CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Erasmus University Digital Repository

Railway timetabling with integrated passenger distribution

Johann Hartleb*[1,2] and Marie Schmidt†[1]

[1]Rotterdam School of Management and Erasmus Center for Optimization in Public Transport, Erasmus University Rotterdam, The Netherlands
[2]Institute for Road and Transport Science, University of Stuttgart, Germany

Abstract

Timetabling for railway services often aims at optimizing travel times for passengers. At the same time, restricting assumptions on passenger behavior and passenger modeling are made. While research has shown that passenger distribution on routes can be modeled with a discrete choice model, this has not been considered in timetabling yet. We investigate how a passenger distribution can be integrated into an optimization framework for timetabling and present two mixed-integer linear programs for this problem. Both approaches design timetables and simultaneously find a corresponding passenger distribution on available routes. One model uses a linear distribution model to estimate passenger route choices, the other model uses an integrated simulation framework to approximate a passenger distribution according to the logit model, a commonly used route choice model. We compare both new approaches with three state-of-the-art timetabling methods and a heuristic approach on a set of artificial instances and a partial network of Netherlands Railways (NS).

Keywords: Transportation; timetabling; public transport; route choice; discrete choice model; passenger distribution

*hartleb@rsm.nl, johann.hartleb@isv.uni-stuttgart.de
†schmidt2@rsm.nl

Highlights

- We propose a novel timetabling approach with integrated passenger distribution model
- Two mixed integer linear programs for this problem are developed
- One uses a linear distribution model, the other a simulation of passenger distribution
- One integrates a linear distribution, the other a simulation of passenger distribution
- We compare our models/programs in experiments to state-of-the-art timetabling methods
- Integrating a passenger distribution model can help to find better timetables

Declarations of interest

none

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1 Introduction

Public transport is important to our society for various reasons, such as increased mobility for the general public or lower air pollution compared to individual transport. Especially the potential of public transport to reduce emissions is recently much discussed in the context of climate change. To be considered an alternative to individual transport, public transport has to be as attractive as possible to passengers. Since decades both researchers and practitioners work on the improvement of public transport from different perspectives using various approaches. Most of them follow the same pattern and design public transport sequentially. First, long term planning decisions are taken, such as stop location planning and, in case of railways, network design. Afterwards, the line routes are designed and the corresponding frequencies of lines are fixed. On the tactical level, a timetable is determined, based on the results of the previous steps. Finally, vehicles and crew are scheduled.

Finding a good timetable is an integral step for providing high quality public transport services to passengers. Next to driving times of vehicles, the timetable determines the transfer times and therefore the travel times of passengers. Since transfer and travel times have a significant effect on the chosen routes of passengers and also their satisfaction with public transport, timetabling is a relevant problem with high practical impact. Also from an algorithmic perspective timetabling is an interesting task. It has been shown that finding a feasible periodic timetable is NP-complete, even when omitting safety constraints. For this reason, research often focuses on efficient solution strategies. In recent years, many publications deal with the question how passenger travel time can be used as an objective to guide the search for good timetables.

At the design of public transport systems a good trade-off between service quality and costs for operating a public transport service has to be found. Since costs are mainly determined by the line plan as well as the vehicle and crew schedule, in a sequential approach many optimization approaches for timetabling solely aim at providing best quality to passengers. Even though the focus is on the quality for passengers, strong assumptions on passenger demand are made. Among them, two assumptions are commonly found: First, all passengers travel on their shortest available route and, second, a predetermined passenger assignment to routes is sufficient to estimate passenger loads in the public transport network. In this context a passenger route defines when and on which lines passengers travel. As summarized in Table 1.1, the impact of each of these two assumptions has been studied individually and improvements could be achieved by considering a passenger distribution on multiple routes and by integrating a shortest route search into optimization, respectively.

Motivated by these findings, we relax both assumptions at the same time. We study the problem of finding a travel-time minimal timetable under the assumption that passengers' route choice can be modeled using a discrete choice model. To our best knowledge, this is the first time that a choice model is used to derive a passenger distribution within a timetable optimization model.

Depending on the quality of all available routes, choice models estimate the probability that a route is chosen by passengers which gives a passenger distribution in the network. We use the logit model to estimate passenger distributions on available routes, which is a commonly used passenger route choice model in transport applications, and incorporate it in an optimization framework for timetabling. Due

	Predetermined route choice	Integrated route choice
Single route	Liebchen (2018) and many others	Gattermann et al. (2016) Borndörfer et al. (2017) Löbel et al. (2019)
Distribution	Parbo et al. (2014) Sels (2015) Robenek et al. (2016)	this paper

Table 1.1: Selection of timetabling papers, categorized by (1) whether a predetermined route choice is assumed or a route choice model is integrated and (2) whether it is assumed that passengers use a single route only or distribute on multiple routes. The mentioned papers are discussed together with other related literature in Section 2.

to the nonlinear structure of the logit model, the mathematical program for this problem is intractable. We present two ways to integrate a passenger distribution on multiple routes into a timetabling model as a linear formulation. Our first model uses a novel linear distribution model. This distribution model is designed to have the same characteristics as the logit model and due to its linear formulation it can easily be incorporated into an optimization model. The second model relies on a simulation of the logit distribution of passengers. By considering multiple scenarios, the distribution of passengers according to a logit model can be approximated within an optimization model that is linear in all variables.

We aim at maximizing the quality of timetables for passengers. In research and practice a variety of ways to evaluate timetables from passengers' perspective is used. However, not all of them are suitable to be used as objective in an optimization program and Hartleb et al. (2019) showed in an empirical comparison that they do not necessarily yield a consistent evaluation. To best reflect the quality of the found solutions, we evaluate all timetables in our experiments with multiple evaluation functions. As objective, the first model uses the concept of absolute travel time to minimize the time spent in the public transport system, which follows common practice in timetabling literature. In the second model the simulated travel times are minimized to also incorporate passengers' preferences that are not captured by absolute travel times only. We discuss theoretical properties of the chosen objective functions of the two models and analyze their influence on the resulting timetable in the experiments. This shows that the absolute travel time, although commonly used in literature, might not be suitable for evaluating timetables when considering multiple alternative routes for passengers.

We compare our models for timetabling with integrated passenger distribution with four timetabling methods motivated by approaches in the literature. Two of these methods assume that a passenger assignment to routes is fixed before optimizing the timetable, using either a single route for all passengers traveling between the same stations or a distribution on multiple routes. Another method finds optimal timetables based on the assumption that passengers use shortest available routes. A fourth approach solves the problem of timetabling with integrated passenger distribution heuristically by iterating between assigning passengers to routes according to the logit model and finding optimal timetables. The experiments show that the two proposed models are capable of finding better solutions than the benchmark approaches. The found timetables performed better with respect to some evaluation functions while

being of comparable quality with respect to other evaluation functions when compared to the timetables found by existing methods. These improvements come at the expense of increased complexity of the models. From this we conclude that the integration of a passenger distribution model has the potential to find better timetables for passengers, but more efficient solution strategies have to be developed.

We want to highlight two contributions of this paper: First, we present a novel timetabling model with an integrated choice model to derive a passenger distribution on multiple routes. We provide and discuss linear representations of the passenger distribution and develop two linear timetabling programs. Second, we show on multiple artificial instances and a partial real-world network the advantages and disadvantages of the novel approaches when compared to state-of-the-art methods.

The remainder is structured as follows. We summarize the literature on passenger distribution models, optimization approaches for timetabling and on the evaluation of timetables in Section 2. In Section 3, the basic models relevant for this paper are introduced and the problem is defined. In Section 4, we develop and discuss two linear timetabling models with an integrated passenger distribution model. Section 5 describes the experimental setup such as considered instances, benchmark methods and used evaluation functions. We report and discuss our results of the experiments in Section 6 and conclude in Section 7.

2 Related literature

2.1 Passenger Route choice

State-of-the-art discrete choice models provide appropriate solutions for describing passengers' behavior concerning mode and route choices (de Dios Ortuzar and Willumsen, 2011). A choice model estimates which alternative is chosen by an individual given the utilities of all alternatives. Ben-Akiva and Lerman (1985) give in their book a comprehensive overview of the theory of choice models. In aggregate form, the chosen routes of individual passengers correspond to a distribution of all passengers in the public transport network. For estimating passenger distributions in public transport applications the logit model is most commonly applied. To adjust to certain requirements, the logit model is constantly developed further, for example Espinosa-Aranda et al. (2018) propose a constrained nested logit model to model passenger distributions on routes in public transport. Since recently, choice models in general and the logit model in specific are applied in optimization approaches for public transport applications. Canca et al. (2019) use it to estimate a passenger distribution and mode choice in the context of transit network planning. The resulting nonlinear program was solved with a neighborhood search based matheuristic. Due to the non-linear structure, exact solution approaches rely mostly on a linearization of the logit model. De-Los-Santos et al. (2017) developed a linear approximation by using that one alternative with fixed utility is available. An overview of common linearizations of the logit model is given by Haase and Müller (2014).

One interpretation of choice theory is that each alternative is perceived differently by people. This is usually modeled by adding an error term to the deterministic utility of alternatives. The error terms are used as an unknown part of the utility in many choice models. They model different sources of uncertainty and imperfect knowledge of analysts, such as unobserved route attributes, unobserved passenger preferences or measurement errors (Ben-Akiva and Lerman, 1985). The distribution of the error

terms determines a certain choice model. For example, independent and identically normal distributed error terms yield a probit model and independent and identical Gumbel distributed error terms yield a logit model. Drawing random terms from a certain distribution, the corresponding choice model can be simulated (Train, 2009). Such a simulation framework was used in Pacheco et al. (2016), where optimal pricing strategies for different parking options considering passenger behavior were computed.

2.2 Timetabling

Timetabling approaches for public transport applications are usually classified in periodic and aperiodic cases. As we aim at finding a periodic timetable, we focus on the periodic timetabling literature. Most formulations are based on the periodic event scheduling problem (PESP) as introduced by Serafini and Ukovich (1989) or the cyclic periodicity formulation (CPF), which is a further development of the PESP model by Nachtigall (1994). While the PESP has one variable for each event modeling points in time, the CPF uses one variable for each activity expressing a time duration.

Serafini and Ukovich (1989) showed that the problem of finding a periodic timetable is NP-complete and many publications focus on finding efficient ways to solve periodic timetabling. Schrijver and Steenbeek (1994) developed a constraint propagation algorithm which later on served as a basis for one of the first successful implementations of a timetable found with methods of Operations Research (Kroon et al., 2009). A powerful heuristic to solve the PESP model is the modulo network simplex algorithm developed by Nachtigall and Opitz (2008). The algorithm is inspired by the simplex algorithm for solving linear programs where a feasible solution is improved in each iteration of exchanging a basis and a non-basis variable. Liebchen (2018) describes how the special structure of a PESP instance can be exploited to derive effective preprocessing techniques that reduce the complexity of the timetabling problem. An overview of common models and solution methods for railway timetabling is given in Borndörfer et al. (2018).

Originally introduced as a feasibility program, the PESP model was quickly extended by objective functions to guide the optimization. Recent publications often aim at designing timetables timetables with minimal passenger travel time or with lowest energy consumption during operation. We refer to Scheepmaker et al. (2017) for a summary of energy efficient timetabling approaches and focus on passenger travel time. However, passenger travel time as objective is usually modeled making two restrictive assumptions on passenger behavior that have been shown to distort the search for an optimal solution.

First, passengers are usually assigned to routes in the transport network before the timetable optimization. With this passenger assignment to routes, the arcs in the network are assigned weights in order to take passenger routes into consideration during optimization in a heuristic way. Many publications have challenged this assumption and shown that the routes passengers use depend on the timetable (Schmidt, 2014) and therefore cannot be reliably determined beforehand. To take passengers reaction on the designed public transport into consideration, Nachtigall (1998) and Siebert and Goerigk (2013) experimented with alternately finding shortest route assignments of passengers and optimizing a timetable given the updated passenger routes. Schmidt and Schöbel (2015) integrated a shortest route search for passengers into the timetabling optimization model and further improved the quality of timetables

found. They used that the exact route of passengers does not need to be known in the aperiodic case since start and end events contain sufficient information for travel time computation. With this trick, the resulting timetabling model with integrated passenger assignment to shortest routes could be solved efficiently. Borndörfer et al. (2017) introduced the shortest route search also for periodic timetabling and found significantly improved transfer waiting times for passengers. A different solution approach to periodic timetabling with integrated shortest route search was described in Gattermann et al. (2016). They used time slices to model departure time preferences and defined a translation of the integrated model to a satisfiability problem. Upper and lower bounds for the timetabling problem with integrated shortest route search are discussed in Schiewe and Schöbel (2018). Recently, Löbel et al. (2019) proposed an adjustment of the modulo simplex algorithm to incorporate a shortest passenger route search during optimization. Assuming that passengers always take the next available train in a high frequency network, Polinder et al. (2019) integrated a route selection of passengers in a PESP model.

Second, for the design of a majority of timetable objective functions it is assumed that passengers only travel on the shortest route. Van der Hurk et al. (2014) concluded from their study of smart card based travel data that this is one of the common misassumptions on passenger behavior. Many publications challenged this assumption and proposed enhanced models to develop better timetables for passengers. As input to their timetabling model, Sels (2015) described a passenger assignment to routes that are at most 20% longer than the potentially shortest route. Robenek et al. (2016) used estimates for utilities of available connections as defined for choice models together with time dependent demand structures to estimate a distribution of passengers. A similar approach was used by Parbo et al. (2014) for deriving passenger distributions, who updated the passenger distribution after each timetable computation. As mentioned in the literature review on passenger choice models in Section 2.1, first choice models were integrated in optimization approaches of other public transport applications. To the best of our knowledge, other choice models for passenger route choice than a shortest route search were not integrated in an optimization framework for timetabling, which is done in this paper.

