How Do Individual Walk Lengths and Speeds, Together with Alighting Flow, Determine the Platform Egress Times of Train Users?

Fabien Leurent, Kang Liang

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 3633293, 22 pages
https://doi.org/10.1155/2022/3633293

Research Article

How Do Individual Walk Lengths and Speeds, Together with Alighting Flow, Determine the Platform Egress Times of Train Users?

Fabien Leurent and Kang Liang

Ecole des Ponts ParisTech, CIRED, Territorial Mobility Chair, Paris, France

Correspondence should be addressed to Fabien Leurent; fabien.leurent@enpc.fr

Received 13 January 2022; Revised 13 April 2022; Accepted 12 May 2022; Published 19 July 2022

Academic Editor: Gabriella Mazzulla

Egress times of railway passengers from train alighting up to station exit typically amount to some tens of seconds, but with much variability even at the train level. Here, we first model the egress time as the ratio of the walk length to the preferred walk speed, under free-flow conditions. Then, we model the possible occurrence of congestion among the users alighting from a train as a traffic bottleneck affecting those passing at a "queue focal point" during a "queued time interval." Analytical formulas are provided for the CDF and PDF of egress times, covering the free-flow case and the congested case. Their computation is straightforward for bivariate Gaussian length-speed walk pair. A maximum-likelihood method is developed, together with a quick estimation procedure. A case study of four contrasted trains serving an urban mass transit station in Paris is reported. One train experienced free-flow alighting conditions, whereas each of the other three had its own bottleneck. The MLE method enabled us to recover all parameters but one, due to an issue of identifiability: the solution was to take the mean walk speed as exogenous.

1. Introduction

The rail mode is the best suited to the mass transit of passengers in big cities, as it can provide a high level of service to very large numbers of passengers (TCQSM, 2013). The busiest lines can flow up to 100,000 passengers per hour and per direction on their trunk links: this is achieved for instance by the RER A line in the Greater Paris Area (RER for Regional Express Railways), owing to a peak frequency of 30+ trains per hour, times large train capacities of about 3,000 passengers (using duplex trains about 210 m long).

Train passenger loads give rise to proportional flows of alighting and boarding passengers at the stations along the line. The boarding flows may experience specific congestion, notably so when some users are not able to board the first train that services the station just after their arrivals—thus being "left-behind" and having to wait for the next train [1]. User exposure to such boarding congestion can be mitigated by selecting one's waiting position along the platform: this position will give rise to the user's longitudinal position on board. Indeed, the train length size is purported to supply passenger capacity all along the train, supposedly in a homogenous way to limit crowding and make the best use of seats. The resulting spreading of passengers along the train will also exert some less direct consequences: it influences the length to walk in the station of alighting and in turn the platform egress time up to the station exit point. Furthermore, as the alighting passengers all egress from the train at about the same instant, specific congestion is likely to occur and to increase the platform egress times.

Up to now, the alighting traffic has been studied in two ways. First, in the perspective of network planning and passenger route choice, the egress times of individual passengers from the train alighting position to the station access point have been modeled in macroscopic models of traffic simulation, static or dynamic, as an average time exogenously specified for each station platform and train service (Cf. [2]). Second, in the perspective of railway operations,

the flow volume and platform clearance times have been studied using microscopic simulation models of pedestrian traffic (e.g., [3]).

This article is focused on the platform egress times of train passengers. We are interested in the egress time as a physical variable involving both a space length to walk and a pedestrian speed. The physical variable is subjected to important variations among the train alighting passengers, due to the different alighting positions along the long platform as well as to the distribution of pedestrian speeds [4]. The article deals with the following three research questions. First, what are the influences of the on-board position and the pedestrian walking speed onto the passenger egress time? Second, when congestion occurs among the alighting passengers, what are its specific effects on their respective egress times? Third, what information on on-board positions and pedestrian speeds can be gained from the observation of platform egress times?

To answer the questions, we put forward a physical and stochastic model of platform egress times for transit users, which involves a statistical distribution of alighting positions and a statistical distribution of free-flow walking speeds. From these assumptions and the alighting flow volume, we derive the possibility of crowding and analyze its consequences on the individual egress times. Analytical properties are established for the statistical distribution of the egress times of a given train at a given station. By assuming either Gaussian or log-normal distributions for the alighting position and walk speed pair, closed-form formulas are established for the probability density function (PDF) of the egress time.

Turning to the issue of traffic observation, we apply the stochastic model to the estimation of train alighting positions and pedestrian walk speeds on the basis of egress time data collected mostly from smart cards (automated fare card or AFC system) together with train arrival times collected by an automated vehicle location (AVL) system. The PDF function is used to constitute a likelihood function according to observed egress times for a statistical population at the train level: by maximum-likelihood estimation, the parameters of the ex-ante distributions of alighting positions and walk speeds can be recovered. As an instance of application, we study the case of the Noisy-Champs station on the eastern part of the RER A line in Paris.

The rest of the article is organized into six sections. Section 2 reviews the related academic literature. Section 3 introduces the physical and stochastic model. Section 4 provides some distributional assumptions and derives specific formulas to compute the PDF and CDF of the Gaussian and log-normal models. Section 5 develops the estimation methodology, from a simple scheme to maximum-likelihood estimation. Then, Section 6 addresses the case study: after describing the traffic scene and the datasets, we provide the estimation results for four contrasted trains. Lastly, Section 7 concludes by stating the article's contribution and pointing to further developments.

2. Related Work

The topics of train length, station platform paths for train users, and passenger egress times have been dealt with for purposes of either train and platform design and the management of platform pedestrian traffic (§2.1), of traffic modeling of pedestrian paths and egress times (§2.2), or the stochastic modeling and statistical analysis of users' transit travel times (§2.3).

2.1. Train Length: From Principle to Effects. According to the Transit Capacity and Quality of Service Manual (TCQSM) [5], the passenger capacity of a railway line involves two factors in a multiplicative relation: line capacity and train capacity [6]. Line capacity is the maximum number of trains that can be operated on a line during a given period: it is typically measured in trains per hour and per track. Train capacity is the maximum number of passengers that can be accommodated with sufficient comfort on board a given train. The longer the train, the higher its passenger capacity: the respective capacities of all cars making up the train add up to the overall train capacity, in the same way as their respective lengths add up to the overall train lengths (up to that of connecting elements). Railway operators are accustomed to schedule single or double trains depending on the expected traffic load.

Not only does the train length enable it to carry a proportional number of passengers, but it is also convenient to provide a proportional number of doors to be used as channels for passengers boarding and alighting. The respective boarding and alighting throughput capacities depend on doorway width (TCQSM). They add up along the doors and the cars to the overall train boarding and alighting capacities. The larger these capacities, the shorter the time required for train dwelling; thus, the larger the line capacity in trains and in turn in passengers [7].

To make the best use of the train capacities—on board, at boarding, and at alighting, it is desirable to split the passenger flows evenly along the train–both on board and on the platforms prior to boarding. In fact, each train making a particular run will have a particular passenger load at each station it will serve, both in volume and in longitudinal distribution. High volume together with density peak in that distribution gives rise to the critical door issue, that is, the door putting the highest requirement on the dwelling time. This is why Liu et al. [8] modeled the number of waiting passengers at each of the 24 waiting positions on line 4 in the Pinganli metro station, Beijing. Using a multinomial logit discrete choice model, these authors postulated that the utility of a given position stems from the expected on-board density and the length from the station entry point to that position. Hoogendoorn et al. [9] further analyzed the effect of spatial densities on passengers' behaviors on the basis of a macroscopic fundamental diagram for pedestrian traffic.

The heterogeneity of passenger loads and crowding conditions depending on train cars—that is, longitudinal positions—has motivated the design of traffic management schemes (TMSs) to spread the flow of users waiting along the platform in a shape adapted to that of the passenger distribution along the arriving train. Zhang et al. [10] tested the provision of crowding information to the users of a metro line in Stockholm, Sweden. Their results indicate that users

can react to the crowding information and adapt their positioning strategy while waiting. Christoforou et al. [3] designed alternative information strategies and assessed their respective effects using a pedestrian microsimulation model. Related pedestrian TMSs for railway platforms include the choreography of alighting and boarding flows: while most platforms use the same platform and train side for both movements, using a dual-sided platform enables the operator to clear out the alighting flow more quickly, as their side is not impeded by boarding candidates, and reciprocally to start the boarding phase sooner and make it more fluent and quicker since it will not be impeded by the alighting flow. For a one-sided platform in a line metro transfer station in Santiago, Chile, Muñoz et al. [11] described the design and simulation assessment of a strategy compelling the users transferring there to come out of the first part of the train, by implementing a one-way gate in the platform width. There again the primary objective was to minimize the train dwelling time. The platform clearing time was also considered a complementary performance indicator.

2.2. On Passenger Paths and Walk Times in Traffic Models. While the platform clearing time can be seen as a maximum time for passenger alighting and exiting the platform, it is of interest mainly to the line operator—as are the train critical door and the train dwell time. On an individual basis, the train users are more interested in their own egress times. These times have long been modeled on an average basis in traffic assignment models to transit networks. Such traffic assignment models are especially purported to simulate individual users along their transit network paths between their origin and destination points [2]. Such path is composed as a sequence of nodes and links along the network, and it travel time is decomposed accordingly. In the first and second generations of transit traffic assignment models (from [12] to [13]) "in-vehicle links" typically go from one station to the next one along the transit line and there is one "in-vehicle node" per station and line direction to depict the dwelling operation; passenger paths involve such transit links for their in-vehicle rides, together with walk links for boarding, alighting, transfers, and station access and egress. By modeling the alighting path as one link, only the average egress time has been modeled. Walk lengths and speeds could be modeled as underlying random variables yielding distributed egress times in some kind of stochastic traffic assignment model, but to our knowledge, there has not been any such modeling attempt, neither for alighting nor for boarding.

In contrast, the influence of congestion onto the wait-for-boarding times has attracted several research contributions in the field of transit assignment modeling. In their macroscopic dynamic assignment models, Poon et al. [14] and Hamdouch and Lawphongpanich [15] addressed capacitated boarding as a traffic bottleneck under FIFO queuing discipline, yielding some delay when the boarding flow volume is in excess of the train residual capacity. The latter authors also suggested modeling the alighting flow in relation to some platform exit capacity, yet without providing an associated mathematical formulation. In both contributions, the platform lengths have not been considered explicitly. Longitudinal distribution has remained implicit in the macroscopic theory of dynamic transit traffic assignment up to Hänseler et al. [16] who introduced longitudinal detail of train platform and other "pedestrian elements" in a macroscopic framework that goes consistent from the station level to the line level and up to the network level. As their model deals with longitudinal positions of transit users in trains and on platforms in an endogenous way, the on-board positions as well as the alighting walk paths and the associated egress times are both distributed and endogenous. The model can also encompass different kinds of traffic congestion: as an instance, the authors considered macroscopic fundamental diagrams of pedestrian traffic to relate local walk speeds along the platform to the local pedestrian densities.

Such fine representation of platform issues in the frame of network traffic assignment bridges much of the gap between the previous generation of macroscopic models and the stream of dynamic microsimulation of transit traffic. At the platform level, Zhang et al. [17] devised a cellular automata microsimulation model of the alighting and boarding processes of passengers, revealing the potential mutual influences between passengers, such as the desire to board and pressure from behind. Haghani and Sarvi [18] used an error-component mixed logit model to analyze the differences of passenger's route choice between an emergency case and a base case. Ji et al. [19] studied the pedestrian choice between stairway and escalator in the transfer station by using a logit model, taking into account quantitative factors and nonquantitative factors. Christoforou et al. [3] modeled the Noisy-Champs railway station in eastern Paris using a crowd dynamics model so as to simulate the effects of passenger orientation strategies on the wait-for-boarding positions. These microsimulation models deal with a specific traffic issue of pedestrian movement on platforms and capture the different influencing factors; they represent the accessing and exiting points, each waiting position, and its corresponding door through which passengers can board in or alight from the train. At the network level, the microscopic model "BusMezzo" of Cats [20] is purported to simulate the bus and train events and on-board crowding for all lines in a network, but it does not consider the length of train vehicles, nor the issue of multiple doors. Specialized microsimulation traffic models including VISSIM, Legion, and MassMotion have been developed on a commercial basis and enable for microscopic detail in both time and space, over a whole network.

2.3. Stochastic Models and Statistical Analysis. Thus, there are some recently developed models of transit traffic simulation that consider spatial detail explicitly along platforms and trains. Assuming fine spatial description on both sides of macro- vs. microsimulation, a salient feature still differentiating macro- and micromodels pertains to the congestion model—based on either a macroscopic law or the dynamic simulation of interactions between entities such as

passengers, vehicle elements, and platform elements. Stochastic modeling constitutes another bridge between micro- and macromodeling: in a stochastic traffic model, physical traffic variables such as length, speed, and time are modeled as random variables with specific distributions. Stochastic modeling therefore lays the ground for the statistical analysis of traffic data.

Stochastic modeling of transit paths was pioneered by Sun et al. [21]: following the path topological decomposition in network traffic assignment models, they analyzed the user time along a transit path as a four-tier sequence of (i) access, (ii) wait, (iii) ride, and (iv) egress. By postulating a specific statistical distribution for each tier time depending on its own physical conditions, the authors provided an estimation method for the parameters of all distributions. Their method was applied to a dataset of individual travel times observed between two validation gates (AFC records of tap-in and tap-out pairs), complemented by the related times of train arrival at and departure from the stations of access and egress (AVL data). Further on, Zhu et al. [4] modeled the access and egress times as the ratios of walk lengths divided by walk speeds, so as to estimate the distribution of pedestrian walk speeds in railway stations. A key element in their model is the passenger-to-train assignment probability. In a parallel work, Leurent et Xie [22] related the walk speed at the individual level on both sides of the ride (access and egress): they succeeded to estimate the length distributions together with the speed distribution owing to specific distributional assumptions of shifted exponential lengths together with uniformly distributed speeds. Gaussian distributed walk speeds were also considered in Xie and Leurent [23].

In this stream of passenger traffic stochastic modeling, boarding congestion was addressed by Zhu et al. [24] who modeled the left-behind phenomena as the failure-to-board one or more trains serving the station: the number of missed trains was modeled as a random variable composed at two levels. Leurent and Jasmin [25] provided a physical model of failure-to-board, postulating FIFO among the awaiting users. Hörcher et al. [26] addressed the influence of on-board crowding conditions onto the line choice of individual users under a specific subnetwork configuration: they focused on the passenger egress times first to assign different usage probabilities to the successive trains on each line, in a Bayesian way based on a postulated PDF for egress times, and then to differentiate between the two lines, again on the basis of Bayesian probabilities. This Bayesian approach involves the time of user exit and that of train departure to obtain the egress time conditionally to that train.

Up to now, no consideration has been paid in this stream to egress congestion or to traffic bottlenecks on either the boarding or alighting sides. Leurent and Xie [27] modeled the on-board positions in relation to both the platform entry point, in the access station, and, in the egress station, the platform exit point, together with an individual walk speed maintained on both platforms. This corresponds to free-flow walking conditions unaltered by any kind of congestion on the access and egress sides.

