

## Research Article

# Splitting Travel Time Based on AFC Data: Estimating Walking, Waiting, Transfer, and In-Vehicle Travel Times in Metro System

Yong-Sheng Zhang<sup>1,2</sup> and En-Jian Yao<sup>2</sup>

<sup>1</sup>MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup>School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Yong-Sheng Zhang; 12114241@bjtu.edu.cn

Received 31 July 2015; Accepted 12 November 2015

Academic Editor: Charalampos Skokos

Copyright © 2015 Y.-S. Zhang and E.-J. Yao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The walking, waiting, transfer, and delayed in-vehicle travel times mainly contribute to route's travel time reliability in the metro system. The automatic fare collection (AFC) system provides huge amounts of smart card records which can be used to estimate all these times distributions. A new estimation model based on Bayesian inference formulation is proposed in this paper by integrating the probability measurement of the OD pair with only one effective route, in which all kinds of times follow the truncated normal distributions. Then, Markov Chain Monte Carlo method is designed to estimate all parameters endogenously. Finally, based on AFC data in Guangzhou Metro, the estimations show that all parameters can be estimated endogenously and identifiably. Meanwhile, the truncated property of the travel time is significant and the threshold tested by the surveyed data is reliable. Furthermore, the superiority of the proposed model over the existing model in estimation and forecasting accuracy is also demonstrated.

## 1. Introduction

Travel time reliability in terms of variation in travel time has attracted more and more attention recently, not only in the road traffic, but also in the metro system. In the metro system, travel time consists of entry walking time, entry waiting time, in-vehicle travel time, transfer travel time, exit walking time, and so forth. As is well known, walking and waiting time are of high variability, while in-vehicle travel time has long been considered punctual (Kusakabe et al. [1]; Sun and Xu [2]; Zhou and Xu [3]), even though it may be delayed by excessive demand, especially during peak hours. Splitting travel time is referred to as estimating the entry walking, entry waiting, and exit walking time in every station, in-vehicle travel time in every section between two successive stations, and transfer travel time in every transfer station. All these kinds of times provide us with the level of services of nontransfer stations, sections, and transfer stations. They can help us calculate the congestion levels in the walking channels and waiting platforms, evaluate transfer efficiency between two lines, find the train delay in the sections, optimize the train schedule

even under the cooperated operation condition, estimate passengers' route choice behavior, and so forth. However, measuring these times is extremely challenging at a network level even when carrying out a field survey in the metro system.

Fortunately, the automatic fare collection (AFC) system is currently widely used to collect smart cards data in the metro system. This data is called AFC data and records every passenger's travel details, including entry station, exit station, and the corresponding times when swiping their cards. Based on AFC data, much research has been done, such as trip generation prediction (Guang et al. [4]; Cai et al. [5]), Origin-Destination (OD) distribution prediction (Cai et al. [6]; Cai et al. [7]; Chapleau et al. [8]; Rahbee [9]), route choice proportion estimation (Sun and Schonfeld [10]; Zhu et al. [11]), and travelers' characteristics extraction (Lee and Hickman [12]). For more application, see Pelletier et al.'s review [13]. Research has also been done to estimate various times using AFC data. Sun and Xu [14] assumed that walking and waiting time were all random, but in-vehicle travel time was punctual since it was determined by fixed

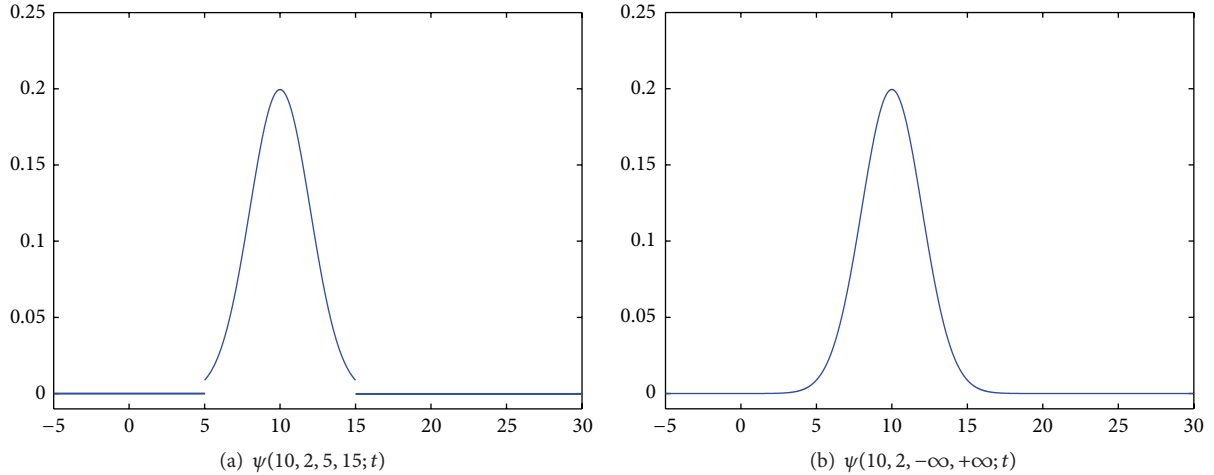


FIGURE 1: The example for the PDFs of truncated normal distribution and general normal distribution.

train schedules. Zhou et al. [15] assumed that, except waiting time following uniform distribution, other types of times yielded to normal distribution, including in-vehicle travel time, and they were estimated by moment estimation method based on the maximum spanning tree. But in this approach, a field survey was still needed to calculate the first station's entry time. In Sun et al.'s research [16], all kinds of times were supposed to independently yield to normal distribution and the parameters were estimated by the Maximum Likelihood Estimation (MLE) approach without any additional surveys. Sun et al. [17] proposed an integrated Bayesian statistical inference framework to characterize the passenger flow assignment model in a complex metro network. In this research, all kinds of times were still assumed to follow normal distribution and a Bayesian approach was designed to estimate all parameters, along with the chosen proportions of feasible routes.

However, the distribution assumptions of all kinds of times are hardly consistent with the facts because the truncated property of the real data is neglected. In practice, all times have their ranges; for example, with respect to the walking time which relates to the travel time of free flows and the congestion degrees, it cannot be very short (e.g., being close to 0) or long (e.g., tending to infinity) which means it is truncated. Meanwhile, on account of the fact that the AFC data only records the travel time which is the period between swiping-in time at the origin station and swiping-out time at the destination station, it cannot indicate which route the passenger has really traveled between the OD pairs with multiple feasible routes (e.g., without loops). Therefore, it is meaningful to select the OD pairs with only one effective route to exactly split the travel time of the route into links and then estimate all kinds of times. Generally, a threshold is set to determine whether the second shortest route is considered or not and then the OD pair with only one effective route can be selected (Sun et al. [16]). However, in the existing research, they usually neglect the random property of the OD pair with only one effective route due to the travel time reliability of the routes and calibrate the threshold exogenously.

Therefore, in order to estimate the walking, waiting, in-vehicle, and transfer times more precisely, this paper proposes an approach in which the contributions are as follows: all kinds of times follow the truncated normal distributions, the random property of the OD pair with only one effective route is considered, and the threshold is calibrated endogenously. The left sections are organized as follows: Section 2 mainly relates to the methodology, in which the property of truncated normal distribution is introduced, the method selecting the OD pair with only one effective route is proposed, and a Bayesian inference formulation is established together with the estimation approach based on Markov Chain Monte Carlo (MCMC) method; in Section 3, based on AFC data, train schedules, topology network, and other data in the Guangzhou Metro, the estimations are discussed and estimation errors between the existing and the proposed models are compared; Section 4 is the conclusions.

## 2. Methodology

It should be mentioned here that, in this paper, the unit of all kinds of times is minute, taking account of the fact that, in terms of the time of the link, it is usually several minutes while for the travel time of the route it is usually tens of minutes.

