



HAL
open science

Timetable rearrangement to cope with railway maintenance activities

Luis Diego Arenas Pimentel, Paola Pellegrini, Said Hanafi, Joaquin Rodriguez

► To cite this version:

Luis Diego Arenas Pimentel, Paola Pellegrini, Said Hanafi, Joaquin Rodriguez. Timetable rearrangement to cope with railway maintenance activities. *Computers and Operations Research*, 2018, 95, pp123-138. 10.1016/j.cor.2018.02.018 . hal-01764963

HAL Id: hal-01764963

<https://hal.science/hal-01764963v1>

Submitted on 15 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Timetable rearrangement to cope with railway maintenance activities

Diego Arenas ^{a, b, d, *}, Paola Pellegrini ^c, Saïd Hanafi ^d, Joaquin Rodriguez ^b

^a Institut de Recherche Technologique Railenium, Valenciennes F-59308, France

^b Univ Lille Nord de France, F-59000 Lille, IFSTTAR, COSYS, ESTAS, Villeneuve Ascq, F-59650, France

^c Univ Lille Nord de France, F-59000 Lille, IFSTTAR, COSYS, LEOST, Villeneuve Ascq, F-59650, France

^d LAMIH UMR CNRS 8201, Université de Valenciennes et du Hainaut Cambrésis, Valenciennes, F-59313, France

Keywords:

Train timetabling

Railway maintenance

MILP

Temporary speed limitations

A B S T R A C T

Maintenance activities on the railway infrastructure are necessary to maintain its functionality and availability. Commonly, the maintenance activities are planned first. Then, the timetable is elaborated respecting the unavailability periods caused by the former. However, sometimes unplanned maintenance activities have to be introduced at short notice, and the timetable must be rearranged to respect the new unavailabilities. In addition, specific trains may be necessary to perform maintenance activities, and they are typically not scheduled in the timetable. In this case, the timetable may need to be further rearranged to integrate the maintenance trains. In this paper, we propose a mixed-integer linear programming formulation that rearranges a timetable to cope with the capacity consumption produced by maintenance activities. It includes the consideration of maintenance trains and other specific constraints, such as temporary speed limitations. In this formulation, the rearrangement of the timetable is optimized based on a microscopic representation of both the infrastructure and the rolling stock. We assess three algorithms founded on this formulation on a real case study in the French railway network and we show their practical applicability.

1. Introduction

The *railway capacity* can be defined as the maximum number of trains that would be able to operate on a given railway infrastructure, during a specific time interval, given the operational conditions (Hachemane, 1997). Currently, most railway systems around the world experience an increasing demand for railway capacity which can only be achieved either by the construction of new infrastructure or by the improvement of the exploitation of the existing one.

Maintenance Activities (MAs) are necessary to maintain the good state of the railway infrastructure, allowing trains to circulate safely and fluidly, thus ensuring the availability of the railway capacity. However, while performing MAs (note that a table of acronyms is provided), the train circulations in the concerned locations are impacted. This impact depends on the type of MA performed, which in general implies circulation interdiction on some track segments and temporary speed limitations on neighbouring

ones. This means that during the performance of MAs the available railway capacity is reduced. Moreover, most MAs require one or more *Maintenance Trains* (MTs). These are particular trains whose rolling stock is specifically equipped to perform maintenance tasks or to transport maintenance materials. The circulations of these MTs may also impact other trains.

Timetables are typically elaborated by considering the track unavailability periods due to MAs. A timetable is *feasible* if all planned train circulations are conflict-free. A *conflict* exists when two trains travelling at the planned speed would concurrently require the same track segment. To guarantee the feasibility of a timetable, the minimum separation between trains is often overestimated during the planning process. This often brings to an inefficient capacity utilization, since more capacity than what strictly necessary is allocated to each train.

Unplanned MAs may be necessary, e.g., due to an accident or tracks malfunction, and may require some rearrangements of the existing timetable. These activities are typically scheduled a few days or few hours in advance. If complex operations are necessary, these are divided into smaller operations that are scheduled separately from one another. In this paper, we assume that this division has already been performed and thus, we deal with the smaller, indivisible activities defined by the infrastructure manager, to which we refer simply as MAs. In the practice, the required timetable re-

* Corresponding author at: Institut de Recherche Technologique Railenium, Valenciennes, F-59308, France.

E-mail addresses: diego.arenas@railenium.eu (D. Arenas), paola.pellegrini@ifsttar.fr (P. Pellegrini), said.hanafi@univ-valenciennes.fr (S. Hanafi), joaquin.rodriguez@ifsttar.fr (J. Rodriguez).

Acronyms

BS	Block Section
DIRiT	Directly impacted Trains
MA	Maintenance Activity
MA_{01-10}	Maintenance locations 1–10
MILP	Mixed Integer Linear Programming
MT	Maintenance Trains
TDS	Track Detection Section
TSL	Temporary Speed Limitation
TSLiT	TSL impacted Trains
OTs	Operational Trains
RECIFE	REcherche sur la Capacité des Infrastructures FERroviaires
RM-{F,B,BF,GF}	The algorithms of RECIFE-MAINT: Full, Battery, Battery-Full and Greedy-Full, respectively
rTDS	reservation TDS

arrangements are usually made either by hand, based on the experience of the dispatchers, or by resorting to some optimization tool, based on macroscopic aspects of the infrastructure.

The objective of this work is, first, to propose a formulation that allows the insertion of unplanned MAs into an existing timetable while guaranteeing its feasibility. More specifically, we present RECIFE-MAINT: a mixed integer linear programming (MILP) formulation to perform rearrangements on planned train circulations while minimizing the scheduled time deviations with respect to the existing timetable. RECIFE-MAINT has been developed as part of the decision support tool named RECIFE (REcherche sur la Capacité des Infrastructures FERroviaires) (Rodriguez et al., 2007). Thanks to the microscopic representation of the infrastructure and rolling stock, a number of specific circulation constraints can be defined in RECIFE-MAINT thus, ensuring the feasibility of the resulting timetable while optimizing the railway capacity utilisation. Additionally, RECIFE-MAINT takes into account specific aspects related to MAs that are often disregarded in the literature, such as temporary speed limitations and planning of MTs.

Moreover, we present three algorithms founded in RECIFE-MAINT to solve the problem of rearranging a timetable to cope with MAs. We use these algorithms in the experimental phase to solve instances based on a case study of the French railway network. Furthermore, we compare the performance of these algorithms with an emulation of the current practice.

The rest of the paper is organized as follows. [Section 2](#) describes the railway infrastructure representation considered in this paper. [Section 3](#) reviews the related scientific literature. [Section 4](#) details the problem addressed in this paper. [Section 5](#) presents the complete formulation of RECIFE-MAINT. [Section 6](#) describes the algorithms we propose. [Section 7](#) introduces a real world case study, which is then used to perform experiments; results are presented and discussed. Finally, [Section 8](#) enlists our conclusions and perspectives.

2. Railway infrastructure representation

The railway infrastructure can be represented in multiple ways. They can be classed into two main groups: macroscopic and microscopic.

On the one hand, macroscopic representations are based on the abstraction of large elements existing in the infrastructure, e.g., stations and lines connecting them. Commonly, they use approximations to define circulation and capacity constraints, e.g., the minimum headway time between two trains on the same track

segment, the maximum number of trains that a station can host at a given time, etc.

On the other hand, in microscopic representations, the actual elements existing in the infrastructure are explicitly modelled, e.g., signals and block sections. This allows the modelling of more sophisticated circulation and capacity constraints, e.g., considering the signalling and interlocking systems.

The approximations which are introduced when considering trains in macroscopic representations often need to be overestimations of reality. For example, this is often the case for the minimum separation between consecutive trains: if the precise routes followed by the two trains in a station are not known, their separation must be at least equal to the highest possibly needed. Indeed, this typically leads to a sub-optimal infrastructure capacity utilization. On the contrary, the high level of details considered in microscopic representations allows the accurate computation of quantities as the minimum separation between trains. This allows the maximization of the capacity utilization, at the cost of a larger amount of data necessary to model an infrastructure.

In microscopic representations, two basic elements compose a railway infrastructure: tracks and switches. Tracks allow the circulation of trains, and switches enable trains to be guided from one track to another.

Ever since the first railway systems were created, different methods were developed to allow trains to circulate in a coordinated and safe manner. Currently, the most widely used method is the signalling system coupled with the route-lock sectional-release interlocking system.

The signalling system is used to direct traffic and keep trains clear of each other at all times. It is based on the division of the railway infrastructure into block sections. A *block section* is a segment of track which can be used by at most one train at a time. The length of a block section is variable; it depends on several factors, e.g., topology of the terrain, braking capacities of the trains authorised to use them, etc. The entrance of a block section is guarded by an entry signal. *Signals* are semaphores that give the train drivers some information about the utilization state of the following block section(s), e.g., a red light, indicates that the next block section is currently used by another train. The driver will then modify the running profile of his train accordingly, i.e., decelerate, stop, etc.

To automate the signalling system, it is necessary to detect the presence of a train on a track segment. Several technologies have been developed for this purpose, e.g., track circuits, axle counters, etc. A *track detection section* (TDS) is a segment of a track where the presence of a train can be automatically detected. A single TDS can form a block section, however, for practical reasons, it is quite common for a block section to contain several TDSs. One of these reasons is strictly related with the route-lock sectional-release interlocking system, which is explained below.

Before a train can use a block section, the latter has to be *reserved* by said train. To reserve a block section, two conditions must be satisfied: First, the block section is not being used, or reserved by another train. Second, all the TDSs inside the block section are set and locked on the desired positions. Should the block section contain several TDSs, all of them are reserved at the same time. This part is called *route-lock*.

As the train runs inside a block section, its presence is detected on the TDSs forming the block. When the train finishes the occupation of a TDS inside a block section, the TDS is released and it can be reserved by another train. The procedure of releasing independently the TDSs forming a block section is known as *sectional-release*. This procedure is particularly useful in locations where TDSs are shared by multiple block sections, e.g., when they include a switch, as it helps reducing the headways between trains and, thus, maximizing the utilization of the railway capacity.

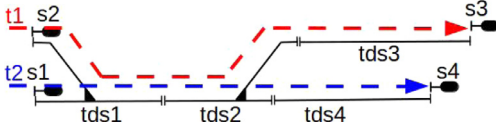


Fig. 1. Example of a microscopic representation of the railway infrastructure. Four types of elements are represented: (I) TDSs: tds1, tds2, tds3 and tds4. (II) Signals: s1, s2, s3 and s4. (III) Block sections: BS_s1-s3, BS_s1-s4, BS_s2-s3 and BS_s2-s4. (IV) Trains: t1, using BS_s2-s3 and t2 using BS_s1-s4.

Fig. 1 shows an example of a microscopic representation of the infrastructure. Note that the block sections used by the trains share 2 TDSs: tds1 and tds2. We use this example to illustrate the route-lock, sectional-release principle. Suppose that t2 follows t1 and t1 reserves BS_s2-s3 just seconds before t2 tries to reserve BS_s1-s4. This means that t2 cannot reserve BS_s1-s4 because part of this block section is currently used by t1. However, as soon as t1 releases tds2, it becomes available. This allows t2 to reserve BS_s1-s4 even if t1 is still occupying tds3. This kind of mechanics can only be modelled using microscopic representations, and allows high accuracy in the scheduling process.

3. Literature review

A large number of works in the scientific literature deal with the timetabling problem. For a comprehensive review, we refer the reader to [Cacchiani and Toth \(2012\)](#) and [Siebert and Gorigk \(2013\)](#).

However, only few contributions deal with the unavailability periods of track segments caused by MAs. We classify them into three categories according to the manner in which they deal with MAs:

- Fixed Timetables, Variable Maintenances (FTVM),
- Variable Timetables, Fixed Maintenances (VTFM) and
- Variable Timetables, Variable Maintenances (VTVM).

In the first category, FTVM, we group the contributions where MAs are scheduled by considering the train timetable, but without modifying it. This means that the generated MAs' plan can have conflicts with the timetable, and these conflicts are left to be solved later. In [Higgins \(1998\)](#), the authors propose a mathematical formulation based on a macroscopic representation of the infrastructure. The objective is to determine the best scheduling for MAs and maintenance crews, while minimizing the disruption to the trains in the timetable and the amount of time needed to complete the MAs. They propose a Tabu Search algorithm. The authors of [Budai et al. \(2004\)](#) propose several heuristic approaches to find near optimal track utilization intervals for carrying out preventive MAs while minimizing the inconvenience for the railway undertaking and the infrastructure utilization time. Their model is based on a macroscopic representation of the infrastructure, and allows to cluster several MAs together into maintenance packages. Another work in this category is [Peng et al. \(2011\)](#). In this work, the authors propose a time-space network model to schedule MAs while minimizing the total travel costs of the maintenance crews as well as the inconvenience for the railway undertaking. They use an iterative heuristic to solve the problem.

