Dynamic Modeling for Metro Passenger Flows on Congested Transfer Routes

Weiyan Mu 1, Xin Wang 1, Chunya Li 2,* and Shifeng Xiong 3

1 School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China
2 School of Mathematics, Physics, and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China
3 NCMIS, KLSC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

* Correspondence: lichunya@sues.edu.cn

Abstract: With the rapid development of urbanization, the metro becomes more and more important for people’s travel in big cities. To quantitatively describe metro passenger flows on congested transfer routes, this paper introduces a dynamic model based on automated data from the automatic fare collection (AFC) and automatic vehicle location (AVL) systems. An expectation maximization (EM) algorithm is proposed to compute the maximum likelihood estimates of unknown parameters in our model. Our model can yield a systematic analysis of one-transfer passenger flows on both population and individual aspects. Important characteristics, including transfer time, boarding probabilities, walking time, passenger-to-train assignment probabilities, and total travel time, can be inferred using only the AFC and AVL data. We provide a case study on the Beijing metro. Detailed analysis results based on our model are given. We also present a cross-validation method to validate our model with real data.

Keywords: automated data; EM algorithm; passenger assignment; peak hours; statistical inference; transfer time

MSC: 62F10; 62P30

1. Introduction

The metro is an important component of the urban transit system. With the advantages of fast speed, punctual operation, and large load, it has become the first choice for most commuters in many big cities. There are over ten million people traveling by metro in Beijing every day [1]. In the meantime, with the rapid development of urbanization, limited numbers of metro lines and trains can no longer meet people’s travel needs. Congestion on platforms and in carriages often occurs, especially in the morning and evening peak periods. To alleviate the problem of congestion, metro administrations should develop reasonable management measures based on effective evaluation of passenger flows. Therefore, it is necessary to study and quantify characteristics of passenger flows, especially during peak hours.

The automatic fare collection (AFC) and automatic vehicle location (AVL) systems in the metro provide a great convenience for collecting data to study passenger flows. We can obtain a passenger’s OD (origin-destination) station, tap-in time, and tap-out time from the AFC data. The route and schedule of each train can be obtained from the AVL data. With the help of AFC data, [2] proposed an estimation model based on a Bayesian inference formulation to enable the study of the main stages affecting passenger travel time, i.e., walking, waiting, transfer, and in-vehicle times. Ref. [3] used AFC and AVL data to propose a comprehensive approach to estimate passenger congestion costs in terms of...
equivalent travel time loss. Other studies on metro passenger flows based on automated AFC and AVL data can be found in [4–8], among others.

Several authors studied passenger-to-train assignment models and related inferential issues concerning passenger flows on no-transfer metro routes based on automated data [9–11]. This paper focuses on congested transfer routes that are more challenging than no-transfer cases. There are limited papers on analysis of metro transfer routes in the literature. Refs. [12,13] developed optimization models for minimizing the transfer waiting time through modifying train schedules. They did not provide methods to infer passenger flows based on automated data. Ref. [14] proposed an equilibrium model based on train operating diagrams and used it to estimate spatial and temporal distributions of the passenger flows at transfer stations. However, this paper did not estimate specific characteristics of passenger flows such as the boarding probability, transfer time, and walking time. Ref. [15] proposed a complete method to estimate the distributions of walking time and waiting time of transfer passengers based on automated data. However, their method is valid only during off-peak periods. Actually, the issue of estimating those distributions becomes much more complicated during peak hours because of the left behind phenomenon. Ref. [16] proposed a data-driven passenger itinerary inference model to infer left behind, route choice, and probabilities of boarding feasible trains. Their model requires a certain assumption on transfer time that is difficult to verify with automated data. Some important characteristics including walking time and total travel time are not provided in [16]. Ref. [11] presented a new passenger-to-train assignment model, and extended it to one-transfer routes. However, they did not provide feasible methods to estimate the input parameters of such a model for transfer cases. Furthermore, Ref. [11] method is based on the data of the same tap-in time, which cannot be straightforwardly applied to dynamic passenger flows.

In this paper we propose a dynamic modeling method for passenger flows on congested transfer routes based on automated data. Our model constructs a probabilistic quantification relationship between a passenger’s tap-in and tap-out times. An expectation maximization (EM) algorithm [17] is proposed to compute the maximum likelihood estimates (MLEs) of unknown parameters in our model. Our model can yield a systematic analysis of one-transfer passenger flows on both population and individual aspects. In particular, inference and uncertainty quantification of transfer time, an important feature on transfer routes, are presented via our model. Other characteristics on a specific one-transfer route including boarding probabilities, walking time, passenger-to-train assignment probabilities, and total travel time can also be inferred. Our method requires only AFC and AVL data on the route under minimal assumptions. We provide a case study on the Beijing metro. Detailed analysis results based on our model are given. Our method can be validated based on real data via cross-validation.

The major contributions of this paper can be summarized as follows.

(i). We provide reasonable statistical modeling and inferential techniques for passenger flows on metro transfer routes, which constitute a comprehensive analysis on passenger flow characteristics, at both the aggregate and individual levels.

(ii). The proposed methods are completely data-driven, with only automated data. We present quantification indices as much as possible by making full use of automated data.

(iii). Based on rigorous mathematical and statistical derivation, we provide a general idea to analyze metro passenger flows. This idea can be extended to other situations.

The rest of this paper is organized as follows. Section 2 provides a literature review. Section 3 introduces the dynamic model for one-transfer routes and provides parameter estimation methods based on AFC and AVL data. Section 4 discusses applications of the proposed model to inferences for passenger flow characteristics. Section 5 provides the case study and validation of our model. Section 6 ends this paper with some conclusions.
2. Research Background

In the literature, research on metro systems is focused on two aspects: the train operation level and the passenger travel level. At the train operation level, the time interval between train departures is an important factor to describe the efficiency of passengers’ travel. A well-organized metro system often requires a reasonable train schedule. Therefore, how to alleviate the metro congestion by adjusting the train schedule has become a hot topic. Ref. [18] proposed a train timetable rescheduling method based on deep learning. This method enables real-time rescheduling of train schedules after random disturbances. Ref. [19] propose a mixed-integer optimization method to generate well-synchronized timetables, thereby reducing the transfer waiting time for transfer passengers. In fact, there are many uncontrollable random factors in metro systems. Optimization of metro system parameters should be based on accurate estimation of the uncertainty caused by these factors, which requires reasonable statistical models. However, such models have not been well developed in the literature, especially for the transfer routes.

