Rescheduling in passenger railways: the rolling stock rebalancing problem

**Gabriella Budai · Gábor Maróti · Rommert Dekker ·
Dennis Huisman · Leo Kroon**

Abstract This paper addresses the Rolling Stock Rebalancing Problem (RSRP) which arises within a passenger railway operator when the rolling stock has to be rescheduled due to changing circumstances. RSRP is relevant both in the short-term planning stage and in the real-time operations.

RSRP has as input a timetable and a rolling stock circulation where the allocation of the rolling stock among the stations at the start or at the end of a certain planning period does not match with the allocation before or after that planning period. The problem is then to modify the input rolling stock circulation in such a way that the number of remaining off-balances is minimal. If all off-balances have been solved, then the obtained rolling stock circulation can be implemented in practice.

For practical usage of solution approaches for RSRP, it is important to solve the problem quickly. Since we prove that RSRP is NP-hard, we focus on heuristic solution approaches: we describe two heuristics and compare them with each other on (variants of) real-life instances of NS, the main Dutch passenger railway operator. Finally, to get further insight in the quality of the proposed heuristics, we also compare their outcomes with optimal solutions obtained by solving an existing rolling stock circulation model.

Keywords Railway planning · Rolling stock rescheduling · Integer linear programming · Heuristic solution methods

G. Budai · R. Dekker · D. Huisman
Econometric Institute, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands

G. Budai
e-mail: budai@ese.eur.nl

R. Dekker
e-mail: rdekker@ese.eur.nl

D. Huisman
e-mail: huisman@ese.eur.nl

G. Maróti (✉) · L. Kroon
Rotterdam School of Management, Erasmus University
Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, The Netherlands
e-mail: gmaroti@rsm.nl

L. Kroon
e-mail: lkroon@rsm.nl

D. Huisman · L. Kroon
Netherlands Railways, Department of Logistics, P.O. Box 2025,
3500 HA, Utrecht, The Netherlands

1 Introduction

The rolling stock planning process of a passenger railway operator is commonly divided into several planning stages. Huisman et al. (2005) distinguish four planning stages, namely strategic, tactical, operational, and short-term planning. After these planning stages, the final plans are carried out and modified if necessary in the real-time operations. Strategic planning deals with long term decisions such as the acquisition of new rolling stock. At the tactical level, the different types of rolling stock are assigned to the different lines of the network. This is typically done once a year. The main goal of operational planning is to find a generic rolling stock circulation with low operational costs and high service quality. This circulation is basically carried out throughout the whole year. However, every calendar day there are minor modifications to the timetable due to some extra trains or due to maintenance of the railway infrastructure. These changes in the timetable require modifications of the rolling

stock circulation as well. This process is carried out in the short-term planning stage. The time horizon of this planning stage ranges from a couple of weeks to a couple of days before the operations.

In a rolling stock circulation, it is essential that at all stations there is always a *balance* between the number of departing rolling stock units and the number of earlier arriving units: a unit can only depart from a certain station if it arrived there earlier. This condition is called the rolling stock *balancing condition*.

1.1 Rolling Stock Rebalancing Problem

This paper deals with the *Rolling Stock Rebalancing Problem (RSRP)*, which is a problem faced in the short-term planning stage as well as during the real-time operations. The input of RSRP consists of the timetable for a given planning period, the available amount of rolling stock, the target rolling stock inventories at the start and at the end of the planning period, and an input rolling stock circulation.

The input rolling stock circulation is feasible, except that it may contain a number of *off-balances*. An off-balance is a deviation from the target inventory level of a certain rolling stock type at a certain station, either at the start or at the end of the planning period. The primary goal is to transform the input rolling stock circulation into a new rolling stock circulation with as few off-balances as possible. As secondary objective, other criteria related to costs, service, and robustness are considered as well. The input rolling stock circulation determines constraints on the allowed modifications of the rolling stock circulation, as will be explained.

1.2 Background information

In order to better motivate our study of RSRP, we first give some background information on the rolling stock planning process at NS (Netherlands Railways), the main passenger railway operator in the Netherlands.

At NS, most trains are operated by electrical self-propelled *train units*, and only a few are operated by a locomotive and carriages. Therefore, we consider only train units in the remainder of this paper. These train units are available in several types. Train units of compatible types can be attached to each other to form longer compositions. Train units of the same type are fully interchangeable.

Each trip in the timetable is assigned a certain *train composition*. A composition describes how many train units of each type are used for the train, and in which order they appear in the train. The practical feasibility of a rolling stock circulation depends to a large extent on the shunting possibilities of the stations. Shunting is a complex problem on its own (see, e.g., Freling et al. 2005; Lentink 2006). Therefore, several key aspects of the shunting processes are taken into

account already in the creation of the rolling stock circulation. Examples are the restrictions on composition changes at certain stations; uncoupling (or coupling) of train units can only take place at the appropriate side of a train.

A first motivation to study RSRP comes from the short-term planning process after the operational planning process. In the operational planning process, a cyclic rolling stock circulation is generated for 7 *generic* weekdays (i.e., Monday to Sunday), where consecutive days of the week fit after each other with respect to the rolling stock *balances* at the stations. Next, in the short-term planning process the timetable must be modified for one or more specific *calendar* days (in the following called the *planning period*), e.g., due to maintenance of the infrastructure certain timetable trips must be canceled. As a consequence, the rolling stock circulation must be modified as well during this planning period. Discussions with planners revealed that it is usually relatively easy for them to come up with a modified rolling stock circulation (the *input* rolling stock circulation) that fulfills all requirements, except that there are certain remaining off-balances. That is, at the start or at the end of the planning period the input rolling stock circulation does not fit with the rolling stock circulation for the generic weekdays *before* the planning period or *after* the planning period. Thus the problem is to modify the input rolling stock circulation such that the number of off-balances is reduced as much as possible. This calls for a solution of RSRP.

A second motivation to study RSRP comes from disruption management in the real-time operations, see Groth et al. (2007). During a disruption of the railway system (e.g., temporarily no railway traffic is possible between two stations due to malfunctioning infrastructure), the timetable is modified, and, as a consequence, the rolling stock circulation must be modified as well. As a result, the train units may not finish their daily duties at the locations where they were planned to. This is a problem if the number of train units ending up in the evening at a certain station differs from the number of train units starting their next day's duty there. To prevent expensive dead-heading trips during the night, the rolling stock circulation must be modified such that the rolling stock is balanced before the night. This real-time version of RSRP is to a large extent equivalent to RSRP in the short-term planning stage. The only difference is that the input rolling stock circulation is not constructed by planners but is caused by unexpected circumstances in the real-time operations. Moreover, in this case the off-balances only occur in the *final* inventories of the planning period.

1.3 Contribution

This is one of the first papers dealing with rolling stock planning problems in the short-term and real-time planning process. In particular, the aspect of solving the off-balances
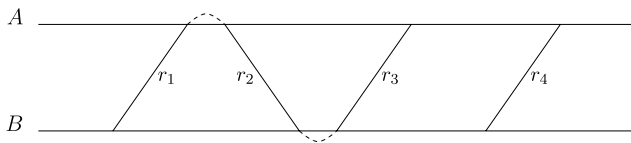
Fig. 1 A time–space diagram with time in the horizontal direction. The *diagonal lines* represent trips r_1 to r_4 between stations A and B

Fig. 2 The rolling stock duties of a black (*solid bold line*) and a gray (*solid dashed line*) train unit on the trips r_1 to r_4 shown in Fig. 1

in a given input rolling stock circulation has not yet been studied before. The need for fast solution approaches in the short-term and real-time planning stages motivated us to focus on two heuristic methods for solving RSRP. Our focus on heuristic approaches was also inspired by the fact that RSRP is NP-hard, as we prove in the appendix of this paper. The heuristics are compared with each other on instances of RSRP based on (variants of) real-life instances of NS. Moreover, to get further insight in the quality of these heuristics, we also compare the results of the heuristics with the results of an existing optimal approach for operational rolling stock planning (see Fioole et al. 2006). The application of the heuristics appears to result in acceptable solutions within short computation times.

The remainder of this paper is organized as follows. In Sect. 2, we give a precise description of RSRP. Section 3 contains a literature overview. In Sect. 4, we describe an algorithm for solving a single off-balance. The heuristics for solving the general RSRP are described in Sect. 5. Computational results are discussed in Sect. 6. In Sect. 7, we draw some conclusions and we outline some directions for further research. Finally, complexity results on variants of RSRP are presented in Appendix A.

2 Problem description

In this section, we define RSRP in more detail. The first part of the input of RSRP consists of the timetable for a certain planning period. The timetable defines a set of *trips*, each of which is characterized by a train number, departure and arrival times, departure and arrival locations as well as an estimated number of passengers. Moreover, most trips (apart from the arrivals in the late evening and some other exceptions) have a *successor* trip, which departs from the same station shortly after the trip (the *predecessor*) has arrived. Figure 1 represents trips r_1 to r_4 between the stations A and B in a time-space diagram. Here trip r_1 has trip r_2 as successor, and trip r_2 has trip r_3 as successor, as is indicated by the dotted lines. Conversely, trip r_2 has trip r_1 as predecessor, and trip r_3 has trip r_2 as predecessor.

2.1 Rolling stock circulation

The second part of the input of RSRP concerns the rolling stock. In particular, we have a list of available rolling stock
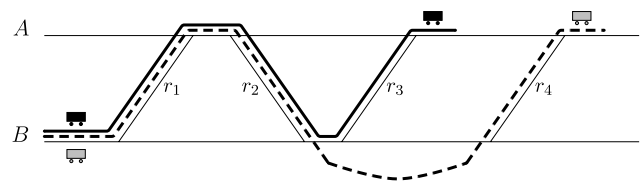
types, the number of available train units per rolling stock type, and an *input rolling stock circulation.*

A rolling stock circulation is represented in terms of a number of rolling stock *duties*. Here a duty is the workload of a single train unit on a single day. It is a chain of tasks where a task is characterized by a trip and by the position of the train unit in the train composition of this trip, e.g., front, middle or rear. Based on the rolling stock duties, the composition of each trip can be determined. Note that a duty may be empty, for example, in the case of a stand-by train unit.

