SORT 38 (2) July-December 2014, 181-214

Integrating network design and frequency setting in public transportation networks: a survey

Francisco López-Ramos*

Abstract

This work reviews the literature on models which integrate the network design and the frequency setting phases in public transportation networks. These two phases determine to a large extent the service for the passengers and the operational costs for the operator of the system. The survey puts emphasis on modelling features, i.e., objective cost components and constraints, as well as on algorithmic aspects. Finally, it provides directions for further research.

MSC: 90B06.

Keywords: Public transport, network design, frequency setting, integrating.

1. Introduction

Rapid population growth in cities has led to traffic congestion. To alleviate this effect, transport agencies have designed public transportation systems, with their operating frequencies and resource capacities continuously being revised. Because of the high cost of construction and exploitation of this resources, it is important to pay attention to issues affecting effectiveness in different planning stages. For instance, the design of the layout of the lines must consider infrastructure budgetary restrictions and coverage demand satisfaction, whereas the line frequencies must be set so that passenger trip requirements are satisfied at reasonable operative costs while not exceeding resource capacities.

Traditionally, both planning phases have been solved sequentially, i.e., when the design of the layout of the lines is determined, the frequencies are assigned to these layouts. This approach may lead the public transportation system to operate in an inefficient manner because the network design phase assigns the demand to the lines without con-

* Pontificia Universidad Católica de Valparaíso, ITRA, 2147 Brasil Avenue, 2362804 Valparaíso, Chile.
francisco.lopez.r@ucv.cl
Received: April 2014
Accepted: May 2014

sidering resource capacities. Therefore, suboptimal designs and congestion are highly likely to be produced. However, an integrated approach may overcome this drawback.

State-of-the-art reviews on public transportation issues (Desaulniers and Hickman, 2007; Guihaire and Hao, 2008; Kepaptsoglou and Karlaftis, 2009; Farahani, Miandoabchi and Szeto, 2013) do not attach the proper importance to optimization approaches that integrate the network design and the frequency setting phases. Moreover, they do not cover the main modelling and solving features involved in these phases. The present review aims at fulfilling these gaps and suggesting lines for further research for both modelling and solving issues.

The rest of the paper is organized as follows. Section 2 defines the planning stages covered in this review. Section 3 reviews all the works which consider some modelling features from both planning stages. Finally, Section 4 presents the conclusions and lines for further research.

2. The transit planning process

The complete transit planning process is made up of the following five phases: network design, frequency setting, timetable development, vehicle scheduling and crew scheduling (Ceder and Wilson, 1986). Traditionally, the five phases have been solved sequentially due to the inefficiency of solving them simultaneously for large-sized networks. The two first phases (i.e., the network design and the frequency setting) determine to a large extent the service for the passengers and the operational costs for the operator of the system. Therefore, state-of-the-art research on transit planning processes has been mainly focused on these two phases.

2.1. The Network Design Problem

The Network Design Problem (*NDP*) involves the allocation of the new public transportation infrastructure, i.e., new stations and its interconnections following a prespecified network layout design rules. This problem considers the construction costs of the new infrastructure, i.e, the new stations, and also its interconnections for railway-based systems. Additionally, the number of new resources to be constructed is limited by an infrastructure budget. Finally, the existence of a network already in operation is considered, in the sense that infrastructure costs are not taken into consideration. Under these assumptions, the aim of the *NDP* is to cover as much demand as possible at reasonable infrastructure construction costs.

2.2. The Frequency Setting Problem

The Frequency Setting Problem (*FSP*) assigns a certain number of vehicles and its services to the existing lines plus the constructed lines so that the expected demand

is covered. The term "service" refers to a complete line cycle performed by a vehicle halting at some or all of the stations on the line. This problem considers the costs and the capacities of the planning resources (vehicles, platform stations, stations and stretches). The costs of the planning resources involve the power consumption costs of the vehicles, the costs of acquisition and maintenance of the vehicles and the salaries of its drivers, mainly. The capacities of the planning resources are related to the maximum number of passengers that a vehicle can hold, the number of available vehicles (fleet size), the maximum number of passengers that a platform station can hold, while they are waiting to board a vehicle, the maximum number of services per unit of time that a station can hold, and the number of vehicles per unit of time that can go through a stretch.

2.3. The passenger transit assignment model

The integration of the *NDP* and *FSP* problems requires a passenger transit assignment model. This model determines how the passengers use the constructed plus the existing lines through a detailed representation of the network. This representation considers in-vehicle traveling time, boarding and alighting from the vehicle times, at-station waiting time, in-vehicle waiting time and walking time. An economical cost is associated with each type of time previously cited so that the total passenger trip time can be compared with the operator costs.

3. Review on the network design and frequency setting problem

The literature on the integration of the network design and frequency setting phases is scant when considering the time elapsed from the first works Lampkin and Saalmans (1967), Silman, Barzily and Passy (1974) and the last work López (2014). Moreover, the research is not well developed because some important modelling features have been discarded and all the modelling features encountered are not considered in the same work. Finally and not least, the solving approaches are either inefficient or unreliable because the goodness of the solution is not guaranteed. In the following subsections, the review will focus the attention on these issues, leaving aside the modelling features not considered in any paper. The later issue will be analysed in Section 4.

3.1. Objective function costs

Tables 1-2 show the main objective function costs considered by the literature works. They are divided into two parts: the costs related to the operator and the costs associated with the users. The operator costs comprise the Infrastructure Resources and the Planning Resources (they are shown in columns 3 and 4), whereas user costs consist of the Travel Times, At-Station Waiting, In-vehicle Waiting, Transfer times, Vehicle Occupancy and Mode Disutility (they are hold on columns 5 to 11). The coloured tick

Table 1: *Objective function cost components considered in the literature (see also the next page).*

Year	Author(s)	Infrastructure Resources	Planning Resources	Unsatisfied Demand	Travel times	At-station Waiting	In-vehicle Waiting	Transfer times	Vehicle Occupancy	Mode Disutility
1967	Lampkin and Saalmans				✓	✓				
1974	Silman et al.				✓	✓			✓	
1979	Dubois et al.				✓	✓				
1981	Hasselström					✓		✓		
1884	Marwah et al.		✓		✓					
1986	Ceder and Wilson		✓		✓	✓		✓		
1987	Van Oudheusden et al.		✓							
1988	Van Nes et al.					✓				
1989	Israeli and Ceder		✓		✓	✓		✓	✓	
1990	Baaj and Mahmassani		✓		✓	✓		✓		
1992	Baaj and Mahmassani		✓		✓	✓		✓		
1992	Israeli		✓		✓	✓		✓	✓	
1994	Shih and Mahmassani		✓		✓	✓		✓		
1995	Baaj and Mahmassani		✓		✓	✓		✓		
1995	Israeli and Ceder		✓		✓	✓		✓	✓	
1996	Bussieck et al.			✓						
1998	Bielli et al.	✓	✓		✓	✓		✓	✓	
1998	Ceder and Israeli		✓		✓	✓		✓	✓	
1998	Pattanaik et al.		✓		✓	✓		✓		
1998	Shih et al.		✓		✓	✓		✓		
1999	Soehodo and Koshi		✓		✓	✓		✓	✓	✓
2000	Rao et al.		✓		✓	✓		✓		
2001	Caramia et al.	?	?	?	?	?	?	?	?	?
2001	Chien et al.		✓		✓	✓				
2002	Bielli et al.	✓	✓		✓	✓		✓	✓	
2002	Fusco et al.	?	?	?	?	?	?	?	?	?

Table 2: *Objective function cost components considered in the literature* (*Continued*).

Year	Author(s)	Infrastructure Resources	Planning Resources	Unsatisfied Demand	Travel times	At-station Waiting	In-vehicle Waiting	Transfer times	Vehicle Occupancy	Mode Disutility
2003	Chakroborty	?				✓		✓		
2003	Ngamchai and Lovell		✓		✓	✓				
2003	Tom and Mohan		✓		✓	✓		✓		
2003	Wan and Lo		✓							
2003	Zhao and Gan				✓	✓		✓		
2004	Agrawal and Mathew		✓		✓	✓		✓		
2004	Carrese and Gori				✓	✓				
2004	Fan and Machemehl		✓	✓	✓	✓		✓		✓
2004	Petrelli		✓		✓	✓		✓		
2005	Cipriani et al.		✓	✓	✓	✓		✓		✓
2005	Lee and Vuchic				✓	✓		✓		✓
2005	Hu et al				✓	✓		✓		
2006	Fan and Machemehl		✓	✓	✓	✓		✓		✓
2006	Zhao				✓	✓		✓		
2006	Zhao and Zheng				✓	✓		✓		
2007	Barra et al.		✓	✓	✓	✓		✓		✓
2007	Borndörfer et al.	✓	✓ / ✓ ?		✓					
2007	Zhao and Zheng		✓		✓	✓		✓		
2008	Fan and Machemehl		✓	✓	✓	✓		✓		✓
2008	Fernández et al.		✓		✓	✓		✓		
2009	Pacheco et al.				✓	✓				
2009	Marín et al.		✓	✓						
2011	Mauttone		✓		✓	✓		✓		
2011	Szeto and Wub				✓	✓		✓		
2012	Cipriani et al.		✓	✓	✓	✓		✓		✓
2012	Shimamoto et al.		✓		✓	✓	✓	✓		
2014	López	✓	✓	✓	✓		✓	✓		✓

marks indicate that the paper considers the feature partially (orange coloured tick mark) or totally (green coloured tick mark). In some papers, there are question marks denoting that it is unknown whether the feature is considered. This problem occurs when the author(s) do(es) not mention its use explicitly. In the following subsections, the reader can find the explanation of each cost with references to the literature works that considered these costs. Additionally, the alternative implementations of these costs are described if any.

