Confrontation of Different Objectives in the determination of train scheduling

David Canca*, Alejandro Zarzo, Encarnación Algaba, Eva Barrena

Abstract

In the railway networks management context, the determination of train schedules is a topic which affects both the level of satisfaction amongst the users and the network performance and profitability. This double influence makes it a widely studied topic in the literature, where nowadays the main lines of research tend to improve the solving methods of the corresponding integer programming problems. However, literature about methods that take both user and service provider point of view jointly into account is sparse. In this work, we tackle the problem of scheduling in middle and long distances networks by means of a non-linear integer programming model which fits the schedules to a dynamic behavior of demand and represents a trade-off between some measures of quality of the service offered and aspects related to the network profitability. The confrontation of different objective selection policies is analyzed in depth in the core part of this study, with the intention of both improving the insight into the proposed model and adding the flexibility of choosing different objectives by knowing in advance their influence on the resulting schedule plan.

Keywords: train scheduling; quality of service; mixed-integer nonlinear programming

1. Introduction: The problem

One of the key aspects of the management of railway networks, widely studied in the literature, is the determination of train schedules, since they would significantly affect both the level of user satisfaction and the network performance and profitability. Currently, in this context, many efforts focus on the process of solving integer programming models. This issue, widely discussed (see Cacchiani, 2006), has specific characteristics depending on the type of service to be provided. So, for short distance networks such as metro or local ones, the frequency is usually high enough, to specify the frequency of service becomes more important than the hourly schedule itself (see Higgins and Kozan, 1998). Railway timetabling is usually the third stage in a hierarchical schema composed by five consecutive steps: Analysis of demand, line planning, scheduling, rolling-stock and crew management (Bussieck et al., 1997, Bussieck, 1998). In the train schedule phase, all arrival and departure times are obtained. This task is usually done subject to the periodicity of the system. Periodic scheduling, initially proposed...
by Voorhoeve, 1993, based on the formulation of Serafini and Ukovich, 1998, has been followed by other authors like Natchtigall, 1994, Natchtigall and Voget, 1997, Liebchen and Möhring, 2002, Liebchen and Peeters, 2002, 2009, and Chierici et al., 2004. Periodic timetables have some advantages in the case of Metropolitan railways since a simpler service is allowed. In fact, regular timetables are easily memorized by users and also can be computed with less effort (Wardman et al., 2004). On the contrary, in the case of middle and long distance networks, the demand is usually captive to the schedules (see Pham 2004, Vansteenwegen and van Oudheusden, 2006). In this kind of scenarios, when passenger demand seems to be dependent on railway scheduling, the periodic regular approach may not be efficient. Even in the situation of subway and local railways, when passenger demand is considered as one of the main factors to determine timetables, periodic scheduling is suboptimal from a user’s perspective. However, an inefficient design of them could lead to undesirable behaviors of the users, whom could choose an alternative transport mode (with a consequent drop of network profitability). Non-periodic timetabling is especially important in long corridors with high traffic densities. This approach allows for obtaining optimal departure and arrival times after the line planning step, see Kwan and Mistry, 2003, Caprara et al., 2006 or Ingolotti et al., 2006.

It should be mentioned that most of the existent approaches do not make realistic hypothesis about the behavior of the demand of each line over a full day of operation. In fact, the analysis of the demand is commonly done only in the first stage of the planning process, generating a full day origin-destination matrix that is used to determine frequencies in the line planning phase. In any case, few authors deal with the problem of relating the scheduling of the units to the quality of service (Natchtigall and Voget, 1997, Vansteenwegen and van Oudheusden, 2006) and the actual capacity of the network (see Burdett and Kozan, 2006 for an analytical approach, Abril et al., 2008, Canca, 2009 and Canca et al., 2009 for several optimization approaches and Barber et al., 2007 for a description of simulation based systems). This is precisely the problem we face in this paper in which we analyze both aspects jointly. So, we first propose a method to characterize the demand in an approximate way by adjusting aggregate demand functions along the whole planning horizon. Secondly, a non-linear integer programming model to suit the timetable to a dynamic behavior of demand is presented. The model enables the analysis of measures of the quality of service, such as average waiting times and the number of passengers waiting, in contrast to measures of the network performance and profitability, such as the number of services or the train capacity. A final analysis of the profitability of the offered services for each line requires a global study of the network. Some aspects like maintenance, rolling stock and crew management should be taken into account (Lindner and Zimmermann, 2005). The complexity of the global problem leads to the search for certain measures which would help to select an appropriate scheduling, subject to a global economic post-study, in charge of the service provider.

