Schedule-based estimation of pedestrian origin-destination demand in railway stations

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

January 8, 2015

Report TRANSP-OR 150108
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*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

Abstract

A framework is outlined for estimating pedestrian demand within a railway station which takes into account the train timetable, ridership data, and various direct and indirect indicators of demand. Such indicators may include e.g. link flow counts, measurements of density and travel times, or historical information. The problem is considered in discrete time and at the aggregate level, i.e., for groups of pedestrians associated with the same user class, origin-destination pair and departure time interval. The formulation of the framework is probabilistic, allowing to explicitly capture the stochastic characteristics of demand. A case study analysis of a Swiss railway 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 information besides the timetable and ridership estimates is available.

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

1 Introduction

Passenger railway systems are experiencing a tremendous growth in many countries around the world. For the last decade, the national rail operator of the US, Amtrak, reports an average annual growth rate in number of daily transported passengers of close to 3% (Puentes et al., 2013). Similar figures are valid for Europe with e.g. 4.1% for Switzerland's national operator SBB-CFF-FFS (Amacker, 2012) or 2.7% for Germany's Deutsche Bahn (Kasparick, 2010). In Asia, annual growth rates have been even higher, with a reported average of 11.0% for Singapore (Land Transport Authority, Singapore, 2012) and 9.2% for South Korea's high-speed train network (Chung, 2012).

Partially in response to that growth, and partially inducing it, transportation systems have been continuously expanded over the past decades (Kallas, 2014). In particular, the frequency and the capacity of trains have been increased. However, one component that has received less attention is that of rail access installations (Schneider, 2012). The capacity of pedestrian facilities has not been a limiting factor for a long time. Today, they are increasingly considered a bottleneck of railway systems, and pedestrian congestion in train stations is becoming a common phenomenon (Ganansia et al., 2014).

During normal operation of a railway station, primarily the arrival and departure of large trains are responsible for the peak usage of pedestrian facilities. Following a train arrival, a potentially large number of passengers alights, and then moves as a dense crowd through the station. These 'pedestrian waves' often provoke congestion in platform access ways (van den Heuvel et al., 2013). Similarly, prior to train departures, outbound passengers accumulate on platforms that serve as waiting areas. If platforms are small, or the number of prospective passengers high, space quickly gets scarce (Schneider, 2012). Such train-induced demand patterns have a significant impact on customer satisfaction, as well as on the performance and safety of a train station. With the densification of train timetables that is ongoing in many countries, and the prospect of more capacious trains, this impact is likely to aggravate.

By increasing the capacity of rail access installations, most negative side-effects of a growing ridership can be alleviated. Unfortunately, the required investment is often prohibitively high. The enlargement of plat-

forms or access ways typically requires a redesign of large parts of a train station, which can seldom be closed during construction. Many studies documenting extensions of railway stations are found in the literature, including examples from the Netherlands (van den Heuvel and Hoogenraad, 2014), France (Ganansia et al., 2014), Portugal (Hoogendoorn and Daamen, 2004) or South Africa (Hermant, 2012).

Given the complexity and cost of an expansion of rail access installations, a diligent planning and dimensioning is indispensable. Key in this process is the assessment of the usage of a railway station. In this work, we propose a methodology for estimating pedestrian demand in a railway station, taking into account the particular demand patterns induced by trains. An explicit integration of the train timetable allows to quantitatively assess its influence on pedestrian traffic in rail access installations. To underline the practical applicability of the framework, a case study involving pedestrian demand in the walking facilities of Lausanne railway station, Switzerland, is presented.

In the long run, we hope that this work contributes to the development of tools and methodologies that allow optimizing the pedestrian infrastructure of railway stations, as well as the train timetable or the track assignment in the context of pedestrian flows.

2 Literature review

Pedestrian behavior in railway stations increasingly attracts the attention of academic research. Broadly, it can be distinguished between empirical studies aiming at characterizing behavior, and those concerned with its mathematical modeling.

In an early study, Daly et al. (1991) investigate the relationship between speed and flow and between flow and travel time in various pedestrian facilities of London's underground system. Lam and Cheung (2000) examine several metro stations as well as pedestrian areas in a shopping center in Hong Kong. Differentiating by trip purpose, flow capacities are evaluated and flow-travel time functions are calibrated. Compared to the results from London, users of Hong Kong's mass transit system are found to be better at dealing with high levels of congestion, which is attributed to anatomical and sociological differences.

In a related study, Lam et al. (1999) investigate the train dwelling time and the distribution of pedestrians on platforms in two stations of Hong Kong's Light Rail Transit system. A behavioral analysis reveals that people are significantly less willing to board a train if the latter is congested, and if the journey to be made is longer. Also focusing on train platforms, Zhang et al. (2008) quantitatively describe the process of alighting and boarding in metro stations in Beijing. A cellular automaton model is developed, calibrated on empirical data and complemented with a literature review. Various empirically observed behavior patterns can be reproduced with high accuracy. Pettersson (2011) investigates the behavior of pedestrians on railway platforms from an architect's perspective. In particular, the effect of signposts, availability of seats and entrances on the distribution of pedestrians along the platform is investigated at the example of a Swedish and a Japanese case study. Concrete recommendations are made regarding how a more homogeneous distribution along a train platform can be attained.

Recently, Ganansia et al. (2014) have studied the use of standard CCTV networks for measuring pedestrian flows in railway stations. Several case studies, including a TGV station and two subway stations in France and Italy, are discussed. It is found that data obtained through such a system is in principle highly useful for a continuous monitoring of the spatio-temporal evolution of pedestrian flows, but also that an a posteriori 'correction' is necessary whenever dense crowds need to be accurately measured. A considerable effort is made towards developing such a calibration model.

Notably using such camera-based data, several researchers have empirically analyzed the influence of train arrivals and departures on pedestrian behavior patterns. For instance, Buchmüller and Weidmann (2008) describe the flows on platform access ways caused by alighting train passengers. Following the same approach, Molyneaux et al. (2014) discuss the concrete example of Lausanne railway station. Characteristic for such train-induced pedestrian arrival flows is the lagged onset of the flow after the arrival of the train, the saturation at a given capacity flow rate, and a subsequent decay (see Section 4.4).

A similar example is provided by van den Heuvel and Hoogenraad (2014), who use automatic fare collection (AFC) data to study various aspects of pedestrian behavior within railway stations. At the example of Utrecht Central Station, the passenger arrival distribution at AFC exit

gates is investigated. In comparison to the arrival pattern observed on platform exit ways in Lausanne (Molyneaux et al., 2014), there are more 'late arrivals', which is attributed to passengers that visit a restroom, restaurant or other service points before leaving the railway station.

In principle, similar empirical relationships can be established for train-induced departure flows, i.e., for the flow patterns of outbound train passengers walking towards platforms. However, empirical evidence shows that such relationships are more complex for reasons such as differences in risk aversion among passengers, or due to constraints imposed by the schedule of tertiary transport modes (van Hagen, 2011). For instance, outbound passengers associated with an interregional train and long headway times typically arrive earlier at a platform than those bound for a regional train with a high frequency. Transfer passengers may simply show up on their respective departure platform after they have arrived at a railway station, giving raise to complex correlation patterns.

In spite of these challenges, a few researchers have proposed a formal relationship between train departures and the number of prospective passengers on the platform prior to departure. Specifically, based on qualitative observations, Tolujew and Alcalá (2004) and Hermant et al. (2010) assume that the accumulation of outbound pedestrians on a train platform prior to departure first follows an S-curve which, once the train has arrived, is 'inverted' by pedestrians that start boarding.

To assess the design of a railway station, the ability to predict the routes taken by pedestrians is crucial. Several studies have been dedicated to this endeavor. Again for the case of a metro station in Hong Kong, Cheung and Lam (1998) investigate the route choice between escalators and stairways leading to a train platform. A relationship between flow and travel time is first established. This characteristic relationship is then used in a choice model allowing to predict the percentage of escalator-users for ascending and descending directions as a function of prevailing traffic conditions. At the example of two Dutch stations, Daamen et al. (2005) have collected route choice data by following passengers through the facility from their origin to their destination. Likewise, a route choice model is estimated allowing to predict the influence of level changes in walking routes on passenger route choice behavior. It is concluded that the various ways of bridging level changes such as ramps, stairs or escalators have a different impact on the attractiveness of a route. Further similar

studies are provided by Srikukenthiran et al. (2014) and Stubenschrott et al. (2014), who consider subway stations in Toronto and Vienna, respectively.

In a more general context, Hoogendoorn and Bovy (2004) develop a model for pedestrian route choice and activity scheduling. Every route and activity schedule is associated with a cost, and it is assumed that pedestrians choose their route and activities such that the perceived utility is maximized. The methodology is applied to a case study of a major Dutch transportation hub.

If a large number of pedestrians is to be considered, an individual consideration of activities can be impractical. Instead, pedestrians may be divided into user classes that are characterized by a common activity or behavior pattern (e.g. Wong et al., 2005). Pursuing a socio-geographical approach, Lavadinho (2012) analyses the behavior of different user classes at the example of Lausanne railway station. In the context of a train station, it can be broadly distinguished between inbound, outbound, transit and local users: Inbound passengers often leave the railway station immediately upon arrival, following the shortest path. Outbound passengers sometimes arrive early at the railway station, spending some spare time waiting, or pursuing any other intermediate activity such as reading a newspaper. For transit passengers, the behavior depends significantly on the duration of the layover, and may consist in hurrying from one platform to another, or in spending time in a restaurant. Local users finally are those that visit a railway station to take advantage of its sales and service points, or simply traverse it.

Most of the aforementioned studies concentrate on individual aspects of pedestrian behavior such as flow-travel time relationships or way finding, or they concern sociological aspects. In the following, several studies are presented that consider more comprehensive and quantitative models of pedestrian behavior in railway stations.

Lee et al. (2001) provide one of the first model-based studies of pedestrian flows in a railway station in the scientific literature. For a major station in Hong Kong's metro system, origin-destination demand and travel times are collected using a large number of human observers. From this data, flow-travel time relationships are derived, which are used in a relatively simple, network-based pedestrian flow model. A comparison between empirical data and model prediction indicates a good performance of the model.

Along the same lines, Daamen (2004) develops a particularly detailed modeling framework for pedestrian flows in railway stations. A multitude of models for describing the processes of queueing, boarding, alighting, waiting, walking as well as route and activity choice are proposed, and jointly implemented. The framework represents a hybrid queueing network/link flow-model operating in discrete space. Various case studies across the Netherlands are considered.

More recently, Kaakai et al. (2007) have developed a related model at the macroscopic level. They consider both discrete processes such as the arrival and departure of trains, as well as continuous processes such as the fill-up of railway platforms by pedestrians awaiting a train, or pedestrian flows in walking facilities. The model is represented as a Petri net and applied to a French case study involving a railway station with a single platform. Hanisch et al. (2003) and Tolujew and Alcalá (2004) qualitatively follow a similar approach, but do not provide a mathematical specification of their model.

At the microscopic level, Xu et al. (2014) develop a model describing pedestrian behavior in a Chinese metro station. The framework is entirely based on a queueing network, i.e., all processes including entering the railway station, passing ticket gates, walking and boarding are represented by queues. The framework is applied to estimate the maximum service rate of a metro station, as well as to determine the optimal inflow rate at the entrance at which this capacity is attained. Pursuing a similar goal, but using a macroscopic approach, Starmans et al. (2014) have conducted a study of Amsterdam Central Station, for which a 'pedestrian transfer chain' model is developed to assess the design and operation of the station.

There are several more studies of pedestrian flows in railway stations that concentrate rather on a high level of accuracy for specific applications than on a methodological contribution. Most of them pursue an agent-based approach and describe various local challenges such as the placement of access gates in Lisbon (Hoogendoorn and Daamen, 2004), the re-design of access ways in Bern (Rindsfüser and Klügl, 2007), the evacuation of a metro station in Beijing (Jiang et al., 2009), the modeling of waiting areas in German railway stations (Davidich et al., 2013) or the design of a new station in South Africa (Hermant, 2012).

While most of the previously mentioned studies use sophisticated models for describing various aspects of pedestrian behavior such as walking,

waiting or boarding, the methods used to estimate pedestrian demand are quite simplistic. Many do not even specify how these estimates are obtained. Other studies rely on flow counts that are converted to origin-destination demand values based on simple rules of thumb, such as assuming a uniform demand over time. Few studies take the train timetable explicitly into account, but if so, only for a single platform.

Despite a considerable interest in pedestrian behavior models for railway stations, there seems to be a lack of dedicated methods for estimating pedestrian demand. Ideally, such a methodology should be able to reproduce the 'demand micro-peaks' (Hermant, 2012) caused by incoming and outgoing trains, i.e., it should explicitly take the train timetable into account. Moreover, it should be able to distinguish between user classes characterized by class-specific behavior.

To develop a corresponding demand estimation model, different approaches seem conceivable. For instance, in the context of a university campus, Danalet et al. (2014) propose an activity choice model based on WiFi traces and individual class schedules, which assume a similar role as the train timetable for a railway station. In principle, an analogous approach could be adapted. However, for most applications involving train stations, disaggregate data is still unavailable. Instead, it is more efficient to estimate origin-destination (OD) demand at the aggregate level.

For problems concerning car traffic, such dynamic OD demand estimation methodologies are well established. Inspired by the seminal work by Cascetta et al. (1993), a large number of statistical methods have been developed in the last two decades (Bera and Rao, 2011). For instance, Wong et al. (2005) provide an example of a multi-class estimation work that is applied to the case of a transit network. More recently, Shao et al. (2014) have presented an estimation model that allows exploring the stochastic characteristics of OD demand. Based on a static formulation of the demand estimation problem, besides the mean, also the variation caused by day-to-day fluctuations of demand is estimated.

By building on the aforementioned achievements, this study aims at providing a dedicated estimation methodology for pedestrian OD demand in railway stations. It is designed in a probabilistic way, so that the stochastic characteristics of OD demand, such as its variance, can be readily captured. Particular emphasis is given to the development of a fully dynamic estimation framework that explicitly considers the train timetable.

