A Rolling Stock Circulation Model for Combining and Splitting of Passenger Trains

Pieter-Jan Fioole

NS Reizigers, Utrecht, The Netherlands,
E-mail: PJ.Fioole@reizigers.ns.nl

Leo Kroon*

NS Reizigers, Utrecht and
Erasmus University, Rotterdam
P.O.Box 1738, 3000 DR Rotterdam, The Netherlands
E-mail: L.Kroon@fbk.eur.nl

Gábor Maróti

CWI, Amsterdam and
NS Reizigers, Utrecht
P.O.Box 94079, 1090 GB Amsterdam, The Netherlands
E-mail: G.Maroti@cwi.nl

Alexander Schrijver

CWI, Amsterdam and
University of Amsterdam
P.O.Box 94079, 1090 GB Amsterdam, The Netherlands
E-mail: Lex@cwi.nl

Abstract

This paper addresses the railway rolling stock circulation problem. Given the departure and arrival times as well as the expected numbers of passengers, we have to assign the rolling stock to the timetable services. We consider several objective criteria that are related to operational costs, service quality and reliability of the railway system.

Our model is an extension of an existing rolling stock model for routing train units along a number of connected train lines. The extended model can also handle underway combining and splitting of trains.

We illustrate our model by computational experiments based on instances of *NS Reizigers*, the main Dutch operator of passenger trains.

*Corresponding author

1 Introduction

This paper addresses the railway rolling stock circulation problem, thereby also taking into account the underway combining and splitting of trains. Given the departure and arrival times as well as the expected numbers of passengers, we have to assign the available rolling stock to the timetable services. We optimize different objective criteria that are related to operational costs, service quality and reliability of the railway system.

The problem we focus on is to determine the rolling stock circulation for a generic week. That is, it arises in an early phase of the railway planning process. In later stages, this generic plan must be adjusted to the specific calendar days, and shunting plans must be created. Finally, the traffic control must carry out these plans, and adapt them in case of disturbances.

A main complicating issue is the fact that the available train units are of different types. This forces one to keep track of the positions of the different train units within each train. This is particularly important in the case of combining and splitting trains.

In this paper we give an integer programming model. It deals with an extension of the problem described by Peeters and Kroon [9]. Their branch-and-price approach strongly uses the fact that trains run up-and-down a number of connected train lines. However, in the case of underway combining and splitting of trains, this property is lost if several branches of the lines have different lengths, as it happens in real-life instances of the Dutch passenger railway operator *NS Reizigers*. The model we are describing here can be applied for such instances. We used the commercial mixed integer programming software CPLEX to solve our model.

So far, the rolling stock plans of NS Reizigers were created manually. It turned out that our model is capable to provide solutions that can be implemented in practice. One of our main achievements is that the rolling stock circulation for the timetable year 2005 is based on solutions of our model.

This paper is organized as follows. We describe the problem in detail in Section 2. A literature overview is given in Section 3. In Section 4 we formulate our model. Section 5 contains our computational results. Finally, in Section 6 some conclusions are drawn.

2 Problem description

2.1 The Noord-Oost case

The train line system that is studied in this paper is the so-called Noord-Oost case, involving the train lines 500, 700, 1600 and 1700 of NS Reizigers. These train lines form the backbone of the Dutch railway system, by connecting the cities Amsterdam (Asd), Schiphol (Shl), Rotterdam (Rtd), and The Hague (Gvc) in the Western part of the Netherlands to the cities Leeuwarden (Lw), Groningen (Gn), and Enschede (Es) in the Northern and Eastern part of the country. Utrecht (Ut), Amersfoort (Amf), Deventer (Dv), and Zwolle (Zl) are important underway stations. (See Figures 1 and 2.)

The train lines are operated by a given number of self-propelled train units with 3 or 4 carriages each. Train units can be composed to a longer train in any possible order. The departure and arrival times of the train services have already been specified in an earlier stage of the planning process. The timetable of the involved lines is more or less periodic with a period of one hour.

Almost all trains in the Noord-Oost are *combined* or *split* at certain locations. For example, a train in the 1600 line arriving in Amersfoort from Enschede is split into two parts. The front part continues to Amsterdam, while the rear part continues to Schiphol. On the way back, the train arriving from Schiphol becomes the front part, it is combined with the train arriving from Amsterdam. Trains in the 1700 line are combined or split in Utrecht. Trains in the 500 line

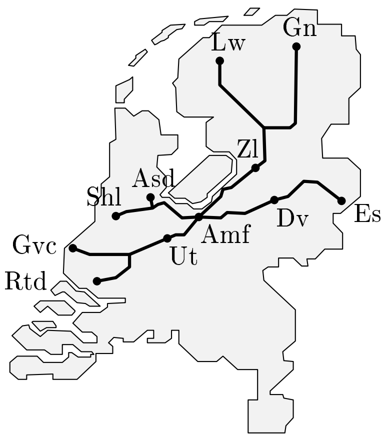

Figure 1: The Noord-Oost train lines on the Dutch railway map

are combined or split both in Utrecht and in Zwolle, and trains in the 700 line are combined or split both in Amersfoort and in Zwolle. Note that the timetable always specifies from which stations the front and the rear part of a combined train arrive, and towards which stations the front and rear part of a split train depart.

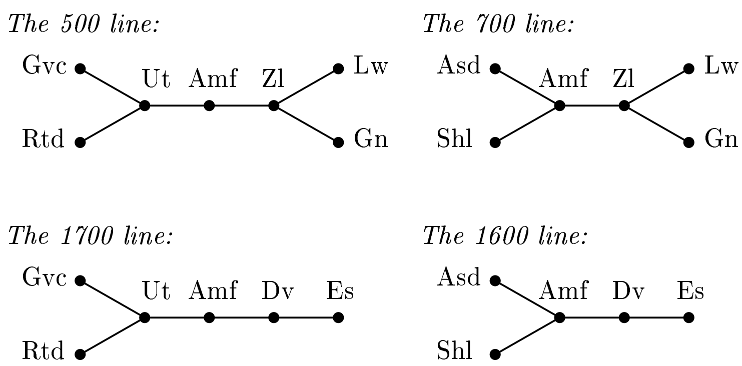

Figure 2: The Noord-Oost train lines

Our task is to assign the train units to the timetable services such that various technical and market constraints are fulfilled. In this study we focus on rolling stock schedules for each day of the week separately, and we use constraints such that solutions for consecutive days can be attached to each other. Computational power is the only limit when considering planning periods longer than one day.

