Optimizing existing railway timetables by means of stochastic programming

Peter Vekas[*], Maarten H. van der Vlerk[†] and Willem K. Klein Haneveld[†]

June 14, 2012

Abstract

We present some models to find the best allocation of a limited amount of so-called running time supplements (extra minutes added to a timetable to reduce delays) on a railway line. By the best allocation, we mean the solution under which the sum of expected delays is minimal. Instead of trying to invent a completely new timetable, our aim is to finely adjust an already existing and well-functioning one. We model this inherently stochastic optimization problem by using two-stage recourse models from stochastic programming, following Vromans [9]. We present an improved formulation, allowing for an efficient solution using a standard algorithm for recourse models. We include a case study that we managed to solve about 180 times faster than it was solved in [9]. By comparing our solution with other, seemingly intuitive solutions, we show that finding the best allocation is not obvious, and implementing it in practice promises a significant improvement in the punctuality of trains. A technique to estimate the model parameters from empirical data and an approximating deterministic problem are also presented, along with some practical ideas that are meant to enhance the applicability of our models.

Introduction

Planning railway timetables is extremely complicated, 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, engines and carriages, not to mention the complex interactions among different trains.
- Additionally, it is a *decision problem under uncertainty*: to decide upon the departure and arrival times of a train, one needs to consider several uncertain factors that may eventually result in some minor or major deviations from the timetable.

In this paper, we will *ignore* the first dimension of the problem, and take a timetable for granted, including the number of trains on a railway line with their order and the planned departure and arrival times at the terminal points of the line. Instead, we will only consider a specific problem arising from the uncertainty of the realized departure and arrival times.

The supplement allocation problem

To avoid delays, the traveling time of a train between two stations is generally planned to be longer than what is expected in an ideal situation. In fact, almost all delays could be eliminated by

[*]Corresponding author. Corvinus University of Budapest, Department of Operations Research and Actuarial Science, Fövám tér 8, 1093, Budapest, Hungary. E-mail peter.vekas@uni-corvinus.hu

[†]University of Groningen, Department of Operations, PO Box 800, 9700 AV Groningen, The Netherlands

simply including a sufficient amount of these “extra minutes” called supplements in the timetable. However, no one would favor the resulting excessively long traveling times. Thus, in practice, the railway company restricts the total amount of supplements to some reasonable extent and tries to distribute it in the timetable in an effective way.

Some numerical methods (e.g. simulation) have been applied for a long time by the Dutch Railways (Nederlandse Spoorwegen, `www.ns.nl`) to facilitate finding the best allocation, and in 2005, Michiel Vromans used stochastic programming to model the problem in his Ph.D. thesis [9], which we took as a starting point to develop our own models. This paper will only apply the framework of stochastic programming to model the supplement allocation problem.

The structure of the paper

Sections 1 and 3 contain stochastic programming models of the supplement allocation problem. We present an improved, brief, recursive formulation of already existing models from literature. In Sections 2, 4 and 5, we apply our models to the Haarlem-Maastricht railway line. The problem in Section 2 was solved in ca. 30 minutes in the literature, and it is described here how we solved it in just 10 seconds. Section 6 elaborates on our ideas about preparing more complex models. Section 7 aims at modeling large delays in a more satisfactory way, by using a heavy-tailed probability distribution. In Section 8, an approximating deterministic non-linear programming problem is presented, which can be solved faster and more easily than the original stochastic model. Section 9 applies maximum likelihood estimation to provide realistic inputs for our models from available statistical data. Section 10 contains suggestions for further research and concludes with a summary. As an appendix, Section A summarizes the definitions and concepts of stochastic programming applied in the paper.

1 The core model: A single line with one train and no cyclicity

Let us consider the following model: there is a railway line with a sequence of $n+1$ stations, and thus n trips between the stations. A train departs from station 0 with no delay. For the sake of

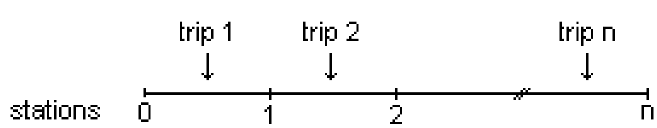

Figure 1: Stations and trips in the core model

simplicity, we assume that it spends no time waiting at the stations (it is very simple to relax this assumption: the periods spent at the stations may be introduced as extra trips in the model), and on each of the trips $i = 1, 2, \ldots, n$, it incurs a so-called *disturbance* ω_i. Following [9], we assume that ω_i is an exponentially distributed random variable with a rate λ_i that is characteristic for the given trip. The disturbance ω_i is the difference between the actual and ideal (i.e., undisturbed) running times on trip i. In the simplest version of the model, the collection of random variables $\{\omega_i\}_{i=1}^{n}$ is assumed to be independent. Since there is a running time supplement x_i allocated to each trip in the timetable, not all disturbances result in delays. The supplement x_i is the difference between the planned and undisturbed running times on trip i. From this interpretation, it follows that if the disturbance on the first trip exceeds the running time supplement then there is a delay (see Figure 2 for an illustration); however, if the running time supplement can compensate for the entire disturbance then the train incurs no delay. In formulas, denoting the delay of the train at station i by d_i,

$$d_1 = \max(\omega_1 - x_1, 0) =: (\omega_1 - x_1)^+.$$

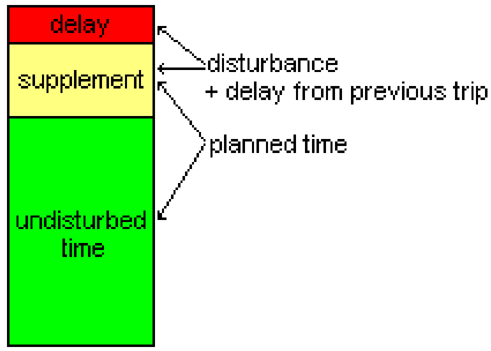

Figure 2: Decomposition of the actual running time on a trip in case of a delay.

