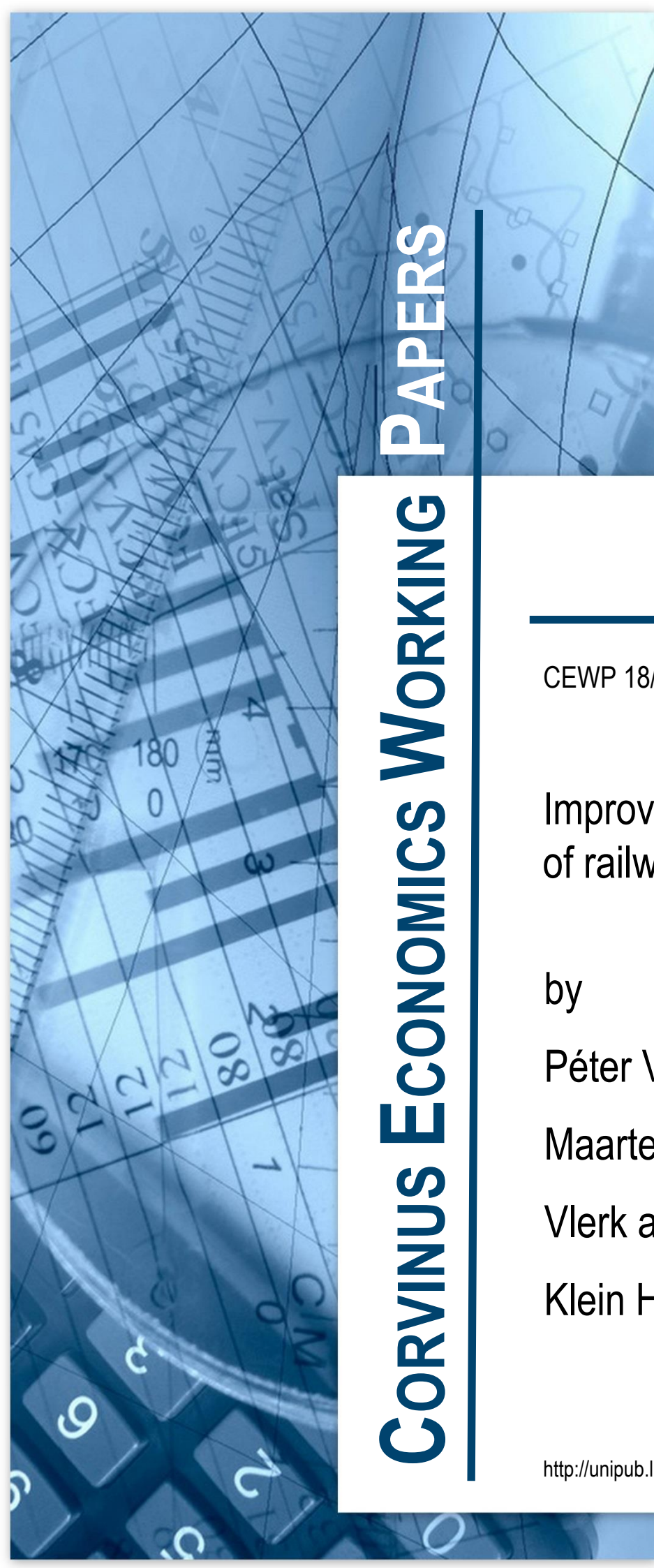




PAPERS

CORVINUS ECONOMICS WORKING



Faculty of Economics

CEWP 18/2015

Improved stochastic optimization
of railway timetables

by

Péter Vékás,

Maarten van der

Vlerk and Willem

Klein Haneveld

<http://unipub.lib.uni-corvinus.hu/2094>

Improved stochastic optimization of railway timetables

P. Vékás* – M. H. van der Vlerk** – W. K. Klein Haneveld**

October 21, 2015

Abstract

We present a general model to find the best allocation of a limited amount of supplements (extra minutes added to a timetable in order to reduce delays) on a set of interfering railway lines. By the best allocation, we mean the solution under which the weighted sum of expected delays is minimal. Our aim is to finely adjust an already existing and well-functioning timetable. We model this inherently stochastic optimization problem by using two-stage recourse models from stochastic programming, building upon earlier research from the literature. We present an improved formulation, allowing for an efficient solution using a standard algorithm for recourse models. We show that our model may be solved using any of the following theoretical frameworks: linear programming, stochastic programming and convex non-linear programming, and present a comparison of these approaches based on a real-life case study. Finally, we introduce stochastic dependency into the model, and present a statistical technique to estimate the model parameters from empirical data.

JEL code: C61

Keywords: stochastic programming, operations research, transportation, railway timetables

* Corvinus University of Budapest, Department of Operations Research and Actuarial Science, MTA-BCE “Lendület” Strategic Interactions Research Group.

Corresponding author, e-mail: *peter.vekas@uni-corvinus.hu*.

** University of Groningen, Department of Operations.

1 Introduction

Planning railway timetables is a complicated task mainly for the following two reasons:

- From one perspective, it is an *organization problem* that involves planning the movement of several trains through several periods by considering the needs of the passengers and the available personnel and rolling stock, not to mention complex interactions among trains.
- Additionally, it is a *decision problem under uncertainty*, as several uncertain factors need to be considered which may result in minor or major deviations from the timetable.

In this paper, we shall *ignore* the first dimension of the problem, taking an initial timetable for granted. Instead, we shall only consider a specific problem arising from the uncertainty of the realized travel times.

1.1 The supplement allocation problem

To avoid delays resulting from unexpected occurrences (e.g. objects on the track, insufficient acceleration, unexpected traffic control decisions etc.), travel times of trains between stations are usually planned to be longer than the corresponding technically feasible travel times. In railway terminology, any planned travel time between two stations consists of the sum of a minimally feasible travel time and a *supplement*. In fact, nearly all delays could be eliminated by introducing a sufficient amount of supplements in the timetable. However, passengers would most likely disapprove of the resulting excessive travel times. Thus, railway companies restrict the total amount of distributable supplements to some reasonable extent in practice and try to distribute it efficiently in the timetable. We emphasize that the aim of this approach is to improve an already existing and well-functioning timetable by introducing minor changes in it instead of attempting to reconstruct it from scratch. Only a few minutes of the given total will be reallocated, which may nevertheless result in large reductions in expected delays, as we shall show later in our paper. As our aim is to increase

the robustness of a timetable against small disturbances that do not completely overturn an existing timetable, we shall only consider relatively small disturbances in our model.

1.2 A brief literature overview

Researchers and railway operators have created various models for the supplement allocation problem outlined in the previous subsection. Most of these models are based on stochastic simulation (e.g. [3]) or max-plus algebra (e.g. [6]), while some of them may be characterized as analytical models (e.g. [7] and [32]). Some researchers in the Netherlands have produced a remarkable amount of related academic research in the recent years including several comprehensive theses on the topic (e.g. [5], [6], [21], [24], [30] and [32]). An especially promising and relatively new modeling approach is the application of the theory of stochastic optimization and in particular, two-stage linear recourse models from stochastic programming (see e.g. [1], [13] and [23]) within the broader discipline of operations research in order to successfully tackle the supplement allocation problem. In contrast to other approaches, stochasticity and optimization are simultaneously inherent components of these models, which may be formulated in an elegant fashion, and a number of related well-researched algorithms are readily available for their efficient solution (see e.g. [10] and [19]). This approach has been pioneered by Kroon and Vromans, who have published several related important contributions in the literature ([14]–[16], [22], [30] and [31]). Our paper aims to present a number of convenient theoretical and practical contributions to this modeling approach.

1.3 Contribution

We construct a general stochastic optimization model which is, in our view, particularly well-suited for an abundance of possible timetabling problems, and present evidence that it possesses several attractive characteristics. In particular, we demonstrate that our model may be simultaneously perceived of as a deterministic linear program (LP), a stochastic linear program (SP) and a closed-form convex non-linear program (NLP), thereby building bridges between this modeling approach and the broader OR community and opening up

a whole new range of possibilities for efficient solutions algorithms. In contrast, only a simple LP solution approach is presented e.g. in [16] and [30], which is rather impractical for complex large-scale models due to computational issues, as we shall demonstrate in a case study. Additionally, we examine the role of the total amount of supplements, and introduce stochastic dependency among the random variables of the model, demonstrating numerical evidence that the simplicity gained by the conventional independence assumption may be very costly in terms of delays in the network. Finally, we provide a statistical estimation technique for both the independent and dependent cases in order to estimate the unknown parameters of the random components of the model using observable statistical data.

2 A general timetable optimization model

2.1 Description

First we present the assumptions of our model informally, to be reformulated into a formal mathematical model later on in our paper:

- The model contains an arbitrary number of *railway instances* (or simply *instances*), which represent scheduled railway connections between pairs of origins and destinations within a day. Several modeled instances may share the same origins and destinations.
- Any instance is scheduled to complete a number of *events* that may include both departures and arrivals at stations or other in-between locations. The model specifies every event time as the number of minutes elapsed since 12 AM (midnight) of the modeled day until the given event. Any event in the model has its own *planned event time* as well as an *actual event time*, where the latter depends on a number of uncertain factors and parameters.

