Introducing exogenous priority rules for the capacitated passenger assignment problem

Stefan Binder [*] Yousef Maknoon [†] Michel Bierlaire [*]

February 9, 2017

Report TRANSP-OR 170209
Transport and Mobility Laboratory
École Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

[*]Transport and Mobility Laboratory (TRANSP-OR), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland, {s.binder, yousef.maknoon, michel.bierlaire}@epfl.ch

[†]Delft University of Technology, Faculty of Technology, Policy and Management, Transport and Logistics Group, Jaffalaan 5, 2628 BX Delft, Netherlands

Abstract

We propose a novel algorithm for the capacitated passenger assignment problem in public transportation where priority lists define the order in which passengers are assigned. The originality of our approach is to define the priority rules exogenously. Separating explicitly theses rules from the assignment procedure allows for a great deal of flexibility to model various priority rules. Computational experiments are performed on a realistic case study based on the the morning rush hours of the timetable of Canton Vaud, Switzerland. The algorithm is able to assign the demand in very low computational times. The results provide evidences that the ordering of the passengers does not have a significant impact on aggregate performance indicators (such as average delay and level of unsatisfied demand), but that the variability at the individual passenger level is substantial. Thanks to its flexibility, our framework can easily be implemented by a railway operator who wishes to evaluate the effects of different policies in terms of passenger priorities.

Keywords: capacitated passenger assignment, public transportation, passenger priority list

1 Introduction

In the context of public transportation, passenger assignment models are used to predict the distribution of passengers over the network. These models thus play a critical role in identifying saturated parts of the network, by detecting crowded vehicles or train lines, for instance. The insights gained by these models can then be used in order to direct investments towards portions of the transportation network where alleviating congestion is most crucial.

One of the main issues of passenger assignment models is to account properly for passenger behavior in case of vehicle saturation. When passengers compete for the limited available capacity of trains, it is of critical importance to decide which passengers can board the train and which cannot. Indeed, consider a situation where a passenger has the choice to board a train or to wait for the next one on the same line. If he boards the train, he will use some of its available capacity and, at a later stop, his presence in the train might prevent another passenger to board. On the contrary, if he decides not to board the train, the available capacity might be sufficient to let the subsequent passenger board.

This simple example highlights why priorities between passengers lie at the core of a behaviorally meaningful passenger assignment model. Most frameworks in the field use a first-in-first-out queue at the stations, coupled with the rule that onboard passengers have priority over those wishing to board. This priority rule is endogenous because it depends on the passenger assignment. The endogeneity forces the existing frameworks to embed the passenger priorities implicitly in the network loading process, making them difficult to modify, in the case where a modeler or the railway operator desires to study the impacts of alternative orderings.

In this paper, we propose a novel framework that considers an exogenous ordering of the passengers. The main contribution of this paper is therefore to separate the passenger ordering from the actual assignment problem. By defining beforehand the rules that govern passenger priorities, complexity is extracted from the assignment problem itself. We use a shortest-path assignment in the illustrative case study presented in this paper, but any route choice model may be used. The fact that the priority lists are completely independent from the assignment process guarantees the flexibility of the framework. Any behavioral or control rule can be used to construct the priority list, including rules that account for the results of a previous assignment, in an iterative context.

The actual assignment algorithm takes three elements as input: (i) a time-dependent origin-destination matrix representing passenger demand, (ii) a space-time graph constructed from the railway timetable, and (iii) the passenger priority list. Passenger flows in the network are obtained by assigning the demand in the order given by the priority list (i.e., a passenger with higher priority will be assigned before a passenger with lower priority), thus settling the issues that arise in case of insufficient train capacity.

We compare five different priority specifications on a realistic case study based on the S-train network of Canton Vaud, Switzerland, and show that the ordering of the passengers does not have a significant impact on aggregate passenger satisfaction indicators. At the individual passenger level however, substantial variability is observed. The algorithm assigns the demand in a few seconds, thus making it practical for real-time evaluation of timetables from the passenger perspective and for a detailed investigation of the distribution of the performance indicators.

The remainder of this paper is structured as follows. Section 2 reviews recent contributions to the field of passenger assignment models. The problem is then formally introduced in Section 3. The generation process for the exogenous passenger priority lists is explained in Section 4, whereas the assignment algorithm itself is described in Section 5. Section 6 reports the results of the computational experiments on the case study. Finally, Section 7 concludes the paper and provides directions for future research.

2 Literature review

The recent literature on passenger assignment models for transit systems is either *frequency-based* or *schedule-based*. In the former approach, transit services are represented by lines with travel frequencies and single vehicles are not explicitly considered. Frequency-based static assignment models are generally suited for urban transportation systems (metro, bus) where the service is so frequent that it can be assumed that a passenger boards the first "attractive" vehicle when waiting at a stop. Seminal works in this field include Spiess and Florian (1989), who introduced the concept of optimal strategies and Nguyen and Pallottino (1988), who formalized the concept in terms of shortest hyperpaths. Many extensions have been proposed in the following years (e.g., de Cea and Fernández, 1993; Cominetti and Correa, 2001). The interested reader can refer to Fu et al. (2012) for an in-depth review of frequency-based passenger assignment models.

Single vehicle loads can only be approximated in frequency-based models. This approximation is especially unsuitable in case of irregular service (which is common in inter-urban systems such as trains or long-distance coaches), as it cannot account for peaks of passengers waiting at the station. In order to model the choice of passengers for a specific run of a specific transportation line, a schedule-based approach is needed. The loads and the performance of each single run can be obtained in such a framework. In schedule-based models, each vehicle is considered individually with its capacity, either implicitly or explicitly. The implicit approach is similar to road network modeling, where link costs are related to link flows through non-decreasing functions. Papers such as Tong and Wong (1999), Nuzzolo, Russo, et al. (2001), and Nielsen (2004) use this approach. By contrast, the explicit schedule-based approach introduces vehicle capacity constraints, thus letting waiting passengers board the arriving train according to its residual capacity. The following papers use the explicit schedule-based approach to assign passengers on transit networks.

Nguyen, Pallottino, and Malucelli (2001) consider the case where timetables are reasonably reliable, and the number and frequencies of transit vehicles are low. For this kind of networks, departure time and route choice are both equivalently important decisions that passengers face. Further, the concept of path available capacity is introduced in order to capture the flow priority aspect (i.e., giving priority to passengers already onboard the transit vehicles with respect to passengers waiting at the station). A traffic equilibrium model of the assignment problem is presented, and a computational procedure based on asymmetric boarding penalty functions is suggested to avoid the explicit enumeration of all paths connecting origins and destinations.

Poon et al. (2004) propose a model that explicitly describes the available capacity of every vehicle at each station, as well as the queuing time for every passenger. The paper focuses on the route choice problem, ignoring other choice dimensions, such as departure time or departure station. In their formulation, route choice for every passenger is modelled by selecting a path that minimizes a generalized cost function consisting of in-vehicle time, waiting time, walking time and line change penalties. The network is loaded (i.e., user equilibrium is achieved) by using a Method of Successive Averages (MSA) algorithm. The authors assume a First-In-First-Out (FIFO) queue discipline at the stations. Depending on the spare capacity of a transit vehicle at a station, the queue is split into passengers that can board the vehicle and passengers that remain at the station and wait for the next vehicle to arrive.

Hamdouch and Lawphongpanich (2008) also propose a user-equilibrium transit assignment model that explicitly considers individual vehicle capacities. For every origin-destination pair, passengers are divided into groups according to their desired arrival time intervals. It is assumed that every passenger group has a travel strategy resulting, at each station and each point in time, in a list of subsequent travel options that are ordered according to the passenger groups' preferences to continue their trip. Passenger preferences are described by the minimization of expected travel costs, made of in-vehicle time, fare and costs associated with early departures from home and/or arrivals outside the desired arrival time interval. When loading a vehicle at a station, onboard passengers continuing to the next station remain in the vehicle and waiting passengers are loaded according to the available remaining vehicle capacity. The paper considers two rules to sequence the boarding procedure of the waiting passengers: first-come-first-serve order and random order. If the vehicle is full, passengers unable to board need to wait for the next vehicle. Demand-supply interactions are defined by a user equilibrium approach and a solution method based on a MSA is proposed.

In Papola et al. (2009), a Dynamic Traffic Assignment model is extended to the case of scheduled services. It allows for explicit vehicle capacity constraints and FIFO queue representation at stations. The authors formulate the deterministic user equilibrium as a fixed point problem in terms of flow temporal profiles. A MSA algorithm is proposed to solve the problem.

Sumalee et al. (2009) propose a dynamic transit assignment model that differentiates discomfort levels experienced by sitting and standing passengers. The probability of getting a seat is captured by a stochastic seat allocation model. The passengers choose departure time and travel route by minimizing the perceived expected disutility, made of walking time, waiting time, early/late arrival penalty, expected perceived in-vehicle time (including congestion effect), transit fare and number of transfers. It is assumed that passengers boarding a vehicle obey the FIFO discipline. Further, standing passengers already on-board have priority over boarding passengers to gain access to available seats. The authors formulate the departure time and route choice problem as a probit stochastic user equilibrium problem and develop a heuristic solution algorithm to find an equilibrium solution.

Nuzzolo, Crisalli, et al. (2012) propose a schedule-based dynamic assignment problem for congested transit networks, explicitly considering vehicle capacities. Its novelty resides in the fact that more complex behavioral choice models are used for passengers, including the choice of departure time, departure station and departure train run. A day-to-day learning process for the passengers is also included in the model. The network loading procedure assigns users on each transit run according to their choice and to the residual capacity of the vehicles arriving at the stop. Again, FIFO queueing discipline is assumed.

Hamdouch, Szeto, et al. (2014) extend the framework of Hamdouch and Lawphongpanich (2008) by considering supply uncertainties. An analytical formulation is provided to ensure that on-board passengers have priority over boarding passengers and waiting passengers are loaded according to FIFO precedence. A user equilibrium model is proposed for the problem, which is solved by a MSA-type algorithm.

Kroon et al. (2015) present a deterministic algorithm to simulate passenger flows in a capacitated network. Passengers are grouped according to origin-destination pairs and arrival time in the system. Passenger flows emerge from the competition between the passengers for the limited capacity of the trains. When more passengers attempt to board a train than the capacity allows, passenger groups are split: some passengers can board the train while the remaining will have to look for an alternative travel route. The authors also assume that passengers who are already on a train when a train calls at a station, and who intend to continue on the same train, have priority over newly boarding passengers.

