

A passenger-pedestrian model to assess platform and train usage from automated data

Hänseler, Flurin S.; van den Heuvel, Jeroen P.A.; Cats, Oded; Daamen, Winnie; Hoogendoorn, Serge P.

DOI

[10.1016/j.tra.2019.12.032](https://doi.org/10.1016/j.tra.2019.12.032)

Publication date

2020

Document Version

Final published version

Published in

Transportation Research Part A: Policy and Practice

Citation (APA)

Hänseler, F. S., van den Heuvel, J. P. A., Cats, O., Daamen, W., & Hoogendoorn, S. P. (2020). A passenger-pedestrian model to assess platform and train usage from automated data. *Transportation Research Part A: Policy and Practice*, 132, 948-968. <https://doi.org/10.1016/j.tra.2019.12.032>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Transportation Research Part A

journal homepage: www.elsevier.com/locate/tra

A passenger-pedestrian model to assess platform and train usage from automated data

Flurin S. Hänseler^{a,*}, Jeroen P.A. van den Heuvel^b, Oded Cats^a, Winnie Daamen^a, Serge P. Hoogendoorn^a

^a Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, the Netherlands

^b NS Stations, the Netherlands



ARTICLE INFO

Keywords:

Public transportation
Crowding
Level-of-service assessment
Investment appraisal

ABSTRACT

We present a transit model that, based on automated fare collection and train tracking data, describes pedestrian movements in train stations and vehicle-specific train ridership distributions. Our approach differs from existing models in that we describe on-board passenger dynamics and pedestrian dynamics at stations in a joint framework. We assume that travelers first decide on the train(s) they will take, and then pursue their journey through the network by successively choosing pedestrian paths, waiting positions on platforms, and specific train cars. Travelers explicitly maximize their travel utility. We model how crowding influences their walking speeds, and how it affects travel utility both at stations and on-board. To illustrate the framework, we present a real case study of a major Dutch rail corridor, for which we collect a rich set of passenger, pedestrian and train operation data. We observe a good agreement of model estimates with empirical observations, and discuss the use of the model for various transit-related problems including level-of-service assessment, crowding estimation, transit optimization, and integrated investment appraisal.

1. Introduction

A key characteristic of transit systems is their level of crowding, i.e., the accumulation of travelers on platforms, access ways, and in trains (Haywood et al., 2017). Typically, the number of travelers in a transit system is subject to strong spatiotemporal fluctuations (Hermant, 2012). Understanding these fluctuations, as well as the underlying interactions between travelers, infrastructure and rolling stock, is of key relevance to improve safety, comfort and efficiency (Raveau et al., 2014).

Significant progress has been made in understanding passenger dynamics at the level of transit networks, as well as in capturing pedestrian dynamics at the level of transit stations. For instance, fare collection and train tracking data are now routinely used to estimate train ridership (Zhu et al., 2017), and transit assignment models are employed to analyze and mitigate disruptions (Cats et al., 2016). Likewise, the measurement, analysis and modeling of pedestrian dynamics in train and metro stations has received considerable research attention (Xu et al., 2014; Hänseler et al., 2016).

Certain problems, such as the estimation of platform usage or the balancing of investment funds between rolling stock and pedestrian facilities, require simultaneous consideration of network-wide and station-specific traveler dynamics. Indeed, the usage of a train or platform depends on the position of access ways, and on the length and stopping position of trains in each station (Kim

* Corresponding author at: Department of Transport & Planning, Delft University of Technology, Stev-inweg 1, 2628 CN Delft, The Netherlands.

E-mail addresses: f.s.hanseler@tudelft.nl (F.S. Hänseler), jeroen.vandenheuvel@nsstations.nl (J.P.A. van den Heuvel), o.cats@tudelft.nl (O. Cats), w.daamen@tudelft.nl (W. Daamen), s.p.hoogendoorn@tudelft.nl (S.P. Hoogendoorn).

<https://doi.org/10.1016/j.tra.2019.12.032>

Received 15 June 2018; Received in revised form 14 November 2019; Accepted 28 December 2019

0965-8564/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

et al., 2014). Such usage patterns are a key determinant of dwell times, platform safety, or station throughput (Szplett and Wirasinghe, 1984; van den Heuvel, 2016). So far, no modeling framework exists that explicitly takes into account the intertwining of traveler dynamics at the network and station level.

The objective of this paper is to present a passenger-pedestrian model that, based on automated transit data, can describe traveler dynamics both in stations and in trains. The solution to this problem facilitates a number of applications, including (i) detailed estimation of platform usage for safety analysis or performance optimization (e.g. by adjusting train stop positions), (ii) analysis of car-specific on-board accumulation for comfort assessment, or for crowding information systems, (iii) investment appraisal when both in-vehicle and at-station benefits are relevant, and, indirectly, (iv) disruption management.

Our approach is inspired by agent-based transit assignment models, which we extend by a pedestrian model to capture walking and waiting behavior in stations. We consider travelers as individual decision makers whose movement along a transit trip is described both while riding and while moving and waiting at stations. Users are assumed to first decide on their transit itinerary, before considering choices at the station level, such as the choice of a specific walking route, or of a waiting position on a platform (Hoogendoorn and Bovy, 2004; Bureau Spoorbouwmeester, 2012). In describing the traveler dynamics, we rely where possible on data, namely automated fare collection (AFC) and train tracking data. Such data is increasingly available, and contributes to a high model accuracy.

In the remainder, we first review the relevant body of literature, present the model framework, and then apply it to a real case study.

2. Background

We briefly present related transit assignment models, focusing on simulation-based and data-driven approaches, as well as pedestrian movement models for train stations. We then propose a new approach that is hybrid in that it considers both the network and station level, and in that it contains both model- and data-driven components.

2.1. Approaches towards modeling transit assignment

Transit assignment models describe the propagation of travelers within a network, taking passenger demand and network characteristics as input. A large number of models have been proposed, for instance within a frequency- or schedule-based framework (Spiess and Florian, 1989; Nuzzolo et al., 2001). In the context of this work, disaggregate approaches are of interest, i.e., those that consider individual travelers and train services. Among them, we distinguish agent-based models and data-driven approaches.

Agent-based assignment models explicitly describe traveler behavior. Relying on simulation, they consider behavioral aspects such as route, train or departure time choice. One of the first agent-based transit models is due to Wahba and Shalaby (2005), who assume that passengers choose their transit itinerary, stop and departure time based on a desired arrival time and a given origin-destination pair. By traveling repeatedly through the network, passengers build up individual knowledge of walking, waiting and in-vehicle cost, on which their decision making is based (see also Teklu, 2008). More advanced frameworks have been proposed by Toledo et al. (2010) and Cats et al. (2016), who focus on dynamic congestion effects e.g. in the form of on-board crowding, denied boardings or service reliability.

Data-driven approaches focus on the assignment of passengers to individual trains by relying on automated transit data (Pelletier et al., 2011). A wide range of approaches has been proposed, with early work employing rule-based methods (e.g. Bagchi and White, 2005), and more recent work focusing on probabilistic frameworks (Hörcher et al., 2017; Zhu et al., 2017). Increasing availability of AFC and train tracking data has enabled several widely cited case studies e.g. in Japan (Kusakabe et al., 2010), China (Sun and Xu, 2012) or Singapore (Sun et al., 2012).

2.2. Station models

At the level of transit stations, pedestrian traffic assignment models are used to describe local travel activities, way finding and movements of passengers. They are henceforth referred to as ‘station models.’ Station models are useful to predict the level-of-service of rail access facilities, and may be loosely divided into macro-, micro- and mesoscopic approaches. Macroscopic station models consider pedestrians as a continuum, and comprise network flow models (Lee et al., 2001), cell transmission models (Hänseler et al., 2014), or models based on Petri-nets (Kaakai et al., 2007). Microscopic approaches consider pedestrians and their interactions individually, including activity-based models (Hoogendoorn and Bovy, 2004; Campanella, 2016), social force-based models (King et al., 2014; Helbing and Molnár, 1995), or cellular automata (Davidich et al., 2013). Mesoscopic station models consider individual pedestrians, but describe their movements in terms of macroscopic relationships (Tordeux et al., 2018). Among the better-known approaches are queueing systems (Xu et al., 2014; Jiang et al., 2015) and hybrid models (Daamen, 2004).

2.3. Research approach

The above-mentioned approaches describe traveler dynamics at the network- and the station-level separately. They neglect how the two levels interact. In practice, at-station pedestrian behavior and in-vehicle passenger behavior are intertwined in several ways:

- The distribution of passengers on platforms influences on-board crowding and vice versa (Lam et al., 1999; Liu et al., 2016).

- Travel utility, a central factor in traveler behavior, comprises components pertaining to both in-vehicle and at-station time (Tirachini et al., 2013; Guo and Wilson, 2011).
- Demand micro-patterns caused by arriving and departing trains yield circular dependencies between congestion in pedestrian facilities and boarding and alighting flows (Hermant, 2012; Hänseler et al., 2017b).

Our aim is to integrate the two levels of traveler dynamics in a way that is computationally feasible for large-scale applications, and consistent. By the latter we mean that characteristics of travelers are preserved along their transit trip, and that they take non-local consequences of their decisions into account (such as how the choice of a train car influences walking distances at their destination and vice versa). This notably precludes a station model (Srikukenthiran, 2015).

To achieve this aim, we envisage an agent-based transit assignment model that fully incorporates a mesoscopic pedestrian model for each station. We assume that fare transaction data and train tracking data are available. Hence, we rely on data for processes that are readily observable (namely train operations and passenger demand), and on behavioral models where direct observations are difficult (such as for platform dynamics). Our focus is on busy railway networks with considerable crowding (i.e., passengers may not be seated, and level-of-service on platforms may be low), but we assume that passengers are able to board any desired train (no denied boardings).

Our main contribution lies in the development of a hybrid approach that consistently describes traveler behavior both on-board and at stations. Our approach is capable of estimating spatiotemporal distributions of crowding levels on platforms and in trains, which is of direct use for safety and comfort considerations. Moreover, it serves as an important building block for real-time applications, for instance for information provision systems.

3. Model framework

The proposed model requires disaggregate travel demand and train tracking data as input. In a hierarchical way, it first assigns each traveler a sequence of train runs, and then a detailed path that specifies which pedestrian route or train vehicles they have used. The model takes the influence of crowding on walking speeds and on path choice explicitly into account. Rich model output allows for detailed analysis of traveler dynamics both at the station and network level.

Fig. 1 shows a scheme of the proposed model.

Specification. We discuss each of its parts in the following. Alongside the more general framework, to illustrate the model, we present an example specification. It is geared towards commuting traffic in passenger rail networks, and focuses on elementary travel activities, namely walking, waiting, and riding, as well as level changes. Travelers are assumed to go straight to a platform when entering the railway system. Upon arrival, they go to some exit unless they seek a connecting train, in which case they walk to the corresponding platform (Davidich et al., 2013). No distinction is made between different fare levels, and travelers are assumed to be capable of using all infrastructures, including stairways and escalators. In Section 4, we will apply the example specification to a case study.

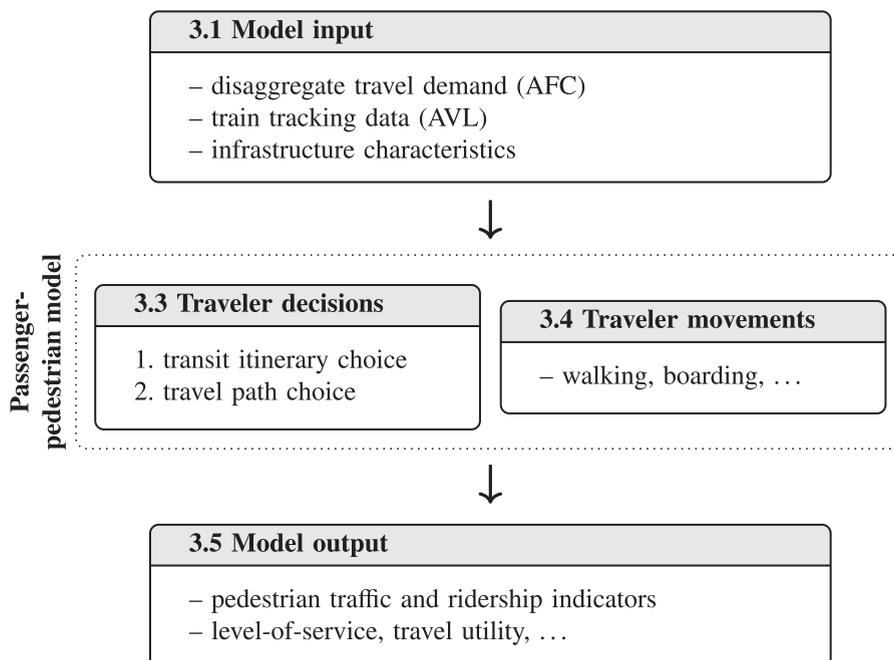


Fig. 1. Scheme of passenger-pedestrian model.

Appendix A provides an overview of the notation. Appendices C and D describe technical aspects relevant for an implementation of the model.

