SOBRAPO

MODEL AND SOLUTION METHOD TO A SIMULTANEOUS ROUTE DESIGN AND FREQUENCY SETTING PROBLEM FOR A BUS RAPID TRANSIT SYSTEM IN COLOMBIA

Fabián Martínez[1], María Gulnara Baldoquín[2] and Antonio Mauttone[3*]

ABSTRACT. We propose a model and solution method to a simultaneous route design and frequency setting problem on a main corridor from one of the Bus Rapid Transit (BRT) Systems of Colombia. The proposed model considers objectives of users and operators in a combinatorial multi-objective optimization framework and takes into account real constraints on the operation of some Colombian BRT systems not found in previous models. The problem is solved heuristically by a Genetic Algorithm which is tailored from an existing work, to consider specific characteristics of the real scenario. The methodology is validated with current data from one of the most important bus corridors in a Colombian BRT system. The results obtained improve the current solutions for this corridor.

Keywords: public transportation, genetic algorithm, BRT.

1   INTRODUCTION

Public transport substantially determines the quality of life of the inhabitants of urban areas. A public transport system well designed and operated not only provides adequate mobility for users but also contributes to solving urban problems such as noise, traffic congestion, lack of public spaces, pollution, etc. Due to the high costs of implementing rapid transit systems based on rails, developing countries (especially in Latin America), as well as industrialized nations have been adopting Bus Rapid Transit (BRT) systems, which require less investment due to some factors as the use of more economical technologies (Hensher & Golob, 2008). Particularly in Colombia, 7 cities have BRT systems implementations, known as SITM (acronym in Spanish of Integrated System of Mass Transportation).

*Corresponding author.

[1] Departamento de Ingeniería Civil e Industrial, Pontificia Universidad Javeriana, Cl. 18 #118-250, Cali, Valle del Cauca, Colombia. E-mail: fabianandresmartinezpacheco@gmail.com

[2] Departamento de Ciencias Matemáticas, Escuela de Ciencias, Universidad EAFIT, Carrera 49 No. 7 Sur-50, Medellín, Antioquia, Colombia. E-mail: mbaldoqu@eafit.edu.co

[3] Facultad de Ingeniería, Universidad de la República, J. Herrera y Reissig 565, Uruguay. E-mail: mauttone@fing.edu.uy

Bus Rapid Transit denotes a high-quality bus-based transit system that delivers fast, comfortable, and cost-effective services at metro-level capacities (Levinson et al., 2002). A BRT corridor is a section of road or contiguous roads served by one or multiple bus routes with a required minimum length of kilometers dedicated to bus lanes. There are some essential features that define a BRT as Dedicated Right-of-Way, Busway Alignment, Off-board Fare Collection and Platform-level Boarding.

All performance indicators of bus transport for monitoring and evaluating urban transport projects in Colombia (Ministerio de Transporte, 2008), reflect a main objective of improving mobility and quality of public transport services in strategic corridors of BRT systems. Although all Colombian cities have a similar regulatory framework, the design and implementation of BRT systems respond to the particularities of each city. According to data from one Survey of Quality of Life of the National Administrative Department of Statistics processed for the study developed in (Yepes et al., 2013), there has not been a strong impact of the BRT on the most vulnerable population. In (Yepes et al., 2013), as well as in the results of various surveys that take place annually, it is suggested that among the major challenges of urban transport systems in Colombia are the improvement of planning and operation of routes and the strengthening of the capacities of managers in optimizing routes.

In this work the problems of designing bus routes and their associated frequencies are studied simultaneously, in a corridor of one of the most representative BRT systems of Colombia, which we name SITM. The model and solution method proposed is based on the work of (Szeto & Wu, 2011); that work, although was not developed for BRT Systems, has some common features with the problem addressed in our study. Some works have been done for BRT systems in Colombia, in relation to the design of routes and frequencies, particularly for the BRT system in Bogotá (Walteros et al., 2015). However, there are features and real assumptions for the SITM that have not been addressed in the literature, according to the extensive review made about models, methods and real applications to simultaneously solve the route design and frequency optimization problems.

The remainder of this paper is organized as follows. This section continues with the related work (Section 1.1) and our contribution (Section 1.2). In Section 2 the problem is formulated, highlighting two subsections: how to determine the transfer stations when it is required by a route, based on the assignment model assumed (Section 2.3), and the solution method to determine the weighting factors in the weighted-sum approach considered for the multi-objective optimization model proposed (Section 2.4). In Section 3, the solution method is discussed, highlighting the differences with the method proposed in (Szeto & Wu, 2011). In Section 4, the proposed model and method are validated with real data of one of the main BRT corridors of the considered SITM. Finally, Section 5 concludes and gives some recommendations to continue this work.

1.1   Related work

The problem of designing a set of bus routes and their associated frequencies in the context of the strategic planning of a public transportation system, has been studied with different terminologies. For example, in (Baaj & Mahmassani, 1991; Cancela et al., 2015) is used the terminology

TNDP (Transit Network Design Problem); in (Guihaire & Hao, 2008; Farahani et al., 2013; Oliveira & Barbieri, 2015) the authors use the terminology that we adopt to refer to this problem: TNDFSP (Transit Network Design and Frequencies Setting Problem). In (Farahani et al., 2013) is used the terminology Transit Network Design Problem (TNDP) exclusively to the design of routes of transit lines including the origins and destinations of the routes and the sequence of links visited.

The problem of designing a set of routes and their frequencies, even treating independently the design of routes and frequencies, is difficult to solve computationally (Borndörfer et al., 2007). The TNDFSP has been tackled in the literature by means of different approaches, there are some representative reviews in (Guihaire & Hao, 2008; Farahani et al., 2013; Cancela et al., 2015).

Usually, the routes should be defined in terms of a given infrastructure of streets and stops and should cover a given origin-destination demand. In this paper, a route (or line) is a sequence of stations (bus stops) with a direction, over a corridor of a BRT system. This means that if $v = (x_1, x_2, \ldots, x_n)$ represents the vector of the n stations along the corridors (between the initial x_1 and final x_n stations), a route is represented either by v or by a subset of v, respecting the order in which the stations are visited; that is, if x_i and x_k are two stop stations on a route with $i < k$, then x_i must appear before x_k. The frequency is the time interval between two consecutive arrivals of a bus performing a route to a bus station. This definition is equivalent to the number of trips assigned to cover a route in a given period, to provide a good level of service. The demand is usually given by an origin–destination (OD) matrix, defined in different ways. For example, in (Cancela et al., 2015) each element d_{ij} of the OD matrix expresses the number of trips from i to j that should be satisfied by time unit in a given time horizon. In our paper, each element d_{ij} expresses the number of users/hour that travel between two stations i and j in a corridor. This number may differ according to the time of the day (peak-hour or not, half day), the type of day (Monday-Friday, Saturday, Sunday, holidays) and seasons of the year as holiday periods or end of the year. In Section 4, is explained how these matrices are estimated for the study case.

In real life, users of a transport system may have different trip strategies. A transit assignment model (or assignment model) describes how users of a public transport system employ the available services between different origins and destinations to plan and make their trips. The definition of this model determines the particular optimization model of routes and frequencies, since it determines the total travel time for a given solution. This assignment problem has been studied as an isolated problem and also as part of other optimization problems (Cancela et al., 2015). The assignment problem can also be divided into categories of congested and uncongested assignment problems (Farahani et al., 2013). The congested transit assignment considers, among other characteristics, the transit passenger capacity restrictions. The consideration of vehicle capacity complicates the development of an assignment model of passenger to routes, since may cause that effective frequencies are not the same as programmed frequencies. Assignment models in which congestion is considered are in (de Cea & Fernández, 1993; Cepeda et al., 2006; Larrain & Muñoz, 2008). In some papers have been used assignment models considering the capacity of the vehicles and the passenger flow on routes (Nguyen & Pallottino 1988; de Cea &

Fernández, 1993; Wu & Florian, 1993; Wu et al., 1994; Cominetti & Correa, 2001; Cepeda et al., 2006). Some authors consider the capacity constraint of the buses as a constraint of the route optimization model, not in the assignment sub-model (Baaj & Mahmassani, 1991; Constantin & Florian, 1995; Leiva et al., 2010; Cancela et al., 2015). Baaj & Mahmassani (1991) introduced the term Load Factor, which is a qualifier that indicates, for each segment of a route, if the buses carry passengers standing. Spiess & Florian (1989) and Fernández (2013) proposed an assignment model on a transit network without congestion, which can be used as part of an optimization model of routes and frequencies, controlling the decision variables in order to respect the load factor. In (Szeto & Jiang, 2014), the assignment formulation proposed by Spiess & Florian (1989) is extended to capture transfer penalty. In our paper, the assignment model assumed is due to (Spiess & Florian, 1989), applied in the context of a simple corridor in a BRT system: a user knows reasonably different routes available for its trip; and he always selects, in his initial station, the first vehicle arriving of a feasible route, operating in the corridor. If the vehicle taken doesn't arrive to the final station j, the passenger goes to the nearest station (k) of the station j that allows transfer with other(s) route(s) to arrive to the destination station j, and takes at k the first vehicle which has a stop in j.

Two of the most common objectives considered in TNDFSP studies, reflecting the interest of the users and operators, are the total travel time (users) and the fleet size (operators). Total travel time (Lampkin & Saalmans, 1967; Dubois et al., 1979; Tom & Mohan, 2003; Borndörfer et al., 2008; Cancela et al., 2015) is composed of: waiting time at station, in-vehicle travel time and transfer time. In-vehicle time is divided in two components: the time of vehicle in movement and the bus stop time in stations. The bus stop time in stations (also known as dwell time), in some real contexts as in our case study, is a constant that differs depending on the arrival station, and it is not considered in the reviewed papers. Few papers take into account other considerations as minimize excess travel time (Ceder & Wilson, 1986) and minimize excess time compared to the minimum path (Carrese & Gori, 2002). One way to model in a more realistic manner the interest of users is to consider minimizing the average deviation of the routes in proportion to the ideal travel time (travel time when there is no time for stops nor timeouts) and weighted by the number of users traveling. It is not the same, for example, the excess travel time, respect to the ideal, of 10 minutes in a trip of 50 minutes than in a trip of 20 minutes. This aspect, according to our review, has not been reflected in proposed models.