For example, a first instance may be a connection from Haarlem to Maastricht by changing trains halfway in Utrecht, another instance may be the subsequent connection in the reverse direction of the same line and a third instance may be a connection from Alkmaar to Hilversum intersecting the line of the first two instances halfway in Amsterdam.

- Pairs of instances may interfere with one another at given *interferences*, where a given instance is scheduled to give priority to another instance, and a prescribed *safety margin* is to be respected both in the timetable and the realizations. The order of passing through of the instances at the interferences is fixed in advance as our model contains *no traffic control decisions* by design. Interferences may represent crossings of lines as well as connections for passengers and rolling stock.
- The *initial departure times* of all instances are assumed to be given. Additionally, any event in the model is associated with its own given (minimally) *feasible event-to-event time*, which is an optimistic estimate of the time elapsed between the previous and current events, to be expected in the absence of any disturbance as determined chiefly by technical parameters (acceleration, speed restrictions etc.), as well as a *supplement*, which increases the corresponding scheduled event-to-event time in order to reduce delays. Planned event times are jointly determined by the initial departure times, the feasible event-to-event times and the supplements.
- The sum of supplements for any instance is bounded from above by a given *quantity of distributable supplements* in order to keep planned travel times reasonably short. An arbitrary number of *additional restrictions* in the form of linear inequalities may be placed on the supplements of the model. Supplements may not be negative, as the events in question would inevitably be delayed otherwise.
- We assume that an instance incurs a random *disturbance* between any consecutive pair of its events. Following the SP paradigm, we assume that random events may be modeled by random variables with a known joint distribution. Actual event times are determined by initial departure times, feasible event-to-event times, disturbances, supplements and possible waiting times at predefined interferences.
- *Delays* are obtained as differences between actual and planned event times and may not be negative. The objective of the model is to minimize the *expected weighted mean delay*, where the *weights of the events delays* are given as parameters.

2.2 Notation

The model requires the following *model dimensions* to be specified a priori:

$m \in \mathbb{N}_+$: number of railway instances,

$n_i \in \mathbb{N}_+$: number of events of instance i ($i = 1, 2, \dots, m$),

$o \in \mathbb{N}$: number of interferences, to be detailed later,

$p \in \mathbb{N}$: number of additional restrictions, to be detailed later.

The derived parameter $n = \sum_{i=1}^m n_i$ will denote the number of events in the model.

We introduce the following *indices* and ordered *index sets*:

$i \in I = \{1, 2, \dots, m\}$: instance index,

$j \in J_i = \{1, 2, \dots, n_i\}$: event index for instance i ,

$(i, j) \in J = \cup_{i=1}^m (\{i\} \times J_i)$: combined instance and event index,

$k \in K = \{1, 2, \dots, o\}$: interference index,

$\ell \in L = \{1, 2, \dots, p\}$: additional restrictions index,

and for the sake of brevity, we will implicitly assume throughout this subsection that these indices fall into the index sets specified above.

For $i \in I$, the following *instance parameters* are assumed to be given a priori:

$t_{i,0} \geq 0$: initial departure times in the original timetable,

$q_i > 0$: quantities of distributable supplements,

For $j \in J_i$, the same holds for the corresponding following *event parameters*:

$f_{i,j} \geq 0$: feasible event-to-event times,

$w_{i,j} \geq 0$: weights of the delays.

We introduce the *normalized delay weights* $\tilde{w}_{i,j} = \frac{w_{i,j}}{\sum_{g=1}^m \sum_{h=1}^{n_i} w_{g,h}}$ and their row vector

$\mathbf{w} = (\tilde{w}_{1,1}, \tilde{w}_{1,2}, \dots, \tilde{w}_{1,n_1}, \tilde{w}_{2,1}, \tilde{w}_{2,2}, \dots, \tilde{w}_{2,n_2}, \dots, \tilde{w}_{m,1}, \tilde{w}_{m,2}, \dots, \tilde{w}_{m,n_m}) \in \mathbb{R}^n$

in order to simplify the notation of the weighted mean delay later on.

Any interference comprises two instances; for each of these, the interference corresponds to a particular event. A corresponding safety margin is necessary, as well, as described earlier.

Thus, for any $k \in K$, we specify the following *interference parameters* a priori:

$i'_k, i_k \in I$: indices of the first and second instances in their order of passing through,

$j'_k, j_k \in J$: associated event indices of the first and second instances in the same order,

$s_k > 0$: safety margins.

The *random variables* $\omega_{i,j}$ represent disturbances in the model. We introduce the vector

$\boldsymbol{\omega} = (\omega_{1,1}, \omega_{1,2}, \dots, \omega_{1,n_1}, \omega_{2,1}, \omega_{2,2}, \dots, \omega_{2,n_2}, \dots, \omega_{m,1}, \omega_{m,2}, \dots, \omega_{m,n_m}) \in \mathbb{R}_+^n$ for simplicity.

The values of the following *decision variables* will be obtained via optimization:

$x_{i,j}$: supplements,

$y_{i,j}$: delays.

We introduce the following column vectors in $\mathbb{R}^n$ in order to keep the notation simple:

$\mathbf{x} = (x_{1,1}, x_{1,2}, \dots, x_{1,n_1}, x_{2,1}, x_{2,2}, \dots, x_{2,n_2}, \dots, x_{m,1}, x_{m,2}, \dots, x_{m,n_m})^T$,

$\mathbf{y} = (y_{1,1}, y_{1,2}, \dots, y_{1,n_1}, y_{2,1}, y_{2,2}, \dots, y_{2,n_2}, \dots, y_{m,1}, y_{m,2}, \dots, y_{m,n_m})^T$.

The values of the following *endogenous variables*, which are included in the model only for the sake of notational simplicity, will be determined by the values of the decision variables and some of the parameters:

$t_{i,j}$: planned event times,

$u_{i,j}$: actual event times.

The following *parameters of the additional restrictions* are given a priori:

$a_{i,j}^\ell \in \mathbb{R}$: coefficient of $x_{i,j}$ in additional restriction ℓ ,

$b^\ell \in \mathbb{R}$: right-hand side constant in additional restriction ℓ .

The following *model objectives*, to be detailed later, will be obtained through optimization:

$\Delta(\mathbf{x}, \boldsymbol{\omega})$: minimum weighted mean delay,

D : minimum expected weighted mean delay.

We shall introduce additional symbols later on in the paper.

2.3 Stochastic programming (SP) formulation

We now formulate the proposed general timetable optimization model as a stochastic linear two-stage recourse model, which is a well-known model class in the literature of stochastic programming. Two-stage recourse models may be written concisely in the following form:

$$\min_{\mathbf{x} \geq \mathbf{0}} \{ \mathbf{c}\mathbf{x} + \mathbb{E}_{\omega} [Q(\mathbf{x}, \omega)] : \mathbf{A}\mathbf{x} = \mathbf{b} \}, \text{ where} \quad (1)$$

$$Q(\mathbf{x}, \omega) = \min_{\mathbf{y} \geq \mathbf{0}} \{ \mathbf{q}\mathbf{y} : \mathbf{T}\mathbf{x} + \mathbf{W}\mathbf{y} = \omega \}.$$

In our model, the first stage will represent a decision made by the planners of the timetable, who must take timetabling constraints into account while they attempt to find an optimal distribution of supplements that results in the lowest possible expected weighted mean delay, and the second stage will describe the decisions of the drivers and railway operators, who must take traffic circumstances into account while aiming to respect the given timetable as precisely as possible. We shall construct a model of this form in a step-by-step manner.

2.3.1 First-stage constraints

The optimal values of the first-stage decision variables $x_{i,j}$ and $t_{i,j}$ ($i \in I, j \in J_i$), which do not depend on the actual realizations of the random variables $\omega_{i,j}$ ($i \in I, j \in J_i$), are determined in this stage while taking into account the following constraints:

- Any planned event time $t_{i,j}$ is the sum of the previous planned event time $t_{i,j-1}$, the corresponding feasible event-to-event time $f_{i,j}$ and the supplement $x_{i,j}$:

$$t_{i,j} = t_{i,j-1} + f_{i,j} + x_{i,j} \quad (i \in I, j \in J_i). \quad (2)$$

- The sums of supplements per instance are bounded from above by the constants q_i :

$$\sum_{j=1}^{n_i} x_{i,j} \leq q_i \quad (i \in I). \quad (3)$$

- The difference between the planned event times of the two instances involved at an

More precisely, this is the canonical representation of a very commonly used special case of two-stage recourse models with deterministic arrays $\mathbf{q}$ and $\mathbf{W}$. All arrays are assumed to be of conformable dimensions. We shall not follow the traditional notational convention of the names of the arrays in this formulation.

