

Research Article

On Individual Repositioning Distance along Platform during Train Waiting

Fabien Leurent and Xiaoyan Xie 

LVMT, UMR-T 9403, Ecole des Ponts, IFSTTAR, UPEM, UPE, 77455, Champs-sur-Marne, France

Correspondence should be addressed to Xiaoyan Xie; xiaoyan.xie@enpc.fr

Received 24 November 2017; Revised 17 April 2018; Accepted 3 May 2018; Published 21 June 2018

Academic Editor: Taku Fujiyama

Copyright © 2018 Fabien Leurent and Xiaoyan Xie. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Out of waiting times spent in rail stations on boarding platforms, some part can be reinvested by the trip-makers to optimize their positions of boarding and save on travel time for the rest of their trips. This paper provides a stochastic model, in which user's journey is decomposed into phases of, successively, walking in the access station, platform positioning, waiting for boarding, train riding, and walking in the egress station. Walking speed and target position are modeled as individual factors, and in-station distances as random variables. Service timetable is exogenous. This makes egress times and exit instants random variables that are characterized by distribution and mass probability functions under closed-forms, for both single and distributed walking speeds. Specific statistical distributions are shown to ease computation. The resulting PDF formulae make likelihood functions of the model parameters. Maximum likelihood estimation is proposed and applied to a case study of commuter rail line in Paris: journeys between stations Vincennes and La Défense along line A of the Regional Express Railways. Based on data from Automated Fare Collection and Automatic Vehicle Location systems and pertaining to an individual user, satisfactory results were obtained.

1. Introduction

Passenger waiting time is one of the most crucial factors for transit design and planning, related to passenger satisfaction and to measure public transport (PT) quality of service [1–4]. However, it was significant at the trip level in urban PT, and the ratio of passenger actual waiting time to actual journey time was between 10 and 30% [5–7]. Furthermore, waiting time is most often spent at standing, often so in a crowded place. That makes passenger less comfortable. In the literature, social sciences tried to reduce perceived or actual waiting time, or emotional and psychological costs, whereas operations research investigated waiting time reduction strategies [8–12], as waiting time could be considered scarce, economic resource and unavoidable [13, 14]. In order to maximize time utility, some passengers allocate waiting time to activities/preoccupations. However, beyond the waiting time investigations, existing research does not sufficiently explore how to reuse the actual waiting time to perform an efficient travel, especially in high-frequency urban rail PT

system. That has received only some attention in cognitive aspect [10, 15–18].

Despite the attention paid so far to individual waiting time reuse in PT station, a related issue seemed to remain unexplored: the reuse of waiting time for passenger repositioning along the boarding platform. Of course, this is of interest for railway-based transit submodes only, since the platform lengths of these modes extend from some dozen meters up to some hundred meters, e.g., up to 200 m in Paris. Assuming that individual walking speeds range from 0.8 to 1.7 m/s, repositioning time from one platform end to another may reach 0.5 to 3.5 min, indeed a valuable gain for passengers in their daily schedules. The influence of waiting time on individual walk on boarding platform can be significant. Such extent makes the “longitudinal” positioning of individual passengers along their trains a matter of significance for their journeys, since it involves walking times along the platform at both access and egress stations, together with waiting time and comfort on the boarding platform as well as in-vehicle in relation to the distribution of passengers along the train. This

pushed frequent passengers (e.g., commuters) to optimize their longitudinal position along the boarding platform with respect to their egress station [19], and also “travel assistant” applications (such as CityMapper, RATP, Paris ci la sortie du Métro) to deliver positioning advice to every passengers, including occasional ones (e.g., tourists).

Railway operators, on their side, consider the longitudinal positioning of candidate riders in relation to train dwelling times: higher passenger densities at some spots along the platform will not only require more boarding time but also slow down the egress of alighting passengers. Dynamic information via travel assistants or specific signage such as with variable color panels has recently been implemented on an experimental basis [20]. From a recent survey in the Paris area of passengers boarding in a suburban train station about their eagerness to reposition themselves in relation to platform crowding, it was shown that the acceptable shift amounts to 2.4 cars on average, about 50 m [19].

The objective of this paper is to estimate passenger longitudinal repositioning distance on boarding platform during train waiting and underlying distributions of individual walking speed and distances. Building upon our previous stochastic models [21–23] based on smartcard data, a stochastic modeling approach about passenger individual repositioning distance with posterior analysis is proposed in this paper, by extending our primary model about passenger longitudinal repositioning [24].

Concerning the proposed passenger flow stochastic model, over the last decade, a new branch of knowledge was emerged, the modern data measurement and data-driven statistical analysis in PT field. Modern data, Automated Fare Collection (AFC) data and Automatic Vehicle Location (AVL) data, among others have become available either for PT service evaluation and improvement, or for passenger flow analysis. Passenger flow study has given rise to the stochastic modeling and statistical estimation of fine individual passenger travel phenomena by trip leg or in station in rail transit system, a closed black-box. Several relevant studies are reviewed in greater details as below; for other applications about AFC data see [25, 26].

The stochastic features of passenger journeys were modeled by [27], who decomposed every journey into separate stages, walking-in, wait, walking-out, and transfer between Origin-Destination (O-D) pairs. A regression model was proposed to generate distributions of walking-in and walking-out times in access and egress stations (gamma distributions) and distributions of waiting times at access and transfer stations (uniform distributions) including the additional waiting time caused by fail-to-board. Based on Poisson arrival assumption, a mixed-weight inference of passenger path choice behavior, gate-to-gate journey time distribution, and crowding metrics on platform were derived from AFC data between two O-D pairs in Beijing. Based on this study, Sun and Schonfeld [28] further developed a schedule-based assignment model by extending the fail-to-board (or left behind) model of [29], and then the transfer penalty. Since simplified assumptions were taken, that constrained the model application.

A key element in the transit model is the probability of matching between a given train run serving the boarding station and a given passenger journey. To generalize the previous works in [27, 28], Zhu et al. [30–33] devised a probabilistic individual Passenger-to-Train Assignment Model (PTAM), in which the matching probability between train runs and individual journeys was emphasized, for rail transit service estimation. A regression model of passenger walking speeds was considered. The PTAM was developed so as to provide estimators of the main state variables of passenger flows both in-station (waiting loads) and in-vehicle (loading snakes) times as well as in passenger in access and egress stations. Within this wide scope, the issue of passenger positioning was mentioned: Zhu [30] suggested modeling it as a Random Variable (RV), yet apparently with no further development. The model was applied on a five-station metro line segment using AFC data and AVL data of HongKong and validated by synthetic data. The authors proposed estimation for several factors, the distributions of passenger walking-in and walking-out speeds by trip leg, the distribution of fail-to-board (left behind) on boarding platform (the number of trains/times a passenger unable to board due to capacity constraint), and in-station crowding level. Nevertheless, this study considered explicitly only distributions of walking-in and walking-out speeds (as normal distributions) and a two-itinerary choice scenario without transfer. To integrate transfer in crowding discomfort analysis, Hörcher et al. [34] combined the probabilistic assignment model of [31–33] with the standing and sitting choice models [35, 36] and applied to a PT network of HongKong. A group of transfers and their choices are modeled in a Bayesian framework at the journey level to model complex trips that could involve transfers hence combine legs (journeys), in order to study route choice between station pairs and reveal the influence of in-vehicle congestion on route disutility to passengers.

In the meanwhile, in more recent studies, econometric methods were also applied for modeling and estimating individual passenger in-station underlying statistical distributions. Inspired from the econometric approach using highway individual vehicle toll data in [37, 38], Leurent and Xie [21] modeled explicitly constant or uniform-distributed passenger walking speed with shifted exponential-distributed distances by trip leg. Individual passenger movement was integrated to train run. Parameters were estimated by the Maximum Likelihood Estimation (MLE) by using AFC data and timetable of the line RER A in Paris area. These two models in [21] were further developed by adding a shifted exponential-distributed waiting time in [22], which made computation tractable and yielded consistent estimates with formers. Xie and Leurent [23] demonstrated that normal-distributed walking speed reduced likelihood function computation complexity. Our stochastic models and related MLE were much analogous to those in [30–33]. Their independent development led to respective assets: the authors had a wide scope of traffic variables and considers fail-to-board an arriving train, which we did not take into account. Nevertheless, on our side, there was a linkage between the tap-in and tap-out (TITO) times based on the individual walking speed,

in addition to the explicit consideration of individual positioning. Moreover, the consideration of passenger individual distances in the model resolved the conservative assumption that the minimum access and egress time were zero (infinite passenger speed) in [27, 28, 31–33]. Simultaneously, [39] provided a stochastic frontier model by extending the model in [40] for estimating Erlang-distributed waiting time and service reliability. Walking times and in-vehicle travel time were normal distributions. Parameters were also estimated by the maximum likelihood method using AFC data and timetable of the Piccadilly line of London's underground network.

Since the study on passenger longitudinal repositioning distance along boarding platform during train waiting was scarce, passenger longitudinal repositioning distance was primarily modeled in our most recent work [24]. By extending this model, a novel approach about passenger individual repositioning distance with posterior analysis is proposed in this paper. To simplify computation complexity and save model optimization time, the passenger longitudinal repositioning distance is modeled explicitly, but implicit for waiting times. It brings about an explicit model that is both physical and probabilistic: longitudinal repositioning distance is modeled in relation to walking distances on the boarding and alighting platforms and to the individual passenger walking speed and the time available to the passenger for walking along the platform up to the train arrival hence prior to boarding it. The influence of longitudinal position is traced out on the passenger journey time between selected points in the access and egress stations, such as tap gates. Analytical formulae are obtained for related Probability Density Functions (PDFs) of tap-out times conditioning on tap-in time and walking speed, or on tap-in time only. They are analytical closed-form PDFs that provide likelihood functions, which are tractable by MLE under a simple specification of the basic statistical distributions. To accomplish that, computation and optimization are detailed. Posterior analysis imitated Bayesian analysis is applied. Train occurrence and passenger trips are reanalyzed.

This paper is organized as follows. Section 2 provides the physical and stochastic model in a bottom up way from elementary bricks by journey phase to individual journey time. The analytical formulae of the probability to take a given train, the distribution of repositioning distances and the distributions of egress times are built. In Section 3, particular statistical distributions are selected as inputs in order to obtain closed-form formulae for the PDFs of interest, thus easing the computation of the model outcomes. Section 4 puts forward the statistical estimation scheme. MLE with optimization is involved on the basis of AFC and AVL datasets in order to estimate the parameters of the statistical distributions. The resulting estimates can then be used for ex-post analysis of individual journeys so as to infer their most likely features. A real case study on the commuter line RER A in Paris area, France, is provided in Section 5. Estimation results are provided for boarding positions, walking speeds and in-station walking distances, and applied to posterior analysis for inferring passenger individual journey details. Lastly, Section 6 assesses the outreach and limitations

of our approach and points to directions for further research.

2. Physical and Stochastic Models of Passenger Repositioning along the Platform

Stochastic models on passenger repositioning between a pair of O-D stations (called also a trip leg) are depicted in this section. This study is built upon the time-space diagram of traffic flow theory, the kinematic theory, and basic econometric theory to understand passenger individual in-station movement related to train run choice by trip leg along an urban rail transit line.

We first state the physical model of one individual passenger making a train journey by availing oneself of train runs and waiting time prior to boarding (Section 2.1). The in-station walking distances are salient features, which are modeled as RVs (Section 2.2). This makes the passenger access to train runs a stochastic process. We derive the probability distributions of passenger-to-train assignment, of the walking (positioning) distance along boarding platform, and of the residual waiting time up to boarding. The egress phase is also a stochastic process, in which the distance that remains to walk along the alighting platform is a RV, the total walking-out time from train alighting to tap-out gate and the tap-out instant from the egress station as well. Lastly, we model the walking speed as a RV and derive the consequences on access and egress processes (Section 2.3). The main notations used in this study are introduced in Table 1.