2.3 Timetable evaluation

As discussed in Section 2.2, the majority of publications in Operations Research use the absolute travel time of passengers on predetermined routes as objective. This is due to the simple structure of this evaluation function which makes it suitable for optimization. In other research areas, timetables are usually evaluated differently. For evaluation purposes in Transport Engineering, often the perceived travel time is used. That is a weighted travel time equivalent that incorporates more factors of influence such as penalties for transfers, fares or adaption time (de Dios Ortuzar and Willumsen, 2011). In contrast to that, commonly applied choice models use an evaluated utility to measure the quality of a timetable for passengers. This evaluated utility is usually a non-linear function of a weighted travel time equivalent such as the perceived travel time. Recently, evaluated utilities are often proposed as replacement for established evaluation functions. For example, de Jong et al. (2007) summarized literature on logsums, an evaluated utility, and showcased the advantages of this evaluation in a case study on high speed trains in the Netherlands. Indeed, Hartleb et al. (2019) showed that timetable evaluation functions do not yield

consistent evaluation results also on realistic networks, although the functions are all designed to evaluate the quality of timetables for passengers. This suggests that timetables should be evaluated from different perspectives.

3 Problem definition

In this section we define the problem of timetabling with an integrated passenger distribution on multiple routes. To this end, we give a basic formulation for both problems: timetabling assuming that a passenger assignment to routes is given, and route choice modeling assuming that a timetable is known.

All formulations are based on an event activity network $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ with a set of events $\mathcal{E}$ and a set of activities $\mathcal{A}$. In this context, an event $i \in \mathcal{E}$ denotes an arrival or a departure of a vehicle at a station, and an activity $ij \in \mathcal{A}$ represents a drive or a wait activity of a vehicle between two events $i \in \mathcal{E}$ and $j \in \mathcal{E}$. Activities can possibly be used to model more than vehicle actions, for example transfer activities of passengers and headway or synchronization constraints between vehicles (Liebchen and Möhring, 2007).

3.1 Passenger distribution

Discrete choice models can be used to describe passengers' behavior concerning route choice when a timetable is known. We use the logit model to estimate a distribution of the passengers on their routes. The passenger routes in the public transport network are represented by paths in the event activity network. A path $p = (i_1, \ldots, i_{m_p})$ is a sequence of events i in the event activity network such that two consecutive events are connected by a drive, wait or transfer activity. We denote the perceived travel time of path p by t_p. The perceived travel time is a weighted linear combination of the influencing factors such as travel time and number of transfers and is often interpreted as a negative utility of the path p.

Let a fixed set P of alternative paths with perceived travel times t_p for all paths $p \in P$ be given. Then, the logit model can be interpreted as a probability function w_p^{lm} that assigns a probability to alternative p, based on the utility of all considered alternatives.

$$w_p^{lm}((t_q)_{q \in P}) = \frac{e^{\beta t_p}}{\sum_{q \in P} e^{\beta t_q}}, \tag{3.1}$$

where $(t_q)_{q \in P}$ is a vector containing the utilities of all paths in the set P. With the scalar $\beta \in \mathbb{R}$ the logit model can be adjusted to suit the specific instance.

3.2 Timetabling

In the literature an instance $\mathcal{I} = (\mathcal{N}, l, u, OD)$ for a timetabling problem usually consists of an event activity network $\mathcal{N}$ with lower and upper bounds l and u on the activities, as well as a demand matrix OD indicating how many passengers wish to travel from each origin to destination. It remains to find arrival and departure times for each vehicle on each line. We focus on the cyclic periodicity formulation for periodic timetabling problems as described in Nachtigall (1994). This integer linear formulation is based

on an event activity network with constraints ensuring that the duration $\delta_{ij} \in \mathbb{Z}_+$ of each activity $ij \in \mathcal{A}$ is between a given lower l_{ij} and upper bound u_{ij}, i.e.,

$$l_{ij} \leq \delta_{ij} \leq u_{ij} \quad \forall ij \in \mathcal{A}. \tag{3.2}$$

We assume that the timetable has an accuracy of one time unit and, therefore, the durations between events are integer valued. To ensure that the durations δ can be transformed to a feasible timetable that assigns a point in time to each event, additionally cycle constraints need to be added to the model (Nachtigall, 1994). It is sufficient to include the cycle constraints for each cycle c in a cycle basis $\mathcal{C}$ of the event activity network. We therefore add

$$\Gamma_c \delta = T \cdot \mu_c \quad \forall c \in \mathcal{C} \tag{3.3}$$

to the constraints, using a integer cycle variable $\mu_c \in \mathbb{Z}$. Each vector Γ_c indicates all edges in forward or backward direction in cycle c and T denotes the length of the period. The objective of most timetabling formulations is to minimize the total travel time of passengers. This is mostly achieved with the help of passenger weights x_{ij} on each activity ij and by minimizing

$$\sum_{ij \in \mathcal{A}} x_{ij} \cdot \delta_{ij}. \tag{3.4}$$

Note that the passenger weights x_{ij} are predetermined by assigning passengers to routes before optimization. The cyclic periodicity formulation for timetabling with predetermined passenger routes uses Constraints (3.2) and (3.3) and is given by

$$\begin{array}{llcll}
\min & \sum_{ij \in \mathcal{A}} x_{ij} \cdot \delta_{ij} & & & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
& \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
& \delta_{ij} & \in & \mathbb{Z}_+ & \forall i \in \mathcal{E} \\
& \mu_c & \in & \mathbb{Z}_+ & \forall c \in \mathcal{C}
\end{array}$$

3.3 Integration of passenger distribution and timetabling

Section 3.1 gives the definition of the logit model to estimate passengers' route choice for a given timetable and Section 3.2 provides a standard model to optimize a timetable for a predetermined passenger route choice. Since the result of one model is the input for the other and vice versa, we aim at developing a model integrating both aspects.

We assume that a finite choice set P_k of n_k possible paths for each OD pair k is given. Each path $p \in P_k$ is a sequence of events in the event activity network that could possibly be taken by the passengers of OD pair k. The passenger weight x_{ij} on each activity ij is not assumed to be predetermined as in the timetabling program introduced in Section 3.2, but we derive it from the distribution on the paths. To

this end, we compute the respective lengths

$$t_p = \sum_{ij \in p} \delta_{ij} \quad \forall p \in P_k,\ \forall k \in OD \tag{3.5}$$

of each path for all OD pairs. Note, that the definition of t_p can easily be extended by additional external influencing factors such as a fare for taking path p or a penalty for each transfer included in path p. Given t_p, we can use the logit distribution w_p^{lm} to compute a share of each OD pair using the path p. Multiplied by the number of passengers o_k of OD pair k, this yields the number of passengers on each activity ij using the path p, which we denote by

$$x_{ij}^p = w_p^{lm}((t_q)_{q \in P_k}) \cdot o_k \quad \forall ij \in p,\ \forall p \in P_k,\ \forall k \in OD. \tag{3.6}$$

Note, that this is an expected value that does not have to be integer. Aggregating these numbers over all paths p for each OD pair, we obtain the number of passengers on each activity

$$x_{ij} = \sum_{k \in OD} \sum_{p \in P_k} x_{ij}^p \quad \forall ij \in \mathcal{A}. \tag{3.7}$$

As in the timetabling formulation from Section 3.2, this number is used in the objective function to find a travel time minimal timetable. We formulate a general optimization problem for timetabling assuming that a passenger distribution can be modeled with a logit model:

$$\begin{array}{llcll}
\min & & & \displaystyle\sum_{ij \in \mathcal{A}} x_{ij} \cdot \delta_{ij} & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
 & \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
 & \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
 & t_p & = & \sum_{ij \in p} \delta_{ij} & \forall p \in P_k,\ \forall k \in OD \\
 & x_{ij}^p & = & w_p^{lm}((t_q)_{q \in P_k}) \cdot o_k & \forall ij \in p,\ \forall p \in P_k,\ \forall k \in OD \\
 & x_{ij} & = & \sum_{k \in OD} \sum_{p \in P_k} x_{ij}^p & \forall ij \in \mathcal{A} \\
 & \delta_{ij} & \in & \mathbb{Z}_+ & \forall i \in \mathcal{E} \\
 & \mu_c & \in & \mathbb{Z}_+ & \forall c \in \mathcal{C} \\
 & x_{ij} & \in & \mathbb{R}_+ & \forall ij \in \mathcal{A} \\
 & x_{ij}^p & \in & [0, o_k] & \forall ij \in \mathcal{A},\ \forall p \in P_k,\ \forall k \in OD \\
 & t_p & \in & \mathbb{Z}_+ & \forall p \in P_k,\ \forall k \in OD
\end{array}$$

Note that this formulation is not tractable due to the chosen passenger distribution function w_p^{lm}. Furthermore, the objective is nonlinear in the variables since the passenger loads x are modeled to be dependent on the durations δ.

4 Models

Already Parbo et al. (2014) argued that the problem from Section 3.3 is "extremely difficult to solve mathematically, since the timetable optimisation is a non-linear non-convex mixed integer problem, with passenger flows defined by the route choice model, where the route choice model is a non-linear non-continuous mapping of the timetable." In this section, we describe two different representations of the route choice model and introduce linear formulations for the problem of finding travel-time minimal routes under the assumption that passengers' routes choice can be modeled using a logit model.

4.1 Model 1 - Timetabling with linear distribution model

The model from Section 3.3 is not tractable because of the integration of the non-linear formulation of the logit model to derive a passenger distribution. In a first model, we use a novel linear passenger distribution model that is developed inspired by characteristics of the logit model. Furthermore, the quadratic objective is linearized. We address these two details in the following two sections and provide a linear formulation for timetabling with integrated passenger distribution model.

4.1.1 Linear distribution model

Using the nonlinear analytic expression of the logit distribution from Equation (3.1) as distribution model in the program of Section 3.3 yields an intractable optimization program. The literature provides multiple linearizations of the logit model for applications in Operation Research. To our best knowledge, these linearizations can be classified into two cases. Either, just the utility of a single alternative is variable while the utilities of all remaining alternatives are fixed. Or, the utilities of all alternatives for customers are fixed and the decision is whether to offer alternatives or not. Since in our case all alternative paths are always available and their utility depends on the timetable, these linearizations are not appropriate.

Therefore, we develop a linear distribution model in order to approximate the logit model. Our model allows all utilities to be flexible in their domain, i.e., $t_p \in [\underline{m}_k, \overline{m}_k]\ \forall p \in P_k$, and satisfies the probability characteristics. For each OD pair $k \in OD$ we require the following five characteristics

Distribution characteristics

$$w_p((t_q)_{q \in P_k}) \in [0,1] \text{ and } \sum_{p \in P_k} w_p((t_q)_{q \in P_k}) = 1 \tag{4.1}$$

Monotonicity

Let $|P_k| > 1$, let $\varepsilon > 0$ and let e_p be the unit vector with a 1 at the position of path p. Then

$$w_p((t_q)_{q \in P_k} + \varepsilon \cdot e_p) < w_p((t_q)_{q \in P_k}) \tag{4.2}$$

Uniform distribution on equivalent alternatives

$$w_p((t, \dots, t)) = \frac{1}{n}, \tag{4.3}$$

where n is the number of alternatives.

Independence of order

Let $\pi_p: P_k \to P_k$ be any permutation on a set of paths P_k that keeps the path p constant, i.e., $\pi_p(p) = p$. Then

$$w_p((t_q)_{q \in P_k}) = w_p((t_{\pi_p(q)})_{q \in P_k}) \tag{4.4}$$

Logit characteristic: absolute differences in utility determine probability

$$w_p((t_q + \hat{t})_{q \in P_k}) = w_p((t_q)_{q \in P_k}) \tag{4.5}$$

This yields a family of linear distribution functions.

Lemma 4.1. *Let $n_k = |P_k|$ be the number of alternative paths and let $\underline{m}_k$ and $\overline{m}_k$ be the minimal and maximal possible length of any considered path in the event activity network for OD pair k, respectively. Then all linear distribution functions fulfilling the five characteristics mentioned above can be characterized according to the three following cases:*

I $n_k = 1$:
If there is just one path p for OD pair k given, then $P_k = \{p\}$ and

$$w_p((t_p)) = 1. \tag{4.6}$$

II $n_k \neq 1$ and $\underline{m}_k = \overline{m}_k$:
If $\underline{m}_k = \overline{m}_k$, all paths have the same fixed length, i.e., $t_p = t_q \ \forall p, q \in P_k$. Then,

$$w_p((t_q)_{q \in P_k}) = w_p((t_p, \dots, t_p)) = \frac{1}{n_k}. \tag{4.7}$$

III $n_k \neq 1$ and $\underline{m}_k \neq \overline{m}_k$:
In the general case all linear functions with the required characteristics have the form

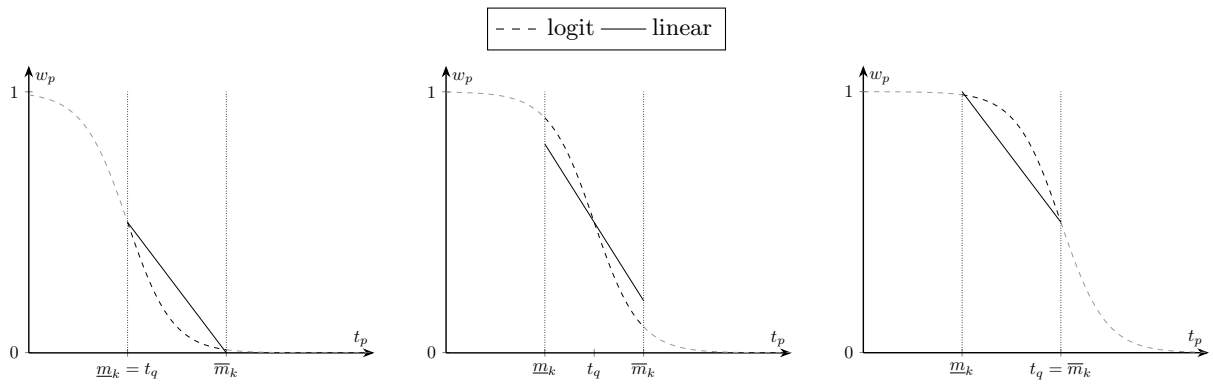
$$w_p((t_q)_{q \in P_k}) = \alpha d_k \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \tag{4.8}$$

with $d_k = -(n_k(\overline{m}_k - \underline{m}_k))^{-1}$ and $\alpha \in (0, 1]$.