The second category, VTFM, includes the contributions in which MAs are fixed and the timetables are created or modified to cope with them. In [Caprara et al. \(2006\)](#), the authors present a graph-based mathematical formulation to solve the timetabling problem while considering MAs. In the author's macroscopic representation of the infrastructure, MAs are modelled as periods of time in which the tracks connecting two stations are unavailable. The authors use a Lagrangian heuristic algorithm. Another example in this category is [Sourd \(2010\)](#). The author presents SIOUCS, a tool that optimizes

the crossing of trains in a line segment with two tracks, where one of the tracks is unavailable due to an incident or MA. This means that train movements are coordinated to use the single track, in the form of batteries. A battery is a group of trains circulating in the same direction. The optimization criterion is the reduction of the total delay experienced by trains. The infrastructure is represented macroscopically, and a headway time is defined for trains forming the same battery. The solution approach is based on dynamic programming and scheduling theory.

The third category, VTVM, collects the works where both timetables and MAs are scheduled at the same time. This category can be viewed as a generalisation of the previous ones. In the formulation presented by [Albrecht et al. \(2013\)](#), trains have a minimal departure time and their routes across the infrastructure can be changed. MAs have a desired start and end dates, they can be divided into subtasks if necessary and they have a lateness cost function. The solution approach is based on a combination of heuristic methods. Basically, it schedules track utilizations (by trains or subtasks belonging to a MA) one at the time. Some random perturbations are added by the algorithm to create a variety of solutions from which one is chosen. The objective function is the minimization of both the total delay of trains and the accumulated costs of MAs. The authors of [Forsgren et al. \(2013\)](#) present a mathematical formulation that is solved using a MILP solver. In their formulation, trains can be cancelled, re-routed and rescheduled. MAs, instead, cannot be cancelled, but they can be rescheduled within a predefined time window. The infrastructure is represented macroscopically through a graph where stations or junctions are vertexes and tracks are edges. The capacity of the stations is also taken into account. The objective is to generate a feasible timetable, while minimizing the number of cancelled trains and the total accumulated delay.

As several other authors, e.g., [Higgins \(1998\)](#) and [Forsgren et al. \(2013\)](#), we agree in the premise that, despite of the increased complexity, better results may be achieved by simultaneously scheduling MAs and train movements, i.e., VTVM. However, the problem we tackle in this paper belongs to the second category: VTFM. The main motivation behind this choice is that, in a typical railway operations planning schema, the process of MA scheduling is performed separately from the timetabling process. This is not only a temporal separation, e.g., MAs are planned before timetabling, but also an organizational one, i.e., different organizations are in charge of the two scheduling processes. In most European countries, as a rule of thumb, the MA schedule has a higher priority and is less likely to be altered to accommodate timetabling requests. From a practical perspective, then, the VTFM approaches are closer to today's reality. In fact, after several interviews with the French infrastructure manager, SNCF Réseau, we confirmed that they are indeed more interested in an approach considering fixed schedules for MAs, at least in a perspective of short-medium term deployment. Differently from the reviewed works, we use a microscopic representation of the infrastructure. Moreover, we consider specific constraints imposed by MAs, as temporary speed limitations and the schedule of MTs. These constraints are never considered in the cited literature and can have a substantial impact on the feasibility, quality and complexity of the solutions.

4. Detailed problem description

As mentioned, the maintenance needs of the infrastructure are commonly defined before the conception of train timetables. Then, the timetables are conceived taking into account the temporal unavailability of track segments caused by MAs. This is often done considering a macroscopic representation of the infrastructure, hence, the separation between pairs of trains is constrained

by an estimation of a minimum separation time. This value ought to be the highest possible one considering the routes of both trains, as discussed in Section 2. Nevertheless in practice, a lower value is used to avoid excessive capacity utilization. In this case, a timetable rearrangement may be necessary before the operations to guarantee feasibility. The same necessity arises when unexpected maintenance needs emerge in the pre-operational phase and additional MAs need to be scheduled into existing timetables.

For safety reasons, some types of MAs impose a *temporary speed limitation* (TSL) on adjacent tracks, e.g., a speed limit of 40 km/h on a track segment where trains normally run at 160 km/h. By applying a TSL on a track segment, the time needed for a train to circulate over this segment is increased. This may require adjustments to the schedule of the train.

Furthermore, the circulation of MTs necessary to perform MAs shall also be taken into account. The insertion of these trains will likely require additional alterations to other trains' circulations. We denominate the trains that are scheduled in the original timetable as *operational trains* (OTs). OTs include several types of trains, e.g., passenger, freight, shunting, empty rides, etc.

There are mainly two types of adjustments that can be made to the trains in a timetable: rescheduling and re-routing. We do not consider train cancellations although it may be an alternative, since in the pre-operational phase this is not always an option. Indeed, train cancellations may require negotiations with the railway undertaking, which are out of the scope of this work.

Re-routing decisions of the same nature can be applied to OTs and MTs. A *route* is the succession of tracks that a train can use to circulate from an origin to a destination. The routes available for a train depend on both the infrastructure and rolling stock characteristics, e.g., some types of rolling stock are not allowed to run in some parts of the infrastructure. In a timetable, trains are scheduled along with their *default route*. A re-routing decision is the assignment of a route different from the default one to a train.

The nature of scheduling decisions is different for OTs and MTs. On the one hand, for OTs, the departure time cannot be advanced, only delayed. The reason for this is that in a pre-operational phase, the timetable is normally already known by the users. For instance, the advance of a departure may lead passengers to miss their train. Another particularity of OTs is that a delay may be only introduced while they are dwelling at some station. This means that an OT can only stop at its predefined intermediate stopping stations. This behaviour is better known as *green wave policy*. There are several reasons to do so: better passengers comfort, lower power consumption, compliance with security and safety issues, etc. Ideally, the modifications to the schedule of OTs are made while trying to preserve the planned times of the original timetable. On the other hand, since MTs are not bind to a timetable, they can be scheduled as it is best convenient. However, MTs have to be present at MA locations when they start, and leave when the MAs are finished. Provided that this holds, MTs can be scheduled to stop at any signal along the track, to be overtaken by OTs.

To summarize, the problem of rearranging a timetable to cope with MAs can be formalized as follows.

Given:

- A railway infrastructure.
- An initial timetable.
- A set of MAs.

Find a working timetable compatible with all MAs where:

- All OTs from the initial timetable are scheduled according to the green wave policy.
- All the required MTs are scheduled.
- All capacity, safety and TSL constraints are respected.
- The scheduled time deviations with respect to the initial timetable are minimized.

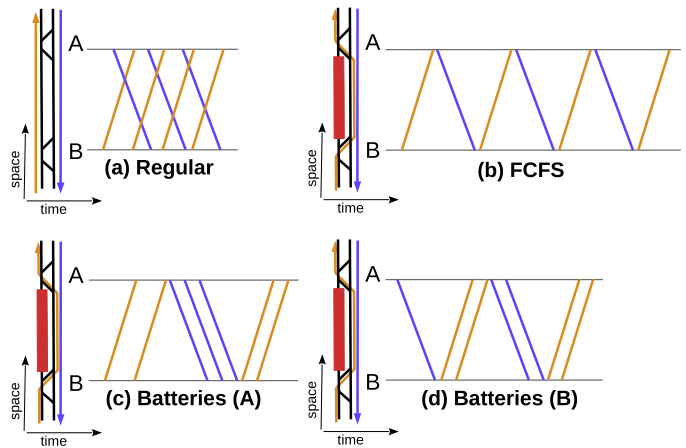


Fig. 2. Example of train rescheduling using FCFS and batteries approaches. In (a) a space-time diagram shows a regular schedule of seven trains in a double-track segment connecting stations A and B. Provided that one track becomes unavailable, new schedules are proposed in which trains are rescheduled by applying FCFS (b) and batteries (c and d) approaches.

For this problem, we think that a microscopic representation of the infrastructure is necessary because of the following reasons:

- MAs can occur in small parts of the infrastructure, therefore a way to identify specific track sections is necessary.
- It must be possible to locally re-route trains. Thus, it is essential to be able to distinguish the different routes available for a train and, hence, to exploit all the possibilities offered in the practice by the interlocking system. Even small differences between routes can have a large impact on the quality of the final timetable.
- The MAs limit the railway capacity available for OTs, hence, this capacity should be used as efficiently as possible.
- In this context of scarce capacity, guaranteeing the feasibility of a timetable is particularly important.

4.1. Current practice: train batteries

The constitution of train batteries is a commonly used approach to deal with the problem presented above. A *train battery* is a group of trains circulating in the same direction. Scheduling trains in batteries allows the maximization of the railway capacity utilization in a single-track segment with traffic running in both directions. Indeed, train batteries allow the reduction of headway times between trains, compared to the situation in which trains travelling in opposite directions cross in, e.g., a first-come first-served (FCFS) order.

Fig. 2 shows an example of train rescheduling. The space-time diagrams represent a small part of a railway infrastructure containing two stations: A and B. These stations are connected by a double-track segment, therefore, in Fig. 2(a) trains are scheduled in both directions without any conflict. However, if one track becomes unavailable, e.g., because an unplanned MA, the trains need to be rescheduled to use the available track without any conflict. Three new train schedules are obtained by applying two approaches: FCFS (Fig. 2(b)) and train batteries (Fig. 2(c) and (d)).

By comparing the different schedules proposed in Fig. 2, we observe that the train schedules obtained by using train batteries (Fig. 2(c) and (d)) effectively rescheduled all trains present in Fig. 2(a) while utilizing the railway capacity more efficiently than the schedule obtained by FCFS (Fig. 2(b)): the same number of trains are scheduled using the infrastructure for a shorter time. Moreover, trains rescheduled in batteries experience a lower delay than the ones rescheduled following a FCFS approach.

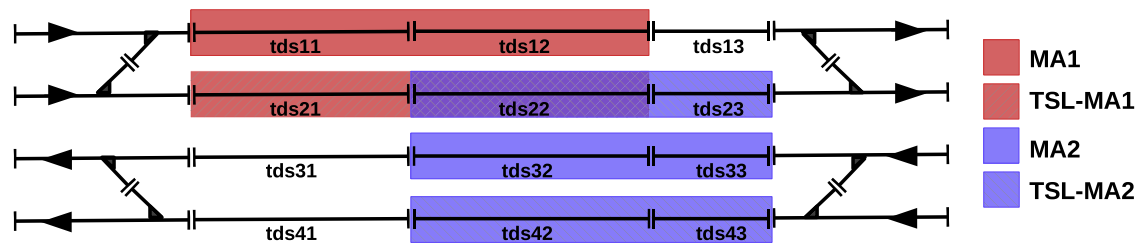


Fig. 3. Example of two MAs affecting a set of TDSs. MA1 directly affects tds11 and tds12, while imposing a TSL in tds21 and tds22. MA2 directly affects tds32 and tds33, and it imposes a TSL in two set of TDSs: First, tds22 and tds23, and second, tds42 and tds43.

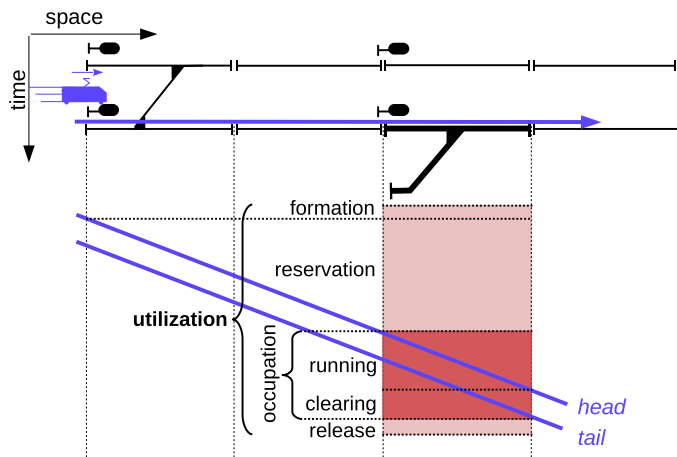


Fig. 4. The utilization time of a train traversing a TDS.

By analyzing the differences between Fig. 2(c) and (d), we can identify three parameters that characterize a schedule obtained by following a batteries approach: The number of train batteries (3 in Fig. 2(c) and 4 in Fig. 2(d)), the number of trains per battery (2,3,2 and 1,2,2,2 in Figs. 2(c) and 2(d), respectively) and which is the direction of the first battery (B to A versus A to B). Note that in some cases it may be convenient to compose a schedule by considering some batteries of only one train (Fig. 2(d)). The settings of these parameters may have a high impact on the quality of the timetable produced.