2.1. Physical Model of a Passenger Journey. Consider here a noncyclic urban rail transit line with J stations. There are several itineraries between station platform and tap gates with different kinds of pedestrian facilities (flat road, stair, escalator, lift, etc.).

Let us consider a transit passenger (user), denoted by index u , with individual walking speed v that is a “cruising speed.” The passenger makes a simple journey by train along a line from access station a to egress station e . The instants of passage at selected points in each station at TITO gates are denoted as h_u^+ and h_u^- (Figure 1), respectively, and called TITO instants.

In access station, the walking-in distance from tap-in gate O to boarding platform entrance A is denoted as s^+ , whereas the walking distance is denoted as s on boarding platform from A to boarding (also waiting) point M. In egress station, the walking-out distance from alighting platform exit E to tap-out gate D is denoted as s^- , whereas the walking distance is denoted as s' on alighting platform from alighting points N to E. Passenger walking paths in stations are called “walking links,” green arrows in Figure 1. The total walking-in and total walking-out times in access and egress stations are $t^+ = t^+ + s/v$ and $t^- = t^- + s/v$, respectively, where $t^+ = s^+/v$ and $t^- = s^-/v$. Passenger's journey time on this trip leg is equal to $h_u^- - h_u^+$.

Let us now consider the longitudinal dimension of platforms. In urban rail transit system, station platforms are long objects of relatively modest width (e.g., several meters),

TABLE 1: Notations.

Notations	Descriptions
Passenger	
u and U	Passenger number and passenger set by station leg
v	Passengers' walking speeds on walking links in access and egress stations
L	Passenger longitudinal repositioning distance
h_u^+ and h_u^-	Passenger u 's tap-in instant at a access gate and tap-out instant at a egress gate on a trip leg
s^+ and t^+	Passenger walking-in distance and time from tap-in gate to platform access point in access station
s	Passenger walking-in distance from platform entrance to platform boarding point in access station
s'	Passenger walking-out distance from platform alighting point to platform exit in egress station
s^- and t^-	Passenger walking-out distance and time from platform egress point to tap-out gate in egress station
t^a and t^e	Total walking-in and total walking-out times in access and egress stations
w_u^i	Passenger apparent waiting time
\tilde{w}_u^i	Passenger residual waiting time on waiting (also boarding) point, $\tilde{w}_u^i = w_u^i - s/v$
$\Delta h_{u,i}^+$	Passenger total time cost in access station for boarding train run i , $\Delta h_{u,i}^+ = t_a + \tilde{w}_u^i$
Train	
i and I	Train run number and train run set along a line
n_u^0 and I_u^0	Number and subset of train runs for passenger u
n_u and I_u	Feasible number and subset of train runs for passenger u
h_i^+ and h_i^-	Train run i 's departure time at access station and arrival time at egress station
v_i	Train run speeds between access and egress stations
Network	
j and J	Station number and station set along a line
a and e	Access station and egress station
Auxiliary variables or functions	
X	Cumulative Distribution Function (CDF) of variable x
\dot{X}	Probability Density Function (PDF) of variable x
$P_i^{h,v}$	Probability to take train run i , conditioning on h and v
$\hat{s}_i^{h,v}$	Threshold value of repositioning distance
$\psi_{i,h}(x, v)$	Kernel function of repositioning on platform
t_p	Studied period, $t_p = [t_1, t_2]$
θ and $\hat{\theta}$	Parameter vector and its estimated value
Θ	Space of parameter vector θ

whereas their lengths range from some dozen meters in tramway stations to a couple of hundred meters in stations of metro, urban or suburban train. In the latter case, passengers are expected to walk up to a given position for waiting and boarding, boarding point M with abscissa s_M along the platform from a "starting point" A (platform entrance) with abscissa s_A . The walking-in distance s on boarding platform amounts to $|s_M - s_A|$. If a passenger does not make a significant longitudinal move on the train, passenger's boarding point (door) M and alighting point (door) N is the same door of the train. Along egress platform, the walking-out distance s' amounts to $|s_E - s_N|$ by using the same scale to measure abscissas along platforms of a given line. Thus, the total distance walked by the passenger on both the access and the egress station platforms amounts to $s + s' = |s_M - s_A| + |s_E - s_N|$. Given boarding platform entrance A and alighting platform exit E , this is a function of the boarding abscissa s_M only, made up of two parts.

Let us denote by $L \equiv |s_E - s_A|$ the maximum distance for passenger longitudinal repositioning, called longitudinal repositioning distance. Postulating that the passenger is well-aware of positions, rational and unimpeded by crowding, then it holds that $L = s + s'$ and the function does not vary with s_M . However, s_M influences the walking times prior to and posterior to running aboard train: passenger u can decide to invest time s/v in order to save on $s'/v = (L - s)/v$ just after alighting. For example, assuming that the platform entrance is at one end of the station platform (that of train head or tail) and the platform exit is at the other end of the train (that of train tail or head).

The time available to u for positioning oneself may be limited by the instant of train departure. Indexing by i the train run of interest to the passenger, let us denote as h_i^+ its departure time from access station and h_i^- its arrival time at egress station. Train travel time from access station to egress station along the train trajectory (red arrows in Figure 1)

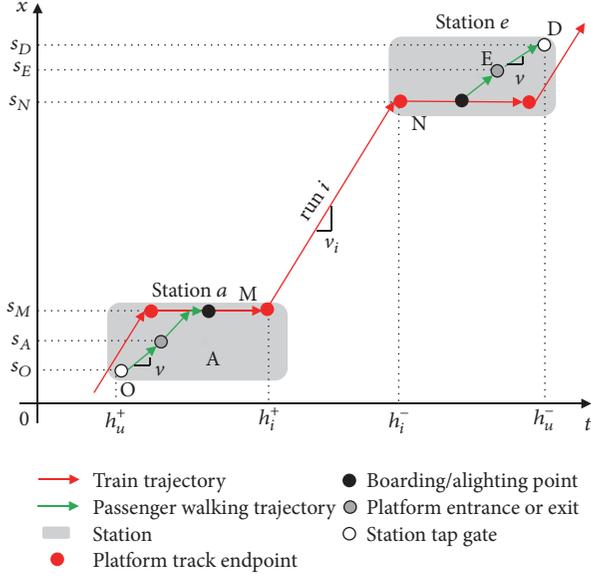


FIGURE 1: Time-space diagram: train and passenger trajectories.

is equal to $h_i^- - h_i^+$. Train travel time is considered as in-vehicle travel time for all passengers who take the same train. The frequency of each service is heterogeneous: Peak Hour (PH) is different from Off-Peak Hour (OPH). Trains along passengers' travel direction consist in a set I for all train runs i during the studied period $t_p = t_2 - t_1$. The traffic conditions and line service quality (punctuality, regularity, etc.) depend on passengers' conditions in all served stations.

Then, on boarding platform, the apparent waiting time of passenger u amounts to $w_u^i = (h_i^+ - h_u^+) - t^+$. By selecting a particular location along the platform, the passenger succeeds to turn part of that apparent waiting time w_u^i into "useful" time for the rest of his trip. His "distance investment" s is limited by both L and the available time before the train departure. Letting $\Delta h_{u,i}^+ \equiv h_i^+ - h_u^+ > 0$ passenger total time cost in access station for boarding train run i , the time available to the passenger for positioning himself along the platform is limited by $t_{u,i} \equiv \Delta h_{u,i}^+ - t^+$. So, the maximum distance for repositioning on boarding platform amounts to $s_{u,i} \equiv vt_{u,i} = v\Delta h_{u,i}^+ - s^+$.

Assume that the passenger is a rational decision-maker willing to minimize his or her exit time and walks along boarding platform as much as possible, yielding $s = \min\{L, v\Delta h_{u,i}^+ - s^+\}$. This optimizing behavior induces an adjustment of the passenger to the temporal prism of opportunities that are opened by the respective times h_i^+ and $h_u^+ + s^+/v$. The waiting time until train departure on boarding point M is called residual waiting time \tilde{w}_u^i and amounts to $(h_i^+ - h_u^+) - (s^+ + s)/v = w_u^i - s/v$.

Thus, the journey time $h_u^- - h_u^+$ is derived by $(s^+ + s)/v + \tilde{w}_u^i + (h_i^+ - h_u^+) + (s^+ + s^-)/v$. The time-space diagram (Figure 1) depicts both passenger trajectory and train trajectory between the access and egress stations.

2.2. Models Conditioning on Distributed Walking Distances. We model firstly passenger journey with distributed walking

distances and constant walking speed; the stochastic model is integrated with respect to walking (positioning) distance s on boarding platform.

2.2.1. In-Station Walking Distances Probabilistic Models. As a general notation, for a Random Variable (RV) x , its Cumulative Distribution Function (CDF) is denoted as X , and its Probability Density Function (PDF) as \dot{X} . Recall that $X(x) = \int_x \dot{X}(z)dz$, where \dot{X} is regular enough.

Urban rail transit stations vary from simple stations at grade providing access to one line only, to complex transit hubs connecting several lines and equipped with several platforms on several floors. Most of them are underground. Whatever the case, the distances between station tap gates and platform entrances or exits extend to some dozen meters at least and up to some hundred meters. Since the walking-in distance s^+ is variable, it is modeled as a RV with CDF S^+ and PDF \dot{S}^+ . The same hypothesis is applied to walking-out distance s^- , as a RV with CDF S^- and PDF \dot{S}^- .

2.2.2. Access Model. A passenger u is characterized by the (h, v) pair, where $h = h_u^+$, that is taken as exogenous.

Train run i is taken if and only if $(h + s^+/v) \in]h_{i-1}^+, h_i^+]$. Then, the probability to take train run i is as follows:

$$\begin{aligned} p_i^{h,v} &\equiv \Pr \left\{ \left(h + \frac{s^+}{v} \right) \in]h_{i-1}^+, h_i^+] \right\} \\ &= S^+ (v\Delta h_{u,i}^+) - S^+ (v\Delta h_{u,i-1}^+). \end{aligned} \quad (1)$$

The positioning distance on boarding platform depends on the train run that is taken and the walking-in distance

$$s_i^{h,v} = \min \{ L, v\Delta h_{u,i}^+ - s^+ \}. \quad (2)$$

Two cases must be distinguished

- (i) either $s_i^{h,v} = L$, that holds iff $s^+ \leq v\Delta h_{u,i}^+ - L$;
- (ii) or $s_i^{h,v} < L$, that holds iff $s^+ > v\Delta h_{u,i}^+ - L$.

The former case of total positioning can happen only if $v\Delta h_{u,i}^+ - L \geq v\Delta h_{u,i-1}^+$. Denoting $\tilde{s}_i^{h,v} = \max\{v\Delta h_{u,i-1}^+, v\Delta h_{u,i}^+ - L\}$, total positioning happens for $s^+ \in]v\Delta h_{u,i-1}^+, \tilde{s}_i^{h,v}]$, with probability

$$\Pr \{s = L \mid h, v, i\} = \frac{S^+ (\tilde{s}_i^{h,v}) - S^+ (v\Delta h_{u,i-1}^+)}{p_i^{h,v}}. \quad (3)$$

In the alternative case of partial positioning $s = x < L$, the associated PDF is

$$\Pr \{s = x \mid h, v, i\} = \frac{\dot{S}^+ (v\Delta h_{u,i}^+ - x)}{p_i^{h,v}}. \quad (4)$$

Bringing together the two cases, a CDF $S_i^{h,v}$ for the positioning distance s is built up, as $S_i^{h,v}(s) = \int_0^{\min\{s, L, v(h_i^+ - h_{i-1}^+)\}} (\dot{S}^+ (v\Delta h_{u,i}^+ - x) / p_i^{h,v}) dx + \mathbf{1}_{\{s \geq L\}} ((S^+ (\tilde{s}_i^{h,v}) - S^+ (v\Delta h_{u,i-1}^+)) / p_i^{h,v})$. This CDF satisfies the fact that

$$S_i^{h,v}(s) = \frac{\left[S^+ \left(v\Delta h_{u,i}^+ \right) - S^+ \left(v\Delta h_{u,i}^+ - \min \{s, L, v(h_i^+ - h_{i-1}^+)\} \right) + 1_{\{s \geq L\}} \left(S^+ \left(\hat{s}_i^{h,v} \right) - S^+ \left(v\Delta h_{u,i-1}^+ \right) \right) \right]}{P_i^{h,v}}. \quad (5)$$

Thus the positioning distance s is an RV conditioning on v and h on the passenger side and on i on the train side. It is endowed with a closed-form CDF that involves the tap-in instant h_u^+ , the train departure time h_i^+ , the CDF S^+ of walking-out distance from tap-in gate to the boarding platform entrance, and the target distance of repositioning L .

