Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Part A 42 (2008) 951–972

TRANSPORTATION RESEARCH PART A

www.elsevier.com/locate/tra

Demand responsive urban public transport system design: Methodology and application

J. Enrique Fernández L. [a,*], Joaquín de Cea Ch. [a], R. Henry Malbran [b,1]

[a] *Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile, Casilla 306, Santiago 22, Chile*
[b] *Secretaría Interministerial de Planificación de Transporte de Chile (SECTRA), Chile*

Received 26 August 2005; accepted 31 December 2007

Abstract

In this paper, we present a methodology for solving the Public Transport Network Design Problem (PTNDP) and describe its application in the context of the Design Study developed in order to propose a new structure for the transit system of the city of Santiago, Chile. Firstly, we briefly define the PTNDP as a multilevel programming problem and discuss the solution method implemented. Then, the application of this methodology to the Santiago transit system is presented, and the main results obtained are analyzed. The new restructured system, based on a hierarchy of specialized services that complement and coordinate their operations and using an integrated fare scheme, is compared with an optimized version (optimal frequencies) of the current one, a set of direct services, mainly based on the operation of independent itineraries, without fare integration. The most important conclusions are the following: (a) the private operating costs and the social costs of the restructured system, using higher standard buses, are considerably lower than the costs of the current system; (b) these cost reductions allow government authorities to introduce an important number of modernizing measures without subsidies and fare increases.

Keywords: Network design; Transit assignment; Transport network equilibrium

1. Introduction

Santiago de Chile, as does several other developing world cities, has an atomized system of public transport services. The largest part is served by long bus itineraries crossing most of the city operating independently. There is also a Metro system of 40 km[2] organized in four corridors. No explicit integration exists between the public transport modes with the exception of a few bus lines that operate as feeders to the end of line Metro

* Corresponding author. Tel.: +56 686 4270; fax: +56 2 686 4818.
E-mail addresses: jef@ing.puc.cl (J. Enrique Fernández L.), hmalbran@sectra.cl (R.H. Malbran).
[1] Tel.: +56 2 671 09 35.
[2] Right now there are 40 km (additional) under construction that will go into operation together with the new integrated public transport system at the end of 2005.

0965-8564/$ - see front matter
doi:10.1016/j.tra.2007.12.008

stations of each of the four corridors. The bus system is in the hands of a large number of small private operators organized in cooperatives that operate 370 individual bus lines, with a total of 8000 buses. Bus lines service concessions are periodically assigned to private operators by tendering. Fares defined by the tendering process are individually charged by each service (metro or bus). Customer using two or more services must pay the corresponding fare to each operator.

The system that serves close to five million daily trips (4.2 million bus trips and 0.8 millions metro trips) is highly inefficient producing low quality services at relative high cost per trip produced; it generates a high rate of accidents and contributes significantly to the congestion of the urban road network and to the high level of air pollution experienced in the city. This situation caused the government to make the decision of restructuring the public transport system, based on the principle of an integrated system of Metro and buses with complementary services and an integrated fare system.

The Public Transport Network Design Problem (PTNDP) can be formulated as a bi-level programming problem containing: (i) the topological design of the network of public transport itineraries, as first level and (ii) the operational design, corresponding to the determination of optimum frequencies and capacities for the itineraries defined, as second level. A methodology is presented and applied to solve this problem for the urban public transport system of Santiago de Chile.

The solution obtained is an integrated system composed of three main categories of services: (i) the metro operating as the core of the system, (ii) a set of corridors operated by high capacity buses, complementing the metro corridors, called trunk services, (iii) several sets of feeder services operated by lower capacity buses and serving the influence area of some main corridor. The methodology applied allowed the definition of the itineraries of the services in each of the last two categories, the frequencies of each defined itinerary and the capacity of the vehicles used in each case.

The paper is organized as follows. Section 2 includes a description of the problem. Section 3 presents the solution methods and algorithms implemented. Section 4 describes the application of the methodology to the Santiago de Chile urban transport system, the main results and their analysis. Finally, Section 5 presents the main conclusions of the Design Study.

2. The Transportation Network Design Problem

The Transportation Network Design Problem (TNDP) can be formulated as a mathematical programming problem (Friesz and Shah, 2001) to find the values of a set of variables, that define the characteristics of the transportation services considered in a network structure and that are used by rational trip makers that maximize their individual utilities. The objective of the problem corresponds, in general, to the maximization of a measure of the social benefit of the system's operation, subject to a set of network, demand and behavioral constraints. The behavioral constraints consider the system user's decisions, with respect to the use or consumption of the set of transportation services offered. They correspond to complex nonlinear relations that must in general be formulated as an independent mathematical programming problem; in some cases this can take the form of an optimization problem, or a system of nonlinear equations (Friesz and Shah, 2001).

There are several possible applications of the general TNDP. The best known and most studied in the literature is the Network Infrastructure Design Problem, where the main variables are the capacities and design characteristics of a given set of streets or roads in a road network and the users are car drivers that use it to travel between geographical locations (O–D pairs). The formulation that is dealt with in this paper corresponds to the Public Transport Network Design Problem (PTNDP), where the variables correspond to the itineraries, frequencies and capacities of a set of public transport services with network structure. In our case the services are provided by urban buses and a metro system.

The mathematical representation of the TNDP can be made by a bi-level (or multilevel) Programming Problem. Because TNDP has a social optimization objective it takes in general the form of a non-cooperative game of the Stackelberg type (1952). On the first, or higher level of the bi-level problem, is the society, represented by the transportation system authority, which wants to find the socially optimum set of transportation services. The second or lower level of this problem represents the system user's behavior that use the services provided, such that their own private utilities are maximized. Thus, specific solutions (decision variables values) are chosen in the first level of the problem (playing the leader's role), assuming some predefined values of the users demands and then, in the second level, users (playing as followers) decisions are simulated for the current values of the transportation service variables. A solution of the problem should provide consistent outputs for both levels in the Stackelberg sense: the decisions taken by the users in order to maximize their individual utilities (second level) should correspond to those assumed by the leader (first level) when choosing the optimum values for the transportation services.

The two levels of the TNDP defined above present especially complex formulations in the PTNDP. The first level, is composed by two sub-levels: (i) the upper sub-level, that finds the optimum service itineraries (physical design), and (ii) the lower sub-level finds the operational characteristics of the service itineraries (operational design). The second level is also especially complex because it must solve a multimodal (bus and metro) transit assignment problem with capacity constraints (De Cea and Fernández, 1993). Given this,

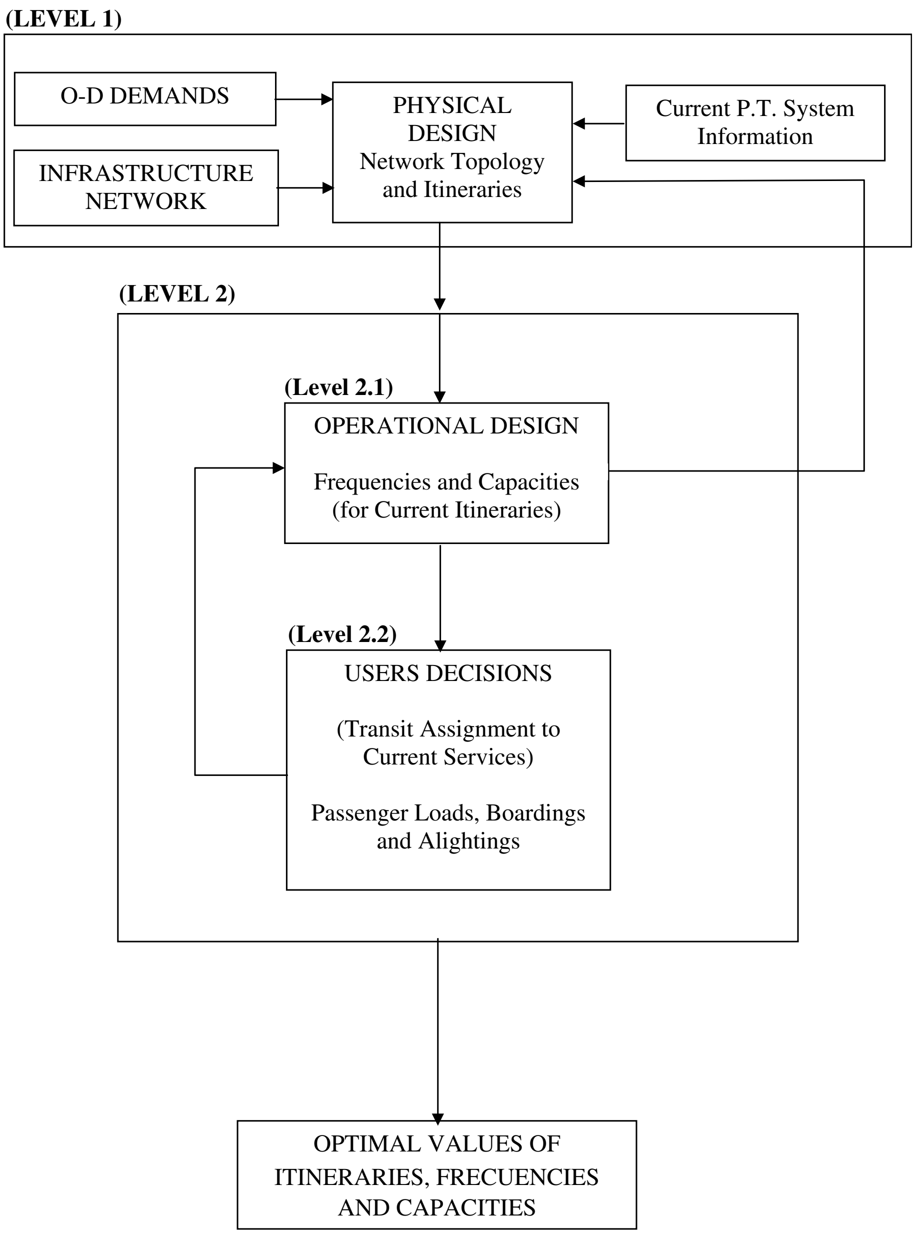

Fig. 1. Public Transport Network Design Problem. Multilevel programming structure.

it is more convenient to redefine the bi-level problem in the following way: in the first level is the transport system physical design problem and in the second level is the operational design problem together with the users behavior determination. The physical design sub-problem in the first level is a difficult discrete combinatorial problem. The second level has itself the structure of a bi-level problem, with the operational design problem on top and a transit assignment problem at the bottom (see Fig. 1).