A constructive proof for Lemma 4.1 is given in Appendix B. We replace the logit model by the linear distribution functions (4.6), (4.7) and (4.8) in their respective cases in the model from Section 3.3. This yields a linearly constrained feasible region of the optimization problem and further ensures that the five characteristics (4.1) to (4.5) hold.

The linear distribution function is defined in the range $[\underline{m}_k, \overline{m}_k]$ for the length t_q of each path q and its slope in that domain can be adjusted with the parameter $\alpha \in (0, 1]$. For example, for $\alpha \to 0$

we approximate the uniform distribution, independent of the path lengths. The higher α, the more do passengers react on differences in path lengths. In experiments we learned that the linear distribution function from Lemma 4.1 tends to distribute passengers more evenly on paths than a logit distribution. Therefore, we use a value of $\alpha = 1$ to scale the linear distribution function in all experiments.



(a) Probability that path p is chosen, given that length t_q of path q is as short as possible

(b) Probability that path p is chosen, given that length t_q of path q is between bounds $\underline{m}_k$ and $\overline{m}_k$

(c) Probability that path p is chosen, given that length t_q of path q is as long as possible

Figure 4.1: Probabilities of a logit and a linear distribution model that path p is chosen, given an alternative path q with fixed length t_q

Figure 4.1 visualizes the probabilities that path p is chosen according to a logit and a linear distribution model, given a second path q with fixed length t_q. To better demonstrate the linear distribution model, three cases for the fixed path length t_q are considered. For example, in Figure 4.1a it is assumed that the length of the alternative path q is as short as possible, i.e., $t_q = \underline{m}_k$. Then, the probability that path p is chosen is at most 0.5 since it cannot be shorter than path q. The higher the length of path q, the higher the probability that path p is chosen, see Figures 4.1b and 4.1c.

This figure also illustrates how the probability of the logit distribution can be over- or underestimated by the linear distribution model. Knowing the length t_q of the alternative path q, a better linear approximation of the logit model is possible. However, since the utilities of all alternatives depend on the timetable, a linear distribution model can only rely on the bounds $\underline{m}_k$ and $\overline{m}_k$.

4.1.2 Linearization of objective function

In the following we denote the set of OD pairs k with $n_k > 1$ and $\overline{m}_k \neq \underline{m}_k$ by OD^*. Using the linear distribution functions from Lemma 4.1 instead of the logit distribution allows us to express the number of passengers on each activity x_{ij} as a linear function of the durations δ_{ij}. We obtain the quadratic integer timetabling program with **I**ntegrated passenger **D**istribution according to a **LIN**ear distribution

model (ID-LIN):

$$\begin{array}{llll} \min & \delta^T A\delta + b^T\delta & & \\ \text{s.t.} & \delta_{ij} & \geq & l_{ij} \quad \forall ij \in \mathcal{A} \\ & \delta_{ij} & \leq & u_{ij} \quad \forall ij \in \mathcal{A} \\ & \Gamma_c\delta & = & T\cdot\mu_c \quad \forall c \in \mathcal{C} \\ & \delta_{ij} & \in & \mathbb{Z}_+ \quad \forall i \in \mathcal{A} \\ & \mu_c & \in & \mathbb{Z}_+ \quad \forall c \in \mathcal{C} \end{array}$$

with the coefficients

$$A_{ij,i'j'} := \sum_{k\in OD^*} \alpha d_k o_k \left(\sum_{\substack{p\in P_k:\\ ij,i'j'\in p}} 1 - \sum_{\substack{p\in P_k:\\ ij\in p}} \sum_{\substack{q\neq p\in P_k:\\ i'j'\in q}} \frac{1}{n_k - 1} \right), \quad \forall ij, i'j' \in \mathcal{A} \tag{4.9}$$

and

$$b_{ij} := \sum_{k\in OD} \sum_{p\in P_k: ij\in p} \frac{o_k}{n_k}, \quad \forall ij \in \mathcal{A}.$$

The derivation of the coefficient matrix A and vector b can be found in Appendix C.

We have a minimization program and the coefficient matrix A can be proven to be negative semi-definite, see Appendix D. That means, the objective function is concave and standard methods for quadratic programs are not expected to be efficient. We therefore apply a linearization to the objective function. To this end, we express the integer variables δ_{ij} as a sum of binary variables,

$$\delta_{ij} = l_{ij} + \sum_{m=0}^{\lfloor \log(u_{ij}-l_{ij}) \rfloor} 2^m \sigma_{ij}^m$$

and linearize the products of binaries $\sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$. The corresponding linearization of the optimization program (ID-LIN) as used for the experiments can be found in Appendix E.

4.2 Model 2 - Simulation of logit model

In a second model we integrate a simulated passenger distribution into the timetabling framework. The simulation is based on an alternative way to compute the logit probabilities. According to Train (2009) it holds that

$$\begin{aligned} w_p((t_q)_{q\in P_k}) &= \frac{e^{\beta t_p}}{\sum_{q\in P_k}^n e^{\beta t_q}} \\ &= Prob\left(t_p + \varepsilon_p \leq \min_{q\in P_k}(t_q + \varepsilon_q) \right), \end{aligned} \tag{4.10}$$

where the ε_p are independent and identically Gumbel distributed. That means, the logit probability that alternative p is chosen equals the probability that the length of path p, deferred by some random value ε_p, is shorter than the length of any alternative path q, deferred by some random value ε_q. Following similar

steps as Pacheco et al. (2016), we use the representation in Equation (4.10) to simulate the logit model by drawing random values for ε. That means, we consider several scenarios $r \in R$, draw a random value ε_{pr} for each path p in each scenario r and add these to the path lengths. This yields a different, randomized path length in each scenario, which we denote by

$$t_{pr} = \sum_{ij \in p} \delta_{ij} + \varepsilon_{pr} \quad \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R.$$

Note, that similar to the path length computation in Equation (3.5) also this modeling can easily be extended by additional factors like fares or a penalty for each transfer. Then, we choose the shortest path in each scenario for each OD pair and denote the travel time for OD pair k in scenario r by

$$t_{kr} = \min_{p \in P_k} t_{pr} \quad \forall k \in OD,\ \forall r \in R. \tag{4.11}$$

This discrete choice of the shortest path in each scenario r yields a distribution of the passengers of OD pair k over the available paths in the path choice set P_k. Since we choose the random terms ε_{pr} to be independent and identically Gumbel distributed, this distribution converges towards a logit distribution for increasing number of scenarios, see Equation (4.10).

Using a binary choice variable z_{pr} which is set to one if and only if path p is the shortest in scenario r, constraint ((4.11)) can be linearized to

$$\begin{aligned}
t_{kr} &\leq t_{pr} && \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
t_{kr} &\geq t_{pr} - (1 - z_{pr}) M_{kr} && \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
\sum_{p \in P_k} z_{pr} &= 1 && \forall k \in OD,\ \forall r \in R
\end{aligned}$$

where

$$M_{kr} = \max_{p \in P_k} \left(\sum_{ij \in p} u_{ij} + \varepsilon_{pr} \right) - \min_{p \in P_k} \left(\sum_{ij \in p} l_{ij} + \varepsilon_{pr} \right)$$

is sufficiently large.

Note, that in case of equality of best randomized path lengths in a scenario, this modeling will do a random assignment of the passenger choice. We obtain the model for timetabling with an **I**ntegrated passenger **D**istribution by **SIM**ulation of the logit model (ID-SIM):

$$
\begin{array}{llcll}
\min & \sum_{k \in OD} o_k \frac{1}{|R|} \sum_{r \in R} t_{kr} & & & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
& \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
& t_{pr} & = & \sum_{ij \in p} \delta_{ij} + \varepsilon_{pr} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
& \sum_{p \in P_k} z_{pr} & = & 1 & \forall k \in OD,\ \forall r \in R \\
& t_{kr} & \leq & t_{pr} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
& t_{kr} & \geq & t_{pr} - (1 - z_{pr}) M_{kr} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
& \delta_{ij} & \in & \mathbb{Z}_+ & \forall ij \in \mathcal{A} \\
& \mu_c & \in & \mathbb{Z} & \forall c \in \mathcal{C} \\
& t_{pr} & \in & \mathbb{R}_+ & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R \\
& t_{kr} & \in & \mathbb{R}_+ & \forall k \in OD,\ \forall r \in R \\
& z_{pr} & \in & \{0,1\} & \forall k \in OD,\ \forall p \in P_k,\ \forall r \in R
\end{array}
$$

The constraints and the objective function of this formulation are linear in the variables.

Obviously, there is a trade-off between solvability of the MILP model (ID-SIM) and accuracy of the simulation. Considering only few scenarios results in a small model which, however, yields a random solution because a path could be privileged or disadvantaged by chance. With an increasing number of scenarios we expect the solution to converge, but also the model size and hence solution time to increase. To choose a setting that balances solvability and accuracy, we ran preliminary experiments with varying numbers of scenarios. Based on this, we choose to use a low number of $|R| = 10$ scenarios and pick the best solution of 10 repetitions instead of using a large number of scenarios. In our experiments this has shown to yield a good trade-off between running time and a high probability to find the best solution. Another advantage of solving each instance multiple times with a small number of scenarios over considering large scenario sets is that the repetitions are independent and therefore easily parallelizable.

4.3 Theoretical comparison of the two models

In this section we compare the two models (ID-LIN) and (ID-SIM) with respect to their objective functions. The objective function of (ID-LIN) is the sum of the absolute travel times of all passengers on their respective routes which are chosen based on the linear distribution function introduced in Lemma 4.1. This distribution assumption implies that not everyone travels on a shortest route with respect to travel times but passengers also make use of routes that have slightly longer travel times than the shortest. Combining the distribution of passengers on multiple routes with an objective to minimize absolute travel time can have undesirable consequences, as it can be seen in the following example.

Example 1. *Consider a network consisting of two stations A and B and one OD pair k that wants to travel from A to B. Assume, there are two possible routes p and q given in the choice set P_k with respective bounds $[10, 22]$ and $[11, 21]$. The example network is illustrated in Figure 4.2.*

We compare three different timetables t^1, t^2 and t^3. The first timetable offers two equally good paths, these are $t^1_p = 11 = t^1_q$. The second and the third timetable have one very short path and one longer

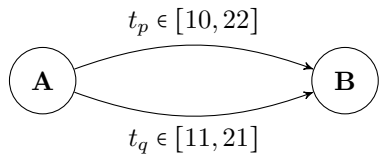


Figure 4.2: Mini network

alternative. These are $t_p^2 = 10$, $t_q^2 = 13$ *and* $t_p^3 = 10$, $t_q^3 = 21$, *respectively.*

These three timetables are evaluated with respect to travel time on the passengers' respective shortest route, travel time when assuming that passengers distribute according to a logit distribution with parameter $\beta = -0.22$ *and travel time when assuming that passengers distribute according to the linear distribution model from Lemma 4.1 with parameter* $\alpha = 1.0$. *The objective values of one passenger of OD pair* k *can be found in Table 4.1.*

	(t_p, t_q)	Shortest route	Logit distribution	Linear distribution
t^1	$(11, 11)$	11	11	11
t^2	$(10, 13)$	10	11.02	11.13
t^3	$(10, 21)$	10	10.90	10.46

Table 4.1: Comparison of travel times of three timetables w.r.t. different distributions

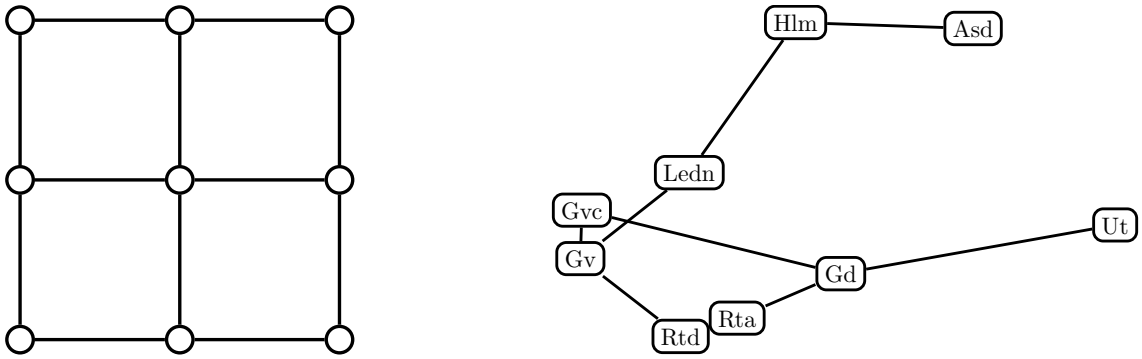
We find, as expected, that the travel time with respect to the shortest path is best in timetables t^2 *or* t^3, *regardless of the length of alternative* q. *Considering travel time according to a linear or logit distribution, timetable* t^2 *is worse than timetable* t^1. *This result is open for discussion as none of the two timetables is obviously better than the other. However, it is striking that the travel time according to a linear or logit distribution is better in timetable* t^3 *than in timetable* t^2. *This result is as unexpected as undesired, but has a very simple explanation. The worse the travel time* t_q *of alternative* q, *the more probable it is that passengers choose to travel via path* p, *which yields a lower total travel time.*

The objective of the second model (ID-SIM) is to minimize the weighted sum of randomized shortest path lengths t_{kr} instead of the absolute travel time as used in the first model (ID-LIN). There, a path just enters the objective function if it is perceived better than any alternative in one scenario. Hence, no considered path in the path choice set can deteriorate, but only improve the objective value. This implies that undesirable effects as demonstrated in Example 1 do not occur when considering the objective function of the program (ID-SIM).

