Integration of timetable planning and rolling stock in rapid transit networks

Luis Cadarso · Ángel Marín

Abstract The aim of this paper is to propose an integrated planning model to adequate the offered capacity and system frequencies to attend the increased passenger demand and traffic congestion around urban and suburban areas. The railway capacity is studied in line planning, however, these planned frequencies were obtained without accounting for rolling stock flows through the rapid transit network.

In order to provide the problem more freedom to decide rolling stock flows and therefore better adjusting these flows to passenger demand, a new integrated model is proposed, where frequencies are readjusted. Then, the railway timetable and rolling stock assignment are also calculated, where shunting operations are taken into account. These operations may sometimes malfunction, causing localized incidents that could propagate throughout the entire network due to cascading effects. This type of operations will be penalized with the goal of selectively avoiding them and ameliorating their high malfunction probabilities. Swapping operations will also be ensured using homogeneous rolling stock material and ensuring parkings in strategic stations.

We illustrate our model using computational experiments drawn from RENFE (the main Spanish operator of suburban passenger trains) in Madrid, Spain. The results show that through this integrated approach a greater robustness degree can be obtained.

Keywords Timetable · Rolling sock · Integration · Rapid transit · Robustness

1 Introduction

The increased passenger demand in urban and suburban areas has led many rapid transit network (RTN) operators to increase the offered capacity and system frequencies. However,

L. Cadarso · Á. Marín (✉)
Universidad Politécnica de Madrid, E.T.S.I. Aeronáuticos, Plaza Cardenal Cisneros, 3, Madrid 28040, Spain
e-mail: angel.marin@upm.es

L. Cadarso
e-mail: luis.cadarso@upm.es

 Springer

this increase may produce system congestion and consequently malfunction. In this way, frequencies and capacities must be planned according to passenger demand trying to satisfy their trip requirements. In RTN frequencies are high and distances are relatively short. Although the distances are not very large, service times are high due to the large number of intermediate stops required to allow proper passenger flow.

For any railway company, its efficiency in competitive markets depends strongly on its ability to define the railway timetabling so the demand meets a train at a desirable time. The railway company must provide the trains with the adequate material and composition so they have enough capacity to passengers feel comfortable in their trips.

Planning processes related to railway systems are fields that are rich in combinatorial optimization problems. Well-known examples of these are operational planning problems such as line planning, timetabling and rolling stock assignment. However, previous planning processes have been usually solved sequentially. This issue may lead to the system to operate in an inefficient way. The integration increases the flexibility and the robustness of the railway system and also makes easier to introduce robustness to the system.

The overall railway planning problem may be summarized by the following steps: the railway network design problem, the line planning problem, the timetabling problem, the rolling stock assignment, the train routing problem, and the crew scheduling problem.

Designing a RTN, or even extending one that is already functioning, is a vital strategic subject due to the fact that it reduces the future traffic congestion, travel time and pollution. The location decisions and the maximum coverage of the demand for public network is the main goal, taking a list of potential rapid transit corridors and stations and budget into account.

The following step after designing a RTN is planning its lines (origin and destination stations, stops and frequencies), from now on called Railway Line Planning (RLP) problem, a tactical planning problem. It is the problem of designing a line system such that all travel demands are satisfied and certain objectives are met, such as maximizing the service towards the passengers or minimizing the operational costs of the railway system.

The general aim of the Railway Timetabling (RT) problem is to provide a timetable for a number of trains on a certain part of the railway network based on the RLP frequencies. One may distinguish between cyclic and noncyclic timetables. A cyclic railway system is easy to remember for passengers because trains belonging to the same line always leave, for example, at $x : 20$ and $x : 50$. The noncyclic timetable is especially relevant on RTN, where the capacity of the infrastructure is limited. Then, the railway operators request to the infrastructure manager their preferred time slots.

Given a train fleet and finding the optimal material and composition assignment to satisfy both the timetable and the demand in a dense RTN is known as the rapid transit Rolling Stock (RS) problem. Material type and composition to be assigned to each train are determined in this phase.

The train routing problem is the process of determining a sequence for each material in the train network once the material assignment for each operation is known. The goal is to obtain sequences that minimize some cost, as the propagated delay, to achieve a robust solution; a different objective might be to maximize maintenance opportunities. To achieve an efficient routing solution, we must first obtain a good RS assignment. Finally, the crew assignment is solved. The crew must be efficiently assigned to perform each obtained train sequence.

The operative planning is aimed to meet the public transport demand, which varies during the hours of a day, the day of a week, from one season to another, etc. In this work we match demand requirements updating frequencies, planning trains' departure times and assigning

RS material and compositions for a given planning period. For RENFE, the planning period is of one day, which is repeated during the working days of a week, so a periodic planning is needed where the data and the decisions must be considered in the context of a space-time network.

The planned frequencies are known from the RLP problem. However, these planned frequencies were obtained without accounting for RS flows through the RTN. In order to provide the problem more freedom to decide RS flows jointly with train scheduling and therefore better adjusting these flows to passenger demand, we impose a minimum and maximum arc frequency values based on the previous obtained frequencies during the RLP. In this way, frequencies will be readjusted providing a more efficient network operation.

The RT model makes decisions about trains departure times accounting for passenger demand during each time period. In an integrated approach, it also decides shunting schedules (empty movements and composition changes schedules). The RS model makes decisions about the RS assignment to trains and shunting in the depot stations. The integrated RT&RS problem can be stated as follows in the context of metropolitan RTN: given the expected numbers of passengers at each arc and during each period, and accounting for shunting operations' schedule design, find the optimal RT and RS assignment.

Major complicating issues are the available shunting capacities. Shunting is related to the need for RS to be parked in shunting areas when the RS is not used for traffic during off-peak hours and for those maneuvers to match train compositions during the planning period. This forces us to combine and split convoys to form trains and to consider the logistics of empty train movements in order to meet depot station capacities. However, empty trains will also be moved to ensure RS availability because RS is a very limited resource during rush hours, when the passenger demand is very high. Another complicating issue is rotation time. Rotations are the maneuvers performed at depot stations to change the direction of motion of the RS.

As stated above, shunting operations are necessary in RTN. However, these operations may sometimes be difficult to execute and they can easily malfunction, causing localized incidents that could propagate through the entire network due to cascading effects. In order to provide robustness to the problem, these operations will be penalized to selectively avoid them and their high malfunction probabilities.

This paper presents a specialized problem describing RTN. The RT&RS Model (RT&RSM) will consider the optimization of trains' departure times, material and composition assignments, empty trains and the optimal management of convoys in the depot stations, all while considering the character and capacities of these type networks.

This paper is organized as follows. A literature overview is given in Sect. 2. We describe the problem for RTN in Sect. 3. In Sect. 4, the mathematical formulation is presented in detail. Section 5 contains the computational results based on a realistic cases provided by RENFE. Finally, we present our conclusions in Sect. 6.

2 State of the art

Caprara et al. (2007) provide an excellent state of art in railway optimization. They develop an introduction, a survey of available literature and a formal statement of the following problems: the RLP problem, the RT problem, the train platforming problem, the RS problem, the shunting problem and the crew planning problem.

The Railway Timetabling (RT) problem provides the train depart scheduling. There are two different possibilities when defining the RT: cyclic and non-cyclic timetables. Some

references in cyclic RT are Nachtingall (1996), Nachtingall and Voget (1997) and Kroon and Peeters (2003); they develop the so-called Periodic Event Scheduling Problem (PESP), based on the paper of Serafini and Ukovich (1989), a set of repetitive events is scheduled under cyclic time windows constraints. They do not consider the passenger demand.