3 Estimation framework

The train station is considered as a mathematical system, for which a model consisting of input variables, state variables, structural equations and measurement equations is defined. To facilitate the discussion of these components, first a general notation is established. Subsequently, the model formulation is presented, and in Section 4, a concrete specification is provided.

3.1 Notation

Walkable space is represented by a directed graph $\mathcal{G} = (\mathcal{N}, \Lambda)$, where $\mathcal{N}$ represents the set of nodes $\nu \in \mathcal{N}$, and Λ the set of directed links $\lambda \in \Lambda$. Nodes through which pedestrians enter and leave the pedestrian facility network are referred to as centroids, and their set is denoted by $\mathcal{C} \subset \mathcal{N}$.

The subset of centroids associated with a platform $p \in \mathcal{P}$ is denoted by $\mathcal{C}_p$, where $\mathcal{P}$ denotes the set of platforms. The subset of centroids $\mathcal{C}_\mathcal{P}$ represents the set of all centroids associated with any platform. Similarly, the set Λ_p^{arr} includes the links representing platform exit ways of platform p, and $\Lambda_\mathcal{P}^{arr}$ the set of all platform exit ways.

Fig. 1 provides an illustration of the proposed space representation. In the case of the upper platform, an East and a West sector are considered, each of which is represented by two centroids. Pedestrian flows within platform areas from and towards each railway track are explicitly modeled. This case is important for the derivation of the model formulation presented in this section. In the case of the lower platform, a single centroid represents each platform sector, and no explicit distinction is made between railway tracks. This case will be relevant for the case study presented in Section 4.

Any two centroids can be connected by a route $\rho \in \mathcal{R}$, defined as a sequence of links $\rho = (\lambda_1^\rho, \lambda_2^\rho, \dots)$. Besides, a set of 'subroutes' $\mathcal{Q}$ is defined, where for each subroute $\varrho \in \mathcal{Q}$ at least one parent route exists. The set of parent routes associated with subroute ϱ is denoted by $\mathcal{R}_\varrho^{sup}$. Subroutes may originate and terminate at any node, $\nu_\varrho^o, \nu_\varrho^d \in \mathcal{N}$, and are useful to incorporate subroute flows obtained from a tracking system. In case the tracking system covers the full network, the set of subroutes is identical to the set of routes.

A connected subnetwork $\mathcal{G}_\alpha = (\mathcal{N}_\alpha, \Lambda_\alpha)$, with $\mathcal{G}_\alpha \subset \mathcal{G}$, is referred to

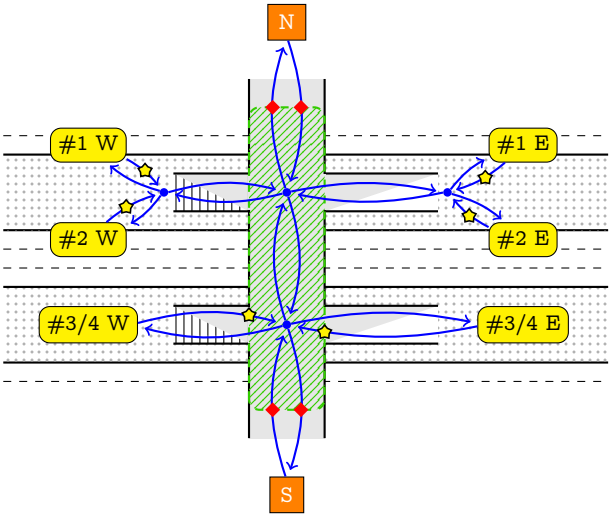


Figure 1: Illustration of network topology at the example of a simple train station. Railway tracks are denoted by dotted lines. Gray areas represent walking facilities, and dotted areas denote joint walking/waiting areas (in this case only platform areas). Levels are bridged by ramps and stairways, denoted by standard floor plan symbols. Platform sectors are represented by centroids shown as yellow rectangles with rounded corners. They may be associated with one or a pair of railway tracks. Further centroids are shown as orange squares, which include sales or service points, or exit/entrance areas. The pedestrian walking network is represented by a graph connecting centroids and intersection nodes (solid blue arrows representing directed links). Pedestrian counters are represented by red diamonds. An exemplary area is shown in shaded green. Yellow stars denote cordons at which train-induced passenger arrival flows are estimated.

as an area α, and the set of all areas by $\mathcal{A}$. The concept of an area is useful to consider the occupation of certain facilities such as waiting halls or platforms. Knowledge of occupation can be helpful for dimensioning such facilities, or to validate the results of the estimation model, as done in this work. Areas are allowed to overlap, and their union is not required to cover the full network.

Each pedestrian is associated with a specific OD pair $\zeta = (\nu_\zeta^o, \nu_\zeta^d)$, where $\nu_\zeta^o, \nu_\zeta^d \in \mathcal{C}$. The set of all OD pairs is denoted by $\mathcal{Z}$. Furthermore, $\mathcal{Z}_\nu^{orig}$ represents the set of OD pairs sharing node ν as their origin, and $\mathcal{Z}_\nu^{dest}$ those having node ν as destination. Also, for each OD pair ζ, the set of connecting routes is denoted by $\mathcal{R}_\zeta$.

Moreover, each pedestrian is associated with a user class β. The set

of classes is represented by $\mathcal{B}$. Without loss of generality, we consider the aforementioned set of four user classes representing inbound, outbound and transfer passengers, as well as local users, i.e., $\mathcal{B} = \{\text{in}, \text{out}, \text{tr}, \text{loc}\}$. In another context, a different specification of user classes is easily applicable.

Regarding the representation of time, the period of interest is divided into a set of discrete intervals $\mathcal{T}$, where each interval $\tau = [t_\tau^-, t_\tau^+]$, $\tau \in \mathcal{T}$, is of uniform length $\Delta t = t_\tau^+ - t_\tau^-$.

Based on the above representation of space, time and user classes, the concept of demand can be defined. The number of travelers leaving the origin v_o^ζ during time interval τ towards destination v_d^ζ associated with class β is represented by $d_{\zeta,\tau}^\beta$, and the corresponding time-space expanded vector by $\boldsymbol{d}_\beta = [d_{\zeta,\tau}^\beta]$, being of size $|\mathcal{Z}||\mathcal{T}|$. The complete demand vector is denoted by $\boldsymbol{d} = [\boldsymbol{d}_{\text{in}}; \boldsymbol{d}_{\text{out}}; \boldsymbol{d}_{\text{tr}}; \boldsymbol{d}_{\text{loc}}]$ and represents the state variable of the demand estimation model. The unit of demand is number of pedestrians.

During the time horizon $\mathcal{T}$, a set of trains $\mathcal{M}$ is considered. The platform serving train $m \in \mathcal{M}$ is denoted by p_m, the corresponding vector of length $|\mathcal{M}|$ by $\boldsymbol{p}$, and the set of trains associated with platform p by $\mathcal{M}_p$.

For a train $m \in \mathcal{M}$, t_m^{arr} and t_m^{dep} denote the arrival and departure time, and w_m^{off} and w_m^{on} the alighting and boarding volume. The corresponding vectors are $\boldsymbol{t}_{\text{arr}} = [t_m^{\text{arr}}]$, $\boldsymbol{t}_{\text{dep}} = [t_m^{\text{dep}}]$, $\boldsymbol{w}_{\text{off}} = [w_m^{\text{off}}]$ and $\boldsymbol{w}_{\text{on}} = [w_m^{\text{on}}]$, which are of size $|\mathcal{M}|$, as well as $\boldsymbol{w} = [\boldsymbol{w}_{\text{off}}; \boldsymbol{w}_{\text{on}}]$ of length $2|\mathcal{M}|$. In the following, boarding and alighting volumes are jointly referred to as train exchange volumes.

A pre-specified aggregated network supply model is assumed to exist that, given the demand, predicts for each user class the expected traffic conditions. Such a model is typically referred to as dynamic traffic assignment (DTA) model. We assume that the vector γ contains all its parameters specifying e.g. the free-flow walking speed distribution or the route choice model. The traffic conditions of each class β are assumed to be described by the generic vector $\boldsymbol{s}_\beta$. While no precise definition of $\boldsymbol{s}_\beta$ is necessary in the current context, it may be thought of as the vector containing the travel time distribution of users of class β on all links $\ell \in \Lambda$ during all time intervals $\tau \in \mathcal{T}$. The corresponding vector $\boldsymbol{s} = [\boldsymbol{s}_{\text{in}}; \boldsymbol{s}_{\text{out}}; \boldsymbol{s}_{\text{tr}}; \boldsymbol{s}_{\text{loc}}]$ represents the state variables of the supply model.

Various demand indicators can be derived from OD demand to facilitate the formulation of the structural and measurement model. In this work, primarily flows and area occupations are of interest. All flows are cumula-

tive over time, i.e., they effectively represent pedestrian counts. Following Edie (1963), occupation is defined as the time-mean average number of users in an area. Therefore, the unit of demand, cumulative flows and area occupation is always 'number of pedestrians'.

Multiple types of link flows are distinguished, namely cumulative link flows, total cumulative link flows, cumulative arrival link flows, and cumulative subroute link flows. Cumulative link flows are associated with a single pedestrian class, whereas total cumulative link flows represent the sum over all pedestrian classes (The same distinction is made between the occupation associated with a specific pedestrian class, and the total occupation.). Cumulative arrival and subroute link flows are associated with arriving passengers and with a specific subroute, respectively.

Besides link flows, the total cumulative origin flow and the total cumulative destination flow associated with a centroid are considered, representing the total 'generation' and 'absorption' of pedestrians of any class at a network node. They are useful for integrating ridership information, sales data and travel surveys. In this work, link flows are cumulative with respect to a single time interval τ, and flows associated with centroids are cumulative with respect to the full time period $\mathcal{T}$.

Furthermore, the total cumulative departure flow of a train platform is of interest, denoting the number of departing pedestrians associated with a platform. Compared to the total cumulative destination flow to a node, the cumulative departure flow takes into account that a platform may consist of several centroids.

With the exception of the train exchange volumes whose notation has already been defined, the following list provides an overview of all considered demand indicators. Their unit is number of pedestrians unless stated otherwise. A mathematical definition of all demand indicators is provided in Section 3.3.

- $\boldsymbol{d}_{\text{tot}} = [\mathrm{d}^{\text{tot}}_{\zeta,\tau}]$: The total demand $\mathrm{d}^{\text{tot}}_{\zeta,\tau}$ associated with OD pair ζ and time interval τ, and the corresponding time-space expanded vector $\boldsymbol{d}_{\text{tot}}$ of length $|\mathcal{Z}||\mathcal{T}|$.

- $\boldsymbol{f}_{\beta} = [\mathrm{f}^{\beta}_{\lambda,\tau}]$: The cumulative link flow $\mathrm{f}^{\beta}_{\lambda,\tau}$ associated with users of class β on link λ during time interval τ, and the corresponding time-space expanded vector $\boldsymbol{f}_{\beta}$ of length $|\Lambda||\mathcal{T}|$.

- $\boldsymbol{f}_{tot} = [f^{tot}_{\lambda,\tau}]$: The total cumulative link flow $f^{tot}_{\lambda,\tau}$ on link λ during time interval τ, and the corresponding time-space expanded vector $\boldsymbol{f}_{tot}$ of length $|\Lambda||\mathcal{T}|$.

- $\boldsymbol{f}_{arr} = [f^{arr}_{\lambda,\tau}]$: The cumulative arrival link flow $f^{tot}_{\lambda,\tau}$ on link λ during time interval τ, and the corresponding time-space expanded vector $\boldsymbol{f}_{arr}$ of length $|\Lambda||\mathcal{T}|$. This part of the link flow is particularly volatile, and useful for the integration of the train timetable in the estimation framework.

- $\boldsymbol{e}_{\beta} = [e^{\beta}_{\varrho,\tau}]$: The cumulative subroute flow $e^{\beta}_{\varrho,\tau}$ associated with users of class β along subroute ϱ leaving node ν^{o}_{ϱ} during time interval τ, and the corresponding time-space expanded vector $\boldsymbol{e}_{\beta}$ of length $|\mathcal{Q}||\mathcal{T}|$.

- $\boldsymbol{e}_{tot} = [e^{tot}_{\varrho,\tau}]$: The total cumulative subroute flow $e^{tot}_{\varrho,\tau}$ along subroute ϱ leaving node ν^{o}_{ϱ} during time interval τ, and the corresponding time-space expanded vector $\boldsymbol{e}_{tot}$ of length $|\mathcal{Q}||\mathcal{T}|$.

- $\boldsymbol{n}_{\beta} = [n^{\beta}_{\alpha,\tau}]$: The occupation $n^{tot}_{\alpha,\tau}$ associated with user class β on area α during time interval τ, and the corresponding time-space expanded vector $\boldsymbol{n}_{\beta}$ of length $|\mathcal{A}||\mathcal{T}|$.

- $\boldsymbol{n}_{tot} = [n^{tot}_{\alpha,\tau}]$: The total occupation $n^{tot}_{\alpha,\tau}$ on area α during time interval τ, and the corresponding time-space expanded vector $\boldsymbol{n}_{tot}$ of length $|\mathcal{A}||\mathcal{T}|$.

- $\boldsymbol{o}_{tot} = [o^{tot}_{\nu}]$: The total cumulative origin flow o^{tot}_{ν} emanating from centroid ν during the time period $\mathcal{T}$, and the corresponding vector $\boldsymbol{o}_{tot}$ of length $|\mathcal{C}|$.

- $\boldsymbol{q}_{tot} = [q^{tot}_{\nu}]$: The total cumulative destination flow q^{tot}_{ν} reaching centroid ν during the time period $\mathcal{T}$, and the corresponding vector $\boldsymbol{q}_{tot}$ of length $|\mathcal{C}|$.

- $\boldsymbol{g}_{tot} = [g^{tot}_{p}]$: The total cumulative departure flow g^{tot}_{p} at platform p during the time period $\mathcal{T}$, and the corresponding vector $\boldsymbol{g}_{tot}$ of length $|\mathcal{P}|$.

- $\boldsymbol{r}_{\beta} = [r^{\beta}_{\nu}]$: The time-mean average ratio r^{β}_{ν} of users associated with class β in the origin flow at centroid ν during the time period $\mathcal{T}$,

and the corresponding vector $\boldsymbol{r}_\beta$ of length $|\mathcal{C}|$. The ratio r_v^β is a dimensionless quantity.