Similarly, the supplement x_i in general is meant to compensate for the delay from the previous trip and the disturbance on trip i, thus the following recursion holds (with $d_0 \equiv 0$, as we assume that there are no negative delays, which is obviously equivalent to stating that it stays waiting at the station if it accidentally arrives too early):

$$d_i = (d_{i-1} + \omega_i - x_i)^+, \quad i = 1, 2, \ldots, n. \tag{1}$$

In the model, we assume that the total amount of running time supplements is limited from above by a scalar $M > 0$, and the railway company would like to distribute it among the trips in such a way that the sum of expected delays at the stations is minimal. Introducing vectors $x, \omega \in \mathbb{R}^n$ with components x_i and ω_i ($i = 1, 2, \ldots, n$), this leads to the following stochastic optimization problem in a natural way:

$$\min_{x \geq 0} \left\{ E_\omega[v(x, \omega)] : x_1 + x_2 + \ldots + x_n \leq M \right\}, \tag{2}$$

$$\text{where} \quad v(x, \omega) = d_1(x, \omega) + d_2(x, \omega) + \ldots + d_n(x, \omega).$$

At this point, defining $v(x, \omega)$ as the optimal value of a linear programming problem would turn (2) into a standard two-stage linear recourse model. This may be accomplished by constructing a second-stage problem in the following way:

$$\begin{aligned} v(x, \omega) = \min_{y \geq 0} \{ y_1 + y_2 + \ \ldots \ + y_n : & \\ y_1 &\geq \omega_1 - x_1, \\ y_2 &\geq y_1 + \omega_2 - x_2, \\ &\vdots \\ y_n &\geq y_{n-1} + \omega_n - x_n \}. \end{aligned} \tag{3}$$

This may be verified in three steps: first of all, for every feasible solution of (3), it can be proved by induction that $y_i \geq d_i$ ($i = 1, 2, \ldots, n$) holds, furthermore, d_i is the sharpest lower bound of y_i ($i = 1, 2, \ldots, n$), since $y_i = d_i$ ($i = 1, 2, \ldots, n$) is trivially a feasible solution of (3), and finally, since the objective function of (3) is monotone increasing in y_i, it follows that in the optimum, $y_i = d_i$ ($i = 1, 2, \ldots, n$), and $v(x, \omega) = d_1(x, \omega) + d_2(x, \omega) + \ldots + d_n(x, \omega)$. We introduce the following $n \times n$ matrix:

$$W = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

Summarizing the preceding results, as the reader may simply verify, the core model is equivalent to the following two-stage recourse model:

$$\min_{x \geq 0} \left\{ E_\omega[v(x,\omega)] : 1x \leq M \right\}, \tag{4}$$

$$\text{where} \quad v(x,\omega) = \min_{y \geq 0} \left\{ 1y : x + Wy \geq \omega \right\}.$$

Because of the non-negativity of the second-stage objective function, this problem trivially has a sufficiently expensive recourse structure (SER), and it also has complete recourse (CR, see Section A for the definitions of SER and CR): by letting, for any $z \in \mathbb{R}^n$, $z^T = (z_1, z_2, \ldots, z_n)$, $y_i = (y_{i-1} + z_i)^+$ $(i = 1, 2, \ldots, n)$, $y_0 \equiv 0$, it is obvious that $y \geq 0$, and it can be verified by induction (just like in the derivation of the properties of (3)) that $Wy \geq z$ holds. Additionally, we tried an alternative formulation as well: instead of a recursion, the difference equation (1) can be solved explicitly for d_i. Since in this representation, the number of second-stage constraints increases quadratically with the number of trips, this formulation finally proved to be computationally inefficient.

2 Example: The Haarlem-Maastricht case study

To test how our model performs on a practical problem, we solve a problem originally presented by Vromans ([9], p.153). In this case study, a total of $M = 10.93$ minutes of running time supplements (typically a fixed percentage of the total traveling time) has to be allocated among the 8 trips of the Haarlem-Maastricht railway line. Waiting times at the stations are not included explicitly in the model, i.e., they are built into the running times. The running time disturbances are assumed to be independent and exponentially distributed, and the means of the distributions are given for each trip. Even though total traveling times are rounded to integers by the Dutch Railways, values of the supplements are given in seconds in practice, so the optimal solutions need not be integers.

This problem was solved in [9] by using a different model formulation, discretizing the random variables and solving the deterministic equivalent (see (14) in Section A) of the resulting two-stage recourse model. To implement our core model from the previous section, we used the model management system SLP-IOR (see [3]). This software contains several solution algorithms for two-stage recourse models. In the beginning, we tried a few algorithms for continuous distributions, and we found the same optimal solution as Vromans did, but it took us ca. 25 minutes, which was hardly any better than his 30 minutes, so we finally decided to discretize the random variables. We applied right-truncation at the 5-minute level to the exponential random variables (truncation is not necessary in principle, however, in our model, it is meant to represent the idea that we only wish to model small disturbances in railway traffic), and used statistical sampling from their joint distribution. We used the resulting empirical distributions in the model instead of the exponential distributions. This technique is called *external sampling*. 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 [6] for a proof and further properties of external sampling). After comparing the solution times of several algorithms, we decided to use a modified version of the L-shaped algorithm (see [7]), which is referred to as Regularized Decomposition (see [5]). We tried to solve the problem with several sample sizes, and finally decided to use 5000 joint realizations, because larger samples did not result in significantly different optimal solutions. Our results and those of Vromans are summarized in Table 1 and Figure 3.