Overall, the stochastic modeling of individual egress times, possibly influenced by congestion in bottleneck form, with distributed walk lengths and speeds, is an original research topic.

The notation table is as follows:

(i) u: an individual user

(ii) w: free-flow walk speed, with CDF W

(iii) ℓ_I: walk length of individual user from train alighting point to platform intermediary point

(iv) ℓ_e: length from platform intermediary point to station exit point

(v) ℓ: individual walk length, with CDF S_w conditionally to w

(vi) τ: walk egress time along ℓ

(vii) ℓ^*: queue focal point

(viii) $[\tau_1^*, \tau_2^*]$: time interval of queuing at ℓ^*

(ix) v^*: queue moving speed

(x) t^*: queued time from ℓ^* to station exit point

(xi) $k \in \{1, 2, 3\}$: index of user subset U_k, respectively, before, during, and after queuing episode at ℓ^*, with associated probability P_k

(xii) T_k: CDF of effective egress time conditionally to k, where T denotes the unconditional CDF

(xiii) A number of alighting users, for train-PIP-SAP triple

3. Physical and Stochastic Model

3.1. Platform Geometry and Walk Lengths. Each station platform on a given railway line has its own geometry. As a spatial object of area type, it has a long, rectangular shape, mirroring the train lengths and the straightness of the infrastructure track. Its longitudinal dimension, typically in the range from 100m to 200m, gives rise to relative longitudinal positions on the platform and in turn on the trains that dwell there. By contrast, the platform width is relatively narrow, typically in the 5m–10-m interval: it is designed to accommodate the flows of alighting passengers and of incoming passengers that wait for train arrivals and constitute waiting stores, yet in a scarce way to spare the urban space. Let us define an intermediary point along the platform, say PIP for platform intermediary point, typically at the dwelling point of the train head (or tail) endpoint.

The platform is endowed with its own points for pedestrian access and egress, each one with a specific longitudinal abscissa with respect to the origin point. These points may be called pedestrian flow injectors, or platform funnels. Let us call them "platform egress points" (PEPs) to focus on the alighting flow.

Considering now the station, it has its own points of passenger access from, and egress to, the outer world: let us call them station access points (SAPs). As for SAPs, we typically consider a point equipped with ticket and card validation gates.

Each platform egress point is connected to one or several SAPs by way of a pedestrian path.

To an alighting passenger, the length to walk from train alighting to SAP, say ℓ, adds up that from PIP to SAP, say ℓ_e, and the length $\ell\prime$ on the platform from the alighting position to the PIP. We shall denote this decomposition as

$$\ell = \ell_e + \ell'. \tag{1}$$

As railway station platforms have long, narrow shapes, for any pair of points along the platform, we shall assimilate the walked length and the longitudinal difference in abscissas between the points (Figure 1). A typical PIP situation is at one endpoint of the platform. For a PIP situated at some intermediary point, the correspondence between ℓ and train alighting positions would be 1 : 2 instead of 1 : 1. For ease of discussion, we shall hereafter assume a PIP situation at a platform endpoint, so as to interpret $\ell\prime$ as the position along the train.

3.2. Free-Flow Pedestrian Speeds and Egress times. Given the PEP and the SAP, the egress length ℓ still depends on the alighting position. As a walk length, it gives rise to the egress time of the train user from his train alighting point to the SAP, denoted by t. The factor linking ℓ to t is the walking pace or its inverse the walk speed denoted by w: notionally,

$$t = \frac{\ell}{w} + t_e^0. \tag{2a}$$

This formula is an idealization. The time lag t_e^0 accounts for specific delay such as on taking an escalator. As for the walk time $\tau \equiv \ell/w$, we take w to be about constant along the length: in other words, it is a cruising speed at the individual level. Such walk speeds are distributed among the transit users, according to physical condition, age, luggage, etc. (TCQSM, 2013). Under this interpretation, w would be a free-flow speed: postulating the individual user not to be impeded by other pedestrians.

Let us denote W the CDF of unimpeded pedestrian speeds for a statistical population of transit users and $\dot{W}$ its PDF. Denote similarly S and $\dot{S}$ the CDF and PDF of walk lengths, respectively. Further statistical description involves the stochastic dependencies between s and w as random variables. Conditionally to walk speed w, we shall denote S_w as the CDF of walk lengths ℓ.

To sum up, under free-flowing, the walk egress time of an individual user is modeled as

$$\tau \equiv t - t_e^0 = \frac{\ell}{w}. \tag{2b}$$

Let also h_0 denote the instant of train arrival to dwell on the platform, taken homogenous among the alighting flow. For every user u, the instant h_u of passing the SAP and the egress time τ_u are straightforwardly related as follows:

$$\tau_u \equiv h_u - h_0. \tag{3}$$

In practice, some caution must be exerted to compare instants h_u and h_0. Assumedly, instants h_u of user exit are measured at the validation gates, with respect to the station clock say, while instants h_0 of train arrivals are measured by the automated vehicle location system, say the train clock. Between the two-timing systems, there may be some time lag, especially so if the AVL time is measured at a fixed sensor located upstream the station. We might denote $h_0' \equiv h_0 - \Delta h_0$ a corrected train arrival instant. Analogously, to focus on walk times, we would tend to decrease every h_u by t_e^0. In the rest of the article, we take such corrections as given and we consider that τ_u corresponds to the walk egress time $\tau \equiv \ell/w$. Under free-flow pedestrian traffic conditions, walk speed w is a free-flow one and τ has a statistical distribution (CDF T and PDF $\dot{T}$) that stems from the joint distribution of walk lengths and walk speeds.

Let us express the free-flow distribution functions of egress times: $T(x) := \Pr\{\tau \le x\} = \Pr\{\ell/w \le x\} = \Pr\{\ell \le w \cdot x\}$ since $w > 0$.

As

$$\Pr\{\ell \le w \cdot x\} = \Pr\{\ell \le w \cdot x \mid w\} \cdot \Pr\{w\} = S_w(w \cdot x)\Pr\{w\},$$

then

$$T(x) = \int S_w(w \cdot x) dW(w). \tag{4a}$$

This formula stands as the stochastic version of physical model (2). The PDF is obtained by partial derivation with respect to x: $\dot{T}(x) := \mathrm{d}/\mathrm{d}x T(x) = \int \mathrm{d}S_w(w \cdot x)/\mathrm{d}x \mathrm{d}W(w)$, so that

$$\dot{T}(x) = \int w \cdot \dot{S}_w(w \cdot x) dW(w). \tag{4b}$$

3.3. On Pedestrian Traffic Conditions and Queuing Phenomena. On exiting the train, the alighting passengers become pedestrians willing to get first to a PEP and then to an SAP. On their walk paths, they may be hindered by other pedestrians, be it due to conflicting directions or to different walk speeds and the inability to overtake a slower walker for lack of available width.

Three kinds of conflicting directions may arise. First, between alighting passengers and boarding candidates just out of the train: under severe platform crowding, the width available in front of a train door for train passengers to alight may be very scarce, leading to a train exit bottleneck: here we do not address that severe kind of congestion.

Second, between alighting passengers and incoming pedestrians just arriving on the platform to board the train or cross the platform from one point to another: as the alighting phase is concentrated in time, while the incoming flow is relatively homogenous over time, the likelihood of such conflicts is small and we shall neglect their effects.

The third kind of conflicting directions would arise among alighting passengers that would cross one another to go to different platform egress points. The potential outcomes will depend on whether the pedestrian density on the platform is low or high. Under low densities, such crossings are easy and the associated elemental delays are negligible. But under high densities, walking is slowed down and it would be very tedious for an individual pedestrian to manage a large number of directional conflicts. Then, the following collective behavior is likely to arise: the alighting

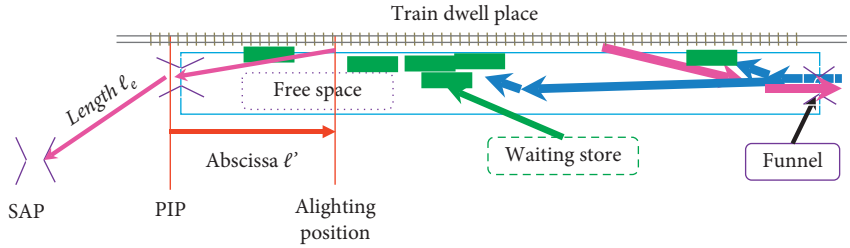

Figure 1: Platform layout.

flow "naturally" splits with respect to the closest egress points, so that each egress point will have its own "catchment area" along the sequence of train doors. In such a case, a targeted PEP will have a large number of egressing passengers and queuing is likely to occur among them.

We shall model neither the light kinds of pedestrian traffic hindrance nor the sharpest kind of overcrowding when train alighting is delayed for lack of space. We only model two traffic regimes of either free-flow conditions at the individual level or a queuing episode upstream of the SAP from some focal point ℓ^*.

As will be reported in the case study, alighting passengers targeting a given PEP, when their number is high, will constitute a pedestrian queue affecting all of its members. We shall model that kind of queue as a traffic bottleneck. Here are the main assumptions:

(A1) There is some focal point ℓ^* upstream the SAP, at which the queue beings at instant h_1^* soon after the instant h_0 of train arrival at the station. To instant, h_1^* corresponds an egress time $\tau_1^* \equiv h_1^* - h_0$.

(A2) At ℓ^*, the queue lasts from h_1^* to h_2^*, with related egress time $\tau_2^* \equiv h_2^* - h_0$. Users involved in it, after passing at ℓ^*, will walk from ℓ^* to SAP (will null length) at queued speed v^*, spending time $t^* = \ell^*/v^*$. Their set is denoted as $\tilde{U}_3$. The first and last egress times of them are $\tau_1 \equiv \tau_1^* + t^*$ and $\tau_2 \equiv \tau_2^* + t^*$.

(A3) Users unaffected by queuing are of two kinds, depending on whether they pass the SAP before $h_1 \equiv h_1^* + t^*$ or after $h_2 \equiv h_2^* + t^*$. Their respective sets are denoted U_1 and U_2, respectively.

(A4) Every un-queued user of the first kind U_1 has free-flow walk speed w from their alighting position ℓ up to SAP, hence free-flow egress time $\tau = \ell/w$. The (ℓ, w) pair also satisfies that $\{\ell - \ell^* \le w.\tau_1^*\}$, that is, either they alight downstream ℓ^* if $\{\ell \langle \ell^*\}$ or upstream it but before the beginning of queuing. Thus,

$$U_1 \equiv \{(\ell, w):\ \ell \le w \cdot \tau_1 \cap \ell - \ell^* \le w \cdot \tau_1^*\}. \tag{5a}$$

(A5) Every un-queued user of the second kind U_2 has free-flow walk speed w from their alighting position ℓ up to SAP, hence free-flow egress time $\tau = \ell/w$. The (ℓ, w) pair also satisfies that $\{\ell - \ell^* \rangle w.\tau_2^*\}$; that is, they pass at ℓ^* after the queue vanished from it. Thus,

$$U_2 \equiv \{(\ell, w):\ \ell w\rangle \cdot \tau_2 \cap\ \ell - \ell^* \rangle w \cdot \tau_2^*\}. \tag{5b}$$

Consequently, the set of users affected by the queue amounts to the complementary set:

$$U_3 \equiv U \setminus (U_1 \bigcup U_2). \tag{5c}$$

It holds that $\tilde{U}_3 \subset U_3$, which also contains users passing at ℓ^* before τ_1^* or after τ_2^* but who join the queue at some point from ℓ^* to SAP due to their own free-flow speed in relation to v^*.

3.4. The Effective Distribution of Platform Egress Times. For a train and a PEP giving rise to a queuing episode, within the statistical population of train-alighting users, the proportion of the three user groups is, respectively,

$$P_1 \equiv \Pr\{\ell \le w \cdot \tau_1 \cap\ \ell - \ell^* \le w \cdot \tau_1^*\}. \tag{6a}$$

$$P_2 \equiv \Pr\{\ell > w \cdot \tau_2 \cap\ \ell - \ell^* \rangle w \cdot \tau_2^*\}. \tag{6b}$$

$$P_3 \equiv 1 - P_1 - P_2. \tag{6c}$$

Users in $U_1 \bigcup U_2$ enjoy free-flow egress times $\tau = \ell/w$. From the definition of their sets and the non-negativity of walk speeds, we have that

In U_1: $\ell \le w.\tau_1$,

$$\tau \le \tau_1. \tag{7a}$$

In U_2: $\ell > w \cdot \tau_2$,

$$\tau > \tau_2. \tag{7b}$$

Then, users with $\tau \in [\tau_1, \tau_2]$ belong to U_3. Among that set, the users are subjected to some queuing effect. We further assume the following:

(A6) Most of the queued users are involved in a traffic bottleneck originating at ℓ^*, and the bottleneck exit flow rate is about constant (at some capacity value).

Then, among those queued users the egress time up to ℓ^* is distributed uniformly from τ_1^* to τ_2^*. In turn, as queued walk time t^* is assumed from ℓ^* to SAP, the egress time from alighting to SAP will be distributed

uniformly from τ_1 to τ_2. This applies strictly to the queued users in $\tilde{U}_3$, and we further assume that:

(A7) It extends to all users involved in U_3.

