A Spatiotemporal Probabilistic Graphical Model Based on Adaptive Expectation-Maximization Attention for Individual Trajectory Reconstruction Considering Incomplete Observations

Xuan Sun [1,2], Jianyuan Guo [1,*], Yong Qin [2,*], Xuanchuan Zheng [1,3], Shifeng Xiong [4], Jie He [1,2], Qi Sun [5] and Limin Jia [2]

[1] School of Traffic and Transportation, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China; xuansun@bjtu.edu.cn (X.S.); 20114089@bjtu.edu.cn (X.Z.); 23120804@bjtu.edu.cn (J.H.)

[2] State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, No. 3 Shangyuancun, Haidian District, Beijing 100044, China; lmjia@bjtu.edu.cn

[3] Beijing Urban Construction Design & Development Group Co., Ltd., No. 5 Fuchengmen North Street, Xicheng District, Beijing 100032, China

[4] NCMIS, KLSC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; xiong@amss.ac.cn

[5] Beijing Metro Network Administration Co., Ltd., No. 6 Xiaoying North Road, Chaoyang District, Beijing 100020, China; sunqi@bmncc.com.cn

* Correspondence: jyguo@bjtu.edu.cn (J.G.); yqin@bjtu.edu.cn (Y.Q.)

Abstract: Spatiotemporal information on individual trajectories in urban rail transit is important for operational strategy adjustment, personalized recommendation, and emergency command decision-making. However, due to the lack of journey observations, it is difficult to accurately infer unknown information from trajectories based only on AFC and AVL data. To address the problem, this paper proposes a spatiotemporal probabilistic graphical model based on adaptive expectation maximization attention (STPGM-AEMA) to achieve the reconstruction of individual trajectories. The approach consists of three steps: first, the potential train alternative set and the egress time alternative set of individuals are obtained through data mining and combinatorial enumeration. Then, global and local potential variables are introduced to construct a spatiotemporal probabilistic graphical model, provide the inference process for unknown events, and state information about individual trajectories. Further, considering the effect of missing data, an attention mechanism-enhanced expectation-maximization algorithm is proposed to achieve maximum likelihood estimation of individual trajectories. Finally, typical datasets of origin-destination pairs and actual individual trajectory tracking data are used to validate the effectiveness of the proposed method. The results show that the STPGM-AEMA method is more than 95% accurate in recovering missing information in the observed data, which is at least 15% more accurate than the traditional methods (i.e., PTAM-MLE and MPTAM-EM).

Keywords: urban rail transit; trajectory prediction; probabilistic graphical model; expectation-maximization algorithm; attention mechanism

1. Introduction

Currently, urban rail transit (URT) has become the preferred public transport mode for residents due to its large capacity and high efficiency. For example, in Beijing, the total number of passengers reached 5.327 billion in 2022, of which 42.5% were transported by URT [1]. Due to the large proportion of transportation, the URT system also faces many problems, such as the fact that it is often difficult to transport passengers in a timely manner during peak traffic hours, which leads to crowding induced by passengers waiting for

trains on platforms and other areas [2–4]. Furthermore, factors such as the capacity of different train types and station layouts also result in uncertain waiting times and complicated travel choices for passengers [2,5–11].

To better monitor the URT system's status and optimize train scheduling, precise access to spatiotemporal characteristics and semantic information of passengers is a prerequisite [2,4,5,10,12–16]. The prediction of individual mobility using data-driven modeling approaches based on Automatic Fare Collection (AFC) data and Automatic Vehicle Location (AVL) data has been a hot research topic in recent years. Meanwhile, individual movement information can also be used for emergency commands or providing personalized recommendation services [10]. Among the studies on individual mobility modeling in urban rail transit, scholars mainly carry out three aspects to achieve accurate prediction of individual movement, namely, travel pattern mining [17,18], route choice model [19–24], and individual trajectory inference [10,15,25–27]. These studies are categorized into network-level, path-level, and train-level according to the scale of the URT system.

In the first aspect of research, unsupervised learning methods (e.g., K-means, LDA) are used to mine job-housing relationships or travel patterns about passengers, which can subdivide passengers into different groups [5,18,28–32]. For example, Cheng et al. [18] developed a topic model to predict passengers' travel destinations, thereby distinguishing between commuters and non-commuters. However, the focus of these methods is generally to construct input features (such as travel days, travel time, etc.), which are primarily used to support macro-level transportation planning or the prediction of new lines, offering limited assistance for operational-level adjustments and strategies [32].

Furthermore, considering the path-level, passengers need to be matched to one physical path between Origin-Destination (OD) pairs. Thus, large research on route choice and assignment models consists of three main methodologies: the Logit model based on labeled data, the clustering model based on unsupervised learning, and the probability-based generative model [3,21,33–36]. The Logit model and its variants are generally established by considering the number of transfers, distance, waiting time, etc. [6,34,37,38]. Some scholars have adopted unsupervised clustering methods for exploration, e.g., Fu et al. [39] combined the AFC data of London Underground with the Gaussian Mixture Model (GMM) within a Naive Bayesian framework to calculate the line selection probability. Wu et al. [36] proposed a fuzzy matching method to assign the passenger flow to each line using the AFC data. Probabilistic generative models appeared almost simultaneously with clustering methods [40]. They are mainly based on Bayes' rule or frequency-based statistical inference methods. Sun et al. [21] proposed a comprehensive Bayesian inference framework that is combined with the Metropolis–Hastings (M-H) algorithm to provide a posterior distribution for route choice. From an application perspective, at the path-level, these researchers are still unable to obtain fine-grained information about individual trajectories and face methodological limitations such as poor stability or over-reliance on survey data.

Moreover, some scholars have expanded individual trajectory reconstruction (ITR) from the path-level to the train-level by integrating AFC data with other data, focusing on models for matching passengers to the train. Current research primarily rely on the Rule-based Method (RM) and the Probabilistic Generative Model (PGM). RM directly utilizes the segmentation and concatenation of AFC and AVL data to mine the matching relationship between passengers and trains [15,26]. However, the spatiotemporal constraints in such methods are considered hard constraints, lacking detailed depictions of passenger behaviors. Studies based on PGM refine the modeling of passengers' left-behind or waiting behaviors at stations [2,12,20,41–43], such as PTAM [20] and LBPMF [42]. However, some essential parameters in these studies still need manual surveys (walking speed, etc.). The improved MPTAM model established by Xiong et al. [12] can automatically fit parameters without resorting to external data. Considering the randomness of boarding choice at the individual level, the error of these researches may be large. Further exploring the

inherent value of data to replace manual surveys presents a worthwhile approach for obtaining passengers' spatiotemporal trajectories to explore.

In summary, the existing methods have the problems of the high cost of manual investigation, large sample randomness, coarse sampling granularity, etc. Therefore, it is extremely challenging to fully explore the hidden information and obtain the unknown state and semantic information (e.g., waiting time, walking time, etc.) of each passenger without relying on any manual investigation.

In this paper, a spatiotemporal probabilistic graphical model based on the adaptive expectation-maximization attention algorithm (STPGM-AEMA) is proposed. The method can effectively recover the rich semantic and state information of each individual trajectory only from Automatic Fare Collection (AFC) data and Automatic Vehicle Location (AVL) data. Specifically, the main contributions of the paper are as follows:

1. A spatiotemporal probabilistic graphical model (STPGM) is proposed with global and local interactive representation to capture the complex spatiotemporal dependencies between individuals and system components (stations or trains) and obtain the individual trajectory at the train level, operating without manual survey data input.
2. Considering the sensitivity of the expectation-maximization (EM) approach to initial parameters, a novel data-driven parameter estimation framework is developed called the Adaptive Expectation-Maximization Attention Algorithm (AEMA). It can autonomously alternate between maximum likelihood estimation and latent variable information interpolation to return the missing information we want while ensuring fast and stable convergence.
3. Actual individual trajectory tracking (ITT) data is used to compare baselines on multiple OD pair datasets, thereby confirming the effectiveness and robustness of the proposed approach, STPGM-AEMA.

The paper is structured into six sections. Section 2 describes the problem of reconstructing individual trajectories with incomplete information. In Section 3, the trajectory inference model is developed, and the methods for parameter estimation are described in Section 4. Section 5 outlines the validation scenarios and compares various methods using real ITT data, followed by an interpretive analysis and a residual analysis of the model results. Finally, Section 6 elucidates the conclusion of the study.

2. Problem Description

In the closed URT system, it is assumed that the passenger i enters into the station s at t and leaves from station s' at t', as exemplified by OD pair on a single line in Figure 1. Only tap-in and tap-out events are recorded with spatiotemporal information from AFC data, and train arrival and departure events are obtained from AVL data. However, due to the low sampling frequency, the sequential events of each passenger, e.g., waiting for boarding event, boarding, and alighting event, and the associated state information, are severely missing in the system. This further results in the inability to obtain accurate system status (e.g., congestion state at platforms or on trains).

The information lost in a single trajectory is usually obtained through the spatiotemporal interpolation method, but it usually cannot satisfy the comparison of semantic information in URTS. Different from traditional methods, this paper aims to capture the missing spatiotemporal events, status, and semantic information in passenger trajectories through data mining, information interaction design, parameter learning, and probabilistic reasoning without manual investigation. This process is called individual trajectory reconstruction (ITR). It is worth noting that the problem of ITR in this paper is a further extension of individual trajectory prediction at the train level.
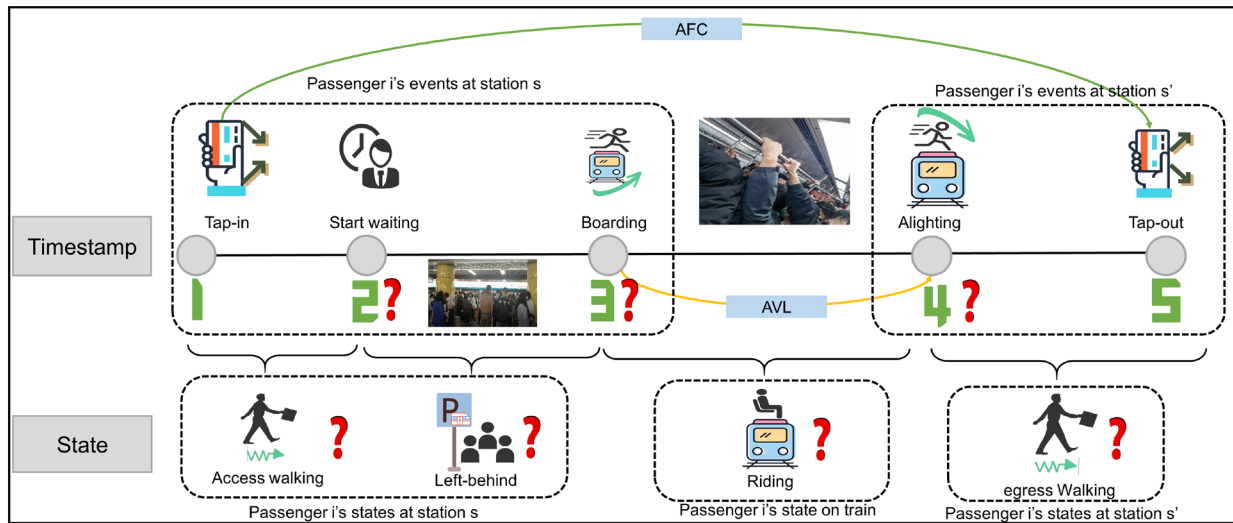
Figure 1. Example of individual travel itinerary.

Further, a set of journey among OD pairs is defined as $\mathbf{X} = \left\{ \mathbf{x}_1, \ldots, \mathbf{x}_I, \ldots, \mathbf{x}_N \right\}$, where $\mathbf{X}_I$ represents the original information that can be obtained, with the index being $I \in 1, \ldots, N$, and the total number of trips being N. Based on this, $\mathbf{D} = \left\{ \mathbf{D}_1, \ldots, \mathbf{D}_I, \ldots, \mathbf{D}_N \right\}$ is defined as a set of observable information, it comprises known itinerary information $\mathbf{X}_I$, system observability data $\mathbf{D}_{\text{sys}}$, and mined information $\mathbf{D}_{\text{mining}}$ encompassing individual trips, train operations, station flows, etc. It can be given as follows:

$$\mathbf{D} = \left\{ \mathbf{D}_I \right\} = \left\{ \mathbf{X}_I, \mathbf{D}_{\text{sys}}, \mathbf{D}_{\text{mining}} \right\} \tag{1}$$

Next, the individual trajectory tr_I is defined as being represented by a sequence of ordered spatiotemporal events E recorded in chronological order and a state vector S. tr_I can be stated as follows:

$$tr_I = \left\{ E, S \right\} = \left\{ \left\{ E_h \right\}, \left\{ S_f \right\} \right\}, h \in [1, M], f \in [1, W] \tag{2}$$

where E_h denotes a single spatiotemporal event, h is the event index, there are M in total. S_f indicates a state set between two adjacent events, including single or multiple status values. The value f is the state set index, there are W in total. Furthermore, a single event E_h is represented in the form of a ternary tuple, containing the characteristics of the moment of occurrence, location, and instantaneous behavior, namely:

$$E_h = \left(\mathrm{T}_{E_h}, \mathrm{L}_{E_h}, \mathrm{B}_{E_h} \right) \tag{3}$$

The passenger travel process consists of two main modes of spatial and temporal transitions, i.e., walking within the station or moving with the train. The state chain of an individual trajectory is defined as follows:

$$S_f = \left\{ S_{station}, S_{train} \right\} \tag{4}$$

where, $S_{station}$ and S_{train} represent the set of states of individuals at the origin/interchange/terminal and on the train, respectively. They can be represented by n-tuples. And every state is a scalar.

The overall trajectories Tr is a set composed of ordered spatiotemporal event sequences as $Tr = \left\{ tr_I \right\} = \left\{ tr_1, tr_2, \ldots, tr_N \right\}$.