Of course, the RV does only exist when the matching probability is strictly positive, i.e., iff $\Delta h_{u,i}^+ > 0$. We denote as the subset I_u of train runs i that are feasible for u , i.e., such that $\Delta h_{u,i}^+ > 0$ and $h_u^+ < h_i^+$.

2.2.3. Egress Model. The positioning distance s that stems from the access model makes an input to the egress model. The distance to walk in the egress station includes $s' = L - s$ on the alighting platform plus s^- . The conditional egress time $t_{i,h,v,s}^e = (L - s + s^-)/v$ satisfies the fact that

$$\Pr \left\{ t_{i,h,v,s}^e \leq x \right\} = \Pr \left\{ L - s + s^- \leq vx \right\}. \quad (6)$$

So $t_{i,h,v,s}^e$ has CDF $T_{i,h,v,s}^e(x) = S^-(vx + s - L)$. It is integrated with respect to s ; then

$$\begin{aligned} T_{i,h,v}^e(x) &= \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} T_{i,h,v,s}^e(x) dS_i^{h,v}(s) \\ &= \frac{\psi_{i,h}(x, v)}{P_i^{h,v}}. \end{aligned} \quad (7)$$

The resulting value $T_{i,h,v}^e(x)$ is a probability conditioning on i, h, v . Let us define $\psi_{i,h}(x, v) \equiv \Pr\{t_{i,h,v}^e \leq x \cap (u \text{ takes } i)\}$. From the definition of conditional probability, we have that

$$\psi_{i,h}(x, v) \equiv \Pr \left\{ t_{i,h,v}^e \leq x \mid (u \text{ takes } i) \right\} P_i^{h,v} = P_i^{h,v} T_{i,h,v}^e(x). \quad (8)$$

On replacing $S_i^{h,v}(s)$ with its expression in (5), we get

$$\begin{aligned} \psi_{i,h}(x, v) &\equiv \left(S^+ \left(\hat{s}_i^{h,v} \right) - S^+ \left(v\Delta h_{u,i-1}^+ \right) \right) S^-(vx) \\ &+ \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} S^-(vx - L + s) \hat{S}^+(v\Delta h_{u,i}^+ - s) ds. \end{aligned} \quad (9)$$

The first term is zero if $\hat{s}_i^{h,v} = v h_{u,i-1}^+$; i.e., $L \geq v(h_i^+ - h_{i-1}^+)$.

Thus the total walking-out time t^e as an RV $t_{i,v,h}^e$ conditioning on v, h , and admissible run $i \in I_u$ is endowed with a closed-form CDF that involves the tap-in instant h_u^+ , the train departure time h_i^+ , the target distance of repositioning L , and both distributions of in-station walk distances S^+ and S^- in access and egress stations, respectively. The product between S^- and \hat{S}^+ in the formula exhibits the influence of v on both the egress and access times. In other words, the individual walk speed links together the egress and access times.

To integrate with respect to i , the CDF $T_{h,v}^e(x) = \sum_{i \in I} P_i^{h,v} T_{i,h,v}^e(x)$ is considered. From (5), it satisfies that

$$T_{h,v}^e(x) = \sum_{i \in I} \psi_{i,h}(x, v). \quad (10)$$

Concerning the tap-out instant h_u^- , conditioning on i , we have $h_u^- = t^e + h_i^-$ and $H_{i,h,v}^-(x) = T_{i,h,v}^e(x - h_i^-)$.

By integrating with respect to i , we get the unconditional CDF $H_{h,v}^-(x) = \sum_{i \in I} P_i^{h,v} T_{i,h,v}^e(x - h_i^-)$, which satisfies the fact that

$$H_{h,v}^-(x) = \sum_{i \in I} \psi_{i,h}(x - h_i^-, v). \quad (11)$$

Thus the tap-out instant h_u^- is an RV conditioning on v and h , which is endowed with a closed-form CDF that involves not only the tap-in instant h_u^+ , the train departure time h_i^+ , the target distance of repositioning L , and the distributions of in-station walk distances S^+ and S^- in access and egress stations but also the instants of train arrival at the egress station h_i^- .

From the CDF of the total walking-out time and tap-out instant, either conditioning on i or not, it is easy to derive the associated PDF by straightforward differentiation with respect to argument x . That is,

$$\begin{aligned} \dot{\psi}_{i,h}(x, v) &\equiv \frac{\partial}{\partial x} \psi_{i,h}(x, v) = \left(S^+ \left(\hat{s}_i^{h,v} \right) \right. \\ &- S^+ \left(v\Delta h_{u,i-1}^+ \right) \left. \right) v \dot{S}^-(vx) \\ &+ v \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} \dot{S}^-(vx - L + s) \\ &\cdot \dot{S}^+(v\Delta h_{u,i}^+ - s) ds. \end{aligned} \quad (12)$$

The associated PDFs are

$$\dot{T}_{i,h,v}^e(x) = P_i^{h,v} \dot{\psi}_{i,h}(x, v); \quad (13a)$$

$$\dot{T}_{h,v}^e(x) = \sum_{i \in I} \dot{\psi}_{i,h}(x, v); \quad (13b)$$

$$\dot{H}_{i,h,v}^-(x) = P_i^{h,v} \dot{\psi}_{i,h}(x - h_i^-, v); \quad (13c)$$

$$\dot{H}_{h,v}^-(x) = \sum_{i \in I} \dot{\psi}_{i,h}(x - h_i^-, v). \quad (13d)$$

It should be noted that the above stochastic model assumes that a passenger has constant velocity motion in access and egress stations. From a practical point of view, this assumption may not be realistic and thus too restrictive. Extensions will be considered in next subsection to consider a distributed walking speed.

2.3. Models Conditioning on Distributed Speed and Walking Distances. When passenger walking speed v is a distributed variable, the stochastic model is integrated with respect to walking speed as well.

2.3.1. In-Station Walking Speed Probabilistic Model. The walking speed v of one passenger u is a notional speed, averaged for that passenger on that journey over a range of travel situations. It may be called an individual “cruising speed” during walking. Although personal walking habits may be consistent, as a fact, the cruising walking speed fluctuates from one occasion to another, for instance, between journeys reiterated on a given access and egress pair from day to day.

In addition to this intraindividual diversity, there is an even larger diversity between individual passengers, since people differ in their respective walking abilities. Young adults can walk faster than elderly people and are likely to be more hurried. People with luggage or young child either walking or in a stroller walks more slowly than the average adult. A rough indication about the statistical distribution of walk speeds for a typical population of transit users in the urban setting was close to a normal distribution with mean of about 0.90 m/s and standard deviation of about 0.20 m/s or to a uniform distribution ranging from 0.58 to 1.24 m/s in the Appendix, the cases with waiting time integrated in walking time.

Assuming that passengers’ walking speeds on walking links in access and egress stations obey the same statistical distribution, it is easy to extend the stochastic model of passenger repositioning to a diversity of walking speeds, by considering walking speed in a given population U of passengers as a RV, with CDF V and PDF \dot{V} .

We still denote S^+ and S^- the distribution functions of in-station walking distances. In fact, they can be expected to have different average values and wider spreads as compared to their individual counterparts. Such differences between intra-individual and inter-individual distributions were classical in the stochastic modeling of socioeconomic behaviors of highway individuals in Chapter 6 of [38].

2.3.2. Access Model. To integrate with respect to v , the probability to take train run i is still conditioning on the tap-in time $h = h_u^+$ and is

$$\begin{aligned} p_i^h &\equiv \int p_i^{h,v} dV(v) \\ &= \int \left(S^+(v\Delta h_{u,i}^+) - S^+(v\Delta h_{u,i-1}^+) \right) dV(v). \end{aligned} \quad (14)$$

Then, conditioning on i and h , the distribution of walking speeds has PDF as follows:

$$\dot{V}_i^h(v) = \frac{\dot{V}(v) p_i^{h,v}}{p_i^h}. \quad (15)$$

2.3.3. Egress Model. Conditioning on i and h , the egress time $t_{i,h}^e$ has CDF $T_{i,h}^e(x) = \int T_{i,h,v}^e(x) \dot{V}_i^h(v) dv = \int (\psi_{i,h}(x, v) / p_i^{h,v}) \dot{V}_i^h(v) dv$, so

$$T_{i,h}^e(x) = \frac{1}{p_i^h} \int \psi_{i,h}(x, v) dV(v). \quad (16)$$

Denoting $\tilde{\psi}_{i,h}(x) \equiv \int \psi_{i,h}(x, v) dV(v)$, it is then easy to integrate with respect to i . As $\{t_h^e \leq x\} = \bigcup_{i \in I} \{t_{i,h}^e \leq x \cap i\}$, there is $\Pr\{t_h^e \leq x\} = \sum_{i \in I} \Pr\{t_{i,h}^e \leq x\} \Pr\{i\}$ from the composition of conditional probabilities. Thus,

$$T_h^e(x) = \sum_{i \in I} T_{i,h}^e(x) p_i^h = \sum_{i \in I} \tilde{\psi}_{i,h}(x). \quad (17)$$

The exit instant h_u^- has CDF as follows:

$$H_h^-(x) = \sum_{i \in I} \tilde{\psi}_{i,h}(x - h_i^-). \quad (18)$$

Thus, both the total walking-out time and the tap-out instant are endowed with closed-form CDF that involve the tap-in instant h and the statistical distribution of s^+ , s^- , and v . This enables us to derive their PDF by differentiation with respect to their argument x . Denoting $\dot{\tilde{\psi}}_{i,h}(x) \equiv (\partial/\partial x) \tilde{\psi}_{i,h}(x)$, it holds that

$$\dot{T}_h^e(x) = \sum_{i \in I} \dot{\tilde{\psi}}_{i,h}(x); \quad (19a)$$

$$\dot{H}_h^-(x) = \sum_{i \in I} \dot{\tilde{\psi}}_{i,h}(x - h_i^-). \quad (19b)$$

3. Distribution Specification for Tractable Computation

The analytical formulae obtained so far involve the integration of specific functions along one or two scalar dimensions, namely one dimension of space (with respect to distance s) and eventually another dimension of speed (with respect to walking speed v). The outcomes can be obtained by numerical integration along the said axes. This can be circumvented by availing ourselves of ad hoc distributions that yield straightforward formulae for the respective PDFs of total walking-out time and tap-out instant, though it is not so simple.

In this section, we firstly put forward specific distributions that are suitable to our purpose (Section 3.1). Then, we provide a lemma (Section 3.2) for the core computation to deal with distributed speed by providing formulae for $x \mapsto \psi_{i,h}(x, v)$ and $x \mapsto \tilde{\psi}_{i,h}(x)$ functions (Section 3.3). Finally, the variations of those two functions are illustrated (Section 3.4).

3.1. Ad Hoc Specification of the Distributions. The obtained general PDF functions constitute the likelihood functions of all assumed parameters. To prepare for further work on the MLE of those parameters, some hints of PDF computation under ad hoc selection of distributions are provided.