In principle, a trip and its successor trip are operated by the same train units. That is, the rolling stock composition of a trip forms the base of the rolling stock composition of the successor trip. However, sometimes one or more train units are coupled or uncoupled between two successor trips. This is called a *composition change*, and requires a *shunting movement*.

Since the time between two successor trips is short, there is only a limited amount of time for a composition change. Therefore, a general rule is that coupling and uncoupling cannot be performed at the same time. Furthermore, in each station composition changes usually may take place only at a pre-defined side of a train, either at the front side or at the rear side, the *shunting side* of the train at that station. Thus the train compositions of a trip and its successor trip are often the same. The difference may be that one or more train units are either coupled onto or uncoupled from the shunting side of the train at that station.

For example, the solid and dashed bold lines in Fig. 2 show the duties of a black and a gray train unit on the trips r_1 to r_4 that were shown in Fig. 1, e.g., the duty of the gray train unit is $(r_1,$ rear$)$–$(r_2,$ front$)$–$(r_4,$ solo$)$. Trips r_1 and r_2 have compositions consisting of both train units. The gray train unit is uncoupled from the train after trip r_2 and stored at the shunt yard of station B. This train unit is used later for trip r_4. Note that trip r_4 is not a successor of trip r_2, since the time between these trips is too long. Clearly, the shunting side of station B is the front side of an arriving train. Therefore, it is possible to uncouple the gray train unit there. If station B would allow shunting only at the rear side of an arriving train, then it would have been impossible to uncouple the gray train unit after trip r_2.

Train units that have been uncoupled from a train are stored at the shunt yard of the station until they are coupled

onto another train. Hence they are not immediately available again, only a certain *re-allocation time* ϱ later they may be coupled again. This is to reserve time for the necessary shunting operations at the station. A typical value for ϱ is 30 min.

Each rolling stock circulation satisfies the following requirements: The length of the composition on a trip r must be under a certain limit $\mu_r^{\max}$ (determined by the relevant platform lengths). Moreover, the trip has to be assigned at least a given number $\mu_r^{\min}$ of carriages in order to cover (a large part of) the passenger demand.

Next, we need the concept of the *inventory*. The inventory of a station at a given time t consists of the train units that are staying idle at that moment at that station. These train units can be coupled to a departing train. Note that a train unit that is idle at a station between two successor trips does not belong to the inventory of that station at that moment. If the rolling stock balancing condition is always satisfied *within* the planning period, then the total inventory of train units at a station and at time t equals the number of duties that begin in that station at the start of the planning period, plus the number of train units that were uncoupled from the composition of a trip (either with or without a successor trip) arriving in that station before t, minus the number of train units that were coupled to the composition of a trip (either with or without a predecessor trip) leaving from that station before t. The inventory of a station can also be recorded per train unit type.

In the inventory, the order of the train units is assumed to be arbitrary. This is in contrast with the trains themselves, where the order of the train units is essential.

2.2 Off-balances

If the rolling stock circulation was modified during a planning period of one or more consecutive days (and not during the previous days nor during the days thereafter), then the result may be a rolling stock circulation that satisfies the rolling stock balancing condition *before* this planning period, *during* this planning period, and *after* this planning period. However, the balancing condition may be violated at the transition moments at the start and at the end of the planning period.

In order to make this more precise, we introduce the following definitions. For a planning period of one or more consecutive days, the *target initial inventory* of a station equals the number of duties per type that ended there just before the start of the planning period. Thus it represents, per type, the number of train units that are available at that station at the start of the planning period. Similarly, the *target final inventory* is the number of duties per type that start at that station just after the end of the planning period. In other words, it represents, per type, the number of train units that are needed there at the end of the planning period.

For example, if the planning period is a Saturday and there is no traffic at night, then the target initial inventory of a certain type at a station equals the number of duties of that type that end at that station on Friday evening. The target final inventory equals the number of duties of that type that start there on Sunday morning.

A station has a *surplus* (or a *deficit*) of a certain train unit type in the *initial* inventory if the number of duties of that type that begin at that station just after the start of the planning period is lower (higher) than the number of duties of that type that end there just before the start of the planning period. Similarly, a station has a *surplus* (or a *deficit*) of a certain type in the *final* inventory if the number of duties of that type that end at that station just before the end of the planning period is higher (lower) than the number of duties of that type that start there just after the end of the planning period.

The total number of remaining *off-balances* in a rolling stock circulation is obtained by adding the surpluses over all stations and over all types. In the end, this number expresses how many train units must be involved in dead-heading trips, i.e., driving empty train units during the night in order to correct the off-balances in all stations. Since dead-heading trips are expensive, this number of remaining off-balances must be reduced as much as possible. Note that the total number of surpluses equals the total number of deficits.

2.3 Rolling Stock Rebalancing Problem

The *Rolling Stock Rebalancing Problem (RSRP)* can now be defined formally as the problem of modifying the input rolling stock circulation during the planning period into a new rolling stock circulation such that (i) the new rolling stock circulation is feasible, (ii) it contains a minimum number of remaining off-balances at the start and at the end of the planning period, and (iii) also certain secondary objectives are taken into account.

In the experiments, we choose for an objective function which is a linear combination of the number of off-balances (with a very high weight), carriage-kilometers, shortage-kilometers and the number of composition changes. Carriage-kilometers express the operational costs of the railway operator. Seat shortage kilometers are computed by taking the expected number of passengers without a seat on a trip, multiplying it by the length of the trip, and adding them up over all trips. The obtained value is a representation of the provided service quality. The number of composition changes counts how many times train units are coupled to or uncoupled from the trains during a stop between two successor trips. A circulation with a smaller number of composition changes is expected to be more robust in the operations, since composition changes are a source of delays.

The heuristic methods that we describe in Sects. 4 and 5 mimic the way planners proceed in practice to modify an

input rolling stock circulation into an improved circulation. They apply certain transformations to the rolling stock circulation in a stepwise approach (these transformations are called *Balancing Possibilities*), each of which reducing the number of off-balances by one. Examples of Balancing Possibilities are given in Sect. 4.1. Thus the input rolling stock restricts the modifications to the rolling stock circulation. As a consequence, the resulting rolling stock circulation will differ only to a limited extent from the input rolling stock circulation.

Furthermore, the operational rolling stock scheduling models described by Fioole et al. (2006) and by Peeters and Kroon (2008) might be adapted in a straightforward way for solving RSRP. In that case, these models can modify the input rolling stock circulation either in an arbitrary way, or they can mimic the application of the Balancing Possibilities to a certain extent. However, their rather long and unpredictable computation times on large and complex instances motivated our research on heuristic solution approaches for solving RSRP.

3 Literature overview

A large number of publications have addressed operational rolling stock planning, see Caprara et al. (2007) for a recent overview. We only mention here Peeters and Kroon (2008) and Fioole et al. (2006). Their models have basically the same specifications as those in this paper, in particular the specifications related to the shunting possibilities in the stations. In the case that trains are not combined or split, Peeters and Kroon (2008) solve the rolling stock circulation problem by applying Dantzig–Wolfe decomposition and Branch-and-Price as solution technique. Fioole et al. (2006) extend the model for splitting and combining of trains. They use the commercial MIP software CPLEX to solve the model.

Compared to operational planning, literature on short-term railway rolling stock planning is scarce. Brucker et al. (2003) consider the problem of routing railway carriages through a railway network. The carriages should be used in timetable services or dead-heading trips such that each timetable service can be operated with at least a given number of carriages, thereby satisfying the passenger demand. The order of the carriages in the trains is not considered. The objective is to minimize a non-linear cost function. The solution approach is based on local search techniques such as simulated annealing.

Ben-Khedher et al. (1998) study the short-term rescheduling problem of the French TGV trains from a revenue management point of view. The rolling stock circulation must be adjusted to the latest demand from the seat reservation system in order to maximize the expected profit.

Lingaya et al. (2002) deal with the effect of an altered timetable and passenger demand on the rolling stock circulation, focusing on the case of locomotive hauled carriages. They explicitly take the order of the carriages in the trains into account and assume that for each train a successor train has already been specified. Several real-life aspects, such as maintenance, are considered as well.

Substantial research has been carried out on aircraft and bus rescheduling. Kohl et al. (2007) and Clausen et al. (2005) give overviews of airline disruption management, including a detailed list of aircraft rescheduling publications and applications. The common solution approaches are based on multi-commodity network flows, thereby applying various exact and heuristic methods. Many of the models incorporate maintenance of the aircraft as well.

Recently, Li et al. (2007) introduced the single depot vehicle rescheduling problem. It is motivated by the problem of updating bus schedules in the case when a single vehicle breaks down. The rescheduling problem is formulated as a minimization problem over a number of vehicle scheduling problems.

A main distinguishing feature of railway (re-)scheduling is that trains may consist of multiple train units and that the order of the train units is to be regarded when they are coupled to each other. In contrast, a single bus or aircraft is to be used for a flight or a bus trip. Thus a model for an airline or a bus application usually cannot be used directly for solving a railway rolling stock problem.

We conclude that, although a large variety of related rolling stock scheduling problems has been described and partly successfully solved, railway rolling stock rescheduling—in particular in the real-time operations—still lacks the appropriate models and solution methods.

4 An off-balance of a single train unit

In this section, we consider the special case of RSRP with an off-balance of a single train unit; we call this problem 1-RSRP. In particular, we assume that the rolling stock circulation realizes the target inventories except at two locations. There is a surplus of one train unit of type t in the *final* inventory of station A, and there is a deficit of one train unit of the same type t in the *final* inventory of station B. Note that deviations from the target *initial* inventories of stations A and B can be handled in a very similar way.