The non-cyclic RT is the common one for dense metropolitan networks. Carraresi et al. (1996) consider the improvement of the effectiveness of the RTN, proposing a model which minimizes passengers' total waiting time by modifying the departure time of the trips. They take as input a passenger assignment which satisfies the vehicles capacities. Brännlund et al. (1998) discretize the time into time slots and subdivide the line into blocks. They use constraints in order to avoid the use of the same block during the same time slot by two different trains. Caprara et al. (2002) define ILP models based on graph representation. They solve it by Lagrangian relaxation to drive bounds in the context of a heuristic procedure. It produces good relaxations and heuristic solutions also for large size test problems. Caprara et al. (2006) in the same context and methodology consider additional constraints: station capacities, prescribed timetable for a subset of trains, track out of order, etc. Cacchiani et al. (2010) study the problem where both passenger and freight trains are run. While the passenger trains have a prescribed timetable that cannot be changed, freight train operators send the infrastructure manager requests to insert new freight trains. The objective is to introduce as many new freight trains as possible by assigning them timetables.

Once the RT has been defined, the RS assignment must be done. An integer programming model is considered by Alfieri et al. (2006) to determine the RS circulation for multiple RS types on a single line and on a single day. They use the concept of a transition graph to deal with this aspect. This concept is based on the assumption that for each trip, the next trip is known a priori. The model described by Alfieri et al. (2006) was extended by Fioole et al. (2006), to include combining and splitting trains, as happens at several locations in the Dutch timetable. They use an extended set of variables to locally obtain an improved description of the convex hull of the integer solutions. Robustness is considered by counting the number of composition changes.

Maróti (2006) focuses on planning problems that arise at NS (the main operator of passenger trains in the Netherlands). He identifies tactical, operational and short-term rolling stock planning problems and develops operations research models for describing them. Then, he analyzes the considered models, investigates their computational complexity and proposes solution methods. The allocation of RS units to French TGV trains is studied by Ben-Khedher et al. (1998). The RS circulation must be adjusted to the latest demand known from the seat reservation system. Therefore, this problem contains a strong re-scheduling component. The objective is to maximize the expected profit for the company.

A locomotive and carriage assignment problem was presented by Cordeau et al. (2000). The authors formulate the problem as a large integer program and use Benders decomposition to solve it. In a subsequent paper by Cordeau et al. (2001), their model was extended by considering various aspects such as maintenance of the RS. A RS circulation problem related to the circulation of ICE train units in the German ICE network was described by Mellouli and Suhl (2007). In this case, the required capacities of the trains are known a priori. Carriages and locomotives first have to be combined into train units of certain pre-specified groups, and these train units then have to be routed through the network in an optimal way. Cadarso and Marín (2011) define a model to study suburban rapid transit RS with convoys formed by three cars of the same type. The trains may be composed of one or two convoys in a dense network to attend to asymmetric demand and scheduling.

Liebchen and Möhring (2008) propose to integrate vehicle scheduling into the task of periodic timetabling. They demonstrate that the modeling capabilities of the PESP are not

limited only to periodic timetabling. They use the well-established technique of minimizing the number of vehicles required to operate a periodic timetable by penalizing waiting times of vehicles. However, they do not assign specific vehicles to trains.

2.1 Contributions

In the mentioned literature, the RT and RS problems are solved in a sequential manner. Our main contribution is that we integrate these planning phases in a model that decides simultaneously trains' schedules and RS material and compositions. Moreover, trains' frequencies are not fixed data. We impose minimum and maximum frequency values related to the one obtained by the RLP problem. In this way, the model recalculates frequency values in order to enable a smooth material flow through the network.

In this work, we use a non-cyclic approach for the RT problem. We use this approach because the RTN we are working on has limited RS resources and station capacities. Passenger demand is associated to arcs and time periods and the RT is developed accounting for passenger willings.

In this integration we also include shunting operations: empty movements, composition changes and rotations. As we are studying RTNs with limited resources and high frequencies, we have to account for these possibilities. The model decides their schedule and material and composition assignment.

Robustness has been introduced into the model by penalizing dangerous shunting operations and also ensuring that swapping operations are possible to perform in strategic depot stations.

Most of the contributions mentioned above assume that the RT problem defines also trains' sequences, that is, for each train, the next train is known. However, in our approach we consider a multicommodity flow model in which commodity flows and their schedules are studied. Once these flows are known, train sequences will be determined in the routing problem.

3 Timetable planning and rolling stock problem in rapid transit networks

In this section, the RT and RS problems are described in detail. First, the train supply is introduced. Next, we introduce passenger demand, and finally, we establish the concept of robustness in the problem.

3.1 Train supply

The railway infrastructure is described by stations and directed arcs between stations. Let S be the set of stations and let A be the set of arcs. Each arc $a \in A$ is from departure station $ds_a \in S$ to arrival station $as_a \in S$. Some stations have depots attached and at those stations it is possible to change the assignment of RS to trains. $SC \subseteq S$ denotes the set of depot stations.

Time is discretized into a set of time intervals T. Each time interval $t \in T$ represents a certain interval in time, for example from 8:00 to 8:01.

Let L be the set of Train Lines. They are defined by an origin and a destination, both depot stations. A train line is a passenger train traveling from a depot station to another depot station stopping at a determined number of intermediate stations. The model decides whether a train line is assigned to a departure time or not, based on the knowledge of passenger demand in each arc and period.

Each train line $l \in L$ starts at depot station $SD_l \in SC$ and ends at depot station $ED_l \in SC$. For each infrastructure arc $a \in A$ we let $LA_a \subseteq L$ denote the set of lines that use arc a. The line l is of length km_l kilometers.

For safety purposes a minimum separation time is imposed between two consecutive trains using the same infrastructure, that is, the headway (in order to avoid congestion and possible incidents). Thus, the headway must be maintained everywhere. For the cases where a central station exists it will be enough to impose the headway in this station for arriving and departing trains. We mean by central station as a station through which every train line comes.

As mentioned before, different types of carriages and locomotives (which can move themselves) exist, and naturally, a carriage cannot move without a locomotive. To enable train movement, carriages are attached to locomotives of the same material. Let M be the set of train materials. A convoy can move in both directions and it is composed of one carriage and two locomotives. The composition assigned to a train line will be given by the number of convoys forming the train line, that is, convoys from the same material can be coupled to form longer trains or uncoupled to form shorter trains. Let C be the set of train compositions. For example, a train line with double composition will have two carriages (Ca) and four locomotives (Lo) in the following sequence: Lo.Ca.Lo-Lo.Ca.Lo.

Each train line l will have to be assigned a departure time $t \in T$ and a RS material $m \in M$ and composition (i.e., number of convoys) $c \in C$.

3.2 Passenger demand

A known demand must be met by a given fleet. We assume that the forecasted travel demand is represented by a given origin destination matrix. However, this data is always difficult to obtain or/and to maintain it updated and does not represent passenger willings accurately. It is usually assumed that each passenger uses a pre-specified path through the network: each passenger is supposed to travel along the shortest-distance path through the network from his/her origin to his/her destination. In this way, with the available origin destination matrix and passenger counts in the network, we can obtain a passenger flow through arcs associated to time periods approximately.

Thus, for passenger demand we use the expected number of passengers between consecutive stations during each time period given by RENFE. The expectation is based on historical data from the autumn of 2008. The first and best choice would be to include the actual passenger behavior, that is, the passenger realization through different itineraries. Although this approach would give almost the true optimum, it is very complicated to include it in the model in a realistic way due to the large number of possibilities involved in RTNs. For example, in RTNs there are a lot of passengers itineraries; including them into the model, associated with the large number of time periods makes difficult to solve the problem efficiently.

The passenger demand for this problem is treated as a passenger flow $pf_{a,\tau}$ through each arc a during time period $\tau \in TD_a$. Let TD_a represent the time periods through which demand is counted in each arc a. We also include into the model the possibility of neglecting passengers from the system. That is, if passengers are not attended during the time periods they are willing to travel, they are denied. τ is a time interval containing different time periods t. That is, $\tau = [t_n, t_{n+k}]$, where t_n represents any time period and k is a scalar to be defined. The scalar k measures the aggregation level of the passenger demand. That is, if k is small we are not aggregating passengers' flows and they must be attended at their desired times or they are denied from the system. If it is greater, we let the model to aggregate the demand.

We assume that passenger demand waits in the station until a train comes. If there is enough capacity, every passenger will go on the train. Otherwise, passengers must wait until the next train coming in the station during time interval τ. If the train does not come, remaining passengers will be denied from the system. We must note that the demand is treated as a passenger flow, so in most cases, the major part of the demand will already come in the train from the previous arc.