*2.1. Truncated Normal Distribution.* On the assumption that random variable  $T$  follows a normal distribution  $T \sim N(\bar{\mu}, \bar{\sigma}^2)$ , if  $T$  lies within the interval  $T \in (a, b)$ ,  $T$  conditional on  $a < T < b$  has a truncated normal distribution  $T \sim TN(\bar{\mu}, \bar{\sigma}^2, a, b)$ , and its probability density function (PDF) is defined as follows, along with the curve shown in Figure 1 (Johnson et al. [18]):

$$\psi(\bar{\mu}, \bar{\sigma}, a, b; t) = \begin{cases} 0, & \text{if } t \leq a \\ \frac{\phi(\bar{\mu}, \bar{\sigma}; t)}{\Phi(\bar{\mu}, \bar{\sigma}; b) - \Phi(\bar{\mu}, \bar{\sigma}; a)}, & \text{if } a < t < b \\ 0, & \text{if } t \geq b, \end{cases} \quad (1)$$

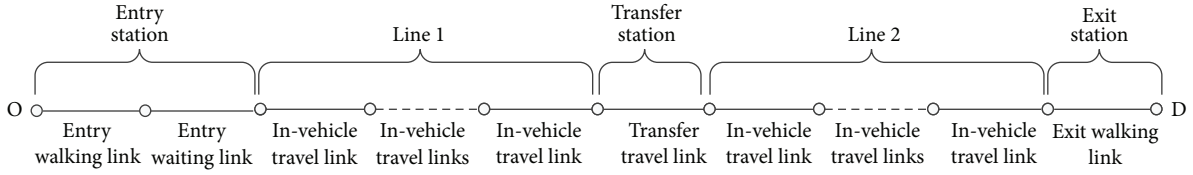


FIGURE 2: The example for the network used in this paper.

where  $\bar{\mu}, \bar{\sigma}$  are the mean and standard derivation of normal distribution,  $\psi(\cdot)$  is the PDF of truncated normal distribution,  $\phi(\cdot)$  is the PDF of normal distribution, and  $\Phi(\cdot)$  is the cumulative density function (CDF) of normal distribution.

Generally, in the metro system, travel time contains entry walking time, entry waiting time, in-vehicle travel time, transfer travel time, exit walking time, and so forth. Suppose that every type of time is represented by a link in the metro network shown in Figure 2. It depicts that a route from O to D may consist of multiple links which represent types of times. The transfer travel link represents the sum of transfer walking and waiting times which are hard to be distinguished as they always appear at the same time for a transfer station. In practice, all these times are truncated within intervals; that is, with respect to the time of link  $i$ , it follows a truncated normal distribution  $T_{li} \sim TN(\bar{\mu}_i, \bar{\sigma}_i^2, a_i, b_i; t_i)$ . For the travel time  $T_k$  of the route  $k$ , it is the sum of all kinds of mentioned times; that is,  $T_k = \sum_{i \in \mathbf{H}_k} T_{li}$ , where  $\mathbf{H}_k$  denotes the set of links constituting route  $k$ . Supposing that the times of every link are independent of each other, the travel time of the route still follows the truncated normal distribution whose probability density function  $\psi(\bar{\mu}'_k, \bar{\sigma}'_k, a'_k, b'_k; t)$  is shown:

$$\begin{aligned} \psi(\bar{\mu}'_k, \bar{\sigma}'_k, a'_k, b'_k; t) &= \sum_{i \in \mathbf{H}_k} \psi(\bar{\mu}_i, \bar{\sigma}_i, a_i, b_i; t) \\ &= \psi\left(\sum_{i \in \mathbf{H}_k} \bar{\mu}_i, \sqrt{\sum_{i \in \mathbf{H}_k} \bar{\sigma}_i^2}, \sum_{i \in \mathbf{H}_k} a_i, \sum_{i \in \mathbf{H}_k} b_i; t\right). \end{aligned} \quad (2)$$

The mean  $\mu'_k$  and the variance  $\sigma_k'^2$  of truncated normal distribution of route  $k$  can be regarded as a perturbation of the mean  $\bar{\mu}'_k$  and the variance  $\bar{\sigma}'_k$  of parent normal distribution, respectively, which can be derived from moment equations:

$$\begin{aligned} \mu'_k &= \int_{\sum_{i \in \mathbf{H}_k} a_i}^{\sum_{i \in \mathbf{H}_k} b_i} t \cdot \psi(\bar{\mu}'_k, \bar{\sigma}'_k, a'_k, b'_k; t) dt \\ &= \int_{\sum_{i \in \mathbf{H}_k} a_i}^{\sum_{i \in \mathbf{H}_k} b_i} t \cdot \psi\left(\sum_{i \in \mathbf{H}_k} \bar{\mu}_i, \sqrt{\sum_{i \in \mathbf{H}_k} \bar{\sigma}_i^2}, \sum_{i \in \mathbf{H}_k} a_i, \sum_{i \in \mathbf{H}_k} b_i; t\right) dt \quad (3) \\ &= \sum_{i \in \mathbf{H}_k} \bar{\mu}_i + \sqrt{\sum_{i \in \mathbf{H}_k} \bar{\sigma}_i^2} \Lambda_1 \end{aligned}$$

$$\begin{aligned} \sigma_k'^2 &= \int_{\sum_{i \in \mathbf{H}_k} a_i}^{\sum_{i \in \mathbf{H}_k} b_i} t^2 \cdot \psi(\bar{\mu}'_k, \bar{\sigma}'_k, a'_k, b'_k; t) dt \\ &= \int_{\sum_{i \in \mathbf{H}_k} a_i}^{\sum_{i \in \mathbf{H}_k} b_i} t^2 \cdot \psi\left(\sum_{i \in \mathbf{H}_k} \bar{\mu}_i, \sqrt{\sum_{i \in \mathbf{H}_k} \bar{\sigma}_i^2}, \sum_{i \in \mathbf{H}_k} a_i, \sum_{i \in \mathbf{H}_k} b_i; t\right) dt \quad (4) \\ &= \sum_{i \in \mathbf{H}_k} \bar{\sigma}_i^2 \cdot (\Lambda_2 - \Lambda_1^2), \end{aligned}$$

where  $\Lambda_1 = -(\phi(0, 1; \beta) - \phi(0, 1; \alpha)) / (\Phi(0, 1; \beta) - \Phi(0, 1; \alpha))$ ;  $\Lambda_2 = -(\beta \cdot \phi(0, 1; \beta) - \alpha \cdot \phi(0, 1; \alpha)) / (\Phi(0, 1; \beta) - \Phi(0, 1; \alpha)) + 1$ ;  $\alpha = (\sum_{i \in \mathbf{H}_k} a_i - \sum_{i \in \mathbf{H}_k} \bar{\mu}_i) / \sqrt{\sum_{i \in \mathbf{H}_k} \bar{\sigma}_i^2}$ ;  $\beta = (\sum_{i \in \mathbf{H}_k} b_i - \sum_{i \in \mathbf{H}_k} \bar{\mu}_i) / \sqrt{\sum_{i \in \mathbf{H}_k} \bar{\sigma}_i^2}$ .

**2.2. The OD Pairs with Only One Effective Route.** Though the AFC data records the travel times between OD pairs, not all travel times data can be used to estimate distribution parameters because there may be multiple effective routes between some OD pairs. This paper just focuses on the OD pairs with only one effective route. Considering the randomness of the travel time of the route, except the OD pairs between which there is only one feasible route (e.g., no loop) due to the topological structure, other OD pairs with only one effective route are random. Suppose that there are more than two feasible routes between a specific OD pair. Before calculating the probability of this OD pair with only one effective route, the routes with the first and second shortest travel time should be generated. Usually, the routes with the shortest travel time and the second shortest travel time can be selected out easily according to K-shortest routes algorithms if they are constant. In the reliable context, researchers paid more attention to the route with the reliable shortest travel time (Khani and Boyles [19]). Taking account of the fact that the scale of metro network is comparatively smaller than the road network, this paper firstly generates amounts of feasible routes (e.g., at most ten routes) for the OD pair based on the route's physical length. Then, it selects the routes with the first and second reliable shortest travel time from the generated routes set  $\mathbf{R}$  according to the below minimization formulation.

