An optimization model for line planning and timetabling in automated urban metro subway networks

Víctor Blanco, Eduardo Conde, Yolanda Hinojosa, and Justo Puerto

ABSTRACT. In this paper we present a Mixed Integer Nonlinear Programming model that we developed as part of a pilot study requested by the R&D company METROLAB®[1] in order to design tools for finding solutions for line planning and timetable situations in automated urban metro subway networks. Our model incorporates important factors in public transportation systems from both, a cost-oriented and a passenger-oriented perspective, as time-dependent demands, interchange stations, short-turns and technical features of the trains in use. The incoming flows of passengers are modeled by means of piecewise linear demand functions which are parameterized in terms of arrival rates and bulk arrivals. Decisions about frequencies, train capacities, short-turning and timetables for a given planning horizon are jointly integrated to be optimized in our model. Finally, a novel Math-Heuristic approach is proposed to solve the problem. The results of extensive computational experiments are reported to show its applicability and effectiveness to handle real-world subway networks.

1. INTRODUCTION

In this paper, we propose a model for line planning and timetabling on general urban subway transportation systems. This study was originated by a real-world problem proposed by Metrolab®, a French R&D company, dealing with the line planning and timetabling of trains of existing subway networks. It was a pilot experience to automatize the decision making process, at the tactical and operational level, of a small section of the Paris subway network.

The development of flexible tools to control, automatically, a transportation system according to a set of indicators of its service quality may have a considerable impact in its efficiency and usefulness. The quality perception of a public transportation system from a customer's point of view is highly dependent of its reliability, comfortability and effectiveness when comparing with alternative transportation means. If just poor-quality connections are offered or the quality-price relationship does not fulfil the passengers' expectations, they may decide to use alternative transportation means. Therefore, the quality of a public transportation system, from the passengers' point of view, is a key objective in its design and management, besides infrastructure constraints, operational limitations or budget considerations [12].

The existing literature on line planning is very extensive (see e.g., [11, 12, 25] and the references therein), including different models which can be classified according to the decisions covered (determination of train routes, frequency setting, or both), infrastructure and operational aspects, objective functions and the way in which passengers' decisions are taken into account in the decision making process. For instance, in [12], line planning models are classified with respect to their objective functions into models with cost-oriented or with passenger-oriented objective functions. Our model will consider both points of view in an attempt to find an equilibrium

between these conflicting objectives which, as mentioned in [14], is an important challenge in a public transportation system.

However, building an effective model is much more complex than selecting the appropriate nature of the objective function. The correct delimitations of the considered features in the context of the transportation system or in the set of customers served by this system is a very important phase in the actual modeling. Often, an initial line planing must be re-engineered motivated by changes on costumers' flows induced by changes in *passengers' route choices*, [12]. This gives rise to a *bilevel optimization problem* with a line planning problem on the upper level and a passenger's route choice problem on the lower level. Moreover, the existence of several decision-makers is not the only difficulty of this problem. There exist uncertainties in the number of passengers that must be served (demands), in the origin-destination pairs that customers want to go across or in the times needed to cover the network links due, for instance, to machine failures or other incidents. In this situation it seems hard to integrate all these elements in a suitable optimization model.

Usually, some simplifications must be assumed in order to obtain operational solutions for realistic situations. The usefulness of the model will be strongly conditioned by these assumptions. Having a valid model for a wide range of scenarios might be a lofty target but one worth aiming for. The resulting approximation can be seen not only as a model to optimize the existing resources in a given transportation system but also as a *what-if* tool to make rational decisions. For instance, it can be used to check the implementation of possible modifications in the structure of the existing network (adding stations, new connections, ...) or in the conditions (demand variation, service disruptions, ...) under which the transportation system is working at present.

Line planning is only one of the planning process's stages. Indeed, following [11, 21, 26], the planning process in public transportation includes several phases that usually are sequentially executed in the following order:

(1) **network design**, where the stations, links and routes of the lines are established,
(2) **line planning**, specifying the frequency and the capacity of the vehicles used in each line (*line concept*, [12]),
(3) **timetabling**, defining the arrival/departure times and
(4) **scheduling**, in which vehicles and/or crews are planned.

The first phase, namely network design, is done at the strategic level and implies a high cost (see e.g., [3]). Moreover, the remaining steps involve decisions at the tactical and operational levels which are conditioned by that design. Thus, it seems appropriate to assess different designs by means of a *what-if analysis* based on the efficiency of the system under a given scenario. In that case, after initializing with a *reasonable* design, the procedure could *optimize* the efficiency of the transportation system using tactical/operational decisions. This may reveal possible weaknesses of the current design and, after fixing some of them, will give rise to a new design. The process could be repeated until a compromise transportation design is found.

In addition to the literature on line planning commented above, one may also find a rich literature on timetabling and scheduling (see e. g. [1, 7, 8, 19, 27] for timetabling models and [6, 10, 28] for scheduling models). Timetabling models are usually classified according to the capacity of the transportation system or the requirements of passengers. Regarding to the scheduling literature, models can be classified into single and multiple depot frameworks [4]. Besides, we can find periodic and time-dependent timetabling and scheduling models and many other variants depending on the considered constraints. Considerable effort has also been made to analyze the practical effects of the different elements or features covered by these models. For instance, as mentioned in [22], periodic timetables easily allow passengers to remember the exact departure times at stations but, in general, they are not fully sensitive to time-varying passenger

demands, which could result in long waiting times and reduced service reliability, particularly under irregular over-saturated conditions.

Once the framework of the transportation system has been delimited, the resulting model should be optimized. However, as pointed out in [26], going through the above-mentioned stages of the planning process in a sequential way, often leads just to suboptimal solutions. Recent research (see [13, 14, 18, 24, 26]) is increasingly oriented towards **integrated planning** in which two or even more of these planning stages are simultaneously addressed. These integrated optimization models are frequently superior to those optimizing sequentially the considered stages, as it has been recognized in the literature (see e.g., [26] and the references therein).

In this paper we propose a new integrated model in which line planning and timetabling are simultaneously optimized using a combination of cost- and passenger-oriented objective function. In our model, time-dependency on demands is considered in addition to two other elements that, as far as we know, have only been addressed separately in the literature: short-turns and interchange stations.

Short-turning is a tactical decision for which some trains can perform short cycles in order to increase frequency in specific sections of a line. In general, due to their analytical complexity, the approaches to manage short-turnings in the context of railways planning are based on particular cases and there are no general models that can be applied without modifications to every situation [5]. Besides, most of the literature analyzing short-turning in a railway context is limited to a single two-way transit line [5, 9, 20, 29]. Our model incorporates decisions concerning the activation of short-turns simultaneously in several lines of the network.

On the other hand, *interchange* or *transfer stations* are shared by several lines in the system allowing passengers to change from one to another line. Papers dealing with interchange stations (see e.g. [15, 16, 30]) usually aim for minimizing the total transfer waiting time of passengers by synchronizing train arrival times at transfer stations. We will manage here the effective flows of passengers at the interchange stations and compute the corresponding effects both in the quality of the service and in technical constraints, as those related to the capacities of the trains.

The final aim is to model a real system, inspired by one initially proposed by Metrolab®, using its more relevant features whilst its computational tractability is preserved. This is an important challenge taking into account the combinatorial, stochastic, multilevel and multiobjective nature of the problem. The resulting outcome is a Mixed Integer Second Order Cone Programming model, which can be solved using off-the-shelf optimization solvers, but only for limited sizes. For larger sizes we propose a Math-Heuristic approach in which the system is decoupled into different lines. Afterward, each subsystem is optimized individually but including in the input data the flows of passengers generated after optimizing other lines. The process is repeated like in a *block coordinate descent* procedure (see e.g., [2]) until a given stoping rule is verified.

The remainder of this paper is organized as follows. Section 2 deals with a detailed description of the optimization problem and its main elements. In Section 3 we present the Mathematical Programming formulation of the problem. The demand function modelling how the flow of passengers entering into certain stations of a line changes according to the effects of the other lines or due to external factors is detailed in Section 4. The usefulness of the proposed model is illustrated in Section 5 with a case study using real data on a section of the Paris subway provided by Metrolab®. Our Math-Heuristic approach is proposed in Section 6 and the corresponding computational results, including its comparison with the exact method, are reported in Section 7 using several network topologies adapted from the literature. The paper ends with a section where some conclusions and future extensions of the model are outlined.

2. Problem description

Let us assume that the technical features of a public transportation network, which is a part of a complex underground train system, are known. Our goal is to model the problem of how to operate different metro lines on this network according to a set of technical requirements and a given structure of the demand requesting for this service. Suppose that a set of routes for the potential lines (*lines pool*) and a set of interchange stations are specified. Furthermore, some other factors involved in the system performance such as passenger flows, set of possible train capacities, maximum number of allowed trips in a given line during the planning horizon, demand fluctuation (e.g., rush/off-peak hours) or stopping time windows, amongst others, are assumed to be also available. The goal is to model such a system using its more relevant features whilst its computational tractability is preserved. In the description of our approach we will consider three main blocks: input data, feasible actions and assessment of a particular solution together with some additional specifications of our model.

2.1. **Input data.** In addition to the input network, including the topological route map and the interchange stations, the main block of input data corresponds to the passenger flow amongst the considered set of stations. Obviously, the number of passengers awaiting in a specific station for the next train which connects to a given destination is a stochastic process. Furthermore, the stochastic processes corresponding to the set of considered stations are interdependent due to the flow relationships amongst the stations which, in addition, may change over time. In our model, given that we deal with a heavily congested subway line, these stochastic processes will be replaced by average rates of the number of passengers. The dynamic dependence of these processes on time should be preserved in some way in the optimization model since it is one of the relevant elements in order to obtain realistic operating solutions. We will do it through a function measuring the intensity of the demand.

In our framework, we assume that two main situations affect to the passengers flows: transitions between stations or lines and arrival of external demand functions. The first one refers to the behaviour of the passengers with respect to the mobility pattern. These data fix an assignment for the destination of the passengers catching a train in a given station, what is known in the literature as *line planning with route assignment*, [12]. The alternative approach of line planning with route choice seems to be more appropriate just in those transportation systems with high density of connections (with alternative paths between two given locations) and low trip frequency. However, this is not the case in our model since these two features rarely appear in underground train systems.

We will use an Origin-Destination (OD)-matrix per line having as entries the proportions of passengers moving between pairs of stations of the line and also a set of values quantifying the proportion of passengers which want to change from one metro line to another in each one of the transfer stations. These proportions may be considered as estimations of the probabilities with which a passenger moves through the network and could be dependent on the dynamic nature of the transportation system. Following [26], in order to model the passengers' flows, OD-data gives rise to more realistic applications than those based on traffic loads since the paths followed by passengers depend strongly on the line concept. As observed by several authors ([12, 26]), the optimization models derived from the management of OD-data are often harder to be solved numerically, and thus, approximated ad-hoc algorithms need to be used to deal with problems of realistic sizes.

The second concept, the arrival of external demand functions, refers to the intensity of use of the transportation network. The external demand models the incoming flow of passengers entering to the system from outside during the planning horizon. These functions determine the relative importance, in the overall planning cost, of the stations used to access the system and

it may change depending on time. Also, rush hours at given time periods, irregular weather conditions, or the celebration of events at certain places close to stations may provoke increasing or decreasing of the incoming flow of passengers taken into account in our model.

These external demand functions, together with the OD-data corresponding to movements between pairs of stations and the proportions of transfer passengers, give rise to a model having time-dependent passenger flows. Time-dependent flows are an appropriate feature of a realistic approach for traffic planning and, as pointed out in [26], at present, there is not much research literature covering integrated optimization planning models under these conditions.

2.2. **Feasible actions.** In our model, the line concept design is specified by choosing the operating frequencies and the train capacities for each line considered in the lines pool. Furthermore, our line concept design allows *short-turning* in some lines, i.e., the possibility of activating, for some or all the trains, and at certain time periods, short cycles, in order to increase the frequency in specific (consecutive) stations suffering from intensive demand. This situation is typical in lines which connect distant residential areas with the city center or economic centers.

In the integrated planning model, the line concept design is optimized together with their corresponding timetable. Both elements define the feasible solutions of the problem once a set of technical constraints, involving passenger flows and leaving/arrival times, is specified.

The line concept design will be the main source of discrete decision variables for the formulation of our problem. As, for instance, the selection, among a finite set of capacities for the trains. On the contrary, the actual number of trips of a given line will be modeled using a finite set of replicas of continuous variables corresponding to potential timetables. On the basis of these decision variables a number of additional auxiliary variables will be considered in the optimization model in order to control the times in which different events happen at each station during the planning horizon. In our model, unlike most of the approaches deriving optimal timetables in transportation systems, the departure times are not discretized and the period of time elapsing between consecutive train departures is not constrained to be constant. Thus, it provides more flexible decisions as well as timetables sensitive to the changeable conditions in the passengers's flow during the planning horizon, turning out, in general, in an aperiodic timetabling.

