Lines spacing and scale economies in the strategic design of transit systems in a parametric city

Fielbaum, Andrés; Jara-Diaz, Sergio; Gschwender, Antonio

DOI
10.1016/j.retrec.2020.100991

Publication date
2020

Document Version
Final published version

Published in
Research in Transportation Economics

Citation (APA)
Fielbaum, A., Jara-Diaz, S., & Gschwender, A. (2020). Lines spacing and scale economies in the strategic design of transit systems in a parametric city. *Research in Transportation Economics*, *90 (2021)*, Article 100991. https://doi.org/10.1016/j.retrec.2020.100991

Research in Transportation Economics 90 (2021) 100991

Contents lists available at ScienceDirect

Research in Transportation Economics

journal homepage: http://www.elsevier.com/locate/retrec

Research paper

Lines spacing and scale economies in the strategic design of transit systems in a parametric city.

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

[a] Universidad de Chile and Instituto Sistemas Complejos de Ingeniería (ISCI), Chile
[b] Delft University of Technology, the Netherlands

ARTICLE INFO

JEL classification:
R40
R48
R49
L91
L92

Keywords:
Transit lines spacing
Parametric city
Scale economies
Lines structures
Optimal
Designd

ABSTRACT

In this paper, we incorporate the spacing of transit lines in addition to frequencies, vehicle sizes and routes in both the design and the analysis of scale economies in transit systems. First, we present a way of looking at lines spacing in a simple parallel-lines-model whose properties regarding optimal design and scale economies are derived. Then we introduce this concept of spacing into the parametric description of a city - that permits the representation of different degrees of mono and polycentrism - in order to analyze the choice between basic strategic lines structures as feeder-trunk, hub-and-spoke or direct services, where lines spacing is optimized jointly with frequencies, vehicle sizes and routes of all lines involved. We show that (a) there is a link between optimal spacing and frequency such that waiting and access costs are equal; (b) the inclusion of spacing increases the range of demand volumes where transit networks that include transfers are preferred; (c) the degrees of mono and polycentrism influences optimal spacing; and (d) introducing spacing increases the degree of scale economies.

1. Introduction

At a strategic aggregate level, the design of urban public transport systems involves the identification of routes, and the calculation of frequencies and vehicle sizes of all lines that form the transit network. Designing the routes of public transport lines is one of the most relevant and complex aspects of transit design. Real cities present a very large number of streets of different hierarchy and many zones involving different activities. Finding optimal line structures in real cities is an NP-hard problem,[1] which makes the search for an appropriate generic structure of lines a very important step prior to the detailed design of the system. Although heuristics and previous applied experience have played an important role in this search, it has been shown that the comparative analysis of preconceived strategic line structures can be quite rewarding, e.g. transit lines spatially organized as feeder-trunk, hub-and-spoke or direct services. Under this approach, the idea is to identify possible strategic lines structures and to study which one has the optimal response for different city models taking into account all resources involved including those contributed by the operators (e.g. fleet, labor, energy) and those contributed by the users (their time). Several authors have proposed different ways of approaching this problem, such as Byrne (1976), Jara-Díaz and Gschwender (2003a), Daganzo (2010), Badia et al. (2014), Chen et al. (2015), Gschwender et al. (2016) or Fielbaum et al. (2016).

The most appropriate arrangement of lines in an urban space from a strategic viewpoint depends on some characteristics of the city, on the volume and distribution of demand in space, and on the characteristics of the transport modes considered. Therefore, one key element in the process of finding the most appropriate strategic line structures is the way in which the city is described, including not only its network topology but also its demand pattern. Simple graphs have been very useful to link and explore these aspects. Díaz et al. (2002), for instance, inspired a huge change in the structure of the lines in Santiago, Chile, in 2007, by using a Y-shaped model of a city where they compare a feeder-trunk system with a direct one, which was refuted later on by Gschwender et al. (2016). Quadrifoglio and Li (2009) and Jara-Díaz and

* Corresponding author.
E-mail address: jaradiaz@ing.uchile.cl (S. Jara-Diaz).