3.1.1. Infrastructure resource costs

Infrastructure resources are related to line segments (also called stretches in the context of a railway based system) and stations. The costs of the stretches and the stations represent a great amount of the operational costs of the system operator. To mention one of them, the construction of one kilometer of stretch in the Spanish commuter train network ($RENFE$) costs between 1-1.6 M€ and is amortized between 30-60 years (Ferropedia, 2014a). So, it costs around 3.8-6 €/km-h. Moreover, if it is an underground system, we must also consider the construction of a tunnel whose value, according to $RENFE$, is significantly higher (around 30 M€/km). However, it is amortized in a larger period of time, 100 years, so it costs 34.25 €/km-h. Despite the importance of these costs, there are few works in the literature that consider the construction and/or maintenance costs of the stretches and stations of the new network infrastructure (Bielli, Carotenuto and Confessore, 1998; Bielli, Caramia and Carotenuto, 2002; Borndörfer, 2007; López, 2014).

3.1.2. Planing resource costs

The planning resources are associated with the public transportation vehicles and the drivers that control these vehicles. The costs of the vehicles and the drivers represent the same order of magnitude as the costs of the network infrastructure. For the $RENFE$, they costs around 26 €/train-km (Ferropedia, 2014b) and comprise the power consumption costs of the public transportation vehicles, the costs of acquisition and maintenance of these vehicles and the salaries of the drivers, mainly. Planning resource costs have been partially considered in the works of Agrawai and Mathew (2004), Barra et al. (2007), Baaj and Mahmassani (1990), Baaj and Mahmassani (1991), Baaj and Mahmassani (1995), Bielli et al. (1998), Bielli et al. (2002), Fan and Machemehl (2006), Fan and Machemehl (2008), Wan and Hong (2003), Marwah et al. (1993), Ceder and Wilson (1986), van Oudheusden et al. (1987), Israeli and Ceder (1989), Ngamchai and Lovell (1993), Pattnaik et al. (1998), Rao, Muralidhar and Dhingra (2000), Shih and Mahmassani (1994), Soehodho and Koshi (1999), Fan and Machemehl (2004), Fernández, de Cea and Malbran (2008), Mauttone (2011), Shimamoto, Schmöcker and Kurauchi (2012), Tom and Mohan (2003), and fully considered in López (2014), Marín, Mesa and Perea (2009), Cipriani et al. (2012), Zhao and Zeng (2007), Cipriani et

al. (2005), Petrelli (2004), Chien, Yang and Hou (2001), Shih, Mahmassani and Baaj (1998). The cost of acquisition and maintenance of vehicles are considered in the vast majority of the works; however, salaries and power consumption costs have been seldom considered. Additionally, there are some works which do not specify whether they use planning resource costs or which planning resource costs are used (Borndörfer, 2007; Fusco, Gori and Petrelli, 2002; Rao et al., 2000).

3.1.3. Unsatisfied demand

Urban public transportation networks aims at covering as much demand as possible. To penalize the uncovered demand, the vast majority of authors use a penalization weight in the objective function that increments the value of the objective function as the amount of unsatisfied demand increases (Barra et al., 2007; Bussieck et al., 1996; Cipriani et al., 2005 and 2012; Fan and Machemehl, 2004, 2006 and 2008; Marín et al., 2009). However, in López (2014) a pedestrian network that connects the O-D demand pairs directly is used. So the unsatisfied O-D pairs go through these links, with a high travel time costs. There are some additional works which do not specify whether they include some uncovered demand costs (Caramia, Carotenuto and Confessore, 2001; Fusco et al., 2002).

3.1.4. Travel time costs

Travel time costs refer to the in-vehicle and the boarding and alighting from the vehicle passenger times. These times are considered in the vast majority of works except in van Oudheusden et al. (1987), van Nes, Hamerslag and Immer (1988) and, possibly, in Caramia et al. (2001), Fusco et al. (2002). The last works do not specify whether they use the travel time costs. The travel times are generally computed using a time value associated with the network link that is expressed in units of time per person and the total amount of passengers going through the link. The overall time is weighted by a constant term which represents the passenger time cost perception. This constant plays an important role on the quality of the solution because it determines the importance of the main passenger costs in the optimum with respect to the operator costs. However, scant literature address that issue.

Mauttone (2011) has study the effect of the time weight by applying the interactive multi-objective optimization method (Ehrgott and Gandibleux, 2002). This method determines a set of non-dominated solutions which approximates to the optimal pareto front. A solution S1 dominates another solution S2, if S1 is no worse than S2 in all objectives and S1 is strictly better than S2 in at least one objective. So, in the study models, it means that all non-dominated solutions have passenger and operators costs which are equal or less than the ones of the dominated solutions and, the operator costs or the passengers costs of the non-dominated solutions are strictly less than the ones of the dominated solutions.

The conference presentation of Codina et al. (2008) also deals with multi-objective analysis but the aim was to determine a value to the cost of time such that all passengers will want to use the available network resources, no matter which operator cost is. Therefore, the authors did not seek for a set of non-dominated solutions but for a set of solutions where passenger costs have much more importance than operator costs.

3.1.5. At-station waiting costs

Waiting at a station is a very unpleasant situation for a user of the public transportation system. On average, the perception cost of the waiting time for a user is three times the time perception costs of the travel times. Thus, it is very important to consider at-station waiting times. Non-exact approaches have implemented these times because the module that evaluates them, assumes a fixed route configuration (i.e., its stretches, stations and operating frequencies are already determined) and, thus, the model is linear. However, in (quasi)-exact approaches the mathematical programming program formulated copes with some non-linearities. For instance, in the absence of congestion and link capacities (see Subsection 3.2.4), there is the product of frequencies and waiting times in particular constraints (Spiess and Florian, 1989). This fact discourages the authors of these works to face to waiting times. The reader is referred to Table 3 to see the distinct approaches of the literature works.

3.1.6. In-vehicle waiting costs

Passengers also experience some waiting when they are in a vehicle that is serving at a station, different from the station where they have boarded or alighted. This waiting time is increasingly significant as the vehicle takes more time in serving other passengers at the station. It is also influenced by the number of intermediate stations between the passenger boarding station and the passenger alighting station. Consequently, this waiting time component merits some consideration. There are just a few works in which this waiting time component is considered (López, 2014; Shimamoto et al., 1993). In both works, the number of passengers waiting in the vehicle are weighted by an average passenger time value per person and then the overall waiting time is penalized by a time cost that represents the passenger cost perception of waiting time.

3.1.7. Transfer time costs

Transfer time is related to the passenger walking due to changing between lines or when there are no stations that connects directly with its origin or destination. There are basically two ways of implementing transfer costs. One approach uses a penalty weight or a constant time associated with each transfer unit (Barra et al., 2007; Baaj and Mahmassani, 1990, 1991 and 1995; Pattnaik, Mohan and Tometc, 1998; Shih and Mahmassani, 1994; Shih et al., 1998; Soehodho and Koshi, 1999; Rao et al., 2000;

Tom and Mohan, 2003; Zhao and Ghan, 2003; Zhao, 2006; Zhao and Zeng, 2006 and 2007; Agrawai and Mathew, 2004; Fan and Machemehl, 2004, 2006 and 2008; Petrelli, 2004; Cipriani et al., 2005 and 2012; Mauttone, 2011; Szeto and Wub, 2011). The other approach attaches the proper time cost associated with a complementary network link, i.e., a pedestrian network (Bielli et al., 1998 and 2002; Ceder and Israeli, 1998; Ceder and Wilson, 1986; Chakroborty, 2003; Fernández et al., 2008; Hasselström, 1981; Hu et al., 2005; Israeli, 1992; Lee and Vuchic, 2005; López, 2014; Shimamoto et al., 1993). There is an additional approach which is worth-mentioning despite of being only related to the Network Design phase. García et al. (2006) considers as transfer time costs not only walking time cost between line platforms but also waiting time to board a vehicle in the transferred line platform. The walking times are considered constant, i.e., not depending in the distance between line platforms, whereas waiting at the transferred line platform is computed as the inverse of twice the frequency of the line, which is given as an input to the model.

3.1.8. Vehicle occupancy

Vehicle occupancy refers to the utilization level of the bus capacity. The bus capacity is an input parameter that is fixed according to a maximum number of seated passengers plus a maximum number of standees. The last amount is computed according to an allowable passenger density. The vehicle occupancy is a relevant feature for passengers because it dictates the comfortability of the passengers in the vehicle. The crowder is the vehicle, the less comfortable are the passengers. This feature is implemented using a penalisation term that weights the number of standees in a vehicle going through the segments of the operating line. Surprisingly, this feature is not considered in recent works (from 2003 until present).

3.1.9. Mode disutility

The mode disutility cost allows considering alternative modes of transportation. This cost is implemented using a combined modal splitting assignment model in which the disutility is expressed using a probabilistic function. The vast majority of works employ a multinomial logit function, except Fan and Machemehl (2004), Fan and Machemehl (2006), Fan and Machemehl (2008) in which a nested logit is used. The probabilistic function is usually employed in an iterative procedure where, first, the routes and frequencies of the transportation system are determined and, then, a network evaluation procedure uses the probabilistic function to evaluate several performance indicators of the built lines (see, for instance, Fan and Machemehl, 2004). These indicators are compared to the indicators of the alternative modes of transportation and, according to this comparison, the current network is modified and re-evaluated until some convergence criteria is met. There is only one work (López, 2014) in which the probabilistic function is indirectly expressed as a deterrence function. The author

demonstrates that in the optimal solution of the resulting bilevel program, the modal demand is distributed according to a logit function. Another interesting work, although applied only to network design, is Marín and García-Rodenas (2009) where the authors also use a logit function to represent the modal demand splitting but in a single level problem.

3.2. Modelling features

Tables 3-4 show the main modelling features considered by the literature works. The modelling features are divided into three parts: the features strictly related to the operator, the features strictly associated with the users and the features which correspond to both operator and passenger agents. The terms "strictly related to" and "strictly associated with" refer to the agent which mainly manages these features. The operator is strictly related to Infrastructure Restrictions, Working Lines, Stretch Capacity, Vehicle Fleet Size, Vehicle Capacity and Time Horizon features (shown in columns 3, 4, 6, 7, 8 and 9 of Table 2). Passengers are strictly associated with the % of Satisfied Demand (shown in column 10 of Table 2). The remaining feature, the Express Services, is related to both operator and passenger agents. Like in the preceding table, the coloured tick marks indicate that the paper considers the feature partially (orange coloured tick mark) or totally (green coloured tick mark). In some papers, there are question marks denoting that it is unknown whether the feature is considered. This problem occurs when the author(s) do(es) not mention its use explicitly. In the following subsections, the reader can find the explanation of each modelling feature with references to the literature works that considered these features. Additionally, the alternative implementations of these features are described if any.