3.1. Model input

Three types of model inputs are required:

Disaggregate travel demand, specifying for each traveler a departure time, an origin, and a destination. Such information is increasingly available from automated fare collection systems. AFC systems can be classified based on: (i) whether they include tap-in only or tap-in and tap-out records, typically corresponding to flat and distance-based fare schemes; (ii) whether fare validation is bound to vehicles or station gates. Depending on these two aspects, destination (for boarding only) and vehicle (for station validation) inferences might be applied. In the absence of supplemental data, methods to infer the alighting station of a given tap-in record (e.g. Trépanier et al., 2017; Munizaga and Palma, 2012; Gordon et al., 2013) and for inferring the vehicle boarded by each passenger (Hörcher et al., 2017; Zhu et al., 2017) may be applied. Alternatively, if no disaggregate AFC data is available, travel demand may also be obtained from a high-level transit assignment model (Nuzzolo et al., 2012).

Train tracking data, describing the itinerary of train runs, their composition and capacity. Such data is often available from signaling and control systems. If train tracking data is unavailable, the planned timetable in combination with an appropriate delay model may be used (Higgins and Kozan, 1998; Goverde, 2007).

Infrastructure characteristics, describing the rail network and rail access facilities, are finally required. Infrastructure characteristics include the topology of train stations, specifying the dimensions and capacity of platforms, platform access ways, and station halls, as well as stopping positions of trains.

3.2. System representation

We consider individual travelers in a continuous-time, discrete-space model. Time is denoted by $t \in (t^-, t^+)$, where t^- and t^+ represent the start and end time of the period of interest. Space is represented by a set of train stations, \mathcal{S} , of which each station $s \in \mathcal{S}$ is associated with a directed graph \mathcal{G}_s representing its rail access facilities.

A station graph $\mathcal{G}_s = (\mathcal{N}_s, \mathcal{L}_s)$ is described by a set of nodes, \mathcal{N}_s , and a set of pedestrian links, \mathcal{L}_s (see Fig. 2). Certain elements of pedestrian facilities, such as stairs or corridors, translate naturally into links, and others naturally into nodes, such as entrance/exit points. For other elements, such as waiting halls or platforms, the decomposition into links and nodes depends on the desired granularity (Løvås, 1994). A node $n \in \mathcal{N}_s$ through which travelers enter or leave the pedestrian network is referred to as origin/destination node, and their set is denoted by $\mathcal{N}_{od}^s \subset \mathcal{N}_s$.

A pedestrian link $\ell: n_\ell^- \rightarrow n_\ell^+$ connects two nodes and has a fixed length $\lambda_\ell > 0$. Each link is associated with an area $a \in \mathcal{A}$, defining a space in which it interacts with other links. The surface size of an area a is denoted by α_a . The set of links associated with area a in station s is denoted by \mathcal{L}_a , with $\mathcal{L}_a \subset \mathcal{L}_s$ and $\mathcal{L}_a \cap \mathcal{L}_{a'} = \emptyset$ if $a \neq a'$.

A traveler $y \in \mathcal{Y}$ is associated with a vector of traveler attributes η_y , a time t_y^o at which they enter the transit system, and an origin and a destination node $n_y^o, n_y^d \in \mathcal{N}_{od}$. Origin and destination nodes typically represent an entrance/exit point at a train station, such as a ticket gate. The set of travelers \mathcal{Y} is assumed to be known a priori.

A train run $r \in \mathcal{R}$ is represented by a sequence of train stations (s_1, s_2, \dots) , where the set of train runs is denoted by \mathcal{R} . Each train run is associated with a set of cars, $\mathcal{C}_r = \{c_1, c_2, \dots\}$. For every visited station s , a train run r is characterized by an arrival time $t_{r,s}^{arr}$, a

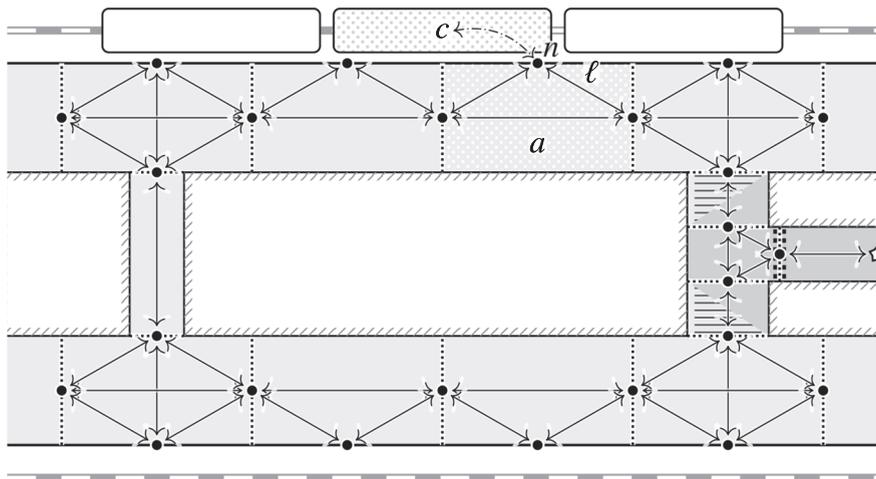


Fig. 2. Space representation of a train station. Walkable space (gray) is partitioned into areas (delimited by dotted lines), and represented by a network of directed links (pairs of links as double-headed arrows) that connect internal nodes (black circles) and origin-destination nodes (stars). A train, represented by individual train cars (rounded rectangles), is dwelling at the upper platform.

departure time $t_{r,s}^{\text{dep}}$ and a platform, which are all assumed known from train tracking data.

3.3. Traveler decisions

To describe traveler decisions, we introduce the notions of a transit itinerary and of a travel path. A transit itinerary i is a sequence of train runs between train stations, $(r_{s_1,s_2}, r'_{s_2,s_3}, \dots)$. A travel path h , on the other hand, is a sequence of path segments, $h = (w_1, w_2, \dots)$, where a segment w is represented either by a train car that carries a traveler between two train stations, or by a pedestrian link.

We assume that travelers first decide on a transit itinerary, subject to which they choose their travel path. This assumption is appropriate in the context of scheduled services with high regularity and no denied boardings.

Let \mathcal{I}_y denote the set of attractive transit itineraries of traveler y . It may be obtained from empirical observations, or from a suitable choice set generation model (e.g. [Cats, 2011](#)). The probability of traveler y to choose a transit itinerary $i \in \mathcal{I}_y$ can be generally expressed as

$$Pr(i|y) = Y(n_y^o, n_y^d, t_y^o, \eta_y), \tag{1}$$

where $Y(n_y^o, n_y^d, t_y^o, \eta_y)$ denotes the transit itinerary choice model. We assume that some appropriate model is available, which may be a simple all-or-nothing assignment model, or a complex choice model that ensures transfers, and minimizes travel fare and travel time.

Subject to the choice of a transit itinerary i , a traveler $y \in \mathcal{Y}$ is assumed to attach a utility to each feasible travel path $h \in \mathcal{H}_i$, with \mathcal{H}_i the set of available travel paths. The path-specific utility can be expressed as

$$\bar{u}_{y,h} = u_{y,h} + \varepsilon_{y,h}, \tag{2}$$

where $u_{y,h}$ denotes the deterministic part, and $\varepsilon_{y,h} \sim EV(0, 1)$ an i.i.d. extreme value type I error. Travel utility typically comprises components of in-vehicle, walking and waiting time as well as transfer penalties, and may depend on various factors, such as personal attributes, or prevailing crowding levels. Note that correlations between path alternatives can in principle be captured by including a correction term as proposed by [Ben-Akiva and Bierlaire \(2003\)](#) or more recently by [Fosgerau et al. \(2013\)](#).

We assume that path utility is made up of link-wise additive components. The deterministic part of the utility then takes the form

$$u_{y,h} = \sum_{w \in h} \hat{u}_{y,h}^w(\mathbf{z}_{y,h}^w), \tag{3}$$

where $\hat{u}_{y,h}^w$ is the utility of traveler y associated with segment w on path h , and $\mathbf{z}_{y,h}^w$ the corresponding vector of attributes. These attributes characterize the travel path and personal preferences and may be subject to learning and adaptation, such as the expected level of on-board crowding, or subject to real-time information, such as the delay of an anticipated train run.

We assume that travelers have full information on their available choice set, and that they select alternatives that maximize their expected utility. The progress of individual travelers is modeled as a sequence of consecutive travel decisions in combination with a learning process, which we specify in the form of an exponential filter ([Hickman and Bernstein, 1997](#); [Cantarella and Cascetta, 1995](#)). Details may be found in [Appendix B](#).

3.3.1. Context and assumptions

We assume that an exogenous, case study-specific transit itinerary choice model of the form of Eq. (1) is available. Railway operators typically provide recommendations of transit itineraries, and often the corresponding choice fractions are empirically known. The specific model we will use in the case study is described by [Banninga et al. \(2016, in Dutch\)](#).

The utility of a travel segment (see Eq. (3)) is assumed to depend on the time spent on each activity. Let \mathcal{B}_w denote the set of activities available on segment w , including e.g. walking, waiting, or ascending a stairway. Taking into account spatiotemporal changes in level-of-service, the utility of traveler y on segment w is specified as

$$\hat{u}_{y,h}^w = \beta_0^{\text{ivt}} \sum_{b \in \mathcal{B}_w} \int_{t_{y,w}^{\text{in}}}^{t_{y,w}^{\text{out}}} \Theta_{y,w}^b(\hat{z}_w(t)) 1_y^b(t) dt. \tag{4}$$

In Eq. (4), $\hat{z}_w(t)$ generically denotes the level-of-service of segment w at time t , and $1_y^b(t)$ is an indicator equaling one if traveler y pursues activity $b \in \mathcal{B}$ at time t . The variable $\Theta_{y,w}^b$ represents a time multiplier that depends on the level-of-service and the travel activity, and β_0^{ivt} is a constant associated with typical in-vehicle conditions. Note that, since the number of transfers is determined by the transit itinerary, a transfer penalty would not influence the travel path choice, and is thus omitted.

In the context of Eqs. (2) and (4), the constant β_0^{ivt} describes the residual deviance, i.e., the amount of unexplained variation. Its value depends on the specification of the time multiplier, on the targeted user group, and is country-specific. To obtain an approximate estimate, a preliminary stated-preference survey among Dutch commuters has been carried out, yielding a value of $\beta_0^{\text{ivt}} = -6.35 \cdot 10^{-3} \text{ s}^{-1}$. The details of this survey are beyond the scope of the present work, and not relevant for the illustration of the proposed modeling framework.

For train cars, the level-of-service is captured by the load factor, i.e., the ratio between car occupancy and seat capacity. Among others, [Wardman and Whelan \(2011\)](#) have studied the in-vehicle time perceived by seated and standing passengers as a function of the load factor. Their specification is adopted in this work and shown in [Fig. 3a](#).

For walking facilities, the level-of-service is captured by the pedestrian density ([Fruin, 1971](#)). At free-flow conditions, waiting and

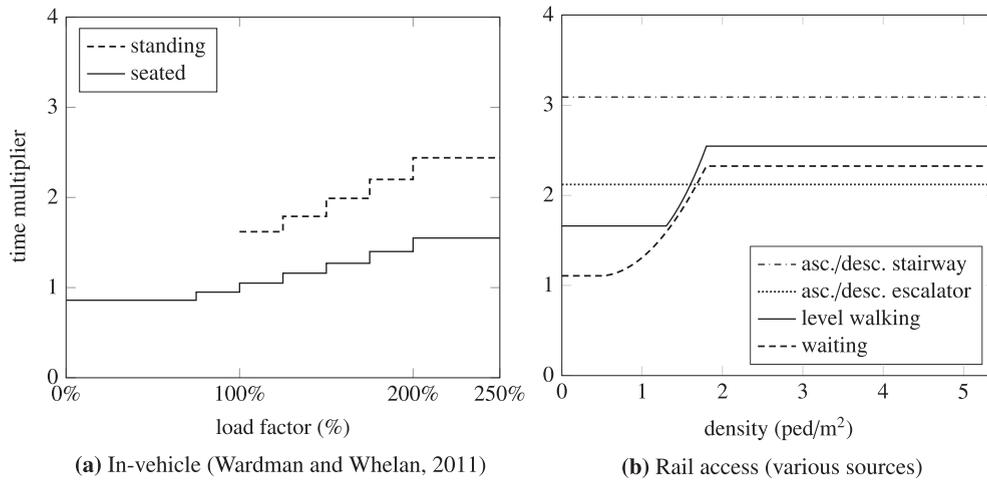


Fig. 3. Example specification of time multipliers as obtained from the literature (Wardman, 2004; Wardman and Whelan, 2011; Douglas and Karpouzis, 2005; Daamen et al., 2005).

walking are valued at $\Theta = 1.66$ and 1.11 units of in-vehicle time, respectively (Wardman, 2004, Table 6, inter-urban rail passengers). Walking is thus given a premium of approximately 50% compared to waiting, attributed to the physical effort (see also Douglas and Karpouzis, 2005). With increasing density, the time multiplier Θ increases. Fig. 3b shows the proposed relationship, which is adopted from Douglas and Karpouzis (2005). For stairways and escalators, no density-dependent specification is available, and hence a constant valuation is assumed. Specifically, a time multiplier of 2.12 and 3.09 is assumed,¹ not distinguishing between the upward and downward usage directions (Daamen et al., 2005).