Summarizing, this paper aims to interpolate missing spatiotemporal events in each travel trajectory and complement the semantic state information through probabilistic inference, which can naturally be represented by conditional probabilities $P\left(Tr \middle| \mathbf{X}\right)$. To achieve optimal estimation of individual itineraries inference, a probability-based framework is proposed. Within this framework, the core of ITR is reduced to an optimization problem, namely seeking the parameter configuration Θ that maximizes the posterior probability in the parameter space. This optimization problem can be formalized as follows:

$$\arg\max P\left(Tr \middle| \mathbf{X}\right) = \arg\max P\left(Tr \middle| \mathbf{D}\right) \propto \arg\max_{\Theta} P(\mathbf{D}, ? \mid \Theta) \tag{5}$$

3. Methodology

How to make the best use of limited information and infer high-fidelity individual trajectories through appropriate design is the key to methodology. A data-driven spatiotemporal probabilistic graphical model inference framework is proposed in the paper, which consists of three steps: potential set mining, modeling, and parameter estimation. The input data sources of the method are as follows: AFC, AVL, and Lines and Stations data. Where AFC records passengers' information, including their origin and destination stations and times of entry/exit. AVL data captures train operation information such as the train's ID, service line, station numbers, and arrival/departure time. Line and Station data provide physical distance and adjacency relations between stations. The outputs are spatiotemporal events and state information involved in individual trajectories. The key steps of the methodology are shown in Figure 2a–c, respectively.
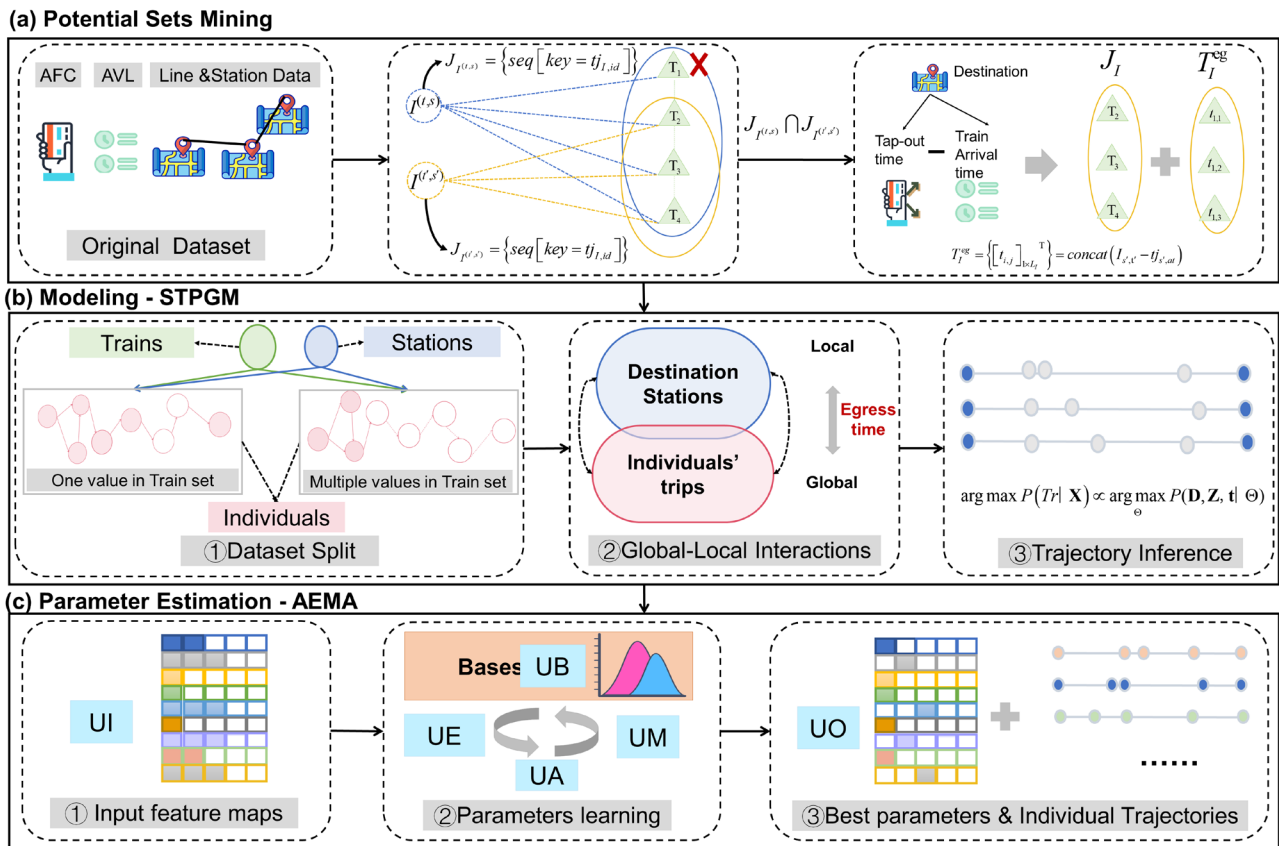
Figure 2. A data-driven spatiotemporal probabilistic graphical model inference framework for ITR.

3.1. Framework

In brief, the steps are as follows:

1. **Potential Sets Mining.** Considering the sequential nature of passengers' behaviors in spatiotemporal events, wherein each event is dependent on the preceding specific event, the get-off-leave-now (GOLN) principle is introduced. A feasible train alternative set for a journey as well as an egress time alternative set at the destination station of the individual are obtained, combined with complex spatiotemporal constraints and a combinatorial enumeration algorithm. This strategy can effectively reduce the space of candidate solutions under the premise of guaranteeing accuracy for subsequent computations.

2. **Modeling.** In order to suppress the bias caused by small-sample randomness, global and local latent variables are introduced to model the complex spatiotemporal dependencies of all trips and observed components (stations, trains) in the URT system. The construction of the model consists of three steps: dataset segmentation, global-local interaction representation, and trajectory inference. The main details of the model are presented in Section 3.3.

3. **Parameter Estimation.** To obtain the optimal parameters of the model and infer the most probable trajectories, an adaptive expectation-maximizing attention (AEMA) parameter learning method is proposed, which integrates a base adaptive embedding unit (UB), which provides automated a priori parameters to the likelihood function. Next, the introduction of the key-value attention computation unit (UA), where train labels can be matched to every individual trajectory. Details of the algorithm are given in Section 4.

3.2. Potential Sets Mining

The subsequent section outlines the necessary constraints and computational formulas for resolving both the set of train alternatives and individual travel alternatives. Finally, the combined enumeration method is used to obtain the collection. Appendix A provides relevant notation definitions.

Constraint 1. *The departure time* $tj_{s,dt}$ *of a potential train tj at the origin station* s *constraint. The departure time* $tj_{s,dt}$ *must be such that between the time period tap-in time* t *and tap-out time* t' *in itinerary* I.

$$I_{s,t} < tj_{s,dt} < I_{s',t'}, tj_{s,at} \neq tj_{s,dt} \ and \ tj_{s,q} \neq tj_{es,q} \tag{6}$$

The process generates a set of potential candidates for the train at the origin s called $J_{I^{(t,s)}}$:

$$J_{I^{(t,s)}} = \left\{ seq \left[key = tj_{id} \right] \right\} \tag{7}$$

Constraint 2. *The departure time* $tj_{s',dt}$ *of the potential train* tj *at the destination station* s' *constraint. The departure time* $tj_{s',dt}$ *must be such that between the time period tap-in time* t *and tap-out time* t' *in itinerary* I.

$$I_{s,t} < tj_{s',dt} < I_{s',t'}, tj_{s',at} \neq tj_{s',dt} \ and \ tj_{s',q} \neq tj_{fs,q} \tag{8}$$

A set of potential candidates for the train at destination s' can be generated called $J_{I^{(t',s')}}$:

$$J_{I^{(t',s')}} = \left\{ seq \left[key = tj_{id} \right] \right\} \tag{9}$$

The set of feasible train choices in the itinerary I can be obtained by taking the intersection, denoted as J_I:

$$J_I = J_{I^{(t,s)}} \bigcap J_{I^{(t',s')}} = \left\{ j = seq \left[key = tj_{I,id} \right]_{1 \times L_I} \right\}, j \in [1, \cdots, L_I] \tag{10}$$

Based on this premise, constructing the egress time sequence set in the itinerary I as T_I^{eg}. Each egress time value $t_{i,j}$ is calculated as the time difference between the tap-out time and the arrival time $j_{s',at}$ of the corresponding train of J_I.

$$T_I^{eg} = \left\{ \left[t_{i,j} \right]_{1 \times L_I}^{\mathrm{T}} \right\} = concat \left(I_{s',t'} - tj_{s',at} \right) \tag{11}$$

3.3. Modeling

The Bayes theorem principle and the backward inference method are introduced to establish a mechanism for global and local interactive representation. After obtaining the

optimal parameters, probabilistic reasoning about individual trajectories is realized. Figure 3 illustrates the trajectory inference framework based on STPGM, where color coding is employed to denote different categories of nodes and edges (refer to the legend for details). Events are represented as nodes, while edges describe potential spatial transition dependencies between state time intervals and events. Shaded nodes correspond to deterministic variables, whereas hollow circles indicate unobservable random variables. Solid and dashed lines distinguish deterministic relationships from uncertain ones, with unidirectional arrows representing causal relationships and bidirectional arrows indicating correlations.

As Equation (3), the set of nodes state as follows:

$$\{E_h\} = \{I_I, W_I, V_I^B, V_I^A, O_I\} \tag{12}$$

where, the events of tap-in I_I, waiting for boarding W_I, boarding V_I^B, alighting V_I^A, and tap-out O_I are represented in sequential order.

As in Equations (4) and (12), for OD pairs on a single line that do not require transfers, the state chain of an individual trajectory is defined as follows:

$$\{S_f\} = \{S_s, S_j, S_{s'}\} = \{(T_{AWT}, T_{WT}, T_{AT}), (T_{RT}), (T_{ET})\} \tag{13}$$

where, $S_s, S_j, S_{s'}$ represents the state of an individual at different spatial locations of the origin station s, train j, and destination station s', respectively. The value of the total time at the origin station T_{AWT} is calculated by summing the access time T_{AT} and the waiting time T_{WT}. T_{RT} denotes the running time on the train and T_{ET} indicates the egress time at the destination station. An individual trajectory tr_I can be represented as follows:

$$tr_I = \left\{ E = \begin{cases} I_I = (08:00:23, TTYB, \text{Tap}-in) \\ W_I = (08:01:28, TTYB, Start\ waiting) \\ V_I^B = (08:04:32, TTYB, Boarding\ Train\ j) \\ V_I^A = (08:28:20, DD, Alighting) \\ O_I = (08:29:03, DD, \text{Tap}-out) \end{cases}, S = \begin{cases} S_s = (249s, 65s, 184s) \\ S_j = (1428s) \\ S_{s'} = (43s) \end{cases} \right\}.$$

The inference tasks of this paper encompass the identification of waiting events at the platform, as well as the boarding and alighting events of passengers, along with a chain of unknown states. To establish the model, two strategies are employed:

1. The data is divided into deterministic dataset $\mathbf{D}_{L_I=1}$ and stochastic dataset $\mathbf{D}_{L_I>1}$ in order to generate prior samples.

2. A global-local interaction module is devised to transform the problem from maximizing the probability of individual trajectories to posterior parameter estimation based on the basis function. Building upon this foundation, boarding and alighting events are inferred by estimating egress time T_{ET}, then determining access time T_{AT} and waiting for the event through MCMC simulation, thereby achieving comprehensive inference of unknown events and latent states in trajectories. The modeling process consists of three steps which are described in detail below.
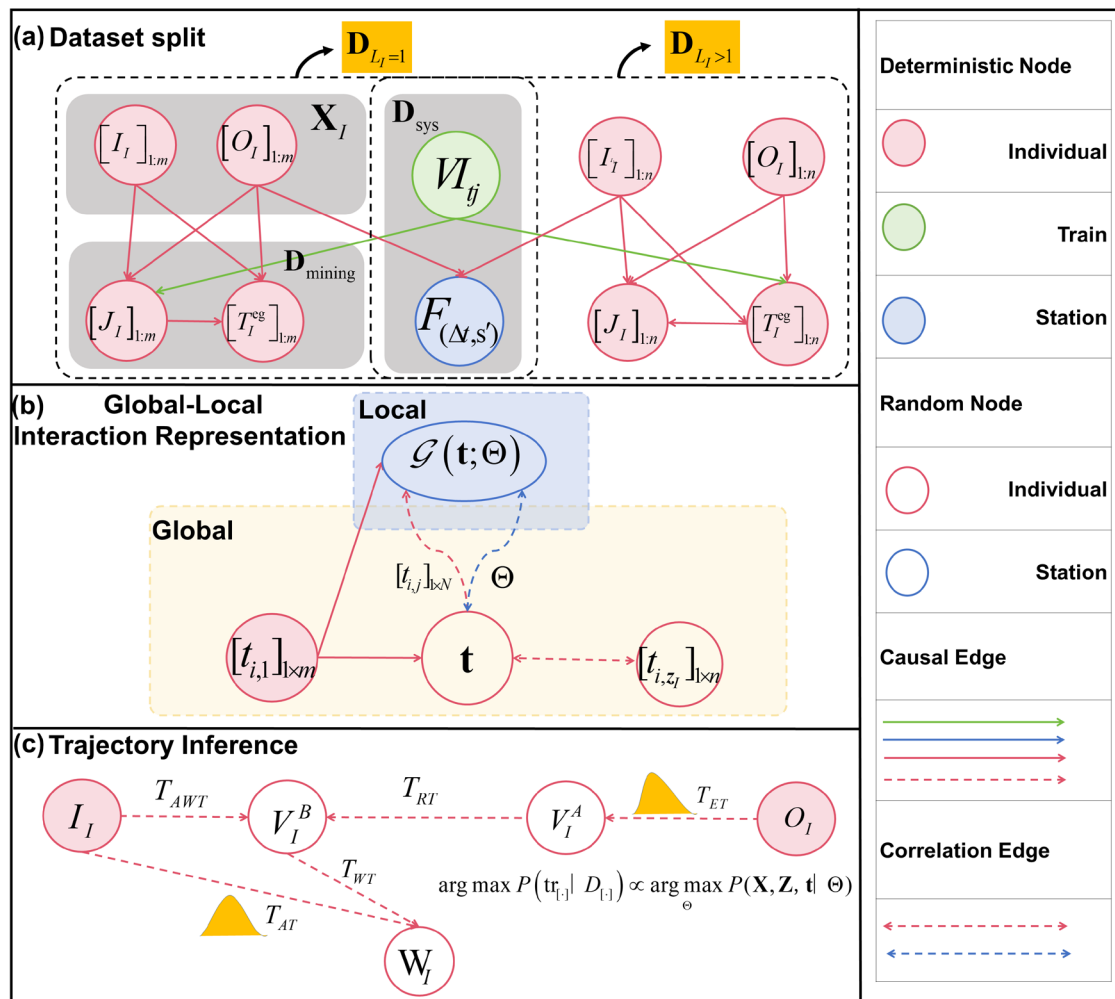
Figure 3. Trajectory inference framework based on STPGM.

3.3.1. Dataset Split

In this paper, the dataset is split into a deterministic dataset $\mathbf{D}_{L_I=1}$ and a stochastic dataset $\mathbf{D}_{L_I>1}$ with multiple alternatives, based on whether the number of options in the train candidate set is greater than one. Consequently, Equation (1) can be modified accordingly:

$$\mathbf{D} = \left\{ (\mathbf{D}_{L_I=1}, I = 1, \ldots, m), (\mathbf{D}_{L_I>1}, I = 1, \ldots, n) \right\} \tag{14}$$

Wherein, the numbers of samples in the deterministic dataset and the stochastic dataset are respectively denoted as m and n, with $m+n=N$. This approach benefits by providing prior data for the training of model parameters from the deterministic dataset $\mathbf{D}_{L_I=1}$, thereby replacing manual surveys and reducing the introduction of system noise.

