in metro networks in large scale. It remains a challenge to develop a comprehensive framework which can solve both mentioned issues simultaneously using only travel time observations. In this sense, the Bayesian computational tools become attractive as it builds a posterior distribution by simply combining likelihood of the observable and our prior knowledge about the model (see Robert (2014) for a review). In a previous study, Hazelton (2010) developed an unified framework which integrates a statistical linear inverse structure with network-based transport model. The author illustrated the performance of this framework by estimating perception parameters in logit route choice models in Leicester, UK. The successful application of this model also inspired us to apply Bayesian inference on metro networks in this paper. Despite calibrating choice models, Bayesian inference also exhibits excellent performance in stochastic traffic assignment modeling (Wei and Asakura, 2013) and vehicle travel time estimation using only Global Positioning System (GPS) data (Westgate et al., 2013). With the help of Bayesian inference and large quantities of travel time observations provided by smart card data , this paper introduces an integrated modeling framework to quantify both network attributes and passenger route choice behavior.

3. Modeling Framework

To associate the observed passenger travel time with link costs and route choices, in this section we first propose a network reconstruction process, which distinguishes transfer stations by services and adds transfer links among different platforms correspondingly. Afterwards, we present a brief description of the integrated inference problem and introduce all model parameters in this framework. Then, we determine route choice set R_w for each O-D pair w. Given actual network configuration and property, in doing so one may apply a brute-force-search (BFS) method or k-shortest path method. After obtaining choice set, a Multinomial Logit (MNL) model is applied to measure the probability of choosing each choice r among the available set R_w given route attributes, where travelers' sensitivity to each attribute is parameter to be estimated. Finally, as a key component of the proposed framework, a Bayesian inference model is built to estimate the unknown parameter vectors, by taking all registered travel time from smart card transactions as observations.

3.1. Network reconstruction

In order to better model passenger travel time and route choice behavior, we reconstruct a metro network following the examples illustrated in Fig. 1. In general, we can model each station as a single node in a sense similar to a map. However, by doing so we essentially miss the transfer cost for interchanging from one service to another (including walking and waiting), which is a crucial component of total travel time. In order to take transfer cost into consideration, we reconstruct a metro network by separating each transfer station as different nodes by services. For example in Fig. 1, nodes marked as "T" represent an identical transfer station in the metro system; however, we distinguish them on each metro service and add transfer links to characterize transfer cost (including waiting) from platforms of one service to another. Essentially, in the case that n ($n \geq 2$) services pass through a single transfer station, C_n^2 transfer links will be created between every pair.



(a) (b)

Figure 1: Reconstructing network by distinguishing transfer stations and adding transfer links

Despite that links could be directed as trains are operated in two-way, we model a metro network as undirected in this study for simplification, assuming that bi-directional travel costs between two adjacent stations are characterized by an identical distribution. In other words, we assume that two reciprocal links have the same properties.

3.2. Problem description

We consider a general metro network $G(N, A)$, consisting of a set of nodes N ($|N| = n$) and a set of links A ($|A| = m$). As we use travel time as cost measure in this study, 'cost' and 'time' as treated the same (interchangeable) throughout the paper. We assume that link travel time $\boldsymbol{x} = (x_1, \cdots, x_a, \cdots x_m)^\top$ are random variables, in the sense that services are not punctual to exact timetables due to various disturbances; as a result, stochastic travel time will be observed as in reality. This is also prerequisite to apply Bayesian statistical inference in our framework. Despite the fact that statistical properties of link travel time can be obtained from large quantities of service operation logs, the detailed train arrival/departure time and trajectory data along the service is seldom available. In this case, the Bayesian inference might be advantageous by taking unknown parameters as random variables and using travel time transactions as observations.

Let W be the set of O-D pairs; R_w denotes the set of possible routes connecting O-D pair w; $\boldsymbol{T}_w$ is the set of individual travel time obtained from those passengers traveling on O-D pair w, which is the final observable in this framework. We denote $\boldsymbol{T} = \bigcup_{w \in W} \boldsymbol{T}_w$ as the union of travel time observations from all O-D pairs. We start by introducing a combination of four parameter vectors, which capture different aspects of a metro system in our model:

- $\boldsymbol{c}$: describing network link costs;

- α : describing link cost variation (coefficient of variation);
- $\boldsymbol{\theta}$: describing passenger route choice behavior;
- m : describing extra cost on waiting/access/egress/failed boarding.

The details of these parameter vectors will be introduces in the following. In principle, our aim in this study is to use available observations (travel time $\boldsymbol{T}$) to infer all the unknown parameters above.

To allow for travel time reliability in our model, we assume that link cost x_a are random variables characterized by normal distribution $\mathcal{N}(c_a, (\alpha c_a)^2)$, of which the standard deviation is in proportion to its mean ($\sigma = \alpha\mu$). We assume that all link costs are independent. Thus, the overall distribution for all links can be written as:

$$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{c}, \mathrm{diag}\{\alpha \boldsymbol{c}\}^2\right), \tag{1}$$

where $\boldsymbol{c} = (c_1, \cdots, c_a, \cdots c_m)^\top$ represents the mean travel time for all links and $\alpha = \sigma/\mu$ is the coefficient of variance. The independent normality assumption of link cost is crucial in our modeling, as it provides us a simplified way to measure route travel time given the additive property of normal distributions.

In modeling passenger route choice behavior in the metro network, we assume that choice probability is characterized by a Multinomial Logit (MNL) model and the representative utility of each route is measured as a linear combination of different route attributes with parameters $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_K)^\top$.

As stated, the smart card system only provides us with the inter-tapping interval for each individual traveler, which is treated as travel time observations. In spite of transfer costs, the inter-tapping interval still involves in the access/egress walking time at boarding/alighting stations respectively, and waiting time at boarding platforms. In order to capture these extra costs, we impose a universal cost y on all O-D pairs and assume it to be characterized by a normal distribution:

$$y \sim \mathcal{N}\left(m, \sigma_y^2\right), \tag{2}$$

where m is an unknown parameter representing the mean of extra time and σ_y is standard deviation of y.

Note that the normal distribution assumption of link travel time is not mandatory in the proposed framework, but it will simplify the following step on calculating route cost. One can replace the normal assumption with any other distributions to facilitate the modeling requirements.

3.3. Generating route choice set

Before modeling passenger route choice behavior, we need to generate a choice set R_w for each O-D pair w, comprising all possible alternatives. In doing so, one may apply different strategies, such as link elimination, labeling and k-shortest-paths. Nevertheless, given the limited size and its simple structure of a metro network, a brute-fore-search (BFS) algorithm could be more advantageous than other methods in generating choice sets in shorter time.

