Optimal Temporal Spreading of Alternative Trains in order to Minimise Passenger Travel Time in Practice

Peter Sels [a,b,c,1], Thijs Dewilde [a], Dirk Cattrysse [a], Pieter Vansteenwegen [a]
[a] KU Leuven, Leuven Mobility Research Centre, CIB,
Celestijnenlaan 300, 3001 Leuven, Belgium
[1] E-mail: sels.peter@gmail.com, Phone: +32 (0) 486 95 67 97
[b] Logically Yours BVBA, Plankenbergstraat 112 bus L7, 2100 Antwerp, Belgium
[c] Infrabel, Traffic Management & Services, Fonsnylaan 13, 1060 Brussels, Belgium

Abstract
In an earlier paper on timetabling, we derived a stochastic objective function for what we consider an ideal timetable: minimal expected passenger travel time in practice. This includes primary delay distributions and a degree of robustness against these via a knock-on delay model. This timetable is ideal for passengers who know the timetable in advance and plan their departure and/or arrival times accordingly. In general, it is considered as a good service to passengers that the different alternatives to get from an origin to a destination are equally spread in time. This is even more important for passengers who are not informed about the timetable or who cannot adapt their time of arrival in the station, for instance because they want to go home as soon as possible after a meeting. Spreading the trains decreases the average waiting time at departure and/or waiting time at arrival. In this paper, we add these terms to the objective function to also provide the advantage of low waiting times before and after the routes in the resulting timetable. We properly balance benefits of spreading the trains with the other objectives included in the objective function. The model solves quickly for a set of 26 trains and 12 sets of alternative routes but still proves to be difficult to solve for bigger instances.

Keywords
Optimal Cyclic Timetabling, Minimal Expected Passenger Time, Mixed Integer Linear Programming,

1 Introduction

It is generally understood, also by train passengers, that in a train service network with several alternative trains connecting an origin station with a destination station, spreading of these alternative trains in time is beneficial for train passengers travelling between these stations. Assuming, for a certain portion of all passengers, a random arrival pattern, it decreases the total waiting time at departure and symmetrically, also at arrival. The mathematical expression for these expected waiting times as a function of the planned heading times at departure (or arrival) between these alternative trains is well known and widely published and used in transport practice. In this paper, this expression is added to the objective function of our Periodic Event Scheduling Problem (PESP) based Mixed Integer Linear Programming (MILP) model (Sels et al., 2013b, 2014). When solving this model, the objective function will be minimised. When doing so, the expected waiting time at departure and arrival are minimised together with the other components that were already present in the objective function: expected ride time, dwell time, transfer time and knock-on time. The result is that we now minimise not only the total *journey time*, being in vehicle plus transfer time, but additionally also the *excess journey time*, being inter-departure plus inter-arrival waiting time (Zhao et al., 2013). Their sum, *actual or observed journey time* (Zhao et al., 2013), is now the new total objective function. This results in improved timetables, as the objective function approximates better the real expected time that passengers experience in practice. Indeed, *excess journey time* is a real expected time component for passengers and should not be forgotten.

Section 2 gives an overview of research into waiting time and alternative approaches to realising temporal spreading during timetable construction. We will conclude that our approach is a novel one. Section 3 formally derives the linear constraints and objective function terms that will be added to our MILP model later. In the process, some statistics on the typical number of alternative trains is given for the network of all Belgian passenger trains. Section 3.4 then replaces the constant departure (or arrival) times to variable ones. Indeed, in the PESP model begin (or end) times of activities are unknown still. As a consequence, some terms become quadratic and must be linearised. Section 4 discusses the successful application on a network consisting of 26 Belgian IC trains. We first limit the objective function to excess journey time without transfer time, then add transfers to the model, then add more OD-pairs for temporal spreading to the model and then attempt to tackle train service networks with more trains. The section shows, as is expected, that there is a noticeable competition between the minimisation of the different time components, especially between excess journey time and transfer time. Section 5 concludes and hints at further work that could potentially improve computation times.

2 Related Research Overview and Comparison

Research concerning the advantages of temporal spreading of alternative trains has been dominated by the analysis of how passengers choose their trains, when they choose to arrive at their station of departure and how this affects their expected inter-departure waiting time. The published results are described in section 2.1. Section 2.2 discusses which subcategories of passengers display random arrival behaviour. While the mentioned research is very interesting, it is very noticeable how inter-*arrival* waiting time has been left mostly unmentioned. We discuss this in section 2.3. Also, less attention has been paid to how this waiting time should be incorporated in models to construct timetables with the desirable property of temporal spreading. The traditional method of timetabling that intends to realise temporal spreading and a newer method are mentioned in section 2.4. Section 2.5 gives a quick overview of the rationale of our novel method.

2.1 Models of Expected Inter-Departure Waiting Time

The most simplistic model of passengers arriving in their station of departure is one where it is assumed that a constant number of passengers f arrives during every unit of time. This implicitly assumes that passengers are not adapting their arrival time to the knowledge of the train departure times as planned in the timetable. This model is called the *random arrival* model. This model has been assumed for studying transit reliability by de Pirey (1971), Barnett (1974), Bly and Jackson (1974) and Friedman (1976).

Welding (1957), Holroyd and Scraggs (1966) and Osuna and Newell (1972) derived that, for passengers who are arriving randomly at their station of departure, the expected waiting time until departure $E(w)$ can be expressed as a function of the *average vehicle planned heading time* $E(h)$ (over all heading times H_i in the cyclic timetable period T) and the *variation coefficient* $C_v(h)$ of the real time heading times:

$$E(w) = E(h)/2 \cdot \left(1 + C_v(h)^2\right). \tag{1}$$

Here, h is the heading time distribution and $E(h)$ is the expected heading time as it can be calculated from the planned timetable as

$$E(h) = \sum_{i=0}^{N-1} p_i \cdot H_i = \sum_{i=0}^{N-1} (H_i/T) \cdot H_i = \sum_{i=0}^{N-1} H_i^2/T, \tag{2}$$

where p_i is the probability of a passenger experiencing heading time H_i in period T. $C_v(h) = \sigma(h)/\mu(h)$ is the ratio of the standard deviation over the mean, both of the heading time distribution in real time. Substitution of the right hand side of equation (2) for $E(h)$ in equation (1) and multiplication by the number of considered randomly arriving passengers f delivers

$$E(f \cdot w) = \frac{f}{2T} \sum_{i=0}^{N-1} H_i^2 \cdot \left(1 + C_v(h)^2\right). \tag{3}$$

The general objective is to minimise the expected waiting time at arrival, $E(w)$, in equation (1). This can be done by either minimising $E(h)$ or $C_v(h)^2$ or a combination of both. Note that, for equal temporal spreading between N vehicles in time T, it holds that $E(h) = H = T/N$ and that otherwise $E(h) > H = T/N$. The second can be understood, from equation (2) and realising that the probability p_i equals H_i/T, so the probability to experience a higher heading time during time T is higher than the probability to experience a lower one. So, when supposing there are no real time delays, $E(h)$ is minimised by equal spreading of headways in the timetable if the amount of vehicles N per period T is fixed or otherwise by increasing the number of vehicles N. The value of $C_v(h)$ is decreased by lowering $\sigma(h)/\mu(h)$, representing the reduction of relative deviations from the timetable headways. This is done by better control of the timing of vehicles. Islam and Vandebona (2010) mention that $C_v(h)$ is often used as a measure for reliability. They study bus systems where *planned* heading times are all equal. In that case $E(h) = H$ and, assuming a fixed amount of vehicles, decreasing $E(w)$ can then only be obtained by reducing $C_v(h)$ by making the timing of the system more reliable. For a train system, often one must, like when having to insert a freight train between a set of N passenger trains, plan unequal heading times H_i. In that case, even in the planning $E(h) > H$ occurs. This summarises the effect of the number of alternative vehicles N, planned vehicle spreading in time $E(w)$ and reliability $C_v(h)$ on passenger waiting time for randomly arriving passengers. In this paper we will try to reduce $E(w)$ by reduction of $E(h)$, with fixed number of trains N, so by the use of optimal spreading only. Reducing $C_v(h)$ is useful but cannot be done via the timetable.

Later, studies appeared where the more complex behaviour of passengers, adapting their arrival time at their station of departure to their chosen vehicle planned departure time is studied (Okrent, 1971; Jolliffe and Hutchingson, 1975; Jackson, 1977; Turnquist, 1978).

Jolliffe and Hutchingson (1975) subdivide passengers in 3 categories. Some fraction q is supposed to manage to arrive coincidental with the vehicle arrival time and are assumed to have no waiting time. The remaining fraction $(1-q)$ of passengers is split into a proportion $(1-q)p$, who arrive so as to minimise expected waiting time and a proportion $(1-q)(1-p)$ arrive randomly.

In this paper, we want to optimise the heading times of all trains a passenger going from a station O to station D can choose, so that the corresponding waiting time stays low. This will be achieved by minimisation of the total journey time of which this waiting time is a component. The effect on passengers arriving randomly is described by equation (1). We suppose they come in a portion r. So their total waiting time is $r \cdot E(w)$. The primary effect on passengers waiting time of passengers that shift their arrival times with time δt when their planned vehicle departure time is shifted in a new timetable with δt is zero. This means that the fixed waiting time of the portion $(1-r)$ of passengers needs not and so will not occur in the objective function waiting time at departure terms of our timetable optimisation model. When evaluating the total waiting time, their fixed waiting time x can possibly be added as fixed amounts, $(1-r) * x$, which is indeed independent of the timetable.

