Transportation Research Part B 142 (2020) 1–18

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Strategies for transit fleet design considering peak and off-peak periods using the single-line model

Sergio Jara-Díaz[a,*], Andrés Fielbaum[b], Antonio Gschwender[a]

[a] Universidad de Chile and Instituto Sistemas Complejos de Ingeniería, Chile
[b] Universidad de OĤiggins, Chile

ARTICLE INFO

Article history:
Received 17 July 2019
Revised 12 September 2020
Accepted 25 September 2020
Available online 10 October 2020

Keywords:
Public transport
Peak off-peak periods
Fleet design
Holding

ABSTRACT

Transit demand and traffic conditions present relevant differences between peak and off-peak periods - e.g. flows, trip lengths, congestion - raising a relevant strategic design choice regarding the potential use of vehicles with different sizes. Here we first revisit the optimal design using a single-line, single-fleet model, showing that buses should always run full at the peak but not always at the off-peak. Then we develop a two-fleet strategy (with different vehicle sizes) where one fleet operates the whole day and the other during the peak period only. This strategy includes holding during the peak (in order to avoid bunching) by imposing equal cycle times for both fleets. The two-fleet operation has slightly lower total costs than one-fleet, but exhibits very different effects on users' and on operators' costs across periods. A sensitivity analysis reveals the role played by various elements and shows that results are robust. Optimal one and two-fleet designs are both better than optimizing each period independently, revealing economies of time scope.

1. Introduction

Urban public transport systems operate daily during periods that exhibit quite different patterns of demand in both space and time, something that should be taken into account at the design stage. Regarding variations in time, the usual approach is to represent the day by considering peak and off-peak periods such that, by definition, the peak exhibits the largest demands (trips per unit time); this, however, is not the only characteristic that varies, as the peak period is usually shorter, trip distances are longer and congestion is present. How to design in an optimal way public transport services taking these differences between periods into account is still an open question; in particular, we still need to understand better how the strategic long-run design, which has been studied in depth for single-period scenarios, is affected.

Despite its relative simplicity, the single-line model has proved useful to analyze general characteristics of public transport systems, helping to identify relevant elements that survive when extensions to networks are considered. This is the case, for example, with the growth of frequency (Mohring, 1972) and vehicle size (Jansson, 1984) at a decreasing rate as demand increases, used by Jara-Díaz and Gschwender (2009) to show the unpleasant effects caused by cost recovery, by Hörcher and Graham (2018) to analyze demand imbalance, and by Jara-Diaz and Tirachini (2013) to study the impact of boarding technology on cycle time, fleet and vehicle size. These and other studies have indeed contributed to unveil relevant properties of transit design, including the paradigmatic Mohring effect (reduction of waiting time as flow increase).

* Corresponding author at: Universidad de Chile Facultad de Ciencias Fisicas y Matematicas, Chile.
E-mail address: jaradiaz@ing.uchile.cl (S. Jara-Díaz).

https://doi.org/10.1016/j.trb.2020.09.012
0191-2615/

In these types of analyses based on a stylized representation of a transit system, a number of simplifications and approximations are used. For instance, Hörcher and Graham (2018) and Evans and Morrison (1997) dismiss time at stops, similar to the first model presented by Mohring (1972) in its seminal paper, which implies a constant cycle time independent of boarding and alighting time. Similarly, acceleration and deceleration of vehicles at stops is typically included in average speed when the number of stops is not optimized, and detailed bus design (e.g. number of doors) is not included besides vehicle capacity (boarding-alighting time considered fixed). Notably, when the variation of operators' costs with vehicle size is introduced making size a design variable, it is done assuming it is continuous (see Jansson, 1980, 1984; Oldfield and Bly, 1988; Jara-Díaz and Gschwender, 2009).

In order to study the role played by the difference between peak and off-peak periods on the optimal strategic design of public transport, Jansson (1984) departed from his own model of one line with passengers distributed homogeneously in space during an isolated period (Jansson, 1980), and attempted a generalization to two periods noting that the size of the fleet is given by peak conditions; this makes the peak bear the burden of fixed cost (fleet acquisition), such that operators' costs during the off-peak are only operational. He imposed, however, the very limiting assumption of equal frequencies in both periods. Jara-Díaz et al. (2017) dropped this assumption and analyzed the two periods' case keeping a single fleet operation. Noting that the fleet size is commanded by the peak conditions, they imposed that vehicle size is also determined by the peak by making buses run full; they obtained that optimal frequencies would differ more than when each period is solved in isolation, while capacity is in-between those obtained independently by period. They noted, however, that vehicle size could be determined by the off-peak period under some conditions that do not seem unusual, because a lower frequency might cause buses to run full. "Although fleet is always calculated according to peak conditions (that present larger flows by definition), bus size could be determined by off-peak conditions if its flow is sufficiently large; this would require a re-examination of the solutions making size equal to the off-peak requirement" (Jara-Diaz et al, 2017, pp. 73).

Both Jansson (1984) and Jara-Díaz et al (2017) studied the two-period case assuming a single fleet, i.e. all vehicles have the same size. Nevertheless, as peak and off-peak periods might exhibit very different conditions, it is quite reasonable to think that using two fleets (of different vehicle size) could introduce relevant degrees of flexibility, reducing total costs (users' plus operators'); buses of different size within the same line are indeed observed in many cities. As discussed by Walker (2012), the off-peak period is much longer than the peak such that the design could well be done considering the off-peak conditions introducing complementary services at the peak with buses of possibly different capacity. For short, the two-fleets system has been both proposed and used but a formal analysis is still missing.

A two-fleet operation can be conceived under different combinations, such as using independent fleets at each period, using only one fleet at the off-peak and complement with a different one at the peak, or using both fleets but at different rates at each period. In this paper, we will compare the single-fleet strategy against the strategy that considers a complementary fleet running at the peak (including a holding rule to avoid differences in cycle times). This comparative analysis will consider as a reference the extremely simple strategy of optimizing each period independently, which is also useful to study the presence of (period) scope economies. To do a proper and fair comparison, we first have to complete the single-fleet analysis by solving the case where buses run full at the off-peak period using a single fleet; second we analyze the strategically optimal design using two fleets, one that will be used during both periods and the other that will be used to complement service at the peak.

Besides the analyses summarized above made by Jansson (1984) and Jara-Díaz et al (2017), there are other models that focus on specific aspects of public transport using more than one period. Chang and Schonfeld (1991), for instance, study the optimal spatial separation between consecutive lines in public transport, minimizing costs that, in one of their specifications, occur in more than one period. Similarly, Tirachini et al. (2010) compare different public transport technologies in a model that optimizes angular density between lines considering costs for various periods. Along a different research line, Glaister and Lewis (1978), De Borger et al. (1996), Proost and Van Dender (2008), Parry and Small (2009) and Basso and Silva (2014) study the problem of optimal fares and subsidies for public and private transport to achieve an optimal modal split considering peak and off-peak periods. Fernández et al. (2005) consider various periods in their analysis of operators costs only, holding bus size fixed.

The paper is organized as follows. In Section 2 we re-formulate the general two-periods single-fleet model presented by Jara-Díaz et al. (2017), indicating the assumptions made, summarizing their most important results, and generalizing them by studying the case in which buses run full at the off-peak. In Section 3 we formulate, solve and analyze the model where two fleets are allowed. In Section 4, we compare the different systems that can be designed to serve two periods, while in Section 5 we run a sensitivity analysis on various elements and discuss the role of omitted ones. Section 6 concludes.

2. Single-fleet serving two periods: generalization of previous results

2.1. The one-period model: a synthesis

Before studying the two-periods model it is useful to begin explaining briefly the one-period model formulated by Jara-Diaz and Gschwender (2003, 2009) that generalizes Jansson (1980, 1984). They studied the optimal frequency f and bus capacity K for a line of length L that is used by Y passengers per hour, homogeneously distributed along the line, all of them traveling a distance l, such that the flow F on each section is given by Yl/L, yielding a load per bus equal to Yl/fL. T is the time it would take a vehicle to tour the whole cycle if there were no bus stops. Time at bus stops is captured by the time t needed by each passenger to board and alight the bus. The (daily) capital cost of a vehicle is given by $c_{BC} + c_{KC}K$; operating costs are given by $E(c_{BO} + c_{KO}K)$, where E is the length of the period. The values of time are p_w (waiting) and p_v (in-vehicle)[1].

Assuming that vehicle capacity is just enough to carry all passengers, and expressing cycle time ($t_c = T + tY/f$) and fleet ($B = ft_c$) as a function of frequency, the total value of the resources consumed (VRC) per day considering a single period of length E can be written as:

$$VRC = B(c_{BC} + Kc_{KC}) + EB(c_{BO} + Kc_{KO}) + \frac{EYp_w}{2f} + EYp_v\frac{t_c l}{L} = \\ T(c_{BC} + Ec_{BO})f + \left[\frac{Y^2lt}{L}(c_{KC} + Ec_{KO} + Ep_v) + \frac{p_wEY}{2}\right]\frac{1}{f} + T(c_{KC} + Ec_{KO})\frac{lY}{L} + tY(c_{BC} + Ec_{BO}) + \frac{p_v lYT_P E}{L} \tag{1}$$

The first expression for VRC in (1) is useful to describe the model: the first term are the capital costs, the second term represents the operating costs, while the third and fourth term are users' waiting and in-vehicle costs, respectively. The second expression emerges by replacing all the variables to write VRC as a function of f only, which yields an optimal frequency given by

$$f = \sqrt{\frac{\frac{Y^2lt}{L}(c_{KC} + Ec_{KO} + Ep_v) + \frac{p_wEY}{2}}{T(c_{BC} + Ec_{BO})}}, \tag{2}$$

which can be synthetically expressed as $f = \sqrt{\frac{G}{A}}$, with $G = \frac{Y^2lt}{L}(c_{KC} + Ec_{KO} + Ep_v) + \frac{p_wEY}{2}$ and $A = T(c_{BC} + Ec_{BO})$. The second derivative confirms that this result yields a minimum of VRC.