Different train speeds are not taken into account in our line concept design because the pilot proposal by Metrolab® only considered constant and fixed speeds between each pair of consecutive stations. Nevertheless, continuous variables modeling the speed of a train between two consecutive stations could be *easily* added to our model as explained in Remark 1.

2.3. **Assessing a planning solution.** As commented above, due to the multi-objective nature of the problem, one of the most difficult modeling issues is that of assessing a feasible solution (line concept+timetable). Maintenance/operational planning costs are usually easy to handle as a part of the objective function. However, in order to consider also the passenger-oriented nature of the objective, the cost induced by the quality of the service provided by the system should be included in the objective function. This will be incorporated into the model quantifying the cost of unmet (non-served) demand, that is, passengers who cannot take a train due to lack of capacity. Non-served passengers contribute a given amount, in terms of costs, due to their confidence loss in the transportation system, their balking rate or a combination of these two and some other factors. Obviously, calibrating these costs is not an easy task, but it may be partially handled by managing a finite set of cost estimates and solving the problem for each one of them in order to evaluate the influence of these hard-to-calibrate parameters in the proposed solution.

2.4. **Specifications of our model.** In the following we list the assumptions that are imposed to derive a suitable Mathematical Programming formulation for our model.

***Planning horizon*::** Our model considers a continuous time interval in which all the events must start and decisions occur. The range of this interval depends on the data collection accuracy and with-in-day variability of the demand changes. This planning horizon is fixed a priori but, as mentioned in Remark 2, our model allows to join two consecutive planning horizons by passing data about numbers of passengers and arrival/departure times obtained from an optimal solution on a given planning period as input data for the next one.

***One direction trips*::** We assume that each line in the pool operates only in one direction, from a given line-header station to the final one. Usual round-trip lines are modeled as two symmetric lines by interchanging the order of the stations. A *trip* consists in making the complete walk along all the ordered stations of a given line, from the head to the final station. Thus, a physical round-trip starting and ending at the same station will be given by two lines sharing the stations but in the opposite order.

***Short-turns*::** We consider that specific sections of consecutive stations in certain lines are allowed to be activated in some of the trips.

***Interchange stations*::** We consider that the lines in the lines pool may share common interchange stations where some of the passengers change of line to go to their final destination.

***Train capacities*::** The model assumes that a finite set of admissible capacities for the trains operating a line is given.

***Safety interval*::** A minimum security time window between consecutive trips in any line is established.

***Maximum number of trips*::** We assume, w.l.o.g., that the maximum number of possible trips in each line is given beforehand. Note that an upper bound of this maximum number can be obtained taking into account the range of the planning horizon and the safety interval. In our formulation, we resort to a set of decision variables controlling the time in which different events happen. These variables must be replicated as many times as the maximum number of trips, although only some of those *trip-variables* are activated and then represent actual trips. Hence, the size of the formulation strongly depends on this maximum number. The idea is to consider that the potentially variable number of trips of the line concept is fixed to the maximum number although, in fact, some of them are really *fake trips*. This trick will ease the task of building constraints to ensure non-overlapping events and the safety interval between consecutive trips in the stations of a given line.

***Piecewise linear cumulative incoming demand*::** We model the accumulated number of passengers arriving to each station up to a given instant during the planning horizon by the so-called *demand function*. With this function we manage variable arrival rates during the planning horizon and bulk arrivals due to special events, like for instance the end of a football match in a close location, the arrival of passengers coming from another transportation mean or, in the interchange stations, the arrival of passengers coming from another line of the subway network. As proposed by Metrolab® for its pilot experience, we consider that this function is a piecewise linear function of time (further details are given in Section 4).

Figure 1 illustrates some of the considered features of the networks under study. There, we represent by circles or squares the nodes corresponding to the stations of two subway lines. An interchange station common to two lines is marked with a black square. We have also included a possible short-turn in the *horizontal* line covering a set of four consecutive squared stations (drawn as a dashed line in the picture).
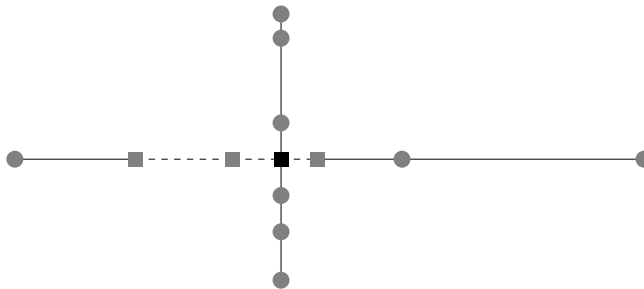
FIGURE 1. Representation of a sample network in our framework.

3. MINLP FORMULATION

In this section we provide a Mathematical Programming formulation for the problem described in Section 2. First of all, we are given an input network, like the one depicted in Figure 1, including the topological route map and the stations, some of them being interchange stations and the lines pool L defined over this network. In addition, short-turning decisions are allowed in some lines of the lines pool. These decisions always concern a set of given consecutive stations of those lines. In what follows and when no confusion is possible, we will refer to both, the set of consecutive stations and the trip concerned by a short-turning decision as a *short-turn*. On the other hand, trips trough the full-length lines will be referred as *whole trips*. In order to introduce the model we start by defining the set of *parameters* describing the remainder of the input data as well as the *decision variables* used to identify a feasible solution.

3.1. **Parameters.** The input parameters for our model are the following:

- $[0, T]$: Planning horizon in which trains start their journeys at a given head of line station. Note that a train may start its last journey on $[0, T]$ while its last stop may occur in a time instant $t > T$.
- L: Set of lines in the network (lines pool). As mentioned above, round-trip lines are considered as two different lines sharing the stations but traversed in opposite directions (rigorously speaking, the stations represent platforms of the corresponding line). Each one of the lines is assumed to be described by its node stations and its directed connections between consecutive stations. Additionally we will denote by LS the set of lines containing a short-turn and by LNS the remainder (those in which none of its proper subsets of stations can be activated as short-turns). Clearly, $L = LS \cup LNS$.
- $N_\ell = \{1, \ldots, n_\ell\}$: Set of stations of line $\ell \in L$. Stations are assumed to be ordered in its travelling direction. Observe that if the lines ℓ and ℓ' correspond to the same round-trip line but traversed in opposite directions they have the same number of stations ($n_\ell = n_{\ell'}$) and station $i \in \ell$ represents the opposite platform to station $n_{\ell'} - i + 1 \in \ell'$.

  Additionally, for each line $\ell \in LS$ we will denote by S_ℓ the set of stations in the (unique) short-turn, being 1_{S_ℓ} the first station of the short-turn and n_{S_ℓ} the last one.

  In order to present a clearer Mathematical Programming model, we will also denote from now on by $\overline{S} = \left\{ (i, \ell) : i \in N_\ell, \ell \in LNS \right\} \cup \left\{ (i, \ell) : i \in N_\ell \backslash S_\ell, \ell \in LS \right\}$, i.e., those stations which are not part of the available short-turns.

- d_i^ℓ: Distance (measured as travel time) between the stations i and $i+1$ of line $\ell \in L$. Recall, as mentioned in Section 2, that we assume that the speed of the trains operating between consecutive station i and $i+1$ is fixed, and thus, this distance is trip independent.

- e_i^ℓ: Stopping time for any train at station i of line ℓ before leaving to station $i+1$. This value represents a time window to perform different operations as the unload/load of passengers from/to the train or the fine-tuning of the train. Without loss of generality, we also consider that the stopping times are trip independent.
- $t_{1 \mapsto 1_{S_l}}$: Distance (measured as travel time plus stopping time at the last and intermediate stations) between the head of line station and the first station of the short-turn for line $l \in LS$. Note that, this parameter is given by the following expression:

$$t_{1 \mapsto 1_{S_l}} = \sum_{r=1}^{1_{S_l}-1} (d_r^l + e_{r+1}^\ell). \tag{1}$$

- IS^ℓ: Minimum safety time interval between consecutive trips in a given line $\ell \in L$.
- $K_\ell = \{1, \ldots, \bar{k}_\ell\}$: Set of trips in line $\ell \in L$. It is worth noting that the exact number of trips in a line is a decision of the line planning process and which is not known beforehand. In fact, some trips will not be actually used on the planning. We will refer to them as *fake trips*. Note also that the maximum number of possible trips in the line $\ell \in L$ can always be upper bounded by $\left\lceil \frac{T}{IS^\ell} \right\rceil + 1$. Typically the value $\bar{k}_\ell$ will be smaller than this bound and should be estimated on the basis of the technical specifications of the network.
- $Q = \{q_1, \ldots, q_{|Q|}\}$: Admissible capacities for trains operating in all the lines. In some cases, a single capacity profile is allowed while in some others a base capacity is available and, by adding extra wagons, it can be doubled, triplicated, etc.

We also need the proportion of passengers, which, not being able to catch the train in a given attempt, still insist on using the system and wait for the next train. In addition, our model corresponds to a line planning with route assignment in which the OD-matrix, whose entries describe the proportions of passengers moving among stations of each line is given.

- α: Proportion of passengers that cannot get on a train because of lack of capacity and await for the next one (proportion of persisting passengers). This parameter is assumed to be independent of stations and lines. It can also model the probability that a passenger gives up from getting on a crowded train and awaits for the next one.
- p_{ij}^ℓ: Proportion of passengers awaiting at station i to go to the station j using the line $\ell \in L$. Note that this parameter can assume positive values only if $i < j$, i.e., if station i is previous to station j.

The following parameters will be used to assess a given planning solution. As explained in Section 2, we propose the aggregation of the maintenance/operational planning costs together with a *quantification* of the social cost whose purpose is to measure the quality of the service provided by the system. Obviously, the correct estimation of these costs is fundamental for a meaningful model.

- b_q^ℓ: Fixed cost for starting a whole trip at line $\ell \in L$ with capacity $q \in Q$. Usually, the larger the capacities, the higher the fixed costs on the trips. These costs model the consumption of resources to prepare a train for a new trip, the energy spent in the trip, the fixed costs of the staff needed to control the train, etc.
- b_{Sq}^ℓ: Fixed cost for starting a short-turn at line $\ell \in LS$ with capacity $q \in Q$. As in the previous cost, larger capacities and lengths of the short-turns usually involve higher costs on the trips. These fixed costs are usually smaller than the corresponding ones for whole trips.
- γ_{ij}^ℓ: Unitary profit for transporting a passenger from the station i to the station j of the line $\ell \in L$. It represent the cost of the ticket paid by a single passenger to use the service for a given origin-destination trip.

- μ_1: Unitary penalty for passengers who cannot get on the first arriving train due to its limited capacity and still insist on using the system. This parameter is an estimation of the loss supported by the service for unsatisfied passengers that still keep waiting to use the subway service.
- μ_2: Unitary penalty for passengers who give up using the network and leave the system after they cannot get on the first arriving train due to its limited capacity. This parameter is an estimation of the loss supported by the service for unsatisfied passengers that are lost by the system.

The above set of parameters are listed in a more compact form in Table 1

Parameter	Description
$[0, T]$	Planning horizon.
L	Lines of the network (Lines pool).
LS	Lines enabling short-turns.
LNS	Lines without short-turns.
$N_\ell = \{1, \ldots, n_\ell\}$	Stations of line $\ell \in L$.
$S_\ell = \{1_{S_\ell}, \ldots, n_{S_\ell}\}$	Stations in the short-turn of line $\ell \in LS$.
$\bar{\mathcal{S}}$	Stations not affected by short-turns.
d_i^ℓ	Distance (measured as time) between stations i and $i+1$ of line $\ell \in L$.
e_i^ℓ	Stopping time that a train spends in the station i of the line $\ell \in L$.
$t_{1 \mapsto 1_{S_l}}$	Distance (measured as time) between the head of line station and the first station of the short-turn for line $\ell \in LS$.
IS^ℓ	Minimum safety time interval between consecutive trips, for line $\ell \in L$.
$K_\ell = \{1, \ldots, \bar{k}_\ell\}$	Set of trips in line $\ell \in L$, some of them being *fake trips*.
$Q = \{q_1, \ldots, q_{|Q|}\}$	Admissible capacities for trains operating in all the lines.
p_{ij}^ℓ	Proportion of passengers moving between stations i and j of line $\ell \in L$.
α	Proportion of persisting passengers.
b_q^ℓ	Fixed cost for starting a whole trip at line $\ell \in L$ with capacity $q \in Q$.
b_{Sq}^ℓ	Fixed cost for starting a short-turn at line $\ell \in LS$ with capacity $q \in Q$.
γ_{ij}^ℓ	Unitary profit for transporting a passenger from the station i to the station j of line $\ell \in L$.
μ_1	Unitary penalty for persisting passengers.
μ_2	Unitary penalty for passengers who give up using the system.

TABLE 1. Parameters of the model.

3.2. **Decision variables.** The following set of decision variables are used in our Mathematical Programming model:

- **Continuous Variables:**
  - $t_1^{k\ell}$: Departure time from the initial station of line $\ell \in L$ at its k-th trip, $k \in K_\ell$.