The route with the reliable shortest travel time is the solution to the minimization formulation as follows:

$$\text{Min } Z_1 = \mu'_k + \lambda \sigma'_k, \quad k \in \mathbf{R}, \quad (5)$$

where  $\lambda$  is a parameter indicating the value of route reliability (measured by standard deviation  $\sigma'_k$ ) relative to the expected

travel time  $\mu'_k$ . And the route with the second reliable shortest travel time is the solution to the same minimization formulation after removing the route with the reliable shortest travel time from the routes set  $\mathbf{R}$ .

Let  $T_1$  denote the reliable shortest travel time and  $T_2$  is the second reliable shortest travel time. They both follow the truncated normal distributions, that is,  $T_1 \sim TN(\bar{\mu}'_1, \bar{\sigma}'_1, a'_1, b'_1; t)$  and  $T_2 \sim TN(\bar{\mu}'_2, \bar{\sigma}'_2, a'_2, b'_2; t)$ , respectively. The probability of the OD pair with only one effective route can be measured by the probability formulation as follows:

$$P(s | Y^s) = P(T_2^s - T_1^s \geq Y^s) + \text{eps} = \int_{Y^s}^{+\infty} \psi\left(\bar{\mu}'_2 - \bar{\mu}'_1, \sqrt{(\bar{\sigma}'_2)^2 + (\bar{\sigma}'_1)^2}, a'_2 - b'_1, b'_2 - a'_1; t\right) dt \quad (6)$$

+ eps

$$Y^s = \alpha \cdot \ln\left(\frac{\bar{\mu}'_1}{60} + 1\right), \quad (7)$$

where  $Y^s$  is the threshold value for OD pair  $s$  which varies with the OD pairs due to the scale effects of OD pairs;  $\alpha$  is the parameter; and eps is an extreme positive value tending to 0 (e.g., 0.000001 in this paper). The purpose to add eps is to avoid the probability 0 of the OD pair with only one effective route, that is, to satisfy the condition  $P(s | Y^s) > 0$  to carry out the estimation method below. The above equations mean that, for the OD pair with multiple feasible routes, the probability of the OD pair with only one effective route is equal to the probability that the difference between the second reliable shortest travel time and the reliable shortest travel time is larger than the threshold value. Obviously, if the threshold value is larger than the upper boundary  $Y^s \geq b'_2 - a'_1$ , then  $P(T_2^s - T_1^s \geq Y^s) = 0$ , and the probability of the OD pair with only one effective route tends to 0; that is,  $P(s | Y^s) = \text{eps}$ ; if the threshold value is smaller than the lower boundary  $Y^s \leq a'_2 - b'_1$ , the probability  $P(T_2^s - T_1^s \geq Y^s)$  is equal to 1 and  $P(s | Y^s) = 1 + \text{eps} \approx 1$ .

**2.3. Bayesian Inference.** Let  $\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}$ , and  $\mathbf{b}$  represent the sets of the parent mean, variance, interval lower boundary, and upper boundary of all kinds of times, respectively, that is,  $\mathbf{T} \sim TN(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b})$ . Together with the threshold  $\mathbf{Y}$ , they need to be estimated based on the AFC data. The complex structure of likelihood function as shown in (9) makes Maximum Likelihood Estimation method hard to be applied when trying to estimate distribution parameters and threshold parameter endogenously in this paper. Therefore, a Bayesian inference formulation is established. The Bayesian posterior distribution of unknown parameters given the travel time observations can be derived based on Bayes' theorem as follows:

$$\begin{aligned} & \pi(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, \mathbf{Y} | \mathbf{T}) \\ & \propto p(\mathbf{T} | \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, \mathbf{Y}) \cdot \pi(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, \mathbf{Y}) \\ & = p(\mathbf{T} | \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, \mathbf{Y}) \cdot \pi(\bar{\boldsymbol{\mu}}) \cdot \pi(\bar{\boldsymbol{\sigma}}) \cdot \pi(\mathbf{a}) \cdot \pi(\mathbf{b}) \\ & \quad \cdot \pi(\mathbf{Y}) \end{aligned} \quad (8)$$

$$\begin{aligned} p(\mathbf{T} | \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, \mathbf{Y}) & \propto \prod_{s \in \mathbf{S}} \psi(\mathbf{T}_s | \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, s) p(s | Y^s) \\ & = \prod_{s \in \mathbf{S}, t \in \mathbf{T}_s} \psi(t | \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, s) p(s | Y^s), \end{aligned} \quad (9)$$

where  $\pi(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, \mathbf{Y} | \mathbf{T})$  denotes the density function of all parameters given the observations;  $\mathbf{S}$  is the set of OD pairs;  $\mathbf{T}_s$  is the travel time observations of OD pair  $s$ ;  $p(\cdot)$  is equal to the likelihood function of all parameters given the travel time observations;  $\pi(\bar{\boldsymbol{\mu}})$ ,  $\pi(\bar{\boldsymbol{\sigma}})$ ,  $\pi(\mathbf{a})$ ,  $\pi(\mathbf{b})$ , and  $\pi(\mathbf{Y})$  are the prior distributions. At last, taking account of the threshold parameter  $\alpha$ , the Bayesian formulation is shown as follows:

$$\begin{aligned} \pi(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}, \alpha | \mathbf{T}) & \propto \prod_{s \in \mathbf{S}, t \in \mathbf{T}_s} [\psi(t | \bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\sigma}}, \mathbf{a}, \mathbf{b}) \\ & \quad \cdot p(s | Y^s) \cdot \pi(\bar{\boldsymbol{\mu}}) \cdot \pi(\bar{\boldsymbol{\sigma}}) \cdot \pi(\mathbf{a}) \cdot \pi(\mathbf{b}) \cdot \pi(\alpha)]. \end{aligned} \quad (10)$$

The prior distribution needs to be determined in advance though the number of travel time observations in metro system is large enough to revise the prior knowledge to a great extent.

**2.4. Estimation Method.** Markov Chain Monte Carlo (MCMC) method allows us to simulate draws that are slightly dependent and are approximately from a posterior distribution. Then, those draws can be taken to calculate quantities of interest for the posterior distribution. In Bayesian statistics, Gibbs Sampler and Metropolis-Hastings (M-H) algorithm (Metropolis et al. [20]; Hastings [21]) are widely used MCMC methods. Without the prior knowledge among parameters and the full conditional distributions for each parameter, M-H algorithm is more appropriate to estimate the parameters in this paper. Considering that it is a high-dimensional problem to solve so many parameters, the variable-at-a-time Metropolis sampling scheme (Metropolis et al. [20]) is used to avoid the large rejection rate perhaps caused by general Metropolis. Let  $\boldsymbol{\zeta}$  be the set of all parameters; that is,  $\boldsymbol{\zeta} = \bar{\boldsymbol{\mu}} \cup \bar{\boldsymbol{\sigma}} \cup \mathbf{a} \cup \mathbf{b} \cup \alpha$ . In the variable-at-a-time scheme, it generates new sample for each coordinate (parameter) in turn in the parameters' set. The estimation approach is as follows.

*Step 1.* Choose a starting value  $\boldsymbol{\zeta}^{(0)} = (\zeta_1^{(0)}, \dots, \zeta_i^{(0)}, \dots, \zeta_N^{(0)})$ , where  $N$  denotes the number of parameters. And generate routes sets for OD pairs based on physical length. Set iteration  $m = 1$ .

*Step 2.* Update the travel time distribution of feasible routes based on the distribution parameters of links and select out the routes with the first and second reliable shortest travel time based on minimization formulation (5).