Bowman and Turnquist (1981) studied passenger arrival patterns at bus stops at seven locations around Chicago. They write that expected passenger waiting time at a transit stop is dependent on three things: the distribution of *passenger arrival times* at that stop, the *planned* arrival times of the *vehicles* at that stop and the deviations in practice from the arrival times of these vehicles. They use a *passenger choice model* that is fitted to measurements of passenger arrival times at bus stops. They study the effect of both decreasing headway times (by adding more vehicles) and decreasing the standard deviation on vehicle arrival times (by increasing reliability in operations) on expected waiting times. They conclude that both can decrease expected waiting time both for the random arrival model and for the passenger choice model. Bowman and Turnquist (1981) suggest that higher heading times will lead to fewer random arrival passengers, since they will want to avoid the higher waiting penalty time. This indicates that our r should in fact be dependent on the number of vehicles N from O to D per timetable period. This is indeed what Fan and Mechemehl (2002) also report from their data analysis. They identified a 10 minute vehicle headway as the transition from random to non-random passenger arrivals.

More recently, the terms *excess journey time (EJT)* and *passenger incidence behaviour* have been added to the transport research vocabulary (Furth and Muller, 2006; Frumin and Zhao, 2012; Zhao et al., 2013). EJT is the difference between the journey time as implied by the published timetable and the actual journey time (Zhao et al., 2013). The waiting time before the first vehicle departs from the moment the passenger arrives at station O and the waiting time after the last vehicle arrives before the passenger leaves station D are the two components making up this excess journey time. The second component can also be seen as inter-departure time of the next transportation system the passenger moves to after the one considered. *Passenger incidence behaviour* is the generic name for how passengers interact with the transportation system (Furth and Muller, 2006). Information about the transportation system can influence how they make choices and use the transportation system. The above mentioned random arrival model and passenger choice model are two different passenger incidence behaviours.

The research mentioned above is mainly focussed on accurately measuring in reality and then fitting and modelling passenger and vehicle arrival distributions and the resulting expected passenger waiting time. However, we are not aware of any research also incorporating any of these models in a timetabling method that automatically minimises passenger waiting time at departure or arrival. This is our focus in this paper.

2.2 Random Arrival Passenger Subcategories

Jolliffe and Hutchingson (1975) mention that the fraction of passengers who are *uninformed or unaware* about timetable departure times of their alternative trains, especially during peak hours, is low. That may be true for some networks, but they are not the only passengers that show random arrival behaviour. Consider table 1. We will traverse it from left to right.

<table>
<tr><td rowspan="6">informed</td><td rowspan="5">caring</td><td rowspan="4">adaptable</td><td rowspan="4">adapting</td><td rowspan="2">in time</td><td>for departure</td><td rowspan="2">$(1-r)$</td><td rowspan="2">passenger choice model</td></tr>
<tr><td>for transfer</td></tr>
<tr><td rowspan="2">over time</td><td>for departure</td><td rowspan="5">r</td><td rowspan="5">random arrival model</td></tr>
<tr><td>for transfer</td></tr>
<tr><td>unadaptable</td><td rowspan="3" colspan="3">not adapting</td></tr>
<tr><td colspan="2">not-caring</td></tr>
<tr><td colspan="3">uninformed</td></tr>
</table>

Table 1: Sub-categories of passengers: fraction (1-r) showing passenger choice model behaviour and fraction r showing random arrival behaviour.

Amongst the *informed* passengers , there are some that do not care or don't want to be too stressed out about getting the first possible train departure. They are informed but are *not adapting* their arrival time to the train departure times available in the timetable. Amongst the passengers who do care, some *cannot* adapt, because they have time constrained obligations before arriving at the train platform. For example they have to bring their kids to the creche at a certain fixed time. We categorise them as *unadaptable*. Amongst the *caring*, *adaptable* passengers, all will be *adapting* but only some will be successful and *in time* for the aimed train departure (at origin or transfer station) and some will be unsuccessful and *over time* for the train departure (again at origin or transfer station). Only the *in time* passengers, with fraction $(1-r)$ should be considered to behave according to the passenger choice model. For this fraction, temporal spreading of alternative trains will not decrease their expected waiting time at departure. The rest, fraction r, behave according to the random arrival model. For them, spreading of alternative trains will decrease their expected waiting time at departure and so is useful. We do not claim that the vertical dimensions of the rows in table 1 are scaled proportionally to the ratios in reality. However, consider that there are quite some types of passengers making up the ratio r. The table can be used to derive the value of r if one can estimate the splitting fractions, at each step, going from left to right through the table.

2.3 Expected Inter-Arrival Waiting Time

Most research focussing on excess journey time mentions inter-departure waiting time but does not mention inter-arrival waiting time. Temporal spreading at train departure is inspired by the idea of randomly arriving passengers, which is the best assumption one can make if no more detailed information is known about passenger incidence behaviour. Similarly, spreading at the arrival side of the transportation system considered is desirable because one wants to minimise waiting time for the transportation system coming after the system a timetable is being constructed for. So at the arrival side, we suppose a waiting time occurs, according to the same model as for the waiting time occurring at the departure side.

2.4 Approaches to Realise Temporal Spreading of Trains

Within the PESP approach to timetabling (Serafini and Ukovich, 1989; Schrijver, 1993; Nachtigall, 1996; Goverde, 1998a,b), the usual approach to enforce constant heading times between trains along the same path is the use of *regularity constraints* (Peeters, 2003; Kroon et al., 2007; Liebchen, 2006, 2007; Caprara et al., 2007; Kroon et al., 2009; Caprara et al., 2011; Sparing et al., 2013). A regularity arc in the PESP model connects two departure times of alternative trains and imposes a fixed heading difference (modulo the timetable period T). We think that sometimes deviation from this optimal spreading should be allowed since minimisation of other expected time components is also important. Secondly, the enforcement of spreading is intended for the portion r of passengers that arrive randomly at their departure station. The portion $(1 - r)$ of people that adapt their arrival time to a particular train departure time have no primary beneficial effect. Hard (regularity) constraints cannot differentiate between two passenger categories while soft spreading can do this via the objective function. Thirdly, in this approach, these regularity constraints are usually only applied to trains which have exactly the same path. So routes which contain transfers do not occur. However this is not a restriction imposed by PESP. We think what matters is that *all alternative train services* leading from O to D should, together, be imposed upon *some degree* of temporal spreading.

When many groups of alternative trains exist, all trains within a group are repetitions of each other and one ends up with a problem called *multi-module PESP* (Galli and Stiller, 2010). For these systems they developed a specific formulation. They claim that the powerful PESP based methods developed for uniform modules generally fail for the multi-module case and demonstrate that their approach helps reduce computation time.

2.5 Minimising Excess Journey Time to Realise Optimal Temporal Spreading of Trains

We think we can formulate the benefit of temporal spreading of alternative trains in terms of the waiting time it saves for passengers. If we minimise the expected waiting time at departure and arrival, together with journey time, the resulting timetable will possess the spreading that is optimal to passengers, not more, not less.

Setting up an integer linear model that achieves minimisation of waiting time, both at departure and at arrival, in the expected passenger domain and applying it on as many trains as possible are the topics of this paper. To the best of our knowledge, no research has been carried out in this direction before.

3 Cost of Waiting for the Next Alternative Train

In section 3.1 the constraints for imposing optimal spreading are derived. Section 3.2 derives the corresponding objective function terms and proves that these will be minimal at equal spreading. Section 3.3 shows that the number of alternative trains stays low in practice.

3.1 Derivation of Basic Constraints, Solving the Unknown Order Problem

Sometimes, a train passenger can choose between multiple train routes that lead him from his origin station O to his destination station D. Some of the alternative routes will consist of a different train or trains visiting the same intermediate stations and some may be composed of different trains that also visit different intermediate stations. Some connections between O and D may require a transfer and some may not. In the sequel, we will simply talk about *alternative trains* or *alternative routes* when we mean the collection of all possibilities to go from O to D using any combination of trains possible. To be able to optimise and also evaluate any train schedule concerning the cost of waiting for alternative routes, we derive this cost analytically here. This cost can then be added to the objective function described by Sels et al. (2013b).

Let $R_{O,D}$ be the set of all route alternatives from station O to D. Say that from O to D, we have a total of $\#R_{O,D} = N$ routes. We define their index set as $I_N = \{0, 1, ..., N-1\}$. We choose the couple of different routes r_i and r_j, where $i \in I_N$ and $j \in I_N$ and $i \neq j$. Let the variables b_i and b_j be these routes' respective planned begin times modulo T in station O. We now choose to develop our derivation for the begin sides of routes, but the derivation for the end sides is entirely similar. To be able to define the times in between two *subsequent* begin times, we need to know the *cyclic order* of these begin times. Since, during timetable optimisation, this is still to be determined in the model, we define a vector $\overline{b}$, with the same N values as vector b, but then sorted in non-decreasing order. Formally, the non-decreasing order is enforced by

$$\forall(O,D): \begin{cases} \forall i \in I_N \setminus \{N-1\}: & \overline{b}_i \leq \overline{b}_{i+1} \\ & \overline{b}_{N-1} \geq \overline{b}_0. \end{cases} \tag{4}$$

Since we need the same scalar values in the vectors $\overline{b}$ and b, we declare that $\overline{b}$ is a permutation, defined by the permutation matrix $p_{i,j}$,of b by imposing

$$\forall(O,D): \forall i \in I_N: \begin{cases} \overline{b}_i = \sum_{j \in I_N} p_{i,j} \cdot b_j \\ \forall j \in I_N: p_{i,j} \in \{0,1\} \\ \sum_{j \in I_N} p_{i,j} = 1 = \sum_{j \in I_N} p_{j,i}. \end{cases} \tag{5}$$

The expected time of waiting for the next departure from O to D will be a function of the time differences in between subsequent $\overline{b}_i, \overline{b}_{i+1}$ values. We call these delta times *supplements* s_i and define them by

$$\forall(O,D): \begin{cases} \forall i \in I_N \setminus \{N-1\} & : & s_i & = & \overline{b}_{i+1} - \overline{b}_i \\ & & s_{N-1} & = & (\overline{b}_0 + T) - \overline{b}_{N-1} \\ \forall i \in I_N & : & 0 & \leq & s_i \leq T - \delta. \end{cases} \tag{6}$$

δ is the smallest time difference in our model, so our time resolution. We used 6 seconds for δ and 1 hour for T, so $T = 600\delta$. The range of s_i from 0 up to, but not including T, is sufficient to allow equation (6) to always have feasible solutions. Indeed, $\forall(i,j) \in I_n : 0 \leq b_i \leq T - \delta$ and so for the non-decreasingly ordered $\overline{b}_i$ it holds that $\forall i \in I_N \setminus \{N-1\}$: $0 \leq \overline{b}_{i+1} - \overline{b}_i \leq T - \delta$. This also allows the case $s_i = 0$. Note that from equation (6), it follows that

$$\forall(O,D): \sum_{i \in I_N} s_i = T. \tag{7}$$

Even though the equations (7) are linearly dependent of the equations (6), for computational reasons, we also impose them in our model.