The TNDFSP is of multi-objective nature, and the weighted sum approach is normally used to model the different aspects to include in the objective function. For example, in (Zhao & Zeng, 2007) is minimized a weighted sum of costs of users and operator. In (Szeto & Wu, 2011) the objective function is a weighted sum of the number of transfers and network travel time; these weighting factors are generated in experimental way, where one of them is fixed, and the other is varied in certain interval. In (Guan et al., 2006), the authors propose a formulation that minimizes route cost, number of transfers and on-board travel time, in a weighted sum objective function with three weighting factors (which sums one). In two of test examples, the weighting factors are defined, and only is justified the factor with value zero. The simplified Hong Kong Mass

Transit Railway network is tested with various combinations of the weighting factors, where it is justified only some combinations where the values are zero or one. In (Cipriani et al., 2012) the objective function is defined as the weighted sum of operator's costs, users' costs and a penalty related to the level of unsatisfied demand. Such weights have been calibrated applying a sensitivity analysis. In (Oliveira & Barbieri, 2015) the two objectives are to minimize both passengers' costs (given by the total number of transfers, waiting and in-vehicle travel times) and operators' costs (the total required fleet to operate the set of routes). The TNDFSP is addressed using a Genetic Algorithm (GA) and the bi-objective nature of the problem is solved using an alternating objective function, which alternates from minimizing users' cost to minimizing fleet at each generation of their GA. In some papers, the objective function is the sum of some other functions (Baaj & Mahmassani, 1991), either it is not specified how the weighting factors are obtained (Walteros et al., 2015) or weights are obtained in an experimental way, even if some applications of case studies are shown (Szeto & Wu, 2011). Another less referenced approach is multicriteria. In (Janarthanan & Schneider, 1986) the authors describe and apply a computer-based multicriteria method using concordance analysis to evaluate alternative transit system designs, including objectives, criteria, normalization methods and selection of weights in the weighted sum method.

The normalization and the weighting, in the weighted sum approach, play an important role in ensuring the consistency of optimal (or near optimal) solutions obtained with the preferences expressed by the decision maker (DM) (Kaplinski & Tamošaitien, 2015). In a real context, it is important to find a solution that is both Pareto (or near Pareto) optimal and also that satisfies to the DM. It is known that the weighted sum approach works well only when the Pareto front is convex. Even in this case, there are different approaches to achieve the weighting factors, some more effective than others. The weighting factors are generally composed by two factors, one being the weight given by the DM and the other, the normalization factor. The normalization is important not only when the objectives have different metrics, in order to obtain a unidimensional numerical form, but also when the range of solutions of the objectives are very different. There are different possible normalization schemas (Haftka & Gürdal, 1992). Some of them have proved to be ineffective and are not practical (Grodzevich & Romanko, 2006). In our paper, the normalization and the weighting in the weighted sum approach are justified and contextualized to the case study presented, and according to our review about TNDFSP studies, this approach has not been used.

The resolution of TNDFSP has been tackled in the literature by means of different approaches, one of them metaheuristics. In that context, Genetic Algorithms (GA) has the most number of applications in TNDFSP (Pattnaik et al., 1998; Bielli et al., 2002; Ngamchai & Lovell, 2003; Tom & Mohan, 2003; Fan & Machemehl, 2004; Hu et al., 2005; Szeto & Wu, 2011). The proposed solution method in this paper is an adaptation of the GA proposed in (Szeto & Wu, 2011).

In (Farahani et al., 2013) ten real-world case studies about of TNDFSP are reviewed, where the size of the network varies substantially from one case to other, even under the same problem catalogue. Other recent studies are in (Szeto & Jiang, 2014; Cancela et al., 2015). None of

the reviewed papers make specific references to models and solution methods about TNDFSP, applied to BRT systems. The only case study found, applied to a simple corridor, is in (Larrain & Muñoz, 2008). However, although the real-world case study for the transport service of the city of Tin Shui Wai, Hong Kong, presented in (Szeto & Wu, 2011; Szeto & Jiang, 2014) is different from our real-world case, is the only paper of the reviewed literature that includes two common aspects with our problem: it considers a network with 28 stations connecting suburban areas to urban areas, size similar to the most corridors of the BRT studied in our paper, and the restriction of subsets of stations where the buses can start and return their trip.

Some of the reviewed papers are related to models and solution methods about TNDFSP applied to BRT systems, in particular on an isolated bus corridor (Larrain et al., 2010; Leiva et al., 2010; Scorcia, 2010; Chiraphadhanakul & Barnhart , 2013). In (Larrain et al, 2010), four parameters are defined for identifying corridor demand profiles, to determine what types of express services would be attractive on a bus corridor given the characteristics of its demand. The authors, using experimental simulations, conclude that a crucial parameter for determining the potential benefits of express services is the average trip length along the corridor and that the incorporation of express services is particularly attractive in corridors with demand profiles that increase or decrease monotonically. In (Leiva et al., 2010), four optimization models are formulated, with and without vehicle capacity constraints and transfers between lines serving a corridor. These models can accommodate the operating characteristics of a bus corridor, given an origin–destination trip matrix and a set of services that are a priori attractive. A real-world case study of a bus corridor in the city of Santiago, Chile is presented. In (Chiraphadhanakul & Barnhart, 2013), the authors seek to modify a given bus schedule on a particular corridor by optimally reassigning some number of bus trips, to operate a limited-stop service in parallel with the local service, which serves every stop along the corridor. Scorcia (2010) proposes a methodology for the design and evaluation of service configurations for limited-stop services overlapping with local services.

Although the TSW-HK case study presented in (Szeto & Wu, 2011) does not consider the simultaneous definition of routes and frequencies for an isolated corridor in a BRT system, it has some aspects common to the BRT studied in this work, that are not identified in other papers reviewed. For example, it considers the definition of subsets of stations for the initial and final station of the routes and establishes a minimum frequency for each enabled route.

Also, our solution method proposed is a heuristic algorithm based on a Genetic Algorithm, in which the crossover and mutation operators proposed have similarities with the operators presented in (Szeto & Wu, 2011); also, we use the same diversity control mechanism for the selection of individuals of each new generation. In addition, Szeto and Wu (2011) demonstrate experimentally, by means of a "t-test", that the solutions offered by their method are robust under uncertain demand, which is important in our study case where the information for the elaboration of the source data does not allow a level of confidence higher than 60%.

1.2 Contribution

The main contributions of this paper are:

1. The solution of a TNDFSP for a public transportation system with some characteristics not considered in the literature, according to the review done. We used a realistic scenario for which improved solutions were found by applying the proposed methodology.

2. The definition of components of the objective function, taking into account one aspect of the interest of the users, not previously contemplated.

3. The argumentation about the coefficients proposed in the weighted-sum approach, in the Multi-Objective Optimization Model considered.

4. Some modifications proposed to the model and solution method presented in (Szeto & Wu, 2011), in order to be able to apply such methodology to our case study.

2 PROPOSED FORMULATION

We propose a model to design simultaneously a set of routes for a corridor of a SITM and their associated frequencies with three objectives which reflect interest of users (two objectives) and operators (one objective).

The following hypotheses are considered:

- A simple BRT corridor of the SITM.

- The corridor is double lane. Each route has exactly the same stops in both directions between two points (origin and destination). There is a subset of stations for the start and return routes, because there are stations where the buses can not turn back towards the station from where it came.

- There are no restrictions about stations where the users can transfer.

- The total travel time considered is composed of: waiting time at station, in-vehicle time and transfer time. In-vehicle time is divided in two components: the time of vehicle in movement and the bus stop time in stations. The bus stop time in stations is a constant that differs depending on the arrival station.

- All users must have at least one route for travel from any station to another without transfer.

- It is assumed the assignment model proposed by (Spiess & Florian, 1989).

- Although the fleet operating on some corridors is not homogeneous, the assignment model adopted does not consider that a user can miss a route because the bus is full, so it is indifferent to consider different types of buses.

Also, some specific restrictions are considered:

- A minimum frequency for the routes operating in a corridor for a specific time slot.

- A maximum capacity of vehicles available and a maximum number of routes to be considered in a corridor.

- A maximum capacity of arrivals allowed at each station (number of vehicles/hour).

2.1   Notation and Modelling

Sets:

U	set of nodes representing corridor stations, $U \subseteq \mathbb{N} \backslash \{0\}$;
Y	set of nodes that allow south-south return (start), $Y \subseteq U$;
V	set of nodes that allow north-north return (final), $V \subseteq U$, $V \cap Y \neq \emptyset$;
i, j, k, e	indices for nodes;
n, m, q, r	indices for routes, $\epsilon \{1, 2, \ldots, R_{\max}\}$.

Parameters:

c_{ij}	time in vehicle moving from i to j (hours);
s_i	average stop time in i (hours);
d_{ie}	number of passengers traveling from i to e (per unit time, at a given horizon);
$f_{\min}$	minimum frequency for routes operating in the corridor (number of vehicles/hour);
W	available fleet size in the corridor;
$R_{\max}$	maximum number of routes in the corridor;
CE_j	maximum number of arrivals to j (number of vehicles/hour);
B_1, B_2, B_3	weighting factors in the objective function for: total travel time of the users (B_1), maximum deviation of travel (B_2) and the number of vehicles assigned to routes (B_3).

Decision variables:

X_{ijn}	1 if route n arrives node $j \neq i$ immediately after node i, and 0 otherwise, indicates the trajectory of the route n from i to j in both directions, if the route n is enabled; when it goes from north to south $i < j$ and when it goes from south to north $j < i$;
X_{0jn}	1 if route n starts at node j and 0 otherwise;
X_{i0n}	1 if route n ends at node i and 0 otherwise;

X_{00n} 1 if route n is disabled and 0 otherwise;

RT^n_{ij} 1 if route n arrives to nodes i and j, consecutives or not in the route n; 0 otherwise;

W^n_{ij} 1 if the route n has a stop in i, no matter that does not stop in j, and has a stop in common between i and j with at least other route which has a stop at j station, 0 otherwise; $j \neq i + 1$; $j \neq i - 1$;

O^n_{ij} 1 if the route n allows travel from i to j, either directly and/or with transfer, 0 otherwise;

f_n frequency of route n, (vehicles/hour);

T_n cycle time of route n, in hours (time of the vehicle moving plus the time of each stop in all the trajectory of the route n, from north to south and south to north);

V_n number of vehicles needed to operate the route;

T_{ie} expected travel time from i to e, (hours);

t^n_{ij} travel time from i to j on the route n, being i and j two stop points on the route, (hours);

T^n_{ij} expected travel time from i to j using n as initial route, where i is a stop point of the route n, j may or may not be a stopping point on the route n, (hours).