2.2. The two-periods model so far (one fleet)

In Jara-Díaz et al. (2017), VRC was expanded to consider peak (P) and off-peak (N) conditions represented by Y_i, l_i, E_i and $T_i (i = P, N)$, with capital costs incurred only at the peak. Then the value of the resources consumed during two periods, VRC_2, is:

$$\begin{aligned} VRC_2 &= B_P(c_{BC} + c_{KC}K) + (B_PE_P + B_NE_N)(c_{BO} + c_{KO}K) + \frac{p_w}{2}\left(\frac{Y_PE_P}{f_P} + \frac{Y_NE_N}{f_N}\right) + \frac{p_v}{L}(l_Pt_{cP}Y_PE_P + l_Nt_{cN}Y_NE_N) \\ &= (f_PT_p + tY_P)(c_{BC} + c_{KC}K) + [(f_PT_P + tY_P)E_P + (f_NT_N + tY_N)E_N][c_{BO} + c_{KO}K] + \frac{p_w}{2}\left(\frac{Y_PE_P}{f_P} + \frac{Y_NE_N}{f_N}\right) \\ &+ \frac{p_v}{L}\left(l_PY_PE_P\left[t\frac{Y_p}{f_P} + T_P\right] + l_NY_NE_N\left[t\frac{Y_N}{f_N} + T_N\right]\right) = VRC_2(K, f_P, f_N) \end{aligned} \tag{3}$$

where the second expression for (3) is formulated as a function of K, f_P and f_N by properly replacing B_i and t_{ci} as functions of the frequencies.

VRC_2 should be minimized subject to capacity constraints for both periods, i.e.

$$\text{a)} \quad K \geq \frac{Y_Pl_P}{f_PL} \text{ and b)} \quad K \geq \frac{Y_Nl_N}{f_NL} \tag{4}$$

It is easy to see that $VRC_2(K,f_P,f_N)$ in Eq. (3) is convex; nevertheless, as we explain in the following paragraph, constraints (4) prevent the minimum to be reached within the interior of the feasible region. The conditions for finding the optimal frequencies and capacities, then, are obtained by a careful analysis of the cases in which constraints (4) are active or not.

As VRC_2 increases with K in (3), it is clear that at least one of the two constraints in (4) has to be active at the optimal solution. In Jara-Díaz et al. (2017), bus size was assumed to be given by the peak, i.e., the peak constraint (4a) was active and the off-peak constraint was not. After replacing this value of K in VRC_2, first order conditions for both frequencies were obtained[2]. Closed forms were impossible to obtain from the resulting equation of order five, but an analysis using coefficients like A and G, introduced earlier in this paper, was enough to prove that peak frequency was larger when considering both periods than when considering the peak period in isolation. This happens because buses are used during both periods, such that the off-peak operating costs (that increases with K) pushed bus size downwards, a result that emerges because K is being optimized as well. A similar analysis for the off-peak frequency was inconclusive, but numerical simulations showed that the off-peak frequency was lower than in the isolated case. Moreover, and precisely because of this (small buses and low frequency), the numerical analysis suggested that buses could get full at the off-peak. In this case, it is constraint (4b) that becomes active.

[1] Writing operators' costs as the sum of capital and operating costs is not necessary when studying only one period; neither is necessary to make the duration E explicit. We do it here to facilitate the explanation of the two-periods case.

[2] Jansson (1984) did not consider bus size K as a factor influencing operators' cost. He did not obtain a system of equations for the optimal frequencies, but he imposed equality between them and found an expression for the optimal single frequency.

Finally, Jara-Diaz et al. (2017) analyzed the crossed effects between periods, proving that $\frac{\partial f_N}{\partial Y_P} < 0$ and $\frac{\partial f_P}{\partial Y_N} > 0$. The first inequality is explained because bus size increases with peak patronage, so buses are more costly to operate; the second inequality happens because off-peak frequency increases with off-peak patronage, so it is costlier to have larger buses.

For summary, the two periods analysis in Jara-Diaz et al. (2017) concluded that: i) peak frequency surpasses the off-peak one by an amount larger than the one obtained when isolated periods are considered; ii) capacity is in-between those obtained independently by period; iii) the larger the difference across different periods' demands, the lower both frequencies; and iv) bus size could be given by off-peak conditions (a case that was not analyzed). Their emphasis was on understanding the strategic design using one fleet only, which is evidently a limitation.

2.3. The generalized two-periods model (one fleet)

For a proper comparison of these results against other possible strategies using two fleets, the single fleet analysis has to be completed first by studying the case where buses run full at the off-peak, a case that was shown numerically possible; intuitively, this could happen because optimal frequency at the off-peak is lower than at the peak such that buses could run at capacity under some conditions. The remainder of this Section is devoted to complete the analysis with the case in which the off-peak constraint is active (buses full at the off-peak), i.e.

$$K = \frac{Y_N l_N}{f_N L} \tag{5}$$

It should be noted that fleet size is still given by the peak, because both cycle time and frequency are larger than at the off-peak (larger flow, trip length and in-motion time). Then VRC_2 in (3) can be written as a function of frequencies using Eq. (5) and the expressions that link fleet size and cycle time with the corresponding frequencies. This yields:

$$VRC_2 = (f_P T_p + tY_P)\left(c_{BC} + c_{KC}\frac{Y_N l_N}{f_N L}\right) + [(f_P T_P + tY_P)E_P + (f_N T_N + tY_N)E_N]\left[c_{BO} + c_{KO}\frac{Y_N l_N}{f_N L}\right] + \frac{p_w}{2}\left(\frac{Y_P E_P}{f_P} + \frac{Y_N E_N}{f_N}\right) + \frac{p_v}{L}\left(l_P Y_P E_P\left[t\frac{Y_p}{f_P} + T_P\right] + l_N Y_N E_N\left[t\frac{Y_N}{f_N} + T_N\right]\right) \tag{6}$$

Following Jara-Díaz et al. (2017) it is useful to re-write Eq. (6) in a more compact form as

$$VRC_2 = A_P f_P + G_P/f_P + A_N f_N + G_N/f_N + \varepsilon f_P/f_N + W \tag{7}$$

where the explicit expressions for A_P, A_N, G_P, G_N, ε and W are

$$A_P = T_P c_{BC} + T_P E_P c_{BO} \tag{8}$$

$$A_N = T_N E_N c_{BO} \tag{9}$$

$$G_P = \frac{p_w Y_P E_P}{2} + \frac{p_v\ l_P E_P t {Y_P}^2}{L} \tag{10}$$

$$G_N = tY_P c_{KC} Y_N \frac{l_N}{L} + tY_P E_P c_{KO} Y_N \frac{l_N}{L} + t{Y_N}^2 E_N c_{KO}\frac{l_N}{L} + \frac{p_w Y_N E_N}{2} + \frac{p_v\ l_N E_N t {Y_N}^2}{L} \tag{11}$$

$$\varepsilon = T_P c_{KC}\frac{Y_N l_N}{L} + T_P c_{KO}\frac{Y_N l_N}{L} E_P \tag{12}$$

$$W = tY_P c_{BC} + tY_P E_P c_{BO} + T_N Y_N c_{KO}\frac{l_N}{L} + \frac{p_v}{L}(Y_P E_P l_P T_P + Y_N E_N l_N T_N) \tag{13}$$

Regarding constraint (4a) there are two possible cases: active or non-active, i.e. buses run either full or not at the peak. The following proposition shows that they indeed run full.

Proposition: the peak constraint is active when the off-peak constraint is.

Proof: by contradiction. Let us assume that constraint (4a) is not active. In that case, we just take the derivatives in Eq. (7) and make them both equal to zero, which yields:

$$f_N = \sqrt{\frac{G_N + \varepsilon f_P}{A_N}} \text{and} f_P = \sqrt{\frac{G_P}{A_P + \varepsilon/f_N}} \tag{14}$$

These expressions can be compared with the frequencies obtained for each period in isolation, i.e. $f_{i1} = \sqrt{\frac{G_{i1}}{A_{i1}}}$, with $A_{i1} = T_i(c_{BC} + E_i c_{BO})$ and $G_{i1} = \frac{Y_i^2 l_i t}{L}(c_{KC} + E_i c_{KO} + E_i p_v) + \frac{p_w E_i Y_i}{2}$.

It is quite direct to observe that:

$A_P = A_{P1}$
$A_N < A_{N1}$: no capital costs c_{BC} at the off-peak.
$G_P < G_{P1}$: costs associated with bus size, $c_{KC} + E_i c_{KO}$, do not appear in G_P (as bus size is given by the off-peak, changing frequency does not reduce the bus size).
$G_N > G_{N1}$: the first two terms in G_N, involving Y_P and costs associated with bus size, do not appear in G_{N1}(choosing large buses is costlier than in the isolated case, because the same buses will run at the peak; so it is better to have smaller buses).

These relationships show directly that

i) peak frequency f_P is lower than f_{P1} (isolated case);
ii) off-peak frequency f_N is larger than f_{N1} (isolated case).

Conclusion ii) implies that bus size is lower than in the isolated off-peak case, which in turn is lower than in the isolated peak case (because K increases with Y). Therefore, the assumption of constraint (4a) not active when (4b) is, leads to a contradiction: at the peak, frequency and bus size are simultaneously lower than in the isolated case, which cannot occur and proves the assumption wrong. **Q.E.D.**

The Proposition implies that buses run full at the peak under every circumstance (it is only the off-peak that may present both cases). Therefore it always holds that

$$f_P = \frac{Y_P l_P}{KL} \tag{15}$$

So (4a) is always active. We know from Jara-Díaz et al (2017) that (4b) might be active as well, but the case in which both constraints are active has not been yet analyzed; let us do it. Combining Eqs. (15) and (5) we obtain an expression for f_P as a function of f_N:

$$f_P = \frac{Y_P l_P}{Y_N l_N} f_N \tag{16}$$

This confirms that the peak frequency has to be larger than the off-peak one, because both $Y_P > Y_N$ and $l_P > l_N$; their ratio replicates the ratio between the corresponding passengers-kilometers traveled per hour. Thus, contrary to Jansson's intuition, it is never optimal that frequencies are equal across periods.