The close similarity of the two solutions is apparent at first sight. The difference is due to the fact that we used different discretizations. The fact that this difference is so small is evidence that the optimal supplement allocation is sufficiently robust with respect to the discretization one uses. The computation time of our algorithm was two orders of magnitude smaller than the 30 minutes of Vromans, indicating that using a combination of an appropriate model and advanced SP solution techniques strongly increases computational efficiency. The solution time remained

	Our solution	Vromans	Mean disturbance
Haarlem - Amsterdam Centraal	0.89	0.85	1.03
Amsterdam Centraal - Duivendrecht	1.02	1.01	0.84
Duivendrecht - Utrecht Centraal	1.43	1.43	1.15
Utrecht Centraal - 's Hertogenbosch	2.68	2.63	2.01
's Hertogenbosch - Eindhoven	1.64	1.71	1.28
Eindhoven - Roermond	2.49	2.57	2.4
Roermond - Sittard	0.77	0.72	1.22
Sittard - Maastricht	0	0	0.87
Computation time	10.4 sec	≈ 30 min	-

Table 1: Optimal supplement values in minutes according to our solution and that of Vromans (see [9])

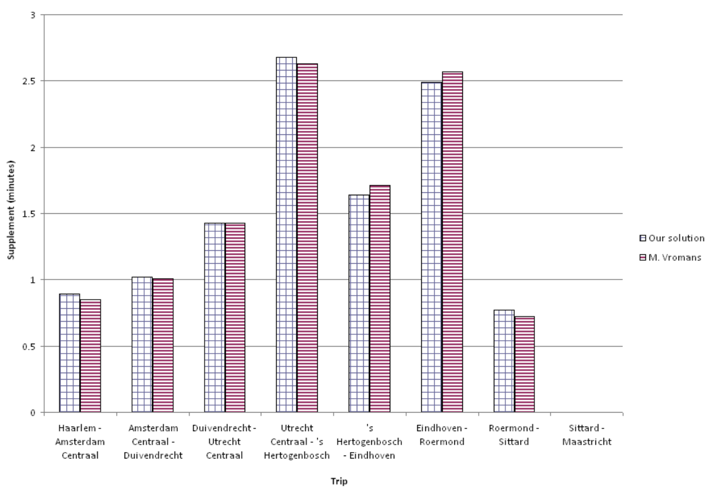

Figure 3: The patterns of our optimal allocation and that of Vromans (see [9])

very small irrespectively of the sample size and the accuracy parameter of the solution algorithm. These results increased our confidence for future model extensions.

Finally, we present an example to point out that relying on intuition mostly leads to poor allocations. A naive analysis might suggest that it must provide a fairly good solution to allocate the running time supplements in such a way that they are proportional to the average disturbances on each trip. In fact, this is what railway companies (including the Dutch Railways) mostly do in real-life timetables as well. One might even think that this problem is not worth dealing with, as "putting some minutes back and forth" does not make any real difference in practice. In order to test whether this approach is correct, we solved the Haarlem-Maastricht model with the decision variables fixed at a proportional allocation, and we also tried a uniform allocation under which the supplements are identical on all trips. The patterns of these allocations are plotted in Figure 4, and the results are summarized in Table 2. As the last row of Table 2 shows, the optimal objective function values under a proportional and a uniform allocation increase by 11.2% and 25.7%, respectively. These are highly significant increases in practice, hence we conclude that using proportional or uniform allocations should better be avoided in real-life timetables.

3 The cyclic model

Throughout the remainder of this paper, the term *number of cycles* and the variable k will refer to the number of times the train in the model passes through the n trips of the railway line. Thus in the core model, $k = 1$. In the optimal solution of the Haarlem-Maastricht case study in the

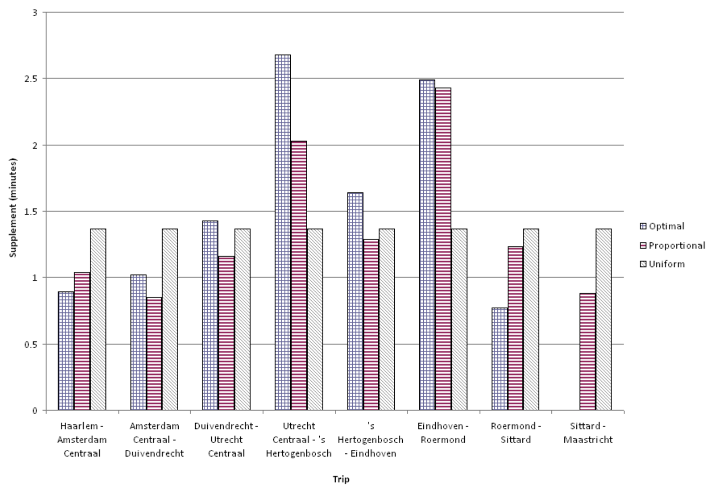

Figure 4: The optimal, proportional and uniform allocation patterns

	Optimal	Proportional	Uniform
Haarlem - Amsterdam Centraal	0.89	1.04	1.365
Amsterdam Centraal - Duivendrecht	1.02	0.85	1.365
Duivendrecht - Utrecht Centraal	1.43	1.16	1.365
Utrecht Centraal - 's Hertogenbosch	2.68	2.03	1.365
's Hertogenbosch - Eindhoven	1.64	1.29	1.365
Eindhoven - Roermond	2.49	2.43	1.365
Roermond - Sittard	0.77	1.23	1.365
Sittard - Maastricht	0	0.88	1.365
Expected total delay	8.46	9.41	10.63
Increase compared to optimal	0%	11.2%	25.7%

Table 2: The optimal, proportional and uniform allocations

previous section, we saw that no supplement was allocated to the last trip. However, one may well suspect that this was only because the model assumed just one cycle, which is obviously not a realistic assumption. This observation motivated us to investigate what happens to the optimal supplement allocation in a cyclic model, as well as whether the new model yields a substantially different optimal solution and whether it is sufficient to use a non-cyclic model in practice. To investigate these questions, the assumption of having a single cycle is relaxed in this section. In the cyclic extension of the model, we still assume that there is only one train, but instead of stopping, now we assume that it turns around at the end of the line (at station n) and continues its journey through all the trips in the reverse order, where it turns around again etc. (Figure 5 is an illustration of this.) This will be modeled analogously to the core model. We assume the running

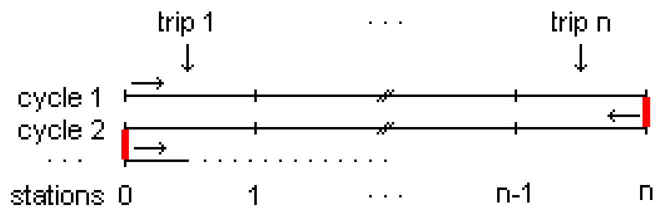

Figure 5: Stations and trips in a cyclic model

time supplements in the odd and even cycles to be equal in the case of each trip, following the principle of symmetric timetable planning applied in the Netherlands (see [9]), which prescribes that the planned running times of the same train in the two opposite directions have to be equal on all trips. (One could easily drop this assumption from the model by doubling the number of decision variables.)