2.2   Mathematical programming formulation

$$\text{min.} \quad z = B_1 \sum_{i \in U} \sum_{e \in U} d_{ie} T_{ie} + B_2 \frac{\sum_i \sum_e (d_{ie} T_{ie}/c_{ie})}{\sum_i \sum_e d_{ie}} + B_3 \sum_{n=1}^{R_{\max}} V_n \tag{1}$$

$$\text{s.t} \quad \sum_{j \in Y \cup \{0\}} X_{0jn} = 1, \qquad \forall n \tag{2}$$

$$\sum_{i \in V \cup \{0\}} X_{i0n} = 1, \qquad \forall n \tag{3}$$

$$X_{ijn} - X_{jin} = 0, \qquad \forall i, \ j \in U, \ \forall n \tag{4}$$

$$\sum_{i \in Y \cup \{0\}, i < j} X_{0in} \geq X_{j0n}, \qquad \forall j \in Y \cap V, \ \forall n \tag{5}$$

$$\sum_{j \in U j > i} X_{ijn} \leq 1, \qquad \forall i \in U \cup \{0\}, \ \forall n \tag{6a}$$

$$\sum_{i \in U i > j} X_{ijn} \leq 1, \qquad \forall j \in U \cup \{0\}, \ \forall n \tag{6b}$$

$$f_n \geq f_{\min}(1 - X_{00n}), \qquad \forall n \tag{7}$$

$$T_n = \left[\sum_{i \in U} \sum_{j \in U, i \neq j} X_{ijn}(c_{ij} + s_i) \right]$$
$$+ \sum_{i \in U}(s_i X_{i0n} + s_i X_{0in}), \qquad \forall n \qquad (8)$$

$$V_n = T_n f_n (1 - X_{00n}), \qquad \forall n \qquad (9)$$

$$\sum_{n=1}^{R_{\max}} V_n \leq W \qquad (10)$$

$$\sum_{n=1}^{R_{\max}} \sum_{i \in U \cup \{0\} i < j} X_{ijn} f_n (1 - X_{00n}) \leq CE_j, \qquad \forall j \in U \qquad (11)$$

$$RT_{ik}^n = X_{ikn} + \sum_{j \in U k > j > i} X_{ijn} RT_{jk}^n, \qquad \forall i, \ k \in U, \ k > i, \ \forall n \qquad (12a)$$

$$RT_{ik}^n = X_{ikn} + \sum_{j \in U i > j > k} X_{ijn} RT_{jk}^n, \qquad \forall i, \ k \in U, \ k < i, \ \forall n \qquad (12b)$$

$$\sum_{n=1}^{R_{\max}} RT_{ij}^n (1 - X_{00n}) \geq 1, \qquad \forall i, \ j \in U, \ i \neq j \qquad (13)$$

$$t_{ik}^n = X_{ikn}(c_{ik} + s_i)$$
$$+ \sum_{j \in U, i < j < k} X_{ijn} RT_{jk}^n (t_{jk}^n + c_{ij} + s_i), \qquad \forall i, \ k \in U, \ k > i, \ \forall n \qquad (14a)$$

$$t_{ik}^n = X_{ikn}(c_{ik} + s_i)$$
$$+ \sum_{j \in U, i > j > k} X_{ijn} RT_{jk}^n (t_{jk}^n + c_{ij} + s_i), \qquad \forall i, \ k \in U, \ k < i, \ \forall n \qquad (14b)$$

$$W_{ij}^n = 1 - \prod_{\forall m, m \neq n} \prod_{k \in U, \ i < k < j} (1 - RT_{ik}^n RT_{kj}^m), \qquad \forall i, \ j \in U, \ j > i, \ \forall n \qquad (15a)$$

$$W_{ij}^n = 1 - \prod_{\forall m, m \neq n} \prod_{k \in U, \ i > k > j} (1 - RT_{ik}^n RT_{kj}^m), \qquad \forall i, \ j \in U, \ j < i, \ \forall n \qquad (15b)$$

$$O_{ij}^n = \max\{RT_{ij}^n, \ W_{ij}^n\}, \qquad \forall i, \ j \ \in \ U, \ j \neq i, \ \forall n \qquad (16)$$

$$T_{ij}^n = \left(\frac{\alpha}{\displaystyle\sum_{m=1}^{R_{\max}} f_m O_{ij}^m (1 - X_{00m})} + t_{ij}^n \right) RT_{ij}^n$$

$$+ \left(\frac{\dfrac{\alpha}{\displaystyle\sum_{m=1}^{R_{\max}} f_m O_{ij}^m (1 - X_{00m})} + t_{ik'}^n}{} \right.$$

$$\left. + \sum_{m=1}^{R_{\max}} \left(\frac{f_m RT_{k'j}^m (1 - X_{00m})}{\displaystyle\sum_{q=1}^{R_{\max}} f_q RT_{k'j}^q (1 - X_{00q})} \right) \right.$$

$$\left. \times \left(\frac{\alpha}{\displaystyle\sum_{q=1}^{R_{\max}} f_q RT_{k'j}^q (1 - X_{00q})} + t_{k'j}^m \right) \right)$$

$$\times W_{ij}^n (1 - RT_{ij}^n) + M(1 - W_{ij}^n)(1 - RT_{ij}^n) \quad \forall i, j \in U, \ j \neq i, \ \forall n \qquad (17)$$

$$T_{ie} = \sum_{n=1}^{R_{\max}} \frac{f_n O_{ie}^n (1 - X_{00n})}{\displaystyle\sum_{m=1}^{R_{\max}} f_m O_{ie}^m (1 - X_{00m})} T_{ie}^n, \qquad \forall i, \ e \in U, \ i \neq e \qquad (18)$$

In (1) three objectives with different weights are considered, to minimize:

- Total expected travel time (in hours), including stop times, initial wait time, the time of the vehicle in movement, and the waiting time for transfer;

- The average deviation of the routes in proportion to the ideal travel time and weighted by the number of users to carry;

- The number of vehicles required to operate routes.

The first objective aims for reducing the travel time of passengers with routes of greater demand, disfavoring users with low demand routes. The second objective is formulated considering that the same additional time for a couple of routes, where one of them takes much longer than another, is not equitable. The deviation should be measured in proportion to the ideal travel time (T_{ie}/c_{ie}) rather than in units of time ($T_{ie} - c_{ie}$). Additionally, it should be weighted by the number of passengers to maximize social benefit. The third objective aims to reduce the number of vehicles assigned to routes in order to improve the rate of passengers per kilometer, which will increase if the same demand is met with fewer vehicles. The units of each component of the objective function (1) are different: (number of passengers*hour), (dimensionless) and (number of vehicles) respectively; in Section 2.4 a method for normalization and weighting of these functions is proposed.

Type restrictions (2) and (3) ensure that each route, if enabled, starts in a station of the set Y and returns in a station of the set V. Type restrictions (4) ensure that each enabled route, has

the same trajectory in both directions. Type restrictions (5) ensure that, for each enabled route, the initial station is before the return station. Type restrictions (6a) and (6b) ensure that each enabled route, on each direction, can stop at most once, in every possible station on the route. Type restrictions (7) ensure that the frequency of each of the routes enabled operating in the corridor is greater than the minimum frequency. Type restrictions (8) determine the cycle time of each route n, according to the appropriate parameters. Type restrictions (9) determine the number of vehicles required to operate each route n, if it is enabled, considering its cycle time and frequency. Type restrictions (10) ensure that the number of vehicles needed to operate the enabled routes does not exceed the number of vehicles available in the corridor.

Type restrictions (11) ensure that the number of arrivals at each station j does not exceed its capacity. It takes into account that the stations visited by each enabled route are the same in both directions. Restrictions (12a) and (12b) guarantee, by calculating the binary variables RT_{ik}^{n},, for each route n, if it allows travel without transfer from any point i to any point k on the route. The two terms reflected in equations (12a) and (12b), the first with a binary variable, and the second with a sum expression, are exclusive and only one of them can take the value 1. When one of them takes the value 1 indicates that the route n allows to travel from i to k without transfer. If the first term is 1, indicates that nodes i and k are consecutive stop stations on the route; if the second takes the value 1, i and k are not consecutive stop stations on the route.

Type restrictions (13) ensure that all users have at least one option for travel from any point i to j without transfer. Type restrictions (14) calculate the travel time from i to k on the route n, using equations (14a) and (14b) being i and k two stop points of the route. Type restrictions (15) determine, for each route n, the value of the binary variables W_{ij}^{n}, if the route arrives to i and has a stop station in common between i and j with at least one other route that arrives to j. Type restrictions (16) determine, expressed in value taken by the binary variables O_{ij}^{n} for each possible route n, and whatever the stations of the corridor i, j, if it allows travel from i to j directly and/or performing transfer.

Type restrictions (17) estimate, expressed in the variables (T_{ij}^{n}), the travel time from i to j if n is taken as initial route (T_{ij}^{n}), whatever the stations (i, j) of the corridor are. The subscript k' in (17) refers to a particular transfer station, determined by the strategy of trip exposed, being $t_{ik'}^{n}$ and $t_{k'j}^{m}$ the travel time on the route n from the station i (respectively the travel time on the route m to the station j) where k' is a transfer station. The procedure for calculating the value of k' is shown in Section 2.1. In the formulation of T_{ij}^{n} in (17), the average waiting time for the arrival of the first vehicle, is given by:

$$\left(\frac{\alpha}{\displaystyle\sum_{m=1}^{R_{\max}} f_m O_{ij}^{m}(1 - X_{00m})} \right)$$

Note that when $O_{ij}^m = 1$ and $X_{00m} = 0$ corresponds to

$$\frac{\alpha}{\sum\limits_{a} f_a}, \alpha > 0 \tag{19}$$

where a belongs to the subset of feasible routes (routes that allow travel from i to j directly and/or with transfer).

In (18) the expected travel time T_{ie} is determined. The probability that a vehicle that operates on each of the routes that allow to travel directly or with transfer from the source station to the destination station, arrives at the station destination, is given by:

$$\frac{f_n O_{ie}^n (1 - X_{00n})}{\sum\limits_{m=1}^{R_{\max}} f_m O_{ie}^m (1 - X_{00q})} = \frac{f_n}{\sum\limits_{a} f_a}, \text{ (for feasible routes } a\text{)} \tag{20}$$

The average waiting times and probabilities, given by equations (19) and (20) respectively, are defined based on the work done by (Spiess & Florian, 1989). In (Spiess & Florian, 1989; Fernández, 2013) different values that can take α are indicated, according to the inter-arrival times of vehicles and the rate of arrival of passengers established for the transport system. In this paper, it is assumed $\alpha = 1$ supposing that the rate of arrival of passengers is uniform and that the distribution of time of inter-arrival of vehicles is exponential, with mean $1/\sum_a f_a$, belonging a to the subset of feasible routes defined).