2.1 Summary and contributions of the present work

In all the reviewed works (except Kroon et al. (2015) — see below), priority is given to passengers already onboard with respect to passengers wishing to board a vehicle, and the priority among boarding passengers is decided using a FIFO or random rule. These priorities are endogenous because the passenger ordering depends on the assignment, and vice versa. The issue with endogeneity is the fact that the logic behind the priorities needs to be "hard-coded" into the assignment algorithm itself. By contrast, we propose exogenous priority rules, which remove the complexity from the network loading procedure and allow the modeler to investigate potential alternative rules very easily. Note that our framework can easily be included in a fixed point formulation to calculate the user equilibrium, in the case where passenger decisions depend on the assignment (if crowding is considered, for instance). It can therefore be seen as an extension of recent contributions to the field of schedule-based passenger assignment models.

A notable exception is the work by Kroon et al. (2015), where passengers are moved through the network according to the available capacity of each vehicle. In case of vehicle saturation, passenger groups are split and the number of passengers that can board the train is proportional to the group size. Our framework is more flexible than this one as it allows to define any priority rule to order the passengers before the assignment.

The contributions of this paper are summarized as follows:

- We propose a flexible network loading framework for the capacitated passenger assignment problem.
- We define exogenous priority lists to order the passengers before the assignment, thus extracting the complexity from the assignment problem.
- We allow total flexibility for the definition of the priority rules.
- We carry out computational experiments on a realistic case study and our framework is able to assign the demand in little time.

3 Problem description

This section is dedicated to the formal description of the problem addressed in this research. We begin by explaining the representation of the demand and by describing the assumptions on passenger behavior that lead to our passenger travel choice model. The representation of the public transportation timetable as a space-time graph is portrayed thereafter. We conclude this section by discussing the interactions between the demand side (passengers) and the supply side (train capacities) of the problem, which lie at the core of a passenger assignment model.

In the remainder of this paper, we use the following notation. Time is discretized into $n + 1$ time intervals of length τ (typically, one minute) and we introduce the set of time steps $H = \{0, \tau, 2\tau, \cdots, n\tau\}$, where $n\tau$ is the considered planning horizon. We model the railway network at a macroscopic level. The set of stations is denoted by S, and the travel time between two consecutive stations $s, s' \in S$ by $t(s, s') \in H$. This travel time is deterministic and independent of the train. Further notations are introduced when needed.

3.1 Passenger travel choice model

Passenger demand is assumed to be known, in the form of an origin-destination (OD) matrix. The latter describes the number of passengers entering the system at a given origin station, at a certain time, and who wish to travel to a given destination station. The availability of such data becomes more and more frequent with the gradual introduction of smart cards in public transportation networks.

Based on the OD matrix, a passenger p is denoted by a triplet (o_p, d_p, t_p), where $o_p \in S$ is the origin station, $d_p \in S$ the destination station, and $t_p \in H$ the desired departure time from the origin. Note that, as we assume deterministic train travel times in our approach, a passenger can equivalently be characterized by the desired arrival time at the destination. We adopt the former representation in the following. The set of all passengers is denoted by P.

We assume that the passengers know the timetable (i.e., all the train departure and arrival times at all stations) and that they plan their path in the network accordingly. Such a path may consist of different trains if there is no direct connection between origin and destination of the passenger. For every passenger, we consider the set $\Omega(o_p, d_p)$ of all paths linking o_p to d_p. A path is defined as a sequence of access, in-vehicle, waiting, transfer and egress movements (refer to Section 3.2 for a definition in terms of arcs in the space-time graph). In order to distinguish different paths, we associate a generalized cost with every alternative (i.e., path) and assume that each passenger chooses the one with the lowest generalized cost. In general, the specification of generalized cost C^p_ω for alternative $\omega \in \Omega(o_p, d_p)$ and passenger p is a function of several attributes:

$$C^p_\omega = C(z^p_\omega, y_p; \beta), \tag{1}$$

where z^p_ω is a vector of attribute values (such as travel time and cost) for alternative ω as viewed by passenger p, and β is a vector of parameters representing the weight of the different attributes. As preferences may vary across the passengers, this general specification also includes a vector y_p of socioeconomic characteristics (e.g., income, age, education) that models the heterogeneity of the population. Our formulation is thus general enough to accomodate heterogenous demand, but in this research we focus on homogenous demand, similarly to Binder et al. (forthcoming). We also assume that the price of the trip is equal among all paths for a given passenger. We therefore consider, for a given passenger p and path ω, the following attributes in the specification of the generalized cost:

- *In-Vehicle Time* (VT^p_ω): time, in minutes, spent by the passenger in one (or more) train(s) along the path,
- *Waiting Time* (WT^p_ω): time, in minutes, spent by the passenger waiting between two consecutive trains at a station along the path (does not consider the waiting time for the first train),
- *Number of Transfers* (NT^p_ω): number of times the passenger needs to change trains along the path,
- *Early Departure* ($ED^p_\omega = \max(0, t_p - t)$): time difference (in minutes) between the desired (t_p) and the actual (t) departure time from origin, if early,
- *Late Departure* ($LD^p_\omega = \max(0, t - t_p)$): time difference (in minutes) between the actual (t) and the desired (t_p) departure time from origin, if late.

Based on the aforementioned description for a given passenger p, the generalized cost of alternative ω is defined as follows:

$$C^p_\omega = VT^p_\omega + \beta_1 \cdot WT^p_\omega + \beta_2 \cdot NT^p_\omega + \beta_3 \cdot ED^p_\omega + \beta_4 \cdot LD^p_\omega, \tag{2}$$

where $\beta_1, \ldots, \beta_4$ are the relative weights of the attributes described above. C^p_ω is in minutes and expresses the generalized travel time of passenger p along path $\omega \in \Omega(o_p, d_p)$. As commonly done in the literature, the weights of the various elements of the generalized travel time are defined relative to the in-vehicle time of the path. We use the values reported in Table 4, obtained from the literature. We assume that passengers have full knowledge of the system and that they choose the fastest path (i.e., the lowest generalized travel time) to travel from origin to destination.

Due to train capacity issues, it is possible that, for some passengers, no feasible alternative exists between origin and destination within the time horizon. We therefore include an artificial "penalty path" for those passengers. This path models the worst possible option to travel from origin to destination. We therefore associate it with the highest possible travel time: the duration of the time horizon.

In case of high demand, capacity issues may also arise because of passenger congestion. In other words, a passenger may fail to board a train on his preferred path, characterized by Eq. (2), because this train has reached its capacity due to other passengers having boarded the train earlier. The consideration of such interactions between passengers is discussed in detail in Section 3.3 and lies at the core of the passenger assignment algorithm designed in Section 5.

3.2 Timetable representation as a space-time graph

The passenger paths are mathematically defined in a graph representation of the timetable. To that end, we introduce a directed space-time graph $G(V, A^p)$, for every passenger p. The set of all trains in the timetable is denoted by K.

The set of nodes $V = N \cup N_O \cup N_D$ consists of three different types of nodes. Starting from an empty graph, we add a space-time node $(s, t, k) \in N$ for each arrival/departure event of train $k \in K$ at/from station $s \in S$ at time $t \in H$. For instance, if train $k' \in K$ leaves station $s_1 \in S$ at time $t_1 \in H$, stops at station $s_2 \in S$ from time $t_2 \in H$ to $t_3 \in H$ and finishes its trip at station $s_3 \in S$ at time $t_4 \in H$, four space-time nodes are added: (s_1, t_1, k'), (s_2, t_2, k'), (s_2, t_3, k') and (s_3, t_4, k'). In addition, N_O and N_D are the sets of time-invariant origin and destination nodes of the passengers. We denote by $s(o)$ and $s(d)$ the station associated with node $o \in N_O$ and $d \in N_D$, respectively.

There are six types of arcs in the graph:

- *Driving arcs* model the movements of trains between stations. From the timetable, we define, for every train $k \in K$, the set of driving arcs A^k_{Dri}. A driving arc connects a departure event at one station $(s, t, k) \in N$ to an arrival event at the following station $(s', t', k) \in N$, with $t' = t + t(s, s')$. By repeating this procedure for every train in the timetable, we construct the set of driving arcs $A_{Dri} = \cup_{k \in K} A^k_{Dri}$.
- *Waiting arcs* model trains waiting at a station for passengers to board or alight. We define from the timetable, for every train $k \in K$, the set of waiting arcs A^k_{Wai}. A waiting arc connects an arrival event at a station $(s, t, k) \in N$ to a departure event from the same station $(s, t', k) \in N$, with $t' = t + w(s)$, where $w(s)$ is the waiting time at station s. By repeating this procedure, we construct the set of waiting arcs $A_{Wai} = \cup_{k \in K} A^k_{Wai}$.
- *Access arcs* model passenger p arriving at the origin. They are given by the set $A^p_{Acc} = \{(o, (s, t, k)) \in N_O \times N | s = s(o) = o_p\}$. Note that, by definition, passenger p can therefore take any train that departs from his origin station.
- *Egress arcs* model passenger p leaving the system at destination. They are given by the set $A^p_{Egr} = \{((s, t, k), d) \in N \times N_D | s = s(d) = d_p\}$.
- *Transfer arcs* model passengers transferring from one train to another in a station, with a minimal transfer time m and a maximal transfer time M. The set of transfer arcs is constructed in the following way: $A_{Tra} = \{((s, t, k), (s, t', k')) \in N \times N | \forall s \in S, \forall k \in K, \forall k' \in K \setminus \{k\}, \forall t, t' \in H : m \leq t' - t \leq M\}$.
- *Penalty arcs* model passenger p not succeeding to find a path from origin to destination. They are given by the set $A^p_{Pen} = \{(o, d) \in N_O \times N_D | s(o) = o_p, s(d) = d_p\}$.

The set of arcs associated with passenger p is given by $A^p = A_{Dri} \cup A_{Wai} \cup A^p_{Acc} \cup A^p_{Egr} \cup A_{Tra} \cup A^p_{Pen}$. With each arc a, a capacity q_a and a weight c^p_a are associated. The capacity of driving and waiting arcs is given by the capacity of the associated train (i.e., the maximum number of passengers that can be in the train at the same time), q_k. The other arcs have infinite capacity. The arcs are weighted according to the generalized cost function introduced in Section 3.1. Table 1 summarizes arc weights and capacities.