Obviously, the first cycle in a cyclic model corresponds to the trips of the core model. After the train arrives at station n, we will assume that an idle period follows (denoted by a thick vertical line in Figure 5), which links the first and second cycles. Since this break acts as a running time supplement in the timetable, we will denote its planned length by x_0. Additionally, we introduce a disturbance ω_{n+1} for this linking period. The indices need some explanation:

- The index i of a disturbance ω_i or a delay d_i will indicate the number of trips and linking periods elapsed since the start of the train from station 0 in the first cycle. Thus, after introducing a linking period between each cycle and its successor analogously to the one between the first and second cycles, one can see that $i = 1, 2, \ldots, k(n+1) - 1$.
- Running time supplement x_j ($j = 1, 2, \ldots, n$) will denote the supplement on trip j (between stations $j-1$ and j, regardless of the cycle), and x_0 (i.e., $j = 0$) will denote the supplement for linking periods (regardless if the linking period is at station 0 or n).

This way, to every index i defined as above, one can uniquely assign an index $j = j(i)$. This notation is consistent with the notation of the core model, and it will be practical in the cyclic model. Then one can easily see that the formula for the delay in the departure of the train from station n in the beginning of the second cycle is as follows:

$$d_{n+1} = (d_n + \omega_{n+1} - x_0)^+.$$

Finally, assuming for now that the planned break length x_0 is the same in each cycle, and with the index j belonging to i determined as above, one gets the following general formula for the delays (with $d_0 \equiv 0$, which could be relaxed by introducing a linking period with a disturbance and a supplement before trip 1 in the first cycle):

$$d_i = (d_{i-1} + \omega_i - x_j)^+, \quad i = 1, 2, \ldots, k(n+1) - 1. \tag{5}$$

Since it is reasonable to assume that the distribution of a random disturbance ω_i only depends on the index j, we assume that ω_i ($i = 1, 2, \ldots, k(n+1) - 1$) is exponentially distributed with a rate

λ_j $(j = 0, 1, \dots, n)$ that is characteristic for the given trip or the linking period. Furthermore, in the simplest version of the model, the collection $\{\omega_i\}_{i=1}^{k(n+1)-1}$ is assumed to be independent. With this notation, the decision problem of the railway company may be modeled as follows:

$$\min_{x \geq 0} \left\{ E_\omega[v(x, \omega)] : \sum_{j=0}^{n} x_j \leq M \right\}, \tag{6}$$

$$\text{where} \quad v(x, \omega) = \sum_{i=1}^{k(n+1)-1} d_i(x, \omega).$$

To transform (6) into a two-stage recourse model, one can construct a second-stage problem just like (3): the only difference is that the variables x_j $(j = 0, 1, \dots, n)$ show a cyclic pattern in the equations now. To describe this pattern, let I denote the $n \times n$ identity matrix and $\tilde{I}$ denote the same matrix flipped around its central vertical axis, and let W be the $k(n+1) \times k(n+1)$ version of the matrix W from Section 1. The cyclic pattern in the model is captured by the $k(n+1) \times (n+1)$ matrix T as follows:

$$T = \begin{pmatrix} 0 & I \\ 1 & 0^T \\ 0 & \tilde{I} \\ 1 & 0^T \\ \vdots & \vdots \end{pmatrix}$$

Using this notation, the cyclic model may be summarized concisely as:

$$\min_{x \geq 0} \left\{ E_\omega[v(x, \omega)] : 1x \leq M \right\}, \tag{7}$$

$$\text{where} \quad v(x, \omega) = \min_{y \geq 0} \left\{ 1y : Tx + Wy \geq \omega \right\}.$$

Finally, exactly as in the core model, one can verify that (7) has complete and sufficiently expensive recourse.

4 Example: Extending the Haarlem-Maastricht model

To answer the questions that motivated us to come up with the cyclic model, we solved the cyclic extension of the Haarlem-Maastricht case study from Section 2. We used $k = 10$ cycles instead of 34, which is the number of cycles in the actual timetable (see `www.ns.nl`), since our experiments indicated that the optimal allocations with 4, 8 and 10 cycles were practically the same. Additionally, we introduced a constraint $x_0 \geq 5$ and increased the total amount of supplements M by 5, since it seemed to be realistic to include a minimum break between two cycles. Since departure delays are not especially important, we assigned second-stage objective coefficients 0.01 to the linking periods, whereas the trips had objective coefficients 1. Just like before, we used the external sampling technique combined with the Regularized Decomposition algorithm to solve the model. The results of the cyclic model and the core model are summarized in Table 3 and Figure 6. Although the optimal allocations are apparently not very far from each other (their mean absolute deviation is 14 seconds), certain differences may be noticed, as well: namely, the supplements in the middle of the line are almost the same, whereas the initial and final trips in the cyclic model get less and more supplements, respectively. Fortunately, the results are in line with common sense: in the case of only one cycle, it is very important to prevent delays in the first trips, since they would be carried on to the rest of the line, whereas in a cyclic model, the roles of the first and last trips are nearly the same, and they are distinguished from the middle trips as being close to the linking periods. One sees that no extra supplement beyond the minimum 5 minutes is allocated to the linking periods, indicating that a break of 5 minutes is sufficient for the train to compensate for delays from the past.