*Step 3.* Draw a candidate  $\boldsymbol{\zeta}^* = (\zeta_1^{(m)}, \dots, \zeta_{i-1}^{(m)}, \zeta_i^*, \zeta_{i+1}^{(m-1)}, \dots, \zeta_N^{(m-1)})$  from a jumping distribution  $J_m(\zeta_i^* | \zeta_i^{(m-1)})$  in sequential order ( $i = 1, \dots, N$ ).

Here, we suppose that the jumping distribution is a normal distribution which is a symmetric distribution; that

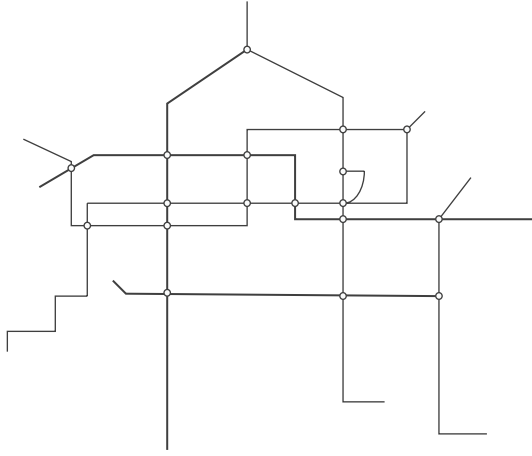


FIGURE 3: Guangzhou Metro Map (2014).

is,  $J_m(\zeta_i^* | \zeta_i^{(m-1)}) = J_m(\zeta_i^{(m-1)} | \zeta_i^*)$ . Generally, the jumping distribution can be denoted as  $J_m(\zeta_i^* | \zeta_i^{(m-1)}) \sim N(\zeta_i^{(m-1)}, \xi^2)$ , where  $\xi^2$  is the proposal variance for the  $i$ th parameter. This is known as Gaussian random walk Metropolis sampling.

*Step 4.* Calculate the acceptance ratio  $\kappa = \min\{1, \frac{\pi'(\zeta_i^*)J_m(\zeta_i^{(m-1)} | \zeta_i^*)/\pi'(\zeta_i^{(m-1)})J_m(\zeta_i^* | \zeta_i^{(m-1)})}{\pi'(\zeta_i^*)/\pi'(\zeta_i^{(m-1)})}\}$ . In this equation, the posterior distributions are calculated as follows:

$$\begin{aligned} \pi'(\zeta_i^*) &= \pi(\zeta_1^{(m)}, \dots, \zeta_{i-1}^{(m)}, \zeta_i^*, \zeta_{i+1}^{(m-1)}, \dots, \zeta_N^{(m-1)} | \mathbf{T}) \\ &\propto \prod_{s \in \mathcal{S}, t \in \mathcal{T}_s} \left[ p(t | \zeta_1^{(m)}, \dots, \zeta_{i-1}^{(m)}, \zeta_i^*, \zeta_{i+1}^{(m-1)}, \dots, \zeta_N^{(m-1)}) \right. \\ &\quad \left. \cdot \prod_{j=1}^{i-1} \pi(\zeta_j^{(m)}) \cdot \prod_{j=i+1}^N \pi(\zeta_j^{(m-1)}) \cdot \pi(\zeta_i^*) \right] \end{aligned} \quad (11)$$

$$\pi'(\zeta_i^{(m-1)}) = \pi(\zeta_1^{(m)}, \dots, \zeta_{i-1}^{(m)}, \zeta_i^{(m-1)}, \zeta_{i+1}^{(m-1)}, \dots, \zeta_N^{(m-1)} |$$

$$\mathbf{T}) \propto \prod_{s \in \mathcal{S}, t \in \mathcal{T}_s} \left[ p(t | \zeta_1^{(m)}, \dots, \zeta_{i-1}^{(m)}, \zeta_i^{(m-1)}, \zeta_{i+1}^{(m-1)}, \dots,$$

$$\zeta_N^{(m-1)}) \cdot \prod_{j=1}^{i-1} \pi(\zeta_j^{(m)}) \cdot \prod_{j=i+1}^N \pi(\zeta_j^{(m-1)}) \cdot \pi(\zeta_i^{(m-1)}) \right],$$

where  $p(t | \zeta_1^{(m)}, \dots, \zeta_N^{(m-1)}) = \psi(t | \zeta_1^{(m)}, \dots, \zeta_{N-1}^{(m-1)}) \cdot p[s | Y^s(\zeta_N^{(m-1)})]$ .  $p[s | Y^s(\zeta_N^{(m-1)})] > 0$  ensures that the denominator is not equal to 0. After canceling out the same parts, the ratio can be simplified as

$$\kappa = \min \left\{ 1, \frac{\prod_{s \in \mathcal{S}, t \in \mathcal{T}_s} [p(t | \zeta_1^{(m)}, \dots, \zeta_{i-1}^{(m)}, \zeta_i^*, \zeta_{i+1}^{(m-1)}, \dots, \zeta_N^{(m-1)}) \cdot \pi(\zeta_i^*)]}{\prod_{s \in \mathcal{S}, t \in \mathcal{T}_s} [p(t | \zeta_1^{(m)}, \dots, \zeta_{i-1}^{(m)}, \zeta_i^{(m-1)}, \zeta_{i+1}^{(m-1)}, \dots, \zeta_N^{(m-1)}) \cdot \pi(\zeta_i^{(m-1)})]} \right\}. \quad (12)$$

*Step 5.* Draw a value  $u$  from the uniform (0, 1) distribution. If  $u \leq \kappa$ ,  $\zeta_i^{(m)} = \zeta_i^*$ ; otherwise,  $\zeta_i^{(m)} = \zeta_i^{(m-1)}$ .

*Step 6.* If  $m < M$ ,  $m = m + 1$ , repeat Steps 2–5; otherwise, stop sampling.

### 3. Results and Discussion

As shown in Figure 3, in the year of 2014, except the APM line, there are 8 lines and 136 stations (including 19 transfer stations) in operation in Guangzhou Metro. APM is the abbreviation of Zhujiang New Town Automated People Mover System which is also operated by Guangzhou Metro Corporation but independent of other lines. People should swipe the smart card again to go through the APM line even if they transfer from other lines. Therefore, it is excluded in this paper. The number of OD pairs is up to 18360. There are about 3 million transactions every day in which the maximum amounts of transactions are about 6000 between OD pair while, for few OD pairs, there are only several transactions. Here, we use one-month data to keep the distribution property, that is, the data in June for estimation and the data in July for testing. There are some hypotheses made here: the

entry and exit walking time relating to the same line at the same station are the same while they are different relating to different lines even at the same (transfer) station; the waiting times relating to different lines even at the same (transfer) station are different but they are the same relating to the same line's different directions; the transfer times in the same transfer station from the same line transferring to another line's different directions are different; in-vehicle travel times of different directions between the same successive stations are the same; and all kinds of times follow the truncated normal distribution.

