

Estimation of pedestrian origin-destination demand in train stations

Flurin S. Hänseler* Nicholas A. Molyneaux*
Michel Bierlaire*

July 3, 2015

Report TRANSP-OR 150703
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

This is a revised version of technical report #150108.

* École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, Switzerland, {flurin.haenseler,nicholas.molyneaux,michel.bierlaire}@epfl.ch

We present a framework for estimating pedestrian demand within a train station. It takes into account ridership data, and various direct and indirect indicators of demand. Such indicators may include link flow counts, density measurements, survey data, historical or other information. The problem is considered in discrete time and at the aggregate level, i.e., for groups of pedestrians associated with the same origin-destination pair and departure time interval. The formulation is probabilistic, allowing to consider the stochasticity of demand. A key element of the framework is the use of the train timetable to better capture demand peaks. A case study analysis of a Swiss train station underlines its practical applicability. Compared to a classical estimator that ignores the notion of a train timetable, the gain in accuracy in terms of RMSE is between 20% and 50%. More importantly, the incorporation of the train schedule allows for prediction when little or no data besides the timetable and ridership information is available.

Keywords: Origin-destination demand; schedule-based estimation; pedestrian flows; public transportation.

1 Introduction

Pedestrian behavior in train stations increasingly attracts the attention of academic research. Broadly, it can be distinguished between empirical studies (Daly et al., 1991; Cheung and Lam, 1998; Pettersson, 2011; Ganansia et al., 2014), and those concerned with its modeling (Lee et al., 2001; Daamen, 2004; Kaakai et al., 2007; Xu et al., 2014). A review of the literature is provided by Mustafa and Ashaari (2015).

Most studies make great efforts in describing behavioral aspects such as walking, waiting or boarding. At the same time, only few methods are proposed in the literature to estimate pedestrian demand from data. Many studies are solely based on theoretical demand scenarios (Hoogendoorn and Daamen, 2004; Rindsfuser and Klügl, 2007; Davidich et al., 2013). Other studies rely on simple assumptions to estimate demand (Kaakai et al., 2007; van den Heuvel and Hoogenraad, 2014), or consider railway stations that serve only a single line (Lee et al., 2001; Xu et al., 2014). Yet knowledge of pedestrian demand is a prerequisite for the analysis of pedestrian flows in train stations, be it for the design of infrastructure, the optimization of operations such as the train-track assignment, or real-time management and control of pedestrian flows.

Several approaches to estimate pedestrian demand in train stations seem conceivable. For instance, an activity-based approach could be pursued (Danalet et al., 2014). However, for most train stations, disaggregate data is still unavailable, or only with low sampling rates and low temporal or spatial resolution. Instead, it is more efficient to estimate origin-destination (OD) demand at the aggregate level. Pedestrians may be divided into user classes that are characterized by a common activity or behavior pattern, which in the context of a train station could be ‘inbound’, ‘outbound’, but also ‘elderly’, ‘in a hurry’, or any combination thereof (Wong et al., 2005; Lavadinho, 2012).

The problem of estimating OD demand has a long history in the context of road networks, for which link flow volumes and other indirect observations of demand are available (van Zuylen and Willumsen, 1980; Cascetta, 1984). Typically, an assignment map is assumed that relates observations to OD volumes, such that the latter can be ‘reverse engineered’. This map may be obtained via a dynamic traffic assignment (DTA) model, combining travel behavior models such as mode, departure time and route choice

models, as well as a network loading model. In the context of pedestrian flows in train stations, mode choice is irrelevant, and the choice of departure time is mainly governed by the train timetable, which is discussed in this work. A variety of approaches have been proposed for route choice (Cheung and Lam, 1998; Hoogendoorn and Bovy, 2004; Daamen et al., 2005), and for network loading models a rich literature is available as well (Lee et al., 2001; Daamen, 2004; Kaakai et al., 2007; Xu et al., 2014; Starmans et al., 2014). The problem of pedestrian OD demand estimation in train stations is thus in principle amenable to ‘classical’ estimation techniques.

A key issue in OD demand estimation is the ratio between the number of unknowns and the number of independent observations, yielding an intrinsically underdetermined problem. Various forms of exogenous information, either in the form of a priori knowledge or structural assumptions, are used to lead the calculation to a unique solution.

For static OD estimation, concepts like gravity (Casey, 1955), entropy maximization (Wilson, 1970; Willumsen, 1981) or information minimization (van Zuylen and Willumsen, 1980) have been used. In most cases, however, an a priori OD trip table (Cascetta and Nguyen, 1988) is used. Other researchers make specific assumptions on the structure of OD trip tables (Bierlaire and Toint, 1995) or the covariance across measurements (Hazelton, 2003).

For dynamic problems, a common approach is to assume a dynamic process for the evolution of demand, such as an autoregressive process in the deviates from historical estimates (Ashok and Ben-Akiva, 2000; Bierlaire and Crittin, 2004; Zhou and Mahmassani, 2007). Recent approaches assume slowly evolving route split fractions in the framework of a ‘quasi-dynamic’ estimator (Marzano et al., 2009; Cascetta et al., 2013), or reduce the dimensionality of the estimation problem by applying principal component analysis (Djukic et al., 2012).

Several researchers consider also the problem of OD demand estimation in transit networks. Early approaches assume a constant average cost along routes (Nguyen et al., 1988), whereas newer studies focus on schedule-based transit network models (Wong and Tong, 1998), of which some additionally consider passenger overload delays (Lam et al., 2003) or data from ICT sensors (Montero et al., 2015). These models can predict the evolution of in- and outflows at stations or the number of passengers in vehicles, but do not provide detailed information on OD demand *within* a train station.

Pedestrian OD demand in train stations is particularly unsteady due to arriving and departing trains that lead to demand ‘micro-patterns’. Moreover, acyclic schedules and unplanned delays make it difficult to use historical OD data, or any other of the aforementioned approaches for dealing with underdetermination. This is where we would like to make a contribution. In this paper, we propose a dedicated methodology for estimating pedestrian OD demand in train stations in general, and we do so by explicitly integrating the train timetable and ridership data in particular.

2 Data sources

To reduce the underdetermination, the use of any relevant, available data is desirable. In comparison to motorized traffic, there are several inherent challenges that make the monitoring of pedestrian traffic difficult. First, pedestrians are not confined to fixed lanes, and can explore space freely. This makes the placement of sensors difficult, and may decrease their accuracy. Second, pedestrians often travel in groups. Special care must be taken to differentiate between individuals, and to take into account effects such as occlusion (Alahi et al., 2014). Third, pedestrian traffic is typically more variable than motorized traffic, implying that sensors are required to capture a large range of traffic levels (cf. Traffic Monitoring Guide, U.S. Department of Transportation, 2013).

In the following, five *types* of data sources are discussed that are relevant for the case of a train station. For a discussion of sensing technologies, including guidance on making the most appropriate choice for selected practical applications, we refer instead to Turner et al. (2007), Bauer et al. (2009) and the aforementioned Traffic Monitoring Guide (U.S. Department of Transportation, 2013).

OD flow data: OD flow data represent direct observations of OD demand that are obtained from surveys, pedestrian tracking systems, or ICT sensors (e.g. Bluetooth and WiFi scanners for smartphones). Such observations are generally expensive to collect and rare (Bauer et al., 2009).

Automatic collection techniques depend on the location of sensors, and their sampling rate. Typically, they do not cover the entire network of interest. Also, if only a subset of pedestrians is detected, this needs to

be corrected by means of sampling rates. Their estimation is difficult, as they are generally time- and location-dependent (Bauer et al., 2009). There exist two ways to deal with that. Either the sampling rates are estimated a priori, or directly within the OD estimation process. While the latter is theoretically more attractive due to its generality, it is also computationally more expensive.

Link flow data: In analogy to car traffic, a pedestrian facility may be thought of as a network of links. Links denote walkways or walkway sections, such as a part of a corridor, stairway, or an escalator. On links, flows can be observed at a physical gate, such as a turnstile or a train door, or at a virtual gate like the entrance of a walkway. Link flow data represent indirect observations of OD demand, and depend on route choice decisions and prevailing traffic conditions.

Link flow data are typically more accurate than OD flow data, but sampling may still be an issue, especially for camera-based detectors (Ganansia et al., 2014). The number and position of detectors plays a crucial role for the observability of OD demand (Gentili and Mirchandani, 2012; Yang and Fan, 2015). Ideally, if a link exclusively serves routes associated with a single OD pair, a well-placed detector may be used to directly infer the OD volumes for that particular OD pair. Also, if detectors are located on links that are adjacent to origin and destination nodes, they may provide an estimate of the generation of that node. On the other hand, if measurements are highly correlated, further sensors may not provide substantial additional information.

Other traffic condition data: Other traffic condition data include densities, walking speeds and point-to-point travel times. They characterize the system response of the traffic network to a given OD demand.

Such data may be obtained from pedestrian tracking systems or ICT sensors. Speed or density measurements can for instance help to identify whether a link is in a congested or uncongested state, and thus to adjust the OD demand in one way or another (Djukic et al., 2015). There exist several ways of including this type of input data in the estimation process. One is to include it in the objective function of the estimation problem. A DTA model is then used to define the relationship between OD demand and traffic conditions, and a match between observation and estimation

is sought. Alternatively, traffic condition data may be used to replace the DTA model altogether. For instance, Montero et al. (2015) use travel times collected from ICT sensors to estimate the travel time distributions. These are then used as time-varying exogenous model parameters.

Train timetable and ridership data: Pedestrian demand within a train station and the train timetable are inextricably intertwined. To establish a formal relationship between the two, the train-track assignment and the train exchange volumes are useful, i.e., the number of boarding and alighting passengers for each train.

Unplanned changes in the train timetable and train-track assignment are common in most railway systems around the world (Higgins and Kozan, 1998; Cule et al., 2011). For highly inter-connected timetables or dense railway traffic, a single delayed train may cause a domino effect of secondary delays due to infrastructure restrictions, connection constraints or logistics. Several approaches are available to predict the actual timetable based on the scheduled one (Goverde, 2007; Yuan and Hansen, 2007).

The train exchange volumes may be inferred from traffic surveys, ticket sales, or train capacities. Additionally, some trains are equipped with door counters that not only allow for an automatic detection of these volumes, but also for an estimation of their distribution across vehicles. Especially on long platforms, such distributional information can be important for an accurate understanding of the usage of a train station.

Other data: Further data sources can be useful to narrow the solution space. Such information typically comes in the form of survey data. In many train stations, service and sales points are found. The number of customer visits to these places may be known, and can be used as an a priori estimate of the corresponding origin or destination flows. Similarly, railway operators often have an idea of the relative share of certain user classes, such as transfer passengers. This information may be used in the form of destination split ratios.

3 Methodological framework

This section presents a methodological framework for estimating pedestrian OD demand based on the notion of a train timetable, and an exemplary specification that is applicable to any suitable train station. Section 4 then presents a case study based on that specification.

3.1 Notation

A recapitulation of important variables is provided in Appendix A.

3.1.1 Time and space representation

The time period of interest is divided into a set of discrete intervals \mathcal{T} , where each interval $\tau \in \mathcal{T}$ is of uniform length Δt .

Walkable space is represented by a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{L})$, where \mathcal{N} is the set of nodes $v \in \mathcal{N}$, and \mathcal{L} the set of directed links $\lambda \in \mathcal{L}$. Certain elements of pedestrian facilities, such as stairs or corridors, translate naturally into links, and others naturally into nodes, like for instance OD areas. For other elements, such as waiting halls or platforms, a decomposition into areas can be made. An area α is associated with a subnetwork $(\mathcal{N}_\alpha, \mathcal{L}_\alpha)$ denoted by \mathcal{G}_α . The set \mathcal{N}_α contains all the nodes corresponding to physical locations in the area, and $\mathcal{L}_\alpha \subset \mathcal{L}$ all links such that their two incident nodes belong to \mathcal{N}_α . Areas are allowed to overlap, and their union is not required to cover the full network.

Fig. 1 illustrates the proposed space representation. Railway tracks are denoted by dotted lines. Levels are bridged by ramps and stairways, denoted by standard floor plan symbols. Platform sectors are represented by centroids shown as rectangles with rounded corners. They may be associated with one or a pair of railway tracks. Further centroids are shown as squares, which include sales or service points, or exit/entrance areas. The pedestrian walking network is represented by solid lines. An exemplary area is shaded.

3.1.2 Demand representation

Nodes through which pedestrians enter and leave the pedestrian network are referred to as centroids, and their set is denoted by $\mathcal{C} \subset \mathcal{N}$.