3.2 Input variables

To estimate time-dependent OD demand, it is desirable to consider multiple sources of data (Cascetta et al., 1993; Sherali and Park, 2001). Some of them, such as link flow counts, are accurate and widely available, whereas e.g. route flow data may be biased and less complete, but useful for resolving structural ambiguities (Hazelton, 2003). Furthermore, travel surveys or previous demand estimates may be relatively inaccurate, but useful for reducing the solution space if the estimation problems allows for multiple solutions.

In the following, we mark biased variables such as measurements by a hat (e.g. $\widehat{\boldsymbol{f}}_\beta$), and 'incomplete' vectors by a prime (e.g. $\boldsymbol{f}'_\beta$). The latter are referred to as reduced vectors, and a reduction matrix $\boldsymbol{R}$ is defined that relates each of them to the corresponding full vector (e.g. $\boldsymbol{f}'_\beta = \boldsymbol{R}_f \boldsymbol{f}_\beta$).

The following assumptions regarding data availability are made:

- The network topology, including the set of OD pairs and routes, the time discretization, as well as the set of trains including the train timetable and the train-track assignment, are assumed to be given. The parameter vector $\boldsymbol{\gamma}$ is also known a priori.

- An a priori estimate of train exchange volumes $\widehat{\boldsymbol{w}}'$ for a subset of trains is assumed to be available, which can be inferred e.g. from traffic surveys, door counts or train capacity.

- Some direct or indirect observations of OD demand are assumed to be available. Direct observations can be obtained from a pedestrian tracking system. Indirect observations usually include measurements of flow, density, travel time or walking speed. Their temporal resolution has to be such that dynamical features of pedestrian flows, such as 'micro-peaking', can be captured, i.e., a maximum aggregation period of the order of a minute is desirable. Here, measurements of total cumulative link flows, $\widehat{\boldsymbol{f}}'_{\text{tot}}$, total cumulative subroute flows, $\widehat{\boldsymbol{e}}'_{\text{tot}}$, and total area occupations, $\widehat{\boldsymbol{n}}'_{\text{tot}}$, are assumed to be available, understanding that other demand indicators can be considered analogously.

- For a subset of centroids, estimates of the total cumulative origin and destination flows, $\hat{\boldsymbol{o}}'_{\text{tot}}$ and $\hat{\boldsymbol{q}}'_{\text{tot}}$, are assumed available. They may be obtained from customer frequentation data at sales and service points, or from manual surveys. Moreover, a priori information on the cumulative class split ratios, $\hat{\boldsymbol{r}}'_{\beta}$, is considered available. For instance, some shop keepers may know what percentage of their clients are train passengers, or railway operators may dispose of an a priori estimate of the overall percentage of transfer vs. non-transfer passengers.
- An a priori estimate of class-specific origin-destination demand $\hat{\boldsymbol{d}}$ is assumed to be available, which is typically obtained from previous estimates and often afflicted with a large uncertainty.

3.3 Structural model

The structural model consists of three parts, namely a set of definitions of traffic-invariant demand indicators, a DTA model as used in 'classical' car traffic problems, and a schedule-based model that considers the arrivals and departures of trains.

Traffic-invariant demand indicators: Several demand indicators can be computed from the demand vector by aggregation.

The total demand is obtained by aggregating over user classes, i.e.,

$$d^{\text{tot}}_{\zeta,\tau} = \sum_{\beta\in\mathcal{B}} d^{\beta}_{\zeta,\tau}. \tag{1}$$

The total cumulative origin and destination flows during the time period $\mathcal{T}$, the total platform departure flow, and the average class split ratios are obtained by aggregating over time and space, i.e.,

$$o^{\text{tot}}_{\nu} = \sum_{\tau\in\mathcal{T}} \sum_{\zeta\in\mathcal{Z}^{\text{orig}}_{\nu}} d^{\text{tot}}_{\zeta,\tau}, \tag{2}$$

$$q^{\text{tot}}_{\nu} = \sum_{\tau\in\mathcal{T}} \sum_{\zeta\in\mathcal{Z}^{\text{dest}}_{\nu}} d^{\text{tot}}_{\zeta,\tau}, \tag{3}$$

$$g^{\text{dep}}_{p} = \sum_{\tau\in\mathcal{T}} \sum_{\nu\in\mathcal{C}_p} \sum_{\zeta\in\mathcal{Z}^{\text{dest}}_{\nu}} \sum_{\beta\in\{\text{out},\text{tr}\}} d^{\beta}_{\zeta,\tau}, \tag{4}$$

$$r^{\beta}_{\nu} = \frac{\sum_{\tau\in\mathcal{T}} \sum_{\zeta\in\mathcal{Z}^{\text{orig}}_{\nu}} d^{\beta}_{\zeta,\tau}}{o^{\text{tot}}_{\nu}}. \tag{5}$$

Dynamic traffic assignment model: The role of the DTA is to estimate the traffic conditions $\boldsymbol{s}$ given a certain demand $\boldsymbol{d}$.

Let $\boldsymbol{\sigma}_\beta(\boldsymbol{d};\boldsymbol{\gamma})$ denote the network supply model associated with user class β, and let the random variable $\boldsymbol{\eta}_s^\beta$ represent the corresponding structural error. The class-specific traffic conditions are then given by

$$\boldsymbol{s}_\beta = \boldsymbol{\sigma}_\beta(\boldsymbol{d}) + \boldsymbol{\eta}_s^\beta. \tag{6}$$

For $\boldsymbol{\eta}_s^\beta$, as well as all other error terms introduced in this section, a zero mean is assumed, i.e., that the corresponding estimator is unbiased (Cascetta et al., 1993).

To obtain link flows, subroute flows and area occupations, two steps are necessary (see e.g. Cascetta and Improta, 2002). First, OD demand is mapped to route flows. The corresponding probabilities are obtained from a route choice model, which is part of the DTA model and generally traffic dependent. It is represented by the class-specific route choice matrix $\boldsymbol{\Delta}_\beta(\boldsymbol{s}_\beta;\boldsymbol{\gamma}) = [\delta^\beta_{(\rho,\tau),(\zeta,\kappa)}]$ of size $|\mathcal{R}||\mathcal{T}| \times |\mathcal{Z}||\mathcal{T}|$. Each of its elements $\delta^\beta_{(\rho,\tau),(\zeta,\kappa)}(\boldsymbol{s}_\beta;\boldsymbol{\gamma})$ represents the probability that a pedestrian associated with user class β, OD pair ζ and departure time interval κ chooses route ρ during time interval τ. Route choice is assumed instantaneous such that generally $\delta^\beta_{(\rho,\tau),(\zeta,\kappa)} = 0$ if $\tau \neq \kappa$.

Second, a dynamic network loading (DNL) model is used to describe the propagation of pedestrians along their routes. The DNL model is also part of the DTA model, and also traffic-dependent. The class-specific mapping from route to link flows, route to subroute flows, as well as from route flows to area occupations, are represented by the assignment matrices $\boldsymbol{A}_\beta(\boldsymbol{s}_\beta;\boldsymbol{\gamma}) = [a^\beta_{(\lambda,\tau),(\rho,\kappa)}]$, $\boldsymbol{B}_\beta(\boldsymbol{s}_\beta;\boldsymbol{\gamma}) = [b^\beta_{(\varrho,\tau),(\rho,\kappa)}]$ and $\boldsymbol{C}_\beta(\boldsymbol{s}_\beta;\boldsymbol{\gamma}) = [c^\beta_{(\alpha,\tau),(\rho,\kappa)}]$ of sizes $|\Lambda||\mathcal{T}| \times |\mathcal{R}||\mathcal{T}|$, $|\mathcal{Q}||\mathcal{T}| \times |\mathcal{R}||\mathcal{T}|$ and $|\mathcal{A}||\mathcal{T}| \times |\mathcal{R}||\mathcal{T}|$, respectively. The entry $a^\beta_{(\lambda,\tau),(\rho,\kappa)}(\boldsymbol{s}_\beta;\boldsymbol{\gamma})$ represents the probability that a pedestrian associated with user class β, route ρ and departure time interval κ reaches link λ during time interval τ; the entry $b^\beta_{(\varrho,\tau),(\rho,\kappa)}(\boldsymbol{s}_\beta;\boldsymbol{\gamma})$ represents the probability that a pedestrian associated with user class β, route ρ and departure time interval κ reaches subroute ϱ during time interval τ, and the entry $c^\beta_{(\alpha,\tau),(\rho,\kappa)}(\boldsymbol{s}_\beta;\boldsymbol{\gamma})$ denotes the 'occupation contribution' of a pedestrian with the same class, route and departure time interval to area α during time interval τ (see Section 4.3).

If $\boldsymbol{\eta}_f^\beta$, $\boldsymbol{\eta}_e^\beta$ and $\boldsymbol{\eta}_n^\beta$ denote the class-specific error terms corresponding to the mappings from OD demand to link flows, subroute flows and area

occupations, the latter can be expressed as

$$\boldsymbol{f}_\beta = \boldsymbol{A}_\beta(\boldsymbol{s}_\beta)\boldsymbol{\Delta}_\beta(\boldsymbol{s}_\beta)\boldsymbol{d}_\beta + \boldsymbol{\eta}_{\mathrm{f}}^\beta, \tag{7}$$

$$\boldsymbol{e}_\beta = \boldsymbol{B}_\beta(\boldsymbol{s}_\beta)\boldsymbol{\Delta}_\beta(\boldsymbol{s}_\beta)\boldsymbol{d}_\beta + \boldsymbol{\eta}_{\mathrm{e}}^\beta \tag{8}$$

and

$$\boldsymbol{n}_\beta = \boldsymbol{C}_\beta(\boldsymbol{s}_\beta)\boldsymbol{\Delta}_\beta(\boldsymbol{s}_\beta)\boldsymbol{d}_\beta + \boldsymbol{\eta}_{\mathrm{n}}^\beta, \tag{9}$$

respectively. Note that Eq. (7), Eq. (8) and Eq. (9) consider the probabilities defined by the assignment models $\boldsymbol{A}_\beta(\boldsymbol{s}_\beta)\boldsymbol{\Delta}_\beta(\boldsymbol{s}_\beta)$, $\mathrm{B}_\beta(\boldsymbol{s}_\beta)\boldsymbol{\Delta}_\beta(\boldsymbol{s}_\beta)$ and $\boldsymbol{C}_\beta(\boldsymbol{s}_\beta)\boldsymbol{\Delta}_\beta(\boldsymbol{s}_\beta)$ as proportions. This is feasible since the demand estimation problem is considered at the aggregate level.

If the dependency on prevailing traffic conditions in Eq. (7), Eq. (8) and Eq. (9) is neglected, the relationship between demand and the derived indicators becomes linear. This implies that the traffic situation is independent of demand, which is generally the case for an uncongested network. Alternatively, if the traffic situation is known a priori e.g. through direct measurements, an estimate of the assignment maps may also be obtained without considering the demand.

If a network is congested and link costs 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 (see Section 3.5).

Based on Eq. (7), Eq. (8) and Eq. (9), the total cumulative link flow, the total cumulative subroute flow, the cumulative arrival link flow and the total area occupation can be defined as

$$\mathrm{f}_{\lambda,\tau}^{\mathsf{tot}} = \sum_{\beta\in\mathcal{B}} \mathrm{f}_{\lambda,\tau}^{\beta}, \tag{10}$$

$$\mathrm{f}_{\lambda,\tau}^{\mathsf{arr}} = \sum_{\beta\in\{\mathrm{in},\mathrm{tr}\}} \mathrm{f}_{\lambda,\tau}^{\beta}, \tag{11}$$

$$e_{\varrho,\tau}^{\mathsf{tot}} = \sum_{\beta\in\mathcal{B}} e_{\varrho,\tau}^{\beta}, \tag{12}$$

$$\mathrm{n}_{\alpha,\tau}^{\mathsf{tot}} = \sum_{\beta\in\mathcal{B}} \mathrm{n}_{\alpha,\tau}^{\beta}. \tag{13}$$

Schedule-based model: The schedule-based model establishes a relationship between OD demand and train exchange volumes.

Alighting volumes are associated with inbound and transfer passengers by means of a class-homogeneous assignment matrix $\boldsymbol{H}(\boldsymbol{t}_{\text{arr}}, \boldsymbol{t}_{\text{dep}}) = [h_{m,(\zeta,\tau)}]$ and a corresponding error ε_{off} such that

$$\boldsymbol{w}_{\text{off}} = \boldsymbol{H}\left(\boldsymbol{d}_{\text{in}} + \boldsymbol{d}_{\text{tr}}\right) + \varepsilon_{\text{off}}. \tag{14}$$

The element $h_{m,(\zeta,\tau)}(t_m^{\text{arr}}, t_m^{\text{dep}})$ quantifies the degree to which demand emanating from the origin v_ζ^o of OD pair ζ during time interval τ is associated with the alighting volume of train m idling on platform p_m during the time period $[t_m^{\text{arr}}, t_m^{\text{dep}}]$. For instance, if at most one train is served by a platform at a time, and if a homogeneous distribution of demand within a time interval is assumed, the entries of the assignment matrix $\boldsymbol{H}$ are given by

$$h_{m,(\zeta,\tau)} = \begin{cases} \left|\left[t_m^{\text{arr}}, t_m^{\text{dep}}\right] \cap \tau\right| / |\tau| & \text{if } v_\zeta^o \in \mathcal{C}_{p_m}, \\ 0 & \text{otherwise,} \end{cases} \tag{15}$$

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

In principle, an analogous approach can be used to consider boarding volumes. However, as mentioned in the literature review, it is difficult to find a meaningful specification of the corresponding assignment matrix. Alternatively, boarding volumes may be considered in terms of cumulative platform departure flows. If $\boldsymbol{\varpi}_\chi$ represents a vector containing structural errors, the vector of cumulative platform departure flows can be expressed as

$$\boldsymbol{g}_{\text{dep}} = \boldsymbol{\chi}(\boldsymbol{w}_{\text{on}}) + \boldsymbol{\varpi}_\chi, \tag{16}$$

where $\boldsymbol{\chi} = [\chi_p]$ denotes an aggregation function defined by

$$\chi_p(\boldsymbol{w}_{\text{on}}) = \sum_{m \in \mathcal{M}_p} w_m^{\text{on}}. \tag{17}$$

Eq. (14) and Eq. (16) are useful to relate train exchange volumes to OD demand, but they provide little information about their distribution across time intervals if the train idling times cover several time intervals.