Using (9) the value of the resources consumed can be expressed as a function of f_N only:

$$VRC_2 = \left(A_P \frac{Y_P l_P}{Y_N l_N} + A_N\right) f_N + \left(G_N + G_P \frac{Y_N l_N}{Y_P l_P}\right) \frac{1}{f_N} + \varepsilon \frac{Y_P l_P}{Y_N l_N} + W \tag{17}$$

As VRC_2 is a convex function of f_N first order condition yields a minimum at

$$f_N^* = \sqrt{\frac{tY_P c_{KC} Y_N \dfrac{l_N}{L} + \dfrac{p_w Y_N (E_N + E_P l_N/l_P)}{2} + [E_P Y_P + E_N Y_N] t Y_N (l_N/L)(c_{KO} + p_v)}{T_N E_N c_{BO} + T_P c_{BC} Y_P l_P/(Y_N l_N) + T_P E_P c_{BO} Y_P l_P/(Y_N l_N)}}, \tag{18}$$

Then, by virtue of Eqs. (16) and (5)

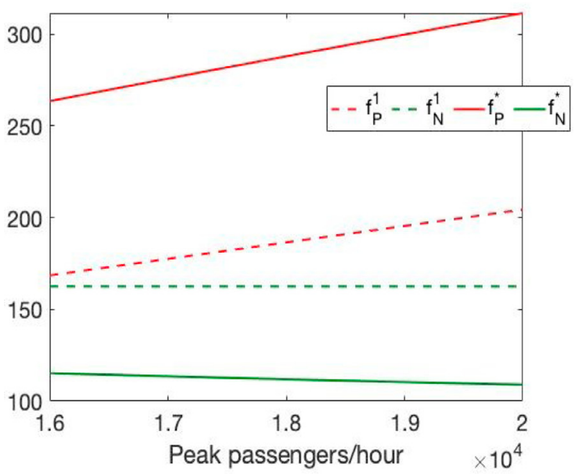
$$f_P^* = Y_P \sqrt{\frac{tY_P c_{KC} \dfrac{l_P^2}{L} + \dfrac{p_w \left(E_N l_P^2/l_N + l_P E_P\right)}{2} + [E_P Y_P + E_N Y_N] t \left(l_P^2/L\right)(c_{KO} + p_v)}{T_N E_N c_{BO} Y_N l_N + T_P c_{BC} Y_P l_P + T_P E_P c_{BO} Y_P l_P}} \tag{19}$$

$$K^* = \sqrt{\frac{Y_N T_N l_N E_N c_{BO} + T_P c_{BC} Y_P l_P + T_P E_P c_{BO} Y_P l_P}{tY_P c_{KC} L + L^2 \dfrac{p_w (E_N + E_P l_N/l_P)}{2} + L[E_P Y_P + E_N Y_N] t (c_{KO} + p_v)}} \tag{20}$$

The comparison between the expressions for f_N^* and f_P^* with those given by the corresponding isolated ones can be done by looking at the (expanded) generic version of the optimal isolated frequencies for one period, shown in Eq. (21):

$$f_{i1} = \sqrt{\frac{Y_i \left(\dfrac{p_w E_i}{2} + tY_i (l_i/L)(p_v E_i + c_{KC} + c_{KO} E_i)\right)}{T_i (c_{BC} + E_i c_{BO})}} \tag{21}$$

The analytical comparisons between optimal frequencies in Eqs. (18) and (19) with the corresponding expressions of Eq. (21), lead to inconclusive results because they depend on the values of all parameters. Instead, a numerical comparison is offered in Fig. 1 using parameters that are representative of Santiago, Chile (shown in the Appendix), varying Y_P from



a)$Y_N = 14{,}000$ passengers/h b)$Y_P = 40{,}000$ passengers/h

Fig. 1. Comparison of peak and off-peak frequencies considering joint optimization against isolated optimization.

16,000 to 20,000 keeping Y_N=14,000 in Fig. 1a, and varying Y_N from 26,000 to 35,000, keeping Y_P=40,000 in Fig. 1b. [3] Ranges are chosen such that the off-peak constraint is always active at the optimum (i.e. Eq. 5 holds).

From Fig. 1 we obtain that, numerically, $f_P^* > f_{P1}$ and $f_N^* < f_{N1}$, the same results obtained in Jara-Díaz et al (2017) (shown analytically for f_P^*), when (4a) was assumed to be active - which we now know is always true at the optimum - and (4b) was assumed to be non-binding - which we now know might happen or not, depending on the parameters. Fig. 1 also shows that the relations analytically found for the crossed effects between frequencies and flows hold in this case as well, i.e.

$$\frac{\partial f_P}{\partial Y_N} > 0, \ \frac{\partial f_N}{\partial Y_P} < 0 \tag{22}$$

As by definition $Y_P > Y_N$, these equations show that when flow i approaches flow j unilaterally, frequency j increases. The intuition behind this is related with the relative values of flows across periods: when they get closer to each other in magnitude, the day becomes more homogeneous and approaches a single extended period, making capital costs weigh less.

Regarding bus size, although Eq. (20) shows that K could increase or decrease when either Y_P or $Y_N$increases, Eq. (22) imply that

$$\frac{\partial K}{\partial Y_P} > 0, \ \frac{\partial K}{\partial Y_N} < 0 \tag{23}$$

In this case, when flow i approaches flow j unilaterally (i.e. when Y_N increases or Y_P decreases) bus size decreases. This could be interpreted in terms of the parameters involved in Eq. (20), particularly for the off-peak flow effect:

- When Y_P increases, bus size has to increase just as in the single period case.
- When Y_N increases, it is better to reduce bus size in order to reduce the time spent at bus stops.

Now that the case in which off-peak buses run full has been presented, we can make an analysis holding $Y = Y_P + Y_N$ fixed, in order to examine the total effects of flow variations when they get closer. These effects can be seen in Fig. 2, where $Y = 40,000$ and Y_N/Y varies from 5% to 48%. As the off-peak flow increases while decreasing the peak flow, off-peak frequency increases because of a direct effect and the corresponding crossed effect (Eq. 22). In the case of peak frequency, it is pushed upwards because of the crossed effect and downwards because of the decrease of Y_P (direct effect); this last effect prevails resulting in a slight decrease. Bus size decreases because both effects represented in Eq. (23) work in the same direction.

Fig. 2 reveals a slight but evident change of slope in the three curves represented when the ratio Y_N/Y reaches 0.42, which is the value where constraint (4b) becomes active, i.e. when the off-peak buses start to be full. This motivates an examination of the conditions on the flows that make constraint (4b) active. Such conditions cannot be found explicitly, as

[3] The chosen figures and ranges are representative of corridors in Santiago, where the most loaded section in the highest demand corridor is crossed by a flow F equal to 15,000 and 5,000 pax/h in one direction during peak and off-peak periods respectively, which means that total demands Y entering the corridor are 60,000 and 40,000 pax/h respectively, considering the average trip and route lengths given in the Appendix (recall that $F = Yl/L$). Santiago concentrates about one third of the Chilean population and is quite extended. According to the last OD survey, 2.8 trips per person are made, 65 % by public transport except in the north-east rich area (27%). If approximated by a circle, the diameter would be some 40 Km. Measured average trip length indeed differs between peak and off-peak but both are quite large.

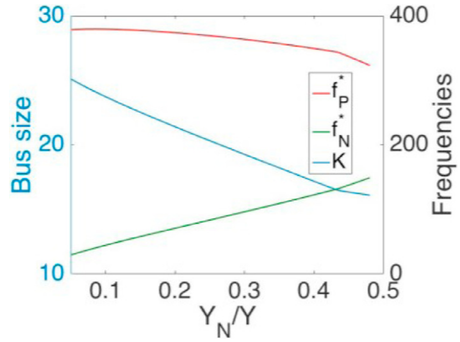


Fig. 2. Bus size, peak and off-peak frequencies as a function of $\boldsymbol{Y_N/Y}$.

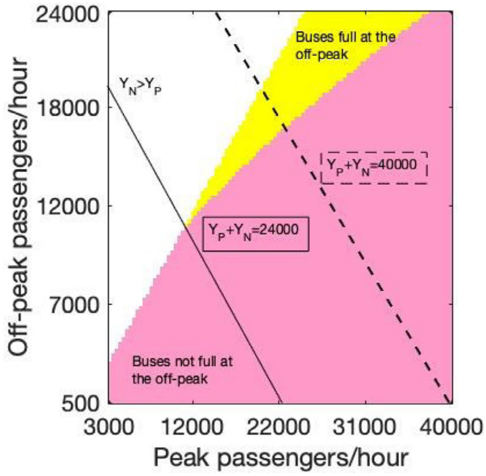


Fig. 3. Flow conditions that make buses full at the off-peak.

they depend on the relationship between frequencies and flows when only constraint (4a) is active, and they are linked by an equation of degree 5 (Jara-Díaz et al., 2017). The conditions on the flows are represented numerically in Fig. 3, where the yellow zone represents the combinations of flows that activate constraint (4b) while the pink zone contains the points where the constraint is not active; within the white zone Y_N is larger than Y_P and does not belong in the analysis.

The dotted diagonal line in Fig. 3 shows the combinations used to construct Fig. 2, i.e. where total demand $Y = Y_P + Y_N = 40,000$. The solid diagonal line represents Y around 24,000, a level at which the yellow zone emerges and begins to grow, meaning that it becomes more likely that off-peak buses run full. This can be explained by looking at the growth of Y_N at every level of Y_P; by virtue of Eq. (23) buses get smaller and eventually get full (for a large enough Y_P) in spite of the increase in off-peak frequency. Note that, irrespective of the peak flow, buses at the off-peak never run full for off-peak flows lower than 10,000 pax/hour.