[1] A combinatorial problem is said to be NP-Hard if there is no algorithm (up to current knowledge) that can solve it optimally in a reasonable time when the size of the input is large (as a real-life transit network). Several specifications of the problem of finding optimal lines structures have been proven to be NP-Hard (see, for instance, Borndörfer et al., 2007).

https://doi.org/10.1016/j.retrec.2020.100991
Received 3 June 2020; Received in revised form 5 September 2020; Accepted 12 October 2020
Available online 21 October 2020

Gschwender (2003a) also used a simple-graph perspective to study this type of problem, while Fielbaum et al. (2017) proposed a not-so-simple graph that has topological properties which were shown to be closer to those of real cities and permits representing different centers structures within a city in a parametric form. Using this representation of a city, Fielbaum et al. (2016) introduced the decision on lines structure as part of the public transport design problem besides frequencies and vehicle sizes for each of the lines in the system. Later on, Fielbaum et al. (2020) showed that, as patronage grows keeping trip distribution constant, the lines structure evolves increasing directness, i.e. the system has to adapt by diminishing distance traveled, number of transfers and number of stops, such that the degree of scale economies increases when lines structure changes (decreasing afterward). But even in this latter case, there is one important limitation that is troublesome: streets are inevitably aggregated because of the simplicity of the network representation, i.e., each arc is actually representing many real arcs. This leads to models that - by construction - ignore the access to the transit network as a variable and treat every arc as a collector of many lines that actually run along different parallel streets. This is not a minor issue, as it also has an evident impact on the calculation of waiting times which are underestimated because the modeled frequencies are the sum over possibly many lines that run in parallel (which, in turn, affects the other design variables). The different strategic line structures are affected in different ways by this omission; for example, structures that involve more transfers may be favored. The impact of lines spacing on waiting times and transfers in addition to the introduction of access time as a new element has also an effect on the analysis of scale economies in transit networks. Our challenge, then, is to incorporate spatial spacing of transit lines as a design variable in a model based on an adequate representation of the city without going into a highly-detailed description of its network.

Transit lines spacing has been indeed introduced as a design variable in transit analysis. Hurdle (1973), Kocur and Hendrickson (1982) and Chang and Schonfeld (1991), have proposed different models aimed at optimizing frequency and spacing between identical parallel transit lines, which introduces access time users' cost besides waiting time costs. Later on Daganzo (2010), Badia et al. (2014) and Tirachini et al. (2010) introduced line spacing or angular distance as one of the variables to optimize on an urban space described either as a regular grid or as a radial grid in a monocentric city. As shown by Fielbaum et al. (2017), these regular models of the underlying urban spaces exhibit topological properties that are far from those present in real cities and, therefore, are not particularly appropriate for the analysis of transport systems design at a strategic level where avenues tend to define the relevant network.[2] Besides, the specific impact of introducing lines spacing - that affects all other design variables - on scale economies is not studied.

In this paper, we contribute to understanding this topic in two steps. First, in Section 2 we formulate and solve a simple parallel transit lines model that incorporates both a cycle time that is sensitive to boarding and alighting, and operators' cost depending on vehicle size, analyzing scale economies. Three results are obtained: (i) within each line optimal frequency and vehicle size follow the same rule as in a one-line model, keeping its properties; (ii) optimal frequency is proportional to optimal density for all levels of patronage, in such a way that the average waiting time cost has to be equal to the average access time cost; and (iii) the simultaneous adjustment of lines spacing and frequency enlarges the degree of scale economies.

The second step - and main contribution of this paper - is to expand the analysis of lines spacing towards the design of transit networks (lines structures) in a general city. For this purpose, in Section 3 we use the parametric representation of a city introduced and applied by Fielbaum et al. (2016, 2017; 2020); as explained above, this model was originally constructed with an aggregate description of streets and, therefore, we now recognize that each arc represents a limited number of parallel streets or avenues such that a new design variable emerges, lines spacing, involving the number of streets that will be used by parallel transit lines. This new variable is optimized jointly with frequencies and vehicle sizes of all lines involved in the design. From this analysis, we get three main conclusions: (i) as patronage increases, optimal spacing tends to diminish while the best lines structures evolve from hub-and-spoke through feeder-trunk and a direct structure with no transfers, towards a set of OD specific lines with no intermediate stops; however, the introduction of spacing as a design variable postpones the emergence of more direct structures, which now occurs for larger demand volumes; (ii) the equality between the costs of access and waiting times is shown to persist; and (iii) adapting lines spacing increases the degree of scale economies. As explained, the analysis presented here requires the transit network to have design flexibility as patronage increases, which makes it more appropriate for bus systems than for subway systems, where infrastructure investment is essentially irreversible. For this reason, the models in subsequent sections will be referred to as bus related.