Observable information is redefined based on node information, as shown in the dashed box on the left side of Figure 3a, taking the observable dataset as an example:

$$\mathbf{D}_{L_I=1} = \left\{ \mathbf{x}_I, \mathbf{D}_{\text{sys}}, \mathbf{D}_{\text{mining}} \right\}_{L_I=1} = \left\{ \left[I_I, O_I \right]_{1:m}, \left[\left(VI_{tj} \right), \left(F_{(\Delta t, s')} \right) \right], \left[J_I, T_I^{\text{eg}} \right]_{1:m} \right\} \tag{15}$$

where, an individual's journey $\mathbf{x}_I$ observations encompass tap-in event I_I and tap-out O_I event, while system observations $\mathbf{D}_{\text{sys}}$ include train operation events VI_{tj} and

outbound passenger flow within a specific time interval $F_{(\Delta t, s')}$. The mined information set $\mathbf{D}_{\mathrm{mining}}$ comprises a feasible train choices set J_I and potential egress time set T_I^{eg}, with their sample sizes remaining consistent. It is important to note that these observable pieces of information are either localized or aggregated. Similarly, this definition $\mathbf{D}_{L_I > 1}$ follows a similar logical framework.

3.3.2. Global-Local Interactive Representation

In the study of passenger journeys between OD pairs under incomplete information, the spatiotemporal dependency is manifested in the dynamics of individual travel events and state information as they evolve over time and space, exerting a significant influence on the local elements of the system. This paper introduces two latent variables to facilitate parameter estimation based on local elements, as elaborated below.

Global variable: latent variables z_I and $\mathbf{t}$. The index position corresponding to the individual egress time $t_{i,j}$ is set as a discrete random hidden variable z_I, following a multinomial distribution. The probability mass function can be expressed as follows:

$$P(z_I = j) = p_{ij}, j = 1, 2, \cdots L_I \tag{16}$$

where, p_{ij} represents the probability of selecting the jth index in T_I^{eg}, and satisfies $\sum_{j=1}^{L_I} p_{ij} = 1$, represents the probability distribution in the ordered sequence $j = 1, 2, \cdots L_I$. The complete hidden variable is denoted as $\mathbf{Z} = \{z_I\}$.

Moreover, in order to effectively characterize the parameter variations throughout the iterative process and disentangle the interdependencies between global and local elements, we propose a set of aggregate vectors referred to $\mathbf{t}$, which are composed of egress time $t_{i,j}$ for all individuals. Consequently, we obtain the following:

$$\mathbf{t} = [t_{i,j}]_{1 \times N} = \mathbf{c}\left([t_{i,1}]_{1 \times m}, [t_{i,z_I}]_{1 \times n}\right), t_{i,1} \in \mathbf{D}_{L_I=1}, t_{i,z_I} \in \mathbf{D}_{L_I > 1} \tag{17}$$

where, $t_{i,1}$ represents the unique egress time value from dataset $\mathbf{D}_{L_I=1}$, with the dimension being $1 \times m$; t_{i,z_I} states the jth(z_I) egress time component from the set T_I^{eg} of $\mathbf{D}_{L_I > 1}$, with the dimension being $1 \times n$; and $\mathbf{c}$ denotes the vector concatenation operation.

Local variable: basis function $\mathcal{G}(\bullet)$. In this paper, the distribution of egress time $\mathcal{G}(\bullet)$ is designed as a local variable. It is represented by a continuous probability distribution form that is integrable $\int_0^\infty x \cdot f(x)dx$ and $\int_0^\infty (x - \mu)^2 \cdot f(x)dx$ absolutely convergent, meaning it possesses finite mean and variance as a basis function $\mathcal{G}(\mathbf{t}; \Theta)$. The general form of representation is provided as follows:

$$\mathcal{G}(\mathbf{t}; \Theta) = \mathcal{G}(\mathbf{t}; \theta, \mu, \sigma^2) \tag{18}$$

where, Θ represents parameters related to the time scale Δt and exit station s', functioning as spatiotemporally adaptive parameters. θ denotes the intrinsic parameters of

the function $\mathcal{G}(\bullet)$ itself, μ determines the central position of the distribution, σ^2 describes the dispersion of data points around the mean, and $\mathbf{t}$ signifies the input value.

Interactive representation mechanism. Figure 3b shows the global-local interactive representation mechanism by basis function $\mathcal{G}(\bullet)$, latent variables z_I, and $\mathbf{t}$. Among them, vector $\mathbf{t}$ plays a key role. As a transmission channel, it not only aggregates the egress time information of all individuals but also provides the required input for updating the parameters of the basis function $\mathcal{G}(\bullet)$.

Specifically, in the process of transferring information from global parameters Θ to local parameters z_I, $\mathbf{t}$ collects the candidate egress time t_{i,z_I} generated by each individual in each iteration, passes these data to the function $\mathcal{G}(\bullet)$, and estimates by MLE to fit the parameters Θ. This step ensures that local parameter updates reflect the latest data in the context of the system.

In turn, the results of parameter optimization are used to construct the query vector and the global variable candidate solution t_{i,z_I} is used as the key component to construct the key-value pair for the next step of similarity comparison (to be described in detail in Section 4). Through this operation, we can re-evaluate and update everyone's t_{i,z_I}, thereby optimizing the performance of the entire system in each iteration.

3.3.3. Trajectory Inference

1. Calculate the maximum probability of individual trajectories. Under the principle GOLN, the problem of calculating the maximum probability of all journeys is equivalent to estimating the best parameters Θ^* of the basis function $\mathcal{G}(\bullet)$ by maximizing the probability of $\mathbf{t}$ under the influence of the latent variable $\mathbf{Z}$, making the observed data most likely to occur. Thus:

$$\arg\max P\left(Tr\middle|\ \mathbf{X}\right) \propto \arg\max_{\Theta} P(\mathbf{D},\mathbf{Z},\mathbf{t}|\ \Theta) \tag{19}$$

It is worth noting that the basis vector $\mathbf{t}$ serves as a conduit, facilitating the process of global-local interaction by aggregating individual travel time information and channeling it to the basis function $\mathcal{G}(\bullet)$ representing local characteristics. Subsequently, parameter updates occur during each iteration to ensure stability in parameter estimation, which is elaborated upon in Section 4.

2. Calculate unknown events and state variables. Figure 3c illustrates the trajectory inference process, indicating that the egress time of the passenger is $T_{ET} = t_{i,z_I}$. Subsequently, the individual's train ID is determined as follows:

$$tj_{I,id} = idx[i, j] \tag{20}$$

where, $idx[\bullet]$ represents a mapping function used to locate an element based on its index number. Further, the spatiotemporal characteristics of the boarding event V_I^B and alighting event V_I^A are established, and the running time duration and access time are computed by $T_{RT} = \mathrm{T}_{V_I^A} - \mathrm{T}_{V_I^B}$ and $T_{AWT} = \mathrm{T}_{V_I^B} - \mathrm{T}_{I_I}$, individually.

The next step is to use the formula quoted by Zhu et al. [20] to calculate the waiting time T_{WT}, which is a typical formula. And the access time of every passenger can be calculated by $T_{AT} = T_{AWT} - T_{WT}$. However, it was found that about 10% of the results of T_{AT} are negative by calculating T_{WT} and T_{AT}, using 26,000 samples from one line of the Beijing Rail Transit System (BRTS) during the peak period. This reveals that this method may not be suitable for BRTS. Building upon the decomposition method outlined in Equation (14), this study defines the waiting duration of the sample set $\mathbf{x}_I \in \mathbf{X}_{L_I=1}$ as follows:

$$T_{WT}\left(\mathbf{x}_I \in \mathbf{X}_{L_I=1}\right) = \frac{E(H_{j \to j+1})}{2} + \frac{\mathrm{Var}(H_{j \to j+1})}{2E(H_{j \to j+1})} \tag{21}$$

where, $H_{j \to j+1}$ states the departure interval between jth and $j+1$th train.

The access time of the sample set $\mathbf{x}_I \in \mathbf{X}_{L_I=1}$ is calculated by means of a piecewise function as follows:

$$T_{AT}\left(\mathbf{x}_I \in \mathbf{X}_{L_I=1}\right) = \begin{cases} T_{AWT} - T_{WT}, & if\ T_{AWT} > T_{WT} \\ T_{AWT}, & else \end{cases} \tag{22}$$

On this basis, the waiting duration of the sample set $\mathbf{x}_I \in \mathbf{X}_{L_I>1}$ calculated by the following:

$$T_{WT}\left(\mathbf{x}_I \in \mathbf{X}_{L_I>1}\right) = T_{AWT} - T_{AT} \tag{23}$$

4. Parameter Estimation

The second challenge addressed in this paper is how to design a likelihood function that maximizes the reconstruction of high-fidelity trajectories for all individuals Tr, considering dependence between $\mathbf{t}$ and Θ. To tackle this, we draw inspiration from the GMM for mixed distributions [35] and the EMA for semantic segmentation [44]. The Adaptive Expectation-Maximization Attention (AEMA) algorithm is proposed, which incorporates the EM algorithm and attention mechanisms.

The idea behind the proposed algorithm is derived from how to establish a data flow mechanism between global variables representing all individuals and local features representing the system, which is crucial for fully data-driven algorithms.

The AEMA algorithm consists of an input and output unit and four main operation units (as shown in Figure 4), namely: Input Unit (UI), Bases Adaptability Embedding Unit (UB), Expectation-Step Unit (UE), Key-Value Attention Calculation Unit (UA), Maximization-Step Unit (UM), and Output Unit (UO). In brief, the UI is the first step of AEMA, aimed at providing the UB with observable dataset inputs. The UB is responsible for dynamically obtaining initial base vectors providing an initial parameter for fitting the distribution. UE, the E-step in the EM algorithm, defines the objective function under prior parameters. UA provides methods for computing the posterior distribution of hidden variables. UM, the M-step in the EM algorithm, aims to maximize until convergence criteria are met. The UO outputs the reconstructed trajectories. Each step is explained below.
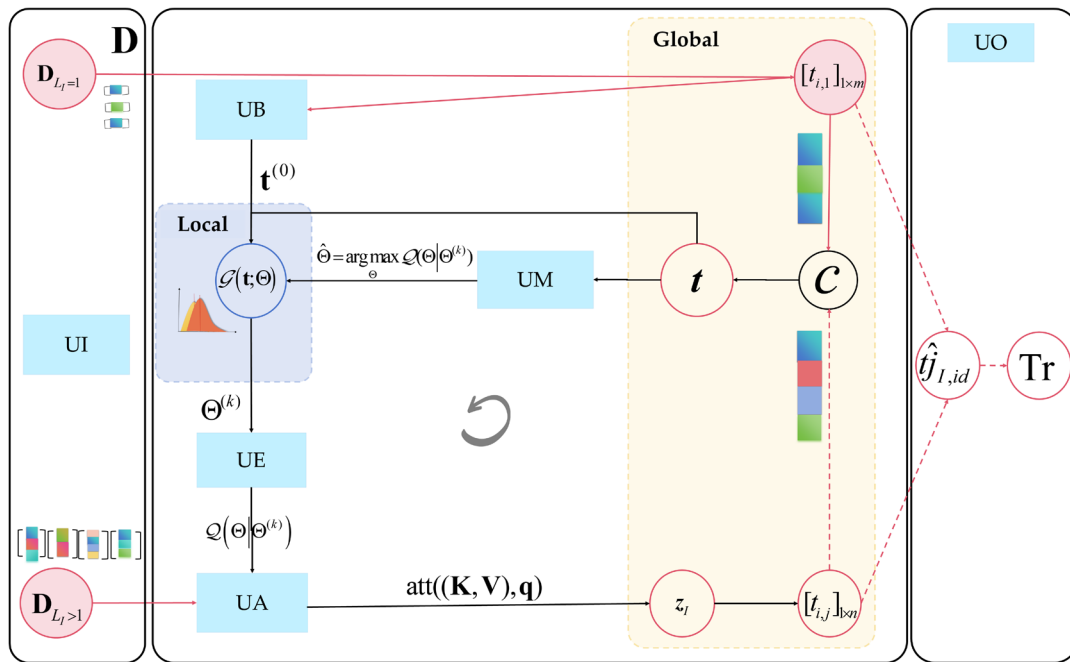
Figure 4. Parametric learning process based on the AEMA algorithm. .

4.1. Input Unit (UI)

UI is responsible for inputting the observable data set and defining the parameters of the core steps. It is important to note that $\{\mathbf{X}, [t_{i,j}], [[z_{i,j}]]\}$ should be considered as complete observation data, with the parameters to be estimated as $\Theta = \{\theta, \mu, \sigma^2, [[p_{ij}]]\}$, and the number of parameters as $3 + N * L_I$. Let the likelihood function be $\mathcal{L}(\Theta) = P(\mathbf{D}, \mathbf{Z}, \mathbf{t} \mid \Theta)$, with the conditional distribution of the latent variables $\mathbf{Z}$ and $\mathbf{t}$ being $P(\mathbf{Z}, \mathbf{t} \mid \mathbf{D}, \Theta^{(k)})$. Where $\Theta^{(k)}$ represents the parameter estimated in the kth iteration. The parameters in the $k+1$th round, Θ^* are thus the target parameter values to be maximized.

4.2. Bases Adaptability Embedding Unit (UB)

Prior information is typically obtained through surveys involving small labeled datasets, followed by fitting the parameters $\Theta^{(0)}$ through maximum likelihood estimation (MLE) [20]. However, the calculation of this mean value μ cannot be adaptively chosen, and different scenarios require different survey data. UB is responsible for acquiring prior information and initializing parameters. It dynamically obtains samples from the dataset $\mathbf{D}_{L_I=1}$ based on station s' and time interval Δt as inputs for prior knowledge, supporting the automated calculation of prior parameters $\Theta^{(0)}$ for using the MLE method, replacing the practice in traditional EM algorithms of randomly initializing model parameters. Let:

$$\Theta^{(0)} = \arg\max_{\Theta} \prod_{i=1}^{m} \mathcal{G}\left(\Theta \big| [t_{i,1}]_{1 \times m}\right), t_{i,1} \in \mathbf{D}_{L_I=1} \tag{24}$$

In summary, the UB exhibits the capability to automatically capture spatiotemporal information thereby enhancing the model's robustness and accuracy in handling intricate spatiotemporal correlations and dynamic patterns. This effectively addresses the issue of sensitivity in parameter initialization.