Note that the proposed network reconstruction processes may produce some redundant alternatives, which are in principle illogical in reality, such as

- route with first link being a transfer link;
- route with last link being a transfer link;
- route containing two consecutive transfer links (appears where more than two services go through the same transfer station).

To better model choice behavior, we identify those routes with previous attributes and discard them when generating the final route choice set R_w.

3.4. Bayesian formulation

In this subsection we derive the Bayesian posterior distribution of parameter vectors given travel time observations. Based on the previous description, the unknown parameters are mean of link travel time $\boldsymbol{c}$, coefficient of variation α of link cost, parameters $\boldsymbol{\theta}$ for the MNL route choice model and average extra cost m. The observables we have are the travel time transactions for each O-D pair obtained from smart card data.

Taken together, applying Bayes' theorem on the unknown parameters and observations will give us the posterior density

$$\pi\left(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m | \boldsymbol{T}\right) = \frac{p\left(\boldsymbol{T} | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) \pi\left(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right)}{p\left(\boldsymbol{T}\right)}, \tag{3}$$

where the denominator $P\left(\boldsymbol{T}\right)$ is the marginal density for T over all unknown parameters

$$p\left(\boldsymbol{T}\right) = \int\int\int\int \pi\left(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) p\left(\boldsymbol{T} | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) \mathrm{d}\boldsymbol{c}\mathrm{d}\alpha\mathrm{d}\boldsymbol{\theta}\mathrm{d}m. \tag{4}$$

With this formulation, $P\left(\boldsymbol{T}\right)$ is in fact a normalizing constant expressed as high-dimensional integrals, being independent of any unknown parameters. Thus, by further assuming that all unknown parameter vectors (and all elements in each vector) are independent, we have

$$\pi\left(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m | \boldsymbol{T}\right) \propto p\left(\boldsymbol{T} | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) \pi\left(\boldsymbol{c}\right) \pi\left(\alpha\right) \pi\left(\boldsymbol{\theta}\right) \pi\left(m\right), \tag{5}$$

where $\pi\left(\delta\right)$ is the prior distribution of unknown parameter δ. Note that the probability of observing travel time $\boldsymbol{T}$ conditional on all unknown parameters equals the likelihood of all parameters given travel time observations, which means $p\left(\boldsymbol{T} | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) = \mathcal{L}\left(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m | \boldsymbol{T}\right)$.

Next, we focus on the likelihood function $\mathcal{L}\left(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m | \boldsymbol{T}\right)$. By distinguishing travel time observations by their O-D pairs, we can re-write the likelihood as

$$\mathcal{L}\left(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m | \boldsymbol{T}\right) = \prod_{w \in W} p\left(\boldsymbol{T}_w | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right). \tag{6}$$

As stated, there often exists more than one possible route for an O-D pair w, so that the probability of observation travel time t from an individual also depends on the the alternative routes he/she may take. Therefore, by applying the formula of total probability on each observation t against all possible routes R_w, the probability of observing travel time t on O-D pair w can be expressed as

$$p_w\left(t | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) = \sum_{r \in R_w} h\left(t | r, \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) f_w\left(r | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right), \tag{7}$$

where $f_w\left(r | \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right)$ is the conditional probability of choosing route r from choice set R_w given all model parameters, and $h\left(t | r, \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right)$ represents the conditional probability of observing travel time t given that route r is taken on O-D pair w.

Based on our primary assumption that link costs all follow normal distribution independently, we know that $t | r, \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m$ also follows a normal distribution given its additive property

$$t | r, \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m \sim \mathcal{N}\left(\sum_{a \in r} c_a + m, \alpha^2 \sum_{a \in r} c_a^2 + \sigma_y^2\right), \tag{8}$$

and thus

$$h\left(t | r, \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) = \frac{1}{\sqrt{2\pi\left(\alpha^2 \sum\limits_{a \in r} c_a^2 + \sigma_y^2\right)}} \exp\left(\frac{\left(t - \left(\sum\limits_{a \in r} c_a + m\right)\right)^2}{2\left(\alpha^2 \sum\limits_{a \in r} c_a^2 + \sigma_y^2\right)}\right). \tag{9}$$

To model passenger route choice behavior, we apply a Multinomial Logit (MNL) choice model, which usually assumes that representative utility V_r on route r is a linear function of route attributes $\boldsymbol{X}_r = \left(X_{r1}, \cdots, X_{rK}\right)^\top$ (which is a function of cost parameters)

$$V_r\left(\boldsymbol{\theta}; \boldsymbol{c}, \alpha, m\right) = \boldsymbol{\theta}^\top \boldsymbol{X}_r = \sum_k \theta_k X_{rk}, \tag{10}$$

where θ_k is sensitivity parameter for attribute X_{rk}. Researchers have tried to quantify the impact of various attributes in determining passenger route choices in metro systems, such as in-vehicle time, waiting time, walking time, number of transfers and occupancy (Raveau et al., 2014). However, we cannot apply previous estimations directly since such behavioral parameters vary enormously from system to system, from city to city. Thus, one of the main purposes of such modeling framework is to infer parameter vector $\boldsymbol{\theta}$ case by case (Hazelton, 2010). Taken together, when traveling on O-D pair w, the conditional probability $f_w(r|\cdot)$ of choosing route r conditional on other parameters $(\cdot)$ is

$$f_w(r|\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m) = \frac{\exp(V_r)}{\sum\limits_{r' \in R_w} \exp(V_{r'})}. \tag{11}$$

Therefore, the likelihood of O-D pair w can be given as

$$p(\boldsymbol{T}_w|\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m) = \prod_{t \in \boldsymbol{T}_w} \left(\sum_{r \in R_w} h(t|r, \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m) f_w(r|\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m) \right). \tag{12}$$

Substituting Eqs. (6) and (12) in to Eq. (5), we can write the posterior probability as

$$\pi(\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m|\boldsymbol{T}) \propto \prod_{w \in W} \left(\prod_{t \in \boldsymbol{T}_w} \left(\sum_{r \in R_w} h(t|r, \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m) f_w(r|\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m) \right) \right) \times \pi(\boldsymbol{c}) \times \pi(\alpha) \times \pi(\boldsymbol{\theta}) \times \pi(m) \tag{13}$$

Before implementing the Bayesian inference framework, we need to specify exact prior distributions $\pi(\delta)$ for the unknown parameters $\boldsymbol{c}$, α, $\boldsymbol{\theta}$ and m. Prior distributions are important if the number of observations is limited. However, for a metro system, the smart card system actually provide us with large quantities of travel time observations, helping us to correct our prior knowledge to a great extent. In practice, it would be better to propose prior distributions based on our experience or existing knowledge about the systems. In the case that we almost have no information about the parameters, a broad distribution such as uniform should be used.