1.1. Lines spacing as a design variable: a simple parallel lines model

In the basic single-line models (Mohring, 1972; Jansson, 1980; Jara-Díaz & Gschwender, 2009), several design variables like frequencies and bus sizes are optimized to minimize the value of the resources consumed by operators and users. This is done on a circular line where passengers are evenly distributed.

In order to analyze the introduction of lines spacing as a design variable, we will first develop a single-line model serving a single origin-destination (OD) pair. This will be extended to optimize spacing at the origin of a number of parallel lines in a model shaped after Kocur and Hendrickson (1982) and Chang and Schonfeld (1991).

Let us consider a single line where buses collect Ypassengers per unit of time from one origin towards a single destination. The value of the resources consumed (VRC) includes the acquisition and operation of a fleet of B buses of size K (assumed to be continuous) and the values of waiting and in-vehicle times $t_w$and t_v respectively. The design variables are B, $K$and the frequency f of the line.

$$VRC = B(c_0 + c_1 K) + p_w Y t_w + p_v\ Y t_v \tag{1}$$

The known parameters are c_0 and c_1 representing operators costs; p_w and p_V representing the values of waiting and in-vehicle times respectively. Also, T is vehicle time in motion needed to run the line in one direction, and $t$is the time needed by each passenger to board and alight from a vehicle ($t/2$ in each case), such that average in-vehicle time t_v is obtained by adding two terms: in motion time of passengers (T), and average in-vehicle time at the destination while other passengers are alighting[3] ($\frac{tY}{4f}$), yielding $t_v = T + \frac{tY}{4f}$. The arrivals of buses and passengers at the origin are assumed to be evenly distributed, such that the average waiting time is $t_w = \frac{1}{2f}$. Fleet size is given by $B = ft_c$, where the cycle time t_c is given by $t_c = 2T + \frac{tY}{f}$. Finally, vehicle size has to bear the load $\frac{Y}{f}$, such that $K = \frac{Y}{f}$. Replacing these in equation (1) and optimizing yields:

$$f^* = \sqrt{\frac{Y}{2Tc_0}\left(\frac{p_w}{2} + tY\left(\frac{p_v}{4} + c_1\right)\right)} \quad \text{and} \quad K^* = \sqrt{2Tc_0 Y\left(\frac{p_w}{2} + tY\left(\frac{p_v}{4} + c_1\right)\right)^{-1}} \tag{2}$$

[2] The intuition behind the bad topological performance of some popular regular representations of a city network lies in elements as the lack of arc hierarchy (grid) or structures that fail for relatively large cities (monocentric).

[3] The second term is built as the average between the passengers that face the best situation (zero time for the first alightment) and the worst situation (the time needed by all passengers to alight, for the last alightment).

which closely resembles results previously obtained by, for example, Jara-Díaz and Gschwender (2009) in a circular single line where users are evenly distributed.

Plugging back expressions (2) into (1) gives us the cost function for this one-line simple system. Using this cost function, it can be easily shown that the degree of scale economies *DSE* (defined as the ratio between the average and the marginal costs) has the generic form found in Fielbaum et al. (2020), i.e. $DSE = 1 + \frac{\beta}{2\alpha Y + \beta + 2\varepsilon Y\sqrt{\alpha + \beta/Y}}$, with $\alpha = Tc_0(8c_1 + 2p_v), \beta = 4Tc_0p_w$ and $\varepsilon = c_0t + p_vT + 2Tc_1$. This means that scale economies are always present ($DSE > 1$) but get exhausted eventually, with *DSE* decreasing asymptotically to 1.

Let us extend this model by considering a rectangular area of width P and length Las shown in Fig. 1, where Ypassengers are homogeneously distributed along Pat the top and travel to a faraway point where a number of vertical parallel lines equally spaced converge (which means that egress costs do not enter the design problem); this way there are $y = Y/P$ users per time-width unit. The spacing between lines is R such that its inverse represents the number of lines per unit width, the spatial density D. The decision (design) variables are the spatial separation of the lines ($R = \frac{1}{D}$) and the frequency of each line (f), considering that each passenger uses the closest line, as represented in Fig. 1, in which $\frac{1}{2}\frac{y}{D}$ passengers board each line from each side, such that $\frac{y}{D} = \frac{Y}{PD}$ is the patronage per hour of each line.