Instead, such information may be obtained from empirical relations, which is discussed at the example of train-induced arrival flows. We denote by $\phi_{\lambda,\tau}(\boldsymbol{w}_{\text{arr}})$ an empirical model predicting the flow of pedestrians on link λ during time interval τ that have alighted from a train. If $\boldsymbol{\varphi}(\boldsymbol{w}_{\text{off}}, \boldsymbol{s}; \boldsymbol{\gamma}) =$

$[\phi_{\lambda,\tau}]$ represents the corresponding time-space expanded vector and ϖ_{ϖ} a structural error, it holds that

$$\boldsymbol{f}_{\text{arr}} = \boldsymbol{\varphi}(\boldsymbol{w}_{\text{off}}, \boldsymbol{s}) + \boldsymbol{\varpi}_{\varphi}. \tag{18}$$

3.4 Measurement model

The measurement model links the structural model to a priori information and measurements. For each data source, a random error term $\boldsymbol{\omega}_{(\cdot)}$ takes into account the uncertainty that the data is afflicted with, and the aforementioned reduction matrices $\boldsymbol{R}_{(\cdot)}$ account for the incomplete coverage of the data collection infrastructure, i.e.,

$$\hat{\boldsymbol{w}}' = \boldsymbol{R}_{w}\boldsymbol{w} + \boldsymbol{\omega}'_{\boldsymbol{w}}, \tag{19}$$
$$\hat{\boldsymbol{f}}'_{\text{tot}} = \boldsymbol{R}_{\text{f}}\boldsymbol{f}_{\text{tot}} + \boldsymbol{\omega}'_{\boldsymbol{f}}, \tag{20}$$
$$\hat{\boldsymbol{e}}'_{\text{tot}} = \boldsymbol{R}_{\text{e}}\boldsymbol{e}_{\text{tot}} + \boldsymbol{\omega}'_{\boldsymbol{e}}, \tag{21}$$
$$\hat{\boldsymbol{n}}'_{\text{tot}} = \boldsymbol{R}_{\text{n}}\boldsymbol{n}_{\text{tot}} + \boldsymbol{\omega}'_{\boldsymbol{n}}, \tag{22}$$
$$\hat{\boldsymbol{o}}'_{\text{tot}} = \boldsymbol{R}_{\text{o}}\boldsymbol{o}_{\text{tot}} + \boldsymbol{\omega}'_{\boldsymbol{o}}, \tag{23}$$
$$\hat{\boldsymbol{q}}'_{\text{tot}} = \boldsymbol{R}_{\text{q}}\boldsymbol{q}_{\text{tot}} + \boldsymbol{\omega}'_{\boldsymbol{q}}, \tag{24}$$
$$\hat{\boldsymbol{r}}' = \boldsymbol{R}_{\text{r}}\boldsymbol{r} + \boldsymbol{\omega}'_{\boldsymbol{r}}, \tag{25}$$
$$\hat{\boldsymbol{d}} = \boldsymbol{d} + \boldsymbol{\omega}_{\boldsymbol{d}}. \tag{26}$$

3.5 Estimation problem

The estimation problem consists in finding the distribution of the OD demand volumes $\boldsymbol{d}^{\star}$ such that (i) actual observations of demand indicators are reproduced at best, (ii) train-induced arrival and cumulative platform departure flows are 'most consistent' with empirical predictions based on the train schedule, 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 (Hazelton, 2000), and depends critically on the assumptions that are made regarding the distribution of the error terms.

Alternatively, if any 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 $\text{dist}_{\text{hist}}\langle\cdot\rangle$. In practice, such an assumption is often made due to the difficulty of estimating the correlation structure and the high cost involved in solving the full problem (Cascetta and Improta, 2002). The estimation problem can then be formulated as

$$\begin{aligned} \boldsymbol{d}^{\star}_{\gamma} = \underset{\boldsymbol{d}\geq 0}{\arg\min}\ \text{dist}_{\text{obs}} \left\langle \begin{pmatrix} \hat{\boldsymbol{w}}' \\ \hat{\boldsymbol{f}}'_{\text{tot}} \\ \hat{\boldsymbol{e}}'_{\text{tot}} \\ \hat{\boldsymbol{n}}'_{\text{tot}} \\ \hat{\boldsymbol{o}}'_{\text{tot}} \\ \hat{\boldsymbol{q}}'_{\text{tot}} \\ \hat{\boldsymbol{r}}' \end{pmatrix}, \begin{pmatrix} \boldsymbol{w}' \\ \boldsymbol{f}'_{\text{tot}} \\ \boldsymbol{e}'_{\text{tot}} \\ \boldsymbol{n}'_{\text{tot}} \\ \boldsymbol{o}'_{\text{tot}} \\ \boldsymbol{q}'_{\text{tot}} \\ \boldsymbol{r}' \end{pmatrix} \right\rangle + \\ \text{dist}_{\text{sched}} \left\langle \begin{pmatrix} \boldsymbol{\varphi}' \\ \boldsymbol{\chi}' \end{pmatrix}, \begin{pmatrix} \boldsymbol{f}'_{\text{arr}} \\ \boldsymbol{g}'_{\text{dep}} \end{pmatrix} \right\rangle + \text{dist}_{\text{hist}} \left\langle \hat{\boldsymbol{d}}, \boldsymbol{d} \right\rangle . \end{aligned} \tag{27}$$

When solving Eq. (27), it is critical not to rely on point estimates. The demand vector $\boldsymbol{d}^{\star}$ is generally distributed, and may follow a complex distribution that is insufficiently described by a single value such as its mean. Its distribution depends both on the variation of input variables, which may be distributed themselves, and on the uncertainty involved in terms of modeling and measurement errors. To approximate the distribution of $\boldsymbol{d}^{\star}$, Monte Carlo sampling can be used.

We close the presentation of the model formulation with three comments:

First, Eq. (27) could be generalized to allow for the estimation of the parameter vector γ, i.e., the a priori information and measurements could be used to 'improve' the estimation of model parameters. Presumably due to the difficult implementation and high computational cost, such applications have not been pursued beyond the stage of formal specification (Cascetta and Improta, 2002).

Second, Eq. (7), Eq. (11) and Eq. (18) establish a relationship between the demand vector $\boldsymbol{d}$ and the train exchange volumes $\boldsymbol{w}$ that is generally non-linear. This can make solving the problem (27) hard. If the a priori estimate of train exchange volumes $\hat{\boldsymbol{w}}$ is accurate, the corresponding vector $\boldsymbol{w}$ may be assumed known a priori, which reduces the computational complexity.

Third, an interaction between demand and supply incurs a dependency of the assignment maps $\boldsymbol{A}_{\beta}\boldsymbol{\Delta}_{\beta}$, $\boldsymbol{B}_{\beta}\boldsymbol{\Delta}_{\beta}$ and $\boldsymbol{C}_{\beta}\boldsymbol{\Delta}_{\beta}$ on prevailing traffic con-

ditions, and thus indirectly on demand. A fixed-point between the network supply model, Eq. (6), and the demand estimation model, Eq. (27), arises. To solve such a problem, typically iterative algorithms are used that alternately update the state variables of the demand and of the supply model (Cascetta and Postorino, 2001; Bierlaire and Crittin, 2006).

If any interaction between demand and supply is neglected, and if the train exchange volumes are assumed to be known a priori, Eq. (27) reduces to a constrained, linear estimation problem for most of the commonly used distance measures. The resulting problem can be solved e.g. with the projected gradient algorithm (Cascetta and Improta, 2002). Such an approach is pursued in the subsequent model specification.

4 Case study

The previous section has left the functional form of the structural and measurement model undefined. To demonstrate the applicability of the proposed framework, a case study of Lausanne railway station is carried out.

4.1 Description

Lausanne railway station is the largest train station in French-speaking Switzerland, serving approximately 120'000 passengers with about 650 arriving and departing trains every weekday (Amacker, 2012). Located at the junction of three national lines, it provides express train service to a variety of destinations across Switzerland and beyond. It also provides access to suburban transportation, notably through a local metro system that can be reached across the train station square, and by means of several bus lines. Fig. 2 shows a schematic map of Lausanne railway station.

The station encompasses the passing tracks #1–9 and the dead end track #70. Track #2 is used by freight trains and through traffic only, as it is not accessible by any platform. Except for platforms #1 and #70, all platforms are accessible from the city solely through two pedestrian underpasses (PU), PU West and PU East. Furthermore, platform #9 is only accessibly from PU West. Longitudinally, the train station is divided into sectors A-D, where the historical ordering from East to West is adopted. The blue graph in Fig. 2 shows the corresponding walking network.

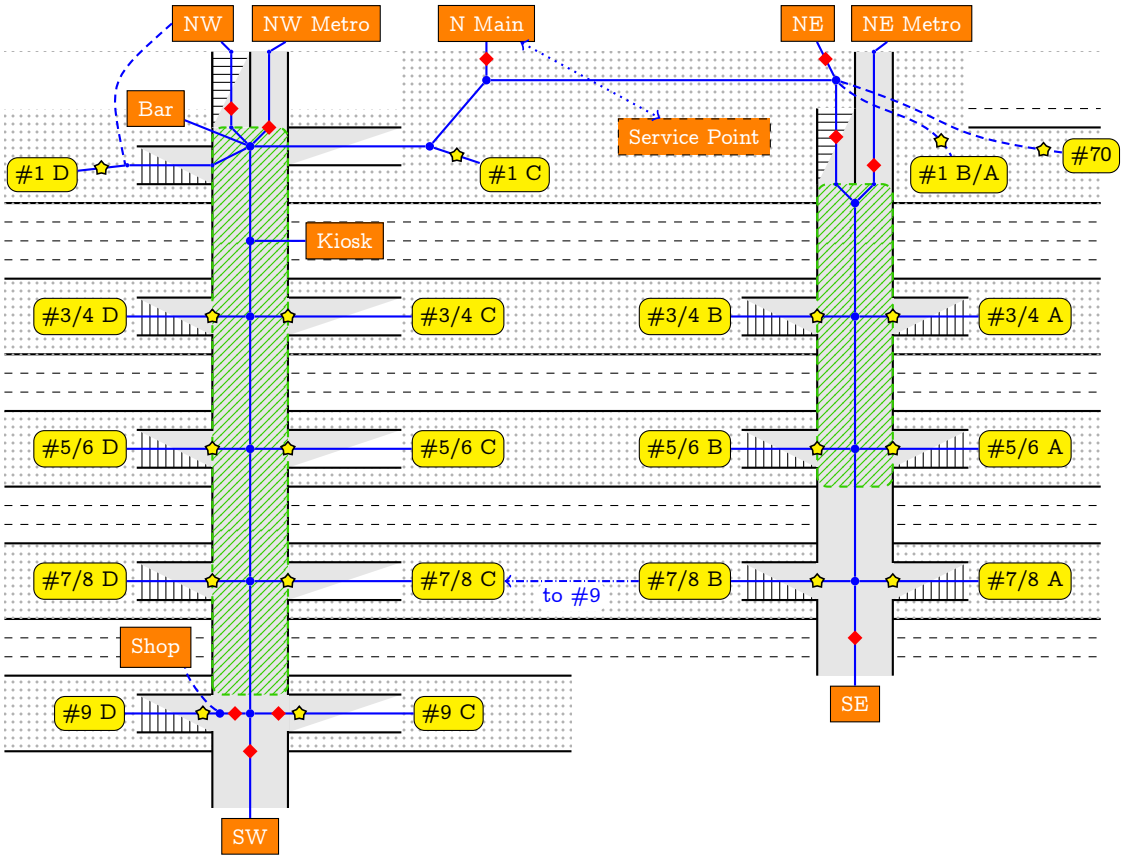


Figure 2: Schematic map of Lausanne railway station, encompassing ten tracks (#1–#9, #70) that are served by platforms #1, #3/4, #5/6, #7/8, #9 and #70. Platforms are connected by two pedestrian underpasses (PU) referred to as PU West and PU East, each partially covered by a pedestrian tracking system (corresponding areas are shaded in green). Dashed lines represent network links that cannot be directly shown on the scheme due to the chosen two-dimensional representation. For a description of the various symbols, please see Fig. 1. Note that all links and flow sensors are directed.

The demand peak over a workday in Lausanne is reached at around 07:45 when several long distance trains arrive and depart in close succession (Gendre and Zulauf, 2010). At this time of the day, more than 500 incoming users alight during a peak minute, whereas a few instants later it can be less than a hundred per minute (Alahi et al., 2013a). Such a periodical concentration of pedestrians is characteristic for the Swiss railway network that aims at bundling train arrivals and departures in major railway stations in order to minimize waiting time for transfer passengers (SBB-Infrastruktur, 2013).

Within-day and day-to-day variation in demand during the period between 07:30 and 08:00 is considered. The chosen temporal aggregation is 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, representing Tuesdays, Wednesdays or Thursdays, have been selected by SBB based on the punctuality of trains and the availability of data sources listed in the following.

Train timetable and train frequentation data: During the time period of interest, a total of 25 trains stop at Lausanne railway station (train timetable not shown). For all trains, the assigned track, and the actual arrival and departure times are known. From ticket sales data, within-train surveys, and infrared-based counts at train doors, an estimate[1] of the boarding and alighting volumes is available (SBB-Personenverkehr, 2011; Olesen, 2006). These estimates date back to the year 2010 and are of varying quality. For use in this study, they have been increased by 15% based on the growth rate recommended by SBB (Gendre and Zulauf, 2010). All boarding and alighting volumes are considered as random normal variables with a standard deviation equal to 19.2% of their mean (Molyneaux et al., 2014).