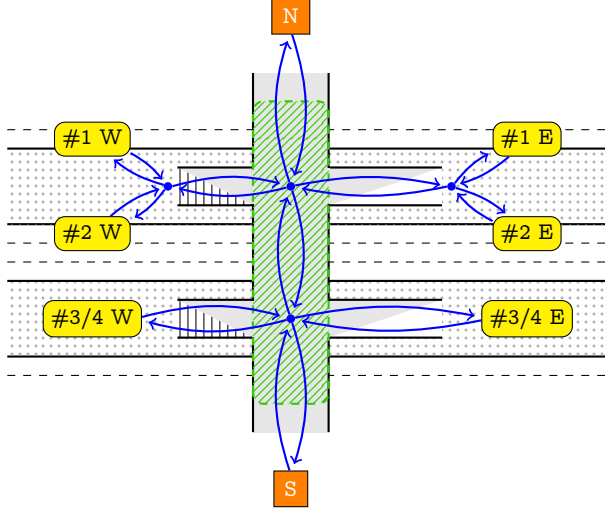


Figure 1: Network topology at the example of a simple train station.

We assume that the network is connected, i.e., any two centroids are connected by at least one route. A route ρ is a sequence of links connecting two centroids, $\rho = (\lambda_1, \lambda_2, \dots)$. A sequence of links that does not connect two centroids is called a subroute. The set of routes and subroutes is denoted by \mathcal{R} and \mathcal{R}_{sub} , respectively.

Each pedestrian is associated with a pair of centroids, denoted by $\kappa \in \mathcal{K}$ where \mathcal{K} is the set of OD pairs. In principle, it is possible to associate pedestrians with further attributes, such as behavior classes (Wong et al., 2005). The necessary extension is straightforward, and useful if personal attributes such as trip purpose, age or gender are available. In the interest of a more readable notation, in this work a single-class formulation is considered.

The number of travelers associated with OD pair κ that depart during time interval τ is represented by $d_{\kappa, \tau}$. We refer to this variable as demand. Demand is by nature stochastic, i.e., it varies from day to day.

A set of demand indicators is derived from OD demand to facilitate the formulation of the structural and measurement models. These include in particular link flows and area accumulations. Flow is cumulative over a time interval τ . Accumulation is defined as the time-mean average number of users in an area (Edie, 1963). Table 1 provides an overview of these indicators. Further demand indicators are defined in Section 3.2 to integrate specific data sources.

Table 1: List of demand and demand indicators. The unit is ‘number of pedestrians per unit of time’, unless stated otherwise.

$\mathbf{d} = [d_{\kappa,\tau}]$	<u>D</u> emand $d_{\kappa,\tau}$ associated with OD pair κ and departure time interval τ , and time-space expanded vector \mathbf{d} of length $ \mathcal{K} \mathcal{T} $.
$\mathbf{f} = [f_{\lambda,\tau}]$	<u>F</u> low $f_{\lambda,\tau}$ entering link λ during time interval τ , and time-space expanded vector \mathbf{f} of length $ \mathcal{L} \mathcal{T} $.
$\mathbf{a} = [a_{\alpha,\tau}]$	T <u>A</u> me-mean average <u>a</u> ccumulation $a_{\alpha,\tau}$ on area α during time interval τ , and time-space expanded vector \mathbf{a} of length $ \mathcal{A} \mathcal{T} $.
$e_{\text{off}} = [e_{\zeta}^{\text{off}}]$, $e_{\text{on}} = [e_{\zeta}^{\text{on}}]$	T <u>R</u> ain <u>e</u> xchange volumes associated with alighting, e_{ζ}^{off} , and boarding, e_{ζ}^{on} , of train ζ , and corresponding vectors e_{off} and e_{on} of length $ \mathcal{K} $. The unit is ‘number of pedestrians’.

3.1.3 Representation of trains

A set of trains \mathcal{Z} is considered. For a train $\zeta \in \mathcal{Z}$, t_{ζ}^{arr} and t_{ζ}^{dep} denote the actual arrival and departure time in the train station. They are assumed to follow a known random distribution that may be obtained empirically, or from any suitable delay model. Each train is associated with an alighting and boarding volume, referred to as train exchange volumes and denoted by e_{ζ}^{off} and e_{ζ}^{on} , respectively. The platform serving train ζ is denoted by π_{ζ} . Each platform $\pi \in \mathcal{P}$, with \mathcal{P} the set of all platforms, is associated with a set of centroids, $\mathcal{C}_{\pi} \subset \mathcal{C}$.

3.2 Data requirements

To distinguish between model estimates and actual observations, ‘biased’ variables such as measurements are marked by a hat (e.g. \hat{f}). Often, such observations are not available for the complete network, or only for certain time intervals. Vectors containing a reduced set of variables are marked by a prime (e.g. f'), and a reduction matrix Δ is defined that relates each of them to the corresponding full vector (e.g. $f' = \Delta_f f$).

For the estimation methodology, availability of the actual train timetable, t_{ζ}^{arr} and t_{ζ}^{dep} for each train $\zeta \in \mathcal{Z}$, and of the corresponding exchange volumes $\hat{e}_{\zeta}^{\text{off}}$, $\hat{e}_{\zeta}^{\text{on}}$ is essential. Moreover, partial indirect observations of demand, for instance in the form of link flows \hat{f}' or area accumulations \hat{a}' , are required. These observations need to be such that demand micro-patterns of individual trains are captured, i.e., an aggregation in the order of min-

utes is desirable. Availability of an a priori estimate of demand \hat{d}' is useful to improve the estimation, but not mandatory.

Example specification: To illustrate the demand estimation methodology, a concrete specification is elaborated. For that purpose, several assumptions are made throughout the document. The general estimation methodology is independent of these assumptions.

Assumption 1 (Data availability) *Available are (i) the actual train timetable, (ii) train exchange volumes, (iii) partial observations of link flows, (iv) aggregate destination split ratios (e.g. from travel surveys), and (v) cumulative origin and destination flows for selected centroids (e.g. from sales data). For validation, (vi) partial observations of area accumulations, and (vii) flows along selected subroutes are available. No historical demand prior is considered.*

To capture the format of these data sources, additional demand indicators are defined in Table 2.

Table 2: Additional demand indicators.

$\mathbf{f}_{\text{sub}} = [f_{\rho,\tau}^{\text{sub}}]$	Subroute flow $e_{\rho,\tau}^{\text{sub}}$ reaching subroute ρ during time interval τ , and time-space expanded vector \mathbf{f}_{sub} of length $ \mathcal{R}_{\text{sub}} \mathcal{T} $. Its unit is ‘number of pedestrians per unit time’.
$\mathbf{f}_{\text{out,cum}} = [f_{\nu}^{\text{out,cum}}]$	Cumulative origin flow $f_{\nu}^{\text{out,cum}}$ emanating from centroid ν during the time period \mathcal{T} , and vector $\mathbf{f}_{\text{out,cum}}$ of length $ \mathcal{C} $. Its unit is ‘number of pedestrians’.
$\mathbf{f}_{\text{in,cum}} = [f_{\nu}^{\text{in,cum}}]$	Cumulative destination flow $f_{\nu}^{\text{in,cum}}$ reaching centroid ν during time period \mathcal{T} , and vector $\mathbf{f}_{\text{in,cum}}$ of length $ \mathcal{C} $. Its unit is ‘number of pedestrians’.
$\mathbf{r}_{\text{avg}} = [r_{\nu}^{\text{avg}}]$	Time-mean average ratio r_{ν}^{avg} of users headed for a platform in the origin flow at centroid ν during time period \mathcal{T} , and vector \mathbf{r}_{avg} of length $ \mathcal{C} $. The destination split ratio r_{ν}^{avg} is dimensionless.

3.3 Structural model

The structural model describes the relationship among the various variables involved in the framework. We consider two parts, namely an assignment

model, and a schedule-based model that considers the arrivals and departures of trains.

3.3.1 Assignment model

A pre-specified aggregate network supply model, referred to as assignment model, is assumed to exist. It is designed to derive the demand indicators from a given demand, depending on a parameter vector \mathbf{y} . If $\Sigma(\mathbf{d}; \mathbf{y})$ denotes the assignment model, and if $\Sigma_{(\cdot)}$ is its output with respect to demand indicator (\cdot) and $\eta_{(\cdot)}$ the corresponding structural error, the aforementioned demand indicators may be expressed as

$$\mathbf{f} = \Sigma_{\mathbf{f}}(\mathbf{d}; \mathbf{y}) + \eta_{\mathbf{f}}, \quad (1)$$

$$\mathbf{a} = \Sigma_{\mathbf{a}}(\mathbf{d}; \mathbf{y}) + \eta_{\mathbf{a}}. \quad (2)$$

In this work, we assume the vector \mathbf{y} to be known a priori, but note that it could also be estimated simultaneously with demand. Such an approach incurs substantial computational cost, and is not commonly pursued in the literature (Cascetta and Improta, 2002).

To implement Eq. (1) and Eq. (2), any suitable supply model may be used. It can be a simple linear mapping, or a detailed commercial DTA model such as PTV Viswalk or Legion for Aimsun. Essential is that basic supply variables like flow and density are provided. Additional information such as user class-specific properties or walking speeds may be useful to improve the estimation.

Internally, most assignment models perform two steps to obtain an estimate of demand indicators. First, OD demand is mapped to route flows by means of a route choice model. For a given OD pair and known link and route attributes, it identifies the route that a traveler would select. The choice of alternatives, and all attributes are assumed to be known. A large number of route choice models are available for that purpose (see e.g. Dial, 1971; Cascetta et al., 1996; Ben-Akiva and Bierlaire, 2003; Frejinger and Bierlaire, 2007). Second, a network loading model is used to describe the propagation of pedestrians along the routes. A large number of models is available in the literature as well (e.g. Løvås, 1994; Helbing and Molnár, 1995; Blue and Adler, 2001; Hughes, 2002; Antonini et al., 2006; Hänseler et al., 2014). To represent heterogeneity among pedestrians, route choice

and network loading are usually expressed by means of probability distributions.

Both route choice and network loading are subject to prevailing traffic conditions, and thus mutually dependent. If the dependency on prevailing traffic conditions is neglected, the relationship between demand and derived indicators becomes linear (Cascetta and Improta, 2002). This holds true for an uncongested network. Alternatively, if the traffic situation is known a priori through direct measurements, an estimate of the assignment maps may also be obtained without considering the demand (see the aforementioned example by Montero et al., 2015).

If on the other hand a network is congested and link travel times are unknown, a problem of circular dependence arises between the demand estimation and the network supply model. One way of dealing with that is by formulating a bi-level optimization problem that explicitly includes traffic equilibrium conditions. Among the most popular studies pursuing such an approach are those by Fisk (1988), Yang (1995) and Florian and Chen (1995). An alternative way to consider the mutual dependency between the demand and supply model is by using a fixed-point formulation (Cascetta and Postorino, 2001; Bierlaire and Crittin, 2006).

Example specification: An assignment model for pedestrian walking facilities in a train station with a low level of congestion is considered. It consists of two independent models for route choice and network loading. For the sake of simplicity, we consider an assignment that is demand-independent.

Following Dial (1971), we adapt a probabilistic route choice model that is suitable for traffic assignment and behaviorally accurate in the context of pedestrian flows (Bierlaire and Robin, 2009).

Assumption 2 (Route choice) *The route choice decision rule is given by a logit model, where the cost of a route is equal to the sum of link traversal times. The set of routes is finite and known.*

Following Mustafa and Ashaari (2015), we assume that walking speed in pedestrian facilities of a train station with a low or medium level of congestion is normally distributed (LOS E or better, Highway Capacity Manual, 2000, Exhibit 18-3).

Assumption 3 (Network loading) *The propagation of pedestrians along routes is described by a demand-invariant walking speed distribution $f_v(v)$. The corresponding cumulative distribution function is denoted by $F_v(v)$.*

The resulting mathematical specification of the assignment model is provided in Appendix B.

3.3.2 Schedule-based model

The schedule-based model establishes a relationship between OD demand and train exchange volumes. It is based on the assumption that the alighting volume of trains served by a specific platform is related to the demand emanating from centroids representing that platform, and vice versa for boarding volumes.