3.2.1. Infrastructure budgetary restriction

As explained in the preceding Subsection 3.1.1, infrastructure resource costs are the leading costs for the operator of the system. Therefore, there is a limitation in the number of infrastructure resources that can be used to construct or expand the present public transportation network. Surprisingly, there are only two works that impose such a limitation (López, 2014; Marín et al., 2009). In both works, the number of stations and stretches that can take part of the new railway lines is subject to a infrastructure budget. The infrastructure resource costs and the budget are expressed as the currency value per unit of time. In some works focusing only on Network Design, that feature is more frequently found (see, for instance, Laporte et al., 2007 and 2001; Marín, 2007; Marín and Jaramillo, 2008 and 2009; Marín and García-Rodenas, 2009). The way the feature is modelled is the same as mentioned in the previous two works.

Table 3: *Modelling features considered in the literature works (see also the next page).*

Year	Author(s)	Infrastructure Restrictions	Working Lines	Express Services	Stretch Capacity	Vehicle Fleet Size	Vehicle Capacity	Time Horizon	% Satisfied Demand
1967	Lampkin and Saalmans					✓			
1974	Silman et al.					✓			
1979	Dubois et al.					✓			
1981	Hasselström					✓			
1884	Marwah et al.					✓	✓		
1986	Ceder and Wilson					✓			
1987	Van Oudheusden et al.					✓			✓
1988	Van Nes et al.					✓		✓	
1989	Israeli and Ceder								
1990	Baaj and Mahmassani					✓	✓		
1992	Baaj and Mahmassani					✓	✓		
1992	Israeli								
1994	Shih and Mahmassani					✓	✓		
1995	Baaj and Mahmassani					✓	✓		
1995	Israeli and Ceder								
1996	Bussieck et al.						✓		
1998	Bielli et al.								
1998	Ceder and Israeli								
1998	Pattanaik et al.						✓		
1998	Shih et al.					✓	✓		
1999	Soehodo and Koshi					✓			
2000	Rao et al.					✓	✓		
2001	Caramia et al.	?	?	?	?	?	?	?	?
2001	Chien et al.					✓			✓
2002	Bielli et al.								
2002	Fusco et al.	?	?	?	?	?	?	?	?

Table 4: *Modelling features considered in the literature works* (*Continued*).

Year	Author(s)	Infrastructure Restrictions	Working Lines	Express Services	Stretch Capacity	Vehicle Fleet Size	Vehicle Capacity	Time Horizon	% Satisfied Demand
2003	Chakroborty	?	?						
2003	Ngamchai and Lovell						✓		
2003	Tom and Mohan						✓		
2003	Wan and Lo				✓	✓	✓		
2003	Zhao and Gan					✓	✓		
2004	Agrawal and Mathew						✓		✓
2004	Carrese and Gori					✓	✓		✓
2004	Fan and Machemehl				✓	✓	✓		
2004	Petrelli						✓		✓
2005	Cipriani et al.				✓	✓	✓		
2005	Lee and Vuchic								
2005	Hu et al.				✓				
2006	Fan and Machemehl				✓	✓	✓		
2006	Zhao					✓	✓		
2006	Zhao and Zheng					✓	✓		
2007	Barra et al.				✓	✓	✓		
2007	Borndörfer et al.				✓		✓		
2007	Zhao and Zheng				✓	✓	✓		
2008	Fan and Machemehl				✓	✓	✓		
2008	Fernández et al.				✓	✓	✓		
2009	Pacheco et al.					✓		✓	
2009	Marín et al.	✓			✓	✓			
2011	Mauttone						✓		✓
2011	Szeto and Wub					✓			
2012	Cipriani et al.				✓	✓	✓		
2012	Shimamoto et al.					✓			
2014	López	✓	✓	✓	✓	✓	✓	✓	

3.2.2. Working lines

Working lines refer to those lines that are already in operation and that can be considered for an extension of the current working network at no infrastructure resource cost. Under this definition of working lines, there is no literature work that imposes such a constraint, except for the work of López (2014). In this work, the infrastructure resources have a zero-cost and, thus, they have no contribution to the objective function value and to the infrastructure budget limitation, as explained in the Subsections 3.1.1 and 3.2.1, respectively.

3.2.3. Express Service design

Express Service design refers to a specific way that vehicles work on a line, usually when lines are longer (in the sense that lines have many intermediate stations between the terminal stations) and there are high levels of congestion at stations. A vehicle performs a express service when it does not halt at some intermediate stations contained in the line cycle. As shown in the studies of Vuchic (1973) and Ercolano (1984), express services allow decreasing the waiting times experienced by the passengers (see Subsection 3.1.5 and 3.1.6 for an explanation of these concepts). For the point of view of operators, this type of service permits savings in the planning resource costs (see Subsection 3.1.2 for a detailed explanation of these costs). Although state-of-the-art setting frequency models consider this feature (Chiraphadhanakul and Barnhart, 2013; Larraín et al., 2013), models that integrate the network design and the frequency setting problems mislead this feature, except for the work of López (2014).

3.2.4. Stretch capacity

The stretch or link capacity is related to the maximum number of vehicles per unit of time that can go through a segment of a line so that overtaking cannot occur. This feature is commonly known as the minimum headway. The headway is expressed as the difference of two consecutive vehicle arrivals at a given station. Therefore, the headway is inversely proportional to the frequency. Literature works implement this feature in three different ways: 1) A lower bound on the line headway (Wan and Hong, 2003; Fan and Machemehl, 2004, 2006 and 2008; Zhao and Zeng, 2007), 2) An upper bound on the line frequency (Cipriani et al., 2005 and 2012; Borndörfer, 2007; Marín et al., 2009; López, 2014) and 3) A maximum service time at stations (Hu et al., 2005). The work of Fernández et al. (2008) does not explain how this feature is implemented.

3.2.5. Vehicle capacity

As explained in Subsection 3.1.8, the capacity of a public transportation vehicle is regarded as the maximum number of seated passengers plus the maximum number

of standees according to an allowable passenger density. This feature is commonly referred to as the line capacity which is expressed as the product of the line frequency and the capacity of the vehicles operating on the line. The line capacity is limited in two distinct ways. Headway-based approaches constraint the link load factor (Marwah et al., 1993; Baaj and Mahmassani, 1990, 1991 and 1995; Israeli, 1992; Shih and Mahmassani, 1994; Shih et al., 1998; Pattnaik et al., 1998; Petrelli, 2004; Rao et al., 2000; Fan and Machemehl, 2004, 2006 and 2008; Carrese and Gori, 2004; Cipriani et al., 2005 and 2012; Zhao and Ghan, 2003; Zhao, 2006; Zhao and Zeng, 2006 and 2007; Barra et al., 2007; Agrawai and Mathew, 2004; Ngamchai and Lovell, 1993; Tom and Mohan, 2003) which is expressed as follows:

$$L_a = \frac{q_a \cdot h^l}{q_v} \tag{1}$$

where parameter q_a is related to the maximum allowable passengers flow on the link, q_v is the vehicle capacity and h^l is the headway of the line l. On the other hand, frequency-based approaches limit the maximum flow load on the link according to the line capacity (Borndörfer, 2007; Bussieck et al., 1996; Fernández et al., 2008; López, 2014; Mauttone, 2011; Wan and Hong, 2003). The line capacity is expressed as the line frequency times the capacity of the vehicle.

3.2.6. Vehicle fleet size

The vehicle fleet size accounts for the maximum number of available vehicles. In general, the vehicles comprised in the fleet are considered to have an acquisition cost. However, López (2014) also considers a subset of vehicles with no acquisition cost due to the fact that this subset of vehicles is already in operation in some working line. This feature is verified using a constraint that limits the total number of vehicles used in the whole set of lines. This number is obtained by means of the product of the line cycle and the frequency of the line.

3.2.7. Time horizon

The time horizon or also the planning horizon refers to the maximum amount of time that all the services performed by a vehicle on a line must be accomplished. This feature is usually related to the peak hour of a working day, when the public system is supposed to be most congested. Surprisingly, there are just a few works in the literature that limit the planning horizon (López, 2014; Pacheco et al., 2009; van Nes et al., 1988).

3.2.8. Minimum amount of demand satisfaction

Some works in the literature aim at covering a minimum amount of demand to justify the investment costs of the operator (Agrawai and Mathew, 2004; Carrese and Gori, 2004; Chien et al., 2001; Mauttone, 2011; Petrelli, 2004; van Oudheusden et al., 1987). It is arguable whether this feature can be indirectly considered including the infrastructure resource costs in combination with the mode disutility in the objective function (see Subsections 3.1.1 and 3.1.9 for an explanation of these costs). Anyway, this feature is also mentioned in the present review to not exclude the cited works.

3.3. Solving techniques

This subsection reviews the solving techniques without attaching importance to algorithmic details. Rather than that, the focus is on the utility and quality of the approaches. Tables 5-6 show the five distinct features of each solving technique. They comprise the solving scheme (column 3), the nature of the approach (column 4), the algorithms involved in this approach (column 5), the way in which the line layout is determined (column 6) and the network size which is capable to solve (column 7). The following subsections go into the details of each feature.

3.3.1. Solving scheme

The solving scheme refers to the sequence in which the network design and the frequency setting phases are solved. In a sequential scheme, the network design is first solved and, then, the frequency setting is conducted having fixed the line layout. The simultaneous scheme solves both phases at the same time or modifies one of these phases having computed the other phase in an iterative fashion. The second variant of the simultaneous scheme is the most used in the literature, whereas a sequential scheme was implemented in the earlier works (Dubois et al., 1979; Lampkin and Saalmans, 1967; Silman et al., 1974). Although some other works in the beginnings of 2000 also implement the sequential scheme (Chakroborty, 2003; Hu et al., 2005; Soehodho and Koshi, 1999).