Let us formulate operators' and users' costs. As transit lines run in parallel, we can work either with R or D. To calculate the value of the resources consumed per unit of time and width, it is necessary to express the different components of *VRC* as a function of the decision variables (f and D): cycle time t_c is given by $t_c = 2T + t\frac{y}{fD}$. Then $B = t_cfD$ vehicles per unit width; $K = \frac{y}{fD}$ (passengers aboard each vehicle); average in-vehicle time $t_v = T + \frac{t}{4}\frac{y}{fD}$ (individual alighting time is $\frac{t}{2}$); average waiting time $t_w = \frac{1}{2f}$; and, following Fig. 1, average access time $t_a = \frac{1}{4Dv_a}$ with v_a the walking speed.

Then the value of the resources consumed by operators and users per unit of time and width is

$$VRC = 2c_0TfD + c_0ty + 2c_1yT + \frac{c_1ty^2}{fD} + p_a\frac{y}{4Dv_a} + p_w\frac{y}{2f} + p_vy\left(T + \frac{t}{4}\frac{y}{fD}\right) \tag{3}$$

Equation (3) shows the effect of each of the two decision variables very clearly. It is evident that analytically D acts very similar to f, as they always appear as a product with the exception of the two terms dealing with their direct specific effects on access time (D) and waiting time (f). Therefore, D is yet another source of scale economies, just as f generates the so-called Mohring effect.

Now we can get the first-order conditions in f and D for *VRC* to be a minimum. Taking the first derivative in (3) with respect to f, making it equal to zero and multiplying times f^2D yields:

$$2c_0Tf^2D^2 - c_1ty^2 - p_w\frac{y}{2}D - p_v\frac{t}{4}y^2 = 0 \tag{4}$$

This implies that

$$f^* = \sqrt{\frac{\frac{y}{D}}{2Tc_0}\left(\frac{p_w}{2} + t\frac{y}{D}\left(\frac{p_v}{4} + c_1\right)\right)} \quad \text{and} \quad K^* = \sqrt{2Tc_0\frac{y}{D}\left(\frac{p_w}{2} + t\frac{y}{D}\left(\frac{p_v}{4} + c_1\right)\right)^{-1}} \tag{5}$$

Note that the same form of equations (2) is replicated here, with y/D replacing Y: within each of the parallel lines of this model, the relationship between frequency (and K) and the number of passengers is the same as in the single-line model. This is an interesting novel result: each of the lines (that carry y/D passengers) operates as in the one-line model, such that frequency (and vehicle capacity) increases with patronage per line.

Regarding D, making the first derivative of (3) equal zero and multiplying times fD^2 we get:

$$2c_0Tf^2D^2 - c_1ty^2 - p_a\frac{y}{4v_a}f - p_v\frac{t}{4}y^2 = 0 \tag{6}$$

The left-hand sides of equations (4) and (6) have three identical terms, such that the remaining terms have to be equal, which means that at the optimum

$$f = uD \quad \text{with} \quad u = 2\frac{p_wv_a}{p_a} \tag{7}$$

Using equations (7), equation (4) can be written as a function of f only:

$$0 = 2c_0T\frac{1}{u^2}f^4 - c_1ty^2 - p_w\frac{y}{2}\frac{f}{u} - p_v\frac{t}{4}y^2 \tag{8}$$

Note that by dividing equation (8) by f^4 it becomes evident that f increases with Y, and so does D because of equations (7).

The explicit solution of equation (8) - degree 4 in f - is extremely long. However, equations (5) and (7) lead to the following properties of the model.

Property 1. *passengers per line and bus size increase as y grows.*

The number of passengers per line $z = \frac{y}{D}$ increases with y, because $\frac{dz}{dy} = \frac{dz}{df}\frac{df}{dy}$; the first factor is positive by equation (5) and the second was already shown to be positive.

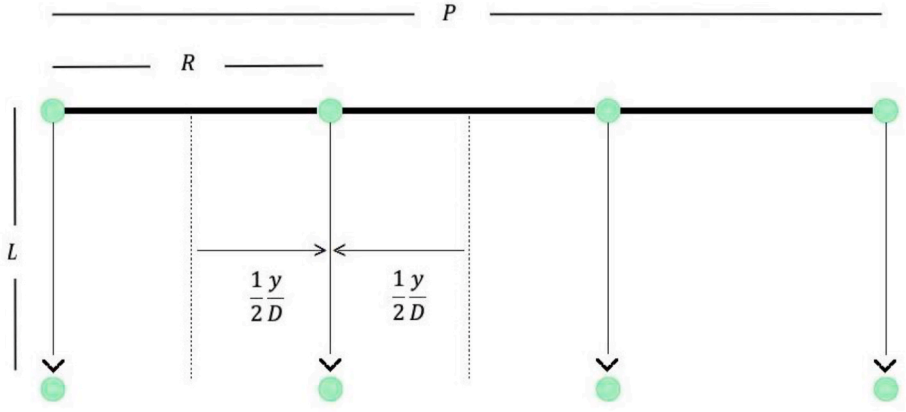

Fig. 1. The parallel-lines transit design problem.