	Cyclic model	Core model
Haarlem - Amsterdam Centraal	0.39	0.89
Amsterdam Centraal - Duivendrecht	0.86	1.02
Duivendrecht - Utrecht Centraal	1.47	1.43
Utrecht Centraal - 's Hertogenbosch	2.4	2.68
's Hertogenbosch - Eindhoven	1.63	1.64
Eindhoven - Roermond	2.55	2.49
Roermond - Sittard	1.16	0.77
Sittard - Maastricht	0.48	0
Linking period	5	-
Objective value in cyclic model	107.99	115.64
Computation time	7 min 45 sec	10.4 sec

Table 3: Optimal solutions of the cyclic and non-cyclic models

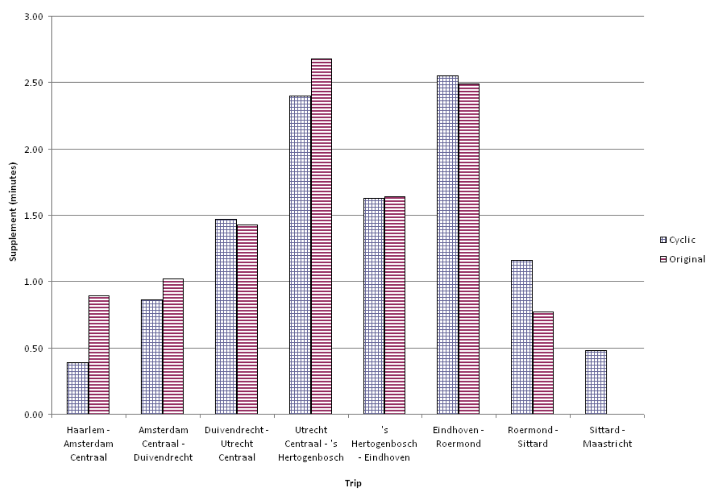

Figure 6: Optimal allocation patterns in the cyclic and non-cyclic models

The fact that cyclicity does not perturb the optimal allocation completely is a sign of robustness, which increased our confidence in the results. Although the computation time was much higher this time than in the core model, it is still tolerable in practice. The optimal objective value of the core model in Table 3 was obtained by solving the cyclic model with the supplements fixed at the optimal solution of the core model. Comparing the objective values, one sees that implementing the optimal solution of a non-cyclic model increases the total amount of expected delays by 7.08%, which is highly significant, so we conclude that it appears to be more convenient to use cyclic models in practice.

5 Sensitivity analysis: The role of the total amount of supplements

The value of the total amount of supplements M is arbitrary in some sense, so the railway company might well raise the question whether the current value is reasonable or not. To investigate this question, we analyzed the sensitivity of the cyclic Haarlem-Maastricht model with respect to the parameter M. Our results are summarized in Table 4 and Figure 7. Our numerical results show that the expected sum of the delays is highly sensitive with respect to changes in the parameter, so increasing the value of the parameter is worth considering in practice. At the same time, our numerical results as well as Figure 7 indicate that the optimal allocation pattern proves to be

M	10	10.93 (original)	11.5	13
Optimal obj. value	138.43	107.99	93.9	66.9

Table 4: The sum of expected delays for different values of M

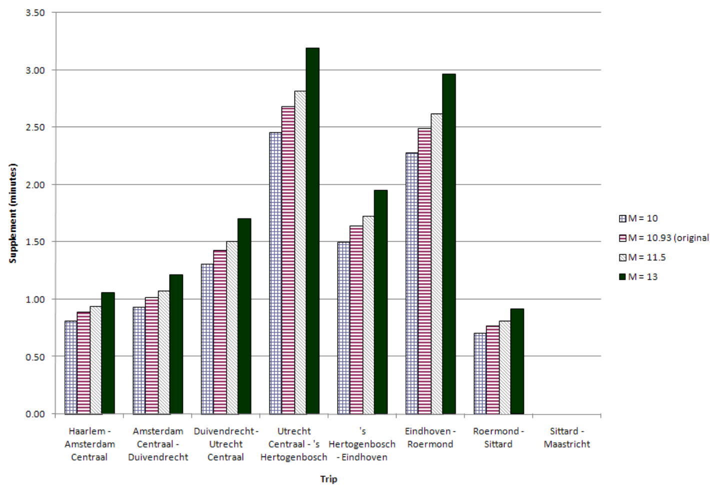

Figure 7: Optimal allocation patterns for different values of M

robust in the sense that the shares of the trips from the total amount remain unchanged under different values of M (even though for extreme values of M which eliminate all delays, obviously alternative optima exist).

6 Modeling a network with several trains

A significant deficiency of the models presented so far is that they only model a single train, whereas in reality, complex interactions among trains have to be taken into account. We classify our ideas on how to model this into two main model types.

6.1 Exact models

A straightforward extension of our models is to include further trains with their own disturbances. These trains can be identical (sharing the same route and supplements) as well as different kinds of trains. Adding identical trains to the model is somewhat less complicated: in this case, there is a planned time difference between the trains, which is constant at all stations. To enforce that the trains keep a certain safety distance at all times, one needs to introduce extra constraints on their delays. Additionally, in a more sophisticated model, one needs to consider trains with routes intersecting the train line in the model. Even though combining cyclicity with a large number of trains resembles the real decision problem better, the complexity of the model becomes enormous even after adding a modest number of cycles and trains, so we looked for a different solution.

6.2 Models with aggregate disturbances

The last remark about complexity foreshadows two serious problems about the exact approach. Firstly, computation times might become explosive, and limits of computational power may be reached before the model can even just approximately represent the problem in its full complexity. More importantly, a more complex model does not necessarily provide better answers to the original question: too many parameters may turn the model into an unpredictable black box, thereby threatening the applicability of the entire model. Additionally, using an exact model, one faces an estimation problem. To understand this, we consider the following formula for d_j in a model with several trains (without distinguishing between the indices i and j as in the cyclic model here):

$$d_j = (d_{j-1} + \zeta_j + \omega_j - x_j)^+, \tag{8}$$

where ζ_j denotes disturbances due to other trains (i.e., waiting times, which have not been present in the models so far) and ω_j now denotes disturbances that are independent from other trains. The problem is that railway companies only measure delays, from which it is impossible to infer about ζ_j and ω_j simultaneously. Thus, these models turn out to be inappropriate in practice, since one may not be able to estimate the model parameters from empirical data. A possible solution of the problem is to merge ζ_j into ω_j. This idea renders the estimation problem solvable, and at the same time, it evades the difficulties of exact models by modeling only one train with a cyclic timetable, with the rest of the trains being implicitly present in the disturbances. Even though this solution assumes that the waiting times for other trains remain unchanged under different supplement allocations, it might still yield a good approximation of the optimal solution.