Trajectory recordings: For the two pedestrian underpasses, individual pedestrian trajectories are available (Lavadinho et al., 2013). These recordings have been obtained by means of a tracking system consisting of ca. 60 visual, depth and infrared sensors (Alahi et al., 2011, 2013b). In spite of this large number of sensors, the resulting trajectories are on average 60%

[1]The actual figures cannot be revealed due to a non-disclosure agreement.

interpolated (Babel, 2014). This is in part due to occlusion, and due to the fact that the sensors only cover approximately 75% of the total walking area in the PUs. For the purposes of this study, this level of interpolation is negligible, as demand is studied at the aggregate level. Critically, all entrance and exit areas of PUs are fully covered by typically three sensors, such that the estimates of total in- and outflows in the pedestrian underpasses are considered accurate.

Link flow observations: For ten links of the pedestrian walking network, directional minute-by-minute counts are available. The surveyed links are marked by red diamonds in Fig. 2.

By comparing the link flow observations to the trajectory recordings described above, the count sensors are found to reach saturation at high flows, i.e., they underestimate the actual throughput if pedestrian density is high. A correction function of the form $f(x) = ax + bx^2$ with $a = 1.065$ (0.950, 1.179) and $b = 0.005515$ (0.00276, 0.00827; numbers in brackets represent 95% confidence bounds) is thus applied a posteriori. For further details on the correction of pedestrian counts, see e.g. Ganansia et al. (2014).

Sales data and historical information: There are three sales points located in PU West, for which the average number of monthly customer visits in 2013 is available (footnote 1 on page 23 applies). The corresponding number of visits during the morning peak hour is estimated by assuming that the customer frequentation is proportional to the overall occupation of the train station, i.e., that 10% of all daily sales are achieved within the morning peak hour (Lavadinho et al., 2013). There are further restaurants and sales points in the train station building, represented by a generic 'service point' in Fig. 2. For this node, no customer frequentation data is available, and it is not considered in the demand analysis.

Besides sales data, some information on user class split ratios is available. According to Anken et al. (2012), the fraction of inbound passengers among alighting train-users is estimated at $r_v^{in} = 91.4\% \pm 4.6\%$, $\forall v \in \mathcal{C}_\mathcal{P}$. Benmoussa et al. (2011) and Lavadinho et al. (2013) report for the fraction of outbound passengers among pedestrians entering the train station from the city a value of $r_v^{out} \approx 95\%$, $\forall v \notin \mathcal{C}_\mathcal{P}$. These estimates date back to the year 2010 and do not specifically apply to the morning peak hour.

4.2 Assumptions of model specification

To apply the estimation framework presented in Section 3, the following assumptions are made:

(i) The considered pedestrian network covers only facilities in which walking is the predominant activity. Pedestrian behavior on platforms is not explicitly modeled. Instead, platform sectors are represented by centroids that are connected to the walking network (see space representation of platform #3/4 in Fig. 1). Furthermore, differences in walking behavior across user classes are considered negligible, yielding a class-homogeneous supply model.

(ii) The level of congestion is low to moderate (LOS E or better, Highway Capacity Manual, 2000, Exhibit 18-3), such that the network supply model is demand-invariant, and such that schedule-based link flows associated with different trains are independent. A detailed analysis of pedestrian trajectories has not revealed any significant evidence of interaction between demand and supply, justifying this assumption in the case of Lausanne railway station (Hänseler et al., 2013).

(iii) The train alighting volumes are a priori known and do not need to be estimated (see closing comment at the end of Section 3).

(iv) Pedestrians are assumed to follow the shortest path, of which there exists at most one per OD pair. During peak periods, in which regular commuters with a good knowledge of the railway station constitute the main user group, this is a valid assumption (Lavadinho, 2012). Besides obviating the need for a route choice model, this assumption implies that all pedestrians leaving a platform are either inbound or transfer passengers (as opposed to e.g. 'train spotters' or some lost travelers).

(v) All error terms follow a univariate normal distribution with zero mean and are independent from each other. This represents a common assumption in practice, both because it yields accurate results and because the resulting problem is relatively easy to estimate (Cascetta and Improta, 2002).

(vi) Multi-destination trips are considered as multiple, independent single-destination trips. For instance, an incoming train passenger that stops at a shop before heading for the city is represented by two independent OD trips, namely 'platform → shop' and 'shop → entrance/exit'. Moreover, it is assumed that each OD pair is associated with exactly one user class, reducing the multi-class estimation problem to a single-class problem. An OD pair $\zeta \in \mathcal{Z}$ is associated with the class of (i) inbound passengers if $\nu_\zeta^o \in \mathcal{C}_\mathcal{P}$ and $\nu_\zeta^d \notin \mathcal{C}_\mathcal{P}$, (ii) outbound passengers if $\nu_\zeta^o \notin \mathcal{C}_\mathcal{P}$ and $\nu_\zeta^d \in \mathcal{C}_\mathcal{P}$, (iii) transfer passengers if $\nu_\zeta^o, \nu_\zeta^d \in \mathcal{C}_\mathcal{P}$, and (iv) local users if $\nu_\zeta^o, \nu_\zeta^d \notin \mathcal{C}_\mathcal{P}$.

4.3 Dynamic traffic assignment model

Based on the assumption of a single route per OD pair, $|\mathcal{R}_\zeta| = 1$, $\forall\, \zeta \in \mathcal{Z}$, the route choice fractions are given by

$$\delta_{(\rho,\tau),(\zeta,\kappa)} = \begin{cases} 1 & \text{if } \rho \in \mathcal{R}_\zeta,\ \tau = \kappa, \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

This implies that the corresponding route choice matrix $\boldsymbol{\Delta}$ is equal to the identity matrix of size $|\mathcal{R}||\mathcal{T}| \times |\mathcal{Z}||\mathcal{T}|$ (with $|\mathcal{R}| = |\mathcal{Z}|$).

To describe the propagation of pedestrians along routes, the walking speed distribution recommended by SBB's dimensioning guidelines (Weidmann, 1992; Buchmüller and Weidmann, 2008)

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

is used. The corresponding probability density and cumulative distribution functions are denoted by $f_v(v)$ and $F_v(v)$, respectively.

Expression (29) has been obtained indirectly from observations of travel times and traveled distances on horizontal walking areas. Here, it is used in the inverse way, i.e., to estimate the distribution of travel times given a distance.

On inclined walking areas or stairways, the velocity of pedestrians differs from that observed on even areas. According to Weidmann (1992), the horizontal speed on stairways averages to 0.61 m/s for pedestrians walking upward, whereas an average of 0.694 m/s is reported for pedestrians walking downward. For ramps with an inclination of 15%, the corresponding

velocities are estimated at 1.07 and 1.40 m/s. It is assumed that the standard deviation of the distribution changes proportionally with the change in average speed. In case of such uneven surfaces, Eq. (29) is modified accordingly.

The assignment fractions for the cumulative link flows, cumulative subroute flows and area occupations can be derived as follows. 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 k for any route during a time interval κ is given by

$$h_\kappa(k) = \begin{cases} \frac{1}{\Delta t} & \text{if } k \in \kappa, \\ 0 & \text{otherwise.} \end{cases} \tag{30}$$

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

$$\begin{aligned} a_{(\lambda,\tau),(\rho,\kappa)} &= \Pr(k \in \kappa, t \in \tau | \rho, \lambda) \\ &= \Pr\left(k \in \kappa, \nu \in \left[\frac{\ell_\rho^\lambda}{t_\tau^+ - k}, \frac{\ell_\rho^\lambda}{t_\tau^- - k}\right]\right), \end{aligned} \tag{31}$$

where k and t represent the (continuous) departure and arrival time, respectively. For the most common case that $\ell_\rho^\lambda > 0$ and $\tau > \kappa$, we obtain

$$\begin{aligned} a_{(\lambda,\tau),(\rho,\kappa)} &= \int_{k=k^-}^{k^+} \int_{\nu=\ell_\rho^\lambda/(t_\tau^+ - k)}^{\ell_\rho^\lambda/(t_\tau^- - k)} f_\nu(\nu) g_\kappa(k) \, d\nu \, dk \\ &= \frac{1}{\Delta t} \int_{k=k^-}^{k^+} F_\nu\left(\frac{\ell_\rho^\lambda}{t_\tau^- - k}\right) - F_\nu\left(\frac{\ell_\rho^\lambda}{t_\tau^+ - k}\right) dk. \end{aligned} \tag{32}$$

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

$$\begin{aligned} a_{(\lambda,\tau),(\rho,\tau)} &= 1 - \Pr\left(k \in \tau, t \notin \tau | \rho, \lambda\right) \\ &= 1 - \Pr\left(k \in \tau, \nu \in \left[0, \frac{\ell_\rho^\lambda}{t_\tau^+ - k}\right]\right) \\ &= 1 - \frac{1}{\Delta t} \int_{k=t_\tau^-}^{t_\tau^+} F_\nu\left(\frac{\ell_\rho^\lambda}{t_\tau^+ - k}\right) - F_\nu(0) \, dk. \end{aligned} \tag{33}$$

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

$$a_{(\lambda,\tau),(\rho,\kappa)} = \begin{cases} 0 & \text{if } \ell_\rho^\lambda = 0,\ \kappa < \tau, \\ 1 & \text{if } \ell_\rho^\lambda = 0,\ \kappa = \tau, \\ \text{Eq. (32)} & \text{if } \ell_\rho^\lambda > 0,\ \kappa < \tau, \\ \text{Eq. (33)} & \text{if } \ell_\rho^\lambda > 0,\ \kappa = \tau. \end{cases} \tag{34}$$

Regarding subroute flows, the probability that a pedestrian associated with route ρ and departure time interval κ reaches subroute ϱ during time interval τ can be expressed as

$$b_{(\varrho,\tau),(\rho,\kappa)} = \begin{cases} a_{(\lambda_\varrho^\varrho,\tau),(\rho,\kappa)} & \text{if } \rho \in \mathcal{R}_\varrho^{\text{sup}}, \\ 0 & \text{otherwise.} \end{cases} \tag{35}$$

The assignment fraction for area occupations 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

$$\sigma(v, \ell_{\text{in}}^{\rho,\alpha}, \ell_{\text{out}}^{\rho,\alpha}, t^-, t^+) = \begin{cases} t^+ - \ell_{\text{in}}^{\rho,\alpha}/v & \text{if } t^- \le \ell_{\text{in}}^{\rho,\alpha}/v \le t^+ \le \ell_{\text{out}}^{\rho,\alpha}/v, \\ \ell_{\text{out}}^{\rho,\alpha}/v - t^- & \text{if } \ell_{\text{in}}^{\rho,\alpha}/v \le t^- \le \ell_{\text{out}}^{\rho,\alpha}/v \le t^+, \\ t^+ - t^- & \text{if } \ell_{\text{in}}^{\rho,\alpha}/v \le t^- \le t^+ \le \ell_{\text{out}}^{\rho,\alpha}/v, \\ (\ell_{\text{out}}^{\rho,\alpha} - \ell_{\text{in}}^{\rho,\alpha})/v & \text{if } t^- \le \ell_{\text{in}}^{\rho,\alpha}/v \le \ell_{\text{out}}^{\rho,\alpha}/v \le t^+, \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

In Eq. (36), 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. (36), the 'occupation contribution' of a pedestrian traveling along route ρ with departure time interval κ in area α during time interval τ is given by

$$\begin{aligned} c_{(\alpha,\tau),(\rho,\kappa)} &= \int_{t=t_\kappa^-}^{t_\kappa^+} \int_{v=0}^{\infty} \frac{\sigma(v, \ell_{in}^{\rho,\alpha}, \ell_{out}^{\rho,\alpha}, t_\tau^- - t, t_\tau^+ - t)}{\Delta t} f_v(v) h_\kappa(t)\, dv\, dt \\ &= \frac{1}{\Delta t^2} \int_{v=0}^{\infty} f_v(v) \int_{t=t_\kappa^-}^{t_\kappa^+} \sigma(v, \ell_{in}^{\rho,\alpha}, \ell_{out}^{\rho,\alpha}, t_\tau^- - t, t_\tau^+ - t)\, dt\, dv. \end{aligned} \tag{37}$$

For an efficient implementation, we note that the assignment fractions (34) and (37) are time-invariant, i.e., for $\tau' = \tau - \kappa$ it holds that

$$a_{(\lambda,\tau),(\rho,\kappa)} = a_{(\lambda,\tau'),(\rho,0)} \text{ and } c_{(\alpha,\tau),(\rho,\kappa)} = c_{(\alpha,\tau'),(\rho,0)}. \tag{38}$$

To further reduce the computational cost involved in computing Eq. (34) and Eq. (37), a maximum travel time TT_{max} is defined. If $\tau' \geq TT_{max}$, it is assumed that $a_{(\lambda,\tau'),(\rho,0)} = 0\ \forall \lambda, \rho$ and $c_{(\alpha,\tau'),(\rho,0)} = 0\ \forall \alpha, \rho$. The threshold TT_{max} is chosen such that the error incurred by this numerical approximation is negligible. For the present case study, a threshold value of $TT_{max} = 10$ min has been found sufficient.

4.4 Schedule-based link flow model

Assumption (iii) in Section 4.2 allows to simplify the schedule-based link flow model in several ways. It obviates the need for Eq. (14), and for a specification of the assignment matrix $\boldsymbol{H}$ used therein. Also, it allows to 'pre-compute' the vector of cumulative platform departure flows and the vector of train-induced arrival flows, denoted by $\hat{\chi}$ and $\hat{\varphi}'$, respectively. The former can be obtained by aggregation over boarding volumes through Eq. (17). The 'pre-computation' of the latter is discussed in the following.

In line with assumption (iv), on links representing platform exit ways, it holds that $f_{\lambda,\tau}^{tot} = f_{\lambda,\tau}^{arr}$, $\forall\, \lambda \in \Lambda_{\mathcal{P}}^{arr}$, $\forall\, \tau \in \mathcal{T}$. Furthermore, assumption (ii) allows to decompose the train-induced arrival flow into a set of independent contributions of individual trains.