Under the above hypothesis, the model will treat the passengers from a centralized point of view (i.e., only the operator criteria are optimized). However, since the proposed problem relates to a suburban RTN, it is obvious that every passenger will have the option to choose any other available company or transportation mode. Thus, the operator has to factor in passenger behavior to avoid losing passengers to other transportation companies.

Passenger flow is directly related with the timetable that was being used when the demand study was made. However, we use this data to change the timetable. First, we must realize that we are studying rapid transit systems where frequencies are very high, so passengers do not care so much the timetable but the frequency. In this way, according to the RLP problem we impose minimum and maximum frequency values in determined arcs of the network. Hence, the new solution will not be so far from the previous one planning arc frequencies. Moreover, a minimum separation time is imposed for every train path departing from each depot station. In this way, the offered capacity, in terms of frequency, in the new solution will be similar to the previous one and enough sparse in time.

The operator will always try to offer comfortable capacity in each train line. It is obvious that when more capacity is offered, more passengers will travel comfortably. However, offering more capacity increases operating costs dramatically. Therefore, the composition assigned to each train line will be a tradeoff between the operating costs and the passengers' comfort level.

For each convoy formed by one material type m, the passenger capacity is known. There is a fixed seating passenger capacity and a variable standing passenger capacity. Multiple possibilities arise when considering the standing passenger capacity. The aim is to obtain adequate passenger capacity for every train path. This may be obtained with different configurations for standing passengers. We define comfortable capacity as the full seating capacity and fewer than 3 pax/m^2 standing. If this capacity is exceeded, the passengers above this level are deemed passengers in excess. If the density of standing passengers is between 3 and 4 pax/m^2, each passenger in excess represents a moderate penalty, because the operator would like to obtain 3.5 pax/m^2. If the density of standing passengers exceeds 4 pax/m^2, passengers in excess are highly penalized because this situation is deemed very uncomfortable.

3.3 Robustness

Railway planning is currently divided into several optimization steps from first strategic decisions to daily operations. It is well known that disintegrated planning produce optimal solutions for each stage but non-optimal global solutions. We integrate, so far separately considered, key planning stages (railway timetabling and rolling stock planning), and we carry on the research on integrating robustness into railway planning. That is, robustness is in part achieved by integrating different planning phases, because in our case the RS assignment and shunting operations depend strongly on RT, so a greater robustness degree can be achieved with the proposed combined approach.

The general meaning of the integration is as follows: consider two consecutive planning stages. In order to obtain a high-quality solution in the second stage, some *slack* must be

introduced in the first stage. In the same way, a robust solution adds *slack* to safeguard against data perturbation. This slack may provide a *smooth interface* between subsequent planning stages. This becomes a connexion between robustness and integration.

Robustness is also introduced penalizing composition changes and empty movements. When a composition change is performed, multiple failures can occur, forcing the train to be parked for a long time and causing an incident. The mechanical system used to perform a composition change is automatic, but it sometimes fails and requires extra time to enact the change. Moreover, during composition changes, the brakes' pneumatic circuit must be joined or separated depending on the performed operation. This is always a difficult and complicating issue, and human resources are required to perform it. Above all, composition change times are overestimated to account for the effects mentioned above and to try to introduce robustness into the system. Finally, if a malfunction has occurred it must be contained to avoid cascading effects. Containment of cascading effects is easier if the incident occurs during off-peak hours when more RS material is available at the depot station and other material can depart from the station to attend to as much of the demand as possible. In our model, this is treated by harshly penalizing composition changes during rush hours (see coefficient $\vartheta_{s,t}$, which depends on the station s and the time period t).

Similarly, empty movements during rush hours complicate network operation because they use the same infrastructure as commercial train services. In addition, they obviously require human resources. Although human resources are always available at depot stations to perform composition changes and empty movements, it is better to keep these resources in the depot station to alleviate possible incidents during rush hours. Therefore, empty movements during rush hours are also heavily penalized. It is also better to avoid (if possible) empty movements to destination depot stations with time-dependent capacities (i.e., stations that are shared with different lines). This idea is represented by the $\theta_{s,s',t}$ coefficient, which penalizes empty movements between depot stations s, s' within departure time period t.

Finally, the system is made more robust by assigning only one material type per line (i.e., for every train service operating the same line, the material must be equal). This constraint allows for all material on one line to be swapped between different train services at depot stations serving that line. Thus, there will be more opportunities to swap train services if an incident occurs, and the propagation of the incident can be mitigated easily.

In order to ensure that at least one swapping operation is always possible at both edges of train lines, we mandate the model to provide solutions with free parkings in depot stations at lines' extremes. In this way, the obtained solution will be more resistant to possible delays because they may be alleviated by swapping trains.

4 The timetable planning and rolling stock model

In this section the RT&RSM integrated formulation is presented. It is based on the formulation presented by Cadarso and Marín ([2011](#)), where the RS assignment was done to a known schedule. Then, the main difference with respect to the cited work is that the timetable is a variable in the problem. That is, given train lines departure times must be determined. In order to match infrastructure manager requirements, headway constraints must be used to ensure a minimum separation time between trains using the same infrastructure. Moreover, frequency values are not fixed. For some arcs this value is upper and lower bounded in order to provide a desired service level. Another important difference is the demand treatment. Passenger flows associated to arcs and time periods are well known. Different demand aggregation levels may be studied with this new formulation depending on the value of the scalar k.

   This model formulation is based on the assumption of that an unique mode of transport exists. Thus, if the offered capacity is not enough, the demand will be neglected assuming that it will use another mode of transport. However, choosing the demand value according to the operator database we can assume that almost the whole demand will use the RTN under the above explained assumptions.

4.1 Sets

- S is the set of stations.
- $SC \subseteq S$ denotes the set of depot stations.
- A is the set of arcs. The arcs represent the infrastructure. Each arc $a \in A$ is from departure station $ds_a \in S$ to arrival station $as_a \in S$.
- L set of train lines defined by departure depot station, arrival depot station and every intermediate arc.
- $LA_a \subset L$ is the set of train lines containing arc a.
- T is the set of time periods. A time period $t \in T$ represents a certain interval in time, for example from 8:00 to 8:01. τ represents a time interval grouping k different time periods t: $\tau = [t_n, t_{n+k}]$, where t_n represents a determined time period.
- $TD_a \subseteq T$ is the set of time intervals through which the demand is counted in each arc a.
- M is the material set.
- C is the composition set.
- $CT \subseteq T$ is the set of time intervals where the trains are counted.
- LM, LO: set of train lines l going in one sense of the line; set of train lines going in the opposite sense of movement.
- $LCS_{l,t}$: set of time periods during which train line l departs from the origin such that it is coming through central station during time period t.
- $LAT_{l,a,\tau}$: set of time periods during which train line l departs from its origin such that it is coming through arc a during τ.
- LA_s: set of train lines arriving in depot station s.
- LD_s: set of train lines departing from depot station s.
- AMF: set of arcs a with a bounded frequency.

4.2 Cost parameters

We consider the following cost parameters in the objective function:

- $c_{l,t}^{m,c}$ is the operating cost for train line l departing during time period t with material $m \in M$ and composition $c \in C$.
- $ec_{m,c}$ is the empty movement operating cost per kilometer with material $m \in M$ and composition $c \in C$.
- $\theta_{s,s',t}$ is a penalty value for empty movements between stations s, s' departing during time period t.
- $\vartheta_{s,t}$ is the cost of performing a composition change in depot station s during time period t. This cost may be increased in order to penalize composition changes in some stations during certain time periods.
- $epc_{a,\tau}^{3-4}$ is the penalty per passenger in excess between 3 and 4 pax/m^2 in arc a during time period $\tau \in TD_a$.
- $epc_{a,\tau}^{4-\text{up}}$ is the penalty per passenger in excess above 4 pax/m^2 in arc a during time period $\tau \in TD_a$.
- $upc_{a,\tau}$ is the penalty per unattended passenger in arc a during time period τ.

- ic_m is the investment or leasing cost for a unit of material $m \in M$.
- $km_{s,s'}$ is the distance in kilometers from s to s'.