A solution of 1-RSRP is called a *Balancing Possibility* (below abbreviated as BP). This terminology is motivated by the two heuristic algorithms in Sect. 5 where BPs serve as building blocks for the solution of the general RSRP.

Finding a BP, i.e., deciding whether or not an instance of 1-RSRP has a feasible solution, is an NP-complete problem.

Fig. 3 A BP for the case that station A has a final deficit, and station B has a final surplus: the train length on trip r is decreased

Fig. 4 Another BP for the case that station A has a final deficit, and station B has a final surplus: the train length on trip r is increased

We prove this in Appendix A by reducing the maximum independent set problem in an undirected graph to 1-RSRP. Therefore, the optimization variant of 1-RSRP is NP-hard.

In this section, we first give some examples of BPs. Then we describe a heuristic algorithm for 1-RSRP based on single commodity network flows. Since 1-RSRP is NP-hard, there does not exist an exact polynomial-time algorithm for 1-RSRP (unless P = NP). The performance of the heuristic algorithm (as part of the heuristic approaches for the general RSRP) is demonstrated in Sect. 6.

4.1 Examples of balancing possibilities

Examples of BPs can be obtained by examining the rolling stock duties. The left-hand side of Fig. 3 shows a time–space diagram of a small railway system with two trains between stations A, B and C. The bold lines show the duties of the train units according to the rolling stock circulation. The target final inventories are indicated by the gray train units. One train unit must be available at stations A and B at the end of the planning period. Since the represented rolling stock circulation has two ending train units in station B, this circulation has a final surplus of one train unit in station B and a final deficit of one train unit in station A.

The right-hand side of Fig. 3 is a possible solution to the balancing problem. Trip r from A to B is operated with one train unit only; the second train unit is uncoupled from the train at station A before leaving towards station B. Of course, this solution is a BP only if trip r can be operated with a single train unit, and if time and shunting capacity at station A are sufficient for carrying out the composition change there (i.e., for uncoupling one of the train units).

Another example is shown in Fig. 4 where the final off-balance of stations A and B is resolved by increasing the train length on trip r by one train unit.

The ideas of increasing and decreasing the length of some trains can also be combined. Such a more elaborate example is given in Fig. 5. It involves stations A, B, C and D as well as five trips denoted by $r_1, \ldots, r_5$. Then, in order to resolve the final deficit of station A and the final surplus of
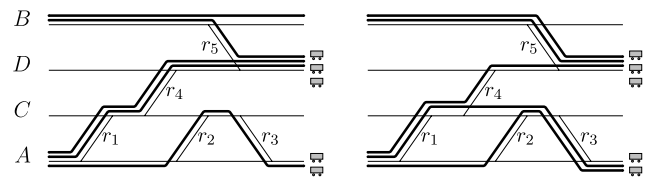
Fig. 5 A more elaborated example of a BP for the case that station A has a final deficit, and station B has a final surplus

station B, one has to modify the compositions on three trips: trips r_3 and r_5 are to be served by two train units, while the train length on trip r_4 has to be reduced to a single train unit. Again, this modified circulation has to agree with the shunting possibilities. Moreover, all modified train lengths must respect the upper and lower bounds for those trips.

The foregoing examples illustrate the fact that a solution without off-balances can often be found by applying a number of basic BPs to the input rolling stock circulation. Based on the input rolling stock circulation, the BPs can be determined. A solution that is constructed in this way is preferable over a solution that is constructed completely from scratch, since it leaves the basic structure of the rolling stock circulation unchanged to a large extent.

4.2 A heuristic algorithm for 1-RSRP

Discussions with planners revealed the desire to change the rolling stock circulation not too deeply when solving an off-balance of a single train unit of a certain type t. This motivates the basic restriction in the heuristic approach to 1-RSRP: *The rolling stock circulation is to be modified in such a way that the circulation of every train unit type differing from t must not be changed.* So, for example, if a trip has composition 'tab' in the rolling stock circulation with train unit types a, b and t, then the algorithm must assign to this trip a train unit of type a and a train unit of type b in this order, and any number of train units of type t before, between and after them. In particular, the modified composition can be 'ab', 'atb', '$attb$', '$tatbt$', etc. However, it cannot be 'taa' or 'tba' since those would change the circulation of types 'a' and 'b'.

The main idea of the heuristic algorithm for 1-RSRP is to represent the problem as a single commodity network flow problem with two additional side constraints. Here the underlying graph structure ensures that in case of composition changes, train units are coupled to or uncoupled from the proper side of the trains. The additional side constraints express that:

(i) the train lengths lie between the given lower and upper bounds, and

(ii) each shunting operation is either coupling of train units or uncoupling of train units, but not both

The algorithm we propose relaxes the additional side constraints (i) and (ii), solves the network flow problem, and checks thereafter whether the obtained flow satisfies the side constraints (i) and (ii). If these side constraints are violated, then the algorithm terminates. This approach is justified by our computational results, where it turns out that none of the several thousand test runs leads to a network flow violating the side constraints.

4.2.1 The graph representation

We represent an instance of 1-RSRP as a network flow problem. To do so, we build up a graph $G = (V, E)$ which is a variant of a usual time–space network occurring in public transport problems. Let us start with an empty graph.

A time moment j is *relevant* at station C if a trip departs at time moment j from C or if a trip r arrives at time moment $j - \varrho$ at C where ϱ is the re-allocation time. In addition, the begin and the end of the planning period are also relevant. Create a *station node* for each pair (C, j) where C is a station and j is a relevant time moment at C. For each pair j, j' of consecutive relevant time moments at station C, draw a *station arc* from the node associated with (C, j) to the node associated with (C, j'). The flow values on the station arcs shall express the current inventories of type t at the stations. Station nodes at the start of the planning period are the *source nodes*; station nodes at the end are the *sink nodes*.

Consider a trip r and suppose the rolling stock circulation assigns composition

$$\underbrace{t \cdots t}_{k_1^{(r)}} t_1 \underbrace{t \cdots t}_{k_2^{(r)}} \cdots t_{\ell_r - 1} \underbrace{t \cdots t}_{k_{\ell_r}^{(r)}} \tag{1}$$

to trip r where $t_1, \ldots, t_{\ell_r - 1}$ denote train unit types differing from t. We assume that the left-hand side of this string corresponds to the front side of the train. That is, train units of type t are assigned to trip r in ℓ_r possibly empty groups, separated by $\ell_r - 1$ train units of other types. The heuristic algorithm shall only modify the integer values $k_1^{(r)}, \ldots, k_{\ell_r}^{(r)}$, and leave the train units $t_1, \ldots, t_{\ell_r - 1}$ unchanged.

For example, if a and b represent train units of types different from t and a train consists of 4 train units in the composition $atbt$, then $\ell_r = 3$ and $k_1^r = 0$ and $k_2^r = k_3^r = 1$.

For each trip r, we create ℓ_r new nodes $u_1^{(r)}, \ldots, u_{\ell_r}^{(r)}$ corresponding to the groups of type t at the departure of trip r, and we create ℓ_r new nodes $v_1^{(r)}, \ldots, v_{\ell_r}^{(r)}$ corresponding to the arrival of trip r. Moreover, we draw the arcs $u_i^{(r)} v_i^{(r)}$ for each $i = 1, \ldots, \ell_r$. We call these arcs *trip arcs*.

Let r' be the successor of trip r and suppose that in the rolling stock circulation train units are uncoupled from the arriving trip r. We also assume that the uncoupling takes place at the rear side of the train. Then our graph representation does not contain the possibility of coupling any train unit to trip r' and we have $\ell_r \geq \ell_{r'}$. Physically, the train is split into two parts at a point that lies in the $\ell_{r'}$th group of the arriving composition. Then the first (i.e., left-most in (1)) $\ell_{r'} - 1$ groups go over unchanged to become the first $\ell_{r'} - 1$ groups of trip r'. The last (i.e., right-most in (1)) $\ell_r - \ell_{r'}$ groups (if any) are uncoupled. Train units in the $\ell_{r'}$th group of trip r can go over to the $\ell_{r'}$th group of trip r' or they can be uncoupled. These possibilities are expressed by the arcs shown in Fig. 6(a) for the case $\ell_r = 3$ and $\ell_{r'} = 2$. Notice that the re-allocation time ϱ is respected. The construction can easily be adjusted if uncoupling takes place at the front side of the arriving trip.

The cases when, according to the rolling stock circulation, train units are added to the departing trip r' and when there is no composition change between trips r and r' are modeled similarly. Examples are shown in Figs. 6(b) and 6(c).

We call an arc from a station node to a node $u_i^{(r)}$ a *coupling arc* and we call an arc from a node $v_i^{(r)}$ to a station node an *uncoupling arc* as they are intended to describe coupling and uncoupling of train units.