3. A two-fleet strategy

So far we have dealt with models that assume that all vehicles have the same size (one fleet), such that the whole fleet is used during the peak and a subset during the off-peak. Given the differences across periods, the use of two fleets of different vehicle sizes is a reasonable alternative to explore. But this approach has, in turn, many alternatives. The simplest one is the design of one fleet for each period, which evidently generates idle capacity and suggests possible advantages if one fleet is used (at least partially) during the other period as well. So, a better alternative could be to design two fleets simultaneously: one that runs alone in the off-peak – the "small" buses - and other that runs as a complement during the peak only to satisfy the (larger) demand. Let us analyze this alternative.

The simultaneous operation of vehicles of different size during the peak means that the large ones would spend more time at stops because they carry more passengers, which would imply that their cycle time could also be larger. This would induce at least three undesirable effects:

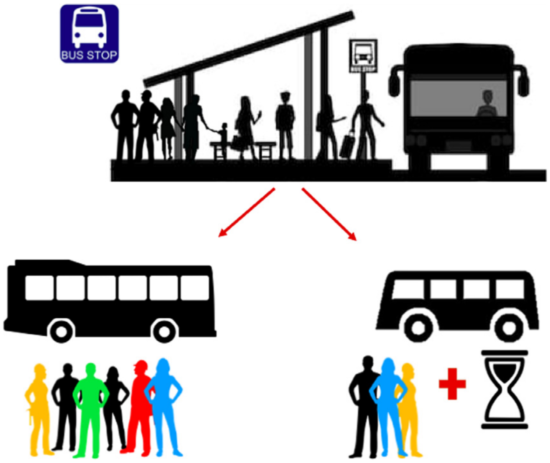


Fig. 4. Graphical description of the holding strategy.

- It is a well-known fact that different time at stops might induce *bunching* (see, for example, Newell, 1974). The explanation in our case is quite simple: if a large bus is followed by a small bus, the temporal headway between them is going to be reduced after each stop, just because the large bus is spending more time there.
- Headways between consecutive buses are not constant in time, which means that the system becomes much more irregular. Headway variability is an index of unreliability and has been shown to be a very relevant (negative) quality factor for users by many researchers (like Friman et al., 1998; Beirão et al., 2007; Redman et al., 2013), such that regularity has merit in itself.
- A particular aspect of the headway variability condition is the chance that some buses might get full, forcing passengers to wait for the next bus. This is also likely to happen when headways change over time because longer headways induce a larger amount of passengers accumulated at the bus stop.

So, letting the system operate uncontrolled is a bad strategy. To avoid the harmful effects aforementioned, some actions need to be applied to induce equal cycle times across fleets in the peak. In this model we will use a *holding* strategy (that has been well studied in other contexts, such as Osuna and Newell, 1972; Daganzo, 2009), simply consisting in forcing small vehicles to have the same time at stops as large vehicles. This can be done by waiting with their doors closed for a fixed time (after the passengers have boarded) before leaving the stop[4], as illustrated in Fig. 4.

Therefore, there is a trade-off when buses of different size are introduced: if they are too similar, buses will be too large at the off-peak; if they are too different, the time spent holding would be large. The system studied in Section 2 can be considered an extreme case in which buses in the off-peak have the same size as those of the peak, possibly running below capacity at the off-peak. Let us now study in detail the other extreme case, in which there is a fleet of small buses, all of them running full at the off-peak; large buses run only at the peak and always run full because there is no cost advantage otherwise. Solving this problem with the holding strategy at the peak is quite complex but provides all the elements to understand neatly the differences between the single-fleet and the two-fleets cases.

To describe the equations that govern the operation of the system, let us introduce first some notation. B_S and B_L will denote the size of the small and large fleets respectively, such that $B_S + B_L$ buses will be running at the peak and B_S at the off-peak. Vehicles capacities are K_S and K_L respectively. We will not describe operational rules at the off-peak, as this period can be represented by the traditional one-line model provided capital costs are correctly adjusted to avoid double-counting. At the peak, headways h_S and h_L are going to be used instead of frequencies (which are not particularly informative in this case); h_i represents the time elapsed between the instant in which the previous vehicle (that might be large or small) closes its doors and the instant in which the current type i vehicle does. In other words, h_S and h_L represent the period during which passengers arrive at the bus stop to board the corresponding vehicle. Note that: 1) under this definition, the time spent by a small bus at the stop with its doors closed is not part of its headway in that stop; it belongs to the headway of the next bus, 2) the headway between two consecutive buses depends only on the size of the second bus. The operation with headways as defined above is represented in Fig. 5 in a space-time diagram with three red large buses, two green small ones, and three stops. Let us describe the operation at the first stop, starting with the large red bus that arrives at b and gets those passengers that arrive between a and c. The following (first small green) bus gets passengers arriving from c to d, and the second small green one receives those passengers arriving between d and f. Note that the time that the first green small bus spends holding (d to e) belongs to the "headway" of the following (green small) bus, that lasts until

[4] A literal application of this rule could raise complaints from the users on board. Nevertheless, it can be changed to some equivalent rules that do not affect the equations deduced here, such as diminishing the speed of the small buses during the peak period.

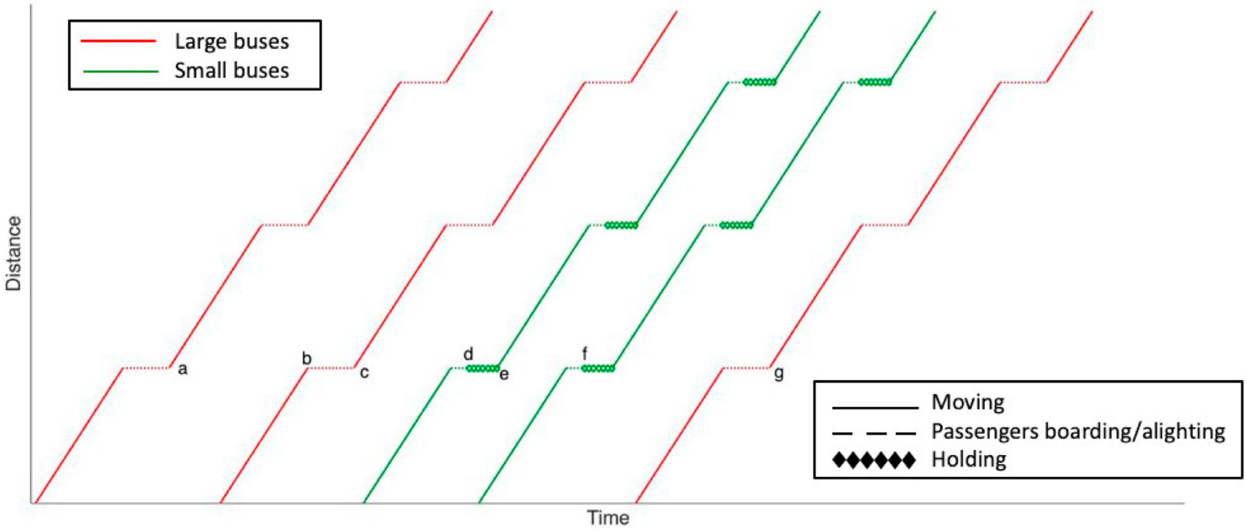


Fig. 5. The holding strategy at the peak.

doors are closed at f, when holding begins. The next red large bus receives the passengers arriving from f to g. This process replicates in the other stops and continues in time.

During the peak period a total of Y_P passengers arrive per unit time such that the number of passengers that ride a particular vehicle type during a cycle can be expressed as $Y_P h_i$; then total time spent at the stops is $tY_P h_i$. Equality of cycle times t_c of both types of vehicles imply

$$T_P + tY_P h_S + H = t_c = T_P + tY_P h_L, \tag{24}$$

where H is the total holding time spent by each small vehicle in a cycle. We will show that the *VRC* can be expressed as a function of B_L and B_S only, as vehicle sizes and headways can be written as functions of the fleet sizes. To do this, let us begin noting that vehicle sizes are given by

$$K_L = Y_P h_L \frac{l_P}{L}, \quad K_S = Y_P h_S \frac{l_P}{L} \tag{25}$$

To calculate the average waiting time, note that if the next bus is of type i, passengers will wait from zero to h_i, such that in average they wait for $h_i/2$. A proportion $\frac{K_i B_i}{K_L B_L + K_S B_S}$ of the users take a bus of type i, which yields an average waiting time given by

$$t_w^- = \frac{B_L K_L}{B_L K_L + B_S K_S}\frac{h_L}{2} + \frac{B_S K_S}{B_L K_L + B_S K_S}\frac{h_S}{2} \tag{26}$$

Eqs. (25) and (26) yield

$$t_w^- = \frac{1}{2}\frac{B_L h_L^2 + B_S h_S^2}{B_L h_L + B_S h_S} \tag{27}$$

Under this formulation waiting time for passengers boarding a small vehicle includes the time waiting for other passengers to board as is usually done, but, once the doors are closed, time until departure is considered as in-vehicle time. Following Jansson (1980) and Jara-Díaz and Gschwender (2003), time waiting for passengers in other stops is always in-vehicle time for those already aboard the bus. Then in-vehicle time can be expressed as:

$$t_v^- = \frac{l_P}{L} t_c = \frac{l_P}{L}(T_P + tY_P h_L) \tag{28}$$

So far, we have expressed the components of the users' costs as functions of both fleet sizes and headways. Operators' costs - that depend on fleet sizes and vehicle sizes - can also be expressed as functions of B_i and h_i using Eq. (25). Two additional equations will allow us to have fleet sizes as the only variables. The first one is a different way of calculating the cycle time:

$$t_c = B_L h_L + B_S h_S \tag{29}$$

To understand where does this come from, imagine a user observing which bus to take during a lapse of time that lasts t_c. This user will observe each of the B_L large buses as the next one coming during a lapse h_L (and equivalently with each small bus). Eq. (29) can also be written as $t_c = (B_L + B_S)\overline{h}$, i.e., cycle time equals total fleet times average headway.