Figure 4 shows the characteristics of the AFC data collected from Tianhe Coach Terminal to Gangding on June 18, 2014. The number of observations is 6328, the average travel time is 11.61 min, and the standard derivation is 3.39 min. This clearly indicates that the travel time follows approximately a truncated normal distribution. Meanwhile, the lower boundary is closer to the average value than the upper boundary. The far longer travel time than the average value may result from the common fact that some passengers miss the trains or just walk slowly. It can be found that this phenomenon exists in most OD pairs according to the AFC data. Therefore, the truncated normal distribution assumption is more suitable for the fact and the estimations should satisfy that the



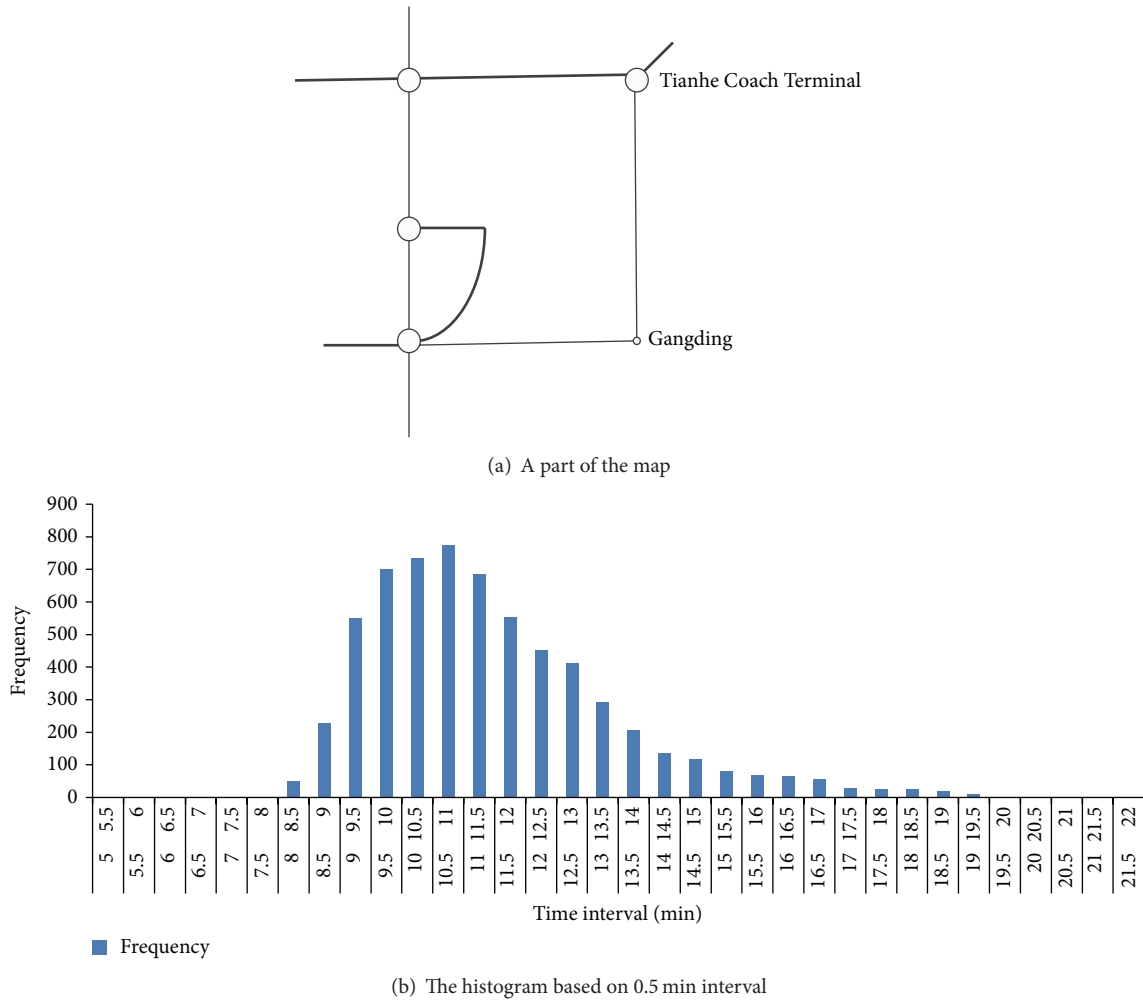


FIGURE 4: An example for the travel time distribution from Tianhe Coach Terminal to Gangding.

difference between the upper boundary and the mean should be larger than the difference between the mean and the lower boundary.

**3.1. Estimation Results.** The prior knowledge is noninformative; thus, all parameters are assumed to follow uniform distributions according to the operational experiences and the train schedules. The in-vehicle travel time distribution can be easily estimated based on the real train schedule. However, the real train schedule is hard to be collected even for the operation corporation. Therefore, in-vehicle travel time distributions are also estimated in this paper and the prior knowledge can be gained from the planned train schedule. The ratio between the standard derivation and the mean of the in-vehicle travel time is recognized as a constant (Sun et al. [17]); that is,  $\sigma^{vt} = \beta \cdot \bar{\mu}^{vt}$ . In order to distinguish the parameters conveniently, we use  $e_e$ ,  $e_w$ ,  $t_t$ , and  $v_t$  to denote entry/exit walking time, entry waiting time, transfer travel time, and in-vehicle travel time, respectively. The total number of parameters is  $156 * 4 + 156 * 4 + 147 * 3 + 1 + 168 * 4 + 1 = 2363$  (156 entry/exit walking links, 156 entry waiting links, 147 in-vehicle travel links, 1

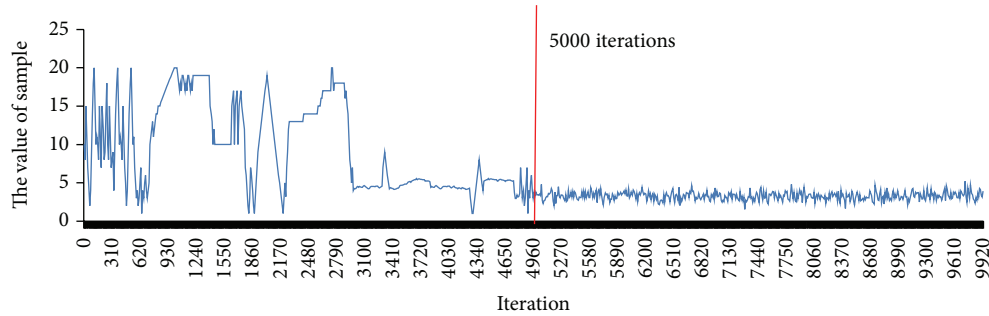
constant measuring the ratio between the standard derivation and the mean of the in-vehicle travel time, 168 transfer travel links, and 1 threshold parameter). The random walk sampling variance  $\xi^2$  is 15 min for entry/exit walking time, 7 min for entry waiting time, 15 min for transfer travel time, 2 min for in-vehicle travel time, and 10 min for threshold parameter, respectively. Though the large quantities of data in Guangzhou Metro system can correct the biased prior knowledge, an appropriate distribution assumption is still needed to save the simulating time. For the same type of the time, the prior distributions of some types of parameters are the same which are exemplified in Table 1. The initial value of each parameter for estimation is the mean.

The splitting model proposed in this paper avoids selecting particular OD pairs in advance, such as the OD pairs with only one feasible route. But the routes between the OD pairs should cover all links. Here, we randomly choose 5000 OD pairs (excluding the OD pair with OD volume smaller than 1000 for a month) and, for each OD pair, 100 travel time observations collected in June, 2014, are chosen. For each OD pair, the observations with 2 times greater or shorter travel time than the average travel time are not drawn for

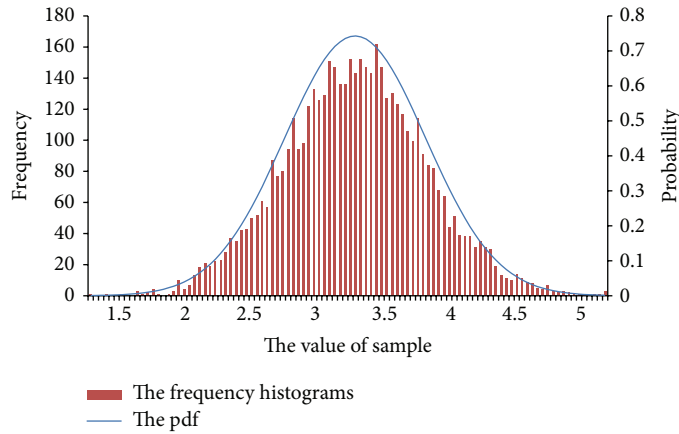
TABLE 1: The prior distributions of some parameters.