We now defined all necessary variables to be able to derive the expected waiting time until first departure in O for D. Note that, in the equations (4), (5) and (6), for notation simplicity we left out the indices O, D for $N, \overline{b}_i, p_{i,j}, b_i$ and s_i and we will also do so in the sequel.

3.2 Derivation of Objective Function Terms representing the Expected Excess Journey Time

The accumulative function of the number of randomly arriving passengers, with assumed rate f per hour, wanting to go from O to D and randomly arriving in O is a sawtooth function. From the time $\overline{b}_i$ up to the time $\overline{b}_{i+1} = \overline{b}_i + s_i$, the number of randomly arriving passengers at O has accumulated from 0 up to $f \cdot s_i$, at which time they all take the train departing on route r_j. The total average time they have to wait altogether for their route departure is given by the number of additional people arriving during time dt at origin station O, namely $f \cdot dt$, multiplied by the time they will have to wait $(s_i - t)$, and this integrated over t ranging from 0 to s_i. This gives

$$u_i \quad = \quad \int_0^{s_i} (s_i - t) \cdot f \; dt = s_i f \int_0^{s_i} dt - f \int_0^{s_i} t \; dt = f \frac{s_i^2}{2}. \tag{8}$$

This integration results in the surface of a rectangular triangle from the sawtooth function, with base s_i and height $f \cdot s_i$. The total expected waiting time for a randomly arriving passenger arriving at O and wanting to go to D before he has taken his first train is

$$\forall (O, D) : U = \sum_{i \in I_N} u_i = f/2 \cdot \sum_{i \in I_N} s_i^2. \tag{9}$$

Note the equivalence of equation (9) with the equation (3) from earlier literature. Indeed $s_i = H_i$ and $f = F/T$. One can prove that the function U in equation (9) is minimal when $\forall i \in I_N : s_i = T/N$. Indeed, using a Lagrange multiplier for constraint (7) that has to be satisfied for all solutions, one gets that the following should be minimised

$$\sum_{i \in I_N} s_i^2 - \lambda \cdot \left(T - \sum_{i \in I_N} s_i \right). \tag{10}$$

So the partial derivatives to all s_i and to λ should all be zero. This means

$$\forall (O, D) : \left\{ \begin{array}{lcll} \forall_{i \in I_N} & : & 2 s_i + \lambda & = \quad 0 \\ & : & T - \sum_{i \in I_N} s_i & = \quad 0. \end{array} \right. \tag{11}$$

From the top half of equation (11), it follows that all s_i are equal and from the bottom part, that they are all equal to T/N. Each second derivative to s_i equals 2 which is positive, so $s_i = T/N$ gives a minimum and not a maximum. □

The $\overline{b}_i$ variables in equations (6) will, when integrated in our complete model, also be connected (via the $p_{i,j}$ permutation matrix variables) to the b_i variables. In our complete model, the b_i variables and also the supplement variables s_i between them also occur in other constraints and objective function terms. These hard and soft constraints may influence the optimal choice of the s_i here. As a consequence, the optimal solution where all s_i are equal, found when nothing apart from equations (4), (5), (6) constrains $\overline{b}_i$, may not be the one that is also found when integrating it in our complete model.

Equation (9) is useable for *evaluation* of a given schedule, but for *linear* optimisation, a function that is linear in the model variables s_i is needed. We will solve this issue here by linearisation of the u_i terms. Every term $u_i = f \cdot s_i^2/2$, can be approximated by a piecewise linear function composed of 2 segments, by sampling the curve $(s_i, f \cdot s_i^2/2)$ in 3 points. Since the a priori optimal spreading of begin times for N routes in a time T is at equidistant intervals of time $\frac{T}{N}$, as proven, we take this delta time as one of the three sample values for s_i. This will lead to a high accuracy approximation of the real cost function in the part of the range of s_i that delivers the most optimal solutions. The other necessary sample points are its lower bound 0 and upper bound T. So, the resulting 3 points are

$$\forall(O,D) : \forall i \in I_N : \begin{cases} (s_{i,0}, u_{i,0}) = (0, 0) \\ (s_{i,1}, u_{i,1}) = \left(\frac{T}{N}, \frac{f}{2}\left(\frac{T}{N}\right)^2\right) \\ (s_{i,2}, u_{i,2}) = \left(T, \frac{f}{2}T^2\right). \end{cases} \tag{12}$$

Equation (12) describes a piecewise linear convex function which can be implemented in a linear programming formulation by addition of a linear inequality for each of the two subsequent segments, forming a convex (s_i, u_i) search space together, as follows

$$\forall(O,D) : \forall i \in I_N : \begin{cases} u_i & \geq & u_{i,0} + \frac{u_{i,1}-u_{i,0}}{s_{i,1}-s_{i,0}} \cdot (s_i - s_{i,0}) \\ & = & 0 + \frac{\frac{fT^2}{2N^2}}{\frac{T}{N}}(s_i - 0) = \frac{fT}{2N}s_i \\ u_i & \geq & u_{i,1} + \frac{u_{i,2}-u_{i,1}}{s_{i,2}-s_{i,1}} \cdot (s_i - s_{i,1}) \\ & = & \frac{fT^2}{2N^2} + \frac{\frac{fT^2}{2}-\frac{fT^2}{2N^2}}{T-\frac{T}{N}}(s_i - \frac{T}{N}) \\ & = & \frac{fT^2}{2N^2} + \frac{N}{(N-1)\cdot T}\left(\frac{fT^2N^2-fT^2}{2N^2}\right)(s_i - \frac{T}{N}) \\ & = & \frac{fT^2}{2N^2}\left[1 + \frac{N(N+1)}{T}(s_i - \frac{T}{N})\right]. \end{cases} \tag{13}$$

Since the units of f, T, N and s_i are respectively $1/time$, $time$, 1 and $time$, the right hand sides of (13) are indeed two costs in units of $time$. The fact that our total objective function is minimised rather than maximised, guarantees that the (s_i, u_i)-points resulting from the model solution will lie *on*, rather than *above* the two line segment function. It is important to realise that the values f, T, N are constant (manifest) to the model, so the piecewise linear functions in s_i can be calculated as known (linear) functions of only the model variable s_i at model setup time.

We have now converted the cost function of second degree in (9) into the necessary linear inequalities (13). Thanks to minimisation of our objective function which contains the terms u_i for every OD-pair, and the enforcement of equation (5) and (6), a non-decreasing order for $\bar{b}_i$ will be selected and as a consequence, the $\geq$-signs in (13) will turn into $=$-signs. This will deliver us the correct, be it linearly approximated cost, instead of only a - possibly weak - upper bound of it.

3.3 N is mostly Small in Practice

We have derived a model extension that takes care of spreading concerns. This extension generates equations and costs when ranging over $i \in I_{O,D}$. For each (O, D)-pair, this range contains $N(N-1)$ elements. The question then is if this N does not get impractically large.

In practice, passengers choose between different routes leading them from their origin station O to their destination station D. Our route choice algorithm, as described in Sels et al. (2011), tries to mimic this behaviour by considering the fastest route. Suppose that this route has length l. We also consider all routes up to length $l(1 + a)$, where $a = 20\%$. A transfer is penalised by attributing a time of $p = 15$ minutes to it. We estimate that these values for a and p are realistic average for passengers. van der Hurk et al. (2014) use exactly the same graph for route generation, which they call *extended network* and for one algorithm, they call STA, also consider trip duration only and also penalise transfers. They could validate routes with a route set that passengers take in reality and report that STA is able to generate 95% of the routes correctly. We can expect that our method in Sels et al. (2011), if validated, would reach similar percentages of realistic routes and so that it uses enough relevant information as input.

In our algorithm, we do not consider departure times nor their current spreading in time, since they are still unknown in the timetable to be computed. We believe that taking this information from the current timetable would lead to too much bias to the current, possibly suboptimal, departure times or spreading. To avoid any bias for passengers to prefer routes with particular departure hours, we suppose an a-priori distribution of departure times as well as passengers that is ideal in that sense. This means we suppose both passengers and departure hours to be uniformly spread in the hour.

Train filling levels which passengers may also consider in practice, play no role in our algorithm. Typically, vehicle assignment and decisions on the numbers of cars per train is indeed done after timetabling.