7 Introducing a heavy-tailed disturbance distribution

In all sections so far, the disturbances in the models have been assumed to be exponentially distributed. This choice seems to be verified by an empirical study quoted in [9], which concluded that measured disturbances on a particular line fit sufficiently well against the exponential distribution. Nevertheless, in this section, a new distribution is introduced for the following two reasons:

- Since distributions can only be estimated with a statistical error, it is indispensable to test the robustness of the model with respect to the choice of distribution.
- More importantly, even though the shape of the pdf of the exponential distribution seems to be intuitively acceptable, computing some probabilities from this pdf shows that the exponential decline in probabilities further and further away from 0 largely contradicts practical experience: e.g., if the mean disturbance on a trip is 1 minute then assuming an exponential distribution, the probability that a disturbance is larger than 10 minutes is as little as $4.5 \cdot 10^{-5}$.

Instead of the exponential distribution, it seems to be appropriate to assume a heavy-tailed one which allows for larger disturbances. As a test, we tried the following pdf:

$$f(x) = \begin{cases} \frac{m^2}{(m^2+x^2)^{\frac{3}{2}}} & \text{if } x \geq 0; \\ 0 & \text{otherwise.} \end{cases}$$

Magnitude of disturbance (mins)	Heavy-tailed	Exponential
> 3	5.42%	5.43%
> 5	2.06%	0.78%
> 10	0.53%	0.006%
> 15	0.23%	$4.74 \cdot 10^{-7}$

Table 5: Probabilities of some large disturbances between Haarlem and Amsterdam Centraal using two different distributions

This distribution has finite mean m but no variance and moment-generating function. We computed the probabilities of some large disturbances in our model between Haarlem and Amsterdam Centraal (the mean disturbance is 1.03 mins) using the two different distributions (see

Table 5). The very slow decline in the tail probabilities of the heavy-tailed distribution is clearly much more acceptable in practice.

We solved the non-cyclic Haarlem-Maastricht model (cyclicity is not really important this time) again with a statistical sample of 5000 realizations from the joint distribution of the new random variables. We set the means of the disturbances equal to the ones in the original model, and additionally, since the purpose of this section is to adequately model large disturbances, we did not apply any truncation this time. Our results are shown in Table 6.

	Exponential	Heavy-tailed
Haarlem - Amsterdam Centraal	0.89	0.76
Amsterdam Centraal - Duivendrecht	1.02	1.02
Duivendrecht - Utrecht Centraal	1.43	1.32
Utrecht Centraal - 's Hertogenbosch	2.68	2.54
's Hertogenbosch - Eindhoven	1.64	1.61
Eindhoven - Roermond	2.49	2.69
Roermond - Sittard	0.77	0.99
Sittard - Maastricht	0	0
Obj. value in exponential	8.46	8.51
Increase in obj. value compared to exponential	0	0.59%

Table 6: Optimal solutions using the exponential and heavy-tailed distributions and the objective function values of these solutions evaluated in the model using exponential distributions.

The two solutions are apparently quite similar. This indicates the robustness of the model with respect to truncation and the choice of distribution, thereby greatly increasing the credibility of our results. The optimal objective function values of the two solutions in the original model are also very close to each other, and the computation times did not differ significantly.

8 An approximating non-linear programming problem

For practical purposes, it may be useful to use another problem that may be solved more easily, without the need of special software designed for stochastic programming problems, yet returns nearly the same solution as the original model. In this section, a problem of this type is presented for the core model, which may easily be generalized to the cyclic model.

Recalling that $d_1 = (\omega_1 - x_1)^+$ and assuming that ω_1 is exponentially distributed with mean m_1, one may verify that for any $x_1 \geq 0$,

$$E(d_1) = m_1 e^{-\frac{x_1}{m_1}}. \tag{9}$$

From Section 1, let us further recall the formula for the delay on an arbitrary trip $j = 2, 3, \ldots, n$:

$$d_j = (d_{j-1} + \omega_j - x_j)^+.$$

In this formula, unfortunately, $d_{j-1} + \omega_j$ is certainly not exponentially distributed even if ω_j is. However, since we saw in Section 7 that the optimal solution was sufficiently robust with respect to assumptions on the distribution, we use this false assumption for the distribution but the correct mean

$$n_j := E(d_{j-1} + \omega_j) = E(d_{j-1}) + m_j \tag{10}$$

to arrive at the expression

$$E(d_j) = n_j e^{-\frac{x_j}{n_j}}. \tag{11}$$

Assuming that the means $\{m_j\}_{j=1}^n$ are known, $E(d_1)$ is known from (9). This and (10) give the value of n_2, which gives the value of $E(d_2)$ by substituting into (11). This can be repeated for all values $j = 3, 4, \ldots, n$ to arrive at a closed form expression for the objective function $E(d_1) +$

$E(d_2) + \ldots + E(d_n)$, which has to be minimized under the original constraints $x_1, x_2, \ldots, x_n \geq 0$ and $x_1 + x_2 + \ldots + x_n \leq M$. This is a convex, deterministic, non-linear optimization problem, which can be solved easily. Additionally, one is not restrained to the exponential distribution: the same procedure can be mimicked entirely e.g. for the heavy-tailed distribution introduced in Section 7 by using the relationship $E(d_1) = \sqrt{x_1^2 + m_1^2} - x_1$, which may be verified to hold for all $x_1 \geq 0$ in this case.

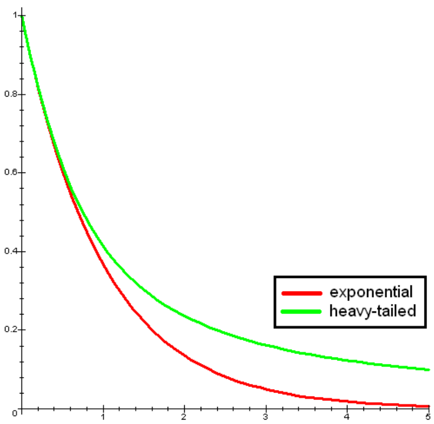

Figure 8: Plots of the function $f(x_1) = E((\omega_1 - x_1)^+)$ in the case $E(\omega_1) = 1$.