Thus, among U_3, the CDF of egress time is

$$T_3(x) = 1_{\{x \geq \tau_1\}} \cdot \frac{\min\{x, \tau_2\} - \tau_1}{\Delta\tau}, \tag{8a}$$

where $\Delta\tau \equiv \tau_2 - \tau_1$. By partial differentiation, the associated PDF is

$$\begin{aligned} \dot{T}_3(x) &\equiv \frac{d}{dx} T_3(x) \\ &= 1_{\{x \in [\tau_1, \tau_2]\}} \cdot \frac{1}{\Delta\tau}. \end{aligned} \tag{8b}$$

For users in U_1, the egress times are distributed with the following conditional CDF:$T_1(x) \equiv \Pr\{\tau \leq x \mid \tau = \ell/w \cap \ell \leq w \cdot \tau_1 \cap \ell - \ell^* \leq w \cdot \tau_1^*\}$, $= (1/P_1)\Pr\{\ell/w \leq x \cap \ell \leq w \cdot \tau_1 \cap \ell - \ell^* \langle w \cdot \tau_1^*\}$, so that

$$T_1(x) = \frac{1}{P_1} \Pr\{\ell \leq \min\{w \cdot x, w \cdot \tau_1, \ell^* + w \cdot \tau_1^*\}\}. \tag{9a}$$

The probability $\Pr\{\ell \leq y \cap w\}$ is simply $\Pr\{\ell \leq y \mid w\} \cdot \Pr\{w\}$, that is, $S_w(y)\Pr\{w\}$. By integration, it comes out that

$$T_1(x) = \frac{1}{P_1} \int S_w(\min\{w.x, w.\tau_1, \ell^* + w.\tau_1^*\}) dW(w). \tag{9b}$$

As P_1 does not depend on x , we get the associated PDF by straightforward differentiation: $\dot{T}_1(x) \equiv d/dx T_1(x) = (1/P_1) \int \mathrm{d}S_w(\min\{w \cdot x, w \cdot \tau_1, \ell^* + w \cdot \tau_1^*\})/\mathrm{d}x \mathrm{d}W(w) = (1/P_1) \int w \cdot \dot{S}_w(w \cdot x) \cdot 1_{\{x \leq \tau_1 \cap w \cdot x \leq \ell^* + w \cdot \tau_1^*\}} dW(w)$, so that

$$\dot{T}_1(x) = \frac{1}{P_1} 1_{\{x \leq \tau_1\}} \int w \dot{S}_w(w.x) 1_{\{w \cdot x \leq \ell^* + w \cdot \tau_1^*\}} dW(w). \tag{10a}$$

Let us distinguish two cases depending on whether $\{x \leq \tau_1^*\}$ or $\{x \rangle \tau_1^*\}$. In the former case, as walk speeds are positive and so is ℓ^* ; then, the condition $\{w \cdot x \leq \ell^* + w \cdot \tau_1^*\}$ holds true for every $w \geq 0$. In the latter case, the condition is equivalent to $w \cdot (x - \tau_1^*) \leq \ell^*$, that is, $w \leq \ell^*/(x - \tau_1^*)$. Thus, defining $\overline{M}_x(y) \equiv \int_y^\infty w \dot{S}_w(w \cdot x) dW(w)$, $\dot{T}_1(x) = (1_{\{x \leq \tau_1\}}/P_1)[1_{\{x \leq \tau_1^*\}} \int w \dot{S}_w(w.x) dW(w) + 1_{\{x \rangle \tau_1^*\}} \int_0^{\ell^*/(x - \tau_1^*)} w \dot{S}_w(w.x) dW(w)], = (1_{\{x \leq \tau_1\}}/P_1)[1_{\{x \leq \tau_1^*\}} \overline{M}_x(0) + 1_{\{x \rangle \tau_1^*\}} (\overline{M}_x(0) - t\overline{M}_x n(\ell^*/(x - \tau_1^*)))]$ and finally

$$\dot{T}_1(x) = \frac{1_{\{x \leq \tau_1\}}}{P_1} \left[\overline{M}_x(0) - 1_{\{x \rangle \tau_1^*\}} \overline{M}_x\left(\frac{\ell^*}{x - \tau_1^*}\right) \right]. \tag{10b}$$

Analogously, for users in U_2^*: the CDF from above, $\overline{T}_2(x) \equiv 1 - T_2(x)$, satisfies that $\overline{T}_2(x) = \Pr\{\tau \rangle x \mid \tau = \ell/w \cap \ell \rangle w \cdot \tau_2 \cap \ell - \ell^* \rangle w \cdot \tau_2^*\}$, $= (1/P_2)\Pr\{\ell \rangle w \cdot x \cap \ell \rangle w \cdot \tau_2 \cap \ell \rangle \ell^* + w \cdot \tau_2^*\}$, $= (1 / P_2)\Pr\{\ell \rangle \max\{w.x, w.\tau_2, \ell^* + w.\tau_2^*\}\}$, so that

$$\overline{T}_2(x) = \frac{1}{P_2} \int \overline{S}_w(\max\{w \cdot x, w \cdot \tau_2, \ell^* + w \cdot \tau_2^*\}) dW(w). \tag{11a}$$

As P_2 does not depend on x , we get the associated PDF by straightforward differentiation: $\dot{T}_2(x) \equiv \mathrm{d}T_2(x)/\mathrm{d}x = -\mathrm{d}\overline{T}_2(x)/\mathrm{d}x = (1/P_2) \int (dS_w)(\max\{w.x, w.\tau_2, \ell^* + w.\tau_2^*\})/dx \, dW(w)$, $= (1_{\{x \rangle \tau_2\}}/P_2) \int w.\dot{S}(w.x).1_{\{w.(x - \tau_2^*) \geq \ell^*\}} dW(w)$. Finally, as $x - \tau_2^* \rangle 0$ when $x \rangle \tau_2$,

$$\dot{T}_2(x) = \frac{1_{\{x > \tau_2 \tau_2\}}}{P_2} \overline{M}_x\left(\frac{\ell^*}{x - \tau_2^*}\right). \tag{11b}$$

We are now able to express the overall CDF and PDF, denoted as $T_U(x)$ and $\dot{T}_U(x)$, respectively, by combining the conditional distributions according to the three cases: $T_U(x) \equiv \Pr\{\tau \leq x\} = \sum_{k=1}^3 \Pr\{\tau \leq x \cap u \in U_k\} = \sum_{k=1}^3 \Pr\{\tau \leq x \mid u \in U_k\} \cdot \Pr\{u \in U_k\} = \sum_{k=1}^3 P_k T_k(x)$, and analogously,

$$\dot{T}_U(x) = P_3 \dot{T}_3(x) + P_1 \dot{T}_1(x) + P_2 \dot{T}_2(x). \tag{12a}$$

Substituting, we obtain that

$$\dot{T}_U(x) = 1_{\{x \in [\tau_1, \tau_2]\}} \cdot \frac{P_3}{\Delta\tau} + 1_{\{x \leq \tau_1\}} \left[\overline{M}_x(0) - 1_{\{x \rangle \tau_1^*\}} \overline{M}_x\left(\frac{\ell^*}{x - \tau_1^*}\right) \right] + 1_{\{x \rangle \tau_2\}} \overline{M}_x\left(\frac{\ell^*}{x - \tau_2^*}\right). \tag{12b}$$

3.5. An Incomplete Congestion Model. In a preliminary approach, we defined the sets of users egressing before or after the queuing episode on the basis of incomplete conditions as follows:

$$\widetilde{U}_1 \equiv \{(\ell, w):\ \ell \le w \cdot \tau_1\}, \tag{13a}$$

$$\widetilde{U}_2 \equiv \{(\ell, w):\ \ell \rangle w \cdot \tau_2\}. \tag{13b}$$

Both definitions make no reference to focal point ℓ^*. They are consistent with the definition of $\widetilde{U}_3$, which is equivalent to $\widetilde{U}_3 \equiv U \backslash (\widetilde{U}_1 \bigcup \widetilde{U}_2)$.

As $U_1 = \{\ell \le w.\tau_1 \cap\ \ell \le \ell^* + w.\tau_1^*\}$, it imposes a stronger condition on lengths ℓ: it may occur that $\ell^* + w.\tau_1^* \langle w.\tau_1$ if $w(\tau_1 - \tau_1^*) \rangle \ell^*$, that is, if $w.t^* \rangle \ell^*$ and if when $w \rangle v^*$. So, users in $\widetilde{U}_1$ but with $\ell^* + w.\tau_1^* \le \ell \langle w.\tau_1$ and $w \rangle v^*$ do not belong to U_1.

On the other side, $U_2 = \{\ell \rangle w \cdot \tau_2 \cap\ \ell \rangle \ell^* + w \cdot \tau_2^*\}$ imposes a stronger condition on lengths ℓ than does $\widetilde{U}_2$: it may occur that $\ell^* + w.\tau_2^* \rangle w.\tau_2$ when $w(\tau_2 - \tau_2^*) \langle \ell^*$, that is, when $w.t^* \langle \ell^*$ hence when $w \langle v^*$. So, users in $\widetilde{U}_2$ but with $\ell^* + w.\tau_2^* \rangle \ell \ge w.\tau_2$ and $w \langle v^*$ do not belong to U_2.

In the incomplete congestion model, we would have the following:

$$\widetilde{P}_1 \equiv \Pr\{\ell \langle w \cdot \tau_1\}$$
$$\widetilde{P}_2 \equiv \Pr\{\ell \rangle w \cdot \tau_2\}$$
$$\widetilde{P}_3 \equiv \Pr\{\ell \in [w \cdot \tau_1, w \cdot \tau_2]\}$$
$$\dot{\widetilde{T}}(x) = 1_{\{x \in [\tau_1, \tau_2]\}} \cdot \widetilde{P}_3/\Delta\tau + 1_{\{x \langle \tau_1\}} \overline{M}_x(0) + 1_{\{x \rangle \tau_2\}} \overline{M}_x(0);$$

hence,

$$\dot{\widetilde{T}}(x) = \overline{M}_x(0) - 1_{\{x \in [\tau_1, \tau_2]\}} \left(\overline{M}_x(0) - \frac{\widetilde{P}_3}{\Delta\tau} \right). \tag{14}$$

It is easy to compare the incomplete congested PDF to the free-flow one in (4b): as $\overline{M}_x(0) = \int w \cdot \dot{S}_w(w \cdot x) dW(w)$, the two PDFs are identical out of $[\tau_1, \tau_2]$, that is, out of the queued interval, on which the free-flow density $\overline{M}_x(0)$ is replaced by its queued counterpart $\widetilde{P}_3/\Delta\tau$.

Between the incomplete and full congested models, the respective PDFs have similar queued parts up to the definition of P_3 instead of $\widetilde{P}_3$, whereas the free-flowing parts are subjected to subdomain restriction in the full model that considers only those speeds above $\ell^*/(x - \tau_k^*)$. These differences vanish when $\{\ell^* = 0\}$: the incomplete congestion model can be seen as a restricted version of the full congestion model such that $\{\ell^* = 0\}$, leading to $t^* = 0$ and $\tau_k^* = \tau_k$, so that $\widetilde{U}_k = U_k$ and $\widetilde{P}_k = P_k$.

3.6. Capacity Issues. Let us denote A as the total number of alighting passengers, for that train and PEP and SAP. Then, there are $A.P_3$ queued users that exit in time length $\Delta\tau$. Defining the flow rate capacity for that train, K_A, it holds that

$$K_A = \frac{A \cdot P_3}{\tau_2 - \tau_1} = A \frac{T_U(\tau_2) - T_U(\tau_1)}{\tau_2 - \tau_1}. \tag{15}$$

In the incomplete congestion model where queuing originates at point $\{\ell^* = 0\}$, we would expect the flowing regime to change from free-flow to queued at τ_1 in a smooth way, yielding equivalent flow rate on both sides of τ_1 between $A.\dot{T}_U(\tau_1)$ from below and K_A from above. This condition is equivalent to

$$\dot{T}_U(\tau_1) = \frac{T(\tau_2) - T(\tau_1)}{\tau_2 - \tau_1}. \tag{16}$$

Since $P_1 = T(\tau_1)$ and $P_2 = 1 - T(\tau_2)$ in the incomplete congestion model, yielding $P_3 = T(\tau_2) - T(\tau_1)$.

Under the bottleneck postulate, formula (16) therefore constitutes a characteristic condition associated with $\{\ell^* = 0\}$. By contrast, a significant flow rate discontinuity at time τ_1 indicates that $\{\ell^* \rangle 0\}$.

Figure 2 depicts the statistical distribution of egress times either free-flow or including a queuing episode at $\ell^* = 0$. The free-flow distribution corresponds to users' arrivals at the potential bottleneck, whereas the effective one corresponds to their departures from the potential bottleneck.

4. Distributional Assumptions

Basically, free-flow egress time τ is modeled as the ratio of space length l and cruising walk speed w. At first glance, statistical independence between l and w looks a reasonable assumption. On second thoughts, however, there may be correlation between them: for instance, hurried train users would both walk faster and position themselves on board so as to alight closer to their platform egress point.

Let us recall the free-flow CDF and PDF of egress times in (4a) and (4b):

$$T(x) = \int S_w(w \cdot x) dW(w),$$
$$\dot{T}(x) = \int w \cdot \dot{S}_w(w \cdot x) dW(w). \tag{17}$$

In (4a) and (4b), we expressed the free-flow CDF and PDF of egress times. Let us now put forward specific distributional assumptions of two kinds: either a bivariate Gaussian distribution for (ℓ, w) or a bivariate Gaussian distribution for $(\widetilde{\ell}, \widetilde{w})$, where $\widetilde{\ell} \equiv \ln(\ell)$ and $\widetilde{w} \equiv \ln(w)$, called the log-normal model.

4.1. Model with Gaussian Components

4.1.1. Basic Definitions and Free-Flow Properties. We may assume Gaussian distributions for lengths and speeds: or more precisely that the (ℓ, w) pair is a bivariate Gaussian vector with $\ell \approx N(m_\ell, s_\ell^2)$, $w \approx N(m_w, s_w^2)$ and $\chi = \mathrm{cov}(\ell, w)$. Given x, the spacings around the coma are too large $y_x \equiv (s_\ell^2 + s_w^2 x^2 - 2\chi x)^{0.5}$, we have that $\ell - x.w \approx N(m_\ell - x m_w, y_x^2)$ so that the free-flow CDF in (4a) takes on the specific form as follows:

$$T(x) = \Phi\left(\frac{x m_w - m_\ell}{y_x} \right), \tag{18a}$$

where Φ denotes the CDF of the reduced Gaussian variable. The associated reduced PDF will be denoted as $\phi(z) \equiv \exp(-(1/2) z^2)/\sqrt{2\pi}$.

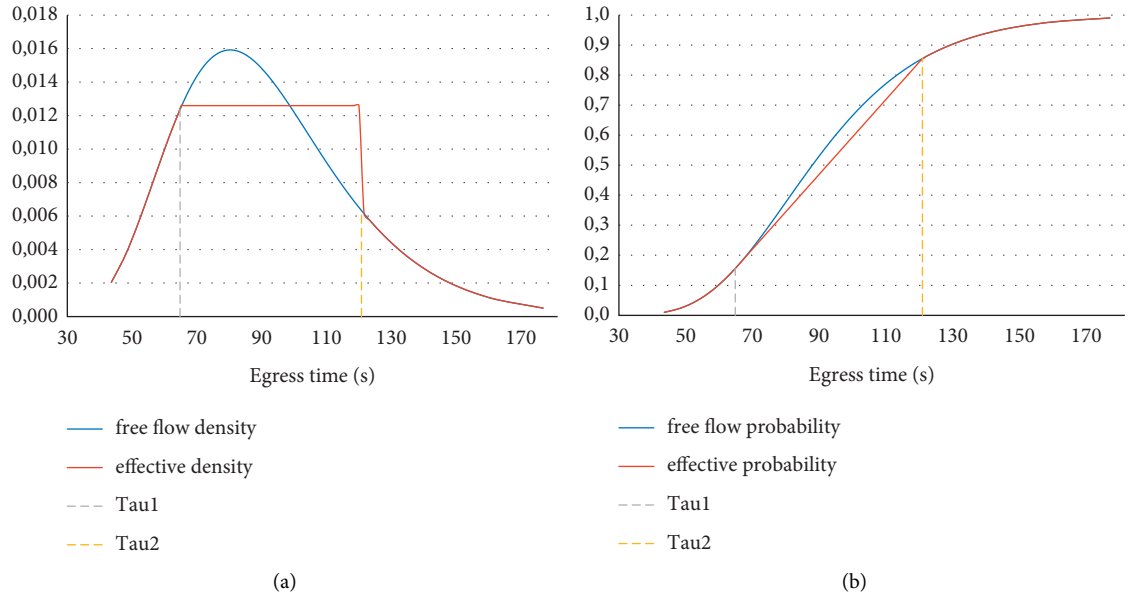

Figure 2: PDF (a) and CDF (b) of egress times.

The resulting distribution of τ is not Gaussian, because the influence of x on the CDF (both through x and y_x) is a complex one.