3. Solution method

The problem represented in Fig. 1 is highly complex and has no close analytical solution. The only possible solution approach is therefore heuristic. In this section, we describe the solution method that was applied to the PTNDP in the case of Santiago de Chile.

Considering the formulation described in the previous section and shown in Fig. 1, the problem is separated in two main parts:

(i) Using the information from the operation of the current public transport system and the network infrastructure, several network topologies for the public transport services are formulated. These are refined using a set of heuristics that are explained below.
(ii) For each topological solution a bi-level problem is solved to calculate frequencies and capacities for each of the itineraries considered.
(iii) Feedback to step (i) until a stable (optimum) solution is obtained for itineraries, frequencies and capacities.

3.1. Network topology and itineraries: physical design

The physical design part of the PTNDP has received rather little attention in the specialized transportation literature. Newell (1979) studied some basic public transport services network characteristics. More recently, some heuristics that allow obtaining reasonable feasible solutions were proposed by Ceder and Wilson (1986), Van Nes et al. (1988), Baaj and Mahmassani (1995) and Lee (1998). In general, the models used are too simple to have practical application or use an analytical approach without integration of practical experience in the operation of public transport systems.

Given that the physical design of a network of public transport itineraries is a combinatorial problem of high complexity, it is practically impossible to analyze the whole solution space. It is necessary therefore to develop some general process to reduce the size of the set of solutions to evaluate. This can be done considering that many feasible solutions do not make much sense in practice and that network structures being analyzed can be considerably reduced by using some general principles based on the knowledge of the technical principles of operation of the system considered and the experience obtained in the operation of real systems in different parts of the world. Based on the observation of urban public transport systems used over the planet, we postulate that there are two types of basic solutions to the topology of an urban public transport system:

(i) Direct services, based on the operation of independent itineraries, of the same basic characteristics, serving directly different city O–D pairs, without explicit operational integration with other services. Itineraries are very long and generally cross most of the city. The main advantage of these systems is that direct services are provided to trip makers, in general, without the need to transfer between different itineraries. The weakness is that the interaction between demand and supply is very inefficient in terms of the use of capacity provided and service production resources; production costs are high, service quality is low and externalities produced are significant. This is the current system provided in Santiago de Chile and most cities of the developing world.
(ii) Integrated systems, based on a hierarchy of specialized services that complement and coordinate their operations. Itineraries are shorter than in case (i) and vehicles are of different characteristics, depending on the type of trip flow levels that the itinerary must serve. Specialization of complementary services allows a lower cost of production and higher quality services and a better control of externalities. This

is derived from the use of economies of scale and scope and the specialization of services to better fit the demand. The only disadvantage with respect to the other system is that transfers are, in general, necessary. This is the type of system that the government wanted to implement in Santiago.

An important inefficiency of the present system is that buses compete with the Metro, because bus lines run parallel to metro lines with no significant integration.[3] Therefore a lot of resources invested in the metro system are sub utilized and additional resources are wasted operating parallel bus services. This socially inefficient operation is incentived by the use of disintegrated fares. If a user wants to alight from a bus and take a faster parallel metro line he must pay a new fare for the transfer. The new integrated system will operate with integrated fares.

The metro system that will count with four lines and a total length of 80 km by the end of 2005, will be the backbone of the new integrated system. A set of main segregated bus corridors will complement this and additional feeder services will complete the system. Even though the second system had been chosen there were several design aspects that needed to be decided:

- The number of main corridors to include in the system, their itineraries and frequencies.
- The number and grouping of the feeder services, their itineraries and frequencies.
- The number, itinerary and frequencies of special services that use more than one corridor to offer a direct service between the most demanded O–D pairs (these special services reduce the number of transfers required in the integrated system).
- The creation of special express services, without intermediate stops to increase commercial speeds during the peak operation.

The methodology used to find the best possible network topology is based in the application of three different types of heuristics that solve the following parts of the problem (Fernández et al., 2003):

- Type 1. Generation of the set of main corridors of buses and the services to operate them.
- Type 2. Generation of feeder services to the main corridors.
- Type 3. Two complementary heuristics for the adjustment and improvement of heuristics type 1 and 2. One of them (fractioning and reduction heuristic) analyzes which is the optimum length of each itinerary and the possibility of fractioning itineraries to obtain a better fit of service capacity to passenger's loads. The adjustments are evaluated on the basis of a simplified calculation of the social costs involved. The second, heuristic eliminates redundant itineraries that do not provide additional accessibility.

It is important to notice that the feedback obtained from the operational design level of the problem allows the revision of the topological solution obtained. Thus, some itineraries included in the topological design can obtain a zero (or lower than a pre-specified value) frequency (or capacity) from the operational design. Those services will be candidates to be eliminated or consolidated (if possible) with other similar services. Given this possibility it is better to include all itineraries that appear to be potentially attractive, that eliminate them in advance.

3.2. Determination of optimum frequencies and capacities: operational design

This level of the problem determines which are the optimum frequencies and vehicle capacities for a given set of itineraries. It is obvious that both variables are interdependent because as the capacity of the vehicles increases, the frequency necessary to carry a given flow of trips decreases. However, because the objective function considers users waiting costs, the relation will not be deterministic and it is possible to obtain solutions in which some of the capacity provided by the vehicles is not used. This can happen when a reduction of

[3] In some parts, a reduced number of lines are used as feeders to the metro. However the service fare is only a little lower (10%) than the sum of both services.

the frequency, in order to increase the vehicles usage, produces an increase in waiting costs that is higher than the corresponding reduction in vehicle operating costs. In such case it is obvious that an adjustment (reduction) in vehicles capacity can improve the solution. Therefore, a limited number of vehicle capacities are examined (two or three per type of service) and capacity adjustments are made if necessary.

The objective of this bi-level Stackelberg problem (see Fig. 1) is the minimization of the total social costs of the urban transport system.[4] Total social cost is calculated considering the following components:

(i) Production costs of the services that are considered directly proportional to the vehicles–km operated by each service per unit of time during the period analyzed (peak or off-peak).
(ii) Travel times perceived by the system users; these are separated in: access times, from the origin of the trip to the closest transit stop; waiting time at the transit stop, that depends on the effective frequencies (De Cea and Fernández, 1993) of the services;[5] transfer times when the trip maker has to use more than one itinerary and in-vehicle travel times. All times considered are evaluated economically using the corresponding value of time, externally provided to the model.
(iii) Congestion externalities are also considered by taking into account cars and buses travel times. This is done by using a multimodal (private–public transport) assignment model to evaluate the solutions (De Cea et al., 2005). In such model, when cars and buses use the same streets congestion interactions are considered: scheduled buses reduce street capacity for cars operation and cars flows reduce buses commercial speed; notice that if average commercial speeds of vehicles are reduced, the fleet size will have to be increased to maintain the same frequencies, which directly affects operating costs.

The road network is represented by a graph $G = (N, A)$, where N is the set of nodes and A the set of arcs, representing city streets. N represents centroids (locations of trips generations) and streets intersections. G provides the infrastructure for the operation of public transport vehicles (buses). The public transport network is represented by a graph $\overline{G} = (\overline{N}, L)$, where the set of nodes $\overline{N}$ is a subset of N, representing transit stops and L is the set of transit lines (buses and metro) that provide the public transport services.

The assignment model used for transit trips uses an auxiliary network $G' = (\overline{N}, S)$, where S is the set of route sections and transit access links. A route section represents the set of all the transit "attractive" lines running between two transfer nodes (De Cea and Fernández, 1993). Let us consider the following notation:

$\overline{A}$	subset of arcs in A used by transit lines, $\overline{A} \supset A$
A'	subset of arcs in A not used by transit lines, $A' \supset A$
l	transit line index
L_a	set of transit lines running over arc a
S^M	set of metro route sections
S^B	set of bus route sections
$A_{\text{acc}}^{\text{tpub}}$	subset of transit access links contained in S
G_a^l	line l operating costs on arc a, per unit of time
c_a'	individual private cars operating cost on arc a
c_a	individual private cars travel time on arc a
c_s	individual transit users travel cost on route section s
$c_{\text{acc}}^{\text{tpub}}$	transit user's access time on access arc, acc
f_a	total private cars flow on arc a, per unit of time
$\bar{f}_a$	total transit users flow on arc a, per unit of time
$\bar{f}_a^l$	transit users flow traveling by line l on arc a
V_{acc}	transit users flow on access arc, acc
V_s	transit users flow traveling by route section s
d_l	service frequency of line l

[4] This is equivalent to the maximization of the social welfare in this case because the demand (O–D matrices) is considered fixed.

[5] Effective frequencies consider that if the system is congested users will have to wait longer at stops because some buses will pass full by the stop.

d_a total number of buses that use arc a, per unit of time
d_s total transit vehicles frequency on route section s

The operational design can be formulated as an integer-programming problem; however important algorithmic and computational advantages can be obtained by representing frequencies and capacities as continuous variables. Therefore most of the models presented in the literature have adopted this approach (Abdulaal and LeBlanc, 1979; Marcotte, 1983; Friesz et al., 1992, 1993; Friesz and Shah, 2001; Meng et al., 2001).