interference may not be less than the prescribed safety margin s_k :

$$t_{i_k, j_k} - t_{i'_k, j'_k} \geq s_k \quad (k \in K). \quad (4)$$

- Additional restrictions may be written as:

$$\sum_{i=1}^m \sum_{j=1}^{n_i} a_{i,j}^\ell x_{i,j} \leq b^\ell \quad (\ell \in L). \quad (5)$$

- Supplements are non-negative:

$$x_{i,j} \geq 0 \quad (i \in I, j \in J_i).$$

2.3.2 Second-stage constraints

The optimal values of the second-stage decision variables $y_{i,j}$ and $u_{i,j}$ ($i \in I, j \in J_i$), which *do depend* on the actual realizations of the random variables $\omega_{i,j}$ ($i \in I, j \in J_i$) (and also on the values of the first-stage decision variables), are determined in this stage while considering the following constraints:

- Any actual event time $u_{i,j}$ is *at least* the sum of the previous actual event time $u_{i,j-1}$, the corresponding feasible event-to-event time $f_{i,j}$ and the disturbance $\omega_{i,j}$ (and it can be even larger due to waiting for a possible interfering instance):

$$u_{i,j} \geq u_{i,j-1} + f_{i,j} + \omega_{i,j} \quad (i \in I, j \in J_i, u_{i,0} \equiv t_{i,0}). \quad (6)$$

- Delays $y_{i,j}$ are the differences of the corresponding actual and planned event times:

$$y_{i,j} = u_{i,j} - t_{i,j} \quad (i \in I, j \in J_i). \quad (7)$$

- Safety margins must be respected by the realized event times just like in the timetable:

$$u_{i_k, j_k} - u_{i'_k, j'_k} \geq s_k \quad (k \in K). \quad (8)$$

- Delays are non-negative:

$$y_{i,j} \geq 0 \quad (i \in I, j \in J_i). \quad (9)$$

2.3.3 A two-stage model with endogenous variables

The model proceeds as follows: the second stage minimizes the weighted mean delay for any given combination of supplements and disturbances while considering the second-stage

constraints of the model, whereas the first stage minimizes the expectation of the minimal weighted mean delay obtained in the second stage by determining the optimal supplements over the set of first-stage constraints, which may be written concisely as

$$D = \min_{\mathbf{x}, \mathbf{t} \geq \mathbf{0}} \{ \mathbb{E}_{\boldsymbol{\omega}} [\Delta(\mathbf{x}, \boldsymbol{\omega})] : \text{constraints (2)–(5)} \}, \text{ where} \quad (10)$$

$$\Delta(\mathbf{x}, \boldsymbol{\omega}) = \min_{\mathbf{y}, \mathbf{u} \geq \mathbf{0}} \{ \mathbf{w}\mathbf{y} : \text{constraints (6)–(8)} \}.$$

Here we introduced the vectors

$$\mathbf{t} = (t_{1,1}, t_{1,2}, \dots, t_{1,n_1}, t_{2,1}, t_{2,2}, \dots, t_{2,n_2}, \dots, t_{m,1}, t_{m,2}, \dots, t_{m,n_m})^T \text{ and}$$

$$\mathbf{u} = (u_{1,1}, u_{1,2}, \dots, u_{1,n_1}, u_{2,1}, u_{2,2}, \dots, u_{2,n_2}, \dots, u_{m,1}, u_{m,2}, \dots, u_{m,n_m})^T.$$

The variables $t_{i,j}$ and $u_{i,j}$ ($i \in I, j \in J_i$) are endogenous and thus redundant in the sense that they are uniquely determined by the vectors $\mathbf{x}$ and $\mathbf{y}$ due to equations (2) and (7).

2.3.4 Reduced form

In this subsection, we eliminate all endogenous variables from the model in order to arrive at a more tractable, albeit less elegant formulation, which will turn out to be useful later on.

As a first step, we solve equation (2) for the planned event times:

$$t_{i,j} = t_{i,0} + \sum_{h=1}^j f_{i,h} + \sum_{h=1}^j x_{i,h} \quad (i \in I, j \in J_i). \quad (11)$$

Subsequently, we introduce the linear function $d_k(\mathbf{x})$ for the *planned time difference between the interfering instances* at interference k and substitute equation (11) in order to eliminate all planned event times from the first-stage interference constraints:

$$\begin{aligned} d_k(\mathbf{x}) &= t_{i_k, j_k} - t_{i'_k, j'_k} \\ &= t_{i_k, 0} + \sum_{h=1}^{j_k} f_{i_k, h} + \sum_{h=1}^{j_k} x_{i_k, h} - t_{i'_k, 0} - \sum_{h=1}^{j'_k} f_{i'_k, h} - \sum_{h=1}^{j'_k} x_{i'_k, h} \quad (k \in K). \end{aligned} \quad (12)$$

Using this notation, we reformulate inequality (4) in the straightforwardly interpretable form

$$d_k(\mathbf{x}) \geq s_k \quad (k \in K). \quad (13)$$

Subtracting $t_{i,j}$ from both sides of inequality (6) and substituting equation (2) yields

$$u_{i,j} - t_{i,j} \geq u_{i,j-1} - t_{i,j-1} + f_{i,j} + \omega_{i,j} - (t_{i,j} - t_{i,j-1}) = u_{i,j-1} - t_{i,j-1} + f_{i,j} + \omega_{i,j} - (f_{i,j} + x_{i,j}).$$

By considering equation (7), we arrive at the following reformulation of inequality (6):

$$y_{i,j} \geq y_{i,j-1} + \omega_{i,j} - x_{i,j} \quad (i \in I, j \in J_i, y_{i,0} \equiv 0). \quad (14)$$

Rearranging and subtracting t_{i_k,j_k} from both sides of inequality (8) yields $u_{i_k,j_k} - t_{i_k,j_k} \geq u_{i'_k,j'_k} - t_{i'_k,j'_k} - (t_{i_k,j_k} - t_{i'_k,j'_k}) + s_k$. By noting equations (7) and (12), we arrive at

$$y_{i_k,j_k} \geq y_{i'_k,j'_k} - d_k(\mathbf{x}) + s_k \quad (k \in K). \quad (15)$$

Using the reduced constraints, we obtain the following concise SP formulation of the model:

$$D = \min_{\mathbf{x} \geq \mathbf{0}} \{ \mathbb{E}_{\omega} [\Delta(\mathbf{x}, \omega)] : \text{constraints (3), (5) and (13)} \}, \text{ where} \quad (16)$$

$$\Delta(\mathbf{x}, \omega) = \min_{\mathbf{y} \geq \mathbf{0}} \{ \mathbf{w}\mathbf{y} : \text{constraints (14) and (15)} \}.$$

This model of the form (1) contains $m+o+p$ and $n+o$ first and second-stage linear constraints, respectively, along with n first-stage as well as n second-stage decision variables. In comparison, the extended form contains $n+m+o+p$ and $2n+o$ first and second-stage linear constraints, respectively, along with $2n$ first-stage and $2n$ second-stage decision variables.

2.4 Determining the optimal delays in the second stage

As we shall point out in Section 2.5, problem (16) always has an optimal solution under a few reasonable assumptions. To demonstrate this, we first proceed by proving an intermediate result, namely that the second stage of problem (16) always has a closed-form optimal solution under two very mild assumptions on the predefined interferences. The two assumptions in question may be summarized informally as follows:

- Out of the two instances involved, neither instance must give priority to more than one other instance at an interference.
- The interferences are not circular so that no deadlock may occur.

These rather technical restrictions are very natural and are without loss of generality, as we shall point out in Section 2.6, where we shall discuss our model assumptions in detail.

Without loss of generality, we have been assuming since Section 2.2 that every interference only involves two instances. If, in reality, three instances are involved at an interference, the situation may be modeled as two interferences each affecting a pair of instances due to the assumption of the fixed order of trains.

To motivate the concept of a closed-form solution of the second-stage problem, which will turn out to be practical later, we note that if the disturbances ω are known, it is obviously optimal to respond to them in a minimalist fashion: the operator does not want more delays than it is strictly necessary due to the safety margins arising at the interferences, and obviously, no event should take place earlier than it is listed in the schedule. As real-time disturbances are encountered, the corresponding real-time "natural" delays can be recorded sequentially. This is possible, as interferences have a fixed order of instances and are non-circular, as we assume. It is therefore not surprising that explicit formulas can be generated for these "natural" delays. Nor is it surprising that these "natural" delays are optimal in the second-stage minimization problem for every set of non-negative weights $\mathbf{w}$.

In order to proceed with the outlined closed-form solution of the second stage, we now formalize the two restrictions listed above by defining the concept of *well-structured interferences*:

Definition 1. *Problem (16) contains well-structured interferences if the following is true:*

- (a) $\nexists k_1, k_2 \in K : k_1 \neq k_2$ and $(i_{k_1}, j_{k_1}) = (i_{k_2}, j_{k_2})$, furthermore
- (b) $\nexists k_1, k_2, \dots, k_c \in K : (i_{k_{h+1}}, j_{k_{h+1}}) = (i'_{k_h}, j'_{k_h}) (h = 1, 2, \dots, c-1)$ and $(i_{k_1}, j_{k_1}) = (i'_{k_c}, j'_{k_c})$.