The second equation comes from the fact that small buses have to fulfill off-peak conditions. The relationship between fleet and vehicle capacity obtained for the single period model by Jansson (1980) and Jara-Díaz and Gschwender (2003) do apply in this case; therefore

$$K_S = \frac{l_N Y_N T_N}{L(B_S - tY_N)} \tag{30}$$

Eqs. (25) and (30) reflect the fact that all buses are full in both periods. Combining those equations yields

$$Y_P l_P h_S \frac{B_S - tY_N}{T_N} = Y_N l_N \tag{31}$$

Eqs. (24)–(31) are sufficient to make VRC a function of fleet sizes only, i.e. $VRC(B_L,B_S)$. In Eqs. (32)–(34) we show the value of the resources consumed during each period, VRC_N and VRC_P, and the capital costs of small buses that are assigned to the whole day, VRC_D. The fact that large buses operate during one period only (and, therefore, depreciate at a slower rate) is reflected by the operating costs coefficients that are multiplied times the duration of the period.

$$VRC_N = B_S E_N\left(c_{BO} + \frac{l_N Y_N t}{L(B_S - tY_N)} c_{KO}\right) + Y_N p_w \frac{T_N}{2(B_S - tY_N)} + Y_N p_v \frac{l_N}{L}\frac{B_S T_N}{B_S - tY} \tag{32}$$

$$\begin{aligned} VRC_P = {} & B_L\left(c_{BC} + E_P c_{BO} + Y_P \frac{l_P}{L}\frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)}(c_{KC} + E_P c_{KO})\right) + \\ & B_S\left(E_P c_{BO} + \frac{l_N Y_N t}{L(B_S - tY_N)} E_P c_{KO}\right) + Y_P \frac{p_w}{2} \frac{B_L\left(\frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)}\right)^2 + B_S\left(\frac{l_N Y_N T_N}{l_P Y_P (B_S - tY_N)}\right)^2}{B_L \frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)} + B_S\left(\frac{l_N Y_N T_N}{l_P Y_P (B_S - tY_N)}\right)} \\ & Y_P p_v \frac{l_P}{L} B_L \frac{B_S Y_N l_N T_N - T_P Y_P l_P (B_S - tY_N)}{Y_P l_P (B_S - tY_N)(tY_p - B_L)} + B_S\left(\frac{l_N Y_N T_N}{l_P Y_P (B_S - tY_N)}\right) \end{aligned} \tag{33}$$

$$VRC_D = B_S\left(c_{BC} + \frac{l_N Y_N t}{L(B_S - tY_N)} c_K\right) \tag{34}$$

First order conditions for B_S and B_L yield a system of equations of degree at least 5, such that no analytical solutions are possible. It is clear, however, that $VRC = VRC_N + VRC_P + VRC_D$ achieves a minimum, because it is a continuous positive function that goes to infinity either if the fleets are too large, i.e. when $B_S \to \infty$ or $B_L \to \infty$, or if they are too small, when $B_S \to tY_N$ or $B_L \to tY_P$, which are the lower bounds for the fleet sizes (recall that, as discussed just before Eq. (1), $B = ft_c = fT + Yt$).[5] A numerical approach is used to analyze the optimal values of fleets and vehicle sizes, using the same parameters as in the one-fleet case, shown in the Appendix (representative of Santiago). Fig. 6 represent the optimal values for these variables as a function of the off-peak (6a, with $Y_P = 40,000$) and the peak (6b, with $Y_N = 8,000$) passengers flows; the figure also shows the values of the corresponding vehicle sizes (6c and 6d), showing interesting results including some intuitive ones, e.g. buses added at the peak (the so-called "large" buses) resulted actually larger than those that run all day (the "small" buses).

As expected, the fleet of small buses and their size increase with the off-peak flow, and the fleet of large buses and their size increase with the peak flow, i.e.

$$\frac{\partial B_S}{\partial Y_N} > 0,\ \frac{\partial K_S}{\partial Y_N} > 0,\ \frac{\partial B_L}{\partial Y_P} > 0,\ \frac{\partial K_L}{\partial Y_P} > 0 \tag{35}$$

Additionally, as the off-peak flow increases not only the fleet of small buses increase but also their size. As these vehicles also run at the peak, a larger off-peak flow reduces both the number of large buses and their size.

$$\frac{\partial B_L}{\partial Y_N} < 0,\ \frac{\partial K_L}{\partial Y_N} < 0 \tag{36}$$

Finally, Fig. 6b and 6d also show that the fleet of small buses and their size are practically insensitive to the peak flow, although actually the fleet decreases and vehicles size increases by a very small amount, i.e.

$$\frac{\partial B_S}{\partial Y_P} \approx 0, \text{with} \frac{\partial B_S}{\partial Y_P} < 0;\ \ \frac{\partial K_S}{\partial Y_P} \approx 0, \text{with} \frac{\partial K_S}{\partial Y_P} > 0 \tag{37}$$

The results reflected by Eqs. (35) and (37) indicate that an increase in the peak flow is covered by having more and bigger large buses, and - in a less relevant way - by having fewer and larger small buses, which can be intuitively explained

[5] This shows the advantage of analyzing the case where small buses are assumed to run full at off-peak. Otherwise Eqs. (30) and (31) would not hold, increasing complexity not only in terms of optimality but also in terms of the intuition behind the numerical results. The advantage of this analysis is further explained in Section 4.

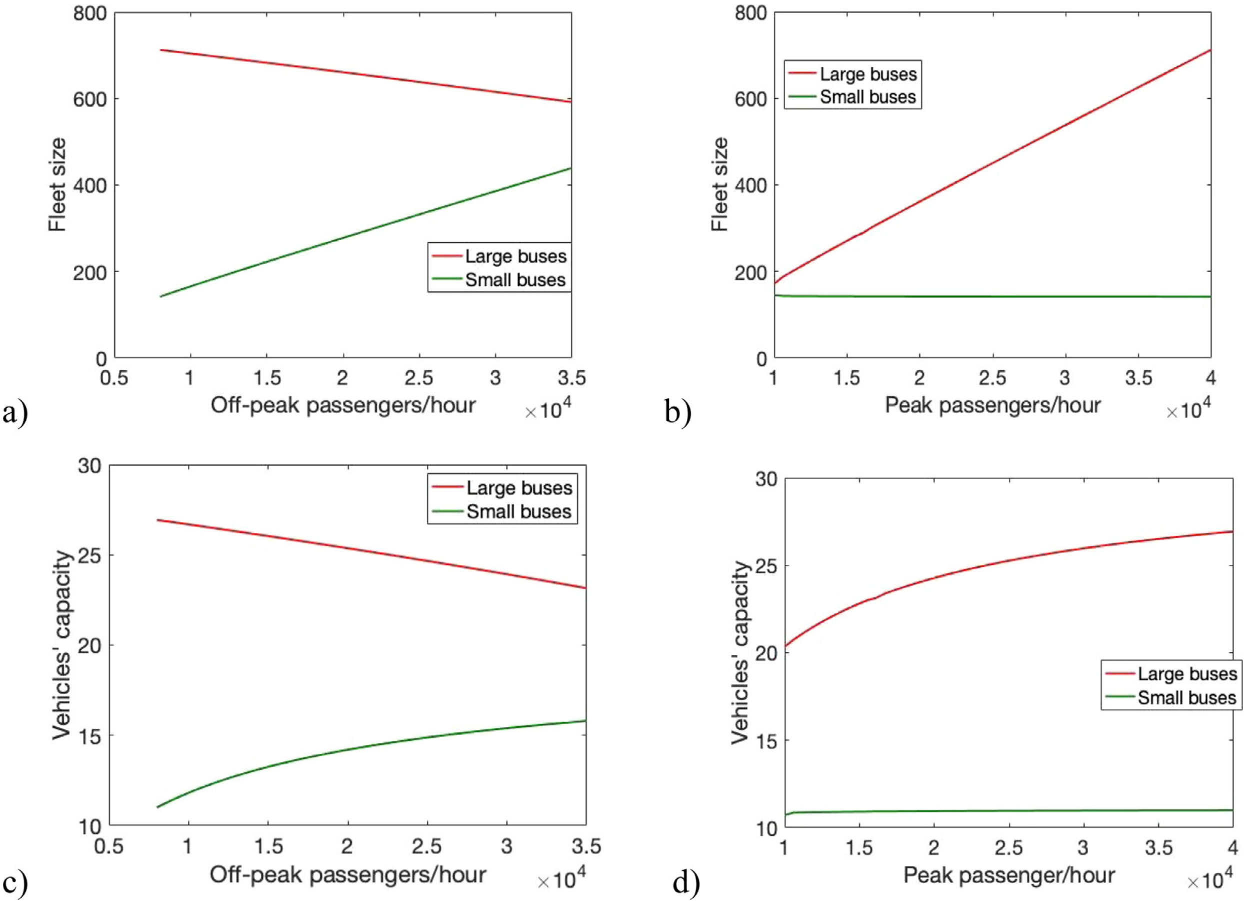


Fig. 6. Optimal fleets and vehicle sizes as a function of flow.

as follows. First, the combination of these changes contributes to increasing the average size of the vehicles: large buses become larger, small buses become larger and the percentage of large buses increases. Second, the model adjusts mostly the large buses in order to avoid an unnecessary impact on the off-peak period, whose conditions remain unchanged. Nevertheless, as the difference between the vehicle sizes of the two types of buses increases, so does the holding time for small buses, which is inefficient; this is why the number of small buses also increases slightly.

The points in which peak and off-peak flows become close in Fig. 6 are also quite interesting. Let us focus on the left side of Fig. 6b, where $Y_P = 10,000$ and $Y_N = 8,000$, but the fleet operating at the peak ($B_S + B_L$) more than doubles the off-peak fleet B_S. This is partially explained by slower buses at the peak due to congestion (peak period's speed is assumed to be ¾ of the off-peak one), and by the time at stops at the peak, commanded by large buses, such that small buses become inefficient because they have to spend time holding. The costs imposed by holding are neatly shown by this seemingly paradoxical situation.