4.3. Expectation-Step Unit (UE)

The marginal likelihood function of a sample is denoted as $p(\mathbf{D}_I, z_I, t_{i,j} \mid \Theta)$, while the conditional distribution probability of the latent variables is represented by $p(z_I, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)})$. By applying Jensen's inequality, the log-likelihood function is:

$$
\begin{aligned}
\ln \mathcal{L}(\Theta) &= \ln \sum_{\mathbf{Z}} p(\mathbf{D}_I, \mathbf{Z}, t_{i,j} \mid \Theta) = \sum_N \sum_{z_I} \ln p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta) \\
&= \sum_N \sum_{z_I} \ln \frac{p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta^{(k)}) \cdot p(z_{i,j}, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)})}{p(z_{i,j}, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)})} \\
&\geq \sum_N \sum_{z_I} p(z_{i,j}, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)}) \ln p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta) \Big/ p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta^{(k)}) \\
&= \sum_N \sum_{z_I} p(z_{i,j}, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)}) \ln p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta) - \sum_N \sum_{z_I} p(z_{i,j}, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)}) \ln p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta^{(k)})
\end{aligned}
\tag{25}
$$

where $\sum_N \sum_{z_I} p(z_{i,j}, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)}) \ln p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta^{(k)})$ is a constant, then the objective function is $\mathcal{Q}\left(\Theta \middle| \Theta^{(k)}\right)$, which is given by the following:

$$
\begin{aligned}
\mathcal{Q}\left(\Theta \middle| \Theta^{(k)}\right) &= \sum_N \sum_{z_I} p(z_{i,j}, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)}) \ln p(\mathbf{D}_I, z_{i,j}, t_{i,j} \mid \Theta) \\
&= \mathrm{E}_{\mathbf{Z}}[\ln P(\mathbf{D}, \mathbf{Z} \mid \Theta) \mid \mathbf{D}, \Theta^{(k)}]
\end{aligned}
\tag{26}
$$

It is evident that we need to calculate the conditional distribution probability $p(z_I, t_{i,j} \mid \mathbf{D}_I, \Theta^{(k)})$ for each individual and use it as the maximization target function $\mathcal{Q}\left(\Theta \middle| \Theta^{(k)}\right)$ in the UM. However, due to the nonlinearity of high-order terms in the objective function, conventional optimization methods may be sensitive to initial parameters and prone to local optima. To more effectively capture the relationship between an individual's egress time sequence set T_I^{eg} and the station's local prior information while reducing the communication cost, this paper draws on a key-value attention mechanism. This involves constructing a query vector representing local parameters and determining attention weights for key-value pairs $\left(\mathbf{k}_I, \mathbf{v}_I\right)$ associated with each individual during the UA step.

4.4. Key-Value Attention Calculation Unit (UA)

In this step, the distribution of the latent variables $\mathbf{Z}$ and $\mathbf{t}$ are calculated. Figure 5 shows the entire calculation process of UA. The specific steps are as follows:
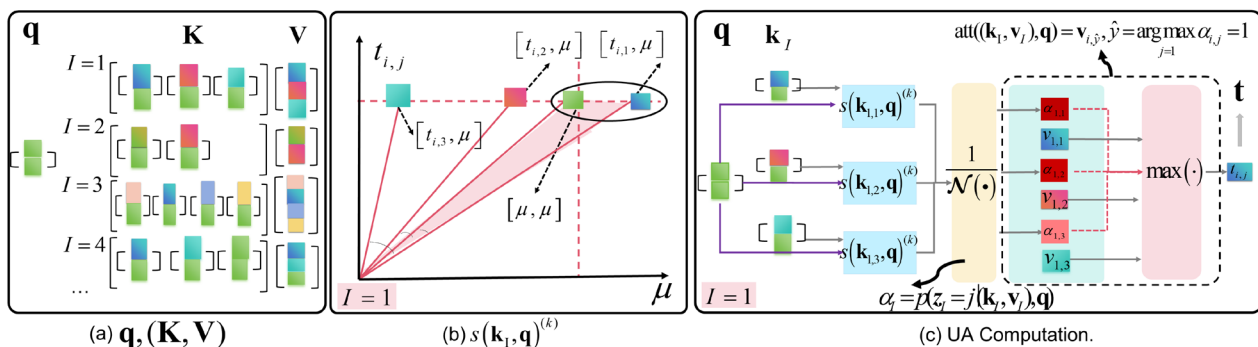

Figure 5. Calculation process of key-value attention unit (UA).

First, the query vector (as shown in Figure 5a). It is constructed by leveraging the broadcasting method commonly used in deep learning to transform scalar values into a vector representation. Specifically, we utilize the parameter u to create a vectorized form denoted as $\mathbf{q}^{(k)}_{T^{eg}_{(\Delta t, s')}}$, which stands for $\mathbf{q}$. Let:

$$\mathbf{q}^{(k)}_{T^{eg}_{(\Delta t, s')}} = \left[\mu^{(k)}, \mu^{(k)} \right]^{T}, \forall \mu \in \Theta \tag{27}$$

In fact, this definition expands the information aggregated at the current station. Furthermore, the initial value $\mathbf{q}^{(0)}$ comes from the sample set $\mathbf{D}_{L_I = 1}$.

Second, key-value pairs (as shown in Figure 5a). Let $\mathbf{k}_I = \left[\left[t_{i,1}, \mu^{(k)} \right]^{T}, \cdots, \left[t_{i,j}, \mu^{(k)} \right]^{T} \right]_{1 \times L_I}$ used to simulate the values that each individual with multiple potential trajectories can take in each round, and let $\mathbf{v}_I = \left[t_{i,1} \quad \cdots \quad t_{i,j} \right]^{T}_{1 \times L_I}$ represent the egress time corresponding to the index position of the latent variable. So there are:

$$(\mathbf{K}, \mathbf{V})^{(k)} = \left\{ \left(\mathbf{k}_I, \mathbf{v}_I \right) \right\}^{(k)}, \ I \in [1, \ldots, n] \tag{28}$$

Third, the scoring function is defined (as depicted in Figure 5b). The role of this function is to compute the correlation between each input vector $\mathbf{k}_I$ and the query vector $\mathbf{q}$. In the standard Scaled Dot Product Model, the dot product operation tends to be more sensitive to a larger value. This study focuses on computing proximity values between vectors $\mathbf{k}_I$ and $\mathbf{q}$ for assigning higher weights accordingly. Hence, cosine similarity based on vector angle principles is chosen as the definition for the scoring function. This approach not only considers individual value relationships with groups but also accounts for self-relationship. Let:

$$s\left(\mathbf{k}_I, \mathbf{q} \right)^{(k)} = \frac{\mathbf{k}_I \cdot \mathbf{q}}{\left\| \mathbf{k}_I \right\| \cdot \left\| \mathbf{q} \right\|} \tag{29}$$

where, the inner product of vectors, denoted by $\mathbf{k}_I \cdot \mathbf{q}$, $\left\| \cdot \right\|$ represents the norm of a vector. A smaller angle indicates higher similarity.

Moreover, calculate the attention function (as shown in Figure 5c). The attention distribution α_I represents the degree of attention the jth component of $\mathbf{k}_I$, given the query vector $\mathbf{q}$. Specifically, when dealing with class-imbalanced data, traditional attention functions such as $Softmax(\cdot)$ often fail to provide sufficient learning opportunities for minority classes because they may be suppressed by dominant classes during computation. To overcome this, we use a normalization function $\mathcal{N}(\cdot)$ to replace the traditional activation function, which can enhance the model's focus on minority class features and learning efficiency. The calculation of α_I amounts to computing the posterior probability distribution of z_I, which is as follows:

$$\alpha_I = [\alpha_{i,j}] = p(z_I = j | (\mathbf{k}_I, \mathbf{v}_I), \mathbf{q}) = \frac{s\left(\mathbf{k}_I, \mathbf{q} \right)}{\mathcal{N}(\cdot)} = s\left(\mathbf{k}_I, \mathbf{q} \right) \Big/ \sum_1^n s\left(\mathbf{k}_I, \mathbf{q} \right) \tag{30}$$

The attention mechanism offers two available forms, hard attention is chosen in this paper. Subsequently, the attention function is defined as:

$$\text{att}((\mathbf{k}_I, \mathbf{v}_I), \mathbf{q}) = \mathbf{v}_{i,\hat{y}} = [t_{i,\hat{y}}], \hat{y} = \arg\max_{j=1} \alpha_{i,j} \tag{31}$$

where $\hat{y}$ is the subscript of the input vector $\mathbf{v}_{i,j}$ with the greatest probability, name is $\arg\max\limits_{j=1} \alpha_{i,j}$.

Finally, the basis variable $\mathbf{t}$ is calculated, let:

$$\mathbf{t} = [t_{i,j}]_{1\times N} = \mathbf{c}\left([t_{i,1}]_{1\times m}, [t_{i,\hat{y}}]_{1\times n}\right), t_{i,1} \in \mathbf{D}_{L_I=1} \tag{32}$$

4.5. Maximization-Step Unit (UM)

The main objective in this step is to solve the optimization problem by utilizing the complete observed variables and obtaining maximum likelihood estimates of the parameters, which is given by Equation (26). Let:

$$\Theta^* = \arg\max_{\Theta} \mathcal{Q}\left(\Theta \middle| \Theta^{(k)}\right) \tag{33}$$

Finally, by alternating iterations among the UE, UA, and UM until convergence is achieved. In this process, three termination conditions are set. (1) The local variable Θ is controlled by the Tolerance parameter ϵ_1 (set to 1×10^{-3}). (2) The global variables convergence tolerance is managed by the tolerance of the objective function ϵ_2 (set to 1×10^{-1}). (3) The overall convergence speed is controlled by setting a maximum number of iterations K, which is set to 50 times. If the algorithm satisfies the tolerance conditions before reaching the maximum number of iterations, it will stop prematurely. That is as follows:

$$\begin{cases} \left\|\Theta^{(k+1)} - \Theta^{(k)}\right\| < \epsilon_1 \text{ or} \\ Q\left(\Theta^{(k+1)}, \Theta^{(k)}\right) - Q\left(\Theta^{(k)}, \Theta^{(k)}\right) < \epsilon_2 \text{ or} \\ k = K \end{cases} \tag{34}$$

4.6. Output Unit (UO)

The train ID serves as a primary criterion for assessing the accuracy of the spatial position of trajectories in ITR. According to Equation (18), it is known that $\hat{tj}_{I,id} = idx[i, \hat{y}]$. Further, the samples for $\mathbf{x}_I \in \mathbf{X}_{L_I>1}$ are simulated to generate values T_{AT} using the NUTS algorithm for MCMC sampling with the pyMC3 library. Based on this, by applying the formulas from Section 3.3.3, we can provide detailed information on the attributes and state features of unknown events involved for all passengers.

5. Experiments

In traditional methods, local features (like egress time distribution) are simulated to verify the accuracy of parameters or the usability of methods [12,42], but they are rarely considered from the perspective of actual individual trajectories. Verifying the method from a bottom-level rather than an aggregate perspective is also one of the contributions of this article.

5.1. Dataset Description

5.1.1. Design of Individual Trajectory Tracking Simulation Experiment

During peak hours, a large number of passengers entering the origin station or transferring at the transfer station will flow into the platform; meanwhile, many passengers will flow out at other destination stations. This may cause local congestion and increase the complexity of the spatiotemporal modeling of travel trajectories. Therefore, in the design stage of the individual trajectory tracking simulation experiment, this study pays special attention to capturing the tidal characteristics of the passenger flow [35]. Take the CY station in Line 6 as an example. The station is located in a suburban area with more residential areas in the neighborhood, and the main service targets are commuters. During the morning peak period, most passengers entering the station from 7:00 to 09:00 are heading towards downtown Beijing, and the passenger inflow during this period accounts for about 50% of the total daily passenger inflow. In contrast, the number of passengers entering the station from 17:00 to 20:00 in the evening is significantly lower, accounting for only about 15% of the total.

In addition, in order to better validate the effectiveness and applicability of the STPGM-AEMA method proposed in this article, several factors are considered in depth in this study. First, this study covers OD pairs with different distances and station types to ensure the comprehensiveness and representativeness of the experimental setup. Moreover, the investigation team is from Beijing Metro Network Management Co. Ltd., Beijing, China, the official regulator of the BRTS. Considering the human and material conditions coordinated by this agency, we carefully selected four typical OD pairs for experimental validation. See Section 5.1.2 for details.

Next, in the actual simulation phase, the investigation team simulates the travel process of actual passengers on specific OD pairs and obtained data on each timestamp on the travel chain (as shown in Figure 6), including tap-in time, time of arrival at the platform, time of train departure at the origin station, boarding time, time of train arrival at the destination station, alighting time, and tap-out time. Simultaneously, the integration of AFC and AVL data was employed to generate travel chain status information encompassing behavioral semantics such as access time to the platform, waiting time, running time, and egress time at the destination station. The timestamp and state data tables are provided in Appendixes B and C, respectively, serving as a genuine and dependable foundation for evaluating the efficacy of the proposed methodology.
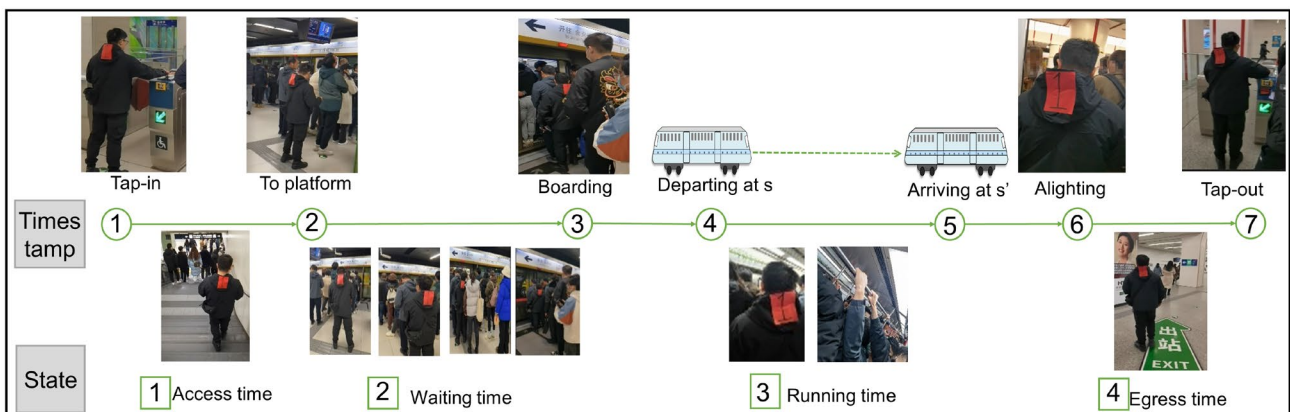

Figure 6. Actual trajectory to obtain experimental records.

