Timetabling for strategic passenger railway planning

Gert-Jaap Polinder [a,b], Marie Schmidt [a,b,*], Dennis Huisman [b,c,d]

[a] *Rotterdam School of Management, Erasmus University, P.O. Box 1738, Rotterdam 3000DR, the Netherlands*
[b] *Erasmus Center for optimisation in Public Transport (ECOPT), Rotterdam, the Netherlands*
[c] *Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam 3000DR, the Netherlands*
[d] *Process quality and Innovation, Netherlands Railways, P.O. Box 2025, Utrecht 3500HA, the Netherlands*

A B S T R A C T

In research and practice, public transportation planning is executed in a series of steps, which are often divided into the strategic, the tactical, and the operational planning phase. Timetables are normally designed in the tactical phase, taking into account a given line plan, safety restrictions arising from infrastructural constraints, as well as regularity requirements and bounds on transfer times.

In this paper, however, we propose a timetabling approach that is aimed at decision making in the strategic phase of public transportation planning and to determine an outline of a timetable that is good from the passengers' perspective. Instead of including explicit synchronization constraints between train runs (as most timetabling models do), we include the adaption time (waiting time at the origin station) in the objective function to ensure regular connections between passengers' origins and destinations. We model the problem as a mixed integer quadratic program and linearize it. Furthermore we propose a heuristic to generate starting solutions. We illustrate the trade-offs between dwell times and regularity of trains in two case studies based on the Dutch railway network.

© 2021 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY license
(http://creativecommons.org/licenses/by/4.0/)

1. Introduction

The public transportation planning process is traditionally subdivided into a number of steps which are assigned to either the strategic, the tactical, or the operational planning phase. According to Huisman et al. (2005), the strategic phase encompasses a time horizon of two to ten years before implementation and includes infrastructure decisions and line planning. The timetabling problem is often allocated to the tactical phase (approximately one year before implementation). Timetabling in the tactical phase takes a given line plan, safety restrictions arising from infrastructural constraints, as well as regularity requirements and bounds on transfer times as input.

This paper, however, focuses on *strategic timetabling*, i.e., the generating of a (preliminary) timetable already in the strategic planning phase. Strategic timetabling can be used to make strategic decisions with respect to timetables, like *"What should the headway times be between consecutive trains at a station?"* and *"Where should good transfer connections between*

---

* Corresponding author at: Rotterdam School of Management, Erasmus University, P.O. Box 1738, 3000DR Rotterdam, the Netherlands.
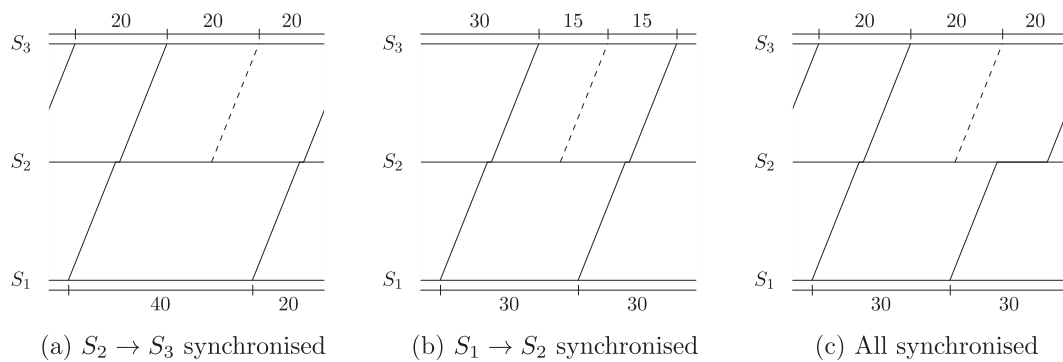*E-mail address:* schmidt2@rsm.nl (M. Schmidt).

(a) $S_2 \to S_3$ synchronised (b) $S_1 \to S_2$ synchronised (c) All synchronised

Fig. 1. Time space diagrams for different synchronisation options.

trains be made?". Due to the location of strategic timetabling early in the planning horizon, it can also be used to evaluate line plans and to point to bottlenecks in the infrastructure (and thus to promising infrastructure investments).

The value of strategic timetabling has been recognized in the practice of transportation planning. Following the example of Switzerland, the initiative *Deutschland-Takt* aims at establishing a so-called 'integraler Taktfahrplan' in Germany from the year 2030 on (Deutschland-TAKT, 2019). The transportation system should be redeveloped in such a way that connections between cities are served every 30 or 60 min, and that better transfer connections are provided. Reversing the current planning practice, the creation of a so-called 'target timetable' should precede and guide infrastructure investment decisions (e.g. in additional tracks between stations, or additional platforms at stations). In the Netherlands, a similar approach is used to evaluate infrastructure investments (Beter and Meer, 2014).

However, to the extent of our knowledge, models from academic research on timetabling as well as software tools for timetabling decision support are aimed at timetabling in the tactical (and operational) planning phase. Therefore, they focus on operational feasibility on a given infrastructure, and are rather restrictive in modeling of quality requirements. While these features are suitable for the more restrictive setting of tactical and operational planning (where changes in infrastructure and major changes in passenger behaviour are not desirable or possible), they are not appropriate to find new and innovative timetabling solutions as is desirable in the strategic planning phase.

In this paper we aim to close this gap by presenting an optimisation approach to strategic timetabling. As common in railway timetabling, we aim at finding a periodic timetable, i.e., we require that the timetable follows a repeating pattern and hence the timetable of a base period is repeated throughout the day. Our objective is to find a periodic timetable that minimizes average perceived travel time for a given line plan. Different from most other timetabling models, we include adaption time in the perceived travel time. Adaption time describes the time difference between the desired departure time of a passenger and his actual departure time. This allows us to omit regularity constraints between runs of the same line (or runs of different lines that run in parallel for part of their route), which are otherwise often used to ensure low adaption times in an indirect way, simply by enforcing regular departures. Note that in case of dense networks and high frequencies, where OD-pairs are served by more than one line, the question which trains should be synchronized with each other becomes far from trivial to answer. In such situations, imposing regularity constraints may lead to sub-optimal solutions or even infeasibility of the timetabling problem (Polinder et al., 2018).

This is illustrated in the following example. Consider three stations S_1, S_2, S_3, and travel demand between all pairs of stations. Assume that the line plan prescribes a line from S_1 via S_2 to S_3 with a frequency of two trains per hour, and a line from S_2 to S_3 with a frequency of one train per hour. If we synchronize all trains between S_2 and S_3, the headway time between the trains on this part of the route will be 20 min, but from S_1 to S_2 the headway times are 20 min and 40 min. This is depicted in the time-space diagram in Fig. 1a, where time and distance are shown on the horizontal and vertical axis, respectively. On the other hand, if we synchronize between S_1 and S_2, we have one headway time of 30 min and two shorter headway times between S_2 and S_3 (Fig. 1b). Perfect synchronization on both parts of the network is possible, but only if one of the trains from S_1 to S_2 waits an additional 10 min at S_2 (Fig. 1c). Which of these solutions is best with respect to average perceived travel times depends on the size of the travel demand between the stations and the perceived value of adaption time compared to in-train time, as will be further illustrated in Section 3.4.

Our timetabling model allows us to find the best trade-off in case of different synchronization options for more complicated networks than the one sketched by explicitly including the adaption time into the perceived travel time, instead of deciding on where to impose regularity constraints heuristically before the optimisation.

We define the *Strategic Passenger-Oriented Timetabling (SPOT)* problem as follows:

Given a railway network consisting of stations and links connecting them, a line plan specifying lines routes and frequencies on the network, and an estimate of the hourly expected demand per origin-destination pair: find a timetable that minimizes average perceived passenger travel time under the assumption that passengers will choose the route with least perceived travel time. Hereby, perceived travel time is measured from the desired departure time on, that is, it includes adaption time.

Since we consider timetabling in the strategic planning phase, we cannot expect to have accurate demand information. In particular, the exact timing of travel requests is unknown. Therefore, we think that in this time frame it is appropriate to model passengers' desired departure times as evenly spread over the period and explicitly use this assumption in our mathematical program for the SPOT problem. Both the assumption that passengers indeed arrive randomly at the station and the assumption that they adapt to the communicated timetable to a large extent can be modeled by a parameter in our objective function which relates the perceived duration of waiting at the origin to in-train time.

We model the SPOT problem as a quadratic mixed integer program that extends the traditional PESP model for periodic timetabling (Serafini and Ukovich, 1989). We linearize the model and develop a heuristic to find a starting solution. We test our approach in two case studies based on the Dutch railway network.

Note that infrastructure constraints can be included in PESP (and thus also in our SPOT model) as headway constraints in a natural way (Liebchen and Möhring, 2007). However, due to the strategic perspective we take, we do not include them in our approach for two reasons: to be able to identify promising infrastructure investments, and to keep the model tractable. In later planning phases (tactical and operational planning), when the timetabling focus shifts towards macroscopic and later microscopic feasibility, such constraints can (and should) be added. Furthermore, in our model we omit upper bounds on transfer times and regularity constraints, since this would artificially restrict the solution space and long transfers and irregular departure patterns will already be penalized in the objective function.

Our contribution in this paper is fourfold. First, we formulate the *Strategic Passenger-Oriented Timetabling (SPOT)* problem for timetabling in the strategic planning phase. Second, we model this problem as a quadratic integer program that integrates timetabling with passenger routing (on perceived-travel-time-minimal routes) and linearize this formulation. Third, we propose a heuristic to construct a starting solution, in order to find good solutions even for complex large instances. Fourth, we test our model on two case studies based on the Dutch railway network, illustrating the trade-offs between the duration of dwell times and regularity of train service.

The remainder of this paper is organized as follows. In Section 2, we describe literature that is related to our study. We discuss the SPOT problem in detail in Section 3. In Section 4 we propose a quadratic integer programming model for SPOT and linearize it. In Section 5 we describe how we solve it. In Section 6 we evaluate our solution approach and perform a case study on two practical instances from Netherlands Railways. Finally, we conclude the study in Section 7.

2. Related work

In this section, we give an overview on related research. In Section 2.1 we describe other attempts to timetabling in the strategic planning phase. Section 2.2 gives a brief overview on periodic timetabling. Section 2.3 describes how passenger routing can be combined with timetabling and how this is done in existing literature.

2.1. Strategic timetabling

Decisions in the public transportation planning process are traditionally assigned to either the strategic, the tactical or the operational planning phase, according to their location in the planning horizon. E.g., according to Huisman et al. (2005), the strategic planning phase at the Dutch railway operator Netherlands Railways (NS) spans a period from 20 to 'few' years before the date of implementation and encompasses rolling stock management, crew management, and line planning. Timetabling and rolling stock scheduling are allocated to the tactical planning phase, and operational and short-term planning include detailed timetabling and rolling stock scheduling, as well as crew scheduling. However, finding a feasible and good timetable for a given line plan becomes increasingly difficult with an increasing utilization of the railway network. We therefore propose to include (a first attempt on) timetabling already in the strategic planning phase, to be able to use it as a tool to evaluate, and possibly reject proposed line plans. Several attempts have been made in the literature to *integrate* line planning and timetabling, see, e.g., in Michaelis and Schöbel (2009); Burggraeve et al. (2017); Schöbel (2017a); Yan and Goverde (2019). We follow a different approach, regarding timetabling as a stand-alone problem (as opposed to integrating it with line planning), but taking into account typical characteristics of the strategic planning phase, like the lack of detailed passenger information and the possibility to extend infrastructure. While in the practice of public transportation planning, strategic attempts on timetabling are not uncommon (see, e.g., Deutschland-TAKT (2019)), research on railway planning normally allocates timetabling to the tactical phase. An omission of infrastructure constraints is not uncommon in railway timetabling, see e.g., Robenek et al. (2017, 2016); Pätzold et al. (2017); Schmidt and Schöbel (2015); Borndörfer et al. (2017), but other than omitting infrastructure constraints, the approaches cited here do not seem particularly tailored to strategic planning.

2.2. Periodic timetabling

The Periodic Event Scheduling Problem (PESP) model that is commonly used for periodic railway timetabling was introduced in Serafini and Ukovich (1989). Overviews on how to model timetabling constraints and extensions in a PESP framework can be found, e.g., in Odijk (1996); Peeters (2003). Liebchen and Möhring (2007) provide a discussion on what can

be included in the PESP framework, and what cannot, like symmetry of timetables and maximum headway times between consecutive trains.

In its original formulation, PESP is a feasibility problem. Approaches to solve it include besides integer programming approaches (Liebchen, 2008) also constraint programming (Kroon et al., 2009), the modulo-simplex heuristic (Nachtigall and Opitz, 2008; Goerigk and Schöbel, 2013), a matching-approach (Pätzold and Schöbel, 2016), using a Satisfiability (SAT) solver after applying a polynomial transformation from PESP to SAT (Großmann et al., 2012; Kümmling et al., 2015), or by combining SAT with machine learning (Matos et al., 2018). If a feasible solution exists, this can often be found rapidly.

The integer program for PESP can be extended with an objective function, to find *good* timetables, as several of the aforementioned approaches do. For example, a weighted sum of the activity durations can be minimized. See, e.g., Peeters (2003); Nachtigall (1999); Liebchen (2008); Liebchen and Peeters (2009); Caimi et al. (2017). More details about the PESP model are provided in Section 4.

2.3. Timetabling and passenger routing

Timetables are evaluated according to different criteria. Following Cacchiani and Toth (2012), the most prominent are that the timetable should be (1) 'efficient' and (2) 'robust'. Overviews on approaches to deal with robustness can be found in Cacchiani and Toth (2012); Lusby et al. (2018). Efficiency can be aimed both at costs and travel time aspects (or a trade-off of both). In the following we give an overview on how the literature addresses one aspect of "efficiency", namely minimizing the passenger travel times, since this is also the objective we use in our model.