For a given pair of access and egress stations, we take v as either a variable or a normal-distributed walking speed, together with shifted exponential-distributed walking distances and the variable L . The distribution functions of v , s^+ , and s^- are specified as follows:

- (i) Individual speed follows a normal (Gaussian) distribution with mean μ and variance σ^2 , since a normal distribution reduces likelihood function complexity [23].
- (ii) Walking-in distance s^+ obeys a shifted exponential distribution with main parameter λ^+ and shift B^+ ; thus CDF $S^+(x) = 1_{\{x \geq B^+\}} [1 - \exp(-\lambda^+(x - B^+))]$ and PDF $\dot{S}^+(x) = 1_{\{x \geq B^+\}} \lambda^+ \exp(-\lambda^+(x - B^+))$.
- (iii) Walking-out distance s^- is also a shifted exponential distribution, with main parameter λ^- and shift B^- ; thus CDF $S^-(x) = 1_{\{x \geq B^-\}} [1 - \exp(-\lambda^-(x - B^-))]$ and PDF $\dot{S}^-(x) = 1_{\{x \geq B^-\}} \lambda^- \exp(-\lambda^-(x - B^-))$.

3.2. Lemma. Let us establish a property for a normal RV $N(\mu, \sigma^2)$ combined with a consumption function that is the product of Kv and an exponential function e^{-lv+k} . The aim is to obtain a straightforward formula for integral function $F(x) \equiv \int_{-\infty}^x Kve^{-lv+k} d\Phi((v - \mu)/\sigma)$.

As $e^{-lv} e^{-(1/2)((v-\mu)/\sigma)^2} = e^{-(1/2)((v-\mu')/\sigma)^2} e^{l\mu - (1/2)l^2\sigma^2}$ for $\mu' = \mu - l\sigma^2$, we can put aside the constant factor $Ke^{k+l\mu - (1/2)l^2\sigma^2}$ and focus on

$$\begin{aligned} \int_{-\infty}^x vd\Phi\left(\frac{v - \mu'}{\sigma}\right) &= \int_{-\infty}^x \frac{v - \mu'}{\sqrt{2\pi}\sigma} e^{-(1/2)((v-\mu')/\sigma)^2} dx \\ &\quad + \int_{-\infty}^x \mu' d\Phi\left(\frac{v - \mu'}{\sigma}\right) \\ &= \frac{\sigma}{\sqrt{2\pi}} \left[-e^{-(1/2)((v-\mu')/\sigma)^2} \right]_{-\infty}^x \\ &\quad + \mu' \Phi\left(\frac{x - \mu'}{\sigma}\right) \\ &= \mu' \Phi\left(\frac{x - \mu'}{\sigma}\right) \\ &\quad - \frac{\sigma}{\sqrt{2\pi}} e^{-(1/2)((x-\mu')/\sigma)^2}. \end{aligned} \quad (20)$$

So the final result is $\int_{-\infty}^x Kve^{-lv+k} d\Phi((v - \mu)/\sigma) = Ke^{k+l\mu - (1/2)l^2\sigma^2} [\mu' \Phi((x - \mu')/\sigma) - (\sigma/\sqrt{2\pi})e^{-(1/2)((x-\mu')/\sigma)^2}]$.

3.3. Core Computation. The PDF formulae (13d), (13b), and (13c) for a constant speed and (19a) and (19b) for

normal-distributed speed involve basic bricks of the form $x \mapsto \psi_{i,h}(x, v)$ and $x \mapsto \tilde{\psi}_{i,h}(x)$, respectively. These are obtained by straightforward differentiation of function $\psi_{i,h}(x, v)$ with respect to x .

As $\psi_{i,h}(x, v) \equiv (S^+(\tilde{s}_i^{h,v}) - S^+(v\Delta h_{u,i-1}^+))S^-(vx) + \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} S^-(vx - L + s)\dot{S}^+(v\Delta h_{u,i}^+ - s)ds$, then $\psi_{i,h}(x, v) \equiv (S^+(\tilde{s}_i^{h,v}) - S^+(v\Delta h_{u,i-1}^+))v\dot{S}^-(vx) + v \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} \dot{S}^-(vx - L + s)\dot{S}^+(v\Delta h_{u,i}^+ - s)ds$. This function is in two parts, left and right, respectively.

The left part is computed firstly and produces the two terms: $(S^+(\tilde{s}_i^{h,v}) - S^+(v\Delta h_{u,i-1}^+)) = e^{-\lambda^+(v\Delta h_{u,i-1}^+ - B^+)} 1_{\{v\Delta h_{u,i-1}^+ \geq B^+\}} - e^{-\lambda^+(\max\{v\Delta h_{u,i-1}^+, v\Delta h_{u,i}^+ - L\} - B^+)} 1_{\{\max\{v\Delta h_{u,i-1}^+, v\Delta h_{u,i}^+ - L\} \geq B^+\}}$ and $v\dot{S}^-(vx) = v\lambda^- e^{-\lambda^-(vx - B^-)} 1_{\{vx \geq B^-\}}$.

So the product of the two terms is ready to compute the left part in $\psi_{i,h}(x, v)$.

To obtain $x \mapsto \tilde{\psi}_{i,h}(x)$, it is necessary to integrate the product over v . The product is nonzero only if $vx \geq B^-$; i.e., $v \geq B^-/x$. Then, it gives rise to two terms with respective formula as follows:

$$(L-a) \quad v\lambda^- e^{-\lambda^-(vx - B^-) - \lambda^+(v\Delta h_{u,i-1}^+ - B^+)} 1_{\{v\Delta h_{u,i-1}^+ \geq B^+\}} \text{ and}$$

$$(L-b) \quad v\lambda^- e^{-\lambda^-(vx - B^-) - \lambda^+(\max\{v\Delta h_{u,i-1}^+, \Delta h_{u,i}^+ - L\} - B^+)} 1_{\{\max\{v\Delta h_{u,i-1}^+, \Delta h_{u,i}^+ - L\} \geq B^+\}}.$$

The first one (L-a) must be integrated for $v \geq B^-/\Delta h_{u,i-1}^+$ only. Denoting $\mathcal{V} = \max\{B^-/x, B^-/\Delta h_{u,i-1}^+\}$, there is $\int_{\mathcal{V}}^{+\infty} v\lambda^- e^{-\lambda^-(vx - B^-) - \lambda^+(v\Delta h_{u,i-1}^+ - B^+)} dV(v) = \lambda^- e^{\lambda^+ B^+ + \lambda^- B^-} \int_{\mathcal{V}}^{+\infty} v e^{-(\lambda^- x + \lambda^+ \Delta h_{u,i-1}^+)v} dV(v)$.

The second term (L-b) is dealt with similarly, yet with distinction between two subdomains depending on which function is greater between $v \mapsto v\Delta h_{u,i-1}^+$ and $v \mapsto v\Delta h_{u,i}^+ - L$.

If $v\Delta h_{u,i-1}^+ \leq v\Delta h_{u,i}^+ - L$, i.e., $L/(h_i^+ - h_{i-1}^+) \leq v$, then, denoting $\mathcal{Y} = \max\{B^-/x, L/(h_i^+ - h_{i-1}^+)\}$, the second term reduces to $\int_{\mathcal{Y}}^{+\infty} v\lambda^- e^{-\lambda^-(vx - B^-) - \lambda^+(v\Delta h_{u,i-1}^+ - L - B^+)} dV(v) = \lambda^- e^{\lambda^+(L+B^+) + \lambda^- B^-} \int_{\mathcal{Y}}^{+\infty} v e^{-(\lambda^- x + \lambda^+ \Delta h_{u,i-1}^+)v} dV(v)$.

In the other case where $v < L/(h_i^+ - h_{i-1}^+)$, the subdomain of integration is empty if $L/(h_i^+ - h_{i-1}^+) < B^-/x$ or $[B^-/x, L/(h_i^+ - h_{i-1}^+)]$ otherwise. The formula is the same as for (L-a).

Concerning the right part, there is

$$\begin{aligned} v \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} \dot{S}^-(vx - L + s)\dot{S}^+(v\Delta h_{u,i}^+ - s)ds &= v\lambda^- \lambda^+ \int_0^{\min\{L, v(h_i^+ - h_{i-1}^+)\}} e^{-\lambda^-(vx - L + s - B^-) - \lambda^+(v\Delta h_{u,i-1}^+ - s - B^+)} 1_{\{vx - L + s \geq B^-\}} \\ 1_{\{v\Delta h_{u,i-1}^+ - s \geq B^+\}} ds &= \sigma v\lambda^- \lambda^+ \int_{B^- + L - vx}^{\min\{L, v(h_i^+ - h_{i-1}^+), v\Delta h_{u,i-1}^+ - B^+\}} e^{-\lambda^-(vx - L + s - B^-) - \lambda^+(v\Delta h_{u,i-1}^+ - s - B^+)} ds \\ &= \sigma \frac{\lambda^- \lambda^+}{\lambda^+ - \lambda^-} v e^{\lambda^+(B^- - v\Delta h_{u,i-1}^+) + \lambda^-(L+B^+)} \left[e^{(\lambda^+ - \lambda^-)s} \right]_{(B^- + L - vx)}^{\min\{L, v(h_i^+ - h_{i-1}^+), v\Delta h_{u,i-1}^+ - B^+\}}, \end{aligned} \quad (21)$$

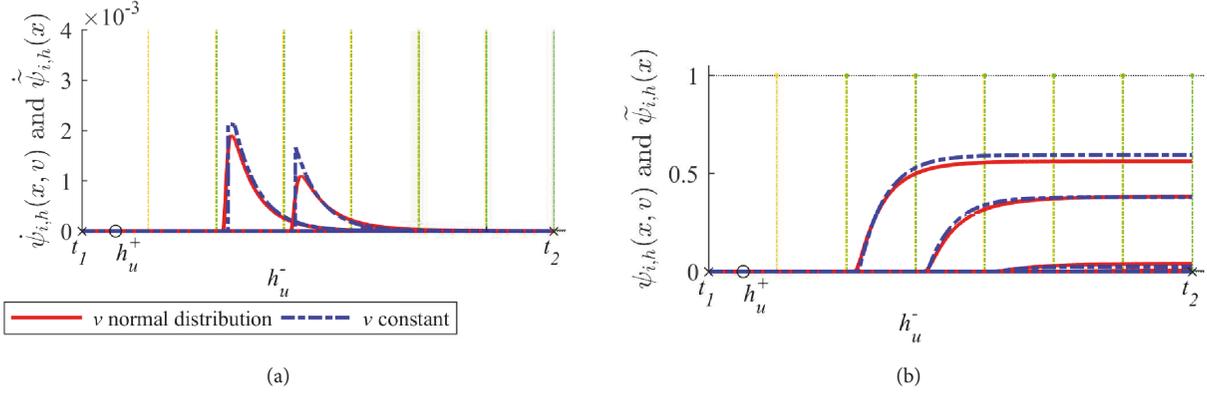


FIGURE 2: With respect to tap-out instant: (a) bricks of PDFs and (b) bricks of CDFs.

where $\sigma \equiv \mathbb{1}_{\min\{L, v(h_i^+ - h_{i-1}^+), v\Delta h_{u,i-1}^+ - B^+\} \geq (B^- + L - vx)^+}$. This gives the right part in $\psi_{i,h}(x, v)$ as a tractable analytical formula. It is thus easy to compute function $\psi_{i,h}(x, v)$.

As for integration over the distribution of walking speeds, the right part breaks into two bricks to which the lemma applies.

3.4. Illustration. By taking into account the previous distributions of v, s^+, s^- , the parameter vector is a six- or seven-fold vector $\theta = \{\mu, \sigma, B^+, \lambda^+, B^-, \lambda^-, L\}$ (without σ for a constant speed $v = \mu$). During a studied period $[t_1, t_2]$ and a given h_u^+ (black circle), Figure 2(a) depicts the variations of functions $x \mapsto \psi_{i,h}(x, v)$ and $x \mapsto \tilde{\psi}_{i,h}(x)$ with respect to tap-out instant h_u^- starting from that of a given train arrival time h_i^- at the egress station (green dashed lines), either conditioning on constant (blue dashed line) or normal-distributed (red solid line) walking speed v . The brown dashed lines represent train departure times in access station. Figure 2(b) provides the corresponding integral functions $x \mapsto \Psi_{i,h}(x, v)$ and $x \mapsto \tilde{\Psi}_{i,h}(x)$. The difference between conditional and unconditional looks minor though discernible. For a given tap-in moment h_u^+ , the functions relate to each train run i which the passenger could take. The most feasible train is the first arrival train at egress station: the latter the train arrival, the smaller the probability.