5.1.2. Data Introduction

As an illustration in Figure 7, typical OD pairs from Line 5 and Line 6 of the BRTS during peak periods were selected. The complete dataset comprises four data groups (D1~D4) collected in March 2023, including two morning peak period OD pairs (TTYB-

DD, CY-DS) and two evening peak period OD pairs (PHY-TTYB, HJL-CY). Table 1 presents information related to the dataset that encompasses three parts: basic information, validation data information, and training data information. Basic information includes route details, such as the number of stations and distance covered. Validation data eliminates 10 trajectories that do not meet requirements, with a total sample size of ITT of 78. It is noteworthy that AFC and AVL data must correspond one-to-one in terms of dates and periods, while ITT data is included in all AFC data for training purposes. Training data involves multiple days, with a total of 3851 AFC samples and 1069 AVL samples.
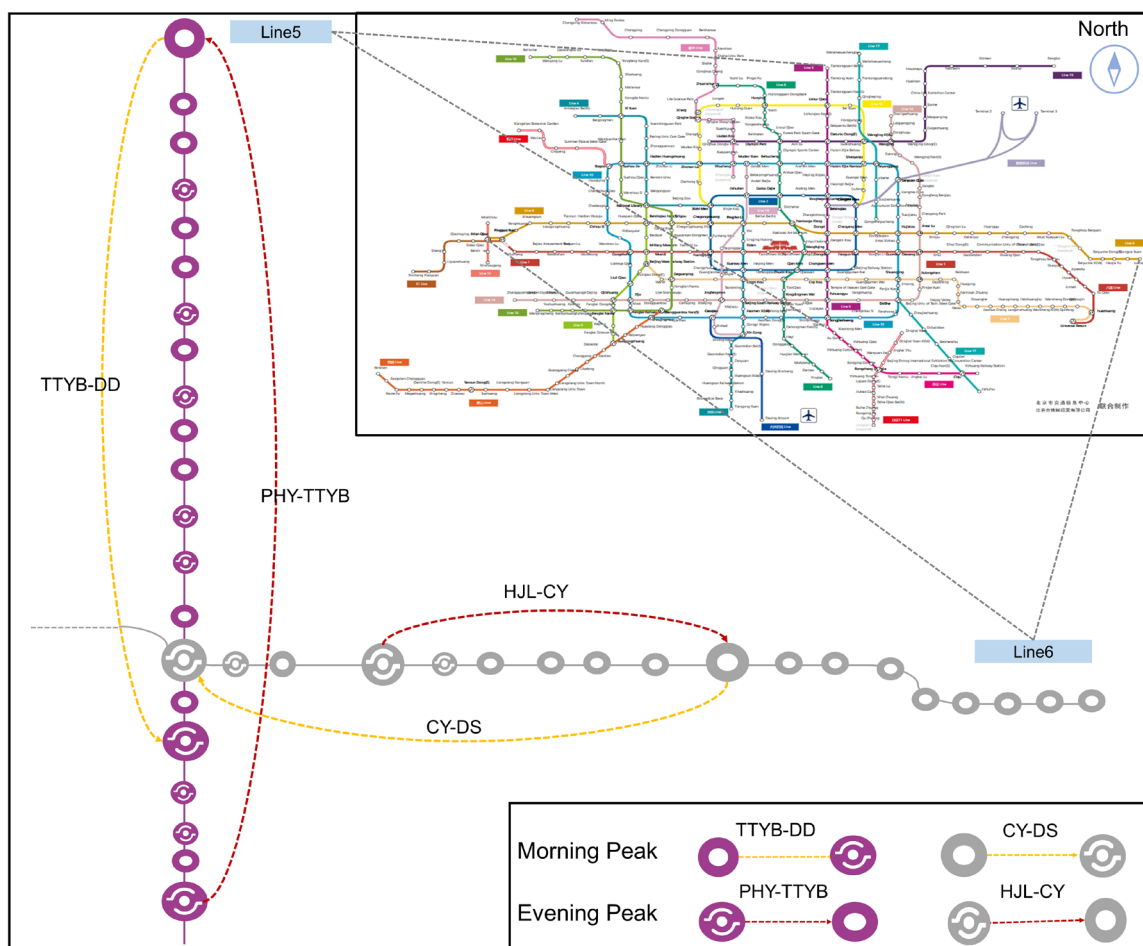

Figure 7. Typical OD pairs.

The parameter learning process involves training utilizing AEMA, followed by performing probabilistic inference and subsequently comparing the inference results with the timestamp and status information of the verification data. The L-BFGS-B optimization algorithm is predominantly employed for parameter learning among various comparison algorithms. All algorithms are compiled and executed on a computer equipped with an Intel(R) Core(TM) i9-10920X CPU processor and 48 GB of memory.

Table 1. Dataset Description.

Description	Dataset	D1	D2	D3	D4
	OD pair	TTYB-DD	CY-DS	PHY-TTYB	HJL-CY
I. Base Information	Line	5	6	5	6
	Station Numbers	17	10	21	7
	Distance(m)	19,700	15,771	24,480	11,859
II. Validation Data	Time Duration	07:00–09:00 a.m.		17:00–19:00	

Day	21 March 2023	22 March 2023	21 March 2023	22 March 2023
ITT Numbers	21	19	20	18
Days	6	6	4	4
III. Training Data AFC Samples	1359	699	206	1587
Train Numbers	340	309	225	195

5.2. Baselines

This paper compares the proposed method, STPGM-AEMA, with traditional rule-based approaches and Bayesian methods at the train-level scale. The methods are outlined as follows:

1. **LTRM (Last Train Rule-based Model):** Spatiotemporal Segmentation of Metro Trips algorithm searching for "BORDER-WALKERS" using the nearest timestamp principle, proposed by Zhang et al. [15], wherein the train's departure time closest to the passenger's tap-out time at the destination station was utilized to determine the train they boarded. Luo et al. [45] also employed this rule to infer passenger trajectories. Furthermore, both studies assumed "speed invariance" as a behavioral postulate.

2. **PTAM-MLE (Passenger-to-Train Assignment Model with MLE)**: Zhu et al. [20] proposed a probabilistic approach, named PTAM, which requires AFC/AVL data and the station's walking speed distribution as inputs. To ensure consistency in measuring speed, this paper replaces it with their later proposed LBPMF [42], where the input is the egress/access time distribution and the likelihood function is expressed accordingly.

3. **MPTAM-EM (Modified Passenger-to-Train Assignment Model with EM)**: A modified model MPTAM was constructed by Xiong et al. [12], and the EM algorithm was proposed for estimating the parameters of the egress time distribution and the boarding probability distribution function, and the likelihood function was formulated by them.

4. **STPGM-EMA (without UB)**: The proposed STPGM-AEMA algorithm forms the basis of this method, which entails the removal of the UB module.

5. **STPGM-AEM (without UA)**: Similarly, the proposed STPGM-AEMA algorithm forms the basis of this method, which entails the removal of the UA module.

5.3. Evaluation Metrics

The present study employs two categories of metrics to assess its accuracy and robustness. The details are given below.

5.3.1. Accuracy Evaluation Metrics

Considering the precision of evaluating the multiclassification problem, a confusion matrix is introduced, and six evaluation metrics are chosen: macro-precision, macro-recall, macro-F1 score, micro-precision, micro-recall, and micro-F1 score, and calculated based on true positives (TP), false positives (FP), and false negatives (FN) across all categories. These metrics serve as indicators for model performance improvement; higher values indicate better results. The formula is as follows:

$$P_{Macro} = \left. \sum_{c=1}^{N} P_c \middle/ N \right. \tag{35}$$

$$R_{Macro} = \left. \sum_{c=1}^{N} R_c \middle/ N \right. \tag{36}$$

$$F1_{Macro} = \sum_{c=1}^{N} F1_c \Big/ N \tag{37}$$

$$P_{Micro} = \sum_{c=1}^{N} TP_c \Big/ \left(\sum_{c=1}^{N} TP_c + \sum_{c=1}^{N} FP_c \right) \tag{38}$$

$$R_{Micro} = \sum_{c=1}^{k} TP_c \Big/ \left(\sum_{c=1}^{N} TP_c + \sum_{c=1}^{N} FN_c \right) \tag{39}$$

$$F1_{Micro} = 2 \cdot \frac{P_{Micro} \cdot R_{Micro}}{P_{Micro} + R_{Micro}} \tag{40}$$

where, c is the index of categories.

5.3.2. Consistency Evaluation Metric

Considering the impact of random classification by the model, this paper introduces Cohen's Kappa coefficient (K) to calculate the overall consistency and random agreement between observed and predicted values. The *K* measures the model's resistance to interference in the presence of a class imbalance, serving as a statistical measure to evaluate credibility. The formula is as follows:

$$K = \frac{P_o - P_e}{1 - P_e} \tag{41}$$

where, the variable P_o represents the observed accuracy, i.e., the proportion of correctly classified instances. P_e denotes the expected accuracy, which refers to the proportion of instances correctly classified by chance. The *K* value ranges between [−1, 1], with higher values indicating a model's more genuine resistance to randomness. A *K* value closer to 1 signifies a model's perfect agreement with reality; $K = 0$ indicates the model's performance is equivalent to random classification; and $K < 0$ suggests the model's performance is even worse than random classification.

5.4. Result

5.4.1. Accuracy Evaluation Results

Figure 8 presents the classification results across four datasets (D1–D4) utilizing optimal parameters, depicted through a Confusion Matrix format. Each column is allocated to a dataset, while each row showcases the efficacy of a specific method applied to that dataset. Predicted labels are displayed along the horizontal axis, with true labels along the vertical axis. Areas of correct classification are marked in green, whereas inaccuracies are highlighted in red, accompanied by percentages that reflect the proportion of correct and incorrect classifications. The intensity of the color signifies proportionality, with darker shades indicating a higher frequency of occurrences. Owing to the consistent outcomes between the STPGM-EMA and STPGM-AEMA methods, their results have been consolidated for representation.
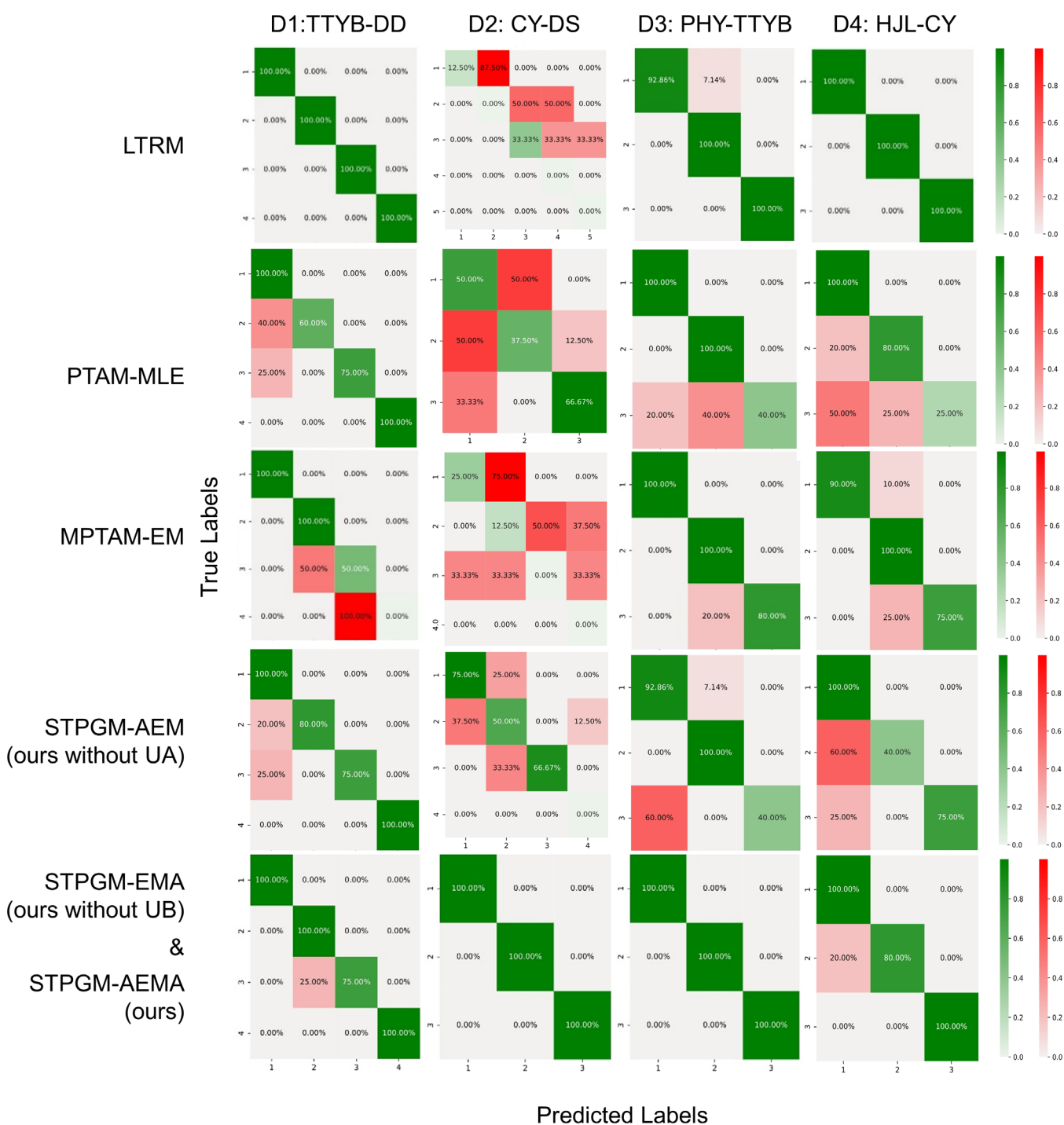
Figure 8. Comparison of the confusion matrix.

The experimental analysis elucidates that while most algorithms fare well in scenarios with a single alternative train option, their efficacy diminishes in contexts with multiple train choices, illustrating a notable challenge in navigating complex classification landscapes. This delineates a direct linkage between the number of potential train choices and the escalation of uncertainty in passenger trajectories, inherently augmenting the likelihood of misclassification. A cross-comparison of various methods reveals that the UA module plays a crucial role in the STPGM-AEMA framework to capture data details. This highlights a direct correlation between the number of potential train selections and an escalation in passenger trajectory uncertainty.

Table 2 shows the results of the accuracy assessment of the different algorithms on each dataset. It is particularly noteworthy that STPGM-AEMA(ours) and STPGM-EMA (ours without UB) perform well on all datasets, while the other algorithms perform poorly at least on the D2 dataset. the prediction accuracy of the STPGM_AEMA method proposed in this article reaches more than 90% on all datasets, showing that the algorithm can cope with scenarios of different complexity levels. Despite some random errors, the overall robustness is good.

Table 2. Results of accuracy evaluation metrics.