4.3 Other parameters

- $pf_{a,\tau}$ passenger flow through arc a during time period τ.
- $q_{m,c}^{3-\text{down}}, q_{m,c}^{3-4}, q_{m,c}^{4-\text{up}}$ passenger capacity of material $m \in M$ and composition $c \in C$ for different capacities: below 3 pax/m², between 3 and 4 pax/m² and above 4 pax/m², respectively.
- χ_m is the fleet size for each material m, i.e. The maximum number of trains that may run at the same time.
- $\underline{f}_a$ minimum frequency in arc a.
- $\overline{f}_a$ maximum frequency in arc a.
- tt_l trip time for train line l.
- st separation time between two consecutive trains.
- r_s rotation time in depot station s.
- e_s coupling time in depot station s.
- d_s uncoupling time in depot station s.
- o_c number of convoys in composition c.
- t_i, t_f initial and ending time periods in the planning period.
- $cap_{s,t}$ capacity of depot station s during time period t.
- $\beta_{l,t',t} = 1$, if train line l, which departed during time period t', is still rolling during time period t; $= 0$, otherwise.
- $\gamma_{s,t',t} = 1$, if a rotation performed in depot station s, which finishes during time period t', was being performed during time period t; $= 0$, otherwise.
- $\xi_{s,s',t',t} = 1$, if an empty movement performed between depot stations s, s', which started during time period t', is still being performed during time period t; $= 0$, otherwise.
- $\mu_{s,t',t} = 1$, if a composition change performed in depot stations s, which started during time period t', is still being performed during time period t; $= 0$, otherwise.

4.4 Variables

The following variables are used in the model:

- $x_{l,t}^{m,c}$: binary variable. $= 1$, if train line $l \in L$ is assigned to departure time $t \in T$ using composition $c \in C$ and material $m \in M$; $= 0$, otherwise.
- $em_{s,s',t}^{m,c}$: binary variable. $= 1$, if an empty movement is performed between stations s and s' departing during time period t with material m and composition c; $= 0$, otherwise.
- $\epsilon_{s,t}^{m,c}$: binary variable. $= 1$, if a coupling operation is performed in station s starting during time period t with material m and composition c; $= 0$, otherwise.
- $\delta\epsilon_{s,t}^{m,c}$: binary variable. $= 1$, if an uncoupling operation is performed in station s starting during time period t with material m and composition c; $= 0$, otherwise.
- $\rho_{s,t}^{m,c}$: binary variable. $= 1$, if a rotation operation is performed in station s finishing during time period t with material m and composition c; $= 0$, otherwise.
- $yt_{s,t}^{m,c}$: integer variable. It denotes the train inventory in depot station s during time period t of material and composition m, c.
- $h_{a,\tau}^{3-4}$: positive variable. It denotes the number of passengers in excess between 3 and 4 pax/m² in arc a during time period $\tau \in TD_a$.
- $h_{a,\tau}^{4-\text{up}}$: positive variable. It denotes the number of passengers in excess above 4 pax/m² in arc a during time period $\tau \in TD_a$.

- $dp_{a,\tau}$: positive variable. It denotes the number of disrupted passengers in arc a during time period τ.
- yn_m: integer variable. It denotes the number of convoys of material m to buy or borrow from other lines.

In the following subsections we introduce the Timetable Planning and Rolling Stock Model (TTP&RSM) formulation.

4.5 Objective Function

$$
\min \sum_{l \in L} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} c_{l,t}^{m,c} x_{l,t}^{m,c} + \sum_{s,s' \in SC} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \theta_{s,s',t} ec_{m,c} km_{s,s'} em_{s,s',t}^{m,c}
$$
$$
+ \sum_{s \in SC} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} \vartheta_{s,t} \left(\epsilon_{s,t}^{m,c} + \delta\epsilon_{s,t}^{m,c} \right) + \sum_{m \in M} ic_m yn_m + \sum_{a \in A} \sum_{\tau \in TD_a} epc_{a,\tau}^{3-4} h_{a,\tau}^{3-4}
$$
$$
+ \sum_{a \in A} \sum_{\tau \in TD_a} epc_{a,\tau}^{4-\text{up}} h_{a,\tau}^{4-\text{up}} + \sum_{a \in A} \sum_{\tau \in TD_a} upc_{a,\tau} dp_{a,\tau}
$$

In the objective function the following terms have been considered, every one representing an economic penalization:

– operating cost for every line with a determined material and composition,
– operating costs for empty movements,
– cost of performing composition changes,
– investment or leasing cost in new material,
– penalty for passengers in excess of comfortable capacity and,
– penalty for unattended passengers.

4.6 Service constraints

$$
\sum_{l \in LA_a} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} x_{l,t}^{m,c} \geq \underline{f}_a \qquad \forall a \in AMF \tag{1}
$$

$$
\sum_{l \in LA_a} \sum_{t \in T} \sum_{m \in M} \sum_{c \in C} x_{l,t}^{m,c} \leq \overline{f}_a \qquad \forall a \in AMF \tag{2}
$$

$$
\sum_{\substack{t2 \in T: \\ t2 \geq t, t2 \leq t+st}} \sum_{l \in LM} \sum_{t1 \in LCS_{l,t2}} \sum_{m \in M} \sum_{c \in C} x_{l,t1}^{m,c} \leq 1 \quad \forall t \in T \tag{3}
$$

$$
\sum_{\substack{t2 \in T: \\ t2 \geq t, t2 \leq t+st}} \sum_{l \in LO} \sum_{t1 \in LCS_{l,t2}} \sum_{m \in M} \sum_{c \in C} x_{l,t1}^{m,c} \leq 1 \quad \forall t \in T \tag{4}
$$

Constraints (1) ensure that a minimum frequency is offered in some arcs. The maximum frequency value for the arcs is imposed by constraints (2). Constraints (3) ensure that the headway is maintained in one of the senses of movement of the RTN infrastructure, that is, we only account for the set of train lines going in that sense (LM). Similarly, constraints (4) ensure the headway for the set of train lines in the other sense of movement (LO). For each train line l and its possible time $t2$ during which it is coming through central station we account for its departure time from the origin ($LCS_{l,t2}$); then we add every train line coming through this station between a time instance t and the time instance $t + st$ and we obey it to be equal or lower to 1.

4.7 Passenger constraints

$$\sum_{l \in LA_a} \sum_{t \in LAT_{l,a,\tau}} \sum_{m \in M} \sum_{c \in C} q_{m,c}^{3-\text{down}} x_{l,t}^{m,c} \geq pf_{a,\tau} - h_{a,\tau}^{3-4} - h_{a,\tau}^{4-\text{up}} - dp_{a,\tau}$$

$$\forall a \in A, \tau \in TD_a \tag{5}$$

$$\sum_{l \in LA_a} \sum_{t \in LAT_{l,a,\tau}} \sum_{m \in M} \sum_{c \in C} (q_{m,c}^{3-4} - q_{m,c}^{3-\text{down}}) x_{l,t}^{m,c} \geq h_{a,\tau}^{3-4} \quad \forall a \in A, \tau \in TD_a \tag{6}$$

$$\sum_{l \in LA_a} \sum_{t \in LAT_{l,a,\tau}} \sum_{m \in M} \sum_{c \in C} (q_{m,c}^{4-\text{up}} - q_{m,c}^{3-4}) x_{l,t}^{m,c} \geq h_{a,\tau}^{4-\text{up}} \quad \forall a \in A, \tau \in TD_a \tag{7}$$

Passenger capacity constraints (5) ensure that for each arc $a \in A$ and each time interval $\tau \in TD_a$, the capacity of the train lines is enough to accommodate the passenger demand $pf_{a,\tau}$ minus the passengers that are in excess (denoted by variables $h_{a,\tau}^{3-4}$, $h_{a,\tau}^{4-\text{up}}$) or unattended (denoted by variable $dp_{a,\tau}$). The available capacity is the add of all possible train lines that will come trough arc a and during each time interval τ. This information is in $LAT_{l,a,\tau}$, where for each pair (a, τ) and train line l, we have the departure time during which it departed from its origin.