4. Statistical Estimation

Previous analytical formulae constitute the cores of our stochastic models. They derive analytical closed-form formulae that provide likelihood functions, which are tractable under a specification of basic statistical distributions. The stochastic models are theoretical constructs that involve human behaviors of trip-making in relation to the dynamic process of train runs. Such a theoretical model can be applied to particular cases, notably so by estimating the values of its parameters so as to make its outcomes replicate observed values well.

As reported by [25, 26], AFC data provide ample information on users' pairs of tap instants in stations, while AVL data provide both departure and arrival times in stations for each train run along its route [41].

In this section, we put forward an approach of MLE (Section 4.1). Its implementation involves a scheme for practical application (Section 4.1). Then, we build upon model estimation by proposing an inference method to enrich the observed data by adding "most likely predictions" of unobserved items (Section 4.1).

4.1. Maximum Likelihood Estimator. Let us assume here that TITO (h_u^+, h_u^-) pairs are observed over a journey between an O-D pair of access and egress stations. The model parameters consist in a six- or seven-fold vector $\theta = \{\mu, \sigma, B^+, \lambda^+, B^-, \lambda^-, L\}$ (without σ for a constant speed $v = \mu$). Knowing $\theta, h = h_u^+$, and train departure and arrival time pairs $\{(h_i^+, h_i^-) : i \in I\}$, the PDF $x \mapsto \dot{H}_{h_u^-}^-(x = h_u^-)$ is a function of tap-out instant h_u^- . Conversely, given (h_u^+, h_u^-) and $\{(h_i^+, h_i^-) : i \in I\}$, the same formula can be interpreted as a function of θ , called the individual likelihood function for one trip of a passenger. It is then denoted as

$$L_u(\theta) = \dot{H}_{h_u^-}^-(h_u^+). \quad (22)$$

The joint observation of a sample $\{(h_u^+, h_u^-) : u \in U\}$ of journeys provides a joint/total likelihood function, denoted $L_U(\theta)$. If the observations are independent, then

$$L_U(\theta) = \prod_{u \in U} L_u(\theta). \quad (23)$$

The MLE consists to set up the value of parameter vector θ so as to maximize the likelihood function of the observed sample, or equivalently to maximize the log-likelihood function as follows:

$$\begin{aligned} \Lambda_U(\theta) &\equiv \ln L_U(\theta) = \sum_{u \in U} \Lambda_u(\theta) \quad \text{and} \quad \Lambda_u(\theta) \\ &= \ln L_u(\theta). \end{aligned} \quad (24)$$

That simplifies the computation of MLE.

The estimator of MLE can be applied to our stochastic models for either an individual passenger observed on several journeys or a set of passengers to differentiate 'intra-' versus 'inter-' individual cases. In the former case [24],

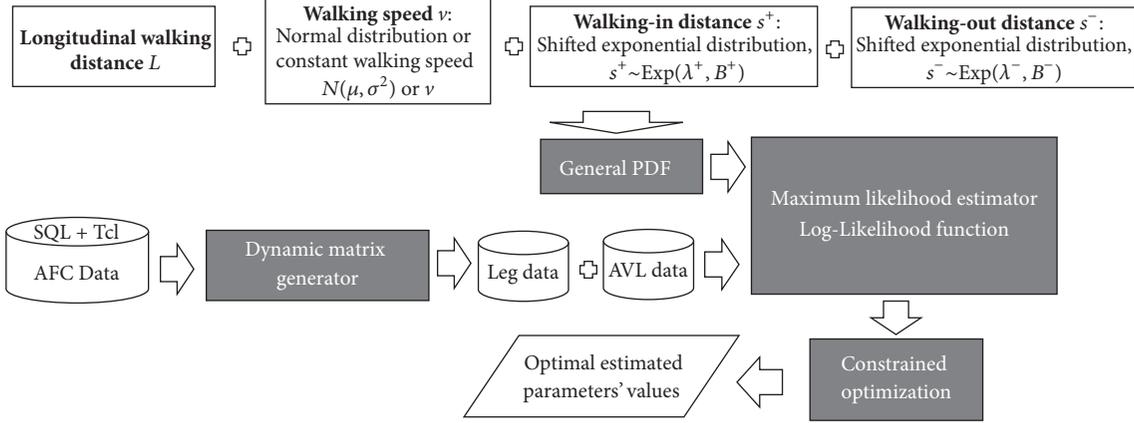


FIGURE 3: Modeling and optimization schema based on MLE.

the underlying statistical distributions are related to the particular passenger: his walking speed as either a constant or a distribution to allow for fluctuations and his own conditions for in-station walking distances and repositioning target L . In the latter case [21–24, 42], the condition is aimed to ensure statistical independence within the sample. The O-D stations must be shared by the passengers so as to give consistency to the L , S^+ , and S^- notions. The former will be further investigated in this paper.

The estimator of MLE is endowed with powerful statistical properties that are well known in econometrics [43]. Although it has often some bias, this bias vanishes when the sample size tends to infinity; and the estimator variance is minimal. An especially valuable property is that the Hessian matrix of the log-likelihood function evaluated at the global optimum, up to a minus sign, contains the estimated covariance by pair of scalar parameters $\Lambda_U(\theta) \equiv \ln L_U(\theta) = \sum_{u \in U} \Lambda_u(\theta)$. As these properties pertain to points of global maximization, a suitable optimization algorithm must be used. Furthermore, the properties rely upon a requirement of parameter-free domain for the RVs in the model. The requirement is satisfied by the normal distribution of speeds, but not by the shift in a shifted exponential distribution. So the application of MLE to our model is somewhat heuristic. However, under given B^+ and B^- , the estimation of the remaining parameters meets the domain condition.

4.2. Optimization. In practice, the estimator searches for the estimate $\hat{\theta}$ within an admissible space Θ with bounds of vector components. To improve the optimization approach, a global pre-estimation of the space is proposed to find the optimal initial point. The pseudocode by nonlinear constrained optimization is shown in Figure 3.

The available AFC dataset was exploited by a specific dynamic O-D matrix inference scheme devised in [44], which extended previous works of [45, 46]. The scheme involves three principal steps as in [22]: (i) extracting the data of a given line from the dataset of the transit network, (ii) data filtering to exclude one-tap individual records and data inferring process for other oddities, and (iii) generating O-D

pairs by scanning individual records to select appropriate trips.

4.3. Guidelines for Posterior Analysis. Based on parameter estimates, we can model each journey in the observed sample. The outcomes fall into three categories: (i) user's attributes of walking speed v and maximum longitudinal repositioning distance L ; (ii) matching probabilities between train runs and passengers; and (iii) the level of probability associated with a pair of TITO times (h_u^+, h_u^-) .

Matching probabilities associated with a given journey may be analyzed in three steps: (i) to identify the number n_u of feasible train runs I_u per trip by the model below from the number n_u^0 of hypothetical train runs compatible with $h_u^+ < h_i^+ < h_i^- < h_u^-$; (ii) among the feasible train run probabilities, to identify the biggest one; (iii) to evaluate the “dominance ratio” of the most likely run, the ratio between the second biggest and the biggest probabilities when the feasible train runs is more than 1.

Furthermore, based on the parameter estimates, we can make inference about some journey items that are not observed per se. Such ex-post analysis in a given journey of a passenger can be performed along the passenger's trajectory in the following way. In this model, assume that in each trip k of passenger, the train run with the biggest probability is taken, but conditioning on estimated average speed $E(v)$. Main terms of passenger trajectory are derived directly by the following calculations:

- (i) mean of truncated exponential distribution, $E(s^+(k))$, between B^+ and B^+ plus maximum value compatible with times and speed $s_{\max}^+ = (h_i^+ - h_u^+)E(v)$
- (ii) repositioning distance, $s(k) = \min(s1, E(L))$, where $s1 = s_{\max}^+ - E(s^+(k))$, based on estimated L
- (iii) residual waiting time, the waiting time in excess of repositioning time, $\hat{w} = \max(0, s1 - E(L))/E(v)$
- (iv) mean of truncated exponential distribution, $E(s^-(k))$, between B^- and B^- plus maximum value compatible with times and speed $s_{\max}^- = (h_i^- - h_u^-)E(v)$.

Hence, tap-out time h_u^- is equal to $h_u^+ + t^+ + \widehat{w}_u^i + L/E(v) + (h_i^+ - h_u^+) + t^-$, where $t^+ = E(s^+(k))/E(v)$ and $t^- = E(s^-(k))/E(v)$. Once the tap-in times are given, the tap-out times are calculated.

5. Case Study of a Commuter Rail Line in Paris Area

The models are applied to a real case study, the busiest urban rail transit line RER A in Paris area, France, on the basis of AFC data provided by IdFM (ex STIF) and AVL data provided by RATP.

After introducing the case, observations of trip-making, and train traffic (Section 5.1), we estimate the model parameters with distributed walking speed for two samples of journeys (Section 5.2). Then, building upon the data and the estimated parameters, we infer the distance and time components for every sampled journey (Section 5.3).

5.1. Case Presentation: Navigo System, Line RER A, and Related Datasets. There are two main systems of urban rail transit in Paris area [47]: (i) the semiclosed metro system including 14 lines, equipped only tap-in gates, (ii) the heavier train system, including the “Transilien” and RER (Réseau Express Régional, the Regional Express Network) systems, and the closed RER system including 5 lines equipped both TITO gates, except for connected transfer stations between RER lines. The Paris transit system is integrated as concerns fares, and there is a unique smartcard called Navigo. Thus, the AFC system of transit in Paris area is called Navigo system. The Navigo system records anonymous passenger information including the smartcard number which is anonymized (with anonymous number that is maintained for only 3 months), the date, the validation instant at tap-in or tap-out gate, the gate ID, and the access or egress station name. During the PHs on workdays, more than 90% of the trips taken PT were home-work or home-study trips using network subscription hence the smartcard.

The line RER A is the busiest urban rail transit line in Paris area and maybe Europe, carrying more than one million passengers every workday [44]. It contains 46 stations in total 109 km and is structured around a central trunk into which are grafted five branches [48]: two eastward branches, northeast terminal Chessy and southeast terminal Boissy; and three westward branches, northwest terminal Cergy, central-west terminal Poissy, and southwest terminal Saint-Germain. The central trunk between stations Vincennes and La Défense passes through the largest underground mass transit hub, Châtelet-Les Halles, and serves the major business district La Défense in France.

The train time headways on the central trunk range from 2 min at peak to 10 min off peak. Our study focused on the O-D pair between Vincennes and La Défense on the central trunk. Each station has a number of entrances in relation to its importance, from 2 at Vincennes to 6 at La Défense. The topological structures of O-D pair Vincennes and La Défense are detailed in Figure 4, in which depicts the line platform at either station and indicates some of the passenger walking paths (green arrows) between tap gates and train

doors. The ‘Copy’ nodes connect to the same destination nodes as the ‘Copied’ one to form in-station itineraries. There are more route choices in egress station. The complexity of route choice in egress station comes from the choices between alighting doors and the platform exit points. There is another source of route choice in egress station, the choices between tap gates and station exits, which could not be considered only by the AFC data. The diversity of passenger paths is a source about the variability of in-station walking distance.