4. One or two fleets? Comparison of the models

In this section we analyze which strategy to face the two-periods design problem is better: the single-fleet design (buses of one size running all periods) against the two-fleets designs, including not only the one developed in Section 4 but also the basic case of one independent fleet per period, whose design corresponds simply to solving twice the classic single period case.

Let us begin comparing the two systems analyzed in this paper, recalling that in the single-fleet case buses may run full or not at the off-peak depending on the relative values of the flows. Numerical results were obtained for $Y_P \in [10,000; 40,000]$ keeping $Y_N = 8,000$, using the value of the parameters shown in the Appendix. By looking at Fig. 3, this means moving within the pink zone, i.e. where buses do not run full at the off-peak, when a single fleet is used. Results show that total costs for the one-fleet system are always larger than the two-fleets case, but by less than 0.7%, as shown in Fig. 7. There, the cost difference diminishes towards the extremes: at the beginning of the curve, both systems yield almost equal

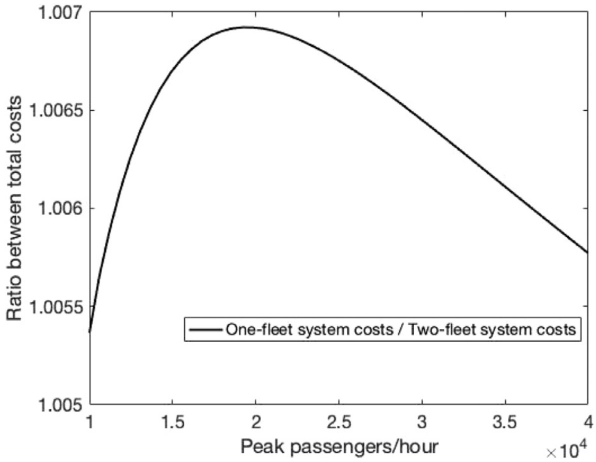


Fig. 7. Ratio between total costs of the two systems as a function of Y_P ($Y_N = 8,000$).

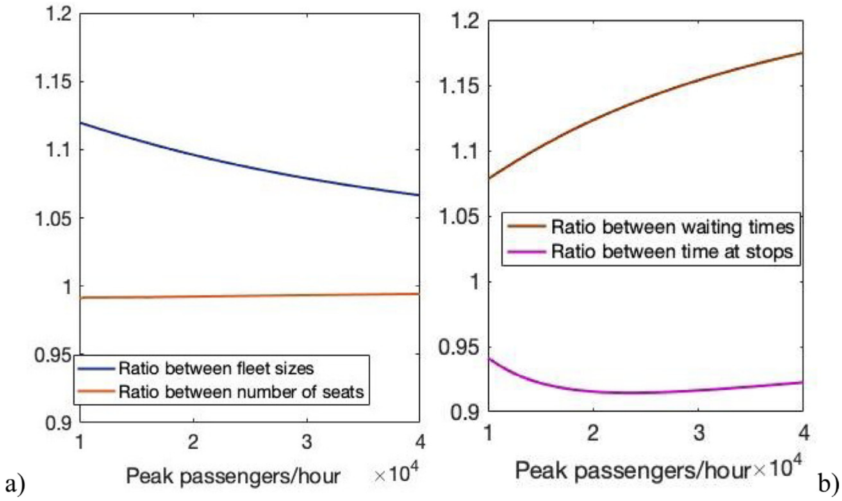


Fig. 8. One-fleet / two-fleet ratios for operating characteristics

costs because peak and off-peak flows are almost equal; when the peak flow is much larger than the off-peak, it is the peak that dominates the cost calculations in both cases.

The small difference in total cost does not provide the full picture of the comparison between the one-fleet versus two-fleets designs. In order to understand better the results shown in Fig. 7, in Fig. 8 we disentangle the differences between the operational characteristics of both strategies; as in Fig. 6, we put the one-fleet characteristics in the numerator and the two-fleet characteristics in the denominator.

Let us view Fig. 8 from the perspective of operators' costs first, recalling that they depend on the fleet size and the number of seats, i.e. B and BK (one-fleet case) or $B_SK_S + B_LK_L$ (two fleets case). Fig. 8a shows that the one-fleet system has to use more buses of a smaller size. Note that technology plays a role here: for instance, if autonomous vehicles are used in the future, the size of the fleet will be less relevant (because of no wages), improving the relative performance of the one-fleet system. Regarding users' costs, in-vehicle time includes time in motion and time at stops; the former ($T_i\frac{l_i}{L}$ in each period i) bears the largest proportion and does not depend on the design decisions, which is why in Fig. 8b we show only the time at stops besides the waiting time ratios. It reveals that the two-fleet design involves less waiting in spite of more time at stops caused by the holding strategy.

In Fig. 9 we expand the analysis done using Fig. 8, by looking at what is happening at each period separately. At the peak, all the vehicles are being used (in both systems), such that we should look only at the components of users' costs. Fig. 9a shows that peak users are much better under the one-fleet system; this is explained because the two-fleets system use less

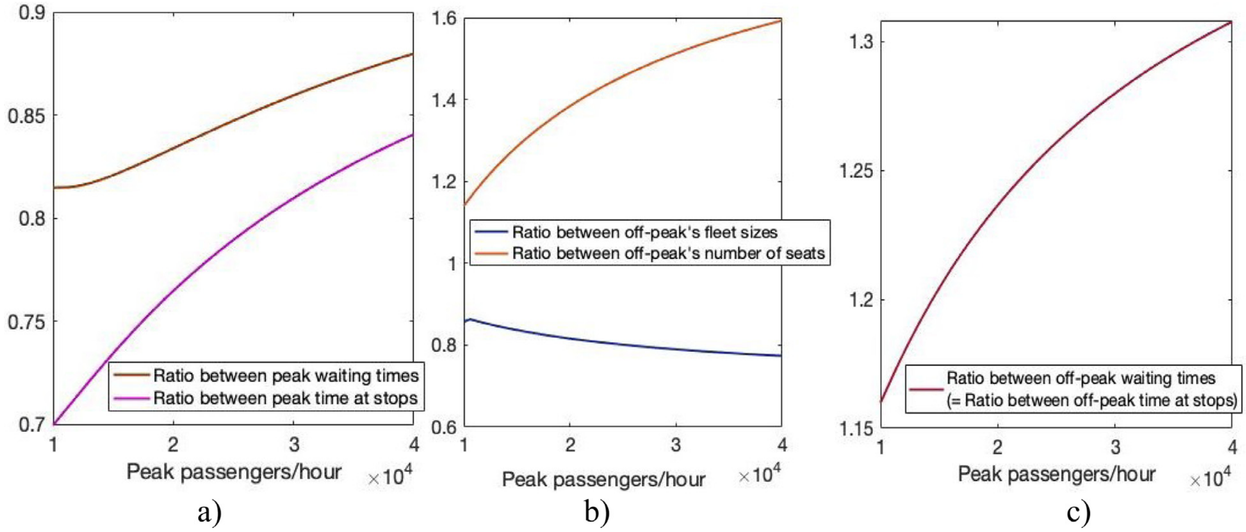


Fig. 9. One-fleet / two-fleet ratios for peak and off-peak indices

vehicles (decreasing the frequency), and spend more time at stops due to the holding strategy (that forces all vehicles to behave as the very large ones).

Regarding the off-peak, Fig. 9b shows that operators' characteristics are reversed with respect to Fig. 8a, i.e., the one-fleet system presents a smaller fleet of larger vehicles. Smaller vehicles are explained by the flexibility of the two-fleets system (as this type of vehicles is not covering the whole demand at the peak), which induces a larger fleet; this, in turn, explains the lower waiting time and time at stops (Fig. 9c) of the two-fleets system, where only one curve is needed as waiting time is $\frac{1}{2f_{Ni}}$ and the time at stops is $\frac{Y_N l_N}{f_{Ni} L}$ for both systems i such that the ratios are identical.

Let us synthesize this analysis by stating that both systems present total costs that are very similar; nevertheless, they work under different conditions for operators and users. Peak users are favored by the one-fleet system, and off-peak users would prefer two fleets. On the other hand operators should provide more buses but less seats if the one-fleet system is used. This implies, for instance, that if there were reasons to assign peak users a larger weight than off-peak ones, decision makers might want to choose a single fleet rather than the slightly better two-fleets system; alternatively, a larger weight on off-peak users would increase the relative advantages of the two fleet-system. It is relevant to recall here that, as proven by Jara-Diaz and Gschwender (2009), a budget constraint imposed on the transit system design results in different weights on users and operators costs, inducing a smaller than optimal fleet of larger than optimal size, modifying the comparison between strategies.

As advanced in the previous Section, one could consider that the small buses do not run necessarily full at the off-peak, i.e. they could be a bit larger, aiming at the reduction of the holding time at the peak, which would increase substantially the mathematical complexity of the model, diminishing intuition. However, as such model lies in-between the two models studied in this paper, we can conclude that if it was compared against the one-fleet model, the curves in Figs. 8 and 9 would keep the same signs but would be closer to 1; the tight differences between the operational characteristics revealed in Figs. 8 and 9 imply that total costs would diminish just slightly.

What about the two-fleet system optimizing each period independently? The system that optimizes each period in isolation does not take advantage of the obvious scope economies associated to running one bus during two periods, but fleets can be adapted to demand exactly. In Fig. 10 we compare the total costs of the three designs using the two-fleets system with holding as the reference to construct the ratios (same parameters as in Fig. 7): "independent periods" system (grey line) and the one-fleet system (green line). As expected, the independent periods design yields the worst results, although differences are quite small for the three systems (less than 1%). As the time spent by passengers in the buses in-motion (i.e. not in the stops) is independent of the design, it can be subtracted to highlight the differences, which yields that optimizing periods independently can increase costs by about 5%. Numerical analysis shows that this difference is explained mostly by operators' costs (losing economies of scope); actually, users' costs are slightly lower in the system that optimizes each period in isolation. The degree of economies of scope (defined as the proportion saved by serving both demands jointly instead of independently) can be as high as 0.018 for 10,000 pax/hr at the peak.