Early approaches to find timetables minimize passenger waiting times by assigning weights, modeling passenger numbers, to activities in the objective function (see the aforementioned references). This approach, however, neglects that passengers choose their routes based on the timetable, which makes it difficult to assign a-priori weights to activities.

Thus, better results can be obtained when timetable and passenger routing are determined simultaneously. Several approaches have been published regarding an integrated approach, both in periodic and aperiodic settings. Schmidt and Schöbel (2015) integrate passenger routing in aperiodic timetabling. Passenger demand is a priori assigned to a departure event, and passengers are routed from that point onwards. For periodic timetabling, a similar approach is taken by Borndörfer et al. (2017). In this approach, train capacities are used to determine frequencies of train lines. Furthermore, many performance criteria are introduced regarding several routing methods. A more recent approach where a viewpoint on applicability in practice is taken is by Schiewe and Schöbel (2020). The authors study the effect of including only a subset of the OD-pairs, in order to reduce the computational effort and to obtain good timetables in a short time. Hartleb and Schmidt (2019) also integrate timetabling with passenger routing, but here passengers are not routed along shortest paths, but a logit distribution is used. An alternative to integrating timetabling and passenger routing in one integer programming model, is to iterate timetabling and passenger routing. Kinder (2008); Lübbe (2009); Siebert (2011); Siebert and Goerigk (2013) determine passenger flows by routing passengers through the network on, for example, shortest paths with respect to the travel time. After this, the timetable is optimized and passenger are rerouted, until a stopping criterion is reached.

The division of the public transportation planning into several sub problems (like line planning, timetabling, vehicle scheduling, etc.) is likely to lead to globally suboptimal solutions. Therefore, there are attempts to also integrate line planning and vehicle scheduling into timetabling with passenger routing (see, e.g., Schöbel, 2017b; Lübbecke et al., 2018). However, for real-life instances this leads to models that are hard to solve, and in these cases each sub problem is solved separately.

None of the aforementioned approaches considers adaption time, although a few approaches exist who explicitly consider this. Some of them focus on a single corridor and schedule the trains such that the average adaption time is minimized (Barrena et al., 2014a; 2014b). In these papers, a mathematical programming model and a large-neighbourhood search algorithm to find good solutions is provided. A similar situation is considered in Zhu et al. (2017), where the authors consider a bi-level model. In the upper layer, a timetable is found based on passenger demand. In a lower layer, passenger arrival times are updated such that passengers arrive shortly before their train departs, in order to minimize adaption time. Another approach to solve timetabling with integrated passenger routing including adaption times is given in Gattermann et al. (2016), where the problem is modelled as a Satisfiability problem and solved with a dedicated solver. Here passengers are assigned to a time slice and a penalty is given if a passenger cannot depart in that time slice and has to adapt to a different slice. Borndörfer et al. (2017) and Hartleb et al. (2019) discuss and compare alternative evaluation functions for passenger-oriented timetabling. Wang et al. (2015) propose an approach to reduce the operation costs of train which models demand as time-dependent and includes route choice. Finally, Yin et al. (2017) include passenger demand and adaption time minimization into an approach to optimize energy efficiency.

In this paper, we integrate timetabling and passenger routing in a mathematical model and include the adaption time of passengers. By discarding the current infrastructural situation, solutions to our model can be used to support decision making in the strategic planning phase of railway planning.

3. Problem definition

3.1. Periodic timetabling

A timetable π is an assignment of points in time π_i to events i, that is: to departures and arrivals of trains at stations. In order to lead to a feasible timetable, the departure times need to be sequenced in time: A train cannot depart from a station before it has arrived there and, if it is scheduled to stop at the station, before it has dwelt there for a minimum amount of time. The arrival time of a train at a station is constrained by its departure time at the previous station and the minimum and maximum driving time between the two stations.

In periodic timetabling, the events reoccur each period. E.g., if the base period is one hour, if a train of line 1 departs from station A at time 7:05, a second train of line 1 will depart from station A at time 8:05, a third one at time 9:05, and so on. It is therefore sufficient to define one *periodic* event at time :05 that represents all departures of line 1 at station A simultaneously. We denote the duration of the base period as T.

The *periodic event-scheduling (PESP)* approach (Serafini and Ukovich, 1989) visualizes the set of periodic events and constraints on these events in a so-called *event-activity network* $G = (V, A)$. Each activity $(i, j) \in A$ represents a constraint and is formulated as a relation between events $i, j \in V$, stating that the time difference between these two events should be in a given (periodic) time interval, bounded by a lower bound ℓ_{ij} and an upper bound u_{ij}, where the word 'periodic' refers to the fact that event times are considered modulo T. This can be formalized by requiring that a timetable is a function $\pi : V \to \{0, \ldots, T - 1\}$ on the departure and arrival events V such that

$$y_{ij}(\pi) = \left(\left(\left(\pi_j - \pi_i - \ell_{ij} \right) \bmod T \right) + \ell_{ij} \right) \in \left[\ell_{ij}, u_{ij} \right] \tag{3.1}$$

for all activities $(i, j) \in A$. We call $y_{ij}(\pi)$ the duration of activity (i, j) under timetable π.

In our event-activity network, drive activities constrain the time that a train may need between a departure at a station and the arrival at a next station, based on minimum and maximum speed of the train. Dwell activities impose a lower bound on the time between the arrival of a train at a station and its subsequent departure (to define an upper bound without imposing a restriction on dwell times, we can set $u_{ij} := T + \ell_{ij} - 1$).

It is common in periodic timetabling to impose more constraints in the form of activities: firstly, a set of constraints that impose *infrastructure* constraints (trains that use a common part of the infrastructure have to keep a safety distance), secondly a set of constraints that is aimed to increase the *quality* of a timetable from the passengers' perspective by enforcing, e.g., synchronization of trains of the same line (or of lines with a similar route), upper bounds on dwell times and coordination of transfers. Overviews on how to model a variety of timetabling constraints in a PESP framework can be found in Odijk (1996); Peeters (2003); Liebchen and Möhring (2007).

In our model, the quality of the timetable is controlled by the objective function 'total perceived travel time'. Therefore, the set of quality constraints is unnecessary in our model. For routing purposes, however, we do include transfer activities in our event-activity network, that connect the arrival event of a train at a station with the departure events of other trains at the same station. The bounds on these activities are set to $[\ell_{ij}, T + \ell_{ij} - 1]$, with ℓ_{ij} representing the minimum time needed for a transfer between arrival i and departure j. In this way, transfer activities do not impose any constraint on the feasibility of a timetable, but can be used in a convenient way to describe passenger routes and to compute their perceived travel time. This is detailed in Section 3.2.

Furthermore, in many cases, in the strategic phase of the timetabling process, infrastructure does not impose hard constraints on the timetable yet. Therefore, we propose to disregard (most of) the infrastructure constraints for timetabling in the strategic phase as well.

Note that without infrastructure constraints and quality constraints, the event-activity network consists of train paths, which are not connected to each other (except for the transfer activities whose bounds are sufficiently loose not to impose any feasibility constraints). Therefore, finding a *feasible* timetable on such an event-activity network is trivial - just choose an arbitrary time for each first event of a train and propagate it along the train path. In other words, the difficulty of solving PESP as a feasibility problem (see Section 2) can be attributed to the infrastructure and quality constraints.

It depends on the planning horizon and the budget for investments available, to which extent infrastructure availability imposes hard constraints on the timetable, or is subject to extension where needed. In this paper, we neglect infrastructure constraints, due to the early planning phase in which we consider the timetabling problem. Adding all or some infrastructure constraints would pose no problem from the modelling perspective, but would make the problem computationally harder to solve.

In the same way, quality constraints may be added to our model if they are considered indispensable. We do believe, however, that in most cases the addition of quality constraints does unnecessary constrain the set of feasible solutions, and that our objective function will find timetables of better quality if no quality constraints are imposed.

Note that, as common in the railway planning literature, we do not yet consider the rolling stock allocation to trips during timetabling, but postpone these questions to the *Rolling Stock Scheduling* problem which is solved in the tactical planning phase and outside of the scope of this paper. In line with that, we do not consider constraints on the maximum number of passengers in a train, or operational costs, as these are largely determined by the rolling stock schedule.

3.2. Passenger demand, route choice, and perceived travel time

We consider passenger demand aggregated over all time periods, by assuming that we have an OD-matrix $\mathcal{OD}$, providing for each OD-pair $k \in \mathcal{OD}$ an estimate of the average number of passengers d_k that want to travel from the origin to the corresponding destination per time period. While demand may differ between different time periods, we assume that within the same time period, demand is uniformly distributed. For each OD-pair $k \in \mathcal{OD}$, we precompute the set of routes among which the passengers may choose. A route r is a directed path through the Event-Activity Network. It consists of a sequence of trip, dwell, and transfer activities, so r is an (ordered) subset of A. The set of routes for OD-pair $k \in \mathcal{OD}$ is denoted by $\mathcal{R}^k$. The set of all routes is

$$\mathcal{R} = \bigcup_{k \in \mathcal{OD}} \mathcal{R}^k. \tag{3.2}$$

For a fixed timetable π, the activity duration y_{ij} of drive, dwell, and transfer activities (i, j) is given by (3.1) in dependence of the chosen timetable. The *pure* travel duration of a route r is simply the sum of all activity durations on the route. To obtain the *perceived* travel duration $Y_r(\pi)$ of route r for timetable π, we penalize each transfer by adding a transfer penalty γ_t, that is

$$Y_r(\pi) := \sum_{a \in r} y_a + \gamma_t \cdot 1_t(a), \tag{3.3}$$

where the function $1_t(a)$ is an indicator function, denoting whether activity $a \in A$ is a transfer activity or not.

However, the route choice of a passenger does not only depend on the perceived route duration $Y_r(\pi)$, but also on the departure time of the route r. A passenger will prefer a route with slightly longer perceived duration, if its departure time is closer to his desired departure time. To formalize this idea, we introduce the notion of the perceived travel time of the passenger, which includes the *adaption time*. The *adaption time* $\mathrm{at}_r^t(\pi)$ for a passenger with desired moment of departure t on route r under timetable π is defined as the time difference between t and the scheduled time for the first event $\sigma(r)$ on route r $\pi_{\sigma(r)}$:

$$\mathrm{at}_r^t(\pi) := \left(\pi_{\sigma(r)} - t \right) \bmod T. \tag{3.4}$$

Consequently, the perceived travel time of a passenger with desired departure time t on route r under timetable π is

$$\hat{Y}_r^t(\pi) := \gamma_w \cdot \mathrm{at}_r^t(\pi) + Y_r(\pi), \tag{3.5}$$

where γ_w is a weighting factor modeling the perception of adaption time in relation to perceived duration of the route.

Note, however, that our demand data does not contain information on the individual desired departure moments of passengers, but only aggregated information in the form of average hourly demand for an OD-pair.

We therefore work with expected passenger numbers per minute. For an OD-pair with an average demand of d_k passengers per time period T, we assume that for each possible departure moment t we have on average $\frac{d_k}{T}$ passengers who would like to depart at time t.

Under the assumption that passenger demand is distributed uniformly over the period, and that each passenger chooses a route which minimizes his perceived travel time, we can thus compute the perceived travel time of a timetable π as

$$\mathrm{tt}(\pi) := \sum_{k \in \mathcal{OD}} \sum_{t=1}^{T} \frac{d_k}{T} \cdot \min_{r \in \mathcal{R}^k} \hat{Y}_r^t(\pi). \tag{3.6}$$

Note that demand will vary among periods. Since d_k is the average demand for OD-pair k per period, $\mathrm{tt}(\pi)$ is the total travel time under timetable π, averaged over all considered periods.

3.3. Time-sliced based reformulation of route choice and objective function

Note that the formulation of the objective (3.6) suggests that to evaluate a timetable, we have to determine $T \times |\mathcal{OD}|$ routes, one for each OD-pair and each moment in time. We can reduce the number of routes to be determined for the evaluation of a timetable considerably by dividing the time period into time slices for each OD-pair. To this end, we denote the set of *relevant* departure events for OD-pair k, that is, all first departure events $\sigma(r)$ of the routes in $r \in \mathcal{R}^k$ by

$$V^k = \bigcup_{r \in \mathcal{R}^k} \{\sigma(r)\}. \tag{3.7}$$

For an OD-pair k and a given timetable π, we divide the period into time slices, according to the departure times π_v of the relevant departure events $v \in V^k$. Let S_v^k denote the time slice between the relevant departure event v and the departure event $v'(\pi) \in V^k$ that immediately precedes v according to timetable π.

Note that while passengers with desired departure time t do not necessarily take the *next* departing route towards their destination, but the one that minimizes the perceived travel time $\hat{Y}_r^t(\pi)$ of which adaption time is only *one* component,

Table 1

Expected average adaption time per passenger for OD-pairs (S_i, S_j) in the timetables shown in Fig. 1. For OD-pair (S_2, S_3), additional waiting time for timetable (c) is added in brackets.

	Timetable		
	(a)	(b)	(c)
OD-(S_1, S_2)	16.67	15	15
pair(S_2, S_3)	10	11	10
(S_1, S_3)	16.67	15	16.67(+10)

for all passengers with departure time in the same time slice S_v^k, the same route will be optimal with respect to perceived travel time. Namely the route r with

$$Y_v^k(\pi) = \min_{r \in \mathcal{R}^k} \gamma_w \cdot \left(\left(\pi_{\sigma(r)} - \pi_v \right) \bmod T \right) + Y_r(\pi), \tag{3.8}$$

the time difference from v to the departure event of that route (weighted with the adaption time factor γ_w) plus the perceived route duration. Therefore, for each OD-pair k, we aggregate demand in each time slice S_v^k.

Let $L_v^k(\pi)$ denote the length of time slice S_v^k in timetable π. Then the number of passengers arriving during time slice S_v^k in timetable π is $\frac{d_k}{T} \cdot L_v^k(\pi)$, and we can compute the average waiting time for a passenger arriving during time slice S_v^k until π_v as $W_v^k(\pi) = L_v^k(\pi)/2$.