The cost of a path in the graph for a passenger is obtained by summing the weights of the arcs along the path. Note that the choice of a path determines the exact train(s) a passenger takes in the network, as well as his departure time from origin (penalizing early or late departure with respect to

Table 1: Arc weights and capacities.

Name	Start node	End node	Weight (c_a^p)	Capacity (q_a)
Driving	(s, t, k)	(s', t', k)	$t' - t$	q_k
Waiting	(s, t, k)	(s, t', k)	$\beta_1 \cdot (t' - t)$	q_k
Access	o_p	(s, t, k)	$\beta_3 \cdot \max(0, (t_p - t)) + \beta_4 \cdot \max(0, (t - t_p))$	∞
Egress	(s, t, k)	d_p	0	∞
Transfer	(s, t, k)	(s, t', k')	$\beta_2 + (t' - t)$	∞
Penalty	o_p	d_p	$n\tau$	∞

the desired departure time). Therefore, our framework is able to model accurately passenger behavior in congested public transportation networks by reproducing effects such as waiting for the next train when failing to board the current train (because of congestion) or adjusting the departure time in order to anticipate congestion.

3.3 Supply-demand interactions in the assignment

The goal of our framework is to obtain accurate passenger flows in a public transportation network where passengers compete for the limited capacity of the trains. When the number of passengers attempting to board a train exceeds its available capacity, it has to be decided eventually which passengers can board the train and which cannot. The remaining ones will have to look for another alternative (e.g., wait for the next train of the line). There are two main paradigms to take this decision:

System optimum Passengers are assumed to collaborate in order to minimize the overall inconvenience (i.e., generalized travel time) of all passengers.

User equilibrium Passengers are assumed to be selfish actors that attempt to minimize their personal inconvenience.

Although a system optimal passenger assignment yields a better experience for everyone *on average*, public transportation users are usually not willing to accept a longer personal travel time for a theoretical "greater good". In this framework, we therefore assume passengers to be selfish and independent and to maximize their personal utility, given by Eq. (2). In order to decide which passengers can board a train in the situation described above, we introduce in the following section exogenous passenger priorities that order the passengers according to different criteria.

4 Exogenous passenger priority lists

We introduce in this section the concept of passenger priority lists that model the relative importance of the passengers. A priority list is characterized by an injective mapping $\Gamma : P \to \mathbb{R}$ that associates each traveler $p \in P$ with a "level of importance" $\Gamma(p)$. It is then assumed that passenger p has priority on passenger q if and only if $\Gamma(p) > \Gamma(q)$. The injection assumption is designed to avoid ties. The challenge for the modeler is to translate actual priority rules into an appropriate mapping Γ. Note that the framework must be general enough to model priority rules that are actually prevailing in practice, or rules that need to be evaluated by an operator, or artificial rules that are used to benchmark the system.

In order to allow for a great deal of modeling possibilities, we propose a probabilistic approach to build the priority list. The importance of passenger p is modeled using a random variable U_p. Therefore, the *probability* that passenger p has priority over passenger q is given by $Pr(U_p \geq U_q)$. Actual mappings can then be obtained by simulation. For each realization $r = 1, \ldots, R$, the mapping $\Gamma^r(p)$ is defined as $\Gamma^r(p) = U_p^r$, where U_p^r is a realization of the random variable U_p. Note that if U_p is a continuous random variable, the probability that two passengers share the same level of importance is zero, so that each mapping is an injection with probability 1. However, due to finite arithmetic, ties may appear occasionally. If so, they may be broken in an arbitrary way. We present below concrete examples of the specification of the random variable U_p. The importance function U_p is composed of two components: a deterministic part (V_p) and a stochastic part (ε_p),

$$U_p = V_p + \varepsilon_p, \forall p \in P. \tag{3}$$

Deterministic part V_p In order to provide modeling flexibility, the deterministic part of the importance function is a function of various attributes of the passenger. In this research, we illustrate the concept using five different specifications, summarized in Table 2.

Table 2: Definitions of the passenger priority specifications.

Specification	Prioritized passengers	V_p
D	Passengers with early desired departure time from origin	$-t_p$
M	Passengers whose generalized travel time increases the most if they miss their first choice	$C^p_{\omega^p_2} - C^p_{\omega^p_1}$
S	Passengers with short origin-destination pairs	$-C^p_{\omega^p_1}$
L	Passengers with long origin-destination pairs	$C^p_{\omega^p_1}$
R	Random	0

- **Specification D** The first specification assumes that passengers with earlier desired departure time from origin are assigned before passengers with later desired departure time. This assumption is consistent with the observation that passengers who arrive earlier at a station board before the others. Also, trains will be filled up from the beginning to the end of their journey. In this case, $V_p = -t_p$, where t_p is the desired departure time of passenger p.

- **Specification M** The second specification assumes that passengers whose generalized travel time increases most if they miss their first travel choice are prioritized. The behavioral motivation behind this assumption is to prioritize passengers who have the most to "regret" if they cannot board their train. More specifically, we construct, for every passenger p, the set of all paths in $G(V, A^p)$ between o_p and d_p, $\Omega(o_p, d_p)$. Then we denote by ω_1^p (ω_2^p) the shortest (respectively, second-shortest) path of passenger p, i.e., $\omega_1^p = \{\omega \in \Omega(o_p, d_p) | C_\omega^p = \min_{\omega' \in \Omega(o_p,d_p)} C_{\omega'}^p\}$ and $\omega_2^p = \{\omega \in \Omega(o_p, d_p) | C_\omega^p = \min_{\omega' \in \Omega(o_p,d_p)\setminus\{\omega_1\}} C_{\omega'}^p\}$. The deterministic part of the importance function is then given by the difference in the generalized cost of the two paths: $V_p = C_{\omega_2^p}^p - C_{\omega_1^p}^p$.

- **Specifications S and L** The third and fourth specifications assume that the train operator wishes to give priority to passengers that have either short or long origin-destination pairs. In this case, the systematic part of the importance function is given by the generalized cost of the shortest path of passenger p, with a different sign. To prioritize short OD pairs, $V_p = -C_{\omega_1^p}^p$, whereas to prioritize long OD pairs, $V_p = C_{\omega_1^p}^p$.

- **Specification R** Finally, a priority specification where the passengers are assigned in a completely random order can be modeled using a constant systematic part of the importance function, so that only the random term matters. For instance, $V_p = 0$.

Specifications D and M are behiavorally driven and therefore quite realistic. Passengers arriving early at the station will probably board before passengers arriving later (specification D). Passengers prioritized in specification M (those who have the most to "lose" if they miss their train) can be assumed to be very aggressive in order to gain priority over other passengers, thus leading to a higher rank in the priority list. It may actually even happen that their second choice is not available either. This specification ignores that fact. Specifications S and L can be considered as control strategies that may be implemented by the train operator. For instance, the operator might want to give priority to "VIP" passengers or passengers on a particular OD for marketing reasons. Finally, specification R can be seen as a benchmark to compare to the other specifications.

Note that the variables influencing the order of the passengers in the five specifications are independent of the assignment. In reality however, the priority lists are not always exogenous. The flexibility of our framework also allows to accomodate endogenous priority lists: Section 5.3 details how it can be embedded in a fixed point formulation.

Stochastic part ε_p The stochastic part is designed to capture various elements, not explicitly modeled by the deterministic part, that may influence the relative "importance" of passengers and, therefore, their priority to board the train. Examples include the exact time of arrival on the platform, or the distance from the platform to the train door. In practice, a distribution must be assumed. It is convenient to assume that the ε_p are independent and identically distributed across p. The mean of this distribution can be assumed to be any arbitrary value (typically zero), as any deviation from the assumption can be captured by the deterministic part of the importance function. The variance, on the other hand, matters a lot, as it controls the relative importance between the deterministic and the stochastic parts of the importance function. For instance, following the assumptions applied in the context of discrete choice models, we assume that ε_p are independent and identically extreme value distributed with location parameter $\eta = 0$ and scale parameter $\mu > 0$, noted $\varepsilon_p \sim EV(0, \mu)$. This assumption is motivated by the analogy with choice models, where decision makers maximize their utility. Here, the importance of passengers is also maximized, to identify passengers with higher priorities. The cumulative distribution function of a random variable $x \sim EV(0, \mu)$ is given by $F(x) = \exp(-\exp(-\mu x))$. Draws v from this distribution are obtained from draws u from the uniform distribution as follows:

$$v = -\frac{1}{\mu} \ln(-\ln u), \qquad u \sim U(0, 1). \tag{4}$$

Note that the variance of the random variable is inversely proportional to the square of the scale parameter: $\frac{\pi^2}{6\mu^2}$. Hence, the scale factor μ captures the variability of the disturbances. Appendix B describes how its value is determined for every specification.

5 Passenger assignment algorithm

Given a public transportation network, we present a stochastic schedule-based passenger assignment algorithm that considers vehicle capacities explicitly. Fig. 1 describes the general framework. The passenger assignment model takes the following information as an input: a train timetable and a passenger OD matrix (both described in Section 3), as well as a priority list Γ for the passengers, as discussed in Section 4. The latter defines the order in which the passenger demand is assigned on the network: a passenger with a higher priority will be assigned before another passenger with lower priority. Based on these inputs, the assignment model computes the passenger flows on every arc of the network. Performance indicators can then be deduced from the passenger flows.