Pedestrian demand within a train station is associated with alighting volumes by an assignment matrix $\mathbf{H} = [h_{\zeta,(\kappa,\tau)}]$ and a corresponding error ε_{off} such that

$$e_{\text{off}} = \mathbf{H}\mathbf{d} + \varepsilon_{\text{off}}. \quad (3)$$

The error ε_{off} takes into account pedestrians that visit a platform by mistake, or e.g. to accompany a passenger. The entry $h_{\zeta,(\kappa,\tau)}$ represents the proportion of pedestrians associated with OD pair κ and departure time interval τ that have alighted from train ζ . It is high if the time interval τ coincides with the idling time $[t_{\zeta}^{\text{arr}}, t_{\zeta}^{\text{dep}}]$ of train ζ on platform π , and if the origin node v_{κ}^o of OD pair κ is associated with the corresponding platform, i.e., if $v_{\kappa}^o \in \mathcal{C}_{\pi}$. Otherwise, it is zero. Under the basic assumption that demand is distributed homogeneously within a demand interval, the entries of the assignment matrix \mathbf{H} are given by

$$h_{\zeta,(\kappa,\tau)} = \begin{cases} \left| \left[t_{\zeta}^{\text{arr}}, t_{\zeta}^{\text{dep}} \right] \cap \tau \right| / |\tau| & \text{if } v_{\kappa}^o \in \mathcal{C}_{\pi}, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where $|\tau|$ represents the length of time interval τ .

In principle, a similar approach may be used to relate OD demand to boarding volumes. However, it is difficult to find a meaningful specification of the corresponding assignment matrix. Prospective passengers often arrive at the platform long before the scheduled departure, which may be

due to constraints imposed by the schedule of tertiary transport modes, or a high risk aversion (van Hagen, 2011). We leave the development of an appropriate arrival process, for instance based on a Poisson distribution, for future research.

For now, boarding volumes may be considered in a temporally aggregated way. We denote by $f_{\pi}^{\text{dep,cum}}$ the cumulative departure flow from platform π during time period \mathcal{T} , given by

$$f_{\pi}^{\text{dep,cum}} = \sum_{\tau \in \mathcal{T}} \sum_{v \in \mathcal{C}_{\pi}} \sum_{\kappa \in \mathcal{K}_{v}^{\text{dest}}} d_{\kappa,\tau} \Delta t, \quad (5)$$

where the set $\mathcal{K}_{v}^{\text{dest}}$ contains all OD pairs with destination v . The corresponding vector $\mathbf{f}_{\text{dep,cum}} = [f_{\pi}^{\text{dep,cum}}]$ is of length $|\mathcal{P}|$.

If ε_{χ} represents a vector containing structural errors, the vector of cumulative platform departure flows can also be expressed by summing over the boarding volumes of the served trains, i.e.,

$$\mathbf{f}_{\text{dep,cum}} = \chi(\mathbf{e}_{\text{on}}) + \varepsilon_{\chi}, \quad (6)$$

where $\chi = [\chi_{\pi}]$ is given by

$$\chi_{\pi}(\mathbf{e}_{\text{on}}) = \sum_{\zeta \in \mathcal{Z}_{\pi}} e_{\zeta}^{\text{on}}, \quad (7)$$

and where \mathcal{Z}_{π} is the set of trains associated with platform π .

Eq. (6) provides no information about the distribution of demand across time. Similarly, Eq. (3) may not provide significant temporal information unless the train idling times are of approximately the same length as the discretization time intervals.

Empirical relations between OD demand and exchange volumes may instead be used to obtain such temporal information. This approach is illustrated at the example of ‘train-induced arrival flows’, and further discussed in the specification below.

We assume there exists an empirical model that predicts the flow on platform exit ways caused by pedestrians that have alighted from a train.

Let $\mathcal{L}_{\pi}^{\text{arr}}$ denote the set of links representing the exit ways of platform π , and $\phi_{\lambda,\tau}(\mathbf{e}_{\text{off}}; \mathbf{y})$ a model that predicts the cumulative flow on link $\lambda \in \mathcal{L}_{\pi}^{\text{arr}}$ during time interval τ based on the arrival times of trains and their alighting volumes. If $\varphi(\mathbf{e}_{\text{off}}; \mathbf{y}) = [\phi_{\lambda,\tau}]$ represents the corresponding time-space expanded vector, it holds that

$$\mathbf{f}_{\text{arr}} = \varphi(\mathbf{e}_{\text{off}}; \mathbf{y}) + \varepsilon_{\varphi}, \quad (8)$$

where ε_φ denotes a structural error, and where the flow vector associated with arrival links is given by

$$\mathbf{f}_{\text{arr}} = \mathbf{\Delta}_{\text{arr}} \mathbf{f} \quad (9)$$

and where the reduction matrix $\mathbf{\Delta}_{\text{arr}}$ is of size $|\mathcal{L}_{\mathcal{P}}^{\text{arr}}| |\mathcal{T}| \times |\mathcal{L}| |\mathcal{T}|$.

Eq. (8) can be used to merely complement, or to replace Eq. (3). If an accurate empirical model is available, Eq. (3) does not provide much additional information, and can be omitted. This is assumed to be the case in the specification below. If on the other hand both Eq. (3) and Eq. (8) are used, a strong correlation among their error terms is likely to exist and needs to be explicitly considered.

Example specification: Our approach is inspired by Benmoussa et al. (2011) and Lavadinho (2012).

Assumption 4 (Schedule-based model) *Flows on platform exit ways consist of independent ‘arrival flows’ induced by trains served by the corresponding platform. These ‘train-induced arrival flows’ follow a piecewise linear model, characterized by a lagged onset of flow after the train arrival, and a constant flow thereafter until all passengers have left the platform.*

Assumption 4 allows to empirically predict the platform exit flows by decomposing them into the independent contributions of each train. Assume that for a train ζ , the arrival flow rate at continuous time t on link $\lambda \in \mathcal{L}_{\mathcal{P}}^{\text{arr}}$ is given by $\tilde{\Phi}_{\zeta,\lambda}(t; e_\zeta^{\text{off}}, t_\zeta^{\text{arr}}, \mathbf{y})$. The platform exit flow during time interval τ on link $\lambda \in \mathcal{L}_\pi^{\text{arr}}$ is then given by

$$\Phi_{\lambda,\tau} = \int_{t \in \tau} \sum_{\zeta \in \mathcal{Z}_\pi} \tilde{\Phi}_{\zeta,\lambda}(t; \hat{e}_\zeta^{\text{off}}, t_\zeta^{\text{arr}}, \mathbf{y}) dt. \quad (10)$$

Let $\tilde{f}_\zeta = \tilde{f}_\pi(e^{\text{off}})$ denote the total exit flow rate associated with platform π if train $\zeta \in \mathcal{Z}_\pi$ with alighting volume e_ζ^{off} has arrived, and let $\Delta t_{\zeta,\lambda}^{\text{lag}}$ be the lag time representing the delay between the arrival of train ζ and the onset of flow on link $\lambda \in \mathcal{L}_\pi^{\text{arr}}$. This lag time takes into account the walking time, a potential lag between the arrival of a train and the time at which train doors open, and random delays. It may be modeled as a normally

distributed random variable, and assumed to depend on the platform only, i.e., $\Delta t_{\zeta,\lambda}^{\text{lag}} = \Delta t_{\pi}^{\text{lag}}$, where $\pi = \pi_{\zeta}$ (Molyneux et al., 2014).

Assuming that the total exit flow rate of platform π is shared according to platform sector split fractions $r_{\zeta,\lambda}^{\text{sec}}$ with $\sum_{\lambda \in \mathcal{L}_{\pi}^{\text{arr}}} r_{\zeta,\lambda}^{\text{sec}} = 1$, the flow rate on link λ associated with train $\zeta \in \mathcal{Z}_{\pi}$ is given by

$$\tilde{\phi}_{\zeta,\lambda}(t) = \begin{cases} r_{\zeta,\lambda}^{\text{sec}} \tilde{f}_{\zeta} & t \in \left(t_{\zeta}^{\text{arr}} + \Delta t_{\zeta,\lambda}^{\text{lag}}, t_{\zeta}^{\text{arr}} + \Delta t_{\zeta,\lambda}^{\text{lag}} + e_{\zeta}^{\text{off}} / \tilde{f}_{\zeta} \right), \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Fig. 2 illustrates the cumulative arrival flow associated with Eq. (11). The solid curve illustrates an observation from Lausanne railway station (Molyneux et al., 2014), and the dash-dotted curve a piecewise linear approximation.

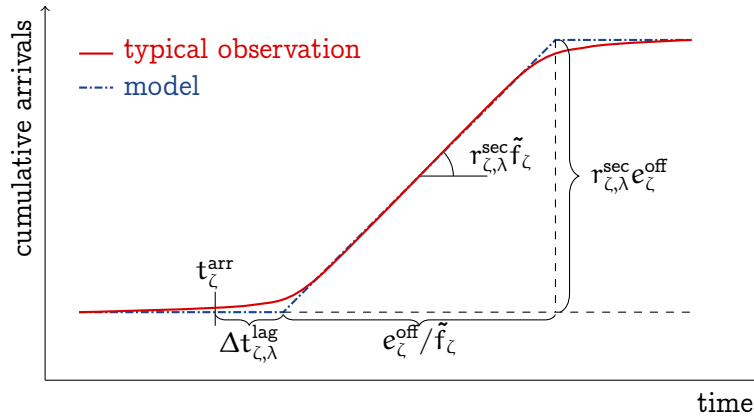


Figure 2: Model for train-induced arrival flows.

The total platform exit flow rate $\tilde{f}_{\pi}(e^{\text{off}})$ is assumed to depend linearly on the alighting volume e^{off} at low values, and to reach saturation at a platform-specific threshold e_{π}^{crit} . If q_{π} , b_{π} and c_{π} represent shape parameters, the total exit flow rate on platform π is given by the stochastic model

$$\tilde{f}_{\pi}(e^{\text{off}}) = \tilde{f}_{\pi}^{\text{det}}(e^{\text{off}}) + \mathcal{N}(0, q_{\pi}), \quad (12)$$

where the deterministic part of the flow rate is specified as

$$\tilde{f}_{\pi}^{\text{det}}(e^{\text{off}}) = \begin{cases} b_{\pi} e^{\text{off}} + c_{\pi} & \text{if } e^{\text{off}} \leq e_{\pi}^{\text{crit}}, \\ b_{\pi} e_{\pi}^{\text{crit}} + c_{\pi} & \text{otherwise.} \end{cases} \quad (13)$$

Eq. (11) and Eq. (12) may be specified based on studies by Weidmann (1992), Buchmüller and Weidmann (2008) and Molyneaux et al. (2014). Alternatively, they can be calibrated on actual data. An example of the latter is provided in Section 4.

3.4 Measurement model

The measurement model links the structural model to a priori information and measurements, which are useful for the estimation of demand, and for validating the obtained results.

For each data source, a random error term takes into account the uncertainty it is afflicted with, and the aforementioned reduction matrices account for the incomplete coverage of the data collection infrastructure, i.e.,

$$\hat{d}' = \Delta_d d + \omega'_d, \quad (14)$$

$$\hat{f}' = \Delta_f f + \omega'_f, \quad (15)$$

$$\hat{a}' = \Delta_a a + \omega'_a, \quad (16)$$

$$\hat{e}'_{\text{on}} = \Delta_{\text{on}} e_{\text{on}} + \omega'_{\text{on}}, \quad (17)$$

$$\hat{e}'_{\text{off}} = \Delta_{\text{off}} e_{\text{off}} + \omega'_{\text{off}}. \quad (18)$$

The measurement errors $\omega_{(\cdot)}$ are generally correlated, both across time and space. Temporal correlation occurs if a sensor is malfunctioning, or if it reaches saturation. Spatial correlation is a concern if two sensors capture similar information, for instance if they are placed nearby. For an efficient statistical inference, these effects need to be taken into account by using an appropriate covariance matrix.

Example specification: For the illustration of the model, the estimation problem is reduced to a formulation that is linear in the unknown demand vector.

The cumulative origin and destination flows as well as the destination

split fractions are obtained by aggregation, i.e.,

$$f_v^{\text{out,cum}} = \sum_{\tau \in \mathcal{T}} \sum_{\kappa \in \mathcal{K}_v^{\text{orig}}} d_{\kappa,\tau} \Delta t, \quad (19)$$

$$f_v^{\text{in,cum}} = \sum_{\tau \in \mathcal{T}} \sum_{\kappa \in \mathcal{K}_v^{\text{dest}}} d_{\kappa,\tau} \Delta t, \quad (20)$$

$$r_v^{\text{avg}} = \sum_{\tau \in \mathcal{T}} \sum_{\kappa \in \mathcal{K}_v^{\text{orig}}, \nu_k^d \in \mathcal{C}_p} d_{\kappa,\tau} \Delta t / f_v^{\text{out,cum}}, \quad (21)$$

where ν_k^d is the destination of OD pair κ , and where $\mathcal{K}_v^{\text{orig}}$ and $\mathcal{K}_v^{\text{dest}}$ denote the set of OD pairs which originate and terminate in centroid v , respectively.