Subsequently, we combine inequalities (9) and (14) yielding

$$y_{i,j} \geq \max(0, y_{i,j-1} + \omega_{i,j} - x_{i,j}) \quad (i \in I, j \in J_i, y_{i,0} \equiv 0). \quad (17)$$

In order to extend the latter inequality with relevant interference constraints, we define a function that identifies the index of the interference where instance i is scheduled to wait before completing its event j :

$$K(i, j) = \{k \in K : i_k = i, j_k = j\}.$$

Assuming well-structured interferences, this function may either return a single value or the empty set. Defining $y_{i_{\emptyset,1}, j_{\emptyset,1}} \equiv d_{\emptyset}(\mathbf{x}) \equiv s_{\emptyset} \equiv 0$ and considering inequalities (9) and (15), we gain an additional lower bound:

$$y_{i,j} \geq y'_{i_{K(i,j)}, j_{K(i,j)}} - d_{K(i,j)}(\mathbf{x}) + s_{K(i,j)} \quad (i \in I, j \in J_i), \quad (18)$$

where we use the function $d_k(\mathbf{x})$ defined by equation (12). This constraint describes possible *secondary delays* that may occur due to interferences, as a delay of the second instance may also delay the first instance that is scheduled to wait at the interference.

We now define a new function that returns the maximum of the right-hand sides of inequalities (17) and (18) in order to merge the two existing lower bounds:

$$L_{i,j}(\mathbf{y}, \mathbf{x}, \boldsymbol{\omega}) = \max(0, y_{i,j-1} + \omega_{i,j} - x_{i,j}, y_{i'_{K(i,j)}, j'_{K(i,j)}} - d_{K(i,j)}(\mathbf{x}) + s_{K(i,j)}) \quad (i \in I, j \in J_i), \quad (19)$$

which provides us with the following combined lower bound:

$$y_{i,j} \geq L_{i,j}(\mathbf{y}, \mathbf{x}, \boldsymbol{\omega}) \quad (i \in I, j \in J_i). \quad (20)$$

We argue that there is always a unique delay vector $\mathbf{y}$ for which the latter holds with equality for all events if the interferences are well-structured, and it may be written as a closed-form function of the given supplements and disturbances:

Theorem 1. *If problem (16) contains well-structured interferences then $\exists! \mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega}) \in \mathbb{R}^n$: $y_{i,j}^*(\mathbf{x}, \boldsymbol{\omega}) = L_{i,j}(\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega}), \mathbf{x}, \boldsymbol{\omega})$ ($i \in I, j \in J_i$). Moreover, $\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega})$ may be expressed in closed form in a recursive way.*

Proof. We sort the set J of all events in a sequence R_h ($h = 1, 2, \dots, n$) successively such that any event is preceded in the list by its immediate predecessor as well as any interfering event that has priority, which is necessary as the delays must be determined sequentially due to the recursive nature of inequality (20). The following is a possible appropriate sequence, as we shall demonstrate:

$$R_h = \min_i \min_j \{(i, j) \in J: \exists g < h : R_g = (i'_{K(i,j)}, j'_{K(i,j)}), \nexists g < h : R_g = (i, j)\} \quad (h = 1, 2, \dots, n, R_0 \equiv (0, 0)). \quad (21)$$

Since the interferences are well-structured, an event may not have more than one associated interfering event with priority, which implies that the number of interferences may not be greater than the number of instances, or formally $o \leq n$. If $o = n$ were true then the interferences would be circular, which is impossible due to the assumption of well-structuredness, so we conclude that $o < n$ so that there is at least one event without an interfering event, and by construction, the first item of the list must be one of these. After removing the first event of the list from J , if the remaining list $J \setminus R_1$ contains no event without an interfering instance with priority then its interfering event with priority must be R_1 , otherwise the remaining list of $n-1$ events with $n-1$ interferences would be circular. Thus R_2 exists and is an event different from R_1 . By repeating the same argument with the remaining list $J \setminus (R_1 \cup R_2)$, we

conclude that R_3 must also be either an event without an interfering event with priority or an event that has either of the interfering events R_1 or R_2 with priority. Repeating the same argument $n-1$ times, we conclude by induction that our list contains every event exactly once. Any event (i, j) is trivially preceded in this list by the previous event $(i, j-1)$ if the latter exists due to the minimization involved with respect to j , which implies that the first item of the list has $j = 1$, and due to $K(R_1) = \emptyset$, $y_{R_1}^*(\mathbf{x}, \boldsymbol{\omega}) = \max(0, \omega_{R_1} - x_{R_1})$, which is a closed-form expression of $\mathbf{x}$ and $\boldsymbol{\omega}$. By construction, any event (i, j) is trivially preceded also by a possible interfering event with priority $(i'_{K(i,j)}, j'_{K(i,j)})$, so $y_{R_2}^*(\mathbf{x}, \boldsymbol{\omega}) = L_{R_2}(\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega}), \mathbf{x}, \boldsymbol{\omega})$ is uniquely determined in closed form, and by induction, we may determine $y_{R_h}^*(\mathbf{x}, \boldsymbol{\omega})$ ($h = 3, 4, \dots, n$) in closed form in a sequential manner. $\square$

On top of this, we argue that the vector $\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega})$ defined by the previous theorem is always an optimal solution of the second stage of problem (16):

Theorem 2. *If problem (16) contains well-structured interferences then*

$$\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega}) \in \arg \min_{\mathbf{y} \geq \mathbf{0}} \{\mathbf{w}\mathbf{y} : \text{constraints (14) and (15)}\}.$$

Proof. The vector $\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega})$ is a feasible second-stage solution as it obviously satisfies $\mathbf{y} \geq \mathbf{0}$ as well as constraints (14) and (15). Assume that $\mathbf{y}'$ is another feasible solution with a better objective function value: $\mathbf{w}\mathbf{y}' < \mathbf{w}\mathbf{y}_{i,j}^*(\mathbf{x}, \boldsymbol{\omega})$. Then $y'_{i,j} < y_{i,j}^*(\mathbf{x}, \boldsymbol{\omega})$ for a specific $i \in I$ and $j \in J_i$ is impossible as it would contradict the feasibility of $\mathbf{y}'$ by violating inequality (20). Thus we conclude that $\mathbf{y}' \geq \mathbf{y}_{i,j}^*(\mathbf{x}, \boldsymbol{\omega})$, implying $\mathbf{w}\mathbf{y}' \geq \mathbf{w}\mathbf{y}_{i,j}^*(\mathbf{x}, \boldsymbol{\omega})$ due to $\mathbf{w} \geq \mathbf{0}$, which is a contradiction. $\square$

2.5 The existence of an optimal solution

The following theorem states that assuming a few very mild technical conditions, our model always has an optimal solution.

Theorem 3. *If problem (16) contains well-structured interferences, $\mathbb{E}(\boldsymbol{\omega}) < \infty$ and the feasible set of its first stage is non-empty then problem (16) has an optimal solution.*

Proof. We showed in the previous section that a finite value of $\Delta^*(\mathbf{x}, \boldsymbol{\omega})$ may be determined explicitly for any $\mathbf{x}$ and $\boldsymbol{\omega}$ under well-structured interferences. As $\Delta^*(\mathbf{x}, \boldsymbol{\omega})$ is the optimal value function of an LP, it follows due to the finiteness of the mean of the vector $\boldsymbol{\omega}$ that $D(\mathbf{x}) := \mathbb{E}[\Delta^*(\mathbf{x}, \boldsymbol{\omega})]$ is finite and convex hence continuous. Because of inequality (3) the feasible set is compact and it is assumed to be non-empty. Hence it follows that problem (16) has at least one optimal solution. $\square$

2.6 Discussion of the model assumptions

We now turn our attention to the discussion of the (sometimes implicit) assumptions of our model, arguing that the simplicity of our formulation comes with little loss of generality, and it may easily be modified to accommodate additional requirements imposed by real-life practice. We proceed one by one with a list of relevant assumptions: I

- *No integrality constraints:* Even though event times are published as integers in timetables, both planned and actual event times involve fractions of minutes in reality, thus it is reasonable not to impose any integrality restrictions on either of these in the model.
- *Non-negative delays:* The assumption that initial departures take place on time is only a technical one, which may be relaxed by introducing an additional departure event with a feasible time of 0 minutes and an associated disturbance, supplement and delay, as we shall do in Section 4. The assumption of no negative delays later on means that instances are not permitted to complete any event before time, and if they accidentally do so then they are required to wait until the planned time of the given event, which is desirable for passengers, who would not like to miss their connections due to premature departures. This may easily be omitted for certain events whenever necessary.
- *Interferences:* The fixed order of passing through at the interferences represents the lack of traffic control decisions in the model, which is reasonable as we wish to improve the initially given timetable with respect to small disturbances and delays, and it lies outside the scope of our model to deal with cases where the whole traffic needs to be

rescheduled due to a major anomaly (e.g. the breakdown of an engine resulting in a one-hour delay of the given train as well as significant delays on other lines). The first part of the well-structuredness of the interferences introduced by definition implies that if an instance is scheduled to wait for two or more other instances at a given event then the interference in question should be split up into two or more separate interferences on the list, while the second part of the definition is very natural as deadlocks would be unreasonable and would completely incapacitate the network.