Entry/exit walking time (min)		Entry waiting time (min)		Transfer time (min)		In-vehicle time (min)		Threshold parameter	
$\bar{\mu}_1^{ec}$	$U(1, 20)$	$\bar{\mu}_1^{ew}$	$U(0, 10)$	$\bar{\mu}_1^{tt}$	$U(1, 20)$	$\bar{\mu}_1^{vt}$	$U(1, 5)$	$\alpha$	$U(0, 20)$
$\bar{\sigma}_1^{ec}$	$U(0, 20)$	$\bar{\sigma}_1^{ew}$	$U(0, 10)$	$\bar{\sigma}_1^{tt}$	$U(0, 20)$	$\beta$	$U(0, 1)$	—	—
$a_1^{ec}$	$U(0, 20)$	$a_1^{ew}$	$U(0, 10)$	$a_1^{tt}$	$U(0, 20)$	$a_1^{vt}$	$U(0, 5)$	—	—
$b_1^{ec}$	$U(1, 50)$	$b_1^{ew}$	$U(0, 30)$	$b_1^{tt}$	$U(1, 50)$	$b_i$	$U(1, 20)$	—	—
...	...	...	...	...	...	...	...	—	—
...	...	...	...	...	...	$\bar{\mu}_{147}^{vt}$	$U(1, 5)$	—	—
$\bar{\mu}_{156}^{ec}$	$U(1, 20)$	$\bar{\mu}_{156}^{ew}$	$U(0, 10)$	...	...	$\beta$	$U(0, 1)$	—	—
$\bar{\sigma}_{156}^{ec}$	$U(0, 20)$	$\bar{\sigma}_{156}^{ew}$	$U(0, 10)$	$\bar{\mu}_{168}^{tt}$	$U(1, 20)$	$a_{147}^{vt}$	$U(0, 5)$	—	—
$a_{156}^{ec}$	$U(0, 20)$	$a_{156}^{ew}$	$U(0, 10)$	$\bar{\sigma}_{168}^{tt}$	$U(0, 20)$	$b_{147}^{vt}$	$U(1, 20)$	—	—
$b_{156}^{ec}$	$U(1, 50)$	$b_{156}^{ew}$	$U(0, 30)$	$a_{168}^{tt}$	$U(0, 20)$	—	—	—	—
—	—	—	—	$b_{168}^{tt}$	$U(1, 50)$	—	—	—	—

Note: ... means the rest; — means not applicable.



(a) Iteration tendency



(b) Histograms and pdf

FIGURE 5: An example for statistical analysis on the samples of the parameter.

estimation in case that some abnormal data is excluded. The parameter  $\lambda$  which indicates the value of standard deviation relative to the mean is assumed to be 1 in this paper. We take 10000 iterations of which the former 5000 iterations are burn-in period. Simulating in the PC with Intel Core i3-2130 CPU at 3.40 GHz and 4.00 GB RAM, the simulation takes about 20 s for each iteration and at last 5000 effective samples for each parameter are drawn. By analyzing the effective samples of the parameters via histograms and the

results' tendency of the iterations which are shown in Figure 5 as an example, they are all converged. From Figure 5(a), we can see that, after the burn-in period, the values of the samples keep stable which means that the iterations are converged. Meanwhile, in Figure 5(b), the histograms express the frequency distribution of the effective sampled parameter values which shows that the samples of the parameter follow the normal distribution. Figure 5 indicates that the drawn samples are effective significantly and the estimations are

TABLE 2: The estimations of some parameters.

Entry/exit walking time (min)		Entry waiting time (min)		Transfer time (min)		In-vehicle time (min)	
Para.	Mean (CI)	Para.	Mean (CI)	Para.	Mean (CI)	Para.	Mean (CI)
$\bar{\mu}_1^{ee}$	3.31 (2.42, 5.02)	$\bar{\mu}_1^{ew}$	2.03 (1.91, 2.52)	$\bar{\mu}_1^{tt}$	4.22 (3.82, 6.06)	$\bar{\mu}_1^{vt}$	2.22 (1.92, 2.82)
$\bar{\sigma}_1^{ee}$	1.22 (1.04, 2.13)	$\bar{\sigma}_1^{ew}$	1.01 (0.92, 1.22)	$\bar{\sigma}_1^{tt}$	1.21 (1.01, 2.02)	$\beta$	0.04 (0.01, 0.10)
$a_1^{ee}$	1.42 (1.11, 2.31)	$a_1^{ew}$	0.05 (0.00, 0.10)	$a_1^{tt}$	2.83 (1.73, 3.65)	$a_1^{vt}$	1.71 (1.43, 2.12)
$b_1^{ee}$	8.03 (5.97, 9.88)	$b_1^{ew}$	6.22 (5.11, 7.31)	$b_1^{tt}$	13.33 (12.11, 14.86)	$b_{147}^{vt}$	3.02 (1.76, 4.38)
...	...	...	...	...	...	...	...
...	...	...	...	...	...	$\bar{\mu}_{147}^{vt}$	1.67 (1.11, 2.51)
$\bar{\mu}_{156}^{ee}$	5.31 (4.51, 6.11)	$\bar{\mu}_{156}^{ew}$	2.65 (1.83, 3.10)	...	...	$\beta$	0.04 (0.01, 0.10)
$\bar{\sigma}_{156}^{ee}$	1.52 (1.21, 2.07)	$\bar{\sigma}_{156}^{ew}$	1.11 (0.92, 1.81)	$\bar{\mu}_{168}^{tt}$	4.55 (3.53, 6.39)	$a_{147}^{vt}$	1.49 (1.05, 2.06)
$a_{156}^{ee}$	3.87 (3.13, 4.78)	$a_{156}^{ew}$	0.11 (0.01, 0.30)	$\bar{\sigma}_{168}^{tt}$	1.67 (1.18, 2.31)	$b_{147}^{vt}$	2.57 (1.99, 3.71)
$b_{156}^{ee}$	10.22 (8.99, 11.81)	$b_{156}^{ew}$	6.97 (6.42, 7.03)	$a_{168}^{tt}$	2.71 (1.65, 3.48)	—	—
—	—	—	—	$b_{168}^{tt}$	17.07 (15.19, 19.02)	—	—

Note: ... means the rest; — means not applicable.

reliable. Considering that too many parameters are estimated in this paper so that it is impossible to list all the results, the estimation results of some parameters are shown in Table 2, including the mean and 95% Bayesian confidence interval (CI) according to the distribution provided by the Bayesian estimation. The mean of the threshold parameter  $\alpha$  is 5.04 and its 95% CI is 4.01, 6.10.

And the prior distributions of the parameters for the same kind of time are referred to as the same as shown in Table 1, but the results in Table 2 show us that the posterior distributions can be significantly different which means that amounts of AFC data can correct the biased prior knowledge. It also can be seen from Table 2 that the lower boundary of the entry waiting time is the smallest (tending to 0) because, in some cases, passengers can catch up the departing train in time without any waiting periods. Meanwhile, the difference between the upper and lower boundaries for in-vehicle travel time and the variance are both the smallest which means that the in-vehicle travel time relating to the train schedule varies the least relative to other kinds of times.

As is shown in Figure 4, from *Tianhe Coach Terminal* to *Gangding*, the truncated characteristic of the observations is significant. According to (2), the distribution of the travel time (unit: min) for this OD pair is  $TN(10.22, 6.30, 8.01, 19.33)$ ; that is, for normal distribution, the mean is 10.22 min and the standard derivation is 2.51 min. For the truncated normal distribution, the mean is 11.06 min and standard derivation is 3.74 min which are obviously closer to the real values (the average value is 11.61 min and the standard derivation is 3.39 min) than the mean and standard

derivation of normal distribution. Meanwhile, the estimated boundaries are both consistent with the real censored data. It shows the superiority of the truncated normal distribution to the normal distribution.