The function $f(x_1) = E((\omega_1 - x_1)^+)$ in the case $E(\omega_1) = 1$ is plotted for both distributions in Figure 8. The plot shows that using an exponential distribution, a supplement of 5 minutes is enough to eliminate all delays, whereas using the heavy-tailed one, some delays persist even in case of a large supplement. The latter is apparently far more realistic than the former.

To test the idea of approximating NLP problems, we solved the Haarlem-Maastricht model with both distributions and arrived at the results in Table 7. It can be noted from the last row of the table that the approximating solutions are really excellent in terms of their objective function values (with an increase of less than 1% compared to the original value). In addition, both models were solved by the built-in Solver add-in of MS Excel within just a couple of seconds. Hence we conclude that these approximating problems appear to be useful in practice.

	Original	Exponential	Heavy-tailed
Haarlem - Amsterdam Centraal	0.89	0.98	0.88
Amsterdam Centraal - Duivendrecht	1.02	1.17	1.12
Duivendrecht - Utrecht Centraal	1.43	1.52	1.47
Utrecht Centraal - 's Hertogenbosch	2.68	2.39	2.28
's Hertogenbosch - Eindhoven	1.64	1.94	1.94
Eindhoven - Roermond	2.49	2.18	2.22
Roermond - Sittard	0.77	0.75	1.00
Sittard - Maastricht	0	0	0
Obj. value in original	8.46	8.54	8.54

Table 7: Optimal solutions of the original SLP and the approximating NLP problems with exponential and heavy-tailed distributions, and the objective function values of these solutions evaluated in the original model.

9 Estimation of the disturbance distributions from empirical data

At the Dutch Railways, realized traveling times are measured at several stations on a regular basis. However, from these data, one can only compute the delays directly, whereas in our models, one needs to specify the disturbance distributions. Therefore, to obtain realistic input data for optimization models, the only possibility is to try to gain inference about the distribution of the non-observable disturbances from data measured by railway companies (i.e., realized and planned traveling times and the values of the supplements). Assuming that the disturbances follow a parametric distribution (e.g. exponential or the heavy-tailed distribution presented in Section 7), this section describes how these data may be used to obtain maximum likelihood estimates of the unknown parameter(s).

Realizations of random variables will be denoted by hats, e.g. $\hat{\omega}_1$. Using this notation and assuming a model with aggregate disturbances, Equation (8) for the delay of the train on trip j now reads as follows for a specific realization:

$$\hat{d}_j = (\hat{d}_{j-1} + \hat{\omega}_j - x_j)^+ \tag{12}$$

This cannot be solved for $\hat{\omega}_j$ if $\hat{d}_j = 0$, since the plus operator is not an invertible function. What one can infer about $\hat{\omega}_j$ from (12) is the following:

- Case 1: If $\hat{d}_j \neq 0$ then $\hat{\omega}_j = \hat{d}_j - \hat{d}_{j-1} + x_j$.
- Case 2: If $\hat{d}_j = 0$ then $\hat{\omega}_j \leq x_j - \hat{d}_{j-1}$.

Assuming that several observations of $\hat{d}_j$ are available, if the model uses truncation then first one needs to exclude the values above the truncation level, and then proceed as follows:

- After partitioning the (possibly truncated) support of ω_j into a predetermined set of disjoint subintervals $\{S_i\}_{i=1}^s$ with positive probabilities, and denoting the probability of an observation being in subinterval S_i ($i = 1, \ldots, s$) by $p_i := P(\omega_j \in S_i)$, one can classify all values as in Case 1 into these subintervals, with n_i being the number of observations in S_i.
- The probability of an observation as in Case 2 is $q_i := F_j(x_j - \hat{d}_{j-1})$ ($i = 1, 2, \ldots, t$), where the index i goes through the t observations falling into Case 2 while it is suppressed on the right hand side of the equation and $F_j(x)$ denotes the cdf of ω_j.

Then the log-likelihood of an IID sample is as follows:

$$L = \sum_{i=1}^{s} n_i \log p_i + \sum_{i=1}^{t} \log q_i.$$

This only depends on the parameter(s) of the distribution, and it can be maximized numerically to obtain the required estimate(s). This estimation procedure completes the process leading from observing the statistical data to an optimal supplement allocation.

10 Ideas for further improvement of the models

In this section, we discuss some further ideas to increase the applicability of our models:

- *Punctuality* means the probability that the delay at a randomly chosen station in a randomly chosen cycle remains below a given threshold. This concept is closely related to the expected sum of delays, but minimizing the two quantities generally yield different optimal allocations. Although our models only minimize the expected sum of delays, they also enable us to compute the punctuality of any solution for any threshold, since the second-stage variables attain the values of the delays in the optimum, as we showed in Section 1. This information is highly valuable for the Dutch Railways, which measures reliability in terms of punctuality instead of average delays, and publishes punctuality statistics regularly at `www.ns.nl`. In fact, some years ago one of the CEO's of the Dutch Railways stepped down because the desired punctuality level had been missed by as little as 1%.)

  For the sake of illustration, we computed in the cyclic Haarlem-Maastricht model with the heavy-tailed distribution from Section 7 that the punctuality in the model was 86.0% using a 3-minute threshold and 91.8% using a 5-minute threshold. These values seem to be very realistic if one compares them to tyipical values measured by the Dutch Railways.

- Instead of the unweighted sum of expected delays, it is more realistic to use a *weighted sum as an objective function*. A weight can be assigned to each station by using ticket sales statistics or some other measure of the importance of the station on the railway line. Additionally, it is reasonable to assume that passengers are not especially worried about departure delays, so the linking periods should have small objective coefficients (yet the coefficients have to be positive so that the second-stage objective function is monotone increasing in all variables: this guarantees that the recourse decision variables attain the values of the delays in the optimum). These weights may also be used to compute a weighted version of punctuality by assuming that different stations are drawn with different probabilities. We have experimented with weighted objective functions, and using different weights resulted in significantly different optimal decisions. For practical applications, we recommend using weighted objective functions.