Bus size increases with y because $\frac{dK}{dy} = \frac{dK}{dz}\frac{dz}{dY}$and both terms are positive.

Property 2. *optimal design implies that the average waiting and access costs are equal in this model.* This flows directly from equation (7), *i.e.* $\frac{p_w}{2f} = \frac{p_a}{4Dv_a}$.

The relationship between the results in equations (5) and (2) (obtained for the single-line model) suggests some conclusions regarding scale economies. First, those sources of scale economies presented for the single-line model remains valid for each of the lines in this model as well: increasing patronage induces a larger frequency and vehicle size, diminishing waiting time, and also induces larger times at stops increasing both in-vehicle time and operators' costs. Second, introducing the separation R between lines works in favor of scale economies because, as shown earlier, R diminishes (D increases) with patronage inducing a reduction in access time. Note that the effect of y on the optimal frequency in equation (5) is "softened" by lines density D, which also increases with patronage such that the Mohring effect is mitigated; as passengers are divided into many lines, buses become smaller when compared to the single-line model, which diminishes the diseconomies of scale provoked by the time at the stops.

Adding spacing as a design variable works in favor of scale economies such that the degree of scale economies *DSE* is expected to be larger than in the single line case; we now show that this is, in fact, the case. In order to compare both models rigorously, one should add access cost to the transit line in the single-line model. As in that model access time does not depend on the design, and therefore is constant per passenger, the total access cost is linearly increasing with Y, which means adding a term of the form λYin the total cost function. This means that both average and marginal costs increase by λ. On the other hand, in the single-line case *DSE* is larger than 1 when access is not considered (i.e. when $\lambda = 0$), which means that average cost AC is larger than marginal cost m; then the new average cost is $AC + \lambda$and the new marginal cost is $m + \lambda$ such that *DSE*, given by $(AC + \lambda)/(m + \lambda)$, decreases with λ. This means that *DSE* has an upper bound for $\lambda = 0$. This upper bound is represented in Fig. 2 (bottom-red line) together with the *DSE* of the new model (upper-blue line), using the simulation parameters shown in the Appendix. The new *DSE* is always larger than the upper bound of the single-line *DSE*, i.e. *DSE* increases when lines spacing is included. Note that scale economies get exhausted eventually. A sensitivity analysis on the parameters maintains these conclusions.

Let us now take a closer look at equation (7) that implies $\frac{f}{D} = 2v_a\frac{p_w}{p_a}$. An analogous property was also found by Hurdle (1973), Schonfeld (1981), Kocur and Hendrickson (1982) and Chang and Schonfeld (1991), where vehicle cost was assumed independent of bus size and total boarding-alighting time was assumed constant (i.e. affecting neither users' in-vehicle time nor cycle time); we have shown that the property remains valid even if those quite strong assumptions are dropped. This means that frequency and lines density grow at the same rate, irrespective of the number of passengers, of the boarding-alighting times, of the distances traveled by buses or passengers, etc. The intuition behind this is quite attractive: the optimal fleet of vehicles of an optimal size could be distributed in a large number of lines with a small frequency or *vice versa*; what the result says is that this trade-off between D and f is resolved by making the average waiting time value equal to the average access time value (Property 2).[4]

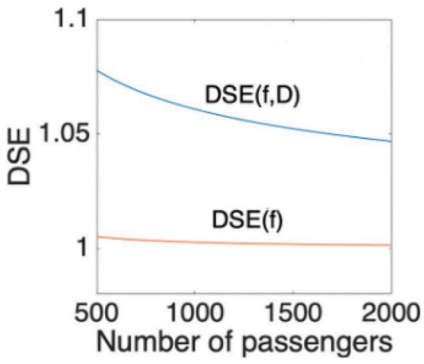

Fig. 2. The effect of lines spacing on the Degree of Scale Economies.