Furthermore, a trip without a predecessor has to be supplied completely with rolling stock from the inventory at the involved station. Arcs that are similar to the coupling arcs described above are introduced for dealing with this situation. Similarly, for a trip without a successor, uncoupling arcs are introduced to allow the complete composition of

Fig. 6 The graph representation for the cases that uncoupling, coupling or no composition change takes place between trips r and r'

(a) Uncoupling   (b) Coupling   (c) No change

(a) The time-space diagram   (b) The graph
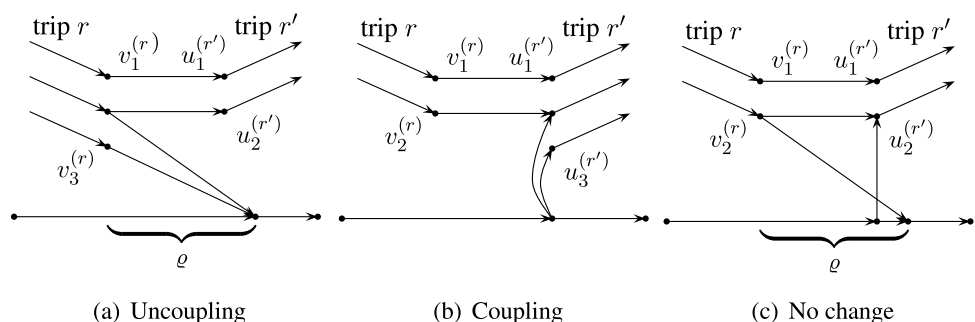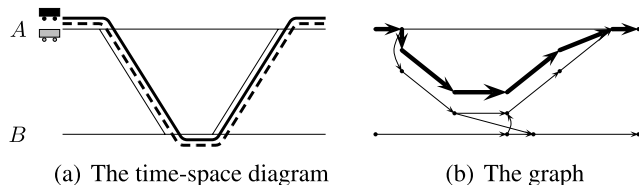
Fig. 7 The graph representation of a small railway network and the flow of the black train unit type: bold arcs have flow value one, other arcs have zero flow value

the trip to be moved to the inventory at the involved station. This completes the definition of the graph G.

4.2.2 The initial flow in the graph

In this section, we describe how the movements of the train units of type t in the input rolling stock circulation correspond to a network flow x in the graph G. Next, the off-balances are solved by modifying this network flow.

In the graph, each trip arc corresponding to group i of trip r gets the corresponding flow value $k_i^{(r)}$. The number of coupled or uncoupled train units of type t is assigned to the coupling and uncoupling arcs. The flow value on a station arc is the inventory of type t at that station during the time interval indicated by the arc. Then the source nodes have a (possibly zero) out-flow, the sink nodes have a (possibly zero) in-flow, and all other nodes satisfy the flow conservation law.

Figure 7 indicates the graph representation for the black train unit type in a small railway network.

The flow value on each arc is non-negative and, depending on the problem specification, it obeys certain upper bounds denoted by the capacity $g(a)$ of each arc a. For example, bounds on the station arcs may express the storage capacity of the stations. In addition, the following two side constraints (2)–(3) must be satisfied. These constraints represent the earlier mentioned constraints (*i*) and (*ii*) in mathematical terms. First, the train length on each trip r obeys the lower and upper bounds $\mu_r^{\min}$ and $\mu_r^{\max}$:

$$\mu_r^{\min} \leq \sum_{i=1}^{\ell_r} \lambda_t \times x\left(u_i^{(r)} v_i^{(r)}\right) + L_r$$
$$\leq \mu_r^{\max} \quad \text{for each trip } r. \tag{2}$$

Recall that $\mu_r^{\min}$ and $\mu_r^{\max}$ represent the minimal and maximal number of carriages of the train on trip r. Furthermore, λ_t denotes the number of carriages of each train unit of type t, and L_r denotes the total number of carriages in those train units on trip r whose types differ from t.

Second, coupling and uncoupling may not take place together between a trip r and its successor r':

$$\sum_{i=1}^{\ell_r} \sum_{\substack{a \in \delta^{\text{out}}(v_i^{(r)}): \\ a \text{ is an uncoupling arc}}} x(a) = 0 \quad \text{or}$$

$$\sum_{i=1}^{\ell_r} \sum_{\substack{a \in \delta^{\text{in}}(u_i^{(r')}): \\ a \text{ is a coupling arc}}} x(a) = 0. \tag{3}$$

Here $\delta^{\text{in}}(v)$ (resp., $\delta^{\text{out}}(v)$) denotes the set of arcs entering (resp., leaving) node v.

Conversely, if a network flow in G satisfies side constraints (2)–(3), then it corresponds to a feasible rolling stock circulation.

4.2.3 Solving the off-balances

Recall our assumption that station A has a final surplus of one train unit and station B has a final deficit of one train unit. That is, the target initial inventories are equal to the out-flow of the source nodes, and the target final inventories are equal to the in-flow of the sink nodes except for stations A and B. In order to resolve this off-balance, we have to find a network flow x' such that

$$\sum_{e \in \delta^{\text{in}}(A)} x'(e) = \sum_{e \in \delta^{\text{in}}(A)} x(e) + 1 \quad \text{and}$$

$$\sum_{e \in \delta^{\text{in}}(B)} x'(e) = \sum_{e \in \delta^{\text{in}}(B)} x(e) - 1,$$

where we identified stations A and B with their sink nodes. At each other node, the in- and out-flow of x and x' must be equal. Furthermore, x' must satisfy the side constraints (2)–(3).

It is well-known in network flow theory that, if such a flow x' exists (without requiring (2)–(3)), then it can be obtained by modifying the flow x along an augmenting path P which is a directed A–B path in the so-called *auxiliary graph* G_x. The auxiliary graph G_x on the node set V is constructed as follows. Let G_x have a *forward* arc uv if $uv \in E$ with $x(uv) < g(uv)$. Let G_x have a *backward* arc vu if $uv \in E$ with $x(uv) > 0$. If there exists a directed A–B path P in the auxiliary graph G_x, then the modified flow x' is defined as follows:

$$x'(uv) = \begin{cases} x(uv) + 1 \\ \quad \text{if the forward arc } uv \text{ is used by path } P, \\ x(uv) - 1 \\ \quad \text{if the backward arc } vu \text{ is used by path } P, \\ x(uv) \quad \text{otherwise.} \end{cases} \tag{4}$$

An arbitrary augmenting path P may lead to a violation of the side constraints (2)–(3). Actually, the feasibility version of 1-RSRP is NP-complete (see Appendix A); therefore, an augmenting path satisfying (2)–(3) cannot be found in polynomial time (unless P = NP). In our heuristic approach, we simply relax the side constraints (2)–(3) by looking for an augmenting path and verifying afterwards whether the updated network flow x' satisfies the side constraints (2)–(3).

If there is no augmenting path at all, then the instance of 1-RSRP is certainly infeasible. If there is an augmenting path and x' fulfills constraints (2)–(3) then the off-balance of stations A and B has been resolved. However, if there exists an augmenting path, but the side constraints are violated, then the algorithm reports that the off-balance could not be resolved. In the latter case, the answer might be wrong. Another augmenting paths might have resulted in satisfied side constraints (2)–(3). However, in our extensive computational tests, we did not find any augmenting path that led to violated side constraints (2)–(3).

The algorithm as described above attempts to find any augmenting path. This reflects that the main objective is to resolve as many off-balances as possible. The additional secondary objective criteria (carriage-kilometers, seat shortage kilometers and the number of composition changes) are taken into account by assigning cost values to the arcs of G. Then, according to classical network flow theory, arc costs in G_x are defined by $c_x(uv) = c(uv)$ if uv is a forward arc and by $c_x(vu) = -c(uv)$ if vu is a backward arc. Now we have to look for a minimum cost augmenting path in G_x.

5 Arbitrary off-balances

This section describes two heuristic algorithms for solving RSRP in the case of arbitrary off-balances. The main idea is to reduce the solution process for an instance of the general RSRP to iteratively solving instances of 1-RSRP. That is, a solution for an instance of the general problem is built up from BPs that each resolve a single off-balance. The first approach iterates greedily: In each iteration, it takes the outcome of the previous iteration as input, and resolves one off-balance in the next iteration. In the second approach, we derive a priori a large number of BPs from the input rolling stock circulation. Then we use an Integer Programming model to combine these BPs into a solution for the general problem.

5.1 An iterative heuristic

This section describes an iterative heuristic for solving RSRP. In each iteration, either a *type switching step* or a *rerouting step* is carried out on the current rolling stock circulation. Both steps try to decrease the number of off-balances in a greedy way. The overall algorithm stops if no step can bring any further improvement.

A *type switching step* considers pairs of train units of different types. The algorithm checks whether exchanging the duties of these train units over the whole planning period results in a feasible rolling stock circulation, and also whether the exchange decreases the number of off-balances. The two train units whose exchange leads to an improvement are in fact switched, yielding an updated rolling stock circulation. Thereafter, another iteration is launched. Type switching steps are straightforward; therefore, they are not further described.

A *rerouting step* looks for a BP that reduces the number of off-balances in the rolling stock circulation by one. It does so by applying the flow-based algorithm for solving 1-RSRP described in Sect. 4 to the current rolling stock circulation. Each rerouting step takes into account the objective criteria that were described in Sect. 4 (i.e., carriage kilometers, seat shortage kilometers, and number of composition changes) in a weighted way. When the flow-based algorithm has found the best possible BP that reduces the total number of off-balances by one, the rolling stock circulation is updated accordingly. Then another iteration is carried out.

5.2 A two-phase heuristic

Here we describe a two-phase heuristic approach for solving RSRP. In Phase 1, we identify a number of elementary BPs, each of which reduces the number of off-balances in the rolling stock circulation by one. Phase 2 selects a subset of the elementary BPs computed in Phase 1 such that carrying out the selected BPs leads to a new feasible rolling stock circulation with less off-balances.

In Phase 1, the elementary BPs are generated by applying the flow-based algorithm for solving 1-RSRP described in Sect. 4 under varying parameter settings. For example, for each combination of a surplus and a deficit at two stations—either in the initial inventory or in the final inventory—the algorithm attempts to generate one or more BPs. The objective criteria that were described in Sect. 4 (i.e., carriage kilometers, seat shortage kilometers, and number of composition changes) are used in a weighted way to express the desirability of each BP in terms of a cost value. Given the set of all BPs that were defined in Phase 1, we choose in Phase 2 those BPs that minimize the weighted sum of the number of remaining off-balances and the total costs of the BPs.