    Since the travel time between consecutive stations and the stopping time at each station are fixed, the departure time from the initial station of a line will be the reference time to calculate the departure time from the rest of stations of the line.

  It is worth mentioned that we will force any fake trip to operate on the line at the same time as the previous true trip. Thus, the departure time from the initial station of the line of a fake trip must coincide with the departure time from the initial station of the previous true trip and therefore, the departure time from the rest of stations must also coincide. Note also that if $\ell \in LS$ some of its trips may be short-turns. When a short-turn is activated the stations not affected by the short-turn are considered as

nodes of a fake trip, being the departure times from these stations fixed to the ones of the previous true whole trip. In order to avoid infeasible solutions due to inconsistent times relating departure times in consecutive stations (in and out the short-turn) the following continuous variables are needed.

  − $w^{k\ell}$: Difference between the actual departure time from the first station of the short-turn of the k-th trip, $k \in K_\ell$, of line $\ell \in LS$ and the time when it should depart from this station taking into account its departure time from the initial station of the line. Note that if the k-th trip is a whole trip then, $w^{k\ell} = 0$.

  Note also that in the definition of these variables we are implicitly assuming that the initial station of the line is not part of the short-turn. If this occurs, we need to take as reference time the departure time from any other station out of the short-turn.

For modeling purpose we need to distinguish between the flow of passengers that get on a train that performs a whole trip and the flow of passengers that get on a train that only performs a short-turn. Note that the first one is defined for any line of the lines pool whilst the second is only defined for those lines containing a short-turn.

  − $f_i^{k\ell}$: Flow of passengers captured in the station $i \in N_\ell$ by the train that covers the k-th trip of the line $\ell \in L$, being $k \in K_\ell$ a whole trip.

  − $g_i^{k\ell}$: Flow of passengers captured in the station $i \in S_\ell \backslash \{n_{S_\ell}\}$ by the train that covers the k-th trip of the line $\ell \in LS$, being $k \in K_\ell$ a short-turn.

Note that the k-th trip of a line $\ell \in LS$ is either, a whole trip or a short-turn. Thus, if $g_i^{k\ell} > 0$ then, $f_i^{k\ell} = 0$ and viceversa. Furthermore, if the k-th trip is a fake trip for line $\ell \in LS$ (resp. $\ell \in LNS$), then $g_i^{k\ell} = 0$, for all $i \in S_\ell \backslash \{n_{S_\ell}\}$ and $f_i^{k\ell} = 0$ for all $i \in N_\ell$ (resp. $f_i^{k\ell} = 0$ for all $i \in N_\ell$).

- **Binary Variables:**
  − The decisions about the capacities of the trains at each trip are modeled using the following variables:

$$y_q^{k\ell} = \begin{cases} 1 & \text{if the } k\text{-th trip of line } \ell \text{ is a whole trip} \\ & \text{with capacity } q \\ 0 & \text{otherwise,} \end{cases} \qquad k \in K_\ell, \, q \in Q, \, \ell \in L.$$

$$y_{Sq}^{k\ell} = \begin{cases} 1 & \text{if the } k\text{-th trip of line } \ell \text{ traverses every} \\ & \text{station of the short-turn } S_\ell \text{ with capacity } q \\ 0 & \text{otherwise,} \end{cases} \qquad k \in K_\ell, \, q \in Q, \, \ell \in LS.$$

Note that if the k-th trip of line $\ell \in LS$ is a short-turn with capacity q then, $y_{Sq}^{k\ell} = 1$ and $y_q^{k\ell} = 0$. On the other hand, if the k-th trip of line $\ell \in LS$ is a whole trip with capacity q ($y_q^{k\ell} = 1$), the train traverses all the stations of the line and, in particular, the stations of the short-turn, and then, $y_{Sq}^{k\ell} = 1$. So, we have the following relationship between the two sets of variables:

$$y_q^{k\ell} \leq y_{Sq}^{k\ell} \quad k \in K_\ell, \, q \in Q, \, \ell \in LS. \tag{2}$$

Observe that we can detect whether the k-th trip is a fake trip for line $\ell \in LS$ (resp. $\ell \in LNS$) by checking if $\sum_{q \in Q} y_{Sq}^{k\ell} = 0 (= \sum_{q \in Q} y_q^{k\ell})$ (resp. $\sum_{q \in Q} y_q^{k\ell} = 0$).

- **Auxiliary Variables**: A set of auxiliary variables, computed from the previous ones, is considered in order to ease the reading of our Mathematical Programming formulation.
  − $t_i^{k\ell}$: Time instant in which a train departs from station i at the k-th trip of line $\ell \in L$. This value can be obtained by adding the travel times between consecutive stations plus the stopping times at the traversed stations. Thus, this time is given by the

following expressions:

$$t_i^{k\ell} = t_1^{k\ell} + \sum_{r=1}^{i-1}(d_r^\ell + e_{r+1}^\ell), \quad i > 1, \ (i,\ell) \in \overline{\mathcal{S}}, \ k \in K_\ell, \tag{3}$$

$$t_i^{k\ell} = t_1^{k\ell} + \sum_{r=1}^{i-1}(d_r^\ell + e_{r+1}^\ell) + w^{k\ell}, \quad i > 1, \ i \in S_\ell, \ k \in K_\ell, \ \ell \in LS. \tag{4}$$

Expressions (3) allow us to appropriately define our auxiliary variables $t_i^{k\ell}$ by updating the time in which a train starts its journey at the first station of the line to the time spent until it leaves station i for stations outside a short-turn. Equations (4) allow us to compute the arrival time of a train to any station of the short-turn. As mentioned before, trips that only cover stations in the short-turn are considered as fake trips for the stations which are not part of the short-turn and their corresponding departure times are supposed to be the same as the times of the previous trip. Variables $w^{k\ell}$ permit to fit departure times of the stations in each short-turn trip. These variables will take value 0 if the train cover the complete line.

- $D_i^\ell(t_i^{k\ell})$: Number of passengers accumulated at station i of line $\ell \in L$ from the origin of the planning horizon up to time $t_i^{k\ell}$. This number is given by a function of the time $t_i^{k\ell}$ and takes into account the external arrivals and also the passengers arriving to the station to change to another line. As mentioned above, a complete description of this *demand function* is given in Section 4.
- $h_i^{k\ell}$: Excess of passengers that where not able to get on the train at station i of the k-th trip of line $\ell \in L$ because a lack of capacity of the train. To compute this variable we distinguish between whole trips in any of the pool of lines and short-turn trips in those lines containing a short-turn. In the last case we assume that the passengers catching a short-turn train with destination to a station outside the short-turn get-off from the train in the last station of the short-turn to catch a whole trip train of the same line. Taking into account these assumptions, the excess of passengers can be computed as follows:

$$h_i^{1\ell} = D_i^\ell(t_i^{1\ell}) - f_i^{1\ell}, \quad \text{for } (i,\ell) \in \overline{\mathcal{S}}, \tag{5}$$

$$h_i^{1\ell} = D_i^\ell(t_i^{1\ell}) - f_i^{1\ell} - g_i^{1\ell}, \quad \text{for } i \in S_\ell \backslash \{n_{S_\ell}\}, \ \ell \in LS, \tag{6}$$

$$h_i^{1\ell} = D_i^\ell(t_i^{1\ell}) - f_i^{1\ell} + \sum_{r=1_{S_\ell}}^{i-1} \sum_{j=i+1}^{n_\ell} p_{rj} g_r^{1\ell}, \quad \text{for } i = n_{S_\ell}, \ \ell \in LS, \tag{7}$$

$$h_i^{k\ell} = D_i^\ell(t_i^{k\ell}) - D_i^\ell(t_i^{(k-1)\ell}) + \alpha h_i^{(k-1)\ell} - f_i^{k\ell}, \quad \text{for } (i,\ell) \in \overline{\mathcal{S}}, \ k = 2, \dots, \bar{k}_\ell, \tag{8}$$

$$h_i^{k\ell} = D_i^\ell(t_i^{k\ell}) - D_i^\ell(t_i^{(k-1)\ell}) + \alpha h_i^{(k-1)\ell} - f_i^{k\ell} - g_i^{k\ell},$$
$$\text{for } i \in S_\ell \backslash \{n_{S_\ell}\}, \ k = 2, \dots, \bar{k}_\ell, \ell \in LS, \tag{9}$$

$$h_i^{k\ell} = D_i^\ell(t_i^{k\ell}) - D_i^\ell(t_i^{(k-1)\ell}) + \alpha h_i^{(k-1)\ell} - f_i^{k\ell} + \sum_{r=1_{S_\ell}}^{i-1} \sum_{j=i+1}^{n_\ell} p_{rj} g_r^{k\ell},$$
$$\text{for } i = n_{S_\ell}, k = 2, \dots, \bar{k}_\ell, \ell \in LS. \tag{10}$$

In the first trip, the excess of passenger is computed by subtracting to the demand the flow of passengers already caught by the train (equations (5)-(7)). For modeling the flow of caught passengers we take into account in (6) that the trip could be either, a whole trip ($f_i^{1\ell}$) or a short-turn ($g_i^{1\ell}$). In the special case of the last station of the short-turn, one needs to add to the excess of passengers, those that get-off from a

train of a short-turn trip to catch a train of a whole trip of the same line (equations (7)). For the rest of the trips (equations (8)–(10)), we take into account the demand accumulated after the previous trip plus the excess of persisting passengers of the previous trip.

- **Semicontinuous Variables**: We also consider a set of semicontinuous variables collecting the excess of passengers only for *true trips*:

$$x_i^{k\ell} = \begin{cases} h_i^{k\ell} & \text{if } k \text{ is a true trip for station } i \text{ of line } \ell \\ 0 & \text{otherwise,} \end{cases} \qquad k \in K_\ell,\, i \in N_\ell,\, \ell \in L.$$

This set of variables will allow us to account in the objective function for the actual excess of passengers, i.e., the ones associated to true trips.

Table 2 summarizes the set of above mentioned variables.

Variable	Description
$t_1^{k\ell}$	Departure time from the initial station of line $\ell \in L$ at its k-th trip.
$f_i^{k\ell}$	Flow of passengers captured in the station i by the train that covers the k-th trip of the line $\ell \in L$, when k is a whole trip.
$g_i^{k\ell}$	Flow of passengers captured in the station $i \in S_\ell \backslash \{n_\ell\}$ by the train that covers the k-th trip of the line $\ell \in LS$, when k only covers the short-turn stations.
$w^{k\ell}$	Difference between the actual departure time from the first station of the short-turn of the k-th trip of line $\ell \in LS$ and the time when it should depart from this station taking into account its departure time from the initial station of the line.
$y_q^{k\ell}$	$\begin{cases} 1 & \text{if the } k\text{-th trip of line } \ell \in L \text{ is a whole trip with capacity } q \\ 0 & \text{otherwise} \end{cases}$
$y_{Sq}^{k\ell}$	$\begin{cases} 1 & \text{if the } k\text{-th trip of line } \ell \in LS \text{ traverses the short-turn } S_\ell \text{ with capacity } q. \\ 0 & \text{otherwise} \end{cases}$
$t_i^{k\ell}$	Departure time from the station i of line ℓ in its k-th trip.
$D_i^\ell(t_i^{k\ell})$	Number of passengers accumulated from instant 0 up to instant $t_i^{k\ell}$ in the station i of line $\ell \in L$.
$h_i^{k\ell}$	Excess of passengers that where not able to get on the train at station i at the k-th trip of line $\ell \in L$ because of a lack of capacity.
$x_i^{k\ell}$	Excess of passengers only if k is a *true* trip for station i of line $\ell \in L$.

TABLE 2. Variables of the model.

Using the variables described above, we give now the formulation of our problem distinguishing the two main elements: (1) an objective function aggregating the economical and social costs of any feasible solution; and (2) a set of technical constraints where the line concept and the timetable are specified in order to serve the estimated flow of passengers over the planning horizon.

3.3. **Objective Function.** In what follows we describe the different **costs** and **rewards** that are considered in the objective function of our model, for a given line $\ell \in L$, in terms of the variables and parameters described above.

- **Capacity cost:**
  It accounts for the costs b_q^ℓ and b_{Sq}^ℓ corresponding to the train capacity $q \in Q$ for every trip of line $\ell \in L$. This capacity is controlled by means of the variables $y_q^{k\ell}$ or $y_{Sq}^{k\ell}$

depending on whether the line contains short-turns or not.

$$\begin{cases} \displaystyle\sum_{k\in K_\ell}\sum_{q\in Q} b_q^\ell y_q^{k\ell} & \text{if } \ell \in LNS, \\[2ex] \displaystyle\sum_{k\in K_\ell}\sum_{q\in Q} b_q^\ell y_q^{k\ell} + \sum_{k\in K_\ell}\sum_{q\in Q} b_{Sq}^\ell (y_{Sq}^{k\ell} - y_q^{k\ell}) & \text{if } \ell \in LS. \end{cases} \qquad (\text{Cap}(\ell))$$

Observe that in the second expression, if proper short-turns with capacity q are activated, one has $y_{Sq}^{k\ell} = 1$ and $y_q^{k\ell} = 0$, and thus, the amount b_{Sq}^ℓ is accounted for. In case of whole trips with capacity q, by (2), one has that $y_{Sq}^{k\ell} = 1$ and $y_q^{k\ell} = 1$. Thus, the second addend is zero, and only b_q^ℓ is accounted for, modeling adequately the capacity costs.