Passenger in excess of comfortable capacity constraints (6)–(7) ensure that passengers in excess are bounded by the different available capacities.

4.8 Rolling stock constraints

$$yt_{s,t-1}^{m,c} + \sum_{l \in LA_s} x_{l,t-tt_l}^{m,c} + o_{c-1} \cdot \epsilon_{s,t-e_s}^{m,c-1} + o_{c+1} \cdot \delta_{s,t-d_s}^{m,c+1} + \rho_{s,t}^{m,c} + \sum_{s' \in SC} em_{s',s,t-et_{s',s}}^{m,c}$$

$$= yt_{s,t}^{m,c} + \sum_{l \in LD_s} x_{l,t}^{m,c} + o_{c+1} \cdot \epsilon_{s,t}^{m,c} + o_{c-1} \cdot \delta_{s,t}^{m,c} + \rho_{s,t+r_s}^{m,c} + \sum_{s' \in SC} em_{s,s',t}^{m,c}$$

$$\forall s \in SC, t \in T, m \in M, c \in C \tag{8}$$

$$\sum_{l \in LD_s} x_{l,t}^{m,c} = \rho_{s,t}^{m,c} + \epsilon_{s,t-e_s}^{m,c-1} + \delta_{s,t-d_s}^{m,c+1} \quad \forall s \in SC, t \in T, m \in M, c \in C \tag{9}$$

$$\sum_{s \in SC} \sum_{c \in C} o_c \cdot yt_{s,t}^{m,c} + \sum_{l \in L} \sum_{t' \in T} \sum_{c \in C} o(c)\beta_{l,t',t} x_{l,t'}^{m,c}$$

$$+ \sum_{s \in SC} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \gamma_{s,t',t} \cdot \rho_{s,t'}^{m,c} + \sum_{s,s' \in SC} \sum_{t' \in T} \sum_{c \in C} o_c \cdot \xi_{s,s',t',t} \cdot em_{s,s',t'}^{m,c}$$

$$+ \sum_{s \in SC} \sum_{t' \in T} \sum_{c \in C} \mu_{s,t',t}(o_{c+1} \cdot \epsilon_{s,t'}^{m,c} + o_c \cdot \delta_{s,t'}^{m,c})$$

$$\leq \chi_m + yn_m \quad \forall m \in M, t \in CT \tag{10}$$

$$\sum_{m \in M} \sum_{c \in C} o_c \cdot yt_{s,t}^{m,c} + \sum_{t' \in T} \sum_{m \in M} \sum_{c \in C} \mu_{s,t',t}(o_{c+1}\epsilon_{s,t'}^{m,c} + o_c\delta_{s,t'}^{m,c})$$

$$+ \sum_{t' \in T} \sum_{m \in M} \sum_{c \in C} o_c \cdot \gamma_{s,t',t} \cdot \rho_{s,t'}^{m,c} \leq cap_{s,t} \quad \forall s \in SC, t \in T \tag{11}$$

$$yt_{s,t_i}^{m,c} = yt_{s,t_f}^{m,c} \quad \forall s \in SC, m, c \in M, C \tag{12}$$

Material and composition conservation constraints (8) ensure the convoys flow balance. These constraints ensure that the train number for every station and material at period $t - 1$, plus the arriving trains, minus the departing ones is equal to the train number at period t.

Rotation and departure constraints (9) ensure that a rotation is performed before each train service departure.

In order to avoid that some of the time (i.e., $t - tt_l, t - e_s, t + r_s$) is outside set T, some variables are fixed to 0 in the mathematical program. This situation affects to constraints (8) and (9).

Fleet capacity constraints (10) ensure that the number of trains used during time $t \in T$ is limited by the size of the fleet. Note that these constraints count the running trains and those ones in depot stations. The variable yn_m specifies the additional trains to borrow from other lines at high cost.

Depot capacity constraints (11) ensure that station capacity is not overpassed. They account for parked and shunting material.

Trains initial and final constraints (12) denote that the number of convoys and compositions available in the stations at the beginning and end of the planning period are equal.

4.9 Variable domain

$$x_{l,t}^{m,c} \in \{0, 1\} \qquad \forall l \in L, \ t \in T, \ m \in M, \ c \in C \qquad (13)$$

$$em_{s,s',t}^{m,c} \in \{0, 1\} \quad \forall s, s' \in SC, \ t \in T, \ m \in M, \ c \in C \qquad (14)$$

$$\epsilon_{s,t}^{m,c} \in \{0, 1\} \qquad \forall s \in SC, \ t \in T, \ m \in M, \ c \in C \qquad (15)$$

$$\delta\epsilon_{s,t}^{m,c} \in \{0, 1\} \qquad \forall s \in SC, \ t \in T, \ m \in M, c \in C \qquad (16)$$

$$\rho_{s,t}^{m,c} \in \{0, 1\} \qquad \forall s \in SC, \ t \in T, \ m \in M, c \in C \qquad (17)$$

$$yt_{s,t}^{m,c} \in Z^+ \qquad \forall s \in SC, \ t \in T, \ m \in M, c \in C \qquad (18)$$

$$h_{a,\tau}^{3-4} \in R^+ \qquad \forall a \in A, \ \tau \in TD_a \qquad (19)$$

$$h_{a,\tau}^{4-\text{up}} \in R^+ \qquad \forall a \in A, \ \tau \in TD_a \qquad (20)$$

$$dp_{a,\tau} \in R^+ \qquad \forall a \in A, \ \tau \in TD_a \qquad (21)$$

$$yn_m \in Z^+ \qquad \forall m \in M \qquad (22)$$

5 Computational experiments

5.1 Introduction

All of our computational experience is for realistic cases drawn from RENFE's regional network in Madrid, also known as "Cercanías Madrid" (Fig. 1). This network is composed of 10 different lines with almost 100 stations. All data are from the year 2008. Approximately one million passengers use "Cercanías Madrid" every day.

The network presented in this study case (Fig. 1) is characterized by its modular structure. That is, in real life it is separated into different and independent modules for operating purposes. Every module has its own infrastructure (stations, depot stations, sections, etc.). In this way, the RS material cannot be easily transferred from one module to another; it

Fig. 1 RENFE Cercanías
Madrid rapid transit network

can be done, but extra human resources and time are needed to accomplish it; these extra resources are represented by the leasing cost in the objective function.

The capacity assigned to each module is decided jointly with the infrastructure manager during the RT phase. That is, the operator designs a planning with given capacities. This planning is presented to the infrastructure manager and the planning may be accepted or rejected. If it is rejected the timetable must be updated in order to match the infrastructure manager's willings. Thus, the operator faces an iterative process to develop the RT and therefore the RS assignment.

Our runs were performed on a Personal Computer with an Intel Core2 Quad Q9950 CPU at 2.83 GHz and 8GB of RAM, running under Windows 7 64Bit, and our programs were implemented in GAMS/Cplex 11.1.

5.1.1 Line C5

Line C5 has 23 stations and 4 depot stations: Mostoles el Soto, Atocha, Fuenlabrada and Humanes. Inside this line we consider the possibility of attending 10 different train lines. These train lines are defined by their initial and final depot stations and every station they come through. They are shown in Table 1. In the first column the line number is shown, in the second one the departure depot station and in the last one the arrival depot station.

There is one material type available (m1), and the trains can be of simple (one convoy) or double (two convoys) composition. The convoys for the material in Line C5 have definite characteristics, as shown in Table 2. The train capacity is divided into seated and standing passengers. The number of seats is fixed, and a density value is defined for standing passengers. In every convoy, for example, for a density of 3 pax/m^2 we would have 240 fixed seats plus 261 standing passengers (i.e., 501 passengers per convoy). However, for a density of 4 pax/m^2, we would have the same 240 seats but a total capacity of 588 passengers.

In a daily planning period from 5:00 a.m. to 1:00 a.m. divided into one minute periods, we have 1200 time periods. The rotation time is 3 minutes at every depot station.