AVL and AFC datasets were made available to us by the line operator RATP and the mobility authority IdFM, respectively, for a period in March 2015 from the 16th to the 29th, excluding the 21st, 22nd, and 23rd. AVL data of RATP includes trains’ arrival and departure times in stations. Out of the AFC data pertaining to the O-D pair Vincennes and La Défense in either direction, we selected one sample per direction, both for a given passenger with maximum number of such journeys during the period. As it turns out, the two busiest cards are identical, with 15 trips in Case 1 from La Défense to Vincennes and 16 trips in Case 2 from Vincennes to La Défense. Figure 5(a) exhibits the TITO times of the 31 trips (5 trips in 1 day, 4 trips in 4 days, 3 trips in 3 days, 2 trips in 0 day, 1 trip in 1 day, and 0 trips in 2 days). All trips are during workdays, except one trip on Sunday March 29th 2015. Indeed, the trip pattern is peculiar since every trip from Vincennes to La Défense is followed by a return trip after a short period of about two hours in the morning or one hour in the evening; those trips would be home-work trips. Figure 5(b) evaluates the related journey times: on average, the journey takes more time from Vincennes to La Défense during PHs than that from La Défense to Vincennes during OPHs.

5.2. Estimations. The admissible spaces of scalar parameters were specified on the basis of field measurement or literature on urban mobility. As for in-station walking distances, a range of [4, 325] m applies to s^+ in Vincennes, and a range of [4, 525] m applies to s^- in La Défense. Platform lengths amount to 225 m in each station: yet we allowed a feasible space of [0, 600] m for L , as it may be bigger than the sum of access and egress station platform lengths 450 m. A feasible range of (0, 4] m/s is imposed on average walking speed v or μ , based on the last household travel survey in Paris area [49]. This walking speed v includes the relative speed of escalator or lift. Based on previous given bounds of variables, reasonable ranges for all estimated parameters are defined, as reported in Table 2.

While the log-likelihood function of the model gets its maximum value, parameters’ estimates and their Standard Deviations (SDs, approximations) are obtained in Table 2. The results show averaged day-to-day dynamic variation of this passenger. It appears that the estimates of the target repositioning distances, \widehat{L}_1 and \widehat{L}_2 , are relatively consistent and fairly precise. The same applies to the distance shifts \widehat{B}_1^+ , \widehat{B}_2^+ , and \widehat{B}_1^- , but \widehat{B}_2^- is less precise. About the main parameters of in-station distances, the estimated values are small for $\widehat{\lambda}_1^+$, $\widehat{\lambda}_1^-$, $\widehat{\lambda}_2^+$, and $\widehat{\lambda}_2^-$: the first of them is not significantly different

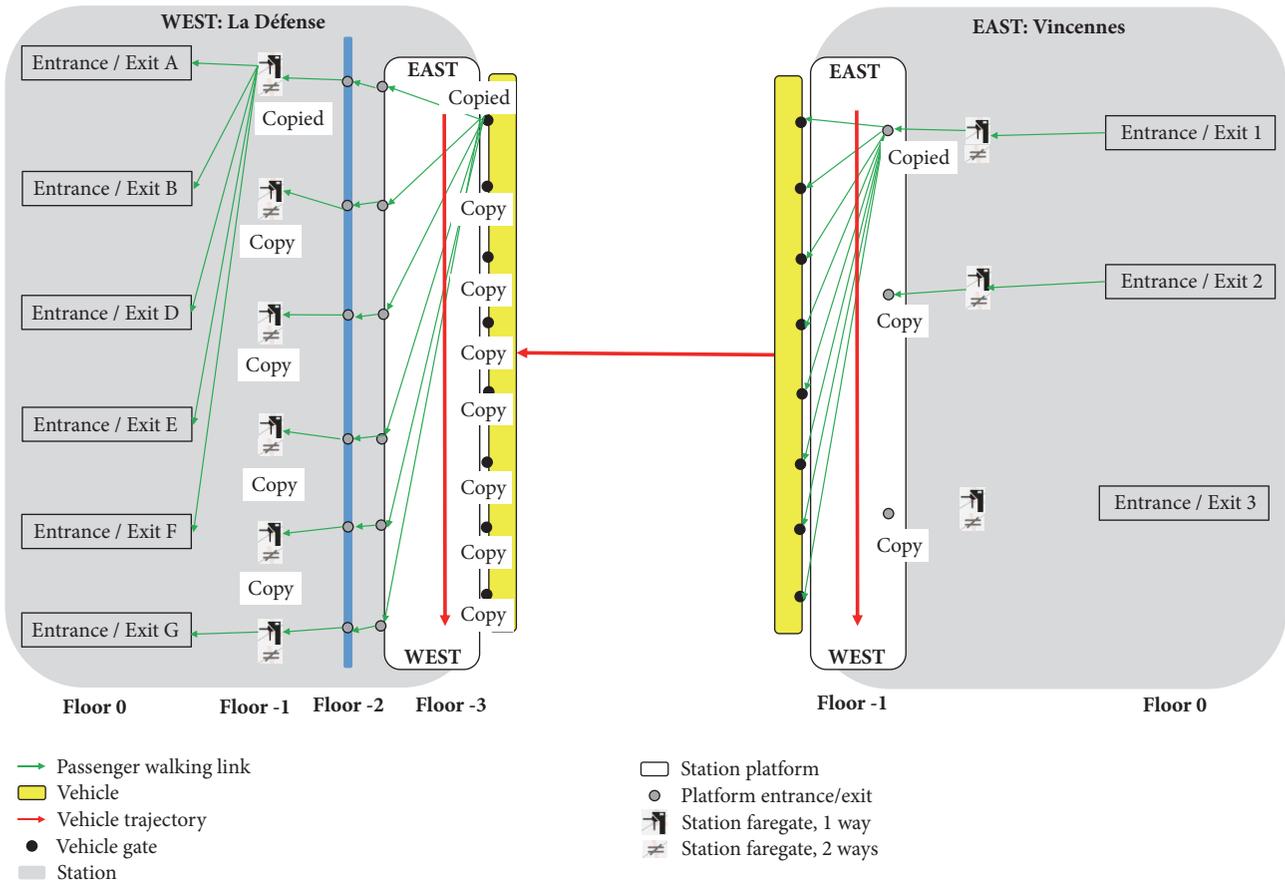


FIGURE 4: Sketch of stations' topological structures: Vincennes and La Défense.

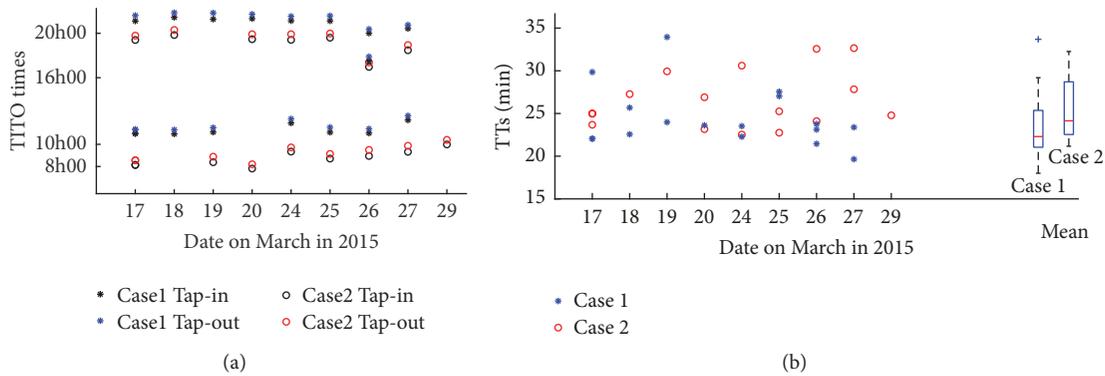


FIGURE 5: Trips made by the busiest passenger during nine days in March 2015.

from zero. As for speed estimates, the average values are fair in Case 2 and tolerable in Case 1. However, the discrepancy between the two cases calls for further investigation, since the data pertain to the same user for whom an identical walking speed is expected in both directions. The parameter of speed dispersion, σ , has a low estimate in Case 2 but a bigger one in Case 1.

In all, the consideration of an individual passenger enabled us to recover meaningful information about his or her trip-making behavior. The target repositioning distances

are significant, at about 101 or 76 m depending on the direction. Combined to the respective estimates of average speeds, these distances correspond to repositioning time of about 60 s in both cases. This value is close to half of a time headway at peak hours. The in-station distances exhibit a mirror effect at Vincennes station (similarity between s_2^+ and s_1^-), but there is a discrepancy at La Défense station between s_1^+ and s_2^- . A potential reason may pertain to high passenger crowding that occurs at La Défense during PHs, which makes individual walking more dependent on crowd dynamics,

TABLE 2: Parameter estimation for a single passenger by MLE.

Indicator (Unit)	Parameter (Unit)	Range	Case 1		Range	Case 2	
			Estimated value	Standard deviation		Estimated value	Standard deviation
ν (m/s)	μ (m/s)	(0, 4]	1.70	0.054	(0, 4]	1.19	0.21
	σ	(0, 4]	0.33	0.059	(0, 4]	0.07	0.01
s^+ (m)	λ^+ (1/m)	(0, 2]	0.017	0.015	(0, 2]	0.73	0.03
	B^+ (m)	(0, 550]	126.86	14.72	(0, 150]	106.29	10.14
s^- (m)	λ^- (1/m)	(0, 2]	0.91	0.042	(0, 2]	0.10	0.01
	B^- (m)	(0, 150]	93.49	2.78	(0, 550]	70.52	28.33
L (m)	L (m)	(0, 600]	101.46	8.28	(0, 600]	76.14	6.35
Log-likelihood function			-47.48		-49.01		

TABLE 3: Indicators' mean values.

Mean value (Unit)	Range	Case 1		Range	Case 2	
		Estimated mean value	Standard deviation		Estimated mean value	Standard deviation
$E(\nu)$ (m/s)	(0, 4]	1.70	0.33	(0, 4]	1.19	0.07
$E(s^+)$ (m)	[4, 525]	186.75	59.89	[4, 325]	107.66	1.37
$E(s^-)$ (m)	[4, 325]	94.59	1.10	[4, 525]	80.52	10.00
$E(L)$ (m)	[3, 600]	101.46	8.28	[3, 600]	76.14	6.35
$E(t^+)$ (min)	NA	1.83	NA	NA	1.51	NA
$E(t^-)$ (min)	NA	0.93	NA	NA	1.28	NA

*NA: not available.

especially so at station egress. The difference between the parameter values in the same station indicates the difference of walking-in and walking-out routes. The difference of values between different stations depicts the station topological structures.

Mean walking speed and mean walking distances are calculated based on parameter estimates and illustrated in Table 3. Passenger's mean walking speed with normal distribution is equal to $\hat{\mu}$ with SD $\hat{\sigma}$. As regards the walking distances with shifted distributions, the mean values are recovered as $\hat{B} + 1/\hat{\lambda}$ with SD $1/\hat{\lambda}$. The mean value of longitudinal walking distance is estimated directly by the model. Table 3 provides also some more indications about in-station mean walking times, derived directly by $E(s^+)/E(\nu)$ and $E(s^-)/E(\nu)$. The results are consistent with the above comments. It suggests that, for a given passenger, walking times may be more reliably estimated than the pair of walking distances and walking speed.

In the Appendix, estimation results for a former model neglecting the repositioning behavior are recalled, for the same O-D pair but a single day of observation and a population set of passengers. The journey is from Vincennes to La Défense, which corresponds to Case 2 here. There is much agreement between the intraindividual estimates of Case 2 and the interindividual estimates of model M1 (normal-distributed walking speed) in the Appendix, except for the shift parameter of in-station distances on the egress side. In fact, $\hat{L}_2 + \hat{B}_2^- \approx 140$ m are closer to $\hat{B}_A^- \approx 204$ m, which seems to confirm that repositioning strategy on the boarding platform constitutes a distance investment by the users for the rest of the journey.

5.3. *Posterior Analysis to Infer Journey Details.* Similar to Bayesian analysis, parameter estimates are used for reanalyzing train run occurrences and passenger trip attributes. It will check the rationality of previous results.