Assumption 5 (Measurement model) *The distribution of train exchange volumes is a priori known, and used to pre-compute cumulative platform departure as well as platform arrival flows.*

The measurement model is given by Eq. (15), as well as by

$$\hat{f}'_{\text{out,cum}} = \Delta_{\text{out}} f_{\text{out,cum}} + \omega'_{\text{out}}, \quad (22)$$

$$\hat{f}'_{\text{in,cum}} = \Delta_{\text{in}} f_{\text{in,cum}} + \omega'_{\text{in}}, \quad (23)$$

$$\hat{r}'_{\text{avg}} = \Delta_r r_{\text{avg}} + \omega'_r. \quad (24)$$

$$\hat{\varphi} = \Delta_{\text{arr}} f + \omega_{\varphi}, \quad (25)$$

$$\hat{\chi} = f_{\text{dep,cum}} + \omega_{\chi}. \quad (26)$$

Eq. (25) and Eq. (26) consider empirical estimates of platform arrival and departure flows, $\hat{\varphi}$ and $\hat{\chi}$, which are pre-computed based on the train timetable and a priori known train exchange volumes. This pre-processing is useful to simplify the solution of the example specification, but not a necessity of the framework.

3.5 Estimation problem

The estimation problem consists in finding the distribution of the OD demand volumes d^* such that (i) actual observations of demand indicators are reproduced at best, (ii) platform arrival and platform departure flows are ‘most consistent’ with empirical predictions based on the train timetable,

and (iii) the resulting estimate matches the historical one in case the estimation problem is underdetermined.

In the most general case, these three objectives are captured by a joint distance measure $\text{dist}\langle\cdot\rangle$. A statistically meaningful specification can be found using pure likelihood methods, or within the Bayesian framework, and depends on the assumptions that are made regarding the distribution of the error terms (Hazelton, 2000).

Alternatively, if the cross-correlation across the three objectives is negligible, the joint distance measure can be replaced by three separate terms $\text{dist}_{\text{obs}}\langle\cdot\rangle$, $\text{dist}_{\text{sched}}\langle\cdot\rangle$ and $\text{dist}_{\text{hist}}\langle\cdot\rangle$. The estimation problem reads then as

$$\mathbf{d}_y^* = \arg \min_{\mathbf{d} \geq 0} \text{dist}_{\text{obs}} \left\langle \left(\begin{array}{c} \hat{e}'_{\text{on}} \\ \hat{e}'_{\text{off}} \\ \hat{f}' \\ \hat{a}' \end{array} \right), \left(\begin{array}{c} e'_{\text{on}} \\ e'_{\text{off}} \\ f' \\ a' \end{array} \right) \right\rangle + \text{dist}_{\text{sched}} \left\langle \left(\begin{array}{c} \varphi' \\ \chi' \end{array} \right), \left(\begin{array}{c} f'_{\text{arr}} \\ f'_{\text{dep,cum}} \end{array} \right) \right\rangle + \text{dist}_{\text{hist}} \langle \hat{\mathbf{d}}, \mathbf{d}' \rangle. \quad (27)$$

While the distance measures in Eq. (27) are mutually independent, internally they may still consider complex error structures that, for instance in the context of least squares, can be taken into account by inner weights.

When solving Eq. (27), it is critical not to rely on point estimates. The demand vector \mathbf{d}^* is generally distributed, and follows a complex distribution that is insufficiently described by a single value such as its mean. The distribution depends both on the variation of input variables, which can be distributed themselves, and on the uncertainty involved in terms of modeling and measurement errors.

To approximate the distribution of \mathbf{d}^* , Monte Carlo sampling may be used. Demand on each day is assumed to represent independent random variables that follow a joint distribution. This is valid as long as seasonal effects are absent and no significant one-off events affect the network.

Example specification: As often done in practice, the correlation between error terms is neglected (Cascetta and Improta, 2002).

Assumption 6 (Error terms) *Each error term $\omega_{(\cdot)}$ follows an independent, univariate normal distribution with zero mean.*

Based on Assumptions 1–6, Eq. (27) reduces to a constrained, generalized least squares (GLS) problem both in the context of maximum likeli-

hood and Bayesian estimation (Cascetta et al., 1993). It consists in finding

$$\begin{aligned} \mathbf{d}_y^* = \arg \min_{d \geq 0} & w_{\text{flow}} \|\hat{\mathbf{f}}' - \mathbf{f}'\|_2^2 + \\ & w_{\text{out}} \|\hat{\mathbf{f}}'_{\text{out,cum}} - \mathbf{f}'_{\text{out,cum}}\|_2^2 + w_{\text{in}} \|\hat{\mathbf{f}}'_{\text{in,cum}} - \mathbf{f}'_{\text{in,cum}}\|_2^2 + w_{\text{ratio}} \|\hat{\mathbf{r}}'_{\text{avg}} - \mathbf{r}'_{\text{avg}}\|_2^2 + \\ & w_{\text{arr}} \|\hat{\boldsymbol{\varphi}} - \mathbf{f}_{\text{arr}}\|_2^2 + w_{\text{dep}} \|\hat{\boldsymbol{\chi}} - \mathbf{f}_{\text{dep,cum}}\|_2^2, \end{aligned} \quad (28)$$

where the parameters $w_{(\cdot)}$ denote weights whose specification is discussed below.

The first term on the RHS of Eq. (28) represents the distance between the observed link flows and those predicted by the model (see Eq. 1, 15). The terms on the second line consider the distance between model prediction and survey data in terms of cumulative origin and destination flows (Eq. 19, 22 and Eq. 20, 23), as well as in terms of destination split ratios (Eq. 21, 24). The two terms on the last line consider the distance to the pre-computed train-induced arrival flows (Eq. 8, 10, 11, 25) and the cumulative platform departure flows (Eq. 5, 6, 26).

For optimal statistical efficiency, the weights $w_{(\cdot)}$ are assumed equal to the reciprocal of the variance of the corresponding error term, i.e., $w_{\text{flow}} = 1/\text{Var}(\boldsymbol{\eta}'_f + \boldsymbol{\omega}'_f)$, $w_{\text{out}} = 1/\text{Var}(\boldsymbol{\omega}'_{\text{out}})$, $w_{\text{in}} = 1/\text{Var}(\boldsymbol{\omega}'_{\text{in}})$, $w_{\text{ratio}} = 1/\text{Var}(\boldsymbol{\omega}'_r)$, $w_{\text{arr}} = 1/\text{Var}(\boldsymbol{\varepsilon}'_\varphi)$ and $w_{\text{dep}} = 1/\text{Var}(\boldsymbol{\varepsilon}'_\chi)$. In practice, these variances are unknown, and need to be estimated (Cascetta and Improta, 2002).

In this work, an active set method (Lawson and Hanson, 1974; Bierlaire et al., 1991) is used to solve the KKT conditions for the resulting non-negative least squares problem, Eq. (28). If several optimal solutions exist, the one with the lowest norm is selected, yielding a solution with maximum entropy (Cascetta et al., 1993).

4 Case study

To demonstrate the applicability of the proposed framework, a case study of Lausanne railway station is carried out. All code developed for the implementation, including the assignment model, is freely available (Hänseler et al., 2015b).

4.1 Description

Lausanne railway station is the largest train station in French-speaking Switzerland, serving approximately 120'000 passengers with about 650 ar-

riving and departing trains every weekday (Amacker, 2012). Fig. 3 shows a schematic map of the station. It encompasses ten tracks (#1–#9, #70) that are served by platforms #1, #3/4, #5/6, #7/8, #9 and #70. Platforms and entrance/exit areas, as well as four service points are represented by centroids. These are connected by a network of walking facilities, which at its heart has two pedestrian underpasses (PU), referred to as PU West and PU East. Each of them is partially covered by a pedestrian tracking system (corresponding areas are shaded). Dashed lines represent network links that cannot be represented in the scheme. Pedestrian counters are represented by diamonds. Stars denote cordons at which platform exit flows are empirically estimated. All links and flow sensors are directed.

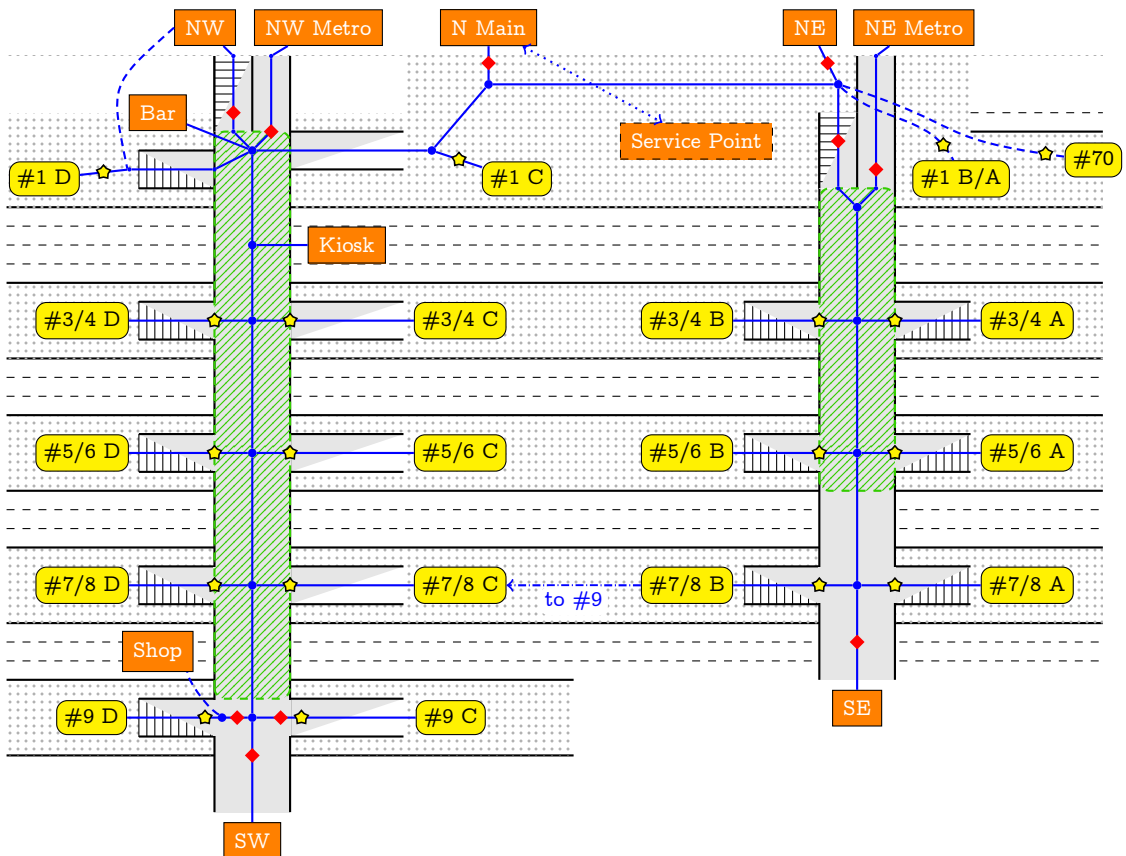


Figure 3: Schematic map of Lausanne railway station.

The demand peak over a weekday is reached at around 07:45 (AM) when

several long distance trains arrive and depart in close succession (Gendre and Zulauf, 2010). In the ensuing analysis, we consider the time period between 07:30 and 08:00 with a temporal aggregation of one minute. Data for a set of 10 ‘reference weekdays’ is available, namely for January 22 and 23, February 6, 27 and 28, March 5, as well as April 9, 10, 18 and 30, 2013. These dates represent a set of typical weekdays (Tue, Wed, Thu) without major disruptions in the railway system, for which the following data sources are available.

OD flow data: Subroute flows are available for the two pedestrian underpasses, in which an elaborate pedestrian tracking system consisting of more than 60 tracking sensors has been installed. Details of the installation, as well as of the accuracy of observations, are described by Alahi et al. (2013). OD flow data is used for validation only.