5 Experimental setup

5.1 Instances

To test and compare our approaches, we run experiments on a number of instances. Each instance $\mathcal{I}$ consists of an event activity network $\mathcal{N}$ with lower and upper bounds l and u and a demand situation.



(a) 3x3 grid infrastructure

(b) Partial network of Netherlands Railways

Figure 5.1: The methods are compared on instances defined on these two infrastructures

The event activity network is derived from information about the public transport network, i.e., stations and tracks, as well as a line plan. Both models (ID-LIN) and (ID-SIM) assume a choice set of paths for each OD pair to be given. How we preprocess the instances and derive a path choice set is described in Appendix F.

5.1.1 Instances on grid network

We consider 32 instances defined on a 3×3 grid network which is depicted in Figure 5.1a. On this network we consider 4 different demand situations, and for each of them several line plans with corresponding event activity networks. The instances are partial instances of a bigger grid network introduced by Friedrich et al. (2017) and made available in an online repository[1]. Due to its structure, the grid infrastructure provides good conditions to find multiple geographically different routes with comparable length for passengers.

5.1.2 Instance on Dutch railway network

To test our approaches on a real-world instance we consider a part of the Dutch railway network as operated by Netherlands Railways (NS). The partial network includes the stations Amsterdam, Den Haag, Den Haag HS, Haarlem, Gouda, Leiden, Rotterdam, Rotterdam Alexander and Utrecht in the Randstad, a metropolitan region in the Netherlands. We consider eight intercity lines operating between the stations. The track network is depicted in Figure 5.1b.

5.2 Timetabling approaches

We compare the timetabling models with integrated passenger distribution (ID-LIN) and (ID-SIM) with three state-of-the-art methods for timetabling: two methods (PS) and (PD) assume a predetermined passengers assignment to routes, and one method (IS) has an integrated passenger routing on shortest paths. Besides the timetabling models (ID-SIM) and (ID-LIN) that integrate the passenger distribution, we also

[1] https://github.com/FOR2083/PublicTransportNetworks/tree/master/Grid_5x5

test and compare a heuristic solution approach (ID-ITR) for timetabling with passenger distribution. These approaches are described in more detail below.

(PS) First, a timetabling model with **P**redetermined passenger assignment on a **S**ingle path is considered. In this model passengers' routes are fixed before the optimization step. We assign passengers to the shortest route with respect to the average bounds $\frac{1}{2}(l_{ij} + u_{ij})$ on edges in the event activity network. This basic version of the timetabling model is the subject of many publications since the development of the PESP model, see for example Nachtigall (1998) or Liebchen (2018). An integer programming formulation is given by Equations (3.2) till (3.4) as described in Section 3.2.

(PD) Second, we implement another model with **P**redetermined passenger routes. In contrast to the model (PS), passengers are **D**istributed on multiple paths according to a logit model with the parameter $\beta = -0.22$ and using average bounds on edges. The value of β is adjusted to model a realistic passenger distribution in the network. We are not aware of a published timetabling approach that explicitly states a predetermined passenger distribution according to a logit model, but this strategy can be compared to those made in Parbo et al. (2014) or Robenek et al. (2016), where passenger distributions were derived from utilities of alternative routes. The underlying integer programming model is all the same as the one in (PS), only the passenger weights are predetermined in a different way.

(IS) Third, we consider a timetabling model with **I**ntegrated **S**hortest path search. The timetable is optimized with the objective of minimizing passenger travel times if passengers choose the shortest path based on the timetable. This approach resembles the idea of the integrated shortest path models described in Siebert and Goerigk (2013), Gattermann et al. (2016) and Borndörfer et al. (2017), for example. An integer programming formulation of this model is attached in Appendix G.1.

(ID-ITR) Fourth, we consider a heuristic approach for timetabling with **I**ntegrated passenger **D**istribution that **IT**e**R**ates between timetable design and passenger distribution. To compute the passenger distribution based on a fixed timetable, we use the logit model with the parameter $\beta = -0.22$. The initial passenger loads are determined by using the average bounds as edge lengths. In all following iterations the realized edge lengths of the timetable are used. This yields fixed passenger loads on each edge in the event activity network in each iteration and a standard timetabling model assuming a predetermined passenger distribution can be solved with the given loads. We iterate until the solution value does not change significantly between two iterations or a maximum number of iterations is reached. Similar iterative approaches for timetabling and passenger route choice are described in Sels et al. (2011) or Parbo et al. (2014), for example. Pseudo code for this method can be found in Appendix G.2.

We refer to these benchmark models by (PS), (PD), (IS) and (ID-ITR), respectively. Table 5.1 indicates whether the route choice is integrated into the methods as well as which kind of route choice model is assumed. By comparing the models (ID-LIN) and (ID-SIM) with the heuristic approach (ID-ITR) and the three benchmark models (PS), (PD) and (IS), we can identify the benefits of integrating (1) passenger route search and (2) simultaneous modeling of a passenger distribution.

	Predetermined route choice	**I**ntegrated route choice
Single route	(PS)	(IS)
Distribution	(PD)	(ID-ITR), (ID-LIN), (ID-SIM)

Table 5.1: Summary indicating which solution approach (1) assumes a predetermined route choice or has an integrated route choice and (2) assumes that passengers use a single route only or distribute on multiple routes

5.3 Implementation

In order to reduce the size of the search space, the domain of the variables μ_c is constrained in all models with the following inequalities.

$$\left\lceil \frac{1}{T} \sum_{ij \in c^+} l_{ij} - \sum_{ij \in c^-} u_{ij} \right\rceil \leq \mu_c \leq \left\lfloor \frac{1}{T} \sum_{ij \in c^+} u_{ij} - \sum_{ij \in c^-} l_{ij} \right\rfloor \quad \forall c \in \mathcal{C}.$$

Here, c^+ and c^- denote the set of edges in cycle c in forward and backward direction, respectively, and l_{ij} and u_{ij} are the lower and upper bounds of activity ij. These well-established inequalities were first described in Odijk (1996).

All mixed integer linear programs are solved with the general-purpose solver Fico Xpress 8.5 on a laptop with 32 GB RAM and an Intel® Core™ i7-6700HQ. For all experiments we use a start solution to warm start the optimization.

5.4 Evaluation of timetables

Different research areas apply different measures to evaluate timetables from the passengers' perspective. We could see in Example 1 that on small networks different evaluation functions can yield different results. This small example suggests two features: First, despite the fact that travel time is commonly used to evaluate timetables, it might not be suitable when considering a passenger distribution on multiple routes. Second, different evaluation measures may consider different timetables to be better although the functions are commonly accepted to serve for the evaluation of timetables. Hartleb et al. (2019) compared multiple timetable evaluation functions for passengers on different instances and indeed found that these functions are often not consistent in their evaluation. We learn that there is no default objective function to be used when optimizing timetables with an integrated passenger distribution. To avoid misinterpretation of the results due to a simplistic or biased evaluation, we evaluate all resulting timetables with four structurally different evaluation functions. As before, we denote the total passenger load of OD pair k with o_k and the perceived travel time on path p with t_p. Let P_k be a set of available paths for OD pair k. The used evaluation functions are

tt_{sp} The total **t**ravel **t**ime of all passengers on their **s**hortest **p**ath:

$$tt_{sp} = \sum_{k \in OD} o_k \sum_{p \in P_k} w_p^{sp} \cdot t_p,$$

where w_p^{sp} is the probability that passengers choose path p assuming that all passengers use their shortest paths only.

tt_{mp} The total **t**ravel **t**ime of all passengers when distributed on **m**ultiple **p**aths according to the logit model:

$$tt_{mp} = \sum_{k \in OD} o_k \sum_{p \in P_k} w_p^{lm} \cdot t_p,$$

where w_p^{lm} is the probability that passengers choose path p assuming that all passengers distribute on their paths according to a logit distribution.

ut_{sum} The evaluated total **ut**tility for all passengers, defined as the weighted **sum** of all logit denominators:

$$ut_{sum} = \sum_{k \in OD} o_k \sum_{p \in P_k} e^{\beta t_p},$$

with $\beta = -0.22$. Derived from the logit model, this measure gives an indication of how useful the public transport service is to the passengers.

ut_{log} The **ut**ility based evaluation function **log**sums for all passengers, defined as the weighted sum of the logarithm of all logit denominators:

$$ut_{log} = \sum_{k \in OD} o_k \cdot \ln\left(\sum_{p \in P_k} e^{\beta t_p}\right),$$

with $\beta = -0.22$. Similar to the evaluated total utility, the logsums are a measure of utility for passengers. Due to the logarithm in this evaluation function, OD pairs have different weights relative to each other than in the evaluated total utility.

All four functions evaluate the quality of timetables from the passengers' perspective. Note that these functions are commonly used for evaluation but due to their structure not all are suitable as objective functions in an optimization program. The first two evaluation functions are *travel time based* and thus to be minimized while the latter two evaluation functions are *utility based* and hence to be maximized. Considering all four evaluation functions allows a thorough investigation and comparison of the timetables and, in this way, of the proposed timetabling methods.

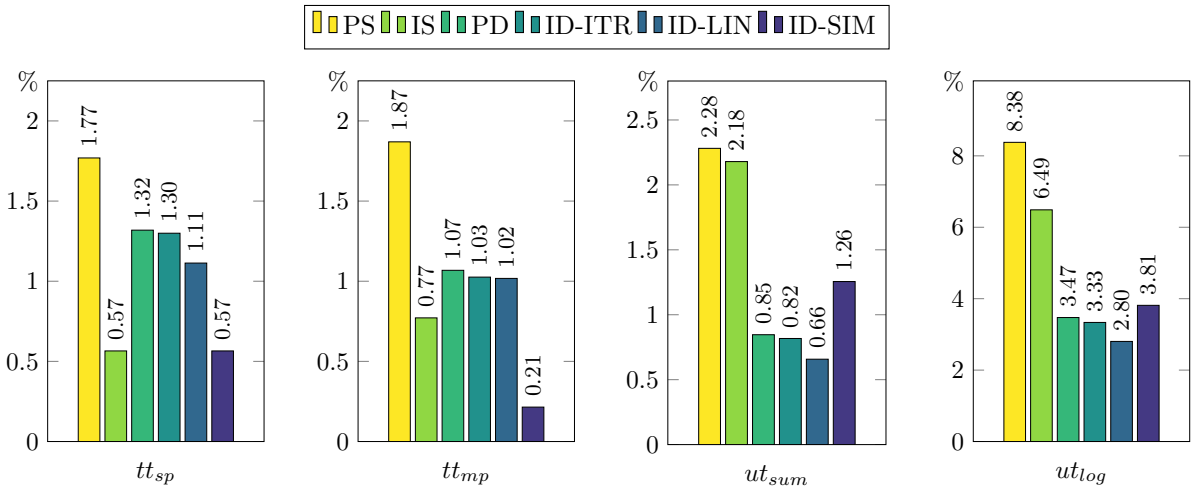
For better comparability we present the relative solution values when compared to an 'ideal solution'. In an ideal solution it is assumed that the travel time on each path for each OD pair is equal to the length of the path with respect to the lower bounds on all edges. For most instances, such an ideal solution does not exist but it is a common measure to see how close solutions are to perfect conditions. More details about ideal solutions and about how they are used in literature can be found in Caimi et al. (2017).

6 Results

In the experiments we showcase the benefits and drawbacks of the timetabling models with integrated passenger distribution (ID-LIN) and (ID-SIM) when compared to existing timetabling approaches.

6.1 Experiments on 32 instances on the grid network

We conduct experiments on 32 instances on the grid network described in Section 5.1.1. On 7 instances all six methods find an ideal solution and on another 4 instances the model (ID-LIN) could not find an optimal solution or could not prove optimality within 10 hours, which is why we exclude these 11 instances from the discussion. In Figure 6.1 we present the evaluation values of the solutions found by the different approaches averaged over the remaining 21 instances on the grid network. This figure shows the average performance of the six methods with respect to the four considered evaluation functions introduced in Section 5.4. All values are given in percent, relative to the evaluation value of an ideal solution.



(a) Relative evaluation values with respect to travel time of all passengers on shortest paths

(b) Relative evaluation values with respect to travel time of all passengers on multiple paths

(c) Relative evaluation values with respect to evaluated total utility for all passengers

(d) Relative evaluation values with respect to total logsums for all passengers

Figure 6.1: The bars show the evaluation values of the six different methods relative to those of an ideal solution, averaged over 21 instances on the grid network.