The decision variables for the continuous operational design problem are the transit line frequencies $\{d_l\}$, the private cars flows $\{f_a^*\}$ and the transit passenger flows $\{V_s^*\}$. The operational design problem can be formulated as

$$\begin{aligned}\operatorname*{Min}_{\{d\}} \quad & \sum_{a\in\bar{A}}(\theta_1 c_a(f_a^*, d_a) + c_a'(f_a^*, d_a))\cdot f_a^* + \sum_{a\in A'}(\theta_1 c_a(f_a^*) + c_a'(f_a^*))\cdot f_a^* \\ & + \theta_2^B \sum_{s\in S^B} c_s(V^*, d_s)\cdot V_s^* + \theta_2^M \sum_{s\in S^M} c_s(V^*, d_s)\cdot V + \theta_3 \sum_{a\in A_{\mathrm{acc}}^{\mathrm{tpub}}} c_{\mathrm{acc}}^{\mathrm{tpub}}\cdot V_{\mathrm{acc}} \\ & + \left(\sum_{a\in\bar{A}}\sum_{l\in L_a} G_a^l(f_a^*, \bar{f}_a^{l*}, d_a, d_a^l)\right) \end{aligned} \tag{1}$$

$$\text{s.t.} \quad d_l \geqslant 0, \quad \forall l\in L \quad \{f_a^*\}, \{V_s^*\} \quad \text{equilibrium flows} \tag{2}$$

where θ_1 and θ_2 are in vehicle travel time values for car and transit users and θ_3 is the access time value for transit.

The main constraints considered by the *equilibrium flows* are: (i) the topology of the network (the set of itineraries defined in the first level), (ii) the trip demands expressed by the O–D trip matrices, and (iii) the user's behavior implicit in the assignment model used.

Three different algorithms were implemented and evaluated in this work for the solution of the operational design problem: (i) Hooke and Jeeves algorithm (Abdulaal and LeBlanc, 1979); (ii) simulated annealing (Friesz et al., 1992, 1993), and (iii) augmented Lagrangean (Meng et al., 2001). The method finally used to obtain the solutions reported in the next sections was Hooke–Jeeves, because in the comparison tests carried out, this method resulted to be the most robust and efficient.

Hooke and Jeeves algorithm (H–J) is a direct search method in the decision variables space. It is based on the exploratory search of improving directions starting from any point of the feasible space. In order to define an improving direction, several initial searches are performed through the space coordinates, on the basis of which an improvement direction vector is defined. Then, an initial advance step is chosen and varied until an improved solution is found. It is a very robust method, because it does not require any good or regularity property of the solution space. It is relatively easy to implement and calibrate. Like other mentioned heuristics applied to the network design problem, the main limitation is that only local optima is assured, therefore the application must be repeated, starting from different initial solutions in order to increase the probability of obtaining the global optimum or a close solution. The authors have worked before applying this method to network design problems and network calibration, with good results (De Cea and Fernández, 1995).

3.3. Equilibrium flows

To obtain the equilibrium flows $\{f_a^*\}, \{V_s^*\}$ required to solve the operational design problem, it is necessary to use a suitable assignment model. This model will provide the trip flows and levels of service (travel times with each of its components enumerated above in (ii)) for each set of itineraries, frequencies and capacities considered. Such flows and levels of service are necessary to evaluate the objective function (1) for any solution analyzed.

It is crucial in this case that the assignment model used considers capacity constraints of the transit services that are optimized. Otherwise, the effect of congestion on social costs will not be correctly considered. The model used was ESTRAUS (De Cea et al., 2005) that has a multimodal assignment module that considers congestion interactions between public and private transport vehicles and a transit assignment with capacity constraints. For given values of itineraries, frequencies and capacities, the model considers that increases in transit O–D flows will produce congestion at transit stops causing an increase in the users waiting time (De Cea and Fernández, 1993). Therefore, the costs of transit congestion are internalized creating in the design model (see Fig. 1) internal incentives to increase frequencies or vehicle capacities. The incentive will be stronger when the relative value of waiting time is high compared to the additional operating cost produced by the increase of frequencies or vehicles size. Normal iterations of the operational design model (level 2.2 in Fig. 1) were performed using only the multimodal transit assignment module of ESTRAUS, over a multimodal public transport network, considering metro, bus and share taxi services. Once a solution was obtained for itineraries, frequencies and vehicles capacities, the values of flows and levels of service (times) where validated running the supply-demand equilibrium version of ESTRAUS.[6] If flows and levels of service values obtained from this validation showed significant differences with those corresponding to the optimization solution the whole process was reiterated.

4. Application to the Santiago system

The methodology described in the previous sections has been used in the context of a Santiago Transit Design Study made for the Chilean government. The main objective of this study was to propose a new integrated design for the Santiago Transit System. The alternatives considered should transform the current transit system, based on the operation of a system of independent itineraries, with an extremely low level of physical and fare integration, into a wholly integrated system, based on a hierarchy of specialized services that complement and coordinate their operations using an integrated fare scheme.[7]

To give a reasonable synthesis of the work done, in this section we will address four aspects: (a) transit network calibration; (b) generation of trunk and feeder services itineraries; (c) determination of optimum frequencies and vehicle capacities and (d) comparison of operational indexes for an "optimized"[8] version of the current system vs. the chosen design solution that will be implemented starting August 2006.[9]

4.1. Transit network calibration

The transit network considered in this work contains 370 bus services,[10] three metro lines and one suburban train line connected to the metro system.

All the information concerning public transport trips and networks, required to implement the design methodology was obtained from existing historical data, complemented by several especial surveys performed during 2001, "calibration year". The information concerning operating costs for both current and future bus operators was obtained from a especial Cost Study (SECTRA, 2003b).[11]

The following surveys were specially performed to complement the historical data available for the design study:

(i) Household origin–destination survey: EOD-2001.[12] A total of 15,000 households were surveyed in the city of Santiago which has a population of 5,772,600 inhabitants and 1,513,900 households.
(ii) On board surveys: Buses-2001[13] and shared-taxis 2001.[14] These surveys were conducted during October and November 2001 and completed in May 2002. In both cases the sample size corresponded to 16% of the travelers of a normal working day, between 6:00 a.m. and 11:00 a.m. The main data collected were:

[6] To obtain a distribution, modal split and assignment simultaneous equilibrium.
[7] The Transit Design study was performed by the Consultant Company Fernández & De Cea Ingenieros Ltd.
[8] Current services frequencies were optimized considering the same objective function (1).
[9] The new system was already concessioned to 15 private companies that will start a transient operation between August 2005 and August 2006, when the new integrated system must start operations. For assigning concessions an international public tendering process was performed in January 2005.
[10] For the optimization process this were represented as 740 one way services.
[11] The "Costs Study" was performed by Fernández & De Cea Ingenieros Ltd.
[12] The survey was carried out by the Catholic University of Chile, for SECTRA.
[13] This part of the Design Study was developed by the company CIS Asociados Consultores en Transporte S.A.
[14] This survey was carried out by Fernández & De Cea Ingenieros Ltd. for the Ministry of Transport.

the origin and destination of trips, origin stop and destination stop of in vehicle trip stages, trip times, and trip makers household income.[15] Additionally, some characteristics of the bus and shared-taxi services (peak and off-peak frequencies, travel times, itineraries) were obtained in order to update the transit network model.

(iii) Origin–destination survey on metro stations: Metro-2001 (conducted by Metro S.A.). This survey considered 15% of the travelers entering a metro station, between 6:30 a.m. and 22:00 a.m. The trip data collected are similar to the one obtained in the surveys described above. In addition, information about car-metro trips has been obtained.

(iv) Bus travel times were measured in 30% of the road network links, including about 80% of the links in the trunk network. For the remaining links of the network, they were classified and typified and a commercial speed was assigned to each of them, for each modeling period.

Combining the trip data obtained from the bus, metro and shared-taxi surveys, the peak and off-peak hourly trip matrices were estimated. The calibration of the network consisted mainly in the determination of the parameters of the generalized cost functions used in the transit assignment model. This was performed using an adaptation of the Hooke and Jeeves algorithm (see Abdulaal and LeBlanc, 1979; SECTRA, 2003a). The objective of the process was to reproduce the flows and travel times observed. Two types of generalized link costs functions (route sections) corresponding to bus lines and metro lines were calibrated with the following functional form:

$$c_s^m = tv^s + p_{\text{wait}} \cdot \left[\frac{\alpha^m}{d_s} + \beta^m \cdot \left(\frac{V_s + \widetilde{V}_s}{K_s} \right)^{n^m} \right] + p_{\text{walk}} \cdot tc^s + \frac{1}{v_{\text{time}}^m} \cdot \text{fare}^s \tag{3}$$

where c_s^m is expressed in time units and m can be equal to bus or metro. tv, tc are the in vehicle travel time and the walking access time; p_{wait}, p_{walk} are the relative weights of waiting and walking access time values, with respect to the in vehicle travel time value; v_{time}^m is the travelers value of time that depends on the mode used.

Within brackets in (3) is the waiting time function. This represents the waiting process that passengers experience at transit stops served by multiple lines. The first term represents the waiting time without congestion, when passengers can always get on the first bus that arrives to the transit stop. This depends on the total frequency, d_s, of the lines serving route section s, and parameter α^m that depends on the buses and passengers arriving processes. The second term represents the waiting time experienced by passengers when the system is congested (De Cea and Fernández, 1993); it depends on: (i) the flow of passengers that demand the service at the transit stop where route section s begins, V_s, (ii) the flow of passengers that already come in the buses that arrive to the transit stop, $\widetilde{V}_s$ (called competing flow), and (iii) the total capacity provided by all lines serving route section s, K_s. α^m, β^m and n^m are the waiting function calibration parameters.

The calibration proceeded as follows: first the values of the waiting function parameters where obtained by micro-simulation. Then, p_{wait}, p_{walk} and v_{time}^m values where calculated by Hook and Jeeves.

The values obtained for the peak period are the following:

– $\alpha^{\text{bus}} = 1$, $\alpha^{\text{metro}} = 0.5$; $\beta^{\text{bus}} = 4.75$, $\beta^{\text{metro}} = 1.50$; $n^{\text{bus}} = 5.81$, $n^{\text{metro}} = 6.0$;
– $p_{\text{wait}} = 1.93$, $p_{\text{walk}} = 3.63$;
– $v_{\text{time}}^{\text{bus}} = 3.62$ (Ch \$/min), $v_{\text{time}}^{\text{metro}} = 7.45$ (Ch\$/min).