The feasibility of a BP for a certain trip depends on the details of the composition of the trip, as well as on the compositions of its predecessor and successor trip. Therefore, it is not possible to determine a priori whether certain combinations of BPs that modify the composition of the same trip result in a feasible rolling stock circulation. To stay on the

safe side, we allow BPs to be selected simultaneously only if each trip gets modified at most once by the selected BPs. Thereby we guarantee that the selected BPs can be implemented in practice indeed.

The BPs with overall minimum cost are selected with the following integer linear programming model. Let E be the set of all BPs that were generated in Phase 1, S the set of all stations, T the set of train unit types, and *Trip* the set of all trips. Let $b_{s,t}^{\text{beg}} \in \{0, \pm 1, \pm 2, \dots\}$ and $b_{s,t}^{\text{end}} \in \{0, \pm 1, \pm 2, \dots\}$ denote the surplus or deficit in the initial and final inventory of type $t \in T$ on station $s \in S$.

Let c_e be the cost of $e \in E$. Furthermore, the parameters $d_{s,t,e}^{\text{beg}}$ (or $d_{s,t,e}^{\text{end}}$) describe the change in the initial (or final) inventory of type $t \in T$ at station $s \in S$ when $e \in E$ is applied. Note that, by definition, $d_{s,t,e}^{\text{beg}}$ and $d_{s,t,e}^{\text{end}} \in \{0, \pm 1\}$. The set Γ_e contains the trips that are modified by $e \in E$.

For each $e \in E$, let x_e be a binary decision variable expressing whether or not e is selected. Then the BP selection problem can be formulated as follows.

$$\text{minimize} \quad \sum_{e \in E} c_e x_e, \tag{5}$$

$$\text{s.t.} \quad \sum_{e \in E} d_{s,t,e}^{\text{beg}} x_e = b_{s,t}^{\text{beg}}, \quad \forall s \in S, \ \forall t \in T, \tag{6}$$

$$\sum_{e \in E} d_{s,t,e}^{\text{end}} x_e = b_{s,t}^{\text{end}}, \quad \forall s \in S, \ \forall t \in T, \tag{7}$$

$$\sum_{e \in E \,:\, v \in \Gamma_e} x_e \leq 1, \quad \forall v \in \text{Trip}, \tag{8}$$

$$x_e \in \{0, 1\}, \quad \forall e \in E. \tag{9}$$

We assume that E contains BPs corresponding to dead-heading trips (i.e., empty repositioning of train units) between each pair of stations. Therefore, the model (5)–(9) always has a feasible solution. The costs of these BPs are equal to the costs of the corresponding dead-heading trips.

The objective function (5) minimizes the total costs of the selected BPs. At each station and for each train unit type, the sum of the changes in the initial (or final) inventory should be equal to the deficit or surplus in the initial (or final) inventory. This is ensured by constraints (6) (or (7)). Constraints (8) guarantee that for each trip at most one BP can be selected that modifies the composition of that trip. Finally, constraints (9) state that the decision variables are binary.

A disadvantage of the two-phase heuristic is that it restricts the solution space: The heuristic forbids two BPs to be selected both if there is a single trip that these BPs try to modify. Although it is very well possible that in practice both BPs fit together, it is hard to check this pairwise feasibility a priori. In order to compensate for this restriction, we apply the two-phase heuristic several times in a row until no further improvement is observed. That is, after each iteration, the BPs corresponding to dead-heading trips—the undesirable BPs—are deleted from the solution, and the two-phase heuristic is carried out once more. From that point of view, the method bears some similarity with column generation techniques: BPs are generated dynamically based on the current rolling stock circulation, and the "master problem" (5)–(9) selects the BPs into an overall solution.

6 Computational tests

In this section, we report our computational results, which are all based on the short-term planning process of NS, the main Dutch passenger railway operator. All test instances are based on timetable of the so-called 3,000 line of NS connecting Den Helder (Hdr) to Nijmegen (Nm). The stations are indicated in Fig. 8. The total length of the line is about 200 km. The line is operated in a cyclic timetable with a frequency of twice per hour in both directions.

Composition changes are possible at the terminals as well as at the intermediate stations Alkmaar (Amr) and Arnhem (Ah). Furthermore, train units may start and finish their daily duties in Amsterdam (Asd) and Utrecht (Ut).

The 3,000 line is served by 11 train units of type VIRM4 and 24 train units of type VIRM6. These are double-deck train units with 4 or 6 carriages, respectively, see Fig. 9. The maximally allowed train length is 12 carriages, thus the VIRM types admit 7 possible compositions, namely 4, 6, 44, 46, 64, 66, and 444.
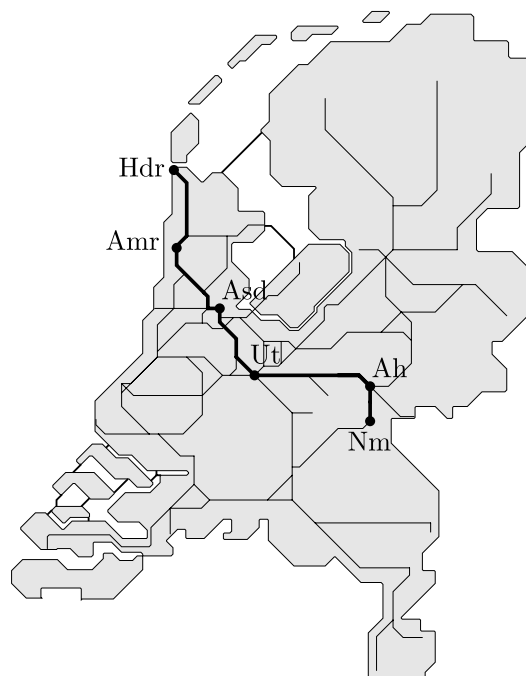
Fig. 8 The 3,000 line connecting Den Helder (*Hdr*) to Nijmegen (*Nm*) via Alkmaar (*Amr*), Amsterdam (*Asd*), Utrecht (*Ut*), and Arnhem (*Ah*)

Fig. 9 A VIRM4 train unit

Table 1 Objective coefficients in the three cost structures

Criterion	Obj-A	Obj-B	Obj-C
Off-balance	1,000.0	1,000.0	1,000.0
Carriage-kilometers	0.050	0.005	0.005
Composition changes	0.010	20.000	0.010
Shortage-kilometers	0.015	0.010	0.015

6.1 Instances

The timetable of the 3,000 line contains about 500 trips on each day. We studied instances for Saturday and Sunday, since these are the typical days of the week on which maintenance of the railway infrastructure takes place. This requires modification of the timetable, and thus of the rolling stock circulation as well.

In the first computational tests, we considered the timetable on Sunday and assumed that a certain part of the trajectory (Amr–Asd or Asd–Ut or Ah–Nm) is closed either until 14:00 or for the entire Sunday. The reduced timetables have about 400 trips; in each of these six timetables, we deleted the closed trips, and we updated the predecessor–successor pairs of the trips.

Moreover, in the case that the infrastructure is blocked for an entire Sunday, it is common in practice to modify the rolling stock circulations both on the day itself and on the previous day. This leads to instances with a planning period of two days: Saturday and Sunday. The 2-day test problems concern about 900 trips.

To illustrate the behavior of the solution methods under different priorities, we considered three different settings for the relative importance of off-balances, carriage-kilometers, shortage-kilometers, number of composition changes. We refer to these settings as Obj-A, Obj-B and Obj-C. Table 1 contains the values of the coefficients in the objective functions. Besides heavily penalizing the remaining off-balances, Obj-A focuses on carriage-kilometers, Obj-B on composition changes, and Obj-C on seat shortages.

The VIRM4 and VIRM6 train units have a limited number of possibilities to be attached to one another. Therefore, in a number of further artificial experiments, we changed the rolling stock types used. We split each VIRM4 and VIRM6 train unit into two identical parts (i.e., VIRM2 and VIRM3). This results in as much as 48 possible compositions for each trip, thereby increasing the complexity of the problem significantly. For these artificial rolling stock types, we considered the same one- and two-day instances and the same solution methods as for the original rolling stock types. All together, this results in 27 instances with VIRM4 and VIRM6 train units, and 27 instances with VIRM2 and VIRM3 train units. We refer to the instances with VIRM4 and VIRM6 train units as V46 and to those with VIRM2 and VIRM3 train units as V23.

6.2 Implementation issues

Throughout this section, Heur-1 denotes the iterative approach described in Sect. 5.1 as well as its results, while Heur-2 denotes the two-phase approach described in Sect. 5.2 as well as its results. To be more precise, the algorithm Heur-2 is composed of several consecutive runs of solving the model (5)–(9), each run using the output of the previous run as input. The iterative process continues until no further improvement is observed. In all test cases this convergence occurred within 4 iterations.

For each of the instances, we computed the input rolling stock circulation by the model of Fioole et al. (2006) with one of the objective functions Obj-A, Obj-B and Obj-C and with the penalty for the off-balances being set to zero. As a consequence, the rolling stock circulations were optimal solutions for the corresponding objective functions, but they also contained positive numbers of off-balances. We applied the heuristics Heur-1 and Heur-2 to resolve these off-balances.

In order to further evaluate the quality of the rebalanced solutions obtained by the heuristics, we also generated optimal solutions by applying the model of Fioole et al. (2006) *with* the penalties on the off-balances. Note that the solutions obtained in this way provide a lower bound to the solutions that are optimally rebalanced based on the input rolling stock circulation, since in the model of Fioole et al. (2006) the solutions are created completely from scratch: The input rolling stock circulation is not taken into account explicitly.

The graphs for Heur-1 have up to 5,800 nodes and up to 6,500 arcs. In the first phase of Heur-2, we generated 10,000 to 30,000 balancing possibilities. Thus the integer program in Phase 2 of Heur-2 has 10,000 to 30,000 variables. Moreover, it has 500 to 1,000 constraints.