Thus, the total contribution to the total perceived travel time of passengers arriving during time slice S_v^k is

$$\frac{d_k}{T} \cdot L_v^k(\pi) \cdot \left(\gamma_w \cdot W_v^k(\pi) + Y_v^k(\pi) \right) \tag{3.9}$$

and we can replace (3.6) by the rewritten objective function

$$\mathrm{tt}(\pi) = \sum_{k \in \mathcal{OD}} \frac{d_k}{T} \sum_{v \in V^k} L_v^k(\pi) \cdot \left(\gamma_w \cdot W_v^k(\pi) + Y_v^k(\pi) \right). \tag{3.10}$$

This allows us to evaluate a timetable π by computing an optimal route per relevant departure event.

3.4. Example

Let us reconsider the example given in Fig. 1, to illustrate the evaluation of different timetables according to Eq. (3.6). We consider the three OD-pairs (S_1, S_2), (S_2, S_3), and (S_1, S_3) and the timetables (a), (b), and (c) from Fig. 1. Note that in each of the three considered timetables, there are three (non-dominated) routes for OD-pair (S_2, S_3), and two (non-dominated) routes for OD-pair (S_1, S_2) and (S_1, S_3). For each OD-pair, the time spent on driving activities is independent of the route chosen, and identical for the three timetable. The difference in the evaluation of the timetables stems purely from the adaption time, and for OD-pair (S_1, S_3) additionally from the dwell time at station S_2, which increases the in-train time of OD-pair (S_1, S_3) on the second route by 10 min under timetable (c).

Let us consider OD-pair (S_1, S_2) in timetable (a). There are two relevant departure events for this OD-pair, σ_1 at time $\pi_{\sigma_1} = 0$ and σ_2 at time $\pi_{\sigma_2} = 40$. Since $T = 60$ in this example, $L_{\sigma_1}^{12} = 20$ and $L_{\sigma_2}^{12} = 40$. Furthermore, $W_{\sigma_1}^{12} = L_{\sigma_1}^{12}/2 = 10$ and $W_{\sigma_2}^{12} = L_{\sigma_2}^{12}/2 = 20$.

Therefore, the expected total adaption time for OD-pair (S_1, S_2) in timetable (a) is $\frac{d_{12}}{60} \cdot (20 \cdot 10 + 40 \cdot 20) = d_{12} \cdot 16.67$, or 16.67 on average per passenger of this OD-pair. In the same way, we compute the adaption time for the other OD-pairs under the three different timetables. The results are summarized in Table 1.

It is now easy to see that it depends on passenger numbers (and the weighting factor γ_w that trades off adaption time and in-train time), which of these timetables is optimal with respect to total perceived travel time. E.g., if $d_{13} = 100$, and $d_{12} = d_{23} = 0$, we have a total adaption time of 15000 and no additional waiting time for timetable (b), while for timetable (a) and (c), the adaption time would be higher.

Furthermore, we observe that for each OD-pair, adaption time is lowest if the relevant departure events are evenly spaced. This is a consequence of the fact that $\sum_{v \in V^k} L_v^k = T$ for all $k \in \mathcal{OD}$ (the time slice lengths add up to the period length) and of the corresponding term in the objective function (3.10) being quadratic in L_v^k, since $W_v^k = L_v^k/2$. A formal proof is given in Appendix C.

3.5. Problem statement

We can now summarize the preceding sections by restating the definition of the *Strategic Passenger-Oriented Timetabling (SPOT)* problem given in Section 1:

Given a railway network consisting of stations and links connecting them, and a line plan specifying lines routes and frequencies on the network, described by a periodic event-activity network as detailed in Section 3.1 and an OD-matrix $\mathcal{OD}$, providing for each $k \in \mathcal{OD}$ an estimate of the average number of passengers d_k that want to travel from the origin to the corresponding destination per time period, find a periodic timetable π that fulfills constraints (3.1) and that minimizes the total perceived travel time tt(π) as defined in (3.10).

Summarizing, our timetabling problems differs from most other timetabling literature (compare Section 2) in the following two aspects:

On the one hand, our problem formulation explicitly aims at finding a timetable in the strategic planning phase. First, other than most publications on timetabling targeted to the tactical planning phase, we purposefully omit *infrastructure* constraints in our experiments, to allow our model to find an ideal timetable for a given line plan, and in doing so to point at bottlenecks and necessary infrastructure improvements. Note however, that from a modeling perspective, the inclusion of infrastructure constraints into our model is straightforward. Second, on a planning horizon of several years, only a coarse estimate of *demand data* can be obtained. In particular, no reasonable estimate on desired departure or arrival times can be made so long in advance. Our model requires only an estimate of daily averages per OD-pair. It assumes that demand is uniformly distributed over the period, but demand may differ between different periods, as is to be expected over a whole day of service.

On the other hand, by the definition of perceived travel time as objective function, which includes *adaption time* in the evaluation of a timetable, we are able to omit many *artificial constraints* that other models need to introduce to ensure *quality* of the timetable. Our objective function balances the need for low transfer times, low in-train waiting times, and equally spaced departures from the origins taking into account passenger numbers on the corresponding connections, and their impact on the total perceived travel time. The omission of artificial, and often arbitrarily set quality constraints expands the space of feasible solutions tremendously, and thus avoids the problem of severely restricted or even infeasible timetabling instances, which occurs increasingly often in view of the growing utilization of railway infrastructure (compare Polinder et al., 2018).

4. Integer programming formulation for SPOT

In this section, we model the SPOT problem as a mathematical program and linearize it. An overview of sets, constants, and variables used in the (linearization of the) mathematical programming for SPOT is given in Table 10 in Appendix B.

4.1. Precomputing passenger routes

We precompute possible routes for each OD-pair. To this end, based on a given line plan (specifying train lines, frequencies, and possible transfer options) we use the method described in Warmerdam (2004). This method first determines all direct travel options. Next, this set of travel options is extended by all options with 1 transfer, then with 2 transfers, and so on, until a predefined maximum number of transfers is reached. The (expected) duration of each of these routes is computed as the sum of the minimum time needed for the trips, increased with a small supplement for stops that have to be made, and multiplied by a percentage (which generally is 5%) to account for uncertainty in the trip durations and to incorporate some robustness against delays. If transfers are included in the route, additional time is added, based on an estimate of transfer times. To estimate transfer times, we proceed as follows: Assume that we consider a path from A to C, with transfer in B. We count the number n_1 of direct connections from A to B (as provided in the line plan) and the number n_2 of direct connections from B to C (as provided in the line plan) and estimate the transfer time as $T/(2\max\{n_1, n_2\})$. After the possible routes have been generated, a check is done whether some routes are dominated by others. In order to check dominance, routes are compared based on their expected duration, geographical length and number of transfers. As an example, if route A is longer with respect to travel time, and has more transfers than route B, it is discarded. Similarly, if two routes are identical, except for the transfer station that is chosen, the transfer station is chosen which has the highest number of trains and other options are discarded. Note that a different method would be possible as well, as long as the paths can be used as input to the model.

4.2. Mathematical programming formulation

We can formalize the SPOT problem as described in Section 3 as follows:

$$\min \sum_{k \in \mathcal{OD}} \frac{d_k}{T} \sum_{v \in V^k} L_v^k \cdot \left(\gamma_w \cdot W_v^k + Y_v^k \right) \tag{4.1a}$$

$$\text{s.t.} \, y_{ij} = \pi_j - \pi_i + T p_{ij} \qquad\qquad \forall \, (i, j) \in A \tag{4.1b}$$

$$\ell_{ij} \leq y_{ij} \leq u_{ij} \qquad\qquad \forall \, (i, j) \in A \tag{4.1c}$$

$$Y_r = \sum_{a \in r} (y_a + \gamma_t \cdot 1_t(a)) \qquad\qquad \forall\, r \in \mathcal{R} \qquad\qquad (4.1\mathrm{d})$$

$$L_v^k = \min_{v' \in V^k \setminus \{v\}} \{\pi_v - \pi_{v'} + T\alpha_{v',v}\} \qquad\qquad \forall\, k \in \mathcal{OD},\ v \in V^k \qquad\qquad (4.1\mathrm{e})$$

$$\alpha_{v,v'} + \alpha_{v',v} = 1 \qquad\qquad \forall\, k \in \mathcal{OD},\ v \in V^k,\ v' \in V^k \setminus \{v\} \qquad\qquad (4.1\mathrm{f})$$

$$W_v^k = \frac{1}{2} L_v^k \qquad\qquad \forall\, k \in \mathcal{OD},\ \forall\, v \in V^k \qquad\qquad (4.1\mathrm{g})$$

$$Y_v^k = \min_{v' \in V^k} \min_{r \in \mathcal{R}_{v'}^k} \{Y_r + \gamma_w \cdot (\pi_{v'} - \pi_v + T\alpha_{v,v'})\} \qquad\qquad \forall\, k \in \mathcal{OD},\ v \in V^k \qquad\qquad (4.1\mathrm{h})$$

$$L_v^k \in [0, T] \qquad\qquad \forall\, k \in \mathcal{OD},\ v \in V^k \qquad\qquad (4.1\mathrm{i})$$

$$W_v^k \in [0, T/2] \qquad\qquad \forall\, k \in \mathcal{OD},\ v \in V^k \qquad\qquad (4.1\mathrm{j})$$

$$Y_r, Y_v^k \in \mathbb{R}_{\geq 0} \qquad\qquad \forall\, r \in \mathcal{R},\ k \in \mathcal{OD},\ v \in V^k \qquad\qquad (4.1\mathrm{k})$$

$$\pi_v \in \{0, \ldots, T - 1\} \qquad\qquad \forall\, v \in V \qquad\qquad (4.1\mathrm{l})$$

$$p_{ij} \in \mathbb{Z}_{\geq 0} \qquad\qquad \forall\, (i, j) \in A \qquad\qquad (4.1\mathrm{m})$$

$$\alpha_{v,v'} \in \{0, 1\} \qquad\qquad \forall\, k \in \mathcal{OD},\ v \in V^k,\ v' \in V^k \setminus \{v\}. \qquad\qquad (4.1\mathrm{n})$$

Constraints (4.1b) and (4.1c) are equivalent to (3.1), with p_{ij} being an integer variable, representing the modulo operator. Constraints (4.1d) determine the perceived duration Y_r of each route r based on the activity durations associated with timetable π, as specified in (3.3). To determine the lengths of time slices in a periodic timetable, we introduce an additional binary variable $\alpha_{v',v}$ to replace the module operator. Then (4.1e) determines L_v^k, the length of time slice S_v^k, as the time distance to the previous periodic departure event in V^k. Hereby, constraints (4.1f) are required to impose an order between events, even when two departures happen at the same time. Constraints (4.1g) and (4.1h) determine average waiting time until the end of time slice S_v^k and perceived travel time on an optimal route from the end of time slice S_v^k, respectively, following Section 3.3. The objective aggregates the contributions of all time slices for all OD-pairs as given in (3.10). Constraints (4.1i)–(4.1n) state the domains of the variables.

4.3. Linearization of the SPOT model

The model stated in (4.1) is non-linear. The objective function is quadratic as it contains the term $W_v^k \cdot L_v^k = \frac{1}{2}(L_v^k)^2$. Next to that, (4.1e) and (4.1h) contain one or two minimums. For our computations, we linearize the objective function and reformulate (4.1e) and (4.1h). For the linearization, we define $Q_{v,v'}$ as the periodic time difference between events v and v', i.e.,

$$Q_{v,v'} = \pi_{v'} - \pi_v + T\alpha_{v,v'}. \qquad\qquad (4.2)$$

Then we can replace (4.1e) by the restrictions

$$L_v^k \le Q_{v',v} \qquad\qquad \forall k \in \mathcal{OD},\, v' \in V^k \setminus \{v\} \tag{4.3a}$$

$$\sum_{v \in V^k} L_v^k = T. \qquad\qquad \forall k \in \mathcal{OD} \tag{4.3b}$$

The first restriction represents the minimum and the second ensures that all time differences add up to T. Note that the latter is already a valid restriction in (4.1).

In order to linearize (4.1h), we introduce new binary variables $z_{v,v',r}^k$, denoting if passengers for OD-pair $k \in \mathcal{OD}$, arriving before event v use route r, starting with event v'. We refer to Appendix A for details. For the linearization of the objective function, we introduce new variables $x_{v,d}^k$, denoting whether $L_v^k \ge d$ or not. For the details, we refer again to Appendix A.

The model stated in this section determines a timetable that minimizes the total perceived travel time of all passengers. No synchronisation constraints are added to the model, instead, the objective is designed such that the optimal spread of trains over time is determined.

5. Solution approach

For real-life instances, even for networks of small size, the size and nature of the models easily exceed the capabilities of commercial solvers to find good solutions. Also due to the complex nature of the models, we do not expect to solve the models to optimality in a reasonable amount of time.

In this section we present the approach that we use in our experiments in order to find good solutions in reasonable time. First of all, we set a time limit TL. Secondly, we can simplify our SPOT model in various ways. These simplifications are described in Section 5.1. Thirdly, we use a heuristic method to generate a feasible starting solution, which is described in more detail in Section 5.2.

5.1. Reduced versions of SPOT

In this section we discuss some simplifications that can be made to the SPOT model, which lead to a reduced model size and therefore possibly speed up the solution process. These can be used as heuristic approaches towards solving the full SPOT model.

The first two simplifications use a subset of the OD-pairs instead of the full set. In the first simplification, we consider only passengers with a direct travel option. Note that passengers from this set do not *need* to travel directly if a better connection is available for them.