The relative evaluation values can be read as follows: For example, a relative value of 1.77 for tt_{sp} in Figure 6.1a of the model (PS) means that the travel time on a shortest connection in the solution of (PS) is on average 1.77 percent longer than the travel time on a shortest connection in an ideal solution. Comparing this to the relative travel time on a shortest connection of the model (IS), 0.57, shows that (IS) performs on average better than (PS) with respect to the travel time on the shortest path. In general, the relative evaluation values show to what extent a solution is worse than an ideal solution, with respect to the used evaluation function. We discuss the results per evaluation function:

Figure 6.1a When evaluating timetables with respect to travel time on the shortest path tt_{sp}, the methods (IS) and (ID-SIM) provide on average the best solutions. This is expected for the method (IS) since its objective is to minimize the total travel time of passengers on their shortest paths. To simulate a logit distribution in the model (ID-SIM), in each scenario the shortest path is chosen, as

modeled in Equation (4.11). It seems that in many scenarios the same path is chosen, which in turn gets assigned high weights in the objective function. The model (ID-LIN) finds solutions with travel times on the shortest route that are on average higher than those of methods (IS) and (ID-SIM) and only slightly lower than those of methods (PD) and (ID-ITR). As discussed in Section 4.1.1, the linear distribution model in (ID-LIN) tends to distribute passengers more evenly on paths than the logit model. Thus, the weights assigned to the shortest paths are lower compared to those in the models (IS) and (ID-SIM). This could be an explanation for the worse performance of (ID-LIN) with respect to the travel time on the shortest path. The remaining three methods (PS), (PD) and (ID-ITR) perform worse with respect to travel time on the shortest path. Compared to the best found solutions, their respective travel times are up to three times as far away from an ideal solution.

Figure 6.1b In the case of evaluating travel time using a logit distribution tt_{mp}, the method (ID-SIM) performs best, which is presumably due to the simulated logit distribution of passengers. The model (ID-LIN) performs on average worse than (ID-SIM) and finds solutions that are only as good as those found by (PD) and (ID-ITR). This indicates that the passenger distribution of the linear distribution model used in (ID-LIN) is different from the distribution according to a logit model which is used for evaluation. Furthermore, we can observe that the method (IS) finds better solutions than (PD) and (ID-ITR), averaged over all 21 instances. This is surprising since the methods (PD) and (ID-ITR) consider a passenger distribution according to a logit model whereas (IS) does not consider any alternatives to the shortest route.

We identify the combination of tt_{mp} as evaluation function and a passenger distribution on multiple routes as reason for this observation. In the model (IS) alternative routes might get assigned high travel times which implies a low utilization of these routes in a subsequent distribution of passengers according to the logit model. As shown in Example 1, this can result in lower total travel times for passengers than when providing low travel times on all alternative routes. Indeed, with all six methods we find solutions on certain instances with negative relative evaluation values with respect to tt_{mp}, implying that the found solutions are 'better' than an ideal solution. This questions whether the total (or average) travel time of passengers, while assuming that passengers distribute over multiple routes in the network, is a valid evaluation function for public transport timetables.

Figure 6.1c The evaluation with respect to evaluated total utility ut_{sum} shows a different pattern. The methods (PD), (ID-ITR), (ID-LIN) and (ID-SIM) clearly outperform the methods (PS) and (IS). The gap to the evaluation value of an ideal solution is more than halved. On average, the method (ID-LIN) finds the best solutions, almost halving the gap to the ideal solution once more compared to the model (ID-SIM). This is contrary to the observations with respect to the travel time based evaluation functions tt_{sp} and tt_{mp} where (ID-SIM) performs better than (ID-LIN), see Figures 6.1a and 6.1b. A similar observation can be made for the model (IS). While it performs very good with respect to the travel time based evaluation functions, (IS) yields solutions that are among the worst with respect to evaluated total utility.

Figure 6.1d Similar observations can be made with the total logsums ut_{log} as evaluation function. Also here, the methods (PD), (ID-ITR), (ID-LIN) and (ID-SIM) find clearly better solutions than the methods (PS) and (IS). However, when evaluating the found timetables with the total logsums, the gaps to an ideal solution are by far larger. Furthermore, the solutions of (IS) are on average rated better than those of (PS), which is not visible with the other utility based evaluation function ut_{sum} in Figure 6.1c.

Cross-figure discussion As indicated in Table 5.1, we consider four different categories of modeling passengers in optimization approaches for timetabling. They result from a combination of (1) whether a predetermined route choice is assumed or a route choice model is integrated into optimization and (2) whether passengers are assumed to use a single route only or to distribute on multiple routes.

With respect to the utility based evaluation functions, ut_{sum} and ut_{log}, our experiments show that the quality of timetables can be considerably improved by considering multiple routes instead of a single route for passengers. All four methods that consider a passenger distribution on multiple routes find solutions with significantly lower gap to an ideal solution than the two models that assume passengers to use a single route only. In comparison, the integration of a passenger route choice model, as opposed to a predetermined route assignment, did not help to improve the quality of the found timetables with respect to the utility based evaluation functions. Only the solutions of (IS) are on average slightly better than those of (PS), but the others were not in comparison to (PD).

With respect to the travel time based evaluation functions, tt_{sp} and tt_{mp}, the methods with an integrated route choice model find better timetables than the corresponding single or multiple route methods assuming a predetermined route choice. Especially the models (IS) and (ID-SIM) could find significantly better timetables with respect to travel time on the shortest path and the latter also on a logit distribution. Considering multiple routes for passengers during optimization instead of only one route yields better solutions with respect to tt_{mp}, but not necessarily with respect to tt_{sp} since there just the shortest path is considered for evaluation. Moreover, although the method (PD) finds on average better solutions than (PS), (PD) is outperformed by all other methods with respect to travel time based evaluation functions. This shows that in our experiments considering multiple routes for passengers is not sufficient to find timetables with best travel times.

We find that considering a passenger distribution on multiple routes mainly improves the utilities, and integrating a passenger route choice model mainly improves the travel times of the found timetables. Furthermore, by integrating a passenger distribution model, it is possible to find solutions with multiple good routes that yield both very good travel times and high utilities for passengers on the considered instances. The model (ID-SIM) provided best solutions with respect to the travel time based evaluation functions and comparable solutions with respect to one utility based evaluation function. The model (ID-LIN) could not perform as well as one state-of-the-art approach with respect to the travel time based evaluation functions, but provided the solutions with best utilities. Thus, by integrating a passenger

distribution model it is possible to find better timetables than the benchmark methods with respect to some evaluation functions while maintaining the quality with respect to some other evaluation functions.

Method	CPU time	no. instances
(PS)	0.4s	21/21
(IS)	5.6s	21/21
(PD)	0.8s	21/21
(ID-ITR)	0.8s	21/21
(ID-LIN)	1952.9s	17/21 (5.68)
(ID-SIM)	1184.0s	19/21 (1.17)

Table 6.1: CPU times and number of instances that were solved within one hour. The remaining gap to the best bound after one hour is given in parentheses.

These improvements by integration of an passenger distribution model come at the expense of significantly larger models. Table 6.1 shows the average solution times of the six different methods on the discussed 21 instances on the grid network. From the running times it is obvious that the two proposed models (ID-LIN) and (ID-SIM) need by far the most time for solving the instances. On average it took almost 20 minutes to solve the model (ID-SIM) and more than 30 minutes to solve the model (ID-LIN), while the other methods were solved within few seconds.

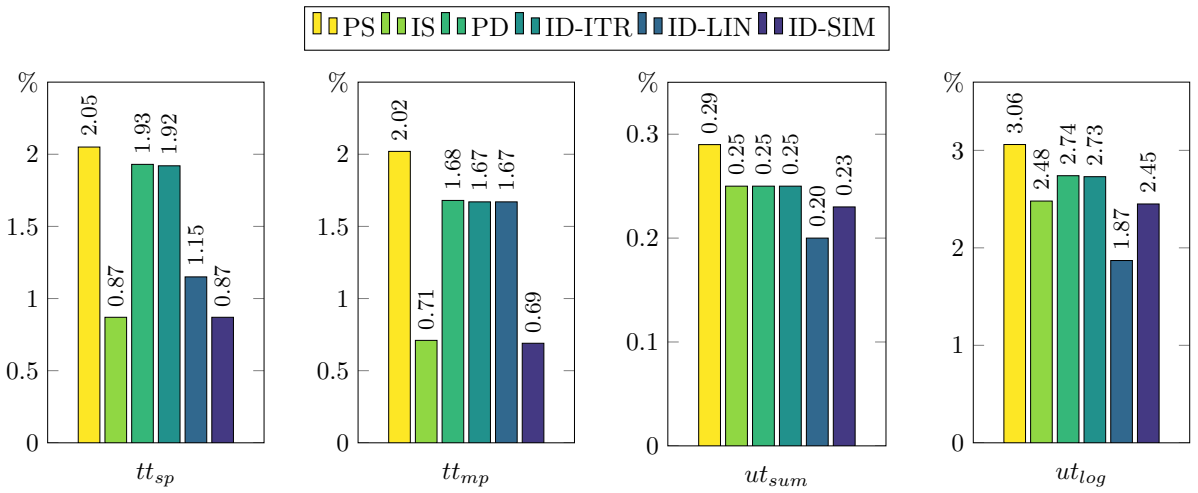
In the second column the number of instances that could be solved within one hour are given. The model (ID-LIN) could only find optimal solutions for 17 of the 21 instances, (ID-SIM) provided optimal solutions for 19 instances. After one hour, the model (ID-LIN) had on average a gap of more than 5% to the best bound whereas the simulation based model was close to an optimal solution with a remaining gap of a little more than 1%. The other four methods were always able to terminate within one hour.

6.2 Experiments on Dutch railway network

We also compare the six different methods on a part of the network of Netherlands Railways, which is described in Section 5.1.2. In Figure 6.2 the evaluation values of all methods are given relative to those of an ideal solution.

In Figures 6.2a and 6.2b we observe that two models with integrated passenger route choice model, (IS) and (ID-SIM), perform best. The gap to an ideal solution is significantly lower compared to the other methods. This is in line with the observation made in the evaluation by the travel time based evaluation functions on the grid instances and demonstrates again the benefits of integrating a passenger route choice model into timetabling optimization. The model (ID-LIN) provides a solution with higher travel times, but it has notably shorter travel times than the remaining methods on the shortest path and comparable travel times assuming a passenger distribution.

The relative evaluation values with respect to the utility based functions in Figures 6.2c and 6.2d suggest that the method (ID-LIN) performs best, as it was observed on the grid instances. In contrast to the instances on the grid network, there is no visible difference between the results of methods that assume that passengers use only a single route and methods that consider a passenger distribution on



(a) Relative evaluation values with respect to travel time of all passengers on shortest paths

(b) Relative evaluation values with respect to travel time of all passengers on multiple paths

(c) Relative evaluation values with respect to evaluated total utility for all passengers

(d) Relative evaluation values with respect to total logsums for all passengers

Figure 6.2: The bars show the evaluation values of the six different methods relative to those of an ideal solution on a partial network of Netherlands Railways

multiple routes. Instead, the method (IS) performs better than the two methods (PD) and (ID-ITR). This might be due to the infrastructure of the network that contains only one cycle and thus hardly provides geographically different routes for passengers, see Figure 5.1b.

We find that the solutions found by (IS) and (ID-SIM) dominate the solutions found by all other methods with respect to the travel time based evaluation functions, where the consideration of multiple routes brings only a slight advantage to the model (ID-SIM). With respect to the utility based evaluation functions the solution found by (ID-LIN) dominates all other solutions. Moreover, the results in Figure 6.2 demonstrate the importance of a thorough evaluation with multiple evaluation functions. Together with the results on the grid network, these experiments illustrate that an evaluation with a single evaluation function is likely to falsify the interpretation.

7 Conclusion

In this paper we study the problem of finding a travel time minimal timetable under the assumption that a distribution of passengers on available routes can be modeled using a discrete choice model. We use the logit model to estimate a passenger distribution and formulate this problem as a mixed integer program. Based on this, we develop two linear models proposing different ways to model the interaction of passenger route choice and timetable design. In the first model we incorporate a novel multidimensional linear passenger distribution model that resembles the characteristics of the logit model. Our second model approximates a logit distribution of the passengers from an integrated simulation framework.

We compare the two timetabling models with integrated passenger distribution with three state-of-the-art methods and a heuristic approach that iterates between timetabling and passenger routing to find travel time optimal timetables for passengers. The experiments are conducted on a set of artificial instances as well as on a part of the network of Netherlands Railways. We provide a thorough comparison of all solutions with respect to four structurally different evaluation functions.

With the integration of a passenger distribution model into a timetabling framework we were able to find better timetables for passengers than the considered state-of-the-art methods. The gap to an ideal solution for passengers could be significantly reduced with respect to some evaluation functions while performing similar with respect to other evaluation functions. In general, the experiments give insights into how the consideration of multiple routes instead of a single route for passengers, and how the integration of route choice instead of a predetermined assignment affect the solution quality.

It is interesting to observe that the different evaluation functions yield different results for the considered methods. This supports the impression that a comprehensive evaluation with multiple functions is useful and necessary to make clear statements about the quality of methods. In particular, we address observations that a commonly used evaluation function for timetables, the total travel time of passengers, in combination with a passenger distribution model yields unexpected results. Our results and a simple example raise the question whether this function is suitable for evaluation or as objective function when considering a distribution of passengers on multiple paths.

The integration of a passenger distribution model in both timetabling models comes at the expense of significantly higher solution times. Future research could deal with the development of solution approaches to be able to solve large instances.

Acknowledgements

We thank Dennis Huisman from Erasmus School of Economics at Erasmus University Rotterdam and Netherlands Railways as well as Markus Friedrich from the Institute for Road and Transport Science at University of Stuttgart for their valuable comments and suggestions throughout this work.

References

Bekhor, S., Ben-Akiva, M. E., and Ramming, M. S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 144(1):235–247.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press.

Borndörfer, R., Hoppmann, H., and Karbstein, M. (2017). Passenger routing for periodic timetable optimization. *Public Transport*, 9(1-2):115–135.

Borndörfer, R., Klug, T., Lamorgese, L., Mannino, C., Reuther, M., and Schlechte, T. (2018). *Handbook of Optimization in the Railway Industry*, volume 268. Springer.

Caimi, G., Kroon, L., and Liebchen, C. (2017). Models for railway timetable optimization: Applicability and applications in practice. *Journal of Rail Transport Planning & Management*, 6(4):285–312.