2. Cumulative passenger demand

Usually, the demand mobility is characterized by means of the so-called Origin-Destination matrices. These kinds of matrices are commonly used in transportation planning analysis, where planners work with different matrices for a design day, each one for a defined time interval, e.g., the peak hour matrix or the total daily demand matrix. The elements of these matrices are usually obtained by means of more or less rigorous extrapolation techniques applied to the data given by mobility surveys (of course in a discrete way). Let \( OD(t) \) be a generic continuous representation of the mobility demand along the daily planning horizon. So, \( OD(t) \) contains information to build the Origin-Destination matrix for a concrete interval.

\[
OD(t) = 
\begin{bmatrix}
0 & f_{12}(t) & \cdots & f_{1S}(t) \\
\vdots & \ddots & \ddots & \vdots \\
f_{i1}(t) & \cdots & 0 & f_{iS}(t) \\
f_{S1}(t) & f_{S2}(t) & \cdots & 0 \\
\end{bmatrix},
\]

(1)

Here, the matrix entry \( f_{ij}(t) \) represents the continuous daily evolution of the travel demand from station \( i \) to \( j \), and \( S \) stands for the total number of stations of a specific line in the railway network (or the full network itself). Although these demand curves can quite differ from a problem to another, a common and relevant characteristic of all of them is the existence of certain demand peaks (local maxima) in certain instants of time along the day. These peaks are associated to rush hours and generally reduced to two or three per day (see Figure 1). In our approach to the
passengers demand, instead of these travel demand functions, we consider the corresponding cumulative or
aggregate demand

\[ F_{ij}(t) = \int_{0}^{t} f_{ij}(s) \, ds, \]

so that, the number of passengers arriving at the \( i \)-th station bound to the \( j \)-th during the time interval \([t_1, t_2]\) is given by,

\[ F_{ij}^{[t_1, t_2]} = F_{ij}(t_2) - F_{ij}(t_1) = \int_{0}^{t_2} f_{ij}(s) \, ds - \int_{0}^{t_1} f_{ij}(s) \, ds = \int_{t_1}^{t_2} f_{ij}(s) \, ds, \]

as it is shown in Figure 1. In this figure can be observed how the variation of the demand changes over time.

![Figure 1. Demand and cumulative demand functions](image)

As the demand is used to obtain train departure times, an analytical expression is needed. The shape of the cumulative demand between each pair of stations suggest an approximation given by a linear combination of a variable number, \( M \), of sigmoid curves.

\[ F_{ij}(t) = \sum_{r=1}^{M} \frac{K_{ij}'}{1 + e^{-\beta_{ij}'(t-x_{ij}')}} \]

The number of terms used in the approximation, i.e. \( M \), symbolizes the number of demand peaks along the day. In each sigmoid function, the parameters \( K_{ij}' \), \( x_{ij}' \) and \( \beta_{ij}' \) stand for the asymptotic value, the deviation with respect to the time and the slope, respectively. This approximation is fully characterized by a number of parameters (including \( M \)) which are determined by solving a set of appropriate least squares minimization problems. In other words, the dynamic adjust of these functions is made by minimizing the sum of quadratic errors as the difference between the proposed demand function \( F_{ij}(t_j) \) and the demand data \( y_{ij}(t_j) \) attained, in a standard day, for each pair of stations \( i, j = 1, \ldots, S, i \neq j \), on certain instants \( t_j \) depending on data availability.

3. Modeling the behavior of trains and passengers

In this section, an optimization model is presented in order to determine the train departure times from the origin station as well as the arrival/departure times at/from each one of the stations, for each train, along the line. As mentioned before, this approach, unlike other models in the existent literature, is based on the cumulative demand approximations described in the above section. We emphasize that this model can be applied to other kind of characterizations of the demand function, even it can be adaptable for its application to problems in which discrete demand functions are used, as the ones obtained typically from mobility surveys. The aim of this formulation is, not only the calculation of train arrival/departure times, but also obtaining measures that allow for the analysis of confronting different objectives.