5. Problem formulation

In this section we present RECIFE-MAINT, a MILP formulation to tackle the problem described in Section 4. It is based on the RECIFE-MILP formulation, presented by Pellegrini et al. (2015). The novel features of RECIFE-MAINT include the introduction of new variables and constraints for the implementation of the green wave policy for OTs, the model of MAs, the planning of MTs and the observance of TSLs. Additional information concerning these new features is given below.

Since we use a microscopic representation of the infrastructure, all elements detailed in Section 2 are present in our formulation, i.e., TDSs, block sections, signals, etc. Hereafter we give a detailed explanation of some terms and additional concepts. Indeed, dealing with a country-wide, microscopic railway infrastructure representation may be computationally unattainable at the actual state of the technology. Hence, it is currently necessary to decompose the network into smaller portions that can be treated in practical computational times. Let a *control zone* be the portion of the infrastructure we consider in the formulation. We define a *train* as the movement of a pre-determined type of rolling stock over the infrastructure to fulfil a required *journey* between an *origin* and a *destination*. It is possible that the origin (destination) of a train's

journey is located outside the control zone. In such case, we consider that the origin (destination) of the journey is the first (last) TDSs which needs to be crossed by the train within the control zone.

As discussed in Section 4, we distinguish two types of trains: OTs and MTs. In the timetable, OTs may be scheduled to perform intermediate stops at predefined locations, e.g., stations. We call these locations *control points*. Note that a minimum *dwell time* at these control points is often imposed. Differently from RECIFE-MILP, OTs must be scheduled according to the green wave policy, which involves a number of modifications on the formulation's constraints.

Depending on the control zone layout and the rolling stock, a train can follow different routes to accomplish its journey. A route is a complete sequence of TDSs which can be traversed by a train to fulfil its journey: it goes from its origin to its destination and passes by its control points, if any. As mentioned in Section 4, each train is associated with a predefined default route.

Concerning MAs, we denote as *directly affected TDSs* the set of consecutive TDSs where the MA takes place. No train circulation is allowed in these TDSs during the realization of a MA. Instead, the *TSL affected TDSs* are the ones, typically contiguous to the directly affected TDSs, where train circulations are allowed during the performance of the MA provided that they respect a TSL. A single MA can define different sets of TSL affected TDSs. Moreover, a single TDS can be TSL affected by more than one MA at the same time. In this case, the lowest speed limitation is imposed. Fig. 3 shows a representation of the elements described above, note that MA2 imposes a TSL in two sets of TDSs, and tds22 is affected by the TSL imposed by both MA1 and MA2. The consideration of MAs represents by itself a number of additions to the RECIFE-MILP formulation, but most importantly, the modelling of TSLs for some TDSs during specific time intervals involves even a bigger set of adaptations and extensions specific to RECIFE-MAINT. In particular, the fact that the time necessary to travers a TDS may depend on the time at which the occupation starts is a major novelty introduced here.

The performance of a MA typically requires the presence of at least one MT. Therefore, the entire schedule of the MT must be defined, i.e., a new train is added to the timetable. This represents an additional capacity consumption in a potentially busy infrastructure. To facilitate the capacity allocation for MTs, and because they are not subject to the same service constraints as OTs, e.g., passenger comfort, MTs are not required to comply with the green wave policy. This means that MTs can be scheduled to stop anywhere as long as a very specific set of timing constraints are met. Indeed, different types of constraints are applied to a MT whether it is arriving at or departing from a MA location. On the one hand, the MT must arrive to the MA location at the exact time as the MA begins, i.e., the final arrival time is constrained. On the other hand, the MT must depart from the MA location as soon as the MA is finished, i.e., the departure time is constrained. In the following, we will refer to these two MTs as *inbound MTs* and *outbound MTs*,

respectively. The presence of MTs represent another set of important differences with respect to RECIFE-MILP, which does not allow the addition of new trains nor contemplates different scheduling modes for different types of trains.

Concerning train circulations, we define the *occupation time* of a train over a TDS as the time elapsed between the entrance of the head of the train into the TDS and the exit of its rear. The occupation time can be divided into two parts: The *running time* is the time elapsed between the entrance of the head of the train in the TDS and its exit. The *clearing time* is the time elapsed between the exit of the head of the train and the exit of its rear. Since a MA can impose a TSL over a set of TDSs, we consider two different running and clearing times: *nominal* and *restricted*. The running and the clearing times are inputs of RECIFE-MAINT and they are calculated by taking into account the characteristics of the tracks (length, gradients, curves, interlocking systems, speed constraints, etc.), the rolling stock (max speed, acceleration, braking capabilities, etc.) and the routes (origin, intermediate stops, destination). When an intermediate stop is to be performed, no dwell time is considered in the running time. Instead, the suitable deceleration and acceleration are computed. The same sequence of TDSs can be used along several routes if different rolling stocks are employed or if different intermediate stops are to be performed.

As introduced in Section 2, when using the signalling and interlocking system, a train must reserve all TDSs contained in a block section before occupying them. The moment in which the reservation of a TDS begins is determined by the signalling system existing in the control zone. Each block section has a reference *reservation TDS* (rTDS). All TDSs belonging to a block section have the same rTDS. The *reservation time* of a TDS starts when the head of the train enters the rTDS of the TDS and lasts until the train starts occupying the TDS. In a signalling system of n aspects, the rTDS is the first TDS of the $n - 2$ preceding block section. Moreover, before a train can reserve a block section, all TDSs inside the block section itself must be set and locked in the desired position. The time needed to perform this task is the *formation time*. Analogously, once a train has cleared a TDS, a *release time* is necessary to set the TDS in its default position. To successfully coordinate train movements while respecting capacity and safety constraints it is imperative to ensure that two trains do not utilize a given TDS at the same time. The *utilization time* of a TDS by a train is defined as the sum of the formation, reservation, occupation and release times. Fig. 4 illustrates the different parts of the utilization time. Utilization times of a TDS by two trains cannot overlap.

5.1. MILP model

The notation we introduce in this section is based on the following formalism: Input values of the problem are underlined, e.g., \underline{at} and \underline{dt} represent an arrival and departure time, respectively, of the initial timetable. The other symbols, i.e., non-underlined, represent either parameters, variables or elements/sets of objects, e.g., trains, TDSs, etc.

5.1.1. Data

Table 1 presents the data used by RECIFE-MAINT. Furthermore, Fig. 5 displays a graphical representation of the main data concerning the TDSs belonging to route r of train t circulating in an infrastructure equipped with a signalling system with 3 aspects.

5.1.2. Variables

Non-negative continuous variables:

for all triplets of $t \in T$, $r \in R_t$ and $v \in \mathcal{V}_r$:

$o_{t,r,v}$: time at which train t starts the occupation of TDS v along route r ;

Table 1
Data used by RECIFE-MAINT.

Infrastructure	\mathcal{V}_r	set of TDSs composing route $r \in \bigcup_{t \in T} R_t$	
	α_r, Ω_r	first and last TDSs of route $r \in \bigcup_{t \in T} R_t$	
	$\mathcal{V}_r^{fbs}, \mathcal{V}_r^{lbs}$	set of TDSs in the first and last block sections of route $r \in \bigcup_{t \in T} R_t$	
	$p_{r,v}, n_{r,v}$	previous and next TDSs w.r.t. TDS v along route $r \in \bigcup_{t \in T} R_t$	
	$ref_{r,v}$	reference TDS for the reservation of TDS v (rTDS) along route $r \in \bigcup_{t \in T} R_t$, depending on the interlocking system	
	$\mathcal{V}_{r,v,v'}^{bet}$	set of TDSs between v and v' along route $r \in \bigcup_{t \in T} R_t$	
	$\underline{bf}_{r,v}, \underline{br}_{r,v}$	formation and release time of the block section which contains TDS v along route $r \in \bigcup_{t \in T} R_t$	
Operational Trains	OT	set of OTs	
	$\underline{bt}_t, \underline{et}_t$	time at which train $t \in OT$ begins its journey, and the end time of t 's journey given \underline{bt}_t , the default route and the intermediate stops	
	C_t	set of control points for train $t \in OT$, including its destination	
	$\underline{at}_t, c, \underline{dt}_t, c$	scheduled arrival and departure times for train $t \in OT$ at control point $c \in C_t$	
	$w_{t,c}$	parametrisable weight associated to one time unit of arrival delay of train $t \in OT$ at control point $c \in C_t$	
	$\underline{dp}_{t,c}$	minimum dwell time for train $t \in OT$ at control point $c \in C_t$ (dwell period)	
	T_t^{bal}	set of trains that result from a rolling stock balance operation, i.e., the turnaround, join or split of train $t \in OT$	
	$\underline{bp}_{t,t'}$	minimum separation time between the arrival of t and the departure of t' , provided that $t' \in T_t^{bal}$, i.e., $t < t'$ (balance period)	
	Maintenance	\mathcal{A}	set of MAs
		$\underline{ab}_a, \underline{ae}_a$	begin and end time of MA $a \in \mathcal{A}$
\mathcal{V}_a^{dir}		set of TDSs directly affected by MA $a \in \mathcal{A}$	
\mathcal{V}_a^{tsl}		set of TSL affected TDSs during MA $a \in \mathcal{A}$	
All trains	MT_a^{inb}, MT_a^{outb}	set of inbound and outbound MTs required for MA $a \in \mathcal{A}$	
	T	set of all trains ($T = OT \cup \bigcup_{a \in \mathcal{A}} (MT_a^{inb} \cup MT_a^{outb})$)	
	R_t	set of routes available for train $t \in T$	
	$\mathcal{V}_{t,r,v}^{occ}$	TDSs occupied by train $t \in T$ along route $r \in R_t$ when the head of t is at the end of TDS v (\emptyset if the rolling stock used for t is shorter than v)	
	$\mathcal{V}_{t,c}^{cop}$	set of TDSs belonging to control point $c \in C_t$ where train $t \in OT$ has a scheduled stop	
	$\underline{nr}_{t,r,v}, \underline{rr}_{t,r,v}$	nominal and restricted running time of train $t \in T$ along route $r \in R_t$ on TDS $v \in \mathcal{V}_r$;	
	$\underline{nc}_{t,r,v}, \underline{rc}_{t,r,v}$	nominal and restricted clearing time of train $t \in T$ along route $r \in R_t$ on TDS $v \in \mathcal{V}_r$	
M	large parametrisable constant		

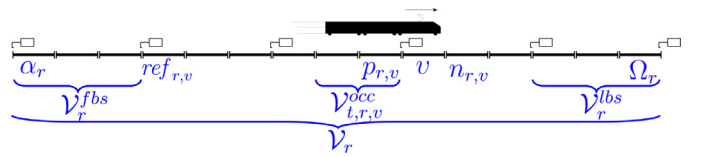


Fig. 5. Main data concerning route r of train t circulating in an infrastructure equipped with a signalling system of 3 aspects.

$l_{t,r,v}$: extra time of train t 's head on TDS v along route r : additional dwell time or delay;

$art_{t,r,v}$: actual running time of train t along route r on TDS v taking into account the possibly imposed TSL and extra time;

$u_{t,r,v}$: utilization time of TDS v by train t along route r ;

for all pairs of $t \in T$ and $v \in \bigcup_{r \in R_t} \mathcal{V}_r$:

$u_{t,v}^{beg}, u_{t,v}^{end}$: time at which TDS v starts and ends being utilized by train t ;

for all pairs of $t \in OT$ and $c \in C_t$:

$d_{t,c}$: delay suffered by train t arriving at control point c ;

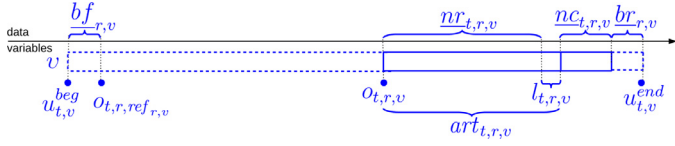


Fig. 6. Main data and variables concerning the utilization of TDS ν belonging to route r of train t when the nominal running and clearing times are used. The dashed rectangles represent reservation and release times. The solid one depicts the running and clearing times, which added up represent the occupation time.

Fig. 6 displays a graphical representation of the main data and variables concerning the utilization of TDS ν belonging to route r of train t . Data are reported above the represented track, while variables are below.