In practice, when routing all passengers through the graph of all Belgian passenger trains, we experience that the number N of alternative routes per OD-pair is usually low. Figures 1 up to 3 represent histograms indicating per number N of found routes for an OD-pair, the number of times that that N occurs over all 18268 OD-pairs derived from ticket sales data in Belgium. Figure 1 shows this for $a = 10\%$, figure 2 for $a = 20\%$ and figure 3 for $a = 30\%$. Naturally, when a is increased, for a particular OD-pair, the number N of found routes cannot decrease. So statistically, when considering all routes, we expect that a higher a leads to a higher average N. The figures 1 to 3 confirm this.

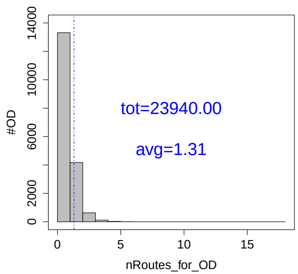

Figure 1: $a = 10\%$

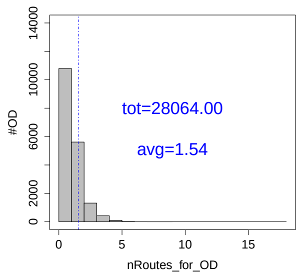

Figure 2: $a = 20\%$

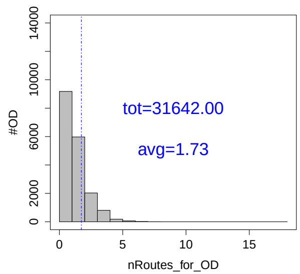

Figure 3: $a = 30\%$

For $a = 10\%$, $a = 20\%$ and $a = 30\%$ respectively only 4, 4 and 6 OD-pairs occur that have a number N between 18 and 32.

The cases with around 32 possibilities for the same OD pair occur between stations around Brussels, where that many trains per hour with equal or similar journey time from O to D do indeed occur. So the number of supplements $N \cdot (N - 1)$, introduced in equations (6), will rarely get big, nor will the same number of boolean variables $p_{i,j}$ in the equations (5) or the number of added inequalities or couples of added inequalities in (13).

3.4 Linearising $p_{i,j} \cdot b_j$

The modelling method described in sections 3.1 and 3.2, has been tested separately with a bunch of random but fixed valued b_i's combined with some values for N. The $p_{i,j}$ indeed give a permutation that makes that $\bar{b}_i$ are sorted in non-decreasing order. With $T = 600\delta$ and $N <= 10$ the model is typically solved in 40 milliseconds. Since N will typically be smaller than 6, this will not be a problem. However, solving such a problem corresponds to determining and calculating the waiting time costs of alternative routes for 1 OD pair only. When considering all trains in Belgium, about 19000 of OD pairs occur simultaneously.

The model should also work when the vector b contains model variables instead of fixed constants. Since the b_j are variables, (since they are also subject to other PESP constraints), the terms $p_{i,j} \cdot b_j$ in equation (5) are quadratic instead of linear. When a linear model is required, this issue is easily solved by introduction of the helper variables $h_{i,j} \in [0, T - \delta]$ where

$$\forall (O, D) : \forall i, j \in I_N : h_{i,j} = p_{i,j} \cdot b_j. \tag{14}$$

Equations (14) are then linearised, using the big-M method (Williams, 1994), to the inequalities

$$\forall (O, D) : \forall i, j \in I_N : \left\{ \begin{array}{lll} (bl - bu)(1 - p_{i,j}) & \leq h_{i,j} - b_j & \leq (bu - bl)(1 - p_{i,j}) \\ bl \cdot p_{i,j} & \leq h_{i,j} & \leq bu \cdot p_{i,j}, \end{array} \right. \tag{15}$$

where bl and bu are the lower respectively upper bounds on all b_j. In our case, $bl = 0$ and $bu = T - \delta$. The top inequality imposes $h_{i,j} = b_j$ when $p_{i,j} = 1$ and imposes nothing new when $p_{i,j} = 0$, while the bottom inequality imposes $h_{i,j} = 0$ when $p_{i,j} = 0$ and imposes nothing new when $p_{i,j} = 1$. This is indeed equivalent to what equations (14) impose. The top line of equation (5) can now be replaced with its linear equivalent

$$\forall (O, D) : \forall i \in I_N : \bar{b}_i = \sum_{j \in I_N} h_{i,j}. \tag{16}$$

The PESP timetabling model is now extended with linear objective function terms and linear constraints to include excess journey time. We wil now apply this model to subsets of the passenger train network in Belgium.

4 Results

This section reports on the results of the application of the model developped in the previous sections on sets of trains extracted from the Belgian timetable as it is planned for 16/12/2014. All experiments in this paper ran Gurobi v5.6.3 on an Intel Xeon CPU E31240 3.3GHz processor with 16GB RAM.

In section 4.1, in order to study the effect of OD spreading separately, we start with a proof of concept where no passenger transfers are considered. A network of 26 trains is used as the test network. Section 4.2 makes a comparison of results obtained with three approaches: (i) our soft spreading technique via the objective function, (ii) the traditional hard spreading technique via hard constraints and (iii) a mixed hard-soft method. In section 4.3, we extend the optimisation by including transfers and verify whether total expected passenger time, and its components (expected ride time, dwell time, transfer time, knock-on time and excess journey time related to temporal spreading) can all still be reduced and if so, to what degree. Section 4.4 does the same but for adding more OD-pairs for spreading optimisation. In section 4.5 the scalability of our method is tested by increasing the network size input to our model.

4.1 Proof of Concept

Table 2 shows the results of optimisations run with a computation time limit of 600 seconds. The OD-pairs which are to be optimised for spreading, are selected by requiring at least 1000 passengers per morning peak for them. This results in 12 OD-pairs. The largest OD-flow occurs between O=Leuven and D=Brussels-Centraal and has F=1751.6 passengers per morning peak. In the case of r=100%, per time unit of time δ of 6 seconds for each of all morning peak hours together, $f = F/T = F/600 = 2.9193$ passengers going to D are assumed to arrive at the origin station O. The 3 Inter City train alternatives for (O,D)=(Leuven, Brussels-Centraal) are IC:A:2, IC:E:2 and IC:F:2. All three connect O and D directly without a transfer. In fact, in these 12 OD-pairs, no single train journey with a transfer occurs.

In the optimisations, the fraction of assumed randomly arriving (and departing) passengers is varied over $r = 0\%, 1\%, 5\%, 10\%, 50\%, 100\%$. For each value of r, a separate optimisation was carried out and a different optimised timetable was computed. Of course, the original timetable, used as reference is the same for all values of r. For each of the 12 OD-pairs that was selected, on the rows s_0, s_1, s_2, the inter-departure times (heading times) at O and inter-arrival times (heading times) at D are given, both for the original timetable (columns *orig O* and *orig D*, italic) and for the optimised one (columns *opt O* and *opt D* * (0,1,5,10,50,100), normal). This gives an idea of the improvement in spreading that was reached. For example, for the case Oostende-Brugge, in Oostende, the original timetable has a spreading of $(s_0, s_1) = (310, 290)$ while the optimised timetables, for many but not all cases of r, deliver the perfect spreading of $(s_0, s_1) = (300, 300)$. For three trains per hour, for example, from Brussel-Zuid to Leuven, we get in the original timetable, at O, $(s_0, s_1, s_2) = (264, 180, 156)$ and at D, $(s_0, s_1, s_2) = (182, 154, 264)$ while in the optimised timetable for $r = 100\%$, we get at O, $(s_0, s_1, s_2) = (195, 154, 264)$ and at D, $(s_0, s_1, s_2) = (200, 200, 200)$, a perfect spreading.

The corresponding excess journey time at O and at D are also mentioned, again both for the original timetable (on the rows u_{orig}) and for the optimised timetable (on the rows u_{opt}). Note that the values on the rows u_{orig}, irrespective of their value for r in their column heading, always refer to the same original timetable. Contrary to this, the values on the rows u_{opt}, refer to a timetable specially optimised for the r value in their column heading. These values indicate how much reduction in excess journey time has been accomplished for a specific OD-pair, for a specific value of r. On the u_{opt} rows a value that is lower than the value above it, in row u_{orig}, is underlined, indicating that a reduction in excess journey time has been achieved. It is marked in bold otherwise. The spreading reached amongst the s_i must be seen as the consequence of the minimisation of the corresponding excess journey time as terms u_i in the objective function according to their relation given by equation (9). We see, as we can expect, that roughly spoken, the higher the value for r and f, the more close the s_i values get to each other, and to T/N.