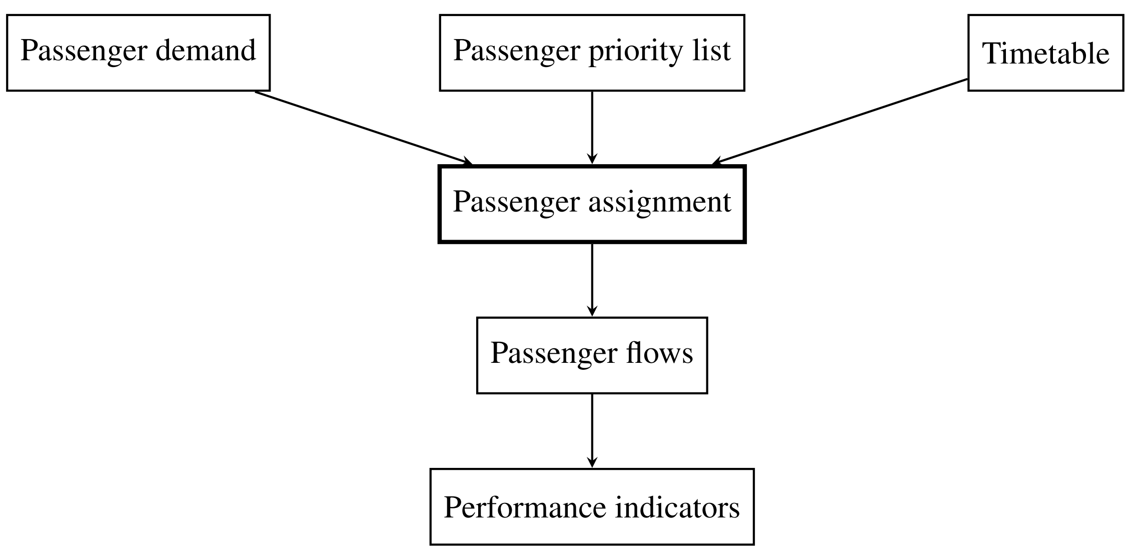

Figure 1: Inputs and outputs of the passenger assignment model.

Note that this framework assumes that crowding does not affect the passenger decisions, nor the priority list. In reality however, passengers may adjust their departure time or wait for the next vehicle if the current one is close to saturation. As a consequence, they can gain or lose priority in the system.

As discussed in Section 5.3, the framework can easily be extended to include the impact of crowding on demand and priorities, using a fixed point formulation.

5.1 Algorithm description

The passenger assignment algorithm assigns the passengers on the network according to their order in the priority list. Algorithm 1 describes the pseudo-code of the method. The algorithm takes as input the elements depicted in Fig. 1 and returns an indicator $f^p : A^p \to \{0, 1\}$, which shows if passenger $p \in P$ uses arc $a \in A^p$ in the assignment. During the initialization phase (lines 1–2), an ordered (according to Γ) copy O of the set of passengers P is made and the flow indicator is set to zero for all arcs and for all passengers. The algorithm iterates as long as the set O is non-empty (line 3). At every iteration, the passenger $p \in O$ with highest priority is selected, and then removed from the set O (lines 4–5). In line 6, we denote by o_p and d_p the origin node and the destination node of passenger p, respectively. We then compute the shortest path (in terms of the generalized cost described by Eq. (2)) in the graph $G(V, A^p)$, from node o_p to node d_p (line 7). $SP \subset A^p$ denotes the set of arcs used by the passenger on his shortest path. For each arc a in the set SP, we update the indicator $f^p(a)$ to reflect the fact that passenger p uses arc a (line 9). Furthermore, if a is a driving or a waiting arc (line 10), we need to verify if the capacity of the arc is reached after the passenger is assigned on the arc (line 11). If it is the case, the arc is removed form the graph of the remaining passengers in O (line 12), that is, for passengers with lower (or equal) priority than p. The penalty arc (with infinite capacity) linking every origin-destination pair guarantees that each passenger can always be assigned.

Algorithm 1: The passenger assignment algorithm.

Input: Set of passengers $p \in P$
    Passenger priority list $\Gamma : P \to \mathbb{R}$
    Passenger graphs $G(V, A^p)$
Output: Passenger flows $f^p : A^p \to \{0, 1\}, \forall p \in P$

1 $O :=$ An ordering of P according to Γ, breaking ties arbitrarily
2 $f^p(a) := 0, \forall a \in A^p, \forall p \in P$
3 **while** $O \neq \emptyset$ **do**
4   Let p be the first passenger in O (i.e., the passenger with highest priority)
5     $O := O \setminus \{p\}$
6   Let o_p and d_p be the origin and destination nodes of p, respectively
7   Obtain $SP \subset A^p$, the set of arcs in the shortest path between o_p and d_p
8   **foreach** $a \in SP$ **do**
9         $f^p(a) := 1$
10   **if** $a \in A_{Dri} \cup A_{Wai}$ **then**
11   **if** $\sum_{p \in P} f^p(a) \geq q_a$ **then**
12                 $A^p := A^p \setminus \{a\}, \forall p \in O$

This algorithm has the advantage of being computationally efficient for two reasons. First, the sorting of the passengers (which determines the priorities in case of competition for the limited capacity) is done beforehand, hence the algorithm sweeps only once over the set of passengers. Second, the traveling strategy of the passengers is implemented as a shortest-path algorithm in each passenger graph. The latter being directed and acyclic, the shortest-path search can be performed in linear time. It is important to note that the route choice model in line 7 of Algorithm 1 can easily be replaced by a more complex one. If a probabilistic model is chosen, the assignment procedure becomes a simulator. The latter is not an issue in our framework, as the system needs to be simulated anyway when the priority lists are generated from a probabilistic model, as described in Section 5.2.

Further, the algorithm is guaranteed to terminate after $|P|$ iterations, thanks to the penalty arc connecting every origin and destination node. Indeed, in the worst case, even if all driving arcs are removed from A^p, passenger p can still use the penalty arc between o_p and d_p.

We conclude this description by noting that removing saturated arcs from the passenger graph of subsequent lower-priority passengers is a way to enforce hard priority rules among passengers. However, slightly different priority lists (e.g., due to stochasticity) may lead to very different passenger flows, due to the fact that capacity constraints are binary. We discuss the variability across realizations for a given priority specification in detail in Section 6.

5.2 Simulation of the assignment procedure

The assignment itself, as described in Algorithm 1, is purely deterministic (because the priority specifications are exogenous). However, both the traveler demand and the priority lists may be stochastic. In that case, the assignment procedure must be embedded in a simulation. Therefore, the passenger assignment algorithm is run (i.e., the system is simulated) multiple times, each time with the same timetable, but stochastically different passenger demand and passenger priority lists. In the following, we denote by r each realization of the simulation algorithm. We also define the flow indicator $f_r^p : A^p \rightarrow \{0, 1\}$, which shows if passenger $p \in P$ uses arc $a \in A^p$ in the r-th realization of the assignment simulation. Based on the flow indicator, we define several indicators to evaluate the performance of the assignment. The distribution of these indicators can then be analyzed in detail, including the calculation of their mean, variance and various quantiles. R stands for the total number of realizations in the following.

5.3 Fixed point formulation

The general framework for the passenger assignment algorithm presented in Fig. 1 assumes that the inputs (passenger demand and priority list) of the model do not depend on its output (passenger flows).

In reality however, passenger flows can have a significant impact on passenger decisions. For instance, crowding has been shown to affect route choice in public transportation systems (see, e.g., Tirachini et al., 2013). Also, it may happen that a passenger cannot board a vehicle that is full, because of the choices of the other passengers. These effects can be taken into account in our framework, by including flow-dependent variables (such as crowding) in the generalized cost function given in equation (2).

Priority rules can also depend on passenger flows, thus introducing endogeneity. The rule giving priority to onboard passengers over boarding passengers, commonly used in the literature, is a classical example of an endogenous priority rule. Our formulation can easily be extended to include in the importance function Γ any variable that influences the passenger order, including performance indicators that result from the assignment (e.g., prioritize passengers who have to take the penalty path, or passengers already onboard a train over passengers wishing to board it).

In the case where the passenger demand and/or the priority list depend(s) on the outcome of the assignment, a fixed point problem needs to be solved. Most frameworks in the literature formulate the passenger demand fixed point problem as a user equilibrium, and use a MSA-type algorithm to solve it (e.g., Poon et al., 2004; Hamdouch and Lawphongpanich, 2008; Hamdouch, Szeto, et al., 2014). The fixed point problem for endogenous priority lists can be formulated similarly, and any heuristic can be used to solve it iteratively.

6 Case study

6.1 Case description

We illustrate our methodology on the network of regional S-trains in canton Vaud, Switzerland, during the morning peak hour. The timetable data used in this case study has been downloaded directly from the official website of the Swiss National Railways (SBB), www.sbb.ch, for the year 2016.

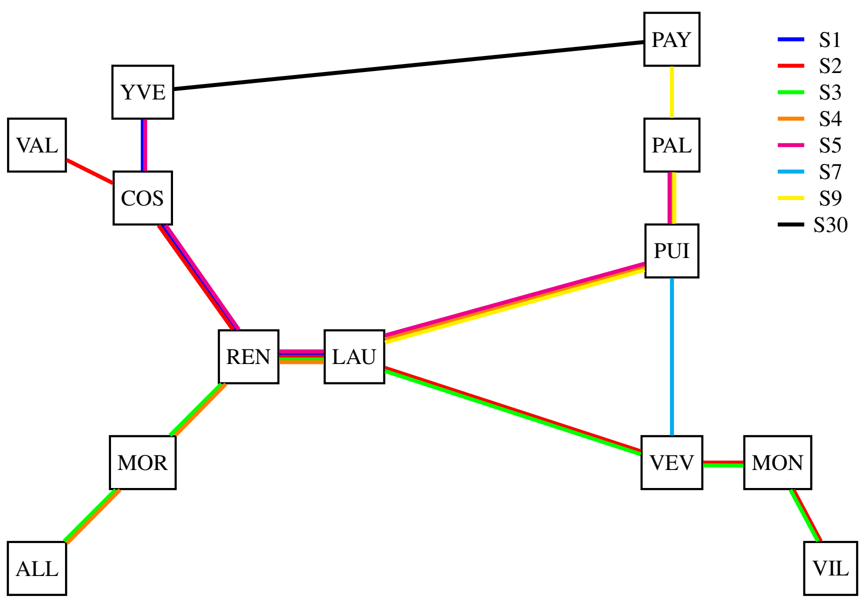

Figure 2: Network of S-trains in canton Vaud, Switzerland (2016).

The reduced network of S-trains is presented in Fig. 2. We consider the 13 main stations in this network, i.e. $S = \{\mathsf{LAU}, \mathsf{REN}, \mathsf{MOR}, \mathsf{ALL}, \mathsf{COS}, \mathsf{VAL}, \mathsf{YVE}, \mathsf{VEV}, \mathsf{MON}, \mathsf{VIL}, \mathsf{PUI}, \mathsf{PAL}, \mathsf{PAY}\}$. The timetable of the morning peak hours, between 5:00am and 9:00am, is used for this case study.