One of the complicating issues in rolling stock management is maintenance of rolling stock units. Maintenance is, however, no requirement in the medium-term rolling stock planning at NS Reizigers. In the Netherlands, timetables and rolling stock circulations are *dense*, thus sensitive to disturbances. So, preset maintenance schedules have little chance to be carried out. Therefore, maintenance is planned in the operational phase.

2.2 Objectives

Creating an appropriate rolling stock schedule means finding a balance between several objectives, such as minimizing: (*i*) the number of carriage kilometers (efficiency), (*ii*) the amount of seat shortages (service), and (*iii*) the number of shunting movements (robustness). The model described in this paper allows to make a trade-off between these objectives. The objectives are explained below in detail.

The operational cost of using the rolling stock depends on traction power, but also on maintenance of the rolling stock: after a certain number of kilometers, each train unit is directed to a maintenance station for a preventive check-up and repair. In both cases, the number of carriage kilometers is an appropriate measure for variable rolling stock costs.

The input of our problem contains the estimated number of first and second class passengers for every timetable service. Good service quality means among others that all passengers, in particular first class passengers, have a seat during their journey. However, during rush hours, there may be more passengers than seats. Outside rush hours, rolling stock capacity is usually sufficient to provide all passengers with a seat. Our measure for the seat shortages is the expected number of passengers without a seat multiplied by the length of the involved trip, and added up for all trips. We call it *seat-shortage kilometers*.

Train compositions can be modified at certain stations by coupling train units to a train or by uncoupling train units from it. Changing the composition of a train is a complex process. It requires a number of shunting movements between the platform area and the shunting area of a station. This may lead to disturbances of the regular train operations. Thus a smaller number of shunting operations may increase the robustness of the railway system. Moreover, the shunting costs are also related to the crew: each shunting movement requires a train driver. On the other hand, changing the train compositions between two rides may allow to use the rolling stock more efficiently and also to decrease seat shortages.

3 Literature review

In the literature, several related problems have been studied, apart from the paper by Peeters and Kroon [9] that was mentioned earlier. These models are often set up for locomotive-hauled railways carriages, and do not consider the order of the carriages (train units). Moreover, no optimization methods were developed so far to handle combining and splitting train units.

Schrijver [10] considers the problem of minimizing the number of train units needed to satisfy passengers' seat demand. The model computes for every trip the number of train units of each type to be used, not taking the order of the units in a train composition into account. Seat shortages and the number of shunting movements are not considered.

Alfieri et al. [2] describe an integer programming model to determine the circulation of rolling stock (including the order of the train units in the compositions) on a single train line and a single day. The objective is to minimize the number of train units needed to satisfy the demand; seat shortages and the number of shunting operations are not taken into account. Their approach was tested on real-life case-studies of NS Reizigers.

Brucker et al. [5] study the problem of routing single railway carriages through a network: The carriages should be used in timetable services or empty trains such that each timetable service can be operated with a given number of carriages. The order of the carriages is not considered. Their solution approach is based on local search techniques like simulated annealing.

Ben-Khedher et al. [4] study the problem of allocating identical train units to the French High Speed Trains. Their rolling stock allocation system is based on a capacity adjustment model that is linked to the seat reservation system and seeks to maximize the expected profit for the company.

Cordeau et al. [6] present a Benders decomposition approach for the locomotive and carriage assignment problem. Computational experiments show that optimal solutions can be found quickly. In a subsequent paper, Cordeau at al. [7] extend their model with various real-life constraints, such as maintenance. Neither of these models considers the order of the carriages in the train compositions in detail.

Lingaya et al. [8] describe a model for operational management of locomotive-hauled railway carriages. They explicitly take the order of the carriages in the trains into account. Several real-life constraints, such as maintenance, are considered. The solution approach is based on a Dantzig-Wolfe reformulation solved by column generation.

Abbink et al. [1] present an integer programming model for distributing the available rolling stock between the train lines. This model does not determine an actual schedule. Instead, seat shortages during the morning rush hours are minimized by allocating rolling stock to the trains running at 8 o'clock in the morning, the busiest moment of the day. This model can provide the number of available train units when the rolling stock scheduling problem is solved for each train line separately.

4 Model Formulation

In this section we first describe the model without combining and splitting trains. However, it is possible to modify train compositions by coupling or uncoupling train units before the train departs, underway, and after the train arrives. Thereafter we extend the basic model in Section 4.4 with combining and splitting of trains.

In our model, we assume that train units stored at a station can be coupled later to any of the departing trains from that station, irrespective of the arrival order of the stored train units. This explains the difference between splitting and uncoupling: In case of splitting, the split parts serve in passenger trains. In particular, the order of the train units in the split parts does matter. However, uncoupled train units are transferred to the shunting area at the station, and there the order of the train units does not matter anymore. The re-allocation time described below reserves time for shunting. The difference between combining and coupling is similar.

4.1 The model without combining and splitting

Let M be the set of train unit types. For each $m \in M$, let n_m denote the number of available train units of type m.

A *composition* of train units is an ordered sequence of elements of M. We assume that the left-hand side of the string corresponds to the front of the train. In the Noord-Oost case, we have two types: units with 3 or 4 carriages. The order of the units in a train composition plays an important role. For example, train compositions 334 and 343 have the same capacity, but they differ in shunting possibilities. The train unit with 4 carriages can be much more easily uncoupled from composition 334 than from composition 343. Let $n_{m,p}$ denote the number of train units of type m in composition p. For a composition p, let $\beta(p) \in \mathbb{Z}^M$ be such that $\beta(p)_m = n_{p,m}$ for each train unit type $m \in M$.