The values in the rows u_{orig} are proportional to their r value. The columns where $r = 0\%$ do in fact result in $u_{orig} = 0 = u_{opt}$. However, we calculated the incidence wait costs for the timetable reached for $r = 0\%$ as if $r = 1\%$ to show what an underestimation

r (%)	*orig* *O*	0(1)	1	5	10	50	100	*orig* *D*	0(1)	1	5	10	50	100
		opt. O, Begin Spread							opt D, End Spread					
Leuven to Brussel-Centraal, F=1751.6, f=2.9193, [0: IC:A:2:5xx, 1: IC:E:2:15xx, IC:F:2:17xx]														
s_0	*152*	63	86	287	287	156	326	*271*	68	118	119	119	198	104
s_1	*175*	64	111	202	202	200	162	*121*	473	401	283	283	248	333
s_2	*273*	473	403	111	111	244	112	*208*	59	81	198	198	154	163
u_{orig}		*1872*	*1872*	*9361*	*18721*	*93607*	*187215*		*1917*	*1917*	*9586*	*19172*	*95861*	*191722*
u_{opt}		**3383**	**2658**	**9889**	**19778**	90406	**211746**		**3384**	**2646**	**9740**	**19480**	90809	**216432**
Brussel-Centraal to Leuven, F=1751.6, f=2.9193, [0:IC:F:1:17xx, 1:IC:A:1:5xx, 2:IC:E:1:15xx]														
s_0	*154*	81	73	151	197	102	196	*182*	74	200	295	179	200	200
s_1	*263*	73	323	149	181	299	206	*154*	450	74	151	221	102	200
s_2	*183*	446	204	300	222	199	198	*264*	76	326	154	200	298	200
u_{orig}		*1845*	*1845*	*9223*	*18446*	*92232*	*184465*		*1847*	*1847*	*9235*	*18470*	*92351*	*184701*
u_{opt}		**3077**	**2208**	**9853**	17641	**101744**	175243		**3120**	**2215**	**9746**	17645	**101599**	175161
Gent-Sint-Pieters to Brussel-Centraal, F=1648.58, f=2.7476, [0:IC:E:1:15xx, 1:IC:A:1:5xx]														
s_0	*156*	451	205	300	178	300	391	*154*	154	396	300	181	299	206
s_1	*444*	149	395	300	422	300	209	*446*	446	204	300	419	301	394
u_{orig}		*3043*	*3043*	*15213*	*30426*	*152131*	*304262*		*3059*	*3059*	*15293*	*30586*	*152928*	*305856*
u_{opt}		**3099**	2721	12364	28818	123644	270040		3059	2726	12364	28620	123645	271565
Brussel-Centraal to Gent-Sint-Pieters, F=1648.58, f=2.7476, [o:IC:E:2:15xx, 1:IC:A:2:5xx]														
s_0	*121*	541	519	198	402	248	333	*151*	86	101	219	227	372	286
s_1	*479*	59	81	402	198	352	267	*449*	514	499	381	373	228	324
u_{orig}		*3353*	*3353*	*16766*	*33532*	*167662*	*335324*		*3083*	*3083*	*15414*	*30829*	*154144*	*308287*
u_{opt}		**4069**	**3791**	13794	27587	127358	250279		**3731**	**3561**	13266	26193	130765	256590
Oostende to Brugge,F=1177.78,f=1.9629, [0:IC:G:1:18xx, 1:IC:A:1:5xx]														
s_0	*310*	66	300	360	300	300	300	*310*	64	300	360	300	300	300
s_1	*290*	534	300	240	300	300	300	*290*	536	300	240	300	300	300
u_{orig}		*1769*	*1769*	*8843*	*17686*	*88432*	*176863*		*1769*	*1769*	*8843*	*17686*	*88432*	*176863*
u_{opt}		**2842**	1767	**9187**	17667	88334	176667		**2860**	1767	**9187**	17667	88334	176667
Brugge to Oostende, F=1177.78, f=1.9629, [0:IC:G:2:18xx, 1:IC:A:2:5xx]														
s_0	*324*	64	457	313	300	312	300	*300*	63	143	289	300	291	300
s_1	*276*	536	143	287	300	288	300	*300*	537	457	311	300	309	300
u_{orig}		*1778*	*1778*	*8890*	*17780*	*88899*	*177798*		*1767*	*1767*	*8833*	*17667*	*88334*	*176667*
u_{opt}		**2860**	**2251**	8850	17667	88475	176667		**2869**	**2251**	**8845**	17667	**88413**	176667
Brugge to Gent-Sint-Pieters, F=1169.78, f=1.9496, [0:IC:A:1:5xx, 1: IC:G:1:18xx]														
s_0	*310*	64	300	240	301	395	300	*300*	68	296	233	300	234	300
s_1	*290*	536	300	360	299	241	300	*300*	532	304	367	300	366	300
u_{orig}		*1757*	*1757*	*8783*	*17566*	*87831*	*175662*		*1755*	*1755*	*8773*	*17547*	*87734*	*175467*
u_{opt}		**2841**	1755	**9124**	17547	**104357**	175467		**2804**	1755	**9211**	17547	**91980**	175467
Gent-Sint-Pieters to Brugge, F=1169.78, f=1.9496, [0: IC:A:2:5xx, 1:IC:G:2:18xx]														
s_0	*307*	51	183	300	313	300	365	*304*	536	436	313	300	313	253
s_1	*293*	549	417	300	287	300	235	*296*	64	164	287	300	287	347
u_{orig}		*1756*	*1756*	*8778*	*17556*	*87781*	*175563*		*1755*	*1755*	*8775*	*17550*	*87749*	*175498*
u_{opt}		**2963**	**2022**	8773	**17580**	87734	**183704**		**2841**	**2115**	**8790**	17547	**87898**	179774
Leuven to Brussel-Zuid, F=1154.47, f=1.9241, [0: IC:A:2:5xx, 1: IC:E:2:15xx, IC:F:2:17xx]														
s_0	*152*	63	86	287	287	156	326	*208*	116	116	200	117	200	162
s_1	*175*	64	111	202	202	200	162	*271*	401	401	117	283	246	106
s_2	273	473	403	111	111	244	112	*121*	83	83	283	200	154	332
u_{orig}		*1234*	*1234*	*6170*	*12339*	*61696*	*123392*		*1264*	*1264*	*6318*	*12636*	*63181*	*126363*
u_{opt}		**2230**	**1752**	**6518**	**13035**	59586	**139560**		**1743**	**1743**	**6435**	**12870**	59759	**142100**
Brussel-Zuid to Leuven, F=1154.47 , f=1.9241, [0:IC:F:1:17xx, 1:IC:A:1:5xx, 2: IC:E:1:15xx]														
s_0	*264*	82	75	153	196	300	195	*182*	74	200	295	179	200	200
s_1	*180*	71	322	148	180	200	205	*154*	450	74	151	221	102	200
s_2	156	447	203	299	224	100	200	*264*	76	326	154	200	298	200
u_{orig}		*1216*	*1216*	*6082*	*12163*	*60817*	*121635*		*1217*	*1217*	*6087*	*12174*	*60868*	*121735*
u_{opt}		**2035**	**1448**	**6480**	11640	**67344**	115495		**2056**	**1460**	**6424**	11630	**66963**	115447
Gent-Sint-Pieters to Brussel-Zuid, F=1086.56, f=1.8109, [0:IC:E:1:15xx, 1:IC:a:1:5xx]														
Brussel-Zuid to Gent-Sint-Pieters, F=1086.56, f=1.8109, [0:IC:E:2:15xx, 1:IC:a:2:5xx]														

Table 2: Inter-departure and inter-arrival times s_i and corresponding excess journey times u per OD-pair, for optimisations over a set of 26 Belgian IC trains over a range of $F \cdot r$ passengers with assumed random incidence behaviour. OD spreading threshold=1000, no transfers, computation time = 600s.

of the value of r by 1% would result in here. One sees that the u_{opt} values in column $r = 0\%$ are much higher than those in column $r = 1\%$. This is of course the case because at $r = 0\%$, no excess journey time terms are present in the objective function. This indicates that there are quite severe waiting times expected for randomly arriving passengers if one does not model the incident waiting time at all. Table 3 shows that, compared to the excess journey time of the original timetable, an increase of 42% is expected. This indicates that in the original timetable, already quite some spreading was achieved and that in turn explains that even if u_{opt} is just a few percentages removed from u_{orig}, the penalties in the objective function on not spreading must already be delivering their result.

Table 3 shows the column sums over the 12 OD pairs, $U_{O,orig} = \sum_{i=0}^{11} u_{O_i,orig}$, $U_{O,opt} = \sum_{i=0}^{11} u_{O_i,opt}$ and similarly for the destinations D, for the different values for r. The sums for origin and destination are also calculated as $U_{O+D,orig} = U_{O,orig} + U_{D,orig}$ and similarly for the optimised schedule with index opt. This table shows that for $r = 0\%$ and $r = 1\%$ we get an increase of excess journey time compared to the original timetable of respectively 42% and 12%. For r larger than 5% our model manages a decrease of between 5% and 7%. We conclude that, for this system of 26 trains and 12 OD pairs, optimisation with our model is able to quickly (in 600 seconds) reduce the excess journey time significantly $(> 5\%)$ compared to a manually constructed timetable as long as the fraction r is also significant $(\geq 5\%)$. If r is very low $(\leq 1\%)$, the excess journey time terms in the objective function have insufficient weight to strongly influence the solution and those few r passengers will experience excess journey times that are higher than in a manual constructed timetable. Other terms then dominate the optimisation of the total time which is still reduced. This proves that our added excess journey time model serves its basic purpose.

sol_i	r(%)	0(1)	1	5	10	50	100
sol_1	MIP gap (%)	16.6	59.4	37.9	65.2	24.9	15.7
	computation time (s)	13	102	48	182	59	242
$sol_{end,opt}$	comp. time=600s $\Rightarrow$ gap (%)	0.65	1.64	1.83	2.62	4.36	5.74
	$U_{O,orig}$	23774	23774	118871	237742	1188710	2377419
	$U_{O,opt}$	**34098**	**26648**	112036	226063	1101965	2217687
	ratio	**1.43**	**1.12**	0.94	0.95	0.93	0.93
	$U_{D,orig}$	23469	23469	117344	234688	1173442	2346884
	$U_{D,opt}$	**32947**	**26386**	110901	223033	1093150	2228194
	ratio	**1.40**	**1.12**	0.95	0.95	0.93	0.95
	$U_{O+D,orig}$	47243	47243	236215	472430	2362152	4724303
	$U_{O+D,opt}$	**67045**	**53033**	222936	449096	2195115	4445882
	ratio	**1.42**	**1.12**	0.94	0.95	0.93	0.94
	excess journey time reduction (%)	**-42**	**-12**	6	5	7	6

Table 3: Total origin-, total destination- and total expected excess journey time U over the OD-pairs of table 2. Corresponding original to optimised timetable excess journey time reductions, MIP gaps and computation times. sol_1 is the first feasible solution and sol_{end} the one at the set time limit.