Assume that for a train m, the arrival flow rate at continuous time t at link λ is denoted by $\tilde{\phi}_{m,\lambda}(t; w_m^{off}, t_m^{arr}, \gamma)$. The estimated cumulative

train-induced arrival flow during time interval τ at link λ is then given by

$$\hat{\phi}_{\lambda,\tau} = \int_{t=t_\tau^-}^{t_\tau^+} \sum_{m \in \mathcal{M}} \tilde{\phi}_{m,\lambda}(t; \hat{w}_m^{\text{off}}, t_m^{\text{arr}}, \boldsymbol{\gamma})\, dt. \tag{39}$$

In two related studies, Benmoussa et al. (2011) and Lavadinho (2012) propose an approximative, piecewise linear specification of the continuous-time flow model $\tilde{\phi}_{m,\lambda}$. Let the variable $Z_p(w^{\text{off}})$ denote the total exit flow rate associated with platform p if a train with alighting volume w^{off} has arrived, and let s_p be the dead time representing the delay between the arrival of the train and the onset of flow at a given set of sensors on each link $\lambda \in \Lambda_p^{\text{arr}}$. This dead time takes into account the walking time between train doors and the sensors under free-flow conditions, a potential lag between the arrival of a train and the time at which train doors open, and random delays. Two exemplary sensor locations are shown for platform #3/4 in Fig. 1, which are indicated by yellow stars and assumed to be approximately at the same distance from the platform.

Assuming that the total exit flow rate of platform p_m is shared according to link-specific split fractions $\xi_{m,\lambda}$ with $\sum_{\lambda \in \Lambda_{p_m}^{\text{arr}}} \xi_{m,\lambda} = 1$, the flow rate on link λ associated with train m is given by

$$\tilde{\phi}_{m,\lambda}(t) = \begin{cases} \xi_{m,\lambda} Z_m & t \in \left(t_m^{\text{arr}} + s_{p_m}, t_m^{\text{arr}} + s_{p_m} + w_m^{\text{off}}/Z_m\right), \\ 0 & \text{otherwise,} \end{cases} \tag{40}$$

where $Z_m = Z_{p_m}(w_m^{\text{off}})$. Fig. 3 illustrates the cumulative arrival flow associated with Eq. (40).

In the following, a calibration of Eq. (40) is discussed at the example of Lausanne railway station. We hereby focus on the methodological aspects, and refer the reader to Molyneaux et al. (2014) for numerical details.

The dead time s_p is modeled as a normally distributed random variable. The dependency of the total platform exit flow rate Z_p on the alighting volume w^{off} is assumed to be linear at low values of w^{off}, and to reach saturation at a platform-specific threshold w_p^{crit}. If σ_p^Z, a_p^Z and b_p^Z represent shape parameters, the total exit flow rate on platform p is given by the stochastic model

$$Z_p(w^{\text{off}}) = Z_p^{\text{det}}(w^{\text{off}}) + \mathcal{N}(0, \sigma_p^Z), \tag{41}$$

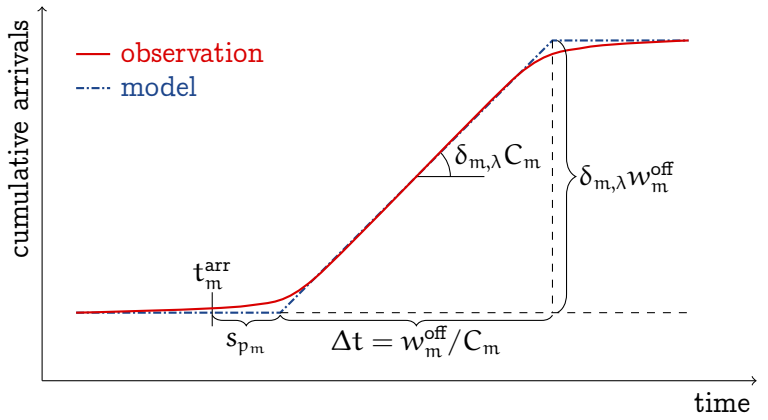


Figure 3: Illustration of a continuous-time, piecewise linear model, Eq. (40), for train-induced arrival flows associated with train m on link λ. The red curve illustrates the actual flow, and the dash-dotted blue curve the piecewise linear approximation.

where the deterministic part of the flow rate is specified as

$$Z_p^{det}(w^{off}) = \begin{cases} a_p^Z w + b_p^Z & \text{if } w \leq w_p^{crit}, \\ a_p^Z w_p^{crit} + b_p^Z & \text{otherwise.} \end{cases} \tag{42}$$

Fig. 4 illustrates Eq. (41) at the example of flow rates recorded on platform #3/4 in Lausanne railway station. Two observations may be made. First, the length of a train N_c, measured in number of passenger cars, does not have a significant influence on the dependency between flow rate and alighting volume. This is explicitly pointed out since the train length is shown below to have a considerable influence on the split fractions. Second, the flow rates are relatively high, such that the total duration of the platform exit flow typically amounts to less than one minute (up to an alighting volume of 333 passengers), and has not been observed to exceed 2 min. Obviously, these values are case-specific and may not hold for other platforms or other train stations.

In Fig. 5, Eq. (40) has been applied to estimate the train-induced passenger arrival flow on platform #5/6 that would be expected for the recorded train timetable of April 10, 2013 (a qualitatively equivalent figure results for most other days). In this example, the alighting volume w_m^{off} of each train m has been inferred from the historical train frequentation data mentioned

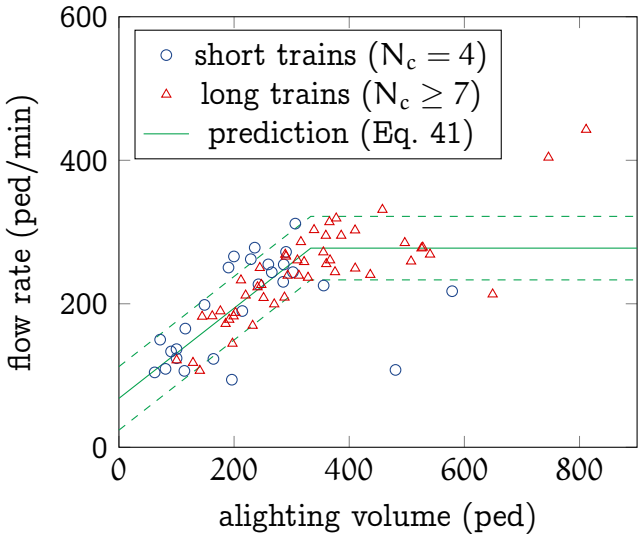


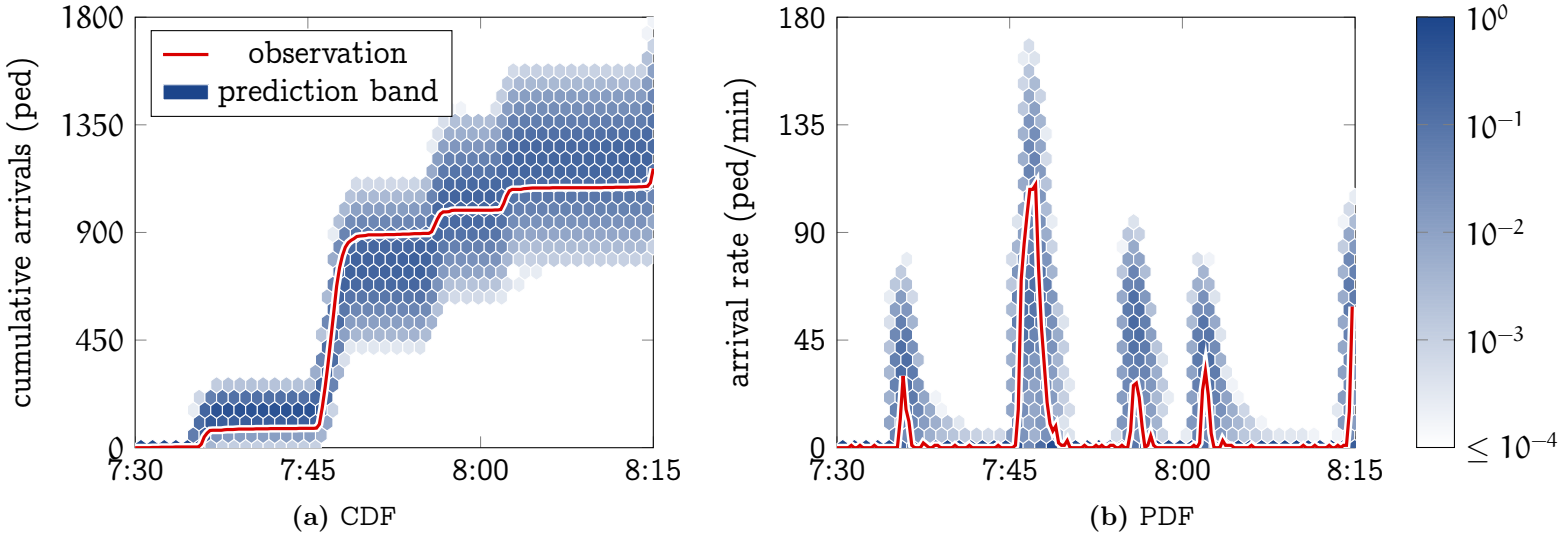
Figure 4: Dependence of total platform exit flow rate Z_p on the alighting volume w^{off} at the example of platform #3/4 in Lausanne railway station. At low volumes, the flow rate increases linearly until a threshold is reached, beyond which the flow rate remains constant. The solid curve denotes the predicted flow rate according to Eq. (41), and the dashed lines the width of the prediction band in terms of $\pm$ one standard deviation.

in Section 4.1. A logarithmic probability density plot shows the expected cumulative arrivals as well as the arrival rate as a function of time. For comparison, the observed flow on that day is shown.

The day-to-day stochasticity of the alighting volumes is high, manifesting itself in a wide prediction band. To obtain the latter in Fig. 5, 7500 Monte Carlo samplings of Eq. (5) are conducted. All parameters have been estimated based on data from platform #3/4, i.e., a clear distinction is made between data used for calibration and validation.

The link-specific split fractions $\xi_{m,\lambda}$ 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. 6 shows measurements of these train-specific split fractions as observed on platform #3/4. The results are grouped by train length and ordered by alighting volumes.

For short trains with $N_c = 4$, mostly the interior platform exit ways in sectors B and C as defined in Fig. 2 are used. This is particularly pronounced if the alighting volume is low. For larger trains with $N_c \geq 7$,



(a) CDF (b) PDF

Figure 5: Comparison of train-induced passenger arrival flow as measured on April 10, 2013 between 7:30 and 8:15, and as estimated using a probabilistic model, Eq. (39) with specification (40), on platform #5/6 in Lausanne railway station.

the lateral exit ways absorb a larger share, and the influence of the alighting volume is smaller.

In the framework of this study, two different specifications of the split fractions for short trains ($N_c = 4$) and long trains ($N_c \geq 7$) are considered. For each case, a multivariate normal distribution is developed, from which the train- and link-specific split fractions $\xi_{m,\lambda}$ can be drawn.

While in this section most model parameters are calibrated from direct measurements, there are also ways of estimating them in case no such data is available. Molyneaux et al. (2014) provide a few examples that rely on earlier studies by Weidmann (1992) and Buchmüller and Weidmann (2008).

4.5 Estimation

Based on assumptions (ii) and (iii) stated in Section 4.2, the estimation problem (27) becomes linear in $\boldsymbol{d}$. The assumption (v) turns it into a constrained, generalized least squares (GLS) problem both in the context of maximum likelihood and Bayesian estimation (Cascetta et al., 1993). It

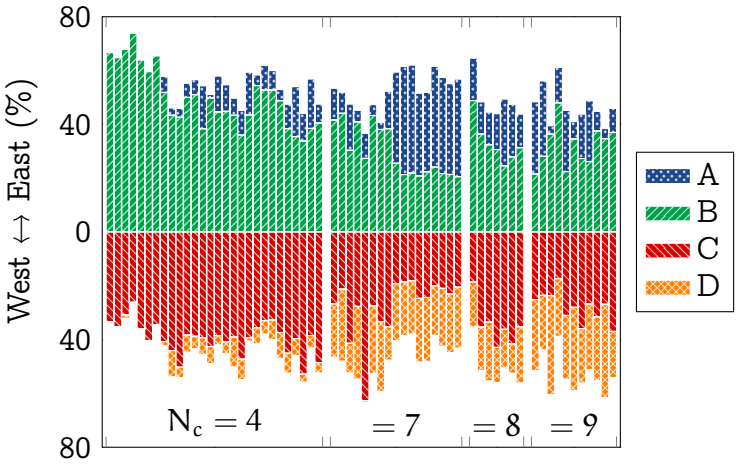


Figure 6: Train-specific split fractions of arrival flows across exit ways on platform #3/4 in Lausanne railway station grouped by train size and ordered by alighting volumes (increasing from left to right). Shorter trains induce only small flows on the lateral exit ways in sectors A and D, whereas longer trains and larger alighting volumes lead to a more homogeneous distribution across sectors.

consists in finding

$$\begin{aligned} \boldsymbol{d}^{\star}_{\gamma} = \arg\min_{\boldsymbol{d}\geq 0} \; & \mu_{\text{flow}} \|\hat{\boldsymbol{f}}'_{\text{tot}} - \boldsymbol{f}'_{\text{tot}}\|_2^2 + \mu_{\text{srf}} \|\hat{\boldsymbol{e}}'_{\text{tot}} - \boldsymbol{e}'_{\text{tot}}\|_2^2 + \mu_{\text{occ}} \|\hat{\boldsymbol{n}}'_{\text{tot}} - \boldsymbol{n}'_{\text{tot}}\|_2^2 \\ & + \mu_{\text{orig}} \|\hat{\boldsymbol{o}}'_{\text{tot}} - \boldsymbol{o}'_{\text{tot}}\|_2^2 + \mu_{\text{dest}} \|\hat{\boldsymbol{q}}'_{\text{tot}} - \boldsymbol{q}'_{\text{tot}}\|_2^2 + \mu_{\text{split}} \|\hat{\boldsymbol{r}}' - \boldsymbol{r}'\|_2^2 \\ & + \mu_{\text{arr}} \|\hat{\varphi}' - \boldsymbol{f}'_{\text{arr}}\|_2^2 + \mu_{\text{dep}} \|\hat{\chi}' - \boldsymbol{g}'_{\text{dep}}\|_2^2 + \mu_{\text{hist}} \|\hat{\boldsymbol{d}} - \boldsymbol{d}\|_2^2, \end{aligned} \tag{43}$$

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

The first three terms on the RHS of Eq. (43) represent the distance between the observed link and subroute flows and those predicted by the model, as well as that between observed and predicted area occupations. The terms on the second line consider the distance between model prediction and survey data in terms of cumulative origin and destination flows, as well as in terms of 'user-class split ratios' (given assumption (vi), the latter may also be referred to as aggregated destination split ratios). The first two terms on the last line consider the distance to the pre-computed train-induced arrival flows and the cumulative platform departure flows. As in Eq. (27), the last term represents the distance to a historical estimate, meant to overcome the underdetermination of the problem.