In order to describe the optimization model, in the following subsections, the notation, the most relevant constraints and the objective function will be introduced.

3.1. Notation

Given a generic line, the notation of sets, parameters and variables, used in the model, is given in the following.
### Sets and parameters

- **$H := \{1, 2, \ldots, S\}$**: Set of stations
- **$N := \{1, 2, \ldots, K\}$**: Set of trains
- **$L := \{1, 2, \ldots, S - 1\}$**: Set of links between each pair of stations
- **$CAP_{\text{min}}$**: Minimum capacity of every train unit (number of seats)
- **$T$**: Planning horizon, usually one day (time units)
- **$\text{Long}(i,i+1)$**: Length of the link corresponding to stations $i$ and $i+1$
- **$\text{Mint}_{\text{stop}}$, $\text{Maxt}_{\text{stop}}$**: Lower and upper bounds for stop times
- **$\text{Mint}_{\text{seg}}(i,k)$**, $\text{Maxt}_{\text{seg}}(i,k)$**: Lower and upper bounds for headways

### Variables

- **$t^k_i$**: Departure time of train $k$ at the $i$-th station
- **$\overline{V}_i(i,i+1)$**: Inverse of speed of train $k$ in the link between stations $i$ and $i+1$
- **$FS^k_i$**: Number of free seats in the $k$-th train when it leaves the $i$-th station
- **$CAP^k_i$**: Capacity of train $k$
- **$t_{\text{stop}}(i,k)$**: Stop time of train $k$ at station $i$
- **$t_{\text{seg}}(i,k)$**: Headway after departure of train $k$ from station $i$

### Passenger arrival variables

- **$N_{ij}^k$**: Number of people who arrive at the $i$-th station with destination to the $j$-th station during the time interval $[t^k_i, t^k_{i+1}]$
- **$\text{NAD}^k_j$**: Number of people who arrive at the $i$-th station during the time interval $[t^k_i, t^k_{i+1}]$
- **$\text{NAO}^k_j$**: Number of people who arrive at any station $i$ ($i \leq j$) with destination to the $j$-th station during the time interval $[t^k_i, t^k_{i+1}]$
- **$ns^k_{ij}$**: Number of people who arrive at the $i$-th station with destination to the $j$-th station during the time interval $[t^k_i, t^k_{i+1}]$ and do not get on the $k$-th train
- **$ne^k_{ij}$**: Number of people who arrive at the $i$-th station with destination to the $j$-th station during the time interval $[t^k_i, t^k_{i+1}]$ and do not get on the $k$-th train

### Passenger waiting variables

- **$E^k_{ij}$**: Number of people who arrive at the $i$-th station with destination to the $j$-th station before $t^k_{i+1}$ (i.e., during the interval $[0, t^k_{i+1}]$) and are waiting on platform before the departure of the $k$-th train
- **$\text{EAD}^k_{ij}$**: Number of people who arrive at the $i$-th station before $t^k_{i+1}$ (i.e., during the interval $[0, t^k_{i+1}]$) and are waiting on platform before the departure of the $k$-th train
- **$\text{EAO}^k_{ij}$**: Number of people who arrive at the $i$-th station ($i \leq j$) before $t^k_{i+1}$ (i.e., during the interval $[0, t^k_{i+1}]$) and are waiting on platform before the departure of the $k$-th train with destination to the $j$-th station
- **$ee_{ij}^k$**: Number of people who arrive at the $i$-th station with destination to the $j$-th one, before $t^k_{i+1}$ (i.e., during the interval $[0, t^k_{i+1}]$) and do not get on the $k$-th train
- **$es_{ij}^k$**: Number of people who arrive at the $i$-th station with destination to the $j$-th one, before $t^k_{i+1}$ (i.e., during the interval $[0, t^k_{i+1}]$) and get on the $k$-th train