Table 1 Train lines in Line C5

Lines	Origin	Destination
l1	Humanes	Atocha
l2	Atocha	Humanes
l3	Fuenlabrada	Atocha
l4	Atocha	Fuenlabrada
l5	Humanes	Móstoles El Soto
l6	Móstoles El Soto	Humanes
l7	Fuenlabrada	Móstoles El Soto
l8	Móstoles El Soto	Fuenlabrada
l9	Atocha	Móstoles El Soto
l10	Móstoles El Soto	Atocha

Table 2 Convoy capacity and length

Material type	Seats	Standing	Density (pax/m^2)	Length (m)
m1	240	261	3	80
		348	4	
		870	10	
m2	277	360	3	80
		480	4	
		1200	10	

5.1.2 Lines C3&C4

Line C3 is composed of 12 stations and 3 depot stations: Chamartín, Atocha and Aranjuez. In line C4, there are 18 stations and 8 depot stations: Parla, Parla Industrial, Getafe Centro, Atocha, Chamartín, Tres Cantos, Alcobendas and Colmenar Viejo. Some depot stations are shared between these lines.

There is one material type available (m2), and the train services can have simple (one convoy) or double (two convoys) compositions (see Table 2). Because the same material is used for both lines, RS can be interchanged between them.

Inside this line we consider the possibility of attending different train lines. These train lines are defined by their initial and final depot stations and every station they come through. They are shown in Table 3. In the first column the line number is shown, in the second one the departure depot station and in the last one the arrival depot station.

5.2 Rolling stock assignment with fixed timetable

Cadarso and Marín (2011) developed a model to assign RS to a set of scheduled trains (fixed timetable) in a RTN. They show how different solutions can be obtained by varying the penalties for excess passengers. These results are summarized in Table 4 for lines C5 and C3& C4, where different penalty cases are studied showing the results obtained. The pair of excess passenger penalties ($epc_{a,\tau}^{3-4}$, $epc_{a,\tau}^{4-up}$) is listed in the second and third columns, the number of convoys used in the proposed solution (#C) in the fourth column, the train service operating costs (TSOC) in the fifth column, the empty movement operating costs (EMOC) in the sixth column, excess passenger costs (PEC) are shown in the seventh column, and the

Table 3 Train lines in lines C3&C4

Lines	Origin	Destination
l1	Parla	Colmenar
l2	Colmenar	Parla
l3	Parla	Alcobendas
l4	Alcobendas	Parla
l5	Parla	Chamartín
l6	Chamartín	Parla
l7	Colmenar	Atocha
l8	Atocha	Colmenar
l9	Alcobendas	Atocha
l10	Atocha	Alcobendas
l11	Aranjuez	Chamartín
l12	Chamartín	Aranjuez

Table 4 Lines C5 and C3&C4 RS model solutions

Line	$epc_{a,\tau}^{3-4}$	$epc_{a,\tau}^{4-up}$	#C	TSOC	EMOC	PEC	#CC
C5	0.5	3	66	78331.44	1635.76	3019	16
	1	5	64	80099.76	1265.04	3554	20
	3	6	65	82896.24	1281.36	4403	18
C3&C4	0.5	3	62	87170.89	3951.18	2751	28
	1	5	63	87837.52	4090.83	2177	34
	3	6	62	88094.56	4090.83	2462	34

number of composition changes (#CC) in the last column. These results were obtained in times on the order of one minute. Moreover, these problems were solved to optimality.

With passengers in excess penalty values the offered capacity can be controlled. That is, if these penalties are big enough, passengers in excess will not arise and great capacity will be offered. However, if they are low enough, a lot of passengers in excess will appear in the model solution because of the low capacity offered.

In this paper, we will choose the same penalties for every arc and time period for simplicity. These penalty values are those ones according to the operator willings based on the results summarized in Table 4.

5.3 Passenger Demand Approach

Passenger demand is defined as a flow through every arc a during each time period τ. In τ's definition the scalar k is implicit. This scalar's value represents the level of aggregation for passenger demand. For example, for $k = 1$ we have the union of two consecutive time periods t, t'; that is, we will have that every passenger flow departing during t, t' will be treated as an aggregated passenger flow.

In this way, depending on k's value, the model size and solution will change. In order to illustrate the RT&RS model size for different k values, Table 5 shows the model size for the Line C5 study case. The RT&RSM numbers of discrete and continuous variables, constraints and non-zero elements are given for the model (RT&RSM). The number of discrete variables

Table 5 Lines C5 and and C3&C4 RT&RS model: the number of variables, constraints and non-zeros

	RT&RSM
# Discrete variables	54130
# Continuous variables	190804–190384–189409–188113–186847
# Constraints	105386–104826–103526–101798–100110
# Non-zero elements	1396843–1601563–1758451–1853187–1902275
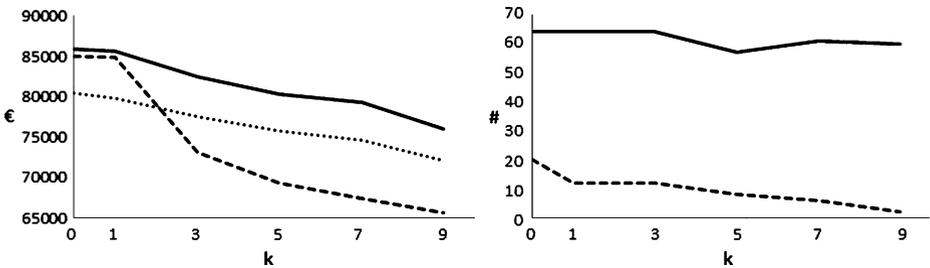
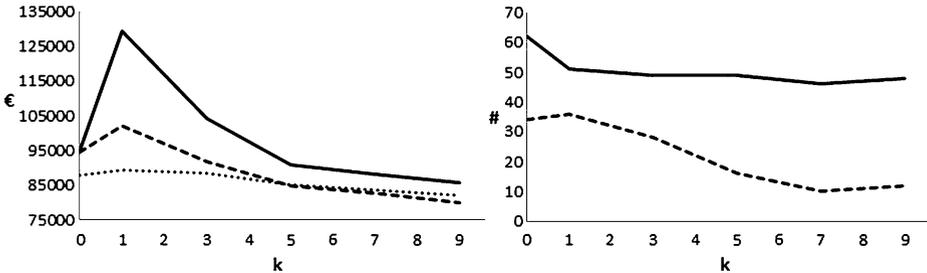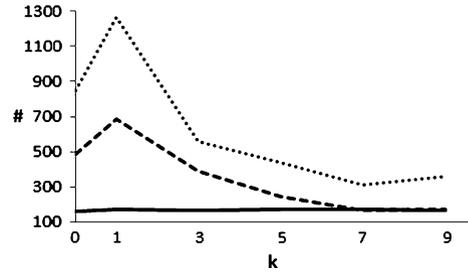**Fig. 2** Different RT&RS solutions for Line C5 obtained by varying $\tau(k)$ value

remains equal for every case because they do not depend on τ. However, the rest of variables, constraints and non-zero elements vary with τ definition.

We have done a study varying τ definition. The different cases we have studied are the following: aggregating two time periods ($k = 1$), four ($k = 3$), six ($k = 5$), eight ($k = 7$) and ten ($k = 9$) time periods. In this way, the model size is presented for every studied case in Table 5 for Line C5.

Every computational experience has been stopped after 1 hour of computational time. For every of them the optimality gap was always lower than 3%. In these computational experiments the robustness parameters have not been included. That is, the parameter $\theta_{s,s',t}$ is set to one and the parameter $\vartheta_{s,t}$ to its nominal value. On the left handside of Fig. 2 the objective function value of the mixed-integer program and of the linear relaxation program is showed for Line C5. The first one is represented by the continuous line and the second one by the discontinuous line. The value corresponding to $k = 0$ is the case where the RS assignment is done for a given and fixed timetable. As k goes up we can see how the liner relaxation becomes weaker; we must mention that the optimality gap (after 1 hour of computation) also increases as k goes up, but never exceeding the 3%. TSOC are represented by the dotted line, which maintains the same behaviour as the objective function for the mixed-integer program. On the right handside, the number of convoys used (continuous line) and the number of composition changes performed (discontinuous line) are shown for every case. The convoys used is maintained almost equal for every case. However, the number of performed composition changes is strongly reduced as k goes up. One could think that decreasing composition changes mandates a bigger fleet. Nevertheless, composition changes decrease as k increases, that is, as the time window's span we are using to attend the passenger demand increases, which means that the model has more freedom to adequate RS flows without performing composition changes.