3.3.2. Approach

The term approach is strongly related to the quality of the solution obtained with the algorithms used in each work. A exact approach means that the solution obtained is an optimum of the optimization model stated in that work. A matheuristic approach refers to a quasi-exact approach in which the solution is not optimal but is certainly close to the optimum, or is only optimal in reduced instances of the model. For instance, in López (2014) instances where only one line is under construction can be solved to optimally. However, for instances with multiple lines under construction a matheuristic is used to reach a near-optimal solution. The heuristic approach stands for the works in which

Table 5: *Solving techniques used in the literature works (see also the next page).*

Year	Author(s)	Solving scheme	Approach	Algorithm(s)	Layout Method	Network size
1967	Lampkin and Saalmans	Sequential	Heuristic	RCA + FSP with RGBSP	Selection	Small
1974	Silman et al.	Sequential	Heuristic	RCA + FSP with GPM	Selection	Small
1979	Dubois et al.	Sequential	Heuristic	NRP + RCA + FSP with GBSP	Selection	Small
1981	Hasselström	Simultaneous	Heuristic	RCA + REA	Select. + Mod.	Medium
1884	Marwah et al.	Simultaneous	Heuristic	NRA + RCA + REA with CLP	Selection	Large
1986	Ceder and Wilson	Simultaneous	Heuristic	RCA + RIA	Select. + Mod.	Small
1987	Van Oudheusden et al.	Simultaneous	Heuristic	CGA + FSP + SCP/SPLP with EA	Selection	Medium
1988	Van Nes et al.	Simultaneous	Heuristic	RCA + REA	Selection	Large
1989	Israeli and Ceder	Simultaneous	Heuristic	RCA + RRA + REA with a CGT	Select. + Mod.	Small
1990	Baaj and Mahmassani	Simultaneous	Heuristic	RCFSA + REA + RIA	Select. + Mod.	Large
1992	Baaj and Mahmassani	Simultaneous	Heuristic	RCFSA + REA + RIA	Select. + Mod.	Large
1992	Israeli	Simultaneous	Heuristic	RCA + RRA + REA with a CGT	Select. + Mod.	Small
1994	Shih and Mahmassani	Simultaneous	Heuristic	RCA + REA + RSA + RIA	Select. + Mod.	Large
1995	Baaj and Mahmassani	Simultaneous	Heuristic	RCFSA + REA + RIA	Select. + Mod.	Large
1995	Israeli and Ceder	Simultaneous	Heuristic	RCA + RRA + REA with a CGT	Select. + Mod.	Small
1996	Bussieck et al.	Simultaneous	Matheuristic	Branch & Bound + VI	Selection	Large
1998	Bielli et al.	Simultaneous	Heuristic	RCA with GA + REA with NN + RIA with GA	Select. + Deter.	Large
1998	Ceder and Israeli	Simultaneous	Heuristic	RCA + RRA + REA with a CGT	Select. + Mod.	Small
1998	Pattanaik et al.	Simultaneous	Heuristic	RCA + REA with GA	Selection	Medium
1998	Shih et al.	Simultaneous	Heuristic	RCFSA + REA + RIA	Select. + Mod.	Large
1999	Soehodo and Koshi	Sequential	Heuristic	RCA + REA + RIA	Select. + Mod.	Medium
2000	Rao et al.	Simultaneous	Heuristic	RCA + REA with GA	Select. + Mod.	Small
2001	Caramia et al.	Simultaneous	Heuristic	REA with NN + RIA with GA	Selection	Medium
2001	Chien et al.	Simultaneous	Heuristic	RCA with GA + RNHSP with GA	Selection	Medium
2002	Bielli et al.	Simultaneous	Heuristic	RCA with GA + REA with NN + RIA with GA	Select. + Mod.	Large
2002	Fusco et al.	Simultaneous	Heuristic	RCA + REA with GA + RIA	Select. + Mod.	Small

Table 6: *Solving techniques used in the literature works* (*Continued*).

Year	Author(s)	Solving scheme	Approach	Algorithm(s)	Layout Method	Network size
2003	Chakroborty	Sequential	Heuristic	RCA with GA + FSP with GA	Selection	Small
2003	Ngamchai and Lovell	Simultaneous	Heuristic	RCA + REA + RIA	Select. + Deter.	Small
2003	Tom and Mohan	Simultaneous	Heuristic	RCA + REA with GA	Selection	Medium
2003	Wan and Lo	Simultaneous	Exact	CPLEX Branch & Bound	Determination	Small
2003	Zhao and Gan	Simultaneous	Heuristic	RCA with SA + HSP with FDS	Selection	Large
2004	Agrawal and Mathew	Simultaneous	Heuristic	RCA + REA with GA	Selection	Large
2004	Carrese and Gori	Simultaneous	Heuristic	RCA + MREA + FREA	Select. + Mod.	Large
2004	Fan and Machemehl	Simultaneous	Heuristic	RCA with Yen-KSP + REA + RIA with SMH	Selection	Large
2004	Petrelli	Simultaneous	Heuristic	RCA + REA with GA + RIA	Selection	Large
2005	Cipriani et al.	Simultaneous	Heuristic	RCA + REA with GA	Selection	Large
2005	Lee and Vuchic	Simultaneous	Heuristic	RCA + RIA + REA	Select. + Mod.	Small
2005	Hu et al.	Sequential	Heuristic	RCA with ACA + FSP with GA	Selection	Large
2006	Fan and Machemehl	Simultaneous	Heuristic	RCA with Yen-KSP + REA + RIA with GA	Selection	Large
2006	Zhao	Simultaneous	Heuristic	RCA with SA + HSP with FDS	Select. + Deter.	Large
2006	Zhao and Zheng	Simultaneous	Heuristic	RCFSP with LSA + RIA with GA	Select. + Deter.	Large
2007	Barra et al.	Simultaneous	Exact	RCSFP with CP	Selection	Small
2007	Borndörfer et al.	Simultaneous	Matheuristic	CGA + GH	Selection	Large
2007	Zhao and Zheng	Simultaneous	Heuristic	RCFSP with LSA + RIA with TS, GS, BS	Select. + Deter.	Large
2008	Fan and Machemehl	Simultaneous	Heuristic	RCA with Yen-KSP + REA + RIA with TS	Selection	Large
2008	Fernández et al.	Simultaneous	Heuristic	RCA + FSP with HJA	Selection	Large
2009	Pacheco et al.	Simultaneous	Heuristic	RCA + RIA with LS/TS	Select. + Mod.	Medium
2009	Marín et al.	Simultaneous	Heuristic	RCA + FSP using CPLEX Branch & Bound	Determination	Small
2011	Mauttone	Simultaneous	Heuristic	RCA with PIA + FSP with GRASP	Selection	Medium
2011	Szeto and Wub	Simultaneous	Heuristic	RCA with GA + FSP with NSH	Selection	Small
2012	Cipriani et al.	Simultaneous	Heuristic	RCA + REA with GA	Selection	Large
2012	Shimamoto et al.	Simultaneous	Heuristic	(VAP + RCA + FSP) with NSGA-II	Selection	Small
2014	López	Simultaneous	Matheuristic	RCA with Yen-KCSP + LSA + SBD	Select. + Deter.	Large

no optimum is guarantee no matter what type of instance of the model is solved. The vast majority of works fall in this category, although some works use mathematical approaches to solve the solving modules (see for instance Hasselström, 1981; van Oudheusden et al., 1987). However, the authors do not demonstrate that the optimum is within the established partition of the solution space. To this end, it seems more appropriate to use exact decomposition techniques as in Marín and Jaramillo (2009) or López (2014).

3.3.3. Algorithm(s)

Most algorithms fall within the category of metaheuristics. To mention some of them, the Greedy Randomized Adaptive Search Procedure (*GRASP*), Simulated Annealing (*SA*) and Genetic Algorithm (*GA*). The last-mentioned metaheuristic is the most used, leaving apart its variant implementations. Within these matheuristics a constructive heuristic is used to build part of the solution. For instance, in Mauttone (2011) this heuristic is called the Pair Insertion Algorithm (*PIA*) where routes are constructed having fixed their frequencies. The *PIA* is driven by means of the outer *GRASP* metaheuristic.

This subsection does not pretend to provide a detailed explanation of the algorithms. Rather than that, the focus is on the type of partition of the solution space employed. Using this criterion, the algorithms within the category of heuristic approaches are divided into: 1) Two-sequential phases, 2) Two-iterative phases, 3) Three-sequential phases, 4) Three-iterative phases and 5) Four-iterative phases. Being the second subcategory the most used.

In two-sequential approaches, a Route Construction Algorithm (*RCA*) builds the skeleton of the routes and then a Frequency Setting Procedure (FSP) assigns frequencies and vehicles to these routes (Chakroborty, 2003; Hu et al., 2005; Lampkin and Saalmans, 1967; Silman et al., 1974). The two-iteration approach has two variants. One variant works in a similar way to the two-sequential approach and the main difference is that both modules interact until some criterion is met (Agrawai and Mathew, 2004; Caramia et al., 2001; Ceder and Wilson, 1986; Hasselström, 1981; Mauttone, 2011; Pattnaik et al., 1998; Pacheco et al., 2009; Rao et al., 2000; Szeto and Wub, 2011; Tom and Mohan, 2003; van Nes et al., 1988; Zhao and Ghan, 2003). In some of these works, the FSP may evaluate some other indicators apart from frequencies and, therefore, the module is called as the Route Evaluation Algorithm (*REA*). The other variant is more sophisticated. Initially, a global feasible solution is determined using a Route Construction and Frequency Setting Procedure ($RCFSP$) and, then, new routes are constructed analyzing this global solution using a Route Improvement Algorithm (*RIA*). These routes are evaluated in the following iteration using again the $RCFSP$ procedure. Both modules interact until some criterion is met (Chien et al., 2001; Zhao and Zeng, 2006 and 2007).