To assess the efficiency of the proposed framework, two estimators are compared. A 'base estimator', representing a simple minimum norm solver taking into account cordon counts only, and a 'full estimator', that additionally considers a 'static' and a 'dynamic' prior. The static prior includes cumulative origin and destination flows obtained from sales data and platform departure flows, as well as user split fractions. The dynamic prior represents pre-computed train-induced arrival flows. Trajectory recordings are only used for validation, for which they are aggregated as minute-by-minute subroute flows and occupations in PU East and PU West. No historical prior is considered. Fig. 7 illustrates the proposed estimation framework.

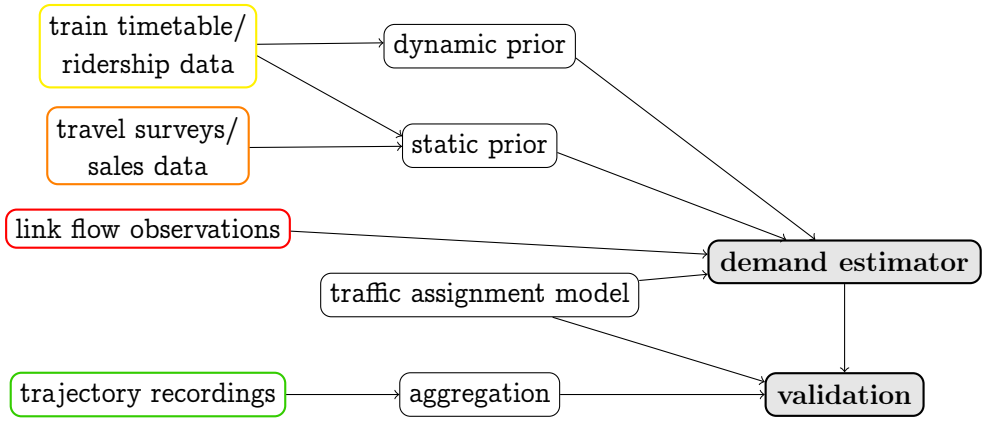


Figure 7: Scheme of the demand estimation framework applied to a case study of Lausanne railway station. The color scheme corresponds to Fig. 2.

For optimal statistical efficiency, the weights $\mu_{(\cdot)}$ are typically assumed equal to the reciprocal of the variance of the corresponding error term (Cascetta and Improta, 2002), i.e., $\mu_{\text{flow}} = 1/\operatorname{Var}(\boldsymbol{\eta}'_{\text{f}}+\boldsymbol{\omega}'_{\text{f}})$, $\mu_{\text{srf}} = 1/\operatorname{Var}(\boldsymbol{\eta}'_{e}+\boldsymbol{\omega}'_{e})$, $\mu_{\text{occ}} = 1/\operatorname{Var}(\boldsymbol{\eta}'_{\text{n}}+\boldsymbol{\omega}'_{\text{n}})$, $\mu_{\text{orig}} = 1/\operatorname{Var}(\boldsymbol{\omega}'_{\text{o}})$, $\mu_{\text{dest}} = 1/\operatorname{Var}(\boldsymbol{\omega}'_{\text{q}})$, $\mu_{\text{split}} = 1/\operatorname{Var}(\boldsymbol{\omega}'_{\text{r}})$, $\mu_{\text{arr}} = 1/\operatorname{Var}(\boldsymbol{\varpi}'_{\varphi})$, $\mu_{\text{dep}} = 1/\operatorname{Var}(\boldsymbol{\varpi}'_{\chi})$ and $\mu_{\text{hist}} = 1/\operatorname{Var}(\boldsymbol{\omega}_{\text{d}})$. In practice, these variances are however unknown, and need to be estimated.

For that purpose, pedestrian trajectory recordings are assumed to represent the truth. This allows to estimate the variance of the errors associated with the pedestrian count data and the train-induced arrival flows inferred from the train timetable. If, without loss of generality, the weight associated with pedestrian count data is set to one, $\mu_{\text{flow}} = 1$, a value of $\mu_{\text{dyn}} = \mu_{\text{arr}} = 0.69$ results for the weight of the dynamic prior.

Regarding the weight of the static prior, $\mu_{\text{stat}} = \mu_{\{\text{orig,dest,split,dep}\}}$, a sensitivity analysis is performed. In the range of $10^{-4} \leq \mu_{\text{stat}} \leq 10^{-1}$, only little variation in the resulting demand estimate is perceivable. 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 inaccuracy of the data sources it contains, the role of the static prior consists mostly in lowering the underdetermination of the problem. Thus, a value of $\mu_{\text{stat}} = 10^{-1}$ is employed in all following considerations.

The size of the estimation problem is given by the number of considered OD pairs, and the number of time intervals. In total, there are 370 feasible routes, and 30 time intervals of interest. To account for artificial transients in the demand estimates 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, for each day an estimation problem with a total of 16280 unknowns has to be solved.

An active set method (Lawson and Hanson, 1974) is used to solve the KKT (Karush-Kuhn-Tucker) conditions for the resulting non-negative least squares problem (43). If several optimal solutions exist, the one with the lowest norm is selected, yielding a solution with maximum entropy (Cascetta et al., 1993).

The current implementation (SciPy NNLS, a Fortran front-end) requires on a standard desktop machine a couple of hours of run time for a single evaluation of Eq. (43), i.e., for a given draw of $\hat{\boldsymbol{w}}$ and $\boldsymbol{\gamma}$. For the Monte Carlo sampling of an individual day, 16–24 evaluations have been found to suffice for generating reproducible and numerically stable results. All computations presented in this work have been conducted using $N = 24$ iterations. Thus, if a sample set of 10 days is to be estimated, 240 Monte Carlo iterations of Eq. (43) are required, resulting in a considerable computational load. By parallelizing the computational framework, and by using a small computer cluster, the overall run time required to estimate the present case study amounts to about three days.

4.6 Results

The temporal evolution of the total demand along the subroutes associated with the pedestrian underpasses PU East and PU West is shown in Fig. 8. Besides the results of the base and full estimators, also the observation from

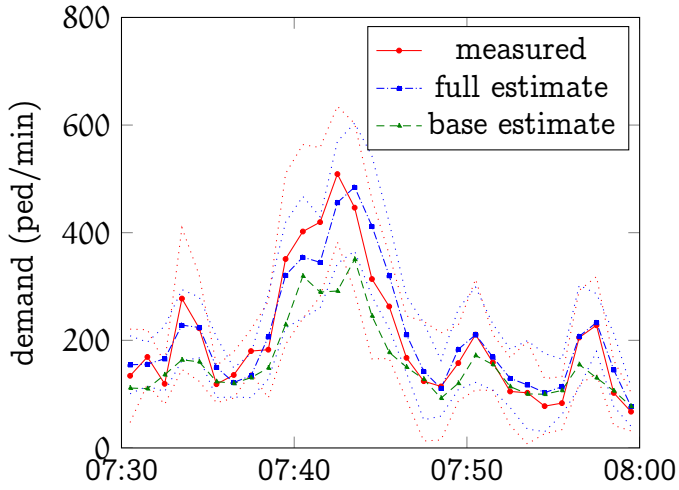
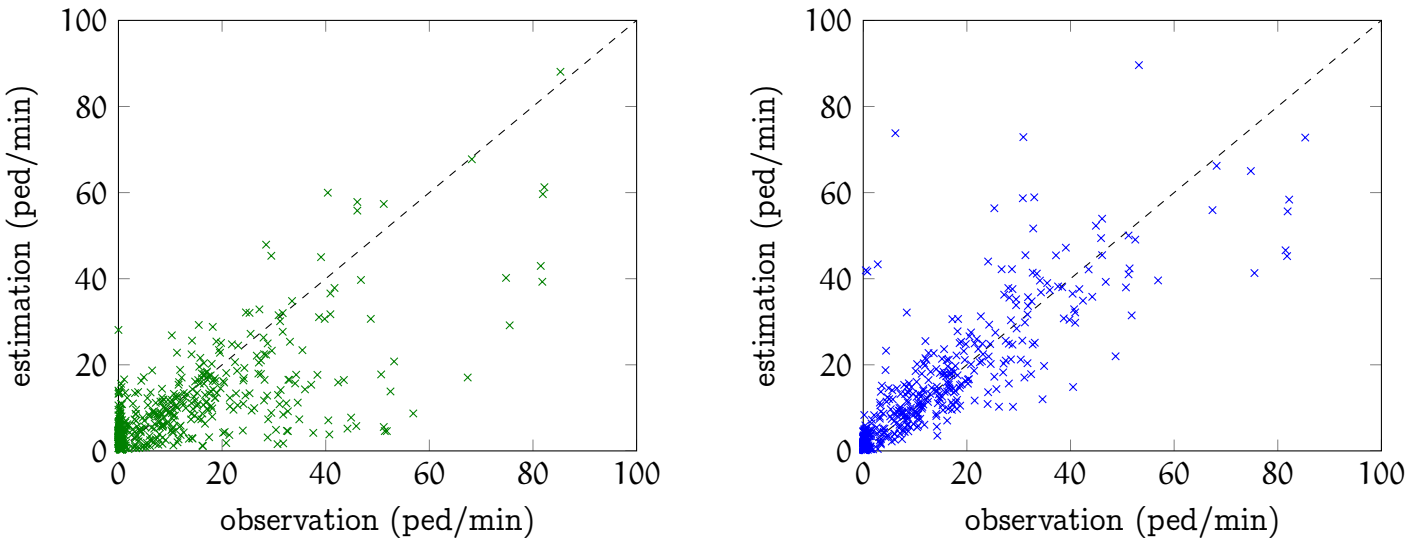


Figure 8: Total demand along subroutes as obtained from pedestrian trajectory recordings, and as predicted by two different estimators: $\mathrm{MAE}_{\mathrm{base}} = 50.74$, $\mathrm{MAE}_{\mathrm{full}} = 30.03$ (-40.8%); $\mathrm{RMSE}_{\mathrm{base}} = 70.47$, $\mathrm{RMSE}_{\mathrm{full}} = 37.56$ (-46.7%). Data: 10-day reference set, 2013.

pedestrian trajectory recordings is shown. The dashed curves represent the standard deviation bands, i.e., the mean $\pm$ one standard deviation, for the measurement and the full estimate. In the considered 30-minute period, the demand along subroutes fluctuates between less than 70 and more than 500 ped/min, i.e., by almost an order of magnitude.

It can be seen that 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. Moreover, the measured demand always lies within the standard deviation bands obtained for the full estimator. Further details regarding the performance of the two estimators can be found in the figure caption, where the mean absolute error (MAE) and the root-mean-square error (RMSE) are provided.

In the introduction to this work, the importance of platform exit flows in the aftermath of a train arrival in the dimensioning of platform access ways has been stressed. Therefore, the ability of the two estimators to reproduce these flows is of interest. Fig. 9 shows a scatter plot of observed and estimated pedestrian arrival flows obtained for platforms #3/4 and #5/6 in Lausanne railway station.



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

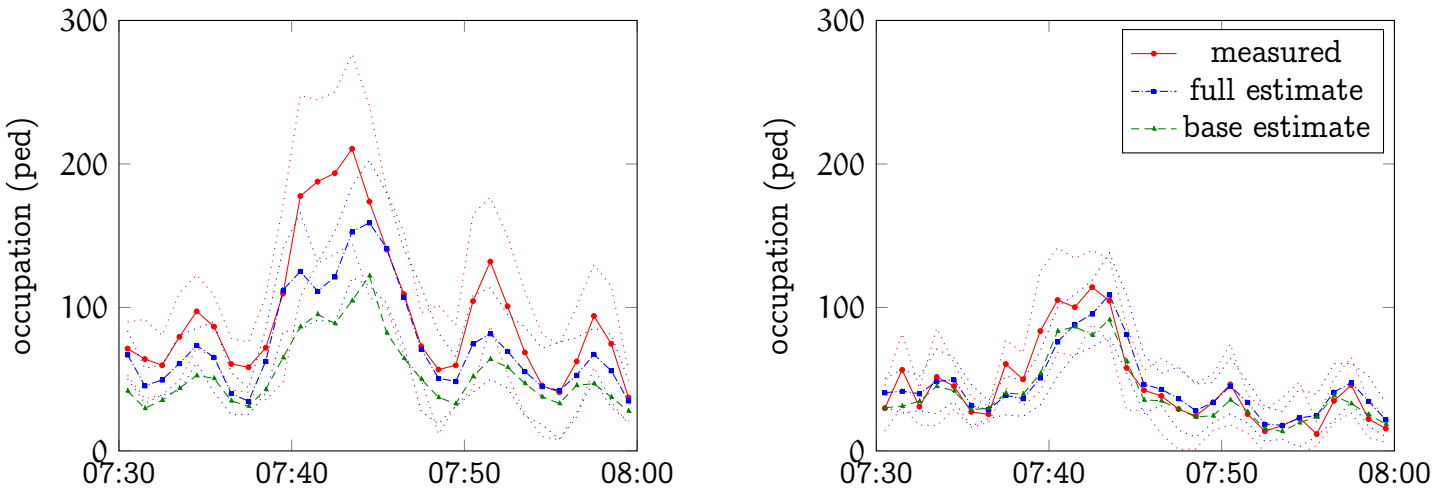
(b) Full estimate: MAE = 5.0457 (-30.26%). RMSE = 8.8952 (-23.35%).