Binary variables:

for all pairs of $t \in T$ and $r \in R_t$:

$$x_{t,r} = \begin{cases} 1 & \text{if train } t \text{ uses route } r, \\ 0 & \text{otherwise.} \end{cases}$$

for all triplets of $t, t' \in T$ such as $index(t) < index(t')$ and $\nu \in \bigcup_{r \in R_t} \mathcal{V}_r \cap \bigcup_{r \in R_{t'}} \mathcal{V}_r$:

$$y_{t,t',\nu} = \begin{cases} 1 & \text{if train } t \text{ utilizes } \nu \text{ before train } t' \text{ (} t < t' \text{),} \\ 0 & \text{otherwise,} \end{cases}$$

for all triplets of $t \in T, a \in \mathcal{A}$ and $\nu \in \bigcup_{r \in R_t} \mathcal{V}_r \cap \mathcal{V}_a^{dir}$:

$$z_{t,a,\nu} = \begin{cases} 1 & \text{if train } t \text{ starts utilizing TDS } \nu \text{ before} \\ & \text{the start of MA } a \text{ (} t < a \text{),} \\ 0 & \text{otherwise.} \end{cases}$$

To cope with the TSL imposed by a MA a , three sets of binary variables are defined for all $\nu \in \mathcal{V}_a^{sl}$. First, variable $f_{t,a,\nu}$ is defined to establish if a train $t \in T$ uses ν before the start of a , and hence before the start of the speed limitation. Although the definition of $z_{t,a,\nu}$ variables is similar to the one of $f_{t,a,\nu}$ variables, remark that the difference is that the involved TDS ν belongs to the set of directly affected TDSs (\mathcal{V}_a^{tds}) for the former and to the set of TSL affected TDSs (\mathcal{V}_a^{tsl}) for the latter. Then, variable $g_{t,a,\nu}$ is defined to identify the case in which t uses ν while a is being performed and the speed limitation is active. Finally, variable $m_{t,\nu}$ is set to determine if t runs through ν during a TSL imposition, i.e., if at least one g variable is set to one. Recall that ν may be affected by TSLs imposed by several MAs, as shown in the example in Fig. 3. If a TSL imposition exists, the restricted running time must be used to calculate the actual running time ($art_{t,r,v}$), otherwise, the nominal running time is used.

For all triplets of $t \in T, a \in \mathcal{A}$ and $\nu \in \bigcup_{r \in R_t} \mathcal{V}_r \cap \mathcal{V}_a^{tsl}$:

$$f_{t,a,\nu} = \begin{cases} 1 & \text{if train } t \text{ starts utilizing TDS } \nu \text{ before the start} \\ & \text{of MA } a \text{ (} t < a \text{),} \\ 0 & \text{otherwise,} \end{cases}$$

$$g_{t,a,\nu} = \begin{cases} 1 & \text{if train } t \text{ uses TDS } \nu \text{ during the realization of MA } a, \\ 0 & \text{otherwise,} \end{cases}$$

for all pairs of $t \in T$ and $\nu \in \bigcup_{r \in R_t} \mathcal{V}_r$:

$$m_{t,\nu} = \begin{cases} 1 & \text{if train } t \text{ runs through } \nu \text{ during the imposition} \\ & \text{of a TSL,} \\ 0 & \text{otherwise.} \end{cases}$$

5.1.3. Objective function

The objective function minimizes the weighted sum of arrival delays at the train's destination and intermediate stops, with respect to the original timetable (1).

$$\min \sum_{t \in OT, c \in C_t} d_{t,c} w_{t,c}. \quad (1)$$

5.1.4. Constraints

Time related constraints

A train t cannot be operated before its earliest operable time \underline{bt}_t (2). The start time of a TDS occupation by train t along a route

r is zero if the route itself is not used (3).

$$o_{t,r,\nu} \geq \underline{bt}_t x_{t,r} \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r. \quad (2)$$

$$o_{t,r,\nu} \leq Mx_{t,r} \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r. \quad (3)$$

A train t starts occupying a TDS ν along a route r after spending in the preceding TDS $p_{r,\nu}$ the corresponding actual running time $art_{t,r,p_{r,\nu}}$ (4). The actual running time of t over a TDS ν , is equal to at least the sum of the extra time of t on ν and its running time, depending on the running time mode to be used on ν (5)–(8). Here, if $m_{t,\nu} = 1$, Constraints (5) and (6) will be active and the restricted running time will be imposed. Otherwise, Constraints (7) and (8) will be active and the nominal running time will be set.

$$o_{t,r,\nu} = o_{t,r,p_{r,\nu}} + art_{t,r,p_{r,\nu}} \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r. \quad (4)$$

$$art_{t,r,\nu} \geq l_{t,r,\nu} + \underline{rr}_{t,r,\nu} x_{t,r} - M(1 - m_{t,\nu}) \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r. \quad (5)$$

$$art_{t,r,\nu} \leq l_{t,r,\nu} + \underline{rr}_{t,r,\nu} x_{t,r} + M(1 - m_{t,\nu}) \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r. \quad (6)$$

$$art_{t,r,\nu} \geq l_{t,r,\nu} + \underline{nr}_{t,r,\nu} x_{t,r} - Mm_{t,\nu} \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r. \quad (7)$$

$$art_{t,r,\nu} \leq l_{t,r,\nu} + \underline{nr}_{t,r,\nu} x_{t,r} + Mm_{t,\nu} \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r. \quad (8)$$

An OT t cannot depart from a control point $c \in C_t$ before its scheduled departure time $\underline{dt}_{t,c}$ (9). Moreover, t must spend an extra time $l_{t,r,\nu}$ on control point c at least equal to its minimum dwell time period $\underline{dp}_{t,c}$ (10). These constraints are active for ν if a route including it is used by t .

$$o_{t,r,n_{r,\nu}} \geq \underline{dt}_{t,c} x_{t,r} \quad \forall t \in OT, c \in C_t, r \in R_t, \nu \in \mathcal{V}_{t,c}^{cop} \cap \mathcal{V}_r \setminus \{\Omega_r\}. \quad (9)$$

$$l_{t,r,\nu} \geq \underline{dp}_{t,c} x_{t,r} \quad \forall t \in OT, c \in C_t, r \in R_t, \nu \in \mathcal{V}_{t,c}^{cop} \cap \mathcal{V}_r. \quad (10)$$

For two OTs t and t' using the same rolling stock, with $t \in T_{t'}^{bal}$ ($t' < t$), at least a balance time period $\underline{bp}_{t',t}$ must exist between the arrival at destination of t' and the departure of t from its origin (11).

$$\sum_{r \in R_t} o_{t,r,\alpha_r} \geq \sum_{r' \in R_{t'}} (o_{t',r',\Omega_{r'}} + art_{t',r',\Omega_{r'}} + \underline{bp}_{t',t} x_{t',r'}) \quad \forall t, t' \in OT, t \in T_{t'}^{bal}. \quad (11)$$

Delay related constraints

As defined in Section 4, an OT t must circulate respecting the green wave policy, therefore, extra times can only be applied in TDSs belonging to t 's control points (12). Moreover, when a route is not selected, there must be no extra times assigned to any TDS belonging to said route (13).

$$l_{t,r,\nu} = 0 \quad \forall t \in OT, r \in R_t, \nu \in \mathcal{V}_r \setminus \bigcup_{c \in C_t} \mathcal{V}_{t,c}^{cop}. \quad (12)$$

$$l_{t,r,\nu} \leq Mx_{t,r} \quad \forall t \in T, r \in R_t, \nu \in \mathcal{V}_r \cap \bigcup_{c \in C_t} \mathcal{V}_{t,c}^{cop}. \quad (13)$$

An OT t having a set of control points C_t , suffers a delay $d_{t,c}$ at least equal to the difference between the actual and the scheduled arrival times at said control point (14). Remark that this implies

that late arrivals are penalized in the objective function, but early arrivals are not. This is a choice we made after interviewing some experts of SNCF Réseau. If required, this may be easily changed to account for early arrivals as well.

$$d_{t,c} \geq \sum_{\substack{r \in R_t: \\ v \in \mathcal{V}_r \cap \mathcal{V}_{t,c}^{sop}}} (o_{t,r,v} + art_{t,r,v} - l_{t,r,v}) - \underline{at}_{t,c} \quad \forall t \in OT, c \in C_t. \quad (14)$$

Maintenance works related Constraints

Inbound MT $t \in MT_a^{inb}$ must arrive at the location of MA a when a itself starts (15). Similarly, outbound MT $t \in MT_a^{outb}$ departs from the location of MA a when a itself finishes (16).

$$\sum_{r \in R_t} o_{t,r,\Omega_r} + art_{t,r,\Omega_r} = \underline{ab}_a \quad \forall a \in \mathcal{A}, t \in MT_a^{inb}. \quad (15)$$

$$\sum_{r \in R_t} o_{t,r,\alpha_r} = \underline{ae}_a \quad \forall a \in \mathcal{A}, t \in MT_a^{outb}. \quad (16)$$

Capacity related constraints

Exactly one route must be used for each train (17). The location coherence between two trains using the same rolling stock must be ensured (18). More precisely, two trains t and t' , where $t \in T_t^{bal}$ ($t' < t$), must use a route $r \in R_t$ and $r' \in R_{t'}$, respectively, such that the last TDS of r' ($\Omega_{r'}$) corresponds to the first TDS of r (α_r).

$$\sum_{r \in R_t} x_{t,r} = 1 \quad \forall t \in T. \quad (17)$$

$$\sum_{\substack{r' \in R_{t'}: \\ \Omega_{r'} = \alpha_r}} x_{t',r'} = \sum_{\substack{r \in R_t: \\ \alpha_r = \Omega_{r'}}} x_{t,r} \quad \forall t, t' \in OT, t \in T_t^{bal}. \quad (18)$$

A train t 's utilization of a TDS v starts as soon as t starts occupying its reference TDS $ref_{r,v}$ along the selected route r , minus the formation time, provided that t does not use the same rolling stock of another train, or v does not belong to the first block section of r (19). Otherwise, the utilization of v starts before that moment. Indeed, the concerned TDS must remain utilized while a turnaround, split or join operation takes place, i.e., a rolling stock balance operation. In such cases, Constraints (19) are defined with a lower than or equal to inequality (\leq).

Moreover, the utilization of v lasts until t finishes utilizing it along any $r \in R_t$ plus the release time (20). The utilization time $u_{t,r,v}$, defined in Constraints (21) and (22), includes: First, the actual running time of all TDSs between $ref_{r,v}$ and v . Second, the extra time on all TDSs v' such that $v \in \mathcal{V}_{t,r,v'}^{occ}$, to take into account the extra time spent by t into the forthcoming TDSs in which the presence of t 's head implies that t also occupies v due to its length. And third, the adequate clearing time, corresponding to the running time mode of t over v . Fig. 7 illustrates an example of the utilization time for a TDS.

$$u_{t,v}^{beg} = \sum_{\substack{r \in R_t: \\ v \in \mathcal{V}_r}} (o_{t,r,ref_{r,v}} - \underline{bf}_{r,v} x_{t,r}) \quad \forall t \in T, v \in \bigcup_{r \in R_t} \mathcal{V}_r: \\ (\nexists t' \in T : t' \in T_t^{bal}) \vee (ref_{r,v} \neq \alpha_r). \quad (19)$$

$$u_{t,v}^{end} = \sum_{\substack{r \in R_t: \\ v \in \mathcal{V}_r}} (o_{t,r,ref_{r,v}} + \underline{br}_{r,v} x_{t,r} + u_{t,r,v}) \quad \forall t \in T, v \in \bigcup_{r \in R_t} \mathcal{V}_r. \quad (20)$$

$$u_{t,r,v} \geq \sum_{v' \in \mathcal{V}_{t,r,v'}^{bet}} art_{t,r,v'} + \sum_{\substack{v' \in \mathcal{V}_t: \\ v \in \mathcal{V}_{t,r,v'}^{occ}}} l_{t,r,v'} + \underline{rc}_{t,r,v} x_{t,r} - M(1 - m_{t,v}) \\ \forall t \in T, r \in R_t, v \in \mathcal{V}_r. \quad (21)$$

$$u_{t,r,v} \geq \sum_{v' \in \mathcal{V}_{t,r,v'}^{bet}} art_{t,r,v'} + \sum_{\substack{v' \in \mathcal{V}_t: \\ v \in \mathcal{V}_{t,r,v'}^{occ}}} l_{t,r,v'} + \underline{nc}_{t,r,v} x_{t,r} - Mm_{t,v} \\ \forall t \in T, r \in R_t, v \in \mathcal{V}_r. \quad (22)$$

The utilization of a TDS v by two trains t and t' must not overlap in the following two cases: First, if t and t' do not use the same rolling stock. Second, if t and t' use the same rolling stock and v does not belong to an extreme (first or last) block section of any route $r \in R_t$ and $r' \in R_{t'}$ (23), (24).