At the level of passenger travel, quantitative indices of passenger travel behavior reflect the current situation in the metro system. Statistical or machine learning methods are used to estimate such indices. Ref. [20] developed an ensemble deep learning approach to forecast the dynamic metro passenger flow. This approach was designed for the entire metro network, and cannot accurately estimate the passenger flow indices on a certain route. Ref. [5] analyzed distributional characteristics of the travel time with AFC data, and proposed a method for estimating path-selecting proportions. Ref. [21] provided a passenger-to-train assignment method, and estimated the walking time and waiting time of metro passengers. Although the two papers analyzed the distributional characteristics of the travel time, the impact of passengers’ tap-in times was ignored in their studies. In fact, passengers who enter the station at different tap-in times face different congestion conditions. Hence, to estimate real-time characteristics of the metro system, we need the study on dynamic models for passenger flows.

Transfer is often an indispensable part of long-distance metro travel. However, the problem of estimating dynamic metrics on congested transfer routes has not received a lot of attention in the literature. This paper aims to fill such a gap, and completely data-driven approaches are proposed in the following sections.

3. Dynamic Models for Transfer Routes

3.1. Likelihood Model of Tap-Out Times

We consider a one-transfer trip in metro systems, which consists of two segments. Let segment 1 and segment 2 denote the segments from the origin to the transfer and from the transfer to the destination, respectively. The time between the tap-in and tap-out of a passenger can be divided into the following five periods:

(i). **Boarding time**: the time from entering the tap-in fare gate to boarding a train on segment 1;

(ii). **Travel time on the train of segment 1**;

(iii). **Transfer time**: the time from getting off the train on segment 1 to boarding the train on segment 2;

(iv). **Travel time on the train of segment 2**;

(v). **Egress time**: the time to walk to the exit fare gate after alighting from the train of segment 2.

The flow chart of such a trip is shown in Figure 1.

For a passenger on the transfer route, denoted by Route A in the following, there may be more than one possible itinerary due to crowded platforms and carriages, especially during peak hours, and it is not determined which train he/she could take with only his/her tap-in and tap-out times. Similar to [11], we need to infer the passenger-to-train assignment and other quantification indices of passenger flows on Route A using the corresponding AVL and AFC data. We then present the following notation and definitions.
Suppose there are \( n \) passengers on Route A. Let \( T_A^{\text{in}} = (t_{A1}^{\text{in}}, t_{A2}^{\text{in}}, \ldots, t_{An}^{\text{in}})^{T} \) and \( T_A^{\text{out}} = (t_{A1}^{\text{out}}, t_{A2}^{\text{out}}, \ldots, t_{An}^{\text{out}})^{T} \) denote the sets of their tap-in and tap-out times, respectively. The egress times of the passengers are independently and identically distributed from the log-normal distribution \( LN(\mu_{A}, \sigma_{A}^{2}) \) with probability density function

\[
f(x|\mu_{A}, \sigma_{A}) = \exp\left\{-\frac{(\ln x - \mu_{A})^{2}}{2\sigma_{A}^{2}}/\left(2\pi\sigma_{A}^{2}x^{2}\right)\right\}^{1/2}, \ x > 0,
\]

where \(-\infty < \mu_{A} < \infty \) and \( 0 < \sigma_{A} < \infty \). This log-normal assumption is widely used and has been verified in the literature (e.g., [15]). Actually, our method in this paper can also be applied to the cases of other egress time distributions such as in [11].

Considering the periodicity of the train schedule, we define a train departure interval as a unit time interval \((0, 1]\), which implies that the departure times of trains at the origin station are \( 1, 2, 3, \ldots \). We first handle the passengers whose tap-in times lie in \((0, 1]\), i.e., \( t_{Ai}^{\text{in}} \in (0, 1], \ i = 1, \ldots, n \). General cases for peak hours will be discussed in Section 3.4.

Mark the first train passengers can take after arriving at the origin station as Train 1 of Route A on segment 2, i.e., \( \text{Train 1} \) on segment 2. Define:

- \( DT_{j}^{(1)} \) : departure time of Train \( j \) from the origin station on segment 1,
- \( AT_{j}^{(1)} \) : arrival time of Train \( j \) at the transfer station on segment 1,
- \( DT_{k}^{(2)} \) : departure time of Train \( k \) from the transfer station on segment 2,
- \( AT_{k}^{(2)} \) : arrival time of Train \( k \) at the terminal station on segment 2,
- \( m_{i} \) : the number of feasible itineraries for the passenger \( i \) of Route A, then \( m_{i} = \#S_{i} \),

where \( S_{i} = \{(j, k) : t_{Ai}^{\text{in}} \leq DT_{j}^{(1)}, AT_{j}^{(1)} \leq DT_{k}^{(2)}, AT_{k}^{(2)} \leq t_{Ai}^{\text{out}} \} \)

\[
= \{(j, k) : AT_{j}^{(1)} \leq DT_{k}^{(2)}, AT_{k}^{(2)} \leq t_{Ai}^{\text{out}} \}.
\]

\( K_{i} \) : the number of feasible itineraries for the passenger \( i \) of Route A on segment 2,

i.e., \( K_{i} = \#\{k : \text{there exists } j \text{ s.t. } (j, k) \in S_{i} \} \), and let \( K = \max_{1 \leq i \leq n} K_{i} \).

\( m_{i}^{(k)} \) : the number of feasible itineraries for the passenger \( i \) of Route A on segment 1 who takes Train \( k \) on segment 2, i.e., \( m_{i}^{(k)} = \#\{j : (j, k) \in S_{i} \} \), \( k = 1, \ldots, K_{i} \),

where \( \#A \) denotes the number of elements in the set \( A \), and let \( m^{(k)} = \max_{1 \leq i \leq n} m_{i}^{(k)} \),

\[
J = \max_{1 \leq i \leq n} m^{(k)}.
\]