By straightforward derivation, $\dot{T}(x) = (m_w/y_x - (1/2)(xm_w - m_\ell)(2s_w^2 x - 2\chi)/(y_x)^3)\dot{\Phi}((xm_w - m_\ell)/y_x)$.

Thus, we get the free-flow PDF of the egress times as

$$\dot{T}(x) = \frac{m_w\left(s_\ell^2 - \chi x\right) + m_\ell\left(s_w^2 x - \chi\right)}{(y_x)^3}\phi\left(\frac{xm_w - m_\ell}{y_x}\right). \quad (18b)$$

Both (18a) and (18b) are easy to compute. The complex influence of x is obvious in (18b), both through ϕ and out of it as a quotient of functions such that the denominator involves an exponentiation to power 3/2.

Of course, the Gaussian postulate is somewhat farfetched for walk speeds as it gives support to some negative values: in each model estimation, we will have to check ex-post that the estimated parameters give rise to "almost certain" positive speeds.

4.1.2. Properties for Trains with Alighting Queues. It is shown in the appendix that the law of ℓ conditionally to w is $N(m_{\ell|w}, s_{\ell|w}^2)$ for $s_{\ell|w} \equiv s_\ell\sqrt{1-\rho^2}$ with $\rho \equiv \chi/(s_\ell \cdot s_w)$ and $m_{\ell|w} \equiv m_\ell + (w - m_w)\beta$ with $\beta \equiv \chi/s_w^2$.

Furthermore, function $\overline{M}_x(y) \equiv \int_y^\infty w\dot{S}_w(w \cdot x)dW(w)$ satisfies that.

$\overline{M}_x(y) = \int_y^\infty (w/s_{\ell|w})\phi((w.x - m_{\ell|w})/s_{\ell|w})(1/s_w)\phi((w - m_w)/s_w)dw = K_x(s_x/s_{\ell|w}) \cdot s_w(m_x\Phi((m_x - y)/s_x) + s_x\phi((y - m_x)/s_x))$, where $s_x \equiv ((x-\beta)^2 s_{\ell|w}^{-2} + s_w^{-2})^{-1/2}$, $m_x \equiv s_x^2((x-\beta)(m_\ell - \beta m_w)s_{\ell|w}^{-2} + m_w s_w^{-2})$ and $K_x = (1/\sqrt{2\pi})\exp(1/2)((m_x/s_x)^2 - (m_\ell - \beta m_w/s_{\ell|w})^2 - (m_w/s_w)^2)$.

These formulas enable us to calculate the PDF of egress times in the congested model.

As for the subset probabilities P_k, we have to calculate P_1 and P_2 numerically:

$P_1 = \int S_w(\min\{w \cdot \tau_1, \ell^* + w \cdot \tau_1^*\})dW(w)$ and $P_2 = \int \overline{S}_w(\max\{w \cdot \tau_2, \ell^* + w \cdot \tau_2^*\})dW(w)$.

The computation needs be done once for each set of parameters.

In the incomplete model, the probabilities $\widetilde{P}_1$ and $\widetilde{P}_2$ are easy to calculate:

$\widetilde{P}_1 = \Pr\{\ell \langle w.\tau_1\} = \Phi((m_w\tau_1 - m_\ell)/\sqrt{s_\ell^2 + s_w^2\tau_1^2 - 2\tau_1\chi})$ and $\widetilde{P}_2 = \Pr\{\ell \rangle w.\tau_2\} = \Phi((m_\ell - m_w\tau_2)/\sqrt{s_\ell^2 + s_w^2\tau_2^2 - 2\tau_2\chi})$.

A similar approach yields fairly good approximations for P_1 and P_2

$P_1 \approx \Pr\{\ell - \ell^* \langle w.\tau_1^*\} = \Pr\{\ell - w.\tau_1^* \langle \ell^*\} = \Phi((\ell^* - m_\ell + m_w\tau_1^*)/\sqrt{s_\ell^2 + s_w^2\tau_1^{*2} - 2\tau_1^*\chi})$, $P_2 \approx \Pr\{\ell - \ell^* \rangle w.\tau_2^*\} = 1 - \Pr\{\ell - w.\tau_2^* \langle \ell^*\} = \Phi((m_\ell - m_w\tau_2^* - \ell^*)/\sqrt{s_\ell^2 + s_w^2\tau_2^{*2} - 2\tau_2^*\chi})$

4.2. Log-Normal Model. When two positive real variables are involved in a product or quotient relationship, it is convenient to model them as bivariate log-normal, because the composed variable will be log-normal, too. We shall indent with tildes the variables to denote their natural logarithms concisely: let then $\widetilde{\ell} \equiv \ln(\ell)$, $\widetilde{w} \equiv \ln(w)$, and $\widetilde{\tau} \equiv \ln(\tau)$. As for parameterization, let $\widetilde{\ell} \approx N(\mu_\ell, \sigma_\ell^2)$, $\widetilde{w} \approx N(\mu_w, \sigma_w^2)$ and $\xi = \operatorname{cov}(\widetilde{\ell}, \widetilde{w})$.

As $\ln(1/w) = -\widetilde{w}$, the log-egress time $\widetilde{\tau}$ satisfies that

$$\widetilde{\tau} = \widetilde{\ell} - \widetilde{w} \approx N\left(\mu_\tau, \sigma_\tau^2\right) \text{ with } \mu_\tau = \mu_\ell - \mu_w \text{ and } \sigma_\tau^2 = \sigma_w^2 + \sigma_\ell^2 - 2\xi. \quad (19)$$

Thus, the egress time is distributed log-normal with parameters μ_τ and σ_τ^2. Its CDF and PDF are, respectively:

$$T(x) = \Phi\left(\frac{\ln(x) - \mu_\tau}{\sigma_\tau}\right). \tag{20a}$$

$$\dot{T}(x) = \frac{1}{\sigma_\tau . x}\phi\left(\frac{\ln(x) - \mu_\tau}{\sigma_\tau}\right). \tag{20b}$$

5. Estimation Methodology at the Train Level

We shall first establish some basic properties and provide a simple, empirical estimation method to recover alighting positions from free-flow egress times, taking as exogenous the distribution of free-flow walk speeds. Then, we devise a "train likelihood function" of observed egress times, enabling for the maximum-likelihood estimation of model parameters: queue focal point ℓ^* and time bounds τ_1^* and τ_2^*, queued walk speed v^*, and the capacity flow rate K^*, as well as the parameters in the joint distribution of the free-flow walk speeds and walk lengths.

5.1. Simple Properties for Independent Lengths and Speeds under Free Flow. Let us establish some properties for free-flow egress times under the simplifying assumption of statistical independence between lengths and speeds.

5.1.1. Signal Analysis. Let us focus on the influence of length ℓ onto egress time $\tau = \ell/w$: we take this influence as the "signal," as opposed to the influence of walk speed w, which is taken as the "noise." Here, we want to assess the importance of the signal in the phenomenon and measure the ratio between the signal and the noise. To do that, we shall decompose the variance $V[\tau]$ with respect to $V[\ell]$ and $V[w^{-1}]$. Postulating here the independence of ℓ and w, it holds that $E[\tau] = E[\ell].E[w^{-1}]$ and $V[\tau] = E[\ell]^2 \cdot V[w^{-1}] + V[\ell] \cdot (E[w^{-1}]^2 + V[w^{-1}])$.

The latter decomposition, after division by $E[\tau]^2$, gives the following relationship between the squared relative dispersions:

$$\gamma_\tau^2 = \gamma_{w^{-1}}^2 + \gamma_\ell^2 + \gamma_\ell^2 \cdot \gamma_{w^{-1}}^2. \tag{21}$$

As first-guess assumptions on the lengths and speeds, let us take the following:

(i) About lengths, that alighting users would spread evenly along a 200m train (resp. 100m long train), then (up to s_e) length ℓ is distributed uniform on [0, 200 m] (resp. 100 m). Then, $E[\ell] = 100$ m and $V[\ell] = (200m)^2/12$ so that $\gamma_{\ell_e}^2 = 1/3$ (resp. the same for a 100-m long train).

(ii) About walk speeds, a normal distribution with mean at 4 km/h and most values from 2 to 6 km/h, that is, spread $6 - 2 = 4$ representing about 4 times the standard deviation. Then, $SD = 1$ km/h and $\gamma_w = 1/4$. As the value is small, it follows that $\gamma_{w^{-1}}^2 \approx \gamma_w^2 \approx 1/16 \approx .07$. Furthermore, $E[w] = 1.1$ m/s and $E[w^{-1}] \approx (1 + \gamma_w^2)/E[w] \approx .95$s/m.

On applying the formulas, we recover an average egress time of $E[\tau] = E[\ell] \cdot E[w^{-1}]$ equal to either 90 s or 45 s depending on train length of 200 m or 100 m. As for relative dispersions, $\gamma_\tau^2 = \gamma_{w^{-1}}^2 + \gamma_\ell^2 + \gamma_\ell^2 . \gamma_{w^{-1}}^2 = (1/16) + (1/3) + (1/3) \times (1/16) = 20/48 = 5/12 \approx 0.40$.

The signal share is $\gamma_\ell^2/\gamma_\tau^2 = (1/3)/(5/12) = 4/5 = 80\%$, and the signal-to-noise ratio is $\gamma_\ell^2/(\gamma_\tau^2 - \gamma_\ell^2) = 80\%/20\% = 4$.

This quick numerical application encourages us to look for the influence of the length signal in the egress time phenomenon, and conversely to utilize observed egress times to infer the associated lengths and the alighting positions behind them.

5.1.2. A Quick Estimation Procedure. From platform geometry and free-flow traffic conditions, it is easy to measure the values of parameters ℓ_e and t_e^0. Furthermore, let us consider an exogenous distribution of walk speeds, so that the statistical moments $E[w^{-1}]$, $E[w^{-2}]$, and $V[w^{-1}]$ are known.

Now, from a dataset of egress times associated with a particular train at the station of interest, it is easy to obtain $E[\tau]$ and $V[\tau]$. From the abovementioned expectation formula, we can then recover the mean alighting position $E[\ell\prime]$ from the mean length $E[\ell]$ as follows: $E[\ell] = E[\ell\prime] + \ell_e = (E[\tau] - t_e^0)/E[w^{-1}]$.

Similarly, from the variance formula of the quotient variable, the variance of alighting positions can be recovered as

$$V[\ell'] = V[\ell] = \left(V[\tau] - (E[\ell])^2 \cdot V\left[w^{-1}\right]\right)/E\left[w^{-2}\right]. \tag{22}$$

This estimation procedure is particularly straightforward for log-normal variables. In this case, from $E[\tau]$ and $V[\tau]$, we obtain $\gamma_\tau^2 = V[\tau]/E[\tau]^2$ and the log-normal parameters: variance of log-egress times is $\sigma_\tau^2 = \ln(1 + \gamma_\tau^2)$ and average log-egress time is $\mu_\tau = \ln(E[\tau]) - (1/2)\sigma_\tau^2$. If the exogenous distribution of speeds is log-normal, too, then we similarly get σ_w^2 and μ_w. Next, in the independent case, the log length is distributed Gaussian with mean $\mu_\ell = \mu_\tau + \mu_w$ and variance $\sigma_\ell^2 = \sigma_\tau^2 - \sigma_w^2$ (the minus sign comes from (19)). It remains to derive first the squared relative dispersion $\gamma_\ell^2 = \exp(\sigma_\ell^2) - 1$, then the average $E[\ell] = \exp(\mu_\ell + (1/2)\sigma_\ell^2)$, and lastly the variance $V[\ell] = E[\ell]^2\gamma_\ell^2$.

The quick estimation scheme pertains to alighting positions on the basis of a prior knowledge of the distribution of walk speeds. Furthermore, it is based on the assumption of independence between lengths and speeds, and it is restricted to free-flow egress times.

5.2. The Likelihood Function of a Train Sample of Egress times. Let us consider the sample of all users alighting from a train at a given station; its size A is the number of alighting users. We index the users with $u \in \{1, 2 \ldots A\}$ in the order of increasing egress times τ_u: thus, our observation dataset is $O = \{\tau_u : u \in \{1, 2 \ldots A\}\}$, and it is an exhaustive sample for that train.

5.2.1. Free-Flow Case. The users that egress under free-flow conditions can be considered independent of the other ones. When all of the alighting users enjoy free-flow, their egress times contain no information about any queuing episode. The set of parameters that can be recovered by statistical estimation then pertains to the joint distribution of walk speeds and lengths, say $\Theta_{FF} \equiv \Theta_{(\ell,w)}$. The PDF of any observed egress time, $\dot{T}(\tau_u \mid \Theta_{FF})$, contains information on Θ_{FF} and constitutes an elementary likelihood function $L_u(\Theta_{FF} \mid \tau_u)$, with associated log-likelihood function $\Lambda_u(\Theta \mid \tau_u) \equiv \ln L_u(\Theta \mid \tau_u) = \ln \dot{T}(\tau_u \mid \Theta)$.

As free-flowing egress times are statistically independent, the train sample gives rise to a train likelihood function under product form:

$$L_O(\Theta) = \prod_{u \in O} L_u(\Theta \mid \tau_u), \tag{23a}$$

with associated log-likelihood function under additive form:

$$\Lambda_O(\Theta) = \sum_{u \in O} \ln \dot{T}(\tau_u \mid \Theta). \tag{23b}$$

5.2.2. Train with Queuing Episode. When queuing occurs among the users alighting of the train, we expect that the egress times of queued users contain some information on the queuing parameters, $\Theta_Q \equiv [\ell^*, \tau_1^*, \tau_2^*, v^*, K^*]$. Then, the overall vector of parameters that may be estimated is $\Theta \equiv [\Theta_{(\ell,w)}, \Theta_Q]$.

All users egressing under free-flow pedestrian conditions can still be taken as mutually independent and independent from the other ones, so that their joint likelihood function is under product form.

It remains to state the likelihood function of queued users. Statistical independence between them is not obvious since some FIFO rule applies within the queue. However, we will also take $\dot{T}(\tau_u \mid \Theta)$ as a likelihood function of Θ and assume that the joint likelihood of queued users is under product form, and further that the queued users are independent of the free-flowing ones. Then, overall,

$$L_O(\Theta) = \prod_{u \in O} L_u(\Theta \mid \tau_u). \tag{24a}$$

And the associated log-likelihood function is under additive form:

$$\Lambda_O(\Theta) = \sum_{u \in O} \ln \dot{T}(\tau_u \mid \Theta). \tag{24b}$$

Given Θ_Q, hence τ_1 and τ_2, the set of alighting users, O, can be split into three subsets $O_1 = \{\tau_u \langle \tau_1\}$, $O_2 = \{\tau_u \rangle \tau_2\}$, and $O_3 = \{\tau_u \in [\tau_1, \tau_2]\}$, with respective sizes A_k that add up to A. From the formula of $\dot{T}_U$, the log-likelihood of the queued users amounts to

$$\Lambda_{O_3} = A_3 \ln \frac{P_3}{\Delta\tau}. \tag{25}$$

While for those users in O_1 and O_2, we have that, respectively, $\Lambda_u(\Theta \mid \tau_u) = \ln[\overline{M}_x(0) - 1_{\{x \rangle \tau_1^*\}} \overline{M}_x(\ell^* / (x - \tau_1^*))]$ if $x \langle \tau_1$, $\Lambda_u(\Theta \mid \tau_u) = \ln \overline{M}_x(\ell^* / (x - \tau_2^*))$ if $x \rangle \tau_2$.