Dataset	Methods	P_{Macro}	R_{Macro}	$F1_{Macro}$	P_{Micro}	R_{Micro}	$F1_{Micro}$
D1: TTYB-DD	LTRM	**1.00×10^0** [1]	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**
	PTAM-MLE	9.46×10^{-1}	8.38×10^{-1}	8.72×10^{-1}	8.57×10^{-1}	8.57×10^{-1}	8.57×10^{-1}
	MPTAM-EM	5.95×10^{-1}	6.25×10^{-1}	6.01×10^{-1}	8.57×10^{-1}	8.57×10^{-1}	8.57×10^{-1}
	STPGM-AEM	9.62×10^{-1} [1]	8.88×10^{-1}	9.16×10^{-1}	9.05×10^{-1}	9.05×10^{-1}	9.05×10^{-1}
	STPGM-EMA	9.58×10^{-1}	9.38×10^{-1}	9.42×10^{-1}	9.52×10^{-1}	9.52×10^{-1}	9.52×10^{-1}
	STPGM-AEMA(ours)	9.58×10^{-1}	9.38×10^{-1}	9.42×10^{-1}	9.52×10^{-1}	9.52×10^{-1}	9.52×10^{-1}
D2: CY-DS	LTRM	2.40×10^{-1}	9.17×10^{-2}	9.44×10^{-2}	1.05×10^{-1}	1.05×10^{-1}	1.05×10^{-1}
	PTAM-MLE	5.13×10^{-1}	5.14×10^{-1}	5.12×10^{-1}	4.74×10^{-1}	4.74×10^{-1}	4.74×10^{-1}
	MPTAM-EM	1.98×10^{-1}	9.38×10^{-2}	1.22×10^{-1}	1.58×10^{-1}	1.58×10^{-1}	1.58×10^{-1}
	STPGM-AEM	5.60×10^{-1}	4.79×10^{-1}	5.10×10^{-1}	6.32×10^{-1}	6.32×10^{-1}	6.32×10^{-1}
	STPGM-EMA	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**
	STPGM-AEMA(ours)	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**
D3: PHY-TTYB	LTRM	8.33×10^{-1}	9.76×10^{-1}	8.77×10^{-1}	9.50×10^{-1}	9.50×10^{-1}	9.50×10^{-1}
	PTAM-MLE	7.56×10^{-1}	8.00×10^{-1}	6.79×10^{-1}	8.50×10^{-1}	8.50×10^{-1}	8.50×10^{-1}
	MPTAM-EM	8.33×10^{-1}	9.33×10^{-1}	8.52×10^{-1}	9.50×10^{-1}	9.50×10^{-1}	9.50×10^{-1}
	STPGM-AEM	7.71×10^{-1}	7.76×10^{-1}	7.02×10^{-1}	8.00×10^{-1}	8.00×10^{-1}	8.00×10^{-1}
	STPGM-EMA	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**
	STPGM-AEMA(ours)	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**
D4: HJL-CY	LTRM	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**	**1.00×10^0**
	PTAM-MLE	8.56×10^{-1}	6.83×10^{-1}	6.90×10^{-1}	7.89×10^{-1}	7.89×10^{-1}	7.89×10^{-1}
	MPTAM-EM	9.05×10^{-1}	8.83×10^{-1}	8.79×10^{-1}	8.95×10^{-1}	8.95×10^{-1}	8.95×10^{-1}
	STPGM-AEM	9.05×10^{-1}	7.17×10^{-1}	7.54×10^{-1}	7.89×10^{-1}	7.89×10^{-1}	7.89×10^{-1}
	STPGM-EMA	9.67×10^{-1}	9.33×10^{-1}	9.45×10^{-1}	9.44×10^{-1}	9.44×10^{-1}	9.44×10^{-1}
	STPGM-AEMA(ours)	9.67×10^{-1}	9.33×10^{-1}	9.45×10^{-1}	9.44×10^{-1}	9.44×10^{-1}	9.44×10^{-1}
Average	LTRM	7.54×10^{-1}	7.59×10^{-1}	7.31×10^{-1}	7.51×10^{-1}	7.51×10^{-1}	7.51×10^{-1}
	PTAM-MLE	7.68×10^{-1}	7.09×10^{-1}	6.88×10^{-1}	7.43×10^{-1}	7.43×10^{-1}	7.43×10^{-1}
	MPTAM-EM	6.33×10^{-1}	6.34×10^{-1}	6.14×10^{-1}	7.15×10^{-1}	7.15×10^{-1}	7.15×10^{-1}
	STPGM-AEM	7.99×10^{-1}	7.15×10^{-1}	7.20×10^{-1}	7.81×10^{-1}	7.81×10^{-1}	7.81×10^{-1}
	STPGM-EMA	**9.81×10^{-1}**	**9.68×10^{-1}**	**9.72×10^{-1}**	**9.74×10^{-1}**	**9.74×10^{-1}**	**9.74×10^{-1}**
	STPGM-AEMA(ours)	**9.81×10^{-1}**	**9.68×10^{-1}**	**9.72×10^{-1}**	**9.74×10^{-1}**	**9.74×10^{-1}**	**9.74×10^{-1}**

[1] Bold denotes the best result, and underline denotes the second-best result. The same is below in Table 3.

The results are consistent with the hypothesis proposed in this paper that the complexity of the operating model and the station structure influence the accuracy of trajectory reconstruction. Specifically, in the D2 dataset, the destination station DS is an interchange and adopts a short-turning operation pattern during peak hours for commuting needs. Although DD is a transfer station, the operation mode of the D1 dataset is a simple mode. While the destination stations in the D3 and D4 datasets are non-transfer stations, their operation mode is also a simple mode. Besides, the passenger flow of the D3 dataset is the lowest. Therefore, the complexity of these four datasets, from high to low, is D2, D1,

D3, and D4. The performance difference of D2 may be due to the operating modes of the origin station CY during peak hours; at the same time, the complex mode of the transfer destination station DS further increases the difficulty of prediction. This comparison reinforces the view that OD's complexity of the scene directly affects the accuracy of the algorithm's prediction of passenger trajectories.

From the average value, the STPGM-AEMA and STPGM-EMA algorithms have demonstrated exceptional performance, with all metrics exceeding 0.95, showcasing significantly superior classification capabilities compared to other methods. Following them are the LTRM and STPGM-AEM algorithms, which, despite performing well in certain scenarios, exhibit relatively lower overall stability, especially when faced with unevenly distributed dataset features or significant variability. In summary, the STPGM-AEMA approach presented in this study demonstrates exceptional performance across both macro and micro metrics, emphasizing the remarkable robustness of the proposed models. This outcome accentuates the precision of the STPGM-AEMA method devised in this research in processing intricate spatiotemporal data and accurately capturing passenger behavior patterns, underscoring its utility in complex urban rail transit analyses.

5.4.2. Consistency Test Result

It is evident that STPGM-AEMA and STPGM-EMA exhibit superior performance (as presented in Table 3), with K values exceeding 0.9. This observation suggests that as the sample size tends toward infinity, the estimator's value can converge to the true parameter value. Subsequently, the LTRM algorithm demonstrates optimal performance on the D1 and D4 datasets but exhibits subpar results on the D2 dataset. It should be noted that for other algorithms, the stability of results may be significantly influenced by variations in walking distances and passenger paths at different entry stations within each dataset.

Table 3. Results of Cohen's Kappa consistency test.

Method	D1: TTYB-DD	D2: CY-DS	D3: PHY-TTYB	D4: HJL-CY	Average
LTRM	1.00×10^0	-1.45×10^{-1}	8.95×10^{-1}	1.00×10^0	6.87×10^{-1}
PTAM-MLE	7.57×10^{-1}	1.52×10^{-1}	6.61×10^{-1}	6.24×10^{-1}	5.48×10^{-1}
MPTAM-EM	7.69×10^{-1}	-1.65×10^{-1}	8.90×10^{-1}	8.30×10^{-1}	5.81×10^{-1}
STPGM-AEM	8.42×10^{-1}	4.14×10^{-1}	5.12×10^{-1}	6.18×10^{-1}	5.96×10^{-1}
STPGM-EMA	9.24×10^{-1}	1.00×10^0	1.00×10^0	9.09×10^{-1}	9.58×10^{-1}
STPGM-AEMA(ours)	9.24×10^{-1}	1.00×10^0	1.00×10^0	9.09×10^{-1}	9.58×10^{-1}

5.5. Results Interpretability Discussion

5.5.1. Potential Train Sets Feature Analysis

Figure 9 contrasts the distribution of potential train sets for typical OD pairs during peak hours (using datasets D2 and D3 as examples) with the distribution of train choices by passengers at the origin station. As shown in Figure 9(a.1–a.4) depict the distribution of potential train sets for passenger journeys entering the station in dataset D2 between 07:00 and 09:00 AM in half-hour increments, where 1–5 represent the number of train options and "P: >" indicates the statistical proportion ranking of train options. For instance, in Figure 9(a.1), "P: 2 > 3 > 1 > 4 > 5" indicates that the proportion of having two train options is the highest at 45.7%, followed by 3 (25%), 1 (14.6%), 4 (12.8%), and 5 (1.8%). Figure 9(b.1–b.4) follow a similar pattern. It is observed that during the morning peak, the statistical values for train options mostly range between 2 and 4, while during the evening peak, options of 1–2 are more prevalent, likely due to the higher frequency of train departures in the morning and relatively sparse intervals in the evening.
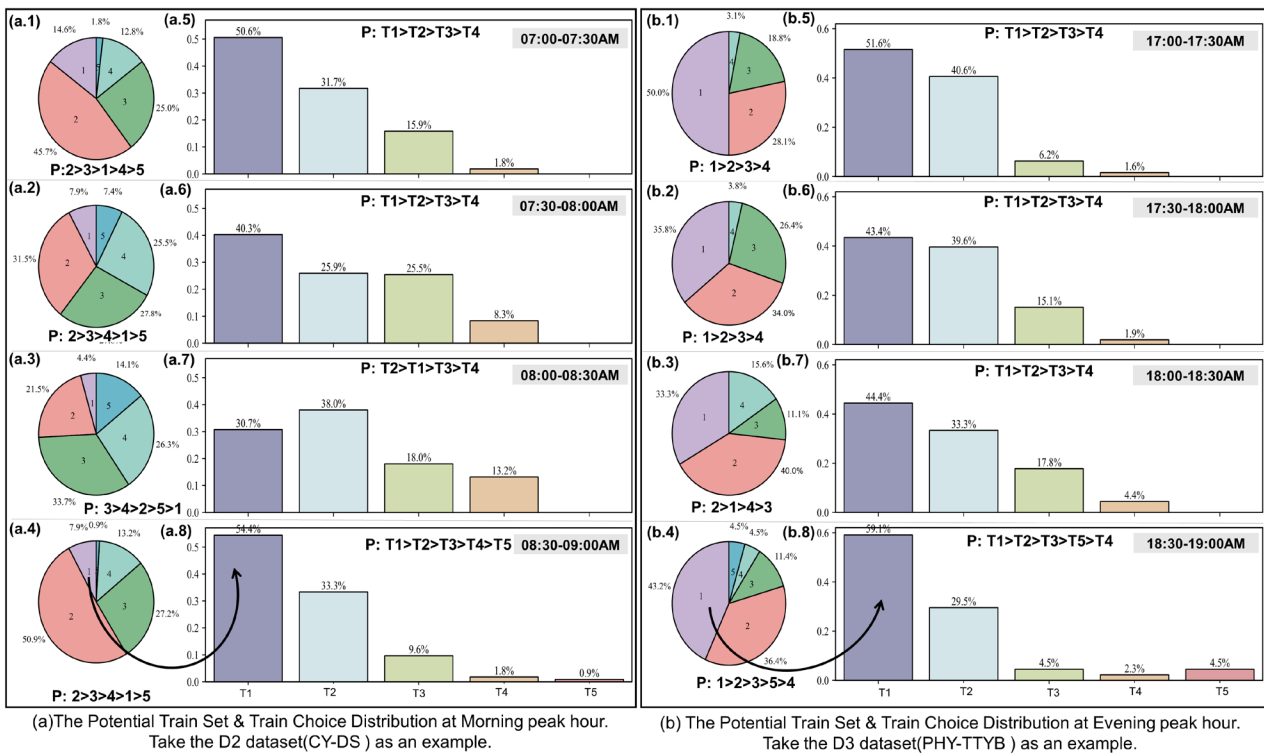
Figure 9. Comparison of typical OD on train selection and train selection distribution.

5.5.2. Analysis of Latent Variable Z Distribution

Figure 9(a.5–a.8) illustrate the distribution of train choices at the origin station for all passenger journeys during the morning peak, in half-hour intervals within dataset D2, representing the distribution of the latent variable Z. T1–T5 denotes feasible train ID, with "P: >" indicating their statistical ranking based on chosen trains. For instance, in Figure 9(a.5), "P: T1 > T2 > T3 > T4" signifies that the first train has the highest selection proportion at 50.6%, followed by T2 (31.7%), T3 (15.9%), and T4 (1.8%). A similar pattern is observed in Figure 9(b.5–b.8). It is evident that during the morning peak, there are variations in train choice probabilities across different time slots; however, a more consistent exponential distribution is apparent during the evening peak hours.

Interestingly, during the evening peak, there is a clear alignment between the ranking of potential train numbers and the sequence of chosen train ID, which is not as evident in the morning peak. For instance, between 08:30 and 09:00 a.m., despite only 7.9% of choices having one train option available, the proportion of selecting the first train reaches 54.4%. In contrast, during the evening peak, when only 43.2% of choices have one train option available, the selection proportion for the first train rises to 59.1%. This discrepancy can be attributed to the higher demand for comfort among evening peak passengers, who prioritize seating and exhibit a slower walking speed compared to morning commuters. In contrast, morning commuters prioritize quick arrival and tend to adopt a "board if possible" behavior.

5.5.3. Analysis of the Changing Process of Attention Mechanism

Figure 10a depicts the variation of value across iterations, with the horizontal axis representing the number of iterations and the vertical axis indicating changes in value μ and pdf of the latent variable t, taking the D2 data set as an example (unit in second). Figure 10b,c display PDF distributions of the initial 0 and final iteration 9, respectively.

The pdf distribution based on $t_{X_{L_I=1}}$ is represented by a blue-filled curve, while that

based on $\mathbf{t}$ is depicted with a green-filled curve. The blue vertical line represents the acquired prior value ($\mu^{(\text{prior})} = 150\text{s}$), serving as a reference for parameter variation, whereas the red vertical line indicates the current iteration round's value μ. It can be observed that the prior value exhibits left-skewness, which decreases from $u^{(0)} = 256\text{s}$ to $u^{(9)} = 189\text{s}$, after learning through UA module and stabilizes thereafter. Notably, Figures 10(b.2–b.5) and 10(c.2–c.5) demonstrate different PDF distribution shapes of vectors at various position indexes (from left to right: T1–T4), corresponding to the first and last rounds, respectively. Evidently, as iterations increase, they tend to align more closely with $u^{(9)}$.