Table 3: List of S-train lines in canton Vaud, Switzerland (2016).

Line	From	To	Departure times						
S1	YVE	LAU	05:28	06:28	07:28	08:28			
	LAU	YVE	05:54	06:54	07:54	08:54			
S2	VAL	VIL	05:10	06:10	07:10	08:10			
	VIL	VAL	05:23	06:23	07:23	08:23			
S3	ALL	VIL		06:07	07:07	08.07			
	VIL	ALL	05:49	06:51	07:51	08:51			
S4	ALL	PAL	05:37	06:37	07:37	08:37			
	PAL	ALL		06:34	07:34	08:34			
S5	YVE	PAL	05:57	06:57	07:57	08:57			
	PAL	YVE		06:06	07:07	08:07			
S7	VEV	PUI		06:09	07:09	08:09			
	PUI	VEV		06:36	07:36	08:36			
S9	LAU	PAY	05:25	06:24	07:24	08:24			
	PAY	LAU	05:40	06:40	07:40	08:40			
S30	PAY	YVE	05:30	06:02	06:30	07:02	07:30	08:02	08:30
	YVE	PAY	05:04	06:04	06:33	07:04	07:33	08:04	08:33

There are 7 bidirectional lines: S1, S2, S3, S4, S5, S7, S9 and S30. We include all trains with a departure time from the beginning of the line between 5:00am and 9:00am. Table 3 reports the first and last station of every train line, along with the departure time from the first station of the line. Overall, there are 65 S-trains considered in the morning peak hours. SBB is currently operating Stadler Flirt train units on lines S1, S2, S3, S4 and S5; and Domino train units on the other lines. Stadler Flirt train units have a capacity of 160 seats and 220 standing people, while Domino units can accomodate 188 sitting and 100 standing people.

We consider a deterministic passenger demand, derived from SBB's annual report of 2015 (Swiss Federal Railways, 2015). Not all data required is available, so we rely on realistic assumptions and approximations to generate synthetic passenger data. We consider a total of 14,920 passengers in our case study. The exact procedures and assumptions used to obtain this number can be found in Appendix A. The generalized travel time of the passengers is computed using the weights given in Table 4. The cost of the penalty arc is the time horizon (four hours). We impose a minimal transfer time m of four minutes and a maximal transfer time M of fifteen minutes.

All of the computational experiments were performed on a computer with a 2.4 GHz Intel Core i7 processor and 8 GB of RAM. The algorithms were implemented in Java. On average, one realization of the assignment algorithm (Algorithm 1) runs in about two seconds. For every priority specification, the passenger assignment is simulated a thousand times (i.e., $R = 1,000$). Note that, for specification M, the computational time is slightly longer — about five seconds — because the set $\Omega(o_p, d_p)$ of all paths from origin to destination of every passenger needs to be constructed to define the priority list. In order to limit the size of $\Omega(o_p, d_p)$, we forbid paths with the following properties: (a) paths passing twice through the same station, (b) paths where passengers transfer twice in the same station, (c) paths where passengers transfer more than twice in total. Using this strategy, and given the fact that $G(V, A^p)$ is very sparse, the size of the set of paths is limited to about 20–100 paths per passenger.

Table 4: Values of weighting factors in the passengers' generalized travel time.

Parameter	Value	Unit	Reference
β_1	2.5	[min/min]	Wardman (2004)
β_2	10	[min/transfer]	de Keizer et al. (2012)
β_3	0.5	[min/min]	Small (1982)
β_4	1	[min/min]	Small (1982)

6.2 Results

The discussion of the results follows a topdown pattern: we begin by presenting aggregate passenger satisfaction indicators, such as average generalized travel time and unsatisfied demand, in Fig. 3. These allow us to compare the overall performance of the five priority specifications. In Fig. 4, we then investigate the variability among passengers for the same priority specification. In other words, instead of aggregating the results over passengers, we now consider each passenger and aggregate the performance indicators over the R realizations of the simulation. Finally, we present performance indicators at the origin-destination level in Figs. 5 and 6.

Fig. 3 presents, for the five priority specifications defined in Table 2, boxplots of the distributions of the following aggregate passenger satisfaction indicators:

- *Average travel time* (in minutes), defined as the average (over $|P|$ passengers) of the generalized travel time, for passengers who do not take the penalty path. More formally,

$$\mathrm{AVG}(r) = \frac{1}{|P|} \sum_{p \in P} \sum_{a \in A^p \setminus A^p_{Pen}} f^p_r(a) \cdot c^p_a, \quad \forall 1 \leq r \leq R.$$

- *Unsatisfied demand*, defined as the number of passengers that are not served by the system, i.e., that need to take the penalty path. Mathematically,

$$\mathrm{UNS}(r) = \sum_{p \in P} \sum_{a \in A^p_{Pen}} f^p_r(a), \quad \forall 1 \leq r \leq R.$$

- *Average passenger delay* (in minutes), defined as the average (over $|P|$ passengers) of the difference between the length of the path taken by a given passenger and the length of the shortest path for this passenger if capacity constraints are ignored. More formally,

$$\mathrm{DEL}(r) = \frac{1}{|P|} \sum_{p \in P} \left(\sum_{a \in A^p} f^p_r(a) \cdot c^p_a - C^p_{\omega^p_1} \right), \quad \forall 1 \leq r \leq R.$$

The boxplots present the 25th (Q_1), 50th and 75th (Q_3) percentiles of the distributions of the three indicators over R realizations. The upper and lower whiskers are located at $\min(\max(\cdot(r)), Q_3 + \frac{3}{2}\mathrm{IQR})$ and $\max(\min(\cdot(r)), Q_1 - \frac{3}{2}\mathrm{IQR})$ respectively, where $\mathrm{IQR} = Q_3 - Q_1$ is the interquantile

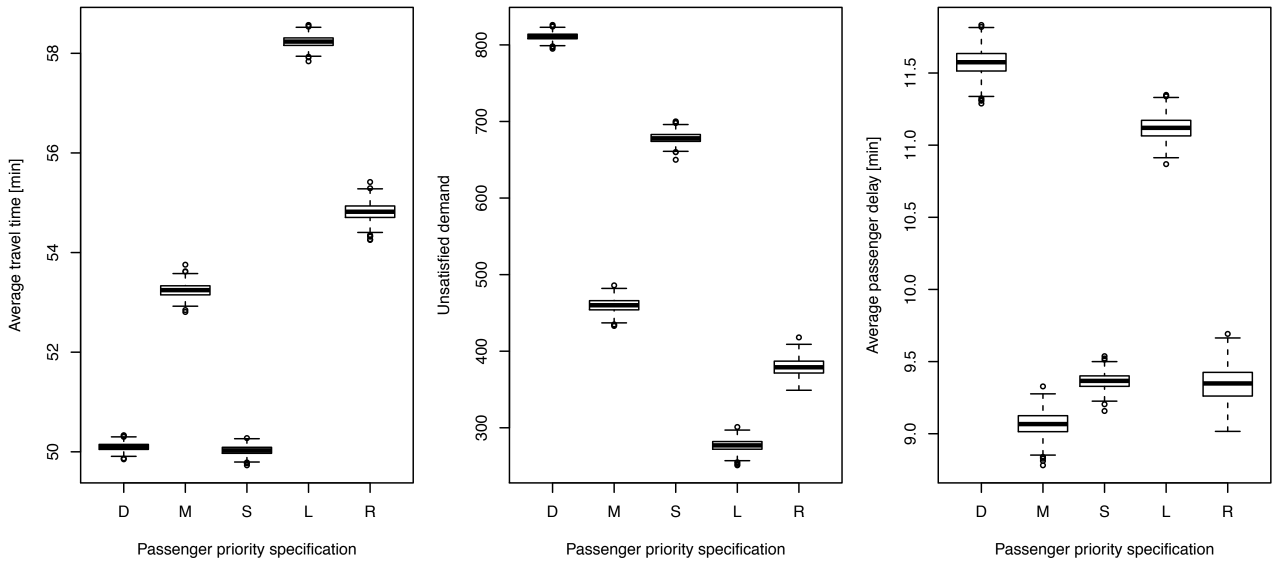

Figure 3: Distributions of aggregate passenger satisfaction indicators AVG(r), UNS(r) and DEL(r); for the five priority specifications.

range. The first observation that can be made for all specifications is the remarkable stability of the aggregate indicators across simulations. Indeed, the interquartile range shows very limited deviation from the median of the distribution and only very few outliers (i.e., realizations situated outside the whiskers) are reported. Even in the pure random case (specification R), the output of the R simulations seems to be fairly robust.

Based on this observation, we can compare the performance of the different priority specifications on an aggregate level. Specifications D and S are the ones with the lowest average travel time. At the same time, they exhibit the highest number of unserved passengers. For priority specification D, 811 passengers (5.4% of the total demand) on average need to take the penalty path, whereas 678 (about 4.5% of the total demand) take it under priority specification S. Priority specifications D and S can be seen as "greedy" assignment specifications, where passengers who arrive first (D) or who travel for a short time (S) are prioritized. It corresponds to a situation where the average travel time is minimized, but the number of passengers who are worse off is substantial. Also, priority specification D has the highest value of average passenger delay. On the other hand, priority specification L has the highest average travel time and the lowest amount of unsatisfied demand. It can therefore be viewed as a priority specification that reduces the worst cases for the passengers. The price to pay is that, on average, the travel time is the highest among all priority specifications. Finally, the performance of priority specification M is a compromise between the "greedy" specifications D and S, and priority specification L. It is also the priority specification with the lowest average passenger delay.

The aggregated passenger satisfaction indicators shown in Fig. 3 exhibit very little variability across R realizations. By contrast, Fig. 4 indicates a much higher variability for the travel time of every individual passenger across the realizations. The thick black line in the five figures (one for each priority specification) represents the average generalized travel time for every passenger (defined as

$\frac{1}{R}\sum_{1\leq r\leq R}\sum_{a\in A^p} f_r^p(a)\cdot c_a^p, \forall p \in P$), ordered by increasing average travel time. Associated with the mean, we depict in light gray the .25 and .75 quantiles of the distributions. Three patterns of variability can be observed: (i) for priority specifications L and R, the variability appears to be fairly equally distributed among all passengers (the interquartile range increases smoothly with the mean travel time); (ii) for priority specifications D and M, passengers with higher average travel time exhibit a significantly higher variability in their travel times than passengers with lower average travel time; (iii) for priority specification S, the variability is extremely low (the assignment is almost deterministic).