The utilization of v may overlap if two trains use the same rolling stock provided that v belongs to an extreme block section of any route available for these trains. As explained before, a TDS v concerned with a turnaround, join or split operation between trains must remain utilized during this time. Therefore, the utilization of v by train t , resulting from another train t' , i.e., $t \in T_{t'}^{bal}$ ($t' < t$), must start no later than the end of v 's utilization by t' (25). This relation is established as an inequality because more than one train may be concerned. For example, consider a join operation in TDS v , where trains t' and t'' form train t ($t \in T_{t'}^{bal} \cap T_{t''}^{bal}$) and t' arrives before t'' , in this case, the utilization of v by t starts no later than v 's end of utilization by t' , which is before the end of v 's utilization by t'' .

$$u_{t,v}^{end} - u_{t',v}^{beg} \leq M(1 - y_{t,t',v}) \\ \forall t, t' \in T, idx(t) < idx(t'), v \in \bigcup_{r \in R_t} \mathcal{V}_r \cap \bigcup_{r' \in R_{t'}} \mathcal{V}_{r'}: \\ (t \notin T_{t'}^{bal} \wedge t' \notin T_t^{bal}) \vee \left(v \notin \bigcup_{r \in R_t} \mathcal{V}_r^{fbs} \cup \mathcal{V}_r^{lbs} \wedge v \notin \bigcup_{r' \in R_{t'}} \mathcal{V}_{r'}^{fbs} \cup \mathcal{V}_{r'}^{lbs} \right). \quad (23)$$

$$u_{t',v}^{end} - u_{t,v}^{beg} \leq M y_{t,t',v} \\ \forall t, t' \in T, idx(t) < idx(t'), v \in \bigcup_{r \in R_t} \mathcal{V}_r \cap \bigcup_{r' \in R_{t'}} \mathcal{V}_{r'}: \\ (t \notin T_{t'}^{bal} \wedge t' \notin T_t^{bal}) \vee \left(v \notin \bigcup_{r \in R_t} \mathcal{V}_r^{fbs} \cup \mathcal{V}_r^{lbs} \wedge v \notin \bigcup_{r' \in R_{t'}} \mathcal{V}_{r'}^{fbs} \cup \mathcal{V}_{r'}^{lbs} \right). \quad (24)$$

$$\sum_{r \in R_t} u_{t,v}^{beg} \leq \sum_{\substack{v \in \bigcup_{r' \in R_{t'}} \{\Omega_{r'}\}}} u_{t',v}^{end} \quad \forall t, t' \in T : t \in T_t^{bal}. \quad (25)$$

Similarly to (23) and (24), the TDS utilizations by a train and a MA must not overlap if the TDS is directly affected by the MA (26), (27).

$$u_{t,v}^{end} - M(1 - z_{t,a,v}) \leq \underline{ab}_a \\ \forall a \in \mathcal{A}, t \in T \setminus (MT_a^{inb} \cup MT_a^{outb}), v \in \mathcal{V}_a^{dir} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (26)$$

$$\underline{ae}_a - Mz_{t,a,v} \leq u_{t,v}^{beg} \\ \forall a \in \mathcal{A}, t \in T \setminus (MT_a^{inb} \cup MT_a^{outb}), v \in \mathcal{V}_a^{dir} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (27)$$

Temporary speed limitation related constraints

A train t can circulate at nominal speed over a TDS v , belonging to the TSL affected TDSs of a MA a , only if v 's utilization by t takes place while a is not being performed, i.e., either if the complete

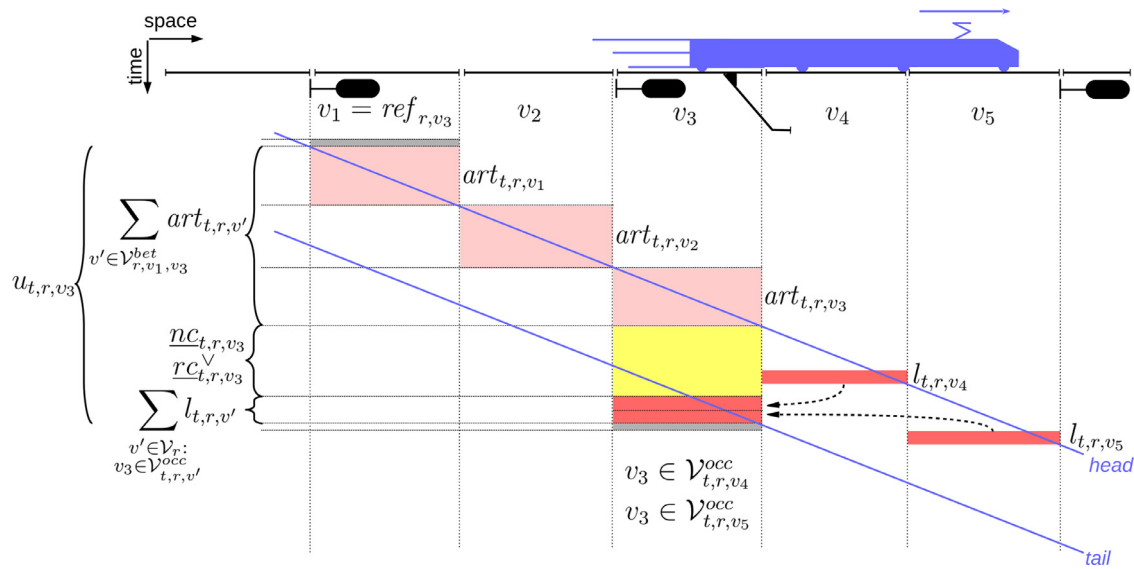


Fig. 7. TDS utilization example: The utilization time of TDS v_3 along route r of train t includes three parts: First, the sum of the actual running times of the TDSs between the rTDS ($ref_{r,v_3} = v_1$) and v_3 . Second, the extra times on v_4 and v_5 because the presence of t 's head in these TDSs implies that t also occupies v_3 due to t 's length, i.e., $v_3 \in \mathcal{V}_{t,r,v_4}^{occ}$ and $v_3 \in \mathcal{V}_{t,r,v_5}^{occ}$. Third, the clearing time, nominal or restricted, of t leaving v_3 .

utilization of v by t takes place before the start of a , or if it takes place after the end of a .

We define Constraints (28) and (29) to identify whether t starts the utilization of v before or after the beginning of a . Next, Constraints (30)–(33) are defined to ensure the proper behaviour of binary variable g . This means that for a train t using a TDS $v \in \bigcup_{r \in R_t} \mathcal{V}_r \cap \mathcal{V}_a^{tsl}$, the value of $g_{t,a,v}$ must be set to 0 only in two cases: First, when the utilization of v by t starts before the beginning of a ($f_{t,a,v} = 1$) and finishes before a starts. Second, when the utilization of v by t starts after the beginning of a ($f_{t,a,v} = 0$) and finishes after the end of a . Otherwise, $g_{t,a,v}$ must be set to 1.

On the one hand, Constraints (30) and (31) are active when $f_{t,a,v} = 1$; then, they ensure that the value of $g_{t,a,v}$ is set to 0 only if $u_{t,v}^{end} < \underline{a}b_a$. On the other hand, Constraints (30) and (31) are active when $f_{t,a,v} = 0$; then, they ensure that the value of $g_{t,a,v}$ is set to 0 only if $\underline{a}e_a < u_{t,v}^{beg}$.

Constraints (34) and (35) ensure that the running time mode of t over v is set to restricted ($m_{t,v} = 1$) if t utilizes v during the performance of one or more MA a that imposes a TSL on v .

$$u_{t,v}^{beg} - \underline{a}b_a \leq M(1 - f_{t,a,v}) \quad \forall t \in T, a \in \mathcal{A} : v \in \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (28)$$

$$\underline{a}b_a - u_{t,v}^{beg} < Mf_{t,a,v} \quad \forall t \in T, a \in \mathcal{A} : v \in \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (29)$$

$$u_{t,v}^{end} - \underline{a}b_a \leq M(1 - f_{t,a,v} + g_{t,a,v}) \quad \forall t \in T, a \in \mathcal{A} : v \in \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (30)$$

$$\underline{a}b_a - u_{t,v}^{end} < M(2 - f_{t,a,v} - g_{t,a,v}) \quad \forall t \in T, a \in \mathcal{A} : v \in \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (31)$$

$$\underline{a}e_a - u_{t,v}^{beg} \leq M(f_{t,a,v} + g_{t,a,v}) \quad \forall t \in T, a \in \mathcal{A} : v \in \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (32)$$

$$u_{t,v}^{beg} - \underline{a}e_a < M(1 + f_{t,a,v} - g_{t,a,v})$$

$$\forall t \in T, a \in \mathcal{A} : v \in \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (33)$$

$$m_{t,v} \geq g_{t,a,v} \quad \forall t \in T, a \in \mathcal{A} : v \in \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (34)$$

$$m_{t,v} \leq \sum_{\substack{a \in \mathcal{A} : \\ v \in \mathcal{V}_a^{tsl}}} g_{t,a,v} \quad \forall t \in T : v \in \bigcup_{a \in \mathcal{A}} \mathcal{V}_a^{tsl} \cap \bigcup_{r \in R_t} \mathcal{V}_r. \quad (35)$$

5.2. Reduction of the number of binary variables

To reduce the number of binary variables y , z , f and g , we apply a boosting technique proposed by Pellegrini et al. (2015), proven to be quite effective in terms of computational time reduction. In particular, we exploit the fact that the topology of a railway infrastructure frequently imposes that precedence relations between trains must be identical on a set of consecutive TDSs. Consider as an example two trains t and t' following each other on a single track segment containing 10 TDSs (v_0 to v_9) without any overtaking area. According to our formulation, for a couple of trains t and t' circulating over this track section, 10 y -variables are defined, one for each TDS (y_{t,t',v_0} to y_{t,t',v_9}). However, if t enters before t' in the first TDS, i.e., $y_{t,t',v_0} = 1$, given that the precedence between t and t' cannot change along the whole single track segment, the value of the remaining y -variables will be 1 as well. Therefore, instead of defining 10 y -variables, we can use only one y_{t,t',v_0-9} . This same principle can be applied as well to z , f and g variables. By doing so, and depending on the infrastructure topology, we can significantly reduce the number of binary variables of RECIFE-MAINT.

6. Solution algorithms

In this section, we present the algorithms that we propose in this paper and that we test in the experiments reported in Section 7. We establish a *whole time limit* for each algorithm. This is the total wall clock time available for an execution. After this time, the best feasible solution (if any) is returned. Note that a solution is considered optimal only if the algorithm completes its optimality proof. Moreover, some of the proposed algorithms are

Algorithm 1: Greedy Algorithm.

Algorithm GreedyAlgorithm()

```
while currTime < TimeLim do
  ord ← generateOrder(T)
  TTord ← constructTimetable(ord)
  objValord ← calc. objVal of TTord
end
bestTT ← TTord' with lowest objValord'
return bestTT
```

Procedure constructTimetable(ord)

```
TT ← empty timetable
foreach t ∈ ord do
  plant ← planTrain(TT, t)
  TT ← add(plant)
end
return TT
```

Procedure planTrain(TT, t)

```
foreach r ∈ Rt do
  schedr ← scheduleRoute(TT, r)
  objContr ← calc. objCont of schedr
end
plant ← r', schedr' with lowest objContr'
return plant
```

Procedure scheduleRoute(TT, r)

```
schedr ← empty schedule
for i ← 1 to |Vr| do
  schedr ← schedule vi ∈ Vr
  if conflict(TT, schedr) then
    i ← solveConflict(TT, schedr)
  end
end
return schedr
```

two-phase algorithms, in these cases, a *first-phase time limit* is also established. More details about the behaviour of the algorithms once the first-phase time limit elapses are given in the respective descriptions.

6.1. RECIFE-MAINT: Full (RM-F)

This algorithm consists in solving the complete formulation described in [Section 5](#) by using a MILP solver.

6.2. RECIFE-MAINT: Battery (RM-B)

This algorithm consists in solving a constrained version of RECIFE-MAINT by using a MILP solver. This constrained formulation is conceived to emulate the train batteries approach. Recall that the batteries approach, as reviewed in [Section 4](#), consists in scheduling train batteries in both directions over a single track segment without considering the microscopic characteristics of the trains' route. As such, RM-B is aimed to serve as a benchmark to compare the performance of the other algorithms with the current practice.

In a pre-processing step of RM-B, we assign a route to each train: First, we verify that the default route does not pass through some MAs' directly affected TDSs. If this is not the case, the default route is assigned to the train. Otherwise, one of the alternative routes of the concerned train is assigned instead. To this end, first, all alternative routes that are directly affected by a MA are discarded, then the route with the most similar TDSs sequence, with respect to the default one, is chosen. If no such route exists, the default route is maintained and assigned to the train, even if this imply that the train will have to wait until the MA is finished for completing its journey.

The constrained version of RECIFE-MAINT we use in this algorithm does not consider alternative routes, i.e., trains can only use their assigned route (either the default one or the new one assigned in the pre-processing step). More precisely, all binary variables related to routing (x) are suppressed from the formulation and the related constraints are updated accordingly.