The model presented is not linear integer mixed, NP-hard, so it is proposed a Genetic Algorithm (GA) for its solution, a modification of the GA presented in the work of (Szeto & Wu, 2011), integrating in this algorithm a heuristic procedure for setting frequencies.

2.3   Determination of the transfer station

As stated in Section 1.1, the assignment sub-model greatly influences the particular optimization model of routes and frequencies. In the proposed formulation, the transfer station selected by users impacts the total travel time for a given solution. In real life, each user chooses the transfer station according to a particular criterion. When a user has several options, he could choose the nearest transfer station, the last transfer station, or any transfer station between both. To avoid different total travel time for a same set of routes with a same frequency setting, we assume that users always choose the closest station to the destination station. It must be clear that unless the determination of the transfer station is random, any criteria used to determine the transfer station is equally valid, depending on the real scenario where it is applied.

The procedure depicted in Algorithm 1 is used to determine the closest station (k') to the destination station, for each route that allows to travel with transfer from the source station (i) to the destination station (j).

Algorithm 1

Let n be the initial route and m the enabled route(s):

For a route that allows travel only with transfer from i to j: $EC = \emptyset$

VK is a vector of zeros of length equal to $R_{\max} - \sum_{n=1}^{R_{\max}} X_{00}$;

    If $i < j$:

        $k' = i$;

        For each enabled route m except the route n

            If the route m arrives to the station j, and has at least one stop station

            in common with the route n between i and j:

                Identify arrival stations in common between i and j with the route

                n and add them to the set EC;

            End

            $k' = \max\{EC\}$;

            Add k' to the vector VK in the position m;

        End

        $k' = \max(VK)$;

    End;

    If $j < i$:

        $k' = i$;

        For each enabled route m except the route n

            If the route m arrives to the station j, and has at least one stop station

            in common with the route n between j and i:

                Identify arrival stations in common between j and i with the route

                n and add them to the set EC;

            End

            $k' = \min\{EC\}$;

            Add k' to the vector VK in the position m;

        End

        $k' = \min(VK)$;

    End;

    Identify the route(s) of transfer in k' by the position of $k' > 0$ in the VK vector.

End.

2.4 Normalization and weightings in the objective function

The normalization approach used is a variation of the method described in (Grodzevich & Romanko, 2006) using the ideal or utopian solution (z^{ideal}) and Nadir (z^{nadir}). Given the above values, a function $f(x)$ is represented on a scale between zero and one, using the function

$$f^{1}(x) = \frac{f(x) - z^{nadir}}{z^{ideal} - z^{nadir}}$$

For the problem posed it is very costly to determine z^{ideal} and z^{nadir}. For this reason, it is proposed to use z^{ideal} as the aspirated value and z^{nadir} as the value offered by the current solution.

z^{ideal} is obtained by taking the objective value offered by the current solution (sa) reduced by a percentage δ. The objective function (1) is transformed into (21):

$$\max z = B_1 \left[\frac{\left(\sum_{i \in U, e \in U} d_{ie} T_{ie} \right) - sa_1}{-\delta sa_1} \right] + B_2 \left[\frac{\frac{\sum_{i \in U, e \in U} (d_{ie} T_{ie} / c_{ie})}{\sum_{i \in U, e \in U} d_{ie}} - sa_2}{-\delta sa_2} \right]$$
$$+ B_3 \left[\frac{\sum_{n=1}^{R_{\max}} V_n - sa_3}{-\delta sa_3} \right] \tag{21}$$

where:

$B_1 + B_2 + B_3 = 1$;

$B_1, B_2, B_3 \geq 0$;

$\delta > 0$;

sa_1: the value of $\sum_{i \in U, e \in U} d_{ie} T_{ie}$ given by the current solution;

sa_2: the value of $\frac{\sum_{i \in U, e \in U} (d_{ie} T_{ie} / c_{ie})}{\sum_{i \in U, e \in U} d_{ie}}$ given by the current solution;

sa_3: the value of $\sum_{n=1}^{R_{\max}} V_n$ given by the current solution.

The values B_1, B_2 and B_3 (weighting factors in the objective function) are not determined experimentally. It is used the Analytic Hierarchy Process (AHP) (Saaty, 1994). In Section 4 the values determined by the company are indicated, using the AHP methodology.

3   SOLUTION METHOD

The proposed solution method is an adaptation from the one developed by (Szeto & Wu, 2011). The main differences between the methods are the problem formulation and the definition of route. In the SITM, a route is a totally ordered set of stations, while for TSW-HK, the stations do not follow strictly an order, since the service does not operate over a corridor (Fig. 1). This poses some differences in the components of the Genetic Algorithm used as well as in the proposed heuristics. In the proposed method, the chromosome representation is the same as the one used by (Szeto & Wu, 2011). However, the initialization method is completely different, because a feasible route set must fulfill the type restrictions (4), (5) and (13). Another difference with respect to the work of (Szeto and Wu, 2011) is that our frequency setting heuristic considers a variable fleet size. This is due to the component of the objective function related to the number of necessary vehicles and the formulation of the type restriction (11) which limits indirectly the maximum frequency on routes. In addition, the initial solution to this problem is not random. In the case of TSW-HK, although there is a type restriction which guarantees that the number of necessary vehicles to operate the routes does not exceed the available number of vehicles, it is always used the whole fleet, since the objective function minimizes the total travel time of users and the maximum frequency is not bounded.

We adopt the two crossing operators proposed by (Szeto & Wu, 2011) and two of the four mutation operators. We use the same diversity control mechanism for the selection of individuals of each new generation. We do not use a stop sequence improvement procedure, since over a trunk corridor of SITM, the routes are defined as totally ordered sets, since for any set of stops there exists only one sequence. Note that for similar instances in terms of number of stations and size of origin-destination matrix, the search space in TSW-HK is much larger than the one corresponding to the SITM, even though the formulation of (Szeto & Wu, 2011) is limited on the maximum number of intermediate stops for each route.
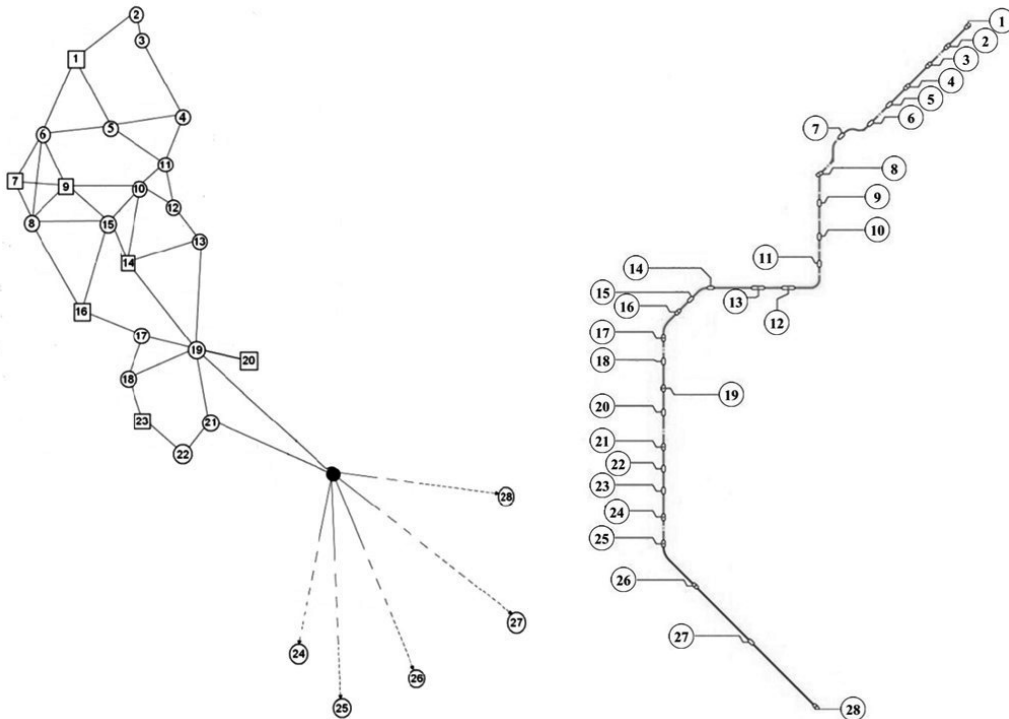
(a) The TSW-HK bus network.   (b) A single corridor in the BRT system.

Figure 1 – The investigated networks in the two study cases.

Figure 2 illustrates the main structure of the Genetic Algorithm (GA) proposed. The initial population is generated randomly. The heuristic for the frequency setting is executed to evaluate the fitness function of each individual. To perform the crossing, the parent individuals are selected by using the roulette method. A number of children equal to the number of individuals in the initial population, are generated by executing one of the proposed crossover operators. Then, the mutation operator is applied to all the children. A repair operator is applied to all the children which do not fulfill the type restriction (13). At last, the surviving individuals from the sets of parents and children are selected by using a diversity control mechanism. At each iteration, the number

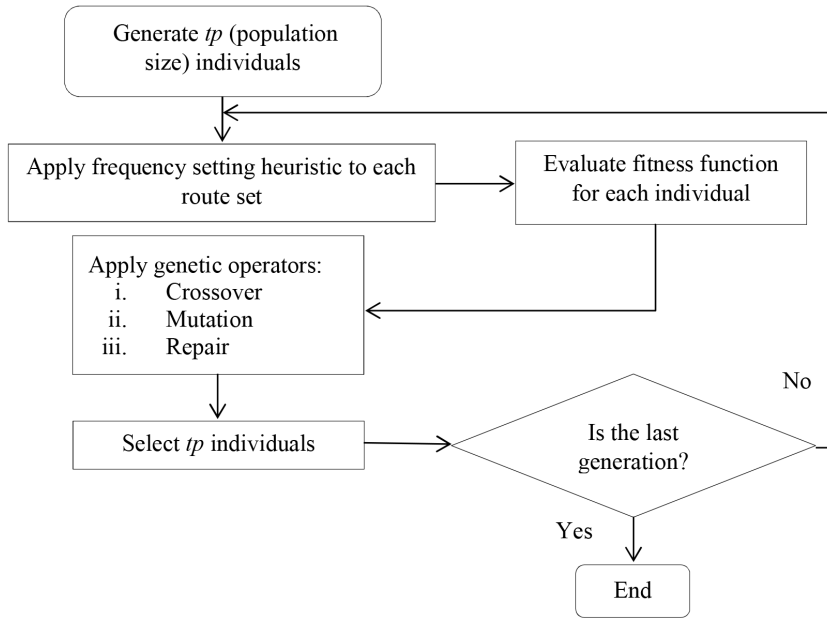of individuals is constant. The process is repeated until a pre-specified number of generations is attained.