It is interesting to note that the variability exhibited by priority specifications S and L is not symmetric. For the former, by giving priority to short origin-destination pairs, it is expected for priority specification S to obtain the lowest interquartile range for low values of average travel time. By contrast, for priority specification L, one might expect low values of interquartile range for higher values of average travel time. This is not the case however. Indeed, all passengers appear to have a certain range of variability. This can be explained by the fact that if priority is given to long-distance passengers, short-distance passengers will usually be able to find another way to travel (incurring an acceptable increase in travel time). If however short-distance travellers "clog up" trains, long-distance travellers may not be able to find another way to travel through the system in any reasonable time.

The difference in variability between the "greedy" priority specifications D and S is also noteworthy. Fig. 3 showed the highest level of unsatisfied demand for these two priority specifications. The involved passengers are very different however. If priority specification S is applied, 227 passengers take the penalty path in every realization. In the case of priority specification D, only 42 passengers will always take the penalty path. By comparing these values to the previously reported average levels of unsatisfied demand, we conclude that, for priority specification S, about 33% of unsatisfied demand is constituted of passengers that have no other choice than taking the penalty path, whereas this number drops to 5% for priority specification D.

These insights may encourage railway operators to enforce a priority specification giving priority to short origin-destination pairs (similar to S). Thanks to an almost deterministic assignment, it is easy to pinpoint the passengers that are worse off under this priority specification. Also, low average travel times and low delays are an advantage of this priority specification.

Results at the origin-destination level are presented in Figs. 5 and 6. Each cell in these plots depicts the average over all passengers on a particular origin-destination pair. Fig. 5 compares the standard deviation (over R realizations) of the generalized travel time, for the five priority specifications. Formally, the value of a cell with origin $o \in S$ and destination $d \in S$ is given by

$$\frac{1}{|P(o,d)|}\sum_{p\in P(o,d)} \sigma\left(\sum_{a\in A^p\setminus A^p_{Pen}} f_r^p(a)\cdot c_a^p\right),$$

where $P(o,d) \subset P$ is the set of passengers traveling from o to d and $\sigma(\cdot)$ indicates the standard deviation of the distribution over R realizations. One can observe that the standard deviation of travel times is similarly distributed across all origin-destination pairs, except for two noteworthy points. First, for priority specification L, the standard deviation is very low for almost all origin-destination pairs, except for passengers ending their journey in REN or LAU (the two main stations of the case study), and for passengers starting their journey in PAY or VAL. This priority specification might therefore not

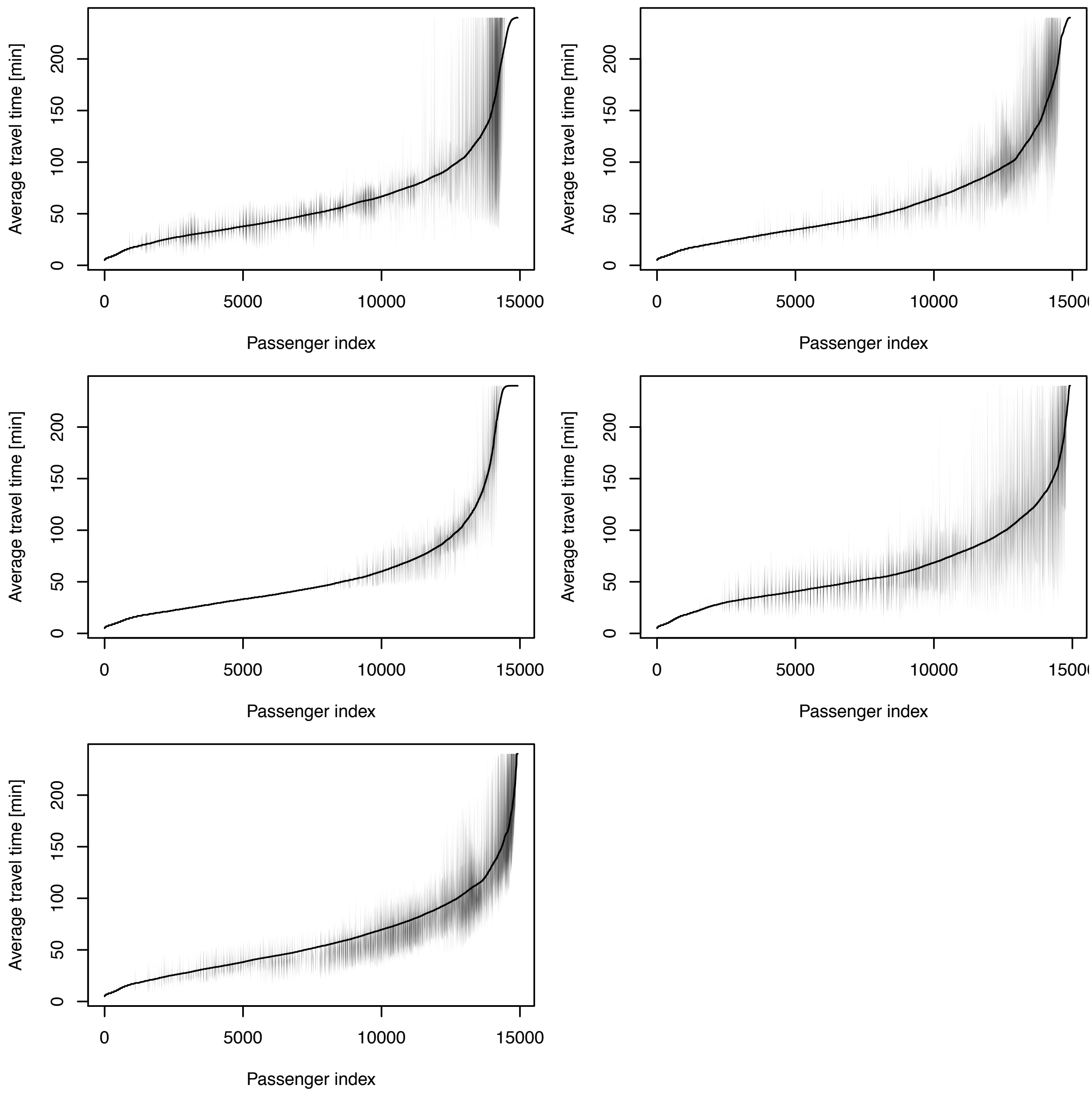

Figure 4: Mean, .25 and .75 quantile (over R realizations) of generalized passenger travel time for the five priority specifications (D top left, M top right, S mid left, L mid right, R bottom left).

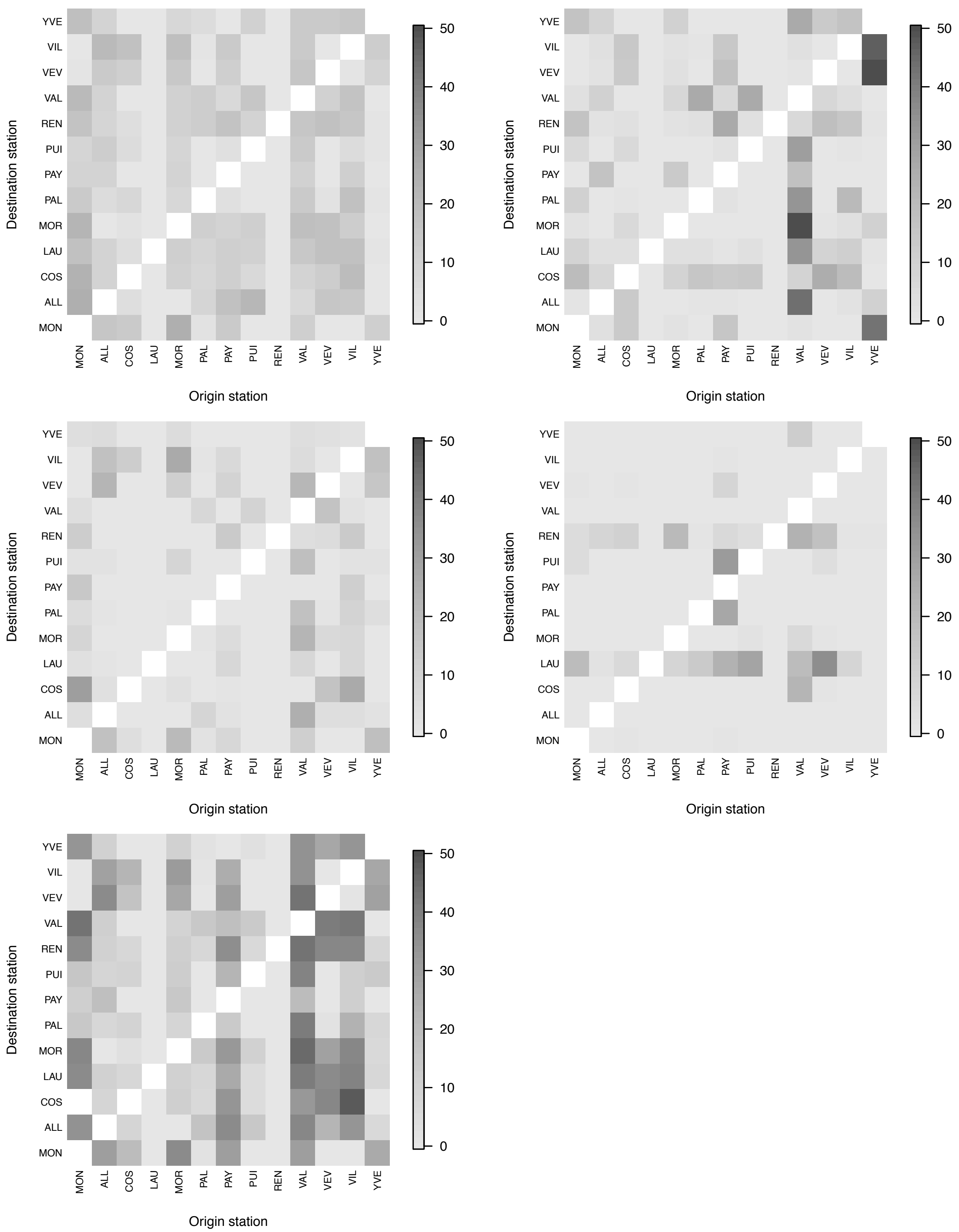

Figure 5: Average (for every origin-destination pair) of standard deviation (over R realizations) of generalized travel time, for the five priority specifications (D top left, M top right, S mid left, L mid right, R bottom left).