Figure 9: Scatter plot of observed vs. estimated arrival flows at platforms #3/4 and #5/6 in Lausanne railway station. Data: 10-day reference set, 2013.

Neither the base nor the full estimator can perfectly reproduce the observed platform exit flows. The total platform exit flow is underestimated by the base model by -18.33%, and overestimated by the full model by 6.98%. As seen from the statistical measures provided in the captions of Fig. 9, the full model performs significantly better than the base model. This is in agreement with a visual impression obtained by comparing Fig. 9a and 9b.

The occupation in the two pedestrian underpasses can also be used to assess the accuracy of the developed framework. Fig. 10a and b show the average number of pedestrians present in PU West and PU East as obtained from pedestrian trajectory data, and as estimated by the base and the full model.

As expected, the occupation in the PUs is found to follow a similar pattern as the demand along subroutes shown in Fig. 8. A maximum occupation of 210 pedestrians is observed between 7:43 and 7:44 in PU West, in which the lowest value is reported at 37 pedestrians between 7:59 and 8:00. In PU East, the overall pattern is similar, but peaks are less distinct and the average occupation is less than half as high. The two



(a) PU West (b) PU East

Figure 10: Occupation in pedestrian underpasses as obtained from pedestrian trajectory recordings, and as predicted by two different estimators: $\text{MAE}_{\text{base}} = 48.53$, $\text{MAE}_{\text{full}} = 24.34$ (-49.9%); $\text{RMSE}_{\text{base}} = 58.83$, $\text{RMSE}_{\text{full}} = 34.86$ (-40.7%). Data: 10-day reference set, 2013.

model estimates are in principle able to reproduce the observed loading patterns, but, at least in PU West, the occupation during peak periods is clearly underestimated. This may have several reasons. High demand can lead to congestion, and thus to prolonged sojourn times in the underpasses. However, this hypothesis seems difficult to justify, as the observed density levels are relatively low and no significant correlation between travel times and demand has been found (Hänseler et al., 2014). It seems more plausible that the high occupation is caused by transfer passengers waiting in the PUs, and by outbound passengers that are either buying a ticket at one of the selling machines, or checking the timetable on one of the boards. Such behavior has been observed in PU West (Lavadinho, 2012), and may be captured by a dedicated assignment model.

Table 1 summarizes the previous findings, listing the RMSE associated with the estimates of subroute flow and occupation in the PUs. Results for the basic estimator, 'basic estimator + static prior', 'basic estimator + dynamic prior', and the full estimator (basic estimator + static prior + dynamic prior) are provided.

According to these results, in particular the incorporation of the dynamic prior leads to a significant improvement as compared to the base

Table 1: Performance in terms of RMSE of the base estimator, an estimator considering additionally the dynamic prior and the static prior, respectively, and of the full estimator. Values denote the relative change in RMSE comparing the subroute flows and the occupation in the pedestrian underpasses to the measurements obtained from a pedestrian tracking system. Data: 10-day reference set, 2013.

	subroute flow	occupation in PUs
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%

model. This implies that the consideration of train-induced arrival flows increases the prediction quality more significantly than sales data, information of user class split ratios and cumulative platform departure flows together. The full model globally performs best, even though the occupation 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.

An impression of the resulting pedestrian movements can be obtained from the flow maps contained in Fig. 11, showing minute-by-minute link flows for the time period between 7:43 and 7:46 on April 30, 2013, as obtained by the full estimator. On that day, the three trains IR 1710, IC 706 and IR 1407 arrive on platform #7 at 7:42:24, platform #5 at 7:42:59, and on platform #3 at 7:43:18, respectively. The traces they leave can be readily discerned in Fig. 11a and 11b. A further train, IR 2517, arrives on platform #1 at 7:44:37 (Fig. 11b and Fig. 11c), following which pedestrian flows decay, as can be seen from Fig. 11d.

The highest value of total demand is found between 7:39 and 7:40, amounting to 557.3 ped/min. A quarter of an hour later, between 7:54 and 7:55, it reaches a low of 112.0 ped/min (figure not shown). Within only a couple of minutes, the average total demand thus varies by almost a factor of 5. In the considered time period, 44.12% of all station visitors represent inbound passengers, 31.18% represent outbound passengers, 16.42% are transfer passengers, and the remaining pedestrians represent local users.

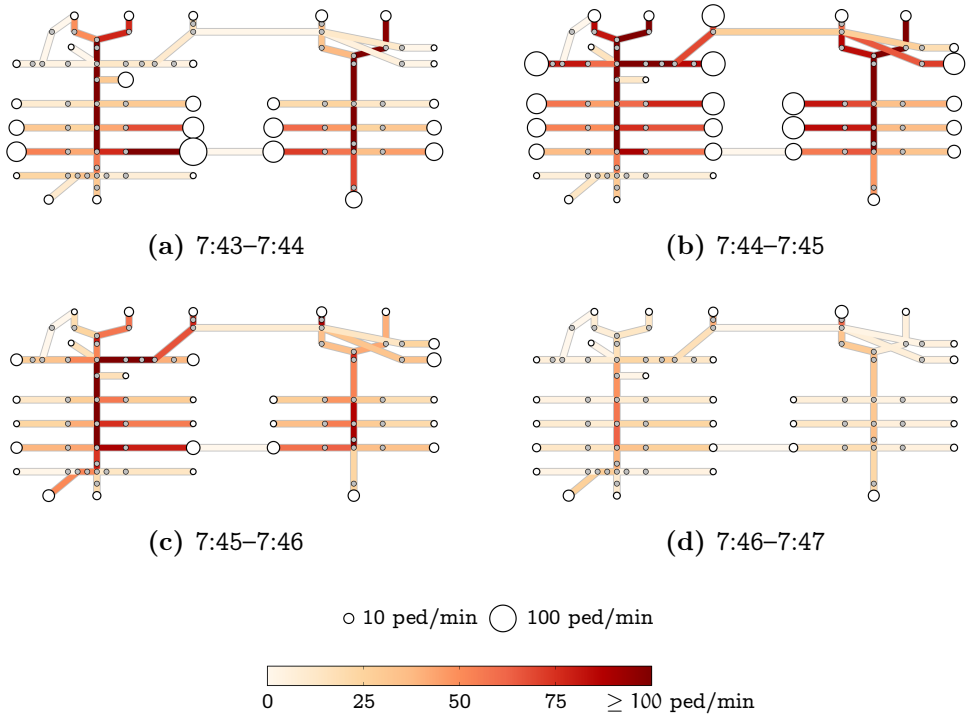


Figure 11: Exemplary pedestrian flow map for Lausanne railway station as estimated for the time period between 07:43 and 07:47 on April 30, 2013 by the full model. The shading of links represents the cumulative link flow over a minute in both directions. Similarly, the diameter of centroids represents the minute-by-minute origin flow.

5 Conclusions

During peak hours, rail access installations in large train stations often reach capacity and may reduce the performance of a transportation system. To optimize their design and operation, there is a general need to better understand pedestrian behavior in railway stations. An increasing effort is made towards this end both by operators of railway networks and academia. However, most researchers and practitioners concentrate on investigating the interaction between pedestrian demand and infrastructure, whereas the estimation of pedestrian demand as such has received relatively little attention so far.

In this study, 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 generally in good agreement with pedestrian tracking data that has been used for validation only. A clear 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 significant, justifying the use of a fully dynamic and probabilistic framework.

In the future, the proposed framework may be improved or extended in several ways. Three examples are given. First, from a practical point of view, the empirical relation between the train timetable and pedestrian movements in railway stations may be further strengthened (Molyneaux et al., 2014). Second, the use of a demand-dependent network loading model may allow for an explicit consideration of congested facilities, in which demand-supply interaction can no longer be neglected (Hänseler et al., 2014). Third, the framework could be employed for real-time traffic monitoring or crowd control (Seer et al., 2008).

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 Amanda Stathopoulos, Bonnie Qian, Nicolas Anken, Quentin Mazars-Simon and Eduard Rojas in implementing the case study of Lausanne railway station is gratefully acknowledged.

References

Alahi, A., Bagnato, L., Chanel, D., Alahi, A., 2013a. Technical report for SBB network of sensors. Tech. rep., VisioSafe SA, Switzerland.

Alahi, A., Jacques, L., Boursier, Y., Vandergheynst, P., 2011. Sparsity driven people localization with a heterogeneous network of cameras. Journal of Mathematical Imaging and Vision 41 (1-2), 39–58.

Alahi, A., Ramanathan, V., Fei-Fei, L., 2013b. 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.

Babel, H. L., 2014. Exploring pedestrian mobility using video tracking data in Lausanne train station. Semester thesis, Ecole Polytechnique Fédérale de Lausanne.

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.

Bera, S., Rao, K. V., 2011. Estimation of origin-destination matrix from traffic counts: The state of the art. European Transport 49, 3–23.

Bierlaire, M., Crittin, F., 2006. Solving noisy, large-scale fixed-point problems and systems of nonlinear equations. Transportation Science 40 (1), 44–63.

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., 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., 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.

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.

Chung, C. Y., 2012. KORAIL - Changing People's Lifestyle. Newspaper article, Business Korea, South Korea.

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.

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 Bilevel O–D Matrix Adjustment Problem. International Transactions in Operational Research 2 (2), 165–179.

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.

Hanisch, A., Tolujew, J., Richter, K., Schulze, T., 2003. Online simulation of pedestrian flow in public buildings. In: Winter Simulation Conference. Vol. 2. IEEE, pp. 1635–1641.

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., Molyneaux, N. A., Thémans, M., Bierlaire, M., 2013. Pedestrian strategies within railway stations: Analysis and modeling of pedestrian flows (PedFlux Mid-Term Report). Internal report (unpublished), Ecole Polytechnique Fédérale de Lausanne.

Hazelton, M., 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.

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.

Hermant, L. F. L., De Gersigny, M. R., Hermann, R., Ahuja, R., 2010. Applying microscopic pedestrian simulation to the design assessment of various railway stations in South Africa. In: Proceedings of the 29th Southern African Transport Conference (SATC 2010). Vol. 16. p. 19.

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.

Jiang, C. S., Deng, Y. F., Hu, C., Ding, H., Chow, W. K., 2009. Crowding in platform staircases of a subway station in China during rush hours. Safety Science 47 (7), 931–938.

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.

Kallas, S., 6 2014. Railways: Paving the way for more growth, more efficiency and service quality in Europe. Press release 14, European Commission for Transport, Brussels, Belgium.

Kasparick, U., 2010. Mobilität in Deutschland 2008: Ergebnisbericht. PN 3849/FE-Nr. 70.801/2006. Bundesministerium für Verkehr, Bau und Stadtentwicklung, Berlin, Germany.

Lam, W. H. K., Cheung, C. Y., 2000. Pedestrian speed/flow relationships for walking facilities in Hong Kong. Journal of Transportation Engineering 126 (4), 343–349.

Lam, W. H. K., Cheung, C. Y., Lam, C. F., 1999. A study of crowding effects at the Hong Kong light rail transit stations. Transportation Research Part A: Policy and Practice 33 (5), 401–415.

Land Transport Authority, Singapore, 2012. Singapore: Average daily public transport ridership (1995-2011), published on www.globalmasstransit.net.

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.

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.

Olesen, A., 2006. HOP(P) Schwiiz. Info Retica 2, 34–35.

Pettersson, P., 2011. Passenger waiting strategies on railway platforms: Effects of information and platform facilities. Master's thesis, KTH.

Puentes, R., Tomer, A., Kane, J., 2013. A new alignment: Strengthening America's commitment to passenger rail. Brookings Institution Press, Washington, DC.

Rindsfüser, G., Klügl, F., 2007. Agent-based pedestrian simulation: A case study of Bern Railway Station. The Planning Review 170, 9–18.

SBB-Infrastruktur, 2013. Le processus de l'élaboration de l'horaire. Internal report (unpublished), Swiss Federal Railways (SBB-CFF-FFS), Bern, Switzerland.

SBB-Personenverkehr, 2011. SBB Clearing Extrapolation Voyageurs HOP. Internal report (unpublished), Swiss Federal Railways (SBB-CFF-FFS), Bern, Switzerland.

Schneider, J., 2012. Network condition report 2012. Annual report, Swiss Federal Railways (SBB-CFF-FFS), Bern, Switzerland.

Seer, S., Bauer, D., Brandle, N., Ray, M., 2008. Estimating pedestrian movement characteristics for crowd control at public transport facilities. In: Intelligent Transportation Systems. IEEE, pp. 742–747.

Shao, H., Lam, W. H. K., Sumalee, A., Chen, A., Hazelton, M. L., 2014. Estimation of mean and covariance of peak hour origin–destination demands from day-to-day traffic counts. Transportation Research Part B: Methodological 68, 52–75.

Sherali, H. D., Park, T., 2001. Estimation of dynamic origin–destination trip tables for a general network. Transportation Research Part B: Methodological 35 (3), 217–235.

Srikukenthiran, S., Shalaby, A., Morrow, E., 2014. Mixed logit model of vertical transport choice in Toronto subway stations and application within pedestrian simulation. Transportation Research Procedia 2, 624–629.

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.

Stubenschrott, M., Kogler, C., Matyus, T., Seer, S., 2014. A dynamic pedestrian route choice model validated in a high density subway station. Transportation Research Procedia 2, 376–384.

Tolujew, J., Alcalá, F., 2004. A mesoscopic approach to modeling and simulation of pedestrian traffic flows. In: Proceedings of 18th European Simulation Multiconference.

van den Heuvel, J. P. A., Daamen, W., Hoogendoorn, S. P., 2013. Using Bluetooth to estimate the impact of congestion on pedestrian route choice at train stations. Traffic and Granular Flow.

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.

Weidmann, U., 1992. Transporttechnik der Fussgänger. Schriftenreihe des IVT Nr. 90. Institute for Transport Planning and Systems, ETH Zürich, Switzerland.

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.

Zhang, Q., Han, B., Li, D., 2008. Modeling and simulation of passenger alighting and boarding movement in Beijing metro stations. Transportation Research Part C: Emerging Technologies 16 (5), 635–649.