Figure 2 – Structure of the proposed Genetic Algorithm.

3.1   Solution representation and generation of the initial population

A chromosome (individual) of the GA represents a set of R_{max} routes. In the SITM, a route r_n is a totally ordered set of stations $i \in U$. The chromosome is a matrix of R_{max} rows and number of columns equal to the number of stations of the corridor. A random feasible individual which fulfills type restrictions (2)-(6) and (13) is generated by using the procedure depicted in Algorithm 2. The process is repeated a number of times necessary to complete the size of the population. In the initialization procedure, first, the start and returning station of each of the R_{max} routes are assigned. It is guaranteed that at least one route has the first station as the starting one and the last station as the returning one. Then, combinations (i, j) are inserted randomly among the R_{max} routes, in order to fulfill the type restriction (13). After sorting the stations on each route and eliminating repeated stations, the types of restrictions (2)-(6) are fulfilled.

3.2   Heuristic procedure for frequency setting

The route frequency setting procedure is executed before the evaluation of the fitness function of the GA. When solving this problem, we should ensure that the number of vehicles necessary to operate the routes does not exceed the available fleet. Also, we should respect the capacity of each station. We formulate this problem as an assignment of no more than W available vehicles among the routes defined in the chromosome of the GA. A solution is a vector of $R_{max} + 1$ positions,

Algorithm 2

For route $n = 1$
    Initial node $= 1$;
    Returning node $=$ last station from set U;
End;
For route $n = 2 \ldots R_{\max}$
    Select randomly an initial node $i \in Y$ and a returning node $j > i \in V$ for route n;
    Next route;
End;
For $i \in U$
    For $j > i \in U$
        If i and j are not present at any route of the individual:
            Select randomly a route from the set of routes having initial station
            before i and returning station after j; insert station i and j;
        End;
        Next j;
    End;
    Next i;
End;
For route $n = 1 \ldots R_{\max}$
    Sort route n and delete repeated stations;
    Next route;
End.

where the first one represents the number of unused vehicles. The second position corresponds to the number of vehicles assigned to r_1 and so on. Taking into account that $f_n = V_n/T_n$ (9), the following procedure is used to generate (if possible) a feasible solution with respect to type restrictions (7), (10), (11) and (13), where m and n are positions in the solution vector. In order to calculate the number of vehicles necessary to operate each route with minimum frequency $f_{\min}$, we create a vector with $R_{\max}$ positions with these values, denoted as VN.

According to the procedure depicted in Algorithm 3, the type restriction (10) always is fulfilled since the sum of all values in the solution vector is equal to W. However, in order to fulfill type restrictions (7), (11) and (13), for each violated restriction we include a penalization in the fitness evaluation and the selection of survival individuals discards unfeasible elements. As a result of this procedure we can obtain disabled routes, i.e., routes without vehicles to operate.

It must be noted that each time Algorithm 3 is called with the same set of routes, it could generate a different set of frequencies and therefore a different fitness value. This is consistent with the search strategy of the overall optimization algorithm, since the goal of the stochastic aspect in Algorithm 3 is to explore solutions to avoid local optima.

Algorithm 3

Let $S = (R, F)$, where R is the set of routes, F the set of frequencies for R, and $F[q]$ the element in the position q from vector F.

Let VN a vector with $Rmax$ positions, where $VN[q]$ is the element in the position q of the vector.

Let VS a vector with $Rmax + 1$ positions, where $VS[k]$ is the element in the position k of the vector; values in VS from position 2 are those of vector VN.

Let nvn be the sum of all values in VN.

$q = 1$

While $q \leq Rmax$

    $VN[q] = min\{k \in \mathbb{N} | k \geq f_{\min} T_q\}$

    $F[q] = VN[q]/Tq$

    $q = q + 1$

End.

If nvn is less than the number of available vehicles W, then $VS[1]$ is equal to $W - nvn$, otherwise is zero.

$vae = nvn - W$;

While $vae > 0$:

    Select randomly a number k from the set $\{k \in \mathbb{N} | 2 \leq k \leq Rmax + 1\}$;

    $VS[k] = VS[k] - 1$;

    $vae = vae - 1$;

End.

Evaluate S using (21);

For $m = 1$

    $VP = \{m + 1 \ldots R_{\max} + 1\}$;

    While $VP \neq \emptyset$:

        Select randomly an element n from set VP;

        $VP = VP - \{n\}$;

        While the value of the objective function (1a) is improved or unchanged

        and the number $VS[m]$ is greater than zero:

            $VS[m] = VS[m] - 1$; $VS[n] = VS[n] + 1$; $F[n-1] = VS[n]/T_{n-1}$;

            evaluate S using (21)

        End;

        While the value of the objective function (1a) is improved or unchanged

        and the number $VS[n]$ is greater than zero:

            $VS[n] = VS[n] - 1$; $VS[m] = VS[m] + 1$; $F[n-1] = VS[n]/T_{n-1}$;

            evaluate S using (21);

        End;

    End;

End

Algorithm 3 – (continuation)

For $m = 2 \ldots R_{\max}$:
    $VP = \{m + 1 \ldots R_{\max} + 1\}$;
    While $VP \neq \emptyset$:
        Select randomly an element n from set VP;
        $VP = VP - \{n\}$;
        While the value of the objective function (1a) is improved or unchanged,
        the number $VS[m]$ is greater than zero and routes $(m-1)$
        and $(n-1)$ are different:
            $VS[m] = VS[m] - 1$; $VS[n] = VS[n] + 1$; $F[n-1] = VS[n]/T_{n-1}$;
            $F[m-1] = VS[m]/T_{m-1}$; evaluate S using (21)
        End;
        While the value of the objective function (21) is improved or unchanged,
        the number $VS[n]$ is greater than zero and routes $(m-1)$
        and $(n-1)$ are different:
            $VS[n] = VS[n] - 1$; $VS[m] = VS[m] + 1$; $F[n-1] = VS[n]/T_{n-1}$;
            $F[m-1] = VS[m]/T_{m-1}$; evaluate S using (21)
        End;
    End;
    Next m;
End
For each pair of identical routes, take the number of vehicles assigned to both and assign them
only to one route;
End.

3.3   Crossover and mutation operators

As in (Szeto & Wu, 2011), we propose two crossing operators: one for crossing routes and other
for crossing stops between routes. Both operators are applied to individuals selected from the
population by using the roulette method. For each crossing, one operator is selected randomly.

The operator for crossing routes exchanges sets of routes between two individuals, as it is done
in (Szeto & Wu, 2011). We generate two numbers k and l, $\{k, l \in \mathbb{N} \mid 1 = k, l = R_{\max}, k = l\}$,
then, the routes between r_k and r_l are exchanged between the selected individuals.

The operator for crossing stops proposed in this work, exchanges sequences of nodes between
two routes having the same return station as the parents, where the sequence of exchanged nodes
is limited by the shortest route (the one having least cost). For instance, in Figure 3 route m
from parent 1 is shorter than route n from parent 2; note that route m starts at node 19. In this
operator, we first select randomly one route from one parent; then, we select randomly from the
other parent, a route with the same return station as the previously selected route. Afterwards, we
select randomly two different stations between the initial and returning stations of the shortest
selected route. Finally, the nodes between the two selected nodes are exchanged between both
routes. Figure 3 illustrates the proposed operator for crossing stops. Route m from parent 1and

route n from parent 2 (having the same return station) are selected randomly. Then, two nodes between the initial and returning stations of the shortest route (route m from parent 1) are selected randomly (nodes 21 and 27 in the example). At last, the nodes between the selected ones are exchanged to generate route m in child 1 and route n in child 2.
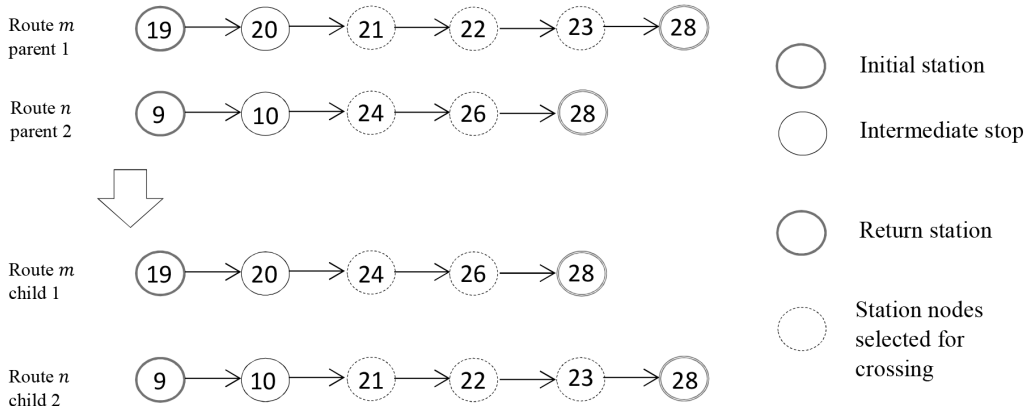
Figure 3 – Crossing of stops.

There are two situations in which the operator for crossing stops cannot be applied. First, when the second parent does not have a route with the same returning point as the one of the route selected from the first parent. Second, when some of the selected routes has consecutive initial and returning stations, since there are not nodes to exchange. If some of these situations arise, the process is repeated. If after a given number of trials, a pair of routes to apply the operator for crossing stops is not found, the crossing of routes is applied. The operator for crossing stops proposed by (Szeto & Wu, 2011) selects randomly sequences of intermediate stops from each route, and the stops are exchanged. This idea cannot be used in our method, since it would generate unfeasible solutions with respect to the type restriction (2).