4.2 Comparing Soft with Hard Spreading

If one wants *perfect temporal spreading* of alternative trains, with all $s_i = T/N$, for a set of given OD pairs, we *additionally* impose

$$\forall (O,D) : \forall i \in I_N : s_i = T/N. \tag{17}$$

sol_i	r(%)	0	1	5	10	50	100
Soft spreading at Origin and Destination							
	comp. time=1200s ⇒ MIP gap (%)	0.83	3.57	7.88	8.18	4.74	5.98
	$U_{O,orig}$	0	27705	138524	277047	1385237	2770474
	$U_{O,opt}$	0	25612	**144050**	**328453**	1189351	2516351
	ratio	1.00	0.92	**1.04**	**1.19**	0.86	0.91
	$U_{D,orig}$	0	27385	136926	273852	1369259	2738519
	$U_{D,opt}$	0	25095	**140854**	**326685**	1177353	2429862
sol_{end}:opt	ratio	1.00	0.92	**1.03**	**1.19**	0.86	0.89
	$U_{O+D,orig}$	0	55090	275450	550899	2754496	5508992
	$U_{O+D,opt}$	0	50707	**284904**	**655138**	2366704	4946213
	ratio	1.00	0.92	**1.03**	**1.19**	0.86	0.90
	excess journey time reduction (%)	0	8	**-3**	**-19**	14	10
	opt. total time reduction (%)	15.42	15.34	15.17	15.03	15.31	14.96
sol_{end}:eval	post-opt. evaluation total time reduction (%)	4.6	6.26	5.68	4.81	3.71	4.92
Hard Spreading at Origin and Destination							
	comp. time=1200s ⇒ MIP gap (%)	68.07	71.56	66.9	64.75	51.64	39.36
	$U_{O,orig}$	0	27705	138524	277047	1385237	2770474
	$U_{O,opt}$	0	21060	105301	210603	1053013	2106026
	ratio	1.00	0.76	0.76	0.76	0.76	0.76
	$U_{D,orig}$	0	27385	136926	273852	1369259	2738519
	$U_{D,opt}$	0	21060	105301	210603	1053013	2106026
sol_{end}:opt	ratio	1.00	0.77	0.77	0.77	0.77	0.77
	$U_{O+D,orig}$	0	55090	275450	550899	2754496	5508992
	$U_{O+D,opt}$	0	42120	210603	421205	2106026	4212052
	ratio	1.00	0.76	0.76	0.76	0.76	0.76
	excess journey time reduction (%)	0	24	24	24	24	24
	opt. total time reduction (%)	6.89	5.64	7.93	6.79	6.84	6.28
sol_{end}:eval	post-opt. evaluation total time reduction (%)	5.24	3.68	6.02	4.96	4.61	3.72
Hard Spreading at Origin and Soft Spreading at Destination							
	comp. time=1200s ⇒ MIP gap (%)	28.15	31.0	35.59	30.45	21.59	13.59
	$U_{O,orig}$	0	27705	138524	277047	1385237	2770474
	$U_{O,opt}$	0	21060	105301	210603	1053013	2106026
	ratio	1.00	0.76	0.76	0.76	0.76	0.76
	$U_{D,orig}$	0	27385	136926	273852	1369259	2738519
	$U_{D,opt}$	0	22979	115029	229787	1176459	2278535
sol_{end}:opt	ratio	1.00	0.84	0.84	0.84	0.86	0.83
	$U_{O+D,orig}$	0	55090	275450	550899	2754496	5508992
	$U_{O+D,opt}$	0	44039	220330	440389	2229472	4384561
	ratio	1.00	0.80	0.80	0.80	0.81	0.80
	excess journey time reduction (%)	0	20	20	20	19	20
	opt. total time reduction (%)	14.55	14.41	14.06	14.24	14.23	14.7
sol_{end}:eval	post-opt. evaluation total time reduction (%)	4.83	3.78	4.07	4.23	3.63	3.27

Table 4: Effects of soft, hard and hard-soft spreading enforcement at Origin and Destination for 26 trains. OD spreading threshold=1000, no transfers.

The existing permutation matrices and objective function terms defined before can still reside in the model and continue to fulfil their purpose of correctly measuring the ordering of train alternatives and excess journey time. To investigate the effects of hard versus soft spreading, we compared three cases, where we allow 1200 seconds of optimisation time for each. The first case is one where we apply soft temporal spreading as before on origin and destination side, so by the model without equations (17). The second case is one where we enforce hard spreading via equations (17) on both O and D side. In the last case we apply hard spreading on the origin side, but soft spreading on the destination side. Note that the MIP gaps of methods with additional hard constraints are typically not reduced as quickly as methods without these. So we compare solutions for these three cases from the fair standpoint of what can be achieved in equal computing times.

Table 4 gives the results of these experiments. First of all, in the top third of the table, for soft spreading at origin and destination, it shows that a total time reduction during optimisation of around 15% is again achieved. However, when doing a post-evaluation with all excess waiting time associated with the 12 OD-pairs, we get a weaker reduction of total expected time from the original to the optimised timetable in the range of 3.71% to 6.26%.

When forcing spreading the hard way, on both origin and destination side, the middle part of Table 4 shows that we get systematic savings of 24 percent in excess journey time. This is the same for all values of r, which is logical. Since the excess journey time is forced to the minimum possible here, this is also higher than the savings of around 15% for the soft spreading technique. However, the reductions in total time range between 3.68% and 6.02%, and so, are similar to the soft spreading technique. This means that the increase in reduction from 15% to 24% in excess journey time has been lost in other time components of the objective function. In fact, more ride and dwell supplements have been added for trains between O and D for all train alternatives to match the strict requirements of equal spreading on both ends. So these trains become 'overstretched'.

To try and avoid this effect of overstretching trains, we now enforce spreading in the hard way on the origin and in a soft way on the destination side. The lower third of table 4 shows the results. We find that the excess journey time reduction is around 14.5% and the total time reduction is in the range 3.27% to 4.83%. This is comparable to the results of the approach with soft enforcement of temporal spreading in the upper third part of the table.

We conclude that the three approaches give similar reductions in total time but the approach of hard constraints on both sides gives a strong bias towards concentrating mainly on the reduction of excess journey time with the negative consequence of increasing expected ride and dwell time more than in the other two cases. For these experiments, which do not consider the component of expected transfer time, as for the highest total expected time reduction at post optimisation evaluation, there is no clear winner method yet. Table 4 shows, for each value of r, the highest percentage for this time reduction as underlined.

4.3 Adding Transfers

In the previous sections, we studied the minimisation of the excess journey time cost. We explicitly removed all transfer time costs from the objective function to avoid they would bias the obtainable reduction percentage of the excess journey time. We now add transfer time terms in the objective function and will see if our model is then also still able to reduce the total expected passenger time, including excess journey time and transfer time. The results are summarised in figure 4. To most clearly show the excess journey time component, in each case, r is set to 100%. Each of the six pictures shows a bar graph for the original timetable on the left and the optimised timetable on the right. We consider the same three cases as in the previous section and so, the organisation of in three rows of figure 4 is similar to that of table 4. The upper third represents the optimisation with soft enforcement of spreading on both O and D side. The middle third shows the optimisation with hard enforcement of spreading on both O and D side. The lower third shows results for hard enforcement of spreading on the O side and soft enforcement of spreading on the D side. The left half shows what is reached during optimisation (on the linearised objective function over the selected portion of OD-pairs for spreading and the selected transfers). The right half shows the evaluation of the original, non-linearised objective function on all OD-pairs and on all transfers and represents the end result on which each method should be evaluated.

In all cases, the dark yellow blocks are summed minimal ride and minimal dwell times.