- **Reward per served passengers:**
  The unitary reward per served passenger computes the estimated revenue received when passengers use a line. It is obtained as the average revenue given by the different transport tickets used by the passengers. In order to compute the overall reward per served passengers, observe that if the k-th trip is a whole trip, the expression $\sum_{r=0}^{i-1} p_{ri} f_r^{k\ell}$ returns the estimated number of passengers getting off at the station i (coming from any other previous station). Note that, in case a short-turn is activated on any line $\ell \in LS$, it is needed to add the rewards of passengers being routed within the short-turn $(\sum_{i\in S_\ell}\sum_{r=1_{S_\ell}}^{i-1} p_{ri}^\ell g_r^{k\ell})$ together with the reward from those which use the short-turn to get-off at the last station and continue with the next whole trip. Then, the overall reward per served passengers is:

$$\begin{cases} \displaystyle\sum_{i\in N_\ell}\sum_{k\in K_\ell}\sum_{r=0}^{i-1} \gamma_{ri}^\ell p_{ri}^\ell f_r^{k\ell} & \text{if } \ell \in LNS, \\[3ex] \displaystyle\sum_{k\in K_\ell}\left(\sum_{i\in N_\ell}\sum_{r=0}^{i-1} \gamma_{ri}^\ell p_{ri}^\ell f_r^{k\ell} + \sum_{i\in S_\ell}\sum_{r=1_{S_\ell}}^{i-1} \gamma_{ri}^\ell p_{ri}^\ell g_r^{k\ell} + \sum_{\substack{r\in S_\ell: \\ r\neq n_{S_\ell}}}\sum_{j=n_{S_\ell}+1}^{n_\ell} \gamma_{rn_{S_\ell}}^\ell p_{rj}^\ell g_r^{k\ell} \right) & \text{if } \ell \in LS. \end{cases}$$
$$(\text{RewPPass}(\ell))$$

- **Costs of non-served passengers:**
  It accounts for the social cost incurred when a passenger cannot get on the train arriving to the station due to its lack of capacity. The unitary cost should aggregate some indicators of the service quality and some subjective measures of the satisfaction degree perceived by the passengers. The overall cost is computed by using the total number of passengers exceeding the capacity of the system at some instant of the planning horizon, scaled by the average penalties of persisting/giving up passengers, as follows:

$$\alpha\mu_1 \sum_{i\in N_\ell}\sum_{k\in K_\ell} x_i^{k\ell} + (1-\alpha)\mu_2 \sum_{i\in N_\ell}\sum_{k\in K_\ell} x_i^{k\ell}, \qquad (\text{NonServed}(\ell))$$

Observe that a high excess of passenger at the end of the planning horizon can be easily avoided by increasing the μ-parameters for the last trip or by adding appropriate constraints involving the x-variables.

The overall cost of using line $\ell \in L$ during the planning horizon can be expressed as:

$$(\text{Cap}(\ell)) - (\text{RewPPass}(\ell)) + (\text{NonServed}(\ell)) \qquad (\text{COST}(\ell))$$

3.4. **Modelling Constraints.** In what follows we describe the constraints linking the variables and parameters in our model. They have been classified in four main blocks: capacity constraints, time control constraints, flow control constraints and passenger surplus constraints.

- Capacities and true/fake trips:

$$\sum_{q \in Q} y_q^{1\ell} = 1, \qquad \ell \in LNS, \tag{C1-1}$$

$$\sum_{q \in Q} y_q^{k\ell} \leq 1, \qquad k > 1, \ell \in L, \tag{C1-2}$$

$$\sum_{q \in Q} y_q^{\bar{k}_\ell \ell} = 1, \qquad \ell \in L, \tag{C1-3}$$

$$y_q^{k\ell} \leq y_{Sq}^{k\ell}, \qquad q \in Q, k \in K_\ell, \ell \in LS, \tag{C1-4}$$

$$\sum_{q \in Q} y_q^{1\ell} + \sum_{q \in Q} y_{Sq}^{1\ell} \geq 1, \qquad \ell \in LS, \tag{C1-5}$$

$$\sum_{q \in Q} y_q^{\kappa_\ell \ell} = \sum_{q \in Q} y_{Sq}^{1\ell} - \sum_{q \in Q} y_q^{1\ell}, \qquad \ell \in LS, \tag{C1-6}$$

$$\sum_{q \in Q} y_{Sq}^{k\ell} \leq 1, \qquad k > 1, \ell \in LS, \tag{C1-7}$$

where $\kappa_\ell = \left\lceil \frac{t_{1 \mapsto 1_{S_l}}}{IS^\ell} \right\rceil + 1$, i.e. the number of short-turn trips of line $\ell \in LS$ which fit within the period of time taken by a train to go from the head of line to the first station of the short-turn.

When short-turns are not allowed for a line, the appropriate definition of the capacity variables is ensured by constraints $(C1-1)$–$(C1-3)$. They enforce that exactly one of the allowed capacities is chosen for the first and the last trip and at most one for the rest of them. Fake (resp. true) trips are identified by trips with capacities equal to (resp. greater than) zero. Thus, constraints $(C1-1)$ and $(C1-3)$ determine that the first and the last trip of each line are true trips. We will see later that this permits the actual trains to be scheduled from the beginning to the end of the planning horizon, providing the users a complete service during that time interval. Note that constraints $(C1-2)$ and $(C1-3)$ are also valid for lines allowing short-turns and then, when short-turns are allowed for a line, the appropriate definition of the capacity variables is warranted by constraints $(C1-2)$–$(C1-7)$. Constraints $(C1-4)$ indicate that when a whole trip is a true trip, it is also a true trip for the short-turn stations. Constraints $(C1-5)$ fix that the first trip (being either, a whole trip or a short-turn) is a true trip. Constraints $(C1-6)$ force trip κ_ℓ to be a true whole trip (resp. a fake trip) if the first trip is a short-turn (resp. if the first trip is a whole trip). These constraints, together with $(C1-5)$ are the equivalent to $(C1-1)$ for lines with short-turns, and they ensure that a real train is scheduled from the beginning of the planning horizon. Finally, constraints $(C1-7)$ are the equivalent to $(C1-2)$ for short-turn trips.
- Time control:

$$t_1^{1\ell} = 0, \qquad \ell \in L, \tag{C2 - 1}$$

$$t_1^{\overline{k}\ell} = T, \qquad \ell \in L, \tag{C2 - 2}$$

$$t_1^{\kappa_\ell \ell} \leq T \left(1 - \sum_{q \in Q} y_{Sq}^{1\ell} + \sum_{q \in Q} y_q^{1\ell}\right), \qquad \ell \in LS, \tag{C2 - 3}$$

$$IS \left(\sum_{q \in Q} y_q^{k\ell}\right) \leq t_i^{k\ell} - t_i^{(k-1)\ell},$$

$$k = 2, \ldots, \overline{k}_\ell, (i, l) \in \overline{\mathcal{S}} \text{ with } k \neq \kappa_\ell \text{ if } \ell \in LS, \tag{C2 - 4}$$

$$t_i^{k\ell} - t_i^{(k-1)\ell} \leq T \left(\sum_{q \in Q} y_q^{k\ell}\right), \quad k = 2, \ldots, \overline{k}_\ell, (i, l) \in \overline{\mathcal{S}} \tag{C2 - 5}$$

$$IS \left(\sum_{q \in Q} y_{Sq}^{k\ell}\right) \leq t_i^{k\ell} - t_i^{(k-1)\ell}, \quad i \in S_\ell, k = 2, \ldots, \overline{k}_\ell, \ell \in LS, \tag{C2 - 6}$$

$$t_i^{k\ell} - t_i^{(k-1)\ell} \leq (T + t_{1 \mapsto 1_{S_l}}) \left(\sum_{q \in Q} y_{Sq}^{k\ell}\right), \quad i \in S_\ell, k = 2, \ldots, \overline{k}_\ell, \ell \in LS, \tag{C2 - 7}$$

$$-t_{1 \mapsto 1_{S_l}} \left(1 - \sum_{q \in Q_\ell} y_q^{k\ell}\right) \leq w^{k\ell}, \quad k \in K_\ell, \ell \in LS, \tag{C2 - 8}$$

$$w^{k\ell} \leq (T + t_{1 \mapsto 1_{S_l}}) \left(1 - \sum_{q \in Q_\ell} y_q^{k\ell}\right), \quad k \in K_\ell, \ell \in LS, \tag{C2 - 9}$$

Constraints $(C2 - 1)$ and $(C2 - 2)$ state that the first and the last trip of each line should exactly start at the first station of the complete line at instant time 0 and T, respectively. If the line does not contain a short-turn, these constraints together with $(C1 - 1)$ and $(C1 - 3)$ ensure that there are trains traveling the line during the whole planning horizon. If the line contains short-turns, recall that constraints $(C1 - 5)$ permit the first trip to be a short-turn trip. In this case, constraints $(C2 - 3)$ together with constraints $(C1 - 6)$ force trip κ_ℓ to be a true whole trip starting at time 0 at the head of the line.

Constraints $(C2 - 4)$ ensure that the arrival times between consecutive trains satisfy the safety time window interval. Constraints $(C2 - 5)$ force a fake trip to operate on the line at the same time as the previous true trip. For lines allowing short-turns constraints $(C2 - 4)$–$(C2 - 7)$ represent the same as the above but taking into account that in this case the trip can be either, a whole trip or a short-turn.

Observe that in $(C2 - 4)$ if $\ell \in LS$ the case $k = \kappa_\ell$ is excluded. The reason is that constraints $(C1 - 6)$ force trip κ_ℓ to be a true whole trip if and only if the first trip $(k = 1)$ is a short-turn and, in this case, constraints $(C2 - 3)$ ensure that trip κ_ℓ starts at the head of the line station at time 0, and then, it is the first whole trip. Finally, constraints $(C2 - 8)$ and $(C2 - 9)$ fix the upper and lower bounds on the values of variables w^{kl} when the k-th trip is a short-turn, enforcing that this variable is 0 if the trip is a whole trip.

- Flow control:

$$f_i^{k\ell} + \sum_{r=1}^{i-1} f_r^{k\ell} \left(\sum_{j=i+1}^{n_\ell} p_{rj}^\ell \right) \leq \sum_{q \in Q} q y_q^{k\ell}, \quad k \in K_\ell, i \in N_\ell, \ell \in L, \qquad (C3-1)$$

$$g_i^{k\ell} + \sum_{r=1_{S_\ell}}^{i-1} g_r^{k\ell} \left(\sum_{j=i+1}^{n_{S_\ell}} p_{rj}^\ell \right) \leq \sum_{q \in Q} q(y_{Sq}^{k\ell} - y_q^{k\ell}), k \in K_\ell, i \in S_\ell \backslash \{n_{S_\ell}\}, \ell \in LS, \qquad (C3-2)$$

$$f_i^{1\ell} \leq D_i^\ell(t_i^{1\ell}), \quad i \in N_\ell, \ell \in L, \qquad (C3-3)$$

$$f_i^{kl} \leq D_i^\ell(t_i^{k\ell}) - D_i^\ell(t_i^{(k-1)\ell}) + \alpha h_i^{(k-1)\ell}, \quad k = 2, \dots, \overline{k}_\ell, i \in N_\ell, \ell \in L, \qquad (C3-4)$$

$$g_i^{1\ell} \leq D_i^\ell(t_i^{1\ell}), \ i \in S_\ell \backslash \{n_{S_\ell}\}, \ell \in LS, \qquad (C3-5)$$

$$g_i^{k\ell} \leq \left(D_i^\ell(t_i^{k\ell}) - D_i^\ell(t_i^{(k-1)\ell}) \right) + \alpha h_i^{(k-1)\ell}, \ k = 2, \dots, \overline{k}_\ell, i \in S_\ell \backslash \{n_{S_\ell}\}, \ell \in LS. \quad (C3-6)$$

The flow of passengers catching a given train is determined by the capacity of the train and the mobility pattern of people. Hence, the effective capacity of the trains arriving to a given station depends of the passengers that caught the train in previous stations of this same trip and whose destination is a subsequent station of the line. Such an effective capacity is warranted, depending on the case (short-turn allowed or not), by constraints $(C3-1)$ and $(C3-2)$. With constraints $(C3-3)$ (resp. $(C3-4)$) we ensure that the flow of passengers captured at a given station by a train that covers the first trip (resp. the k-th trip for $k > 1$) is at most the demand of passengers accumulated at station i since the beginning of the planning horizon (resp. since the instant in which the previous train departed from that station plus the passengers that were not able to get on the previous train because of lack of capacity and wait for the next one). Constraints $(C3-5)$ and $(C3-6)$ are the analogous ones for short-turn trips.