Link flow data: Ten links of the pedestrian walking network, marked by diamonds in Fig. 3, are equipped with sensors that provide directed link counts with a resolution of one minute. To account for sensor saturation, observations are post-processed using a quadratic correction function (Ganansia et al., 2014)

Traffic condition data: Pedestrian trajectories obtained from the aforementioned tracking system allow to compute the accumulation in pedestrian underpasses, which is also used for validation.

Train timetable and ridership data: During the time period of interest, a total of 25 trains stop at Lausanne railway station (see Hänseler et al., 2015a, for the train timetable). The actual arrival and departure time and the assigned track are known for each train and day. An average estimate of boarding and alighting volumes is available from ticket sales data, within-train surveys, and infrared-based counts at train doors (SBB-Personenverkehr, 2011). These estimates date back to the year 2010 and are increased by 15% using the official growth rate (Gendre and Zulauf, 2010). They are considered as random normal variables with a standard deviation equal to 19.2% of their mean (Molyneaux et al., 2014).

Other data: For the sales points located in PU West (see Fig. 3), an estimate of the number of customer visits is available. There are further restaurants and sales points in the train station building, represented by a generic ‘service point’, for which however no information is available, and which are not considered in the demand analysis. Besides sales data, information on destination split ratios is available (Benmoussa et al., 2011; Anken et al., 2012; Lavadinho et al., 2013).

4.2 Assessment

To assess the efficiency of the proposed framework, two estimators are compared. A ‘base estimator’, representing a minimum norm solver taking into account link flow data only, and a ‘full estimator’, that additionally considers a ‘static’ and a ‘dynamic’ prior (see Fig. 4).

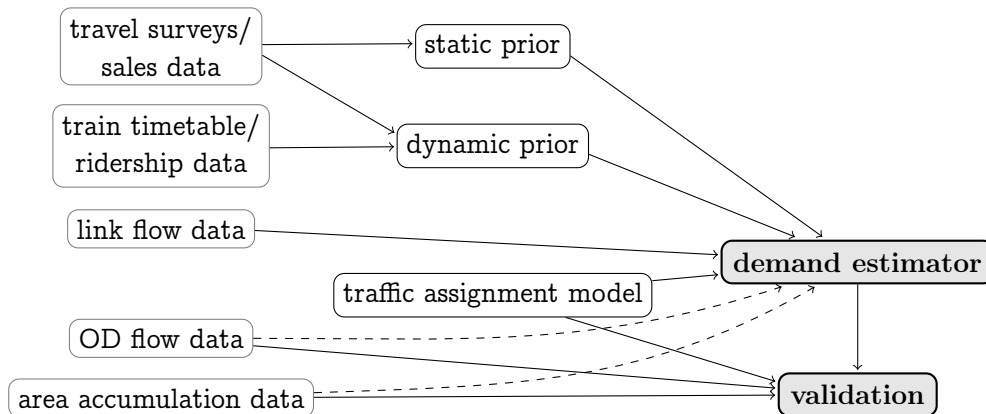


Figure 4: Scheme of the specification of the demand estimation framework.

The static prior includes cumulative origin and destination flows obtained from sales data and platform departure flows, as well as destination split fractions. The dynamic prior represents pre-computed train-induced arrival flows. OD flow data and traffic condition data are used for validation. In a real context, once the specification is successfully validated, these two data sources would also be integrated in the estimation process to improve the quality of the estimate (dashed arrows in Fig. 4).

4.3 Parametrization

The pedestrian walking network of Lausanne railway station disposes of a unique shortest path between every OD pair. During peak periods, regular commuters constitute the largest user group, which are familiar with the facilities (Lavadinho, 2012; Ton, 2014). Following Lavadinho (2012), they almost exclusively travel along these shortest paths, obviating the need for a route choice model.

To describe the propagation of pedestrians along routes, the walking speed distribution proposed by Weidmann (1992),

$$v \sim \mathcal{N}(1.34 \text{ m/s}, 0.34 \text{ m/s}), \quad (29)$$

is used. It holds for even walking areas; link lengths on inclined areas or stairways need to be adjusted beforehand (Weidmann, 1992). The validity of speed distribution (29) has been empirically verified based on the available trajectory recordings, which show no significant signs of demand-supply interaction (Hänseler et al., 2015a).

The schedule-based model is parametrized empirically (Molyneaux et al., 2014). Fig. 5 shows the total exit flow rates observed for platform #3/4, as well as the corresponding model fit. At low volumes, the flow rate increases linearly until a threshold is reached, beyond which the flow rate remains constant. The solid curve denotes the predicted flow rate according to Eq. (12), and the dashed lines the width of the prediction band in terms of \pm one standard deviation.

Two observations may be made. First, the length of a train, measured in number of passenger cars, n_{car} , does not have a significant influence on the flow rate. This is explicitly pointed out since the train length is shown below to have a considerable influence on the platform sector split fractions $r_{\zeta,\lambda}^{\text{sec}}$. Second, the flow rates are relatively high, such that the duration of flow is typically below one minute (up to an alighting volume of 333 passengers), and never exceeds 2 min.

Based on this specification, Fig. 6 shows the predicted exit flow for platform #5/6 on April 10, 2013, as well as the corresponding observation. The prediction band results from 7500 Monte Carlo samplings of Eq. (6). The alighting volumes e_{ζ}^{off} of each train ζ are inferred from the historical ridership data mentioned in Section 4.1. A logarithmic probability density plot shows the expected cumulative arrivals as well as the arrival rate as a

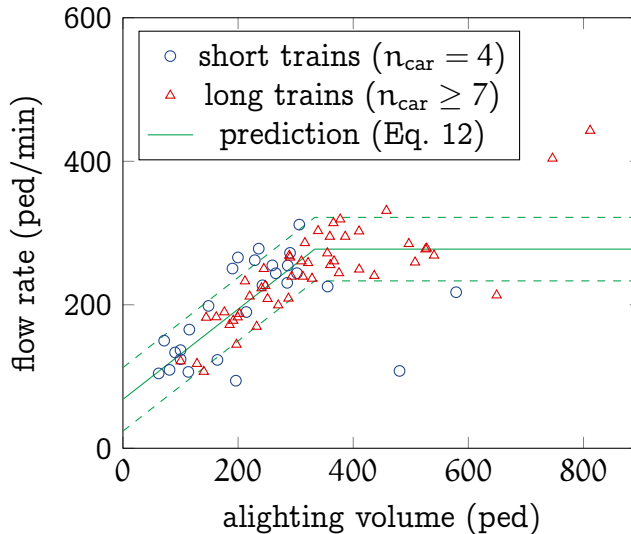


Figure 5: Total platform exit flow rate \hat{f} on platform #3/4.

function of time. A good agreement between observation and prediction is found, although the prediction band is relatively wide. This indicates that the variation in alighting volumes across days is high.

The split fractions $r_{\zeta,\lambda}^{\text{sec}}$ depend on various factors such as the length of a train, its position along a platform, the distribution of passengers within a train, as well as their immediate next destination. Fig. 7 shows measurements from platform #3/4. The results are grouped by train length and ordered by alighting volumes. For short trains with $n_{\text{car}} = 4$, mostly the interior platform sectors B and C are used. This is true particularly if the alighting volume is low. For larger trains with $n_{\text{car}} \geq 7$, the lateral sectors absorb a larger share, and the influence of the alighting volume is smaller.

In the framework of this study, two different specifications of the platform sector split fractions for short trains ($n_{\text{car}} = 4$) and long trains ($n_{\text{car}} \geq 7$) are considered. For each case, a multivariate normal distribution is developed, from which the train- and link-specific platform sector split fractions $r_{\zeta,\lambda}^{\text{sec}}$ can be drawn (Molyneaux et al., 2014).

The weights of demand indicators in Eq. (28) are determined based on the premise that pedestrian trajectory recordings represent the truth. Given the accuracy of the trajectory recordings, and their high spatial and temporal resolution compared to the other data sources, this assumption

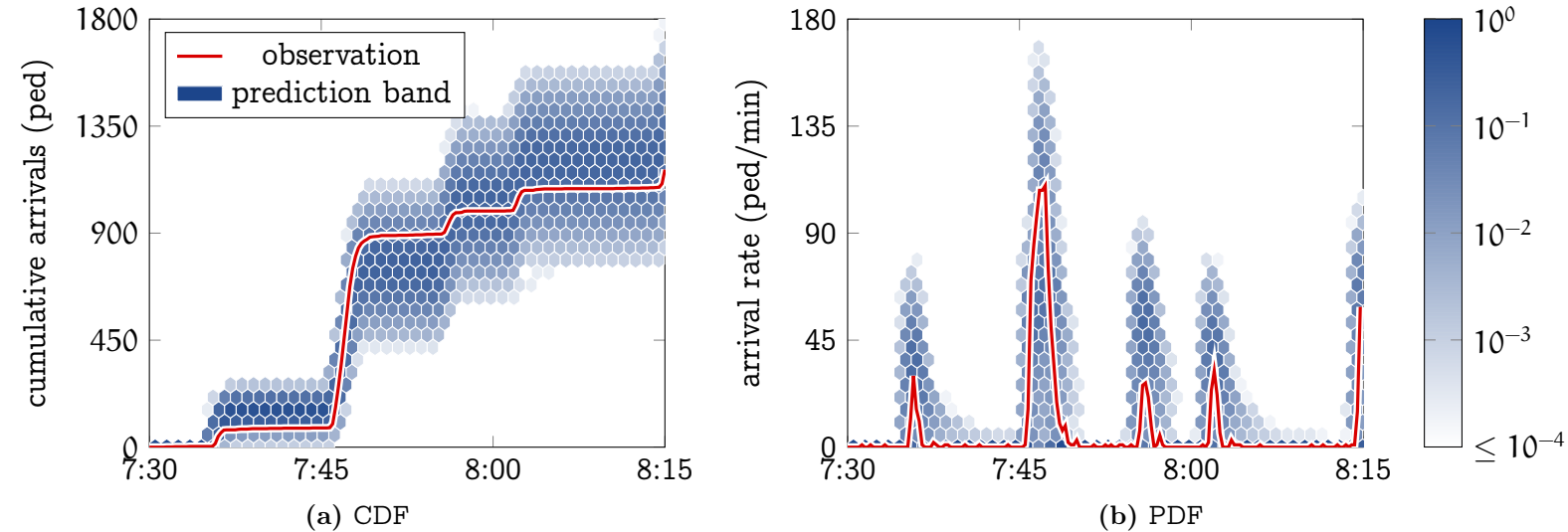


Figure 6: Train-induced passenger arrival flow for April 10, 2013.

seems justifiable. It allows to estimate the variance of the errors associated with the link flow data and the platform exit flows inferred from the train timetable. If, without loss of generality, the weight associated with link flow data is set to one, $w_{\text{flow}} = 1$, a value of $w_{\text{dyn}} = w_{\text{arr}} = 0.69$ results for the weight of the dynamic prior.

Regarding the weight of the static prior, $w_{\text{stat}} = w_{\{\text{out,in,ratio,dep}\}}$, only little variation in the resulting demand estimate is perceivable in the range $10^{-4} \leq w_{\text{stat}} \leq 10^{-1}$. For lower values, due to numerical errors, its influence on the model estimate vanishes completely; for values larger than 10^{-1} , its influence grows rapidly. Given the relative inaccuracy of the data sources it contains, the static prior mainly serves to narrow the solution space. Thus, a value of $w_{\text{stat}} = 10^{-1}$ is used.

The resulting size of the estimation problem is given by the number of considered OD pairs, and the number of time intervals. To account for artificial transients during a potential ‘heat-up’ of the estimation, the computations include an additional 7 minutes both at the beginning and the end of the 30-minute analysis period. Therefore, an estimation problem with a total of 16’280 unknowns has to be solved per day. For each day, 24 Monte Carlo samplings of Eq. (28) are conducted, which is sufficient to generate reproducible and numerically stable results.