We have \( m_{i} = \sum_{k=1}^{K} m_{i}^{(k)} \) and \( 1 \leq m_{i}^{(1)} \leq \cdots \leq m_{i}^{(K)} \). For example, if \( S_{i} = \{(1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\} \), we have \( m_{i} = 6 \), \( K_{i} = 3 \), \( m_{i}^{(1)} = 1 \), \( m_{i}^{(2)} = 2 \), and \( m_{i}^{(3)} = 3 \).
Let $\rho_{jk}(t^m_i)$ denote the probability that a passenger whose tap-in time is $t^m_i$ takes Train $j$ on segment 1 and Train $k$ on segment 2 for $j = 1, \ldots, J$, $k = 1, \ldots, K$. If the passenger takes Train $k$ on segment 2 and $j > m^k_i$, then we let $\rho_{jk}(t^m_i) = 0$, $j = m^k_i + 1, \ldots, J$, $k = 1, \ldots, K$. Moreover, if $K > K_i$, then we let $\rho_{jk}(t^m_i) = 0$, $j = 1, \ldots, J$, $k = K_i + 1, \ldots, K$. Therefore, for each passenger on Route A entering the station at time $t^m_i \in (0, 1]$, he/she has at most $JK$ feasible itineraries, and the corresponding probabilities are $\rho_{11}(t^m_i), \ldots, \rho_{1K}(t^m_i), \ldots, \rho_{J1}(t^m_i), \ldots, \rho_{JK}(t^m_i)$.

Let
\[
\eta_j(t^m_i) = \sum_{k=1}^{K} \rho_{jk}(t^m_i), \quad j = 1, \ldots, J,
\]
denote the probability that a passenger whose tap-in time is $t^m_i$ takes Train $j$ on segment 1. Let
\[
\gamma_k(t^m_i) = \sum_{j=1}^{J} \rho_{jk}(t^m_i), \quad k = 1, \ldots, K,
\]
denote the probability that a passenger whose tap-in time is $t^m_i$ takes Train $k$ on segment 2. Furthermore, let $\rho_{kj}(t^m_i)$ denote the conditional probability that a passenger whose tap-in time is $t^m_i$ takes Train $k$ on segment 2 conditionally on the event that he/she takes Train $j$ on segment 1, and let $\rho_{jk}(t^m_i)$ denote the conditional probability that a passenger whose tap-in time is $t^m_i$ takes Train $j$ on segment 1 conditionally on the event that he/she takes Train $k$ on segment 2. By the conditional probability formula, we have
\[
\rho_{kj}(t^m_i) = \frac{\rho_{jk}(t^m_i)}{\sum_{k=1}^{K} \rho_{jk}(t^m_i)}, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K,
\]
\[
\rho_{jk}(t^m_i) = \frac{\rho_{jk}(t^m_i)}{\sum_{j=1}^{J} \rho_{jk}(t^m_i)}, \quad j = 1, \ldots, J, \quad k = 1, \ldots, K.
\]

Let $E_{ik}$ denote the event that passenger $i$ takes Train $k$ on segment 2. Given the tap-in time of passenger $i$, the distribution function of $t^{out}_{Ai}$ can be written as
\[
P(t^{out}_{Ai} \leq t) = \sum_{k=1}^{K} P(E_{ik}) P(t^{out}_{Ai} \leq t | E_{ik})
\]
\[
= \sum_{k=1}^{K} \left( \sum_{j=1}^{J} \rho_{jk}(t^{in}_{Ai}) \right) P(t^{out}_{Ai} \leq t | E_{ik})
\]
\[
= \sum_{k=1}^{K} \left( \sum_{j=1}^{J} \rho_{jk}(t^{in}_{Ai}) \right) P(t^{out}_{Ai} - AT^{(2)}_k \leq t - AT^{(2)}_k | E_{ik})
\]
\[
= \sum_{k=1}^{K} \left( \sum_{j=1}^{J} \rho_{jk}(t^{in}_{Ai}) \right) F_A(t^{out}_{Ai} - AT^{(2)}_k),
\]
where $F_A$ is the cumulative distribution of $LN(\mu_A, \sigma_A^2)$. Taking the derivative of the above equation with respect to $t^{out}_{Ai}$, we have the tap-out time $t^{out}_{Ai}$ which follows a mixture probability distribution with density
\[
P_A(t^{out}_{Ai}) = \sum_{j=1}^{J} \sum_{k=1}^{K} \rho_{jk}(t^{in}_{Ai}) f(t^{out}_{Ai} - AT^{(2)}_k | \theta_A),
\]
where $\theta_A = (\mu_A, \sigma_A^2)$. The cumulative distribution function of $t^{out}_{Ai}$ can be written as
\[
F_A(t^{out}_{Ai}) = \sum_{j=1}^{J} \sum_{k=1}^{K} \rho_{jk}(t^{in}_{Ai}) F_A(t^{out}_{Ai} - AT^{(2)}_k),
\]
where $F_A$ is the cumulative distribution of $LN(\mu_A, \sigma_A^2)$. Taking the derivative of the above equation with respect to $t^{out}_{Ai}$, we have the tap-out time $t^{out}_{Ai}$ which follows a mixture probability distribution with density
where \( \theta_A = (\mu_A, \sigma_A)^T \) and \( f \) is defined in (1). According to (6), the joint density of \( T_A^{out} = (t_{A1}^{out}, t_{A2}^{out}, \ldots , t_{Am}^{out})^T \) can be given. Here, we write it as the following likelihood form,

\[
L_A(\varphi_A \mid T_A^{out}) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{K} \sum_{k=1}^{\infty} \rho_{jk} \left( t_{Ai}^{in} \right) f \left( t_{Ai}^{out} - A T_{k}^{(2)} \mid \theta_A \right) \right\},
\]

where \( \varphi_A = (\theta_A^T, \rho^T)^T, \theta_A = (\mu_A, \sigma_A)^T, \) and \( \rho(\cdot) = (\rho_{11}(\cdot), \ldots, \rho_{JK}(\cdot))^T. \)

It can be seen from (7) that, by maximizing it over the parameter space, we can only obtain the estimators of \( \sum_{j=1}^{J} \rho_{jk}(t), k = 1, \ldots, K, \) i.e., the probabilities of taking the trains on segment 2, while the value of each \( \rho_{jk}(t), j = 1, \ldots, J, k = 1, \ldots, K \) is inestimable. For example, assuming \( J = K = 2 \), we have \( \rho(t) = (\rho_{11}(t), \rho_{12}(t), \rho_{21}(t), \rho_{22}(t))^T. \)

By maximizing (7), we can obtain estimates of \( \rho_{11}(t) + \rho_{21}(t) \) and \( \rho_{12}(t) + \rho_{22}(t) \), and cannot estimate \( \rho_{11}(t), \rho_{12}(t), \rho_{21}(t), \rho_{22}(t) \). To overcome such an identifiability issue, we introduce a method based on auxiliary data, and discuss it in the next subsection.