A *trip* is a sequence of consecutive train movements during which the train composition cannot change. Let T be the set of trips. A trip t is characterized by its departure station $s_d(t)$, arrival station $s_a(t)$, departure time $\tau_d(t)$ and arrival time $\tau_a(t)$.

The set of compositions that are allowed for trip t is denoted by P_t. When determining the sets P_t, the following aspects are taken into account: (i) the capacity of the train should be sufficient to provide a certain service level, and (ii) the length of the train may not exceed a certain upper bound, mainly determined by the lengths of the platforms along the trip.

Let $\nu(t)$ denote the successor trip of t: the train that carries out trip t carries out trip $\nu(t)$ immediately after t, although some units may be (un)coupled. The set T_0 denotes the set of trips that have no defined previous trip, while T_1 denotes the set of trips without defined subsequent trip. We assume that for a trip $t \in T_0$, all train units serving t must be coupled just before t, while for a trip $t \in T_1$, all train units must be uncoupled just after t.

Let Γ_t denote the set of pairs (p, p') such that $p \in P_t$, $p' \in P_{\nu(t)}$, and such that the composition change $p \to p'$ after trip t is allowed. Thus the shunting possibilities are encoded in the sets Γ_t. In instances of NS Reizigers, a general constraint is that either some units may be uncoupled or some units may be coupled but not both. Coupling and uncoupling train units is only allowed at the front or rear side of a train composition, depending on the infrastructure of the station. For example, the train unit of length 4 cannot be uncoupled from a composition 343. On the other hand, if the station allows shunting at the rear side of the train, then a composition change from 334 to 33 is possible.

In practice, a train unit, that has been uncoupled from a train, cannot be coupled immediately onto another train: a certain re-allocation time $\varrho(s)$, depending on the station, has to be respected. We model this as follows. Train units that are stored temporarily at a station can be coupled to a departing train immediately. However, an uncoupled train unit appears at the station as a train unit available only $\varrho(s)$ minutes later. We illustrate this schematically in Figure 3. The station s is represented by a time-line, and we consider a pair of arrival and departure events. (Un)coupled units go through the arcs that connect the time-line to the arrival or departure event nodes.

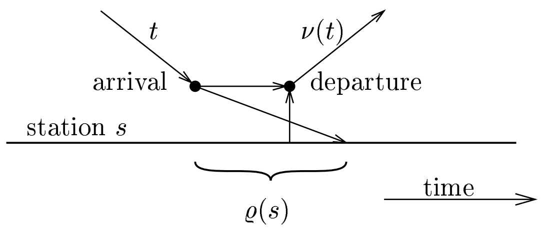

Figure 3: Coupling and uncoupling train units

The main variables are the following:

$X_{t,p} \in \{0,1\}$	whether composition p is used for trip t.
$Z_{t,p,p'} \in \{0,1\}$	whether trip t has composition p and trip $\nu(t)$ has composition p'.

The variable $N_{t,m}$ denotes the number of train units of type m that are used on trip t. We also introduce the variables $C_{t,m}$ and $U_{t,m}$: $C_{t,m}$ expresses the number of train units of type m that have been coupled to the train composition right *before* trip t, and $U_{t,m}$ expresses the number of train units of type m that have been uncoupled from the train composition right *after* trip t.

The variable $I_{t,m}$ denotes the number of train units of type m that are stored (i.e., available for being coupled to a train) at station $s_d(t)$, immediately after the departure of trip t. Note that this inventory value might not be equal to the number of train units that are physically present at the station: Train units that are used both for trip t and $\nu(t)$ are not considered to be stored between the arrival of t and the departure of $\nu(t)$. Moreover, uncoupled train units increase the local inventory after the re-allocation time, while coupled units decrease it immediately.

The variable $I^0_{s,m}$ denotes the number of train units of type m that start the day at station s. The set of stations is denoted by S. Now the basic part of the model reads as follows:

$$\min F(X, Z, N) \tag{1}$$

subject to

$$\sum_{p \in P_t} X_{t,p} = 1 \qquad \forall t \in T \tag{2}$$

$$X_{t,p} = \sum_{p' \in P_{\nu(t)}:\ (p,p') \in \Gamma_t} Z_{t,p,p'} \qquad \forall t \in T \setminus T_1;\ p \in P_t \tag{3}$$

$$X_{\nu(t),p'} = \sum_{p \in P_t:\ (p,p') \in \Gamma_t} Z_{t,p,p'} \qquad \forall t \in T \setminus T_1;\ p' \in P_{\nu(t)} \tag{4}$$

$$N_{t,m} = \sum_{p \in P_t} n_{p,m} X_{t,p} \qquad \forall t \in T;\ m \in M \tag{5}$$

$$C_{\nu(t),m} = \sum_{\substack{(p,p') \in \Gamma_t: \\ n_{p',m} > n_{p,m}}} (n_{p',m} - n_{p,m}) \cdot Z_{t,p,p'} \qquad \forall t \in T \setminus T_1;\ m \in M \tag{6}$$

$$U_{t,m} = \sum_{\substack{(p,p') \in \Gamma_t: \\ n_{p,m} > n_{p',m}}} (n_{p,m} - n_{p',m}) \cdot Z_{t,p,p'} \qquad \forall t \in T \setminus T_1;\ m \in M \tag{7}$$

$$C_{t,m} = N_{t,m} \text{ and } U_{t,m} = 0 \qquad \forall t \in T_0;\ m \in M \tag{8}$$

$$U_{t,m} = N_{t,m} \text{ and } C_{t,m} = 0 \qquad \forall t \in T_1;\ m \in M \tag{9}$$

$$\begin{aligned} I_{t,m} = I^0_{s(t),m} &- \sum_{\substack{t' \in T:\ s_d(t') = s_d(t), \\ \tau_d(t') \leqslant \tau_d(t)}} C_{t',m} \\ &+ \sum_{\substack{t' \in T:\ s_a(t') = s_d(t), \\ \tau_a(t') \leqslant \tau_d(t) - \varrho(s_d(t))}} U_{t',m} \qquad \forall t \in T,\ m \in M \end{aligned} \tag{10}$$

$$n_m = \sum_{s \in S} I^0_{s,m} \qquad \forall m \in M \tag{11}$$

$$X_{t,p} \in \{0,1\} \qquad \forall t \in T;\ p \in P_t \tag{12}$$

$$N_{t,m}, C_{t,m}, U_{t,m}, I_{t,m} \in \mathbb{R}_+ \qquad \forall t \in T;\ m \in M \tag{13}$$

$$I^0_{s,m} \in \mathbb{Z}_+ \qquad \forall s \in S;\ m \in M \tag{14}$$

$$Z_{t,p,p'} \in \mathbb{R}_+ \qquad \forall t \in T;\ (p,p') \in \Gamma_t \tag{15}$$