The estimation result of the threshold parameter is also significant with the mean 26.04 (the unit of the threshold is minute) and 95% CI (24.01, 28.10). Based on (7), the threshold for different OD pair is curved in Figure 5. In order to test the estimation for the threshold parameter and the threshold equation structure, a field survey was carried out in June, 2014. The survey mainly focused on passengers' travel characteristics in metro system and at last 10000 effective samples were collected. With respect to the threshold, 8 scenarios were designed as follows: for the OD pair, if the shortest travel time was 10 min, 20 min, 30 min, 60 min, 90 min, 120 min, 150 min, and 180 min, respectively, how much more time you can tolerate. And the average value for every scenario was calculated and then displayed as the histogram in Figure 6. It can be seen from Figure 6 that the estimated curve can significantly fit for the surveyed data which demonstrates that the estimated threshold parameter and the threshold function structure are both reliable. Exactly, the Mean Absolute Percent Error (MAPE) between the estimated values and real data is 3.46% which shows that the model has a good goodness-of-fit to the real data.

According to the estimated threshold, the distribution of the number of the OD pairs with only one effective route based on the probability intervals is shown in Figure 7. It shows that none of OD pairs have the probability 0 to have only one effective route which is a response to the estimation



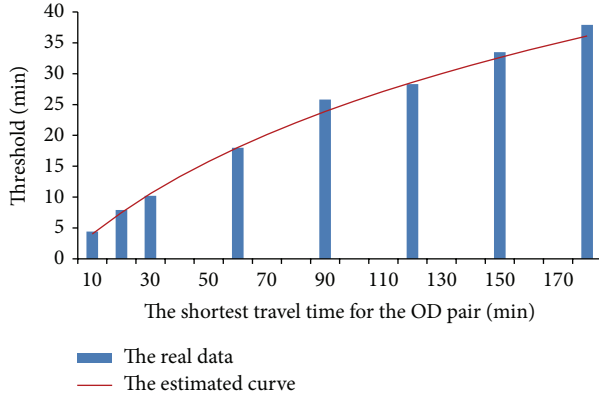


FIGURE 6: A test for the threshold.

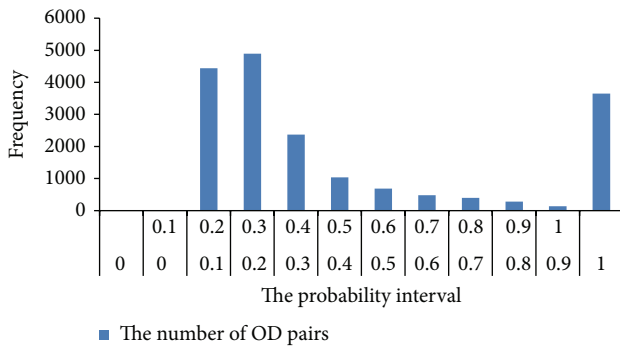


FIGURE 7: The distribution of the number of OD pairs with only one effective route.

condition (referring to the condition that the denominator is not equal to 0). There are 3647 OD pairs whose probabilities are equal to 1 which is due to the fact that there is only one feasible route or the threshold is greater than the upper boundary for those OD pairs. These OD pairs can be used to test the forecasting performance.

**3.2. Performance Test.** The estimated results show their convergence properties based on statistical test, but without the real data test, such as forecasting performance test. In order to test the mobility and estimating effectiveness of the parameters, this paper establishes an index measuring the errors between the estimated results and real data. In advance, some representative OD pairs are selected out based on the condition that, for example, if a specific OD pair  $s \in \mathbf{S}_2$ , then  $P(s | Y^s) = 1$ . Thereafter, the set  $\mathbf{S}_2$  is made of the representative OD pairs. The measuring index  $F$  can be calculated by the following equations:

$$F = \frac{\sum_{s \in \mathbf{S}_2} F_s}{N(\mathbf{S}_2)}, \quad (13)$$

$$F_s = \sum_{i=0}^{i=I_s-1} \text{abs} \left[ \int_{\gamma_i}^{\gamma_{i+1}} \psi(t | \bar{\mu}'_1, \bar{\sigma}'_1, \mathbf{a}'_1, \mathbf{b}'_1) dt - p(\gamma_i, \gamma_{i+1}) \right], \quad (14)$$

$$p(\gamma_i, \gamma_{i+1}) = \frac{\sum_{t \in \mathbf{T}_s} \delta_t}{N(\mathbf{T}_s)}, \quad (15)$$

$$\text{if } \gamma_i \leq t < \gamma_{i+1}, \delta_t = 1; \text{ otherwise, } \delta_t = 0,$$

where  $F_s$  is the error for OD pair  $s$ ;  $I_s$  is the number of travel time intervals for OD pair  $s$ ;  $p(\gamma_i, \gamma_{i+1})$  denotes the ratio between the number of travel time observations within the interval  $(\gamma_i, \gamma_{i+1})$  and the total number of observations; the interval boundary needs to satisfy  $\gamma_0 \leq \min(\mathbf{T}_s) < \gamma_1$  and  $\gamma_{I_s-1} \leq \max(\mathbf{T}_s) < \gamma_{I_s}$ ;  $N(\mathbf{S}_2)$  denotes the size of the OD pairs set  $\mathbf{S}_2$ ;  $N(\mathbf{T}_s)$  denotes the number of the travel time observations in set  $\mathbf{T}_s$ .

According to the above statistical analysis, there are 3647 OD pairs which satisfy the equation  $P(s | Y^s) = 1$ . The test data is a month sized data which was collected in July, 2014, in which the OD pair with the OD volume smaller than 1000 is deleted to make sure of the distribution property. At last, 1121 OD pairs are left to carry out the test. Based on (13)~(15), the measuring index  $F$  is calculated for every OD pair with interval length  $\gamma_{i+1} - \gamma_i = 0.5$  min. Meanwhile, in order to evaluate the forecasting accuracy, an existing model is simulated in advance. The difference between the existing model and the proposed model is formulated as follows: for the existing model, all kinds of times are assumed to follow normal distributions and the selected OD pairs satisfy the condition that the ratio between the shortest mean time and the second shortest mean time is smaller than 0.5 (Sun et al. [16]). The error based on the proposed model ( $F_s(p)$ ) is regarded as the horizontal axis and the error based on the existing model ( $F_s(e)$ ) is regarded as the vertical axis. The scattered points are shown in Figure 8.

In Figure 8, the solid spots represent the errors ( $F_s$ ) and the solid line indicates that the error based on the proposed model is equal to the error based on the existing model; that is,  $F_s(p) = F_s(e)$ , for the OD pair  $s$ . If the spot lies above the line, it means the error based on the proposed model is smaller than that based on the existing model; that is,  $F_s(p) < F_s(e)$ , for the OD pair  $s$ . Otherwise,  $F_s(p) > F_s(e)$ . From Figure 8, we can see that, for all selected OD pairs, the error spots lie above or lie on the line which indicates that the error based on the proposed model is not greater than that based on the existing model; that is,  $F_s(p) \leq F_s(e)$ . And the average error for all selected OD pairs based on the proposed model is 0.19 while that based on the existing model is 0.30. The above results show that the proposed model can improve significantly estimation accuracy on estimating the distributions of route travel time, station walking time, station waiting time, section in-vehicle time, and transfer station transfer time, respectively.