The second simplification is motivated by the observation that in practice, the distribution of OD-pair sizes is very skewed: only a few OD-pairs are very large and cover a large part of the passengers. We expect that the timetable largely depends on the large OD-pairs, and that including the remaining OD-pairs would only lead to some minor changes to the timetable. To choose a subset of OD-pairs, we introduce a parameter $\lambda \in [0, 100]$. We then include the OD-pairs which are largest in passenger size such that in total at least $\lambda\%$ of the passengers is included. If λ is small enough, only a few OD-pairs are included, while a large part of passengers is taken into account. Note that when we combine the first two simplifications in our experiments, we include the OD-pairs which are largest in passenger size among the ones who have a direct travel option such that in total at least $\lambda\%$ of the passengers with a direct travel option are included.

The third simplification is to require in the model that passengers always take the first relevant train leaving from the station: in that case they are not allowed to wait for a later departure. Note that also in this simplification, the order of trains departing from a station is not fixed. The intuition is that for the majority of the passengers, waiting for a later train is in general not beneficial. Therefore, we expect this to be a simplification that does not sacrifice much in terms of quality of the solution, while still reducing the complexity of the model significantly. To implement this simplification, the first minimum in (4.1h) is taken over $v' \in \{v\}$ instead of $v' \in V^k$. Equivalently, we could set $z_{v,v',r}^k = 0$ if $v \ne v'$ in (A.7).

As a fourth possible simplification, we choose to include only direct routes and do not allow for transfers. To achieve this, one could set the penalty γ_t to a very large value, thus allowing transfers, but making them very expensive. We chose to reduce the sets $\mathcal{R}_{v'}^k$, such that it includes only direct routes. This implies that OD-pairs for who no direct travel option exists cannot be included.

5.2. Heuristic generation of a starting solution

In this section we describe a heuristic approach to solve the integer program for the SPOT problem. When trying to solve SPOT to optimality, the heuristically generated solution can be used as starting solution for the IP solver and in this way, help to speed up the solution procedure.

First of all, we take only a subset of the passengers into account when generating this heuristic solution. That is, in line with the second simplification mentioned in the previous section, we take as few OD-pairs as possible, such that $\lambda\%$ of the passengers are incorporated. Based on this subset of passengers, we compute a timetable by a heuristic procedure.

Table 2
Overview of the integrality of variables in the heuristic.

Step	Π	$\mathcal{P}$	$\mathcal{A}$	$\mathcal{X}$	$\mathcal{Z}$	Target gap (%)
1	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	0.0
2	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$ and $\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	1.0
3	$\mathbb{Z}$	$\mathbb{Z}$	$\mathbb{Z}$ and $\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	1.0
4	Fixed	Fixed	Fixed	$\mathbb{Z}$	$\mathbb{Z}$	0.0

In order to state our approach, we first group the variables of the integer programming model for SPOT into the following sets:

$$\Pi = \{\pi_v \; : \; v \in V\} \tag{5.1a}$$

$$\mathcal{P} = \left\{ p_{ij} \; : \; (i,j) \in A \right\} \tag{5.1b}$$

$$\mathcal{A} = \left\{ \alpha_{v,v'} \; : \; k \in \mathcal{OD}, \; v \in V^k, \; v' \in V^k \setminus \{v\} \right\} \tag{5.1c}$$

$$\mathcal{X} = \left\{ x_{v,d}^k \; : \; k \in \mathcal{OD}, \; v \in V^k, \; d = 1, 2, \ldots, T \right\} \tag{5.1d}$$

$$\mathcal{Z} = \left\{ z_{v,v',r}^k \; : \; k \in \mathcal{OD}, \; v \in V^k, \; v' \in V^k, \; r \in \mathcal{R}_{v'}^k \right\}. \tag{5.1e}$$

The first two sets contain variables that relate to the timetable itself. The variables in the set $\mathcal{A}$ are used to determine time differences between two events correctly. Finally, the sets $\mathcal{X}$ and $\mathcal{Z}$ are introduced in the linearization of our model, and are related to the passenger routing.

To generate a good starting solution, we consecutively solve partial relaxations of the SPOT model, as outlined below. Since for all steps we require that the variables in the sets Π and $\mathcal{P}$ are integer, the timetabling constraints (4.1b), (4.1c), (4.1l), and (4.1m) are fulfilled. Thus, in each step, we find a feasible timetable.

Note that as soon as a timetable is fixed, it is possible to evaluate it according to objective function (4.1a). To this end, for each OD-pair $k \in \mathcal{OD}$ we compute the lengths Y_v^k of perceived-travel-time-minimal routes from each relevant departure event $v \in V^k$ to the destination by solving a shortest path problem. Furthermore, we order the relevant departure events, and thus compute the time difference between π_v and the departure time of the previous departure, L_v^k, as well as the average waiting time for these passenger $W_v^k = L_v^k/2$. This allows us to compute the objective value of the timetable as specified in (4.1a).

We evaluate each timetable generated in the heuristic as described in (4.1a). The best solution according to this evaluation is stored as the incumbent and only replaced when a better solution is found.

The steps of the heuristic are detailed below. To give a quick overview, Table 2 displays for each step what type the variables are in that step, i.e., whether they are continuous ($\mathbb{R}$) or integer ($\mathbb{Z}$) or mixed.

Each step is solved with a time limit, which is based on the overall time limit *TL*. Furthermore, the solution for each step is used as a warm start for the next step. The heuristic is a variant on the 'Relax-and-Fix' heuristic, as explained in Belvaux et al. (1998); Wolsey (1998).

Step 1 In the first step, a solution is found that is feasible with respect to all timetabling constraints. Therefore, we relax all variables to continuous variables, except for the timetabling related variables, i.e., those in Π and $\mathcal{P}$. This model is solved to optimality, or until a time limit of $TL/10$ is reached. Note that this is a relatively easy problem, as no infrastructure constraints are considered in the strategic timetabling problem.

Step 2 In this step, we improve the time differences between trains to get a better passenger routing, by changing a subset $\mathcal{A}'$ of the variables in $\mathcal{A}$ into integers, which we initialize as $\mathcal{A}' = \emptyset$.

In order to determine this set, we check for each pair of trains t_1 and t_2 whether their geographical routes overlap. If so, let $v_1, v_2 \in \Pi$ be the departure events of trains t_1 and t_2, respectively, at their first shared station. Then $\mathcal{A}' = \mathcal{A}' \cup \{\alpha_{v_1,v_2}, \alpha_{v_2,v_1}\}$.

We change all variables in $\mathcal{A}'$ into integers and set $\mathcal{A} = \mathcal{A} \setminus \mathcal{A}'$ and $\mathcal{A}' = \emptyset$. Then we re-optimize with a time limit of $TL/10$ or until an optimality gap of 1.0% is reached.

Step 3 In the previous step, a part of the α variables is changed into integers, but the majority is still continuous. In this step we iteratively change the remaining variables in $\mathcal{A}$ into integers, starting with the variables of which the value in the incumbent solution is not close to integer.

In each iteration of this step, we define the set of variables that are to be changed to integers as

$$\mathcal{A}' = \{\alpha \in \mathcal{A} \mid 0.05 \leq \text{val}(\alpha) \leq 0.95\}. \tag{5.2}$$

Here, val(α) denotes the value this variable α attains in the incumbent solution.

Again, we change the nature of all variables in $\mathcal{A}'$ to integers, we set $\mathcal{A} = \mathcal{A} \setminus \mathcal{A}'$, $\mathcal{A}' = \emptyset$ and re-optimize with a time limit of $TL/10$ or until an optimality gap of 1.0% is reached. This is continued until $|\mathcal{A}'| \leq 50$, in which case we set $\mathcal{A}' = \mathcal{A}$ in order to limit the number of iterations. Furthermore, this ensures that after these loops all α-variables have integer values.

Step 4 In this step, we fix all variables in Π, $\mathcal{P}$ and $\mathcal{A}$ to the value they attain in the incumbent solution (according to the evaluation with all OD-pairs). Next, we change all variables in $\mathcal{X}$ and $\mathcal{Z}$ to integers and reoptimize this model to optimality.

In order to better understand the heuristic, we highlight the rationale behind it. As headway constraints are not considered, the timetabling part is relatively easy in our model. Therefore we first find a timetable that is feasible with respect to the timetabling constraints, and include the passenger routing part only as a continuous relaxation. As this is a relatively easy task, we try to find an optimal solution here. This can however lead to a bad timetable, as the time differences between events can be determined incorrectly, due to the continuous nature of the variables in $\mathcal{A}$. Therefore, in the next steps we try to improve this.

First, we determine where train lines meet for the first time. By making the corresponding variables integer, we aim at better spreading different train lines over time. The expectation here is that by changing only a few variables to integers, a good gain in terms of quality can be obtained, without making it too difficult. The places where train lines meet are these places where frequencies on the tracks can change and therefore the expectation is that these are crucial decisions. The next step turns the remaining variables into integers. By experiments we found that the values of the majority of the variables in $\mathcal{A}$ is very close to integer, and that the remaining variables are generally rather close to 0.5. Therefore we select these variables that have $0.05 \leq \text{val}(\alpha) \leq 0.95$, i.e., that are not too close to integer values. By selecting these variables, we expect to make a large step towards an overall feasible solution. Iteratively these variables are changed to integers. When there are not many variables left, we change the remaining variables into integers in order to limit the number of iterations needed. Finally, for the best found timetable, we determine the best routing and the heuristic finishes.

6. Computational results

In this section, we apply our approach to two instances based on the network operated by Netherlands Railways which we describe in Section 6.1. We use these instances to computationally evaluate the use of a heuristic starting solution in Section 6.2.1 and to investigate the effect of solving restricted versions of our model in Section 6.2.2. In the case studies in Section 6.3 we look more in detail into the solutions created by the SPOT model, discuss our findings, giving some insights on how our approach can be used in strategic railway planning.

In all experiments we discretize time to minutes and use a period length of one hour, i.e., $T = 60$.

For the perceived travel time, values for adaption time and transfer penalty have to be set (γ_w and γ_t). According to De Keizer et al. (2015), the resistance for a transfer depends on many factors, but on average a penalty of 23 min (including 2 min of transfer time) is appropriate. We use a minimum of 3 min for a transfer, so we chose to use a value of $\gamma_t = 20$ in our models. We want to put emphasis on the regularity of trains to reduce adaption time, but not over-stress it because it already appears in the objective as a quadratic term, so we use $\gamma_w = 3$.

In our implementation and when reporting objective values in this section, we only report the 'additional time'. That means, we subtract constant terms from the objective function to improve numerical stability. For trip time, that implies that we subtract the minimal duration of the shortest possible trip for that OD-pair. If some OD-pair needs at least one transfer, we subtract a penalty, and only penalize additional transfers. Finally, for the constant for the adaption time, we assume that departure events are spread evenly over time, which leads to the lowest possible adaption time (see Appendix C for a proof), and subtract the corresponding value for the adaption time. This also explains why we do not report relative gaps. If we do not subtract the constant terms, all gaps would be relatively small. In our experiments, the lower bound is often close to zero and hence relative gaps are very large.

Our computations are carried out on a machine with an Intel Xeon Silver 4110 2.10 GHz processor and with 96 GB of RAM installed. The integer programs are solved by Cplex 12.9.0 under default settings, using up to 15 parallel threads.

6.1. Instances

The instances that we use in this study are real-life instances of Netherlands Railways (NS), the largest operator of passenger trains in the Netherlands. The first instance is the so-called 'A2-corridor', a network that contains 5 Intercity lines, that all share part of their route. The second instance is the 2019 Intercity network of Netherlands Railways (NS). In the remainder of this section, we describe the two instances in more detail.

6.1.1. A2 corridor

The first instance we consider in our study contains the so called 'A2-corridor', which is the part of the Dutch railway network between Eindhoven (Ehv) and Amsterdam Centraal (Asd). The line plan for this instance is shown in Fig. 2a. The used abbreviations for the stations are mentioned mentioned in Table 3.
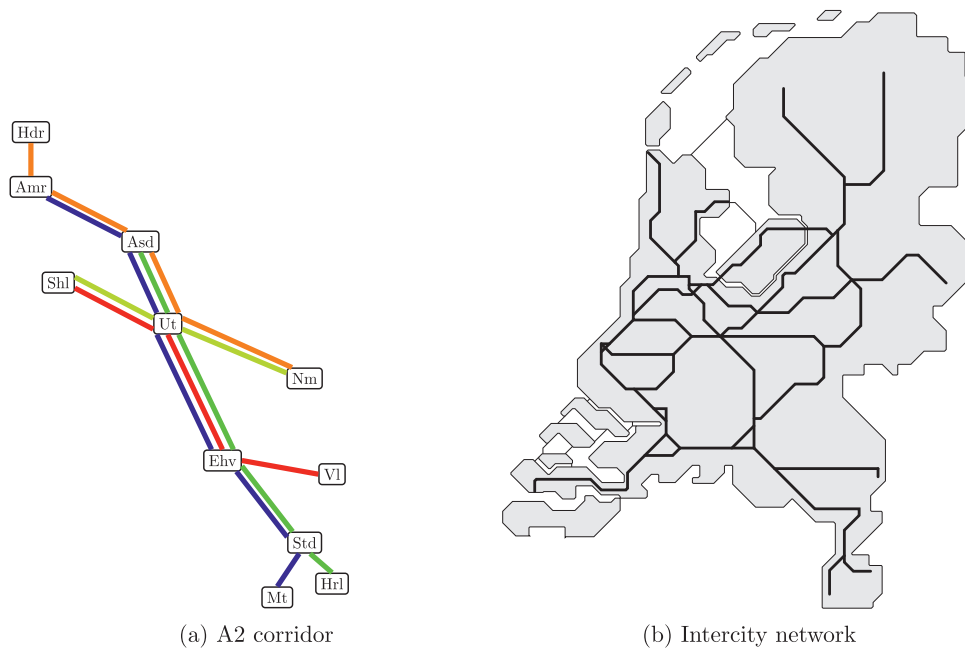
(a) A2 corridor   (b) Intercity network

Fig. 2. Overview of the used instances.

Table 3
Abbreviations of the stations.