- Passenger surplus:

In order to compute only the surplus of passengers of a true trip, we use the set of semi-continuous variables $x_i^{k\ell}$:

$$x_i^{k\ell} = \begin{cases} h_i^{k\ell} \times \sum_{q \in Q} y_q^{k\ell} & \text{if } (i,l) \in \overline{S} \text{ or } (i = n_{S_\ell}, \ell \in LS), \\ h_i^{k\ell} \times \sum_{q \in Q} y_{Sq}^{k\ell} & \text{if } i \in S_\ell \backslash \{n_{S_\ell}\}, \ell \in LS, \end{cases}$$

which can be linearized as follows:

$$x_i^{k\ell} \geq h_i^{k\ell} - M_i^\ell \left(1 - \sum_{q \in Q_\ell} y_q^{k\ell} \right), (i,l) \in \overline{S} \text{ or } (i = n_{S_\ell}, \ell \in LS), \qquad (C4-1)$$

$$x_i^{k\ell} \geq h_i^{k\ell} - M_i^\ell \left(1 - \sum_{q \in Q_\ell} y_{Sq}^{k\ell} \right), i \in S_\ell \backslash \{n_{S_\ell}\}, \ell \in LS, \qquad (C4-2)$$

being M_i^ℓ a large enough constant bounding the surplus of passengers at any station i of line $\ell \in L$.

3.5. **A compact MINLP formulation.** According to the decision variables, the objective function and the constraints described above, the following Mathematical Programming formulation

is valid for our line planning and timetabling model:

$$\min \sum_{\ell \in L} \mathrm{COST}(\ell)$$

s.t. (C1), (C2), (C3) and (C4),

$$
\begin{array}{lll}
0 \le t_1^{k\ell} \le T, & k \in K_\ell,\, \ell \in L, & \\
f_i^{k\ell} \ge 0, & k \in K_\ell,\, i \in N_\ell,\, \ell \in L, & \text{(P)} \\
g_i^{k\ell} \ge 0, & k \in K_\ell,\, i \in S_\ell \backslash \{n_{S_\ell}\},\, \ell \in LS, & \\
w^{k\ell} \in \mathbb{R}, & k \in K_\ell,\, \ell \in LS, & \\
x_i^{k\ell} \ge 0, & k \in K_\ell,\, i \in N_\ell,\, \ell \in L, & \\
y_q^{k\ell} \in \{0,1\}, & k \in K_\ell,\, q \in Q,\, \ell \in L, & \\
y_{S_q}^{k\ell} \in \{0,1\}, & k \in K_\ell,\, q \in Q,\, \ell \in LS. &
\end{array}
$$

Observe that although the above formulation seems to be separable by lines in L, the lines are linked through the demand functions $D_i^\ell(t)$ (constraints (C3 − 3)-(C3 − 6)) which represent the accumulated flow of passengers awaiting for a train at a given station i of line $l \in L$ at time instant t. As we will describe in Section 4, such a flow is affected not only by the line ℓ but also by other lines through passengers changing of lines at transfer stations. This function introduces new variables and non linear constraints into the above formulation.

Several extensions may be easily accommodated within the above model as highlighted in the following remarks:

Remark 1. *In our model the speed of trains is considered to be constant during the whole journey. However, one can easily modify expressions* (1)*,* (3) *and* (4) *using variables $v_i^{k\ell} > 0$ to decide the speed of the train during a trip k of line ℓ between stations i and $i+1$. For instance, let ρ_i be the physical distance between stations i and $i+1$ and let $\omega_i^{k\ell}$ represent the inverse of the speed, i.e., $\omega_i^{k\ell} = \frac{1}{v_i^{k\ell}}$, then one could replace the travel times d_i^ℓ by $\rho_i \times \omega_i^{k\ell}$ in expressions* (1)*,* (3) *and* (4)*. By adding to the objective function a cost assessing the resource consumption due to speed changes one can have a more general model preserving the structure of the one stated above. Similar modifications can be also considered by enabling the model to decide about variable stopping times at any station.*

Remark 2. *The model allows us to join two consecutive planning horizons by passing data about numbers of passengers and arrival/departure times obtained from an optimal solution on the first planning period as input data for the second one, and so on. In particular, the passengers that may remain at station i of line $\ell \in L$ at the end of the first planning horizon can be considered as passengers at station i to use line ℓ at the beginning of the second planning horizon, and so on. This information has to be incorporated in order to compute the demand function, as we will see in the next section.*

4. The Demand function

One of the main goals of our model is to incorporate, in the design of the line planning and timetabling of an existing network, information about the flow of passengers moving through the network during the planning horizon. Clearly, the flow of passengers arriving to a given station is a random variable. Thus, we will incorporate to the model an estimation of its average value.

In order to model the number of passengers entering to the transportation system through a given station of a fixed line, we use the so-called *demand function*, which maps at a given instant t the accumulated number of passengers wanting to catch a train at this station (from

the beginning of the planning horizon). Here, the estimation process should be carefully done in order to capture the essential behaviour of the demands served by the system.

Different shapes for the function are possible within this framework to approximate the demand. The choice of such a shape is a crucial step in the modeling process since one has to find an equilibrium between obtaining accurate estimations and providing manageable mathematical programming formulations. Once again, motivated by our pilot experience with Metrolab®, we use a piecewise linear approximation whose slope is fixed for any given station $i \in N_\ell$, but whose breakpoints and discontinuities may change according to the flow induced by external block of arrivals and by the rest of the lines. For a given line $\ell \in L$ and a station $i \in N_\ell$, we estimate the demand function as follows:

$$D_i^\ell(t) = \beta_{0i}^\ell + \beta_i^\ell t + J_{i\ell}^E(t) + \sum_{\ell' \neq \ell, \ell' \ni i} J_{i\ell\ell'}^I(t), \tag{D}$$

for $t \in [0, \widehat{T}_\ell]$, with $\widehat{T}_\ell = T + \sum_{r=1}^{n_\ell - 1}(d_r^\ell + e_{r+1}^\ell)$, i.e., the maximum time in which the train can reach the last station of the line, and where:

β_{0i}^ℓ is the number of passengers awaiting a train of line $\ell \in L$ in the station $i \in N_\ell$ at the beginning of the planning horizon.

β_i^ℓ is the average rate of passengers arriving to the station $i \in N_\ell$ of line $\ell \in L$ by unit of time.

$J_{i\ell}^E(t)$ is the sum of the external block of arrivals of passengers up to the instant t to the station $i \in N_\ell$ of line $\ell \in L$.

$J_{i\ell\ell'}^I(t)$ is the sum of the block arrivals of passengers up to the instant t to the station $i \in N_\ell$ of line $\ell \in L$ from line $\ell' \in L$.

In what follows we will refer to $[0, \widehat{T}_\ell]$ as the *extended planning horizon for line ℓ*. Thus, the demand function at a given time instant t in station $i \in N_\ell$ of line $\ell \in L$, consists of three parts. The first one is a linear part, in which, from an initial number of passengers, β_{0i}^ℓ, the number increases by a rate β_i^ℓ. However, such a base estimation may be modified either by external block of arrivals (second part), $J_{i\ell}^E(t)$, or by passengers coming from other interacting lines at interchange stations (third part), $J_{i\ell\ell'}^I(t)$.

As can be seen, in the formulation (P), the demand function $D_i^\ell(t)$ is used exclusively to access flows in the set of instants $t = t_i^{k\ell}$ for $i \in N_l, \ell \in L$ and $k \in K_\ell$. Each of the time instants in which the demand function needs to be evaluated induces some sets of inequalities and variables as those described in subsections 4.1 and 4.2 (for external and internal arrivals).

4.1. **External Arrivals: J^E.** We consider that we are given both a set of breakpoints representing time instants when the block of arrivals occur and the amounts of passengers entering to the system at these instants for each station $i \in N_\ell$. That is, we assume that a set of sorted instants $se_1^{i\ell} < \cdots < se_{re^{i\ell}}^{i\ell}$ as well as discontinuity flow jumps associated to each of those instants $\Psi_1^{i\ell}, \ldots, \Psi_{re^{i\ell}}^{i\ell}$ are known, i.e., a block arrival of $\Psi_r^{i\ell}$ is assumed at time instant $se_r^{i\ell}$, for $r = 1, \ldots, re^{i\ell}$. The external arrivals represent block of arrivals of passengers for instance, due to the end of a football match in a place close to one of our stations. For the sake of readability and without loss of generality, we will assume that $se_0^{i\ell} = 0$, $se_{re^{i\ell}+1}^{i\ell} = \widehat{T}_\ell$, and $\Psi_0^{i\ell} = \Psi_{re^{i\ell}+1}^{i\ell} = 0$, i.e., the first and last time instants of external arrivals coincide with the beginning and the end of the extended planning horizon, and the discontinuity jumps at those instants are null. Given a time instant $t \in [0, \widehat{T}_\ell]$, we use the following set of binary variables to determine whether t belongs to the interval $[se_r^{i\ell}, se_{r+1}^{i\ell})$ for $r = 0, \ldots, re^{i\ell}$:

$$\delta_{ri\ell}^E(t) = \begin{cases} 1 & \text{if } t \in [se_r^{i\ell}, se_{r+1}^{i\ell}), \\ 0 & \text{otherwise}, \end{cases} \quad i \in N_\ell, \ell \in L.$$

Note that with these settings, the accumulated discontinuity flow jumps of external arrivals for a station $i \in N_\ell$ of line $\ell \in L$ can be modeled using the following constraints

$$J_{i\ell}^E(t) = \sum_{r=0}^{re^{i\ell}} \left(\sum_{r' \leq r} \Psi_{ir'}^\ell \right) \delta_{ri\ell}^E(t), \quad i \in N_\ell, \ell \in L,$$

$$se_r^{i\ell} \delta_{ri\ell}^E(t) \leq t < se_{r+1}^{i\ell} \delta_{ri\ell}^E(t) + \widehat{T}_\ell (1 - \delta_{ri\ell}^E(t)), \quad r = 0, \ldots, re^{i\ell}, i \in N_\ell, \ell \in L,$$

$$\text{(D}_\text{E}\text{)}$$

$$\sum_{r=0}^{re^{i\ell}} \delta_{ri\ell}^E(t) = 1, \quad i \in N_\ell, \ell \in L,$$

$$\delta_{ri\ell}^E(t) \in \{0, 1\} \quad r = 0, \ldots, re^{i\ell}, i \in N_\ell, \ell \in L.$$

The reader may observe that the first constraint allows us to appropriately define J^E by accumulating the flow jumps previous to a given instant t. The second set of constraints permits to determine the intervals in which t lies and the third constraint ensures that only one of these intervals is identified for each t.

4.2. **Internal Arrivals.** The main difference between internal and external block of arrivals is that in the latter, the breakpoints and the discontinuity flow jumps are known, while for internal arrivals, this information depends on the decision variables of the problem. Recall that internal block of arrivals occur when passengers get off a train in an interchange station to transfer to another line. Thus, the time instants of those arrivals are part of the decision problem. Therefore, we need to state a set of equations showing the existing relationships between the decision variables and the breakpoints and jumps in the flow of passengers they provoke.

In what follows we describe the modeling issues concerning the breakpoint times and flow jumps of internal arrivals.

- **Breakpoint times at line $\ell \in L$.**

    Note that the instants, $si_r^{i\ell\ell'}$, in which an internal flow jump occurs by the arrivals of trains coming from line $\ell' \in L$ to an interchange station $i \in N_\ell \cap N_{\ell'}$ can be computed in terms of the time $t_i^{r\ell'}$ as:

    $$si_r^{i\ell\ell'} = t_i^{r\ell'} - e_i^{r\ell'}, \quad r = 1, \ldots, \bar{k}_{\ell'}$$

    that is, the time instant in which the r-th trip of line $\ell' \in L$ departs to the next station $(t_i^{r\ell'})$ minus the waiting time at the station $i \in N_{\ell'}$ $(e_i^{r\ell'})$, that is, the instants in which the trains arrive to the interchange station of line $\ell' \in L$.