In Table 1, travel time values obtained from the network calibration are compared with average incomes corresponding to Metro and Bus users.

[15] Income information was used to obtain average values and for validation and consistency analysis, because only one type of user was considered in the assignment model.

Table 1
Income vs. users value of time (calibration year 2001)

Mode	Average income (Ch$)[*]	Value of time (Ch $/min.)
Bus users	299,000[a]	3.62
Metro users	569,313[b]	7.45

[a] Calculated using income data from the 2001 EOD home survey.
[b] Calculated using income data obtained from the origin–destination survey on metro stations. ([*]) 1US$=590 Ch$.

Table 2
Peak period transfers and trip stages (peak hour, calibration year 2001)

Period from 7:30 to 8:30 bus + metro networks	Observed trips[a]	Modeled trips
Total number of trips	524,674	524,674
Total number of trip stages	635,576	618,968
Total number of transfers	110,902	94,294
Average number of transfers per trip	0.21	0.18

[a] Calculated from the O–D surveys data.

Bus users in vehicle value of time is a 15% of income and for metro users value of time is a 16.2% of income. Using the value obtained for p_{walk}, access walk time value corresponds to a 55% of income for bus users and a 59% for metro users. Finally using the value obtained for p_{wait}, waiting time values are a 29% and a 31% of income for bus and metro users.

The walking access times were assumed fixed and were obtained from the access links length and walking speeds. The travel time on route section s, tv^s, was obtained from the simultaneous equilibrium of cars and buses, that is performed every time that a new set of bus frequencies is obtained.

To validate the calibrated network, for each modeling period, the transit trip matrix was assigned to the multimodal transit network (metro and bus) using the equilibrium transit assignment model from ESTRAUS. As an example, we show some results corresponding to the morning peak hour.[16] In these results, "modeled" values were obtained by assigning the trip matrix to the calibrated transit network using the assignment model, while "observed" values are those obtained from trip surveys.

Table 2 presents the comparison between total observed and modeled trip stages and transfers. The total number of trips corresponds to the observed O–D matrix used as input data for the transit network assignment to the calibrated network. As can be observed the number of transfers modeled fits reasonably well with those obtained from the O–D survey. It can be noticed that the number of transfers in the current system is relatively low. This is because current services present a very low integration, and travelers must pay a new fare each time that they change from one bus line to other or between buses and metro. Only the transfers within the metro system are free.[17] Therefore, in the current system transfers are made only when there is practically no other alternative to reach a destination. Finally, Table 3 presents some aggregated level of service statistics corresponding to the peak hour, obtained from the calibrated assignment model. They represent the operational characteristics of the current system.

As can be observed from the table, average waiting times are very low in the Santiago transit system, because service frequencies are in general very high.[18]

Figs. 2–4 show a comparison between observed and modeled flows for the three metro lines (lines 1, 2 and 5[19]). Fig. 5 shows the same comparison for the most important transit corridor in the city (Alameda).[20]

[16] Detailed results from the validation process can be obtained upon request to the authors.
[17] However the metro system has only three lines and the number of transfers is also low.
[18] There are 8.000 buses operating in the system and the main corridors have frequencies between 200 and 700 buses per hour.
[19] Because historical reasons coming from the original metro system plan the third line corresponds to line 5.
[20] The same comparison was made to validate the results of the assignment for all the bus lines currently operating in the system.

Table 3
Modeled system levels of service (peak hour, calibration year 2001)

Period from 7:30 to 8:30 bus + metro networks	System total (min)	Average per trip (min)
Waiting time	2,400,986	4.6
Access + transfer time	3,592,860	6.9
In vehicle travel time	15,884,620	30.8
Total travel time	21,878,466	42.4

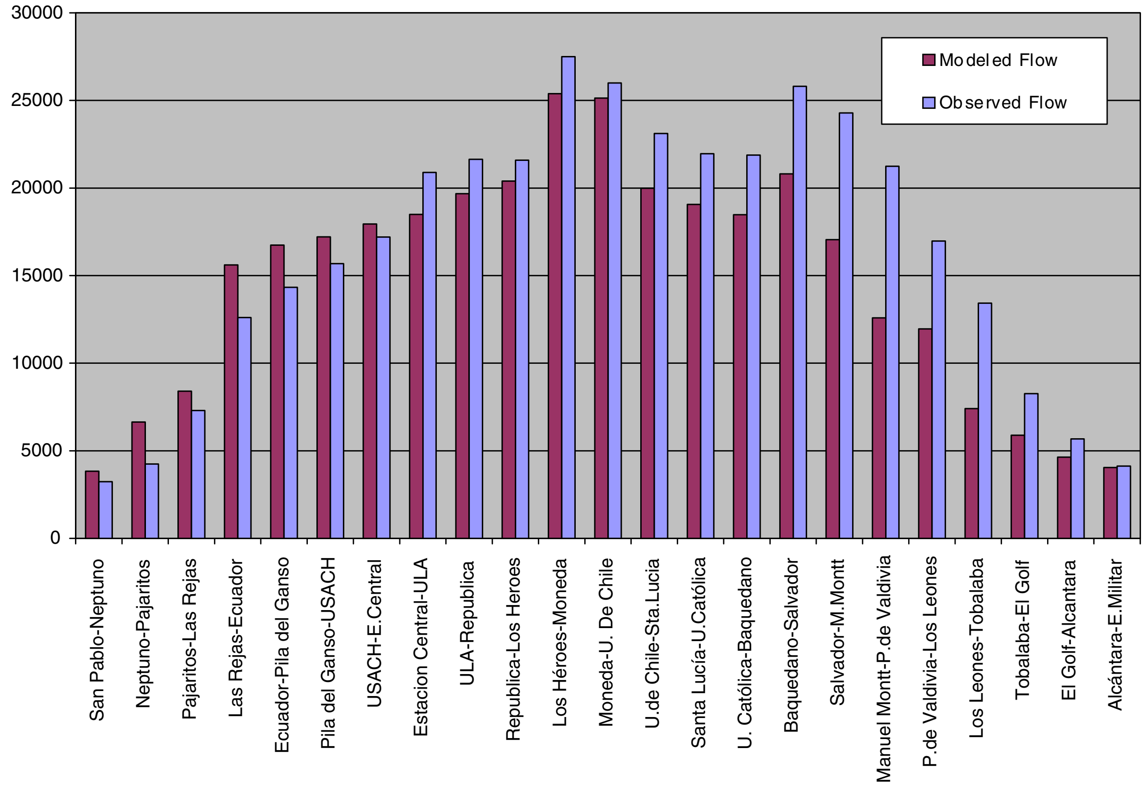

Fig. 2. Metro Line 1 west-east, morning peak hour, year 2001. Modeled vs. observed flows (pas/h).

4.2. Generation of trunk and feeder itineraries and services

The physical design of the transit network was performed using a set of heuristics mentioned in Section 3.1. Applying the approach described in Section 3.1 the following steps were performed:

(i) A potential bus trunk network was initially defined. This will complement the operation of the metro network to constitute the backbone of the new system. A main constraint for this definition is that the new bus trunk corridors will have to use the same existing infrastructure. Although corridors operating characteristics will be improved (new pavement, operational segregation and new transit stops) the streets topology will be basically the same.
   – First, all corridors with flows higher than some pre-specified value (2500; 5000 and 10,000 pas/h) were identified and included in the potential trunk network. Bus and metro services were considered complementary; therefore, trunk bus lines parallel to metro lines were eliminated of consideration.
   – Then, based on continuity and operational factors the network was revised and completed. The network obtained is shown in Fig. 6 together with the metro system that will operate in year 2006, basic simulation year for the new system.

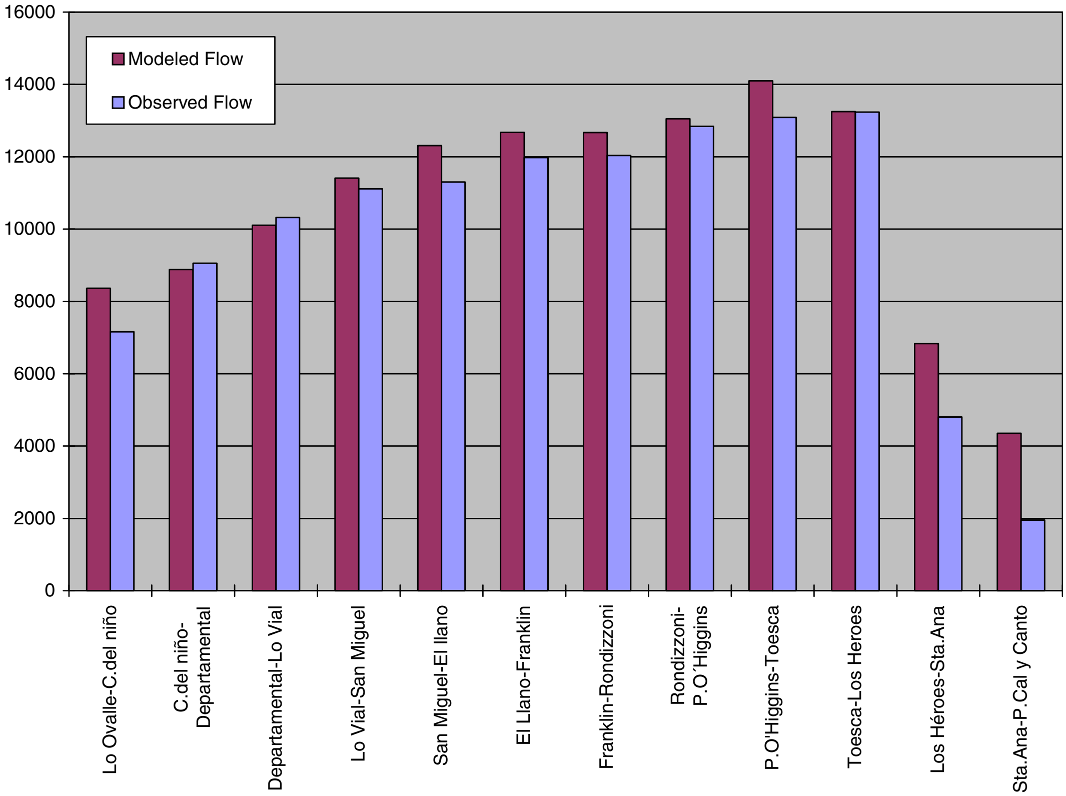

Fig. 3. Metro Line 2 south-north, morning peak hour, year 2001. Modeled vs. observed flows (pas/h).