Abbreviation	Name	Abbreviation	Name
Ah	Arnhem	Mt	Maastricht
Amr	Alkmaar	Nm	Nijmegen
Asb	Amsterdam Bijlmer ArenA	Sgn	Schagen
Asd	Amsterdam Centraal	Shl	Schiphol
Ehv	Eindhoven	Std	Sittard
Hdr	Den Helder	Ut	Utrecht Centraal
Hrl	Heerlen	Vl	Venlo
Ht	's Hertogenbosch		

The instance consists of 5 train lines, each with a frequency of 2 trains per hour in both directions, leading to 20 trains in total. The blue and green lines serve the whole corridor, i.e., between Ehv and Asd, whereas the orange line only serves Asd - Ut, and the red line only serves Ut - Ehv. The line depicted in lime color does not serve the corridor itself, but it is important in this instance to determine good frequencies on the remainder of the network that is not part of the corridor itself.

The reason to study this instance is that the 'A2-corridor' has very high passenger numbers and it has been subject to intensive study in practice recently, since Intercity-frequencies increased from four to six trains per hour here. In Asd and Ehv, four of the six trains continue to Amr and Std, respectively. This raises the question what the headway times should be between consecutive trains, both on the corridor itself and on the remainder of the network. As an example, if the headway times between all consecutive trains on the corridor is 10 min upon arrival in Ehv, and trains do not get additional dwell time there, the pattern between Ehv and Std will be irregular, headway times alternate between 10 and 20 min. In order to get a more regular pattern, trains would have to get a longer dwell time in Ehv. We study these kind of situations to find out what is the best solution from a perspective of total perceived passenger travel time.

In this instance, we consider only OD-pairs that travel either in the southbound or the northbound direction, and not in both directions. For example, OD-pairs Nm to Ehv and vice versa are not considered, as they would have to travel via Ut. In total the network has 34 relevant stations and we consider 891 OD-pairs. The average number of routes per OD-pair is 6.3, with a maximum of 24 routes. The Event-Activity Network contains 1344 events and 1700 activities, of which 376 are transfer activities.

6.1.2. Intercity network

As second instance, we consider the 2019 Intercity Network of NS. In this network, there are many OD-pairs without a direct connection. We thus expect that arrival and departure times at important transfer stations will be influenced by the need to make good transfer connections for these passengers. The network includes 95 stations and 76 trains in total. The geographical network is depicted in Fig. 2b. There are 8870 OD-pairs.

Table 4
Details about the heuristic procedure.

	A2-corridor	Intercity network only direct	Intercity network all OD pairs
# OD-pairs	17 (1.9%)	55 (2.8%)	70 (0.8%)
Time (s):			
Step 1	1	14	108
Step 2	4	342	978
Step 3-1	19	3613	3616
Step 3-2	55	3620	3611
Step 3-3	–	3606	2882
Step 4	0	1	1
Total:	79	11,195	11,195

The corresponding event-activity network contains 3816 events and 6578 activities, 2442 of which are transfer activities. On average, each OD-pair has 11.8 travel options, with a maximum of 280 possible options.

It is interesting to observe that when we consider passengers with a direct travel option only (compare Section 6.2.2), only 1960 OD-pairs remain, but these cover 79.1% of the passengers. In that case, 1760 of the transfer activities are not needed, which makes the model simpler to solve. But even in this case, 682 transfer activities are kept, since it may be beneficial for the passengers to transfer, even if there is a direct connection available.

6.2. Evaluation of the solution approach

In this section, we evaluate our solution approach. In Section 6.2.1 we evaluate computationally the benefit of generating a heuristic starting solution instead of a cold start. In Section 6.2.2 we explore the effects of, next to using the heuristic, solving several restricted versions of the SPOT model on the quality of the timetable.

6.2.1. The benefit of generating a starting solution

To evaluate the benefits of using a starting solution, we compare running times of the linearized SPOT model, with and without starting solution. We do this on three different cases: the A2-corridor instances, the intercity network instance with OD-pairs which have a direct travel option, and the intercity network instance with all OD-pairs. In this experiment, we do not use any of the reductions described in Section 5.1.

For generating a starting solution, we set $\lambda = 30$, i.e., at least 30% of the passengers in the network are included. Given the distribution of the OD-pair sizes (only a few OD-pairs account for a large portion of the passengers) and after performing several tests, this turned out to be a reasonable number to use for this purpose. This leads to including only a small subset of the OD-pairs in the model, while still ensuring that a large portion of the passengers is covered. We then employ the heuristic procedure described in Section 5.2.
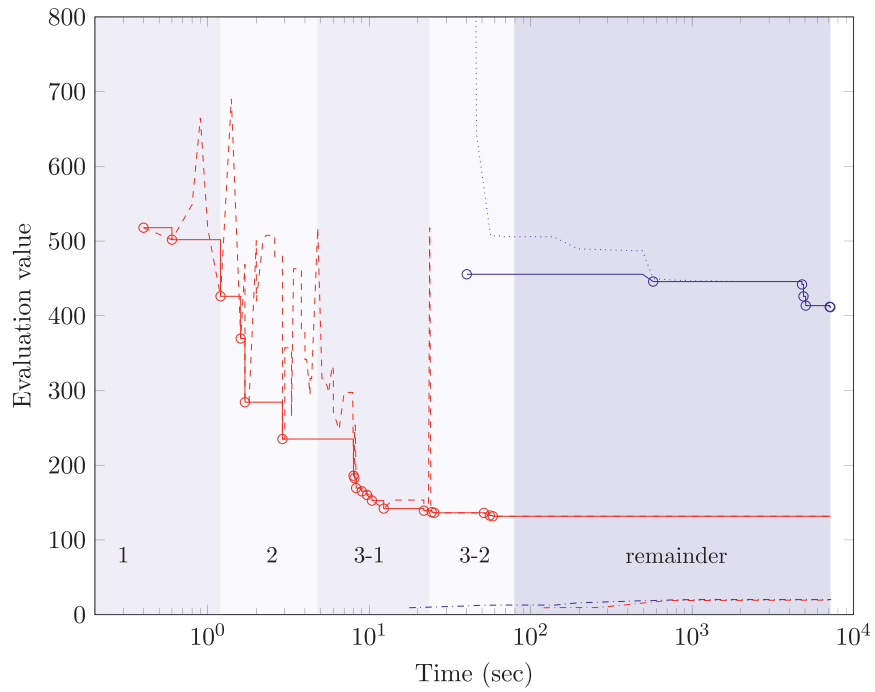
To guide the search when no heuristic starting solution is generated, we first spend 20% of the allotted time on a model where all dwell times are set to their lower bound. The remaining 80% of the time is spent on solving the full model.

For the A2-corridor, we set a total time limit of 2 h for the computations. For the Intercity case, we set a total time limit of 10 h.
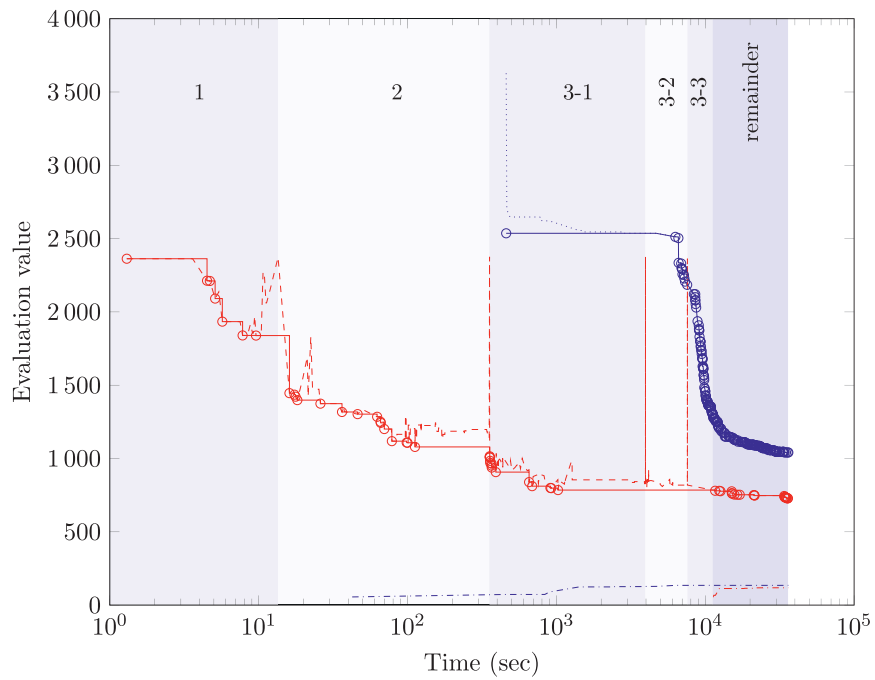
The results of our computations are shown by means of convergence plots in Fig. 3. The horizontal axes display time in seconds on a logarithmic scale. Note that the heuristic solves a strongly restricted problem with a subset of the passengers, and therefore the objective values of the heuristic and the objective value of the full model cannot be compared. Therefore, every time a new timetable is found in the solving process, its objective value (4.1a) is evaluated based on the full set of OD-pairs, in the way described in Section 5.2. Fig. 3 plots these evaluation values over time.

In the convergence plots, each individual plot displays six lines. The dashed red line corresponds to the evaluation value of the timetables found by the heuristic. As this model has incomplete information and is a partial relaxation, not every timetable is necessarily an improvement over the previously found timetable and therefore the dashed red line is not monotonically decreasing. In fact, every time the solver restarts in a new step of the heuristic, it may find an arbitrarily bad timetable, thus causing the peaks in the dashed red line. The evaluation value of the best timetable found so far is shown by the solid red line. The points in time in which a better timetable (according to the evaluation value based on all OD-pairs) is found is indicated by a circle. Once the heuristic finishes, it is fed to the linearized SPOT model as a starting solution. From that point on, a lower bound is available, which is shown as a dash-dotted red line.

To indicate how much time is consumed for the different steps of the heuristic, we indicate the time taken by the different steps in Fig. 3 by means of shaded bars. Each bar displays the step of the heuristic as well. As the third step iteratively changes a subset of the α variables, we also display the iteration number, i.e., 3-2 denotes step 3, iteration 2. Only a few iterations are needed to turn all α variables into integers. Step 4 is not displayed because this interval is too short to be visible on the logarithmic time scale we use. More details on how much time is spent on each step can be found in Table 4. This table also displays the number of OD-pairs used for the heuristic, and its percentage of the total number of OD-pairs in the instance.
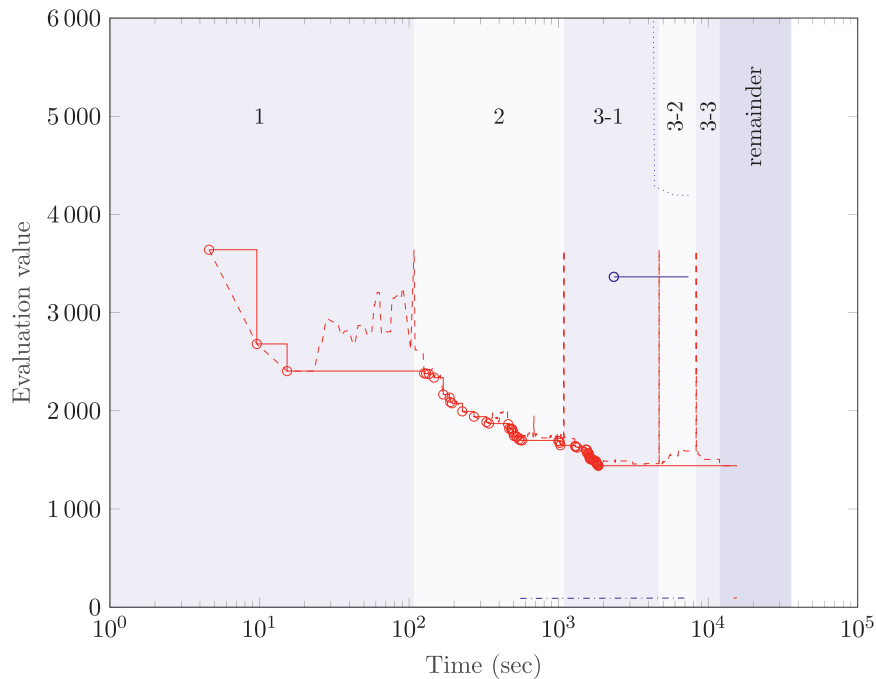
(a) A2-corridor

(b) IC network, only direct passengers

Fig. 3. Convergence plots for the comparison of solution approaches. Red lines display the heuristic approach, blue the standard approach. Solid line: best solution found so far. Dash-dotted line: lower bound. Dashed red line: Current objective value. Dashed blue line: Objective value according to solver. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(c) IC network with all passengers

Fig. 3. Continued

Table 5
Comparison between using a heuristic starting solution or not.

	With heuristic	No heuristic
Instance:	A2-corridor	
Heuristic objective value:	131.6	
Final objective value	131.6	411.5
Lower bound	19.4	19.1
Instance:	IC network - direct passengers	
Heuristic objective value:	784.6	
Final objective value	728.3	1041.2
Lower bound	118.4	123.1
Instance:	IC network - all passengers	
Heuristic objective value:	1440.6	
Final objective value	1440.6	3364.1
Lower bound	98.5	98.9

For the solution process without starting solution, the dotted blue line displays the objective value according to the solver. Note that here we do not use a relaxation nor restrict to a subset of the passengers. However, it still is possible that in intermediate solutions, the chosen passenger routes are suboptimal for the chosen timetable. The solid blue lines shows the evaluation value of the timetable found. We observe that in the beginning, there can be quite a difference between objective value reported by the solved and evaluation value, but soon better routes are found and the objective value reported by the solver decreases. The blue dash-dotted line displays the lower bound according to the solver.