The posterior distribution can provide not only point estimations for the unknown parameters but also their Bayesian confidence interval and Bayesian p-values for the purpose of hypothesis tests. However, in practice, it is usually impossible to get analytic estimations given its complex formulation. In the next section, we show a computational way to obtain the posterior distribution.

4. Solution Algorithm

If the conditional distribution can be written in closed form, ideally one can compute the marginal posterior distribution for each individual parameter analytically by calculating integrals. However, in our case, this approach is essentially impossible due to the difficulties in deriving the posterior distribution in Eq. (13) given its complicated formulation, the high-dimensional nature of the parameter space and in particular the normalizing integrals appearing in the dominator of Eq. (3). For such problems, in practice one usually uses the Monte Carlo Markov Chain (MCMC) approach to construct an updating algorithm to generate $\delta^{(t+1)}$ once we know $\delta^{(t)}$ (Robert and Casella, 2004). The Metropolis-Hastings (M-H) algorithm is a widely applied MCMC method, which enables us to sample candidate from a posterior distribution without knowing the closed form (Metropolis et al., 1953; Hastings, 1970). In each iteration, the M-H algorithm will generate a candidate from a pre-defined proposal distribution and then determine whether to accept it by calculating acceptance probability, which is a function of the ratio between target distribution density of the new candidate and the current sample respectively. By this means, we clear out the normalizing constant during the sampling. On the other hand, the Markov chain also shows advantages in a way that its stationary distribution is the target (or posterior) distribution we want to sample. Therefore, after obtaining enough realizations $\delta^{(1)}, \cdots, \delta^{(M)}$, one can estimate property I of parameter δ using

$$\hat{I} = \frac{1}{M-B} \sum_{i=B+1}^{M} f\left(\delta^{(i)}\right), \tag{14}$$

where B is a fixed number representing the burn-in period and M is the total number of samples. The burn-in samples are discarded as they might be biased given the arbitrarily chosen initial value $\delta^{(0)}$. After the burn-in period, the marginal distribution of the Markov chain is converging to its stationary state. To better determine the length of burn-in period, researchers have proposed different techniques in the literature (Geweke, 1992). The real characteristics of parameter δ can be measured using samples drawn from the posterior distribution π after the burn-in stage.

Given the high-dimensional nature of the studied problem, we perform a Variable-at-a-Time Metropolis sampling scheme (Metropolis et al., 1953). In doing so, we combine all parameter vectors in the posterior distribution as a full vector

$$\boldsymbol{\delta} = \left(\boldsymbol{c}^\top, \alpha, \boldsymbol{\theta}^\top, m\right)^\top = (c_1, \cdots, c_N, \alpha, \theta_1, \cdots \theta_K, m)^\top = (\delta_1, \cdots, \delta_{N+K+2})^\top . \tag{15}$$

The variable-at-a-time Metropolis then performs Metropolis sampling scheme on each coordinate of the parameter space in sequence, in the meanwhile other coordinates (parameters) remain fixed. Essentially, we may take an arbitrary proposal distribution $q\left(\delta_i^*|\delta_i^{(t)}\right)$ to draw samples for the i^{th} coordinate, and accepting new candidate δ_i^* based on M-H algorithm. However, in practice choosing an appropriate proposal distribution is crucial to performing the sampling process effectively. For simplicity, we apply a Gaussian random walk Metropolis (RWM) proposal to generate new candidates in a sequential order, in which

$$\delta_i^* = \delta_i^{(t)} + \epsilon_i^{(t)}, \tag{16}$$

where $\epsilon_i^{(t)} \sim \mathcal{N}\left(0, \sigma_i^2\right)$ and σ_i is the proposal standard deviation for the i^{th} coordinate. In other words, conditioning on the current sample, the new candidate follows a normal distribution $\delta_i^*|\delta_i^{(t)} \sim \mathcal{N}\left(\delta_i^{(t)}, \sigma_i^2\right)$. Thus, for the symmetric Gaussian distribution proposal, we have $q\left(\delta_i^*|\delta_i^{(t)}\right) = q\left(\delta_i^{(t)}|\delta_i^*\right)$, which simplifies the acceptance probability in M-H algorithm to

$$\mathcal{A}\left(\delta_i^*, \delta_i^{(t)}\right) = \min\left\{1, \frac{\pi'\left(\delta_i^*\right) q\left(\delta_i^{(t)}|\delta_i^*\right)}{\pi'\left(\delta_i^{(t)}\right) q\left(\delta_i^*|\delta_i^{(t)}\right)}\right\} = \min\left\{1, \frac{\pi'\left(\delta_i^*\right)}{\pi'\left(\delta_i^{(t)}\right)}\right\}, \tag{17}$$

where $\pi'\left(\delta_i'\right)$ is the target (posterior) probability by changing only the i^{th} parameter to δ_i' and keeping other parameters as their latest updated values. In other words, by updating parameters in sequential order, $\pi'\left(\delta_i'\right)$ is calculated as the posterior density

$$\pi'\left(\delta_i'\right) = \pi\left(\delta_1^{(t+1)}, \cdots, \delta_{i-1}^{(t+1)}, \delta_i', \delta_{i+1}^{(t)}, \cdots, \delta_{N+K+2}^{(t)}|\boldsymbol{T}\right). \tag{18}$$

Taken together, we summarize the variable-at-a-time Metropolis algorithm as the following processes:

Variable-at-a-Time Metropolis Sampling

(V1) Specify initial samples $\boldsymbol{\delta}^{(0)} = \left(c_1^{(0)}, \cdots, c_N^{(0)}, \alpha^{(0)}, \theta_1^{(0)}, \cdots, \theta_K^{(0)}, m^{(0)}\right)^\top$; set $t \leftarrow 1$.

(V2) For $\boldsymbol{\delta}^{(t)}$, sampling new value $\delta_i^{(t+1)}$ conditional on its current value $\delta_i^{(t)}$ in sequential order $(i = 1, \cdots, N + K + 2)$ using M-H sampling scheme (see following).

(V3) If $t < M$, set $t \leftarrow t + 1$ and return to Step (V1); Otherwise, stop sampling.

In order to avoid generating candidate from a high-dimensional distribution directly, the variable-at-a-time generate new sample for each coordinate in turn in Step (V2). In doing so, new candidate on each coordinate is sampled based on the following M-H scheme.

Metropolis-Hasting Sampling

(M1) Sample candidate value δ_i^* using the Gaussian random walk proposal (see Eq. (16)).