– Once the trunk network was obtained, a set of itineraries were defined over the trunk corridors. In the first iterations only single corridor itineraries were defined. Afterwards, this design was revised with the objective of reducing unnecessary transfers, including itineraries that use more than one corridor. The final solution, presented in this paper includes 49 trunk services defined over 19 corridors; 19 correspond to itineraries that operate over only one trunk corridor; 20 operate over two corridors 10 over three corridors and 1 over 4 corridors. As can be observed in Table 2. the current system has some 0.21 transfers/trip in the morning peak hour. While this number went up to 1.56 in the initial solutions that included only single corridor itineraries, in the final solution, with multiple corridor itineraries, this figure was reduced to an average of 0.76 transfers/trip.

(ii) The study region was divided in separate and complementary local feeder areas. In each of these areas, secondary networks were defined for the operation of feeder (local) services. The initial definition of these areas was based on administrative considerations. For instance, a local area corresponds to a county or aggregation of several counties that present important local interactions.
– First, local networks within each area were defined including all the streets that presented transit services in the current system. This approach makes use of the knowledge of traditional transit operators with respect to transit demand location and operational conditions.
– Then, feeder (local) itineraries were obtained starting from an initial solution that considered, for each local area, all the portions of existing itineraries that used the secondary network of the area.
– The set of initial feeder itineraries obtained were then connected to the trunk corridors and those sections that run parallel to trunk services were eliminated.
– In successive stages of the analysis this initial solutions were modified by: (a) applying the fractioning and reduction heuristics mentioned in Section 3.1, new itineraries were added to the system; (b) using feedback from the operational design level; itineraries with a frequency lower than an acceptable

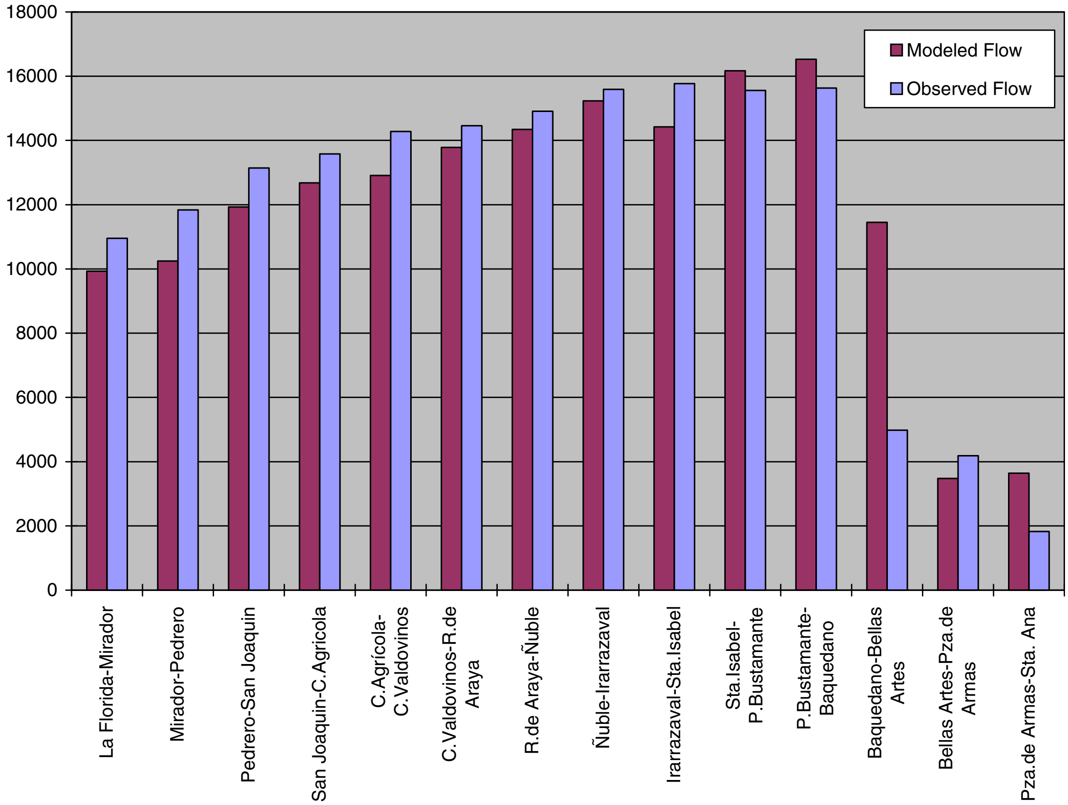

Fig. 4. Metro Line 5 south-north, morning peak hour, year 2001. Modeled vs. observed flows (pas/h).

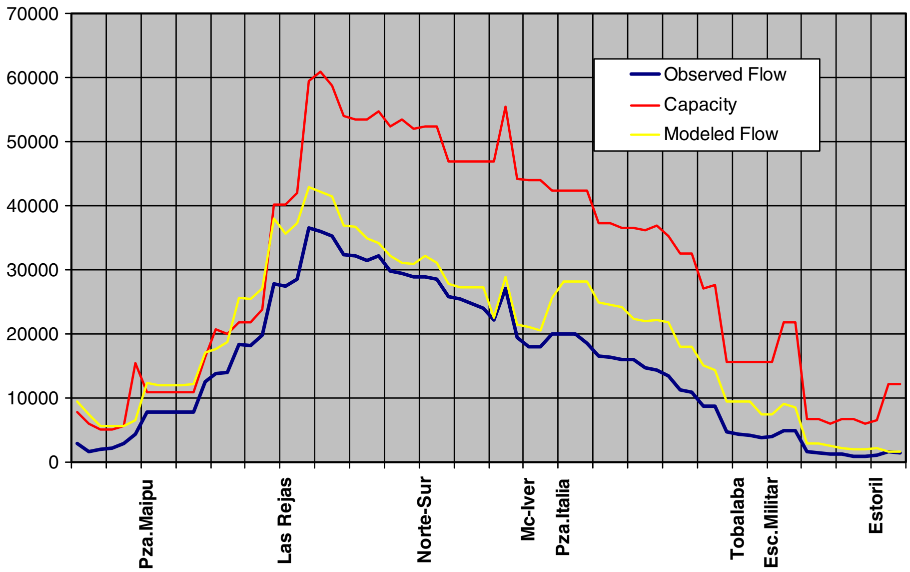

Fig. 5. Alameda west-east, morning peak hour, year 2001. Capacity, Modeled and observed flows (pas/h).

predefined limit were eliminated; (c) using the analysts knowledge of the streets layout and infrastructure, itineraries were marginally adapted to allow a more efficient operation of vehicles.

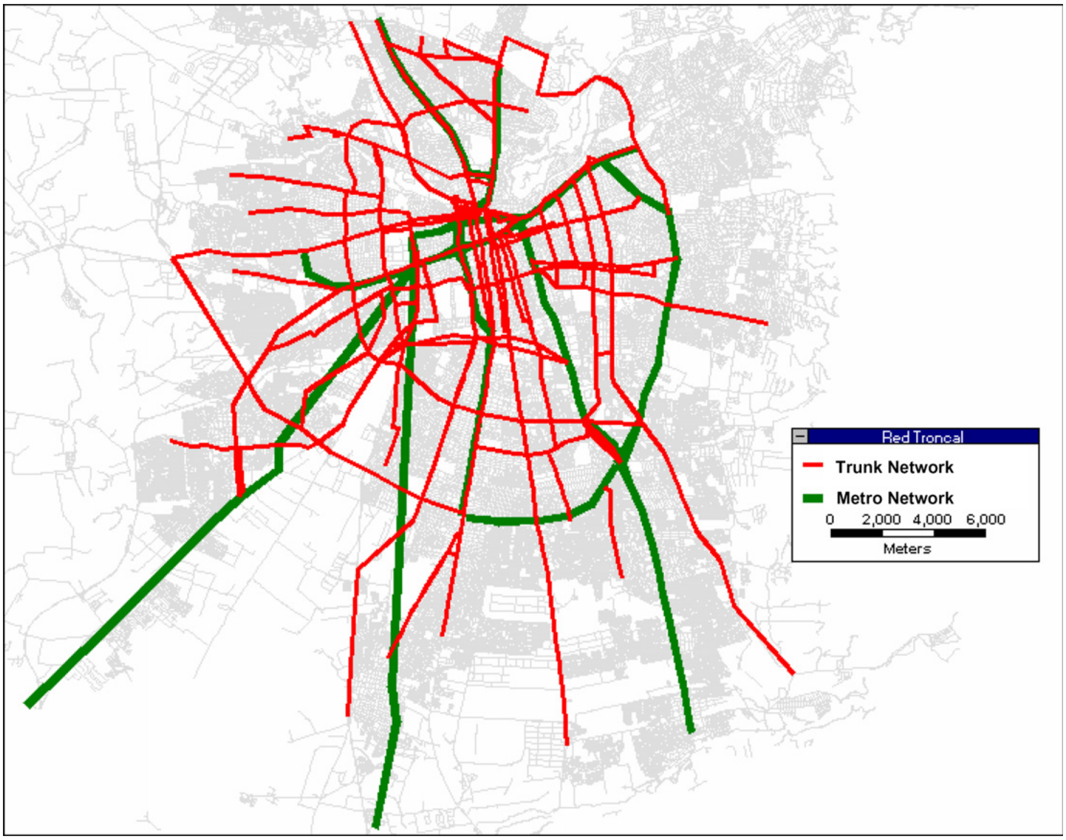

Fig. 6. Trunk network of main transit corridors. Morning peak hour, year 2001.