In fact, the involvement in the queue tends to erase the information on $\Theta_{(\ell,w)}$ in τ_u, analogously to the absence of information on Θ_Q in the egress times of free-flowing users. If the queuing episode is long and involves a vast majority of alighting users, then we expect the sample to carry information mostly on Θ_Q but little if any on $\Theta_{(\ell,w)}$. Thus, a wise estimation strategy could be to set up $\Theta_{(\ell,w)}$ on the basis of prior knowledge and to focus on Θ_Q as the "active" set of parameters for that train.

5.3. Maximum-Likelihood Estimation. Maximum-likelihood estimation is a fundamental method for statistical estimation, owing to both theoretical properties and tractability. Given a sample of observations, it consists in maximizing numerically the likelihood function associated with the sample (or, more conveniently, its logarithm called the log-likelihood function), with respect to the vector of parameters.

The train log-likelihood function is quite tractable for standard optimization algorithms. Yet some caution is in order about discontinuities in the function: changing the queuing parameters may change the assignment of observed times from free-flow regime to queuing and reversely, thereby changing the associated elementary log-likelihood function from one specification to another one, at the risk of discontinuity.

5.3.1. MLE of the Model with Gaussian Components. In the model with Gaussian lengths and speeds, both the CDF and PDF of τ, and in turn L_u in the free-flow case as well as in the queued case, depend on the five parameters: m_w, s_w^2, m_ℓ, s_ℓ^2, and χ. These are involved together starting from relationship (4a); that is, $\{\ell_e \leq w \cdot \Delta x\}$. As this relation puts ℓ and w on the same level, in a linear way, it will make the (ℓ, w) joint distribution identifiable only up to some scale factor that will affect the mean parameters at order 1 and the variance-covariance parameters at order 2. These influences are easy to trace out in both CDF and PDF formulas (18a) and (18b).

In turn, the system of five first-order optimality conditions for likelihood maximization with respect to the five parameters will be underdetermined: only four out of five parameters may be identified. Our intuition here is to take the average walk speed as given and to restrict the application of MLE to the other four parameters.

5.3.2. MLE of the Log-Normal Model under Free Flow. In the absence of queuing, the bivariate log-normal specification of pair (ℓ, w) yields a simple log-normal model of the egress time $\tau = \ell/w$. The associated pair of logs, $(\widetilde{\ell}, \widetilde{w})$, is bivariate normal with $\mu_\ell, \sigma_\ell^2, \mu_w, \sigma_w^2$, and ξ as parameters. Then, the log-egress time $\widetilde{\tau} = \widetilde{\ell} - \widetilde{w}$ is normal with mean $\mu_\tau = \mu_\ell - \mu_w$ and variance $\sigma_\tau^2 = \sigma_w^2 + \sigma_\ell^2 - 2\xi$.

In the application of MLE to a sample of free-flow egress times τ_u hence of $\widetilde{\tau}_u$, only two parameters μ_τ and σ_τ^2 are identifiable. The optimality conditions of the MLE in that case are well known:

$$\mu_\tau = \frac{1}{A}\sum_{u\in O}\widetilde{\tau}_u, \tag{26a}$$

$$\sigma_\tau^2 = \frac{1}{A}\sum_{u\in O}\left(\widetilde{\tau}_u - \mu_\tau\right)^2. \tag{26b}$$

At that stage, the line of attack in the quick estimation procedure is appropriate: it is a wise strategy to take μ_w and σ_w^2 as given and to focus on the estimation of walk length parameters, namely, μ_ℓ and composite parameter $\sigma_\ell^2 - 2\xi$, as the respective influences of σ_ℓ^2 and ξ would be hard to disentangle.

6. Case Study

6.1. The Site and Its Observation

6.1.1. Site Location and Platform Geometry. Line A of the Regional Express Railways (RER) is the busiest urban rail transit line in the Paris region "Ile-de-France" and maybe Europe, carrying more than 1 M passengers on every workday. The line is serviced by duplex trains about 210 m long, each with 10 cars (and per car 3 doors each 2m wide) and a nominal capacity of 2,800 passengers (assuming 4 p/m_2 of standing space). Along that line, the Noisy-Champs station (48°50′34.55″ N, 02°34′55.06″ E) is located in the eastern part of the Paris conurbation on the edge of the Noisy-le-Grand and Champs-sur-Marne communes. Its attendance was of 4.4 million travelers in year 2015. On weekdays, there are significant flows of travelers, especially workers and students coming to work in the "Cité Descartes," a district of offices, high schools, and activity parks. Figure (3(a)) shows the outline of the station and the facilities around the Noisy-Champs train station. To the west of the station (Noisy-le-Grand side) is a residential area and to the east (Champs-sur-Marne side), there are offices and residences as well as the Cité Descartes of many high schools and universities. On peak periods of weekdays, important flows of passengers arrive to, and exit from, the east side of the station, which provides direct access to park and ride facilities and to bus stops.

Regarding the shape of the station, the platform extends over 230 m along an east-west axis on each side of the two rail tracks: each side has a width of about 5 meters. The northern platform is utilized by train runs in the direction from Paris to Marne-la-Vallée, whereas the southern platform serves for the inverse direction. Per platform side, there are two PEPs (platform egress points) at the east and west endpoints. On the northern platform, there are stairs and escalators that connect the PEPs to the SAPs (station access points) with validation gates for tap-in and tap-out. On the southern platform, there are only stairs to connect the PEPs to the SAPs.

This study is focused on the northern platform and more specifically the station access point to the Cité Descartes. From the geographical situation of the station within the metropolitan area, employment, and activities are relatively scarce eastwards compared to those westwards and even to local opportunities (jobs and schools): thus, on the northern platform (from Paris to Marne-la-Vallée EuroDisneyland), the boarding flow is much lower than the alighting flow and the alighting users are not impeded by the boarding candidates.

Figure 3(b) details the geometry of the eastern station access point. The platform is located on level -1 and the validation gates on level 0. There is an escalator for one or two people abreast and a parallel staircase about 2.5 meters wide. As the two vertical elements are parallel and of limited height (about 4 meters), their respective pedestrian times are quite similar (from field observation, also in accordance with Wardrop's 1st principle): this is why no distinction will be made between them. When a train arrives and disembarks the alighting flow, passengers destined to the City Descartes use the exit and a traffic jam may arise on the platform in front of the escalator (Figure 3(d)). After the travelers climbed the escalator or the staircase, the validation gates are about five meters at the bottom. The distance from train head to gates is from 25 to 30 meters depending on the validation gates. There are three entrance gates and six exit gates. During peak hours passengers who took the escalators and stairs may queue in front of the six exit gates. The different positions of those validation gates may induce further difference in the egress times.

6.1.2. AFC and AVL Datasets. We made use of two datasets obtained from two systems of automated fare collection (AFC) and automatic vehicle location (AVL), respectively. Both datasets were constituted for all weekdays from the 16th to the 29th of March 2015, that is, 10 days in total. The time stretch enabled us to pinpoint peak periods of weekdays unaffected by disturbances.

In the Paris region, the AFC information system (named SIDV) records all validations at fare gates in stations for rail modes or on board for buses and trams. Every gate has a particular index, and every card has one unique number. Thus, per validation, the AFC record contains spatial-temporal attributes of line, station, gate, time, user, etc. Our AFC dataset contains 4,675,672 validations along line A, including 83,740 validations in Noisy-Champs; it pertains to a total of 723,185 travelers, among whom 21,822 passing by Noisy-Champs.

The AVL system uses track circuits to detect events of train passage at given points: the resulting train timestamp (geolocation and instant) is transmitted by radio. Three kinds of train and track events are monitored: either outside stations or in them—the ARR and DEP types. ARR and DEP provide the exact instants of arrival and of departure of trains in stations. Our AVL dataset records 6,608 train runs on the RER A line, among which 2,954 stop by Noisy-Champs.

Based on the within-day variations of validations at Noisy-Champs, five daily subperiods were identified: morning peak (7 : 30–9:30) with an average flow of 221 passengers per 15 min (standard deviation of 58), evening peak (17 : 30–19 : 30) with an average flow of 234 passengers per 15 min (standard deviation of 62), together with three off-peak periods: 5 : 50–7:30, 9 : 30–17 : 30 and 19 : 30–23 : 00,

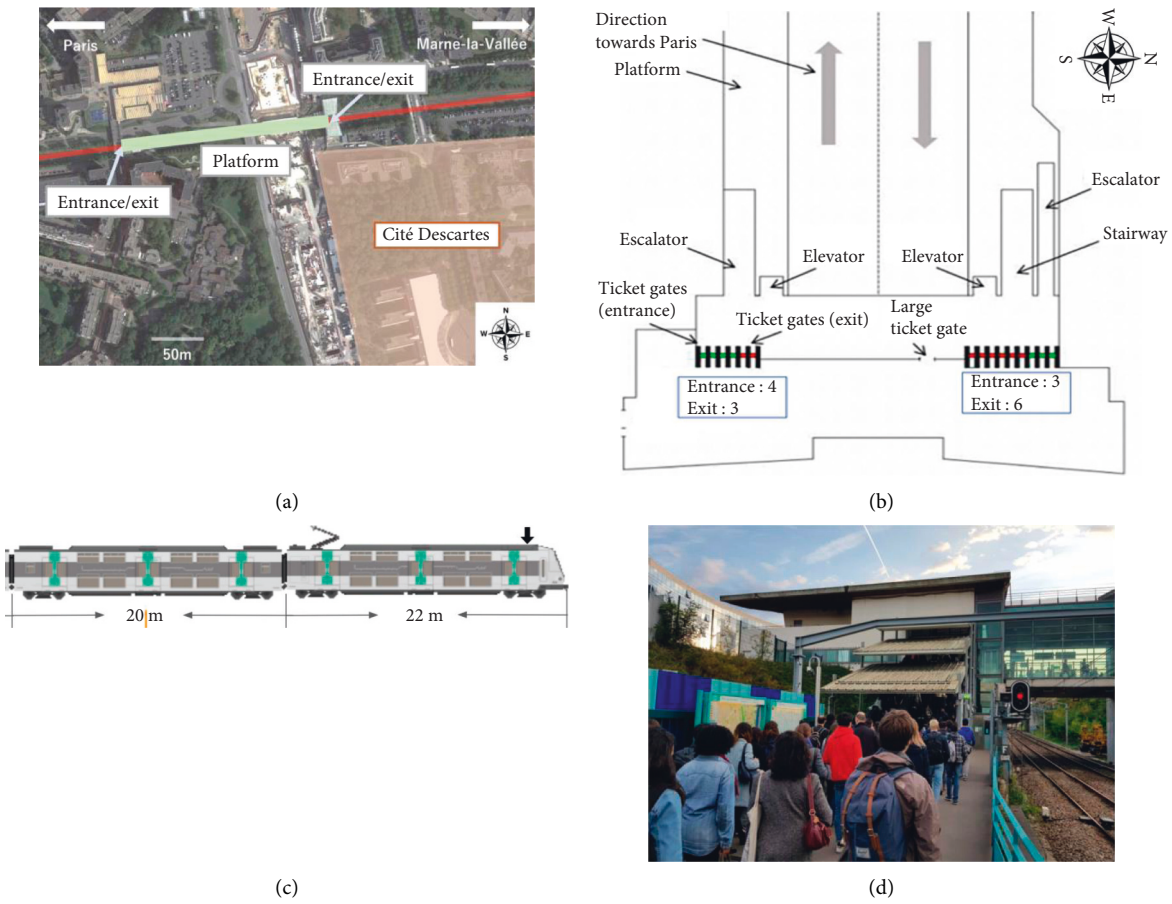

FIGURE 3: (a) Satellite photo of the site, (b) eastern side exit/entrance of the station, (c) train head, first two cars, and their doors (around green bands), and (d) eastern PEP of northern platform.

during which the average validation flows amount to 17, 62, and 38 passengers per 15 min, respectively (standard deviations of 21, 28, and 31).

6.1.3. Empirical Evidence from Field Survey. We also designed a specific field survey, which was carried out by four engineering students on Friday 25/09/2020 from 8 a.m. to 9 a.m. The purpose was to measure passenger egress times according to different train cars, to describe congestion events and understand their mechanisms. A specific four-fold protocol was set up as follows.

The first part consisted in observing the distribution of egress times. Students were posted near to the validation gates in order to count the number of passengers tapping out at each second after the opening of doors. Four trains were observed, yielding a sample of 405 individual times. The average headway of these 4 trains is 4.1 minutes. Among the observed egress times, from the shortest value of 19 s to the longest one of 242 s, the mode was 60 s, the average 73.9 s, and standard deviation 35 s (see Figure 4). The second part consisted in following randomly selected alighting passengers from their car to the validation gates so as to measure their walk time. The results were consistent with the first part. The third part was devoted to counting the number of alighting passengers passing by a specific point along the platform, namely, between cars 4 and 5 from train head (out of ten). The fourth and last part focused on queuing phenomena. Two queues were observed: one at the foot of the escalator (PEP) and the other in front of the gates (SAP). At the PEP, queuing occurred around 15s from train arrival and lasted 32 s on average (among the trains). At the gates, some queuing occurred around 50s from train arrival [28].

6.1.4. An Ex-Ante Convention to Identify Queuing Time Ranges. Based on both field observation and the analysis of AFC and AVL data, we estimated on a provisional basis an exit flow capacity at validation gates of about 10 people in 5 seconds, that is, 2 people per second. Once the flow rate reaches this threshold, the queuing phenomenon appears. All the gates are used and waiting lines appear at both the gates and the escalator. Then, based on provisional exit flow capacity, a "provisional queuing interval" was determined in the following way: from AFC data, the entire egress period was sliced in 5s sub-intervals and we took as "initial queuing instant" the start of the first 5s slice containing 10+ and as "final queuing instant" the end of the last 5s slice with 10+ validations. From these educated guesses, we derived times τ_1, τ_2, $\tau_2 - \tau_1$ (duration of congestion) for every train in the AVL and AFC datasets. The resulting distribution of τ_1 has mean of 57.6 s and standard deviation of 8.7s. That of τ_2 has a

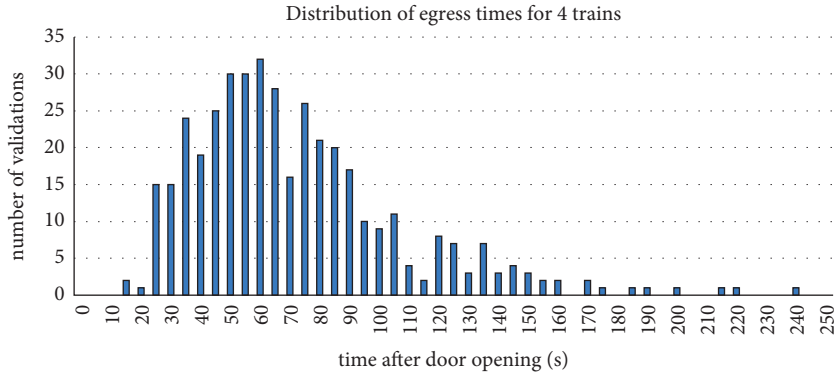

Figure 4: Distribution of observed egress times.

mean of 81.0s and a standard deviation of 20.1 s. That of congestion duration has a mean of 23.4 s and a standard deviation of 21.1 s.