### 3.2. Likelihood Model Combined with Auxiliary Data

Let Route B represent the route from the origin to the transfer. Passengers on Route A share segment 1 with those on Route B. Therefore, it can be assumed that they have the same boarding probabilities on segment 1. In this subsection, we introduce the AFC data on Route B as auxiliary data that help to estimate each component of \( \rho \) in (7).

Let \( T_B^{in} = \left( t_{B1}^{in}, t_{B2}^{in}, \ldots , t_{Bn}^{in} \right)^T \) and \( T_B^{out} = \left( t_{B1}^{out}, t_{B2}^{out}, \ldots , t_{Bn}^{out} \right)^T \) denote the vectors of passengers’ tap-in and tap-out times on Route B, respectively. Like in Section 3.1, we assume \( t_{Bi}^{in} \in [0, 1] \) for \( i = 1, \ldots , n_B \). Assume further that the distribution of egress time on Route B is \( LN(\mu_B, \sigma_B^2) \). Given \( t_{Bi}^{in} \), the tap-out time \( t_{Bi}^{out} \) follows the following mixture probability distribution with density

\[
p_B(t_{Bi}^{out}) = \sum_{j=1}^{J} f \left( t_{Bi}^{out} - A T_{j}^{(1)} \mid \theta_B \right) \left( \sum_{k=1}^{K} \rho_{jk} \left( t_{Bi}^{in} \right) \right),
\]

where \( \theta_B = (\mu_B, \sigma_B)^T \), and \( f \) is defined in (1). The likelihood function based on \( T^{out} = \left( T_A^{out^T}, T_B^{out^T} \right)^T \) is

\[
L(\varphi \mid T^{out}) = \prod_{i=1}^{n} \left\{ \sum_{j=1}^{J} \sum_{k=1}^{\infty} \rho_{jk} \left( t_{Ai}^{in} \right) f \left( t_{Ai}^{out} - A T_{k}^{(2)} \mid \theta_A \right) \right\} \cdot \prod_{i=1}^{n_B} \left\{ \sum_{j=1}^{J} \sum_{k=1}^{\infty} \rho_{jk} \left( t_{Bi}^{in} \right) f \left( t_{Bi}^{out} - A T_{k}^{(1)} \mid \theta_B \right) \right\},
\]

where \( \varphi = (\theta_A^T, \theta_B^T, \rho^T)^T. \) Maximum likelihood estimators (MLEs) of unknown parameters \( \varphi \), including each \( \rho_{jk}(t), j = 1, \ldots, J, k = 1, \ldots, K \), can be computed by maximizing (8). The use of \( T_B^{in} \) and \( T_B^{out} \) overcomes the identifiability issue with only the data on Route A.

Note that \( \rho(\cdot) \) are unknown functions, and that it is hard to find their MLEs. We use the quadratic logistic function to parameterize them, and assume
where

\[
\rho_{jk}(t|\beta) = \frac{\exp(\beta_j^t)}{1 + \sum_{j=1}^{K-1} \sum_{k=1}^K \exp(\beta_j^t) + \sum_{k=1}^{K-1} \exp(\beta_k^t)},
\]

\[
\rho_{jk}(t|\beta) = \frac{1}{1 + \sum_{j=1}^{K-1} \sum_{k=1}^K \exp(\beta_j^t) + \sum_{k=1}^{K-1} \exp(\beta_k^t)},
\]

where \( \tilde{t} = (1, t, t^T) \), \( \beta = (\beta_{1,1}, \ldots, \beta_{j,1}^{(k-1)})^T \), and \( \beta_j^k = (\beta_{0,j,k}, \beta_{1,j,k}, \beta_{2,j,k})^T \) for \( j = 1, \ldots, I, k = 1, \ldots, K \), \( (j, k) \neq (I, K) \). Plugging (9) into (8), we have

\[
L(\varphi_t | T^{\text{out}}) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} f\left( t_{\text{out}}^{Ai} - AT_k^t \mid \theta_A \right) \left( \sum_{j=1}^{J} \rho_{jk}(t_{\text{in}}^{Ai} | \beta) \right) \right\}^{U_{ik}} \cdot \prod_{i=1}^{n} \left\{ \sum_{j=1}^{J} f\left( t_{\text{out}}^{Bi} - AT_j^t \mid \theta_B \right) \left( \sum_{k=1}^{K} \rho_{jk}(t_{\text{in}}^{Bi} | \beta) \right) \right\}^{V_{ij}},
\]

where \( \varphi_t = (\theta_A^{(1)}, \theta_B^{(1)}, \beta^{(1)})^T \) is a \((3JK + 1)\)-dimensional vector of unknown parameters.

### 3.3. The EM Algorithm

This section proposes the EM algorithm [17] to compute MLEs of \( \varphi_t \) in (10). We first introduce two types of hidden variables as follows.

\( U_{ik} \) : the indicator of the event that passenger \( i \) on Route A took Train \( k \) on segment 2,

i.e., \( U_{ik} = 1 \) if this passenger took Train \( k \) on segment 2; \( U_{ik} = 0 \) otherwise,

\( V_{ij} \) : the indicator of the event that passenger \( i \) on Route B took Train \( j \),

i.e., \( V_{ij} = 1 \) if this passenger took Train \( j \); \( V_{ij} = 0 \) otherwise.

According to (10), the complete likelihood function can be written as

\[
L_C(\varphi_t | T^{\text{out}}, U, V) = \prod_{i=1}^{n} \prod_{k=1}^{K} \left\{ f\left( t_{\text{out}}^{Ai} - AT_k^t \mid \theta_A \right) \left( \sum_{j=1}^{J} \rho_{jk}(t_{\text{in}}^{Ai} | \beta) \right) \right\}^{U_{ik}} \cdot \prod_{i=1}^{n} \left\{ \sum_{j=1}^{J} f\left( t_{\text{out}}^{Bi} - AT_j^t \mid \theta_B \right) \left( \sum_{k=1}^{K} \rho_{jk}(t_{\text{in}}^{Bi} | \beta) \right) \right\}^{V_{ij}},
\]

where \( U = \{U_{ik}\}_{i=1,\ldots,n,k=1,\ldots,K} \) and \( V = \{V_{ij}\}_{i=1,\ldots,n,j=1,\ldots,J} \). The corresponding complete log-likelihood function is

\[
l_C(\varphi_t | T^{\text{out}}, U, V) = \sum_{i=1}^{n} \sum_{k=1}^{K} U_{ik} \left\{ \log\left( f\left( t_{\text{out}}^{Ai} - AT_k^t \mid \theta_A \right) \right) + \log\left( \sum_{j=1}^{J} \rho_{jk}(t_{\text{in}}^{Ai} | \beta) \right) \right\} + \sum_{i=1}^{n} \sum_{j=1}^{J} V_{ij} \left\{ \log\left( f\left( t_{\text{out}}^{Bi} - AT_j^t \mid \theta_B \right) \right) + \log\left( \sum_{k=1}^{K} \rho_{jk}(t_{\text{in}}^{Bi} | \beta) \right) \right\}.
\]

The steps of the EM algorithm are as follows.