3.4. Traveler movements

Traveler movements are considered to be the result of a stochastic service system formed by the transit network and its users. This stochastic system governs the times at which a traveler y enters and leaves a segment w , denoted by $t_{y,w}^{in}$ and $t_{y,w}^{out}$, respectively.

The service times depend on the state of the corresponding facility element, denoted by the vector \mathbf{x} . For a pedestrian area a , the state vector groups the number of travelers m_ℓ on each associated link $\ell \in \mathcal{L}_a$, $\mathbf{x}_a = \{m_\ell | \ell \in \mathcal{L}_a\}$. For a train car c , the state $\mathbf{x}_c = \{m_c^{stand}, m_c^{seated}\}$ comprises the number of seated and standing passengers denoted by m_c^{stand} and m_c^{seated} , respectively.

In pedestrian areas, the service time of a pedestrian link is equal to the time required to traverse it. Specifically, for a traveler y on link $\ell \in \mathcal{L}_a$, the traversal time is given by

$$\Delta t_{y,\ell}^{walk} = \lambda_\ell / v_{y,\ell}, \tag{5}$$

where $v_{y,\ell}$ denotes the corresponding walking speed. Walking speed is influenced by various factors such as age, gender, the timetable and prevailing network conditions. It may be expressed as a state-dependent random variable that is governed by a potentially direction-dependent or stochastic fundamental diagram (Nikolić et al., 2016; Hänseler et al., 2017a)

$$v_{y,\ell} = FD_{y,\ell}(\mathbf{x}_a, \eta_y), \tag{6}$$

where $FD_{y,\ell}$ denotes the density-speed relationship associated with traveler y on link ℓ . Typically, under free-flow conditions, walking speeds vary substantially across pedestrians. As more pedestrians occupy space, frictions among pedestrians increase, reducing velocity and its variability (Tregenza, 1976). Various specifications of Eq. (6) are available in the literature, also for inclined areas such as stairs or escalators (Weidmann, 1992).

Upon being served by a facility element, travelers move to a next facility element provided that space is available, and that the resulting flows at interfaces such as doors do not exceed capacity. If capacity limits are reached, travelers are temporarily refrained from proceeding. They experience waiting time while continuing to consume space in their current facility element. A similar consideration holds for the assignment of seats to passengers, where standing incurs lower travel utility. Details may be found in Appendix C.

3.4.1. Specification

Walking speed is specified as a composite model consisting of an individual component and a mean density-speed relationship. The individual component $\Psi_y \in \mathbb{R}^+$ takes into account heterogeneity in walking speed across travelers. It is modeled as a random

¹ Daamen et al. (2005) estimate time spent on escalators and stairways as $1.28 \times$ and $1.86 \times$ as costly as level walking, hence $1.28 * 1.66 = 2.12$ and $1.86 * 1.66 = 3.09$.

Table 1

Parameters for density-speed relationship and flow capacity in various walking facilities as reported by Weidmann (1992, W), Cheung and Lam (1998, CL), Daamen (2004, D) and Bodendorf et al. (2014, B).

	v^{free} [m/s]	γ [1/m ²]	k^{jam} [1/m ²]	ϕ^{crit} [1/ms]
level walkway	1.41 ^D	1.913 ^W	5.4 ^W	1.29
stairway (↑/↓)	0.610 ^W /0.694 ^W	3.722/3.802 ^W	5.4 ^W	0.85/0.98
escalator (↑/↓)	0.893 ^{CL} /1.052 ^{CL}	15.30/5.965 ^{CL}	5.4 ^W	1.25 ^B

Speeds shown for stairways and escalators pertain to their horizontal component. For escalators, the speeds apply to walking pedestrians; for standing escalator users, the speed is given by the mechanical speed of the escalator, amounting typically to 0.5 m/s (Bodendorf et al., 2014). As in Singh et al. (2016), it is assumed that 30% randomly drawn travelers walk on escalators, while the remaining 70% stand. (For interpretation of the critical flow ϕ^{crit} , see Appendix C. Values provided for level walkways and stairways follow from the corresponding parametrization of the fundamental diagram, whereas the value indicated for escalators represents an empirical estimate from Bodendorf et al. (2014).)

variable with Gaussian distribution $\Psi \sim \mathcal{N}(1, 0.215)$, which is appropriate in absence of socioeconomic information (Daamen, 2004). A limitation of this approach is that it does not consider longitudinal differences in free-flow speed. Importantly, if a departure is imminent, prospective train passengers would likely accelerate their pace to reach a connection in time. To account for this, free-flow speeds are resampled for travelers that have missed a train in a previous iteration and whose assigned speed is below average.

To describe the density-dependence of mean walking speed, the fundamental diagram by Weidmann (1992) is used. Unlike most other specifications, the Weidmann-relationship has been calibrated for various facilities, including stairways, ramps and horizontal walkways. It specifies the walking speed of traveler y on pedestrian link ℓ as

$$FD_{y,\ell} = \Psi_y v_\ell^{\text{free}} \left(1 - e^{-\gamma_\ell \left(\frac{1}{k_a} - \frac{1}{k^{\text{jam}}} \right)} \right), \tag{7}$$

where the prevailing pedestrian density in area a , $\ell \in \mathcal{L}_a$, is given by

$$k_a = \frac{1}{\alpha_a} \sum_{\ell \in \mathcal{L}_a} m_\ell. \tag{8}$$

The parameter k^{jam} denotes jam density, v_ℓ^{free} the free-flow speed, and γ_ℓ is a shape parameter. Table 1 provides facility-specific values.

3.5. Model output

As with any simulation-based model, there is flexibility in defining the desired model outputs. In the context of model validation, two groups of data sources may be distinguished:

Pedestrian traffic data, describing passenger movements at train stations. Such information is useful to quantify the safety and level-of-service of rail access facilities, for instance in terms of flows or densities. Various measurement technologies exist to observe such data in practice (Bauer et al., 2009; U.S. Department of Transportation, 2013).

Passenger ridership data, describing the usage of trains. Such data is useful to describe on-board crowding, and to estimate the resulting service level. Automated measurements of ridership may be based on in-vehicle Wi-Fi sensors, door count systems, or axle load sensors (Kim et al., 2014; Nielsen et al., 2014).

4. Case study

We carry out a case study of the Utrecht-Schiphol rail corridor in the Netherlands (see Fig. 4). This case study is of interest for two reasons:

1. The corridor is amongst the busiest in Europe, with a continuously high need for maintenance (Nederlandse Spoorwegen, 2016). We discuss how the framework can provide clues for investment appraisal by comparing travel cost associated with in-vehicle and at-station time.
2. The corresponding platform in Utrecht has been declared ‘overloaded in the near future’ by the responsible government agency (Kleinhout, 2016). There is an elevated risk of passengers falling on tracks, and generally a high level of congestion on platform exit ways. To monitor the situation, the central part of the platform has been equipped with a pedestrian tracking system at a cost of 200,000 € (Vaatstra, 2017). We investigate to what extent our approach can describe platform usage by comparing model prediction and sensor observation.

4.1. Case study description

The Utrecht–Schiphol corridor connects the largest train station in the Netherlands with Amsterdam's airport, a major transportation hub in Europe. In the train timetable of 2017, direct train service is available four times per hour with a capacity of up to



Fig. 4. Railway network of the Netherlands, with the Utrecht-Schiphol train corridor highlighted.

1128 seats per train. The total journey amounts to between 30 and 33 min, depending on the train run.

The direct connection serves four train stations, namely Utrecht Centraal (henceforth referred to as Utrecht), Amsterdam Bijlmer Arena (Bijlmer), Amsterdam Zuid (Amsterdam Zuid), and Schiphol Airport (Schiphol). These train services originate upstream of Utrecht and terminate in Schiphol. Such ‘corridor trains’ are served by platform 5/7 in Utrecht, platform 2/3 in Bijlmer, and platforms 3/4 both in Amsterdam Zuid and Schiphol. ‘Corridor passengers’ start and end their train journey at one of the aforementioned stations. In contrast, ‘auxiliary passengers’ board upstream of Utrecht, or use trains that do not serve the Utrecht-Schiphol route. They are taken into account to reproduce the prevailing crowding conditions, and amount to less than 20% of total passengers on corridor trains.

Fig. 5 shows the configuration of the platforms used by corridor passengers, as well as the train stopping positions for various train lengths. Platforms are longitudinally partitioned into areas of the same length as train cars, i.e., into approximately 30 m long segments. Laterally, they are subdivided when necessary to account for spatial heterogeneity, in particular if built facilities are present along the centerline of the platform. Escalators and their service direction are indicated by arrows.

Travelers check in and out at ticket validation terminals that are placed above or below the platform level. These terminals are either turnstiles as in the case of Utrecht, Bijlmer and Amsterdam Zuid, or free-standing validators as in Schiphol. The walking space of the entire area restricted by ticket validators is modeled. In Bijlmer and Amsterdam Zuid the corresponding areas outside platforms are relatively small. In contrast, in Utrecht and Schiphol, they are vast, and free-flow conditions are assumed. This is a simplifying assumption, motivated by our focus on platforms and the typically dominant crowding levels on platform access ways.

On working days, demand reaches a maximum during the morning peak period between 7 and 9 a.m., which is studied in the following analysis (van den Heuvel et al., 2016). To ensure that the crowding experienced by travelers is as realistic as possible, an additional 30 min are considered before and after. During the resulting three-hour period, on average between 35,000 and 40,000 passengers use approximately 140 trains serving any of the aforementioned platforms. Rolling stock consists primarily of three train types, namely suburban trains comprising four to eight cars, and interregional single- and bi-level trains with up to 12 cars. They differ in terms of passenger and door capacity. Passenger capacity amounts to 45–55 seats per car for suburban trains, 63 for interregional, and 84–94 for bilevel interregional train cars (NS Reizigers, 2004). Crush capacities are approximately twice as high, but in practice only reached during extreme events which are not of interest in this case study. Each train car is equipped with two doors, associated with flow capacities of 1.02, 0.88 and 1.12 ped/s per door for sub-urban, interregional and bilevel interregional trains, respectively (Wiggenraad, 2001).

To describe transit itinerary choice, a proprietary model developed specifically for the Dutch railway network is used (Hoogenraad et al., 2013; van den Heuvel and Hoogenraad, 2014). For every origin-destination pair and departure time, it generates a set of relevant train itineraries, associates a generalized travel time with each alternative, and estimates choice probabilities based on empirically calibrated distribution functions. These choice probabilities implicitly take into account the chance of catching a train at the origin station. In case of multi-leg journeys, the probability of reaching a transfer connection is considered, and for larger

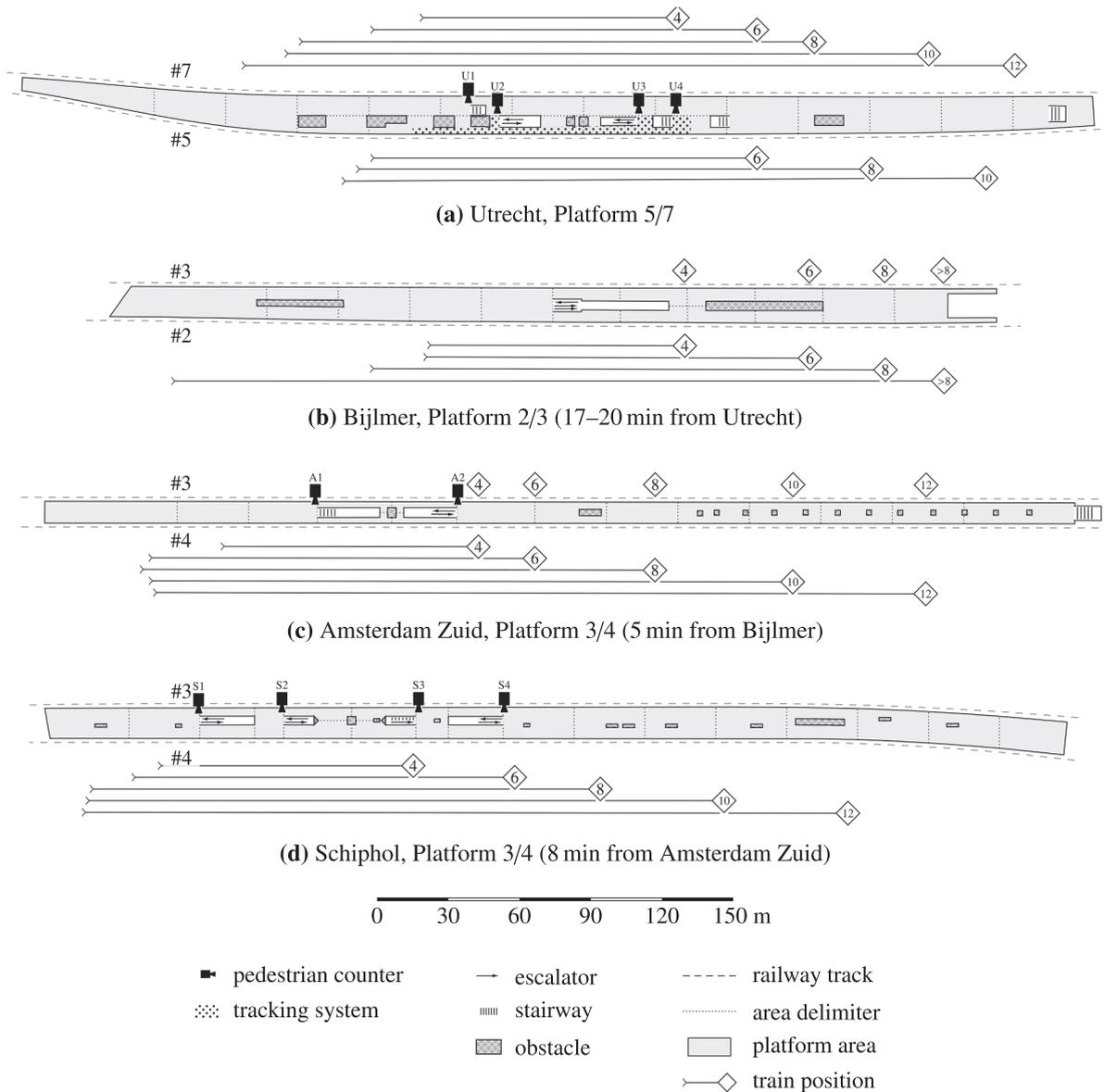


Fig. 5. Platforms of the Utrecht-Schiphol corridor. Train movement is from left to right, top to bottom.

stations also potential time spent on subsidiary travel activities such as shopping. Train itineraries with multiple legs include in particular Utrecht and Schiphol, with a high share of transfers to and from corridor trains. Transfer passengers using any of the four corridor stations are thus explicitly taken into account.