### Passenger getting on variables

- **$S^k_{ij}$**: Number of people who get on the $k$-th train at the $i$-th station with destination to the $j$-th one
Number of people who get on the $k$-th train at the $i$-th station

Number of people who get on the $k$-th train with destination to the $j$-th station

We assume known the length and the minimum and maximum speed limitations in every link, $l \in L$. As stated above, the aim of this paper is centered on obtaining measures related to the quality of service and the network profitability as well as the way of obtaining an adequate timetable. With this goal, we will fix some variables in the original model (see Canca, 2009) and focus the discussion on the set of constraints needed to obtain such kind of measures. Therefore, in this work, we will not include additional constraints to determine the number of trains needed or the exact train capacity (number or carriages that each train should have) because both are used here as experimental design parameters.

### 3.2. Constraints

The first set of equations allow us to compute the number of passengers who arrive at the $i$-th station, with destination to the $j$-th one to take the $k$-th train, as the difference between the cumulative demand between the instant $t_i^k$ and $t_i^{k-1}$, using the approximation for the cumulative demand described in section 2,

$$N_i^k = F_j(t_i^{k-1}, \hat{d}_i) - F_j(t_i^k, \hat{d}_i), \quad \forall i, j = 1, \ldots, S, \ i < j, \ k = 1, \ldots, K. \quad (5)$$

$$NAD_i^k = \sum_{[j \in H : j > i]} N_j^k, \quad \forall i = 1, \ldots, S, \ k = 1, \ldots, K. \quad (6)$$

$$NAO_j^k = \sum_{[i < H : i < j]} N_i^k, \quad \forall j = 2, \ldots, S, \ k = 1, \ldots, K. \quad (7)$$

On the other hand, the number of people who have arrived is the sum of those passengers who will manage to get on the $k$-th train ($n_i^k$) and those who will not achieve to do it ($e_i^k$). Notice that all posterior stations to the $i$-th station are considered as possible destination stations. Without losing generality, we assume that the equations will be written enumerating the stations in a consecutive order. In the reverse way, it is sufficient to express the sum for $j < i$, if the same ordinals are used. Therefore,

$$N_i^k = n_i^k + e_i^k, \quad \forall i, j = 1, \ldots, S, \ i < j, \ k = 1, \ldots, K. \quad (8)$$

The number of passengers who wait for train $k$, having arrived at the platform before $t_i^{k-1}$, is the sum of those who manage to get on train $k$ and those who do not achieve it, (equation (9)). Also, this variable should be aggregated by destination and origin, obtaining variables $EAD_i^j$ (equation (10)) and $EAO_j^k$ (equation (11)) necessary to balance people waiting between two consecutive trains.

$$E_i^k = n_i^k + e_i^k, \quad \forall i, j = 1, \ldots, S, \ i < j, \ k = 1, \ldots, K. \quad (9)$$

$$EAD_i^j = \sum_{[j \in H : j > i]} E_i^j, \quad \forall i = 1, \ldots, S - 1, \ k = 1, \ldots, K. \quad (10)$$

$$EAO_j^k = \sum_{[i < H : i < j]} E_i^k, \quad \forall j = 2, \ldots, S, \ k = 1, \ldots, K. \quad (11)$$

Hence, just at the moment in which train $k$ leaves the $i$-th station, the number of people who stay on the platform satisfy equations (12) and (13).

$$EAO_j^{k+1} = EAO_j^k + NAO_j^k - SAO_j^k, \quad \forall j = 2, \ldots, S, \ k = 1, \ldots, K - 1. \quad (12)$$

$$EAD_i^{k+1} = EAD_i^k + NAD_i^k - SAD_i^k, \quad \forall i = 1, \ldots, S - 1, \ k = 1, \ldots, K - 1. \quad (13)$$

In a similar way to waiting variables, the number of passengers who get on the $k$-th train is composed of those who have arrived in the interval $[t_i^{k-1}, t_i^k]$ and get on the $k$-th train and those who having missed the former train ($k-1$-th train) manage to get on this $k$-th train (from each pair $i, j$ of origin-destination stations), (see equation (14)). In this...
case, it also becomes necessary to aggregate people by destination, $SAD^k$, and origin, $SAO^k$, as it is shown in equations (15) and (16), respectively.