(M2) Compute acceptance probability using Eq. (17). The target (posterior) distributions are calculated as:

$$\begin{aligned}\pi'\left(\delta_i^*\right) &= p\left(\boldsymbol{T}|\delta_i^*, \delta_{-i}^{(t)}\right)\pi\left(\delta_i^*, \delta_{-i}^{(t)}\right)/p\left(\boldsymbol{T}\right)\\ &\propto p\left(\boldsymbol{T}|\delta_1^{(t+1)}, \cdots, \delta_{i-1}^{(t+1)}, \delta_i^*, \delta_{i+1}^{(t)}, \cdots, \delta_{N+4}^{(t)}\right)\pi\left(\delta_i^*\right),\end{aligned} \tag{19}$$

and

$$\begin{aligned}\pi'\left(\delta_i^{(t)}\right) &= p\left(\boldsymbol{T}|\delta_i^{(t)}, \delta_{-i}^{(t)}\right)\pi\left(\delta_i^{(t)}, \delta_{-i}^{(t)}\right)/p\left(\boldsymbol{T}\right)\\ &\propto p\left(\boldsymbol{T}|\delta_1^{(t+1)}, \cdots, \delta_{i-1}^{(t+1)}, \delta_i^*, \delta_{i+1}^{(t)}, \cdots, \delta_{N+4}^{(t)}\right)\pi\left(\delta_i^{(t)}\right),\end{aligned} \tag{20}$$

where $\delta_{-i}^{(t)} = \left(\delta_1^{(t+1)}, \cdots, \delta_{i-1}^{(t+1)}, \delta_{i+1}^{(t)}, \cdots, \delta_M^{(t)}\right)$ is parameter set from the latest updated coordinates except the i^{th} (i.e. $\delta_i^{(t)}$). The normalizing constant $p\left(\boldsymbol{T}\right)$, together with $\pi\left(\delta_{-i}^{(t)}\right)$ in both numerator and dominator, can be canceled out when calculating $\pi'\left(\delta_i^*\right)/\pi'\left(\delta_i^{(t)}\right)$.

(M3) Sample a value $\delta_i^{(t+1)}$ according to the following:

$$\delta_i^{(t+1)} = \begin{cases}\delta_i^* & \text{with probability } \mathcal{A}\left(\delta_i^*, \delta_i^{(t)}\right)\\ \delta_i^{(t)} & \text{otherwise.}\end{cases} \tag{21}$$

The variable-at-a-time Metropolis is a good genetic choice for high-dimensional problems as our case, since it keeps only one dimension (i.e. the i^{th} coordinate) as variable each time; while the general Metropolis moving all coordinates at once, resulting in large rejection rate. For each unknown parameter, the algorithm outputs a collection of iteration-stamped samples, whose stationary distribution is its marginal posterior distribution.

5. Case Study

For the purpose of model illustration and verification, in this section we apply the proposed modeling framework on Singapore's Mass Rapid Transit (MRT) network. The Bayesian inference model is built on real-world travel time (between tapping-in and tapping-out) observations collected on one day (19th, March, 2012) in Singapore as an example.

5.1. Singapore MRT network

We only consider the arterial network of Singapore's metro systems by removing extensions and light rail transit services. Fig. 2 shows the basic structure of the adapted network, which consists of four services (shown in different colors) and 88 stations. The reconstructed network contains 99 nodes and 107 links, of which 95 are in-vehicle links and 12 are transfer links. In this network, most transfer stations connect only two services. In the center of the figure we can see a special case that three services pass through the same transfer station — Dhoby Ghaut. For this special case, three transfer links will be created.

Table. 1 lists all transfer links and the corresponding platforms they connect.

5.2. Route choice behavior

In order to generate route choice set R_w, we performed BFS method described in the modeling framework and removed all redundant alternatives. After obtaining choice set R_w, we used an Multinomial Logit model route choice model as defined in Eq. (11) to computed route choice probability. A variety of studies on passenger route choice behavior have been conducted based on field survey data in the literature (Guo and Wilson, 2007; Wardman and Whelan, 2011; Raveau et al., 2014). For instance, Raveau et al. (2014) studied route choice behavior in two metro networks — London Underground and Santiago Metro, by taking various attributes into consideration, including different time components, transfer experience, level of crowdedness, network topology and other social-demographic characteristics.

Figure 2: Adapted MRT network of Singapore used in this study (source: http://exploresg.com/mrt/)

Table 1: Transfer links in Singapore MRT network

station	platform A	platform B
Bishan	NS17	CC15
Buona Vista	EW21	CC22
City Hall	EW13	NS25
Dhoby Ghaut	NS24	CC1
Dhoby Ghaut	NS24	NE6
Dhoby Ghaut	NE6	CC1
HarbourFront	NE1	CC29
Jurong East	EW24	NS1
Outram Park	EW16	NE3
Paya Lebar	EW8	CC9
Raffles Place	EW14	NS24
Serangoon	NE12	CC13

In fact, our modeling framework provides us with enough flexibility to apply diverse types of utility function in the choice model. Nevertheless, in this study we only examined a simple example, in which the representative utility V_r of route r is completely characterized by two attributes ($K = 2$): (1) total in-vehicle travel time $X_{r1} = \sum_{a \in r \setminus r_t} c_a$, and (2) total transfer time $X_{r2} = \sum_{a \in r_t} c_a$, quantifying route utility as

$$V_r = \theta_1 \sum_{a \in r \setminus r_t} c_a + \theta_2 \sum_{a \in r_t} c_a, \tag{22}$$

where route r is considered as a set of links and r_t is the set of transfer links in route r. In this formulation we do not take transit fares and waiting time of the first stage into consideration, because fare is only computed by the shortest alternative in distance (i.e. transit fares are the same for different route alternatives) and waiting time of the first stage is assumed to be the same across all trips. Therefore, both terms can be cancelled our in the utility function. Note that V_r is also a function of unknown parameters. Under the above assumptions, the probability of choosing route r conditional on other parameters is given by

$$f_w\left(r \mid \boldsymbol{c}, \alpha, \boldsymbol{\theta}, m\right) = \frac{\exp\left(\theta_1 \sum_{a \in r \setminus r_t} c_a + \theta_2 \sum_{a \in r_t} c_a\right)}{\sum_{r' \in R_w} \exp\left(\theta_1 \sum_{a \in r' \setminus r'_t} c_a + \theta_2 \sum_{a \in r'_t} c_a\right)}. \tag{23}$$

5.3. Priors

In the Bayesian inference framework, prior distributions should be given in closed form as chosen by the experimenter. Prior knowledge is crucial to inferring parameters when we have limited number of observations. In our case, as all metro users have to use their smart cards to tap-in/-out for the purpose of fare payment, large quantities of travel time observation is produced, stamped with both spatial and temporal information. Although the large number of observations can help us to correct our prior knowledge on the unknown parameters to a great extent, we still may benefit from an appropriate prior distribution.