The so obtained optimal solution can be interpreted as including train batteries, in which the values of the three parameters characterizing a schedule obtained by using the batteries approach ([Section 4.1](#)) are also optimal: the number of train batteries, the number of trains per battery and the direction of the first battery. For this reason we can assert that the optimal solution obtained by

RM-B may also be interpreted as the result of an optimal battery approach.

6.3. RECIFE-MAINT: Battery-Full (RM-BF)

It is a two-phase algorithm where RM-B is used to obtain a feasible solution, then this solution is used to initialize RM-F.

After the first-phase time limit elapses, the best solution obtained by RM-B is used to initialize RM-F. If no feasible solution is found, the execution of the first-phase is extended until a feasible solution is found or the whole time limit expires. However, if before the first-phase time limit elapses the optimality of a solution is proven, the first phase ends and the remaining time is devoted to the second-phase.

6.4. RECIFE-MAINT: Greedy-Full (RM-GF)

RM-GF is a two-phase algorithm in which the first phase uses a heuristic algorithm that allows re-routing to generate an initial solution for RM-F.

The heuristic algorithm in the first phase is based on a greedy approach that plans trains one at the time. Once a train is planned, its route and schedule cannot be changed. [Algorithm 1](#) describes the main structure of this greedy algorithm.

More precisely, the greedy algorithm can be divided into three main activities: timetable construction, train planning and route scheduling.

Timetable construction: A timetable is built by planning trains, one by one, following a predefined *train order*. Once a train is processed, its *plan* (route and schedule) cannot be modified. Once all trains are planned in the timetable, the objective function value corresponding to the resulting timetable is calculated.

Train planning: The best feasible schedule attainable with each route of the train is computed, and the corresponding objective function contribution value (*objCont*) is calculated. Then, the route with the lowest *objCont* is assigned to the train along with its respective schedule.

Route scheduling: Given a route, the schedule is generated in a sequential way, i.e., the utilisation of each TDS in the route is scheduled progressively, from the first to the last one. When a conflict arises, i.e., superposition of a TDS's utilization time with a MA or another previously planned train, the *solveConflict* function adds an extra time to the closest previous TDS where the train is allowed to experience delay, then, the scheduling is resumed starting

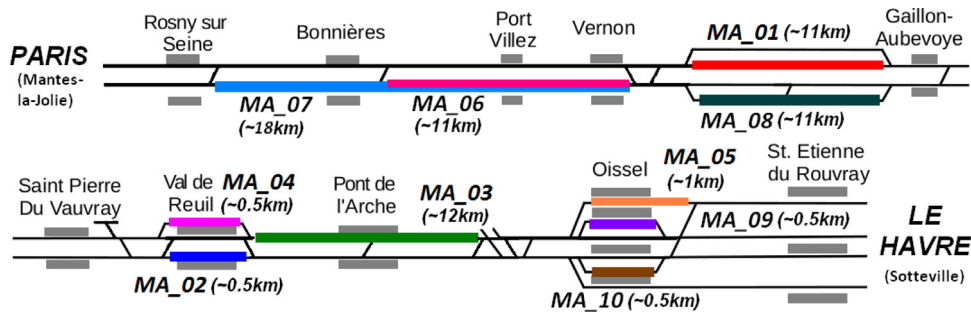


Fig. 8. Section of the Paris - Le Havre line containing MAs.

from said TDS. The scheduling process is finished once the utilisation of the last TDS of the route is set.

Note that the train order may have a high impact on the objective function value of the generated timetable. Indeed, some trains, when planned early in the process, can cause heavy delays to the trains planned later. To mitigate the effect of the train order, we include in the greedy algorithm a mechanism to explore several possibilities. A timetable is generated with each train order, and the one with the lowest objective function value is chosen as the solution of the greedy algorithm. The number of train orders explored depends on the first-phase time limit. Indeed, as long as the first-phase lasts, new random train orders are automatically generated. However, we also define a set of train orders that are considered by the algorithm before any random order. The train orders we propose are described below in the same sequence as they are considered by the algorithm:

- Trains with longer routes are planned first: These trains potentially have more conflicts with the rest because of the length of their routes. Hence, the total delay experienced by these trains might be reduced if planned first.
- Trains with shorter routes are planned first: Ideally, by planning these trains first, the number of conflicts among them will remain low. Hence, probably a significant number of trains are planned without any delay before trains with longer routes are planned.
- Trains with fewer stops are planned last: Due to the green wave policy that must be respected by OTs, these trains have a limited number of locations where an extra time may be introduced. Moreover, the objective function sums the accumulated delay at every stop of each train. Hence, trains with fewer stops potentially contribute a lower value to the objective function.
- Trains with more stops are planned last: These trains have more possible locations where an extra time may be introduced. Therefore, these trains are more flexible for rescheduling. Ideally, the extra times are only applied, if necessary, to its latest stops before the conflict.

Two additional considerations are applied when generating a new train order: First, MTs are always planned before any other train. This is done because of the hard timing constraints related to MTs, i.e., MTs must arrive at, and depart from, the MA location at the exact same time as the MA starts and ends, respectively (Constraints (15) and (16) in Section 5). Second, trains with rolling stock re-utilisation are planned right after MTs. This is done to easily avoid deadlock situations.

Typically, several orders are considered during the first-phase time interval, but if the first-phase time limit elapses before the first timetable is obtained, the execution of this phase is extended until a solution is generated or the whole time limit expires.

7. Experiments and results

We set up two rounds of experiments using the case study described in Section 7.1. The objectives of the first round of experiments are: Initially, to evaluate the capacity of RECIFE-MAINT and the proposed algorithms to obtain solutions to instances of the problem with different characteristics. Then, to assess the solutions improvement with respect to the solutions applied in current practice, i.e., the solutions obtained by RM-B. Finally, to obtain insights about the features of difficult instances and the capability of the algorithms to deal with them. The second round is designed to test the performance of the algorithms when dealing with very large instances.

The experiments are performed in a computer with eight Intel Xeon 3.5 Ghz processors and 128GB RAM. The MILP solver used to solve RECIFE-MAINT is IBM CPLEX MILP solver v 12.6. The whole time limit is set to 1 h for all algorithms. The first-phase time limit is set to 15 min for the two-phase algorithms (RM-BF and RM-GF). These time limits are the same for both rounds of experiments.

7.1. Case study

We model a section of the Paris - Le Havre line which is used by mixed traffic: intercity, regional, high speed and freight trains. The control zone, represented in Fig. 8, covers a distance of approximately 70 km, between the stations of Rosny sur Seine and St. Etienne du Rouvray. We acknowledge the fact that the choices made by our algorithms may have an impact beyond the considered control zone, i.e., in other parts of the network. Indeed, if an algorithm delays a train departure, this may imply a delay of the train itself up to its destination, which the algorithm ignores if out of the control zone. Depending on the choices made, this impact may be substantial or non-existent. A way to account for such impacts may be to iterate macroscopic decisions over the whole network and microscopic ones in each control zone composing the network itself, which is, however, out of the scope of this paper.

The timetable of a weekday in this control zone contains around 220 trains. No information is available on the priority of these trains. Hence, we consider weights equal to 1 for all trains at all control points. We define a set of ten unplanned MAs to be performed (one per instance) by mimicking some MAs actually performed on this line in 2012. The locations of these MAs are shown in Fig. 8. When a MA is performed, all the adjacent tracks are subject to a TSL of 40 km/h, the maximum speed allowed in this line is 160 km/h.

Each MA requires the presence of one MT to be performed. The MTs must traverse a route connecting the location of the concerned MA and a shunting yard. We consider two shunting yards, one at Mantes-la-Jolie, between Rosny sur Seine and Paris, and one at Sotteville, between St. Etienne du Rouvray and Le Havre. For MA_01, MA_03, MA_04, MA_05 and MA_09, the inbound MTs depart from Mantes-la-Jolie and the outbound MTs arrive to Sot-

Table 2

Characteristics of the instances: First round of experiments. (Ins.: Number of instances, Tra.: Median number of trains, DIRiT: Median number of directly impacted trains, TSLiT: Median number of TSL impacted trains, Cont.Vars: Median number of continuous variables, Bin.Vars: Median number of binary variables, Constrs: Median number of constraints.)

	Ins.	Tra.	DIRiT	TSLiT	Cont.Vars	Bin.Vars	Constrs
all	300	48	5	5	158 K	6 K	344 K
MA_01	30	44	6	6	153 K	5 K	330 K
MA_02	30	48	6	2	162 K	6 K	346 K
MA_03	30	48	8	10	157 K	6 K	353 K
MA_04	30	50	6	1	161 K	6 K	341 K
MA_05	30	46	6	4	153 K	5 K	322 K
MA_06	30	45	6	8	155 K	5 K	352 K
MA_07	30	47	7	8	159 K	6 K	376 K
MA_08	30	48	0	6	161 K	6 K	349 K
MA_09	30	48	3	8	162 K	6 K	345 K
MA_10	30	47	3	5	157 K	5 K	325 K

teville. For the other MAs, the inbound MTs depart from Sotteville and the outbound MTs arrive to Mantes-la-Jolie.

7.2. First round of experiments

7.2.1. Instances

Each instance is defined by a combination of: a MA location, a MA duration and a time horizon. The ten MA locations are described in Section 7.1 and shown in Fig. 8. Three *durations* are considered for each MA location: 60, 90 and 120 minutes. We consider time horizons of four hours, starting at randomly selected times. A *time horizon* determines the set of trains present in the corresponding instance, i.e., all trains that enter the control zone between the start and end times of the time horizon. The MA is planned in the middle of the time horizon. Ten different time horizons are randomly drawn for each couple MA location - MA duration, i.e., we tackle 300 instances in total.

Table 2 summarizes the main characteristics of the tackled instances. The first part concerns all 300 instances while the second part classes the instances depending on the MA location. We report the number of instances, the median number of trains present, the median number of trains impacted by the MA, the median number of continuous and binary variables and the median number of constraints in RECIFE-MAINT. Indeed, there are two ways in which MAs can impact trains in the existing timetable:

- Directly impacted trains (DIRiT): Trains in this category have their default route directly affected by a MA, i.e., the trains are

scheduled to use a track segment while the MA is performed on the track segment itself. Trains in this category need to be either re-routed or delayed until the MA finishes.

- TSL impacted trains (TSLiT): Trains in this category have their default route directly affected by a TSL due to a MA, i.e., the trains are scheduled to use a track segment during a TSL period due to a MA.

7.2.2. Results

The results of the first round of experiments are reported in Table 3. The same instance classifications as described for Table 2 are used: all instances and instances classed by MA location. This table reports the results in terms of four performance indicators: % Feasible Sol.: reports the percentage of instances where a feasible solution is obtained by the algorithms. % Optimal Sol.: reports the percentage of instances where the optimal solution is found and proven by the algorithms. Sol. Improv.^a (%) and Sol. Improv.^b (%): These columns report the mean percentage of improvement to the objective function value with respect to RM-B, for instances where all algorithms find a feasible solution within the given time limit and for instances where RM-F does not manage to do so, respectively. Note that RM-B obtains a feasible solution for all instances in 672 s in average. We indicate in bold the best performance across the algorithms according to the solution improvement indicator.

First of all, we remark that both RM-BF and RM-GF obtain at least a feasible solution for all instances within the whole time limit fixed. Instead, RM-F fails to produce a feasible solution for 13% of the instances. More specifically, as shown in the lower part of the table, RM-F has more difficulty to deal with MA_03 and MA_07 instances, where a feasible solution is found for only 50% of the instances. Even with a higher percentage of success, MA_06 and MA_08 also appear to be very problematic to RM-F. In general, the number of optimal solutions proven by the three algorithms is similar, although a slightly better performance is reported for RM-BF and RM-GF (68% of all instances versus the 61% of RM-F).

Concerning the performance of the algorithms in terms of solution improvement with respect to RM-B, we compare separately the instances where all three algorithms find at least a feasible solution and those instances where RM-F does not. Recall that RM-B is considered as a reference as it is a better performing proxy of the current practice, as explained in Section 6.2.

Consider first the 87% of instances where all three algorithms find at least a feasible solution. The solutions returned by RM-BF and RM-GF are better than the ones found by RM-B, in mean of

Table 3

Results of the first round of experiments using RM-F, RM-BF, and RM-GF. (% Feasible Sol.: Percentage of instances where at least a feasible solution is obtained, % Optimal Sol.: Percentage of instances where the optimal solution is found and proven, Sol. Imp. (%): Mean percentage of improvement to the objective function value with respect to the solution obtained by RM-B, ^a for instances where all algorithms find a feasible solution, ^b for instances where RM-F does not find a feasible solution.) Bold indicates the best performance across the algorithms according to the solution improvement indicator.