In (Szeto & Wu, 2011), four mutation operators are proposed: (a) insert, (b) delete, (c) swap and (d) transfer. In this work, we use only the insert and delete operators (Fig. 4) since the application of the other two operators would generate solutions which violate the type restrictions (2) and (3). The used operators generate a perturbation in a single route selected randomly from the individual. The insert (delete) operator inserts (deletes) a node to (from) the selected route. In both operators, the node to be inserted or deleted is selected randomly from the positions between the initial and returning routes; it could not belong to the route. The route remains unchanged if the node to delete does not belong to the route or if the node to insert already belongs to the route. A route with consecutive initial and returning stations cannot be mutated. For each mutation, the specific operator is determined randomly. The mutation probability is a parameter of the algorithm.

Once the crossover and mutation operators are applied, the type restriction (13) could be violated in the children individuals. For this reason, a repair operator is proposed.

3.4 Repair operator

The repair operator works as follows. For each individual which violates the type restriction (13), each missing combination of stops (i, j) is inserted in routes selected randomly from the set of routes having initial station before node i and returning station after node j. Then, repeated nodes from each route of repaired individuals are eliminated.
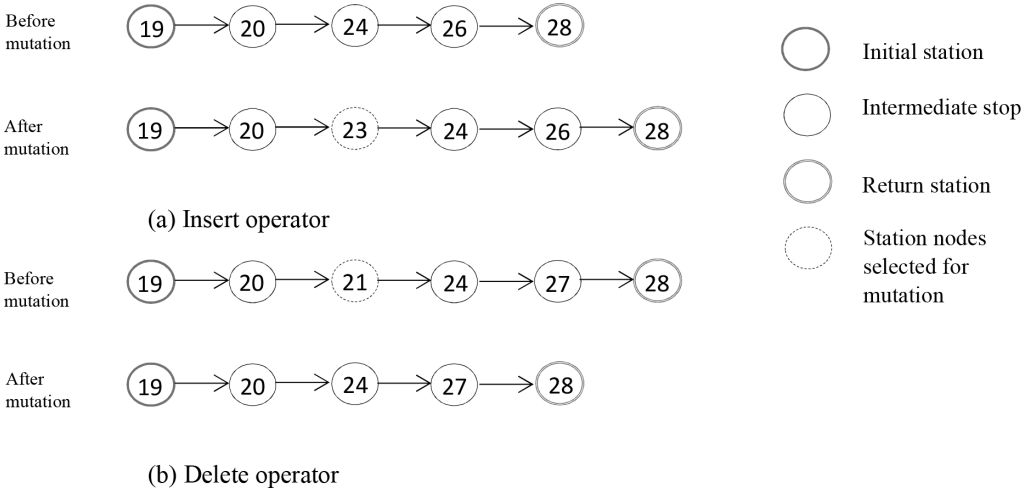

(a) Insert operator

(b) Delete operator

Figure 4 – Mutation operators.

3.5 Control of diversity

In order to control the diversity of the population, we implement the mechanism of survival probability assignment proposed in (Szeto & Wu, 2011), which is based on the DCGA (Diversity Control Genetic Algorithm) proposed in (Shimodaira, 2001). It works by sorting decreasingly the individuals of the population with respect to their fitness values and selecting randomly according to probability $P_{(s)}$ (see expression 22), which is based in the Hamming distance with respect to the individual with highest fitness in the population. The Hamming distance between two sequences of characters with same length is the number of differences between each position of the sequence (it measures the minimum number of substitutions required to change from one sequence to the other). This method favors individuals with good fitness values and high contribution to diversity.

$$P_{(s)} = \left\{ (1 - c) \frac{h}{L} + c \right\}^{a} \tag{22}$$

In (22), h is the Hamming distance between the solution s and the best individual from the population. L is the length of the chromosome, $c \in [0.0; 1.0]$ and a is a real positive number. A suitable selective pressure is obtained by tuning c and a. For a problem with many local optima, a small selective pressure can be induced by reducing the value of c and/or by increasing the value of a.

The Hamming distance was defined originally for binary codes. For that reason, its computation for solutions of our problem is not immediate. For that reason, Szeto & Wu (2011) propose defining the Hamming distance as the number of different pairs of consecutive nodes between two routes, each one from a different individual. The DCGA proposed by Shimodaira (2001) is applied in this work by applying the procedure depicted in Algorithm 4.

Algorithm 4

Delete repeated and/or not feasible individuals in M;
Select the best individual I_{best} from M to be part of the following generation:
$NG = \{I_{best}\}$. $M' = M \setminus \{I_{best}\}$;
tp: size of population;
While there are individuals in M' and the number of individuals in NG is not greater than tp:
    e = random number in [0, 1];
    I_s = best individual of M';
    If $e < P_{(I_s)}$:
        include I_s in the new generation NG;
    End
    $M' = M' \setminus \{I_s\}$;
End
While the number of individuals in NG is not greater than tp:
    Create randomly an individual following the initialization procedure.
    If after setting its frequencies it fulfills the type of restrictions (7), (11) and (13),
    it is included in NG;
End.

By selecting solutions according to the Hamming distance with respect to the best solution (and not considering survival individuals identical in each generation), we reduce the risk of losing information generated by moderately good individuals, which could be used afterwards to find the global optimum. The loss of good solutions depends on the selective pressure induced by the setting of values to parameters a and c. A reduction in the selective pressure induced by this mechanism of diversity control, affects negatively the efficiency of the GA. However, the definition of routes in the corridor of SITM is a strategic decision which does not require a quick response.

4   VALIDATION OF THE PROPOSED METHOD

Data was provided by the transit operator of the corridor, including current frequencies and maximum number of routes. The corridor comprises 28 stations (Fig. 5), where:

$$U = \{1, 2, \ldots 28\};$$
$$Y = \{1, 3, 9, 19, 24\};$$
$$V = \{11, 19, 24, 28\};$$

Other data considered in the case study comprises:

- Capacity of the stations, expressed in number of vehicles per hour.

- Origin-Destination (OD) matrix, expressed in passengers per hour. The number of passengers traveling between each pair of stations over a corridor depends on the time of day (peak, off-peak), the type of day (Monday-Friday, weekends) and the season of the year. The OD matrix provided by the operator corresponds to the time period between 6:00 and 7:00 in the morning of a representative working day. These data correspond to the period of highest demand in the system (Ministerio de Transporte, 2008). The EMME software (INRO, 2017) for transit planning was used by the operator to obtain an OD matrix update, by applying a growth factor method based on a historical OD matrix.

- Average stopping (dwell) time at each station (in seconds).

- Travel time of vehicles (in hours) between each pair of consecutive stations of the corridor, in the same time horizon as the OD matrix. That time can be obtained by two different ways. First, in terms of the average speed and the distance between stations, ignoring such factors like traffic flow and lights, real-time control procedures and driver behavior, among others. The second way is based on the average of observed (historical) travel time in the time horizon; this considers all the factors influencing the parameter of interest. From the point of view of our optimization model, the travel time between two consecutive stations is constant (the same for every different solution to the problem). Therefore, the travel time is relevant only for computing frequencies, since the reduction in the expected travel time and the average deviation of the routes depend on the number of stops, the stopping time at each station, the number of transfers and the waiting time. For this reason, measures which consider minimization of number of stops and/or number of transfers, maximizing the number of direct trips or setting limited-stop routes, are good objective functions when we aim to maximize the level of service subject to a level of profit of the operators.

We used a computer with Core i7 processor and 4 GB of RAM. Table 1 shows results of applying the AHP methodology in order to weight the objectives. The consistency ratio (CR) value indicates the validity of the comparisons between each objective made by the Decision Maker (DM). In this instance, the objective function (1) only allows solutions that adapt to the specific needs of the DM by this weighting of the objectives.

Currently there are three routes operating in the corridor. The first one stops at every station. The second one traverses the whole corridor and stops only at one station. The last one covers only a portion of the corridor. The attributes of each route are given in Table 2. We set as minimum allowable frequency, the value corresponding to the current route with minimum frequency, that is $f_{min} = 8.33$ vehicles per hour (Route 2).

For the current solution, the expected total travel time of the 15186 passengers is 7025 hours, using the whole fleet (52 vehicles) and with average deviation of routes greater than 0.67. These values are used in the normalization of the objective function explained in Section 2.3: $sa_1 = 7025$ (hours), $sa_2 = 1.6704$, $sa_3 = 52$ (vehicles).
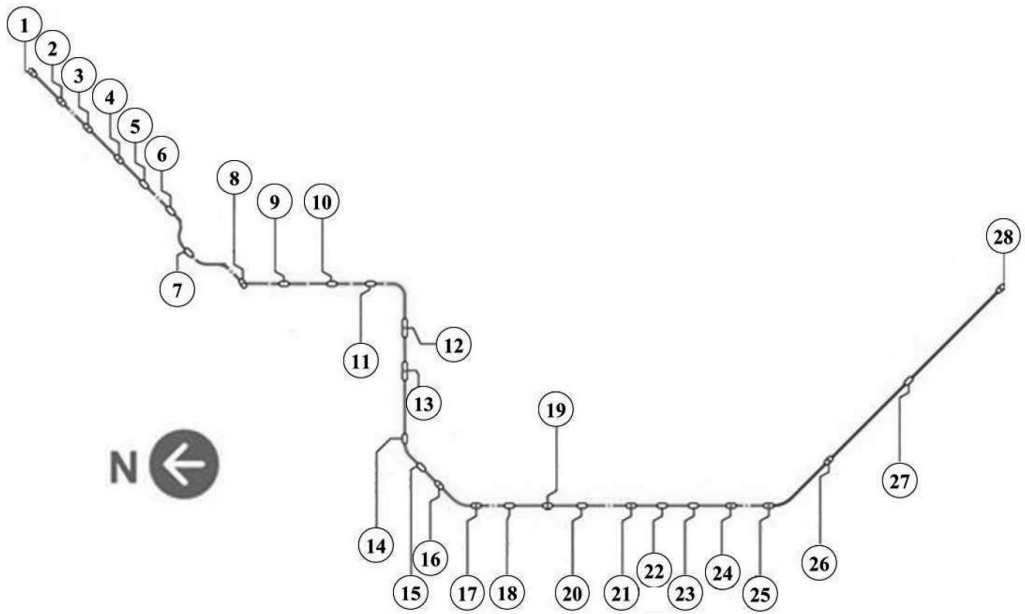
Figure 5 – Corridor of SITM.

Table 1 – Weighting of the objectives. A (total expected travel time), B (average deviation of the routes in proportion to the ideal travel time), C (number of vehicles required to operate the routes).