Previous travel experience in Singapore's metro network indicates that travel time between two successive stations is about 2min (Sun et al., 2012). We therefore assume a normal prior with $\mu = 2$min and $\sigma = 1$min on average link cost c_a (for all links), giving that $\pi\left(c_a\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(c_a - 2\right)^2\right)$. Given the independent link cost assumption, the total prior for all links can be expressed as $\pi\left(\boldsymbol{c}\right) \propto \exp\left(-\frac{1}{2} \sum_{c_a \in \boldsymbol{c}} \left(c_a - 2\right)^2\right)$. Here we do not assign different priors to distinguish in-vehicle links and transfer links.

Extra cost y in a metro trip is also estimated based on previous study. We proposed that $m \sim \mathcal{N}\left(4, 1\right)$ — a normal distribution with mean 4min and variance 1min^2. For the variance of extra cost, we take an empirical value that $\sigma_y^2 = 1.5$min^2.

In terms of coefficient of variation α and route choice parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2\right)$, we almost have no available prior information to make a first guess. Therefore, we assigned uniform priors on these three parameters: $\alpha \sim \mathcal{U}\left(0, 1\right)$ and $\theta_i \sim \mathcal{U}\left(-4, 0\right)$ for $i = 1, 2$.

5.4. Summary

The final parameter vector $\boldsymbol{\delta}$ contains $N + K + 2 = 111$ elements. In each iteration, the variable-at-a-time Metropolis updates these parameters in turn. We implemented the sampling algorithm described in previous section using MATLAB. To avoid biased travel time observations, we discarded observations from O-D pairs with less than 100 transactions and selected a subset (by choosing 100 observations randomly) from each O-D pair in the remaining data sets as final observable $\boldsymbol{T}_w$. The size of O-D pair set is $|W| = 1897$; hence, total number of travel time observations used in this study is 189,700. We employed Gaussian random walk Metropolis proposals on all the unknown parameters; however, in order to build a well-mixed chain of realizations for each parameter, we chose different proposal standard deviations to allow for their characteristics. For instance, the proposal standard deviations of in-vehicle links and transfer links are chosen as 0.2min and 0.5min, respectively. The initial values and the corresponding proposal standard deviations for all parameters are listed in Table. 2. In fact, the initial value $\delta_i^{(0)}$ for each parameter is chosen as the mean of its prior distribution. We conducted computation experiments on a PC with an Intel Core i7 3.40GHz CPU and 16GB RAM. Considering the large size of O-D pairs, calculating posterior density (or log-posterior density) becomes computationally intensive. It takes about 30sec for each iteration of the variable-at-a-time sampling.

The sampling is run for $M = 30000$ iterations, of which $B = 5000$ are discarded as the burn-in period. We observed significant serial correlation in the sampled values of each coordinate δ_i. Fig. 3 shows the autocorrelation plots for chains of α, θ_1, θ_2 and m. Despite a good acceptance rate for all chains, we still found that the realizations are strongly dependent. To avoid such correlation, one may use thinning approach to get spaced samples. For example, one may retain every 50th value generated to obtain a subset with correlation less than 0.1. However, given the considerable cost in obtaining each sample, in this study we did not thin the results (Geyer, 1992).

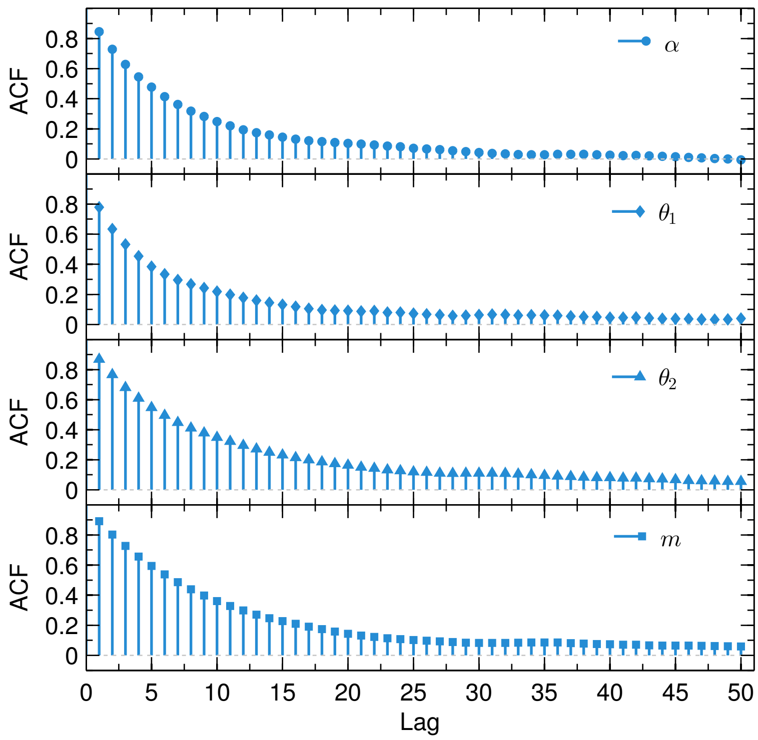


Figure 3: Autocorrelation plots for chains of α, θ_1, θ_2 and m.

As stated, we started the MCMC algorithm using the initial value and proposal standard deviation for each parameter as given in Table. 2. In total, 25000 effective samples for each parameter were drawn. The Bayesian analysis provides us with not only a point estimator but also a distribution to construct Bayesian confidence interval. The last two columns of Table. 2 show the final results of our inference based on those effective samples, including the mean and 95% Bayesian conference interval (CI). As can be seen, the large number of travel time observations have vastly corrected our biased prior knowledge of the system.

We display the estimation results of link cost on the EW service (shown in green in Fig. 2) in Fig. 4. The dots depict the mean values of each link and the corresponding errorbar shows the 95% Bayesian confidence interval. As a guide, the two insets show the distribution of c_1 and c_{28}, which are the first and last links on the East-West service. Despite the same initial values and proposal standard deviation were used in the inference process, the MCMC algorithm has successfully distinguished cost attributes for different links.