5. Sensitivity analysis and the role of other elements

A sensitivity analysis regarding selected variables is shown in Fig. 11, including period duration, running time, trip and line length, boarding time, and values of time and operators' costs. The first global conclusion is that the results presented

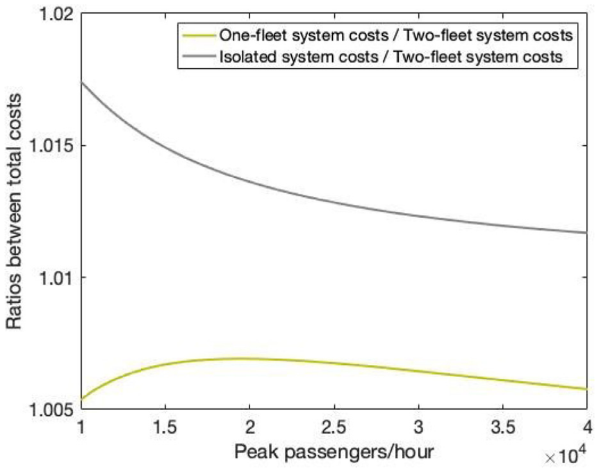


Fig. 10. Comparison of the three systems.

earlier are robust in the sense that two-fleets operation remains superior to one fleet, but by a very little margin. The second global conclusion is that in the neighborhood of the chosen Santiago-like parameters the effects are monotonic, always increasing or decreasing the small advantages of the two-fleet strategy. So it is worth analyzing the specific effects of each parameter in order to gain intuition regarding the operation of both systems.

The first row in Fig. 11 shows the effect on the cost ratio (one-fleet/two-fleet) as three parameters grow: length of the peak period (shorter than the off-peak), running vehicle time at the off-peak (shorter than at the peak, due to streets congestion), and average length of users' trips at the off-peak (shorter than at the peak). Then in all three cases the growth of the corresponding parameter means that their difference between the two periods diminishes, such that the resulting monotonic reduction in the cost ratio is rather intuitive.

The fourth figure shows that the advantage of the two-fleets system increases with route length. This happens because the time at stops becomes less relevant, i.e. the disadvantage of this system due to the holding diminishes. This also explains the effect of in-vehicle time value in the sixth figure, as it increases with discomfort on board such that the time spent holding weighs more, increasing users' costs of the two-fleets system. We have already shown that the one-fleet system presents higher waiting times, which explains the effect of the waiting time value (sixth figure) directly.

The seventh and eighth figures deal with some of the operators costs' parameters. As explained above, the one-fleet system use more buses, which is why it responds worse to an increase on the capital cost per bus. Although the one-fleet system has a lower total number of seats, it also responds worse when the operating cost per seat increases, which highlights the advantages of keeping the large buses out of the streets during the off-peak. The last figure deals with boarding and payment technology, that influences the value of t - the average boarding-alighting time - as studied by Jara-Díaz and Tirachini (2013); the figure reveals that when t increases, the time spent holding also increases, affecting mostly the two-fleets system.

Finally, it is worth looking at the comparison of total costs when flows are not as large as in a corridor but distributes along more than one street or avenue, as in this case observed frequencies will be lower and the two-fleet design should be more advantageous. In Fig. 12 we show the numerical results obtained for $Y_P \in [4,000;\ 20,000]$ keeping $Y_N = 2,000$ and using the value of the parameters shown in the Appendix. This means moving along a horizontal line close to the bottom of Fig. 3 within the pink zone when a single fleet is used (buses do not run full at the off-peak). In terms of total costs, the advantage of the two-fleets system remains and increases slightly regarding the large flows case, but varying between 0.72% and nearly 1.1 %. Again, the cost difference diminishes towards the extremes: when peak and off-peak flows are closer, and when the peak flow dominates the cost calculations in both cases. The main conclusions remain.

As discussed earlier, the single-line model has been used as a stylized representation of a transit system that permits an analytical formulation of the basic relationships that command that system, in order to study some structural properties. As shown in Jara-Diaz and Gschwender (2003), including all elements or detailed technical relationships make it difficult to focus on specific basic aspects of transit design, as is the case when analyzing two periods, were some variables are not subject to the designer's scrutiny or left out of the description. We offer here a discussion on how they could impact the results obtained.

- Optimization of the bus stops: some models (e.g. Mohring, 1972) optimize the distance between bus stops, in order to find the best equilibrium between buses' commercial speed and users' walking times. For a given strategy, our results hold as long as the same stops are used during both periods, which is the usual case in most systems in the world.
- Crowding: Jara-Díaz and Gschwender (2003), as well as Hörcher and Graham (2018), include the discomfort induced by crowding in their single-line, single-period models, by multiplying the in-vehicle unitary costs p_v by $1 + \theta k/K$, where

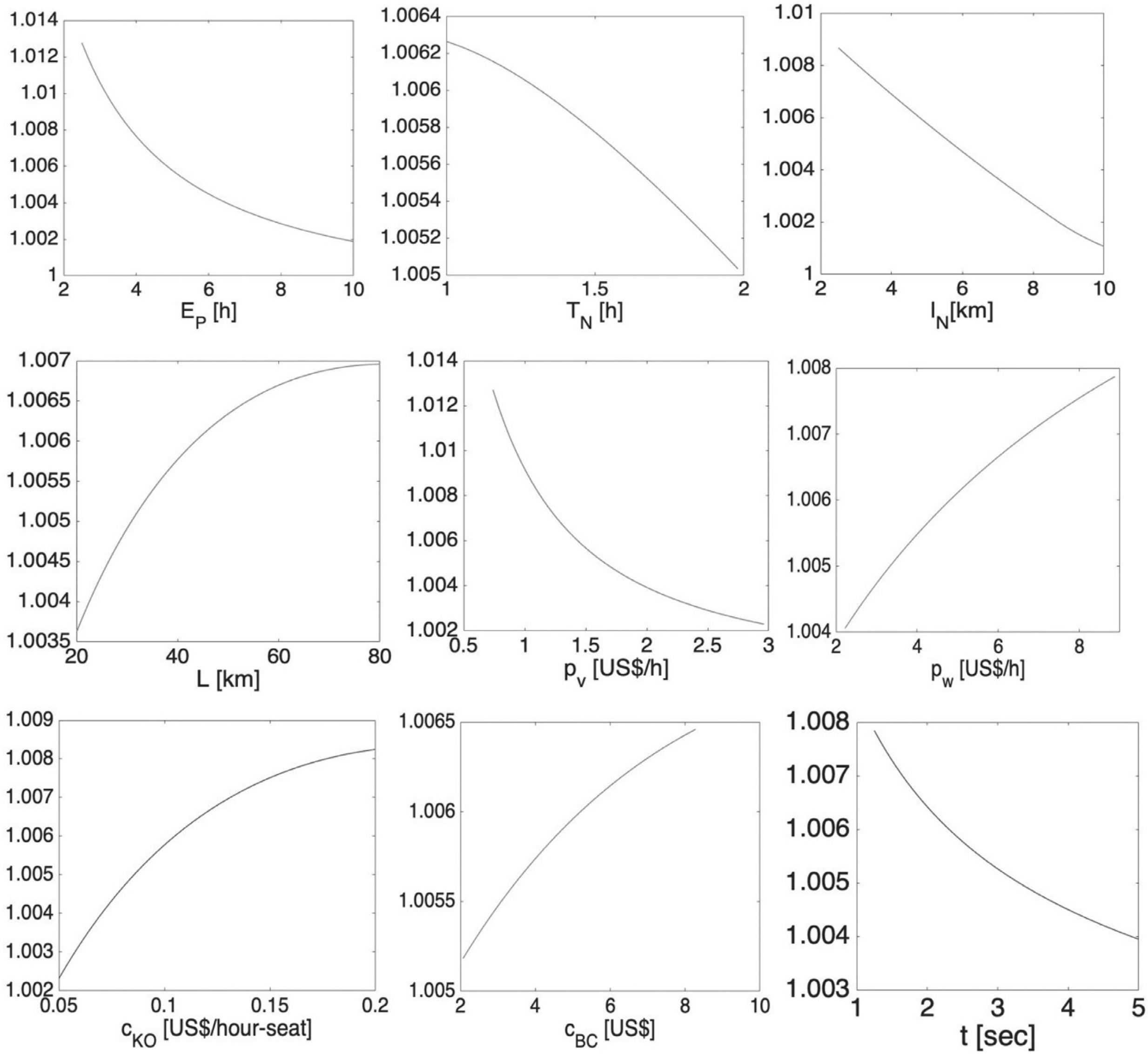


Fig. 11. Total costs of the one-fleet system / total costs of the two-fleets system, for variations of selected parameters around the values used.

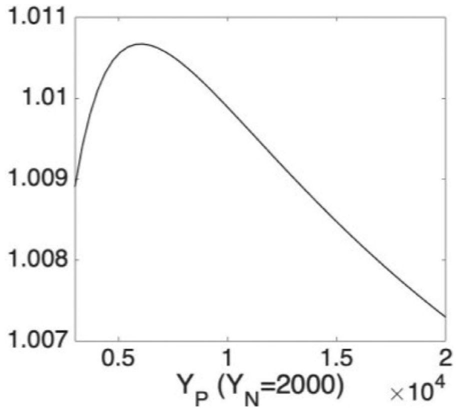


Fig. 12. Ratio between total costs (one-fleet / two-fleets) as a function of Y_P ($Y_N = 2,\ 000$).

k is the actual load of the vehicles and θ is another parameter. As expected, vehicle size is affected depending on the relative values of θ and the corresponding operators' cost parameter, which could yield an inactive capacity constraint depending on the relative value of θ and the marginal effect of K on operators' costs. This could happen (equations 4) in the two-periods case for the single-fleet strategy; if at least one of the constraints is active (which is expected, as buses usually run full during the peak), the analyses studied here remains valid. Under the two-fleets strategy vehicles run full at both periods such that introducing discomfort by crowding will likely affect sizes of both types of vehicles. The relative variation of total costs under the two strategies is likely to depend, again, on θ and the marginal effect of K on operators' costs.