The minimum separation time imposed between two consecutive trains using the same infrastructure is given by the network operator and the infrastructure manager. For Line C5 this value is of 3 minutes equivalent to 3 time periods. The value of this parameter is directly

Fig. 3 Passengers in excess for the arcs with the greatest passenger flow in Line C5

Fig. 4 Different RT&RS solutions for Lines C3&C4 obtained by varying $\tau(k)$ value

related with the choice of k. We must ensure that trains are coming through the different arcs and time periods τ enough separated in order to reduce passengers in excess. That is, if we choose τ's length lower than the time separation, it may occur that many passengers are neglected. In Fig. 3 passengers in excess and neglected passengers for the two consecutive arcs with the greatest passenger flow are showed for Line C5. The frequency values for these arcs are also represented. The continuous line represents these frequency values which are maintained around the same value for every tested case. The discontinuous and dotted lines represent the sum of neglected and in excess passengers for two different arcs. We must mention that neglected passengers appeared only for the case where $k = 1$. In order to develop more computational experiments for the Line C5 we choose the case where τ is exactly two times the value of the separation time, that is, $k = 5$. In this case we can see how passengers in excess are halved compared to the initial case where the timetable was fixed.

For Lines C3&C4 the same approach has been used, that is, 1 hour of computational time for each study case. For every of them the optimality gap was also always lower than 3%. On the left handside of Fig. 4 the objective function value of the mixed-integer program and of the linear relaxation program is showed for Lines C3&C4. The first one is represented by the continuous line and the second one by the discontinuous line. The value corresponding to $k = 0$ is the case where the RS assignment is done for a given and fixed timetable. TSOC are represented by the dotted line, which maintains the same behaviour as the objective function for the mixed-integer program. On the right handside, the number of convoys used (continuous line) and the number of composition changes performed (discontinuous line) are shown for every case.

In Lines C3&C4 the separation time is of 5 minutes. As we said before, if we choose τ's length lower than the time separation, it may occur that many passengers are neglected. For this reason the objective function value in Fig. 4 is very high for k ranging from 1 to 3. In Line C5 this issue was not a big problem because the separation time was lower and even for $k = 1$ there were few neglected passengers. In order to develop more computational

Table 6 RT&RS model solution compared to RS with fixed timetable solution and current RENFE solution

Line	Solution	#C	TSOC	EMOC	PEC	#CC	Savings (%)
C5 ($k = 5$)	RT&RS	57	75792.24	1892.72	2464	8	30.7
	RS	64	80099.76	1265.04	3554	20	27.4
	Current	74	109765.20	2232.12	874	0	–
C3&C4 ($k = 9$)	RT&RS	48	81865.91	2086.29	1372	12	41.2
	RS	63	87837.52	4090.83	2177	34	35.6
	Current	72	136633.86	6083.88	1550	0	–

experiments for Lines C3&C4 we choose the case where τ is exactly two times the value of the separation time, that is, $k = 9$.

In Table 6 different solutions are listed for Lines C5 and C3&C4. For each of them, in the first row the solution obtained by the proposed RT&RS model is shown (for Line C5 $k = 5$ and for Lines C3&C4 $k = 9$). In the second one, the solution obtained with the RS assignment with fixed timetable. In the last one, the current solution operated by RENFE. In the third column the number of used convoys (#C) is listed; this number is lowered compared to the current solution operated by RENFE and to the one obtained by RS assignment with fixed timetable. In this way, maintenance costs are also reduced because they depend on the number of used convoys. In the fourth column operating costs (TSOC) are shown, which are also slightly reduced. In the fifth column, empty movement costs (EMOC) are listed; they are reduced compared to the current solution; however, in the RT&RS solution these costs are increased compared to the RS assignment with fixed timetable. Passengers in excess costs are increased compared to the current solution; however, these costs include passengers above 3 pax/m^2 and RENFE considers passengers in excess when there are passengers exceeding 3.5 pax/m^2 (see Cadarso and Marín 2011). In the current solution there are no composition changes (#CC); however, in order to enable a schedule using fewer convoys they arise for the studied cases; for the case where the timetable is optimized the number of composition changes is lower due to the great freedom that supposes the timetable optimization. In the last column savings in operating costs with respect to the current solution are shown.

5.4 Maintenance costs

An important cost to the operator is the maintenance cost. For example, for this material, there is a daily fixed cost of, say, 400 €. Given the importance of the number of convoys used in the network, different solutions must account for that number.

If we compare any of the proposed solutions in Table 6 with the current solution provided by RENFE's operators, we can see that the number of used convoys is smaller.

However, the model does not account for this maintenance costs. These costs may be included in the model formulation through an economic penalty representing maintenance costs per used convoy in the objective function. This cost is represented by a parameter mc_m which depends on the material type. The complete term in the objective function representing daily maintenance costs is as follows: $\sum_{s,m,c \in S,M,C} mc_m o_c yt_{s,t_i,m,c}$. It is the sum of the number of convoys of each material in every station at the beginning of the planning period multiplied by the daily maintenance cost.

In Table 7 the obtained results are presented. For each study case, the RT&RSM with maintenance costs solution is listed first. Secondly the RT&RSM solution. As maintenance

Table 7 Lines C5 and C3&C4 RT&RS model with and without maintenance costs comparison

Solution	#C	TSOC	EMOC	PEC	#CC
Line C5 RT&RS with maintenance costs	54	76690.32	1833.44	3069	28
Line C5 RT&RS	57	75792.24	1892.72	2464	8
Line C3&C4 RT&RS with maintenance costs	43	84674.94	1242.93	2207	14
Line C3&C4 RT&RS	48	81865.91	2086.29	1372	12

Table 8 Comparing non-robust and robust solutions

Case	TSOC	EMOC	#EMRH	PEC	#CC	#CCRH
Line C5 NoRob	75792.24	1892.72	15	2464	8	3
Line C5 Rob	76966.72	650.64	10	3311	6	2
Lines C3&C4 NoRob	81865.91	2086.29	9	1372	12	8
Lines C3&C4 Rob	82847.21	1816.62	1	1658	8	5

costs depending on the number of convoys used have been included, the number of convoys used is lower than in the case where they are not. However, TSOC are increased; this may be due to the fact that as fewer convoys are available, capacity cannot be adjusted as in the case where more convoys were available, so more kilometers must be rolled by the fleet. PEC present an expected behavior because as capacity is not so well adjusted they go up. Finally, in order to allow using a smaller fleet size more composition changes are performed.

5.5 Robust solutions

Robustness parameters are obtained from operators. For example, in line C5, the Atocha depot station is shared among more than 5 different lines. This causes the capacity of C5 material to vary strongly during the planning period. Robustness parameters are chosen to try to avoid (if possible) composition changes and empty movements with the destination Atocha.

As we have stated above, robustness may be introduced with different approaches. Some of these approaches could prevent dangerous empty movements and composition changes. Dangerous empty movements and composition changes are recognized as those occurring during rush hours. To illustrate this effect, two different cases are shown. In the first case, no robustness is introduced (NoRob) (i.e., $\theta_{s,s',t} = 1$ and $\vartheta_{s,t}$ is equal to its nominal value for all possible cases). In the second case, robustness is introduced (Rob), doubling the values of these parameters at rush hours from 7:00 a.m. to 10:00 a.m. and from 2:00 p.m. to 5:00 p.m. The computational results are shown in Table 8. In both cases, the rest of the parameters are those used in the previous examples.