4.2. Available data

A rich set of data sources is available, describing travel demand, transit operations, and pedestrian traffic at stations:

Automatic fare collection data is available for passengers carrying a smart card, comprising check-in and check-out stamps at their station of origin and destination. The average penetration rate of smart cards for both check-ins and check-outs amounts to 96% for Utrecht and Amsterdam Zuid, 91% for Bijlmer, and 67% for Schiphol. The reduced values in Bijlmer and Schiphol are due to print-at-home tickets provided for nearby events and single-use tickets used by international travelers, respectively. Penetration rates specifically for the morning peak period are not known, but assumed to be higher due to a large share of smart card holders among commuters. Specifically for this study, to fulfill prevailing privacy requirements, the fare collection data needs pre-processing. Details are found in [Appendix D](#).

Train tracking data is available, specifying for each train the realized arrival and departure times, assigned platforms, as well as the rolling stock type. The stopping position of trains is dependent on the number of cars, and can be inferred (see [Fig. 5](#)).

Pedestrian flow counts are available on selected platform access/exit routes in Utrecht, Amsterdam Zuid and Schiphol, as indicated by camera symbols in Fig. 5. Flow sensors are bi-directional, providing minute-by-minute counts for access and exit flows. In Utrecht, access ways connecting platform 5/7 to the station hall are surveilled, whereas access ways to two underground tunnels are not monitored. In Amsterdam Zuid, the two central access ways are surveilled, while a third, more lateral route is not equipped. For platform 3/4 in Schiphol, all access routes are equipped.

Pedestrian density measurements are available on platform 5/7 in Utrecht for a 100 m-stretch on the narrow side of the platform, as mentioned at the beginning of the section. The covered part is indicated by the dotted area in Fig. 5. The density measurements are obtained by a network of 25 stereo cameras, and are available on a minute-by-minute basis at a spatial granularity of approximately 10 m².

The above data is available for the time period between March 8 and March 31, 2017. Demand and transit operation data are used for model estimation, and pedestrian traffic data is used for validation. The analysis focuses on working days, and specifically on Tuesdays, Wednesdays and Thursdays, leaving a set of eleven days for detailed investigation.

4.3. Results

Up to 24 independent simulation runs are carried out per day, with 250 learning iterations each. While not optimal in terms of computational cost, this configuration is sufficient to reliably estimate all performance indicators of interest, which is our main goal. The computation time of one iteration on a single core of a 2.3 GHz-CPU amounts to approximately one minute, and multiple runs can be efficiently parallelized (see Hänseler, 2017, for a freely available implementation). While all available days are considered to study day-to-day fluctuations, special focus is given to the analysis of the morning peak period of March 22. On this day, all sensors have been functioning, and the train traffic was not disrupted. The consideration of a single day allows to study fast dynamics, such as platform exit flows after a train arrival, which otherwise get averaged out by deviations in train arrival times across days.

4.3.1. Consistency

To investigate the consistency of smart card data and the transit itinerary choice model, we first compare cumulative flows on monitored platform access routes. Fig. 6 shows for each station the total flows on escalators and stairways that are equipped with a count sensor. Individual curves represent observations and estimates of March 22, and the gray band the observation range associated with the entire set of days, i.e., the band spanned by the smallest and largest observed cumulative flow curve. Fig. 6a shows that flows on the monitored platform access routes in Utrecht are overestimated by 11.2%. This may be due to errors in the transit itinerary and travel path choice facilitated by a large number of transfer passengers, or sensor saturation leading to reduced flow observations. In contrast, platform exit flows in Schiphol (Fig. 6c) are underestimated by -19.2%, which is expected due to the lower penetration rate of smart cards in that station. In Amsterdam Zuid (Fig. 6b), the inflow is over- and the outflow underestimated, which is consistent with the overestimation in Utrecht and the underestimation in Schiphol. Absolute errors are generally smaller due to less freedom in local pedestrian route choice, and a lower share of transfers.

Overall, observation and estimation agree up to a degree that is comparable to the width of the fluctuation band, implying that the available smart card and pedestrian count data are consistent. Demand micro-peaks, discernible by the step-wise character of inflow curves, corroborate the consistency of alighting and walking dynamics and observed timetable realizations.

4.3.2. Validation

As mentioned, the platform serving corridor trains in Utrecht, and in particular the narrow areas adjacent to track 5, have been declared prone to overloading. Likewise, the two central access ways in Amsterdam Zuid represent a well-known capacity bottleneck. The performance in these two areas is studied in the following to assess the validity of the model.

Fig. 7a and b show exit flows on the central stairway and escalator in Amsterdam Zuid. In each panel, the observed and estimated

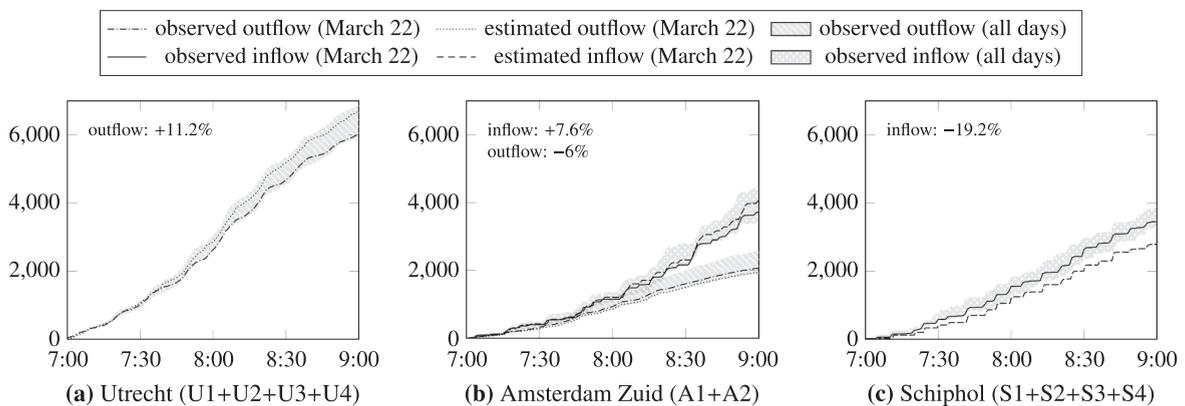


Fig. 6. Platform access and egress flows on monitored routes as observed and estimated for March 22, 2017 (individual curves), with observation range representing weekdays between March 8 and March 30 (gray band). In brackets, sensor IDs are shown (see Fig. 5 for their location).

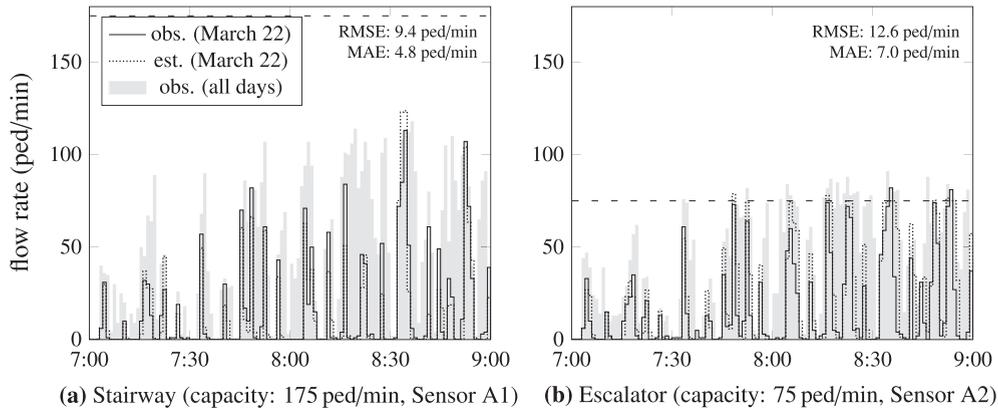


Fig. 7. Exit flow from platform 3/4, Amsterdam Zuid, along the two central exit routes (see Fig. 5 for position of sensors). Observations and model estimates are shown with a granularity of 60 s. Dashed lines represent capacities (obtained from Table 1).

flow rates on March 22 are shown as curves, while the observation range of the 11-day dataset is shown as a shaded area. An aggregation period of 60 s is used, imposed by the measurement data. Root-mean-square (RMSE) and mean absolute errors (MAE) are provided to quantify the agreement between model estimates and observations.

The observed stairway flow, shown in Fig. 7a, does not reach capacity during the entire observation period. In contrast, the escalator flow is saturated on multiple occasions, as depicted in Fig. 7b. The location of flow peaks is relatively well reproduced by the model, with minor differences in their magnitude. One reason for the latter is rooted in simplifications in the modeling process. For instance, it is likely that a person approaching a congested exit way anticipates the discomfort and decelerates in advance, thereby smoothing the observed flow. Such behavior is not considered by the model. At a coarser time aggregation, such as at the widely used five-minute level (Ross, 2000), differences between model prediction and observation decrease rapidly, although while averaging out flow peaks.

Fig. 8 shows pedestrian densities on the equipped platform area in Utrecht. Prior to train arrivals, prospective passengers start accumulating on platforms, yielding a gradual increase in density. Upon arrival, incoming passengers alight, leading to a short peak, followed by a rapid decrease in pedestrian density. Maximum observed densities are slightly above 1 ped/m². On other days, density values exceeding 1.2 ped/m² are observed (see shaded area), although their occurrence is rare and can typically be explained by specific circumstances such as train cancellations.

A longitudinal comparison of the density profiles along the equipped 100 m long area (Fig. 8a–c) shows only small differences, which however is put in perspective by the total platform length of 430 m. Both RMSE and MAE indicate good agreement between estimates and observations, in particular if the substantial day-to-day fluctuations are considered (gray band). Certain differences in peak densities are perceivable, which may be due to the presence of first and second-class cars with differing loads. Also, it can be seen that the model predicts a faster accumulation of passengers prior to train departures. This deviation is likely due to subsidiary travel activities, such as getting a coffee before departure, which are not modeled.

If the entire platform length is considered, longitudinal differences in density become apparent in the model estimate. Fig. 9 shows the evolution of density for the five-minute interval on March 22 for which the maximum density has been measured. During the considered time span, two trains of ten cars arrive on both sides of the platform. At 8:04:47, train #3522 arrives on platform 7 (wider half of platform), with an estimated 567 alighting and 683 boarding passengers. Despite the large passenger volume,

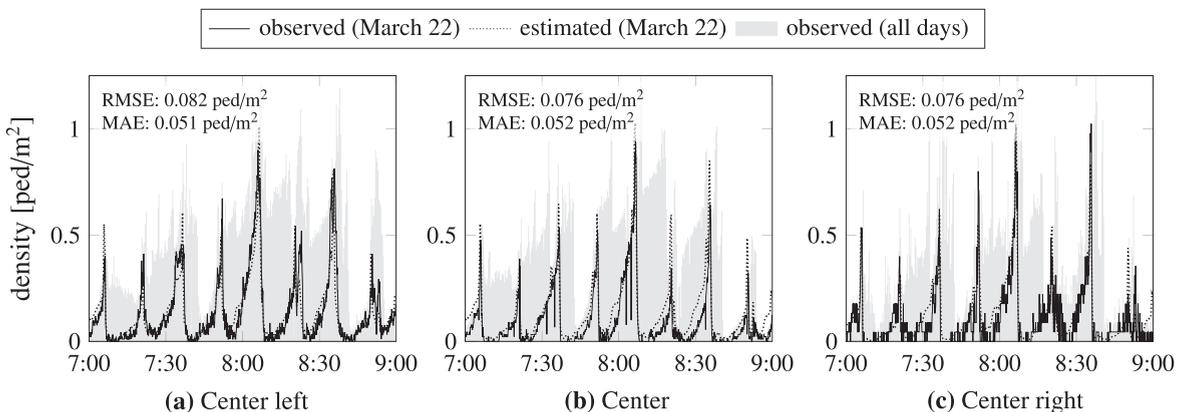


Fig. 8. Density evolution on monitored area on platform 5/7, Utrecht.