- **Discontinuity flow jumps at line $\ell \in L$ coming from line $\ell' \in L$.**

    The volume of the internal block of arrivals to an interchange station can be derived from the passenger flows controlled by our decision variables. To compute it we will also need a set of values quantifying the proportion of passengers which want to change from one metro line to another one in all the interchange stations. Let $\tau_i^{\ell\ell'}$ be the proportion of the passengers that get off a train in an interchange station $i \in N_\ell \cap N_{\ell'}$ of the line $\ell' \in L$ to change to line $\ell \in L$. Thus, the flow jump at an instant $si_r^{i\ell\ell'}$ for $r \in K_{\ell'}, \ell \in L, i \in N_\ell$ is given by

    $$\Phi_{ir}^{\ell\ell'} = \begin{cases} \tau_i^{\ell\ell'} \sum_{j<i} p_{ji}^{\ell'} f_j^{r\ell'} & \text{if } (i, \ell') \in \overline{S} \\ \tau_i^{\ell\ell'} \left(\sum_{j<i} p_{ji}^{\ell'} f_j^{r\ell'} + \sum_{j<i; j \in S_{\ell'}} p_{ji}^{\ell'} g_j^{r\ell'} \right) & \text{if } i \in S_{\ell'}, \ell' \in LS \end{cases}$$

Let $S_{i\ell\ell'}^I = \left\{ si_0^{i\ell\ell'}, \cdots, si_{\bar{k}_{\ell'}+1}^{i\ell\ell'} \right\}$ be the set of sorted breakpoint times at a given interchange station $i \in N_\ell \cap N_{\ell'}$ of line $\ell \in L$ caused by line $\ell' \in L$. In order to make clearer the formulation, we will assume, w.l.o.g., that $si_0^{i\ell\ell'} = 0$, $si_{\bar{k}_{\ell'}+1}^{i\ell\ell'} = \widehat{T}_\ell$, and $\Phi_{i0}^{\ell\ell'} = \Phi_{i\bar{k}_{\ell'}+1}^{\ell\ell'} = 0$, i.e., the first and last time instants of internal arrivals coincide with the beginning and the end of the extended planning horizon, and the flow jumps at those times are null.

Now, we are ready to model the internal flow jumps induced by line $\ell' \in L$ onto the station $i \in N_\ell \cap N_{\ell'}$ of the line ℓ. We proceed analogously as in the external arrivals but incorporating a set of binary variables (δ^I) to identify in which time interval between two consecutive breakpoints a given instant t belongs to:

$$J_{i\ell\ell'}^I(t) = \sum_{r=0}^{\bar{k}_{\ell'}} \left(\sum_{r' \leq r} \Phi_{ir'}^{\ell\ell'} \right) \delta_{ri\ell\ell'}^I(t), \quad i \in N_\ell \cap N_{\ell'}, \ell \in L, \ell' \in L,$$

$$si_r^{i\ell\ell'} \delta_{ri\ell\ell'}^I(t) \leq t < si_{r+1}^{i\ell\ell'} \delta_{ri\ell\ell'}^I(t) + \widehat{T}_\ell(1 - \delta_{ri\ell\ell'}^I(t)), \ r = 0, \ldots, \bar{k}_{\ell'}, \ i \in N_\ell \cap N_{\ell'}, \ell \in L, \ell' \in L,$$

$$\sum_{r=0}^{\bar{k}_{\ell'}} \delta_{ri\ell\ell'}^I(t) = 1, \quad i \in N_\ell \cap N_{\ell'}, \ell \in L, \ell' \in L,$$

$$\delta_{ri\ell\ell'}^I(t) \in \{0,1\} \ r = 0, \ldots, \bar{k}_{\ell'}, \ i \in N_\ell \cap N_{\ell'}, \ell \in L, \ell' \in L. \tag{D_I}$$

Note that the two first sets of the above inequalities are nonlinear since $\mathbf{\Phi}$, $\mathbf{S^I}$ and $\delta^{\mathbf{I}}$ are decision variables of our problem.

As mentioned above, the demand function $D_i^\ell(t)$ is only applied to a certain (finite) set of time instants. Then, it can be incorporated into the model by adding the variables $D_i^{k\ell} \equiv D_i^\ell(t_i^{k\ell})$, δ^E and δ^I, as well as their corresponding linear and nonlinear constraints in (D_E) and (D_I). We also include in the model (P) the following two sets of valid inequalities:

$$\delta_{ri\ell}^E(t_i^{k+1\ell}) \leq \sum_{r'=0}^r \delta_{r'i\ell}^E(t_i^{k\ell}), \qquad\qquad k \in K_\ell, \, , r = 0, \ldots, re^{i\ell}, \, i \in N_\ell, \ell \in L, \tag{15}$$

$$\delta_{ri\ell\ell'}^I(t_i^{k+1\ell}) \leq \sum_{r'=0}^r \delta_{r'i\ell\ell'}^I(t_i^{k\ell}), \qquad k \in K_\ell, \, r = 0, \ldots, \bar{k}_{\ell'}, \, i \in N_\ell \cap N_{\ell'}, \ell \in L, \ell' \in L. \tag{16}$$

Inequalities (15) and (16), ensure that trip $k+1$ goes through station i after trip k. Although these inequalities are redundant with the rest of the constraints in the model, they considerably improve the performance of the solver over the branch-and-bound tree induced by the mathematical programming model.

5. Case study: The Metrolab® pilot experience

We illustrate the MINLP model introduced in Section 3 together with the demand function described in Section 4 with a simplified version of the network of our pilot experience for Metrolab®.

Consider the simple network example depicted in Figure 2. This network has a similar topology to the one provided by Metrolab® to calibrate our model in the pilot experience. It represents a small section of the Paris subway network.

The network consists of two *bidirectional lines*(1-2-3-4-5 and 1'-2'-3-4'-5'), each of them with five stations and sharing one of them (station 3) which acts as an interchange station. Both lines allow short-turns on the planning. The edges of the lines which take part of the short-turn of each line are marked with dashed lines ($2 - 3 - 4$ for one of the lines and $2' - 3 - 4'$ for the other).
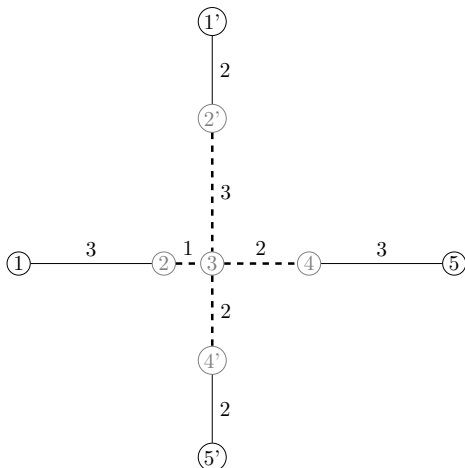
FIGURE 2. Network of Case study 5.

The travel time between consecutive stations is written next to each edge of the network. The stopping times are fixed to 30 seconds for stations different from the head and the final ones (e_i^ℓ-parameters). Also the minimum safety time between consecutive trips is 2 minutes (IS^ℓ-parameter). Hence, using our notation $|L| = 4$ (two lines, but each of them in two directions), $|LS| = 4$ and $|LNS| = 0$. The lines will be numbered from 1 to 4, where 1 and 2 correspond with the horizontal line, left–right and right–left, respectively, and lines 3 and 4 are identified with the vertical line, up–down and down–up, respectively. The set of available capacities for the trains are 800 and 1600 passengers.

For illustration purposes, we consider a planning horizon of $T = 20$ minutes (from 7:30 to 7:50), and a maximum number of trips $K_\ell = 7$ for the four lines. We assume that all the passengers that cannot get on a train because it is full await for the next train ($\alpha = 1$) and that the unitary penalty for persisting passengers is fixed to $\mu_1 = 0.1875$ (except for the last trip in which we fix it to $10 \times \mu_1 = 1.875$ to avoid a high excess of passengers at the end of the planning horizon). We also assume that 40% of the passengers that get-off at station 3 interchange to the other lines (τ-parameter).

The O-D matrix and rewards per passenger are detailed in Tables 3 and 4, respectively.

p_{ij}^ℓ	1	2	3	4	5
1	0	0.40	0.35	0.20	0
2	0.40	0	0.60	0.35	0
3	0.35	0.6	0	0.95	0
4	0.20	0.35	0.95	0	1
5	0.05	0.05	0.05	1	0

p_{ij}^ℓ	1'	2'	3	4'	5'
1'	0	0.40	0.35	0.20	0
2'	0.40	0	0.60	0.35	0
3	0.35	0.60	0	0.95	0
4'	0.20	0.35	0.95	0	1
5'	0.05	0.05	0.05	1	0

TABLE 3. O-D matrix of Case study 5: Lines 1-2 (left) and 3-4 (right).

For the demand function, we consider the arrival rates of passengers per minute and the initial number of passengers at each station of each line detailed in Table 5.

The timetabling for the first line obtained after running our model in Gurobi 8.0 [23] is provided in Table 6. The reported solution was obtained after 12 hours running time with a MIP GAP of 1.51%. There, we detail for each trip (k, indicating with 'S' those trips which are

γ_{ij}^ℓ	1	2	3	4	5
1	0	0.3	0.4	0.6	1
2	0.3	0	0.1	0.3	1
3	0.5	0.2	0	0.2	1
4	0.6	0.3	0.1	0	0
5	0.9	0.6	0.4	0.3	0

γ_{ij}^ℓ	1'	2'	3	4'	5'
1'	0	0.2	0.5	0.7	1
2'	0.2	0	0.3	0.5	1
3	0.4	0.2	0	0.2	0
4'	0.7	0.5	0.3	0	0
5'	0.9	0.7	0.5	0.2	0

TABLE 4. Rewards of Case study 5: Lines 1-2 (left) and 3-4 (right).

	Lines									
	$\ell = 1$					$\ell = 2$				
Stations (i)	1	2	3	4	5	5	4	3	2	1
β_{0i}^ℓ	50	50	50	50	0	50	50	50	50	0
β_i^ℓ	10	100	120	90	0	10	160	180	150	0
	$\ell = 3$					$\ell = 4$				
Stations (i)	1'	2'	3	4'	5'	5'	4'	3	2'	1'
β_{0i}^ℓ	50	50	50	50	50	50	50	50	50	50
β_i^ℓ	10	150	170	160	0	10	100	180	150	0

TABLE 5. Coefficients of the Demand functions of Case study 5.

short-turns), its optimal capacity, the departing times of each train of each line (DepTime), and also the flow estimations at each of the stages: the number of passengers that get off the train (Get-Off), the number of passengers that get-on the train ($f_i^{k\ell}$ for whole trips or $g_i^{k\ell}$ for short-turns), the excess of passengers ($h_i^{k\ell}$), the passengers surplus of true trips ($x_i^{k\ell}$) and the actual load of the train.

Note that trips 3 and 5 are fake trips (they have zero capacity), so the optimal number of trips for the planning is 5. Two of them, the second and third true trips (trips 2S and 4S in Table 6) are short-turns, while the remainder are whole trips with different capacities. We can observe that fake trips operate on the line at the same time as the previous true trip. For instance, trip 3 departs from station 1, which doesn't belong to the short-turn, at the same time than the previous true whole trip (trip 1) and from station 2, which belongs to the short-turn, at the same time than the previous true trip for this station (short-turn trip 2). Note also that the μ_1-parameter considered for the last trip of the lines causes, in this case, that the excess of passengers is equal to 0 at the end of the planning horizon.

The departing times of each true trip of all the lines from each of the stations are detailed in Figure 3. The horizontal axis is the time horizon, while the vertical axis represent spatial lengths (stations). We also report over each of the itineraries, their optimal capacities. Shortest itineraries running only over a subset of stations are short-turns. One can observe that the time difference between consecutive trips is not constant, as expected from the asymmetry of the lines and the transfer of passengers between them at different time instants. Then, as mentioned before, we deal with an aperiodic timetabling.