In three-sequential approaches, A Network Reduction Procedure (*NRP*) is first used to reduce the number of links to be considered in the *RCA* algorithm. The remaining steps are similar to the two-sequential approaches (Dubois et al., 1979; Soehodho and Koshi, 1999). The three-iterative phase approaches have three variants which differ from the type of combination used among the preceding mentioned approaches. One variant combines the first variant of the two-iterative phases approach and the three-sequential phases approach. So, a *NRP* is first used and then the *RCA* and *REA* interact until some criterion is met (Marwah et al., 1993). Another variant extends the first variant of the two-iterative-phases. This extension entails to add the *RIA* algorithm after the application of the *REA* algorithm (van Oudheusden et al., 1987; Israeli and Ceder, 1989; Israeli, 1992; Ceder and Israeli, 1998; Bielli et al., 1998 and 2002; Fusco et al., 2002; Ngamchai and Lovell, 1993; Carrese and Gori, 2004; Fan and Machemehl, 2004, 2006 and 2008; Petrelli, 2004; Lee and Vuchic, 2005; Shimamoto et al., 1993; Israeli and Ceder, 1995). In some of these works, the names of the solving blocks are altered because they are more sophisticated. Moreover, the order in which they are used may be also modified. The remaining variant is the most sophisticated approach. It combines the two variants of the two-iterative approaches in such a way that the *RCFSA* is first called, then the *REA* module and finally the *RIA*. The three solving blocks interact until a criterion is met (Baaj and Mahmassani, 1990, 1991 and 1995; Shih et al., 1998).

Moving on to matheuristic approaches, the literature is very scant. The works of Bussieck et al. (1996) and López (2014) assume a predefined set of routes/corridors that are evaluated within two distinct mathematical programming environments in such a way that the output is the near-optimal set of routes and frequencies. The remaining work of Borndörfer (2007) employs a different scheme. This scheme also assumes a predefined set of routes but then likely fractional routes and frequencies are determined using mathematical programming techniques. The likely fractional routes are then rounded using a greedy heuristic.

Finally, it is mentioned the work of Wan and Hong (2003). To the best of the author knowledge, it is the only work that implements a strictly exact approach. This approach consists in formulating the model as a mixed-integer linear programming problem that is directly solved by CPLEX.

The remaining not mentioned acronyms in column 5 of Table 3 are explained in Table 7 of Appendix 1.

3.3.4. Layout method

The term layout refers to the structure of the line, i.e., which are the stretches (links) and stations of the line, and the way the layout is determined is referred to layout method. The literature on this issue can be classified into the following four categories: selection, selection + modification, determination and selection + determination.

Selection is the leading layout method because it requires less computational effort. This method assumes a fixed layout on a set of candidate lines, determined by a route

construction procedure without considering passengers and frequencies issues, and the aim is to chose a subset of these lines meeting frequency and passenger requirements.

Selection + modification is an extension of the selection method. It consists of the following two stages: a selection stage which selects a preliminary subset of good lines and a modification stage in which the selected lines are improved by inserting/deleting links or merging lines which are similar (see the references on Table 3 where the sixth column contains the label "Select. + Mod.").

Determination is the hardest computational method because it makes no simplistic assumption of the layout. It considers a set of potential links and stations and the aim is to allocate them to the lines. This is carry out by imposing appropriate constraints in the mathematical programming formulation of the model (see Wan and Hong, and Marín et al., 2009).

Selection + determination is a very rare method which has been only found in López (2014). Its aim is to obtain significant better layout solutions than the ones found by the selection (+modification) method but not spending too much time. It consists of the following two stages: a route construction procedure which determines a set of candidate line corridors (a chain of line segments or stretches) and a determination procedure which allocates stations to them. The latter is carried out by means of a mathematical programming approach.

3.3.5. Network size

The network size refers to the biggest instance that can be solved by the solving techniques. This size has been computed counting the number of nodes, links, o-d demand pairs and number of lines under construction of the biggest study case. As shown in column 6 of Table 3, early works are only capable of solving small-sized networks and, as time has gone by, the solvable size of the network has been increased. In the early 80s, Hasselström (1981) succeeded in solving medium-sized networks. One decade after, van Nes et al. (1988) managed to solve large-sized networks. However, the author used a model rather simplistic. Shih et al. (1998) were the firsts in solving a more detailed model in real-sized networks. Despite this success, the model was still far from reality. Moreover, the model was solved heuristically as in the previous mentioned works.

The initial works on matheuristics (Borndörfer, 2007; Bussieck et al., 1996) demonstrated the effectiveness of mathematical techniques in combination with heuristics, although the employed models were rather simplistic, again. Until not very recently, matheuristics have not been demonstrated to solve large instances with more realistic models (López, 2014).

4. Conclusions and further research

This survey has shown the literature works on the integration of the network design and the frequency setting phases in public transportation networks. These phases correspond to the two first stages of the Transit Planning Process (Ceder and Wilson, 1986). The survey has put emphasis on both modelling issues, i.e., objective cost components and constraints; as well as on solving approaches.

Through Tables 1-6 and their corresponding explanations in Subsections 3.1-3.2, we have seen a great variety of works covering different aspects from the point of view of modelling features and solving techniques. However, there are four major concerns. First, non of these works integrates all the mentioned modelling features. Second, many of these modelling features are not fully or properly covered. Third, other key modelling features have been omitted. Finally, the solving techniques do not guarantee an accurate solution, or are limited to the modelling features being considered. In the following subsections, some lines for further research concerning issues 3 and 4 are provided.

4.1. Modelling issues

Modelling issues concern the platform capacity, the operational capacity of stations, the dwell time, the design with multiple demand scenarios and robustness and recovery in Rapid Transit Network Design. All of them affect to both operator and passengers agents and, thus, they are not considered in separate subsections. They are explained in the following subsections.

4.1.1. Platform capacity

The capacity of a platform is related to the maximum number of passengers that a platform can hold while passengers are waiting to board the vehicle. This capacity comes into play in congested scenarios, i.e., in the peak hours. Its implementation is rather cumbersome because there are several variables interconnected, i.e., the operating frequency on the line, the average passenger waiting time at the station and the arrival pattern of the passengers. Additionally, the relationships of these elements are non-linear and non-convex. Codina et al. (2013) have modeled the station capacity in the context of a bus bridging network where a number of lines are given as inputs. The authors imposed the following constraint:

$$\sum_{l \in L_b} \zeta_a^{l,b}(v_a^{b,l}, v_{x(a)}^{b,l}, z^l) \leq \frac{H}{\eta} \bar{N}_b^{pax} \tag{2}$$

where L_b is a set containing the indexes of the lines serving at the station b, $\zeta_a^{l,b}$ is the total passenger waiting time at the station b before boarding a vehicle working on line l. This time depends on the total number of passengers boarding a vehicle serving at

the station b throughout the time period H under consideration ($v_a^{b,l}$), the total number of vehicle waiting in the vehicle serving at the station b ($v_{x(a)}^{b,l}$) and the total number of services on the line l (z^l). The remaining parameters η and $\bar{N}_b^{pax}$ represent the ratio between the passenger queue length exceeded a fraction 1 - α of the line and the average queue length and the capacity of the station b, respectively. The authors also define a general formula for computing the total waiting time $\zeta_a^{l,b}$ as follows:

$$\zeta_a^{l,b}(v_a^{b,l}, v_{x(a)}^{b,l}, z^l) = v_a^{b,l} P_a^b(z^l) \xi_a \left(\frac{v_a^{b,l}}{c\, z^l - v_{x(a)}^{b,l}} \right) \tag{3}$$

where function P_a^b is the average waiting time per passenger and service without congestion effects, ξ_a is a function that considers the congestion and c is a parameter denoting the vehicle capacity. Function P_a^b was approached using the Allen-Cuneen's formula (Allen, 1998), whereas function ξ_a was determined empirically using bulk service queue simulation models. These models establish the relationship between bus stop load factor and passenger waiting times. Finally, constraint (2) was included in a mixed-integer non-linear programming problem. This problem was solved using a specific heuristic consisting of a fixed-point iteration algorithm based on the method of successive averages (*MSA*). The main drawback of this methodology is that it does not consider link capacities (see Subsection 3.2.4 for its explanation). Thus, it is difficult to integrate the network design phase.

4.1.2. Operational capacity of stations

The operational capacity of a station refers to the maximum number of services per unit of time that can be operated at a station. This capacity influences over the operational frequencies of the lines and the dwell times of the vehicles serving at the station. Additional variables may be involved in certain types of bus stations (Codina et al., 2013). Figure 1 shows a type of bus station, known as bay station, in which two queues

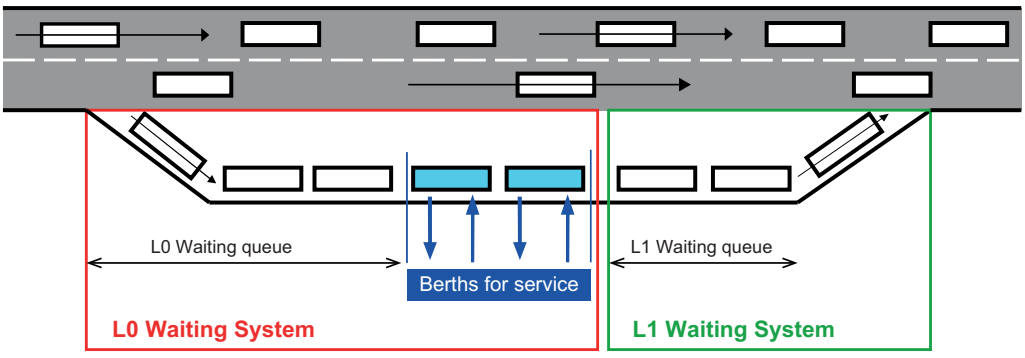

Figure 1: *A schematic representation for a bus stop according to Codina et al. (2013).*

emerge. One queue represents the waiting time of buses willing to enter the berth place (labeled as $\mathscr{L}^0$ queue), whereas the other queue models the waiting time of buses willing to exit the berth (labeled as $\mathscr{L}^1$ queue).