Class	% Feasible Sol.			% Optimal Sol.			Sol. Imp. ^a (%)			Sol. Imp. ^b (%)	
	RM-F	RM-BF	RM-GF	RM-F	RM-BF	RM-GF	RM-F	RM-BF	RM-GF	RM-BF	RM-GF
all	87	100	100	61	68	68	-46	28	40	-11	34
MA_01	100	100	100	50	60	57	-4	16	16	-	-
MA_02	100	100	100	97	100	100	91	91	91	-	-
MA_03	50	100	100	17	20	20	-48	12	27	-6	25
MA_04	100	100	100	73	97	93	44	49	49	-	-
MA_05	100	100	100	97	97	100	69	69	69	-	-
MA_06	77	100	100	17	13	23	-39	22	44	-14	32
MA_07	50	100	100	3	7	10	18	25	36	-14	42
MA_08	97	100	100	67	83	80	-990	2	1	0	0
MA_09	100	100	100	87	100	100	53	58	58	-	-
MA_10	100	100	100	100	100	100	59	59	59	-	-

28% and 40%, respectively. Instead, the solutions obtained by RM-F are in mean 46% worse than the ones produced by RM-B.

By closely analysing the results of RM-F, we observe that this algorithm is systematically outperformed by both RM-BF and RM-GF, except for those instances where all three algorithms find the optimal solution, either proving the optimality or not. This happens in 179 instances. The performance difference is especially marked in MA_03, MA_06 and MA_08 instances, where RM-F obtains solutions considerably worse than RM-B. Note that for MA_08 instances, the mean value of solution improvement is particularly low (−990%). From a detailed exam, it emerges that this is due to one specific instance where RM-F obtains a solution 14,000% worse than the one obtained by RM-B: the first solution has a total arrival time delay of 1,164,560 s, the second of 8299 s, for a MA duration of 120 min. By using the geometric mean instead of the arithmetic one, we observe that the mean value of solution improvement of RM-F for MA_08 instances is −35%, which is indeed closer to the global performance of RM-F. The overall low performance of RM-F with respect to the two-phased algorithms (RM-BF and RM-GF) shows that, as expected, it is very useful to provide an initial solution to the MILP solver to obtain satisfactory results in a reasonable amount of time.

By observing the differences in performance between RM-BF and RM-GF, we remark that for most instances groups both algorithms have a very similar mean improvement value. Yet, for MA_03, MA_06 and MA_07 instances, the solutions obtained by RM-GF are better than the ones obtained by RM-BF. Indeed, we observe that the overall solution improvement of RM-GF is in mean 12% better than the one of RM-BF.

Consider now the 13% of instances where RM-F cannot find a solution within the time limit. RM-GF again outperforms RM-BF. This time the difference between the improvements over RM-B achieved by these algorithms is in mean 45%. Furthermore, RM-BF obtains, in mean, worse solutions than the ones returned by RM-B. This is due to the existence of a first-phase time limit. Recall that RM-BF uses the best solution obtained by RM-B within the first-phase time limit (15 min) to initialize the MILP solver for the complete formulation. Therefore, whenever RM-BF is outperformed by RM-B is because the second phase of RM-BF starts with a sub optimal solution of RM-B. In the 27 instances (9%) in which this is the case, RM-B uses the remaining time more efficiently than RM-BF, which suffers for the very large search space.

We perform several Wilcoxon rank-sum tests. We obtain that for instances where all three algorithms find a feasible solution, the solutions of all algorithms are significantly better than those obtained by RM-B with a confidence level of 0.95. Moreover, we can also state that, for all instances, the solution improvements made by RM-GF are significantly higher than the improvements made by RM-BF with a confidence level of 0.95. Hence, we conclude that the proposed two-phase algorithms are indeed applicable, since they always provide solutions to the real case study. Moreover, we can state that they are well performing, since they outperform the optimized version of the approach used in the current practice.

By analysing the number of impacted trains, i.e., DIRiTs and TSLiTs, and comparing these with the GAP of the solutions produced by RM-GF, our best performing algorithm, we can obtain some insights on the characteristics of the instances which may be more difficult to solve than others. The GAP indicates the percentage difference between the best feasible solution found by the algorithm within a run and the best bound identified. The higher the GAP, the further the solution may be from the optimum. We consider the GAP as an indicator of the difficulty of an instance for the algorithm: if the GAP is equal to 0 (the optimality is proven), we consider the instance easier than if a positive GAP is returned. The higher the GAP, the more difficult the instance. Fig. 9 presents

four boxplots showing, for all instances classed by MA location, the distributions of: The number of DIRiTs (Fig. 9(a)), the number of TSLiTs (Fig. 9(b)), the product of DIRiTs and TSLiTs (Fig. 9(c)) and the GAP of solutions obtained by RM-GF (Fig. 9(d)). Each box represents the distribution of the observations corresponding to the 30 instances of each MA location. The horizontal line within the box represents the median of the distribution, while the extremes of the box represent the first and third quartiles, respectively; the whiskers show the smallest and the largest non-outliers in the data-set and dots correspond to the outliers. Note that the medians of the distributions in Fig. 9(a) and (b) are also reported in Table 2.

By observing the distributions of DIRiTs in Fig. 9(a), we remark that MA_01 to MA_07 instances have a similar number of DIRiTs, with a median between 6 and 9, whereas MA_08 to MA_10 instances have a lower count of DIRiTs. By comparing these distributions with the distributions of the GAPs obtained by RM-GF (Fig. 9(d)) it is not evident to establish a clear relation between the number of DIRiTs and the difficulty of the instances. For example, we observe that MA_01 and MA_05 instances have a similar distribution of DIRiTs. However, MA_01 instances are more difficult to solve: for 13 out of 30 MA_01 instances the GAP is strictly greater than the median value 0, whereas the solutions for all MA_05 instances have a GAP of 0.

Although the relation between the TSLiTs count and the difficulty of the instances seems slightly more perceptible, as we can observe by comparing the distributions shown in Fig. 9(b) and (d), this relation is not always clear. For example, MA_06 and MA_09 instances have a similar distribution of TSLiTs, but MA_09 instances are significantly easier, as 100% of the solutions are proven to be optimal.

When considered independently, neither DIRiTs nor TSLiTs are able to clearly express the difficulty of an instance. However, when coupled by multiplying their values, a more apparent relation emerges. Fig. 9(c) shows the distribution corresponding to this product. We observe that the distributions with higher product of DIRiTs and TSLiTs indeed correspond to the most difficult groups of instances shown in Fig. 9(d): MA_03, MA_07, MA_06 and MA_01, in descending order. This is corroborated by the high value of the correlation coefficient of 0.78, which gives the quality of a least squares fitting to the product and GAP distributions (Edwards, 1976). The correlation coefficient computed for DIRiTs and TSLiTs is 0.47 and 0.62, respectively.

7.3. Second round of experiments

7.3.1. Instances

In the second round of experiments, the size of the time horizon is set to eight hours. The duration of MAs is set to four hours, which corresponds to one of the largest MA durations commonly allowed during daytime operations on the control zone under study. Five time horizons are randomly set for each MA location, hence, we tackle a total of 50 instances.

Table 4 reports the main characteristics of these large instances. The first part of the table concerns all 50 instances while the second part classes the instances depending on the MA location. As we did in Table 2 for the first round of experiments, for each instance class, we report the number of instances, the median number of trains present, the median number of trains impacted by the MA, the median number of continuous and binary variables and the median number of constraints in RECIFE-MAINT.

7.3.2. Results

The results are summarized in Table 5 in the same form as in Table 3. Note that we do not report the results for RM-F. The reason for this is that RM-F is able to find a feasible solution for only 20% of the instances (10 out of 50) within the whole time

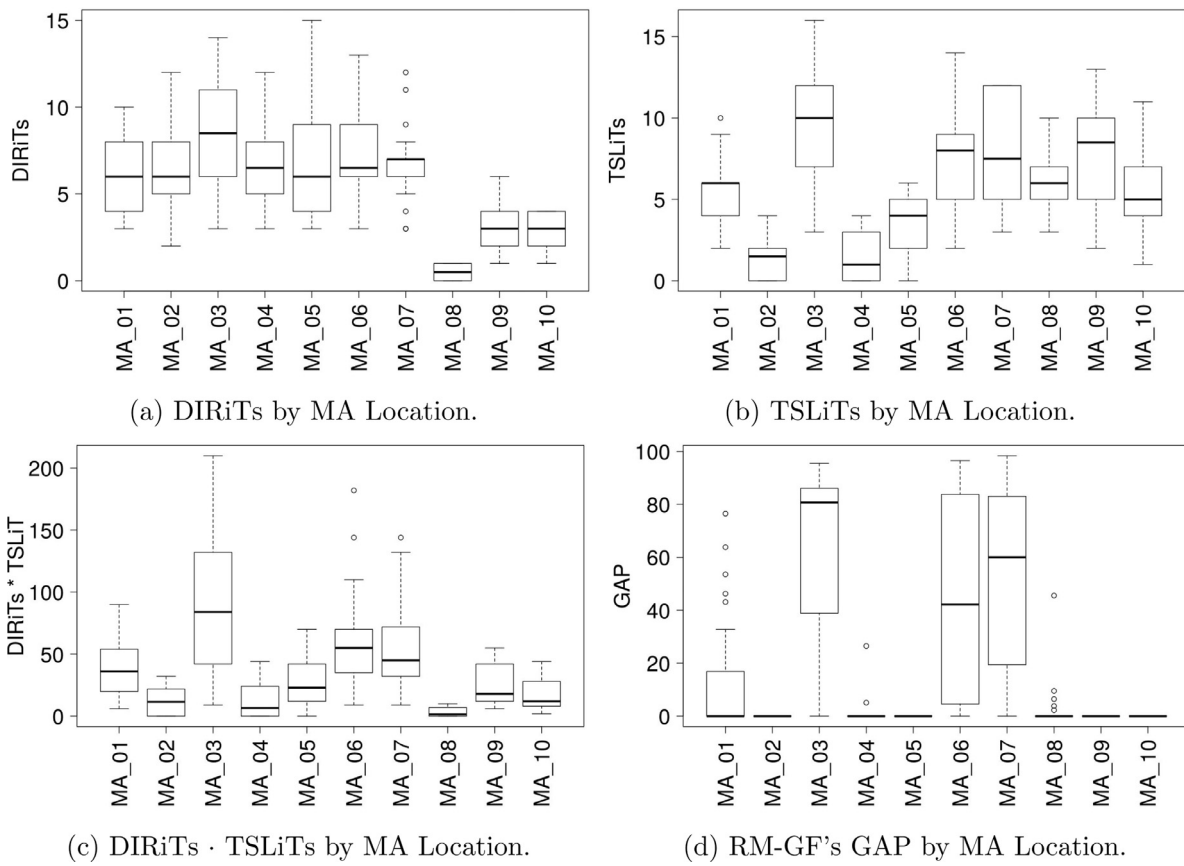


Fig. 9. The distribution of the number of DIRiTs (a), the number of TSLiTs (b), the product of DIRiTs and TSLiTs (c) and the GAP of solutions obtained by RM-GF (d), for all instances classed by MA location.

Table 4

Characteristics of the Instances: Second round of experiments. (Ins.: Number of instances, Tra.: Median number of trains, DIRiT: Median number of directly impacted trains, TSLiT: Median number of TSL impacted trains, Cont.Vars: Median number of continuous variables, Bin.Vars: Median number of binary variables, Constrs: Median number of constraints.).

	Ins.	Tra.	DIRiT	TSLiT	Cont.Vars	Bin.Vars	Constrs
all	50	98	15	14	324 K	21 K	844 K
MA_01	5	102	18	18	326 K	22 K	872 K
MA_02	5	101	15	2	324 K	21 K	835 K
MA_03	5	98	20	24	332 K	22 K	899 K
MA_04	5	99	16	3	336 K	21 K	851 K
MA_05	5	103	17	9	351 K	22 K	892 K
MA_06	5	95	15	15	310 K	18 K	819 K
MA_07	5	89	15	18	321 K	20 K	894 K
MA_08	5	88	1	12	310 K	20 K	804 K
MA_09	5	98	8	18	314 K	21 K	803 K
MA_10	5	94	7	10	313 K	19 K	771 K

Table 5

Results of the second round of experiments using RM-BF, and RM-GF. (% Feasible Sol.: Percentage of instances where at least a feasible solution is obtained, % Optimal Sol.: Percentage of instances where the optimal solution is found and proven, Sol. Imp. (%): Mean percentage of improvement to the objective function value with respect to the solution obtained by RM-B.) Bold indicates the best performance across the algorithms according to the solution improvement indicator.

Class	% Feasible Sol.		% Optimal Sol.		Sol. Imp. (%)	
	RM-BF	RM-GF	RM-BF	RM-GF	RM-BF	RM-GF
all	100	100	12	10	-115	29
MA_01	100	100	0	0	3	-28
MA_02	100	100	0	0	10	89
MA_03	100	100	0	0	-92	27
MA_04	100	100	0	0	9	4
MA_05	100	100	20	20	18	49
MA_06	100	100	0	0	-2	18
MA_07	100	100	0	0	-200	38
MA_08	100	100	0	0	0	-16
MA_09	100	100	40	40	35	16
MA_10	100	100	60	40	68	71

limit. Note that RM-B obtains a feasible solution for all instances in 1099 s in average.