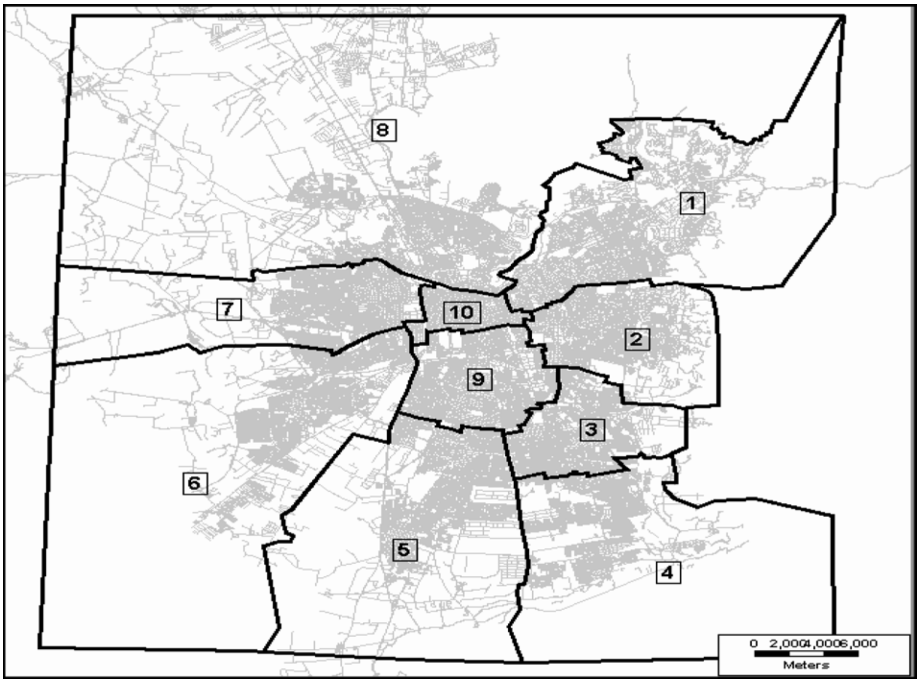

Fig. 7. Feeder (local) services areas.

The final system design considers 10 local feeder areas shown in Fig. 7. Within them a total of 91 feeder (local) itineraries were defined. As an example the feeder itineraries corresponding to local area 1 are shown in Fig. 8.

Frequencies and capacities of the bus services were optimized at the operational design level (see Section 4.3). In the case of the metro services, the capacities and frequencies were exogenously defined taking into account the infrastructure network and fleet of metro cars that will be available for year 2006 (see Fig. 5). For the simulation year, the trip matrices were obtained projecting the 2001 matrices obtained from the O–D survey.

4.3. Determination of optimum frequencies and vehicle capacities

The H-J algorithm was applied to optimize frequencies for the current system of bus lines. The convergence of the algorithm is shown in Figs. 9 and 10 for two different starting solutions: HJ-1 started from a solution in which all services headways were put in 7.5 min and in 2.5 min for HJ-2. The objective function (OF) considered included the following costs: (i) cars operating costs, (ii) cars users travel times values, (iii) transit

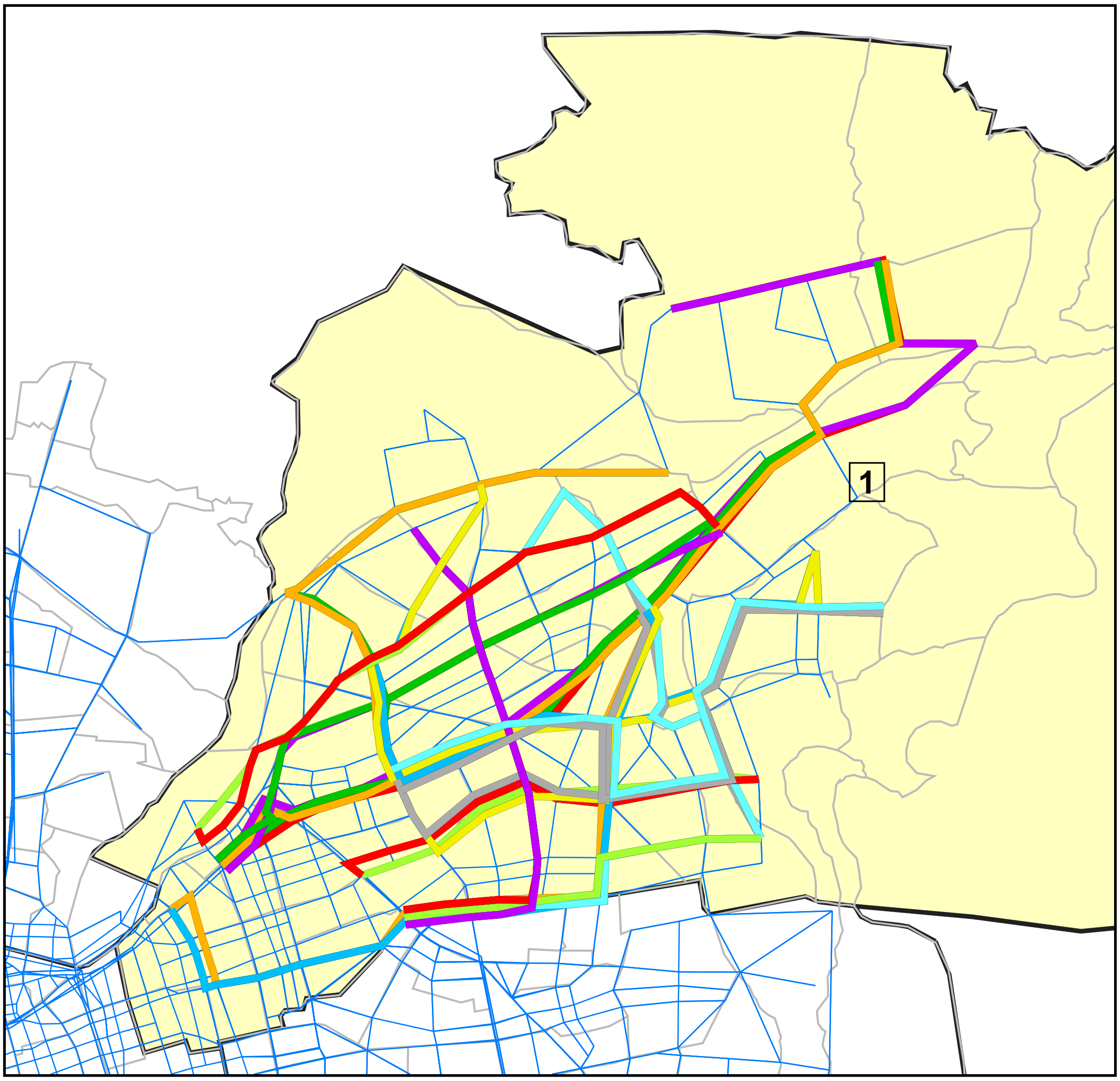

Fig. 8. Feeder (local) services area 1.

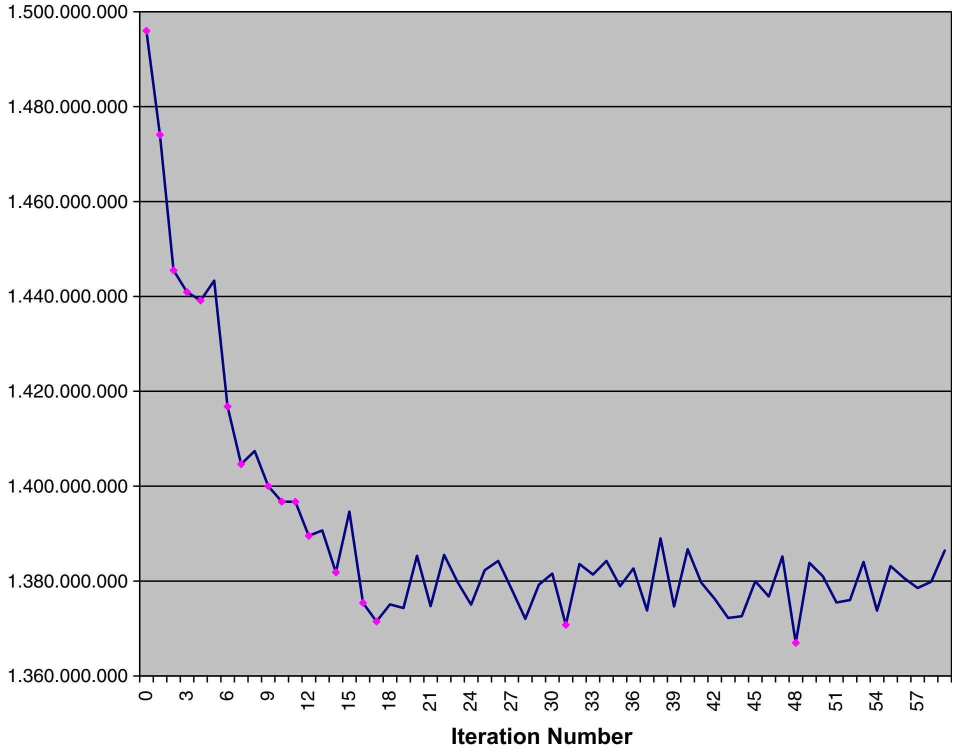

Fig. 9. HJ-1, Convergence of the objective function value (from $h = 7.5$ min).