The computations have been carried out on a PC equipped with a Pentium IV 3.0 GHz processor and 1 GB internal memory. For solving the model in Phase 2 of Heur-2 and for solving the model of Fioole et al. (2006) we used CPLEX 9.0 with the modeling software ILOG Opl Studio 3.7. The heuristic algorithms have been implemented in the Perl language (Heur-1) and in the C language (Heur-2).

6.3 The quality of the solutions

The results of the computational experiments of the heuristics Heur-1 and Heur-2 are presented in Tables 2 and 3.

Table 2 Results for the V46 instances. For each instance, 'IOB' denotes the number of off-balances in the input rolling stock circulation, 'OB' the number of remaining off-balances, 'Rest' the contribution of carriage-kilometers, seat shortages and composition changes, 'Obj' the objective value. 'Optimal' stands for the optimal solution obtained by CPLEX. For the optimal solution, 'ST' denotes the solution time (in seconds)

Instance		Heur-1			Heur-2			Optimal			
Name	IOB	OB	Rest	Obj	OB	Rest	Obj	OB	Rest	Obj	ST
46-C1-H-A	13	4	4,048	8,048	7	4,123	11,123	0	3,907	3,907	20
46-C1-H-B	8	2	624	2,624	2	449	2,449	0	545	545	24
46-C1-H-C	8	2	1,252	3,252	4	394	4,394	0	395	395	21
46-C1-F-A	11	8	3,707	11,707	7	3,771	10,771	3	3,660	6,660	15
46-C1-F-B	9	3	590	3,590	4	437	4,437	3	490	3,490	18
46-C1-F-C	13	5	3,772	8,772	5	996	5,996	3	379	3,379	15
46-C1-W-A	11	6	8,451	14,451	5	8,363	13,363	0	7,840	7,840	73
46-C1-W-B	9	1	1,875	2,875	4	1,184	5,184	0	1,165	1,165	74
46-C1-W-C	13	3	5,713	8,713	4	2,480	6,480	0	814	814	80
46-C2-H-A	11	3	3,747	6,747	4	4,048	8,048	0	3,654	3,654	19
46-C2-H-B	13	0	865	865	3	440	3,440	0	583	538	16
46-C2-H-C	12	3	1,766	4,766	4	381	4,381	0	371	371	19
46-C2-F-A	11	6	2,850	8,850	10	2,733	12,733	3	2,832	5,832	12
46-C2-F-B	11	4	516	4,516	5	328	5,328	3	428	3,428	12
46-C2-F-C	11	9	278	9,278	10	277	10,277	3	286	3,286	12
46-C2-W-A	11	3	7,917	10,917	8	7,546	15,546	0	7,088	7,088	103
46-C2-W-B	11	1	1,809	2,809	5	1,019	6,019	0	1,121	1,121	79
46-C2-W-C	11	6	3,268	9,268	8	2,375	10,375	0	728	728	86
46-C3-H-A	11	3	3,790	6,790	5	3,751	8,751	0	3,753	3,753	17
46-C3-H-B	12	0	849	849	2	456	2,456	0	557	557	17
46-C3-H-C	13	2	1,991	3,991	4	740	4,740	0	384	384	16
46-C3-F-A	11	6	2,935	8,935	7	2,931	9,931	4	2,961	6,961	12
46-C3-F-B	12	6	703	6,703	6	380	6,380	4	424	4,424	11
46-C3-F-C	11	7	774	7,774	7	298	7,298	4	302	4,302	11
46-C3-W-A	11	5	7,741	12,741	6	7,724	13,724	0	7,342	7,342	97
46-C3-W-B	12	2	2,142	4,142	7	1,066	8,066	0	1,178	1,178	71
46-C3-W-C	11	5	2,878	7,878	6	2,408	8,408	0	757	757	194

These tables also show the numbers of unresolved off-balances and the objective function values in the optimal solutions obtained by CPLEX for the model of Fioole et al. (2006).

Each row in the tables represents one instance. The instance names indicate the types used (V46 or V23) and the closed part of the infrastructure due to maintenance (C1 for Ah–Nm, C2 for Asd–Ut and C3 for Amr–Asd). Also, 'H' (or 'F') indicates that the planning period is a Sunday and that the blockage takes half the Sunday (or the full Sunday, respectively). The character 'W' stands for the entirely blocked Sunday and for a planning period of the whole weekend. Finally, the objective function is added (A for Obj-A, etc.).

Tables 2 and 3 show that on many instances Heur-1 results in significantly less remaining off-balances than Heur-2. This can be explained partially by the fact that in Heur-2 two selected BPs may not touch the same trip, even if the technical and market requirements would allow using both of them. Yet, in some cases the greedy method in Heur-1 terminates with a higher number of off-balances than Heur-2. Moreover, Heur-2 appears to be able to balance the four optimization criteria better than Heur-1. Indeed, the contribution of carriage-kilometers, shortage-kilometers and

Table 3 Results for the V23 instances. Here we use the same notations as in Table 2

Instance		Heur-1			Heur-2			Optimal			
Name	IOB	OB	Rest	Obj	OB	Rest	Obj	OB	Rest	Obj	ST
23-C1-H-A	23	2	3,928	5,928	6	4,007	10,007	0	2,961	2,961	730
23-C1-H-B	16	0	964	964	1	849	1,849	0	521	521	2,619
23-C1-H-C	23	5	6,568	11,586	10	1,276	11,276	0	302	302	310
23-C1-F-A	23	9	3,101	12,101	10	3,952	13,952	6	2,808	8,808	104
23-C1-F-B	22	6	762	6,762	7	440	7,440	6	451	6,451	133
23-C1-F-C	25	10	4,473	14,473	15	696	15,696	6	289	6,289	116
23-C1-W-A	23	3	7,894	10,894	4	7,561	11,561	0	5,824	5,824	715
23-C1-W-B	22	1	2,077	3,077	4	1,139	5,139	0	878	878	367
23-C1-W-C	25	4	5,805	9,805	8	2,350	10,350	0	595	595	1,173
23-C2-H-A	21	4	3,345	7,345	7	3,260	10,260	0	2,782	2,782	2,758
23-C2-H-B	20	2	870	2,870	2	522	2,522	0	542	542	247
23-C2-H-C	21	4	5,144	9,144	8	1,113	9,113	0	284	284	2,096
23-C2-F-A	24	10	2,171	12,171	10	2,851	12,851	6	2,128	8,128	114
23-C2-F-B	17	7	909	7,909	6	1,284	7,284	6	391	6,391	253
23-C2-F-C	28	10	408	10,408	16	218	16,218	6	217	6,217	114
23-C2-W-A	24	6	6,918	12,918	6	6,742	12,742	0	5,167	5,167	1,878
23-C2-W-B	17	2	2,299	4,299	4	1,469	5,469	0	782	782	333
23-C2-W-C	28	6	2,505	8,505	8	2,288	10,288	0	528	528	6,810
23-C3-H-A	20	5	3,224	8,224	3	3,634	6,634	0	2,917	2,917	348
23-C3-H-B	21	0	1,337	1,337	1	2,108	3,108	0	555	555	272
23-C3-H-C	24	4	2,265	6,265	7	314	7,314	0	301	301	392
23-C3-F-A	22	9	2,669	11,669	15	2,383	17,383	8	2,363	10,363	132
23-C3-F-B	18	8	990	8,990	9	340	9,340	8	831	8,371	111
23-C3-F-C	24	9	3,868	12,868	16	241	16,241	8	240	8,240	204
23-C3-W-A	22	5	7,628	12,628	7	6,855	13,855	0	5,404	5,404	597
23-C3-W-B	18	2	2,181	4,181	5	1,048	6,048	0	802	802	367
23-C3-W-C	24	5	8,186	13,186	7	2,309	9,309	0	551	551	1,264

composition changes to the objective function (this is the 'Rest' column in Tables 2 and 3) is often higher for Heur-1 than for Heur-2. This is particularly true for the instances with Obj-C.

It can be observed that the quality of the solutions highly depends on the structure of the input rolling stock circulation. Table 4 summarizes the differences in numbers of off-balances obtained by the heuristic methods in comparison with the optimal solutions; the average differences are referred to as Δ_1 and Δ_2. These values are much smaller for Obj-B than for Obj-A and Obj-C. In Obj-B, the number of composition changes has the largest weight. Therefore, the rolling stock circulation for this objective contains a very small number of couplings and uncouplings. Then the heuristic methods find ways to resolve many off-balances.

Table 4 The average difference between the off-balances in the heuristic solutions (OB_1 for Heur-1 and OB_2 for Heur-2) and in the optimal solution (OB_{opt})

		$\Delta_1 := OB_1 - OB_{opt}$	$\Delta_2 := OB_2 - OB_{opt}$
V46	Obj-A	3.78	5.44
V46	Obj-B	1.00	3.11
V46	Obj-C	3.56	4.67
V46	Total	2.78	4.41
V23	Obj-A	3.67	5.44
V23	Obj-B	0.89	2.11
V23	Obj-C	4.11	8.33
V23	Total	2.89	5.30

The rolling stock circulations for Obj-A and Obj-C are obtained by penalizing the carriage-kilometers and shortage-kilometers more heavily. The resulting larger number of composition changes is disadvantageous for both heuristic methods.

This can be explained partially as follows: If at a certain spot in the current rolling stock circulation no composition change takes place, then the models may allow either coupling or uncoupling of train units there in the modified circulation. However, a coupling in the rolling stock circulation cannot be converted directly into uncoupling, nor vice versa. This can be done only in two iterations. Thus the more composition changes take place in the input rolling stock circulation, the less flexibility is allowed for the models to solve the off-balances.