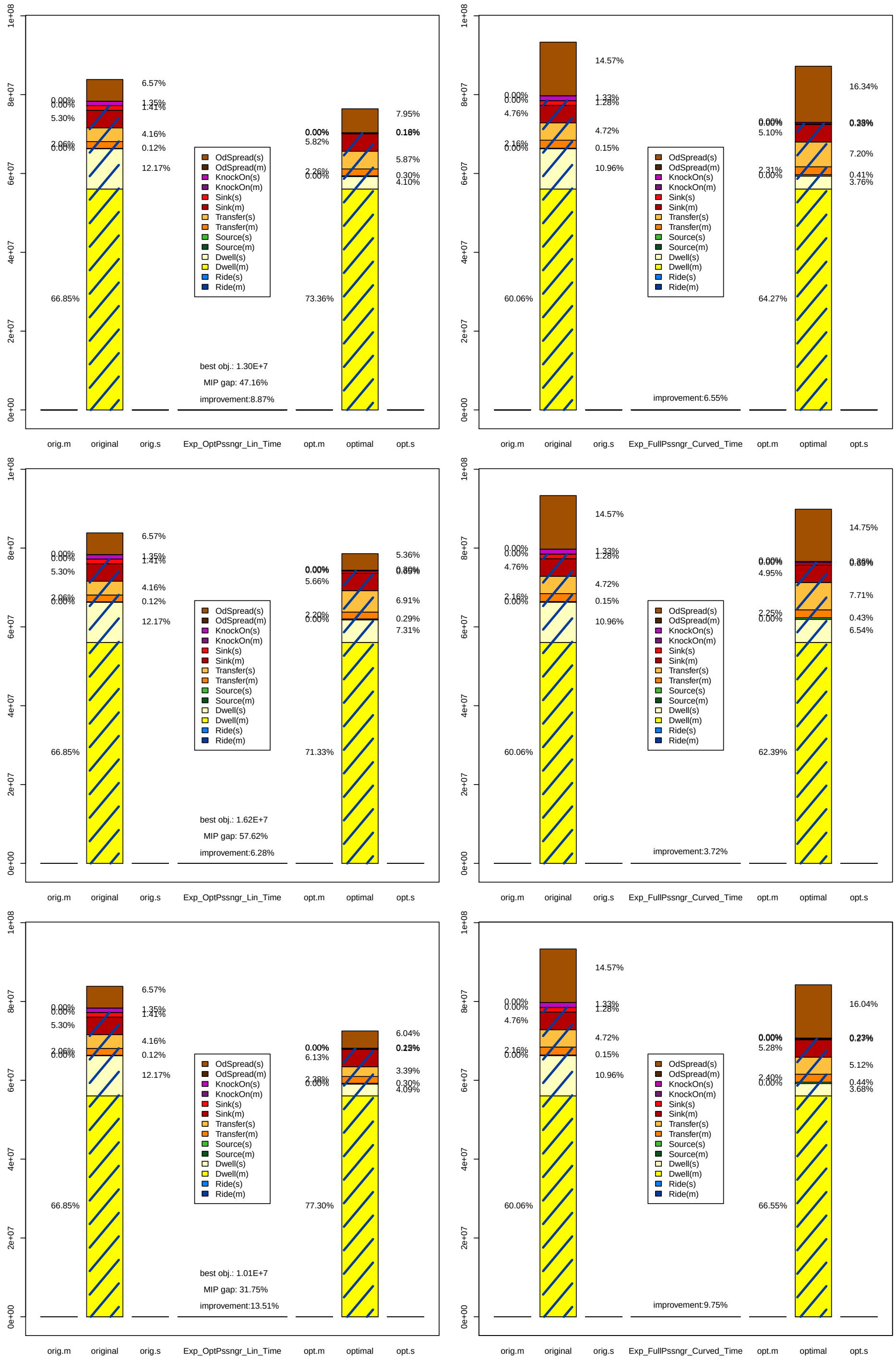

Figure 4: Effects of soft, hard and hard-soft spreading enforcement at Origin and Destination for 26 trains and adding transfers. OD Spreading threshold=1000, transfer threshold=210, computation time=1200s. Left bars are for the original timetable and right bars for the optimised timetable. Note the bigger ride and dwell expected time components on the middle row, representing hard spreading on O and D side.

These are constant and cannot be reduced during optimisation. Any blue shaded block represents an action that has its time convoluted with the 'preceding' ride action (Sels et al., 2011). The (blue shaded) light yellow part represents the summed convoluted ride and dwell supplement times. The (blue shaded) green block corresponds to time attributed to passengers entering and the red block to passengers leaving the transportation system. Orange (blue shaded) blocks represent transfer time. Dark orange is the minimum transfer time and is supposed to be 3 minutes everywhere. Light orange stands for the total expected passenger transfer time due to the transfer supplements. This includes a penalty of 1 hour in case the transfer is missed. In the left column, there are 745 transfers contributing, which are the ones considered during optimisation. In the right column, all transfers are considered during evaluation. The (not blue shaded) light purple colour indicates expected knock-on time, also passenger weighted (Sels et al., 2013a), between subsequent trains on the same infrastructure resource. Brown blocks stand for excess journey time for O and D for all 12 OD pairs together, in the left column. All 7601 OD-pairs are evaluated for excess journey time in the right column. By optimising the timetable, only the light yellow, light orange, light green, light red, purple and brown parts can be shrunk The dark yellow, dark orange, dark green, dark red parts correspond to action minima and are the same in any timetable, so cannot be shrunk.

In the upper-left figure, from the original timetable to the optimised one we get a total time reduction of 8.87% during optimisation. If we call the left bar 100, the right bar represents 100-8.87=91.13. The brown excess journey time component is slightly increased with respect to the original timetable, in absolute terms, from 6.57 out of 100 to 7.95% of 91.13 = 7.2. The transfer time is also growing from 4.16 to 5.87% of 91.13 so 5.35. However, the light-yellow-coloured ride and dwell supplements are strongly reduced from 12.17 out of 100 to 4.10% of 91.13 so 3.74. The decrease of ride and dwell supplements more than compensates the slight increase of expected excess journey time and expected transfer time. The net result is that total expected passenger journey time decreases.

In the left figure of the middle row, representing hard spreading enforcement on original and destination side, we see that the total expected time reduction is slightly less, 6.28%. So the height of the right bar-graph corresponding to the optimised timetable now represents 100-6.28=93.72. However, the brown block representing excess journey time is now reduced, from 6.57% down to 5.36% of 93.72=5.02. The orange transfer component grows from 4.16% to 6.91% of 93.72=6.48, a bigger increase compared with soft spreading enforcement. The ride and dwell supplements are reduced from 12.17 to 7.31% of 93.72 so 6.85 which is substantially more than the 3.74 reached with soft spreading enforcement. Soft enforcement reaches 8.87% total expected time reduction while hard enforcement reaches only 6.28%. Since total time reduction is the end goal, from this observation, soft enforcement is clearly preferable. The full evaluation in the right half of the middle row of figure 4 shows that also when including OD pairs and transfers with fewer passengers than the ones considered in optimisation, the results are similar and even amplified. The reduction is 6.55% for soft enforcement but only 3.72% for hard enforcement of spreading.

On the bottom row, we find the figures corresponding to the optimisation with hard spreading enforcement on the origin and soft enforcement on the destination side. The reduction of total time the optimisation achieves is a remarkable 13.51%. This is significantly better than the other approaches which reached 8.87% and 6.28% total expected time reduction. The excess journey time goes from 6.57 on 100 to 6.04 on (100-13.51)=86.49 so 5.22. In absolute terms this is very close to the 5.02 strict spreading on both sides achieves.

Expected transfer time is reduced from 4.16 on 100 to 3.39 on 86.49 so 2.93. This is almost the double reduction of the 5.35 and 6.48 we get in the other spreading approaches. Ride and dwell supplements are reduced to 4.90% of 86.49 so 3.54. Again this is better than the 3.74 and the 6.85 we saw in the other approaches. So we get better optimisation results when enforcing more hard constraints. Normally, when keeping the same objective function, imposing extra constraints to a model cannot result in a better 'optimal' solution. But here, the solver handles the model with the extra constraints *faster* than the one without. Indeed, in our cases the optimal solution is not reached yet by the solver, which is indicated by the MIP gaps presented in the left half of figure 4. From top to bottom, at 1200 seconds, soft spreading achieves a gap of 47.16%, hard spreading a gap of 57.62% and mixed spreading a gap of 31.75%. When we increase the computation time limit to 3600 seconds, the respective gaps become 41.9%, 58.33% and 31.54%. So between 1200 seconds and 3600 seconds, only the soft spreading model still significantly improves its solution. But even the best result for mixed spreading at 1200 seconds is not beaten at 3600 seconds by any of the two other methods. Note that In all cases, both for 1200 seconds and 3600 seconds, the bound is the same (within a margin of 0.3%), so gaps are comparable.

In the evaluation of all OD-pairs on excess journey time and all transfers in the right figure on the lower row, we notice a 9.75% net reduction is achieved (in 1200 seconds). Again, this is significantly better than the 6.55% and 3.72% of the other approaches. One of the reasons is that transfer time is now slightly reduced instead of significantly growing. We conclude that, once transfers are included, spreading the hard way on the origin and spreading the soft way on the destination (or probably also the reverse) is the preferable tactic for achieving the best results in the shortest time on this network of 26 trains.

4.4 Adding more OD-pairs for Spreading

For the same network of 26 Inter City trains, we now investigate the sensitivity of solution quality and computation time to the amount of OD-pairs in the model. We reduced the OD-threshold for consideration in the model of 1000 passengers per morning peak, in steps of 100, as such optimising over more and more OD-pairs. The transfer threshold is set back to 2000, hoping for low computation times. For each OD-pair, we enforce strict optimal spreading on the origin side and soft spreading via the objective function on the destination side, since this seemed the most promising technique in the previous section.

sol_i	OD-threshold for spreading	900	800	700	600	500	400
	# OD-pairs in opt.($N \geq 1$)	18	24	36	56	90	128
	# OD-pairs in opt. ($N > 1$)	14	16	22	24	30	31
sol_1	MIP gap (%)	23.2	50.1	38.2			
	computation time (s)	52	107	253			
$sol_{end,opt}$	MIP gap (%)	13.2	14.3	17.13			
	computation time (s)	1200	1200	1200	2400	2400	2400
	excess journey time reduction (%)	20.24	20.03	20.72			
	opt. total time reduction (%)	14.69	14.59	14.10			
$sol_{end,eval}$	transfer time reduction (%)	-240	-234	-239			
	post-opt. evaluation total time reduction (%)	4.38	4.65	4.72			

Table 5: Effects of adding more OD-pairs for temporal spreading to the optimisation model. 26 Inter City trains. Computation times are 1200s and 2400s.

The results are given in table 5. For the cases F=900, 800 and 700, models are constructed which perform temporal spreading of at least 2 alternative trains of respectively

14,16 and 22 OD-pairs. These models could all be optimised in 1200 seconds. Over these cases, the MIP gap achievable in 1200 seconds rises, yet the reduction of total expected waiting time stays fairly stable around 14%. The end result reduction when evaluating over all OD-pairs and all transfers increases somewhat from 4.38% over 4.65% to 4.72%.