6.2. Model Estimation. We selected 4 consecutive trains in the evening peak on March 16, 2015. Their respective time of arrival is 18 : 32 : 14, 18 : 35 : 03, 18 : 46 : 33, and 18 : 59 : 09. These trains have various profiles and characteristics that make them typical cases for study. We applied the model with Gaussian components to the 4 trains. Both the free-flow (FF), incomplete congestion (IC), and full congestion (FC) models were tested.

6.2.1. Train at 18 : 35 : 03. At Noisy-Champs on March 16, 2015, the 18 : 35 eastwards train followed the previous one with a $3'$ interval that is relatively short. The alighting flow includes $A \equiv 76$ passengers only. The empirical distribution of their egress times, depicted as the green plot in Figure 5(a), exhibits one primary statistical mode around 60 s and a second, minor mode around 100 s, with an associated subpopulation probability of about 15% (from the .007 density level times the 80–110 s range, minus the tail of the distribution associated with the primary mode). Such secondary mode involving about 10 passengers may correspond to a specific group, for instance a bunch of students coming back to their residences. To circumvent the randomness of such events, we decided to focus on the primary mode and subpopulation in the following way. All egress times up to threshold $\widetilde{\tau} \equiv 80$s are assumed to belong to the primary subpopulation, and their individual values are kept in the sample: there are $A\prime \equiv 46$ of them. As for values above $\widetilde{\tau}$, only one-third of them is taken to belong to the primary subpopulation: their number is $\widetilde{A} \equiv (A - A')/3$. While this number is known, we do not consider the individual values and keep to the information that these values are greater than $\widetilde{\tau}$. The associated log-likelihood amounts to $\widetilde{A} \cdot \ln[1 - T(\widetilde{\tau} \mid \Theta)]$.

Then, the total log-likelihood function of the primary sample amounts to

$$\widetilde{\Lambda}(\Theta) \equiv \widetilde{A} \cdot \ln[1 - T(\widetilde{\tau} \mid \Theta)] + \sum_{u:\, \tau_u > \widetilde{\tau}} \ln \dot{T}(\tau_u \mid \Theta). \tag{27}$$

Looking for parameter vector $\Theta_{(\ell,w)} \equiv [m_\ell, s_\ell, m_w, s_w, \chi]$, we set up initial values of [130, 60, 1.2, 0.25, 0], meaning (i) constant $m_w \equiv 1.2$ m/s to maintain identifiability, (ii) prior knowledge of .25 m/s about the standard deviation of walking speeds, (iii) null covariance χ as first round postulate of independence between lengths and speeds, (iv) prior assumption of alighting positions uniformly distributed along the 200m long train, shifted by $s_e \equiv 30$ m from train head to validation gates.

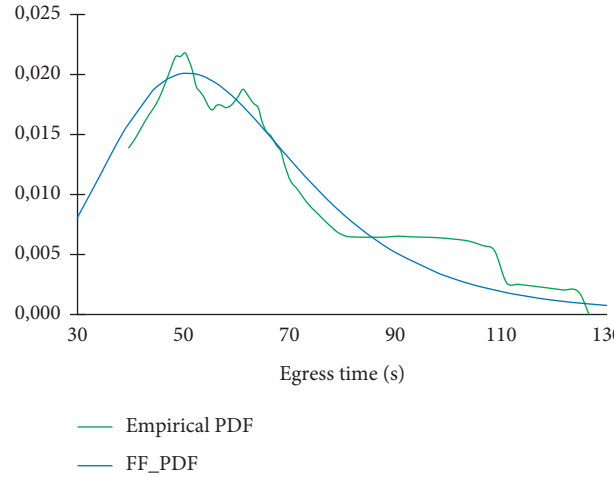

Figure 5: PDF of egress times for a train at 18:35. Note: empirical data are obtained by applying a sliding window over 10 egress times. FF means "free flow."

By maximizing the log-likelihood function with respect to $[m_\ell, s_\ell, s_w]$, starting from value $\widetilde{\Lambda}_0$ =-255.0, an optimal value of $\widetilde{\Lambda}^*_{FF}$ =-212.8 was obtained at point [68.04, 18.28, 0.302]. The $\widetilde{s}_w$ estimate looks quite reasonable. About lengths, estimate $\widetilde{m}_\ell$ of 68 m minus shift s_e of 30 m yields an average alighting position of $\widetilde{\ell}\prime \equiv \widetilde{m}_\ell - s_e = 38$ m, which corresponds to the third and last door of the second car along the train. Then, taking $\widetilde{\ell}\prime \pm 2\widetilde{s}_\ell$ as a quick estimate for a 95% confidence interval of alighting positions, the resulting interval $[1.5m, 74.7m]$ corresponds broadly to the first four cars out of the 10 car train, that is, most of its first half.

Further attempts to estimate $[m_\ell, s_\ell, s_w, \chi]$ starting from initial point [130, 60, 0.25, 0] conducted to the same optimal value and [68.04, 18.28, 0.302, 0.00]; that is, the MLE estimate of the covariance parameter χ is zero: thus, the train

sample strongly supports the assumption of statistical independence between walk lengths and speeds among the alighting passengers.

Based on the estimated values together with the independence property, we can recover some properties of the egress time RV and analyze the signal-to-noise ratio (cf. §5.1.1). As $\gamma_w \approx 25\%$, we can safely approximate $\gamma_{1/w} \approx \gamma_w \approx 25\%$ and derive $E[w^{-1}] \approx (1+\gamma_w^2)/E[w] \approx .89$ s/m.

In turn, $E[\tau] = E[\ell] \cdot E[w^{-1}] \approx 60.3$ s.

$\gamma_\tau^2 = \gamma_{w^{-1}}^2 + \gamma_\ell^2 + \gamma_\ell^2 . \gamma_{w^{-1}}^2 = .14$; hence, $\gamma_\tau = .37$ and $SD[\tau] \approx 22.6$s. Thus, the share of γ_ℓ^2 in γ_τ^2 amounts to 52%, leading to a signal-to-noise ratio of 1.06.

6.2.2. Train at 18 : 59 : 09. At Noisy-Champs, the 18 : 59 train has a $13'$ headway and disembarks an alighting flow of $A \equiv 196$ passengers. The empirical profile of egress times (green curve in Figure 6) strongly suggests the occurrence of queuing. However, we began by estimating the basic free-flow model (FF) with a bivariate Gaussian distribution of the length-speed pair. Initial values of $[130, 60, 1.2, 0.25, 0]$ were set up for parameter vector $\Theta_{(\ell,w)} \equiv [m_\ell, s_\ell, m_w, s_w, \chi]$. By maximizing the log-likelihood function with respect to $[m_\ell, s_\ell, s_w]$, starting from value $\Lambda_{FF}^{(0)}$ =-996.12, an optimal value of Λ_{FF}^* =-922.65 was obtained at point [106.96, 20.21, 0.254]. Adding the χ parameter, we obtained a slightly improved value of -922.56 at point [107.03, 23.77, 0.285, 1.858]. The log-likelihood improvement is very small and does not justify considering a nonzero covariance.

Turning to the incomplete congestion model (IC), we set the $[\tau_1, \tau_2]$ interval to [52s, 116s], according to the provisional exit capacity of 2 people per second. At starting point $[m_\ell, s_\ell, m_w, s_w, \chi] = [130, 60, 1.2, 0.25, 0]$, the initial value of the IC log-likelihood amounts to -998.27. Optimization with respect to $[m_\ell, s_\ell, s_w]$ yielded an optimal value Λ_{IC}^* = -918.86 at point [100.65, 14.16, 0.289]. The 3 point improvement over the free-flow model would strongly support the addition of one parameter to the free-flow model. Considering the FF model as an IC model with $\tau_1 = \tau_2$ set to 52s, adding the complementary parameter $\tau_2 - \tau_1$ with value set to 64s constitutes a statistically significant improvement.

The next step is to estimate the full congestion model. We considered not only the speed and length parameters but also the queuing parameters $\Theta_Q \equiv [\ell^*, \tau_1^*, \tau_2^*, \nu^*]$. Starting from the IC estimates together with $\Theta_Q^{(0)} \equiv [0, 52, 116, 0.80]$, hence from initial log-likelihood value of Λ_{IC}^*, by a composite estimation process alternating steps of automated optimization with manual search, we obtained an optimal value of $\Lambda_{FC}^* = -913.7$ for the parameter vector of $\Theta_{(\ell,w)}^{(FC)} = [102.20, 15.594, 0.283, 0]$ and $\Theta_Q^{(FC)} = [4.00, 61.65, 107.65, 0.92]$. By the way, the resulting estimates of $[\tau_1, \tau_2] = [66s, 112s]$ taken together with $\ell^* \equiv 0$ provide an IC model with a log-likelihood of -918.01, which improves on the previously estimated Λ_{IC}^* by almost one point. From this re-estimated IC model to the optimized FC model, the log-likelihood is improved by more than 4 points, thus providing satisfying justification for considering a queuing focal point ℓ^* greater than 0.

Let us now comment on the estimates in the FC model. That of the standard deviation of walk speeds is fairly standard. About lengths, estimate m_ℓ of 102-m minus shift s_e of 30-m yields an average alighting position of $\overline{\ell\prime} \equiv m_\ell - s_e$ = 72m, which corresponds to the medium tier of the fourth car along the train. Then, taking $\overline{\ell\prime} \pm 2s_\ell$ as a quick estimate for a 95% confidence interval of alighting positions, the resulting interval [41m, 103m] corresponds broadly to cars 3 to 5 out of the 10 car train. This may correspond to the first two cars being occupied mostly by other passengers destined downstream Noisy-Champs, as well as to some relocation behavior before boarding the train by Noisy-Champs users in order to benefit from better on board comfort (more available seats and less crowded standing spaces). Coming to the queuing characteristics, the queue focal point at 4 meters from the validation gates corresponds to the exit point of the vertical element (escalator and stairways), while queuing would arise at its entry point according to our field survey: this difference may be linked to the fact that people have little if any possibility to overtake one another on the vertical element, therefore making its exit point a replica of its entry point. The queue moving speed of .92 m/s is about one-fourth less than the average speed, which looks consistent.

By depicting the CDFs of the travel time distributions in the FF, IC, and FC models, respectively (in the right-hand part of Figure 6), a typical bottleneck pattern arises: the free-flow CDF mimics the cumulated flow of user arrivals in a bottleneck, whereas the IC and FC CDFs each mimics a cumulated flow of bottleneck exits. From the PDF curves (on the left-hand side of the figure), it is obvious that the congested models are much better fits to the empirical observations than the free-flow one. From either congested model, it is possible to estimate exit capacity based on the train data rather than on the provisional basis. From the FC model, the bottleneck flow rate amounts to $A \times .015 s^{-1} =$ 2.94 people per second.

From the free flow to the full congestion estimates of $\Theta_{(\ell,w)}$, the FC length parameters correspond to alighting positions closer to train head (by 6m on average) and more concentrated (SD reduced from 20.2 m to 15.6 m). Both walk speed distributions comply to the same exogenous average of 1.2 m/s; under null covariance, the estimated SD of FC is 10% higher than the FF one, but the FF estimation with nonzero covariance yields the same speed SD estimate as the FC. This suggests that nonzero covariance in the FF model may point out to the occurrence of queuing.

6.2.3. Train at 18 : 32 : 14. The 18 : 32 train has a $10'$ interval and disembarks 156 alighting passengers at Noisy-Champs. The empirical profile of egress times (in Figure 7) suggests the occurrence of queuing in the 70–100-s interval. Yet we first applied the free-flow model to estimate the vector of parameters of the joint distribution of walk speeds and lengths, $\Theta_{(\ell,w)} \equiv [m_\ell, s_\ell, m_w, s_w, \chi]$. Starting from initial values of [100, 20, 1.2, 0.3, 0] and keeping m_w = 1.2 m/s for the sake of identifiability, we

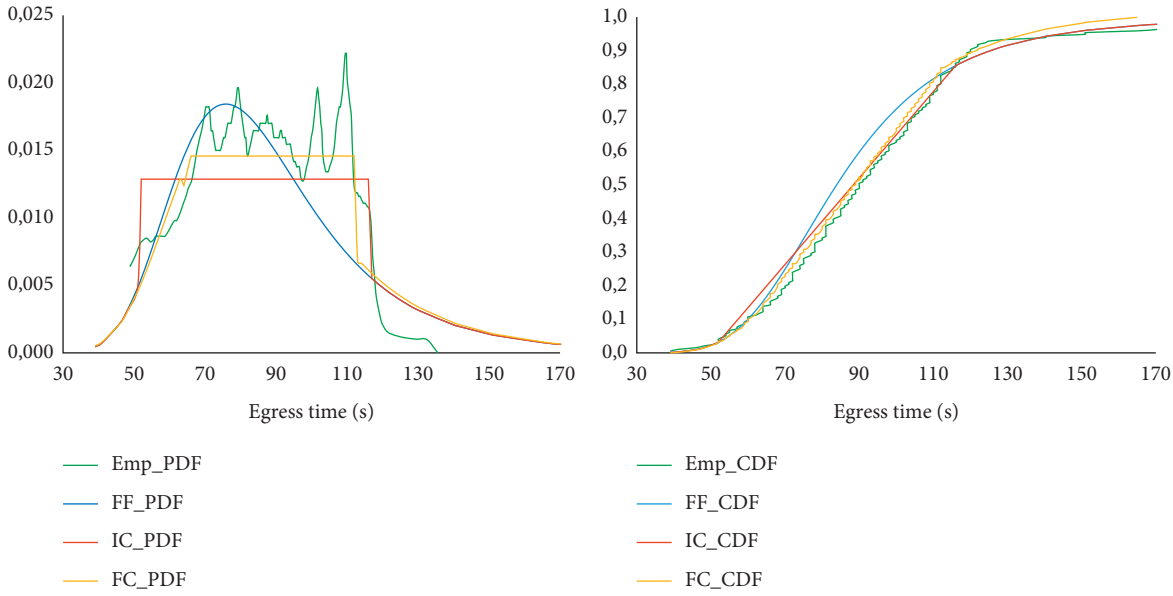

Figure 6: PDF (left side) and CDF (right side) of egress times for the train at 18 : 59.

obtained the MLE estimates of [96.11, 21.85, 0.281, 0], with a log-likelihood of -741.88. The length parameters, as well as the speed SD, are fairly consistent with those of the 18 : 59 train.

Coming to the incomplete congestion model, the $[\tau_1, \tau_2]$ interval is set to [48s,68s], based on our ex-ante convention. The MLE of $\Theta_{(\ell,w)}$ yielded an optimized vector of parameters of [96.66, 20.76, 0.280, 0] for $[m_\ell, s_\ell, s_w, \chi]$, with a log-likelihood value of -739.21. These estimates are very close to those of the FF model.