	A	B	C	Ratio Scale of Priority
A	1	5	6	$B_1 = 0.71$
B	1/5	1	3	$B_2 = 0.2$
C	1/6	1/3	1	$B_3 = 0.09$
	1.37	6.33	10	Consistency Ratio (CR) = 0.08

Table 2 – Current routes and frequencies.

Route	Number of vehicles assigned	Frequency (vehicles/hour)	Stations
1	22	9.47	1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16, 17,18,19,20,21,22,23,24,25,26,27,28
2	18	8.33	1,2,3,4,8,12,16,17,19,21,24,25,26,28
3	12	8.63	1,2,3,4,9,10,14,15,17,19

The number of routes enabled in a given individual depends on the number of available vehicles, the capacity of the stations and the minimum allowed frequency. In order to state the maximum number of routes $R_{\max}$ over the corridor, we propose the following formulation:

$$R_{\max} = \min\left\{\max\{k \in \mathbb{N} \mid k = CE_{\min}/f_{\min}\}, \min\{k \in \mathbb{N} \mid k = W/nv_{\max}\}\right\},$$

where $CE_{\min}$ is the capacity of the station with minimum capacity and $nv_{\max}$ is the number of vehicles necessary to operate the longest route (it passes by all stations and stops at all of them) with minimum frequency. The term $\max\{k \in \mathbb{N} \mid k = CE_{\min}/f_{\min}\}$ considers the fact that a feasible solution should allow the maximum number of routes and all of them can arrive to the station(s) with lowest capacity. In our case $\max\{k \in \mathbb{N} \mid k = 42/8.33\} = 5$. The term $\min\{k \in \mathbb{N} \mid k = W/nv_{\max}\}$ considers the fact that the available fleet should be sufficient to enable the maximum number of routes, thus ensuring the fulfillment of minimum frequency, assuming that not all routes will use $nv_{\max}$ vehicles. In our case $nv_{\max} = 20$, $\min\{k \in \mathbb{N} \mid k = 52/20\} = 3$, therefore $R_{\max} = 3$.

We set the constant δ (normalization of the objective function) as 0.05. The size of the population is 10. We generate the same number of children in the crossover operator and all of them are mutated, where each operator (insert and delete) has the same probability (0.5) of being applied. For setting these parameters, we took the work of (Szeto & Wu, 2011) as reference. Then, we tuned the parameters a and c. Figure 6 shows different combinations of them. The higher is the area below the curve ($P_{(s)}$ vs. h/L), the higher is the selective pressure. For $c = 1$ the selection of survivals is totally elitist (only the best individuals survive). For $c = 0$ and $a = 1$, the probability of selection is directly proportional to h/L. Assuming that a and c are real numbers in the range $[0, 1]$ with at least four significant numbers, we could generate more than 99 million of different combinations.

In order to adjust parameters a and c, we tested different combinations having a value for $P_{(s)}$ less than 0.01 when $h = 0$, considering that if we allow surviving the best individual, the probability of selection $P_{(s)}$ should be small. Among the combinations tested, the one which exhibited best performance corresponds to $a = 0.9999$ and $c = 0.0001$. Figure 7 illustrates the convergence process of the GA; we can note that convergence is attained after 300 generations. This execution takes approximately 180 minutes.

To adjust the population size (tp) we varied its value between 10 and 20 individuals. For those extreme values, the solutions obtained do not exhibit significant differences. However, for $tp = 20$ the number of generations needed to attain convergence increases, therefore, also the execution time increases. For that reason, we set the value $tp = 10$.

4.1   Results

Table 3 shows results obtained by running the GA with the parameters adjusted as explained above. Table 4 shows the best configuration of routes and frequencies obtained.

The configuration of routes shown in Table 4 has two routes. The first one stops at every station of the corridor. The second one passes by every node but it stops only at some stations, skipping
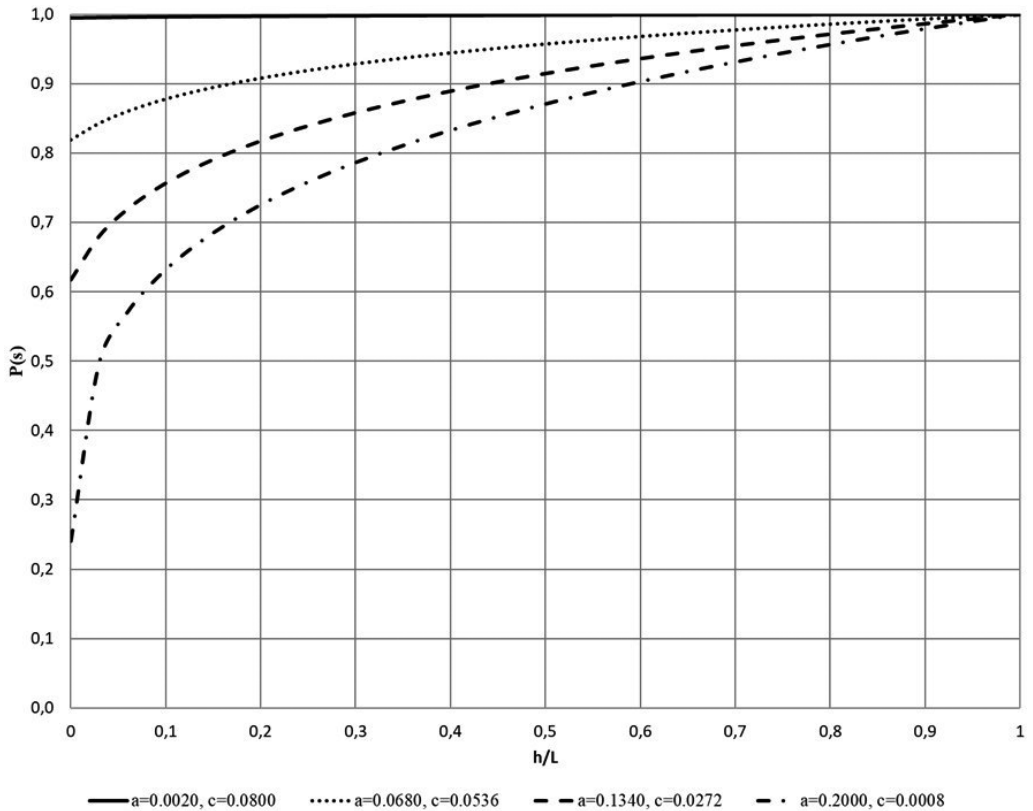
Figure 6 – Surviving probability of an individual for different combinations of a and c.

Table 3 – Solution obtained with the proposed method.

	Average last generation	Best individual	Current solution
Objective function (21)	0.9547	0.9597	0.0000
Expected total travel time (thousand of hours)	6.7620	6.7610	7.0252
Average deviation of routes	1.50	1.49	1.67
Required number of vehicles	52	52	52

stations 5, 6, 7, and 20. In fact, these stations have the smallest flow of passengers. The obtained solution reduces overall expected travel time up to 264.2 hours and reduces the weighted average deviation of routes with respect to the ideal travel time up to 11%, using the same number of vehicles and less routes. In our solution, the average waiting time at the station for all passengers (except those who travel from and to 5, 6, 7 and 20) is of 2.7 minutes. It is worth mentioning that results are sensitive to the weighting of the different objectives as well as to the hypothesis assumed regarding passenger behavior.
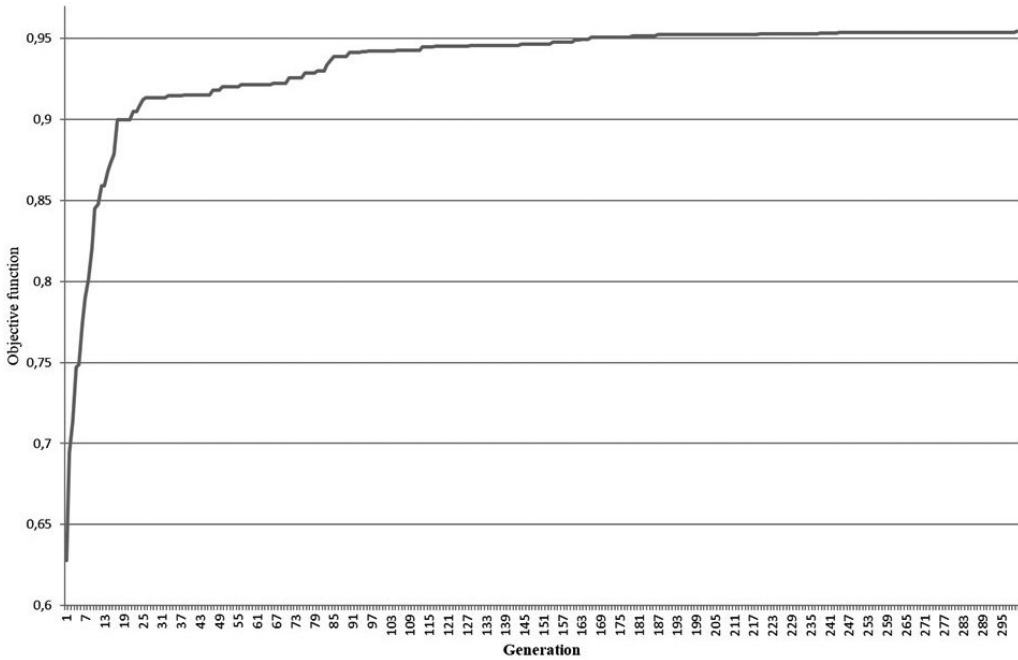
Figure 7 – Convergence of the proposed Genetic Algorithm.

Table 4 – Best configuration of routes and frequencies obtained by the proposed method.

Route	Number of vehicles assigned	Frequency (vehicles/hour)	Stations
1	30	12.9199	1,2,3,4,5,6,7,8,9,10,11,12,13,14, 15,16, 17,18,19,20,21,22,23,24,25,26,27,28
2	22	9.6305	1,2,3,4,8,9,10,11,12,13,14,15,16,17,18, 19,21,22,23,24,25,26,27,28

5   CONCLUSIONS AND FUTURE WORK

In this work, we proposed a mathematical model and a solution method to the problem of determining simultaneously routes and frequencies for a main corridor of SITM in Colombian cities. Both problem formulation and solution method are adapted from the work of (Szeto & Wu, 2011), proposed originally for the city of Tin Shui Wai, Hong Kong. The results obtained with the methodology implemented are consistent with the stated objectives, their weightings and the hypothesis assumed concerning passenger behavior, which are different from the one used by (Szeto & Wu, 2011). We validated the methodology using real data from one of the main corridors of SITM in one Colombian city. The results obtained improve the current solution in terms of level of service to the users, which is recognized by the operator of the corridor.