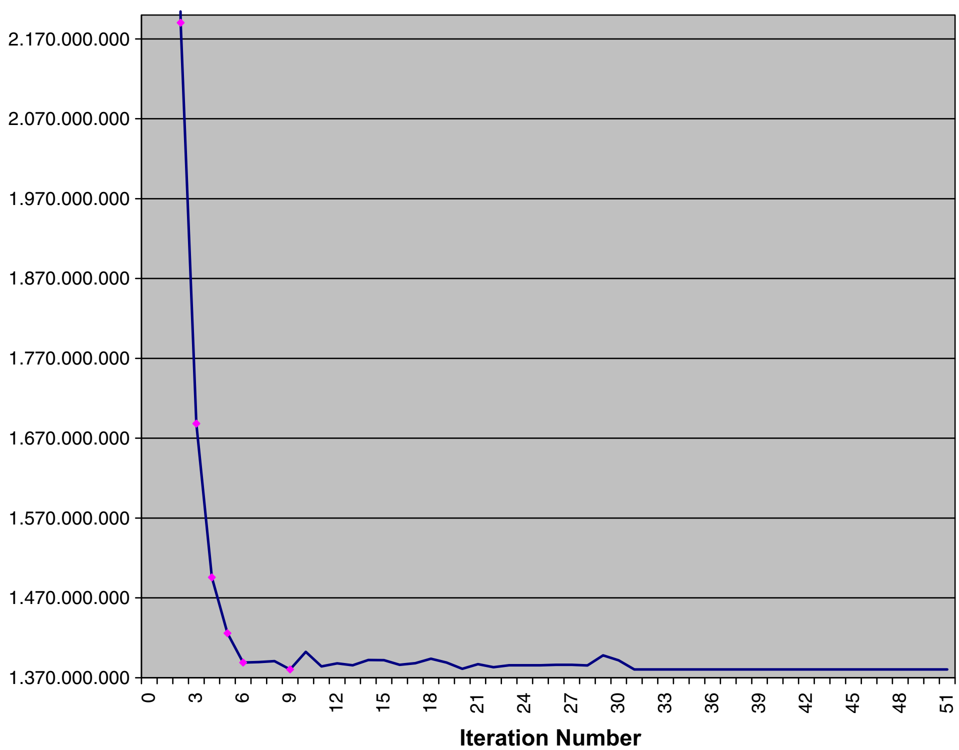

Fig. 10. HJ-2, Convergence of the objective function value (from $h = 2.5$ min).

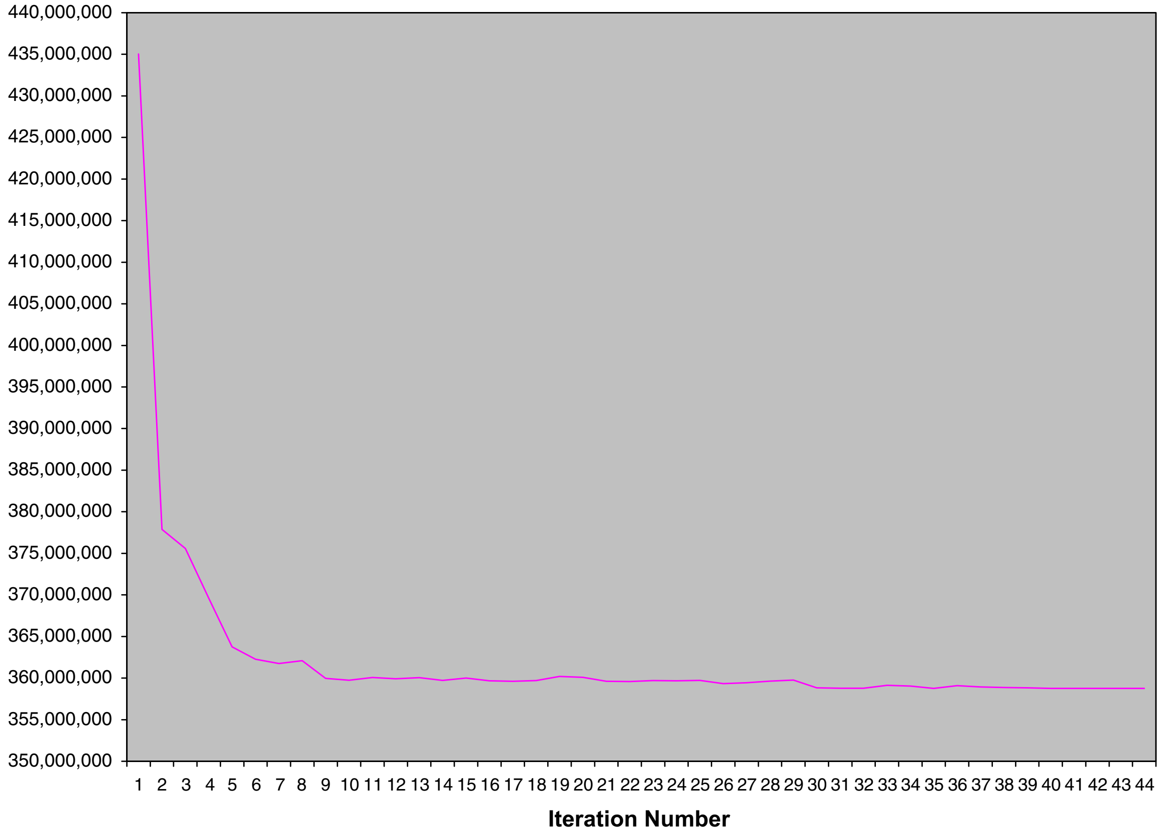

Fig. 11. HJ-IS convergence of the objective function value.

operators operating costs, (iv) transit users travel times values (separated in access, waiting and in vehicle). For HJ-1, the best solution, with an OF value of Ch$ 1,367,000,000, was obtained after 48 iterations and for HJ-2 the optimum solution, with an OF value of Ch$ 1,380,000,000 was obtained after only nine iterations. The solutions are similar in terms of the pattern of frequencies values obtained when these frequencies are aggregated by corridors. However when analyzed line by line the solutions can be significantly different because there is a great number of common lines in many corridors.

The current situation without optimization of frequencies has an OF value of Ch$ 1,419,553,000. Therefore the optimization process saves Ch$52,553,000, for each peak hour.

The OF value of Ch$ 1,367,000,000 obtained in HJ-1 can be decomposed in Ch$ 967,324,896, corresponding to car users operating costs pus travel time values, and Ch$ 399,675,104 corresponding to transit operators costs plus transit users travel time values.

H-J was also applied to find optimal frequencies corresponding to the new integrated services defined in the previous section. The convergence of the algorithm (HJ-IS) can be seen in Fig. 11. In the figure, the value of the OF only considers the costs corresponding to the operation and use of the system of new integrated bus services. The best OF value obtained for the new integrated system was of Ch$ 349,084,358, for the peak hour with a significant reduction of bus operators costs. These cost results are summarized in Table 8 and are analyzed in the next section.

5. Main results

In this section, we present a comparison of the operational and economic characteristics of the "optimized base", corresponding to the current system with optimized frequencies for the 2005 demands vs. the final solution obtained from the application of the design methodology presented in the preceding sections.

Table 4
Supply indexes, off-peak (year 2005)

Index	Optimized base situation	New integrated system
Number of services		
Metro, metro-trains	7	7
Trunk services	–	49
Feeder services	–	91
Current services	370	
Totals	377	147
Kilometers of services[a]		
Metro, metro-trains	222	222
Trunk services	–	1853
Feeder services	–	1812
Current services	20,688	
Totals	20,910	3888
Fleets		
Trunk services		
Buses capacity = 160 pas/bus	–	1212
Buses capacity = 120 pas/bus	–	1058
Buses capacity = 80 pas/bus	–	455
Feeder services		
Buses capacity = 80 pas/bus	–	1153
Buses capacity = 40 pas/bus	–	827
Current services	8405	–
Totals	8405	4704
Equivalent buses (80 pas/bus)	8405	6031

[a] Sum of lengths by type of service.

Fare system: The base system operates with a flat fare of Ch$290[21] for bus services and Ch$330 for metro services; combinations of any kind of service are subject to the payment of a new fare.[22] The new integrated system will operate with integrated but differentiated fares that vary between Ch$290 and Ch$390, depending on the trip type (only local trips; trips using only trunk or/and metro services; trips using feeder and trunk or/and metro services; and services using feeder services before and after the trunk or/and metro services). These fares were calculated considering, in a simplified way, the length of the trips and applying the constraint of maintaining the same mean fare in both scenarios. Two explicit political constraints formulated by the government were that: the new system should not produce fare increases, neither subsidies to the private operators should be necessary.

Tables 4 and 5 present a synthesis of supply indexes for peak and off-peak hours of both compared situations. The main services characteristics shown are the following:

(i) *Optimized base*. The base system has 370 two way bus services, operated by a total bus fleet of 8405 buses during the peak hour and 5.416 in the off-peak (average capacity of 80 pas/bus). The bus lines operate over a set of routes covering a total of 20,688 km. Therefore, the average cycle length per service is of 56.5 km.

(ii) *New integrated system*. The extension of the service network is dramatically reduced to 147 services (public transport routes or lines) covering a total of only 3.888 km. The average service cycle length was reduced to 26.2 km (37.8 for trunk services and 19.9 for the feeder services); less than half of the base

[21] All the monetary data is given in Chilean $ ("pesos"). A US$ corresponds to approximately Ch$590.
[22] Actually, only transfers between Metro lines are free.

Table 5
Supply indexes, off-peak (year 2005)

Index	Optimized base situation	New integrated system
Number of services		
Metro, metro-trains	7	7
Trunk services	–	49
Feeder services	–	91
Current services	370	
Totals	377	147
Kilometers of services[a]		
Metro, metro-trains	222	222
Trunk services	–	1853
Feeder services	–	1812
Current services	20,688	
Totals	20,910	3888
Fleets		
Trunk services		
Buses capacity = 160 pas/bus	–	318
Buses capacity = 120 pas/bus	–	432
Buses capacity = 80 pas/bus	–	271
Feeder services		
Buses capacity = 80 pas/bus	–	323
Buses capacity = 40 pas/bus	–	388
Current services = 80 pas/bus	5416	–
Totals	5416	1732
Equivalent buses (80 pas/bus)	5416	2071

[a] Sum of lengths by type of service.

case. The total fleet size is reduced to 4704 buses (with capacities of 40; 80; 120 and 160 pas/bus) in the peak period. This corresponds to 6031 equivalent 80 pas buses. A significant 25% reduction from the base situation. The equivalent fleet reduction is stronger for the off-peak: from 5.416 to 2.071 (a 62% reduction).