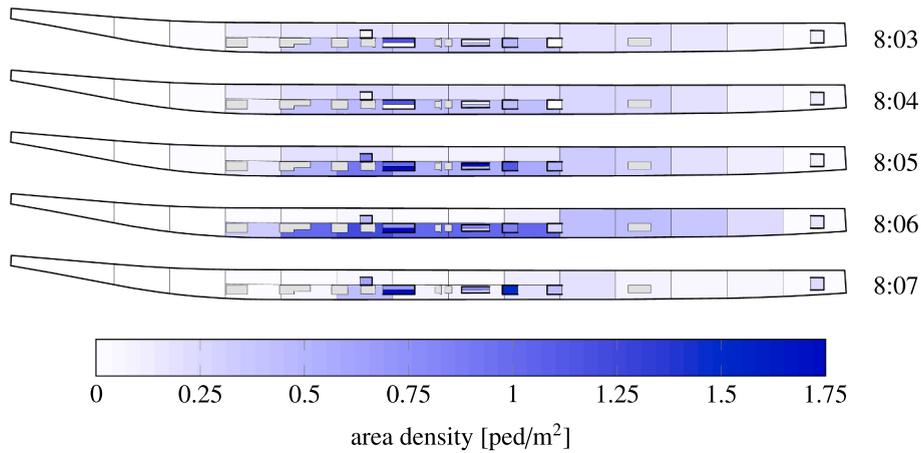


Fig. 9. Estimated density distribution during 5 min-peak period on March 22, 2017, on platform 5/7, Utrecht.

pedestrian densities remain well under 1 ped/m², with the exception of exit ways. At 8:06:11, train #3020 arrives on platform 5 (narrow side), with 691 alighting and 688 boarding passengers. Due to the narrow platform area and a concentration of passengers around the central platform, densities reach significantly higher values. These estimates, both regarding the lateral and longitudinal evolution of density, are in agreement with qualitative observations during peak periods.

4.3.3. Travel utility and train load distributions

Beyond the prediction of pedestrian traffic indicators, the model can be used to investigate the perceived travel utility, or passenger load distributions on-board trains. Fig. 10 shows the estimated travel cost grouped by origin-destination pairs, assuming a value of time of 9.62 €/h for Dutch commuters (Pel et al., 2014). The average of all available weekdays is shown, and error bars denote the corresponding standard deviation. It is apparent that, especially for short trips, the contribution of walking and waiting is significant, and may exceed in-vehicle cost. This is in line with related studies which also consider the travel cost contribution of walking and waiting, although typically without taking the impact of station crowding on walking speeds and comfort into account (e.g. Cats et al., 2016). For transit systems operating close to capacity, consideration of such dependencies is important for an accurate appraisal of infrastructural or operational measures, such as the evaluation of rolling stock investments against investments in rail access facilities.

Fig. 11 shows the estimated passenger distribution of train #3522 when arriving in Bijlmer, Amsterdam Zuid and Schiphol. Train #3522 has one of the highest riderships, particularly between Utrecht and Bijlmer, and is normally operated with ten cars. The estimated distribution pertains to the average of all considered weekdays, except for March 30 on which only eight cars have been available. Naturally, as trains move along the corridor, the percentage of auxiliary passengers decreases.

The load distributions when arriving at Bijlmer (Fig. 11a) and Amsterdam Zuid (Fig. 11b) are high and nearly uniform, with differences of less than 13% between the train car with the highest and lowest load. This is expected, as walking efforts play a smaller role compared to the influence of in-vehicle crowding when passenger loads are high. The highest load in Bijlmer is found in cars #7 and #8, offering direct access to the single platform exit route. The estimated distribution in Amsterdam Zuid is bimodal, with one peak pertaining to the central platform exit ways, and one towards the front of the train associated with the lateral exit way at the head of the platform. When reaching Schiphol, the level of in-vehicle crowding is low, implying lower in-vehicle cost, and thus a higher influence of walking cost. Differences in the load distribution are more pronounced, with the occupancy in the busiest car

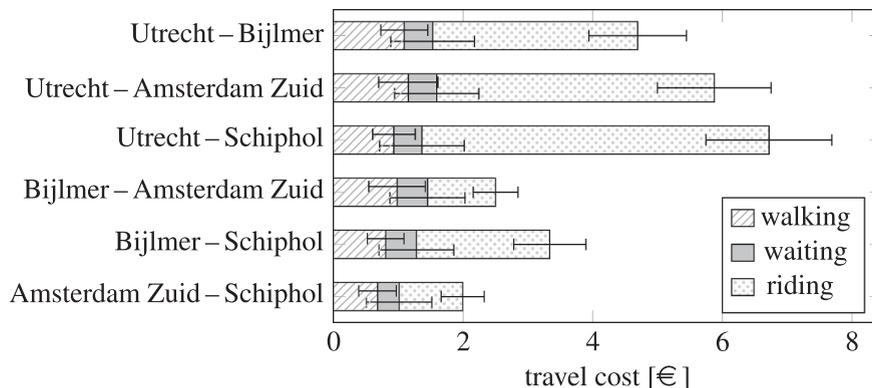


Fig. 10. Average travel cost by OD-relation and travel activity for the 11-day dataset.

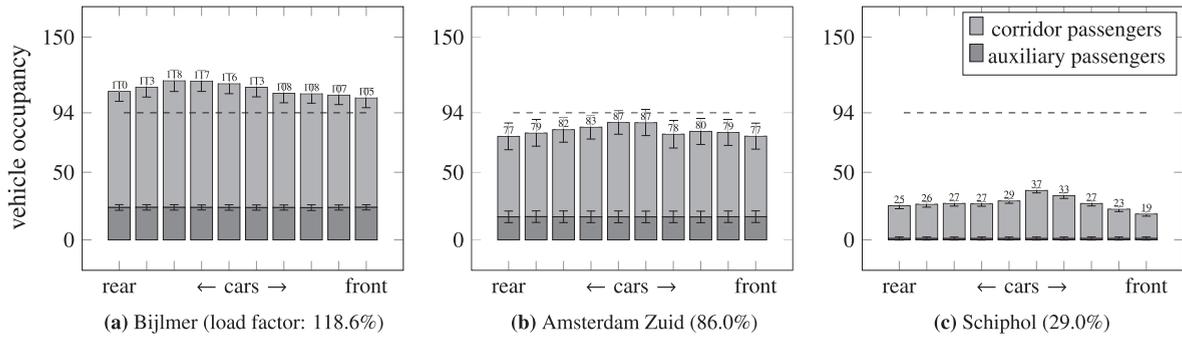


Fig. 11. Average passenger load distribution of train #3522 when reaching Bijlmer, Amsterdam Zuid and Schiphol for the 11-day dataset. The dashed line denotes the seat capacity, amounting to 94 seats per car.

being almost twice as high as in the least utilized car. Passenger load is particularly low in cars at the front of the train, which are not adjacent to any exit way.

5. Discussion

We first assess the proposed model by commenting on its strengths and limitations, and then discuss its scope of application.

5.1. Strengths and limitations

The results show that the proposed framework is capable of producing passenger flows and platform densities that are in good quantitative agreement with observations, and able to realistically describe generalized travel costs and car load distributions.

This is encouraging since the required model inputs – automated fare collection and train tracking data – are increasingly available. The proposed simulation approach allows to potentially account for measurement errors if systematic correlations between measurement errors and passenger load levels can be established. Most station models require inputs that are less readily available, such as origin-destination demand at the station level, or pedestrian flow and density counts, which in this work have been used for model validation only. Likewise, transit assignment models may require case-specific assumptions, specifying for instance walking and waiting times, or the likelihood of successful connections in busy train stations. In contrast, many of the parameters required by the proposed framework are more general, and can be specified based on the literature.

Nevertheless, the issue of optimally specifying trip assignment models is yet to be tackled. Notably, the influence of crowding on travel behavior is fundamental, and still subject to research (Pel et al., 2014). To capture such impacts, the current specification borrows a wide range of parameters from experiments performed in various geographical areas. This is not ideal, and yet yields satisfactory results. A dedicated calibration based on revealed preference data is likely to further improve the model performance, as would be a detailed testing and ‘calibration’ of the space discretization (Openshaw, 1984).

A point of criticism may pertain to the hierarchical choice of transit itinerary and travel path, i.e., the assumption that travelers decide on a (sequence of) train runs before choosing individual pedestrian links and train cars. Such a structure is appropriate for rail systems as studied in this work, and practical since it allows for a seamless integration of empirical itinerary choice models. Yet for applications with large shares of multi-leg trips or when capacity limits result in denied boardings, it is unrealistic to assume that the itinerary is determined beforehand. In such cases, a sequential choice model also for transit itineraries is more appropriate (Spiess and Florian, 1989). A framework for such a choice structure in principle exists (e.g. Cats et al., 2016), although the consistent specification of an integrated choice model requires further research.

5.2. Scope of application

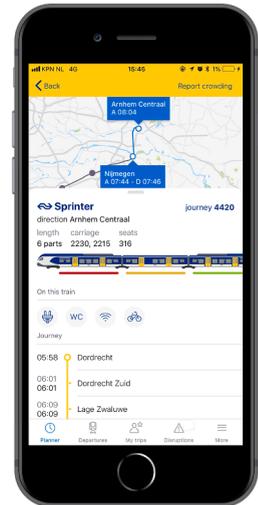
The proposed model is well-suited for problems that require consideration of passenger dynamics along the entire transit travel path. In the following, we highlight three fields of application.

Scenario evaluation and investment appraisal. Scenarios that affect both transit operations and station facilities can be evaluated. Empirical research has shown, for instance, that local adjustments in train stopping positions can significantly reduce dwell times when boarding and alighting process are time-critical (van den Heuvel, 2016). The analysis of such measures is not amenable to classical station models, as they typically assume uniform passenger distributions along a platform (Daamen, 2004; Vaatstra, 2017). Besides adjusting train stop positions, the model may be used to optimize transfer times, to identify platform sections requiring more space for pedestrian traffic, or more generally to determine performance and safety bottlenecks. Similarly, it is suitable for balancing investments between the network and the station level, as discussed previously.

Provision of crowding information. To yield a more balanced use of platforms and trains, transit operators increasingly employ crowding information systems. We illustrate this using two examples from the Netherlands. Fig. 12a shows a 180-meter long luminous band that indicates the occupancy of an arriving train. Crowding levels are directly measured by infrared counters that have



(a) 'Intelligent Platform Bar' (Station 's-Hertogenbosch, Picture: Edenspiekermann)



(b) Travel companion (NS)

Fig. 12. Two examples of information provision systems for on-board crowding. Green indicates that a train car is largely empty, orange means 'semi-crowded,' and red means 'full.' (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to be installed at each train door, which in practice is prohibitively expensive. Fig. 12b shows a smartphone application providing crowding information based on historical observations. While more viable from an economical view point, the usefulness of such crowding estimates is limited by the lack of real-time information. Our model, in contrast, can provide vehicle-specific crowding estimates based on readily available AFC data in real-time.

Disruption management. In conjunction with a rail operations and demand estimation model, the proposed framework can be used as a building block for a disruption management system. It can readily estimate passenger loads in train cars, pedestrian densities on platforms, and flows on access ways. Such information is useful to develop and evaluate mitigation strategies, both for off- and on-line usage. For instance, given a delayed train, one may assess holding strategies to increase the likelihood of successful connections. Moreover, one may change the stopping position to alleviate platform crowding. To evaluate such measures, the model can obtain performance indicators such as travel time losses or number of missed connections.

5.3. Demonstration of scenario assessment

We illustrate the model's capability to support the assessment of 'what if' scenarios using two examples involving Utrecht Centraal Station:

'No escalators': We assume that both escalator pairs on platform 5/7 fail and are closed for pedestrian traffic. Such a disruption may result from a local electricity outage and following repair work.

'Lateral train stopping': We assume that all trains on platform 5, irrespective of their length, stop at the far right, i.e., at the stopping position for trains with 10 cars (see Fig. 5). This may be a planning scenario aimed at reducing pedestrian congestion along the central platform sections.

We analyze the two scenarios for the morning peak period of March 22, 2017. On this day, train #3020 arrives with a reduced number of cars during the busiest period (arrival at 08:06:10, operated with 7 cars), emphasizing the influence of the change in stopping position.

Fig. 13 shows pedestrian densities at 08:06:30 on platform 5/7 for the base case and the two scenarios. Caused by the outage of the nearby escalators, the 'no escalators' scenario shows a high density on the central stairways. The 'lateral train stopping' scenario reveals an increased use of the right half of the platform, and less elevated densities in the narrow, central platform areas. Hence, from a crowd management safety viewpoint, shifting train stopping positions may indeed be worthwhile in that specific case.