(1) Given the initial values \( \varphi_t^{(0)} = (\theta_A^{(0)T}, \theta_B^{(0)T}, \beta^{(0)T})^T \) of the parameters.
where

\[
l_1(\theta_A) = \sum_{i=1}^{n} \sum_{k=1}^{K} C_{ik}^{(q)} \log f \left( \frac{\rho_{\text{out})_{ik}} - AT_{k}^{(2)} | \theta_A} \right),
\]

\[
l_2(\theta_B) = \sum_{i=1}^{n_0} \sum_{j=1}^{J} D_{ij}^{(q)} \log f \left( \frac{\rho_{\text{out})_{ij}} - AT_{j}^{(1)} | \theta_B} \right),
\]

\[
l_3(\beta) = \sum_{i=1}^{n} \sum_{k=1}^{K} C_{ik}^{(q)} \log \left( \sum_{l=1}^{L} \rho_{\text{in})_{il}} (t_{\text{in})_{il} | \beta} \right) + \sum_{i=1}^{n_0} \sum_{j=1}^{J} D_{ij}^{(q)} \log \left( \sum_{k=1}^{K} \rho_{\text{in})_{ik}} (t_{\text{in})_{ik} | \beta} \right),
\]

and

\[
C_{ik}^{(q)} = E \left( U_{ik} | \varphi_{(q)}^{q} \right) = \frac{f \left( \frac{\rho_{\text{out})_{ik}} - AT_{k}^{(2)} | \varphi_{(q)}^{q} \right)}{\sum_{k=1}^{K} f \left( \frac{\rho_{\text{out})_{ik}} - AT_{k}^{(2)} | \varphi_{(q)}^{q} \right)} \left( \sum_{l=1}^{L} \rho_{\text{in})_{il}} (t_{\text{in})_{il} | \beta) \right) \right),
\]

\[
D_{ij}^{(q)} = E \left( V_{ij} | \varphi_{(q)}^{q} \right) = \frac{f \left( \frac{\rho_{\text{out})_{ij}} - AT_{j}^{(1)} | \varphi_{(q)}^{q} \right)}{\sum_{j=1}^{J} f \left( \frac{\rho_{\text{out})_{ij}} - AT_{j}^{(1)} | \varphi_{(q)}^{q} \right)} \left( \sum_{k=1}^{K} \rho_{\text{in})_{ik}} (t_{\text{in})_{ik} | \beta) \right) \right).
\]

(III) Maximization step: maximizing \( Q \left( \varphi_{(q)}^{q} | \varphi_{(q)}^{(q-1)} \right) \) yields the \((q + 1)\)th iteration of parameter \( \varphi_{(q+1)} = \arg \max l_1(\theta_A) \), \( \theta_{(q+1)} = \arg \max l_2(\theta_B) \), and \( \beta_{(q+1)} = \arg \max l_3(\beta) \).

(IV) Iteration: repeat steps (II) and (III) until convergence.

It can be seen that the EM algorithm simplified the original optimization problem through introducing hidden variables. Note that its performance relies on the choice of the initial values. Multiple initial values can be used to better approximate the global solution. Our experience based on numerical simulations shows that the proposed EM algorithm is efficient to compute the MLEs.

3.4. Periodic Modeling

In this paper, we focus on modeling for the transfer routes during peak hours. With train departure intervals of equal lengths, the passenger flow has stable performance. Therefore, the distribution \( f \) of egress times remains, and the boarding probabilities \( \rho(t) \) behave periodically. Specifically, consider train intervals \( (DT_{j}^{(1)}, DT_{j+1}^{(1)}], \ldots, (DT_{j+l-1}^{(1)}, DT_{j+l}^{(1)} \) in such a period. The functions \( \rho(t) \) in any two of these intervals are identical. We construct a periodic model for \( \rho(t) \) to analyze the AFC data of passengers who enter the metro system during peak hours.

Without loss of generality, let \( j = 1, DT_{k}^{(1)} = k - 1 \) for \( k = 1, \ldots, l + 1 \). For \( t \in (DT_{1}^{(1)}, DT_{l+1}^{(1)} = (0, l), \) assume that
\[
\rho_{jk}(t|\beta) = \frac{\exp(\beta'_{jk}\{\dot{t}\})}{1 + \sum_{j=1}^{J-1} \sum_{k=1}^{K} \exp(\beta'_{jk}\{\dot{t}\}) + \sum_{j=1}^{J-1} \sum_{k=1}^{K} \exp(\beta'_{jk}\{\dot{t}\})},
\]
\[
\rho_{1J}(t|\beta) = \frac{1}{1 + \sum_{j=1}^{J-1} \sum_{k=1}^{K} \exp(\beta'_{jk}\{\dot{t}\}) + \sum_{j=1}^{J-1} \sum_{k=1}^{K} \exp(\beta'_{jk}\{\dot{t}\})},
\]
\[
\rho_{JK}(t|\beta) = \frac{1}{1 + \sum_{j=1}^{J-1} \sum_{k=1}^{K} \exp(\beta'_{jk}\{\dot{t}\}) + \sum_{j=1}^{J-1} \sum_{k=1}^{K} \exp(\beta'_{jk}\{\dot{t}\})},
\]

where \( \{\dot{t}\} = (1, \dot{t}, \dot{t}^2)^T \) and \( \dot{t} = t - \lfloor t \rfloor \) if \( t - \lfloor t \rfloor > 0 \), \( \dot{t} = 0 \) otherwise. \( \lfloor \cdot \rfloor \) denotes the floor function. \( \beta = (\beta_{11}^T, \ldots, \beta_{j(k-1)}^T) \), and \( \beta_{jk} = (\beta_{0,jk}, \beta_{1,jk}, \beta_{2,jk}) \) for \( j = 1, \ldots, J \), \( k = 1, \ldots, K \), \( (j,k) \neq (J,K) \) are unknown parameters.