Lastly, we applied the full congestion model to estimate both the walking parameters $\Theta_{(\ell,w)}$ and the queuing parameters Θ_Q. The estimated values are [95.600, 21.70, 0.279, -0.07] for $[m_\ell, s_\ell, s_w, \chi]$ and [3.190, 63, 83, 0.798] for $[\ell^*, \tau_1^*, \tau_2^*, \nu^*]$, with a log-likelihood value of −738.638. Again, the estimate of $\Theta_{(\ell,w)}$ is very close to those in the FF and IC models. The nonzero value of covariance χ is close to zero, and it may be neglected with no loss of statistical significance (log-likelihood level at -738.77). As for queuing parameters, the queuing focal point is located at 3.2 m from the validation gates and the average queue moving speed of 0.798 m/s is both plausible and consistent with the estimation for the 18 : 59 train.

Applying the ex-post analysis of §5.1.1 to the FC outcomes, the distribution of w^{-1} has a relative dispersion of 23% and a mean of 0.88 s/m. Then, the underlying distribution of free-flow egress times has a relative dispersion of 0.32 and mean of 84s and SD of 26.88. Length variations contribute to 47% of egress time variations, yielding signal-to-noise of 90%—on the other side of 1 compared to the 18 : 35 train.

Looking into the PDF profiles, the congestion range of [63s, 83s] estimated under the FC model looks more relevant than the provisional convention of [46s, 105s]. The associated bottleneck capacity amounts to $156 \times .017 s^{-1} = 2.6$ people per second.

Overall, the 18 : 32 train yields a large alighting flow giving rise to some queuing, in a lighter way than the 18 : 59 train. Modeling the light queuing by the IC and FC models, rather than neglecting it by keeping to the FF model, provides an improvement of 2 or 3 points in log-likelihood. This falls in a gray area between "little significance" and "marked significance."

6.2.4. Train at 18 : 46 : 33. The 18 : 46 train has a service interval of 11′ and disembarks 230 alighting passengers. Under the free-flow model with m_w set to 1.2 m/s, the estimated parameters are [104.37, 27.84, 0.357, 0.761] for $\Theta_{(\ell,w)} \equiv [m_\ell, s_\ell, s_w, \chi]$, with a log-likelihood value of -1163.56. The length parameters are consistent with our previous findings, whereas the speed SD is higher by one-fourth. The non-null though small estimate of the covariance parameter suggests the occurrence of congestion, which is further substantiated by the empirical PDF profile in Figure 8(a).

We then resorted to the incomplete congestion model, by setting χ to zero and conventional congestion range of [44s, 120s]. The resulting estimate is $[m_\ell, s_\ell, s_w] = [98.29, 22.01, 0.376]$, with a log-likelihood value of -1160.31—a 3 point improvement upon the FF model. The length parameters have estimates consistent with the FF model and the other trains. The speed SD estimate is consistent with the FF model only.

Turning to the full congestion model and making the queuing parameters endogenous, the log-likelihood was further increased to -1158.91 – indeed a small improvement, again suggesting light congestion for that train arrival. The walking and queuing parameters have respective estimates of $[m_\ell, s_\ell, s_w, \chi] = [102.5, 24.36, 0.352, -0.10]$ and $[\ell^*, \tau_1^*, \tau_2^*, \nu^*] = [4.340, 56.06, 87.07, 0.880]$. Once again the queue focal point is located about 4m from the validation

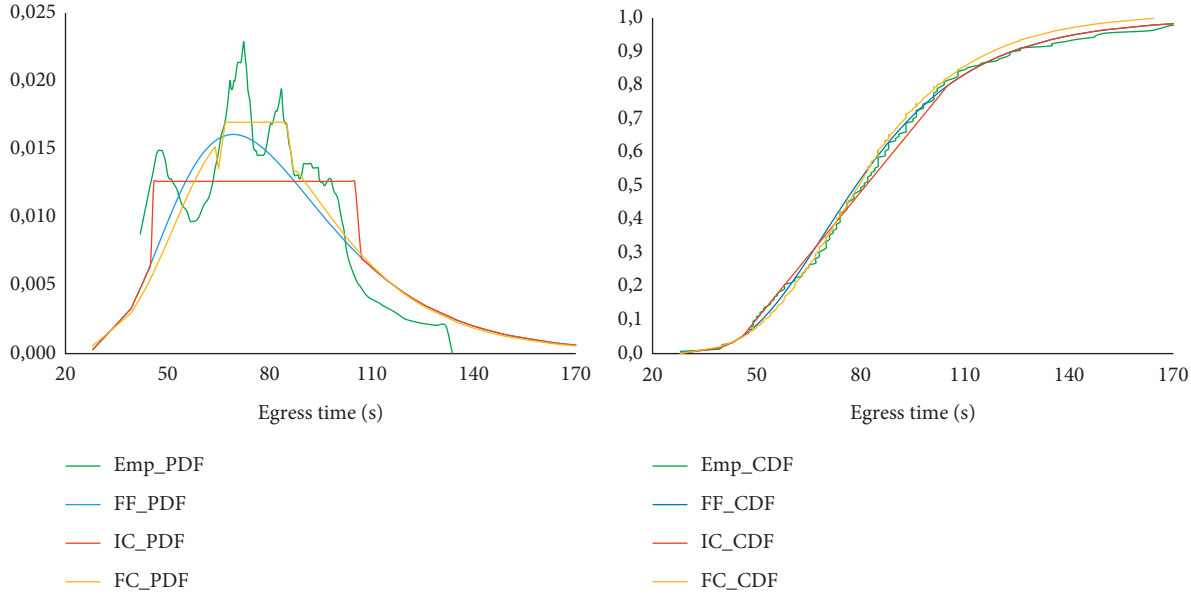

Figure 7: PDF (a) and CDF (b) of egress times for the train at 18 : 32 : 14.

gates, and the queue moving speed is close to 0.9 m/s. The queuing interval of [56 s, 87 s] is much shorter than under our provisional convention. As for the walking parameters, the length parameters are consistent with our previous findings both for that train and the other ones; the speed SD is similar to its FF and IC counterparts.

Applying the ex-post analysis in §5.1.1 to the FC outcomes, the underlying free-flow distribution of w^{-1} has a relative dispersion of 29% and a mean of 0.90 s/m. Then, the underlying distribution of free-flow egress times has relative dispersion of 0.38 and mean of 92.25s and SD of 35s. Length variations contribute to 39% of egress time variations, yielding signal-to-noise of 33%—that is, speed variations would be twice more influent than length variations in the variations of free-flow egress times.

6.3. Synthesis. We presented the estimation of the three traffic models FF/IC/FC for four trains with different levels of alighting flow yet taken from the same time period–half an hour in the evening peak on a given working day. The presentation was ordered so as to demonstrate first the FF model on the basis of the 18 : 35 train and then the congested models using the 18 : 59 train for which it is most advantageous to model congestion in an explicit way. These two trains constitute the extreme points of a range, which encompasses the two other trains at 18 : 32 and 18 : 46.

On MLE computation. For every train, we applied the traffic model in a progressive way, from FF to IC and then to FC. This enabled us to compare the resulting estimates for that train and to characterize the queuing phenomenon progressively. The application of MLE to the FF model is easy: using the standard optimization algorithms provided in Excel as well as in *Python* libraries, the convergence was straightforward. Also endowed with straightforward convergence is the application of MLE to the incomplete congested model under exogenous queuing parameters—the ex-ante determination of the $[\tau_1, \tau_2]$ interval under the educated guess of a 2p/s exit capacity. But the endogenization of the queuing parameters introduces discontinuities in the log-likelihood function, thereby making the numerical optimization a much more demanding task. In practice, using an Excel spreadsheet for each train, we resorted to a heuristic alternation of automated search (using the Excel solver) and manual adjustment to get to the "optimal points," where local optimization was obtained. We got satisfied with the resulting estimations because they induced fairly good reproduction of the PDF and CDF profiles of the empirical distributions of the egress times.

At the train level, the model estimation enables us to recover the underlying distribution of walking lengths and speeds, under preset mean speed to ensure identifiability. The estimates of walk speed SD range from 0.28 to 0.38 m/s, with some interplay with the covariance parameter. From the free-flow model applied to the fluid train at 18 : 35, the statistical independence between walk speeds and lengths is strongly supported. Conversely, a nonzero covariance estimate would capture some part of the queuing phenomenon. This may be useful when applying the free-flow model in order to detect the occurrence of congestion—therefore calling for the application of congested models.

As for the distribution of walk lengths, the mean and SD estimates of the congested models are consistent at the train level: they differ from the corresponding free-flow estimates by a reduction in both the mean and SD, meaning that neglecting the queuing phenomenon by applying only the FF model will lead to biased results.

The estimates of the mean length parameters seem to vary in a systematic way depending on the train: larger alighting flow comes along with larger mean length, meaning larger alighting positions from the train head in the

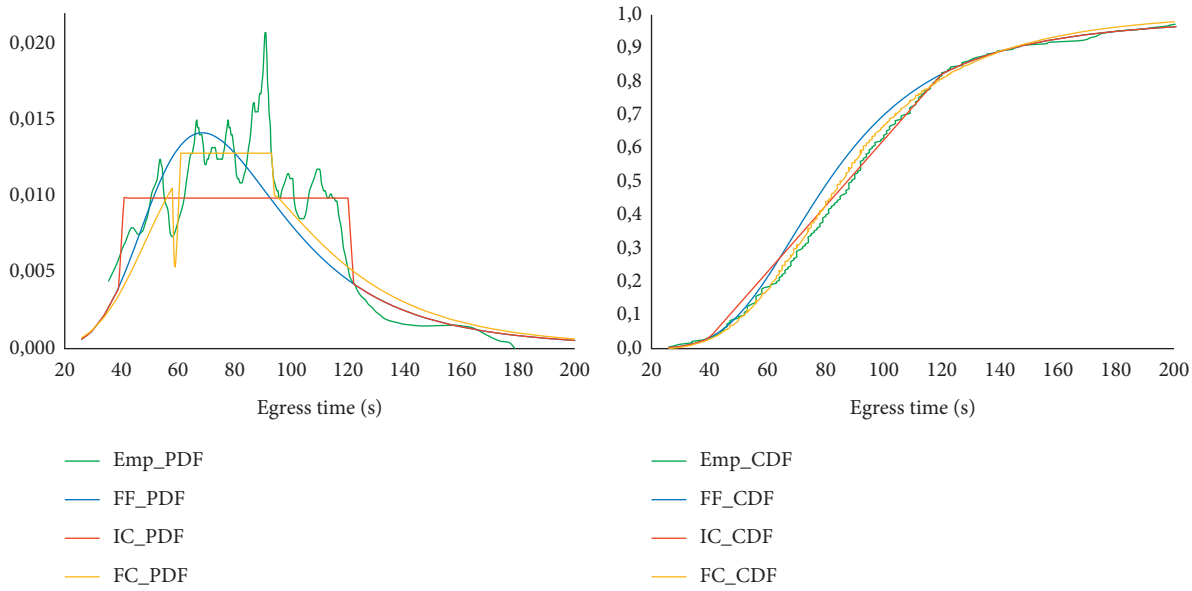

Figure 8: PDF (a) and CDF (b) of egress times for the train at 18 : 46 : 33.

platform configuration at Noisy-Champs. This suggests that on-board positions are influenced both by the exit conditions and the passenger load conditions along the train.

Coming to the *queuing phenomenon*, the bottleneck behavior has been evidenced by three out of the four trains in the half hour period under study. The time range of queuing depends on the train: the larger alighting flow decreases the queue beginning time and increases its end time. While empirical data exhibit significant instantaneous variations in the exit flow rate, the bottleneck postulate enables us to identify the average exit capacity in a straightforward way. From the two most congested trains, this capacity falls in the range of [2.5, 3] p/s; that is, it is much higher than our ex-ante convention of 2 p/s.

The notion of queuing focal point ℓ^* is supported by the estimation results of the FC model. The estimated positions at about 4m from the validation gates correspond to the exit point of the vertical element, which combines an escalator and a stairway. The queue moving speed v^* was estimated consistently between the three congested trains, around 0.9 m/s, that is, one-fourth less than the mean free-flow walk speed. The queue focal point and the queue moving speed are strongly complementary parameters: their ratio $t^* = \ell^*/v^*$ is a propagation time to transport the queuing time range from ℓ^* to length 0.

7. Conclusion

7.1. Summary

Physical Theory and Stochastic Model. Individual egress times from train alighting to station exit constitute a statistical population at the train level, with much variability across the individuals. We modeled the magnitude and variations of egress times as a random variable and captured its dependencies onto underlying factors of (i) walk length, (ii) pedestrian speed, (iii) and possibly congestion among the alighting passengers (in the form of a traffic bottleneck). As the train is long, the alighting positions are stretched out over it, giving rise to distributed walk lengths. Also distributed among the individual users are the walk speeds in the "free-flow" regime. In turn, so is the egress time under free flow (FF). We provided a stochastic model of FF egress times, with explicit analytical formulas for its CDF and PDF depending on the joint distribution of walk lengths and FF speeds. As for congestion, we modeled it in the form of a traffic bottleneck based on a queue focal point ℓ^*, a congested interval $[\tau_1^*, \tau_2^*]$ at that point and a queue moving speed v^* up to the station exit. Decomposing the train population of egress times depending on whether the user would pass at ℓ^* before, after, or during the queued interval, an explicit analytical formula was obtained for the PDF of the egress time in the so-called full congestion (FC) model. Thus, both the FF and FC models are endowed with analytical formulas. Between them, an intermediary model called "incomplete congestion" (IC) assimilates the queue focal point to the station egress point.

By postulating a bivariate Gaussian distribution for the walk pairs of length and FF speed, straightforward computable formulas are available for the CDF and PDF of the FF egress time and for the PDF of the IC/FC egress time.

Estimation Methodology. The physical and stochastic model of egress times pertains to a triple of train, platform exit point and station access point, since the distribution of walk lengths depends on the positions on board the train of the alighting users, as well as on the walk pathway topology. Under free flow, we devised a simple estimation method to recover the length distribution from that of egress times, under exogenous distribution of walk speeds. More generally, a maximum-likelihood estimation (MLE) method was devised on the

basis of the PDF and CDF formulas to construct the likelihood function of the model parameters depending on an observed egress time.

Case Study. The model and its MLE method were applied to the train station "Noisy-Champs" in Paris. The four trains serving the eastwards platform during half an hour on the evening peak of a typical workday were studied on an individual basis. As their respective alighting flows are contrasted, one gives rise to free-flow conditions, whereas the other three experience a queuing episode. The datasets of egress times were constituted from AFC records of passenger station exit times, related to AVL records of train platform arrival times. Estimation results were reported and commented, as well as the MLE applicability.

7.2. Outreach, Limitations, and Further Research. The model is sensitive to train characteristics: notably, the probability distribution of the alighting positions and also the alighting flow volume. It is also sensitive to platform and station features, through the distribution of walk lengths as well the walk pathway and its width available to pedestrians. FF walk speeds are featured as preferred speeds at the individual level, that is, cruising speeds rather than instantaneous speeds.

The traffic theory in the model pertains to a specific kind of congestion among the train-alighting users, with little or no disruption by other users waiting for boarding or staying on board.