- *Known distributions:* The joint probability distribution of the disturbances may be estimated using measured delay data, which are meticulously collected by railway companies in practice. We shall elaborate on this problem in detail in Section 5.
- *Weighting of delays:* Delays may be arbitrarily weighted in the model: being more important for passengers, arrival events may have greater weights than departure events (as later in Section 4), and additionally, the importance of the individual stations may also be accounted for (e.g. based on ticket sales statistics).
- *Cyclicity and symmetry:* Several railway companies use so-called *cyclic timetables* where planned event times of consecutive connections on a given line are shifted by a common cycle time (e.g. half an hour or one hour) in relation to one another. Cyclic timetable planning is described in detail and concise mathematical formulations are given in the literature (see e.g. [14], [15], [16], [21], [22]). Even though cyclicity is not the particular focus of our model, this type of constraints may easily be incorporated in it by using the additional restrictions introduced in Section 2.2, since planned event times are determined by the supplements according to equation (11). The *principle of symmetry* means that running times on identical line segments should be identical whenever possible, which may likewise be modeled by simple additional restrictions.
- *Objective functions:* Instead of minimizing expected delays, most railway companies focus on maximizing the closely related *punctuality*, which is usually defined as the

percentage of delays falling below a given threshold (or alternatively, *unpunctuality*, which equals one minus the punctuality statistic). This formulation leads to models based on chance constraints or integrated chance constraints (see e.g. [11] and [12]). Another solution is to use progressively increasing penalty factors for delays falling into different intervals. Following [16], we minimize expected delays, while the punctuality of a solution may easily be calculated from the second-stage solution of model (16).

- *Two-stage formulation:* Due to the sequential nature of real-life delays, which is mirrored by the sequential approach of Theorem 1, no delay depends on subsequent disturbances in the list introduced in the proof of the theorem. Otherwise it would be necessary to construct a more complicated, multi-stage recourse model (see e.g. [1], [10], [13] or [23]).
- *Independence of disturbances from past delays:* Our model implicitly assumes that disturbances do not depend on previously accumulated delays, even though it could be reasonable to assume an increased speed in case of a delay, within reasonable limits. This phenomenon is a possible topic of future research.

The same model may be applied naturally to any other methods of public transportation involving intermediate stops and interferences (e.g. buses, underground railways etc.). More generally, our model may be applied to any timing problem involving the organization of one or more series of possibly interfering, repetitive tasks with random completion times.

3 Alternative solution frameworks

An investigation of the computational aspects of our model presented in Section 2 is highly relevant due to the high complexity of real-life railway networks, which necessitate the use of fast and reliable algorithms in practical applications. As we shall show in this section, problem (16) lies in the intersection of deterministic linear, stochastic linear and convex

For example, a 3-minute threshold is used by the Dutch Railways, and the company sets a target value of this statistic at a national level.

non-linear programming, which provide an abundance of viable solution algorithms. Henceforth, we shall assume $\boldsymbol{\omega}$ to be discrete with the finite set of r distinct possible values $\Omega = \{\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_r\}$, where $\mathbb{P}(\boldsymbol{\omega} \notin \Omega) = 0$.

3.1 Linear programming (LP)

Since problem (16) is an ordinary LP for a fixed value of the random vector $\boldsymbol{\omega}$, it follows by introducing separate delay vectors $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(r)} \in \mathbb{R}^n$ for every realization, writing the first-stage objective function as the probability-weighted sum of the optimal weighted mean delays and merging the two stages that problem (16) is equivalent to the following large-scale deterministic LP:

$$D = \min_{\mathbf{x}, \mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \dots, \mathbf{y}^{(r)} \geq \mathbf{0}} \left(\sum_{h=1}^r \mathbb{P}(\boldsymbol{\omega} = \boldsymbol{\omega}^{(h)}) \mathbf{w} \mathbf{y}^{(h)} : \text{constraints (3), (5), (13), (14), (15)} \right). \quad (22)$$

The resulting LP has $(r+1)n$ decision variables and $d(n+p+2q)$ linear constraints (excluding non-negativity constraints of the decision variables), which may result in enormous dimensions due to the combined effects of a large number of realizations and events even in the case of a small-scale practical application. Even though the deterministic equivalent LP problem may be solved by any LP solution algorithm in principle, this approach requires brute force while ignoring the special structure of the problem and the availability of a closed-form expression for the delay vector.

3.2 Stochastic programming (SP)

There are several advanced SP solution algorithms that benefit from the special structure of two-stage linear recourse models instead of applying LP on the large-scale deterministic equivalent (see e.g. [19]). As we shall demonstrate later, this approach may be substantially more efficient due to being better adapted to this specific model type. A particularly effi-

Even though some SP solution techniques can tackle the case of an infinite set of distinct possible values, these algorithms are not efficient for large-scale problems in practice, and these cases may be approximated by the less general case described here, thus we are not considering the more general case in this paper.

cient algorithm of this type is the *Regularized Decomposition* (RD) algorithm described in [26] and implemented in the SP model management environment SLP-IOR (see [9]). Regularized Decomposition is a variant of the *L-shaped* algorithm described in [29], which iteratively constructs linear approximations of the first-stage objective function based on its subgradient, which may be computed from the dual optimal solution of the second stage. It has been proved that it yields an optimal solution of problem (1) in finitely many iterations, assuming that there is one. We remark that there are algorithms for continuous distributions (see e.g. [19]), as well, but we have decided against their use due to their relative computational inefficiency.

3.3 Convex non-linear programming (NLP)

We have shown in Section 2.4 how to construct a closed-form expression for the optimal delay vector $\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega})$ as a function of the supplements and the random disturbances. Hence we may eliminate the second stage of problem (16), replace the delay vector by its optimal value and reformulate the remaining first stage as follows:

$$D = \min_{\mathbf{x} \geq \mathbf{0}} \left(\sum_{h=1}^r \mathbb{P}(\boldsymbol{\omega} = \boldsymbol{\omega}^{(h)}) \mathbf{w} \mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega}^{(h)}) : \text{constraints (3), (5), (13)} \right). \quad (23)$$

Due to equation (19) and Theorem 1, the functions $\mathbf{y}^*(\mathbf{x}, \boldsymbol{\omega}^{(h)})$ ($h = 1, 2, \dots, r$) arise as pointwise maxima of functions that are linear in $\mathbf{x}$, so they are themselves convex in $\mathbf{x}$, and so is the objective function of problem (23). Hence this framework amounts to the minimization of a closed-form convex function over a convex polyhedral feasible set, for which an abundance of widely used convex programming algorithms and software solutions are available (see e.g. [2] or [27]).

4 A real-life application with numerical results

In order to test the methods described in this paper on a real-life application, we constructed a model with $m = 7$ instances based on actual timetables of the Haarlem-Maastricht (Haa-Maa) and Alkmaar-Hilversum (Alk-Hil) lines of the Dutch Railways and some data provided

by the railway company, which we extended by a few additional assumptions. The two modeled lines intersect each other in Amsterdam (Ams) at the stations Amsterdam Centraal (Ams-C) and Amsterdam Sloterdijk (Ams-S). Figure 1 is a schematic illustration of the two intersecting lines. The case study contains $n = 67$ events in total. The parameters of the instances (event counts, initial departure times as minutes after midnight, quantities of distributable supplements and feasible event-to-event times in minutes) are summarized in Table 1.



Figure 1: Map of the Haarlem-Maastricht and Alkmaar-Hilversum lines, intersecting in Amsterdam

The model contains $q = 7$ interferences for scheduled passenger transfers at these stations as well as rolling stock connections between the endpoint of an instance and the starting point of the following instance in the reverse direction of the same line. The parameters of the interferences (their instance and event indices and prescribed safety margins in minutes) are

Source of image: ©User: Lencer / Wikimedia Commons / CC-BY-SA-3.0, with additional modifications.