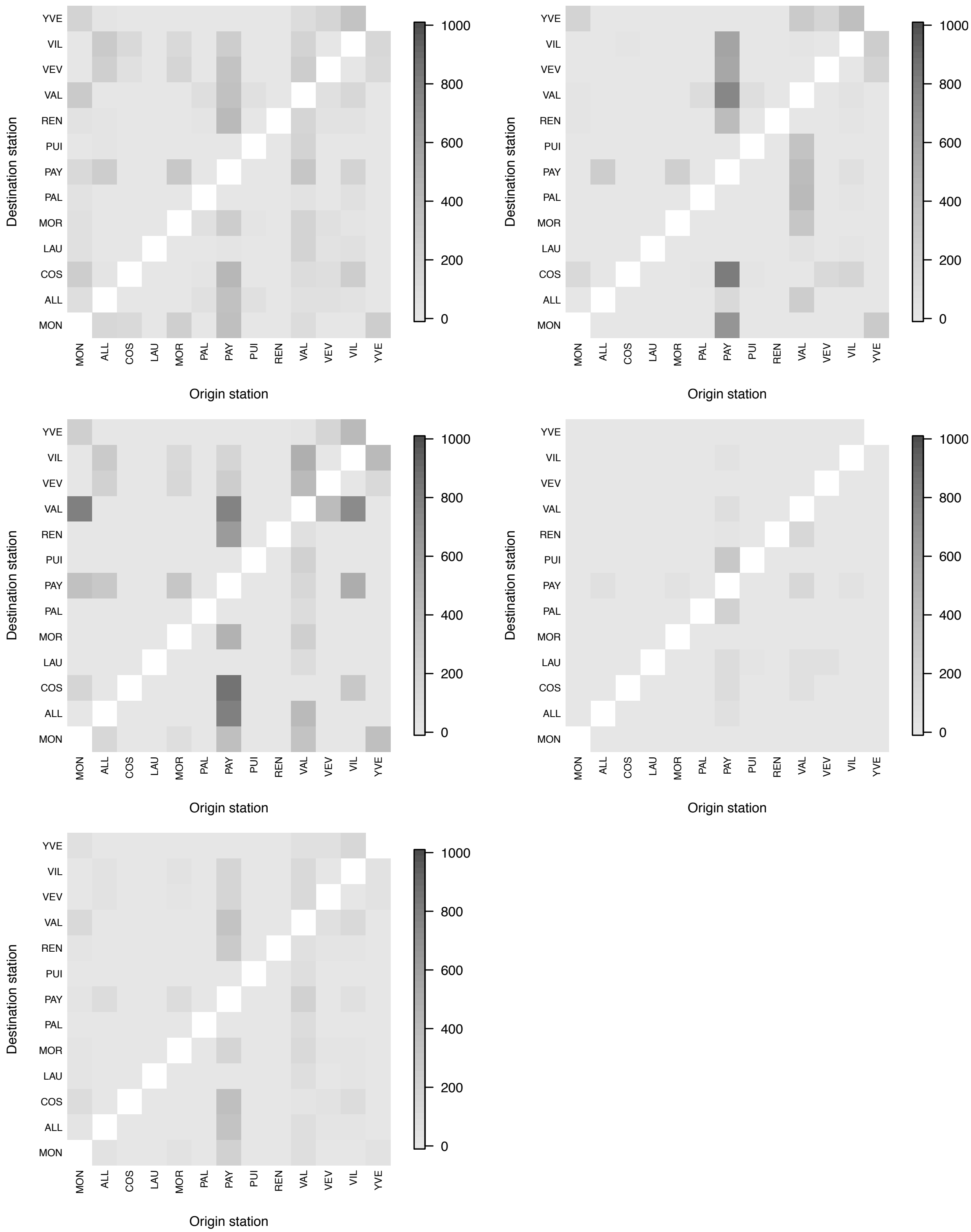

Figure 6: Average (for every origin-destination pair) of average (over R realizations) unsatisfied demand, for the five priority specifications (D top left, M top right, S mid left, L mid right, R bottom left).

be appropriate in the morning rush hours, where passengers are assumed to travel towards the main stations, with little uncertainty on the travel time they can expect. Second, the origin-destination pairs originating in station LAU or REN all have a standard deviation of 0, for any priority specification, showing that there is more than enough capacity in the trains for them.

Fig. 6 compares the average (over R realizations) unsatisfied demand, for the five priority specifications, at the origin-destination level. Mathematically, the value of a cell with origin $o \in S$ and destination $d \in S$ is given by

$$\frac{1}{|P(o,d)|} \sum_{p \in P(o,d)} \left(\frac{1}{R} \sum_{1 \leq r \leq R} \mathrm{UNS}(r) \right).$$

A clear pattern can be observed: origin-destination pairs originating from station PAY or VAL have a much higher level of unsatisfied demand. This pattern is even present for priority specifications L and R, where the overall level of unsatisfied demand is lowest. The explanation behind this phenomenon lies probably in the structure of the network: PAY and VAL are both stations situated at the periphery (see Fig. 2), therefore the train offer is less developed than in more central stations of the network. In accordance with what was observed in Fig. 5, there is no unsatisfied demand originating from station LAU or REN. In addition, the origin-destination pairs originating in the following stations are always served: PAL and PUI for priority specification S, COS and YVE for priority specification L.

6.3 Discussion and comparison with the literature

In conclusion, results show a remarkable stability of aggregate passenger satisfaction indicators (such as passenger delay, mean travel time or level of unsatisfied demand) from one simulation run to the next. When comparing the five different priority specifications, a trade-off is highlighted: giving priority to long-distance passengers minimizes the level of unsatisfied demand, while giving priority to short-distance or early passengers minimizes the average travel time of all passengers. At the individual passenger level however, substantial variability is observed. By analyzing in detail the outcome of the passenger assignment model for every priority specification, we are able to identify clearly who gains and who loses in each case.

Because of the variety of methodologies in the literature, and because of the size of solvable instances in the reported case studies, it is difficult to compare our results with a framework of the literature explicitly. In comparison with other works, our framework is one of the few that have been extensively tested on a case study of a realistic size. For instance, works such as Nguyen, Pallottino, and Malucelli (2001), Sumalee et al. (2009), Hamdouch, Ho, et al. (2011), and Hamdouch, Szeto, et al. (2014) provide numerical case studies, acknowledging that algorithmic enhancements are necessary for application to a real case study.

In the following, we discuss how our framework performs with respect to three major issues identified by Hamdouch, Szeto, et al. (2014) to assess the applicability of a model in real-life examples: computational resource requirements, convergence conditions and computational time.

Our formulation is very lean in terms of memory usage, as space-time nodes are only included in the graph for departure and arrivals of transit vehicles from/to a station. By contrast, some formulations in the literature require variables indexed by station, time, transit line and travel strategy

(Hamdouch, Szeto, et al., 2014). Obviously though, the size of the graph grows as the number of considered transit vehicles increases, thus effectively capping the total number of transit vehicles that can be considered. The number of passengers is not limiting in our formulation, as the size of the space-time graph representing the timetable does not depend on passenger demand.

As discussed in Section 5.1, the assignment algorithm we propose is guaranteed to terminate after one sweep over the set of passengers. It needs however to be embedded in a fixed point formulation if one assumes endogenous passenger demand or priority rules. Convergence issues that arise in models using MSA-type algorithms to find equilibrium solutions (e.g., Poon et al., 2004; Hamdouch and Lawphongpanich, 2008; Hamdouch, Szeto, et al., 2014) can therefore not be excluded in this case.

Low computational time may be the most important asset for the applicability of a methodology in a real-life context. Our framework performs very well in this respect. Thanks to its simplicity, and due to the fact that the priority lists are exogenous, our framework is very fast: one run of the assignment model takes about two seconds to simulate for the case study. As described throughout this paper, our model can nevertheless easily be complexified (e.g., including crowding or more involved route choice models). In that case, a trade-off between computational time and model complexity would appear. Our formulation relies on simulation to account for the stochastic nature of the priority lists. Obviously, the computational time increases with the number of required simulations. The results showed that aggregate passenger satisfaction indicators are very stable across simulations, thus limiting the number of simulations required to compute them. A higher number of simulations would however be required if one is interested in the fate of individual passengers.

7 Conclusion

In this work, we introduce a new framework for the schedule-based passenger assignment problem in capacitated public transportation networks. The originality of the framework is the use of an exogenous passenger priority list in order to decide which passengers can board the train and which cannot. The priority lists are generated in advance, based on priority rules (e.g., prioritize long-distance passengers). Stochasticity appears in the generation process of the priority rules. Therefore, the assignment is run (i.e., the system is simulated) multiples times, in order to obtain the distribution of the performance indicators. Although the framework allows for various route choice models, the use of a shortest path model allows for an efficient implementation of the framework.

Extensive computational experiments were performed on a realistic case study of the S-train network of Canton Vaud, Switzerland. One run of the assignment model takes about two seconds to simulate for our case study. Considering this short computational time, and the fact that a low number of simulations is sufficient (thanks to the stability of aggregate results), train operators can use the framework in practice to evaluate timetables from the passenger perspective in real-time. It can also be used in order to evaluate priority policies that the operator would like to enforce (e.g., giving priority to a set of "VIP" origin-destination pairs).

The assignment algorithm is able to reproduce passenger behavior properly, even though it has some limiting assumptions. As a follow-up to this work, several extensions of the algorithm are possible. An interesting extension would be to include the level of crowding of vehicles, and to solve the additional fixed point problem to obtain the passenger flows. Further, more priority rules could be defined and analyzed. For instance, designing a priority rule not based on a booking system, to model the fact that a passenger already onboard a train should have priority over a passenger wishing to board it, would be a very interesting challenge. Also, we assumed that the stochastic parts of the importance function ordering the passengers are independent and identically distributed across the passengers. Relaxing this assumption may allow to model more complex behavioral assumptions. Finally, a natural extension of this work is the inclusion of the algorithm in an iterative optimization-simulation framework, where, in each iteration, the timetable is updated and then evaluated by the assignment model.