Fig. 14 shows the total number of passengers on platform 5/7 as a function of time. Six service levels are distinguished, revealing varying comfort and security (Fruin, 1971). The 'no escalators' scenario shows that, due to the restrained platform egress, passengers remain longer on platforms and experience lower service quality, including unsafe conditions at times. Between 8:00 and 8:30, total egress times increase by several minutes. Disruption management may take this into account, for instance by holding connecting trains. In contrast, lateral train stopping positions hardly affect passenger accumulation and egress times, with a slight increase in the level-of-service.

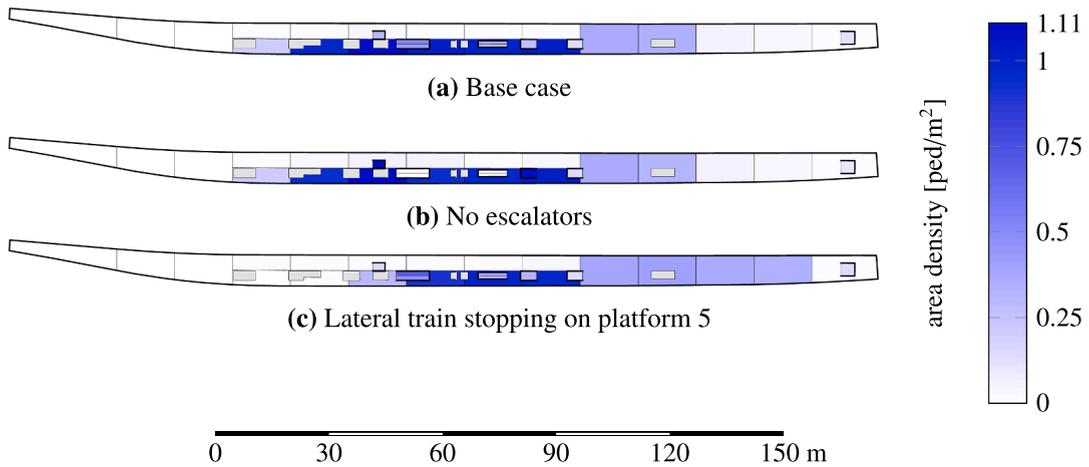


Fig. 13. Estimated density distribution at 8:06:30 on March 21, 2017, on platform 5/7, Utrecht, for the base case and two scenarios.

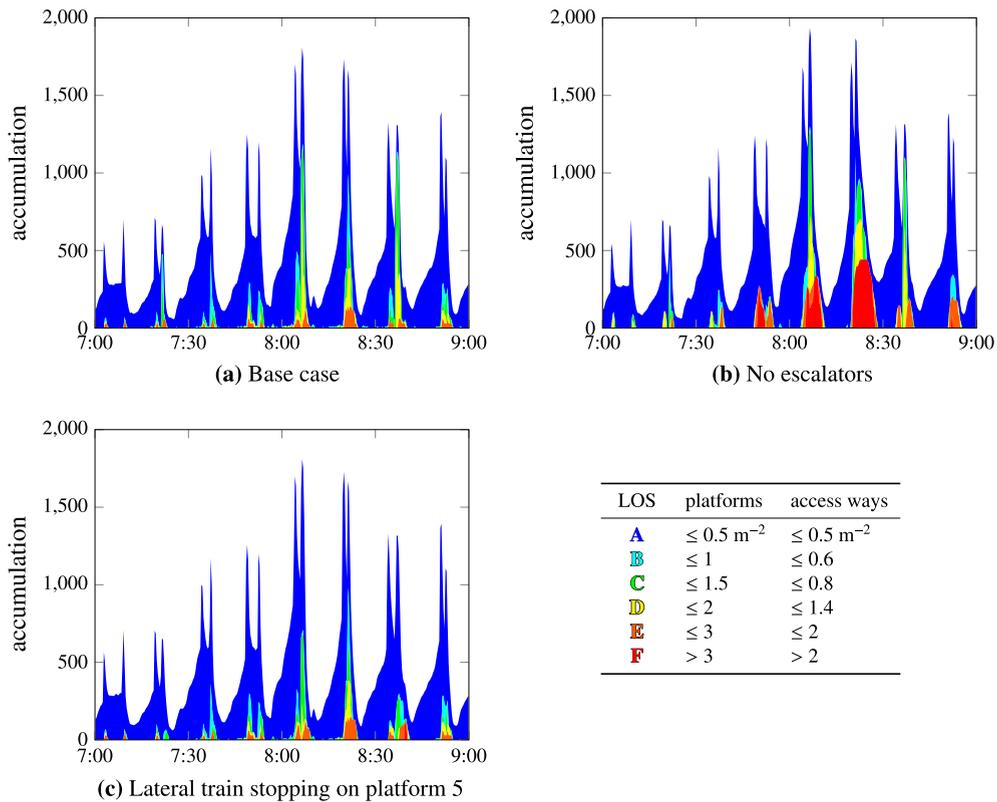


Fig. 14. Level-of-service distribution on platform 5/7 of Utrecht Centraal for the morning peak period of March 22, 2017.

6. Conclusions

A passenger-pedestrian model has been presented that, based on automated fare collection and train tracking data, simultaneously describes traveler behavior at stations and on-board. Travelers are assumed to choose pedestrian paths, platform waiting positions and individual train cars in a way that maximizes their utility. They explicitly consider the downstream consequences of choices, such as the anticipated variation of crowding in a train car, or the walking distance at their destination. The necessary knowledge is obtained within a day-to-day learning framework.

The interaction between travelers, rail access facilities and rolling stock is represented at the aggregate level. Passenger crowding, both in-vehicle and at stations, is endogenous, and its impact on individual travelers is explicitly modeled. The developed pedestrian model is applicable to multi-directional flows, as they may emerge on a walkway or a stairway, and considers static and dynamic

facility capacities, such as the maximum accumulation on a platform area, or the capacity of an escalator.

The model is applied to a case study of a major rail corridor in the Netherlands. The studied two-hour peak period involves approximately 35,000 passengers and 140 trains, depending on the day. Besides fare collection and train tracking data, a rich data set consisting of pedestrian counts and density observations on platforms is collected. In conjunction with an empirically calibrated transit itinerary choice model, the proposed framework is able to realistically reproduce level-of-service indicators such as flows and densities at the station level, as well as to plausibly predict further performance indicators such as ridership distributions. Such results are difficult to achieve with 'classical' transit assignment or station models, or they require additional assumptions, e.g. regarding platform waiting distributions. The computational speed is about two orders of magnitude faster than real-time, which makes the model applicable to extensive networks, or for time-critical applications.

In the long term, we see considerable potential in the application of the model for integrated investment appraisal, real-time estimation of crowding conditions, or for use in disruption management.

CRedit authorship contribution statement

Flurin S. Hänseler: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Jeroen P.A. van den Heuvel:** Conceptualization, Data curation, Formal analysis, Writing - review & editing. **Oded Cats:** Conceptualization, Methodology, Writing - review & editing, Project administration, Funding acquisition. **Winnie Daamen:** Methodology, Writing - review & editing. **Serge P. Hoogendoorn:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition.

Acknowledgement

This work was performed as part of the TRANS-FORM project funded by NWO grant agreement 438.15.404/298 (JPI Urban Europe ERA-NET COFUND Smart Cities and Communities initiative), and as part of the ALLEGRO project funded by European Research Council grant agreement 669792 (European Union Horizon 2020 Framework for Research and Innovation). Part of this work was carried out on the Dutch national e-infrastructure with the support of SURF Cooperative.

Appendix A. Notation

$1_{(\cdot)}$	indicator function
$a \in \mathcal{A}$	area
$b \in \mathcal{B}$	travel activity
$c \in \mathcal{C}$	train car
d	destination
$FD_{y,\ell}(\mathbf{x}_a)$	fundamental diagram
$\mathcal{G} = (\mathcal{N}, \mathcal{L})$	pedestrian network
$h \in \mathcal{H}$	travel path
$i \in \mathcal{I}$	transit itinerary
k_a	density
$\ell \in \mathcal{L}$	link
m_ℓ, m_c	traveler accumulation
$n \in \mathcal{N}$	node
o	origin
$r \in \mathcal{R}$	train run
$s \in \mathcal{S}$	train station
$t, \Delta t$	time, duration
$\tilde{u}_{y,h}$	path utility
$u_{y,h}$	deterministic part of utility
$v_{y,\ell}$	walking speed
$w \in \mathcal{W}$	path segment
$\mathbf{x}_a, \mathbf{x}_c$	state vector
$y \in \mathcal{Y}$	traveler
$z_{y,h}^w$	path segment attribute
α_a	area surface size
β	utility coefficient
γ	shape parameter
$\varepsilon_{y,h}$	random part of utility
$\boldsymbol{\eta}_y$	vector of traveler attributes
λ	length
v^{free}	free-flow speed
$Y(m_y^o, m_y^d, t_y^o, \boldsymbol{\eta}_y)$	itinerary choice model
$\Theta_{y,w}$	time multiplier
Ψ_y	speed multiplier

Appendix B. Recursive travel path choice

When reaching the end of a path segment w , a traveler is assumed to choose a next path segment $w' \in \mathcal{W}_w$ that maximizes the expected remaining travel utility. The set \mathcal{W}_w comprises all available path segments in the particular decision context, i.e., either walking links or train cars for a prospective train trip. The decision process is Markovian in that the choice of the next segment depends only on the current one, given an a priori chosen transit itinerary and individual traveler attributes (Bertsekas, 2017).

The specification of the choice set \mathcal{W}_ℓ is straightforward for walkway choices, or platform waiting section choices for prospective train passengers. Let $\mathcal{N}_{\text{transit}} \subset \mathcal{N}_s$ denote the set of transit nodes, i.e., nodes through which transit services are accessible. For a traveler on any interior walking link, $\ell \in \mathcal{L}$ with $n_\ell^+ \notin \mathcal{N}_{\text{transit}}$, the choice set is equivalent to the set of downstream adjacent links, $\mathcal{W}_\ell = \{\ell' \mid \ell' \in \mathcal{L}, n_{\ell'} \equiv n_\ell^+\}$.

In case the current path segment leads to a transit node, a train car choice arises. Let \mathcal{P}_s denote the set of platforms of station s , and $\mathcal{N}_{\text{transit}}^p \subset \mathcal{N}_{\text{transit}}$ the set of transit nodes associated with platform p . The choice set \mathcal{W}_ℓ of a traveler y on link ℓ with $n_\ell^+ \in \mathcal{N}_{\text{transit}}^p$ who wants to board run r on platform $p \in \mathcal{P}$ is given by

$$\mathcal{W}_\ell = \{c \mid c \in \mathcal{C}_r, 1_{a,c}^{\text{adj,board}} = 1\}, \tag{B.1}$$

where the indicator $1_{a,c}^{\text{adj,board}}$ equals one if area a and train car c are adjacent, and zero otherwise. The resulting choice set is contingent on the definition of adjacency. We present a definition of adjacency based on geometrical considerations below.

The inverse problem pertains to the platform section choice of alighting passengers. The choice set for a traveler riding on-board train car c alighting in train station s is given by

$$\mathcal{W}_c = \{\ell \mid \ell \in \mathcal{L}_a, 1_{c,a}^{\text{adj,alight}} = 1\}, \tag{B.2}$$

where $1_{c,a}^{\text{adj,alight}}$ is an indicator that equals one if train car c and pedestrian area a are adjacent, and zero otherwise.

Based on Eq. (2) and the aforementioned choice sets, the probability of traveler y on path segment w to choose segment $w' \in \mathcal{W}_w$ is given by

$$Pr(w'|w, y) = \frac{e^{g_{y,w'}}}{\sum_{w' \in \mathcal{W}_w} e^{g_{y,w'}}}, \tag{B.3}$$

where the value of path segment $w' \in \mathcal{W}_w$ expected by traveler y is

$$g_{y,w'} = \ln \sum_{f \in \mathcal{F}_{w',d}^i} e^{E(t_{y,f})}. \tag{B.4}$$

The set $\mathcal{F}_{w',d}^i$ comprises all available subpaths connecting path segment w' to the destination node n_d that are associated with transit itinerary i . The operator $E(\cdot)$ provides the expected value, for which typically downstream attributes need to be predicted. The logit-structure of the choice model, Eq. (B.3), follows from the assumed distribution of the random utility component. The log sum in Eq. (B.3) is a direct consequence of a travel segment generally being associated with multiple travel paths.

Eq. (B4) implicitly requires estimation of downstream attributes of each choice alternative. To that end, an exponential filter is employed (Cantarella and Cascetta, 1995), which forecasts attribute z in iteration κ by convex combination of the attribute forecast in the previous iteration $\kappa - 1$, and of the value realized in that same iteration $\kappa - 1$,

$$z_{\kappa}^{\text{forecast}} = \zeta z_{\kappa-1}^{\text{experienced}} + (1 - \zeta) z_{\kappa-1}^{\text{forecast}}, \tag{B.5}$$

where $\zeta \in (0, 1)$ denotes the learning rate. Wahba and Shalaby (2005) suggest a learning rate of 0.7 for a similar application, which is adopted in this work. This specification implies that passengers consistently weight recent experiences more than older ones, stimulating exploration in path choice behavior. The choice of the learning rate mainly influences the speed of convergence, which may be further explored if computational considerations are crucial.