We describe the objective function in detail in Section 4.2. Constraints (2) state that, for each trip, exactly one allowed composition of train units is to be used. Constraints (3) and (4) guarantee a correct link between the compositions that are used for consecutive trips. Constraints (5) describe the connection between the composition that is used for a certain trip and the numbers of train units of the different types that are used on this trip. Constraints (6) – (9) specify the number of coupled and uncoupled train units. Constraints (10) describe the inventory of train units in the different stations. To compute the value $I_{t,m}$, we consider the number of train units that are stored at the station at the begin of the day, then we increase or decrease it whenever a train unit was uncoupled or coupled until the departure time of trip t.

Constraints (11) handle the correct allocation of the available train units to the initial inventories at the different stations. Finally, constraints (12) – (15) describe the binary or non-negative character of the variables.

One easily verifies the following lemma.

Lemma 1. *In any feasible solution to the model (1) – (15), the variables $U_{t,m}$, $C_{t,m}$, $I_{s,t}$ and $Z_{t,p,p'}$ are integral.*

Proof. The variables $U_{t,m}$, $C_{t,m}$, $I_{s,t}$ and $Z_{t,p,p'}$ are uniquely determined by the integer variables $X_{p,m}$ and $I^0_{s,m}$. □

The composition changes could be modelled without the variables $Z_{t,p,p'}$ but using additional linear constraints on the variables $X_{t,p}$. Such a model is described e.g. in Alfieri et al. [2]. One easily proves that the linear relaxation of their model defines a polytope that strictly contains the linear relaxation of our model. Therefore the branch and bound procedure in our model works with better lower bounds. Moreover, according to Lemma 1, we never have to branch on a variable $Z_{t,p,p'}$. Thus we can expect that the number of nodes in the branch and bound tree becomes smaller. On the other hand, we have to solve larger linear programs. State-of-the-art commercial software such as CPLEX can, however, easily cope with linear programs that appear in our applications. The model described above can be solved (nearly) to optimality for instances of NS Reizigers that were computationally intractable by earlier models.

4.2 Objective function

As was mentioned earlier, the objective function contains three major elements: (*i*) carriage kilometers, (*ii*) seat-shortage kilometers, and (*iii*) the number of shunting movements. Given the above defined decision variables, these elements can be computed easily.

For the total number of carriage kilometers CKM, the following holds:

$$CKM = \sum_{t \in T} \sum_{m \in M} \ell_t \cdot c_m \cdot N_{t,m} \tag{16}$$

where ℓ_t is the length of trip t and c_m is the number of carriages in train unit type m.

The total number of seat-shortage kilometers SKM can be determined as follows:

$$SKM = \sum_{t \in T} \sum_{p \in P_t} \ell_t \cdot s_{t,p} \cdot X_{t,p}. \tag{17}$$

Here $s_{t,p}$ denotes the expected number of seat shortages when composition p is used for trip t. The shortage is computed by comparing the forecasted number of passengers to the capacity of the train compositions.

Coupling and uncoupling at the same time is not allowed for a composition change from p to p'. Therefore the total number of shunting movements SHM

can be determined by adding all the variables $Z_{t,p,p'}$ with $\beta(p) \neq \beta(p')$. That is,

$$SHM = \sum_{t \in T} \sum_{\substack{(p,p') \in \Gamma_t: \\ \beta(p) \neq \beta(p')}} Z_{t,p,p'} \tag{18}$$

We minimize a non-negative linear combination of these criteria, the weight factors reflecting their relative importance. In order to guide the different elements of the objective function, we also introduce upper bounds on CKM, SKM and SHM.

Note that the quality of the solutions can be measured by several other characteristics. For example, the largest seat shortage on a trip should not be too high (e.g. 20% of the passengers of a trip). Seat shortages in the first class can be penalized more heavily than in the second class. Another criterion is to keep the number of shunting movements during rush hours small. The objective function and the constraints can easily be modified in order to take these additional criteria into account.

4.3 New integer decision variables

The variables $X_{t,p}$ describe the exact train composition for each trip t. Now it turns out that, if a certain (binary) decision has been made with respect to the length of a train, then the decision with respect to the detailed composition of the train will be binary automatically. That is, the number of binary decision variables becomes smaller: in the Noord-Oost case, about 9,900 binary variables $X_{t,p}$ are replaced by about 5,700 new binary variables.

The new decision variables determine the *number* of train units per type that is allocated for the trips, without specifying the exact order in the train. For a trip t, we define the set of vectors in $\mathbb{Z}_+^M$

$$B_t = \{\beta(p) \mid p \in P_t\}$$

The set B_t contains the possible train lengths for trip t. We introduce the binary variables

$Y_{t,b} \in \{0,1\}$ whether combination $b \in B_t$ is used for trip t.

The connection to the already defined variables $X_{t,p}$ is described by

$$Y_{t,b} = \sum_{p \in P_t:\ b=\beta(p)} X_{t,p} \qquad \forall t \in T;\ b \in B_t. \tag{19}$$

The variables $Y_{t,b}$ correspond to the higher-level decisions like capacity allocation, while the variables $X_{t,p}$ fill in the fine details of the solution.