Characteristics of the solutions and solving process are reported in Table 5. For each instance, the left column displays the objective values after the time limit has passed, for the case where a heuristic starting solution is used. It shows the value of the heuristic solution, the value of the final solution, and a lower bound (provided by CPLEX). The right columns displays the values for solving the model without warm-start, and shows the final objective value and lower bound.

We see in Fig. 3 and Table 5 that although neither the full model, nor the solution approach applying the heuristic could find an optimal solution, in all three cases, better solutions are found when generating a heuristic starting solution first. Even stronger, the heuristic finds a good solution, before the full model finds the first feasible solution.

6.2.2. Solving reduced versions of SPOT

In this section we experiment with solving different reduced versions of SPOT in the second stage. The rationale behind this is that on the one hand, within the same time limit, we may be able to get closer to optimality when working on a
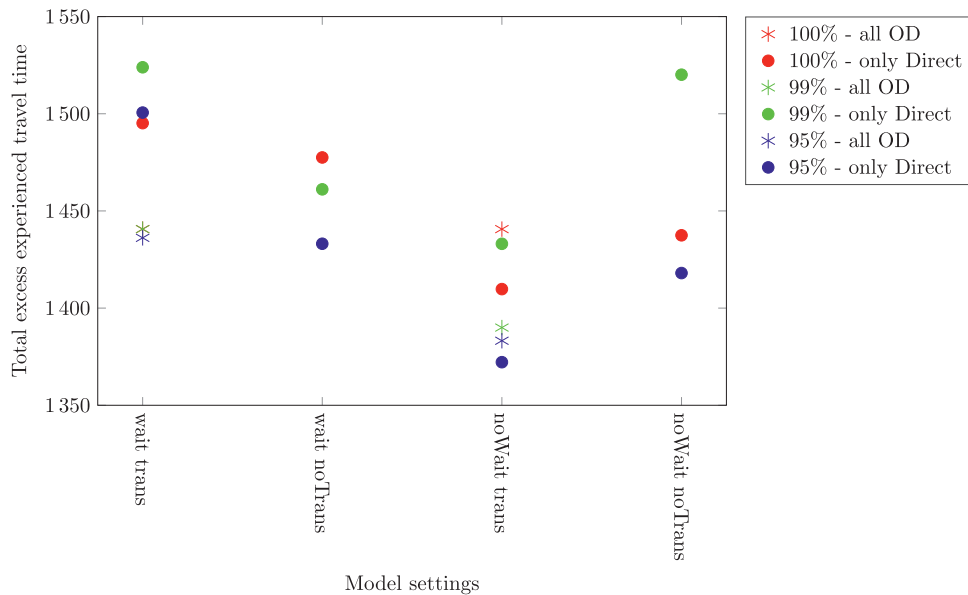
Fig. 4. Evaluation values for Intercity network for different parameter settings.

reduced problem version. On the other hand, we hope that when reducing in the 'right' way, little relevant information is lost, such that the timetable we find is good when evaluated for the full problem.

To reduce the model, we experiment with four different parameters: (1) We take only passenger who have a direct travel option (only Direct), or all passengers (all OD); (2) We take $\lambda \in \{95, 99, 100\}$, i.e., we restrict the number of passengers that we take into account; (3) We either allow passengers to transfer (trans) or not (noTrans); (4) We either force passengers to take the first departing train (noWait), or let them wait for a later train (wait). When we combine (1) and (2), we include the OD-pairs which are largest in passenger size among the ones who have a direct travel option such that in total at least λ% of the passengers with a direct travel option are included. See also Section 5.1 for detailed explanations of the simplifications.

We follow the two-stage approach motivated in the previous section: generate a heuristic start solution, then use this to warm-start the reduced linearized SPOT model. Note that take the above parameters already into account when constructing the heuristic starting solution. E.g., when we do not consider OD-pairs with transfer options (only Direct), only these are considered in the heuristic. Furthermore, when we solve the reduced linearized SPOT model with a subset of the passengers, e.g., only 95% of the passengers, the heuristic uses a subset of this subset, that is, 30% of this 95% of the passengers.

In the following we compare the evaluation values for our approach under different parameter setting on the Intercity network instance. We chose this instance because this is the most difficult instance to find good solutions for. Note that none of the approaches finds a provably optimal solution within the given time limit.

In order to properly compare the resulting timetables, we have evaluated each of them considering all passengers with full route choice. The corresponding evaluation values are displayed in Fig. 4.

Note that when including all OD pairs, but not allowing transfers, this leads to many OD-pairs not having a transfer option in the model. Therefore, we leave out these situations.

As can be seen in this figure, no single parameter choice seems to lead to clearly superior results. As a tendency, in this instance it appears that the combination of 'noWait' and 'Trans', i.e., forcing passengers to take the first train that is leaving, and allowing transfers, leads to lower objective values. A possible explanation for this is that the cases in which waiting at the origin station is beneficial will be very limited, especially since we are looking at an Intercity network in which trains run at the same speed. The option 'noWait' hence leads to a smaller model without sacrificing much in terms of quality. Next to this, transfer options can significantly improve route choices and providing them seems to be relevant.

Overall, we conclude that reducing the set of OD-pairs or the travel options on the Intercity network in the proposed way is possible without leading the solution procedure towards *bad* solutions. Since these reductions allow speed-ups in the solution procedure, working on reduced models can therefore even lead to slightly better solutions if the solution time is limited.

6.3. Case studies

In this section, we look into the solutions that our approach finds in more detail. In Section 6.3.1 we focus on the A2-corridor and demonstrate the trade-offs that our model is able to make, between additional dwell times and regularity of train service. In Section 6.3.2 we demonstrate how our method makes choices regarding transfer connections between trains
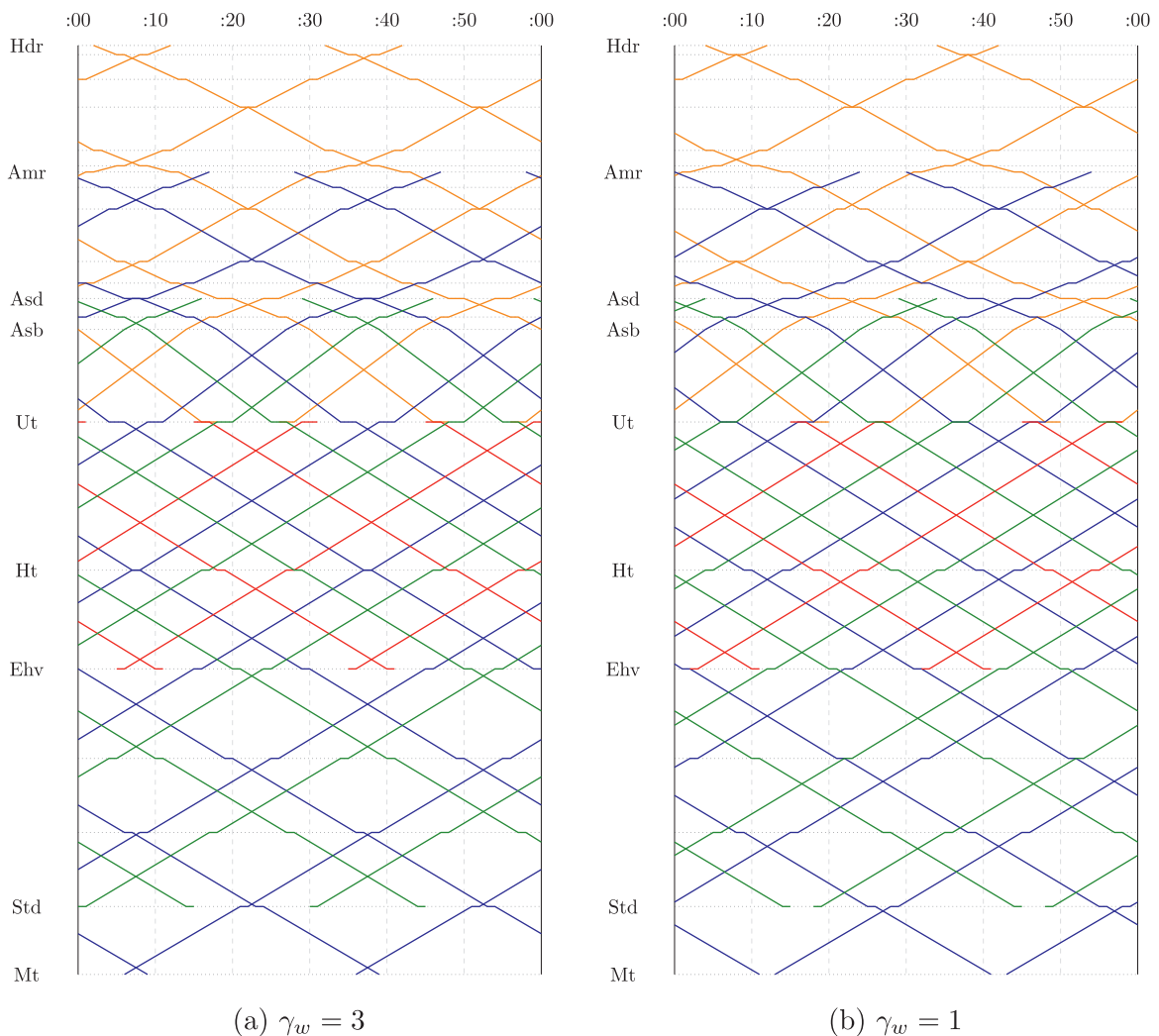
(a) $\gamma_w = 3$ (b) $\gamma_w = 1$

Fig. 5. Time space diagrams for Hdr to Mt.

on the intercity network. Note that the solutions discussed in this section are the best found solutions within the time limit of the approach, and thus not necessarily optimal. In this section, we again use the two-stage approach by employing our heuristic algorithm first, and using the result as a warm start for solving the full linearized SPOT model. For the second stage we do not use any of the reductions mentioned in Sections 5.1 and 6.2.2.

6.3.1. Balancing regularity and dwell times

We illustrate the outcome of our approach on the A2-corridor instance in more detail. Specifically, we focus on what happens at the two locations Asd and Ehv where frequencies change from six to four trains per hour. Remember that, when passing from the corridor with frequency six to the part of the network with frequency four, if the trains arrive with 10 min headway time between each pair of consecutive trains and no additional dwell time at the station is allowed, the headway times will be irregular outside the corridor, alternating between 10 and 20 min. If we want the patterns to be regular both on and off the corridor, additional dwell times are required at Asd and Ehv, in order to make the transition between the different frequencies. In order to shed light on the trade-offs at different values of adaption time, we visualize two timetables for the A2-corridor. To find the first one, we ran our solution approach for different values of parameter γ_w, which relates the perceived duration of adaption time to in-train time. In the first situation $\gamma_w = 3$, thus adaption time is considered to be less pleasant than being in the train itself. This will put a higher emphasis on the regularity of trains. In the second situation $\gamma_w = 1$, thus adaption time is valued equal to in-train time.

Time space diagrams of the timetables for both situations are shown in Fig. 5. Time is shown on the horizontal axis, distance on the vertical axis, where also the relevant stations are shown. Each train line is plotted with a different color. More detailed timetables for the stations Asd and Ehv are shown in Tables 6 and 7. Interesting to note is that the arrival

Table 6
Timetable for northbound trains at Asd.

From	To	$\gamma_w = 3$		$\gamma_w = 1$	
		Arrival	Departure	Arrival	Departure
Mt	Amr	:07	:09	:14	:16
Hrl	Asd	:16	–	:04	–
Nm	Hdr	:24	:26	:24	:26
Mt	Amr	:37	:39	:44	:46
Hrl	Asd	:46	–	:34	–
Nm	Hdr	:54	:56	:54	:56

Table 7
Timetable for southbound trains at Ehv.

From	To	$\gamma_w = 3$		$\gamma_w = 1$	
		Arrival	Departure	Arrival	Departure
Shl	Vl	:10	–	:10	–
Asd	Hrl	:20	:21	:20	:21
Amr	Mt	:29	:30	:31	:32
Shl	Vl	:40	–	:40	–
Asd	Hrl	:50	:51	:50	:51
Amr	Mt	:59	:00	:01	:02

Table 8
Timetable for southbound trains at Ledn.

From	Arrive	To	Depart
Shl	:00	Gvc	:01
Hlm	:02	Rtd	:03
Hlm	:16	Gvc	:17
Shl	:18	Rtd	:19
Shl	:30	Gvc	:31
Hlm	:32	Rtd	:33
Hlm	:46	Gvc	:47
Shl	:48	Rtd	:49

pattern in Asd with $\gamma_w = 1$ is perfectly regular, all headway times between consecutive trains are exactly 10 min. However, when continuing towards Amr, the pattern becomes very irregular, as there is no additional dwell time added, and the headway times now alternate between 10 and 20 min. Also in Fig. 5b, this irregularity is clearly visible. With $\gamma_w = 3$, the irregularity north of Asd is reduced to headway times of 13 and 17 min. In this case the headway times on the corridor are no longer equal and vary between 8 and 13 min.

The arrival headway times at Ehv are fairly regular, they vary between 9 and 11 min, for both values of adaption time. Also here, departure headway times are not perfectly regular. Instead, for $\gamma_w = 3$, they are even as large as 21 min, which is larger than the departure headway times at Asd. This clearly shows that in this case, the emphasis lies on reducing waiting times for passengers already in the train, and not on reducing adaption time for boarding passengers at Asd and subsequent stations.

Interesting to note in Fig. 5, is what happens around Ut, for both values of γ_w. The red line comes from Ehv, and goes towards Shl, it hence leaves the corridor at Ut. At Ut, the orange line, coming from Nm, enters the corridor to guarantee the six train per hour frequency again. Interestingly, in both timetables the orange line departs at Ut around the same time that the red line enters, i.e., in the regular pattern that is present in the corridor, these two lines replace each other.

This instance clearly shows that trade-offs are made between regularity on the corridor and on the branches on the one hand, and additional dwell times at border stations at the other hand. It shows that the value of regularity and the value of short waiting and transfer times depend on the relative value of adaption time versus in-train time γ_w.