The rolling stock circulations of the V23 instances have on average 22.03 off-balances which is much more than the average of 11.15 of the V46 instances. However, for identical objectives the differences Δ_1 and Δ_2 of Table 4 are rather similar for V23 and V46. For Heur-1, the average differences in the number of off-balances are 2.78 and 2.89 for V46 and V23, respectively, while for Heur-2 these averages are 4.41 and 5.30. Since both heuristic algorithms solve a larger part of the initial off-balances of the V23 instances than of the V64 instances, they perform relatively better on the V23 instances than on the V46 instances. This is not surprising. The shorter train units give much more possibilities for adjustments without violating the constraints on the minimal and maximal lengths of the trains.

As said, Heur-2 is composed of several successive runs of the two-phase method of Sect. 5.2. In the first, second, and third run, the average differences Δ_2 over all V46 instances are 5.11, 4.44 and 4.41, while they are 6.74, 5.48 and 5.30 over all V23 instances. That is, the multiple consecutive runs of the two-phase heuristic indeed improved the performance of Heur-2. However, no further improvement was reached in a subsequent fourth run.

6.4 Computation times

Recall that the main motivation for using heuristic algorithms is the need for a quick solution process. In that sense, the heuristic algorithms are quite successful, since algorithm Heur-1 has a running time of 1–2 min for each run, and algorithm Heur-2 has a computation time of 4–5 s for each run.

In Tables 2 and 3, we give the computation times of the exact optimization method of CPLEX. Although this involves a relatively small instance of NS, it already shows how unpredictably the solution times grow when increasing the problem size. The V46 instances are easily solved for a one-day planning period (these instances are denoted by 46-C*-H-* and 46-C*-F-*, where * represents an arbitrary character) within 10–20 s. The two-day instances (denoted by 46-C*-W-*) require more than 5 times more CPU time on average. The V23 instances have a much more complex combinatorial structure due to the higher number of possible compositions. The solution time ranges from 2 min to 45 min, and a particular two-day instance requires nearly 2 h.

7 Conclusions

In this paper, we formulated the Rolling Stock Rebalancing Problem (RSRP). This is a relevant problem both in the short-term planning stage of a passenger railway operator and in the real-time operations.

Due to changes in the timetable (e.g., planned maintenance of the railway infrastructure or unplanned disruptions of the railway system), the input rolling stock circulation for a certain planning period contains a number of off-balances. RSRP is the problem of modifying the input rolling stock circulation such that the number of remaining off-balances is minimal, in combination with other criteria related to efficiency, service and robustness.

Two heuristics have been developed to solve RSRP. The performances of these heuristics are compared with each other and with the performance of the exact solution method of Fioole et al. (2006) that is used at NS, the main Dutch passenger railway operator. The comparison of the results is carried out on some (variants of) real-life instances of NS. These instances varied in size and complexity.

From the results presented in Sect. 6 we can conclude that both heuristics are very fast, even if the complexity of the instances is increased by extending the planning period or by increasing the number of feasible compositions per trip. The results also show that both heuristics can be used effectively, not only for solving larger size short-term planning problems, but also as a basis for solving real-time rescheduling problems in the case of a disruption of the railway system. In our further research we will especially focus on the latter application.

Appendix A: Complexity results

In this section, we prove that it is an NP-complete problem to decide whether an instance of the Rolling Stock Rebalancing Problem has a feasible solution, even if only a single station has a surplus in the final inventory and another station has a deficit in the final inventory. Subsequently, we extend the proof and show that the problem remains NP-complete in the case of an off-balance of a single train unit.

Throughout this section, the railway networks are drawn in time–space diagrams. Stations are represented by horizontal time-lines, where time increases to the right. The trips correspond to diagonal lines between the time-lines. Train stops are indicated by dots. Dotted arcs connect the arrivals of the trips to the departures of their successors.

Furthermore, there are two rolling stock types P and Q which can be combined with each other in one train. Each trip is operated by a single train unit of type P (in the figures represented by thick solid lines), by a single train unit of type Q (thick dotted lines) or by a two-train unit composition PP (thick dashed lines). Here the left-hand side of the string PP corresponds to the front of the train.

At each station, train units can be coupled or uncoupled only at one side of the train, indicated as the *shunting side* of the station. This shunting side of the station, either left or right, is indicated by [L] and [R]. The shunting side of a departing (or arriving) trip is defined as the shunting side of its departure (or arrival) station.

In the figures, for some trips the departure or arrival station is not indicated. These anonymous stations are all different. Trips to or from anonymous stations have a single train unit of a certain type in the rolling stock circulation. Anonymous stations with a departing trip have an initial inventory 1 for this type and 0 for the other type; the final inventory is 0 for both types. Similarly, anonymous stations with an arriving trip have an initial inventory 0 for both types; the final inventory is 1 for the type of the arriving train unit, and 0 for the other type.

A.1 Building blocks for the proofs: the gadgets

A *gadget* is a part of the railway network as shown in Fig. 10. The figure indicates 8 named stations (α, β, γ, δ, ε, α_1, α_2, ω_1, and ω_2), their shunting sides, the trips (among them trips s_1, s_2, t_1, and t_2), and the compositions of the trains on these trips in the input rolling stock circulation.

The initial and final inventories of stations β, γ, δ, and ε are 0. Stations α_1, α_2, ω_1, and ω_2 have undefined initial and final inventories in type P and they have zero initial and final inventory in type Q. Next, Lemma A.1 describes a number of restrictions on the feasible rolling stock circulations for a gadget.
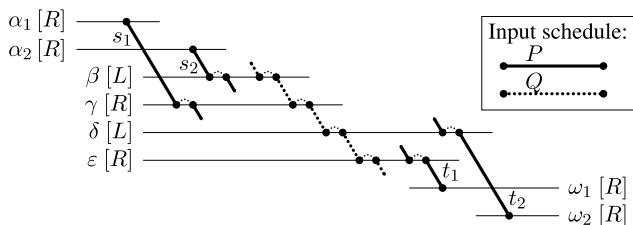

Fig. 10 A gadget

Lemma A.1 *Consider a rolling stock circulation for a gadget that satisfies the aforementioned constraints on the train lengths and on the shunting sides of the trains, as well as the constraints on the initial and final inventories. Then the following statements hold*:

(i) *Trip s_1 has composition PP if and only if trip t_1 has composition PP.*

(ii) *Trip s_2 has composition PP if and only if trip t_2 has composition PP.*

(iii) *At most one of the trips t_1 and t_2 can have composition PP.*

Proof

(i) If trip s_1 has composition PP, then a train unit must be uncoupled from it at station γ. This train unit must be coupled to the right-hand side of the train unit that travels from γ towards ε. Then the train unit of type P must be uncoupled at station ε. Actually, this is the only possibility to lead the uncoupled train unit to either ω_1 or ω_2. Moreover, this is the only way to get composition PP for trip t_1.

(ii) Similar to (i).

(iii) Two extra train units of type P can reach stations ω_1 and ω_2 only if the trip between γ and δ has composition PQP. However, this would violate the upper bound on the train length. □

In the remainder of this section, we use the simplified symbol in Fig. 11 for a gadget. The main purpose of a gadget is to bring an additional train unit either from α_1 to ω_1, or from α_2 to ω_2, but not both.

### A.2 Resolving an off-balance of k train units

Consider an undirected graph $G = (V, E)$ with $V = \{1, \ldots, n\}$ and let k be a positive integer with $k \leq n$. We build an instance of RSRP that is feasible if and only if G contains an independent set of size k. Here an *independent set* is a subset of nodes such that no pair of them is joined by an edge. It is well known that deciding whether a graph with n nodes has an independent set with k ($k \leq n$) nodes is NP-complete (Karp 1972).

Create two stations A and Z. For every node $v \in V$ with d_v neighbors, we create $d_v + 1$ stations $S_1^v, \ldots, S_{d_v+1}^v$. The shunting side of all these stations is [R].

For each $v \in V$, insert a trip from station A to station S_1^v and insert a trip from station $S_{d_v+1}^v$ to station Z. For each trip from station A to station S_1^v, create a predecessor trip from an anonymous station to A. For each trip arriving at Z,

Fig. 11 A simple symbol for a gadget
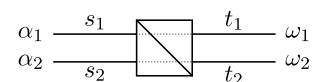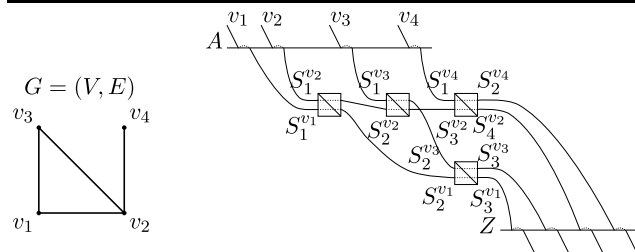
Fig. 12 An example of the construction of the network

insert a successor from Z to an anonymous station. All trips so far are operated with a single train unit of type P in the rolling stock circulation.

For each node $v \in V$ with neighbors $u_1, \ldots, u_{d_v}$ in G, assign stations $S_1^v, \ldots, S_{d_v}^v$ to the edges $u_1 v, \ldots, u_{d_v} v$, bijectively in an arbitrary way. For each edge $uv \in E$ with $u < v$, add a gadget as follows. Let S_i^u and S_j^v be the stations assigned to edge uv. Create four new stations β, γ, δ, and ε, set $\alpha_1 = S_i^u$, $\alpha_2 = S_j^v$, $\omega_1 = S_{i+1}^u$, $\omega_2 = S_{j+1}^v$ and insert all the trips described in the definition of a gadget. A station S_j^v with $1 < j < d_v + 1$ belongs to exactly two gadgets and has one arriving and one departing trip. The departing trip is the successor of the arriving trip.

This completes the railway network. Its size is polynomial in n as it contains $O(n^2)$ trips between $O(n^2)$ stations. The network for a small graph G is shown schematically in Fig. 12.

The rolling stock circulation satisfies the following inventory constraints. The initial and final inventories for type Q are 0 (except for some anonymous stations inside the gadgets). For type P, the initial and final inventories of stations S_j^v are 0. Station A has initial and final inventory k, while station Z has initial and final inventory 0. The initial and final inventories of the β-, γ-, δ-, and ε-stations of the gadgets are all zero.