Theorem 2. *Consider the model (1) – (15) extended by the binary variables $Y_{t,b}$ and by the constraints (19). Replace the constraint (12) in this model by $0 \leqslant X_{t,p} \leqslant 1$. Then this relaxed mixed integer program has an integral optimal solution whenever it has a feasible solution.*

Proof. Let $(X, Y, N, C, U, I^0, I, Z)$ be an optimal solution to the relaxed problem: the variables $Y_{t,b}$ are binary, while the variables $X_{t,p}$ may have fractional values. Consider the directed graph with node set $\{(t,p)\colon p \in P_t\}$ and arc set $\{((t,p),(\nu(t),p'))\colon (p,p') \in \Gamma_t\}$. Set the capacity of an arc to 0 if the corresponding $Z_{t,p,p'}$ is zero, and set the capacity to 1 otherwise. The values $Z_{t,p,p'}$ form a network flow in this graph. Thus there exists an integer valued network flow $\widehat{Z}_{t,p,p'}$ with the same amount of flow. Define the values $\widehat{X}_{t,p}$ according to the constraints (3) – (4).

We shall prove that the variables $Y_{t,b}$ and $I^0_{t,m}$ uniquely determine the variables $C_{t,m}$, $U_{t,m}$ and $I_{t,m}$ as well as the objective criteria *CKM*, *SKM* and *SHM*. Then it is easy to verify that $(\widehat{X}, Y, N, C, U, I^0, I, \widehat{Z})$ is an integral feasible (and optimal) solution to the relaxed model.

One easily derives that for each trip t and train unit type m, we have $N_{t,m} = \sum_{b \in B_t} b_m Y_{t,b}$. Consider now any trip t. The integrality of the variables $Y_{t,b}$ implies that a variable $X_{t,p}$ can have a positive value only if the composition p contains as many units of each type as the variables $N_{t,m}$ indicate. Then

(3) and (4) imply that all members of the set $\{(p,p') \in \Gamma_t \mid Z_{t,p,p'} > 0\}$ have the same differences $n_{m,p'} - n_{m,p}$ for every $m \in M$ (namely $N_{\nu(t),m} - N_{t,m}$). Therefore $C_{t,m} = \max\{N_{\nu(t),m} - N_{t,m}, 0\}$ and $U_{t,m} = \max\{N_{t,m} - N_{\nu(t),m}, 0\}$. It follows that the variables $I_{t,m}$ are also uniquely determined.

Clearly, the number of carriage kilometers *CKM* only depends on the variables $Y_{t,b}$. Moreover, the number of seat-shortage kilometers *SKM* is also determined by the variables $Y_{t,b}$. Indeed, all compositions that correspond to a certain combination $b \in B_t$ have the same capacity and thus the same amount of seat shortages (denoted by $s_{t,b}$) in trip t. Therefore

$$SKM = \sum_{t \in T} \sum_{b \in B_t} \ell_t \cdot s_{t,b} \cdot Y_{t,b}.$$

Finally, each trip has a contribution $\sum_{\beta(p) \neq \beta(p')} Z_{t,p,p'}$ to the number of shunting movements *SHM*. This contribution is 1 if $N_{t,m} \neq N_{\nu(t),m}$ for some train unit type m, and 0 otherwise. It follows that the number of shunting movements *SHM* is determined by the variables $Y_{t,m}$. $\square$

4.4 Combining and Splitting

In order to handle combining and splitting trains, we use an extension of the basic model. Here we only describe splitting trains. Combining two trains can be described in a similar way.

Let T^s be the set of trips after which the corresponding train is split into two trains. Then for each trip $t \in T^s$, there are two trips $\nu^1(t)$ and $\nu^2(t)$ that take place immediately after splitting trip t. Recall that the departure order of the trips $\nu_1(t)$ and $\nu_2(t)$ is determined by the timetable.

Let Γ_t^s be the set of 3-tuples (p, p_1, p_2) such that $p_1 \in P_{\nu^1(t)}$, $p_2 \in P_{\nu^2(t)}$, $p \in P_t$, and p is the concatenation of the strings p_1 and p_2. We introduce a variable $Z^s_{t,p,p_1,p_2} \in [0,1]$ for each $t \in T^s$ and $(p, p_1, p_2) \in \Gamma_t^s$. These variables are linked to the other variables similarly to constraints (3)–(4):

$$X_{t,p} = \sum_{p_1,p_2:\ (p,p_1,p_2)\in\Gamma_t^s} Z^s_{t,p,p_1,p_2} \qquad \forall p \in P_t \tag{20}$$

$$X_{\nu^1(t),p_1} = \sum_{p,p_2:\ (p,p_1,p_2)\in\Gamma_t^s} Z^s_{t,p,p_1,p_2} \qquad \forall p_1 \in P_{\nu^1(t)} \tag{21}$$

$$X_{\nu^2(t),p_2} = \sum_{p,p_1:\ (p,p_1,p_2)\in\Gamma_t^s} Z^s_{t,p,p_1,p_2} \qquad \forall p_2 \in P_{\nu^2(t)} \tag{22}$$

It is possible to modify the compositions further by allowing to couple or uncouple units when the train is split. This can easily be modelled by modifying the sets Γ_t^s. The variables $C_{t,m}$ and $U_{t,m}$ as well as the objective function should then also be adjusted by requiring constraints similar to (6) – (7).

Then Lemma 1 extends to the following lemma.

Lemma 3. *In any feasible solution to the model (1)–(15), (20)–(22), the variables $U_{t,m}$, $C_{t,m}$, $I_{s,t}$, $Z_{t,p,p'}$ and Z^s_{t,p,p_1,p_2} have integral values.*

In the case of combining and splitting, we can introduce the decision variables $Y_{t,b}$ described in Section 4.3. Unfortunately, Theorem 2 does not extend: Easy (but artificial) examples show that introducing the variables $Y_{t,b}$ and relaxing the variables $X_{t,p}$ may lead to a feasible model which does not have any feasible integral solution. However, when we tested the model with relaxed integrality of the variables $X_{t,p}$ on practical instances, we always found fully integral optimal solutions. Nevertheless, in our computations we declared both variables $Y_{t,b}$ and $X_{t,p}$ to be integral with different branching priorities (see Section 5.1).

4.5 Further constraints

Our model can be adjusted in order to incorporate further practical requirements.

When creating schedules for every day of the week separately, we should make sure that the solutions for separate days can be attached to each other.