Under this configuration, the following formula applies to the operational capacity of the station:

$$\sum_{l \in L_b} z^l \leq \hat{Z}_b(v,z) \tag{4}$$

where set L_b contains the identifiers of the lines operating at station b, variable z^l indicates the number of services performed on line l and function $\hat{Z}_b(v,z)$ accounts for the following expression:

$$\hat{Z}_b(v,z) \triangleq H \min\left(\frac{(1-\epsilon)s_b}{\kappa_b(v,z)}, \right. \tag{5}$$
$$\left., \frac{\mathscr{L}^0}{\eta^0(\kappa_b(v,z)+\omega_b^0(v,z))}, \frac{\mathscr{L}^1}{\eta^1\omega_b^1(v,z)}\right)$$

where parameter H denotes the planning time horizon, s_b indicates the number of available berth places, $\kappa_b(v,z)$ accounts for the total dwell time, $\mathscr{L}^{0/1}$ stands for the respective queue lengths (i.e., the maximum number of vehicles that can hold the corresponding queues), $\eta^{0/1}$ denotes the occupancy factors at 95 % of the operating time and, finally, $\omega_b^{0/1}(v,z)$ indicates the total waiting time of busses at the respective queues.

Formula (5) can be easily accommodated for railway-based systems. It suffices to omit the quotient expressions inside the big parenthesis which are associated with the bus queues, and to set parameter s_b to 1. In both applications, railway and bus systems, the resulting expressions are not linear because the dwell time $\kappa_b(v,z)$ is in the denominator of the quotient. This limitation was overcome using a specific heuristic as explained in the preceding subsection. Moreover, it was pointed out that this methodology cannot be directly used because the link capacity is omitted.

4.1.3. Dwell time

The dwell time is related to the time in a station spent by the vehicle allowing passengers to board or alight from the vehicle. This time plays also an important role in congested scenarios and in lines with many service stations. The dwell time involves the times on braking and opening the doors when the vehicle enters the station, the passenger times on boarding and alighting when the vehicle is stopped, and the times on closing the doors and accelerating when the vehicle leaves the station.

The computation of passenger times in the dwell time depends on the type of transportation system under consideration. It is also assumed that passengers behave in a

rational manner, i.e., each passenger waits until its predecessor has boarded or alighted. In railway-based systems, first, in-vehicle passengers alight and, then, waiting at-station passengers board. So, only the highest time is added to the dwell time. Whatever the application is, the total type of movement time is divided by the number of vehicle doors dedicated to the movement.

The implementation of the dwell time is rather cumbersome due to non-linearities emerging from the passenger time costs. These costs are related to the alighting, boarding and waiting in-vehicle times. The alighting and boarding times cannot be considered proportional to the number of passengers performing such movements because each passenger does not experience the same time. For instance, the first passenger in alighting from the vehicle starts this movement immediately after the opening of the doors. However, the second passenger needs to wait until the first passenger has alighted and so on. The same reasoning can be done for the boarding movement. To overcome this issue, a portion of the dwell time must be used to weigh the passenger flow associated with these movements. As for the waiting in-vehicle time, the passenger flow related to this waiting must be weighted by the dwell time. So, all these three products are non-linear. The work of Codina et al. (2013), mentioned in the previous subsections, models this feature. Non-linearities are overcome by freezing some of their values in a first optimisation problem and, then, by updating these values, properly. This mechanism is repeated within a fixed-point procedure based on the method of successive averages (*MSA*). This methodology was devised for frequency setting in bus systems but it can be adapted for railway systems with a few changes. As pointed out in the two preceding subsections, the main drawback is that it does not consider link capacities. Thus, it is difficult to integrate the network design phase in the authors approach.

4.1.4. Design with multiple demand scenarios

The literature works have only considered an o-d demand matrix for a given period of the day. This period is usually associated with the peak hours. However, this approach may lead some o-d demand pairs emerging in different periods of the day without coverage. Therefore, different o-d demand matrices should be considered in such a way that the resulting network is consistent with all solution scenarios. Marín and Jaramillo (2008) implement this feature but only for the network design phase. For each period under consideration, a mixed-integer linear programming problem is solved. This problem considers lines being constructed in previous periods. In the first period, some lines already in operation may be taken into account. The infrastructure resource costs used in previous periods are subtracted from the current period under consideration, thus, encouraging passengers to use the already allocated infrastructure. The main drawback of this approach is that the number of lines under construction are not limited. So, if the o-d demand matrices are very different in each period, i.e., different o-d demand pairs appear in each period, an exorbitant number of new lines are likely to be constructed.

4.1.5. Failure to board

The term failure to board applies to congested scenarios where passengers cannot board the first vehicle arriving at their waiting station due to a lack of residual capacity. This feature has been addressed in Kurauchi et al. (2003) and Codina et al. (2013). The first work presents an innovative passenger assignment model that assumes fixed line layouts and operating frequencies, and uses two additional nodes and links accounting for the lack of residual capacity in the vehicle. The additional nodes evaluate the failure to board using a given probability function. Having evaluated this function, the passengers who were unable to board are redirected to the destination, whereas the succeed passengers are transferred to the boarding links. The probability function is expressed using an absorbing Markov chain which is embedded into a hyperpath structure. The solving approach seeks for the minimum hyperpath within a method of successive averages (*MSA*). This approach was devised for a particular transportation problem and, thus, it cannot be use as a design tool. The other work seems more interesting from the application point of view because it considers the layout of the lines and carries out the frequency setting. The failure to board is indirectly modeled using a function (denoted as ξ_a, where link a accounts for a boarding link) that determines the increment in waiting time due to congestion. The reader is directed to Subsection 4.1.1 for further details.

4.1.6. Robustness in rapid transit network design

Robustness is one of the most complex features from both modelling and computational points of view. It consists of designing a network as much robust as possible so that it is not "very" affected by a vehicle breakdown or an unexpected increase in demand in sections (links) of the network. The robustness issue has been previously considered in Laporte et al. (2011), Marín et al. (2009) and Cadarso and Marín (2012), among others. All these works are based on mathematical programming approaches and the differences rely on the type of robustness measure considered and the way it is incorporated in the model.

In Laporte et al. (2011) and Cadarso and Marín (2012), robustness is only considered from the point of view of the user, i.e., when a failure/congestion occurs in some critical edge, passengers using that edge must have some alternative routes. Laporte et al. (2011) examine separately different scenarios by changing the value of the parameters and using different types of constraints related to robustness. Moreover, they pointed out how the design of the network is affected by each parameter and type of constraint. In contrast, Cadarso and Marín (2012) analysed the different scenarios at the same time and minimized the differences between the optimal network designs in each scenario considering only one type of constraints defined by Laporte et al. (2011).

Marín et al. (2009) extend the concept of robustness to account for the point of view of the operator. A network is robust if a failure or congestion occurring in some critical edge affects the less number of vehicles. To this end, the authors defined an iterative approach in which two mathematical programming problems are solved. The first one

is a network design problem with the same mathematical structure as the one presented in Laporte et al. (2011), the only difference is that the authors make only use of one type of robustness constraint (as in Cadarso and Marín, 2012). The second problem is a frequency setting model with flows expressed by paths (routes) instead of links. In each iteration, alternative routes and planning configurations are sought by fixing to 1 the active routing variables of the previous iterations. The algorithm stops when there is no infrastructure budget. This scheme is repeated twice, one accounting Robustness for the point of view of the user and the other focusing robustness on the point of view of the operator. In this way, a more variety of network designs are found and can be analysed.

All these works can be integrated with the aforementioned features with minor changes. However, algorithmic improvements must be done in order to solve real-sized networks.

4.1.7. Recovery in rapid transit network design

This feature is complementary to Robustness and its aim is to provide an alternative service to those passengers affected by a disruption on their usual transportation system. The literature on this topic have mainly been addressed to some of the final stages of the Transit Planning Process (see, for instance, Cadarso, 2013 or Cadarso and Marín, 2014), and scant literature focus on some of the first stages (see, for instance, Codina et al., 2013).

Cadarso (2013) and Cadarso and Marín (2014) developed an integrated timetable and rolling stock model where the term "rolling stock" refers to vehicle scheduling in Railway systems. In that model, passengers on cancelled services, due to disruptions on some links of their operating lines, are reassigned to new emergency services (ES). These ES services may take place on the line where the disruption occurred, but the ES must begin(end) after(before) the disrupted link. Additionally, an alternative system (the underground) may also carry out some ES service as long as part of their line itineraries are close to the disrupted links of the train system.

Codina et al. (2013) devised a Bus Bridging model which provides service to all passengers affected by disruptions on a railway-based system. The model mainly focuses on the frequency setting phase but it can also determine which lines will provide service. Moreover, the model takes into consideration all the main effects of congestion at a bus station, i.e., waiting time to board a vehicle considering lack of residual capacity, waiting time on entering the berth and waiting time on exiting the station (see Subsections 4.1.1, 4.1.2 and 4.1.5 for further details).

Cadarso (2013) and Cadarso and Marín (2014) works cannot being directly integrated into a network design and frequency setting model but some ideas can be taken from. The work of Codina et al. (2013) seems more suitable but the complexity of the resulting model will considerably be increased in views of their solving approach.

4.2. Solving issues

Solving issues are focused on the generation of more attractive line corridors, convergence enhancements of exact decomposition approaches and the non-convexity in variable demand models. The first issue affects both heuristic and exact approaches, whereas the other two issues are strictly related to exact or certain matheuristic approaches. They are explained in the following subsections.

4.2.1. Generation of more attractive line corridors

The generation of a set of input line corridors to the optimization model represents a key aspect for a good network design. To date, state-of-the art works use a k-shortest path algorithm which determines a preliminary set of line corridors. This set is then reduced, evaluating a certain number of restrictions related to the length of the corridors and, possibly some user behaviour rules (see, for instance, Fan and Machemehl, 2004, 2006 and 2008). This approach may discard a large set of good corridors in the running of the k-shortest path because the restrictions are evaluated afterwards. This drawback is overcome in López (2014). However, there is still an important limitation. The amount of demand is still not considered in the running of the k-shortest path. This limitation affects the in-vehicle waiting time. A k-shortest path algorithm enumerates the line corridors in an increasing fashion. First, it constructs the k-shortest corridor and then seeks for the first least long corridor. At this point, the amount of demand comes into play because part of the enlargement of the previous corridor is only justify iff at least one additional o-d demand pair flow uses part of this new corridor. Therefore, the previous allocated o-d demand flows will experience a delay due to boarding and alighting of this (these) additional o-d demand pair(s) within the section of the corridor in which these movements occur.