Fig. 5 displays the results of Bayesian inference on α, m, θ_1 and θ_2. In all the panels, the solid lines depict the kernel density estimates of parameters. As comparison, the dashed lines depict their prior distributions. The coefficient of variation α is characterized by a centralized distribution, the mean of

Table 2: Parameter description and estimation

parameter	prior	σ_i proposal	$\delta_i^{(0)}$	mean	95% Bayesian CI
α	$\mathcal{U}(0,1)$	0.005	0.500	0.168	[0.167,0.169]
θ_1	$\mathcal{U}(-4,0)$	0.050	-2.000	-0.462	[-0.473,-0.451]
θ_2	$\mathcal{U}(-4,0)$	0.050	-2.000	-0.959	[-0.988,-0.931]
m	$\mathcal{N}(4,1)$	0.010	4.000	3.270	[3.255,3.283]
c_1	$\mathcal{N}(2,1)$	0.200	2.000	3.651	[2.584,3.718]
c_2	$\mathcal{N}(2,1)$	0.200	2.000	2.947	[2.880, 3.013]
c_3	$\mathcal{N}(2,1)$	0.200	2.000	3.660	[3.591, 3.728]
c_4	$\mathcal{N}(2,1)$	0.200	2.000	3.107	[3.038, 3.169]
...					
c_{107}	$\mathcal{N}(2,1)$	0.500	2.000	5.247	[5.151, 5.333]

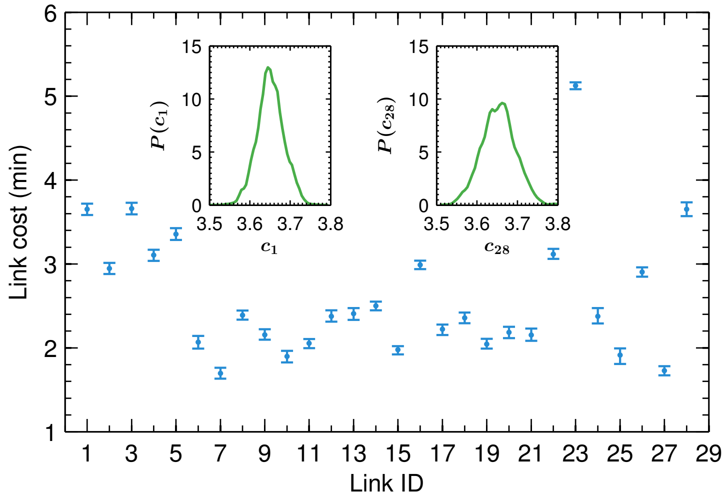


Figure 4: Link cost estimation (mean and 95% Bayesian confidence interval) for the EW line (shown in green in Fig. 2). Link with ID n represents in-vehicle link between station EW n and EW $n+1$.

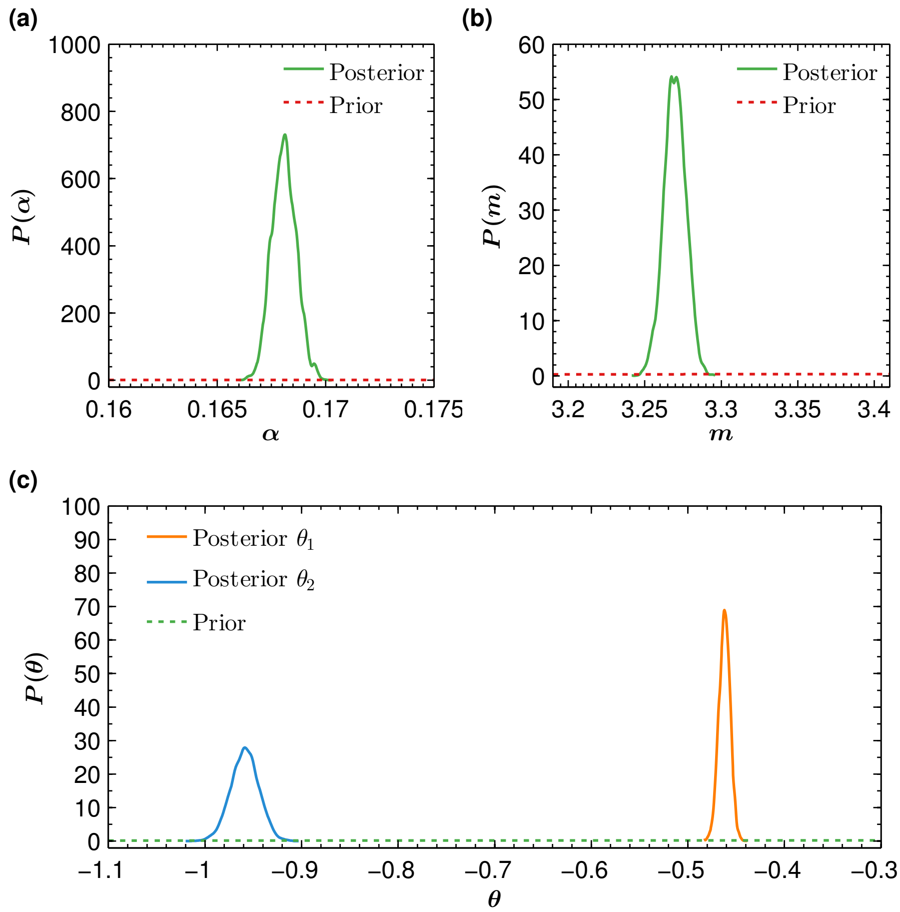


Figure 5: Prior and posterior density for (a) α, (b) m, and (c) θ_1 and θ_2.

which is far from its initial value. The posterior mean and standard deviation are 0.1681 and 0.0006, respectively. Although we expect that a flat normal prior $\mathcal{N}(4,1)$ could characterize m, in contrast the estimation process gives us a more centralized distribution shown in panel (b), with a very small standard deviation of about 0.007min, suggesting that in average passengers may spend about 3min in total as extra cost. In fact, the reason we did not get a distribution with higher variance is that m only capture the mean of extra cost y, while the variance of extra cost is assumed to be known as $\sigma_y^2 = 1.5\text{min}^2$. Thus, the result is consistent with our expectation, suggesting that we may use a more appropriate prior distribution to characterize m.

Essentially, by combining the estimation results on link cost $\boldsymbol{c}$, coefficient of variation α and extra cost m, operators and agencies can better estimate travel time and its variability for all O-D pairs in the network, helping metro users to better plan their trips. Both users and agencies can benefit from such information.

Passenger route choice behavior is reflected in parameter $\boldsymbol{\theta}$. Panel (c) in Fig. 5 shows the distribution of θ_1 and θ_2, respectively. The same uniform prior is also plotted as a guide. As can be seen, the Bayesian inference has significantly distinguished the effect of transfer time from in-vehicle time. The posterior mean of θ_1 is -0.462 and its standard deviation is 0.006. For θ_2, the posterior mean is -0.959 and the posterior standard deviation is 0.015. The significant difference between θ_1 and θ_2 suggests that metro users value transfer time more than in-vehicle time. The result conforms to previous survey-based studied in London Underground and Santiago Metro (Raveau et al., 2014). In addition, the inference framework also provides Bayesian confidence interval for both θ_1 and θ_2.