Specification. The adjacency indicators for pairs of train cars and pedestrian areas are defined based on their proximity, and the corresponding choice sets are constructed by simulation-based sampling.

It is assumed that the stopping position of each train in every station is deterministically known. Consider a train dwelling in a station as depicted in Fig. 15. The stopping position of the train head, and its distance from the front and rear of an adjacent area a are denoted by q_a^+ and q_a^- , respectively (see Fig. 2). The position of the front and rear of car c relative to the head of the train are denoted by q_c^+ and q_c^- , respectively.

Assuming a uniform distribution of pedestrians within an area a , the longitudinal position of a traveler relative to the tip of the train is described by $q \sim \mathcal{U}(q_a^-, q_a^+)$. If a traveler boards the closest train car, the adjacency indicator for boardings from area a to car c is given by

$$1_{a,c}^{\text{adj,board}} = \begin{cases} 1 & \text{if } q \in [q_c^-, q_c^+] \\ 0 & \text{otherwise.} \end{cases} \tag{B.6}$$

The choice set resulting from Eq. (B.1) includes for every boarding traveler one alternative, contingent on the sampled position q .

The same approach is used to determine the adjacency indicator for alighting passengers. For each traveler, a random 'position' $q \sim \mathcal{U}(q_c^-, q_c^+)$ is drawn, and the alighting indicator is specified as

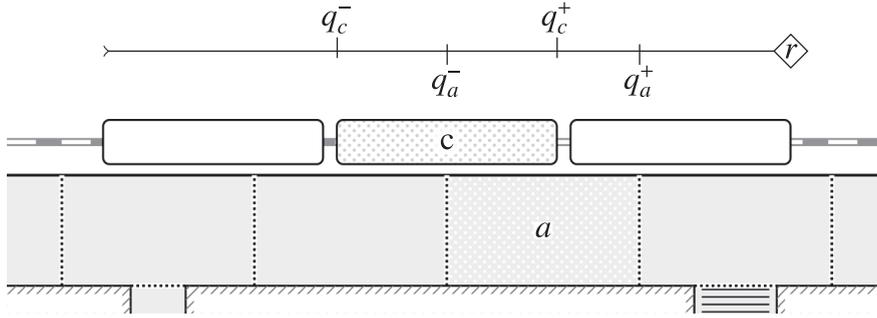


Fig. 15. Position of idling train with respect to platform, indicating the position of the train head (diamond), and its distance from the highlighted pedestrian area and train car.

$$1_{c,a}^{adj,alight} = \begin{cases} 1 & \text{if } q \in [q_a^-, q_a^+], \\ 0 & \text{otherwise.} \end{cases} \tag{B.7}$$

Appendix C. Infrastructure and rolling stock capacity

In the propagation of travelers across the network, several capacity constraints are considered: For pedestrian areas, a maximum number of pedestrians is defined, which is denoted by σ_a^{occ} and referred to as occupation capacity. Similarly, for the interface between two areas a and a' a flow capacity is considered, which is denoted by $\sigma_{a,a'}^{flow}$. For each train car, we define a seat and standing capacity, σ_c^{seat} and σ_c^{stand} with units of pedestrians, and a capacity determining maximum boarding and alighting exchange rates per car, σ_c^{ex} with units of pedestrians per time.

Upon being served by a facility element w , travelers may move to a next facility element $w' \in \mathcal{W}_w$, provided that two conditions are met. First, space must be available in the target facility element w' , which in case of a pedestrian area a at time $t \in \mathcal{T}$ implies

$$\sum_{\ell \in \mathcal{L}_a} m_\ell(t) \leq \sigma_a^{occ}. \tag{C.1}$$

For a train car c , the corresponding constraint is

$$m_c^{seat}(t) + m_c^{stand}(t) \leq \sigma_c^{seat} + \sigma_c^{stand}. \tag{C.2}$$

Second, flow capacities at area interfaces and at train doors may not be exceeded. Assuming an aggregation period τ , and denoting by $1_{y,a,a'}(t)$ an indicator that equals 1 if traveler y moves from area a to a' at time t and 0 otherwise, the capacity limit for the interface between area a and a' at time t can be expressed as

$$\frac{1}{\tau} \int_{t-\tau}^t \sum_{y \in \mathcal{Y}} 1_{y,a,a'}(t) + 1_{y,a',a}(t) dt \leq \sigma_{a,a'}^{flow}. \tag{C.3}$$

The aggregation period needs to be large enough such that $\sigma_{a,a'}^{flow} \tau \geq 1$, and small enough for the interface flow to be sufficiently restrictive in case of demand fluctuations. Ideally, the capacity thresholds $\sigma_{a,a'}^{flow}$ are chosen in conjunction with τ to obtain a meaningful specification. For capacity limits during boarding and alighting, the corresponding flow constraint for train car c is

$$\frac{1}{\tau} \int_{t-\tau}^t \sum_{y \in \mathcal{Y}} 1_{y,c}^{board}(t) + 1_{y,c}^{alight}(t) dt \leq \sigma_c^{ex}, \tag{C.4}$$

where $1_{y,c}^{board}(t)$ and $1_{y,c}^{alight}(t)$ denote corresponding indicators for boarding and alighting, respectively. If capacity limits are reached, travelers are temporarily refrained from proceeding.

A similar consideration holds for the assignment of seats to passengers. The number of seated passengers in train car c is restricted by its seat capacity, i.e.,

$$m_c^{seat}(t) \leq \sigma_c^{seat}. \tag{C.5}$$

Various priority schemes can be envisaged for space and seat assignment, ranging from random to first-come-first-served or more complex approaches.

Specification. From Table 1 and the functional form of the assumed density-speed relationship, occupation and flow capacities can be readily derived. The occupation capacity of area a is given by

$$\sigma_a^{occ} = \lfloor k^{jam} \alpha_a \rfloor. \tag{C.6}$$

The flow capacity associated with the interface between area a and a' is

$$\sigma_{a,a'}^{flow} = \lambda_{a,a'}^{int} \phi^{crit}, \tag{C.7}$$

where $\lambda_{a,a'}^{int}$ denotes the length of the interface between the two areas, and ϕ^{crit} the corresponding critical flow listed in Table 1. The aggregation period τ over which the flow capacity constraint is enforced, is set to 5 s. For a detailed investigation of the aggregation period and its influence on capacity estimates, see Bodendorf et al. (2014).

A first-come-first-served processing order is assumed when advancing pedestrians across area interfaces in presence of congestion, or when assigning seats to standing passengers. Alighting and boarding are assumed to take place consecutively (Wiggenraad, 2001). These assumptions primarily affect individual waiting times, and have only indirect consequences for total process times.

Appendix D. Processing of smart card data

Raw smart card data contains for each trip a check-in and check-out ‘gate group,’ as well as corresponding check-in and check-out times. Gate groups comprise gates located within the same part of a station. In Amsterdam Zuid, for instance, gate groups distinguish transactions at the main and lateral access way, but not at a finer level. For other stations, gates are partitioned similarly. The grouping is required by privacy protection regulations, as it complicates the association of fare collection and video surveillance data.

The transit itinerary choice model, Eq. (1), attaches to each check-in/check-out pair a set of transit itineraries and corresponding choice frequencies. This allows associating with each gate transaction a set of potential trains and corresponding choice probabilities. For privacy reasons, only the latter is used in this study, i.e., augmented smart card data with inferred train assignment probabilities, without any notion of origin-destination (OD) information. Similarly, for each pair of trains, transfer probabilities are available.

For the model framework proposed in this study, individual trip information is required. To reconstruct such information, train-specific OD demand is estimated, from which individual trips are synthesized. This process is straightforward, although cumbersome due to a variety of special cases. In the following, the main concept is described. For more detailed information, the source code or the enclosed documentation may be consulted (Hänseler, 2017).

For a given train serving the UtrechtSchiphol corridor, let

$$\mathbf{x} = [x_{up,U} \ x_{up,B} \ x_{up,A} \ x_{up,S} \ x_{U,B} \ x_{U,A} \ x_{U,S} \ x_{B,A} \ x_{B,S} \ x_{A,S}]^T$$

denote the vector of OD demand, where $x_{o,d}$ represents the passenger demand from station o to station d , with $o \in \{up, U, B, A, S\}$ and $d \in \{U, B, A, S\}$. The set elements $\{U, B, A, S\}$ denote the stations of Utrecht, Bijlmer, Amsterdam Zuid and Schiphol, and ‘up’ is an auxiliary station comprising all stations upstream of Utrecht.

For each train, the privacy-cleared smart card data allows inferring the number of boarding and alighting passengers in station $s \in \{U, B, A, S\}$, denoted by the variables $board_s$ and $alight_s$, respectively. The ridership of a train arriving in Utrecht may additionally be inferred, and is denoted by $ride_U$.

To overcome the underdetermination, average OD split fractions are used. The variable $frac_{o,d}$ denotes the fraction of passengers boarding a train at station o that alight at station d , such that $\sum_{d' \in \mathcal{D}_o} frac_{o,d'} = 1$, where \mathcal{D}_o denotes the set of destination stations served by the train downstream of station o . Such estimates are available for all relevant station pairs, and represent annual averages.

Based on this data, a constrained least squares problem of the form

$$arg \min_{\mathbf{x}} \|\mathbf{B}_1 \mathbf{x} - \mathbf{b}_1\|_2 + \mu \|\mathbf{B}_2 \mathbf{x} - \mathbf{b}_2\|_2 \text{ subject to } \mathbf{x} \geq 0, \tag{D1}$$

is solved for each corridor train, where

$$\mathbf{B}_1 = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{b}_1 = \begin{bmatrix} ride_U \\ alight_U \\ alight_B \\ alight_A \\ alight_S \\ board_U \\ board_B \\ board_A \end{bmatrix}$$

and

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 - frac_{U,B} & -frac_{U,B} & -frac_{U,B} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -frac_{U,A} & 1 - frac_{U,A} & -frac_{U,A} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 - frac_{B,A} & -frac_{B,A} & 0 \end{bmatrix}$$

with $\mathbf{b}_2 = [0 \ 0 \ 0]^T$.

Eq. (D.1) represents a standard OD demand estimation problem (Lawson and Hanson, 1974; Bierlaire et al., 1991). The first term ensures consistency of OD demand with local boarding and alighting volumes, and the second term reduces the solution space by leading the solver towards a solution that reproduces historical OD split ratios. The weight μ is chosen such that it is large enough for the annual OD ratios to be numerically relevant, and small enough not to compromise train-specific boarding and alighting estimates.

Once train-specific OD ratios are estimated, individual trips are synthesized in a sequential manner. This procedure is based on check-in logs, consisting of a gate group and a check-in time, and essentially entails four steps: (i) A train is sampled from the augmented smart card data; (ii) A destination among the stations served by that train is sampled based on the estimated train-specific OD ratios; (iii) Based on the train-specific ratio of transfer passengers and alighting passengers at that station, a connecting train or a

local exit gate group is sampled; (iv) In case a connecting train is sampled, step (ii) is applied again, otherwise the synthesis of the current trip is terminated.

For trains serving only a single station within the considered corridor, boarding passengers are assigned a generic downstream station as destination, and alighting passengers a generic upstream station as origin.