Canca, D., De-Los-Santos, A., Laporte, G., and Mesa, J. A. (2019). Integrated railway rapid transit network design and line planning problem with maximum profit. *Transportation Research Part E: Logistics and Transportation Review*, 127:1–30.

De-Los-Santos, A., Laporte, G., Mesa, J. A., and Perea, F. (2017). The railway line frequency and size setting problem. *Public Transport*, 9(1-2):33–53.

de Dios Ortuzar, J. and Willumsen, L. G. (2011). *Modelling transport*. John Wiley & Sons.

Espinosa-Aranda, J. L., García-Ródenas, R., López-García, M. L., and Angulo, E. (2018). Constrained nested logit model: formulation and estimation. *Transportation*, 45(5):1523–1557.

Friedrich, M., Hartl, M., Schiewe, A., and Schöbel, A. (2017). Angebotsplanung im öffentlichen verkehr–planerische und algorithmische lösungen. *Heureka'17*, 9.

Gattermann, P., Großmann, P., Nachtigall, K., and Schöbel, A. (2016). Integrating passengers' routes in periodic timetabling: A sat approach. In *OASIcs-OpenAccess Series in Informatics*, volume 54. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Haase, K. and Müller, S. (2014). A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research*, 232(3):689–691.

Hartleb, J., Schmidt, M., Friedrich, M., and Huisman, D. (2019). A good or a bad timetable: Do different evaluation functions agree? *ERIM Report Series*.

van der Hurk, E., Kroon, L., Maróti, G., and Vervest, P. (2014). Deduction of passengers' route choices from smart card data. *IEEE Transactions on Intelligent Transportation Systems*, 16(1):430–440.

de Jong, G., Daly, A., Pieters, M., and van der Hoorn, T. (2007). The logsum as an evaluation measure: review of the literature and new results. *Transportation Research Part A: Policy and Practice*, 41(9):874–889.

Kroon, L., Huisman, D., Abbink, E., Fioole, P.-J., Fischetti, M., Maróti, G., Schrijver, A., Steenbeek, A., and Ybema, R. (2009). The new dutch timetable: The OR revolution. *Interfaces*, 39(1):6–17.

Liebchen, C. (2018). On the benefit of preprocessing and heuristics for periodic timetabling. In *Operations Research Proceedings 2017*, pages 709–714. Springer.

Liebchen, C. and Möhring, R. H. (2007). The modeling power of the periodic event scheduling problem: railway timetables—and beyond. In *Algorithmic methods for railway optimization*, pages 3–40. Springer.

Löbel, F., Lindner, N., and Borndörfer, R. (2019). The restricted modulo network simplex method for integrated periodic timetabling and passenger routing.

Nachtigall, K. (1994). *A branch and cut approach for periodic network programming*. Inst. für Mathematik.

Nachtigall, K. (1998). Periodic network optimization and fixed interval timetables. *Habilitationsschrift, Deutsches Zentrum für Luft– und Raumfahrt, Institut für Flugführung, Braunschweig, Germany.*

Nachtigall, K. and Opitz, J. (2008). Solving periodic timetable optimisation problems by modulo simplex calculations. In *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Odijk, M. A. (1996). A constraint generation algorithm for the construction of periodic railway timetables. *Transportation Research Part B: Methodological*, 30(6):455–464.

Pacheco, M., Sharif Azadeh, S., and Bierlaire, M. (2016). A new mathematical representation of demand using choice-based optimization method. In *16th Swiss Transport Research Conference*, number EPFL-CONF-218673.

Parbo, J., Nielsen, O. A., and Prato, C. G. (2014). User perspectives in public transport timetable optimisation. *Transportation Research Part C: Emerging Technologies*, 48:269–284.

Polinder, G.-J., Schmidt, M., and Huisman, D. (2019). A new approach to periodic railway timetabling. In *Rail Norrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA), Norrköping, Sweden, June 17th–20th, 2019*, number 069, pages 998–1014. Linköping University Electronic Press.

Robenek, T., Maknoon, Y., Azadeh, S. S., Chen, J., and Bierlaire, M. (2016). Passenger centric train timetabling problem. *Transportation Research Part B: Methodological*, 89:107–126.

Scheepmaker, G. M., Goverde, R. M., and Kroon, L. G. (2017). Review of energy-efficient train control and timetabling. *European Journal of Operational Research*, 257(2):355–376.

Schiewe, P. and Schöbel, A. (2018). Periodic timetabling with integrated routing: An applicable approach.

Schmidt, M. and Schöbel, A. (2015). Timetabling with passenger routing. *OR Spectrum*, 37(1):75–97.

Schmidt, M. E. (2014). *Integrating routing decisions in public transportation problems*. Springer.

Schrijver, A. and Steenbeek, A. (1994). Dienstregelingontwikkeling voor railned (timetable construction for railned), rapport cadans 1.0. *CWI Center for Mathematics and Computer Science, Amsterdam. In Dutch.*

Sels, P. (2015). Towards a better train timetable for Denmark reducing total expected passenger time. In *Proceedings of the 13th Conference on Advanced Systems in Public Transport (CASPT).*

Sels, P., Dewilde, T., Cattrysse, D., and Vansteenwegen, P. (2011). Deriving all passenger flows in a railway network from ticket sales data. *Proceedings of Rail Rome.*

Serafini, P. and Ukovich, W. (1989). A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581.

Siebert, M. and Goerigk, M. (2013). An experimental comparison of periodic timetabling models. *Computers & Operations Research*, 40(10):2251–2259.

Train, K. E. (2009). *Discrete choice methods with simulation.* Cambridge university press.

Appendices

A Notation

Greek letters

α	Scaling parameter for linear distribution function
β	Scaling parameter for logit model
δ_{ij}	Duration of activity $ij \in \mathcal{A}$
$\varepsilon_{p,r}$	Random term to vary the path length of path p in scenario r
Γ	Indicator matrix for forward and backward edges in cycles
μ_c	Auxiliary variable for cycle constraints
σ_{ij}^m	Variable for binary representation of δ_{ij}
$\gamma_{ij,i'j'}^{m,m'}$	Variable for linearization of product $\sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$

Latin upper case letters

A	Coefficient matrix of program (ID-LIN)
$\mathcal{A}$	Set of activities in the event activity network
$\mathcal{C}$	Cycle basis in the event activity network
(ID-LIN)	Timetabling model with integrated linear passenger distribution model
$\mathcal{E}$	Set of events in the event activity network
$\mathcal{I}$	Timetabling instance
(ID-ITR)	Iterative heuristic for timetabling problem with passenger distribution
(ID-SIM)	Timetabling model with simulated passenger distribution model
M_{ij}	Set of indices for binary representation of δ_{ij}
M_{kr}	Auxiliary number for linearization of choice constraints
(PD)	Timetabling method with predetermined passenger distribution
$\mathcal{N}$	Event activity network
OD	Set of OD pairs
P_k	Set of alternative paths for OD pair k
R	Set of scenarios
(IS)	Timetabling model with integrated shortest path search
T	Length of period
(PS)	Timetabling method with predetermined passenger assignment to one path

Latin lower case letters

b	Coefficient vector of program (ID-LIN)
c	Index for cycles in the event activity network
d_k	Auxiliary scalar for the linear distribution of OD pair k
ut_{sum}	Evaluation function: evaluated total utility
i, j	Indices for events in the event activity network
ij	Index for activity from events i to event j in the event activity network
k	Index for OD pairs
l_{ij}	Lower bound on activity $ij \in \mathcal{A}$
ut_{log}	Evaluation function: total logsums
m	Index for binary representation of δ_{ij}
$\underline{m}_k$	Length of shortest path for OD pair k w.r.t lower bounds
$\overline{m}_k$	Length of longest path for OD pair k w.r.t upper bounds
n_k	Number of alternative paths for OD pair k
o_k	Number of passengers of OD pair k
p, q	Indices for paths in the event activity network
r	Index for scenarios
t_p	Length of path p
t_{pr}	Length of path p in scenario r
t_{kr}	Length of shortest path for OD pair k in scenario r
tt_{mp}	Evaluation function: travel time assuming a logit distribution of passengers
tt_{sp}	Evaluation function: travel time assuming shortest paths for passengers
u_{ij}	Upper bound on activity $ij \in \mathcal{A}$
w_p	Probability that path p is chosen
x_{ij}	Passenger load on activity $ij \in \mathcal{A}$
x_{ij}^p	Passenger load of path p on activity $ij \in \mathcal{A}$
z_{pr}	Binary variable indicating whether path p is the shortest in scenario r

B Linear distribution function with characteristics of logit model

To prove Lemma 4.1, we consider the three cases

I $n_k = 1$:
It obviously follows by the property 'certain event' in Equation (4.1) that

$$w_p((t_p)) = \sum_{p \in P_k} w_p((t_p)) = 1.$$

This does not conflict with any other required characteristic.

II $n_k \neq 1$ and $\underline{m}_k = \overline{m}_k$:
If $\overline{m}_k = \underline{m}_k$, all paths have to have the same fixed length, i.e., $t_p = t_q \ \forall p, q \in P_k$. Then, it follows by the property 'uniform distribution on equivalent alternatives' in Equation (4.3) that

$$w_p((t_q)_{q \in P_k}) = w_p((t_p, \dots, t_p)) = \frac{1}{n_k}.$$

This also does not conflict with any other required characteristic.

III $n_k \neq 1$ and $\underline{m}_k \neq \overline{m}_k$:
To show that all linear functions with the five desired characteristics are of the stated shape, we take a linear function

$$w_p((t_q)_{q \in P_k}) = \alpha_0^p + \sum_{q \in P_k} \alpha_q^p t_q \quad \forall p \in P_k$$

in its general form and restrict it by adding the desired characteristics to it.

Logit characteristic: absolute differences in utility determine probability, Equation (4.5)
To obtain a linear distribution function with the characteristics of a logit distribution, we require that the probabilities do not depend on the values of the utilities but on their absolute differences only. Thus, we get for each path p

$$\begin{aligned} & w_p((t_q + \hat{t})_{q \in P_k}) = w_p((t_q)_{q \in P_k}) && \forall \hat{t} \in \mathbb{R} \\ \Leftrightarrow \quad & \alpha_0^p + \sum_{q \in P_k} \alpha_q^p (t_q + \hat{t}) = \alpha_0^p + \sum_{q \in P_k} \alpha_q^p t_q && \forall \hat{t} \in \mathbb{R} \\ \Leftrightarrow \quad & \sum_{q \in P_k} \alpha_q^p \hat{t} = 0 && \forall \hat{t} \in \mathbb{R} \\ \Leftrightarrow \quad & \sum_{q \in P_k} \alpha_q^p = 0 \end{aligned}$$

That means, to obtain a linear distribution function with the logit characteristic, all coefficients of the utilities t_p have to sum up to zero.

Uniform distribution on equivalent alternatives, Equation (4.3)
We add the requirement that the distribution function should be a uniform distribution in case all

alternatives have the same utility. We therefore require that

$$\begin{aligned} & w_p(t,\dots,t) = \frac{1}{n_k} \quad \forall p \in P_k \ \forall t \in \mathbb{R} \\ \Leftrightarrow \quad & \alpha_0^p + \sum_{q \in P_k} \alpha_q^p t = \frac{1}{n_k} \quad \forall p \in P_k \ \forall t \in \mathbb{R} \end{aligned}$$

Plugging in the result that the coefficients sum up to zero, this leaves the second condition

$$\alpha_0^p = \frac{1}{n_k} \quad \forall p \in P_k.$$

Since this equation has to hold for all paths $p \in P_k$, it follows that

$$\alpha_0^p = \alpha_0^q \quad \forall p, q \in P_k$$

and we define $\alpha_0 \coloneqq \alpha_0^p$ for any path $p \in P_k$. This yields that all linear functions with the characteristics in Equations (4.3) and (4.5) are of the form

$$w_p((t_q)_{q \in P_k}) = \sum_{q \in P_k} \alpha_q^p t_q + \frac{1}{n_k}$$

with

$$\sum_{q \in P_k} \alpha_q^p = 0.$$

Independence of order, Equation (4.4)

Next, we consider the condition 'independence of order' of alternatives. Assume the probability of path $p \in P_k$ is to be determined. Then, its probability depends on the quality of all alternatives, but it should be independent of which alternative takes which of these values. We consider any permutation π_p that permutes two paths $q_1, q_2 \neq p \in P_k$ and keeps all other paths constant. Then,

$$\begin{aligned} & w_p((t_q)_{q \in P_k}) = w_p((t_{\pi_p(q)})_{q \in P_k}) \\ \Leftrightarrow \quad & \sum_{q \in P_k} \alpha_q^p t_q + \frac{1}{n_k} = \sum_{q \in P_k} \alpha_q^p t_{\pi_p(q)} + \frac{1}{n_k} \\ \Leftrightarrow \quad & \alpha_{q_1}^p t_{q_1} + \alpha_{q_2}^p t_{q_2} = \alpha_{q_1}^p t_{q_2} + \alpha_{q_2}^p t_{q_1} \\ \Leftrightarrow \quad & \alpha_{q_1}^p (t_{q_1} - t_{q_2}) = \alpha_{q_2}^p (t_{q_1} - t_{q_2}) \\ \Leftrightarrow \quad & \alpha_{q_1}^p = \alpha_{q_2}^p \end{aligned}$$

The last equivalence holds since we require this for all values of t_{q_1} and t_{q_2}. Since we require this to hold for all permutations π_p, in particular for those that permute two arbitrary paths, the coefficients α_q^p have to be equal for all $q \in P_k$, i.e.,

$$\alpha_{q_1}^p = \alpha_{q_2}^p \quad \forall q_1, q_2 \neq p \in P_k.$$

This condition is sufficient to obtain the independence of order with any permutation π_p. Together with the characteristic

$$\sum_{q \in P_k} \alpha_q^p = 0$$

from the property 'uniform distribution on equivalent alternatives' we get

$$\alpha_p^p = -\sum_{\substack{q \in P_k \\ q \neq p}} \alpha_q^p = -(n_k - 1)\alpha_q^p$$

$$\Leftrightarrow \alpha_q^p = \frac{-\alpha_p^p}{(n_k - 1)}.$$

Note, that this is well defined since we discuss here only the cases where $n_k \neq 1$. This means, we can express all α_q^p by the single parameter α_p^p. Defining $\alpha_p \coloneqq \alpha_p^p$ we can plug this into the linear probability function and get

$$\begin{aligned} w_p((t_q)_{q \in P_k}) &= \alpha_p t_p - \sum_{\substack{q \in P_k \\ q \neq p}} \frac{\alpha_p t_q}{(n_k - 1)} + \frac{1}{n_k} \quad \forall p \in P_k \\ &= \alpha_p \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \quad \forall p \in P_k \end{aligned}$$

Monotonicity, Equation (4.2)
With the monotonicity of the distribution function follows

$$\begin{aligned} & w_p((t_q)_{q \in P_k} + \varepsilon \cdot e_p) < w_p((t_q)_{q \in P_k}) \\ \Leftrightarrow \quad & \alpha_p \left(t_p + \varepsilon - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} < \alpha_p \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \\ \Leftrightarrow \quad & \alpha_p \varepsilon < 0 \\ \Leftrightarrow \quad & \alpha_p < 0 \end{aligned}$$

Note, that for $\alpha_p = 0$ we obtain the uniform distribution

$$w_p((t_q)_{q \in P_k}) = \frac{1}{n_k} \quad \forall p \in P_k$$

which would not fulfill the required monotonicity.