Finally, we plot the joint posterior density for (θ_1, θ_2) in Fig. 6. To estimate the joint density, we fixed all other parameters as the mean values of their effective samples (as provided in Table. 2) and took only θ_1 and θ_2 as parameters. Clearly, the maximum value can be found around $(-0.462, -0.959)$; however, the density decreases at different speed given different parameter direction. The figure provides us with two implications. On one hand, the peaked shape of joint posterior distribution shows that the density is sensitive to the oscillation of both θ_1 and θ_2, suggesting that changing route choice parameters

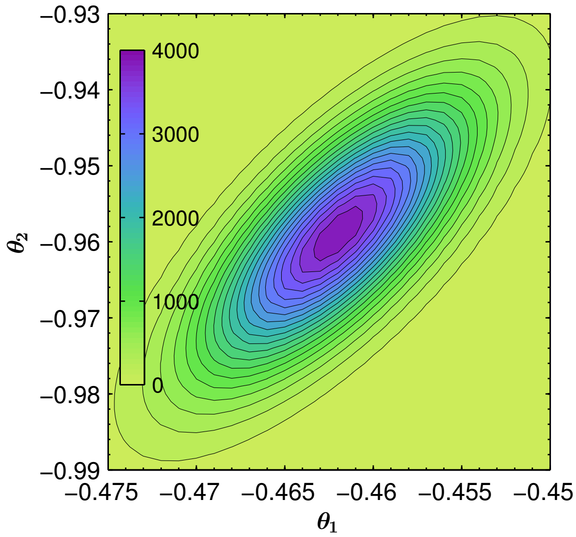


Figure 6: Contour plot of the joint posterior density for θ_1 and θ_2 when other parameters are set to be mean values of their effective samples.

arbitrarily may strongly influence the assignment results. This also indicates that the proposed choice model exhibits great potential in capturing passenger route choice behavior. On the other hand, one may see that the slowest decrease could be achieved by increasing/decreasing (θ_1, θ_2) simultaneously. This suggests that, instead of sampling each parameter separately, we may modify the Metropolis algorithm to obtain the samples of θ_1 and θ_2 collectively by considering their correlation. By doing so, we may get a faster convergence of the MCMC chains with less computation time.

In fact, in this numerical example we employed a simple model containing only two parameters to characterize passenger route choice behavior. For this special case, only in-vehicle time and transfer time are considered as important attributes influencing passenger perception. However, essentially one may take more attributes into consideration in route choice modeling, such as level of crowdedness, number of transfers (Raveau et al., 2014) and path correlation correction terms (Cascetta et al., 1996; Ben-Akiva and Bierlaire, 1999). The proposed framework has the capacity to handle a more sophisticated route choice model.

By using the route choice parameter $\boldsymbol{\theta}$, we computed the probability $f_w(r|\boldsymbol{c}, \alpha, \boldsymbol{\theta}, m)$ of choosing route r for each O-D pair w. After integrating f_w into O-D passenger demand, we obtained the flow assignment results in the studied network. We depict the assignment profile in Fig. 7. In this figure, panel (a) and (b) show the estimated link flow profiles based on passenger demand before 12p.m in both directions, while panel (c) and (d) illustrate the flow assignment of passenger demand after 12p.m. As can be seen, flow assignment shows strong heterogeneity given the specific passenger demand pattern.

6. Conclusion and Discussion

In this paper, we have made use of large quantities passenger travel time observations in a metro network to develop an integrated Bayesian approach to infer both network attributes and passenger route choice behavior. The advantage of this framework lies in the Bayesian statistical paradigm, which requires limited/partial information as input, but provides comprehensive posterior knowledge of the system.

Travel time reliability has been documented extensively in terms of urban road transport; however, as another major component of public transit, metro system attracts little attention in previous literature. One possible explanation is that metro systems have dedicated infrastructure. On the other hand, this may also due to the lack real-world travel time and route choice observations, making researchers underestimate its reliability issues: metro services have long been assumed punctual to timetables. The emergence of smart card ticketing systems, as implemented in Singapore, provide us great opportunities

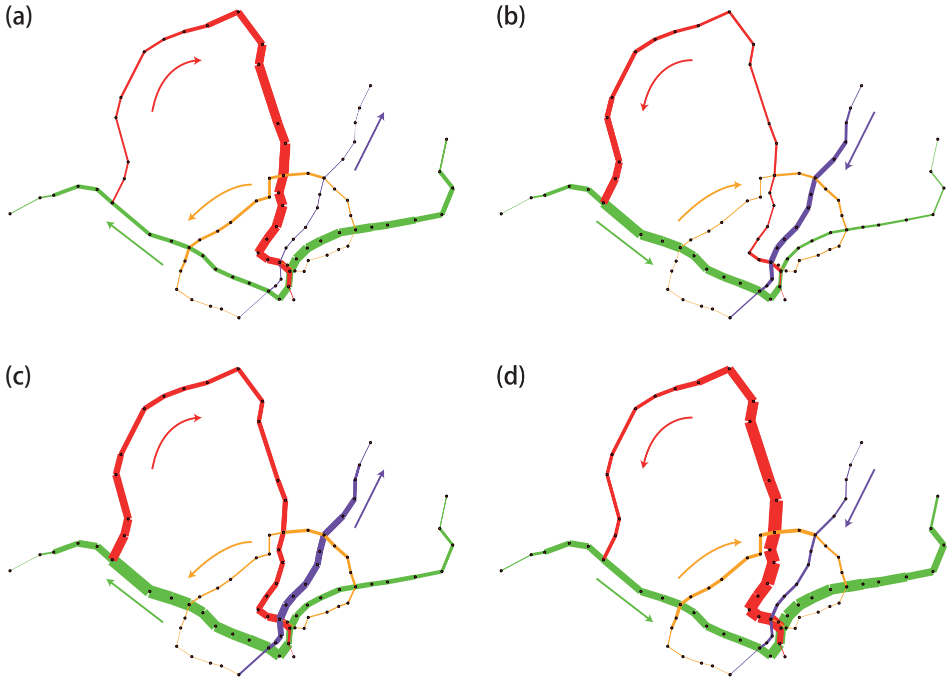


Figure 7: Passenger flow assignment in MRT network (a-b) before and (c-d) after 12p.m.

to understand travel time reliability than ever before. The inference results for link travel time and coefficient of variation offered by the proposed Bayesian framework could be applied in real-world scenarios to better predict travel time and its variability, providing metro users with better travel information.