Based on the sampled data and parameter estimates, the journey elements are inferred following the lines given in Section 4.3. About the matching probabilities, Figures 6(a) and 6(b) depict the numbers of hypothetical n_u^0 and feasible n_u train runs per trip, respectively. Figure 6(b) confirms that the postulate in the Bayesian analysis will be a good approximation. Figure 6(c) shows the maximum probability to take a train run among all n_u feasible train runs. The maximum probability for a tip to take a train is equal to 1 when there are only one feasible train. For the rest, the results show that most of the maximum probabilities to take a train are bigger than 70.66%, but one is 1×10^{-16} . on March 17th. This exception can be caused by the bias of AVL data measurement. Figure 6(d) calculates the dominance ratios among n_u while $n_u > 1$. Since the ratios are so small, smaller than 1×10^{-3} , it shows that there is only one train that is feasible while $n_u > 1$ as well.

A disaggregate analysis of passenger individual trip attributes is proposed by using the passenger trip model proposed in Section 4.3. Passenger trajectories are reproduced by the simple inference model based on previous estimated values. Figure 7 gives each observed trip and its referred result, Case 1 in the left column and Case 2 in the right column. To confront the real tap-out times, Figures 7(a1) and 7(a2) compare the inferred tap-out times (red circle) to the observed ones (blue asterisk). Since the differences (black cross) between them are very small, the agreement is very good, which indicates that the stochastic model has good

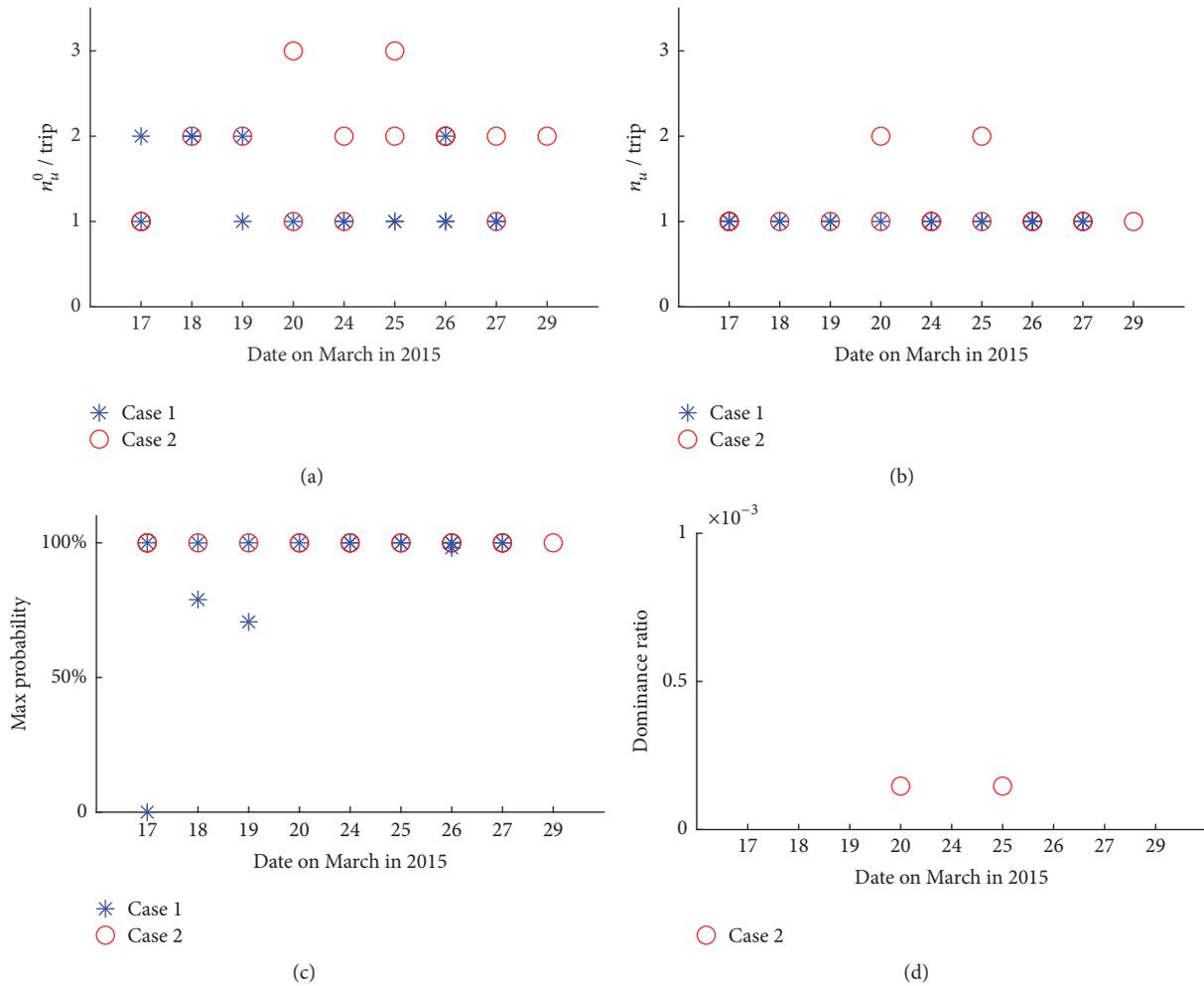


FIGURE 6: Run number and run probabilities per trip.

predictive ability. Figures 7(b1) and 7(b2) about positioning distances show that the total repositioning could be achieved in all but one occurrence. Lastly, Figures 7(c1) and 7(c2) illustrate the residual waiting times obtained by inference: the values are higher at La Défense (Case 1) than those at Vincennes (Case 2), certainly because Case 1 occurs mostly off peak while Case 2 occurs mainly at peak.

6. Conclusions and Perspectives

This section assesses the outreach and limitations of our model and points to directions for further research.

6.1. Summary and Outreach. This paper provided a stochastic model of passenger trip-making along a transit journey by urban rail line, with explicit representation of individual positioning along the boarding platform and the optimizing behavior to save on travel time for the rest of the trip.

The behavioral postulate was appropriate for passengers well aware of the trip conditions at their egress station. This fits well commuters—hence the vast majority of transit users at peak periods—and also customers availing themselves of “travel assistant” applications on their smartphones.

The stochastic model was easy to use in the perspective of simulation, as it followed the physical sequence of phases in a journey path (walking in, platform positioning and waiting, train riding, and walking out). It could readily be applied as a submodel in the frame of a traffic assignment model to a transit network.

While the simulation ability was demonstrated in the case study, the paper was primarily oriented to the estimation perspective: analytical formulae were given to characterize the statistical distributions of egress times and exit instants that stem from the set of modeling assumptions. The CDF and PDF formulae conveyed the influences of individual attributes and behavior (speed and target relocation distance), along with those of local conditions, i.e., in-station distances for the walking phases.

In the estimation perspective, we used the PDF formulae as likelihood functions for the model parameters. We put forward particular yet realistic enough specifications for the statistical distributions, so as to make numerical computation more tractable.

An application was carried out to an O-D pair of stations along a busy rail line in Paris. AFC and AVL data were extracted for the trips of an individual user over a two-week

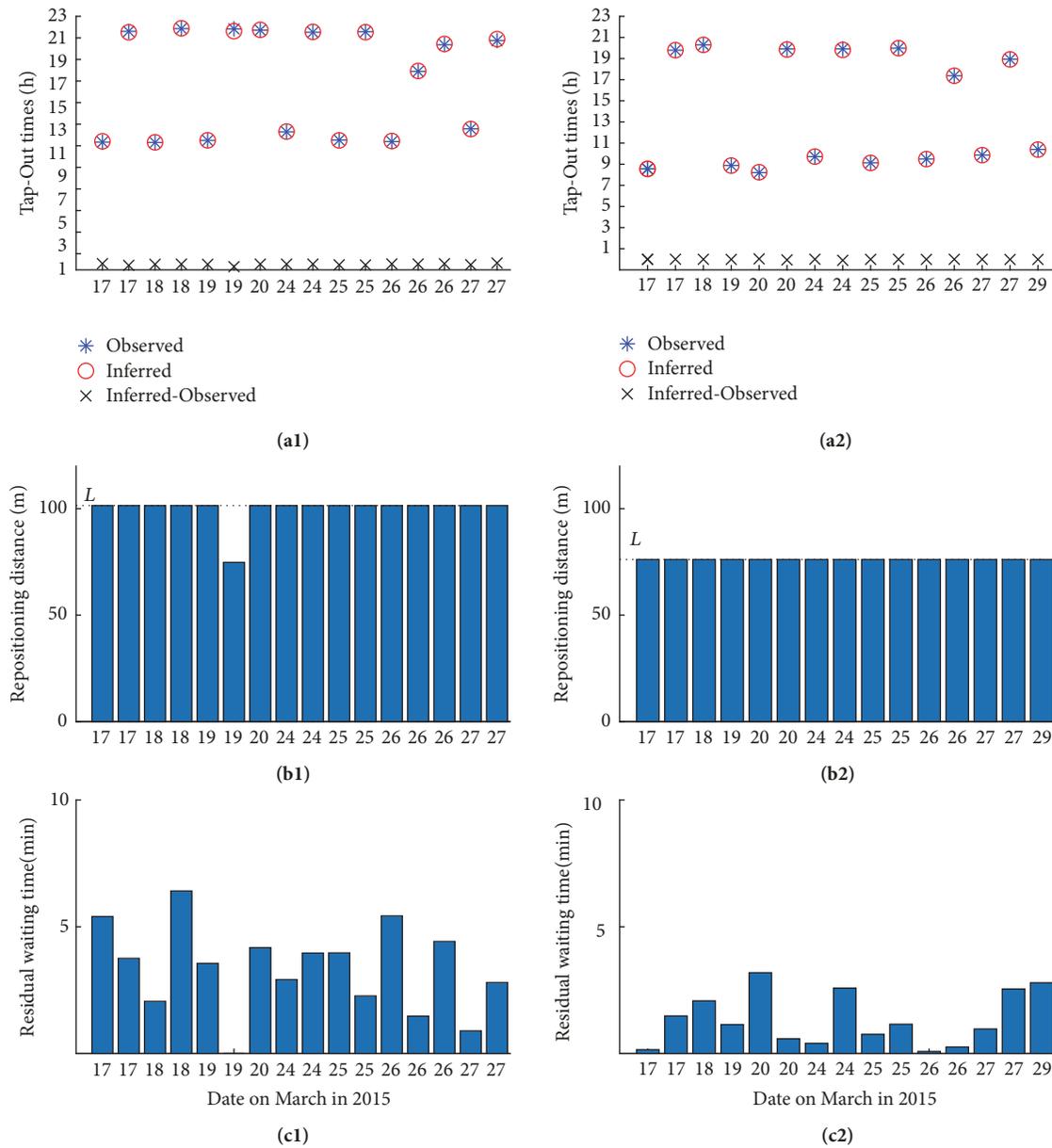


FIGURE 7: Inferred results about: (a) tap-out time; (b) repositioning distance; and (c) residual waiting time.

period. Valuable information was recovered from statistical estimation and posterior inference, notably a time saving of about 1 min owing to platform repositioning along 70 or 100 m, depending on the journey direction. This indicated that the estimation scheme was able to capture fine phenomena and also that the repositioning phenomenon had a limited importance on the journey travel time of this individual user.

This also demonstrated once again the positioning behavior of train users along boarding platforms, which was of interest to railway operators for passenger traffic management under severe crowding. The consequences for station layout and flow orientation were traced out in related work [19].

6.2. Potential Applications and Limitations. Where AFC and AVL datasets are available, our model can easily be applied to

estimate the distance and time components of users' journeys in a gate-to-gate setting which represents quality of service better than just the service quality of the train runs. The identification of positioning will make the estimation of the remaining time components more realistic and reliable. The individual behavior, as postulated and estimated on the basis of empirical data, can easily be simulated for users' journeys whatever the availability status of observations is, because the behavioral structure is endowed with replicability.

Despite its finesse, the stochastic model in its current version does not capture congestion phenomena: neither in-vehicle crowding or the potential restrictions (i.e., the probability of fail-to-board and the related issue of passengers "left behind" by trains), the crowding of platforms and its potential influence on individual positioning, nor the

TABLE 4: Estimated parameters' values by MLE in M1 and M2.