Clearly, (9) is a special case of (11) with \( l = 1 \). Replacing (9) with (11) in (10), we obtain the periodicity model of \((T_{in}^{A_i}, T_{out}^{A_i}, T_{in}^{B_i}, T_{out}^{B_i})\), where \( t_{in}^{A_i} \in (0, l], i = 1, \ldots, n \) and \( t_{in}^{B_i} \in (0, l], i = 1, \ldots, n_0 \). The method to estimate the unknown parameters for \( l = 1 \) in previous subsections can also be used here.

Figure 2 is a process flow chart of the proposed methodology. The first part presents the model and the method for estimating an unknown parameter in the model. The second part provides applications of the model and estimated parameters, which will be discussed in the next section.

**4. Applications of Analysis of Transfer Routes**

In this section, we present some applications of our dynamic model in the previous sections for the analysis of transfer routes during peak hours. First of all, we should point out that the parameters in (8) can be viewed as measures that describe some aspects of characteristics of transfer passenger flows. Therefore, our estimators can directly be used to infer these measures. Furthermore, our method can be used to infer the transfer time and to predict the travel time on the transfer route.

**4.1. Inference for the Transfer Time**

Recall that a passenger’s transfer time is the time from getting off the train on segment 1 to boarding the train on segment 2. It contains transfer walking time and waiting time.
at the platform of the transfer station, and thus is an important index to quantify the congestion condition of the transfer channel and segment 2. Based on AFC and AVL data, this subsection studies inferences for transfer times via the proposed dynamic model.

First, we consider the population aspect, i.e., inference for the transfer time when only the tap-in time is given. By definition, the transfer time $T_{\text{transfer}}$ of a passenger takes its value on a finite set $\{DT_k^{(2)} - AT_j^{(1)}, j = 1, \ldots, J, k = 1, \ldots, K\}$. Given the passenger’s tap-in time $t_{\text{in}}^A$, it is not hard to derive the distribution of the transfer time

$$P\left(T_{\text{transfer}} = DT_k^{(2)} - AT_j^{(1)} \mid t_{\text{in}}^A\right) = \rho_{jk}(t_{\text{in}}^A) \quad j = 1, \ldots, J, k = 1, \ldots, K.$$ 

Therefore, the conditional mean and conditional standard deviation of $T_{\text{transfer}}$ are given by

$$E\left(T_{\text{transfer}} \mid t_{\text{in}}^A\right) = \sum_{j=1}^{J} \sum_{k=1}^{K} \rho_{jk}(t_{\text{in}}^A) \left(DT_k^{(2)} - AT_j^{(1)}\right), \quad (12)$$

$$sd\left(T_{\text{transfer}} \mid t_{\text{in}}^A\right) = \left\{ \left(\sum_{j=1}^{J} \sum_{k=1}^{K} \rho_{jk}(t_{\text{in}}^A) \left(DT_k^{(2)} - AT_j^{(1)}\right)^2 - \left(E\left(T_{\text{transfer}} \mid t_{\text{in}}^A\right)\right)^2 \right)^{1/2} \right\}. \quad (13)$$

The MLEs of $\rho_{jk}(t_{\text{in}}^A)$ obtained in Section 3 can be used to compute these quantities.

Next, we consider the individual aspect, i.e., inference for the transfer time when one’s tap-in time. The prediction method can also be used to validate our method via cross-validation (CV).

This subsection discusses how to predict the tap-out time of a transfer passenger when his/her tap-in time is given. Hence, the total travel time can be predicted given one’s tap-in time. The prediction method can also be used to validate our method via cross-validation (CV).

By (6), given $t_{\text{in}}^A$, the conditional density function of the tap-out time $t_{\text{out}}^A$ of the passenger is
\[ P_A(t_{\text{out}}^A) = \sum_{j=1}^{I} \sum_{k=1}^{K} \rho_{jk}(t_{\text{in}}^A) f(t_{\text{out}}^A - AT_{k}^{(2)} | \theta_A). \]  

(16)

When the parameters are known, the predictor of \( t_{\text{out}}^A \) that minimizes the mean squared prediction error is the following conditional mean:

\[ E(t_{\text{out}}^A | t_{\text{in}}^A) = \sum_{j=1}^{I} \sum_{k=1}^{K} \rho_{jk} \left( t_{\text{in}}^A \right) \int t f\left( t - AT_{k}^{(2)} | \theta_A \right) dt. \]  

(17)

Plugging the MLEs of the parameters into (17), we get the empirical version of the best predictor. With the MLEs, the conditional distribution of \( t_{\text{out}}^A \) can be estimated via (16), and this can yield prediction intervals of \( t_{\text{out}}^A \). Specifically, given a confidence level \( \alpha \in (0, 1) \), the \( 100(1 - \alpha)\% \) prediction interval of the tap-out time of a passenger with the tap-in time \( t_{\text{in}}^A \) can be given by

\[ [c(\alpha/2), c(1 - \alpha/2)], \]  

(18)

where \( c(\alpha) \) represents the lower \( \alpha \)-quantile of the estimator of the distribution (16).

In addition, the accuracy of the above prediction method on a test sample can be used to validate the proposed model. When there does not exist a test sample, the CV method [22] can be applied. We randomly divide the current AFC data of Route A into two parts. One is used as the training sample to train the model parameters by the method in Section 3, and the other part acts as the test sample that is used to evaluate the prediction accuracy.

5. Case Study
5.1. Data Description

There are 24 metro lines of the total mileage 727 km in Beijing at the end of 2020. With an increase of 499 km between 2010 and 2020, the Beijing metro has entered an era of rapid growth [23]. The congestion problem broadly exists in the Beijing metro during peak hours. Analysis of metro data is necessary to study passenger flows on congested routes, which is helpful for metro management.

In this section, the proposed dynamic model is applied to analyze the passenger flow on the following route (Route A) in the Beijing metro during peak hours:

- Segment 1: from Wuzixueyuanlu Station to Hujialou Station (Line 6, 8 stops),
- Segment 2: from Hujialou Station to Liangmaqiao Station (Line 10, 3 stops),

as shown in Figure 3.