As service reliability is highly determined by passenger demand (such as disruption caused by huge demand in peak hours), passenger flow assignment problem in a complex metro network is particularly important with respect to providing good services and sharing profit among operators. On the other hand, knowing the number of passengers traveling on each link at given time is also a central question in disruption/emergency scenarios, where operators have to make quick response such as introducing shuttle bus services (Jin et al., 2013, 2014). Previous studies use discrete choice analysis extensively to predict passenger choice behavior. However, such a model requires preference data and still displays great variability in real-world estimation. In this context, revealing route choice from observed passenger travel time, can be more advantageous (Kusakabe et al., 2010; Sun and Xu, 2012; Zhou and Xu, 2012; Zhu et al., 2014). Applying the inference results on real passenger demand, link flow profile can be estimated in temporal scale, helping us to infer temporal train loading and measure level of crowdedness. The results could also be used to reveal transfer demand to help us identify critical transfer stations/platforms/facilities, providing valuable information to operators and agencies to better design and operate the whole metro system.

Our results also have a number of potential implications for both practice and research. First, link travel time and its variability is characterized using real-world travel time observations from smart card transactions. This data-driven approach can be widely applied in other analyses. Second, the proposed cost estimation framework may help operators identify the bottleneck of a metro network; the route inference solution may contribute to better understand transit demand patterns, more accurate profit sharing and more effective disruption/emergency response. Third, by applying this framework, we can further reveal other service satisfactory indicators, such as the availability of seats, the standing and walking times; hence, the results of this paper can applied on various choice modeling problems, serving future decision making processes.

Acknowledgement

The research is funded by the National Research Foundation of Singapore, the funding authority of the Future Cities Laboratory (FCL), Singapore-ETH Centre (SEC) and Singapore-MIT Alliance for

Research and Technolog (SMART). We thank Land Transport Authority of Singapore for providing the CEPAS smart card data.

References

Bagchi, M., White, P. R., 2005. The potential of public transport smart card data. Transport Policy 12 (5), 464–474.

Barry, J. J., Newhouser, R., Rahbee, A., Sayeda, S., 2002. Origin and destination estimation in new york city with automated fare system data. Transportation Research Record: Journal of the Transportation Research Board 1817, 183–187.

Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. Springer, pp. 5–33.

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Proceedings of the 13th International Symposium on Transportation and Traffic Theory. Pergamon Lyon, France, pp. 697–711.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. Vol. 4. Oxford University Press, pp. 169–193.

Geyer, C. J., 1992. Practical markov chain monte carlo. Statistical Science 7 (4), 473–483.

Guo, Z., Wilson, N., 2007. Modeling effects of transit system transfers on travel behavior: Case of commuter rail and subway in downtown boston, massachusetts. Transportation Research Record 2006 (1), 11–20.

Hastings, W. K., 1970. Monte carlo sampling methods using markov chains and their applications. Biometrika 57, 97–109.

Hazelton, M. L., 2008. Statistical inference for time varying origin-destination matrices. Transportation Research Part B: Methodological 42 (6), 542–552.

Hazelton, M. L., 2010. Bayesian inference for network-based models with a linear inverse structure. Transportation Research Part B: Methodological 44 (5), 674–685.

Jang, W., 2010. Travel time and transfer analysis using transit smart card data. Transportation Research Record: Journal of the Transportation Research Board 2144, 142–149.

Jin, J. G., Tang, L. C., Sun, L., Lee, D.-H., 2014. Enhancing metro network resilience via localized integration with bus services. Transportation Research Part E: Logistics and Transportation Review 63, 17–30.

Jin, J. G., Teo, K. M., Sun, L., 2013. Disruption response planning for an urban mass rapid transit network. In: Transportation Research Board (TRB) 92nd Annual Meeting. Transportation Research Board.

Kusakabe, T., Iryo, T., Asakura, Y., 2010. Estimation method for railway passengers' train choice behavior with smart card transaction data. Transportation 37 (5), 731–749.

Lee, D.-H., Sun, L., Erath, A., 2012. Study of bus service reliability in singapore using fare card data. In: 12th Asia Pacific ITS Forum & Exhibition. Kuala Lumpur, Malaysia.

Li, Z., Tirachini, A., Hensher, D. A., 2012. Embedding risk attitudes in a scheduling model: Application to the study of commuting departure time. Transportation Science 46 (2), 170–188.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of state calculations by fast computing machines. Journal of Chemical Physics 21 (6), 1087–1092.

Park, J. Y., Kim, D. J., Lim, Y., 2008. Use of smart card data to define public transit use in seoul, south korea. Transportation Research Record: Journal of the Transportation Research Board 2063, 3–9.

Pelletier, M. P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies 19, 557–568.

Raveau, S., Guo, Z., Muoñz, J. C., Wilson, N. H. M., 2014. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. Transportation Research Part A: Policy and Practice 66, 185–195.

Robert, C. P., 2014. Bayesian computational tools. Annual Review of Statistics and Its Application 1 (1), 153–177.

Robert, C. P., Casella, G., 2004. Monte Carlo Statistical Methods, 2nd Edition. Springer-Verlag, New York, USA.

Strathman, J., Hopper, J., 1993. Empirical analysis of bus transit on-time performance. Transportation Research Part A: Policy and Practice 27 (2), 93–100.

Sun, L., Jin, J. G., Lee, D.-H., Axhausen, K. W., Erath, A., 2014. Demand-driven timetable design for metro services. Transportation Research Part C: Emerging Technologies 46, 284–299.

Sun, L., Lee, D.-H., Erath, A., Huang, X., 2012. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system. In: ACM SIGKDD International Workshop on Urban Computing. ACM, Beijing, China, pp. 142–148.

Sun, Y., Xu, R., 2012. Rail transit travel time reliability and estimation of passenger route choice behavior. Transportation Research Record: Journal of the Transportation Research Board 2275, 58–67.

van Nes, R., van Oort, N., 2009. Line length versus operational reliability. Transportation Research Record: Journal of the Transportation Research Board 2112 (-1), 104–110.

Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: Evidence and lessons from british experience. Transport Reviews 31 (3), 379–398.

Wei, C., Asakura, Y., 2013. A Bayesian approach to traffic estimation in stochastic user equilibrium networks. Transportation Research Part C: Emerging Technologies 36, 446–459.

Westgate, B. S., Woodard, D. B., Matteson, D. S., Henderson, S. G., 2013. Travel time estimation for ambulances using Bayesian data augmentation. The Annals of Applied Statistics 7 (2), 1139–1161.

Zhou, F., Xu, R., 2012. Passenger flow assignment model for urban rail transit based on entry and exit time constraints. Transportation Research Record: Journal of the Transportation Research Board 2284, 57–61.

Zhu, W., Hu, H., Huang, Z., 2014. Calibrating rail transit assignment models with genetic algorithm and automated fare collection data. Computer-Aided Civil and Infrastructure Engineering 29 (7), 518–530.