6.3.2. Insights on the intercity network

The Intercity network of Netherlands Railways contains many train lines that are linked to each other, because they share part of their route and because they can offer good connections. We have analysed the timetable that was evaluated best in Section 6.2.2 (that is, the one found for parameter setting 'only Direct', $\lambda = 95\%$, 'trans', 'noWait') to demonstrate how our method can be used to generate insights on the desired timetable. For the purpose of this paper, we analyse the timetable at two different stations, namely Leiden (Ledn) and Zwolle (Zl).
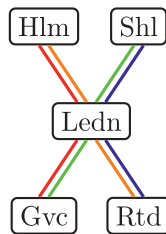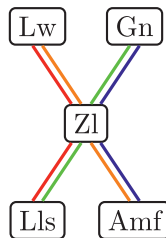
Fig. 6. Network at Ledn.

Fig. 7. Network at Zl.

Table 9
Timetable for southbound trains at Zl.

From	Arrive	To	Depart
Gn	:28	Lls	:29
Lw	:30	Amf	:32
Gn	:57	Amf	:02
Lw	:59	Lls	:00

Leiden (Ledn) This station is located in the western part of the Netherlands and can be reached from four different directions, which are Haarlem (Hlm), Schiphol (Shl), The Hague (Gvc) and Rotterdam (Rtd). Fig. 6 displays a schematic overview of the line plan around Ledn. Each of the neighbouring stations is displayed, and a coloured line indicates a train line serving this station twice per hour per direction. As can be seen, there are direct trains between each of the northern stations and each of the southern stations. The timetable of the southbound trains is detailed in Table 8.

Two interesting things are worth noting here. First of all, the trains depart from Ledn two min apart from each other, and in a fixed order: the trains towards Gvc depart first. Furthermore, all dwell times at Ledn are exactly 1 min, i.e., no dwell times is prolonged. Therefore, the trains need to arrive in Ledn following an alternating pattern. They arrive in groups of two, but in subsequent groups, the order is reversed: In the first group, first the train from Shl arrives, and after that the train from Hlm, while for the second group this order is reversed. However, passengers from the first train always have a transfer connection to the second train of exactly 3 min. That means, in the first group of trains, passengers from Shl can travel quickly to Rtd and Gvc, while passengers from Hlm towards Gvc have to wait much longer. In the next group of trains, this order is reversed, now passengers from Shl to Gvc have to wait.

Secondly, the headway times between trains on a leg of this network are never exactly 15 min. For the trains coming from Hlm and for the trains towards Gvc and Rtd, it varies between 14 and 16 min. For the trains coming from Shl, headway times are even as large as 18 min. This illustrates that regularity is not a necessary condition for optimality.

Zwolle (Zl) Even though this station is situated in the less-populated northern part of the country, it is a crucial station in the network as all trains going further north have to pass Zl. Secondly, trains come and go in four main directions, Groningen (Gn), Leeuwarden (Lw), Amersfoort (Amf) and Lelystad (Lls). The relevant directions in this Intercity network are displayed in Fig. 7. It is a similar situation to the Leiden station. However, in Zl, the frequency of the lines is only one train per hour. In Table 9, the arrival and departure times of the trains in the southbound direction are shown.

The trains arrive in Zl in two groups, approximately at :00, and approximately at :30, with the train from Gn arriving earlier than the train from Lw in both cases. The train from Gn to Lls arrives in Zl at :28. Passengers can either stay in their train and continue towards Lls, or transfer to the train to Amf, leaving at :32. In contrast, the passengers in the train from Lw to Amf do not have a connection to Lls and should take the direct train half an hour later. However, half an hour later, the dwell time of the train coming from Gn is prolonged, so now passengers can actually transfer between both trains. Note that, so far, we have seen few cases in which dwell time was prolonged. However, in this case, a slightly prolonged dwell time of the train in Gn to Amf can greatly reduce the travel time for passengers fron Lw to Amf, as they would have to wait

approximately 30 min for their transfer connection otherwise. In the timetable at Ledn, we have not seen these prolonged dwell times, as this would be less beneficial there due to the higher frequencies.

The above discussion indicates how SPOT can be used to design timetables, and give valuable insights in the strategic timetabling phase. We observe that in Ledn and Zl, the solution is often close to regular, but that exceptions from these patterns can improve the timetables in some cases. This implies that we may overlook good timetables, when imposing regularity constraints. Especially in Ledn, we see that these irregular patterns allow the alternating order of trains, and in turn the alternating connections between trains.

Secondly, we see that for some stations longer dwell times are good to ensure transfers, especially if the alternative for missing the transfer would be a long waiting time, when frequencies are low.

7. Conclusions and further research

In this paper we introduce the Strategic Passenger Oriented Timetabling (SPOT) problem. This problem aims at finding a timetable pattern which is optimal for passengers, explicitly including adaption time into the perceived passenger travel time. In our approach to solve the SPOT problem, we formulate a quadratic integer program. We linearize it and we propose and test an approach for solving it. We have shown in our case studies, how the solutions generated by the SPOT model can be used to learn about desirable patterns at key points of the network.

Due to the strategic nature of the problem at hand, we formulate the SPOT problem without including headway constraints, so that the underlying timetabling problem is relatively simple. However, the inclusion of adaption time in the model formulation leads to a quadratic objective, making the model harder to solve again. We achieve improvement with respect to the solution time by warm-starting the model with a heuristically achieved solution. Still, in none of our instances we were able to prove optimality of the solution found, with a lower bound far off the best solution found. It may be promising to investigate further solution methods, possibly working directly on the quadratic formulation of the program, or to investigate how better bounds can be determined.

As a strategic timetable is computed several years in advance, the passenger numbers used in the SPOT model are only rough estimates of future passenger demand. It would be interesting to study how sensitive the 'optimal' timetable is to these demand estimates. If (computationally) possible, it may be worthwhile to consider not only one, but several future demand scenarios, and to find a timetable that minimizes overall perceived travel time in expectation over these scenarios.

A second aspect related to demand (uncertainty) is the question whether the timetable will actually be *feasible* under different future demand scenarios, in the sense that it will provide enough capacity to transport all passengers. The capacity of a 'train' depends on the train composition (number and type of train units combined), which is determined in the rolling stock schedule in tactical or operational planning phase, when there are better demand estimates available. Still, it could be an valuable addition to our model to require that maximum train capacities may not be exceeded (over various demand scenarios), or that a feasible rolling stock schedule must exist.

Most timetabling models for the tactical planning phase do not include adaption time into the perceived travel time, thus implicitly assuming that passengers will fully adapt to the timetable and not suffer from inconveniently placed departure times. To overcome this questionable assumption, the SPOT model could also be applied to a setting that includes headway constraints in the PESP model. However, we expect that these will make the model even harder to solve. In addition, our current model formulation uses the assumption that passengers arrive uniformly distributed over the period for the definition of passenger groups and average waiting times. To include more detailed passenger information that may become available when entering the tactical or operational planning phase, different modeling approaches may be needed.

A different idea on how to move towards the tactical planning phase is to consider the timetable obtained with SPOT as an ideal timetable, and adjust it, where needed, to make it 'fit' with infrastructure requirements. We introduce one approach to do so in Polinder et al., 2020: a Lagrangian-based method is used to modify train departure and arrival times to make the timetable found by the SPOT model operationally feasible on the given infrastructure, while modifying departure and arrival times as little as possible. The obtained feasible timetable is then evaluated to compute the resulting average perceived passenger travel time. Feedback is sent to the Lagrangian-based heuristic to improve the obtained timetable from the passenger perspective, while still respecting infrastructure constraints.

Appendix A. Linearization

The SPOT model in (4.1) contains a quadratic objective and has several minimums in the constraints. In Section 4.3, constraints (4.1e) are linearized. In this section, we linearize the remainder of the model.

A1. Objective

The objective function (4.1a) is a quadratic function. Using that $W_v^k = L_v^k/2$, the objective can be written as

$$\text{Minimize} \quad \sum_{k \in \mathcal{OD}} \frac{d_k}{T} \sum_{v \in V^k} \frac{\gamma_w}{2} \left(L_v^k\right)^2 + L_v^k \cdot Y_v^k. \tag{A.1}$$

We linearize this expression by writing L_v^k as a sum of binary variables, defined as

$$x_{v,d}^k = \begin{cases} 1 & \text{if } L_v^k \geq d \\ 0 & \text{otherwise} \end{cases} \qquad \forall k \in \mathcal{OD}, v \in V^k, d \in \{1, \ldots, T\}. \tag{A.2}$$

For a stronger formulation, we can impose the additional restrictions that

$$x_{v,d}^k \leq x_{v,d-1}^k \qquad \forall k \in \mathcal{OD}, v \in V^k, d \in \{2, \ldots, T\}. \tag{A.3}$$

Using these new variables, we can write

$$L_v^k = \sum_{d=1}^T x_{v,d}^k, \qquad \left(L_v^k\right)^2 = \sum_{d=1}^T (2d - 1) \cdot x_{v,d}^k. \tag{A.4}$$

Substituting this in (A.1) results in a multiplication of binary variables $x_{v,d}^k$ by bounded variables Y_v^k. This can be resolved by introducing new variables $R_{v,d}^k = Y_v^k \cdot x_{v,d}^k$. The objective then becomes to minimize

$$\sum_{k \in \mathcal{OD}} \frac{d_k}{T} \sum_{v \in V^k} \sum_{d=1}^T \left[\frac{\gamma_w}{2} (2d - 1) \cdot x_{v,d}^k + R_{v,d}^k \right], \tag{A.5}$$

and additional restrictions have to be added to correctly determine the value for $R_{v,d}^k$:

$$R_{v,d}^k \leq u_v^k \cdot x_{v,d}^k \tag{A.6a}$$

$$R_{v,d}^k \geq l_v^k \cdot x_{v,d}^k \tag{A.6b}$$

$$R_{v,d}^k \leq Y_v^k - l_v^k \cdot \left(1 - x_{v,d}^k\right) \tag{A.6c}$$

$$R_{v,d}^k \geq Y_v^k - u_v^k \cdot \left(1 - x_{v,d}^k\right), \tag{A.6d}$$

where l_v^k and u_v^k are the lowest and highest possible values for Y_v^k respectively.

A2. Minimums

Constraints (4.1e) and (4.1h) both contain a minimum, which we can linearize. (4.1e) is already linearized in Section 4.3. Constraints (4.1h) are replaced by the following set of restrictions for every $k \in \mathcal{OD}$ and every $v \in V^k$:

$$Y_v^k \leq Y_r + \gamma_w \cdot Q_{v,v'} \qquad\qquad \forall v' \in V^k, r \in \mathcal{R}_{v'}^k \tag{A.7a}$$

$$Y_v^k \geq Y_r + \gamma_w \cdot Q_{v,v'} - M_v^k \cdot \left(1 - z_{v,v',r}^k\right) \qquad\qquad \forall v' \in V^k, r \in \mathcal{R}_{v'}^k \tag{A.7b}$$

$$\sum_{v' \in V^k} \sum_{r \in \mathcal{R}_{v'}^k} z_{v,v',r}^k = 1. \tag{A.7c}$$

We introduced new binary variables $z_{v,v',r}^k$, which correspond to the route that is chosen. That means, if $z_{v,v',r}^k = 1$, passengers wait from event v to v' (which can be the same), and take route r, starting at v'. For computational stability, the newly introduced constants M_v^k have to be chosen as small as possible, but still large enough to make the second set of constraints redundant if $z_{v,v',r} = 0$, i.e., we can take

$$M_v^k = \gamma_w \cdot T + \max_{r \in \mathcal{R}^k} \left\{\overline{Y}_r\right\} - \max_{r \in \mathcal{R}^k} \left\{\underline{Y}_r\right\}, \tag{A.8}$$

where $\overline{Y}_r, \underline{Y}_r$ denote the highest and lowest possible value for the variable Y_r respectively.

As we are minimizing the perceived passenger travel time, we can exclude the newly introduced constraints (A.6a), (A.6c) and (A.7a) in order to reduce the model size.

To summarize, in the linearization we take several steps. First of all, the objective (4.1a) is replaced by (A.5). Here, additional variables $x_{v,d}^k$ and $R_{v,d}^k$ are introduced, with additional restrictions (A.3), (A.6b) and (A.6d). Secondly, the minima are replaced by linear restrictions. First of all, for notational reasons we defined $Q_{v,v'}$ in (4.2). By using this, we replace (4.1e) by (4.3), and (4.1h) by (A.7b)–(A.7c).

Table 10

List of sets, constants, and variables for the (linearized) SPOT model.

Sets	
V	The set of events (indexed by v, i or j)
A	The set of activities (indexed by a or (i, j))
$\mathcal{OD}$	The set of all OD-pairs (indexed by k)
$\mathcal{R}$	The set of all routes (indexed by r)
$\mathcal{R}^k$	The set of routes for OD-pair k (indexed by r)
V^k	The set of relevant departure events for OD-pair K (indexed by v or v')
Constants	
T	The cycle period
d_k	The number of passengers for OD-pair $k \in \mathcal{OD}$
γ_w	The objective coefficient for adaption time
γ_t	The penalty for using a transfer
Variables	
π_v	The timing of event $v \in V$
p_{ij}	A modulo parameter used for the shift from one cycle period to another, for activity $(i, j) \in A$
y_{ij}	The duration of activity $(i, j) \in A$
Y_r	The duration of route $r \in \mathcal{R}$
L_v^k	The number of minutes before event v, in which no other departure event for OD-pair k takes place
W_v^k	The expected waiting time for passenger for OD-pair k, for who event v is the first departure event
Y_v^k	The perceived travel time for passengers of OD-pair k, from the timing of event v onwards
$\alpha_{v,v'}$	A binary variable ensuring the correct determination of the time difference between event v and v'
$Q_{v,v'}$	The time difference between event v and v'
$x_{v,d}^t$	A linearization variable, indicating whether $L_v^k \geq d$ or not
$z_{v,v',r}^k$	A linearization variable, indicating which route is chosen
$R_{v,d}^k$	A linearization variable, replacing $Y_v^k \cdot x_{v,d}^k$

Appendix B. List of symbols

This appendix summarizes all notation used in the paper in Table 10.