Some pairs of oppositely directed instances on identical lines have different event counts as we modeled both express and local trains. Instead of modeling both departures and arrivals at all stops, we only modeled one event per stop (only departures except for arrivals at final stations).

i	Endpoints	n_i	$t_{i,0}$	q_i	$f_{i,1}$	$f_{i,2}$	$f_{i,3}$	$f_{i,4}$	$f_{i,5}$	$f_{i,6}$	$f_{i,7}$	$f_{i,8}$	$f_{i,9}$	$f_{i,10}$	$f_{i,11}$	$f_{i,12}$	$f_{i,13}$	$f_{i,14}$
1	Haa-Maa	14	42	11.12	0	2.8	3.74	4.67	5.61	3.74	7.48	21.5	28.97	21.5	15.89	12.15	17.76	13.08
2	Maa-Haa	12	235	10.73	0	14.95	14.02	12.15	18.69	19.63	29.91	16.82	8.41	4.67	5.61	8.41	–	–
3	Alk-Hil	13	0	5.10	0	3.74	4.67	4.67	3.74	2.8	2.8	1.87	4.67	5.61	5.61	14.02	18.69	–
4	Hil-Alk	7	95	4.58	0	20.56	11.21	5.61	6.54	11.21	10.28	–	–	–	–	–	–	–
5	Alk-Hil	7	189	4.51	0	7.48	14.02	6.54	5.61	12.15	18.69	–	–	–	–	–	–	–
6	Hil-Alk	7	275	4.58	0	20.56	11.21	5.61	6.54	11.21	10.28	–	–	–	–	–	–	–
7	Alk-Hil	7	369	4.51	0	7.48	14.02	6.54	3.74	14.02	18.69	–	–	–	–	–	–	–

Table 1: Parameters of the instances

summarized in Table 2.

k	Station	i'_k	j'_k	i_k	j_k	s_k
1	Ams-C	3	11	1	5	5
2	Ams-S	2	11	7	5	5
3	Maa	1	14	2	1	5
4	Hil	3	13	4	1	5
5	Alk	4	7	5	1	5
6	Hil	5	7	6	1	5
7	Alk	6	7	7	1	5

Table 2: Parameters of the interferences

There are no additional restrictions in the model ($r = 0$). Before discretization, we assumed the disturbances to be independent right-truncated exponential random variables with original means $\mu_{i,j} = \mathbb{E}(\tilde{\omega}_{i,j})$ given by $\mu_{i,1} = 0.2$, $\mu_{i,j} = 0.07f_{i,j}$ ($j > 1$) and a truncation threshold of ten minutes: $\omega_{i,j} = \min(10, \tilde{\omega}_{i,j})$ ($i \in I, j \in J_i$). We assigned (non-normalized) delay weights of zero to initial departures, two to intermediate departures with transfer opportunities (interferences) as well as to final arrivals and one to other intermediate departures, as initial departure delays are largely irrelevant, and the punctuality of transfers and final arrivals has a key importance.

In this example, the sequence of events R_h ($h = 1, 2, \dots, n$) starts as

$$R_1 = (1, 1), R_2 = (1, 2), R_3 = (1, 3), R_4 = (1, 4), R_5 = (3, 1), R_6 = (3, 2) \text{ etc.,}$$

We include truncation in the model following [16] as delays could explode otherwise, whereas the aim of the model is to increase the robustness of timetables against regular relatively small disturbances.

where a switch to the third instance occurs in the fifth term due to the first interference, which prescribes that the fifth event of the first instance has to wait for the eleventh event of the third instance before moving on. For the sake of illustration, we include here some associated delay formulas in recursive form:

$$\begin{aligned}
y_{1,1}^*(\mathbf{x}, \boldsymbol{\omega}) &= (\omega_{1,1} - x_{1,1})^+, \\
y_{1,2}^*(\mathbf{x}, \boldsymbol{\omega}) &= (y_{1,1}^*(\mathbf{x}, \boldsymbol{\omega}) + \omega_{1,2} - x_{1,2})^+, \\
y_{1,3}^*(\mathbf{x}, \boldsymbol{\omega}) &= (y_{1,2}^*(\mathbf{x}, \boldsymbol{\omega}) + \omega_{1,3} - x_{1,3})^+, \\
y_{1,4}^*(\mathbf{x}, \boldsymbol{\omega}) &= (y_{1,3}^*(\mathbf{x}, \boldsymbol{\omega}) + \omega_{1,4} - x_{1,4})^+, \\
y_{1,5}^*(\mathbf{x}, \boldsymbol{\omega}) &= \max((y_{1,4}^*(\mathbf{x}, \boldsymbol{\omega}) + \omega_{1,5} - x_{1,5})^+, y_{3,11}^*(\mathbf{x}, \boldsymbol{\omega}) - d_1(\mathbf{x}) + 5), \\
&\dots
\end{aligned}$$

4.1 Numerical solution

We applied statistical sampling from the joint distribution of the disturbances via Monte Carlo simulation, using the resulting empirical distributions in the model instead of the original exponential distributions. This technique is called *external sampling* ([28]). Assuming some fairly mild technical conditions, it may be proved that the optimal solution and the optimal objective function value of the approximating problem are strongly consistent statistical estimators of those of the original problem (see [28] for a proof and further properties of the method), so the optimal solution of the sequence of approximating problems with increasing sample sizes converges to the true optimal solution with probability one. Additionally, following [16], we applied the technique of *antithetic variables* (see e.g. [25]) in order to increase the precision of the discrete approximation for a given sample size. We solved the problem with several sample sizes, and finally decided to use $r = 1000$ joint realizations, as increasing the sample size only altered the optimal solution negligibly beyond this value. We solved problem (16) in the model management system SLP-IOR using the Regularized Decomposition algorithm for SP models. Additionally, for the sake of comparison, we also

We define the plus operator as $a^+ = \max(a, 0)$ ($a \in \mathbb{R}$).

solved the equivalent LP and NLP formulations presented in Section 3 in the computer algebra system Maplesoft Maple 2015 ([18]) using a built-in sparse interior point algorithm for the LP and the Sequential Quadratic Programming algorithm for the NLP formulation. The three algorithms arrived at the same optimal solution with computation times of 33.27 minutes (LP), 1.49 minutes (SP) and 1.06 minutes (NLP). These results indicate that using more advanced formulations instead of solving the equivalent LP as done earlier in the literature (in e.g. [16] and [30]) pays off in terms of computational efficiency, especially for larger models of even more complex real-life networks. In the remainder of our paper, we shall stick to the SP formulation, as we shall focus on non-computational issues.

In addition to the three solution frameworks examined so far, we tested an additional, rather heuristic solution to present evidence that relying on intuition may lead to poor allocations. A naive analysis might suggest that it provides a fairly good solution to allocate the supplements proportionally to the mean disturbances associated with the events, which is often done by railway companies in real-life timetables. In order to test this approach, we fixed the values of the decision variables in the SP model to follow a proportional allocation, which increased the optimal objective function value by 13.0%. This is a highly significant increase in practice, suggesting that using proportional allocations should be avoided in real-life timetables.

4.2 Analysis of the role of distributional assumptions

Our choice of the exponential distribution was motivated by several previous publications including [6], [16], [30] and [32], which described experiments indicating that observed disturbances fit well to exponential distributions. Nevertheless, as a high level of sensitivity of the optimal solution with respect to the chosen distributional form would indicate a low level of practical reliability, it is important to investigate the effect of our choice on the optimal solution. Therefore, in addition to the truncated exponential distributions, we performed a sensitivity analysis using the following distributional assumptions for $\omega_{i,j}$ ($i \in I, j \in J_i$, with

On a computer with a 2.8 GHz Intel Core i7 processor and 4 GB RAM.

the parameters $\mu_{i,j}$ the same as in Table 1):

- Untruncated exponential with mean $\mu_{i,j}$,
- Pareto with two degrees of freedom and cdf $F_{i,j}(x) = 1 - \left(\frac{\mu_{i,j}}{x + \mu_{i,j}}\right)^2$ ($x \geq 0$).
- Uniform on $(0, 2\mu_{i,j})$.

We selected these three distributions in order to test the effects of omitting truncation, introducing heavy tails and a lack of positive skewness, respectively, and calibrated them to have equal means so as to ensure their comparability. We summarize the results of our analysis in Table 3. The results indicate that the lack of truncation and heavy tails increase

	Distributional assumption			
	Trunc. exp.	Untrunc. exp.	Pareto	Uniform
Objective function	0.88	0.89	1.06	0.54
Unpunctuality (per 1000)	87.01	86.95	117.94	15.58
RMSD compared to original	0	0.09	0.14	0.12
Correlation between new and original solutions	1	0.99	0.97	0.98

Table 3: Main results for different disturbance distributions

the mean delay and unpunctuality, whereas the lack of skewness decreases them due to increased predictability. More importantly, the RMSD (root-mean-square difference) and the correlation coefficient of the new and original first-stage solution vectors indicates a very high level of similarity of the optimal solutions of the modified problems to the original one, which implies a high degree of robustness of the optimal solution of our model even if the exponential distribution is not appropriate or truncation is dropped from the model.

4.3 Analysis of the role of the total amount of supplements

As increasing the amount of supplements results in decreased delays at the cost of increased planned travel times, there is a natural trade-off between delays and planned travel times, to be regulated by a reasonable choice of the amount of distributable supplements. For this reason, it is important to analyze whether the actual values of the supplements used by the

Dutch Railways, amounting to 7% of the planned travel times, are reasonable. To address this question, we constructed the following modified version of problem (16):

$$D = \min_{\mathbf{x} \geq \mathbf{0}, \gamma \geq 0} \{ \mathbb{E}_{\boldsymbol{\omega}} [\Delta(\mathbf{x}, \boldsymbol{\omega})] + \beta M(\gamma) : \sum_{j=1}^{n_i} x_{i,j} \leq \gamma q_i \ (i \in I), \text{ constraints (5) and (13)} \}$$

where $M(\gamma) = \gamma \sum_{i=1}^m \sum_{j=1}^{n_i} \tilde{w}_{i,j} q_i$ and $\Delta(\mathbf{x}, \boldsymbol{\omega}) = \min_{\mathbf{y} \geq \mathbf{0}} \{ \mathbf{w}\mathbf{y} : \text{constraints (14) and (15)} \}$.