- Spatial imbalances: besides the temporal differences studied here, there are differences between passenger flows in opposite directions of the transit line. Including both in the analysis is a quite complex task. For example, Hörcher and Graham (2018) use the classical single-line model adjusted for a simple (spatially) unbalanced scheme. To obtain conclusions they have to omit (for simplicity) time at stops, which is crucial for our analysis; their omission simplify matters as it makes the quadratic in flow term disappear because that term is commanded by boarding-alighting time (see our Eq. 1). With flow imbalance fleet size and vehicles' capacity will be influenced not only by Y_P and Y_N but also by the maximum load at each period. However, waiting and in-vehicle costs excluding time at stops still depend on the total number of passengers and not on their distribution; time at stops increases when the demand is unbalanced, because passengers in the most loaded direction (the majority) would spend more time at stops and the others would spend less. As time at stops is particularly relevant in the two-fleets strategy (due to the holding and the presence of larger vehicles), it is expected that this strategy becomes more affected.

6. Synthesis and conclusions

In this paper we analyzed and compared two strategies to design frequencies and bus sizes for a single-line scheme that faces two periods, peak and off-peak. One strategy that uses a single fleet - in which only vehicles of one size serve both periods -, and another strategy that uses two fleets - in which small vehicles operate during both periods, complemented by larger vehicles during the peak.

The one-fleet strategy was based on the analysis of the single-line two-periods model introduced by Jara-Diaz et al. (2017), where bus size was determined by the peak. It was completed here by adding the case in which bus size is determined by the conditions in the off-peak period in spite of a demand that is lower than in the peak by definition. We concluded that buses always have to run full at the peak but not always at the off-peak; the conditions under which each case holds were analyzed. Numerical and analytical results show that previous findings from Jara-Díaz et al. (2017) where K was optimized as well, hold also in this new context:

- Off peak frequency is lower and peak frequency is larger than in the respective isolated cases; the size of the vehicles lies in between those of both cases.
- Peak frequency increases if off-peak flow increases, but off-peak frequency decreases if peak flow increases; the more similar the flows, the larger the frequencies.

The two-fleet strategy exhibits more flexibility to respond to the differences between flows. However, two vehicle sizes at the peak imposed a new difficulty: different time at stops waiting for passengers to board and alight. To face this problem and avoid having different cycle times, a holding strategy was imposed, namely that the smaller vehicles should wait at the bus stops (after passengers board the vehicle) a lapse of time that makes the time at stops equal to that of the larger vehicles.

Explicit equations involving the design variables were establish and manipulated in order to express the value of the resources consumed as a function of both fleets. The analysis revealed that if the off-peak flow increases, more and larger small buses are necessary (they are the only ones serving the off-peak flow), which yields less and smaller large buses to serve the same peak flow together with the more small vehicles available. If the peak flow increases, average vehicle size increases by having larger buses (for both categories) and by increasing the percentage of large buses, which is achieved by having more large vehicles and less small vehicles; changes in the small vehicles are mild, such that the off-peak service (whose flow is unchanged) remains similar. This small change happens to be better than no change at all, because it controls the increase in holding time.

Comparison between these two systems showed that the two-fleets strategy with holding is slightly superior. Although they yielded very similar total costs, they are caused by different operational attributes. The one-fleet system is better for peak users and worse for off-peak users, while presenting a larger fleet with less total seats (smaller vehicles in average). Regarding operators, their cost is smaller with two fleets, such that a financial constraint could reinforce this strategy (Jara-Diaz and Gschwender, 2009). These differential effects across the different agents involved are relevant for decision makers, which could assign different importance to the different groups. Examining the effect of financial constraints and weights assigned to different users is indeed something worth exploring in the future.

These systems were also compared with a design based upon the optimal operation of each period independently which happened to be the worst option, mainly because of larger operators' costs. This intuitive fact verifies the existence of scope economies regarding heterogeneity between peak and off-peak periods. A sensitivity analysis on various parameters (e.g.

route and trip lengths, boarding time, time values and operating costs) and on the flow levels revealed that the results and conclusions obtained are quite robust.

The results obtained in Jara-Díaz et al. (2017) and here reveal several structural effects induced by considering the differences in passengers' flows, distance travelled and vehicle speeds between peak and off-peak periods. As this has been done on a single line, a relevant direction for future research is studying these issues on a network. By doing so, an important methodological difference will be the emergence of line structures (the spatial arrangement of lines) as a design variable, that could or could not stay unchanged across periods. As that design is much more complex (as shown, for example, by Borndörfer et al., 2007) than single-dimensioned continuous variables like frequencies, it is worth studying second-best solutions, namely optimizing one-period in isolation and adapting the other period to that solution.

Finally, the strategic results presented here have been obtained assuming that vehicle size K is a continuous variable. As buses are available in a limited number of sizes (say small, medium size, large and very large) it might well be the case that two fleets operating during both periods becomes the best strategy, which is yet another path to explore.

CRediT authorship contribution statement

Sergio Jara-Díaz: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Andrés Fielbaum:** Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Antonio Gschwender:** Formal analysis, Methodology, Validation, Visualization, Writing - review & editing.

Acknowledgements

This research was partially funded by Fondecyt, Chile, Grant 1200157, and Conicyt, Chile, Grant PIA-Basal AFB180003. We are grateful for the comments of three unknown referees that helped us improve the exposition of the technical aspects of the paper. Remaining errors are our responsibility.

Appendix. Parameters for simulations

Parameter	Value
E_P	5 hours
E_N	13 hours
T_P	2 hours
T_N	1.5 hours
t	2.5 seconds
l_P	10 km
l_N	5 km
L	40 km
c_{BC}	4.14 US$
c_{KC}	0.45 US$/seat
c_{BO}	1.32 US$/hour
c_{KO}	0.1 US$/hour-seat
p_v	1.48 US$/hour
p_w	4.44 US$/hour

References

Basso, L., Silva, H., 2014. Efficiency and substitutability of transit subsidies and other urban transport policies. Am. Econo. J. 6 (4), 1–33.
Beirão, G., Cabral, J.S., 2007. Understanding attitudes towards public transport and private car: a qualitative study. Transp. Policy 14 (6), 478–489.
Borndörfer, R., Grötschel, M., Pfetsch, M., 2007. A column-generation approach to line planning in public transport. Transp. Sci. 41 (1), 123–132.
Chang, S., Schonfeld, P., 1991. Multiple period optimization of bus transit systems. Transp. Res. Part B 25 (6), 453–478.
Daganzo, C., 2009. A headway-based approach to eliminate bus bunching: systematic analysis and comparisons. Transp. Res. Part B 43 (10), 913–921.
De Borger, B., Mayeres, I., Proost, S., Wouters, S., 1996. Optimal pricing of urban passenger transport: a simulation exercise for Belgium. J. Transp. Econ. Policy 30 (1), 31–54.
Evans, A.W., Morrison, A.D., 1997. Incorporating accident risk and disruption in economic models of public transport. J. Transp. Econ. Policy 31, 117–146.
Fernández, E., De Cea, J., De Grange, L., 2005. Production costs, congestion, scope and scale economies in urban bus transportation corridors. Transp. Res. Part A 39 (5), 383–403.
Friman, M., Edvardsson, B., Garling, T., 1998. Perceived service quality attributes in public transport: inferences from complaints and negative critical incidents. J. Public Transp. 2 (1), 4.
Glaister, S., Lewis, D., 1978. An integrated fares policy for transport in London. J. Public Econ. 9 (3), 341–355.
Hörcher, D., Graham, D.J., 2018. Demand imbalances and multi-period public transport supply. Transp. Res. Part B 108, 106–126.
Jansson, J.O., 1980. A simple bus line model for optimisation of service frequency and bus size. J. Transp. Econ. Policy 14, 53–80.
Jansson, J.O., 1984. Transport System Optimization and Pricing. Wiley, Chichester.
Jara-Díaz, S., Fielbaum, A., Gschwender, A., 2017. Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. Transp. Res. Part A 106, 65–74.
Jara-Díaz, S., Gschwender, A., 2003. Towards a general microeconomic model for the operation of public transport. Transport Reviews 23 (4), 453–469.
Jara-Díaz, S., Gschwender, A., 2009. The effect of financial constraints on the optimal design of public transport services. Transportation 36 (1), 65–75.
Jara-Díaz, S.R., Tirachini, A., 2013. Urban bus transport: open all doors for boarding. J. Transp. Econ. Policy 47, 91–106.
Mohring, H., 1972. Optimization and scale economies in urban bus transportation. Am. Econ. Rev. 62, 591–604.

Newell, G.F., 1974. Control of pairing of vehicles on a public transportation route, two vehicles, one control point. Transp. Sci. 8 (3), 248–264.
Oldfield, R.H., Bly, P.H., 1988. An analytic investigation of optimal bus size. Transp. Res. 22B, 319–337.
Osuna, E., Newell, G., 1972. Control strategies for an idealized public transportation system. Transp. Sci. 6 (1), 52–72.
Parry, I., Small, K., 2009. Should urban transit subsidies be reduced? Am. Econ. Rev. 99 (3), 700–724.
Proost, S., Van Dender, K., 2008. Optimal urban transport pricing in the presence of congestion, economies of density and costly public funds. Transp. Res. Part A 42 (9), 1220–1230.
Redman, L., Friman, M., Gärling, T., Hartig, T., 2013. Quality attributes of public transport that attract car users: a research review. Transp. Policy 25, 119–127.
Tirachini, A., Hensher, D., Jara-Díaz, S., 2010. Comparing operator and users costs of light rail, heavy rail and bus rapid transit over a radial public transport network. Res. Transp. Econ. 29 (1), 231–242.
Walker, J., 2012. Human Transit. Island Press, Washignton DC.