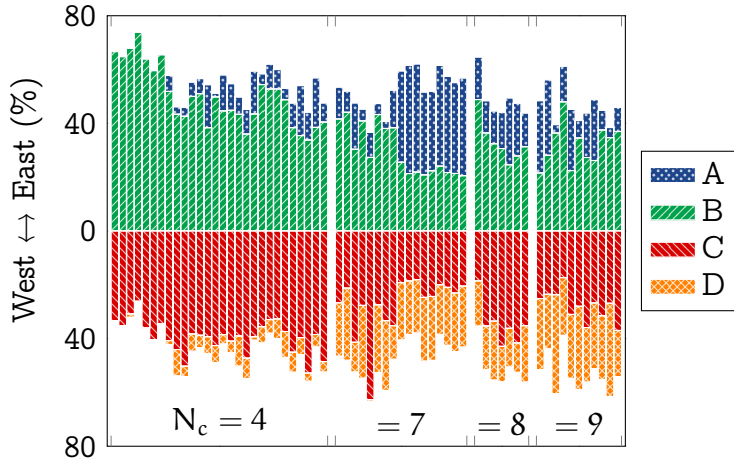
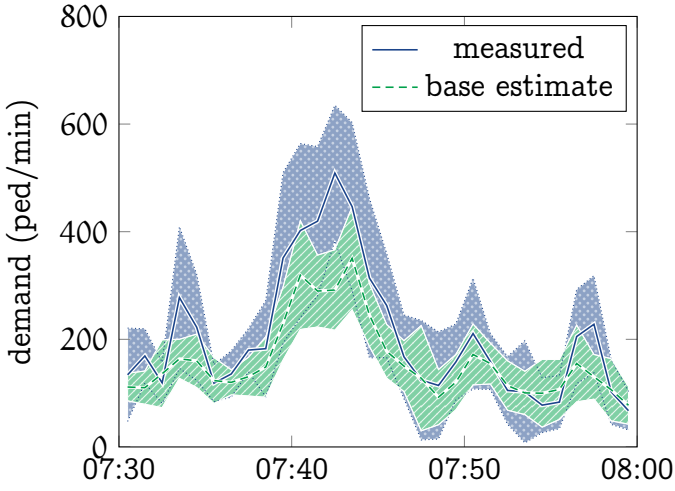


Figure 7: Split fractions of arrival flows across sectors on platform #3/4 grouped by train size and ordered by alighting volumes (increasing from left to right).

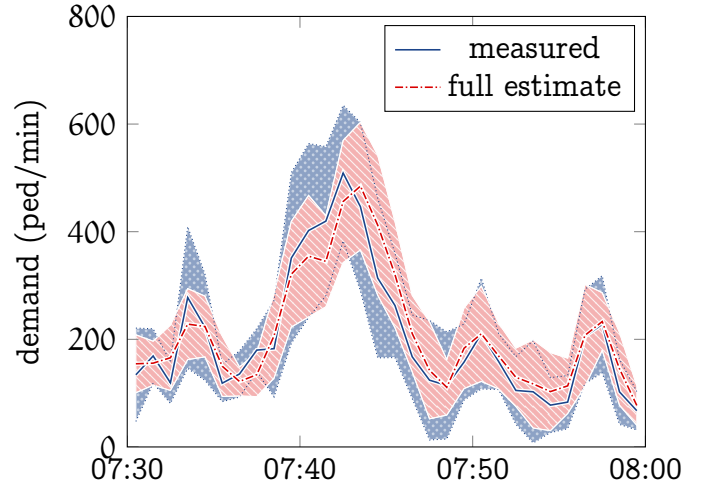
4.4 Results

The temporal evolution of total demand in PU East and PU West is shown in Fig. 8, in which the base and full estimator are compared to actual observations. Both the expected values, and the prediction bands are shown. In the considered period, the demand fluctuates between less than 70 and more than 500 ped/min, i.e., by almost an order of magnitude. Both estimators are capable of following the overall trend. The base estimator, however, tends to underestimate the peaks, and underestimates the cumulative demand by more than 20%. The full estimate mostly represents an accurate guess of the peak amplitudes, and yields an error of less than 4% for the overall demand. It performs between 40.8% (MAE) and 46.7% (RMSE) better than the base estimator. A similar finding results by investigating the accumulation in PU West and PU East, for which the improvements for MAE and RMSE amount to 49.9% and 40.7%, respectively (no figure shown). The lowest average accumulation of 56.6 pedestrians is reached between 7:59 and 8:00, and the maximum of 261.5 pedestrians between 7:43 and 7:44.

The ability of the two estimators to reproduce platform exit flows can be assessed in Fig. 9, showing a scatter plot of observed and estimated flows for platforms #3/4 and #5/6. The accuracy of these estimates is important for the dimensioning of platform access ways. The total platform exit flow



(a) Base estimate (MAE = 50.74, RMSE = 70.47)



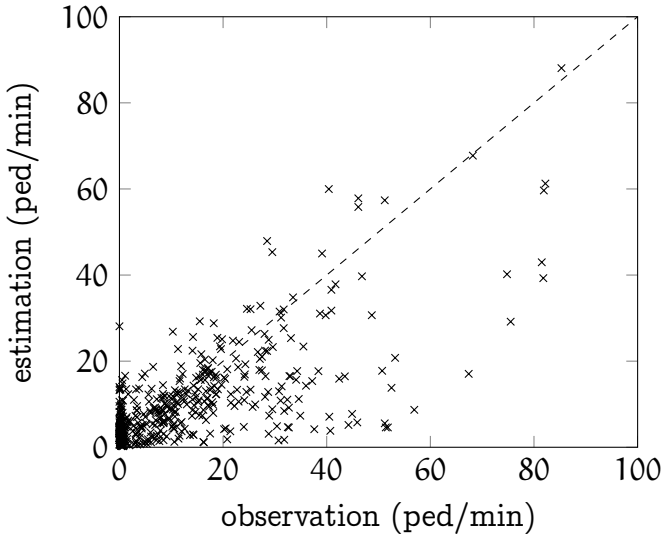
(b) Full estimate (MAE = 30.03, RMSE = 37.56)

Figure 8: Total demand in PU East and PU West. Data: 10-day reference set, 2013.

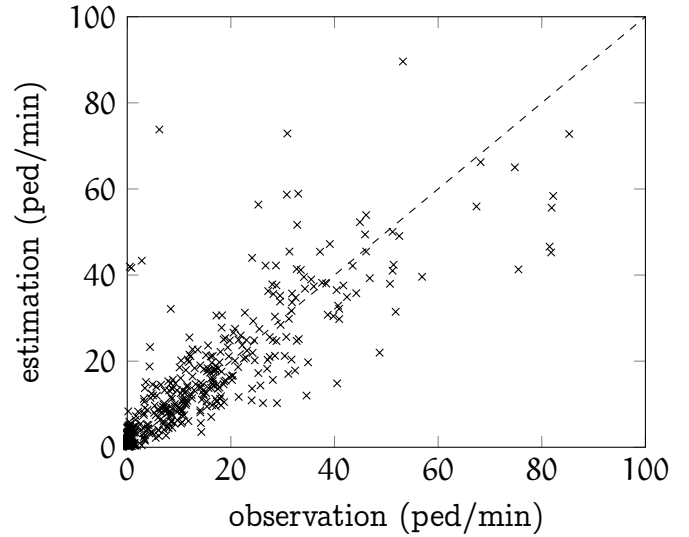
is underestimated by the base model by -18.33%, and overestimated by the full model by 6.98%. The improvement in MAE and RMSE amounts to 30.26% and 23.35%, respectively.

Table 3 provides the RMSE for subroute flows and area accumulations in PU East and PU West for different estimators. Compared to the base model, in particular the incorporation of the dynamic prior leads to a significant improvement. In fact, the consideration of the train timetable increases the prediction quality more than sales data, information of destination split ratios and cumulative platform departure flows together. The full model globally performs best, even though the accumulation estimate is slightly worse than in the case with a dynamic, but no static prior. Similar findings result if instead of RMSE another statistical measure, such as MAE, is used.

For further details of this case study, see Hänseler et al. (2015a).



(a) Base estimate (MAE = 7.24, RMSE = 11.61)



(b) Full estimate (MAE = 5.05, RMSE = 8.90)

Figure 9: Arrival flows at platforms #3/4 and #5/6. Data: 10-day reference set, 2013.

Table 3: RMSE for subroute flows and accumulation in PU East and PU West. Data: 10-day reference set, 2013.

	subroute flow	accumulation
Base estimate	3.52 ped/min	58.84 ped
Estimate with static prior (STAT)	+2.43%	-15.03%
Estimate with dynamic prior (DYN)	-15.59%	-41.76%
Full estimate (STAT + DYN)	-31.07%	-40.74%

5 Conclusions

A framework for the time-dependent estimation of pedestrian origin-destination demand within a train station has been presented. Besides direct and indirect demand indicators such as flow counts or sales data, the train timetable is explicitly taken into account. This is achieved by establishing an empirical relation between the arrival of a train and the subsequent flow of alighting passengers on platform exit ways. The formulation of the framework is such that it can be applied to various types of railway stations and may be used with different data sources.

A case study of the morning peak period in Lausanne railway station has been presented. The obtained results are in good agreement with pedestrian tracking data that has been used for validation. A significant performance gain has been shown to exist when the train timetable is used in the estimation process. Moreover, spatial and temporal fluctuations, both intra- and inter-day, have been investigated and are shown to be important, justifying the use of a fully dynamic and probabilistic framework.

We can think of mainly two ways to extend the proposed framework. An obvious way relates to its application to real-time applications, such as traffic monitoring or crowd control. Another way is to focus on the improvement of the presented model specification. Clearly, the empirical relation between the train timetable and pedestrian movements can be strengthened, or a demand-dependent network loading model could be integrated. Arguably the most pressing issue, however, is the explicit consideration of correlation among measurements, which could significantly improve the statistical inference.

Acknowledgement

Financial support by SNSF grant #200021-141099 ‘Pedestrian dynamics: flows and behavior’ and by SBB-CFF-FFS in the framework of ‘PedFlux’, as well as support by Nicolas Anken, Quentin Mazars-Simon, Bonnie Qian, Eduard Rojas and Amanda Stathopoulos in implementing the case study of Lausanne railway station is gratefully acknowledged.

A List of variables (cf. Table 4)

Table 4: List of recurrent variables.

$\tau \in \mathcal{T}, \Delta t$	time
$\mathcal{G} = (\mathcal{N}, \mathcal{L}), \mathcal{C} \subset \mathcal{N}$	graph, centroids
$v \in \mathcal{N}, \lambda \in \mathcal{L}, \alpha \in \mathcal{A}$	node, link, area
$\rho = (\lambda_1, \lambda_2, \dots), \rho \in \mathcal{R}$	route
$\kappa \in \mathcal{K}$	OD pair
$\zeta \in \mathcal{Z}, t_{\zeta}^{\text{arr}}, t_{\zeta}^{\text{dep}}$	train, arrival and departure time
$\pi \in \mathcal{P}$	platform
$\mathbf{d} = [d_{\kappa, \tau}]$	demand
$\mathbf{f} = [f_{\lambda, \tau}]$	flow
$\mathbf{a} = [a_{\alpha, \tau}]$	accumulation
$\mathbf{e}_{\text{on}} = [e_{\zeta}^{\text{on}}], \mathbf{e}_{\text{off}} = [e_{\zeta}^{\text{off}}]$	exchange volumes
$\mathbf{r} = [r_{v, \tau}]$	destination split ratio
Δ	reduction matrix
$\Sigma(\mathbf{d}; \mathbf{y})$	assignment model
\mathbf{y}	parameter vector
$\chi = [\chi_p], \varphi = [\phi_{\lambda, \tau}]$	schedule-based estimates
$\eta, \varepsilon, \omega$	errors
$r_{\zeta, \lambda}^{\text{sec}}$	platform sector split ratio
w	estimation weight

B Assignment model for walking facilities

This section outlines an assignment model for walking facilities in an uncongested train station. In accordance with Assumptions 2 and 3 in Section 3.3.1, the prevailing traffic conditions are demand-independent.

Route choice: The outcome of the route choice model is represented by a route choice matrix $\mathbf{R}(\mathbf{y}) = [r_{(\rho, \tau'), (\kappa, \tau)}]$ of size $|\mathcal{R}| |\mathcal{T}| \times |\mathcal{Z}| |\mathcal{T}|$. An element $r_{(\rho, \tau'), (\kappa, \tau)}(\mathbf{y})$ denotes the probability that a pedestrian associated with OD pair κ and departure time interval τ chooses route ρ during time interval τ' . Route choice is instantaneous such that $r_{(\rho, \tau'), (\kappa, \tau)} = 0$ if $\tau \neq \tau'$.