4.2.2. Convergence enhancements of exact decomposition approaches

This issue is related to the Benders Decomposition (*BD*). The *BD* (Benders, 1962) is a classical decomposition algorithm applied to many large optimization models successfully. This decomposition consists of a reformulation of the model in which two problems called the Master Problem (*MP*) and the SubProblem (*SP*) are iteratively solved until a duality gap is small enough. The *MP* is a relaxation of the original problem in which the interdependencies between the operator and the passengers are discarded and the active dependencies are iteratively appended in the form of Benders cuts. These cuts may be Optimality Benders Cuts (*OBCs*) when the dual of the *SP* has a solution. Otherwise, Feasibility Benders Cuts are added to the *MP*. The *SP* represents the passenger assignment model, thus it is a continuous problem. When the dual form of the *SP* is degenerated, i.e., it has multiple optimal solutions (which is our case), the performance of the algorithm decreases dramatically Magnanti and Wong (1981). To overcome this limitation, the authors in Magnanti and Wong (1981) propose a

new Benders scheme in which an additional problem per iteration is solved to obtain better *OBCs* that enhance the algorithm convergence. This scheme is later improved in Papadakos (2008), so that the generator of *OBCs* is faster to solve. However, the computation of both generators requires an initial *core point*. This point refers to a point strictly in the interior of the feasible region of the *MP*, excluding the *OBCs*. The obtaining of this core point is non-trivial, in general, and the quality of the *OBCs* depends heavily on that point. Recently, it have been demonstrated that goods *OBCs* can be obtained from a problem that integrates the *SP* and the generator of *OBCs* Sherali and Lunday (2011). Moreover, the resulting problem does not require a core point but a strictly positive point and a weight factor, which can be easily obtained. However, the quality of the resulting *OBCs* still depends heavily on these two parameters. Thus, this enhancement is not reliable.

4.2.3. Non-convexity in variable demand models

The competition among several modes of transportation is correctly formulated as a Bilevel Programming Problem (*BPP*) (López, 2014). The outer level represents a trade-off between operator and passengers agents, whereas the inner level involves only the passenger agent. This *BPP* is solved using an adaptation of the Benders Decomposition (Codina and López, under review). In this adaptation, the Master Problem (*MP*) approaches the original problem iteratively using new types of Benders cuts coming from a more complex SubProblem (*SP*). This *SP* is a reduced *BPP* involving only the continuous variables and their restrictions of the original *BPP*. The two levels of this reduced *BPP* share the same constraints, so the inner level can be first solved and, then, using its solution, a linking constraint is added to the outer level so that it can be solved afterwards. This *BD* encounters serious problems due to inherent non-convexities being raised in the original *BPP*. These non-convexities are detected when non-valid Benders cuts appear (Saharidis and Ierapetritou, 2009). A non-valid Benders cut refers to a cut being generated in the *SP* of current Benders iteration that does not constraint the *MP* (i.e., when the cut is added to the *MP* and the *MP* is resolved, the solution does not change). The authors in Saharidis and Ierapetritou (2009) suggested to add exclusion cuts to the *MP* when a non-valid Benders cut arises. However, the generation of these cuts is rather cumbersome for *NDFSP* problems. Moreover, it requires an exorbitant number of restrictions that slow down the resolution of subsequent *MP* problems.

5. Acknowledgements

The author expresses gratitude to the support of projects TRA2008-06782-C02-02 and TRA2011-27791-C03-01/02 from the Spanish Government.

Appendix 1

Table 7: *Description of the literature methods used to solve the Network Design and Frequency Setting Problem.*

Method	Description
ACA	Ant Colony Algorithm
BS	Bisection Search
CGA	Corridor Generation Algorithm
CGT	Column Generation Technique
CLP	Continuous Linear Problem
EA	Erlenkotten Algorithm
FDS	Fast Descend Search
FREA	Feeder Route Evaluation Algorithm
FSP	Frequency Setting Procedure
GA	Genetic Algorithm
GBSP	Gradient Based Search Procedure
GPM	Gradient Projection Program
GRASP	Greedy Randomized Adaptive Search Procedure
GS	Greedy Search
HJA	Hooke & Jeeves Algorithm
HSP	Headway Setting Procedure
KCSP	K-Constrained Shortest Paths
KSP	K-Shortest Paths
LS	Local Search
LSA	Line Splitting Algorithm
MREA	Main Route Evaluation Algorithm
NSGA-II	Non-dominated Sorting Genetic Algorithm of Type-II
NN	Neuronal Networks
NRP	Network Reduce Procedure
NSH	Neighborhood Search Heuristic
PIA	Pair Insertion Algorithm
RCA	Route Construction Algorithm
RCFSA	Route Construction and Frequency Setting Algorithm
REA	Route Evaluation Algorithm
RSA	Route Selection Algorithm
RGBSP	Random Gradient Based Search Procedure
RIA	Route Improvement Algorithm
RNHSP	Route Nested and Headway Setting Procedure
RRA	Route Reduction Algorithm
SA	Simulated Annealing
SBD	Specialized Benders Decomposition
SCP	Set Covering Problem
SMH	Several Metaheuristics
SPLP	Simple Plant Location Problem
TS	Tabu Search
VAP	Vehicle Assignment Procedure
VI	Valid Inequalities

References

Agrawai, J. and Mathew, T. V. (2004). Transit route design using parallel genetic algorithm. *Journal of Computing in Civil Engineering*, 18, 248–256.

Allen, A. O. (1998). *Probability, Statistics and Queueing Theory*. Academic Press, New York.

Barra, A., Carvalho, L., Teypaz, N., Cung, V.D. and Balassiano, R. (2007). Solving the transit network design problem with constraint programming. *Proceedings of the 11th World Conference in Transport Research*.

Baaj, M. H. and Mahmassani, H. S. (1990). TRUST: a Lisp program for the analysis of transit route configuration. *Transportation Research Record*, 1283, 125–135.

Baaj, M. H. and Mahmassani, H. S. (1991). An AI-based approach for transit route system planning and design. *Journal of Advanced Transportation*, 25, 187–210.

Baaj, M. H. and Mahmassani, H. S. (1995). Hybrid route generation heuristic algorithm for the design of transit networks. *Transportation Research Part C*, 3, 31–50.

Benders, J. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4, 238–252.

Bielli, M., Carotenuto, P. and Confessore, G. (1998). A new approach for transport network design and optimization. *Proc., 38th the Conf. Congress of the European Regional Science Association*, Vienna, Austria.

Bielli, M., Caramia, M. and Carotenuto, P. (2002). Genetic algorithms in bus network optimization. *Transportation Research Part C: Emerging Technologies*, 10, 19–34.

Borndörfer, R., Grotschel, M. and Pfetsch, M. E. (2007). A column-generation approach to line planning in public transport. *Transportation Science*, 41, 123–132.

Bussieck, M. R., Kreuzer, P. and Zimmermann, U. T. (1996). Optimal lines for railway systems. *European Journal of Operational Research*, 96, 54–63.

Cadarso, L. and Marín, A. (2012). Recoverable robustness in rapid transit network design. *Procedia-Social and Behavioral Sciences*, 54, 1288–1297.

Cadarso, L., Marín, A. and Maroti, G. (2013). Recovery of disruptions in rapid transit networks. *Transportation Research Part E: Logistics and Transportation Review*, 53, 15–33.

Cadarso, L. and Marín, A. (2014). Recovery of disruptions in rapid transit networks with origin-destination demand. *Procedia-Social and Behavioral Sciences*, 111, 528–537.

Caramia, M., Carotenuto, P. and Confessore, G. (2001). Metaheuristics techniques in bus network optimization. *NECTAR Conference no 6 European Strategies in the Globalising markets, Transport Innovations, Competitiveness and Sustainability in the Information Age*, 16–18, Helsinki, Finland.

Carrese, S. and Gori, S. (2004). An urban bus network design procedure. *Transportation Planning*. In: M. Patriksson and M. Labbé (eds.), 64, 177–195.

Ceder, A. and Israeli, Y. (1998). User and operator perspectives in transit network design. *Transportation Research Record*, 1623, 3–7.

Ceder, R. B. and Wilson, N. H. (1986). Bus network design. *Transportation Research Part B*, 20, 331–344.

Cipriani, E., Fusco, G., Gori, S. and Petrelli, M. (2005). A procedure for the solution of the urban bus network design problem with elastic demand. *Advanced OR and AI Methods in Transportation Proc.*, 10th Meeting of the EURO Working Group on Transportation, Publishing House of Poznan Univ. of Technology, Poland, 681–685.

Cipriani, E, Gori, S. and Petrelli, M. (2012). Transit network design: a procedure and an application to a large urban area. *Transportation Research Part C*, 20, 3–14.

Codina, E., Marín, A., González, L. and Ramírez, P. (2008). Bus bridging for rapid transit network incidences. *Congreso Latino-Iberoamericano de Investigación Operativa (CLAIO)*, Cartagena, Colombia.

Codina, E., Marín, A. and López, F. (2013). A model for setting services on auxiliary bus lines under congestion. *TOP*, 21, 48–83.

Codina, E. and López, F. (Under review). Solving the strategy based congested transit assignment model using smoothing approximations. *Journal of Global Optimization*.

Chakroborty, P. (2003). Genetic algorithms for optimal urban transit network design. *Computer-Aided Civil and Infrastructure Engineering*, 18, 184–200.

Chien, S., Yang, Z. and Hou, E. (2001). Genetic algorithm approach for transit route planning and design. *Journal of Transportation Engineering*, 127, 200–207.

Chiraphadhanakul, V. and Barnhart, C. (2013). Incremental bus service design: combining limited-stop and local bus services. *Public Transport*, 5, 53–78.