For the cases F=600, 500 and 400, models are constructed that perform spreading on respectively 24, 30 and 34 OD-pairs with more than one alternative train. However, the imposed time limit of 2400 seconds did not suffice to solve these models.

4.5 Scaling Up to a Larger Network

We tried the approach of soft spreading for both O and D on a network of all 74 Inter City trains as they are planned for 16/12/2014. The thresholds for OD-pairs to be considered in optimisation was again set to 1000 and the transfer threshold to 210 passengers per morning peak. The required gap was set to 95%. This did yield a resulting timetable after about 26.7 hours. Optimisation reduced total expected passenger travel time by 2.76% but post evaluation on all OD-pairs for spreading and all transfers gave an increase of 6.48%. Decreasing the required gap below 95% could possibly yield a better timetable, but at the expense of more computation time. Increasing to systems with more than 74 hourly passenger trains did not result in a solution yet. Since our MILP model without excess journey time *can* be solved for all Belgian passenger trains in about *2 hours* (Sels et al., 2014), we conclude that specifically our excess journey time model is not easily scalable yet to larger networks.

5 Conclusions and Further Work

For passengers with random arrival behaviour, we integrated the expected excess journey time cost into our previous PESP based timetabling MILP model (Sels et al., 2014). This also required derivation and integration of extra constraints in our model. Our objective function which represented total expected journey time is now extended to also include excess journey time for all passengers. This paper shows that this model is indeed still able to reduce the total journey time, while it also reduces each of the separate components: ride-dwell supplements, knock-on time and excess journey time as well.

We were unable to solve the model for the network of all passenger trains in Belgium though. Possibly restricting the range of integers present in the formulation of the excess journey time cost could be a remedy for this problem. Alternatively, our high computation times could potentially be reduced by the addition of special cycle sets or a column generation approach. Also, a multi-module PESP formulation approach could be explored.

Incorporating the secondary effect of the waiting time that passengers experience when they miss the train they intended to depart with could make the model even more realistic.

The missed transfer penalty of cycle time T, currently assumed in the model, could now also be adapted as follows. If the transfer takes the transferring passenger on an OD-route that is contained in the optimised OD-pair spreading set, the N trains can be assumed to be well-spread, certainly if hard spreading is imposed. So the penalty for missing a train can be assumed to be about T/N in the average case. This would make a more realistic transfer model and the transfer time component would make up a smaller portion of the total time. Since it typically still grows during optimisation, the total time would be further reduced.

References

Barnett, A., 1974. On Controlling Randomness in Transit Operations. Transportation Science. 8, 102–116.

Bly, P. H., Jackson, R. L., 1974. Evaluation of Bus Control Strategies by Simulation. Tech. rep., Transport and Road Research Laboratory.

Bowman, L., Turnquist, M., 1981. Service Frequency, Schedule Reliability and Passenger Wait Times at Transit Stops. Transportation Research 15A (6), 465–471.

Caprara, A., Kroon, L., Monaci, M., Peeters, M., Toth, P., 1 2007. Passenger Railway Optimization. Handbooks in Operations Research and Management Science. Vol. 14. Elsevier, Amsterdam, 129-187.

Caprara, A., Kroon, L., Toth, P., 2011. Optimization Problems in Passenger Railway Systems. Wiley Encyclopedia of Operations Research and Management Science 6, 3896–3905.

de Pirey, Y., 1971. Simulation of a Bus Line as a Means to Analyse and Improve Bus Transit Reliability. Master's thesis, Northwestern University, Department of Civil Engineering, Evanston, Illinois.

Fan, W., Mechemehl, B., june 2002. Characterising Bus Transit Passenger Waiting Times. Proceedings of 2nd Material Speciality Conference of the Canadian Society for Civil Engineering, June 5-8, Montreal, Quebec, Canada.

Friedman, R. R., 1976. Statistical Models of the Mean and Standard Deviation of Passenger Wait Time in Bus Transit. Master's thesis, Northwestern University, Transportation Center, Evanston, Illinois.

Frumin, M., Zhao, J., 2012. Analysing Passenger Incidence Behaviour In Heterogeneous Transit Services Using Smartcard Data and Schedule Based Assignment. Transportation Research Record: Journal of the Transportation Research Board. 2274 (1), 52–60.

Furth, P. G., Muller, T. H. J., march 2006. Service Reliability and Hidden Waiting Time: Insights from AVL Data. Transportation Research Records.

Galli, L., Stiller, S., october 2010. Strong formulations for the Multi-module PESP and a quadratic algorithm for Graphical Diophantine Equation Systems. Proceeding EAS'10 of the 18th annual European conference on Algorithms: Part I, 338–349.

Goverde, R., 1998a. Optimal Scheduling of Connections in Railway Systems. Paper presented at the 8th WTCR, Antwerp, Belgium.

Goverde, R., 1998b. Synchronization Control of Scheduled Train Services to minimize Passenger Waiting Times. Proceedings of 4th TRAIL Annual Congress, TRAIL Research School, Delft, The Netherlands.

Holroyd, E. M., Scraggs, D. A., 1966. Waiting Times for Buses in Central London. Traffic Engineering Control. 8, 158–160.

Islam, M. K., Vandebona, U., October 2010. Reliability Analysis of Public Transit Systems using Stochastic Simulation. Proceedings of the 33rd Australian Transport Research Forum Conference, 29 September - 1 October, Canberra, Australia.

Jackson, W. H., 1977. Bus Passenger Arrival Times. Master's thesis, University of Toronto, Department of Civil Engineering, Toronto, Ontario.

Jolliffe, J. K., Hutchingson, T. P., 1975. A Behavioral Explanation of the Association between Bus and Passenger Arrivals at a Bus Stop. Transportation Science. 9, 248–282.

Kroon, L., Dekker, R., Vromans, M., 2007. Cyclic Railway Timetabling: A Stochastic Optimization Approach. Algorithmic Methods for Railway Optimization. Lecture Notes in Computer Science, 41–66.

Kroon, L., Huisman, D., Abbink, E., Fioole, P.-J., Fischetti, M., Maróti, G., Schrijver, A., Ybema, R., 2009. The New Dutch Timetable: The OR Revolution. Interfaces 39, 6–17.

Liebchen, C., 2006. Periodic Timetable Optimization in Public Transport. Ph.D. thesis, Technische Universität Berlin.

Liebchen, C., 2007. Periodic Timetable Optimization in Public Transport. Operations Research Proceedings 2006, 29–36.

Nachtigall, K., 1996. Periodic network optimization with different arc frequencies. Discrete Applied Mathematics 69, 1–17.

Okrent, M. M., 1971. Effects of Transit Service Characteristics on Passenger Waiting Time. Master's thesis, Northwestern University, Department of Civil Engineering, Evanston, Illinois.

Osuna, E. E., Newell, G. F., 1972. Control Strategies for an Idealized Public Transportation System. Transportation Science. 6 (1), 57–72.

Peeters, L., 1 2003. Cyclic Railway Timetable Optimization. Ph.D. thesis, Erasmus Resaerch Institue of Management (ERIM), Rotterdam.

Schrijver, A., 1993. Spoorwegdienstregelingontwikkeling (Timetable Construction). Technical Report, CWI Center for Mathematics and Computer Science, Amsterdam.

Sels, P., Dewilde, T., Cattrysse, D., Vansteenwegen, P., Feb. 2011. Deriving all Passenger Flows in a Railway Network from Ticket Sales Data. Proceedings of 4th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailRome2011, February 16-18, Rome, Italy.

Sels, P., Dewilde, T., Cattrysse, D., Vansteenwegen, P., 2013a. Knock-On Delay Modelling and Optimisation: The Passenger Perspective. In: Proceedings of the 3rd International Conference on Models and Technologies for Intelligent Transport Systems (MT-ITS 2013), December 2-4, Dresden, Germany. pp. 1–10.

Sels, P., Dewilde, T., Cattrysse, D., Vansteenwegen, P., May 2013b. Expected Passenger Travel Time as Objective Function for Train Schedule Optimization. Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailCopenhagen2013, May 13-15, Copenhagen, Denmark.

Sels, P., Dewilde, T., Cattrysse, D., Vansteenwegen, P., 2014. Reducing the Passenger Travel Time in Practice by the Automated Construction of a Robust Railway Timetable, under review. Transportation Research Part B.

Serafini, P., Ukovich, W., 1989. A Mathematical Model for Periodic Scheduling Problems. SIAM Journal on Discrete Mathematics 2, 550–581.

Sparing, D., Goverde, R. M., Hansen, I. A., 2013. An Optimization Model for Simultaneous Periodic Timetable Generation and Stability Analysis. Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailCopenhagen2013, May 13-15, Copenhagen, Denmark.

Turnquist, M. A., 1978. A Model for Investigating the Effects of Service Frequency and Reliability on Bus Passenger Waiting Time. Transportation Research Records. 663, 70–73.

van der Hurk, E., Kroon, L., , Gábor, M., Vervest, P., december 2014. Deduction of Passengers' Route Choice from Smart Card Data. IEEE-ITSC2013 on Intelligent Transport Systems, to appear 2014.

Welding, P. I., 1957. The Instability of Close Interval Service. Operational Research Society. 8 (3), 133–148.

Williams, H., 1 1994. Model Building in Mathematical Programming, 3rd Edition. Vol. 1. John Wiley and Sons Ltd, England.

Zhao, J., Frumin, M., Wilson, N., Zhao, Z., 2013. Unified Estimator for Excess Journey Time under Heterogeneous Passenger Incidence Behaviour using Smartcard Data. Transportation Research, Part C. 34, 70–88.