One possibility is to create *cyclic* solutions: at every station, the inventory values at the beginning of the day are equal to the inventory values at the end of the day. Then a cyclic solution can be applied repeatedly. A reason to handle so is that the generic Tuesday, Wednesday and Thursday timetables are nearly identical and that the expected numbers of passengers on these days are also very similar.

Stations may have a bounded storage capacity. This can be expressed by imposing upper bounds on the inventory variables $I_{t,m}$. (Note that in this case, the integrality constraint of the initial inventory variables $I^0_{s,m}$ can be relaxed without decreasing the objective function value.) More realistically, we can give bounds on the number of carriages that are stored after the departure of trip t:

$$\sum_{m \in M} c_m \cdot I_{t,m} \leqslant B \tag{23}$$

where c_m denotes the number of carriages in train type m.

The *continuity requirement* states that for each timetable service (which is represented as a sequence of trips with the same train number), at least one train unit should follow the complete route of that train. As a consequence, a passenger can travel along the complete route of the timetable service without changing seats underway.

In the case of the Noord-Oost lines, timetable services have one (un)coupling possibility on their way. This makes that the continuity requirement is easy to formulate in terms of already declared variables. For example, consider the 1600 line in Figure 2. Train units departing from Amsterdam (Asd) and Schiphol (Shl) are combined in Amersfoort and they go together to Enschede (Es). The front part of the combined train arrived from Schiphol, while the rear part arrived from Amsterdam. Coupling or uncoupling is only possibly in Deventer (Dv): adding units to the front of the train, or removing units from the rear. Then the continuity constraint for one train between Asd–Es and Shl–Es amounts to

$$\sum_{m \in M} U_{\mathrm{Amf-Dv},m} \leqslant \sum_{m \in M} N_{\mathrm{Asd-Amf},m} - 1 \tag{24}$$

Indeed, we make sure that the number of train units uncoupled in Dv is less than the number of train units in the train Asd–Amf. Moreover, none of the train units that arrived from Shl can be uncoupled (as being at the front of the train), thus the direct connection Shl–Es is always provided.

4.6 Special structure of the instances

We can exploit the special structure of the instances as follows. If a train after trip t_1 is split in Utrecht (Ut), one part goes to The Hague (Gvc) (trip t_2), the other part goes to Rotterdam (Rtd) (trip t_3). These parts turn back to Utrecht (trips t_4 and t_5) immediately where they are combined again and leave Utrecht (trip t_6) together. When a train is split or combined in Utrecht, no train units can be coupled or uncoupled. Coupling and uncoupling is allowed in Rotterdam and The Hague. (See Figure 4.)

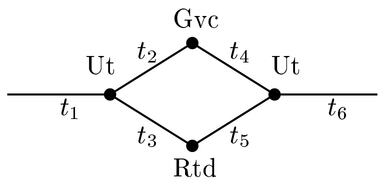

Figure 4: Splitting and combining in Utrecht

We define binary decision variables $Z^{\mathrm{Ut}}_{t_1,p_1,p_2,p_3,p_4,p_5,p_6}$ having value 1, if for each $i = 1, \dots, 6$, trip t_i has composition p_i. We delete from our model all the variables $X_{t,p}$, $Y_{t,b}$ and $Z_{t,p,p'}$ that are attached to the trips t_2, t_3, t_4 and t_5. We also delete the variables Z^s (and Z^c) we used to describe splitting (and combining) at trips t_1 (and t_6). Instead, we introduce constraints that are similar to (3) and (4):

$$X_{t_1,p} = \sum_{p_2,\ldots,p_6} Z^{\mathrm{Ut}}_{t_1,p,p_2,p_3,p_4,p_5,p_6} \qquad \forall p \in P_{t_1}$$

$$X_{t_6,p} = \sum_{p_1,\ldots,p_5} Z^{\mathrm{Ut}}_{t_1,p_1,p_2,p_3,p_4,p_5,p} \qquad \forall p \in P_{t_6}.$$

In order to keep track of the number of (un)coupled units in Rotterdam and The Hague, we use integer variables $U_{t_2,m}$, $U_{t_3,m}$, $C_{t_4,m}$, and $C_{t_5,m}$, and express them as weighted sums of the variables Z^{Ut} similarly to (6) and (7). The variables Z^{Ut} appear in the objective function with weights that indicate the carriage kilometers, seat-shortage kilometers and the number of shunting movements for the deleted trips $t_2, \ldots, t_5$.

Finally, we introduce the binary variables D^{Gvc} (and D^{Rtd}) that express whether or not a composition change happens between trips $t_2 - t_4$ (and between trips $t_3 - t_5$). These variables can also be written as sums of the variables Z^{Ut}.

This aggregation leads to a model with a higher number of decision variables, since the number of possible 6-tuples $(p_1, \ldots, p_6)$ is large, over 700 in our case. However, we eliminated a number of auxiliary variables $Z_{t,p,p'}$, Z^s, and Z^c attached to the trips $t_2, \ldots, t_5$, and also all decision variables describing the compositions on these trips.

Moreover, one can derive from the shunting constraints in the Noord-Oost case that if we relax the integrality constraint of the variables Z^{Ut}, then appropriate rounding on these variables immediately leads to a feasible integral solution. (Note that during our computations, we needed such rounding very rarely: once among about 100 instances.) Although the solution obtained in this way may not be optimal among all integral solutions, the difference in the objective function is very small. Therefore we declare the new variables Z^{Ut} binary with lowest branching priority: we can expect that with this priority setting, a branch and bound algorithm finds integral suboptimal solutions faster. We analyze the effect of this aggregation on the solution time in Section 5.1.

5 Computational results

We implemented the model described in the previous sections for the Noord-Oost lines for one day. This line system contains 167 timetable services, we divided them into 665 trips, according to the possibilities to change the compositions underway. The Noord-Oost lines are operated by about 50 train units of length 3 and 35 train units of length 4.

For our computations we used the modelling software ILOG Opl Studio 3.7 and the mixed integer programming solver ILOG CPLEX 9.0 on a PC with an Intel Pentium IV 3.0 GHz processor and with 512 MB internal memory. Our goal was to find feasible solutions of good quality in reasonable time. If necessary, we stopped our computations after a couple of hours of CPU time.