Distribution characteristics, Equation (4.1)

Finally, we add the distribution characteristics

$$w_p((t_q)_{q\in P_k}) \in [0,1] \text{ and } \sum_{p\in P_k} w_p((t_q)_{q\in P_k}) = 1.$$

To fulfill the first characteristic, we consider the cases with the highest and lowest possible probability. The lowest probability for path p is achieved, if path p is as long as the upper bound $\overline{m}_k$ and all other paths $q \neq p \in P_k$ are as short as as the lower bound $\underline{m}_k$(due to monotonicity and the probability of the certain event). Then, we require

$$\begin{aligned}
& w_p((t_q)_{q\in P_k}) \geq 0 && \forall t_q \in [\underline{m}_k, \overline{m}_k],\ \forall q \in P_k \\
\Leftrightarrow\quad & \alpha_p \left(t_p - \frac{1}{(n_k-1)} \sum_{\substack{q\in P_k \\ q\neq p}} t_q \right) + \frac{1}{n_k} \geq 0 && \forall t_q \in [\underline{m}_k, \overline{m}_k],\ \forall q \in P_k \\
\Leftrightarrow\quad & \alpha_p \left(\overline{m}_k - \frac{1}{(n_k-1)} \sum_{\substack{q\in P_k \\ q\neq p}} \underline{m}_k \right) + \frac{1}{n_k} \geq 0 \\
\Leftrightarrow\quad & \alpha_p \left(\overline{m}_k - \underline{m}_k \right) + \frac{1}{n_k} \geq 0 \\
\Leftrightarrow\quad & \alpha_p \geq \frac{-1}{n_k \left(\overline{m}_k - \underline{m}_k \right)}
\end{aligned} \tag{B.1}$$

Similarly, the highest probability for path p is achieved, if path p is as short as possible and all other paths $q \neq p \in P_k$ are as long as possible, i.e., $t_p = \underline{m}_k$ and $t_q = \overline{m}_k$ $\forall q \neq p \in P_k$. Then, we require

$$\begin{aligned}
& w_p((t_q)_{q\in P_k}) \leq 1 && \forall t_q \in [\underline{m}_k, \overline{m}_k],\ \forall q \in P_k \\
\Leftrightarrow\quad & \alpha_p \left(\underline{m}_k - \frac{1}{(n_k-1)} \sum_{\substack{q\in P_k \\ q\neq p}} \overline{m}_k \right) + \frac{1}{n_k} \leq 1 \\
\Leftrightarrow\quad & \alpha_p \left(\underline{m}_k - \overline{m}_k \right) + \frac{1}{n_k} \leq 1 \\
\Leftrightarrow\quad & \alpha_p \geq \frac{-(n_k-1)}{n_k \left(\overline{m}_k - \underline{m}_k \right)}
\end{aligned} \tag{B.2}$$

Note, that both lower bounds (B.1) and (B.2) on α_p are negative and well defined as we consider the case where $\overline{m}_k \neq \underline{m}_k$. The second lower bound (B.2) is less strict, equality can only be achieved for $n_k = 2$. This also implies that we always have $w_p < 1$ by construction of the probability function, unless $n_k = 2$. In total we obtain

$$\alpha_p \in \left[\frac{-1}{n_k(\overline{m}_k - \underline{m}_k)}, 0 \right)$$

In this range for α_p the function w_p can be tuned.

The second distribution characteristic requires that the probability of the certain event equals one. This is

$$\begin{aligned}
& \sum_{p \in P_k} w_p((t_q)_{q \in P_k}) = 1 \\
\Leftrightarrow \quad & \sum_{p \in P_k} \left(\alpha_p \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \right) = 1 \\
\Leftrightarrow \quad & \sum_{p \in P_k} \alpha_p t_p - \sum_{p \in P_k} \frac{\alpha_p}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q + 1 = 1 \\
\Leftrightarrow \quad & \sum_{p \in P_k} t_p \left(\alpha_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} \alpha_q \right) = 0
\end{aligned}$$

Here, we factorized the path length t_p. Since the sum has to vanish for all $t_p \in [\underline{m}_k, \overline{m}_k]$, each of the addends has to vanish, which yields

$$\begin{aligned}
\Leftrightarrow \quad & \alpha_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} \alpha_q = 0 \qquad \forall p \in P_k \\
\Leftrightarrow \quad & \alpha_p = \alpha_q \quad \forall p, q \in P_k
\end{aligned}$$

That means, for each OD pair k, the parameter α_p are equal for all probability functions w_p with $p \in P_k$. We therefore define $\alpha_k \coloneqq \alpha_p$ for any $p \in P_k$. This yields the linear distribution functions

$$w_p((t_q)_{q \in P_k}) = \alpha_k \left(t_p - \frac{1}{(n_k - 1)} \sum_{\substack{q \in P_k \\ q \neq p}} t_q \right) + \frac{1}{n_k} \quad \forall p \in P_k$$

As from now, we will use a scaling factor $\alpha \in (0, 1]$ and define

$$d_k \coloneqq \frac{-1}{n_k(\overline{m}_k - \underline{m}_k)} \quad \forall k \in OD$$

such that

$$\alpha_k = \alpha d_k \quad \forall k \in OD.$$

For $n_k > 1$ and $\overline{m}_k \neq \underline{m}_k$, all linear probability functions fulfilling the given criteria can be written as

$$w_p((t_q)_{q \in P_k}) = \alpha d_k \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q \in P_k: \\ q \neq p}} t_q \right) + \frac{1}{n_k}$$

with $\alpha \in (0, 1]$.

C Derivation of A and b

For the derivation of the coefficient matrix A and the coefficient vector b of the program (ID-LIN), we split the set of OD pairs into three disjoint sets:

$$OD = OD^* \uplus OD_1 \uplus OD_=, \tag{C.1}$$

where OD_1 denotes the set of OD pairs k with $n_k = 1$, $OD_=$ denotes the set of OD pairs k with $n_k > 1$ and $\overline{m}_k = \underline{m}_k$, and OD^* denotes the set of remaining OD pairs k with $n_k > 1$ and $\overline{m}_k \neq \underline{m}_k$. Then, we plug the Constraints (3.5), (3.6) and (3.7) into the objective function (3.4). In Constraint (3.6) we use the respective linear distribution for the OD sets OD_1, $OD_=$ and OD^* as given in Lemma 4.1.

$$\begin{aligned}
&\sum_{(i,j)\in\mathcal{A}} x_{ij}\delta_{ij} \\
&\overset{(3.7)}{=} \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD} \sum_{\substack{p\in P_k:\\ ij\in p}} x^p_{ij}\delta_{ij} \\
&\overset{(C.1)}{=} \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} x^p_{ij}\delta_{ij} + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} x^p_{ij}\delta_{ij} + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} x^p_{ij}\delta_{ij} \\
&\overset{(3.6)}{=} \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \left[\alpha d_k \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q\right) + \frac{1}{n_k}\right] o_k\delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} o_k\delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k}\delta_{ij} \\
&= \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \alpha d_k \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q\right) o_k\delta_{ij} + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{1}{n_k} o_k\delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k}\delta_{ij} \\
&\quad + \sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k}\delta_{ij} \\
&= \underbrace{\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \alpha d_k \left(t_p - \frac{1}{n_k - 1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q\right) o_k\delta_{ij}}_{=:quadr(\delta_{ij})} + \underbrace{\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{1}{n_k} o_k\delta_{ij}}_{=:lin(\delta_{ij})}
\end{aligned}$$

Here, we can see that the quadratic part $quadr(\delta_{ij})$ of the objective function only depends on the OD pairs $k \in OD^*$ that have multiple paths with possibly different lengths at choice. All OD pairs $k \in OD_1$ with just one optional path per period and all OD pairs $k \in OD_=$ with multiple paths of same fixed lengths are uniformly distributed on their respective path(s) and their total travel time is added to the objective function. These addends

$$\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_1} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k}\delta_{ij}$$

and

$$\sum_{(i,j)\in\mathcal{A}} \sum_{k\in OD_=} \sum_{\substack{p\in P_k:\\ ij\in p}} \frac{o_k}{n_k}\delta_{ij}$$

are hidden in the linear part $lin(\delta_{ij})$ of the objective function. To derive a closed form for the coefficients, we will consider the quadratic and the linear part of the objective function separately.

$$\begin{aligned}
&quadr(\delta_{ij})\\
&= \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \sum_{\substack{p\in P_k:\\ ij\in p}} \alpha d_k \left(t_p - \frac{1}{n_k-1} \sum_{\substack{q\in P_k:\\ q\neq p}} t_q \right) o_k \delta_{ij}\\
&\overset{(3.5)}{=} \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \alpha d_k o_k \sum_{\substack{p\in P_k:\\ ij\in p}} \left(\sum_{i'j'\in p} \delta_{i'j'} - \frac{1}{n_k-1} \sum_{\substack{q\in P_k:\\ q\neq p}} \sum_{i'j'\in q} \delta_{i'j'} \right) \delta_{ij}\\
&= \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \alpha d_k o_k \sum_{\substack{p\in P_k:\\ ij\in p}} \sum_{i'j'\in p} \delta_{i'j'}\delta_{ij}\\
&\quad + \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \alpha d_k o_k \sum_{\substack{p\in P_k:\\ ij\in p}} \left(-\frac{1}{n_k-1} \sum_{\substack{q\in P_k:\\ q\neq p}} \sum_{i'j'\in q} \delta_{i'j'}\delta_{ij} \right)\\
&= \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \alpha d_k o_k \sum_{i'j'\in\mathcal{A}} \sum_{\substack{p\in P_k:\\ ij\in p,\\ i'j'\in p}} \delta_{i'j'}\delta_{ij}\\
&\quad - \sum_{ij\in\mathcal{A}} \sum_{k\in OD^*} \frac{\alpha d_k o_k}{n_k-1} \sum_{\substack{p\in P_k:\\ ij\in p}} \sum_{i'j'\in\mathcal{A}} \sum_{\substack{q\in P_k:\\ q\neq p,\\ i'j'\in q}} \delta_{i'j'}\delta_{ij}\\
&= \sum_{ij\in\mathcal{A}} \sum_{i'j'\in\mathcal{A}} \sum_{k\in OD^*} \alpha d_k o_k \left(\sum_{\substack{p\in P_k:\\ ij,i'j'\in p}} 1 - \frac{1}{n_k-1} \sum_{\substack{p\in P_k:\\ ij\in p}} \sum_{\substack{q\neq p\in P_k:\\ i'j'\in q}} 1 \right) \delta_{i'j'}\delta_{ij}
\end{aligned}$$

This yields the proposed matrix A as given in Equation (4.9).

$$A_{ij,i'j'} = \sum_{k \in OD^*} \alpha d_k o_k \left(\sum_{\substack{p \in P_k: \\ ij,i'j' \in p}} 1 - \sum_{\substack{p \in P_k: \\ ij \in p}} \sum_{\substack{q \neq p \in P_k: \\ i'j' \in q}} \frac{1}{n_k - 1} \right)$$

Next, the linear term will be computed.

$$lin(\delta_{ij}) = \sum_{ij \in \mathcal{A}} \sum_{k \in OD} \sum_{\substack{p \in P_k: \\ ij \in p}} \frac{1}{n_k} o_k \delta_{ij}$$

This yields directly the proposed vector b.

$$b_{ij} = \sum_{k \in OD} \sum_{p \in P_k : ij \in p} \frac{o_k}{n_k}$$

D Proof of negative semi-definiteness of coefficient matrix A

To prove that the coefficient matrix A is negative definite, it is sufficient to prove that $\delta^T A\delta \leq 0$ holds for all $\delta \in \mathbb{R}^{|\mathcal{A}|}$. For this purpose, we rewrite the term $\delta^T A\delta$ using the definition of the path length t_p.