A Data description

The number of passengers in the network has been estimated in the following manner. In 2015, Switzerland had 8,237,666 inhabitants, from which 761,446 lived in Canton Vaud[1], leading to a ratio of roughly 1:10. SBB's annual report (Swiss Federal Railways, 2015) indicates $1.21 \cdot 10^6$ passenger journeys in 2015 for whole Switzerland. By assuming that the growth rate remained the same than the previous years, this gives $1.25 \cdot 10^6$ passenger journeys in 2016. Applying the 1:10 ratio, we arrive to a demand volume of 125,000 daily passenger journeys in Canton Vaud. However, not all these journeys are realized using S-trains. Since almost all of the trains in Canton Vaud have to pass through its capital city Lausanne, we can derive the ratio between S-trains and other trains passing trough Lausanne, of 40-60. This leaves us with 50,000 daily passenger journeys using S-trains in Canton Vaud. SBB's annual report also provides hourly passenger distributions for regional services from Monday to Friday. According to this report, 30% of these journeys are realized between 5:00am and 9:00am, which gives roughly 15,000 daily passenger journeys in the morning peak hours for the S-train network in Canton Vaud.

Table 5: Hourly distribution of passenger demand, taken from Swiss Federal Railways (2015).

Hours	Percentage	Demand
5:00am - 6:00am	3%	1,500
6:00am - 7:00am	7%	3,500
7:00am - 8:00am	13%	6,500
8:00am - 9:00am	7%	3,500
Morning peak hours	30%	15,000

Table 5 reports the hourly rates of passenger demand given by the SBB report, as well as the demand obtained for our case study. Finally, this demand has been smoothed into minutes by using a non-homogeneous Poisson process to compute the desired departure time from origin for every passenger. Due to the randomness of the process, the total number of passengers in the network is 14,920.

In order to generate realistic origin-destination flows, we make the following assumptions. Lausanne is the biggest city in Canton Vaud and all the lines except S7 and S30 pass through its station (see Fig. 2). Lausanne is also used by many users to transfer to long-distance trains. We therefore assume it has the largest probability of being a destination, with half of the demand going to this station ($p(D = LAU) = 0.5$). Renens is assumed to be the city with the second highest probability of being a destination, because it is the closest station to two big Swiss universities, and also based on the network structure (all lines except S7, S9 and S30 stop there). We assume that 20% of the demand travels to Renens ($p(D = REN) = 0.2$) in the morning peak hours. We assume that the rest of the stations have equal probabilities of being a destination: $p(D = i) = 0.3/11, \forall i \in S \setminus \{LAU, REN\}$. This probability is rather small, in accordance with the assumption that most of the morning peak

[1] https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/effectif-evolution/population.assetdetail.194607.html

demand is traveling towards work and/or school places, situated in bigger cities such as Lausanne and Renens. Finally, the probability of being an origin station is uniformly distributed among all the stations (except the already selected destination station): $p(O = i) = 1/12, \forall i \in S$.

B Scale factor

The scale factor μ captures the variability of the disturbances in a choice model. There are two limiting cases that result from extreme values of μ:

- As $\mu \rightarrow 0$, the variance of the disturbances approaches infinity (see Eq. (4)). In this case, the choice model (3) provides no information, so all alternatives are equally likely. In our framework, this translates to ordering the passengers completely randomly: as $\varepsilon_p \rightarrow \infty$ in (3), the systematic part of the importance function becomes irrelevant and passenger are sorted only using random disturbances.
- As $\mu \rightarrow \infty$, the variance of the disturbances approaches zero (see Eq. (4)) and a deterministic choice model is obtained because all the information in included in the systematic utilities. In our framework, this translates to ordering the passengers deterministically: as $\varepsilon_p \rightarrow 0$ in (3), the remaining part of the importance function is the systematic part and passengers are sorted accordingly.

These two limiting cases show that a choice of μ has to be made, and this for every priority specification, in order to control the level of randomness of the ordering. Obviously, this choice will always be arbitrary. In our case, we apply the procedure described in Algorithm 2 to determine an appropriate scale factor. The main idea is to compare the order of the passengers if they are ordered by the complete importance function (U_p), or only by its systematic part (V_p). We then impose the correlation between these two orderings to be greater than a given parameter α. We obtain the scale factors reported in Table 6 for a correlation of 95%.

Table 6: Scale parameters for the different priority specifications, with a correlation of $\alpha = 0.95$.

Priority specification ρ	μ_ρ
D	0.10
M	0.28
S	0.16
L	0.16

Algorithm 2: Determination of an appropriate scale factor for the different priority specifications.

Input: Set of passengers $p \in P$
Priority rule $\rho \in \{D, M, S, L\}$
Required correlation α
Output: Scale factor μ_ρ

$\mu_\rho := 0$
while $\mu < 10$ **do**
  foreach $p \in P$ **do**
    $V_p := V_p(\rho)$, according to Table 2
    $\Delta(p) = V_p$
    Draw $r \sim U(0, 1)$
    $\varepsilon_p = -\frac{1}{\mu_\rho} \ln(-\ln r)$
    $U_p = V_p + \varepsilon_p$
    $\Gamma(p) = U_p$
  Let V_Δ be the vector of systematic utilities of passenger in P ordered according to Δ
  Let V_Γ be the vector of systematic utilities of passenger in P ordered according to Γ
  if $\mathrm{corr}(V_\Delta, V_\Gamma) \geq \alpha$ **then**
    return μ_ρ
  else
    $\mu_\rho := \mu_\rho + 0.01$

References

Binder, S., Y. Maknoon, and M. Bierlaire (forthcoming). "The multi-objective railway timetable rescheduling problem." In: *Transportation Research Part C: Emerging Technologies*. Accepted for publication.

Cominetti, R. and J. Correa (2001). "Common-Lines and Passenger Assignment in Congested Transit Networks." In: *Transportation Science* 35.3, pp. 250–267.

De Cea, J. and E. Fernández (1993). "Transit Assignment for Congested Public Transport Systems: An Equilibrium Model." In: *Transportation Science* 27.2, pp. 133–147.

De Keizer, B., K. Geurs, and G. Haarsman (2012). "Interchanges in Timetable Design of Railways: A Closer Look at Customer Resistance to Interchange between Trains." In: *40th European Transport Conference*. Glasgow.

Fu, Q., R. Liu, and S. Hess (2012). "A Review on Transit Assignment Modelling Approaches to Congested Networks: A New Perspective." In: *Procedia - Social and Behavioral Sciences* 54, pp. 1145–1155.

Hamdouch, Y., H. Ho, A. Sumalee, and G. Wang (2011). “Schedule-Based Transit Assignment Model with Vehicle Capacity and Seat Availability.” In: *Transportation Research Part B: Methodological* 45.10, pp. 1805–1830.
Hamdouch, Y. and S. Lawphongpanich (2008). “Schedule-Based Transit Assignment Model with Travel Strategies and Capacity Constraints.” In: *Transportation Research Part B: Methodological* 42 (7-8), pp. 663–684.
Hamdouch, Y., W. Szeto, and Y. Jiang (2014). “A New Schedule-Based Transit Assignment Model with Travel Strategies and Supply Uncertainties.” In: *Transportation Research Part B: Methodological* 67, pp. 35–67.
Kroon, L., G. Maróti, and L. Nielsen (2015). “Rescheduling of Railway Rolling Stock with Dynamic Passenger Flows.” In: *Transportation Science* 49.2, pp. 165–184.
Nguyen, S. and S. Pallottino (1988). “Equilibrium Traffic Assignment for Large Scale Transit Networks.” In: *European Journal of Operational Research* 37.2, pp. 176–186.
Nguyen, S., S. Pallottino, and F. Malucelli (2001). “A Modeling Framework for Passenger Assignment on a Transport Network with Timetables.” In: *Transportation Science* 35.3, pp. 238–249.
Nielsen, O. A. (2004). “A Large Scale Stochastic Multi-Class Schedule-Based Transit Model with Random Coefficients.” In: *Schedule-Based Dynamic Transit Modeling: Theory and Applications*. Ed. by N. H. M. Wilson and A. Nuzzolo. Vol. 28. Boston, MA: Springer US, pp. 53–77.
Nuzzolo, A., U. Crisalli, and L. Rosati (2012). “A Schedule-Based Assignment Model with Explicit Capacity Constraints for Congested Transit Networks.” In: *Transportation Research Part C: Emerging Technologies* 20.1, pp. 16–33.
Nuzzolo, A., F. Russo, and U. Crisalli (2001). “A Doubly Dynamic Schedule-Based Assignment Model for Transit Networks.” In: *Transportation Science* 35.3, pp. 268–285.
Papola, N., F. Filippi, G. Gentile, and L. Meschini (2009). “Schedule-Based Transit Assignment: New Dynamic Equilibrium Model with Vehicle Capacity Constraints.” In: *Schedule-Based Modeling of Transportation Networks: Theory and Applications*. Boston, MA: Springer US, pp. 1–26.
Poon, M., S. Wong, and C. Tong (2004). “A Dynamic Schedule-Based Model for Congested Transit Networks.” In: *Transportation Research Part B: Methodological* 38.4, pp. 343–368.
Small, K. A. (1982). “The Scheduling of Consumer Activities: Work Trips.” In: *The American Economic Review* 72.3, pp. 467–479.
Spiess, H. and M. Florian (1989). “Optimal Strategies: A New Assignment Model for Transit Networks.” In: *Transportation Research Part B: Methodological* 23.2, pp. 83–102.
Sumalee, A., Z. Tan, and W. H. Lam (2009). “Dynamic Stochastic Transit Assignment with Explicit Seat Allocation Model.” In: *Transportation Research Part B: Methodological* 43 (8-9), pp. 895–912.
Swiss Federal Railways (2015). *SBB: Facts and Figures 2015*.
Tirachini, A., D. A. Hensher, and J. M. Rose (2013). “Crowding in Public Transport Systems: Effects on Users, Operation and Implications for the Estimation of Demand.” In: *Transportation Research Part A: Policy and Practice* 53, pp. 36–52.
Tong, C. and S. Wong (1999). “A Stochastic Transit Assignment Model Using a Dynamic Schedule-Based Network.” In: *Transportation Research Part B: Methodological* 33.2, pp. 107–121.
Wardman, M. (2004). “Public Transport Values of Time.” In: *Transport Policy* 11.4, pp. 363–377.