We compared our solutions to a manually created rolling stock plan. In Table 1 we give the values of the three objective criteria in this reference solution. We made experiments with several objective functions. The weight factors for carriage kilometers (CKM), seat-shortage kilometers (SKM) and the number of shunting operations (SHM) are also given in Table 1. According to the preferences of NS Reizigers we gave a relatively low weight factor to CKM in the objective function but required the constraint

$$CKM \leqslant 318,000.$$

Moreover, in some experiments we looked for solutions with CKM as small as possible while still providing significantly better service than the reference solution does.

The first two columns of Table 2 (under 'Without aggregation') contain the dimensions of the mixed integer program as well as the size of the reduced MIP, created by CPLEX in the preprocessing phase.

	CKM	*SKM*	*SHM*
value in practice	317,853	555,215	135
weight in Obj1	1	1	5,000
weight in Obj2	1	1	1,000
weight in Obj3	1	1	100
weight in Obj4	1	10	1000
weight in Obj5	1	10	100

Table 1: Objective criteria in the practical solution and our weight factors

	Without aggr.		With aggr.	
	MIP	Reduced MIP	MIP	Reduced MIP
# integer var.	17,776	–	32,197	–
# contin. var.	33,901	–	24,560	–
# variables	51,677	34,228	56,757	39,251
# constraints	32,308	17,896	28,149	15,348
# non-zeros	195,184	136,105	287,286	177,622

Table 2: Dimensions of the MIP's: number of variables (integer and continuous), constraints and non-zeros in the matrix

5.1 Speeding up the solution process

Finding feasible solutions of good quality turned out to be quite time-consuming. In order to speed up the solution process, we applied various techniques.

Fine-tuning the parameters of CPLEX had a large impact on the solution times. We used the barrier method to solve the root node of the branch and bound tree. Then we applied the dual simplex method for any other node. Using the built-in *heuristics* frequently, applying *probing*, *perturbing* the objective function, and using *branching priorities* (explained below) turned out to be particularly helpful.

Using priorities when branching

The variables in our model have a hierarchical structure. The variables $N_{t,m}$ describe the number of train units of a given type m used for trip t. The variables $Y_{t,b}$ represent decisions one level lower, while the variables $X_{t,p}$ specify the finer details. As we observed, the objective criteria are basically functions of the variables $Y_{t,b}$.

It turned out to be advantageous to branch first on the variables $N_{t,m}$, then on the variables $Y_{t,b}$, and at last on the variables $X_{t,p}$. (We declared the variables $N_{t,m}$ integral.) This can be explained as follows. The values $N_{t,m}$ describe a rough estimate of the solution. Nevertheless, early decisions on train lengths determine a large part of the objective function. Therefore, the branch and bound process can work out the exact train compositions subject to good lower bounds. We mentioned in Section 4.4 that, once the $Y_{t,b}$ variables are integral, the $X_{t,p}$ variables are likely integral, too.

Giving a higher branching priority to the decision variables corresponding to trains in the rush hours also led to an improvement of the solution time. Seat shortages mostly occur in such trains, thus very good lower bounds on the seat shortages can be computed in the early stage of the algorithm. Then the rolling stock assignment for the rush hours extends to the rest of the day, providing good suboptimal solutions relatively quickly.

Exploiting the structure of the instances

When applying the aggregation described in Section 4.6, the mixed integer program contains more variables but less constraints. The dimensions of the mixed integer program is given in the two last columns of Table 2 (under 'With aggregation'). We noted in Section 4.6 that the vast majority of the 16,565 variables Z^{Ut} are almost always integral, they have the lowest branching priority. This means that the mixed integer program has only 15,724 "essential" integer variables which is smaller than in the model without aggregation.

The linear programming relaxation with aggregation provides slightly better bounds than in the case of no aggregation. Although the difference turned out to be small, about 0.1 - 0.9% of the optimal objective value, even this improvement resulted in speeding up the solution process. The effect of using

CKM	SHM	Aggr	Obj1	Obj2	Obj3	Obj4	Obj5
318,000	–	Yes	1.69% 3000	0.22% 3200	opt 2000	opt 1900	opt 2700
318,000	–	No	2.56% 5400	0.76% 5000	opt 5000	0.12% 4300	opt 4400
318,000	120	Yes	2.30% 3500	0.50% 7100	1.24% 3400	2.22% 6500	2.32% 5100
318,000	120	No	2.06% 4800	0.67% 6400	— —	11.20% 6300	13.16% 5000
318,000	90	Yes	4.80% 6000	2.67% 7100	— —	11.43% 6400	— —
318,000	90	No	4.90% 5700	16.11% 5300	— —	— —	— —

Table 3: For each test instance: i) optimality gap after 7200 seconds, ii) time elapsed till the best solution was found (in seconds)

the aggregation is presented in Table 3. In three test instances, we imposed the constraint $CKM \leqslant 318,000$, and different bounds, if any, on SHM. We tested all instances with and without aggregation, minimizing the 5 objective functions in Table 1. In each row, we give the optimality gap proved in 2 hours of computation, and after how many seconds the best solution was found. A field without numbers means that no feasible solution was found within 2 hours. For these instances, we applied the heuristic approach in Section 5.2. We can see that the aggregation significantly improves the performance of the branch and bound process.

5.2 Heuristic approaches

Comparing the objective value of the linear relaxation to the best lower bound proved during the branch and bound process, the difference turned out to be small, at most 2%. Solutions with small optimality gap indicate that our model is a quite tight description of the convex hull of the integral solutions. This justifies the following heuristic method.

Search around the LP optimum

For some instances, depending on the objective function and on the bounds of the objective criteria, no feasible solution could be found in 2 hours of computation or the quality of the solutions was not satisfactory. In these cases, we applied the following method.