$$
\begin{aligned}
&\delta^T A\delta \\
&= \sum_{ij\in\mathcal{A}} \sum_{i'j'\in\mathcal{A}} \sum_{k\in OD} \alpha d_k o_k \left(\sum_{\substack{p\in P_k: \\ ij,i'j'\in p}} 1 - \frac{1}{n_k-1} \sum_{\substack{p\in P_k: \\ ij\in p}} \sum_{\substack{q\in P_k: \\ q\neq p, \\ i'j'\in q}} 1 \right) \delta_{i'j'}\delta_{ij} \\
&= \sum_{k\in OD} \alpha d_k o_k \left(\sum_{p\in P_k} \sum_{\substack{ij\in\mathcal{A} \\ ij\in p}} \sum_{\substack{i'j'\in\mathcal{A} \\ i'j'\in p}} \delta_{i'j'}\delta_{ij} - \frac{1}{n_k-1} \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} \sum_{\substack{ij\in\mathcal{A}: \\ ij\in p}} \sum_{\substack{i'j'\in\mathcal{A} \\ i'j'\in q}} \delta_{i'j'}\delta_{ij} \right) \\
&= \sum_{k\in OD} \alpha d_k o_k \left(\sum_{p\in P_k} \underbrace{\sum_{ij\in p} \delta_{ij}}_{t_p} \underbrace{\sum_{i'j'\in p} \delta_{i'j'}}_{t_p} - \frac{1}{n_k-1} \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} \underbrace{\sum_{ij\in p} \delta_{ij}}_{t_p} \underbrace{\sum_{i'j'\in q} \delta_{i'j'}}_{t_q} \right) \\
&= \sum_{k\in OD} \alpha d_k o_k \left(\sum_{p\in P_k} t_p^2 - \frac{1}{n_k-1} \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} t_p t_q \right) \\
&= \sum_{k\in OD} \frac{\alpha d_k o_k}{2(n_k-1)} \left(2 \sum_{p\in P_k} (n_k-1) t_p^2 - 2 \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} t_p t_q \right) \\
&= \sum_{k\in OD} \frac{\alpha d_k o_k}{2(n_k-1)} \left(\sum_{p\in P_k} (n_k-1) t_p^2 - \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} 2 t_p t_q + \sum_{p\in P_k} (n_k-1) t_p^2 \right) \\
&\overset{(*)}{=} \sum_{k\in OD} \frac{\alpha d_k o_k}{2(n_k-1)} \left(\sum_{p\in P_k} (n_k-1) t_p^2 - \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} 2 t_p t_q + \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} t_q^2 \right) \\
&= \sum_{k\in OD} \frac{\alpha d_k o_k}{2(n_k-1)} \sum_{p\in P_k} \left((n_k-1) t_p^2 - \sum_{\substack{q\in P_k: \\ q\neq p}} 2 t_p t_q + \sum_{\substack{q\in P_k: \\ q\neq p}} t_q^2 \right) \\
&= \sum_{k\in OD} \frac{\alpha d_k o_k}{2(n_k-1)} \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} \left(t_p^2 - 2 t_p t_q + t_q^2 \right) \\
&= \sum_{k\in OD} \frac{\alpha d_k o_k}{2(n_k-1)} \sum_{p\in P_k} \sum_{\substack{q\in P_k: \\ q\neq p}} (t_p - t_q)^2,
\end{aligned}
$$

where we used in the starred equation that

$$\begin{aligned}
&\sum_{p\in P_k}(n_k-1)t_p^2\\
=&n_k\sum_{p\in P_k}t_p^2-\sum_{p\in P_k}t_p^2\\
=&\sum_{p\in P_k}\sum_{q\in P_k}t_q^2-\sum_{p\in P_k}t_p^2\\
=&\sum_{p\in P_k}\sum_{\substack{q\in P_k:\\ q\neq p}}t_q^2.
\end{aligned}$$

Since $(t_p-t_q)^2$, o_k and α are positive, n_k is at least 2 and d_k is negative, it follows that

$$\delta^T A\delta \leq 0.$$

In our application, δ_{ij} represents the duration of each activity $ij \in \mathcal{A}$ in the event activity network, so we require them to be positive. However, in general no such restriction is required and the implication $\delta^T A\delta \leq 0$ holds for all $\delta \in \mathbb{R}^{\mathcal{A}}$. This implies that the matrix A is negative semi-definite for all practical relevant instances and the objective function is concave.

E Linearized formulation of Model (ID-LIN)

Using the binary representation

$$\delta_{ij} = l_{ij} + \sum_{m \in M_{ij}} 2^m \sigma_{ij}^m$$

with $M_{ij} \coloneqq \{0, \dots, \lfloor \log_2(u_{ij} - l_{ij}) \rfloor\}$ and replacing the product of the binary variables $\sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$ by $\gamma_{ij,i'j'}^{m,m'}$, we obtain a linearized version of the model (ID-LIN)

$$\begin{array}{llcll}
\min & \displaystyle\sum_{ij \in \mathcal{A}} \sum_{i'j' \in \mathcal{A}} A_{ij,i'j'} \left(\sum_{m' \in M_{i'j'}} 2^{m'} l_{ij} \sigma_{i'j'}^{m'} + \sum_{m \in M_{ij}} 2^m l_{i'j'} \sigma_{ij}^m + \sum_{m \in M_{ij}} \sum_{m' \in M_{i'j'}} 2^{m+m'} \gamma_{ij,i'j'}^{m,m'} \right) + \sum_{ij \in \mathcal{A}} b_{ij} \delta_{ij} & & & & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
& \Gamma_c \delta & = & \mu_c \cdot T & \forall c \in \mathcal{C} \\
& \gamma_{ij,i'j'}^{m,m'} & \leq & \sigma_{ij}^m & \forall ij, i'j' \in \mathcal{A} \colon A_{ij,i'j'} < 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \gamma_{ij,i'j'}^{m,m'} & \leq & \sigma_{i'j'}^{m'} & \forall ij, i'j' \in \mathcal{A} \colon A_{ij,i'j'} < 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \gamma_{ij,i'j'}^{m,m'} & \geq & \sigma_{ij}^m + \sigma_{i'j'}^{m'} - 1 & \forall ij, i'j' \in \mathcal{A} \colon A_{ij,i'j'} > 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \sigma_{ij}^m & \in & \{0,1\} & \forall ij \in \mathcal{A},\ \forall m \in M_{ij} \\
& \gamma_{ij,i'j'}^{m,m'} & \in & \{0,1\} & \forall ij, i'j' \in \mathcal{A},\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
& \mu_c & \in & \mathbb{Z}_+ & \forall c \in \mathcal{C}
\end{array}$$

For the sake of simplicity we omitted the constant term $\sum_{ij \in \mathcal{A}} \sum_{i'j' \in \mathcal{A}} A_{ij,i'j'} l_{ij} l_{i'j'}$ in the objective function.

To prove that $\gamma_{ij,i'j'}^{m,m'} = \sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$ in an optimal solution to the ILP, two cases are considered

- If $A_{ij,i'j'} < 0$, then $\gamma_{ij,i'j'}^{m,m'} \leq \min\{\sigma_{ij}^m, \sigma_{i'j'}^{m'}\}$. That means, as soon as $\sigma_{ij}^m = 0$ or $\sigma_{i'j'}^{m'} = 0$ it follows that $\gamma_{ij,i'j'}^{m,m'} = 0$. Otherwise $\gamma_{ij,i'j'}^{m,m'} = 1$ can be chosen, which is preferable since $A_{ij,i'j'} < 0$.
- If $A_{ij,i'j'} > 0$, then $\gamma_{ij,i'j'}^{m,m'} \geq \max\{0, \sigma_{ij}^m + \sigma_{i'j'}^{m'} - 1\}$. That means, as soon as $\sigma_{ij}^m = 1$ and $\sigma_{i'j'}^{m'} = 1$ it follows that $\gamma_{ij,i'j'}^{m,m'} = 1$. Otherwise $\gamma_{ij,i'j'}^{m,m'} = 0$ can be chosen, which is preferable since $A_{ij,i'j'} > 0$.

It follows that $\gamma_{ij,i'j'}^{m,m'} = \sigma_{ij}^m \cdot \sigma_{i'j'}^{m'}$ for $A_{ij,i'j'} \neq 0$. Note that for this linearization only constraints for non-zero coefficients $A_{ij,i'j'}$ have to be installed. By definition of $A_{ij,i'j'}$, this can only happen if two activities ij and $i'j'$ are contained in paths for the same OD-pair.

To decrease the search space in the computations we additionally add the constraints

$$\begin{array}{lcll}
\gamma_{ij,i'j'}^{m,m'} & = & \gamma_{i'j',ij}^{m',m} & \forall ij < i'j' \in \mathcal{A} \colon A_{ij,i'j'} \neq 0,\ \forall m \in M_{ij},\ \forall m' \in M_{i'j'} \\
\gamma_{ij,ij}^{m,m'} & = & \gamma_{ij,ij}^{m',m} & \forall ij \in \mathcal{A} \colon A_{ij,ij} \neq 0,\ \forall m < m' \in M_{ij} \\
\gamma_{ij,ij}^{m,m} & = & \sigma_{ij}^m & \forall ij \in \mathcal{A} \colon A_{ij,ij} \neq 0,\ \forall m \in M_{ij}
\end{array}$$

F Path choice sets

In this section we describe how we preprocess the instances to derive a path choice set based on the line plan and the line frequency.

The quality of the path choice sets is of major importance for the quality of the results. There are two possibilities when designing a choice set. One should either take all alternatives into account and let 'the choice model decide that the choice probabilities of unrealistic options are low or zero' or take 'into account only subsets of the options which are effectively chosen in the sample' (or a heuristic approximation of that) (de Dios Ortuzar and Willumsen, 2011). Since a large path choice set implies a large number of variables and constraints in both models, the first option of including all alternatives is impractical in this setting. However, if the choice sets are too small, they might not contain all routes that are important for passengers and the timetable will be constructed neglecting some important connections.

There are multiple path generation algorithms available in literature to derive realistic path choice sets. Although most of them focus on road networks, the concept is often applicable to public transport networks as well. An overview of different algorithms is given in Bekhor et al. (2006). The magnitude of good alternative paths in such choice sets turned out to be very large complicating the solution process of the models. Since these choice sets often contain many routes that are identical to some extent, we use a simple heuristic to find small choice sets of independent paths that are of comparable quality as described in Sels (2015). For each OD pair we iteratively add the shortest path in the event activity network to the path choice set and delete all visited vertices. This procedure is repeated until origin and destination are not connected any more or the found path is too long or too inconvenient to be a possible alternative. Furthermore, we set the lengths of all transfer activities to the upper bounds and apply the same procedure to the event activity network again to find a choice set of paths with minimal number of transfers. We take the union of these two sets and remove dominated paths.

This heuristic has shown to provide a small choice set for each OD pair with a representative selection of paths. The size of the choice sets is small enough for usage in an optimization framework. At the same time, all paths are good alternatives for the passengers and they are independent of each other.

G Reference methods

G.1 Integer programming formulation for model (IS)

We implement the following mixed integer linear program for timetabling with integrated shortest route search for passengers.

$$
\begin{array}{llcll}
\min & & \sum_{k \in OD} o_k t_k & & \\
\text{s.t.} & \delta_{ij} & \geq & l_{ij} & \forall ij \in \mathcal{A} \\
& \delta_{ij} & \leq & u_{ij} & \forall ij \in \mathcal{A} \\
& \Gamma_c \delta & = & T \cdot \mu_c & \forall c \in \mathcal{C} \\
& t_p & = & \sum_{ij \in p} \delta_{ij} & \forall k \in OD,\ \forall p \in P_k \\
& t_k & = & \min_{p \in P_k} t_p & \forall k \in OD,\ \forall p \in P_k, \\
& \delta_{ij} & \in & \mathbb{Z}_+ & \forall ij \in \mathcal{A} \\
& \mu_c & \in & \mathbb{Z} & \forall c \in \mathcal{C} \\
& t_p & \in & \mathbb{R}_+ & \forall k \in OD,\ \forall p \in P_k \\
& t_k & \in & \mathbb{R}_+ & \forall k \in OD
\end{array}
$$

The minimum is linearized with the following set of constraints

$$
\begin{array}{rcll}
t_{kr} & \leq & t_p & \forall k \in OD,\ \forall p \in P_k \\
t_{kr} & \geq & t_p - (1 - z_p) M_k & \forall k \in OD,\ \forall p \in P_k \\
\sum_{p \in P_k} z_p & = & 1 & \forall k \in OD \\
z_p & \in & \{0, 1\} & \forall k \in OD,\ \forall p \in P_k
\end{array}
$$

where

$$
M_k = \max_{p \in P_k} \sum_{ij \in p} u_{ij} - \min_{p \in P_k} \sum_{ij \in p} l_{ij}.
$$

G.2 Pseudo code for method (ID-ITR)

We give the pseudo code of the algorithm used for the iterative method (ID-ITR).

Algorithm 1: Pseudo code for method (ID-ITR)

Input: Instance $\mathcal{I} = (\mathcal{N}, l, u, OD)$

Output: Durations δ, cycle numbers z and passenger distribution x

Initialize: Set edge lengths to average bounds $\delta_{ij} \coloneqq \frac{1}{2}(l_{ij} + u_{ij})$;

while *Termination criterion not reached* **do**

- Compute passenger weight on edges with Equations (3.5), (3.6) and (3.7) and the logit model in Equation (3.1) as distribution function;
- Find timetable with given weights with model from Equations (3.2), (3.3) and (3.4);

Return (δ, z, x);

The termination criterion is reached if one of the following three conditions is met:

1. The improvement in objective value between two iterations is marginal, i.e., if

$$\frac{v_i - v_{i-1}}{v_i} < 0.01,$$

   where v_i denotes the objective value of the i-th iteration.

2. The number of iterations exceeds 4.

3. The total CPU time exceeds 5400 seconds.