Furthermore, with the solution obtained, the estimated demands at the departing times of each one of the trips at the interchange station are drawn in Figure 4. Observe that in that picture we draw in the horizontal axis, the departing times for such an interchange station at the different trips. In the vertical axis we depict the accumulated demand at those time instants. Note that the jumps in the demand induced by passengers that interchange to that line occur

k: Capacity	i	DepTime	Get-Off	$f_i^{k\ell}$ $(g_i^{k\ell})$	$h_i^{k\ell}$	$x_i^{k\ell}$	Load
1: 800	1	07:30:00	0.00	50.00	0.00	0.00	50.00
	2	07:33:30	20.00	400.00	0.00	0.00	430.00
	3	07:35:00	257.50	627.50	101.50	101.50	800.00
	4	07:37:30	746.13	725.00	0.00	0.00	778.88
	5	07:40:30	778.88	0.00	0.00	0.00	0.00
2S: 1600	2	07:39:34	0.00	606.94	0.00	0.00	606.94
	3	07:41:04	364.17	1231.59	0.00	0.00	1474.36
	4	07:43:34	1474.36	0.00	638.18	0.00	91.93
3: **0**	1	07:30:00	0.00	0.00	0.00	0.00	0.00
	2	07:39:34	0.00	0.00	0.00	0.00	0.00
	3	07:41:04	0.00	0.00	0.00	0.00	0.00
	4	07:43:34	0.00	0.00	638.18	0.00	0.00
	5	07:40:30	0.00	0.00	0.00	0.00	0.00
4S: 800	2	07:43:12	0.00	364.02	0.00	0.00	364.02
	3	07:44:42	218.41	603.05	0.00	0.00	748.66
	4	07:47:12	748.66	0.00	1014.15	0.00	48.35
5: **0**	1	07:30:00	0.00	0.00	0.00	0.00	0.00
	2	07:43:12	0.00	0.00	0.00	0.00	0.00
	3	07:44:42	0.00	0.00	0.00	0.00	0.00
	4	07:47:12	0.00	0.00	1014.15	0.00	0.00
	5	07:40:30	0.00	0.00	0.00	0.00	0.00
6: 1600	1	07:46:15	0.00	162.60	0.00	0.00	162.60
	2	07:49:45	65.04	655.05	0.00	0.00	752.61
	3	07:51:15	449.94	1297.33	0.00	0.00	1600.00
	4	07:53:45	1494.25	1494.25	109.44	109.44	1600.00
	5	07:56:45	1600.00	0.00	0.00	0.00	0.00
7: 800	1	07:50:00	0.00	37.40	0.00	0.00	37.40
	2	07:53:30	14.96	373.99	0.00	0.00	396.43
	3	07:55:00	237.48	641.05	0.00	0.00	800.00
	4	07:57:30	747.38	446.03	0.00	0.00	498.65
	5	08:00:30	498.65	0.00	0.00	0.00	0.00

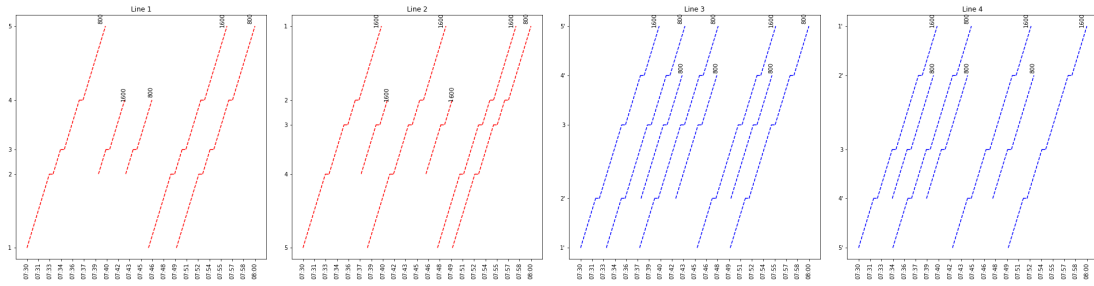TABLE 6. Results for the first line of the Case study 5.

FIGURE 3. Timetables of the different trips for each line of the Case study of Section 5.

during the whole time interval between departing times of the interchange station, but we only account for the accumulated amount when the new train departs.

Costs/Rewards	Line 1	Line 2	Line 3	Line 4	All lines
$(\mathrm{Cap}(\ell))$ Complete Lines	409.08	294.00	294.00	294.00	1291.08
$(\mathrm{Cap}(\ell))$ Short-Turns	102.27	55.99	66.00	66.00	290.26
$(\mathrm{RewPPass}(\ell))$ Complete Lines	1736.16	2514.53	3365.67	2814.05	100430.41
$(\mathrm{RewPPass}(\ell))$ Short-Turns	541.70	534.52	743.69	780.74	2448.67
$(\mathrm{NonServed}(\ell))$	44.85	0	0	0	44.85
OVERALL COST	-1721.65	-2699.06	-3749.37	-3234.79	**-11404.88**

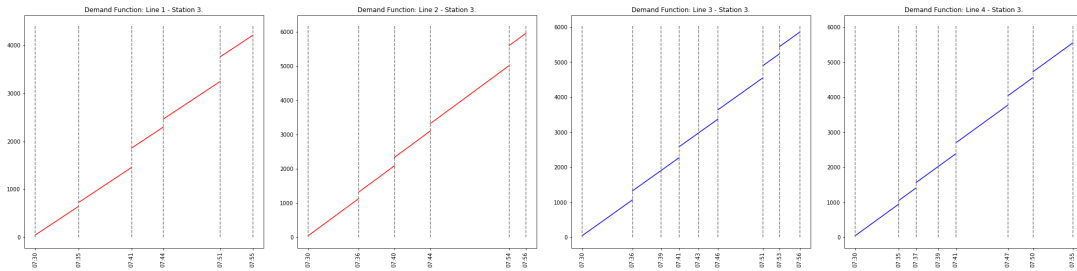TABLE 7. Summary of solution obtained for the Case study of Section 5.

FIGURE 4. Accumulated Demands obtained for the Case study of Section 5.

A summary of the costs and rewards obtained for the reported solution is given in Table 7. Since an aggregated function of the costs is minimized in our model, the negative overall cost obtained (-11404.88) can be seen as a positive global reward for the optimal planning.

Problem (P) depends on the demand that occur in each line and on the correlation between lines induced by interchanging stations. For that reason, to have an exact model the demands by lines must be considered as variables and hence, the problem that considers all the lines at the same time becomes very hard by obvious reasons. Among them, the problem is a Mixed Integer Non Linear Programming Program whose continuous relaxation is neither convex nor concave because of the bilinear constraints in ($\mathrm{D_E}$) and ($\mathrm{D_I}$) (which are needed to be linearized, turning into weak continuous relaxations of the problem); and in addition, the number of variables increases considerably with each line that is jointly considered in the problem. Therefore, in order to provide an efficient procedure which is able to handle the model for realistic instance sizes in reasonable time (recall that the pilot Case study 5 required 12 hours of CPU time to obtain a nearly optimal solution), we propose a Math-Heuristic approach. This algorithm solves sequentially Problem (P) for one line, say $\ell \in L$, at a time (by fixing appropriately the demand parameters after each run) to approximate the actual solution of the global problem. We will denote such a problem as Problem (P)$_\ell$. This Math-Heuristic approach is developed in the following section.

6. A MATH-HEURISTIC APPROACH

In this section we propose an optimization-based heuristic approach to solve (P). Note that the MINLP formulation stated for the problem involves many continuous and discrete variables, as well as a large number of constraints. In particular, the number of δ-variables that allows us to model the internal jumps of the demand function, described in the previous section, is $O(K^2 N^{Ic} L^2)$, where K is the maximum number of trips and N^{Ic} is the number of interchange stations. That bound implies a high complexity to solve even small-size instances of the problem.
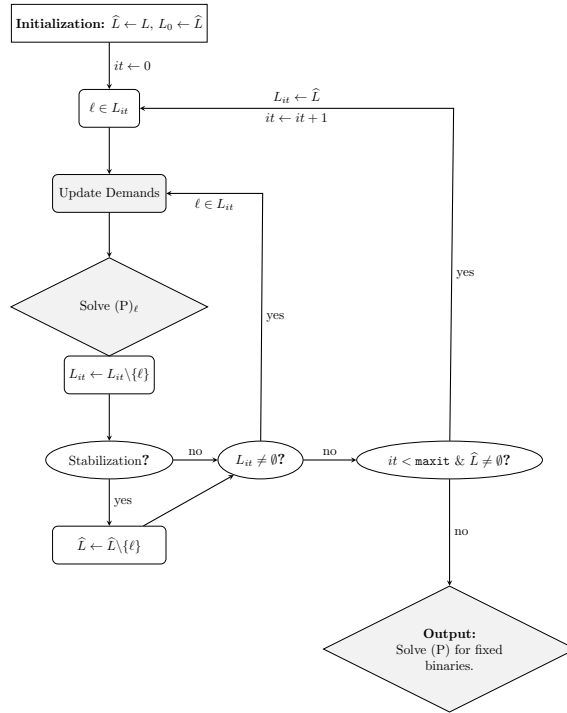
FIGURE 5. Flowchart of the Math-Heuristic approach.

.

In our algorithm we apply a *divide-and-conquer* strategy by splitting the models by lines, that is, we solve sequentially the model for a single line $\ell \in L$ at a time. The output of that solution together with the solutions of those previously solved are passed to the next line to be solved. Hence, the flow information provided by the lines in $L\backslash\{\ell\}$ (transfer passenger flows and train arrival times at the interchange stations) are considered as input for solving the problem $(P)_\ell$ for the fixed line $\ell \in L$. In this way, the demand function is updated and a new line is solved using the same approach. This procedure continues until a stopping rule evaluating the progress of the overall costs is satisfied, i.e. when the solution stabilizes or until a maximum number of iterations is reached. This approach allows us to avoid considering all the δ-variables together in a single model, which is one of the main drawbacks of the exact algorithm. A compact flow chart of the algorithm is drawn in Figure 5.

In the initial iteration we consider the whole set of lines to solve, $\widehat{L} = L$, and the time instants in which the demand is affected by passengers coming from an interchange station, $\mathbf{S^I}$ (it is assumed that, for the first line solved, there are no passengers coming from other lines). Then, for each line $\ell \in L$, we update the demand function taking into account the flows and breakdown times obtained when solving the previous lines and then, we solve the problem for the single line ℓ with such an input information. Observe that after the initial line is solved (in which the connecting flows are set to zero), the optimal flows give us information about the passengers that get on and get off from the trains of the line, what affect the rest of the lines.

Once all the lines have been solved, we repeat the procedure `maxit` times, to avoid getting trapped in locally optimal solutions. Let $z^{\ell(it)}$ be the objective function value of problem $(P)_\ell$ at iteration it. If at an iteration $it <$ `maxit`, this value stabilizes, that is, the relative deviation

in the objective value with respect to the previous iteration of line $\ell \in L$, namely $\frac{|z^{\ell(it)} - z^{\ell(it-1)}|}{z^{\ell(it)}}$ is not significant enough (smaller than a tolerance ε), the corresponding solution for this line is kept fixed for the remainder iterations. Thus, the iterative procedure terminates either, after the iteration when the solution stabilizes with some degree of accuracy, ε, (for all the lines) or when a maximum number of iterations is reached. Finally, to obtain a feasible solution (upper bound), we solve (P) by fixing the binary variables obtained during the heuristic procedure. Such a problem becomes a continuous Linear Programming problem which is efficiently solved.
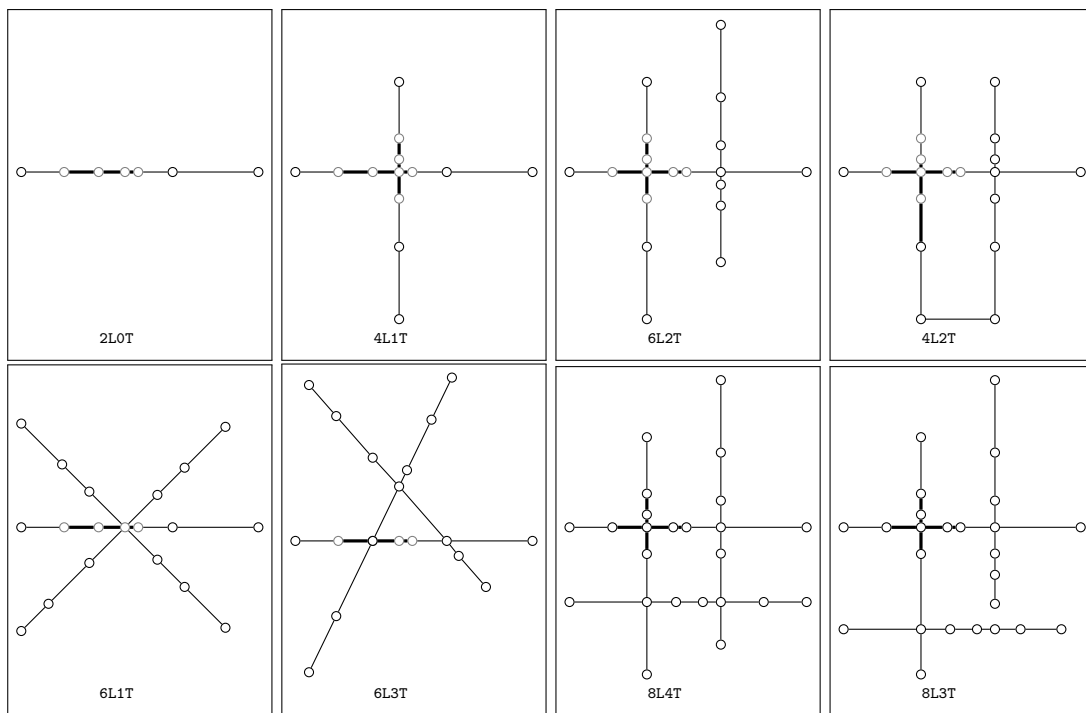
7. Experiments

Figure 6. Network Topologies for the Experiments.

We have run a series of computational experiments in order to test the model and the performance of our two approaches. We have considered the networks topologies drawn in Figure 6. The names of the networks are in the form $XLYT$ where X is the number of lines and Y the number of interchange stations. We consider networks with $2, 4, 6$ and 8 lines, each of them with 7 stations except 4L2T that contains 14 stations. For each of the 8 networks, we tested our approach in two situations: no short-turns are allowed (that is, with our notation $|LS| = 0$ and short-turns are allowed in some of the lines (thickest lines/gray nodes correspond with allowed short-turns for the lines) to evaluate the effect of short-turning. The construction of the different networks is motivated by the representation of the most common situations in real-world subway networks (see [17]). The description of the networks is provided in Table 8.