The model involves a simple topological configuration for the triple of train dwelling position, platform exit point, and station access point (validation gates). When dealing with a platform exit point situated at an intermediary position on the dwelling length, the alighting positions must be considered with respect to that point.

As for the estimation methodology, it requires the identification of the abovementioned triple in the dataset of train egress times. Such dataset enables one to identify all except one components in the parameter vector of the walk-length pair distribution, and all of the queuing parameters when queuing occurs.

Our four cases of model application demonstrate the model ability to simulate the empirical distribution of egress times in an efficient way. Both the walking features and the queuing characteristics were uncovered, with outcomes endowed with much plausibility. The assumption of a bivariate Gaussian distribution of walk lengths and free-flow speeds is merely instrumental: it enables for straightforward interpretation of the estimation results as well as for easy computation. More general distribution may also be considered: yet the computational cost would be higher.

The discontinuity of the likelihood function for the congested model is more problematic. It may be attacked from three different sides: either by refining the physical theory to smooth out the contours of the queuing episode, or by refining the stochastic theory to allow for some speed variability at the individual level, or by utilizing a more sophisticated optimization algorithm to deal with discontinuity and local optima (e.g., a genetic algorithm).

Other directions for further research include the following:

(i) The critical review of the likelihood function, with special attention to the queuing part and the associated assumptions of independence

(ii) The consideration of nonparametric estimation for the walk and speed distribution

(iii) The theorization of other kinds of passenger congestion, involving the alighting users in interaction with other platform users that wait for boarding or simply go through it form an entry point to another exit point, and maybe also with those train users remaining on board

Appendix

In the model with Gaussian components, the bivariate vector (ℓ, w) has mean vector $m = (m_\ell, m_w)$ and covariance matrix $\Sigma = \begin{bmatrix} s_\ell^2 & \chi \\ \chi & s_w^2 \end{bmatrix}$, which is inverted as $\Sigma^{-1} = 1/\Delta \begin{bmatrix} s_w^2 & -\chi \\ -\chi & s_\ell^2 \end{bmatrix}$, wherein $\Delta \equiv \det(\Sigma) = s_\ell^2 s_w^2 - \chi^2$. Define $\rho \equiv \chi/(s_\ell \cdot s_w)$ such that $\Delta = s_\ell^2 s_w^2 (1 - \rho^2)$.

Now, letting $\omega \equiv [(\ell, w) - m]^T \Sigma^{-1} [(\ell, w) - m]$, the joint PDF of the (ℓ, w) pair is

$$f(\ell, w) = \frac{1}{2\pi\sqrt{\Delta}} \exp\left(-\frac{1}{2}\omega\right). \tag{A.1}$$

A. Distribution of Lengths Conditionally to Speed

Conditionally to w, the ω function reduces to a function of ℓ only, such that

$$\begin{aligned} \omega.\Delta &= s_w^2 (\ell - m_\ell)^2 + s_\ell^2 (w - m_w)^2 - 2\chi(\ell - m_\ell)(w - m_w) \\ &= s_w^2 \left(\ell_e^2 - 2\ell\left\{m_\ell + (w - m_w)\chi s_w^{-2}\right\}\right) + \left[s_w^2 m_\ell^2 + s_\ell^2 (w - m_w)^2 - 2\chi m_\ell (w - m_w)\right]. \end{aligned} \tag{A.2}$$

Letting $s^2_{\ell|w} \equiv \Delta/s^2_w = s^2_\ell(1-\rho^2)$ and $m_{\ell|w} \equiv m_\ell + (w - m_w)\beta$ for $\beta \equiv \chi/s^2_w$, the conditional variable ℓ_w has PDF proportional to $\exp(-(1/2)\,\omega)$, hence to $\exp(-(1/2)(\ell - m_{\ell|w}/s_{\ell|w})^2)$: thus, it is a Gaussian variable, ℓ_w is $N(m_{\ell|w}, s^2_{\ell|w})$.

B. Auxiliary Function

Let us now consider function $\overline{M}_x(y) \equiv \int_y^\infty w\dot{S}_w(w\cdot x)dW(w)$: it holds that

$$\overline{M}_x(y) = \int_y^\infty \left(w/s_{\ell|w}\right)\phi\left(\left(w.x - m_{\ell|w}\right)/s_{\ell|w}\right)\frac{1}{s_w}\phi\left(\frac{w-m_w}{s_w}\right)dw$$
$$= K_x\frac{s_x}{s_{\ell|w}.s_w}\left(m_x\Phi\left(\frac{m_x-y}{s_x}\right) + s_x\phi\left(\frac{y-m_x}{s_x}\right)\right), \tag{B.1}$$

whereas $s_x \equiv ((x-\beta)^2 s^{-2}_{\ell|w} + s^{-2}_w)^{-(1/2)}$, $m_x \equiv s^2_x((x-\beta)(m_\ell - \beta m_w)/s^2_{\ell|w} + m_w/s^2_w)$, and $K_x = (1/\sqrt{2\pi})\exp(1/2)((m_x/s_x)^2 - (m_\ell - \beta m_w/s_{\ell|w})^2 - (m_w/s_w)^2)$.

Proof. Consider first the product $\phi((w.x - m_{\ell|w})/s_{\ell|w})\phi((w-m_w)/s_w) = (1/2\pi)\exp - (1/2)(((w.x - m_{\ell|w})/s_{\ell|w})^2 + ((w - m_w)/s_w)^2)$.

It holds that

$$\left(\frac{w.x - m_{\ell|w}}{s_{\ell|w}}\right)^2 + \left(\frac{w-m_w}{s_w}\right)^2 = \left(\frac{w(x-\beta) - (m_\ell - \beta m_w)}{s_{\ell|w}}\right)^2 + \left(\frac{w-m_w}{s_w}\right)^2$$
$$= \left(\frac{w - (m_\ell - \beta m_w)/(x-\beta)}{s_{\ell|w}/(x-\beta)}\right)^2 + \left(\frac{w-m_w}{s_w}\right)^2$$
$$= w^2\left((x-\beta)^2 s^{-2}_{\ell|w} + s^{-2}_w\right) - 2w\left(\frac{(m_\ell - \beta m_w)(x-\beta)^{-1}}{s^2_{\ell|w}(x-\beta)^{-2}} + \frac{m_w}{s^2_w}\right) + \left(\frac{m_\ell - \beta m_w}{s_{\ell|w}}\right)^2 + \left(\frac{m_w}{s_w}\right)^2. \tag{B.2}$$

Hence, the definition of s_x and m_x.

Thus,

$$\left(\frac{w.x - m_{\ell|w}}{s_{\ell|w}}\right)^2 + \left(\frac{w-m_w}{s_w}\right)^2 = \left(\frac{w}{s_x}\right)^2 - 2\frac{w}{s_x}\frac{m_x}{s_x} + \left(\frac{m_\ell - \beta m_w}{s_{\ell|w}}\right)^2 + \left(\frac{m_w}{s_w}\right)^2$$
$$= \left(\frac{w-m_x}{s_x}\right)^2 - \left(\frac{m_x}{s_x}\right)^2 + \left(\frac{m_\ell - \beta m_w}{s_{\ell|w}}\right)^2 + \left(\frac{m_w}{s_w}\right)^2. \tag{B.3}$$

Thus, $\phi((w.x - m_{\ell|w})/s_{\ell|w})\phi((w-m_w)/s_w) = (1/\sqrt{2\pi})\phi((w-m_x)/s_x)\exp(1/2)((m_x/s_x)^2 - ((m_\ell - \beta m_w)/s_{\ell|w})^2 - (m_w/s_w)^2)$, hence the definition of K_x.

By substitution, we get that

$$\int_y^\infty w\cdot\dot{S}_w(w\cdot x)dW(w) = K_x\frac{s_x}{s_{\ell|w}\cdot s_w}\int_y^\infty (w/s_x)\phi((w-m_x)/s_x)dw. \tag{B.4}$$

The part $\int_y^\infty (w/s_x)\phi((w-m_x)/s_x)dw = m_x\int_y^\infty (1/s_x)\phi((w-m_x)/s_x)dw + \int_y^\infty ((w-m_x)/s_x)\phi((w-m_x)/s_x)dw$

$$
\begin{aligned}
&= m_x\left[\Phi\left(\frac{w-m_x}{s_x}\right)\right]_y^{\infty} + \int_y^{\infty}(w-m_x)/s_x\,\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\right)\left(\left(\frac{w-m_x}{s_x}\right)^2\right)dw\ ,\\
&= m_x\left[1-\Phi\left(\frac{y-m_x}{s_x}\right)\right] - s_x\left[\phi\left(\frac{w-m_x}{s_x}\right)\right]_y^{\infty}, \qquad \text{(B.5)}\\
&= m_x\Phi\left(\frac{m_x-y}{s_x}\right) + s_x\phi\left(\frac{y-m_x}{s_x}\right).
\end{aligned}
$$

Data Availability

Data are available on request from the authors.

Additional Points

A physical theory of platform egress times, involving walking features together with a pedestrian traffic bottleneck. Analytical model of egress time PDF, either free flow or congested, depending on the joint distribution of walk lengths and free-flow speeds, together with queuing characteristics. Straightforward numerical computation for bivariate Gaussian pair of length and speed. Maximum-likelihood estimation method at the train level to uncover the underlying walking features as well as the queuing characteristics. Application report for four contrasted trains of urban mass transit in the Paris urban area.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J.-D. Schmöcker, M. G. H. Bell, and F. Kurauchi, "A quasi-dynamic capacity constrained frequency-based transit assignment model," *Transportation Research Part B: Methodological*, vol. 42, no. 10, pp. 925–945, 2008.

[2] G. Gentile and K. Noeckel, *Modelling Public Transport Passenger Flows in the Era of Intelligent Transport Systems*, Springer Tracts on Transportation and Traffic, vol. 10, 641 pages, Berlin, Germany, 2016.

[3] Z. Christoforou, P.-A. Collet, B. Kabalan et al., "Influencing longitudinal passenger distribution on railway platforms to shorten and regularize train dwell times," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2648, no. 1, pp. 117–125, 2017.

[4] 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.

[5] Kittelson, *Transit Capacity and Quality of Service Manual*, Vol. 165, published by the Transportation Research Board, , Washington, DC, USA, 2013.

[6] T. Parkinson and I. Fisher, *Rail transit capacity*, Vol. 13, TCHRP Report, Transportation Research Board, , Washington, DC, USA, 1996.

[7] F. Leurent, "Transport capacity constraints on the mass transit system: a systemic analysis," *European Transport Research Review*, vol. 3, no. 1, pp. 11–21, 2011.

[8] Z. Liu, D. Li, and X. Wang, "Passenger distribution and waiting position selection model on metro station platform," in *Proceedings of the 2016 International Conference on Civil, Structure And Environmental Engineering*, pp. 35–41, Atlantis Press, Guangzhou, China, April 2016.

[9] S. P. Hoogendoorn, W. Daamen, V. L. Knoop, J. Steenbakkers, and M. Sarvi, "Macroscopic fundamental diagram for pedestrian networks: theory and applications," *Transportation Research Part C: Emerging Technologies*, vol. 94, pp. 172–184, 2018.

[10] Y. Zhang, E. Jenelius, and K. Kottenhoff, "Impact of real-time crowding information: a Stockholm metro pilot study," *Public Transport*, vol. 9, no. 3, pp. 483–499, 2017.

[11] J. C. Muñoz, J. Soza-Parra, A. Didier, and C. Silva, "Alleviating a subway bottleneck through a platform gate," *Transportation Research Part A: Policy and Practice*, vol. 116, pp. 446–455, 2018.

[12] R. B. Dial, "Transit pathfinder algorithms," *Highway Research Record*, vol. 205, pp. 67–85, 1967.

[13] H. Spiess and M. Florian, "Optimal strategies: a new assignment model for transit networks," *Transportation Research Part B: Methodological*, vol. 23, no. 2, pp. 83–102, 1989.

[14] M. H. Poon, S. C. Wong, and C. O. Tong, "A dynamic schedule-based model for congested transit networks," *Transportation Research Part B: Methodological*, vol. 38, no. 4, pp. 343–368, 2004.

[15] Y. Hamdouch and S. Lawphongpanich, "Schedule-based transit assignment model with travel strategies and capacity constraints," *Transportation Research Part B: Methodological*, vol. 42, no. 7-8, pp. 663–684, 2008.

[16] F. S. Hänseler, J. P. A. van den Heuvel, O. Cats, W. Daamen, and S. P. Hoogendoorn, "A passenger-pedestrian model to assess platform and train usage from automated data," *Transportation Research Part A: Policy and Practice*, vol. 132, pp. 948–968, 2020.

[17] Q. Zhang, B. Han, and D. Li, "Modeling and simulation of passenger alighting and boarding movement in Beijing metro stations," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 5, pp. 635–649, 2008.

[18] M. Haghani and M. Sarvi, "Human exit choice in crowded built environments: investigating underlying behavioural differences between normal egress and emergency evacuations," *Fire Safety Journal*, vol. 85, pp. 1–9, 2016.

[19] X. Ji, J. Zhang, and B. Ran, "A study on pedestrian choice between stairway and escalator in the transfer station based on floor field cellular automata," *Physica A: Statistical Mechanics and Its Applications*, vol. 392, no. 20, pp. 5089–5100, 2013.

[20] O. Cats, J. West, and J. Eliasson, "A dynamic stochastic model for evaluating congestion and crowding effects in transit systems," *Transportation Research Part B: Methodological*, vol. 89, pp. 43–57, 2016.

[21] L. Sun, D. H. Lee, A. Erath, and X. Huang, "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*, pp. 142–148, Beijing, China, August 2012.

[22] 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.

[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] Y. Zhu, H. N. Koutsopoulos, and N. H. M. Wilson, "Inferring left behind passengers in congested metro systems from automated data," *Transportation Research Part C: Emerging Technologies*, vol. 94, pp. 323–337, 2018.

[25] F. Leurent and J. Thomas, *A Probabilistic Traffic Model to Estimate Fail-To-Board Probabilities in Transit Lines on the Basis of AFC and AVL Data*, Open Archive HAL, CNRS, France, 2019, https://hal.archives-ouvertes.fr/hal-02462453.

[26] 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.

[27] F. Leurent and X. Xie, "On individual repositioning distance along platform during train waiting," *Journal of Advanced Transportation*, vol. 2018, Article ID 4264528, 18 pages, 2018.

[28] A. Iida, S. Tamura, Y. Yamashita, and J. Zhang, *Etude du temps de sortie des trains. TransitLab team project report*, 43 pages, ENPC, Paris, 2021.

[29] L. Meschini, G. Gentile, and N. Papola, "A frequency based transit model for dynamic traffic assignment to multimodal networks," in *Proceedings of the 17th International Symposium on Transportation and Traffic Theory*, vol. 2, no. 3, pp. 2-3, London, UK, July 2007.

[30] E. C. Paul, *Estimating Train Passenger Load from Automated Data Systems: Application to London Underground*, Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA, 2010.

[31] V. Trozzi, G. Gentile, M. Bell, and I. Kaparias, "Dynamic user equilibrium in public transport networks with passenger congestion and hyperpaths," *Procedia - Social and Behavioral Sciences*, vol. 80, pp. 427–454, 2013.