Experimental results demonstrate that the variables of the key-value pairs $(\mathbf{K}, \mathbf{V})$ exhibit a tendency to align more closely with the matrix $\mathbf{q}$, indicating that the UA module enables interactive learning of both passenger egress time $t_{i,j}$ and destination station time distribution $\mathcal{G}(\mathbf{t};\Theta)$.
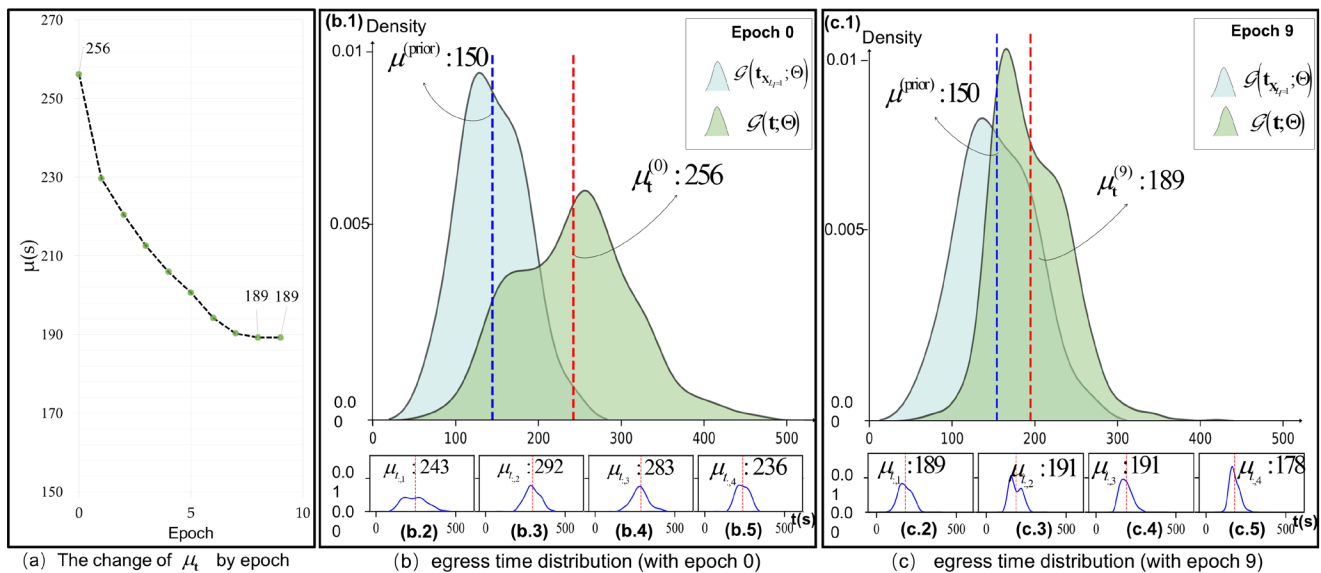

Figure 10. Q value and PDF distribution changes.

5.5.4. Individual Trajectory Visualization

The events and state values involved in the reconstructed trajectory are visualized using 3 samples of the D2 dataset, as shown in Figures 11 a and 11b, respectively. Figure 11a shows the reconstructed trajectory of individual "ID19", where the tap-in and tap-out events are known and marked as "be known" in black font. Other events are inferred and marked "be inferred" in red font. It can be seen that passenger "ID19" has already inferred that he boarded train 2 at station CY at $08:21:14$, and the interpretation of other event information is similar. Of course, if the trajectories of all passengers are displayed, the congestion and distribution of passengers waiting on the platform can be further analyzed, this is not the focus of this article. Figure 11b shows the inferred state information of each individual. It can be seen that the egress times are indeed relatively similar, which in turn confirms the effectiveness of the model STPGM proposed in this article. Furthermore, detailed error analysis is discussed in depth in the next section.
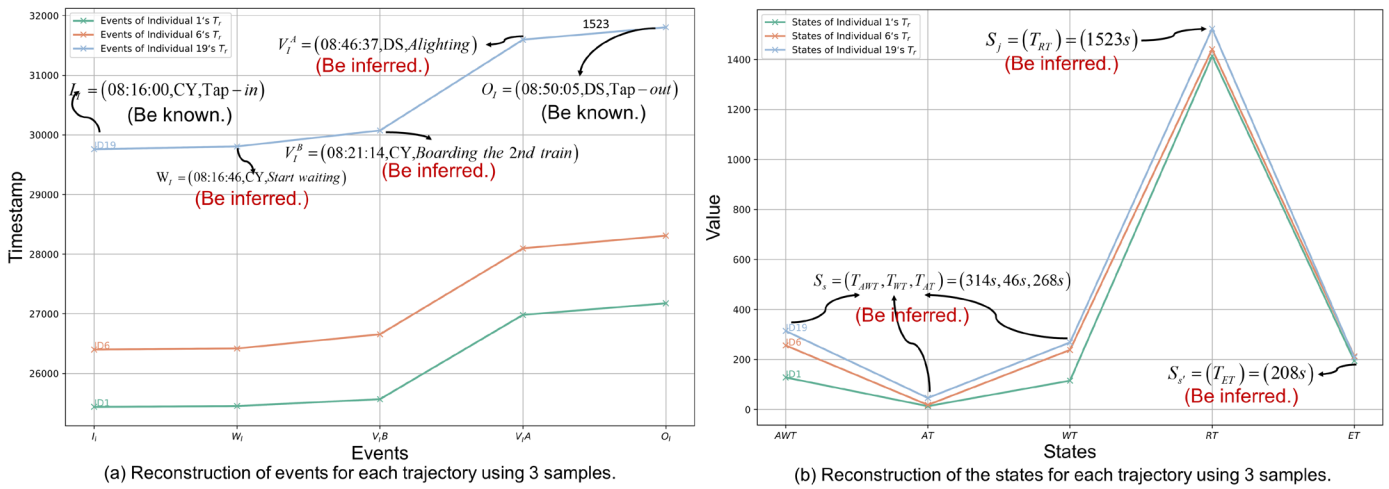
Figure 11. Individuals' reconstructed trajectories visualization.

5.5.5. Residual Analysis of Trajectory Reconstruction Fragments

The reconstruction error of the involved ITT data, encompassing temporal attributes T_{E_h} and state characteristics S_f of the event, is assessed based on residuals $err(\bullet) = tr_I - \widehat{tr_I}$ (unit in second). Further, $err(\bullet) > 0$ indicates predicted occurrence times earlier than the actual events, while positive residuals suggest later predictions. Regarding state values, $err(\bullet) > 0$ denote underpredictions, whereas positive values indicate overpredictions.

Figure 12 employs Q-Q plots to demonstrate the normality of each event-time variable. The horizontal axis represents theoretical quantiles of the probability distribution, and the vertical axis reflects percentiles of residual values. The red line represents the regression line satisfying either $T_{E_h} = \widehat{T_{E_h}}$ (in Figure 12a–e) or $S_f = \widehat{S_f}$ (in Figure 12f–j). Two grey dashed lines indicate a 95% confidence interval, with individual residual $err(\bullet)$ denoted by points on the plot.

The analysis in Figure 12 reveals that, when considering the deviation from events, it is evident that, apart from $err(T_{I_I})$ being influenced by system errors, deviations in $T_{V_I^B}$ and $T_{V_I^A}$ occur due to disregarding the time spent onboarding and alighting. To enhance the model, future research can incorporate a deviation correction coefficient. Notably, the largest deviation value about T_{W_I} primarily stems from insufficient observational information and significant randomness. Regarding state values, aside from errors $err(T_{RT})$ resulting from individual heterogeneity's perception bias, most values are predominantly positive due to their association with calculation methods. It is noteworthy that the error value $err(T_{ET})$ is minimal, thus confirming the efficacy of our proposed method. Furthermore, except for $err(T_{I_I})$, a majority of data points fall within an interval range while exhibiting residuals close to normality, validating the effectiveness of the ITR method proposed in this paper. The Q-Q plots effectively visualize the normality or deviation of residuals, offering a statistical basis to assess the model's performance in reconstructing travel event timelines.

$err(\bullet), (s)$
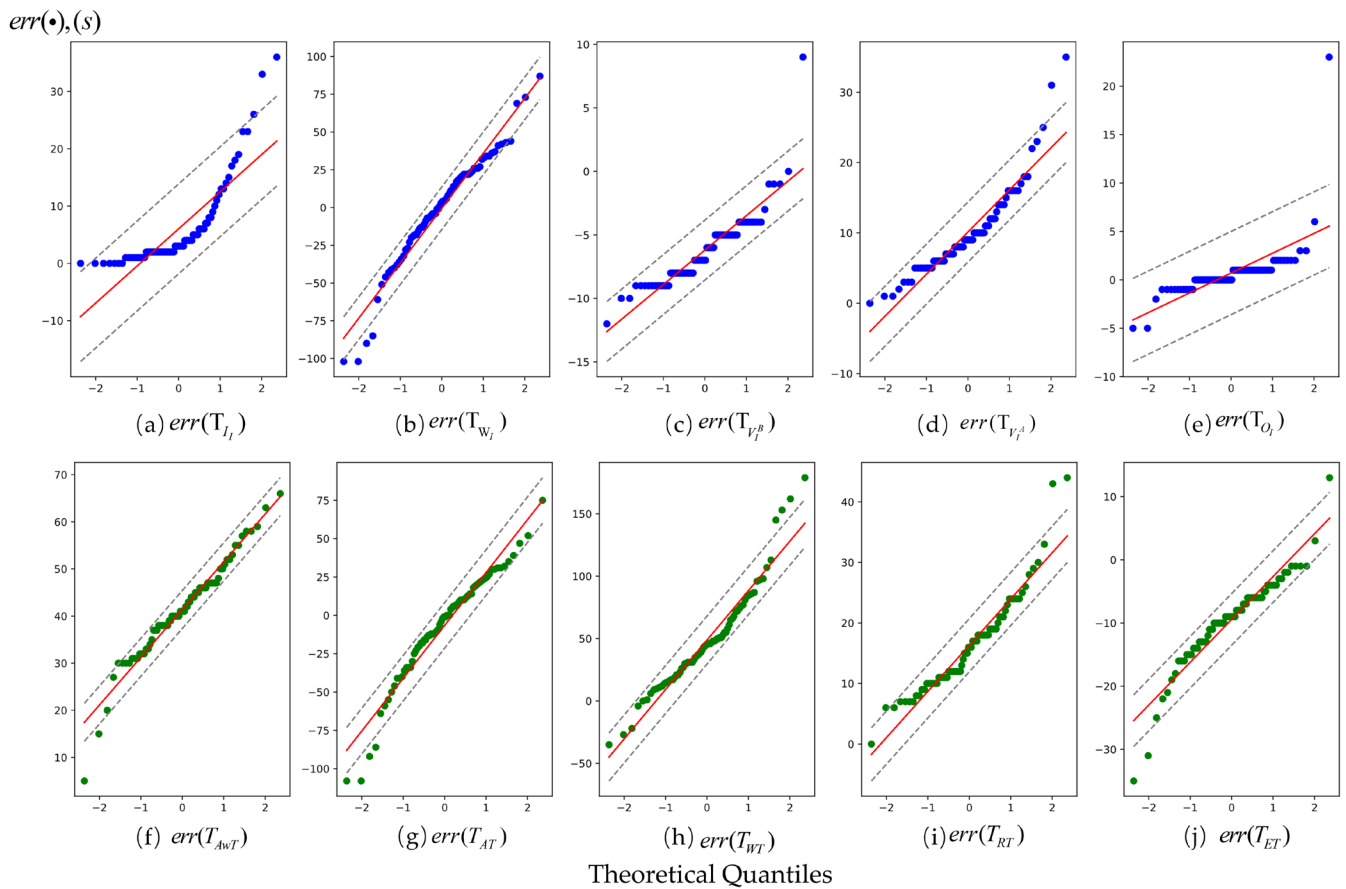
Figure 12. Event and state residuals Q-Q plots of trajectory fragments.

6. Conclusions

In this paper, an automatic inference method for ITR has been proposed, namely STPGM-AEMA, which aims to infer missing information from incomplete information. The method effectively recovers rich semantic and state information about each individual trajectory using only AFC and AVL data. A GOLN rule is introduced in the model as a bridge from observed data to inferred information. On this basis, an information interaction representation module for global and local latent variables was designed, which effectively promotes autonomous communication of information between individuals and the system, eliminating dependence on manual survey data. Secondly, the proposed parameter learning algorithm AEMA enhances the EM algorithm by adaptively introducing a priori parameters and a key-value attention mechanism. It not only improves the stability and convergence speed of parameters but also automatically samples the walking time of individual and egress time distributions to deal with missing data problems. In addition, combined with ITT data, three methods and two ablation experimental methods were comparatively analyzed. The results show that the proposed STPGM-AEMA method performs well in terms of accuracy and robustness, and the accuracy can reach 0.95 (95%), which is at least 15% more accurate than the traditional methods (i.e., PTAM-MLE and MPTAM-EM).

It is worth noting that interpretability analysis was performed on key parts of the STPGM-AEMA method, including potential set feature mining analysis, latent variable distribution analysis, the role of the attention mechanism, and temporal residual analysis. On this basis, some possible directions for improvement could be as follows: (1) addressing the limitations of the proposed model in estimating individual trajectories between OD pairs with insufficient data, as any lack of prior information will adversely affect the

utility of the UB module; (2) Currently, a simple normalization function is used in the UA module. In future research, the application of other activation functions (Leaky ReLU, weighted Softmax, etc.) in multi-class imbalance problems can be explored to enhance model fitting capabilities; (3) Extend the model formulation to include route choice probability, passenger type, station type, or operation strategies as additional model parameters; and (4) Although the AEMA algorithm proposed in this paper employs offline training, the average training time for a single dataset in this study is approximately 15.79 s, which adequately satisfies the requirements for fast trajectory reconstruction of complete samples within a single OD pair. Future work can explore the possibility of integrating real-time sample generation and correction modules to achieve real-time personal travel trajectory prediction. Certainly, this requires significant extensions to existing models.