The time to traverse link λ during time interval τ is denoted by $\Delta t_{\lambda,\tau}^{\text{trav}}(\mathbf{y})$. The travel time on route ρ during time interval τ is given by

$$U_{\rho,\tau}(\mathbf{y}) = V_{\rho,\tau} + \psi, \quad (30)$$

where $\psi \sim \text{EV}(0, \vartheta)$ with ϑ a calibration parameter contained in \mathbf{y} , and where the sum of link travel times is given by

$$V_{\rho,\tau}(\mathbf{y}) = \sum_{\lambda \in \rho} \Delta t_{\lambda,\tau}^{\text{trav}}. \quad (31)$$

For OD pair κ , the likelihood that a user chooses route $\rho \in \mathcal{R}_\kappa$ is then given by

$$r_{(\rho,\tau),(\kappa,\tau)}(\mathbf{y}) = \frac{\exp(-\vartheta V_{\rho,\tau})}{\sum_{\rho' \in \mathcal{R}_\kappa} \exp(-\vartheta V_{\rho',\tau})}. \quad (32)$$

Network loading: The network loading model defines mappings from route flows to link flows and area accumulations. Table 5 defines the corresponding assignment matrices.

Table 5: List of considered network loading maps.

$\mathbf{B} = [\mathbf{b}_{(\lambda,\tau'),(\rho,\tau)}]$	The link flow assignment matrix $\mathbf{B}(\mathbf{y})$ is of size $ \mathcal{L} \mathcal{T} \times \mathcal{R} \mathcal{T} $. The entry $\mathbf{b}_{(\lambda,\tau'),(\rho,\tau)}(\mathbf{y})$ represents the probability that a pedestrian associated with route ρ and departure time interval τ reaches link λ during time interval τ' .
$\mathbf{C} = [\mathbf{c}_{(\alpha,\tau'),(\rho,\tau)}]$	The area accumulation assignment matrix $\mathbf{C}(\mathbf{y})$ is of size $ \mathcal{A} \mathcal{T} \times \mathcal{R} \mathcal{T} $. The entry $\mathbf{c}_{(\alpha,\tau'),(\rho,\tau)}(\mathbf{y})$ denotes the expected contribution of a pedestrian associated with route ρ and departure time interval τ to the accumulation of area α during time interval τ' .

Based on these definitions, we may write

$$\Sigma_f(\mathbf{d}; \mathbf{y}) = \mathbf{B}(\mathbf{y}) \mathbf{R}(\mathbf{y}) \mathbf{d}, \quad (33)$$

and

$$\Sigma_a(\mathbf{d}; \mathbf{y}) = \mathbf{C}(\mathbf{y}) \mathbf{R}(\mathbf{y}) \mathbf{d}, \quad (34)$$

respectively.

Let the distance along a route ρ up to the beginning of link λ be denoted by ℓ_ρ^λ . Furthermore, let the departure times of pedestrians within a

time interval be distributed uniformly, i.e., the distribution of continuous departure time t for any route during a time interval τ is given by

$$h_\tau(t) = \begin{cases} \frac{1}{\Delta t} & \text{if } t \in \tau, \\ 0 & \text{otherwise.} \end{cases} \quad (35)$$

Assuming that each pedestrian is walking at a constant speed, the probability for a person on route ρ that departs during time interval τ to arrive on link λ during time interval τ' is given by

$$\begin{aligned} b_{(\lambda, \tau'), (\rho, \tau)} &= \Pr(t \in \tau, t' \in \tau' | \rho, \lambda) \\ &= \Pr\left(t \in \tau, v \in \left[\frac{\ell_\rho^\lambda}{t_{\tau'}^+ - t}, \frac{\ell_\rho^\lambda}{t_{\tau'}^- - t}\right]\right), \end{aligned} \quad (36)$$

where $t_{\tau'}^-$ and $t_{\tau'}^+$ represent the bounds of time interval τ' , and where t and t' represent the continuous departure and arrival time, respectively. For the most common case that $\ell_\rho^\lambda > 0$ and $\tau' > \tau$, we obtain

$$\begin{aligned} b_{(\lambda, \tau'), (\rho, \tau)} &= \int_{t=t_{\tau'}^-}^{t_{\tau'}^+} \int_{v=\ell_\rho^\lambda/(t_{\tau'}^+ - t)}^{\ell_\rho^\lambda/(t_{\tau'}^- - t)} f_v(v) g_\tau(t) \, dv \, dt \\ &= \frac{1}{\Delta t} \int_{t=t_{\tau'}^-}^{t_{\tau'}^+} F_v\left(\frac{\ell_\rho^\lambda}{t_{\tau'}^- - t}\right) - F_v\left(\frac{\ell_\rho^\lambda}{t_{\tau'}^+ - t}\right) \, dt. \end{aligned} \quad (37)$$

Similarly, if $\ell_\rho^\lambda > 0$ and $\tau = \tau'$, we obtain

$$\begin{aligned} b_{(\lambda, \tau), (\rho, \tau)} &= 1 - \Pr(t \in \tau, t' \notin \tau | \rho, \lambda) \\ &= 1 - \Pr\left(t \in \tau, v \in \left[0, \frac{\ell_\rho^\lambda}{t_{\tau'}^+ - t}\right]\right) \\ &= 1 - \frac{1}{\Delta t} \int_{t=t_{\tau'}^-}^{t_{\tau'}^+} F_v\left(\frac{\ell_\rho^\lambda}{t_{\tau'}^+ - t}\right) - F_v(0) \, dt. \end{aligned} \quad (38)$$

Thus, the probability that a user associated with route ρ and departure time interval τ reaches link λ during time interval τ' is given by

$$b_{(\lambda, \tau'), (\rho, \tau)} = \begin{cases} 0 & \text{if } \ell_\rho^\lambda = 0, \tau < \tau', \\ 1 & \text{if } \ell_\rho^\lambda = 0, \tau = \tau', \\ \text{Eq. (37)} & \text{if } \ell_\rho^\lambda > 0, \tau < \tau', \\ \text{Eq. (38)} & \text{if } \ell_\rho^\lambda > 0, \tau = \tau'. \end{cases} \quad (39)$$

Remark: The probability that a pedestrian associated with route ρ and departure time interval τ reaches subroute ϱ during time interval τ' can be expressed as

$$g_{(\varrho, \tau'), (\rho, \tau)} = \begin{cases} b_{(\lambda_\varrho^o, \tau'), (\rho, \tau)} & \text{if } \varrho \in \mathcal{R}_\rho^{\text{sub}}, \\ 0 & \text{otherwise,} \end{cases} \quad (40)$$

where $\mathcal{R}_\rho^{\text{sub}}$ denotes the set of subroutes contained in route ρ , and λ_ϱ^o the first link of subroute ϱ .

The assignment fraction for area accumulations can be derived accordingly. Let us consider an area α , and let us assume that each route enters and leaves area α at most once. Let v be the constant, individual speed of a person traveling along route ρ , $\ell_{\text{in}}^{\rho, \alpha}$ the distance along the route ρ to the entrance of area α and $\ell_{\text{out}}^{\rho, \alpha}$ the corresponding distance to its exit. Consequently, $t_{\text{in}} = \ell_{\text{in}}^{\rho, \alpha}/v$ is the time after departure at which a person with speed v enters area α and $t_{\text{out}} = \ell_{\text{out}}^{\rho, \alpha}/v$ the corresponding time at which it is exited. If a route ρ does not cross area α , then $\ell_{\text{in}}^{\rho, \alpha} = \infty$. If we consider a time interval $[t^-, t^+]$ after departure, the expected sojourn time for this person with constant speed v inside the area α within the interval is given by

$$\sigma(v, \ell_{\text{in}}^{\rho, \alpha}, \ell_{\text{out}}^{\rho, \alpha}, t^-, t^+) = \begin{cases} t^+ - \ell_{\text{in}}^{\rho, \alpha}/v & \text{if } t^- \leq \ell_{\text{in}}^{\rho, \alpha}/v \leq t^+ \leq \ell_{\text{out}}^{\rho, \alpha}/v, \\ \ell_{\text{out}}^{\rho, \alpha}/v - t^- & \text{if } \ell_{\text{in}}^{\rho, \alpha}/v \leq t^- \leq \ell_{\text{out}}^{\rho, \alpha}/v \leq t^+, \\ t^+ - t^- & \text{if } \ell_{\text{in}}^{\rho, \alpha}/v \leq t^- \leq t^+ \leq \ell_{\text{out}}^{\rho, \alpha}/v, \\ (\ell_{\text{out}}^{\rho, \alpha} - \ell_{\text{in}}^{\rho, \alpha})/v & \text{if } t^- \leq \ell_{\text{in}}^{\rho, \alpha}/v \leq \ell_{\text{out}}^{\rho, \alpha}/v \leq t^+, \\ 0 & \text{otherwise.} \end{cases} \quad (41)$$

In Eq. (41), the first line corresponds to the case where a person reaches the area within the time interval, but does not exit it. The second line is the inverse case. The third line represents the case where a person stays within the area during the full time period. Finally, the fourth line represents the case where a pedestrian enters and leaves the area during the period of interest, and the fifth case the situation where a pedestrian is not present in area α during the time interval at all.

Using Eq. (41), the expected contribution of a pedestrian traveling along route ρ with departure time interval τ to the accumulation of area α during

time interval τ' is given by

$$\begin{aligned} c_{(\alpha,\tau'),(\rho,\tau)} &= \int_{t=t_{\tau}^-}^{t_{\tau'}^+} \int_{v=0}^{\infty} \frac{\sigma(v, \ell_{\text{in}}^{\rho,\alpha}, \ell_{\text{out}}^{\rho,\alpha}, t_{\tau'}^- - t, t_{\tau'}^+ - t)}{\Delta t} f_v(v) h_{\tau}(t) dv dt \\ &= \frac{1}{\Delta t^2} \int_{v=0}^{\infty} f_v(v) \int_{t=t_{\tau}^-}^{t_{\tau'}^+} \sigma(v, \ell_{\text{in}}^{\rho,\alpha}, \ell_{\text{out}}^{\rho,\alpha}, t_{\tau'}^- - t, t_{\tau'}^+ - t) dt dv. \end{aligned} \quad (42)$$

For an efficient implementation, we note that the assignment fractions (39) and (42) are time-invariant, i.e., for $\Delta\tau = \tau' - \tau$ it holds that

$$b_{(\lambda,\tau'),(\rho,\tau)} = b_{(\lambda,\Delta\tau),(\rho,0)} \quad \text{and} \quad c_{(\alpha,\tau'),(\rho,\tau)} = c_{(\alpha,\Delta\tau),(\rho,0)}. \quad (43)$$

To further reduce the cost involved in computing Eq. (39) and Eq. (42), a maximum travel time Π_{max} is defined. If $\Delta\tau \geq \Pi_{\text{max}}$, it is assumed that $b_{(\lambda,\Delta\tau),(\rho,0)} = 0 \forall \lambda, \rho$ and $c_{(\alpha,\Delta\tau),(\rho,0)} = 0 \forall \alpha, \rho$. The threshold Π_{max} is chosen such that the error incurred by this numerical approximation is negligible.

References

- Alahi, A., Bierlaire, M., Vandergheynst, P., 2014. Robust real-time pedestrians detection in urban environments with low-resolution cameras. *Transportation Research Part C: Emerging Technologies* 39, 113–128.
- Alahi, A., Ramanathan, V., Fei-Fei, L., 2013. Socially-aware large-scale crowd forecasting. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2203–2210.
- Amacker, K., 2012. SBB Facts and Figures. Annual report, Swiss Federal Railways (SBB-CFF-FFS), Bern, Switzerland.
- Anken, N., Hänseler, F. S., Bierlaire, M., 2012. Flux piétonniers dans la gare de Lausanne: Vers l'estimation d'une matrice OD à l'aide des extrapolations voyageurs des CFF. Internal report (unpublished), Ecole Polytechnique Fédérale de Lausanne.
- Antonini, G., Bierlaire, M., Weber, M., 2006. Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological* 40 (8), 667–687.