## 4. Conclusions

Travel time reliability in the metro system, referring to the variation in travel time, has attracted more and more attention recently, as it can support evaluating the level of services, calculating route choice proportion, optimizing the train schedule, and so forth. In the metro system, the walking time, waiting time, and delayed in-vehicle time mainly contribute

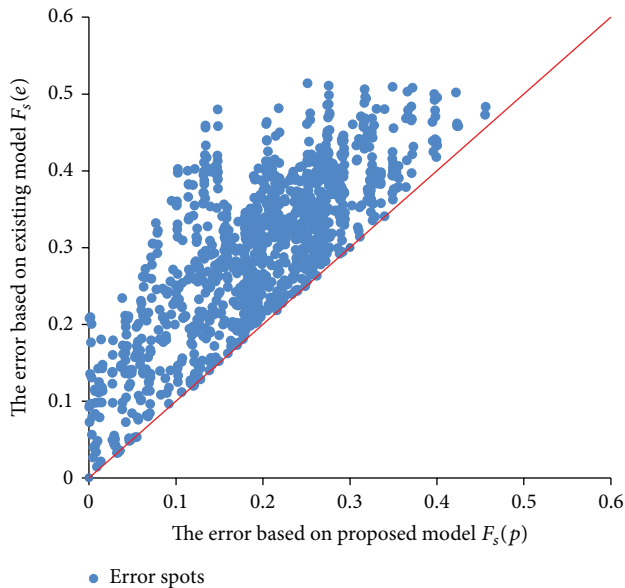


FIGURE 8: The scattered spots figure.

to the variation of the travel time of the travel route. Thus, estimating the walking time in the station, waiting time on the platform, transfer travel time in the transfer channel, and in-vehicle travel time between two successive stations in the metro system is the key to estimate the travel time of the route, especially for the OD pairs between which there are various effective routes. Based on AFC data, huge amounts of passenger's travel times between OD pairs can be collected. But the walking, waiting, transfer, and real in-vehicle travel times cannot be directly gained and the travel times between some OD pairs cannot be directly assigned to the routes as there are more than effective routes between the OD pairs.

Some research has been found in studying the estimation method for various times in the metro system. However, the truncated property of the data is neglected and the threshold which is used to select the OD pairs with only one effective route is calculated exogenously without considering the random property of the travel times of feasible routes. The truncated property of the data is derived from a common sense that all kinds of times will not be too long or short which depends on, for example, the length of the walking channel, walking speed, train schedule, and so forth. Therefore, a new estimation model is proposed in this paper based on Bayesian inference formulation by integrating the probability equation of the OD pair with only one effective route, in which all kinds of times are assumed to follow truncated normal distributions and the threshold is estimated endogenously. The probability of the OD pair with only one effective route is derived from the relationship between the reliable shortest travel time of the route and the reliable second shortest travel time of the route for the OD pair, which varies with the travel time distribution. Considering that the new model contains complex integrations, Markov Chain Monte Carlo (MCMC) method is designed to estimate all parameters, in which Gaussian random walk Metropolis proposals are employed on all the unknown parameters.

Based on the AFC data collected from Guangzhou Metro, the truncated property of the travel time is demonstrated and the proposed model is estimated. The results show that the drawn samples are converged and all parameters are identifiable. The mean value and standard derivation value of the truncated normal distribution are closer to the real average value and standard derivation value than those of normal distribution, which indicates that the truncated normal distribution assumption is more reliable. The threshold is tested by the surveyed data and the MAPE indicates the effectiveness of the threshold parameter estimation and the threshold function structure assumption. Furthermore, relating to the estimating and forecasting accuracy, in terms of every selected OD pair, the error based on the proposed model is not greater than that based on the existing model, and, for all selected OD pairs, the average error based on the proposed model is smaller significantly than that based on the existing model. Therefore, both the estimations' statistical effectiveness and the forecasting accuracy indicate the superiority of the proposed model over the existing model.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This research is supported by the Fundamental Research Funds for the Central Universities (no. 2014YJS082).

## References

- [1] T. Kusakabe, T. Iryo, and Y. Asakura, "Estimation method for railway passengers' train choice behavior with smart card transaction data," *Transportation*, vol. 37, no. 5, pp. 731–749, 2010.
- [2] Y. Sun and R. Xu, "Rail transit travel time reliability and estimation of passenger route choice behavior," *Transportation Research Record*, no. 2275, pp. 58–67, 2012.
- [3] F. Zhou and R. Xu, "Passenger flow assignment model for urban rail transit based on entry and exit time constraints," *Transportation Research Record*, vol. 2284, pp. 57–61, 2012.
- [4] Z. Guang, E. Yao, and Y. Zhang, "Prediction of urban railway station's entrance and exit passenger flow based on fuzzy clustering analysis," in *Proceedings of the International Conference on Railway Engineering*, pp. 70–75, Beijing, China, September 2012.
- [5] C. Cai, E. Yao, M. Wang, and Y. Zhang, "Prediction of urban railway station's entrance and exit passenger flow based on multiply ARIMA model," *Journal of Beijing Jiaotong University*, vol. 38, no. 2, pp. 135–140, 2014.
- [6] C. Cai, E. Yao, S. Liu, Y. Zhang, and J. Liu, "Holiday destination choice behavior analysis based on AFC data of urban rail transit," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 136010, 7 pages, 2015.
- [7] C. Cai, E. Yao, Y. Zhang, and J. Liu, "Forecasting of passenger flow's distribution among urban rail transit stations based on

- AFC data,” *China Railway Science*, vol. 36, no. 1, pp. 126–132, 2015.
- [8] R. Chapleau, M. Trepanier, and K. K. Chu, “The ultimate survey for transit planning: complete information with smart card data and GIS,” in *Proceedings of the 8th International Conference on Survey Methods in Transport: Harmonisation and Data Comparability*, Annecy, France, May 2008.
- [9] A. B. Rahbee, “Farecard passenger flow model at Chicago transit authority, Illinois,” *Transportation Research Record*, vol. 2072, pp. 3–9, 2008.
- [10] Y. Sun and P. M. Schonfeld, “Schedule-based route choice estimation with automatic fare collection data for rail transit passengers,” in *Proceedings of the 93rd Annual Meeting of Transportation Research Board*, CD-ROM, Transportation Research Board of the National Academies, Washington, DC, USA, January 2014.
- [11] W. Zhu, H. Hu, and Z. Huang, “Calibrating rail transit assignment models with genetic algorithm and automated fare collection data,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 7, pp. 518–530, 2014.
- [12] S. G. Lee and M. D. Hickman, “Travel pattern analysis using smart card data of regular users,” in *Proceedings of the 90th Annual Meeting of the Transportation Research Board*, CD-ROM, Transportation Research Board of the National Academies, Washington, DC, USA, January 2011.
- [13] M.-P. Pelletier, M. Trépanier, and C. Morency, “Smart card data use in public transit: a literature review,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 557–568, 2011.
- [14] Y. Sun and R. Xu, “Rail transit travel time reliability analysis and passenger route choice behavior estimation using automated fare collection data,” in *Proceedings of the 91st Annual Meeting of the Transportation Research Board*, CD-ROM, Transportation Research Board of the National Academies, Washington, DC, USA, January 2012.
- [15] F. Zhou, J. Shi, and R. Xu, “Estimation method of path-selecting proportion for urban rail transit based on AFC data,” *Mathematical Problems in Engineering*, vol. 2015, Article ID 350397, 9 pages, 2015.
- [16] L. Sun, J. G. Jin, D. H. Lee, and K. W. Axhausen, “Characterizing travel time reliability and passenger path choice in a metro network,” in *Proceedings of the 94th Annual Meeting of the Transportation Research Board*, CD-ROM, Transportation Research Board of the National Academies, Washington, DC, USA, January 2015.
- [17] L. Sun, Y. Lu, J. G. Jin, D.-H. Lee, and K. W. Axhausen, “An integrated Bayesian approach for passenger flow assignment in metro networks,” *Transportation Research Part C: Emerging Technologies*, vol. 52, pp. 116–131, 2015.
- [18] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, John Wiley & Sons, New York, NY, USA, 1994.
- [19] A. Khani and S. D. Boyles, “An exact algorithm for the mean-standard deviation shortest path problem,” *Transportation Research B: Methodological*, vol. 81, part 1, pp. 252–266, 2015.
- [20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [21] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.





# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