Our set of networks include the one proposed by Metrolab® (4L1T) to analyze the viability of automatizing the management of the Paris subway network. Actually, the parameters of these networks were designed based on that case and are available at `http://bit.ly/InputData_`

Network	Description
2L0T	One bidirectional line with no interchange stations.
4L1T	Two bidirectional lines with a single interchange station.
6L2T	Three bidirectional lines. One of the bidirectional lines with two interchange stations, each of them interchanging for a different bidirectional line (2-by-2).
4L2T	Two bidirectional lines. One of the bidirectional lines with two interchange stations, each of them interchanging from the same second bidirectional line.
6L1T	Three bidirectional lines with one interchange station for different lines (3-by-3).
6L3T	Three bidirectional lines with three 2-by-2 interchange stations.
8L4T	Four bidirectional lines with four 2-by-2 interchange stations.
8L3T	Four bidirectional lines with three 2-by-2 interchange stations.

TABLE 8. Description of the network topologies considered in our computational experiments..

Network	|**LS**|	**Math-Heuristic**		**MINLP** (P)		
		BestObj	**CPU**	**BestObj**	**CPU**	**GAP** (%)
2L0T	0	145702	< 0.1	145573	7	0.09
	2	114145	11	112916	TL	1.08
4L1T	0	206729	132	206242	TL	0.24
	4	152890	631	152016	TL	0.57
4L2T	0	348267	3522	347224	TL	0.30
	4	333102	4892	332665	TL	0.13
6L1T	0	276961	1130	276545	TL	0.15
	2	235488	1521	234589	TL	0.38
6L2T	0	249038	606	248854	TL	0.07
	4	203080	2882	203080	TL	0.00
6L3T	0	248988	520	248979	TL	< 0.01
	2	217004	2688	216909	TL	0.04
8L3T	0	404627	432	404147	TL	0.12
	4	362916	1467	362908	TL	< 0.01
8L4T	0	404469	1191	404211	TL	0.06
	4	374067	1144	374064	TL	< 0.01

TABLE 9. Computational Results

SubwayPlanning. All the models were coded in Python 3.6, and solved using Gurobi 8.0 [23] in a Mac OSX with an Intel Core i7 processor at 3.3 GHz and 16GB of RAM.

7.1. **Results.** The results of our computational experience are shown in Table 9. In this table the results are organized in five blocks of columns. The first block reports the network topology (using the name convention explained above). The second block , "|**LS**|", stands for the number of allowed short-turns on the considered network. The third block gathers the results obtained with the Math-Heuristic approach described in Section 6, whereas the fourth one reports the corresponding results obtained with the exact MINLP Formulation described in Section 3. In both cases, "**BestObj**" is the value of the objective function and "**CPU**" is the time, in seconds, required to meet the stopping criterion or to reach the time limit (a maximum CPU time of 10 hours has been set for solving the MINLP formulation), respectively. Those instances for which the time limit was reached are indicated with TL in that column, in whose case, the best obtained solution is reported. Finally, the block "**Gap** (%)" is the percentage gap between the objective function value obtained in the Math-Heuristic approach and the objective function value obtained when solving the MINLP Formulation. It is worth noting that to improve the performance of

the MINLP solver, we have initialized it with an initial solution given by the one provided by the Math-Heuristic approach. This implies that the solution of the exact approach is always as good as the Math-Heuristic and the %Gap is a measure of the improvement given by the exact method over the proposed heuristic.

The results reported in Table 9 show the remarkable performance of our Math-Heuristic algorithm. In all cases, it achieves rather good solutions with gaps with respect to the exact method smaller than 1.08% and in rather competitive computing times (one order of magnitude smaller than the exact MINLP solver). Note that the time limit of 10 hours was reached in all the instances except in the simplest one. The results are particularly exceptional in the most complex topologies using eight lines where the improvement of the exact method is negligible: gaps are smaller than 0.12% and computing times of the Math-Heuristic are, on average, around 2% of those required by the MINLP. Based on our experiment we conclude that the Math-Heuristic algorithm is a good compromise to solve the line planning and timetabling problem considered in this paper. Apart from the relative deviations between the solutions obtained with the math-heuristic and the exact approach, Gurobi was not even able to solve the problems within the time limit in all the instances except the simplest one, while our math-heuristic is able to obtain good quality feasible solution of the problem in reasonable CPU times. Finally, it is worth mentioning that the consideration of short-turns in the network considerably increases the computational effort needed to solve the problem, but at the same time reduces the overall cost considered in our objective function (an average reduction of 15% was observed under the considered parameters). For instance, the exact approach required 7 seconds to solve to optimality the instance 2L0T, without short-turns. However, it was not able to certify optimality for the same instance with short-turns, within the time limit of 10 hours. The comparison with respect to the Math-Heuristic is less dramatic and we report an average relative deviation of 57% in CPU time between the instances with and without short-turnings.

8. Conclusions and Future work

This paper considers a general model for line planning and timetabling problems on subway networks that was originated by a collaboration with Metrolab®, a French R&D company analyzing the viability of automating the metro of Paris. We propose a new Mathematical Programming- based decision making tool in this context. A number of different elements have been taken into account in the construction of both, objective function and constraint set of the proposed integrated line planning and timetabling model. The model aims to reflect as most real factors as possible of this complex transportation environment, but still being possible its computational suitability. Our outcome is a Mixed Integer Non Linear Programming formulation taking into account several quality and operation measures for feasible planing. We incorporate to the model an integrated cost- and passenger-oriented objective function, time-dependent demands and interchange stations together with several strategic decisions on the planning, as the starting times for each trip, the activation of short-turns, the determination of the number of trips in a journey, the selection of capacities for the trips, and the determination of the optimal number of trips. This representation of the problem may allow us to use the optimization model as a (*what-if*) tool to assess potential technological or managerial innovations that may be introduced into the system. In order to assess such innovations, one can decouple the whole transportation system and consider just a (small) section of it where the main effects of such a new element produce a higher impact. Such a reduction process resembles the one followed in our paper to illustrate the case study. Apart from using a MINLP solver for solving small instances of the problem, we also develop a novel Math-Heuristic algorithm which allows us to solve realistic instances. Extensive computational experiments on different network topologies based on the one provided by Metrolab® on the metro of Paris, show that the Math-Heuristic

algorithm performs remarkably well as compared to the exact resolution of the Mathematical Programming formulation, being an adequate tool to solve this type of problems.

Many future enhancements of the contributions made here could be cited now. Some of them could refer to the integration of new elements to the model such as variable speed for the trains or more general demand functions to model the number of passengers awaiting at the stations. Also, the stochastic nature of the parameters of the problem could be managed by incorporating the uncertainty on the passenger flows to the problem, and then using tools from Stochastic Programming to derive a model and different solution approaches for the problem. On the other hand, an interesting research line could be to obtain theoretical conditions under which one was able to ensure the convergence of our iterative numerical approach. It would be interesting to find conditions ensuring the convergence of the Math-Heuristic approach, even in simpler models, since they may give rise to new versions of the algorithm integrating different stopping rules or ways to decouple the system in parts.

References

[1] E. Barrena, D. Canca, L.C. Coelho, and G. Laporte. Exact formulations and algorithm for the train timetabling problem with dynamic demand. *Computers and Operations Research*, 44:66–74, 2014.

[2] D.P. Bertsekas. *Nonlinear Programming, 2nd ed.* Athena Scientific, Belmont, Massachusetts, 1999.

[3] V. Blanco, J. Puerto, A.B. Ramos. Expanding the Spanish high-speed railway network. *Omega*, 39:138–150, 2011.

[4] S. Bunte and N. Kliewer. An overview on vehicle scheduling models. *Public Transp*, 1:299–317, 2009.

[5] D. Canca, E. Barrena, G. Laporte, and F.A. Ortega. A short-turning policy for the management of demand disruptions in rapid transit systems. *Annals of Operations Research*, 246:145–166, 2016.

[6] D. Canca, A. Zarzob, E. Algaba, and E. Barrena. Confrontation of different objectives in the determination of train scheduling. *Procedia Social and Behavioral Sciences*, 20:302–312, 2011.

[7] A. Caprara, M. Fischetti, and P. Toth. Modeling and solving the train timetabling problem. *Operations Research*, 50(5):851–861, 2002.

[8] A. Caprara, M. Monaci, P. Toth, and P.L. Guidac. A lagrangian heuristic algorithm for a real-world train timetabling problem. *Discrete Applied Mathematics*, 154:738–753, 2006.

[9] C.E. Cortés, S. Jara-Díaz, and A. Tirachini. Integrating short turning and deadheading in the optimization of transit services. *Transportation Research Part A*, 45:419–434, 2011.

[10] Andrea Dariano, Dario Pacciarelli, and Marco Pranzo. A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research*, 183(2):643 – 657, 2007.

[11] G. Desaulniers and M.D. Hickman. Public transit. In G. Laporte and C. Barnhart, editors, *Transportation*, Handbooks in Operations Research and Management Science, chapter 2, pages 69–127. INFORMS, 2007.

[12] M. Goerigk and M. Schmidt. Line planning with user-optimal route choice. *European Journal of Operational Research*, 259(2):424–436, 2017.

[13] M. Goerigk and A. Schöbel. Improving the modulo simplex algorithm for large-scale periodic timetabling. *Computers & Operations Research*, 40(5):1363–1370, 2013.

[14] V. Guihaire and J.-K. Hao. Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10):1251–1273, 2008.

[15] M. Heidari, S.M. Hosseini-Motlagh, and N. Nikoo. A subway planning bi-objective multi-period optimization model integrating timetabling and vehicle scheduling: a case study of Tehran. *Transportation*, 2018.

[16] L. Kanga, X. Zhua, H. Suna, J. Puchingerc, M. Ruthmaire, and B. Huf. Modeling the first train timetabling problem with minimal missed trains and synchronization time differences in subway networks. *Transportation Research Part B*, 93:17–36, 2016.

[17] G. Laporte, J.A. Mesa, F.A. Ortega, and F. Perea. Planning rapid transit networks. *Socio-Economic Planning Sciences*, 45:95–104, 2011.

[18] G. Laporte, F.A. Ortega, M.A. Pozo, and J. Puerto. Multi-objective integration of timetables, vehicle schedules and user routings in a transit network. *Transportation Research Part B*, 98:94–112, 2017.

[19] Y. Lee and C.Y. Chen. A heuristic for the train pathing and timetabling problem. *Transportation Research Part B*, 43:837–851, 2009.

[20] J.A. Mesa, F.A. Ortega, and M.A. Pozo. Effective allocation of fleet frequencies by reducing intermediate stops and short turning in transit systems. In R.K. Ahuja, R.H. Möhring, and C.D. Zaroliagis, editors, *Robust and Online Large-Scale Optimization*, Lecture Notes in Computer Science, Vol. 5868, pages 293–309. Springer-Verlag, Berlin Heidelberg, 2009.

[21] M. Michaelis and A. Schöbel. Integrating line planning, timetabling, and vehicle scheduling: A customer-oriented heuristic. *Public Transport*, 1(3):211–232, 2009.

[22] H. Niu, X. Zhou, and R. Gao. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. *Transportation Research Part B*, 76:117–135, 2015.

[23] Optimization and Gurobi. Gurobi optimizer reference manual, 2018.

[24] F. Ortega, M.A. Pozo, and J. Puerto. On-line timetable rescheduling in a transit line. *Transportation Science*, 52(5):1035–1296, 2018.

[25] A. Schöbel. Line planning in public transportation: Models and methods. *OR Spectrum*, 34:491–510, 2012.

[26] A. Schöbel. An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies*, 74:348–365, 2017.

[27] L. Sun, J.G. Jin, D.H. Lee, K.W. Axhausen, and A. Erath. Demand-driven timetable design for metro services. *Transportation Research Part C*, 46:284–299, 2014.

[28] S. Tekin, S. Köfteci, M.M. Aydin, and M.S. Yildirim. Trip optimization for public transportation systems with linear goal programming (lgp) method. *Sigma Journal of Engineering and Natural Sciences*, 36(4):921–933, 2018.

[29] W. Weerawat and K. Chumkad. A new operations approach for bangkok metro green line using short turn operation patterns. *Journal of Rail Transport Planning and Management*, 8:207–219, 2018.

[30] R.C..W. Wong, T.W.Y. Yuen, K.W. Fung, and J.M.Y. Leung. Optimizing timetable synchronization for rail mass transit. *Transportation Science*, 42(1):57–69, 2008.

IEMath-GR, Universidad de Granada.
E-mail address: vblanco@ugr.es

Dep. Statistics & OR, Universidad de Sevilla.
E-mail address: educon@us.es

Dep. Applied Economics I, Universidad de Sevilla.
E-mail address: yhinojos@us.es

Dep. Statistics & OR, Universidad de Sevilla.
E-mail address: puerto@us.es