These results are obtained as a consequence of the rationalization of the system of bus services: (i) shorter services with capacities that better fit to flows profiles, (ii) elimination of redundant parallel services to metro lines, increasing the occupancy rate of existing metro services, (iii) complementation of bus (trunk and feeder) and metro services, and (iv) integrated fare system without transfers penalization.

Tables 6 and 7 present a synthesis of trips and level of service characteristics for peak and off-peak hours of both compared situations.

(i) The new system shows a significant increase in the use of the metro system for both peak and off-peak (more than 100% increase for the peak periods and about 70% for off-peak period). This is consistent with the significant reductions in the bus fleet size indicated above.
(ii) A 10% reduction in mean travel times is obtained in the new system during the peak hour and a 15% during the off-peak. Average waiting times are almost identical in both systems during the peak and increases in 40% for the new system during off-peak. During the peak the bus fleet reduction is compensated by the metro services; during the off-peak there is a significant fleet and therefore frequencies reduction. This is compensated by a strong reduction in vehicles operating costs and air pollution. Access or walking times increase in 19% during the peak as a consequence of the increase in transfers and experience a 15% reduction during the off-peak. This access time reduction during the off-peak is due to a better coverage of bus services in the city periphery.

Table 6
Trips indexes (year 2005)

Index	Optimized base situation		New integrated system	
	Peak	Off-peak	Peak	Off-peak
Trip stages				
Metro, metro-trains	185,083	106,531	375,358	180,963
Trunk services	–	–	393,768	125,573
Feeder services	–	–	179,906	38,855
Current services	472,041	144,104		
Totals	657,124	250,635	949,031	345,391
Transfers				
Metro–metro[a]	60,077	29,726	82,739	41,872
Bus–metro	10,477	10,062	122,037	37,757
Metro–bus	6057	3448	57,618	32,475
Bus–bus	39,519	8271	145,989	34,295
Totals	116,130	51,508	408,383	146,399

[a] Including Metro-train.

Table 7
Travel times and transfers (year 2005)

Index	Optimized base situation		New integrated system	
	Peak	Off-peak	Peak	Off-peak
Mean travel time (min)	33.67	29.34	30.32	26.39
Mean waiting time (min)	4.38	4.48	4.43	6.29
Mean walking time (min)[a]	7.61	6.26	9.02	5.33
Average number of transfers	0.21	0.26	0.76	0.74

[a] Access + transfers.

(iii) The results show that the main disadvantage of the new system is the increase in the number of transfers, from 116.130 (0.21 transfers per trip) for the base system during the peak hour, to 408.383 (0.76 transfers per trip) in the new integrated system. For the off-peak the increase is from 51.508 to 146.399. There is a significant increase in the bus-metro combinations (from 16.534 to 179.655 for the peak hour) and in the bus-bus combinations (from 39.519 to 145.989 also for peak hour). However this increase in the number of transfers is necessary for the rationalization of the system and was kept to the lowest value possible through the introduction of combined bus services in trunk corridors.

Finally, Table 8 shows some information about social costs (at this level we consider only the social operating costs and travelers time costs). The value of time used to evaluate operating system costs was provided by the Government Planning Office (equivalent to 1 US$ per hour).The operating costs of the metro system are not included because they are the same for the base and new systems. The main results obtained are the following:

(i) The total number of bus–km decreases about 45% during peak hours (from 141,353 km to 77,105 km) and about 66% during off-peak hours (from 84,558 to 29,059), in the integrated system with respect to the base situation.
(ii) As a result of the fleet and vehicle–km reductions social operating costs where reduced in about 49% during peak hour (from Ch$ 88,769,840 to Ch$ 45,649,142) and 67% during off-peak hour (from Ch$ 30,948,130 to Ch$ 10,221,474). Considering the sum of operators plus users travel costs, the total social cost of the system during the peak hour will go from Ch$ 399,675,104 to Ch$ 349,084,358 and for the off

Table 8
Total social costs of transit system (year 2005)

Index	Peak		Off-peak	
	Optimized base situation	Integrated system	Optimized base situation	Integrated system
Vehicle–km				
Trunk services (160 pas/bus)	–	19,980	–	5976
Trunk services (120 pas/bus)	–	17,738	–	7001
Trunk services (80 pas/bus)	–	7275	–	4432
Feeder services (80 pas/bus)	–	18,870	–	5298
Feeder services (40 pas/bus)	–	13,242	–	6352
Current services	141,353	–	84,558	–
Totals	141,353	77,105	84,558	29,059
Operational costs ($)				
Trunk services (160 pas/bus)	–	12,367,340	–	2,354,569
Trunk services (120 pas/bus)	–	10,405,390	–	2,506,351
Trunk services (80 pas/bus)	–	4,073,934	–	1,427,228
Feeder services (80 pas/bus)	–	11,850,581	–	1,938,923
Feeder services (40 pas/bus)	–	6,951,897	–	1,994,402
Current services	88,769,840	–	30,948,130	–
Totals	88,769,840	45,649,142	30,948,130	10,221,474
Time costs ($)				
Travel time (in vehicle)	227,669,900	204,912,940	73,042,024	65,633,444
Waiting time	29,589,430	29,913,911	11,148,910	15,642,464
Walking time[a]	51,468,503	60,951,180	15,583,690	13,261,803
Transfers penalty[b]	2,177,431	7,657,185	965,767	2,744,989
Totals	310,905,264	303,435,216	100,740,391	97,282,700

[a] Including access and transfers.
[b] Estimated from number of tranfers.

peak the reduction will be from Ch$ 131,688,521 to Ch$ 107,504,174. The savings, that when expanded to the year amount to an equivalent of US$ 217 MM, will go to improve the financial situation of the metro system that, as we saw above, will substantially increase its utilization. The fares will have to pay also the costs associated to the new integrated fare system, based on a contact less card, the new ITS technologies that will be implemented with the new integrated system and infrastructure improvements in corridors and transit stops.

(iii) Total passengers travel time costs are marginally reduced in a 2.5% during peak hour and a 3.5% in off-peak.

(iv) One of the important benefits not shown here are the significant reductions in air pollution that will be obtained as a consequence of the reduction of vehicle–km and the improvement of vehicles technical characteristics. This will allow obtaining significant improvements in the air quality of the city.

6. Conclusions

We have presented in this work a methodology for the physical and operational design of transit systems. From a theoretical point of view, the mathematical models involved can be considered state of the art in the subject. From a practical point of view, the methodology has been implemented in an efficient software package and applied to a real case: the transit system of the city of Santiago. This is a considerably large-scale application, and no computational problems appeared, using a normal PC computer.

The results of this application are quite encouraging. Starting from an existing transit system, a restructured one has been obtained which is of better quality, more efficient (has quite a lower social operating costs and significant less air pollution) and privately sustainable, with the same level of fares values of the old system. This means the new system will operate without the need of subsidies. The savings from the significant reduction obtained in the operating costs of buses will be used to pay operational cost of the new metro services (corresponding to the new lines), to implement improvements to the system, and to finance all the new requirements of an integrated system (for instance, a centralized fare collection system, modern equipment to control and managing bus fleets, construction of intermodal terminals, transit segregated corridors, new stops, etc.).

Acknowledgement

This work was supported by SECTRA and the Universidad Católica de Chile.

References

Abdulaal, M.Y., LeBlanc, L.J., 1979. Continuous equilibrium network design models. Transportation Research Part B 13, 19–32.

Baaj, M., Mahmassani, H., 1995. Hybrid route generation heuristic algorithm for the design of transit networks. Transportation Research Part C 3 (1), 31–50.

Ceder, A., Wilson, N., 1986. Bus network design. Transportation Research Part B 20 (4), 331–344.

De Cea, J., Fernández, J.E., 1993. Transit assignment for congested public transport systems: an equilibrium model. Transportation Science 27 (2), 133–147.

De Cea, J., Fernández, J.E., 1995. Diseño de un sistema de servicios alimentadores bus-metro. Apuntes de Ingeniería 18 (1), 51–64.

De Cea, Ch., Fernández, J.E., Dekock, V., Soto, A., 2005. Solving network equilibrium problems on multimodal urban transportation networks with multiple user classes. Transport Reviews 25 (3), 293–317.

Fernández, J.E., De Cea, J., Norambuena, I. 2003. Una Metodología para el Diseño Topológico de Sistemas de Transporte Público Urbano de Pasajeros. In: Proceedings of the XI Congreso Chileno de Ingeniría de Transporte, Santiago de Chile, November 2003.

Friesz, T.L., Cho, H., Mehta, N.Y., Tobin, R., 1992. Simulated annealing methods for network design problems with variational inequality constraints. Transportation Science 26 (1), 18–26.

Friesz, T.L., Tobin, R., Shah, S., Mehta, N., Anandalingam, Y., 1993. The multiobjective network design problem revisited: a simulated annealing approach. EJOR 65, 44–57.

Friesz, T.L.Y., Shah, S., 2001. An overview of non-traditional formulations of static and dynamic network design. Transportation Research Part B 35 (1), 5–22.

Lee, Y. 1998. Analysis and optimization of transit network design with integrated routing and scheduling. Unpublished Ph.D. Thesis, University of Pennsylvania.

Marcotte, P., 1983. Network optimization with continuous control parameters. Transportation Science 17, 181–197.

Meng, Q., Yang, H.Y., Bell, M.G.H., 2001. An equivalent continuous differentiable model and locally convergent algorithm for the continuous network design problem. Transportation Research Part B 35 (1), 83–105.

Newell, G.F., 1979. Some issues related to the optimal design of bus routes. Transportation Science 13, 20–35.

Stackelberg, H.V., 1952. The Theory of Market Economy. Oxford University Press, England.

SECTRA, 2003a. Análisis Modernización Transporte Público V Etapa (Design Study).

SECTRA, 2003b. Análisis Modernización Transporte Público V Etapa (Costs Study).

Van Nes, R., Hamerslag, R., Immers, B.H., 1988. Design of public transport networks. Transportation Research Record 1202, 74–83.