Author Contributions: Conceptualization, X.S. and Y.Q.; Methodology, X.S., J.G., Y.Q. and S.X.; Software, X.S.; Validation, X.Z. and J.H.; Formal analysis, X.S., J.G. and S.X.; Investigation, X.Z. and J.H.; Data curation, Q.S.; Writing—original draft, X.S.; Writing—review and editing, J.G. and J.H.; Supervision, Q.S. and L.J.; Project administration, L.J.; Funding acquisition, X.Z. and S.X. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Xuanchuan Zheng was employed by the company Beijing Urban Construction Design & Development Group Co., Ltd. Author Qi Sun was employed by the company Beijing Metro Network Administration Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A. Notations and Definitions

Symbol	Definition
i	Passenger index.
s, s'	The origin and destination stations of an individual trip, respectively.
t, t'	The tap-in and tap-out time of an individual trip index, respectively.
$I^i_{(t,s)\to(t',s')}$	Itinerary index, where $(t,s)\to(t',s')$ indicates that passenger i enter the origin station s at t and leave the destination station s' at t', abbreviated as I.
$I_{s,t}$, $I_{s',t'}$	The tap-in and tap-out time of an individual itinerary.
tj	Train index of all trains between OD pairs.
$tj_{I,id}$	A unique symbol representing the ID of train in journal I.
$tj_{s,at}$, $tj_{s,dt}$, $tj_{s',at}$, $tj_{s',dt}$	The arrival time and departure time at origin station s/destination station s' of the train tj, respectively.
$tj_{fs,q}$, $tj_{s,q}$, $tj_{s',q}$, $tj_{es,q}$	The order number of the train tj at first station fs, station s, station s' and last station es.
$J_{I^{(t,s)}}$, $J_{I^{(t',s')}}$	The set of feasible train choices at origin station s/destination station s' for an itinerary I, respectively.
J_I	The set of feasible train choices for an itinerary I, denoted as $\{1,2,\cdots,j\}$, with the index being j, L_I represents the ordered sequence of train options available. The total length of this sequence is, with the dimension being $1\times L_I$.

$T_{i^{l(s')}}^{\text{eg}}$	The potential set of egress time for an itinerary I, denoted as $\left\{t_{i,1}, t_{i,2}, \cdots, t_{i,j}\right\}$, abbreviated as T_I^{eg}.
$t_{i,j}$	The jth potential egress time value for passenger i in itinerary I.

Appendix B. The Recorded Data from the Trajectory Simulation Experiment

Date	2023/3/21			
PID	15****36			
Itinerary index	234			
L1	**S1**	**Tap-in Time**	**To platform**	**Boarding Time**	**Train Departure Time**
6	623	17:34:50	17:36:55	17:38:40	17:38:50
L2	**S2**	**Train Arrival Time**	**Alighting Time**	**Tap-out time**	
6	633	17:58:56	17:59:05	17:59:59	

Appendix C. The Table Presents the States of the Trajectory Calculation Results

Date	2023/3/21				
PID	15****36				
Itinerary index	234				
OD pair	**Access time(s)**	**Train ID**	**Waiting time(s)**	**Riding time(s)**	**Egress time(s)**
623-633	125	1222	105	1206	54

References

1. Beijing Transport Institute. *Beijing Transport Development Annual Report*; Beijing Transport Institute: Beijing, China, 2023.
2. Zhu, Y.; Koutsopoulos, H.N.; Wilson, N.H.M. Passenger itinerary inference model for congested urban rail networks. *Transp. Res. Part C Emerg. Technol.* **2021**, *123*, 102896. https://doi.org/10.1016/j.trc.2020.102896.
3. Akhtar, M.; Moridpour, S.; Bazant, M. A Review of Traffic Congestion Prediction Using Artificial Intelligence. *J. Adv. Transp.* **2021**, *2021*, 8878011. https://doi.org/10.1155/2021/8878011.
4. Peftitsi, S.; Jenelius, E.; Cats, O. Modeling the effect of real-time crowding information (RTCI) on passenger distribution in trains. *Transp. Res. Part A Policy Pract.* **2022**, *166*, 354–368. https://doi.org/10.1016/j.tra.2022.10.011.
5. Mo, B.; Zhao, Z.; Koutsopoulos, H.N.; Zhao, J. Individual Mobility Prediction in Mass Transit Systems Using Smart Card Data: An Interpretable Activity-Based Hidden Markov Approach. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 12014–12026. https://doi.org/10.1109/tits.2021.3109428.
6. Zhou, W.; Han, B.; Feng, C. A review of passenger flow assignment model and algorithm for urban rail transit network. *Syst. Eng. Theory Pract.* **2017**, *37*, 440–451.
7. Su, G.H.; Si, B.F.; Zhao, F.; Li, H. Data-Driven Method for Passenger Path Choice Inference in Congested Subway Network. *Complexity* **2022**, *2022*, 5451017. https://doi.org/10.1155/2022/5451017.
8. Mai, T.; Bui, T.V.; Nguyen, Q.P.; Le, T.V. Estimation of recursive route choice models with incomplete trip observations. *Transp. Res. Part B Methodol.* **2023**, *173*, 313–331. https://doi.org/10.1016/j.trb.2023.05.004.
9. Wu, L.; Yang, L.; Huang, Z.; Wang, Y.; Chai, Y.; Peng, X.; Liu, Y. Inferring demographics from human trajectories and geographical context. *Comput. Environ. Urban Syst.* **2019**, *77*, 101368. https://doi.org/10.1016/j.compenvurbsys.2019.101368.
10. Zhao, J.J.; Zhang, L.T.; Ye, K.J.; Ye, J.X.; Zhang, J.; Zhang, F.; Xu, C.Z. GLTC: A Metro Passenger Identification Method across AFC Data and Sparse WiFi Data. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 18337–18351. https://doi.org/10.1109/tits.2022.3171332.
11. Li, C.; Xiong, S.; Xiong, H.; Sun, X.; Qin, Y. Logistic model for pattern inference of subway passenger flows based on fare collection and vehicle location data. *Appl. Math. Model.* **2024**, *130*, 472–495. https://doi.org/10.1016/j.apm.2024.03.013.
12. Xiong, S.; Li, C.; Sun, X.; Qin, Y.; Wu, C.F.J. Statistical estimation in passenger-to-train assignment models based on automated data. *Appl. Stoch. Models Bus. Ind.* **2021**, *38*, 287–307. https://doi.org/10.1002/asmb.2660.
13. Cheng, J.; Zhang, X.; Luo, P.; Huang, J.; Huang, J. An unsupervised approach for semantic place annotation of trajectories based on the prior probability. *Inf. Sci.* **2022**, *607*, 1311–1327. https://doi.org/10.1016/j.ins.2022.06.034.
14. Wang, J.Y.; Wu, N.; Lu, X.X.; Zhao, W.X.; Feng, K. Deep Trajectory Recovery with Fine-Grained Calibration using Kalman Filter. *IEEE Trans. Knowl. Data Eng.* **2021**, *33*, 921–934. https://doi.org/10.1109/tkde.2019.2940950.
15. Zhang, F.; Zhao, J.; Tian, C.; Xu, C.; Liu, X.; Rao, L. Spatiotemporal Segmentation of Metro Trips Using Smart Card Data. *IEEE Trans. Veh. Technol.* **2016**, *65*, 1137–1149. https://doi.org/10.1109/tvt.2015.2409815.

16. Wang, Y.; Zhang, Z.; Wei, H.; Yin, G.; Huang, H.; Li, B.; Huang, C. A Novel Fault-Tolerant Scheme for Multi-Model Ensemble Estimation of Tire Road Friction Coefficient with Missing Measurements. *IEEE Trans. Intell. Veh.* **2024**, *9*, 1066–1078. https://doi.org/10.1109/tiv.2023.3336048.

17. Zhao, J.J.; Qu, Q.; Zhang, F.; Xu, C.Z.; Liu, S.Y. Spatio-Temporal Analysis of Passenger Travel Patterns in Massive Smart Card Data. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3135–3146. https://doi.org/10.1109/tits.2017.2679179.

18. Cheng, Z.; Trépanier, M.; Sun, L. Probabilistic model for destination inference and travel pattern mining from smart card data. *Transportation* **2020**, *48*, 2035–2053. https://doi.org/10.1007/s11116-020-10120-0.

19. Shang, P.; Li, R.; Guo, J.; Xian, K.; Zhou, X. Integrating Lagrangian and Eulerian observations for passenger flow state estimation in an urban rail transit network: A space-time-state hyper network-based assignment approach. *Transp. Res. Part B Methodol.* **2019**, *121*, 135–167. https://doi.org/10.1016/j.trb.2018.12.015.

20. Zhu, Y.; Koutsopoulos, H.N.; Wilson, N.H.M. A probabilistic Passenger-to-Train Assignment Model based on automated data. *Transp. Res. Part B Methodol.* **2017**, *104*, 522–542. https://doi.org/10.1016/j.trb.2017.04.012.

21. Sun, L.J.; Lu, Y.; Jin, J.G.; Lee, D.H.; Axhausen, K.W. An integrated Bayesian approach for passenger flow assignment in metro networks. *Transp. Res. Part C Emerg. Technol.* **2015**, *52*, 116–131. https://doi.org/10.1016/j.trc.2015.01.001.

22. Zhou, F.; Xu, R.H. Model of Passenger Flow Assignment for Urban Rail Transit Based on Entry and Exit Time Constraints. *Transp. Res. Rec.* **2012**, *2284*, 57–61. https://doi.org/10.3141/2284-07.

23. Rahbar, M.; Hickman, M.; Mesbah, M.; Tavassoli, A. Calibrating a Bayesian Transit Assignment Model Using Smart Card Data. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 1574–1583. https://doi.org/10.1109/tits.2018.2852726.

24. Yu, L.; Liu, H.; Fang, Z.; Ye, R.; Huang, Z.; You, Y. A new approach on passenger flow assignment with multi-connected agents. *Phys. A Stat. Mech. Its Appl.* **2023**, *628*, 129175. https://doi.org/10.1016/j.physa.2023.129175.

25. Kusakabe, T.; Iryo, T.; Asakura, Y. Estimation method for railway passengers' train choice behavior with smart card transaction data. *Transportation* **2010**, *37*, 731–749. https://doi.org/10.1007/s11116-010-9290-0.

26. Sun, L.; Lee, D.-H.; Erath, A.; Huang, X. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of MRT system. In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, China, 12 August 2012; pp. 142–148.

27. Zhang, Y.-S.; Yao, E.-J. Splitting Travel Time Based on AFC Data: Estimating Walking, Waiting, Transfer, and In-Vehicle Travel Times in Metro System. *Discret. Dyn. Nat. Soc.* **2015**, *2015*, 539756. https://doi.org/10.1155/2015/539756.

28. Zhang, T.; He, W.; Huang, J.; He, Z.; Li, J. Interactive visual analytics of moving passenger flocks using massive smart card data. *Cartogr. Geogr. Inf. Sci.* **2022**, *49*, 354–369. https://doi.org/10.1080/15230406.2022.2039775.

29. Lin, M.; Huang, Z.; Zhao, T.; Zhang, Y.; Wei, H. Spatiotemporal Evolution of Travel Pattern Using Smart Card Data. *Sustainability* **2022**, *14*, 9564. https://doi.org/10.3390/su14159564.

30. Cats, O.; Ferranti, F. Unravelling individual mobility temporal patterns using longitudinal smart card data. *Res. Transp. Bus. Manag.* **2022**, *43*, 100816. https://doi.org/10.1016/j.rtbm.2022.100816.

31. Goulet-Langlois, G.; Koutsopoulos, H.N.; Zhao, Z.; Zhao, J. Measuring Regularity of Individual Travel Patterns. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1583–1592. https://doi.org/10.1109/tits.2017.2728704.

32. Wu, P.; Xu, L.; Li, J.; Guo, H.; Huang, Z. Recognizing Real-Time Transfer Patterns between Metro and Bus Systems Based on Spatial–Temporal Constraints. *J. Transp. Eng. Part A Syst.* **2022**, *148*, 04022065. https://doi.org/10.1061/jtepbs.0000721.

33. Zhang, Y.; Yao, E.; Wei, H.; Zheng, K. A constrained multinomial Probit route choice model in the metro network: Formulation, estimation and application. *PLoS ONE* **2017**, *12*, e0178789. https://doi.org/10.1371/journal.pone.0178789.

34. Hussain, E.; Bhaskar, A.; Chung, E. Transit OD matrix estimation using smartcard data: Recent developments and future research challenges. *Transp. Res. Part C Emerg. Technol.* **2021**, *125*, 103044. https://doi.org/10.1016/j.trc.2021.103044.

35. Yu, C.; Li, H.; Xu, X.; Liu, J. Data-driven approach for solving the route choice problem with traveling backward behavior in congested metro systems. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *142*, 102037. https://doi.org/10.1016/j.tre.2020.102037.

36. Wu, J.J.; Qu, Y.C.; Sun, H.J.; Yin, H.D.; Yan, X.Y.; Zhao, J.D. Data-driven model for passenger route choice in urban metro network. *Phys. A Stat. Mech. Its Appl.* **2019**, *524*, 787–798. https://doi.org/10.1016/j.physa.2019.04.231.

37. Raveau, S.; Muñoz, J.C.; de Grange, L. A topological route choice model for metro. *Transp. Res. Part A Policy Pract.* **2011**, *45*, 138–147. https://doi.org/10.1016/j.tra.2010.12.004.

38. Raveau, S.; Guo, Z.; Muñoz, J.C.; Wilson, N.H.M. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. *Transp. Res. Part A Policy Pract.* **2014**, *66*, 185–195. https://doi.org/10.1016/j.tra.2014.05.010.

39. Fu, Q. Modelling Route Choice Behaviour with Incomplete Data: An Application to the London Underground. Ph.D. Thesis, The University of Leeds, Leeds, UK, 2014.

40. Chen, X.; Cheng, Z.; Jin, J.G.; Trépanier, M.; Sun, L. Probabilistic Forecasting of Bus Travel Time with a Bayesian Gaussian Mixture Model. *Transp. Sci.* **2023**, *57*, 1516–1535. https://doi.org/10.1287/trsc.2022.0214.

41. Qu, H.; Xu, X.; Chien, S. Estimating Wait Time and Passenger Load in a Saturated Metro Network: A Data-Driven Approach. *J. Adv. Transp.* **2020**, *2020*, 4271871. https://doi.org/10.1155/2020/4271871.

42. Zhu, Y.; Koutsopoulos, H.N.; Wilson, N.H.M. Inferring left behind passengers in congested metro systems from automated data. *Transp. Res. Part C Emerg. Technol.* **2018**, *94*, 323–337. https://doi.org/10.1016/j.trc.2017.10.002.

43. Tuncel, K.S.; Koutsopoulos, H.N.; Ma, Z. An Unsupervised Learning Approach for Robust Denied Boarding Probability Estimation Using Smart Card and Operation Data in Urban Railways. *IEEE Intell. Transp. Syst. Mag.* **2023**, *15*, 19–32. https://doi.org/10.1109/mits.2023.3289969.
44. Li, X.; Zhong, Z.; Wu, J.; Yang, Y.; Lin, Z.; Liu, H. Expectation-Maximization Attention Networks for Semantic Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
45. Luo, Q.; Lin, B.; Lyu, Y.; He, Y.; Zhang, X.; Zhang, Z. Spatiotemporal path inference model for urban rail transit passengers based on travel time data. *IET Intell. Transp. Syst.* **2023**, *17*, 1395–1414. https://doi.org/10.1049/itr2.12332.