Figure 3. Transfer route from Wuzixueyuanlu Station to Liangmaqiao Station.

The data used to analyze Route A are the AFC and AVL data from 7:00 a.m. to 9:00 a.m. in ten weekdays in 2021. The auxiliary data are from:

- Route B. from Wuzixueyuanlu Station to Hujialou Station (Line 6, 8 stops),
during the same period. The sample sizes of Route A and Route B are 2494 and 4084, respectively. Figure 4 shows the frequency histogram of Route A passengers' tap-in and tap-out times during peak hours.

Figure 4. Frequency histograms of tap-in and tap-out times of Route A passengers in peak hours.

According to the data, the departure interval between two trains at Wuzixueyuanlu is about 3 min. The train travel times of segments 1 and 2 are about 21 and 6 min, respectively. The train whose departure time is 7:03 is set as Train 1 of segment 1. The first train departure interval is (7:00, 7:03). After removing several large outliers, we find that each passenger who enters in the interval taps out before 7:41. Therefore, all possible trains for these passengers are those shown in Table 1, which implies $J = K = 3$. We can further infer that the set of all feasible itineraries is

$$S = \{(1, 1), (1, 2), (1, 3), (2, 2), (2, 3), (3, 3)\}$$  \quad (19)

from this table. By periodicity, this holds for all passengers whose tap-in times are from 7:00 to 9:00.

Table 1. Train schedule.

<table>
<thead>
<tr>
<th>Train</th>
<th>Wuzixueyuanlu Station</th>
<th>Hujiayou Station</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Departure Time</td>
<td>Arrival Time</td>
</tr>
<tr>
<td>1</td>
<td>7:03:00</td>
<td>7:24:00</td>
</tr>
<tr>
<td>2</td>
<td>7:06:00</td>
<td>7:27:00</td>
</tr>
<tr>
<td>3</td>
<td>7:09:00</td>
<td>7:30:00</td>
</tr>
</tbody>
</table>
5.2. Parameter Estimation

Based on the AFC and AVL data on Route A and Route B, we build the likelihood model and use the proposed EM algorithm to compute MLEs of unknown parameters. The MLEs are $(\hat{\mu}_A, \hat{\sigma}_A)^T = (0.2329, 0.5620)^T$ and $(\hat{\mu}_B, \hat{\sigma}_B)^T = (0.3003, 0.5063)^T$. From this, we can get the mean and standard deviation of the distribution of egress time on Route A are 4.4346 and 0.9008, respectively; those for Route B are 4.6047 and 0.8296, respectively. Figure 5 shows the corresponding probability density functions. There is not much difference between the egress times on the two routes.

![Figure 5. Density function curves of egress time.](image)

In the meantime the EM algorithm yields the MLEs of $\beta$ in the logistic model (11), which further give the probabilities $\rho_{jk}(t_{\text{in}})$, $(j, k) \in S$ in (19). Their curves with $t_{\text{in}}$ varying in the first train departure interval are shown in Figure 6. It can be seen from Figure 6 that passengers are more likely to take train combinations $(1, 1)$ and $(1, 2)$ when they tap in the station earlier, and that they are more likely to take train combinations $(2, 2)$ and $(3, 3)$ when their tap-in times are closer to the train departure time. Overall, the later a passenger’s tap-in time is, the more likely he/she takes the later trains on segment 1 and segment 2. For an in-depth study, we also need to discuss the boarding probabilities and conditional boarding probabilities on segment 1 and segment 2. Curves of related boarding probabilities in (2)–(5) are presented in Figure 7. Some interesting phenomena can be found from the two figures. It can be seen from (a) and (b) in Figure 7 that, as expected, the probabilities of boarding the first (third) train on both segments 1 and 2 become smaller (larger) as $t_{\text{in}}$ increases. From Figure 6, or (a) in Figure 7, we can find that passengers who tap in after 7:02 have very small probabilities of boarding the first train on segment 1. This may make the values of the conditional probabilities $\rho_{k|1}(t_{\text{in}})$ shown by (c) in Figure 7 inaccurate. From (d) in Figure 7, the conditional probabilities $\rho_{2|2}(t_{\text{in}})$ have an increasing trend as $t_{\text{in}}$ goes to 7:03. A possible reason is that a passenger who enters late but boards the second train has a fast walking speed, and thus he/she has larger probability of boarding his/her first feasible train (Train 2) on segment 2. From (e) and (f) in Figure 7, we can see that the trajectories of the conditional probabilities $\tilde{\rho}_{j|k}(t_{\text{in}})$ are consistent with our expectation. Specifically, given the train which passengers take on segment 2, we can deduce that passengers with an earlier tap-in time have a higher probability of taking Train 1 on segment 1; passengers with a later tap-in time have a higher probability of taking Train 2 or Train 3 on segment 1.
Plugging the MLEs into (6), we can get the conditional density functions of $t_{\text{out}}$ conditionally on $t_{\text{in}}$. Figure 8 presents these function curves given $t_{\text{in}} = 7:00:02, 7:01:00, 7:02:00,$ and $7:03:00$, respectively. They are all mixture density functions, and have more than one peak. According to (17), we can also derive the conditional mean of $t_{\text{out}}$, shown in Figure 9. As expected, it is an increasing function with a non-linear pattern. By periodicity, it can be extended to the situation when $t_{\text{in}}$ lies between 7:00 and 9:00. This curve can be used to predict the tap-out time of a passenger given his/her tap-in time.