We can see the differences between both cases in Table 8. In the second column, the train service operating costs (TSOC) are shown. In the robust case, these costs are greater because more train services are set to double composition; this cost may represent the robustness cost. In the third column, the empty movement costs (EMOC) are shown, which are lower than those of the no robustness case. If we pay attention to empty movements during rush hours (#EMRH), we can see that the number of empty movements is reduced, achieving one of our goals. In a similar way, robustness is introduced in the solution by avoiding composition changes during rush hours (#CCRH). However, introducing robustness into the

Table 9 Comparing robust and swapping robustness solutions for lines C5 and C3&C4

Case	TSOC	EMOC	#EMRH	PEC	#CC	#CCRH
Line C5 Rob	76966.72	650.64	10	3311	6	2
Line C5 S-Rob	77302.08	646.24	3	2728	12	3
Line C3&C4 Rob	82847.21	1816.62	1	1658	8	5
Line C3&C4 S-Rob	82697.69	2884.74	11	1225	10	6

model makes passengers in excess costs (PEC) to rise up. In this case, robustness costs remain partially on passengers.

Another way of introducing robustness into the model is to ensure swapping operations. The model is obeyed to provide solutions with free parkings in depot stations at lines' edges. In this way, the obtained solution will be more resistant to possible delays because they may be alleviated by swapping trains. This robustness (S-Rob) may be achieved by reducing strategic depot stations capacity in order to always allocate some convoys to perform swapping operations.

We compare robust (Rob) and swapping robust (S-Rob) solutions in Table 9. Robust solutions are those ones obtained applying the previous explained approach. Swapping robust solutions are those ones obtained by ensuring capacity in certain depot stations in order to perform swapping operations if necessary. As we can see, TSOC are showed in the second column. EMOC in the third one. #EMRH, in the fourth column, are strongly reduced for Line C5 but they are strongly increased for Lines C3&C4. These two opposite solutions mean that in Line C5 there is enough capacity to continue operating even when we reduce capacity in order to ensure swapping operations. However, for Lines C3&C4 this is not true, capacity is not enough and many empty movements must be done during rush hours. In this way, we are gaining robustness through swapping opportunities but we are also getting worse robustness through empty movements: a trade off must be selected depending on the available infrastructure. In the next column, PEC are showed. They are reduced due to the fact that more trains are set to double composition, and that more composition changes are needed; both facts make to fit better capacity to passenger flows. Composition changes (#CC) are increased because capacity for the S-Rob case is lower. Similarly, composition changes during rush hours (#CCRH) are increased.

6 Conclusions

In the paper an integration of the railway timetabling and rolling stock assignment problem is formulated in the emerging area of urban rapid transit networks. The planned frequencies are known from the railway line planning problem. However, the proposed model updates frequencies providing a more efficient network operation.

Services constraints have been included ensuring that frequencies for some arcs are maintained in a determined frequency window. Moreover, these constraints also ensure that the headway is maintained for every train line in the network. In the passenger constraints for the demand, passengers in excess and unattended passengers are considered; we include different capacities depending on standing passengers' density. Finally, in the rolling stock constraints shunting operations, fleet and depot capacities, and symmetry constraints have been included.

In the model operating and leasing costs are minimized jointly with penalties for robustness. Robustness has been defined through penalizing difficult shunting operations. Another robust approach has also been included: ensuring parkings in strategic depot stations in order to perform swapping operations. Maintenance costs have been included adding to the objective function maintenance daily costs for each material and convoy.

RENFE data have been used for the computational tests for the rapid transit network in Madrid. The presented service network design model considers the possible train lines in that network. The possible train lines are defined as those services between depot stations departing during each time period. Among them, the model chooses the optimal services (i.e.: the number of scheduled services are about 370 and 410 for Lines C5 and C3&C4, respectively). Computational results show how the rolling stock assignment depends on the efficiency of the railway timetable. We do not only change the timetable; also updating frequency values we introduce a greater freedom. That is, comparing with the rolling stock assignment with fixed timetable, we can see how better solutions can be obtained improving the overall solution efficiency and getting a greater robustness degree. We have also studied the possibility of minimizing maintenance costs for every material. However, using this approach, solutions with a lower degree of robustness are obtained because more composition changes are needed, which deteriorates the network operation.

The proposed model treats passenger demand from a centralized point of view. We represent passenger demand as a passenger flow through arcs. In order to decide the new timetable, passenger flow is aggregated in time, that is, every flow through a determined arc during a certain time interval is aggregated into the same time period. We have tried different time intervals' lengths to aggregate passenger flows showing that it must be longer than trains' headway. However, as this time length is increased, passengers' behavior is worsen in the model formulation, so a tradeoff time length must be selected.

A natural extension of this research is the recoverability of the train system in case of disruptions. In these cases, the modification of the train timetabling, the cancellation and the outfit of new services is studied during the disrupted time period, trying to minimize rolling stock deviations from the original planning and minimizing passengers' dissatisfaction with the new planning.

References

Alfieri, A., Groot, R., Kroon, L. G., & Schrijver, A. (2006). Efficient circulation of railway rolling stock. *Transportation Science*, *40*, 378–391.

Brännlund, U., Lindberg, P. O., Nou, A., & Nilsson, J. E. (1998). Railway timetabling using Lagrangian relaxation. *Transportation Science*, *32*, 358–369.

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

Cacchiani, V., Caprara, A., & Toth, P. (2010). Scheduling extra freight trains on railway networks. *Transportation Research. Part B: Methodological*, *44*(2), 215–231.

Cadarso, L., & Marín, A. (2011). Robust rolling stock in rapid transit networks. *Computers & Operations Research*, *38*(8), 1131–1142.

Caprara, A., Fischetti, M., & Toth, P. (2002). Modeling and solving the train timetabling problem. *Operations Research*, *50*, 851–861.

Caprara, A., Monaci, M., Toth, P., & Guida, P. L. (2006). A Lagrangian heuristic approach to real-world train timetabling problems. *Discrete Applied Mathematics*, *154*, 738–753.

Caprara, A., Kroon, L., Monaci, M., Peeters, M., & Toth, P. (2007). Passenger railway optimization. In C. Barnhart & G. Laporte (Eds.), *Handbook in operations research and management science* (vol. *14*). Elsevier: Amsterdam.

Carraresi, P., Malucelli, F., & Pallottino, S. (1996). Regional mass transit assignment with resource constraints. *Transportation Research. Part B: Methodological*, *30*(2), 81–98.

Cordeau, J. F., Soumis, F., & Desrosiers, J. (2000). A Benders decomposition approach for the locomotive and car assignment problem. *Transportation Science*, *34*, 133–149.

Cordeau, J. F., Desaulniers, G., Lingaya, N., Soumis, F., & Desrosiers, J. (2001). Simultaneous locomotive and car assignment at VIA Rail Canada. *Transportation Research. Part B: Methodological*, *35*, 767–787.

Fioole, P. J., Kroon, L. G., Maróti, G., & Schrijver, A. (2006). A rolling stock circulation model for combining and splitting of passenger trains. *European Journal of Operational Research*, *174*(2), 1281–1297.

Kroon, L. G., & Peeters, L. W. P. (2003). A variable trip time model for cyclic railway timetabling. *Transportation Science*, *37*, 198–212.

Liebchen, C., & Möhring, R. H. (2008). The modeling power of the periodic event scheduling problem: railway timetables—and beyond. In *Lecture notes in economics and mathematical systems: Vol. 600 (II). Computer-aided Systems in Public Transport* (pp. 117–150).

Maróti, G. (2006). *Operations research models for railway rolling stock planning*. PhD thesis, Eindhoven, Technische Universiteit Eindhoven.

Mellouli, T., & Suhl, L. (2007). Rotation planning of locomotive and carriage groups with shared capacities. In F. Geraets, L. G. Kroon, A. Schöbel, D. Wagner, & C. Zaroliagis (Eds.), *Algorithmic methods for railway optimization*. Berlin: Springer.

Nachtingall, K. (1996). Periodic network optimization with different arc frequencies. *Discrete Applied Mathematics*, *69*, 1–17.

Nachtingall, K., & Voget, S. (1997). Minimizing waiting times in integrated fixed interval timetables by upgrading railway tracks. *European Journal of Operational Research*, *103*, 610–627.

Serafini, P., & Ukovich, W. (1989). A mathematical model for periodic event scheduling problems. *SIAM Journal on Discrete Mathematics*, *2*, 550–581.