\begin{align}
S^k_i &= nS^k_j + eS^k_j, \quad \forall i, j = 1, \ldots, S, \ i < j, \ k = 1, \ldots, K. \\
SAD^k &= \sum_{\{j \in \mathcal{H} : j > i\}} S^k_j, \quad \forall i = 1, \ldots, S, \ k = 1, \ldots, K. \\
SAO^k &= \sum_{\{e > j : j \in \mathcal{H}\}} S^k_j, \quad \forall j = 2, \ldots, S, \ k = 1, \ldots, K.
\end{align}

Now, it is possible to balance the capacity of each train at each one of the stations, using the variables defined above and the free seat variable.

\begin{align}
FS^k_i &= FS^k_1 + SAO^k_i - SAD^k, \quad \forall i = 2, \ldots, S, \ k = 1, \ldots, K. \\
FS^k_i &= CAP_k - SAD^k, \quad k = 1, \ldots, K.
\end{align}

Finally, a relationship between arrival/departure times at/from stations along the line is required. Notice that here, we use the inverse of speeds in order to maintain linear expressions. Equation (20) ensures that the arrival of train $k + 1$ is carried out according to the safety specifications for each train and station (see also Equation (21)).

\begin{align}
t^{k+1}_i &= t^i + t_{\text{stop}}(i+1, k) + Long(i, i+1) \bar{V}_i(i, i+1), \forall k \in \mathcal{N}, \ i \in S. \\
t^{k+1}_i + t_{\text{stop}}(i+1, k) &\geq t^i + t_{\text{seg}}(i, k), \forall k \in \mathcal{N}, \ i \in S. \\
\text{Mint}_{\text{stop}}(i, k) &\leq \text{Max}_{\text{stop}}(i, k), \quad \text{Mint}_{\text{seg}}(i, k) \leq \text{Max}_{\text{seg}}(i, k), \forall k \in \mathcal{N}, \ i \in S.
\end{align}

3.3. Objectives and methodology

With the aim of obtaining a measure of the quality of the service from the user’s point of view, we propose the study of a nonlinear main-function that minimizes the total average waiting time (AWT) during the whole planning horizon (equation (22)). Other objectives, related to the profitability of the offered service, are also considered by the inclusion of constraints following the $\varepsilon$-constrains method (Marglin, 1967). So, for each line and a fixed global capacity, the optimum AWT is obtained by varying parametrically the maximum number of trains in which the global capacity is distributed, i.e., a Pareto frontier will be obtained for each fixed capacity (Figures 3 and 4).

\begin{align}
EAD^k, EAO^k, FS^k_i, t^i, NAD^k, NAO^k, SAD^k, SAO^k &\geq 0, \ \forall i, j = 1, \ldots, S, \ i \neq j, \ k = 1, \ldots, K \\
CAP_k &\geq 0, \ k = 1, \ldots, K, \quad E^i_j = 0, \ \forall i, j = 1, \ldots, S, \ i \neq j.
\end{align}

Moreover, in these scenarios, the average occupancy percentage per train (AOP) and the average percentages of people getting on each train (ATP), in comparison to people who arrive at each station, are calculated. The first one is, at the same time, a measure of quality of service in terms of comfort and a measure of resource usage (referred to the capacity). The second one is a measure of the revenue, the more people getting on, the greater the network profitability and attended demand. Therefore, both percentages are relevant from the user and the service provider point of view. The confrontation of these two measures and the resulting AWT is done in a second phase of the decision making process in order to provide a mechanism for selecting an appropriate global capacity level and the resulting scheduling as a function of the operating cost (Lindner and Zimmermann (2005)). The implementation of this model has been built in GAMS and solved using COINBONMIN (see Bonami et al., 2005).