References

- Bagchi, M., White, P.R., 2005. The potential of public transport smart card data. *Transp. Policy* 12 (5), 464–474.
- Banning, J., Guis, N., Siderius, P., 2016. Slimmer voorspellen treinkeuze met smartcard data. In: *Colloquium Vervoer-splanologisch Spuurwerk*, pp. 1–13.
- Bauer, D., Brändle, N., Seer, S., Ray, M., Kitazawa, K., 2009. Measurement of pedestrian movements: A comparative study on various existing systems. In: Timmermans, H. (Ed.), *Pedestrian Behavior: Models, Data Collection and Applications*. Emerald Group Publishing, Bingley, U.K., pp. 325–344.
- Ben-Akiva, M.E., Bierlaire, M., 2003. Discrete choice models with applications to departure time and route choice. In: Hall, R. W. (Ed.), *Handbook of Transportation Science*. Springer, Berlin, Germany, pp. 7–37.
- Bertsekas, D.P., 2017. *Dynamic programming and optimal control*, vol. 1. Athena Scientific, Cambridge, MA, U.S.A.
- Bierlaire, M., Toint, P.L., Tuytens, D., 1991. On iterative algorithms for linear least squares problems with bound constraints. *Linear Algebra Appl.* 143, 111–143.
- Bodendorf, H., Osterkamp, M., Seyfried, A., Holl, S., 2014. Field studies on the capacity of escalators. *Transp. Res. Proc.* 2, 213–218.
- Bureau Spoorbouwmeester, 2012. Het Stationsconcept (“The station concept”; in Dutch). Edited by Schaap, P.M. Available online <http://www.spoorbeeld.nl/bureau-spoorbouwmeester/> (last accessed on November 11, 2019).
- Campanella, M.C., 2016. Microscopic modelling of walking behaviour. Ph.D. thesis. Technische Universiteit Delft.
- Cantarella, G.E., Cascetta, E., 1995. Dynamic processes and equilibrium in transportation networks: towards a unifying theory. *Transp. Sci.* 29 (4), 305–329.
- Cats, O., 2011. Dynamic modelling of transit operations and passenger decisions. Ph.D. thesis. KTH Royal Institute of Technology.
- Cats, O., West, J., Eliasson, J., 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. *Transp. Res. Part B: Methodol.* 89, 43–57.
- Cheung, C.Y., Lam, W.H.K., 1998. Pedestrian route choices between escalator and stairway in MTR stations. *J. Transp. Eng.* 124 (3), 277–285.
- Daamen, W., 2004. Modelling passenger flows in public transport facilities. Ph.D. thesis. Delft University of Technology.
- Daamen, W., Bovy, P.H.L., Hoogendoorn, S.P., 2005. Influence of changes in level on passenger route choice in railway stations. *Transp. Res. Rec.: J. Transp. Res. Board* 1930 (1), 12–20.
- Davidich, M., Geiss, F., Mayer, H.G., Pfaffinger, A., Royer, C., 2013. Waiting zones for realistic modelling of pedestrian dynamics: a case study using two major German railway stations as examples. *Transp. Res. Part C: Emerg. Technol.* 37, 210–222.
- Douglas, N., Karpouzis, G., 2005. Estimating the passenger cost of station crowding. In: 28th Australasian Transport Research Forum, Sydney, Australia. pp. 1–11.
- Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. *Transp. Res. Part B: Methodol.* 56, 70–80.
- Fruin, J.J., 1971. *Pedestrian planning and design*. Metropolitan Association of Urban Designers and Environmental Planners, Michigan, USA.
- Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H.M., Attanucci, J.P., 2013. Automated inference of linked transit journeys in London using fare-transaction and vehicle location data. *Transp. Res. Rec.: J. Transp. Res. Board* 2343, 17–24.
- Goverde, R.M.P., 2007. Railway timetable stability analysis using max-plus system theory. *Transp. Res. Part B: Methodol.* 41 (2), 179–201.
- Guo, Z., Wilson, N.H.M., 2011. Assessing the cost of transfer inconvenience in public transport systems: a case study of the London Underground. *Transp. Res. Part A: Policy Practice* 45 (2), 91–104.
- Hänseler, F.S., 2017. Python implementation of integrated trip assignment model. <https://github.com/flurinus/IntTripAssignment>.
- Hänseler, F.S., Bierlaire, M., Farooq, B., Mühlematter, T., 2014. A macroscopic loading model for time-varying pedestrian flows in public walking areas. *Transp. Res. Part B: Methodol.* 69, 60–80.
- Hänseler, F.S., Bierlaire, M., Scarinci, R., 2016. Assessing the usage and level-of-service of pedestrian facilities in train stations: a Swiss case study. *Transp. Res. Part A: Policy Practice* 89 (7), 106–123.
- Hänseler, F.S., Lam, W.H.K., Bierlaire, M., Lederrey, G., Nikolić, M., 2017a. A dynamic network loading model for anisotropic and congested pedestrian flows. *Transp. Res. Part B: Methodol.* 95, 149–168.
- Hänseler, F.S., Molyneux, N.A., Bierlaire, M., 2017b. Estimation of pedestrian origin-destination demand in train stations. *Transp. Sci.*
- Haywood, L., Koning, M., Monchambert, G., 2017. Crowding in public transport: who cares and why? *Transp. Res. Part A: Policy Practice* 100, 215–227.
- Helbing, D., Molnár, P., 1995. Social force model for pedestrian dynamics. *Phys. Rev. E* 51 (5), 4282–4286.
- Hermant, L.F.L., 2012. Video data collection method for pedestrian movement variables & development of a pedestrian spatial parameters simulation model for railway station environments. Ph.D. thesis. Stellenbosch University.
- Hickman, M.D., Bernstein, D.H., 1997. Transit service and path choice models in stochastic and time-dependent networks. *Transp. Sci.* 31 (2), 129–146.
- Higgins, A., Kozan, E., 1998. Modeling train delays in urban networks. *Transp. Sci.* 32 (4), 346–357.
- Hoogendoorn, S.P., Bovy, P.H.L., 2004. Pedestrian route-choice and activity scheduling theory and models. *Transp. Res. Part B: Methodol.* 38 (2), 169–190.
- Hoogenraad, J.H., De Vos, A.F., De Vries, J.J., 2013. NS-project AURORA voorspelt voor elke trein het aantal reizigers (NS project AURORA predicts for each train the number of passengers). *Statistiek en Operationele Research (STATOR)* 14 (3-4), 28–31.
- Hörcher, D., Graham, D.J., Anderson, R.J., 2017. Crowding cost estimation with large scale smart card and vehicle location data. *Transp. Res. Part B: Methodol.* 95, 105–125.
- Jiang, Y., Zhu, J., Hu, L., Lin, X., Khattak, A., 2015. A G/G(n)/C/C state-dependent simulation model for metro station corridor width design. *J. Adv. Transp.* 50, 273–296.
- Kaakai, F., Hayat, S., El Moudni, A., 2007. A hybrid Petri nets-based simulation model for evaluating the design of railway transit stations. *Simulat. Model. Practice Theory* 15 (8), 935–969.
- Kim, H., Kwon, S., Wu, S.K., Sohn, K., 2014. Why do passengers choose a specific car of a metro train during the morning peak hours? *Transp. Res. Part A: Policy Practice* 61, 249–258.
- King, D., Srikukenthiran, S., Shalaby, A., 2014. Using simulation to analyze crowd congestion and mitigation at Canadian subway interchanges: case of Bloor-Yonge Station, Toronto, Ontario. *Transp. Res. Rec.: J. Transp. Res. Board* 2417, 27–36.
- Kleinhou, E., 2016. Overbelastverklaring nabije toekomst: Utrecht perronspoor 5. Technical Report (published on April 24, 2017), ProRail.
- Kusakabe, T., Iryo, T., Asakura, Y., 2010. Estimation method for railway passengers’ train choice behavior with smart card transaction data. *Transportation* 37 (5), 731–749.
- Lam, W.H.K., Cheung, C.Y., Lam, C.F., 1999. A study of crowding effects at the Hong Kong light rail transit stations. *Transp. Res. Part A: Policy Practice* 33 (5), 401–415.
- Lawson, C. L., Hanson, R. J., 1974. *Solving least squares problems*. Vol. 161. Society for Industrial and Applied Mathematics.
- Lee, J.Y.S., Lam, W.H.K., Wong, S.C., 2001. Pedestrian simulation model for Hong Kong underground stations. In: *Intelligent Transportation Systems*. IEEE, pp. 554–558.
- Liu, Z., Li, D., Wang, X., 2016. Passenger distribution and waiting position selection model on metro station platform. In: *International Conference on Civil, Structure, Environmental Engineering*, pp. 1–7.
- Løvås, G.G., 1994. Modeling and simulation of pedestrian traffic flow. *Transp. Res. Part B: Methodol.* 28 (6), 429–443.

- Munizaga, M.A., Palma, C., 2012. Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. *Transp. Res. Part C* 24, 9–18.
- Nederlandse Spoorwegen, 2016. NS Annual Report 2016. NS Groep N.V.
- Nielsen, B.F., Frølich, L., Nielsen, O.A., Filges, D., 2014. Estimating passenger numbers in trains using existing weighing capabilities. *Transport. A: Transp. Sci.* 10 (6), 502–517.
- Nikolić, M., Bierlaire, M., Farooq, B., de Lapparent, M., 2016. Probabilistic speed-density relationship for pedestrian traffic. *Transp. Res. Part B: Methodol.* 89, 58–81.
- NS Reizigers, 2004. Kenmerken rollend materieel voor treinpersoneel. Nederlandse Spoorwegen.
- Nuzzolo, A., Crisalli, U., Rosati, L., 2012. A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transp. Res. Part C: Emerg. Technol.* 20 (1), 16–33.
- Nuzzolo, A., Russo, F., Crisalli, U., 2001. A doubly dynamic schedule-based assignment model for transit networks. *Transp. Sci.* 35 (3), 268–285.
- Openshaw, S., 1984. The modifiable areal unit problem. *Geo Books*, Norwick.
- Pel, A.J., Bel, N.H., Pieters, M., 2014. Including passengers' response to crowding in the Dutch national train passenger assignment model. *Transp. Res. Part A: Policy Practice* 66, 111–126.
- Pelletier, M.-P., Trépanier, M., Morency, C., 2011. Smart card data use in public transit: a literature review. *Transp. Res. Part C: Emerg. Technol.* 19 (4), 557–568.
- Raveau, S., Guo, Z., Muñoz, J.C., Wilson, N.H., 2014. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. *Transp. Res. Part A: Policy Practice* 66, 185–195.
- Ross, J., 2000. *Railway Stations: Planning, Design and Management*. Architectural Press.
- Singh, J., Nandakumar, K., Wynter, L., 2016. Online agent-based pedestrian simulation for complex environments: a hybrid simulation-video analytics approach. In: *AAMAS Workshop on Transportation Applications of Equilibrium, Incentives and Game Theory*.
- Spies, H., Florian, M., 1989. Optimal strategies: a new assignment model for transit networks. *Transp. Res. Part B: Methodol.* 23 (2), 83–102.
- Srikukenthiran, S., 2015. Integrated microsimulation modelling of crowd and subway network dynamics for disruption management support. Ph.D. thesis. University of Toronto.
- Sun, L., Lee, D.H., Erath, A., Huang, X., 2012. 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*. ACM, pp. 142–148.
- Sun, Y., Xu, R., 2012. Rail transit travel time reliability and estimation of passenger route choice behavior: analysis using automatic fare collection data. *Transp. Res. Rec.: J. Transp. Res. Board* 2275, 58–67.
- Szplett, D., Wirasinghe, S.C., 1984. An investigation of passenger interchange and train standing time at LRT stations: (i) alighting, boarding and platform distribution of passengers. *J. Adv. Transp.* 18 (1), 1–12.
- Teklu, F., 2008. A stochastic process approach for frequency-based transit assignment with strict capacity constraints. *Networks Spatial Econ.* 8 (2-3), 225–240.
- Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. *Transp. Res. Part A: Policy Practice* 53, 36–52.
- Toledo, T., Cats, O., Burghout, W., Koutsopoulos, H.N., 2010. Mesoscopic simulation for transit operations. *Transp. Res. Part C: Emerg. Technol.* 18 (6), 896–908.
- Tordeux, A., Lämmel, G., Hänseler, F.S., Steffen, B., 2018. A mesoscopic model for large-scale simulation of pedestrian dynamics. *Transp. Res. Part C: Emerg. Technol.* 93, 128–147.
- Tregenza, P., 1976. *The design of interior circulation*. Van Nostrand Reinhold, New York, USA.
- Trépanier, M., Tranchant, N., Chapleau, R., 2017. Individual trip destination estimation in a transit smart card automated fare collection system. *J. Intelligent Transp. Syst.* 1–14.
- U.S. Department of Transportation, 2013. *Traffic monitoring guide*. Tech. rep., Federal Highway Administration.
- Vaatstra, I., 2017. Capaciteitsanalyse Utrecht Centraal (transferknooppunt perron 5/7). Technical Report K20170020- 53433699-2, ProRail.
- van den Heuvel, J., 2016. Field experiments with train stopping positions at Schiphol Airport Train Station in Amsterdam, Netherlands. *Transp. Res. Rec.: J. Transp. Res. Board* 2546, 24–32.
- van den Heuvel, J.P.A., Hoogenraad, J.H., 2014. Monitoring the performance of the pedestrian transfer function of train stations using automatic fare collection data. *Transp. Res. Proc.* 2, 642–650.
- van den Heuvel, J.P.A., Ton, D., Hermansen, K., 2016. Advances in measuring pedestrians at Dutch train stations using Bluetooth, WiFi and infrared technology. In: *Traffic and Granular Flow '15*. Springer, pp. 11–18.
- Wahba, M., Shalaby, A., 2005. Multiagent learning-based approach to transit assignment problem: a prototype. *Transp. Res. Rec.: J. Transp. Res. Board* 1926, 96–105.
- Wardman, M., 2004. Public transport values of time. *Transp. Policy* 11 (4), 363–377.
- Wardman, M., Whelan, G., 2011. Twenty years of rail crowding valuation studies: evidence and lessons from British experience. *Transp. Rev.* 31 (3), 379–398.
- Weidmann, U., 1992. *Transporttechnik der Fussgänger*. Schriftenreihe des IVT Nr. 90. Institute for Transport Planning and Systems, ETH Zürich, Switzerland.
- Wiggenraad, P., 2001. Alighting and boarding times of passengers at Dutch railway stations. TRAIL Research School, Delft.
- Xu, X.Y., Liu, J., Li, H.Y., Hu, J.Q., 2014. Analysis of subway station capacity with the use of queueing theory. *Transp. Res. Part C: Emerg. Technol.* 38, 28–43.
- Zhu, Y., Koutsopoulos, H.N., Wilson, N.H., 2017. A probabilistic passenger-to-train assignment model based on auto-mated data. *Transp. Res. Part B: Methodol.* 104, 522–542.