In this modified problem, the total amount of distributable supplements for each instance is scaled by a factor $\gamma > 0$, and their weighted mean $M(\gamma)$ is penalized in the first-stage objective function by a linear term with an exogenous penalty coefficient $0 < \beta < 1$. Thus the modified first-stage objective function is the sum of the expected weighted mean delay and an additional penalty term, which is the product of a penalty coefficient and the weighted mean of the planned excess travel times arising due to the use of supplements. Here the penalty parameter expresses how many times planned excess times are more undesirable than unexpected delays. It is natural to assume that $\beta < 1$, as passengers obviously prefer expected excess times to unexpected ones. Due to the previously explained trade-off between expected and unexpected excess times, the first term is decreasing and the second one is increasing in the scale parameter γ , which implies that there is a value γ^* that optimizes the trade-off, and its value depends on the penalty parameter β . Thus by finding the parameter value β^* that results in $\gamma^* = 1$, we may determine which penalty parameter results in the same optimal quantities of distributable supplements as the ones currently used in practice. We shall call this value β^* the *implied penalty parameter* of the current timetable. We solved the modified problem using different values of the parameter β and recorded the optimal values γ^* for each run. The numerical results shown in Table 4 indicate the implied penalty parameter $\beta^* \approx 0.56$, or in other words, the current timetable assumes that unexpected excess travel times are somewhat less than twice as undesirable as expected ones.

β	0.2	0.3	0.4	0.5	0.56	0.6	0.7	0.8
γ^*	1.55	1.34	1.18	1.07	1	0.96	0.86	0.75

Table 4: Optimal scale parameters for different values of the penalty parameter

If e.g. it is more reasonable to assume that unexpected excess travel times are *exactly* twice

as undesirable as expected ones, than the current values of supplements should be increased by a factor of 1.07 according to Table 4.

4.4 Analysis of the effect of dependent disturbances

As the independence assumption is rather unrealistic, we now investigate the effect of possible dependence among the disturbances. To this end, a parametric multidimensional distribution needs to be fully specified. This is greatly simplified by the use of copulas (see e.g. [20]), which are cumulative probability distribution functions over the multidimensional unit hypercube, as it is known from Sklar’s theorem (again, see [20]) that any multivariate probability distribution is uniquely determined by the associated marginal distributions and a copula function. Out of several parametric families of copulas that are known in the literature, we propose the n -dimensional Gaussian copula due to its computational tractability. It is defined as $C(z_1, z_2, \dots, z_n) = \Phi_{\mathbf{R}}[\Phi^{-1}(z_1), \Phi^{-1}(z_2), \dots, \Phi^{-1}(z_n)]$ ($0 < z_1, z_2, \dots, z_n < 1$), where $\Phi_{\mathbf{R}}$ and Φ^{-1} denote the joint cdf of the n -dimensional standard normal distribution with correlation matrix $\mathbf{R}$ and the quantile function (inverse cdf) of the univariate standard normal distribution, respectively. To make our analysis one-dimensional, we used a simple correlation matrix with $R_{g,h} = \rho$ ($g \neq h, g, h \in \{1, 2, \dots, n\}, 0 \leq \rho \leq 1$) and $R_{g,g} = 1$ ($g \in \{1, 2, \dots, n\}$), and solved the original SP problem with the additional assumption of this Gaussian copula joining the disturbances, while leaving their marginal distributions unchanged. In the external sampling algorithm, we sampled from the Gaussian copula using the Cholesky decomposition of the matrix $\mathbf{R}$ (this technique is described in [4]).

Our numerical results for different correlation coefficients are summarized in Table 5.

Correlation coefficient (ρ)	0	0.2	0.4	0.6	0.8	1
Relative increase in unpunctuality	0%	1%	4%	7%	10%	14%
Correlation between new and original solutions	1	0.99	0.97	0.94	0.9	0.82
RMSD compared to original	0	0.09	0.16	0.22	0.28	0.36

Table 5: Main results for different levels of dependence

The relative increases in unpunctuality indicate that higher levels of dependence result in

significantly diminishing punctuality and it is clear from the correlations and RMSD values in the last two rows of Table 5 that the optimal solution vectors diverge largely from the original one if the correlation coefficient increases. Thus we conclude that the underlying dependence should not be ignored in practice. We shall introduce a statistical estimation procedure that enables the modeler to account for dependence in the next section.

5 Statistical estimation of the parameters of the disturbance distributions

Most railway companies keep detailed delay statistics. Even though delays are directly observable and easily measurable, our model and its predecessors require distributional assumptions for the underlying disturbances and not the delays. Thus a statistical procedure is necessary to complete the path which leads from raw data to an optimal timetable. Assuming a parametric joint distribution of the disturbances for convenience, the problem of estimating its parameters should be addressed in a methodologically appropriate way. We remark that no specialized estimation procedure has been described in the literature so far to the best of our knowledge.

Additionally, our presentation aims to correct an important deficiency of previously known models (e.g. [16] and [30]) only considering independent disturbances, as independence is highly unrealistic due to common external environmental, operational and traffic factors (e.g. storms, power outages or track maintenance works) that may simultaneously generate increases in several disturbances within a day.

To construct an appropriate estimation procedure, first we notice the following relationship by considering Theorem 1 and equation (19):

$$\begin{aligned} y_{i,j}^* &= \max \left(0, y_{i,j-1}^* + \omega_{i,j} - x_{i,j}, y_{i_{K(i,j)}', j_{K(i,j)}'}^* - d_{K(i,j)}(\mathbf{x}) + s_{K(i,j)} \right) \\ &= \max \left(y_{i,j-1}^* + \omega_{i,j} - x_{i,j}, (y_{i_{K(i,j)}', j_{K(i,j)}'}^* - d_{K(i,j)}(\mathbf{x}) + s_{K(i,j)})^+ \right) \quad (i \in I, j \in J_i), \end{aligned}$$

An analysis tool developed by the Dutch Railways for analyzing railway log files and determining the intensities of exponentially distributed disturbances is mentioned cursorily without explanation in [16].

which yields n equations for the n disturbances.

For simplicity, we introduce $S(i, j) := (y_{i'_{K(i,j)}, j'_{K(i,j)}}^* - d_{K(i,j)}(\mathbf{x}) + s_{K(i,j)})^+ \ (i \in I, j \in J_i)$ for the secondary delay caused by waiting for an interfering instance with priority, arriving at

$$y_{i,j}^* = \max \left(y_{i,j-1}^* + \omega_{i,j} - x_{i,j}, S(i, j) \right) \quad (i \in I, j \in J_i).$$

The above relationship implies the following for a particular disturbance term:

$$\begin{aligned} \omega_{i,j} &\leq S(i, j) + x_{i,j} - y_{i,j-1}^* && \text{if } y_{i,j}^* = S(i, j), \\ \omega_{i,j} &= y_{i,j}^* - y_{i,j-1}^* + x_{i,j} && \text{if } y_{i,j}^* > S(i, j). \end{aligned} \tag{24}$$

This is all that may be inferred about the disturbances based on measured delay data.

Denoting by $y^{*(g)}(i, j)$ ($i \in I, j \in J_i, g \in \{1, 2, \dots, d\}$) the g^{th} observed value of the delay of event j of instance i out of a sample of d realizations (days) and assuming that delays associated with distinct events are independent from one another as well as from all delays measured on different days, we have an IID sample of d observations for every random disturbance with the following associated log-likelihood:

$$\begin{aligned} LL_{i,j}(\boldsymbol{\theta}_{i,j}) &= \sum_{\{g: y_{i,j}^{*(g)} = S(i,j)\}} \log F_{i,j} \left(\boldsymbol{\theta}_{i,j}, S(i, j) + x_{i,j} - y_{i,j-1}^{*(g)} \right) + \\ &\quad \sum_{\{g: y_{i,j}^{*(g)} > S(i,j)\}} \log f_{i,j} \left(\boldsymbol{\theta}_{i,j}, y_{i,j}^{*(g)} - y_{i,j-1}^{*(g)} + x_{i,j} \right) \quad (\boldsymbol{\theta}_{i,j} \in \Theta_{i,j}, i \in I, j \in J_i), \end{aligned}$$

where the symbols $F_{i,j}$, $f_{i,j}$, $\boldsymbol{\theta}_{i,j}$ and $\Theta_{i,j}$ denote the cdf and pdf, the parameters and the range of possible parameter values of the individual disturbance distributions, respectively. These n functions may be maximized one by one in order to arrive at the estimated parameter vectors $\hat{\boldsymbol{\theta}}_{i,j}$ for every $(i, j) \in J$.