4. Computational results

The proposed methodology has been tested in a piece of the C5 line belonging to the Madrid suburban railway network. Firstly, we have used surveys data (provided by RENFE) to adjust the cumulative demand functions among six stations: Móstoles Soto, Móstoles, Las Retamas, Alarcón, San José and Cuatro Vientos. As we mentioned
above, this adjust procedure is carried out from a minimization of the sum of quadratic errors obtained as the difference between the sigmoid approximation and the data obtained for a standard day. In all the experiments, we suppose carriages with a fixed number of seats, set to 40. The model is solved increasing the global capacity from 400 to 2000 seats (400 seats each time). Due to the discrete nature of the capacity, we first divide the total capacity by an integer number of wagons and then the obtained total carriage number is distributed in one, two, five and ten trains, respectively. Therefore, we obtain twenty timetables in form of time-space diagrams (see Lindner, 2000) where time is represented on the horizontal axis and the distance from the first station on the vertical axis. Figure 2 shows a couple of examples corresponding to the minimum and maximum capacity for five and ten trains. In order to analyze the full range of results, some extreme and unrealistic values are also included (e.g., one train and 50 wagons). Each picture shows simultaneously the number of trains, depicted as thick vertical lines, optimum departure/arrival times from/at each station as well as a part of the cumulative demand functions between each pair of stations. It can be observed how the departure times are closed to the demand peaks.

On Figure 3, the relationship among the total number of trains, the AWT and the AOP on certain points is given, whereas on Figure 4, instead of the above percentages, the ATP is shown.
In these figures, it is noticed that when the number of trains increases, then the AWT decreases. We can notice on the graphics that there exist situations in which fewer carriages, distributed in a larger number of trains, give way to a shorter AWT. This suggests an analysis of the different ways of distributing the global capacity mentioned above. On each curve, both percentages have the same behavior, i.e., both of them increase as the number of trains increases. However, a characteristic of these percentages is that they are opposed with respect to the total number of offered seats. Hence, for a constant number of trains, the bigger the total number of trains, the smaller the AOP and larger the ATP. Reaching a balance between both percentages would help to find appropriate solutions. With that aim, from Figures 3 and 4, both percentages will be represented on the \( z \)-axis, generating the two surfaces on Figure 5. The level curves of both surfaces as well as the curves representing the relationship between the total number of trains and the AWT are also plotted on the \( xy \)-plane.

Figure 6 shows the projections of both surfaces onto \( xy \)-plane. Points on the intersection curve are close to those on the curve associated with a total capacity of 30 carriages. On this curve, both percentages vary between 75% and 80% and when the number of trains increases, the AWT decreases. So, taking into account the AOP and the ATP, 30 carriages would be the most suitable global capacity. Starting at the Pareto frontier for 30 carriages, the service provider can determine the convenient number of trains considering the cost per train-km (see Lindner and Zimmermann, 2005) and the penalty of the increase in the AWT (related to the cost per passenger-km).
5. Conclusions

In this paper, we present a non-linear integer programming model to decide how many trains should be scheduled to attend a given demand and compute the departure time of each scheduled train for all their tour stops. We consider a demand continuous representation in terms of a continuous demand matrix for the full day operation instead of a set of origin-destination matrices. Each matrix element represents the evolution of the travel demand for a pair of stations along the daily planning horizon.

The cumulative demand corresponding to each of these functions is then approximated by a linear combination of sigmoidal functions. The parameters of each term are determined using a least square procedure. The optimal timetable determination uses these cumulative demand functions for each origin-destination pair. Next, we apply a to calculate trains arrival and departure times at each station. These times are adjusted according to the above described cumulative demand functions by means of passenger arrival behavior.
Setting the global capacity and following the \( \varepsilon \)-constrains method, Pareto frontiers are calculated for the bi-objective problem which takes into account the number of trains and the passenger average waiting time (AWT). This will discard non-efficient solutions and reduce the search field for an optimal solution. The model is also applied to compute different measures of quality of service from both a user’s and a service provider perspective. The average getting on each train (ATP), that represents the average number of people that can get on the first leaving train after their arrival and the average train occupancy (AOP), related to the profitability of each train service, respectively. A trade-off between both parameters provides an equilibrium curve that, in conjunction with the AWT, allow the service provider to take a decision on the adequate number of trains and the most suitable train capacity.

Acknowledgments

This research work has been partially supported by the Excellence program of the Andalusian Government, under grant P09-TEP-5022. We are also grateful to RENFE (Madrid suburban railway network) for providing data and information relevant to the realistic adjustment of the model.

References