The *target* inventories differ from these inventories at two points. The target final inventory of station A in type P is 0, the target final inventory of station Z in type P is k. Hence rebalancing the train units means that k train units of type P must be routed from station A to station Z.

Note that the inventory and train length constraints do not leave much choice for feasible rolling stock circulations. Each trip has either the same composition as in the rolling stock circulation or it receives the original composition extended by a single train unit of type P.

Theorem A.2 *Graph $G = (V, E)$ contains an independent set of size k if and only if the instance of RSRP constructed above has a feasible solution.*

Proof Suppose that G contains the independent set $\{v_1, \ldots, v_k\}$. Then a solution of the instance of RSRP can be obtained as follows. Couple the k train units of type

P at station A to the k trips that depart towards stations $S_1^{v_1}, \ldots, S_1^{v_k}$.

Consider any gadget that connects stations $S_j^{v_i}$ and $S_{j+1}^{v_i}$ for some indices i and j. We adjust the input circulation inside the gadget as follows. The trips of this gadget that are incident to stations $S_j^{v_i}$ and $S_{j+1}^{v_i}$ get composition PP. We also make all the necessary modifications to route the additional train unit through the gadget from $S_j^{v_i}$ to $S_{j+1}^{v_i}$. The adjustment of the gadgets can be done simultaneously since there is no edge between the nodes $v_1, \ldots, v_k$. Then all k excess train units reach station Z where they can be uncoupled. Therefore, the instance of RSRP is feasible.

Conversely, consider a solution of the instance of RSRP. At station A, k train units of type P are coupled to trips towards stations, say, $S_1^{v_1}, \ldots, S_1^{v_k}$. These train units pass through all gadgets that are related to the nodes $v_1, \ldots, v_k$ and end up at station Z. Then the nodes $v_1, \ldots, v_k$ form an independent set in G, as otherwise Lemma A.1 (iii) would be violated.   □

Corollary A.3 *It is an NP-complete problem to decide whether an instance of RSRP has a solution with zero off-balances.*

A.3 Resolving an off-balance of one train unit

Here we extend the construction described in the previous section. Thereby we prove that the maximum independent set problem can be reduced to 1-RSRP, the special case of RSRP with an off-balance of one train unit.

Let $G = (V, E)$ be an undirected graph with $|V| = n$, and let k be a positive integer with $k \leq n$. Consider the railway network constructed in the previous section. It is represented in Fig. 13 by stations A and Z and by the gray box.

Create $k + 1$ new stations $\alpha_1, \ldots, \alpha_k$ and ω. Insert $4k$ trips as follows (see Fig. 13). Create a trip from α_i to α_{i+1} for each $i = 1, \ldots, k$ (where $\alpha_{k+1} = \omega$) and insert their predecessors and successors from and to the anonymous stations. Also insert a trip that departs from station α_i and that returns to the same station and has no predecessor or successor. All these new trips are operated by a single train unit of type P in the rolling stock circulation.

Insert k additional gadgets $g_1, \ldots, g_k$. Let $s_1^{(i)}$, $s_2^{(i)}$, $t_1^{(i)}$ and $t_2^{(i)}$ denote the s_1-, s_2-, t_1-, and t_2-trips of gadget g_i. For each $i = 1, \ldots, k$, trips $s_1^{(i)}$, $s_2^{(i)}$, $t_1^{(i)}$, and $t_2^{(i)}$ (and their predecessor or successor trips from or to anonymous stations) connect gadget g_i to stations A, Z and α_i as shown in Fig. 13. The figure also indicates the compositions on these trips in the rolling stock circulation.

The railway network we constructed has polynomial size in n since it contains $O(n^2)$ trips and $O(n^2)$ stations.

The initial and final inventories for train units of type Q are 0 except for some anonymous stations inside the gadgets. For type P, the initial inventory of station A is k, at
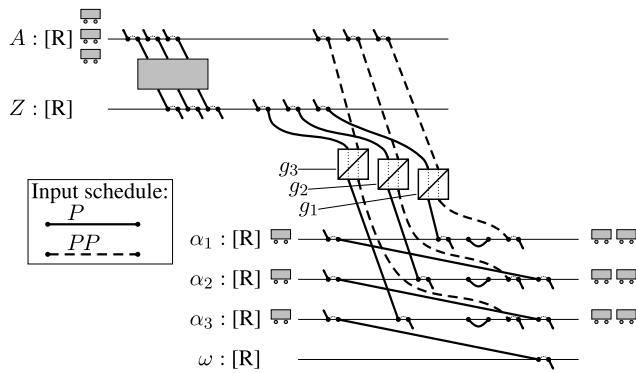
Fig. 13 One train unit to be routed ($k = 3$). At the start and the end of the time-lines, we give the initial and final inventories of type P realized by the rolling stock circulation

stations $\alpha_1, \ldots, \alpha_k$ it is 1, and at stations Z and ω it is 0. The final inventory of stations A, Z and ω is 0, while at stations $\alpha_1, \ldots, \alpha_k$ it is 2. Anonymous stations have initial and final inventory zero or one. All other stations (i.e., the β-, γ-, ε-, and δ-stations of the gadgets) have zero initial and final inventories.

The goal is to decrease the final inventory of station α_1 in type P by one and to increase the final inventory of station ω in type P by one.

Note that, the same as in the previous section, the shunting, inventory and train length constraints strongly restrict the possible rolling stock circulations. Each trip must be assigned the same composition as in the initial rolling stock circulation, possibly with a train unit of type P coupled or uncoupled at the appropriate side. In particular, the circulation of the train units of type Q does not change at all.

Theorem A.4 *Graph G has an independent set of size k if and only if the input circulation can be modified to decrease the final inventory of station α_1 in type P by one and to increase the final inventory of station ω in type P by one.*

Proof To increase the final inventory at station ω by one, the trip from station α_k to station ω must get a composition PP. Then the trip from station α_k returning to station α_k has no train unit to serve unless an extra train unit arrives earlier from gadget g_k. That is, trip $s_1^{(k)}$ from station Z to gadget g_k and trip $t_1^{(k)}$ from gadget g_k to station α_k must get composition PP, too. Then trip $s_2^{(k)}$ from station A to gadget g_k and trip $t_2^{(k)}$ from gadget g_k to station α_k must get composition P only, see Lemma A.1.

To correct the final inventory at station α_k, the trip from station α_{k-1} to station α_k must get composition PP, similar to the trip from station α_k to station ω. By repeating the above argument, all trips from station α_{i-1} to station α_i ($i = 2, \ldots, k$) must get composition PP. Thus all trips from

gadgets i to station α_i and all trips from station Z to station α_i ($i = 1, \ldots, k$) must get composition PP instead of composition P. Thus the k train units that start at station A can reach station Z. Invoking Theorem A.2 completes the proof. □

Corollary A.5 *It is an NP-complete problem to decide whether an instance of RSRP has a solution with zero off-balance even in the case of an off-balance of a single train unit.*

References

Ben-Khedher, N., Kintanar, J., Queille, C., & Stripling, W. (1998). Schedule optimization at SNCF: From conception to day of departure. *Interfaces*, *28*, 6–23.

Brucker, P., Hurink, J., & Rolfes, T. (2003). Routing of railway carriages. *Journal of Global Optimization*, *27*(23), 313–332.

Caprara, A., Kroon, L., Monaci, M., Peeters, M., & Toth, P. (2007). Passenger railway optimization. In C. Barnhart & G. Laporte (Eds.), *Handbook in OR & MS* (Vol. 14, pp. 129–187). Amsterdam: Elsevier.

Clausen, J., Larsen, A., & Larsen, J. (2005). *Disruption management in the airline industry: concepts, models and methods* (Technical Report). Informatics and Mathematical Modeling, Technical University of Denmark, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby. http://www2.imm.dtu.dk/pubdb/p.php?3763.

Fioole, P. J., Kroon, L. G., Maróti, G., & Schrijver, A. (2006). A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research*, *174*, 1281–1297.

Freling, R., Lentink, R. M., Kroon, L. G., & Huisman, D. (2005). Shunting of passenger train units in a railway station. *Transportation Science*, *39*, 261–272.

Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., Maróti, G., & Nyhave Nielsen, M. (2007). *Disruption management in passenger railway transportation* (Econometric Institute Report EI 2007-05). The Netherlands: Erasmus University Rotterdam. https://ep.eur.nl/handle/1765/8527.

Huisman, D., Kroon, L. G., Lentink, R. M., & Vromans, M. J. C. M. (2005). Operations research in passenger railway transportation. *Statistica Neerlandica*, *59*(4), 467–497.

Karp, R. M. (1972). Reducibility among combinatorial problems. In R. E. Miller & J. W. Thatcher (Eds.), *Complexity of computer computations* (pp. 85–103). New York: Plenum Press.

Kohl, N., Larsen, A., Larsen, J., Ross, A., & Tiourine, S. (2007). Airline disruption management—perspectives, experiences and outlook. *Journal of Air Transport Management*, *13*(3), 149–162.

Lentink, R. M. (2006). *Algorithmic decision support for shunt planning*. Ph.D. thesis, Erasmus University Rotterdam, The Netherlands.

Li, J., Mirchandani, P. B., & Borenstein, D. (2007). Vehicle rescheduling problem: Model and algorithms. *Networks*, *50*(3), 211–229.

Lingaya, N., Cordeau, J. F., Desaulniers, G., Desrosiers, J., & Soumis, F. (2002). Operational car assignment at VIA rail Canada. *Transportation Research B*, *36*, 755–778.

Peeters, M., & Kroon, L. G. (2008). Circulation of railway rolling stock: A branch-and-price approach. *Computers & Operations Research*, *35*(2), 538–556.