First, we observe that both algorithms, RM-BF and RM-GF, obtain a feasible solution for all instances. Nonetheless, the mean number of optimal solutions proven is considerably reduced with respect to the first round of experiments. Indeed, as the time horizon and MA duration grow larger, the median number of trains, DIRiTs and TSLiTs increase as well, between two and three times with respect to the first round. As postulated in the discussion of the results of the first round of experiments, the product of these is an effective indicator of the difficulty of the instances. This relation is mildly corroborated in this round: we performed a similar

analysis as the one described in Section 7.2.2, obtaining the correlation coefficients of 0.27, 0.48 and 0.51 for DIRiTs, TSLiTs and the product of them, respectively, with regard to the GAP obtained by RM-GF. The highest coefficient again is the one linking the GAP to the product of the two numbers of affected TDSS. The lower values of these correlation coefficients can be explained because of the contribution of another factor that undoubtedly increases the difficulty of the instances, that is, the size of the formulation. Indeed the size of an instance's formulation has an impact in the difficulty of solving the instance itself. By observing the size of the

formulations in the first round of experiments (Table 2) and the percentage of optimal solutions obtained (Table 3) we notice that the size of the formulations are always manageable. Hence, the impact of the different indicators is quite evident in terms of the correlation coefficient. Instead, as shown in Tables 4 and 5, the size of the instances' formulations in the second round of experiments is almost tripled and the percentage of optimal solutions is considerably lower than in the first round. Here, apparently, the formulations become too large to be managed within the whole-time limit and the solution space is not efficiently explored, disregard of the difficulty due to the affected TDSS.

Concerning the performance of the algorithms, we notice that the overall performance of RM-BF, in terms of solution improvement with respect to RM-B, is greatly reduced in this round. This is particularly true when dealing with the most difficult instances of MA_03, MA_06 and MA_07, which have the highest product of DIRiTs and TSLiTs. Note that no outliers are present among the solutions.

Despite the increased size and difficulty of the instances in this round, the mean percentage of solution improvement obtained by RM-GF is 29%, which remains close to the one obtained in the first round. Moreover, RM-GF outperforms RM-BF in this round as well, the difference being in mean 144%. This outcome is confirmed by the Wilcoxon rank-sum test, in which we observe that the results obtained by RM-GF are significantly better than the results of both RM-B and RM-BF with a confidence level of 0.95. However, we remark that for MA_01 and MA_08 instances, the mean percentage improvement of RM-GF with respect to RM-B is negative.

By comparing the mean solution improvement obtained by the algorithms for MA_01 and MA_03 instances, we observe that RM-BF outperforms RM-GF for MA_01 instances but is greatly outperformed for MA_03 instances. Note that the product of the medians of DIRiTs and TSLiTs for MA_01 and MA_03 instances are 324 and 480, respectively, which means that MA_03 instances are, in principle, more difficult to solve than those of MA_01. Indeed, MA_03 instances are more difficult to solve than MA_01 for RM-BF, as the median GAP are 99.8% and 93.3% respectively. However, RM-GF have more difficulties to solve the easier MA_01 instances as for these it obtains a median GAP of 99.8%, compared to 97.8% for those of MA_03. This may mean that RM-GF has difficulties to deal with some particularities of MA_01 instances.

To understand this difficulty we thoroughly examined the MA_01 instances: we analysed their microscopic characteristics and we compared the timetables produced by the greedy component of RM-GF with those obtained by RM-B. By doing so, we identified two main issues, both related to the train orders used by the algorithm. Recall that although several orders are considered by the greedy algorithm (Section 6.4), they all start with the MTs and immediately after, the trains with rolling stock re-utilisation.

The first issue is explained with an example of one particular instance. In this instance, we find that a train with a rolling stock re-utilisation operation at Gaillon Aubevoye station, let it be t , arrives at this station with a delay due to the TSL of MA_01. Another train t' , which is initially scheduled to pass through Gaillon Aubevoye station just before t and has no alternative routes, is planned after t because of the imposed order, therefore t' is delayed for at least the entire duration of the rolling stock re-utilisation operation. The delay imposed to t' in turn causes more delays to other trains planned afterwards, thus originating a timetable with severe overall delay. We discover that when this same instance is solved by RM-B, t is actually delayed elsewhere, allowing t' to pass through Gaillon Aubevoye first, thus avoiding the multiple delays occasioned to other trains. In this specific instance, t' will always be delayed by t because of the order imposition of the greedy algorithm. This same circumstance may occur in more instances with other trains having rolling stock re-utilisation

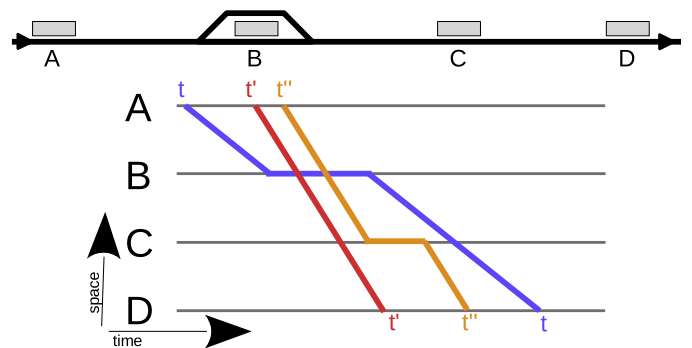


Fig. 10. Example of circular dependency between trains. Three trains t , t' and t'' run between stations A and D. t is overtaken by t' and t'' in station B, the only available overtaking location. First, if the departure of t from A is delayed, it causes a delay on t' . Next, by delaying t' , t'' is delayed as well. Finally, because t must be overtaken by t'' in B, and t'' has a scheduled stop in C, whenever t'' is delayed it causes a delay on t .

or MTs, and cannot be avoided due to the fact that these two types of trains are always planned first in the current version of the greedy algorithm.

The second issue is less evident since it is given by a circular dependency relation between three or more trains that is neither known *a priori* nor easy to identify. Consider the example given in Fig. 10: trains t , t' and t'' present a circular dependency between them such as delaying t causes a delay on t' , delaying t' causes a delay on t'' and delaying t'' causes a delay on t . Consider as well that delaying one of these trains, might also cause delays to other trains outside this circular dependency. Ideally, no delay should be added to the trains within the circular dependency, but if necessary, the best strategy might be to add a small amount of delay to each one of them to maintain the balance. However, this is not how the core of the greedy algorithm works, as one train is planned at each time regardless of the potential conflicts it may cause to the trains planned afterwards. Taking this into account, another strategy could be to identify the most crucial train among those in the circular dependency, e.g., the one that causes more delay to other trains outside the circular dependency, and place it in the train order always before the others. But the current version of the greedy algorithm does not consider train dependencies when generating the train orders, and although random orders are also generated, these do not guarantee that an adequate order is proposed.

These shortcomings clearly affect the overall performance of RM-GF for some groups of instances, and this effect is increased because of the size of the instances in this round of experiments. By assessing the improvements achieved by the second phase of the algorithms for all groups of instances, we discover that for MA_01 instances, the mean improvement achieved by the second phase is only 3% for RM-BF and RM-GF. This evidences the importance of carefully designing the first-phase algorithm, as it can play a critical role in the final outcome of the complete algorithm. However, this does not mean that we can diminish the role of the second phase of the algorithms, since important improvements are obtained, e.g., a mean improvement of 57% is achieved by the second phase of RM-BF for MA_10 instances. The mean solution improvements obtained by the second phase of RM-BF and RM-GF for all instances in this round of experiments are 14% and 19% respectively.

8. Conclusion

In this paper, we presented RECIFE-MAINT: a MILP formulation to solve the problem of rearranging a timetable to cope with main-

tenance activities while considering specific aspects such as maintenance trains and temporary speed limitations.

There is a limited number of approaches proposed in the literature to deal with this problem. All of them are based on macroscopic representations of the infrastructure, which often use time overestimations to guarantee feasibility, thus reducing the available railway capacity. Instead, RECIFE-MAINT uses a microscopic representation of the infrastructure and guarantees the feasibility of the produced timetables while optimizing the railway capacity utilization.

We proposed three algorithms implementing RECIFE-MAINT, which we tested on a real case study in France. Additionally, we considered an algorithm that emulates the current practice, which we used as a benchmark to compare the performance of the algorithms we proposed.

The results showed that our algorithms were able to produce rearranged feasible timetables in a reasonably short amount of time, even when dealing with very large instances. Moreover, the resulting timetables were better, in terms of overall delay reduction, than the timetables obtained by applying the algorithm emulating the current practice. This not only validates the practical applicability of our solution approach but it also shows the level of improvements that can be expected by applying RECIFE-MAINT for solving real-size instances of the problem. Additionally, the analysis of these results allowed us to make conjectures about the characteristics of the instances which impact their difficulty.

In future research we will, first of all, work on the reduction of the processing time needed to solve RECIFE-MAINT. Indeed, the results showed that by using a two-phase algorithm the solution time is significantly reduced. However, the results also showed that the first phase should be carefully designed to deal with specific characteristics of some instances, that may heavily affect its performance. Moreover, a comprehensive study on the potential conflicts between trains may allow a reduction of the number of available routes to consider. This may have a high impact on the reduction of the computational time required to solve large instances.

Our long term perspectives include the extension of our formulation to allow MAS' rescheduling. In our current formulation, train circulations must be rearranged to adapt to a fixed schedule of MAS. By allowing MAS' rescheduling, new possibilities of rearrangement would become available, including the schedule of MAS to a time period in which they would impact traffic as little as possible.

References

- Albrecht, A., Panton, D., Lee, D., 2013. Rescheduling rail networks with maintenance disruptions using problem space search. *Comput. Oper. Res.* 40 (3), 703–712. doi:10.1016/j.cor.2010.09.001.
- Budai, G., Huisman, D., Dekker, R., 2004. Scheduling preventive railway maintenance activities. In: *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, 5, pp. 4171–4176. doi:10.1109/ICSMC.2004.1401185.
- Cacchiani, V., Toth, P., 2012. Nominal and robust train timetabling problems. *Eur. J. Oper. Res.* 219 (3), 727–737. doi:10.1016/j.ejor.2011.11.003. Feature Clusters.
- Caprara, A., Monaci, M., Toth, P., Guida, P.L., 2006. A Lagrangian heuristic algorithm for a real-world train timetabling problem. *Discrete Appl. Math.* 154 (5), 738–753. doi:10.1016/j.dam.2005.05.026. {IV} ALIO/EURO Workshop on Applied Combinatorial Optimization.
- Edwards, A.L., 1976. The correlation coefficient. In: *An Introduction to Linear Regression and Correlation*. W. H. Freeman, pp. 33–46. Chapter 4
- Forsgren, M., Aronsson, M., Gestrelus, S., 2013. Maintaining tracks and traffic flow at the same time. *J. Rail Transp. Plan. Manag.* 3 (3), 111–123. doi:10.1016/j.jrtpm.2013.11.001.
- Hachemane, P., 1997. Évaluation de la capacité de réseaux ferroviaires. *École Polytechnique Fédérale de Lausanne Ph.D. thesis*.
- Higgins, A., 1998. Scheduling of railway track maintenance activities and crews. *Oper. Res. Soc.* 49, 1026–1033.
- Pellegrini, P., Marlière, G., Pesenti, R., Rodriguez, J., 2015. RECIFE-MILP: an effective MILP-based heuristic for the real-time railway traffic management problem. *IEEE Trans. Intell. Transp. Syst.* 16 (5), 2609–2619. doi:10.1109/TITS.2015.2414294.
- Peng, F., Kang, S., Li, X., Ouyang, Y., Somani, K., Acharya, D., 2011. A heuristic approach to the railroad track maintenance scheduling problem. *Comput.-Aided Civ. Infrastruct. Eng.* 26 (2), 129–145. doi:10.1111/j.1467-8667.2010.00670.x.
- Rodriguez, J., Delorme, X., Gandibleux, X., Marlière, G., Bartusiak, R., Degoutin, F., Sobieraj, S., 2007. RECIFE: models and tools for analyzing rail capacity. *RTS-Rech. Transp. Secur.* 95, 129–146.
- Siebert, M., Goerigk, M., 2013. An experimental comparison of periodic timetabling models. *Comput. Oper. Res.* 40 (10), 2251–2259. doi:10.1016/j.cor.2013.04.002.
- Sourd, F., 2010. Sioucs: un outil d'aide la dcision pour la gestion des circulations. *Revue Gnrale des Chemins de Fer* 12, 20–26.