Desaulniers, G. and Hickman, M. D. (2007). Public Transit, chapter 2. *Handbooks in Operations Research and Management Science*, 14.

Dubois, D., Bell, G. and Llibre, M. (1979). A set of methods in transportation network synthesis and analysis. *Journal of Operations Research Society*, 30, 797–808.

Ehrgott, M. and Gandibleux, X. (2002). Multiobjective combinatorial optimization-theory methodology, and applications. *Multi-Criteria Optimization-State of the Art Annotated Bibliographic Surveys*, Kluwer Academic Publishers, Dordrecht, 369–444.

Ercolano, J. (1984). Limited-stop bus operations: an evaluation. *Transportation Research Record*, 994, 24–29.

Fan, W. and Machemehl, R. B. (2004). Optimal transit route network design problem: algorithms, implementations and numerical results. Research SWUTC/04/167244-1, University of Texas.

Fan, W. and Machemehl, R. B. (2006). Optimal transit route network design problem with variable transit demand: genetic algorithm approach. *Journal of Transportation Engineering*, 132, 40–51.

Fan, W. and Machemehl, R. B. (2008). Tabu Search Strategies for the Public Transportation Network Optimizations with Variable Transit Demand. *Computer-Aided Civil and Infrastructure Engineering*, 23, 502–520.

Farahani, R. Z., Miandoabchi, E. and Szeto, W. Y. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, 229, 281–302.

Fernández, J., de Cea, J. and Malbran, H. R. (2008). Demand responsive urban public transport system design: Methodology and application. *Transportation Reserach Part A*, 42, 951–972.

Ferropedia (2014a). Infrastructure construction costs. http://www.ferropedia.es/wiki/Costos_de_construcción_de_infraestructura. Last accessed 1 August 2014.

Ferropedia (2014b). Railway costs: Services. http://www.ferropedia.es/wiki/Costes_del_ferrocarril:_servicios. Last accessed 17 April 2014.

Fusco, G., Gori, S. and Petrelli, M. (2002). A heuristic transit network design algorithm for medium size towns. *Proceedings of 9th Euro Working Group on Transportation*, 652–656.

García, R., Garzón-Astolfi, A., Marín, A., Mesa, J. A. and Ortega, F. A. (2006). *Analysis of the Parameters of Transfers in the Rapid Transit Network Design*. Schloss Dagstuhl, ISBN 978-3-939897-00-2.

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

Hasselström, D. (1981). Public transportation planning: a mathematical approach. PhD dissertation, Univ. of Gothenburg, Gothenburg, Sweden.

Hu, J., Shi, X., Song, J. and Xu, Y. (2005). Optimal design for urban mass transit network based on evolutionary algorithm. *Lecture notes in computer science*. In: L. Wang, K. Chen and Y. S. Ong (eds.), 3611, Springer, Berlin-Heidelberg.

Israeli, Y. and Ceder, A. (1989). Designing transit routes at the network level. *Proceedings of the First Vehicle Navigation and Information Systems Conference*, IEEE Vehicular Technology Society, 310–316.

Israeli, Y. and Ceder, A. (1995). Transit route design using scheduling and multiobjective programming techniques. *Computer-Aided Transit Scheduling. Lecture Notes in Economics and Mathematical Systems*. In: Daduna, J.R., Branco, I. and Piax ao, J. (Eds.), Springer-Verlag, Heidelberg 430, 56–75.

Israeli, Y. (1992). Transit route and scheduling design at the network level. Doctoral dissertation, Technion Israel Institute of Technology.

Kepaptsoglou, K. and Karlaftis, M. (2009). Transit route network design problem: review. *Journal of Transportation Engineering*, 135, 491–505.

Kurauchi, F., Bell, M. G. H. and Schmöcker, J-D. (2003). Capacity constrained transit assignment with common lines. *Journal of Mathematical Modelling and Algorithms*, 2, 309–327.

Lampkin W. and Saalmans P. D. (1967). The design of routes, service frequencies and schedules for a municipal bus undertaking: a case study. *Operational Research Quarterly*, 18, 375–397.

Laporte, G., Marín, A., Mesa, J. A. and Ortega, F. (2007). An integrated methodology for the rapid transit network design. *Algorithmic Methods for Railway Optimization*, 4359, 187–199.

Laporte, G., Marín, A., Mesa, J. A. and Perea, F. (2011). Designing rapid transit network design with alternative routes. *Journal of Advanced Transportation*, 45, 54–65.

Larraín, H., Muñoz, J. C. and Giesen, R. (2013). How to design express services on a bus transit network. Webseminar on BRT Center for excellence, Santiago de Chile.

Lee, Y-J. and Vuchic, V. R. (2005). Transit network design with variable demand. *Journal of Transportation Engineering*, 131, 1–10.

López, F. (2014). Conjoint design of railway lines and frequency setting under semi-congested scenarios. Thesis Report. Universitat Politècnica de Catalunya Barcelona-Tech, Barcelona.

Magnanti, T. and Wong, R. (1981). Accelerating Benders decomposition: algorithmic enhancement and model selection criteria. *Operations Research*, 29, 464–484.

Marín, A. (2007). An extension to urban rapid transit network design. *TOP*, 15, 231–241.

Marín, A. and Jaramillo, P. (2008). Urban rapid transit network capacity expansion. *European Journal of Operational Research*, 191, 45–60.

Marín, A. and Jaramillo, P. (2009). Urban rapid transit network design: accelerated Benders decomposition. *Annals of Operation Research*, 169, 35–53.

Marín, A., Mesa, J. A. and Perea, F. (2009). Integrating robust railway network design and line planning under failures. *Lectures Notes in Computer Science*, 5868, 273–292.

Marín, A. and García-Rodenas, R. (2009). Location of infrastructure in urban railway networks. *Computers & Operations Research*, 36, 1461–1477.

Marwah, B. R., Farokh, S., Umrigar, S. and Patnaik, S. B. (1984). Optimal design of bus routes and frequencies for Ahmedabad. *Transportation Research Record*, 994, 41–47.

Mauttone, A. (2011). Models and Algorithms for the optimal design of bus routes in public transportation systems. Thesis Report, Montevideo.

Ngamchai, S. and Lovell, D. J. (2003). Optimal time transfer in bus transit route network design using a genetic algorithm. *Journal of Transportation Engineering*, 129, 510–521.

Pacheco, J., Álvarez, A., Casado, S. and Gonzalez-Velarde, J. L. (2009). A tabu search approach to an urban transport problem in northern Spain. *Computers & Operations Research*, 36, 967–979.

Papadakos, N. (2008). Practical enhancements to the Magnanti-Wong method. *Operational Research Letters*, 36, 444–449.

Pattnaik, S. B., Mohan, S. and Tom, V. M. (1998). Urban bus transit route network design using genetic algorithm. *Journal of Transportation Engineering*, 124, 368–375.

Petrelli, M. (2004). A transit network design model for urban areas. In: Brebbia C. A. and Wadhwa L. C. (eds.), Urban transport X, WIT Press, U.K., 163–172.

Rao, K. V., Muralidhar, S. and Dhingra, S. L. (2000). Public transport routing and scheduling using genetic algorithms. *Computer-Aided Scheduling of Public Transport*, Berlin, Germany, 21–23 June 2000.

Saharidis, G. K. and Ierapetritou, M. G. (2009). Resolution method for mixed integer bi-level linear problems based on decomposition technique. *Journal of Global Optimization*, 44, 29–51.

Sherali, H. D. and Lunday, B. J. (2011). On generating maximal nondominated Benders cuts. *Annals of Operational Research*, 210, 57–72.

Shih, M. and Mahmassani, H. S. (1994). A design methodology for bus transit networks with coordinated operations. Tech. Rep. SWUTC/94/60016-1, Center for Transportation Research, University of Texas, Austin.

Shih, M., Mahmassani, H. S. and Baaj, M. (1998). A planning and design model for transit route networks with coordinated operations. *Transportation Research Record*, 1623, 16–23.

Shimamoto, H., Schmöcker, J-D. and Kurauchi, F. (2012). Optimisation of a bus network configuration and frequency considering the common lines problem. *Journal of Transportation Technologies*, 2, 220–229.

Silman, L. A., Barzily, Z. and Passy, U. (1974). Planning the route system for urban buses. *Computers and Operations Research*, 1, 210–211.

Soehodho, S. and Koshi, M. (1999). Design of public transit network in urban area with elastic demand. *Journal of Advanced Transportation*, 33, 335–369.

Spiess, H. and Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. *Transportation Research-Part B*, 23, 83–102.

Szeto, W. Y. and Wub, Y. (2011). A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. *European Journal of Operational Research*, 209, 141–155.

Tom, V. M. and Mohan, S. (2003). Transit route network design using frequency coded genetic algorithm. *Journal of Transportation Engineering*, 129, 186–195.

van Oudheusden, D. L., Ranjithan, S. and Singh, K. N. (1987). The design of bus route systems-An interactive location allocation approach. *Transportation*, 14, 253–270.

van Nes, R., Hamerslag, R. and Immer, B. H. (1988). The design of public transport networks. *Transportation Research Record*, 1202, 74–83.

Vuchic, V. R. (1973). Skip-stop operation as a method for transit speed increase. *Traffic Quarterly*, 27, 307–327.

Wan, Q. K. and Hong, K. Lo. (2003). A mixed integer formulation for multiple-route transit network design. *Journal of Mathematical Modelling and Algorithms*, 2, 299–308.

Zhao, F. and Ghan, A. (2003). Optimization of Transit Network to minimize transfers. Report BD-015-02. Research Center Florida Department of Transportation.

Zhao, F. (2006). Large-scale transit network optimization by minimizing user cost and transfers. *Journal of Public Transportation*, 9, 107–129

Zhao, F. and Zeng, X. (2006). Optimization of transit network layout and headway with a combined genetic algorithm and simulated annealing method. *Engineering Optimization*, 38, 701–722.

Zhao, F. and Zeng, X. (2007). Optimization of user and operator cost for large scale transit networks. *Journal of Transportation Engineering*, 133, 240–251.