- Our models have implicitly assumed so far that the running time disturbance on a trip does not depend on the delay accumulated so far. However, it appears reasonable to assume that *drivers increase the speed of the train in case of a delay*, at least within reasonable limits. A possible way of modeling this phenomenon may be to assume that the disturbance on a trip consists of a stochastic component as before, reduced by the product of some fixed coefficient and the delay of the train accumulated in the past. However, in this case, one has to either invent the value of the coefficient or develop a technique to estimate the coefficient and the parameter(s) of the disturbance distribution simultaneously.

- By allowing *different disturbance distributions for the same trip in different cycles*, our models can make a distinction between rush and idle hours. Additionally, the assumption of independent disturbances may be dropped from the model to make it more realistic. Given the necessary statistical data, the technique described in Section 9 may be used to estimate the parameters.

Our results seem to support the idea that using stochastic programming models may be an effective way of optimizing real-life railway timetables.Besides railway timetables, the methods described in this paper can be applied to several other means of transportation, as well. More generally, our ideas may be useful in any timing problem that is aimed at improving the organization of a series of tasks repeated multiple times.

A Two-stage recourse models in stochastic programming

Two-stage recourse models were first formulated by George B. Dantzig in 1955 in [2]. Nowadays, they constitute one of the most fundamental and widely applied model classes in the field of stochastic programming. We included this section for readers who are not especially familiar with these models. The interested reader may find ample information on the subject e.g. in the textbooks [1] and [4] as well as the website [8].

A.1 Definitions and properties

Let ω be a random vector with a known distribution and support $\Omega \subset \mathbb{R}^m$. Assuming that $x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$, $\omega \in \mathbb{R}^m$ and the remaining vectors and matrices in (13) are of conformable dimensions, we define the two-stage linear recourse model with random right-hand side as the following optimization problem:

$$\min_{x \geq 0} \; \{cx + E_\omega[v(x,\omega)] : Ax \leq b\}, \text{where} \tag{13}$$

$$v(x,\omega) = \min_{y \geq 0} \{qy : Tx + Wy \leq \omega\}.$$

The following two regularity conditions are used in the paper:

- We say that (13) has a complete recourse structure (CR) if for all $z \in \mathbb{R}^m$ there exists a $y \geq 0$, possibly depending on z, such that $Wy \leq z$.
- In addition, (13) has a sufficiently expensive recourse structure (SER) if there is no $z \in \mathbb{R}^m$ such that $\min_{y \geq 0} \{qy : Wy \leq z\} = -\infty$.

These conditions are sufficient for the function $v : \mathbb{R}^{n+m} \to \mathbb{R} \cup \{-\infty, \infty\}$ to be finite for all arguments. Additionally, under the assumptions of CR, SER and all components of $E_\omega[|\omega|]$ being finite, it is well known that $Q(x) = E_\omega[v(x,\omega)]$ is a finite-valued, convex function. Moreover, it is polyhedral (piecewise linear) if the distribution of ω is discrete, and everywhere differentiable if the distribution of ω is continuous. Since the objective function of (13) only differs from $Q(x)$ in a linear function, it follows that (13) amounts to the minimization of a finite, convex function over a polyhedral set.

A.2 Solution algorithms

Even though (13) possesses all the nice properties of a convex optimization problem, solving it by means of general convex programming algorithms is extremely demanding, if not impossible, for computational reasons. Nevertheless, if Ω has finitely many elements ($\Omega = \{\omega^1, \omega^2, \ldots, \omega^S\}$ and $p_s = P(\omega = \omega^s)$ $(s = 1, 2, \ldots, S)$) then by introducing a collection of recourse decision vectors $y^1, y^2, \ldots, y^S \in \mathbb{R}^p$, it is easy to see that (13) is equivalent to the following deterministic linear programming problem (called the deterministic equivalent):

$$\min_{x, y^1, \ldots, y^S \geq 0} \begin{array}{llllllll} \{cx & + \; p_1 \cdot qy^1 & + \; p_2 \cdot qy^2 & + & \ldots & + \; p_S \cdot qy^S : & & \\ Ax & & & & & & \leq & b \\ Tx & + \; Wy^1 & & & & & \leq & \omega^1 \\ Tx & & + \; Wy^2 & & & & \leq & \omega^2 \\ \vdots & & & & \ddots & & \vdots & \vdots \\ Tx & & & & & + \; Wy^S & \leq & \omega^S\}. \end{array} \tag{14}$$

Even though LP problems are easy to solve, a significant disadvantage of this formulation is its enormous size. Fortunately, there exist some advanced solution algorithms that benefit from the special structure of two-stage recourse models instead of solving the large-scale deterministic equivalent as an LP problem. As we point out in an example in Section 1, by being better adapted to this specific model type, the application of these algorithms may improve computational efficiency substantially. The only solution algorithm we consider in this paper is a variant of the so-called *L-shaped algorithm* (see [7]). It can be proved that this algorithm finds an optimal solution of (13) in finitely many iterations, assuming that there is one.

References

[1] J.R. Birge and F. Louveaux. *Introduction to stochastic programming.* Springer-Verlag, New York, 1997.

[2] G.B. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.

[3] P. Kall and J. Mayer. SLP-IOR: An interactive model management system for stochastic linear programs. *Math. Program.*, 75(2):221–240, 1996.

[4] A Prékopa. *Stochastic Programming,* volume 324 of *Mathematics and its Applications.* Kluwer Academic Publishers Group, Dordrecht, 1995.

[5] A. Ruszczyński. A regularized decomposition method for minimizing a sum of polyhedral functions. *Math. Program.*, 35:309–333, 1986.

[6] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczynski and A. Shapiro, editors, *Stochastic Programming*, chapter 6, pages 353–425. Elsevier, 2003. Handbooks in Operations Research and Management Science, vol. 10.

[7] R. Van Slyke and R.J-B. Wets. L-shaped linear programs with applications to control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17:638–663, 1969.

[8] Stochastic Programming Community Home Page. `www.stoprog.org`.

[9] M.J.C.M. Vromans. *Reliability of Railway Systems.* PhD thesis, ERIM, Erasmus University Rotterdam, The Netherlands, 2005.

[10] Website of the Dutch Railways (Nederlandse Spoorwegen). `www.ns.nl`.