- Ashok, K., Ben-Akiva, M. E., 2000. Alternative approaches for real-time estimation and prediction of time-dependent origin–destination flows. *Transportation Science* 34 (1), 21–36.
- Bauer, D., Brändle, N., Seer, S., Ray, M., Kitazawa, K., 2009. Measurement of pedestrian movements: A comparative study on various existing systems. *Pedestrian Behavior: Models, Data Collection and Applications*.
- Ben-Akiva, M. E., Bierlaire, M., 2003. Discrete choice models with applications to departure time and route choice. *Handbook of Transportation Science* 32.
- Benmoussa, M., Ducommun, F., Khalfi, A., Kharouf, M., Koymans, A., Nguyen, M., Raies, A., Vidaud, M., Birchler, C., 2011. Analyse des flux piétonniers en gare de Lausanne. Tech. rep., Ecole Polytechnique Fédérale de Lausanne.
- Bierlaire, M., Crittin, F., 2004. An efficient algorithm for real-time estimation and prediction of dynamic OD tables. *Operations Research* 52 (1), 116–127.
- Bierlaire, M., Crittin, F., 2006. Solving noisy, large-scale fixed-point problems and systems of nonlinear equations. *Transportation Science* 40 (1), 44–63.
- Bierlaire, M., Robin, T., 2009. Pedestrians choices. *Pedestrian Behavior. Models, Data Collection and Applications*, 1–26.
- Bierlaire, M., Toint, P. L., 1995. Meuse: An origin-destination matrix estimator that exploits structure. *Transportation Research Part B: Methodological* 29 (1), 47–60.
- Bierlaire, M., Toint, P. L., Tuytens, D., 1991. On iterative algorithms for linear least squares problems with bound constraints. *Linear Algebra and its Applications* 143, 111–143.
- Blue, V. J., Adler, J. L., 2001. Cellular automata microsimulation for modeling bi-directional pedestrian walkways. *Transportation Research Part B: Methodological* 35 (3), 293–312.

- Buchmüller, S., Weidmann, U., 2008. Handbuch zur Anordnung und Dimensionierung von Fussgängeranlagen in Bahnhöfen. IVT Projekt Nr. C-06-07. Institute for Transport Planning and Systems, ETH Zürich, Switzerland.
- Cascetta, E., 1984. Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. *Transportation Research Part B: Methodological* 18 (4), 289–299.
- Cascetta, E., Improta, A. A., 2002. Estimation of travel demand using traffic counts and other data sources. *Applied Optimization* 63, 71–91.
- Cascetta, E., Inaudi, D., Marquis, G., 1993. Dynamic estimators of origin-destination matrices using traffic counts. *Transportation Science* 27 (4), 363–373.
- Cascetta, E., Nguyen, S., 1988. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological* 22 (6), 437–455.
- Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A., 1996. A modified logit route choice model overcoming path overlapping problems: Specification and some calibration results for interurban networks. In: *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*. Pergamon, pp. 697–711.
- Cascetta, E., Papola, A., Marzano, V., Simonelli, F., Vitiello, I., 2013. Quasi-dynamic estimation of o-d flows from traffic counts: Formulation, statistical validation and performance analysis on real data. *Transportation Research Part B: Methodological* 55, 171–187.
- Cascetta, E., Postorino, M. N., 2001. Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks. *Transportation Science* 35 (2), 134–147.
- Casey, H. J., 1955. Applications to traffic engineering of the law of retail gravitation. *Traffic Quarterly* 9 (1), 23–35.
- Cheung, C. Y., Lam, W. H. K., 1998. Pedestrian route choices between escalator and stairway in MTR stations. *Journal of Transportation Engineering* 124 (3), 277–285.

- Cule, B., Goethals, B., Tassenoy, S., Verboven, S., 2011. Mining train delays. In: *Advances in Intelligent Data Analysis X*. Springer, pp. 113–124.
- 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. *Transportation Research Record: Journal of the Transportation Research Board* 1930 (1), 12–20.
- Daly, P. N., McGrath, F., Annesley, T. J., 1991. Pedestrian speed/flow relationships for underground stations. *Traffic Engineering & Control* 32 (2), 75–78.
- Danalet, A., Farooq, B., Bierlaire, M., 2014. A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. *Transportation Research Part C: Emerging Technologies* 44, 146–170.
- 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. *Transportation Research Part C: Emerging Technologies* 37, 210–222.
- Dial, R. B., 1971. A probabilistic multipath traffic assignment model which obviates path enumeration. *Transportation Research* 5 (2), 83–111.
- Djukic, T., Barceló, J., Bullejos, M., Montero, L., Cipriani, E., van Lint, H., Hoogendoorn, S. P., 2015. Advanced traffic data for dynamic od demand estimation: The state of the art and benchmark study. *Transportation Research Record: Journal of the Transportation Research Board*.
- Djukic, T., Flötteröd, G., van Lint, H., Hoogendoorn, S. P., 2012. Efficient real time od matrix estimation based on principal component analysis. In: *Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on*. IEEE, pp. 115–121.
- Edie, L. C., 1963. Discussion of traffic stream measurements and definitions. *Proceedings of the Second International Symposium on the Theory of Traffic Flow*, 139–154.

- Fisk, C. S., 1988. On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research Part B: Methodological* 22 (1), 69–73.
- Florian, M., Chen, Y., 1995. A Coordinate Descent Method for the Bi-level O–D Matrix Adjustment Problem. *International Transactions in Operational Research* 2 (2), 165–179.
- Frejinger, E., Bierlaire, M., 2007. Capturing correlation with subnetworks in route choice models. *Transportation Research Part B: Methodological* 41 (3), 363–378.
- Ganansia, F., Carincotte, C., Descamps, A., Chaudy, C., 2014. A promising approach to people flow assessment in railway stations using standard CCTV networks. In: *Transport Research Arena*, Paris.
- Gendre, G., Zulauf, C., 2010. Gare de Lausanne: Analyse des flux piétonniers. Internal report (I-PM-LS; unpublished), Swiss Federal Railways (SBB-CFF-FFS), Lausanne, Switzerland.
- Gentili, M., Mirchandani, P. B., 2012. Locating sensors on traffic networks: Models, challenges and research opportunities. *Transportation Research Part C: Emerging Technologies* 24, 227–255.
- Goverde, R. M. P., 2007. Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B: Methodological* 41 (2), 179–201.
- Hänseler, F. S., Bierlaire, M., Farooq, B., Mühlematter, T., 2014. A macroscopic loading model for time-varying pedestrian flows in public walking areas. *Transportation Research Part B: Methodological* 69, 60–80.
- Hänseler, F. S., Bierlaire, M., Molyneaux, N. A., Scarinci, R., Thémans, M., 2015a. Modeling pedestrian flows in train stations: The example of Lausanne railway station. *Proceedings of the 15th Swiss Transport Research Conference*.
- Hänseler, F. S., Molyneaux, N. A., Bierlaire, M., 2015b. Python implementation of pedestrian OD demand estimator for train stations. URL <https://github.com/flurinus/DemEstMeth>

- Hazelton, M. L., 2000. Estimation of origin–destination matrices from link flows on uncongested networks. *Transportation Research Part B: Methodological* 34 (7), 549–566.
- Hazelton, M. L., 2003. Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice* 37 (10), 811–822.
- Helbing, D., Molnár, P., 1995. Social force model for pedestrian dynamics. *Physical Review E* 51 (5), 4282–4286.
- Higgins, A., Kozan, E., 1998. Modeling train delays in urban networks. *Transportation Science* 32 (4), 346–357.
- Highway Capacity Manual, 2000. Transportation Research Board. Washington, DC.
- Hoogendoorn, S. P., Bovy, P. H. L., 2004. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological* 38 (2), 169–190.
- Hoogendoorn, S. P., Daamen, W., 2004. Design assessment of Lisbon transfer stations using microscopic pedestrian simulation. In: *Computers in railways IX (Congress Proceedings of CompRail 2004)*. pp. 135–147.
- Hughes, R. L., 2002. A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodological* 36 (6), 507–535.
- Kaakai, F., Hayat, S., El Moudni, A., 2007. A hybrid Petri nets-based simulation model for evaluating the design of railway transit stations. *Simulation Modelling Practice and Theory* 15 (8), 935–969.
- Lam, W. H. K., Wu, Z. X., Chan, K. S., 2003. Estimation of transit origin–destination matrices from passenger counts using a frequency-based approach. *Journal of Mathematical Modelling and Algorithms* 2 (4), 329–348.
- Lavadinho, S., 2012. Compréhension fine des stratégies piétonnières en gare de Lausanne. Internal report (unpublished), Swiss Federal Railways (SBB-CFF-FFS), Lausanne, Switzerland.

- Lavadinho, S., Alahi, A., Bagnato, L., 2013. Analysis of Pedestrian Flows: Underground pedestrian walkways of Lausanne train station. Internal report (unpublished), VisioSafe SA, Switzerland.
- 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.
- Løvås, G. G., 1994. Modeling and simulation of pedestrian traffic flow. *Transportation Research Part B: Methodological* 28 (6), 429–443.
- Marzano, V., Papola, A., Simonelli, F., 2009. Limits and perspectives of effective o–d matrix correction using traffic counts. *Transportation Research Part C: Emerging Technologies* 17 (2), 120–132.
- Molyneaux, N. A., Hänseler, F. S., Bierlaire, M., 2014. Modeling of train-induced pedestrian flows in railway stations. *Proceedings of the 14th Swiss Transport Research Conference*.
- Montero, L., Codina, E., Barceló, J., 2015. Dynamic od transit matrix estimation: formulation and model-building environment. In: *Progress in Systems Engineering*. Springer, pp. 347–353.
- Mustafa, M., Ashaari, Y., 2015. Assessing pedestrian behavioral pattern at rail transit terminal: State of the art. In: *Proceedings of the International Civil and Infrastructure Engineering Conference*. Springer, pp. 1245–1254.
- Nguyen, S., Morello, E., Pallottino, S., 1988. Discrete time dynamic estimation model for passenger origin/destination matrices on transit networks. *Transportation Research Part B: Methodological* 22 (4), 251–260.
- Pettersson, P., 2011. Passenger waiting strategies on railway platforms: Effects of information and platform facilities. Master’s thesis, KTH.
- Rindsfuser, G., Klügl, F., 2007. Agent-based pedestrian simulation: A case study of Bern Railway Station. *The Planning Review* 170, 9–18.

- SBB-Personenverkehr, 2011. SBB Clearing Extrapolation Voyageurs HOP. Internal report (unpublished), Swiss Federal Railways (SBB-CFF-FFS), Bern, Switzerland.
- Starmans, M., Verhoeff, L., van den Heuvel, J. P. A., 2014. Passenger transfer chain analysis for reallocation of heritage space at Amsterdam Central station. *Transportation Research Procedia* 2, 651–659.
- Ton, D., 2014. NAVISTATION: A study into the route and activity location choice behaviour of departing pedestrians in train stations. Master's thesis, Delft University of Technology.
- Turner, S., Middleton, D., Longmire, R., Brewer, M., Eurek, R., 2007. Testing and evaluation of pedestrian sensors. Monograph, Transportation Research Board.
- U.S. Department of Transportation, 2013. Traffic monitoring guide. Tech. rep., Federal Highway Administration.
- 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. *Transportation Research Procedia* 2, 642–650.
- van Hagen, M., 2011. Waiting experience at train stations. Ph.D. thesis, Universiteit Twente.
- van Zuylen, H. J., Willumsen, L. G., 1980. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological* 14 (3), 281–293.
- Weidmann, U., 1992. *Transporttechnik der Fussgänger*. Schriftenreihe des IVT Nr. 90. Institute for Transport Planning and Systems, ETH Zürich, Switzerland.
- Willumsen, L. G., 1981. An entropy maximising model for estimating trip matrices from traffic counts. Ph.D. thesis, University of Leeds.
- Wilson, A. G., 1970. *Entropy in urban and regional modelling*. Pion Press.
- Wong, S. C., Tong, C., 1998. Estimation of time-dependent origin-destination matrices for transit networks. *Transportation Research Part B: Methodological* 32 (1), 35–48.

- Wong, S. C., Tong, C. O., Wong, K. I., Lam, W. H. K., Lo, H. K., Yang, H., Lo, H. P., 2005. Estimation of multiclass origin-destination matrices from traffic counts. *Journal of Urban Planning and Development* 131 (1), 19–29.
- Xu, X. Y., Liu, J., Li, H. Y., Hu, J. Q., 2014. Analysis of subway station capacity with the use of queueing theory. *Transportation Research Part C: Emerging Technologies* 38, 28–43.
- Yang, H., 1995. Heuristic algorithms for the bilevel origin-destination matrix estimation problem. *Transportation Research Part B: Methodological* 29 (4), 231–242.
- Yang, Y., Fan, Y., 2015. Data dependent input control for origin-destination demand estimation using observability analysis. *Transportation Research Part B: Methodological* 78, 385–403.
- Yuan, J., Hansen, I. A., 2007. Optimizing capacity utilization of stations by estimating knock-on train delays. *Transportation Research Part B: Methodological* 41 (2), 202–217.
- Zhou, X., Mahmassani, H. S., 2007. A structural state space model for real-time traffic origin-destination demand estimation and prediction in a day-to-day learning framework. *Transportation Research Part B: Methodological* 41 (8), 823–840.