5.3. Inference for $T_{\text{transfer}}$

In this subsection, we infer the transfer times of passengers on Route A. First, consider the population aspect. According to Table 1, all possible values of $T_{\text{transfer}}$ are in \{2.4 min, 5.4 min, 8.4 min\}. Let $T^*_1 = 2.4$ min, $T^*_2 = 5.4$ min, and $T^*_3 = 8.4$ min. We have $P(T_{\text{transfer}} = T^*_1 | t_{\text{in}}) = \rho_{11}(t_{\text{in}}) + \rho_{22}(t_{\text{in}}) + \rho_{33}(t_{\text{in}})$, $P(T_{\text{transfer}} = T^*_2 | t_{\text{in}}) = \rho_{12}(t_{\text{in}}) + \rho_{23}(t_{\text{in}})$, and $P(T_{\text{transfer}} = T^*_3 | t_{\text{in}}) = \rho_{13}(t_{\text{in}})$. Figure 10 shows the curves of $P(T_{\text{transfer}} = T^*_i | t_{\text{in}})$, $i = 1, 2, 3$ with respect to the tap-in time $t_{\text{in}}$. It can be seen that, when $t_{\text{in}}$ is close to 7:03, $P(T_{\text{transfer}} = T^*_1 | t_{\text{in}})$ has an increasing trend. A possible reason is that, some passengers in a hurry picked up speed when seeing the train coming, and when they missed the train, they were more likely to walk fast through the transfer channel. By (12) and (13), we show the conditional mean and standard deviation of $T_{\text{transfer}}$ in Figure 11. Figure 11 also indicates that the passengers whose tap-in times are close to 7:03 have a shorter $T_{\text{transfer}}$ on average, which is consistent with the above inferring result. Specifically, passengers who tap in the origin station earlier are more likely to be on Train 1 of segment 1. They have less time to wait and are therefore less likely to be in a hurry to transfer, resulting in a longer transfer time. Conversely, passengers who tap in the origin station later will usually spend some time waiting for a train at the origin station, so they will have to walk faster if they want to catch Train 2 in segment 2, thus reducing their transfer times.

Next, consider the individual aspect. By (14), we compute the six passenger-to-train assignment probabilities $p_{ijk}$, $(j, k) \in S$ for given $t_{\text{in}}^i$ and $t_{\text{out}}^i$, and show their contour maps in Figure 12. Moreover, by (15), we present the contour map of the average transfer time for given $t_{\text{in}}^i$ and $t_{\text{out}}^i$ in Figure 13. From these figures, we can infer the behavior of an individual passenger whose tap-in and tap-out times are known on the transfer route. For example, assuming that passenger $i$ has a tap-in time of 7:02:30 and a tap-out time of 7:42:00, we can draw the conclusion of $p_{i11} < p_{i12} < p_{i22} < p_{i13} < p_{i33} < p_{i23}$ from Figure 12.
Figure 7. Curves of $\eta_{j}(t^{in})$, $\gamma_{k}(t^{in})$, $\rho_{k|j}(t^{in})$, and $\tilde{\rho}_{j|k}(t^{in})$. 
Figure 8. Conditional density functions of $t^{out}$ conditionally on different $t^{in}$'s.

Figure 9. Curve of conditional means of $t^{out}$ conditionally on $t^{in}$.

Figure 10. Curves of $P(T_{transfer} = T_i^* | t^{in})$, $i = 1, 2, 3$. 
Figure 11. Curves of $E(T_{\text{transfer}} | \mu_{\text{in}})$ and $sd(T_{\text{transfer}} | \mu_{\text{in}})$.

Figure 12. Contour maps of $p_{1,1}$, $p_{1,2}$, $p_{1,3}$, $p_{2,2}$, $p_{2,3}$, $p_{3,3}$.
5.4. Validation via CV

We use the CV method to validate the proposed model. The AFC data of passengers on Route A are randomly divided into a training set and a test set with proportions of 80% and 20%, respectively. By (18), we use the training data to construct the 95% prediction interval of the tap-out time for each tap-in time during the peak hours; see Figure 14. It can be seen that about 95% of the test data lie within the corresponding prediction intervals. This provides a validation of our model. Furthermore, an interesting phenomenon can be found from Figure 14: when a passenger’s tap-in time is close to 7:00 or 9:00, the time when the peak hours begin or end, his/her tap-out time becomes closer to the lower limit of the prediction interval. The reason is that, such a passenger has a relatively short travel time since the congestion status is not so serious at that time. We can also see that the passengers whose tap-in times lie between 7:30 and 8:30 have more even tap-out times between the upper and lower limits of the prediction interval.

6. Conclusions

This paper proposes a dynamic logistic model to study passenger flow characteristics on one-transfer metro routes. Estimates of parameters in our model can be computed by the EM algorithm based on automated AFC and AVL data. A number of passenger flow characteristics including transfer time, boarding probabilities, and travel time can be inferred and quantified with the proposed model. We also present a CV method to validate our model. Real applications to the Beijing metro show the effectiveness of the proposed model.
The proposed model and its derived inferential methods possess the following advantages. First, they are completely data-driven, with only automated data. We do not require additional information including survey data, speed information of passengers, and layouts of stations. Second, our methods provide a comprehensive analysis on passenger flow characteristics. Using reasonable statistical modeling and inferential techniques, we present quantification indices as much as possible by making full use of automated data. The obtained results can be used for the optimal management of passenger flows. For instance, we can estimate the transfer times on many transfer routes, and consider adjusting the train schedule or implementing passenger limiting at the transfer stations with long transfer times. Third, the basic idea of the dynamic logistic model is to derive the conditional distribution of a passenger’s tap-out time conditionally on his/her tap-in time. This is a very general idea, and can be followed in more complex situations.

There are several future directions:
(i) Extensions of the proposed method to the situations where there are multiple transfers and/or multiple route selections between the origin and destination;
(ii) Generalizations of the proposed method, possibly combined with deep learning or other machine learning methods; for analyzing the whole metro network;
(iii) Combinations with optimization methods for improving metro management.

Author Contributions: Conceptualization, W.M. and S.X.; methodology, W.M.; software, X.W. and C.L.; validation, X.W. and C.L.; formal analysis, W.M. and X.W.; investigation, C.L.; resources X.W.; data curation, X.W.; writing—original draft preparation, W.M. and X.W.; writing—review and editing, W.M., C.L. and S.X.; visualization, X.W.; supervision, C.L.; project administration, C.L.; funding acquisition, S.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work is partially supported by the National Key R&D Program of China (Grant Nos. 2021YFA1000300, 2021YFA1000301, and 2021YFA1000303) and the National Natural Science Foundation of China (Grant No. 12171462).

Data Availability Statement: Data sharing not applicable.

Acknowledgments: The authors gratefully acknowledge the editors and reviewers for their professional comments.

Conflicts of Interest: The authors declare no conflict of interest.

References
2. Zhang, Y.S.; Yao, E.J. Splitting travel time based on AFC data: Estimat- ing walking, waiting, transfer, and in-vehicle travel times in metro system. Discret. Nat. Soc. 2015, 2015, 1–11. [CrossRef]
8. Jiang, A.; Sun, Q.; Peng, Z.; Wei, X.; Huang, Q.; Tong, M. Metro OD matching and statistical research based on AFC data. J. Phys. Conf. Ser. 2020, 1624, 032051. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.