Parameter (Unit)	Range	M1		M2	
		Estimated value	Standard deviation	Estimated value	Standard deviation
M1: μ (m/s); M2: a (m/s)	(0, 4] and $a < b$	0.90	0.0023	0.58	0.07
M1: σ ; M2: b (m/s)	(0, 4]	0.20	0.0007	1.24	0.13
λ^+ (1/m)	(0, 2]	0.78	0.0048	0.32	0.18
B^+ (m)	(0, 150]	83.49	0.7314	72.32	6.81
λ^- (1/m)	(0, 2]	0.13	0.0002	0.18	0.02
B^- (m)	(0, 550]	204.30	0.7998	264.46	26.66
Log-likelihood function		-109.85		-138.18	

TABLE 5: Indicators' mean values in M1 and M2.

Parameter (Unit)	Range	M1		M2	
		Estimated mean value	Standard deviation	Estimated mean value	Standard deviation
$E(\nu)$ (m/s)	(0, 4]	0.90	0.04	0.91	0.19
$E(d^+)$ (m)	[4, 325]	84.77	1.29	75.41	3.09
$E(d^-)$ (m)	[4, 525]	212.22	7.93	270.32	5.66

crowding of platform access points and especially egress points, which may entail queuing and delay among exiting passengers.

6.3. Further Developments. So the consideration of crowding phenomena makes up a first direction for further research on passenger behavior along a rail journey.

A second direction is to devote more attention to the in-station phases, especially so for vertical pedestrian elements that influence individual speed under congestion as well as free-flow conditions. While these issues are well known in the micro-simulation of pedestrian traffic (cf. the Legion and Viswalk modeling software, among others), their estimation on the basis of AFC and AVL data is an open issue.

A third direction for research is to extend the stochastic model with platform repositioning to more complex trip patterns that involve transfers: this is our next objective.

Lastly, more detailed data of users' trajectories are available from smartphones owing to applications that monitor geolocation data from one or several sources—GPS, GSM, or beacons Wifi or Bluetooth. Indeed, location data collected every second say and with fine GPS or Galileo accuracy constitute ideal material for the refined analysis of passenger trip-making. Such research remains to be done for large underground transit stations, where satellite or beacon signals are impeded or modified by local layout—corridors, walls, floors, and ceilings.

Appendix

The results of previous models without passenger longitudinal walking distance in [23] are presented. The two stochastic models M1 and M2 are normal-distributed and uniform-distributed walking speeds models. The tested datasets are all passengers' trips from Vincennes to La Défense on March 16th 2015. Table 4 shows the parameters' values and Table 5 introduces the indicators' mean values.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the Research and Education Chair on “the Socio-Economics and Modeling of Urban Transit,” operated by Ecole des Ponts ParisTech (ENPC) in partnership with the Mobility Authority in the Paris area (IdFM, Île-de-France Mobilités, ex STIF), to whom the authors are grateful. The authors also thank the Autonomous Operator of Parisian Transit (RATP) who provided the train data.

References

- [1] TCQSM. Transit Capacity and Quality of Service Manual (TCQSM). 3rd edition. TRB's Transit Cooperative Highway Research Program (TCRP) Report 165, 2013.
- [2] V. Benezech and N. Coulombel, “The value of service reliability,” *Transportation Research Part B: Methodological*, vol. 58, pp. 1–15, 2013.
- [3] H. Niu, X. Zhou, and R. Gao, “Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints,” *Transportation Research Part B: Methodological*, vol. 76, pp. 117–135, 2015.
- [4] F. Leurent, E. Chandakas, and A. Poulhès, “A traffic assignment model for passenger transit on a capacitated network: Bi-layer framework, line sub-models and large-scale application,” *Transportation Research Part C: Emerging Technologies*, vol. 47, no. 1, pp. 3–27, 2014.
- [5] J. Chan, *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data [M, thesis]*, Department of Civil and Environmental Engineering, 2007.

- [6] M. Friman, "Affective dimensions of the waiting experience," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 13, no. 3, pp. 197–205, 2010.
- [7] Y. Fan, A. Guthrie, and D. Levinson, "Waiting time perceptions at transit stops and stations: Effects of basic amenities, gender, and security," *Transportation Research Part A: Policy and Practice*, vol. 88, pp. 251–264, 2016.
- [8] M. M. Rahman, S. C. Wirasinghe, and L. Kattan, "Users' views on current and future real-time bus information systems," *Journal of Advanced Transportation*, vol. 47, no. 3, pp. 336–354, 2013.
- [9] C. Brakewood, G. S. Macfarlane, and K. Watkins, "The impact of real-time information on bus ridership in New York City," *Transportation Research Part C: Emerging Technologies*, vol. 53, pp. 59–75, 2015.
- [10] G. S. Mishra, P. L. Mokhtarian, and K. F. Widaman, "An empirical investigation of attitudes toward waiting on the part of Northern California commuters," *Travel Behaviour and Society*, vol. 2, no. 2, pp. 78–87, 2015.
- [11] M. A. Ramli, V. Jayaraman, K. H. Chee et al., "Improved estimation of commuter waiting times using headway and commuter boarding information," *Physica A: Statistical Mechanics and its Applications*, vol. 501, pp. 217–226, 2018.
- [12] J. B. Ingvardson, O. A. Nielsen, S. Raveau, and B. F. Nielsen, "Passenger arrival and waiting time distributions depending on train service frequency and station characteristics: A smart card data analysis," *Transportation Research Part C*, 2018.
- [13] G. S. Becker, "A theory of the allocation of time," *The Economic Journal*, vol. 75, no. 299, p. 493, 1965.
- [14] D. Lois, A. Monzón, and S. Hernández, "Analysis of satisfaction factors at urban transport interchanges: Measuring travellers' attitudes to information, security and waiting," *Transport Policy*, 2017.
- [15] A. Bouzir, H. Bouguerra, S. Rjiba, and S. Benammou, "Passengers' activities in multimodal transportation stations," in *Proceedings of the International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA)*, pp. 76–81, 2017.
- [16] A. Millonig, M. Sleszynski, and M. Ulm, "Sitting, waiting, wishing: Waiting time perception in public transport," in *Proceedings of the 15th International IEEE Conference on Intelligent Transportation*, 2012.
- [17] N. Ohmori, T. Hirano, and N. Harata, "Passengers' waiting behavior at bus and tram stops," *Traffic and Transportation Studies*, pp. 520–531, 2008.
- [18] M. van Hagen, *Waiting Experience at Train Station [Doctoral thesis]*, Eburon Uitgeverij, 2011.
- [19] Z. Christoforou, P. Collet, B. Kabalan et al., "Influencing Longitudinal Passenger Distribution on Railway Platforms to Shorten and Regularize Train Dwell Times," *Transportation Research Record*, vol. 2648, pp. 117–125, 2017.
- [20] Z. Yizhou, *Real Time Crowding Information (RTCI) Provision [Ph.D. thesis]*, KTH, 2015.
- [21] F. Leurent and X. Xie, "Exploiting smartcard data to estimate distributions of passengers' walking speed and distances along an urban rail transit line," *Transportation Research Procedia*, vol. 22, pp. 45–54, 2017.
- [22] X. Xie and F. Leurent, "Estimating distributions of walking speed, walking distance, and waiting time with automated fare collection data for rail transit," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2648, no. 2, pp. 134–141, 2017.
- [23] X. Xie and F. Leurent, "Comparison of passenger walking speed distribution models in mass transit stations," *Transportation Research Procedia*, vol. 27, pp. 696–703, 2017.
- [24] F. Leurent and X. Xie, "On passenger repositioning along station platform during train waiting," *Transportation Research Procedia*, vol. 27, pp. 688–695, 2017.
- [25] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: a literature review," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 557–568, 2011.
- [26] T. Kusakabe and Y. Asakura, "Combination of smart card data with person trip survey data," in *Public Transport Planning with Smart Card Data*, pp. 73–92, CRC Press, 2016.
- [27] Y. Sun and R. Xu, "Rail transit travel time reliability and estimation of passenger route choice behavior," *Transportation Research Record*, no. 2275, pp. 58–67, 2012.
- [28] Y. Sun and P. M. Schonfeld, "Schedule-based rail transit path-choice estimation using automatic fare collection data," *Journal of Transportation Engineering*, vol. 142, no. 10, Article ID 04015037, pp. 36–51, 2016.
- [29] J. D. Schmöcker, *Dynamic capacity constrained transit assignment [Ph.D. thesis]*, Imperial College London, London, UK, 2006.
- [30] Y. Zhu, *Passenger-to-train assignment model based on automated data [M. Thesis]*, S.M. Massachusetts Institute of Technology, 2014.
- [31] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "A probabilistic Passenger-to-Train Assignment Model based on automated data," *Transportation Research Part B: Methodological*, vol. 104, pp. 522–542, 2017.
- [32] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "Inferring left behind passengers in congested metro systems from automated data," *Transportation Research Procedia, Transportation Research Part C: Emerging Technologies*, vol. 23, Supplement C, pp. 362–379, 2017.
- [33] Y. Zhu, *Passenger-to-Itinerary Assignment Model Based on Automated Data [Ph.D. thesis]*, 2017.
- [34] D. Hörcher, D. J. Graham, and R. J. Anderson, "Crowding cost estimation with large scale smart card and vehicle location data," *Transportation Research Part B: Methodological*, vol. 95, pp. 105–125, 2017.
- [35] A. Sumalee, Z. Tan, and W. H. K. Lam, "Dynamic stochastic transit assignment with explicit seat allocation model," *Transportation Research Part B: Methodological*, vol. 43, no. 8-9, pp. 895–912, 2009.
- [36] F. Leurent, "Modelling seat congestion in transit assignment," in *Proceedings of the 17th ISTTT conference*, 2006.
- [37] F. Leurent, "A model of travel and stop times, with application to the estimation of travel time functions," in *Proceedings of the 7th Meeting of the EURO Working Group on Transportation*, Roma, Italy, 2000.
- [38] F. Leurent, *Modèle désagrégés du trafic. Book "Les collections de l'inrets, Outils et Méthodes N10"*, 2001.
- [39] A. M. Wahaballa, F. Kurauchi, T. Yamamoto, and J. Schmöcker, "Estimation of Platform Waiting Time Distribution Considering Service Reliability Based on Smart Card Data and Performance Reports," *Transportation Research Record*, vol. 2652, pp. 30–38, 2017.
- [40] D. Aigner, C. A. Lovell, and P. Schmidt, "Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics*, vol. 6, no. 1, pp. 21–37, 1977.

- [41] B. Almadani, S. Khan, R. Tarek et al., "Automatic Vehicle Location and Monitoring System based on Data Distribution Service," *Procedia Computer Science*, vol. 37, pp. 127–134, 2014.
- [42] Xie. Xiaoyan and Fabien. Leurent, "Spatial and temporal analyses about distributions of walking speed and distances," in *Proceedings of the 44th European Transport Conference*, Barcelona, Spain, 2016.
- [43] G. G. Judge, R. C. Hill, W. E. Griffiths et al., *Introduction to The Theory And Practice of Econometrics*, Wiley, 1st edition, 1987.
- [44] X. Xie, F. Leurent, and V. Aguiléra, "Commuter's individual travel time estimation: using AFC data of Paris transit system," in *COST Action TU1004 Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, Paris, France, 2015.
- [45] J. Zhao, M. Frumin, N. Wilson, and Z. Zhao, "Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data," *Transportation Research Part C: Emerging Technologies*, vol. 34, pp. 70–88, 2013.
- [46] J. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 376–387, 2007.
- [47] V. Aguiléra, S. Allio, V. Benezech, F. Combes, and C. Milion, "Using cell phone data to measure quality of service and passenger flows of Paris transit system," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 198–211, 2014.
- [48] RATP, 2015, <http://www.ratp.fr>.
- [49] Omnil, *Enquête globale de transport, La mobilité en Île-de-France: la marche*, 2013, <http://www.omnil.fr>.