Appendix C. Proof of lower bound on adaption time

Lemma 1. *Let $k \in \mathcal{OD}$ be an OD-pair and V^k the set of relevant departure events, let T denote the period length. Then the adaption time contribution for OD-pair k, $\sum_{v \in V^k} \frac{d_k}{T} \cdot L_v^k \cdot W_v^k$ is bounded below by $d_k \cdot \frac{T}{2|V^k|}$, that is*

$$\sum_{v \in V^k} \frac{d_k}{T} \cdot L_v^k \cdot W_v^k \geq d_k \cdot \frac{T}{2|V^k|}.$$

Proof. Remember that $W_v^k = \frac{1}{2} L_v^k$ and $\sum_{v \in V} L_v^k = T$, thus the adaption time contribution of OD-pair k can never be smaller than

$$\min \frac{d_k}{T} \cdot \sum_{v \in V^k} \frac{1}{2} (L_v^k)^2 \tag{C.1}$$

$$\text{s.t.} \sum_{v \in V^k} L_v^k = T \tag{C.2}$$

$$L_v^k \geq 0 \qquad\qquad\qquad \forall v \in V^k. \tag{C.3}$$

Consider any feasible solution $L^k = (L_v^k)_{v \in V^k}$ to this optimization problem. We can write $L_v^k = \frac{T}{|V^k|} + \delta_v$ with $\sum_{v \in V^k} \delta_v = 0$ due to (C.2) and obtain

$$\frac{d_k}{T} \cdot \sum_{v \in V^k} \frac{1}{2} (L_v^k)^2 = \frac{1}{2} \frac{d_k}{T} \cdot \sum_{v \in V^k} \left(\frac{T}{|V^k|} + \delta_v \right)^2$$

$$= \frac{1}{2} \frac{d_k}{T} \cdot \sum_{v \in V^k} \left(\left(\frac{T}{|V^k|} \right)^2 + 2 \frac{T}{|V^k|} \delta_v + \delta_v^2 \right)$$

$$= \frac{1}{2} \frac{d_k}{T} \cdot \left[\left(\sum_{v \in V^k} \left(\frac{T}{|V^k|} \right)^2 \right) + \left(2 \frac{T}{|V^k|} \sum_{v \in V^k} \delta_v \right) + \left(\sum_{v \in V^k} \delta_v^2 \right) \right]$$

$$= \frac{1}{2} \frac{d_k}{T} \cdot \left[|V^k| \cdot \left(\frac{T}{|V^k|} \right)^2 + 2 \frac{T}{|V^k|} \underbrace{\sum_{v \in V^k} \delta_v}_{=0} + \underbrace{\sum_{v \in V^k} \delta_v^2}_{\geq 0} \right]$$

$$\geq \frac{1}{2} \frac{d_k}{T} \cdot |V^k| \cdot \left(\frac{T}{|V^k|} \right)^2 = d_k \cdot \frac{T}{2|V^k|}$$

□

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.trb.2021.02.006.

CRediT authorship contribution statement

Gert-Jaap Polinder: Investigation. **Marie Schmidt:** Supervision, Writing - review & editing. **Dennis Huisman:** Supervision, Writing - review & editing.

References

Barrena, E., Canca, D., Coelho, L.C., Laporte, G., 2014. Exact formulations and algorithm for the train timetabling problem with dynamic demand. Comput. Oper. Res. 44, 66–74. doi:10.1016/j.cor.2013.11.003.

Barrena, E., Canca, D., Coelho, L.C., Laporte, G., 2014. Single-line rail rapid transit timetabling under dynamic passenger demand. Transp. Res. Part B 70, 134–150. doi:10.1016/j.trb.2014.08.013.

Belvaux, G., Boissin, N., Sutter, A., Wolsey, L.A., 1998. Optimal placement of add /drop multiplexers static and dynamic models. Eur. J. Oper. Res. 108 (1), 26–35. doi:10.1016/S0377-2217(97)00021-0.

Beter & Meer, 2014. Voorstel voor de operationele uitwerking van de lange termijn spooragenda. https://www.rijksoverheid.nl/documenten/rapporten/2014/03/28/bijlage-1e-operationele-uitwerking-prorail-en-ns-beter-en-meer, Online; Accessed on 2020-01-06.

Borndörfer, R., Hoppmann, H., Karbstein, M., 2017. Passenger routing for periodic timetable optimization. Public Transp. 9 (1–2), 115–135. doi:10.1007/s12469-016-0132-0.

Burggraeve, S., Bull, S.H., Vansteenwegen, P., Lusby, R.M., 2017. Integrating robust timetabling in line plan optimization for railway systems. Transp. Res. Part C 77, 134–160.

Cacchiani, V., Toth, P., 2012. Nominal and robust train timetabling problems. Eur. J. Oper. Res. 219 (3), 727–737. doi:10.1016/j.ejor.2011.11.003.

Caimi, G., Kroon, L., Liebchen, C., 2017. Models for railway timetable optimization: applicability and applications in practice. J. Rail Transp. Plan. Manag. 6 (4), 285–312. doi:10.1016/j.jrtpm.2016.11.002.

De Keizer, B., Kouwenhoven, M., Hofker, F., 2015. New insights in resistance to interchange. Transportation Research Procedia 8, 72–79. doi:10.1016/j.trpro.2015.06.043. Current practices in transport: appraisal methods, policies and models - 42nd European Transport Conference Selected Proceedings.

Deutschland-TAKT, 2019. TAKT website. Accessed on 2019-12-30.

Gattermann, P., Großmann, P., Nachtigall, K., Schöbel, A., 2016. Integrating passengers' routes in periodic timetabling: a SAT approach. In: Goerigk, M., Werneck, R. (Eds.), 16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, pp. 3:1–3:15. doi:10.4230/OASIcs.ATMOS.2016.3.

Goerigk, M., Schöbel, A., 2013. Improving the modulo simplex algorithm for large-scale periodic timetabling. Comput. Oper. Res. 40 (5), 1363–1370. doi:10.1016/j.cor.2012.08.018.

Großmann, P., Hölldobler, S., Manthey, N., Nachtigall, K., Opitz, J., Steinke, P., 2012. Solving periodic event scheduling problems with SAT. In: Jiang, H., Ding, W., Ali, M., Wu, X. (Eds.), Advanced Research in Applied Artificial Intelligence. In: Lecture Notes in Computer Science, 7345. Springer Berlin Heidelberg, pp. 166–175.

Hartleb, J., Schmidt, M., 2019. Railway timetabling with integrated passenger distribution. ERIM report series research in management Erasmus Research Institute of Management.

Hartleb, J., Schmidt, M., Friedrich, M., Huisman, D., 2019. A good or a bad timetable: do different evaluation functions agree? ERIM report series research in management Erasmus Research Institute of Management.

Huisman, D., Kroon, L.G., Lentink, R.M., Vromans, M.J., 2005. Operations research in passenger railway transportation. Stat. Neerl. 59 (4), 467–497.

Kinder, M., 2008. Models for periodic timetabling. Technische Universität Berlin Diploma thesis.

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

Kümmling, M., Großmann, P., Nachtigall, K., Opitz, J., Weiß, R., 2015. A state-of-the-art realization of cyclic railway timetable computation. Public Transp. 7 (3), 281–293. doi:10.1007/s12469-015-0108-5.

Liebchen, C., 2008. The first optimized railway timetable in practice. Transp. Sci. 42 (4), 420–435. doi:10.1287/trsc.1080.0240.

Liebchen, C., Möhring, R., 2007. The modeling power of the periodic event scheduling problem: Railway timetables–and beyond. In: Geraets, F., Kroon, L., Schöbel, A., Wagner, D., Zaroliagis, C. (Eds.), Algorithmic Methods for Railway Optimization: International Dagstuhl Workshop, Dagstuhl Castle, Germany, June 20–25, 2004, 4th International Workshop, ATMOS 2004, Bergen, Norway, September 16–17, 2004, Revised Selected Papers. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–40. doi:10.1007/978-3-540-74247-0_1.

Liebchen, C., Peeters, L., 2009. Integral cycle bases for cyclic timetabling. Discrete Optim. 6 (1), 98–109. doi:10.1016/j.disopt.2008.09.003.

Lübbe, J., 2009. Passagierrouting und Taktfahrplanoptimierung. Technische Universität Berlin Diploma thesis.

Lübbecke, M., Puchert, C., Schiewe, P., Schöbel, A., 2018. Integrating line planning, timetabling and vehicle scheduling. integer programming formulation and analysis. In: Conference on Advanced Systems in Public Transport.

Lusby, R., Larsen, J., Bull, S., 2018. A survey on robustness in railway planning. Eur. J. Oper. Res. 266 (1), 1–15. doi:10.1016/j.ejor.2017.07.044.

Matos, G.P., Albino, L.M., Saldanha, R.L., Morgado, E.M., 2018. Solving periodic timetabling problems with SAT and machine learning. In: Conference on Advanced Systems in Public Transport.

Michaelis, M., Schöbel, A., 2009. Integrating line planning, timetabling, and vehicle scheduling: a customer-oriented heuristic. Public Transp. 1 (3), 211.

Nachtigall, K., 1999. Periodic network optimization and fixed interval timetables.. Universität Hildesheim Habilitation thesis.

Nachtigall, K., Opitz, J., 2008. Solving periodic timetable optimisation problems by modulo simplex calculations. In: Fischetti, M., Widmayer, P. (Eds.), 8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany doi:10.4230/OASIcs.ATMOS.2008.1588.

Odijk, M., 1996. A constraint generation algorithm for the construction of periodic railway timetables. Transp. Res. Part B 30 (6), 455–464. doi:10.1016/0191-2615(96)00005-7.

Pätzold, J., Schiewe, A., Schiewe, P., Schöbel, A., 2017. Look-ahead approaches for integrated planning in public transportation. In: D'Angelo, G., Dollevoet, T. (Eds.), 17th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2017). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, pp. 17:1–17:16. doi:10.4230/OASIcs.ATMOS.2017.17.

Pätzold, J., Schöbel, A., 2016. A matching approach for periodic timetabling. In: Goerigk, M., Werneck, R. (Eds.), 16th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2016). Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany, pp. 1:1–1:15. doi:10.4230/OASIcs.ATMOS.2016.1.

Peeters, L., 2003. Cyclic Railway Timetable Optimization. Erasmus University Rotterdam Ph.D. thesis.

Polinder, G.J., Kroon, L., Aardal, K., Schmidt, M., Molinaro, M., 2018. Resolving infeasibilities in railway timetabling instances. Available at https://ssrn.com/abstract=3106739.

Polinder, G.J., Cacchiani, V., Schmidt, M., Huisman, D., 2020. An Iterative Heuristic for Passenger-Centric Train Timetabling with Integrated Adaption Times. Technical Report. Erasmus Research Institute of Management.

Robenek, T., Azadeh, S.S., Maknoon, Y., Bierlaire, M., 2017. Hybrid cyclicity: combining the benefits of cyclic and non-cyclic timetables. Transp. Res. Part C 75, 228–253. doi:10.1016/j.trc.2016.12.015.

Robenek, T., Maknoon, Y., Azadeh, S.S., Chen, J., Bierlaire, M., 2016. Passenger centric train timetabling problem. Transp. Res. Part B 89, 107–126.

Schiewe, P., Schöbel, A., 2020. Periodic timetabling with integrated routing: toward applicable approaches. Transp. Sci. doi:10.1287/trsc.2019.0965.

Schmidt, M., Schöbel, A., 2015. Timetabling with passenger routing. OR Spectrum 37 (1), 75–97. doi:10.1007/s00291-014-0360-0.

Schöbel, A., 2017. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. Transp. Res. Part C 74, 348–365.

Schöbel, A., 2017. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. Transp. Res. Part C 74, 348–365. doi:10.1016/j.trc.2016.11.018.

Serafini, P., Ukovich, W., 1989. A mathematical model for periodic scheduling problems. SIAM J. Discrete Math. 2 (4), 550–581. doi:10.1137/0402049.

Siebert, M., 2011. Integration of routing and timetabling in public transportation. Georg-August-Universität Göttingen M.Sc. thesis.

Siebert, M., Goerigk, M., 2013. An experimental comparison of periodic timetabling models. Comput. Oper. Res. 40 (10), 2251–2259. doi:10.1016/j.cor.2013.04.002.

Wang, Y., Tang, T., Ning, B., van den Boom, T.J., Schutter, B.D., 2015. Passenger-demands-oriented train scheduling for an urban rail transit network. Transp. Res. Part C 60, 1–23. doi:10.1016/j.trc.2015.07.012.

Warmerdam, J., 2004. Specificaties TRANS toedeler. Technical Report. QQQ Delft.

Wolsey, L., 1998. Integer Programming. Wiley Series in Discrete Mathematics and Optimization. Wiley.

Yan, F., Goverde, R.M., 2019. Combined line planning and train timetabling for strongly heterogeneous railway lines with direct connections. Transp. Res. Part B 127, 20–46.

Yin, J., Yang, L., Tang, T., Gao, Z., Ran, B., 2017. Dynamic passenger demand oriented metro train scheduling with energy-efficiency and waiting time minimization: mixed-integer linear programming approaches. Transp. Res. Part B 97, 182–213. doi:10.1016/j.trb.2017.01.001.

Zhu, Y., Mao, B., Bai, Y., Chen, S., 2017. A bi-level model for single-line rail timetable design with consideration of demand and capacity. Transp. Res. Part C 85, 211–233. doi:10.1016/j.trc.2017.09.002.