For future work, we identify three main lines which deserve attention. First, a study of the passenger behavior observed in the scenario where the methodology is applied would allow for tailoring the assignment sub-model adopted. Moreover, that sub-model could include different types of behavior, for example considering stochastic perception of passenger travel time. Second, assign more penalization to the perception of dwell time (the time buses stop at the stations) would enrich the passenger behavior model. Users that perform short trips will be indifferent to that time, whereas passenger performing long trips will be sensitive and therefore, they will prefer express services (Larraín, 2013). Third, taking into account that in this study the proposed methodology is applied only to a trunk corridor, the joint optimization of both trunk and feeder routes would be more realistic, since relevant interplays are observed between these different types of routes. Note that a portion of the total demand for public transportation, is generated (either produced or attracted) at zones reachable only by feeder routes. Finally, applying the proposed methodology to other SITM scenarios would enrich the study, contributing to the generalization of the proposal.

REFERENCES

[1] BAAJ MH & MAHMASSANI HS. 1991. An AI-based approach for transit route system planning and desing. *Journal of Advanced Transportation*, **25**(2): 187–210.

[2] BIELLI M, CARAMIA M & CAROTENUTO P. 2002. Genetic algorithms in bus network optimization. *Transportation Research Part C: Emerging Technologies*, **10**(1): 19–34.

[3] BORNDÖRFER R, GRÖTSCHEL M & PFETSCH M. 2007. A column-generation approach to line planning in public transport. *Transportation Science*, **41**(1): 123–132.

[4] BORNDÖRFER R, GRÖTSCHEL M & PFETSCH M. 2008. Models for line planning in public transport. In: *Computer-Aided Systems in Public Transport* [edited by M. HICKMAN, P. MIRCHANDANI & S. VOSS] Springer Berlin/Heidelberg, 363–378.

[5] CANCELA H, MAUTTONE A & URQUHART ME. 2015. Mathematical programming formulations for transit network design. *Transportation Research Part B: Methodological*, **77**: 17–37.

[6] CARRESE S & GORI S. 2002. An urban bus network design procedure. *Applied Optimization*, **64**: 177–196.

[7] DE CEA J & FERNÁNDEZ E. 1993. Transit assignment for congested public transport systems: An equilibrium model. *Transportation Science*, **27**(2): 133–147.

[8] CEDER A & WILSON NHM. 1986. Bus network design. *Transportation Research Part B: Methodological*, **20**(4): 331–344.

[9] CEPEDA M, COMINETTI R & FLORIAN M. 2006. A Frequency-Based Assignment Model for Congested Transit Networks with Strict Capacity Constraints: Characterization and Computation of Equilibria. *Transportation Research Part B: Methodological*, **40**(6): 437–459.

[10] CHIRAPHADHANAKUL V & BARNHART C. 2013. Incremental bus service design: combining limited-stop and local bus services. *Public Transport*, **5**: 53–78.

[11] CIPRIANI E, GORI S & PETRELLI M. 2012. Transit network design: A procedure and an application to a large urban area. *Transportation Research Part C: Emerging Technologies*, **20**(1): 3–14.

[12] COMINETTI R & CORREA J. 2001. Common-lines and passenger assignment in congested transit networks. *Transportation Science*, **35**(3): 250–267.

[13] CONSTANTIN I & FLORIAN M. 1995. Optimizing frequencies in a transit network: a nonlinear bi-level programming approach. *International Transactions in Operational Research*, **2**(2): 149–164.

[14] DUBOIS D, BEL G & LLIBRE M. 1979. A set of methods in transportation network synthesis and analysis. *The Journal of the Operational Research Society*, **30**(9): 797–808.

[15] FAN W & MACHEMEHL RB. 2004. Optimal transit route network design problem: Algorithms, implementations, and numerical results. Technical Report No. SWUTC/04/167244-1. Austin, Texas: Center for Transportation Research, University of Texas.

[16] FARAHANI RZ, MIANDOABCHI E, SZETO WY & RASHIDI H. 2013. A review of urban transportation network design problems. *European Journal of Operational Research*, **229**(2): 281–302.

[17] FERNÁNDEZ A. 2013. Modelos matemáticos de asignación de tránsito, aplicación a la red metropolitana de la ciudad de México y sus efectos en el STC-Metro. Tesis de Maestría en Ciencias, Departamento de Matemáticas Aplicadas e Industriales, Universidad Autónoma Metropolitana-Iztapalapa.

[18] GRODZEVICH O & ROMANKO O. 2006. Normalization and Other Topics in Multi-Objective Optimization. Proceedings to the *Fields-MITACS Industrial Problem Solving Workshop*, Toronto, Canada.

[19] GUAN J, YANG H & WIRASINGHE S. 2006. Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B: Methodological*, **40**(10): 885–902.

[20] GUIHAIRE V & HAO J-K. 2008. Transit Network Design and Scheduling: a Global Review. *Transportation Research Part A: Policy and Practice*, **42**: 1251–1273.

[21] HAFTKA RT & GÜRDAL Z. 1992. Elements of Structural Optimization. Springer Netherlands.

[22] HENSHER DA & GOLOB TF. 2008. Bus rapid transit systems: a comparative assessment. *Transportation*, **35**(4): 501–518.

[23] HU J, SHI X, SONG J & XU Y. 2005. Optimal design for urban mass transit network based on evolutionary algorithms. In: Advances in natural computation [edited by L. WANG, K. CHEN & Y. ONG], Springer Berlin/Heidelberg, 429–429.

[24] INRO. 2017. INRO Software. Available at: `http://www.inrosoftware.com`. Last accessed March 1st, 2017.

[25] JANARTHANAN N & SCHNEIDER J. 1986. Multicriteria evaluation of alternative transit system designs. *Transportation Research Record*, **1064**: 26–34.

[26] KAPLINSKI O & TAMOŠAITIEN J. 2015. Analysis of normalization methods influencing results: a review to honour Professor Friedel Peldschus on the occasion of his 75th birthday. *Procedia Engineering*, **122**: 2–10.

[27] LAMPKIN W & SAALMANS PD. 1967. The design of routes, service frequencies, and schedules for a municipal bus undertaking: A case study. *Operational Research Quarterly*, **18**(4): 375–397.

[28] LARRAÍN H. 2013. Diseño de Servicios Expresos para Buses. PhD Thesis. Pontificia Universidad Católica de Chile.

[29] LARRAÍN H & MUÑOZ J. 2008. Public transit corridor assignment assuming congestion due to passenger boarding and alighting. *Networks and Spatial Economics*, **8**(2-3): 241–256.

[30] LARRAIN H, GIESEN R & MUÑOZ J. 2010. Choosing the right express services for bus corridor with capacity restrictions. *Transportation Research Record*, **2197**: 63–70.

[31] LEIVA C, MUÑOZ J, GIESEN R & LARRAÍN H. 2010. Design of limited-stop services for an urban bus corridor with capacity constraints. *Transportation Research Part B: Methodological*, **44**(10): 1186–1201.

[32] LEVINSON HS, ZIMMERMAN S, CLINGER J & RUTHERFORD CS. 2002. Bus rapid transit: An overview. *Journal of Public Transportation*, **5**(2): 1–30.

[33] MINISTERIO DE TRANSPORTE. 2008. Seguimiento y Evaluación del Transporte Urbano, Indicadores. Colombia. Available at: `http://portal.mintransporte.gov.co:8080/transporte_urbano/indicadores.asp`. Last accessed March 1st, 2017.

[34] NGAMCHAI S & LOVELL DJ. 2003. Optimal time transfer in bus transit route network design using a genetic algorithm. *Journal of Transportation Engineering*, **129**(5): 510–521.

[35] NGUYEN S & PALLOTTINO S. 1988. Equilibrium Traffic Assignment for Large Scale Transit Networks. *European Journal of Operational Research*, **37**: 176–186.

[36] OLIVEIRA R & BARBIERI C. 2015. Efficient transit network design and frequencies setting multi-objective optimization by alternating objective genetic algorithm. *Transportation Research Part B: Methodological*, **81**(2): 355–376.

[37] PATTNAIK SB, MOHAN S & TOM VM. 1998. Urban bus transit route network design using genetic algorithm. *Journal of Transportation Engineering*, **124**(4): 368–375.

[38] SAATY T. 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, **1**(1): 83–98.

[39] SCORCIA H. 2010. Design and evaluation of BRT and limited-stop services. Master Thesis. Massachusetts Institute of Technology.

[40] SHIMODAIRA H. 2001. A diversity-control-oriented genetic algorithm (DCGA): Performance in function optimization. Proceedings of the *2001 Congress on Evolutionary Computation*, Seoul, Korea.

[41] SPIESS H & FLORIAN M. 1989. Optimal strategies: A new assignment model for transit networks. *Transportation Research Part B: Methodological*, **23**(2): 83–102.

[42] SZETO WY & JIANG Y. 2014. Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. *Transportation Research Part B: Methodological*, **67**: 235–263.

[43] SZETO WY & WU Y. 2011. A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. *European Journal of Operational Research*, **209**(2): 141–155.

[44] TOM VM & MOHAN S. 2003. Transit route network design using frequency coded genetic algorithm. *Journal of Transportation Engineering*, **129**(2): 186–195.

[45] WALTEROS JL, MEDAGLIA AL & RIAÑO G. 2015. Hybrid Algorithm for Route Design on Bus Rapid Transit Systems. *Transportation Science*, **49**(1): 66–84.

[46] WU JH & FLORIAN M. 1993. A Simplicial Decomposition Method for the Transit Equilibrium Assignment Problem. *Annals of Operations Research*, **44**: 245–260.

[47] WU JH, FLORIAN M & MARCOTTE P. 1994. Transit Equilibrium Assignment: A Model and Solution Algorithms. *Transportation Science*, **28**(3): 193–203.

[48] YEPES T, JUNCA JC & AGUILAR J. 2013. La integración de los sistemas de transporte urbano en Colombia, una reforma en transición. FEDESARROLLO, Centro de Investigación Económica y Social. Available at: `http://www.fedesarrollo.org.co/wp-content/uploads/2011/08/La-integraci%C3%B3n-de-los-sistemas-de-transporte-urbano-en-Colombia-Findeter.pdf`. Last accessed March 1st, 2017.

[49] ZHAO F & ZENG X. 2007. Optimization of user and operator cost for large scale transit networks. *Journal of Transportation Engineering*, **133**(4): 240–251.