For the framework of dependent disturbances conjoined by a Gaussian copula, which we introduced in Section 4.4, we follow Joe ([8]), who advocates a two-stage procedure of first estimating the parameters of the marginal distributions, to be followed by the estimation of the copula parameters (the correlation matrix $\mathbf{R}$ in our case), arguing that the resulting estimators are consistent. To estimate the correlation matrix $\mathbf{R}$ this way, first the disturbances must be transformed to follow standard normal distributions under the assumption of the estimated marginal distributions by applying the transformation

$$\begin{aligned}
\nu_{i,j} &:= \Phi^{-1}(F_{i,j}(\hat{\boldsymbol{\theta}}_{i,j}, \omega_{i,j})) \quad (i \in I, j \in J_i), \text{ transforming (24) into} \\
\nu_{i,j} &\leq \Phi^{-1} \left[F_{i,j}(\hat{\boldsymbol{\theta}}_{i,j}, S(i,j) + x_{i,j} - y_{i,j-1}^*) \right] =: T(i,j) \quad \text{if } y_{i,j}^* = S(i,j), \\
\nu_{i,j} &= \Phi^{-1} \left[F_{i,j}(\hat{\boldsymbol{\theta}}_{i,j}, y_{i,j}^* - y_{i,j-1}^* + x_{i,j}) \right] =: U(i,j) \quad \text{if } y_{i,j}^* > S(i,j).
\end{aligned} \tag{25}$$

To estimate the correlation coefficient of two transformed disturbances $\nu_j, \nu_{j'}$ ($\mathbf{j}, \mathbf{j}' \in J$), from a sample, we follow the procedure described in [17], and construct a log-likelihood function by considering the four possible cases depending on whether $y_j^* = S(\mathbf{j})$ and/or $y_{j'}^* = S(\mathbf{j}')$ hold for a given observation:

$$\begin{aligned}
LL_{\mathbf{j}\mathbf{j}'}(\rho_{\mathbf{j}\mathbf{j}'}) = & \sum_{\{g: y_j^{*(g)} = S(\mathbf{j}) \text{ and } y_{j'}^{*(g)} = S(\mathbf{j}')\}} \log \Phi_{12}(\rho_{\mathbf{j}\mathbf{j}'}, T(\mathbf{j}), T(\mathbf{j}')) + \\
& \sum_{\{g: y_j^{*(g)} = S(\mathbf{j}) \text{ and } y_{j'}^{*(g)} > S(\mathbf{j}')\}} \log \Phi_1(\rho_{\mathbf{j}\mathbf{j}'}, T(\mathbf{j}), U(\mathbf{j}')) + \\
& \sum_{\{g: y_j^{*(g)} > S(\mathbf{j}) \text{ and } y_{j'}^{*(g)} = S(\mathbf{j}')\}} \log \Phi_2(\rho_{\mathbf{j}\mathbf{j}'}, U(\mathbf{j}), T(\mathbf{j}')) + \\
& \sum_{\{g: y_j^{*(g)} > S(\mathbf{j}) \text{ and } y_{j'}^{*(g)} > S(\mathbf{j}')\}} \log \Phi(\rho_{\mathbf{j}\mathbf{j}'}, U(\mathbf{j}), U(\mathbf{j}')) \quad (-1 \leq \rho_{\mathbf{j}\mathbf{j}'} \leq 1, \mathbf{j}, \mathbf{j}' \in J),
\end{aligned}$$

where Φ , Φ_1 , Φ_2 and Φ_{12} are the joint cdf of the bivariate standard normal distribution, its two partial derivatives and its mixed second-order partial derivative, respectively. These functions may be maximized one by one to obtain the estimated pairwise correlation coefficients and eventually the estimated matrix $\hat{\mathbf{R}}$. This way we obtain the estimated joint probability distribution by Sklar's theorem. We finally remark that the estimation of the joint distribution requires individual delay data, whereas we only possess aggregate data from the Dutch Railways, so we have only tested the above estimation procedure under the independence assumption on real-life data.

Bibliography

References

- [1] Birge, J.R. and Louveaux, F. (1997). *Introduction to stochastic programming*. Springer-Verlag, New York.
- [2] Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- [3] Demitz, J., Hübschen, C., Albrecht, C. (2004). Timetable stability – using simulation to ensure quality in a regular interval timetable. In Allan, J., Brebbia, C.A., Hill, R.J., Sciutto, G., Sone, S. (editors), *Computers in Railways IX*, pp. 549–562. WIT Press.
- [4] Gentle, J.E. (2009). *Computational Statistics*, pp. 315–316. Springer-Verlag, New York.
- [5] Goossens, J.H.M. (2004). *Models and algorithms for railway line planning models*. Ph.D. thesis, University of Maastricht, the Netherlands.
- [6] Goverde, R.M.P. (2005). *Punctuality of railway operations and timetable stability analysis*. Ph.D. thesis, Technical University Delft, The Netherlands.
- [7] Huisman, T., Boucherie, R.J. (2001). Running times on railway sections with heterogeneous train traffic. *Transportation Research Part B*, 35:271–292.
- [8] Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.
- [9] Kall, P. and Mayer, J. (1996). SLP-IOR: An interactive model management system for stochastic linear programs. *Mathematical Programming.*, 75(2):221–240.
- [10] Kall, P. and Mayer, J. (2005). *Stochastic linear programming: Models, theory, and computation*. Springer-Verlag.
- [11] Klein Haneveld, W.K. and Van der Vlerk, M. H. (1999). Stochastic Integer Programming: General models and algorithms. *Annals of Operations Research*, 85:39–57.
- [12] Klein Haneveld, W.K. and Van der Vlerk, M.H. (2006). Integrated Chance Constraints: Reduced Forms and an Algorithm. *Computational Management Science*, 3(4): 245–269.
- [13] Klein Haneveld W.K. and Van der Vlerk, M. H. (2014). *Stochastic programming*. Lecture notes. Rijksuniversiteit Groningen, the Netherlands.

- [14] Kroon, L.G., Peeters, L.W.P. (2003). A variable trip time model for cyclic railway timetabling. *Transportation Science*, 37:198–212.
- [15] Kroon, L.G., Dekker, R., Vromans, M.J.C.M. (2007). Cyclic railway timetabling: a stochastic optimisation approach. In: Geraets, F., Kroon, L.G., Schöbel, A., Wagner, D., Zaroliagis, C. (editors), *Algorithmic Methods in Railway Optimization*, vol. 4359 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 41–66.
- [16] Kroon, L.G. et al. (2008). Stochastic improvement of cyclic railway timetables. *Transportation Research Part B*, vol. 42, issue 6, pp. 553–570.
- [17] Lyles, R.H., Williams, J.K., Chuachoowong, R. (2001). Correlating two viral load assays with known detection limits. *Biometrics*, 57:1238–1244.
- [18] *Maple User Manual*. Toronto: Maplesoft, a division of Waterloo Maple Inc., 2005-2014.
- [19] Mayer, J. (1998). *Stochastic linear programming algorithms: A comparison based on a model management system*. Gordon and Breach Science Publishers.
- [20] Nelsen, R.B. (2007). *An Introduction to Copulas*. Springer-Verlag.
- [21] Peeters, L.W.P. (2003). *Cyclic railway timetable optimization*. Ph.D. thesis, ERIM, Erasmus University Rotterdam, the Netherlands.
- [22] Peeters, L.W.P. and Kroon, L.G. (2001). A cycle based optimization model for the cyclic railway timetabling problem. In *Computer aided scheduling of public transport*, pp. 275–296., Springer-Verlag, Berlin, Germany.
- [23] Prékopa, A. (1995). *Stochastic Programming*, vol. 324 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht.
- [24] Retel Helmrich, M.J. (2006). *Improving robustness of a railway timetable through stochastic optimisation of supplementary times*. M.Sc. thesis, Erasmus University Rotterdam, Rotterdam School of Economics, The Netherlands.
- [25] Ross, S.M. (2012). *Simulation* (fifth edition). Elsevier.
- [26] Ruszczyński, A. (1986). A regularized decomposition method for minimizing a sum of polyhedral functions. *Mathematical Programming*, 35:309–333.
- [27] Ruszczyński, A. (2006). *Nonlinear optimization*. Princeton University Press.
- [28] Shapiro, A. (2003). Monte Carlo sampling methods. In Ruszczyński, A. and Shapiro, A.

- (editors), *Stochastic Programming*, chapter 6, pp. 353–425. Elsevier, vol. 10. of *Handbooks in Operations Research and Management Science*.
- [29] Van Slyke, R. and Wets, R.J-B (1969). L-shaped linear programs with applications to control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17:638–663.
- [30] Vromans, M.J.C.M. (2005). *Reliability of Railway Systems*. Ph.D. thesis, ERIM, Erasmus University Rotterdam, The Netherlands.
- [31] Vromans, M.J.C.M., Kroon, L.G., Dekker, R. (2006). Reliability and heterogeneity of railway services. *European Journal of Operational Research*.
- [32] Yuan, J. (2006). *Stochastic modelling of train delays and delay propagation in stations*. Ph.D. thesis, Technical University Delft, The Netherlands.