We first solved a relaxation of the problem, dropping the integrality requirements for some variables (e.g. taking the linear relaxation itself). Then we extracted information from the optimal (fractional) solution to the relaxation, and based on it, we added extra constraints to the original model. In the simplest form, we required $X_{t,p} = 0$ whenever this variable had value zero in the optimal solution to the relaxed model. As another example, we allowed a composition change (that is, extra shunting) after trip t only if the optimal solution to the relaxation had a variable $Z_{t,p,p'} > 0$ with $\beta(p) \neq \beta(p')$.

Reducing the solution space in this way allowed us to find feasible solutions within 2 hours for each instance. The optimality gap was, however, still over 10% in some instances. Nevertheless, even these solutions were in all objective criteria better than the reference solution from practice.

Local search

Once a feasible solution σ was found, we ran our model in some "neighborhood" of this solution. We defined the neighborhood by constraints like saying that for each trip, the composition may be at most one carriage shorter or longer than in σ. In case of smaller neighborhoods, the reduced solution space could be enumerated relatively quickly, but often they did not lead to any improvement. For larger neighborhoods we allowed running times up to 4 hours. After a few local search steps, we were able to find feasible solutions with optimality gaps of 2–5%.

We applied this local search to minimize only the carriage kilometers while imposing upper bounds on seat-shortages per trip, and on the number of shunting movements. We found solutions with as few as 297,000 – 303,000 carriage kilometers. Compared to 318,000 carriage kilometers in the reference solution, it amounts to a reduction in operational costs by nearly 6%.

5.3 Comparison to the reference solution

In this section we compare the values of the objective criteria in our solutions to those in the manually created reference solution. Figure 5 shows all rolling stock schedules we created where the number of carriage kilometers is not higher than in the reference solution. The x-axis represents the seat-shortage kilometers, the y-axis represents the numbers of shunting operations. We drew a small circle with appropriate x and y coordinates for solutions we found. The curved line indicates the linear programming lower bound. Solutions that lie far away from this line or that are dominated by other solutions were obtained by emphasizing additional optimization criteria: by penalizing first class seat-shortages and shunting during rush hours more heavily.

As mentioned earlier, all the solutions in Figure 5 obey the constraint $CKM \leqslant 318,000$. The solutions inside the dashed circle in Figure 5 have only 297,000 – 303,000 carriage kilometers.

6 Conclusions

In this paper, we described a model for the weekly rolling stock planning. Our model can express various technical and market requirements. We tested the model on fairly large instances of the Dutch passenger railway operator NS Reizigers. Our model turned out to be robust enough to cope with very different weights on the objective criteria. When solved by a commercial MIP solver, it provided rolling stock schedules with different objective characteristics in a couple of hours of computation time. The decision makers can choose a solution

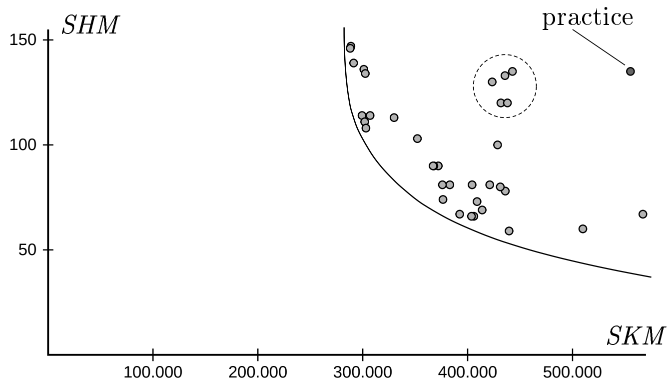

Figure 5: Shortage-kilometers and number of shunting operations in the solutions with at most 318,000 carriage-kilometers

that matches the practical requirements best. The planners at NS Reigers agreed that the solutions to our model are in any respect better than the manually created plans. Starting with the timetable year 2005, the weekly rolling stock schedule of NS Reizigers is partly computed by this model.

Acknowledgement

This research was supported by the Human Potential Programme of the European Union under contract no. HPRN-CT-1999-00104 (AMORE).

References

[1] E.J.W. Abbink, B.W.V. van den Berg, L.G. Kroon, and M. Salomon. "Allocation of Railway Rolling Stock for Passenger Trains". *Transportation Science*, 38(1) (2004), 33-42.

[2] A. Alfieri, R. Groot, L.G. Kroon, and A. Schrijver. "Efficient Circulation of Railway Rolling Stock". ERIM Research Report, ERS-2002-110-LIS. Erasmus University Rotterdam, The Netherlands (submitted to *Transportation Science*).

[3] C. Barnhart, N.L. Boland, L.W. Clarke, E.L. Johnson, G.L. Nemhauser, and R.G. Shenoi. "Flight String Models for Aircraft, Fleeting and Routing", *Transportation Science*, 32(3) (1998) 208-220.

[4] N. Ben-Khedher, J. Kintanar, C. Queille, and W. Stripling. "Schedule Optimization at SNCF: From Conception to Day of Departure", *Interfaces*, 28 (1998), 6-23.

[5] P. Brucker, J. Hurink, and T. Rolfes. "Routing of Railway Carriages: A Case Study", *Osnabrücker Schriften zur Mathematik*, Reihe P, Heft 205 (1998).

[6] J.F. Cordeau, F. Soumis, and J. Desrosiers. "A Benders Decomposition Approach for the Locomotive and Car Assignment Problem". *Transportation Science*, 34 (2000), 133-149.

[7] J.F. Cordeau, F. Soumis, and J. Desrosiers. "Simultaneous Assignment of Locomotives and Cars to Passenger Trains", *Operations Research*, 49 (2001), 531-548.

[8] N. Lingaya, J.F. Cordeau, G. Desaulniers, J. Desrosiers, and F. Soumis. "Operational Car Assignment at VIA Rail Canada", *Transportation Research B*, 36 (2002), 755-778.

[9] M. Peeters and L.G. Kroon. "Circulation of Railway Rolling Stock: a Branch-and-Price Approach", ERIM Research Report, ERS-2003-055-LIS. Erasmus University Rotterdam, The Netherlands (submitted to *Transportation Science*).

[10] A. Schrijver. "Minimum Circulation of Railway Stock", *CWI Quarterly*, 6 (1993) 205-217.