2. Lines spacing in a synthetic city model

2.1. Strategic lines in the parametric city

The role of lines spacing when deciding the most appropriate transit lines structure as part of the design requires the definition of a network and a demand pattern. Here we will use the model proposed and applied by Fielbaum et al. (2016, 2017), which has proven to be a useful city model for the analysis of transport systems. The city is composed by n zones - each one containing a subcenter and a periphery - and a CBD. The CBD is linked to each subcenter, and subcenters are linked to their peripheries and to their neighbor subcenters, as represented in Fig. 3a, where some spatial parameters are shown. Demand pattern is synthesized in Fig. 3b and represents morning peak, such that the CBD only attracts trips, peripheries only generate them and subcenters do both. There are three very relevant demand parameters: α, β and γ, that (grossly speaking) represent respectively the proportion of trips that are attracted by the CBD, by the own subcenters and by the rest of the subcenters (with $\widehat{\alpha}$ and $\widehat{\gamma}$ defined proportional to $\alpha$and γ) such that monocentric, polycentric and dispersed cities can be modeled when the respective parameters approach one; a and b are the proportion of trips generated at the peripheries and subcenters respectively. T_0 and g are spatial parameters. For a detailed explanation of the model, see Fielbaum et al. (2017).[5]

Four basic line structures are proposed over this urban scheme: feeder-trunk (FT), hub and spoke (HS), no transfers (NT) and no stops (NS).[6] They are represented in Fig. 4 showing only lines emerging from the "south" as they are all radially symmetric, together with a circular line when it exists. Note that, to simplify the figures, some arcs are drawn each per "type" of line; for instance, there are 3 black lines per zone in FT, and 3 blue lines per zone in HS, all reaching the opposite subcenters. A brief description follows:

- FT: "feeder" lines bring passengers from the peripheries to their own subcenter. There exist several "trunk" lines connecting each subcenter with the CBD and with some opposing subcenters, and a circular line.
- HS: from each periphery several lines depart passing through the CBD, that acts as the hub, arriving at some opposing subcenters. There is an additional circular line.[7]

[4] The equality between waiting and access costs (Property 2) was shown by Chang and Schonfeld (1991) to extend to operators cost as well. However, when their simplifying assumptions regarding vehicle costs and boarding-alighting time are dropped, this fails to be true even in the single-line case explained at the beginning of this section, and it remains untrue in the parallel-lines model.

[5] The conceptual model is suitable for any transit mode. However, adapting rail systems' infrastructure to an optimal design as patronage grows requires the consideration of capital costs associated to the length of the lines (Fawaz & Newell, 1976; Marín & García-Ródenas, 2009; Lovett et al. 2013) which is not done here.

[6] In previous papers, the No Transfers and No Stops structures had been named Direct (DIR) and Exclusive (EXC) respectively.

[7] When frequencies are optimized, the "green" lines that go from the CBD to each subcenter vanish as they always obtain null frequency.

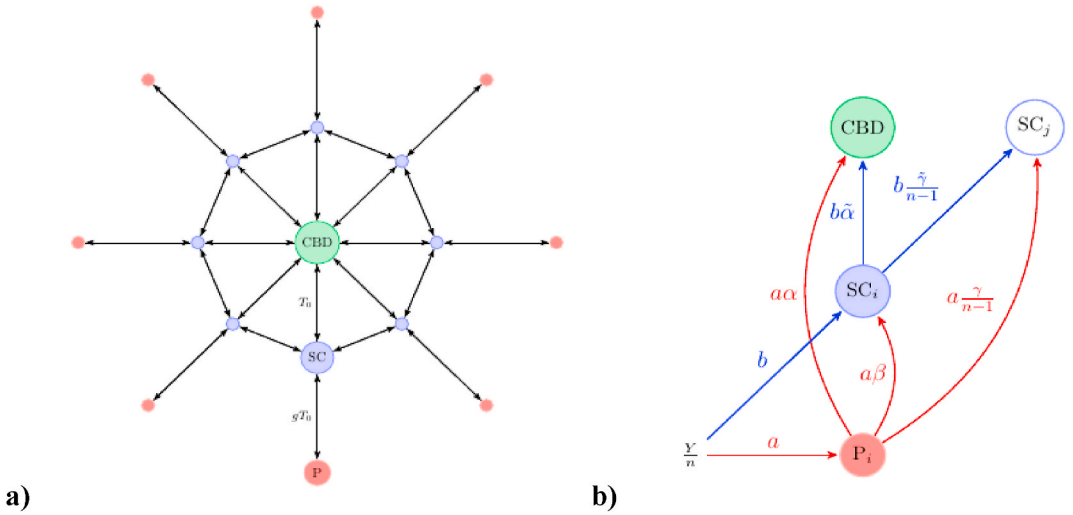

Fig. 3. Parametric representation of a city: network and demand pattern.

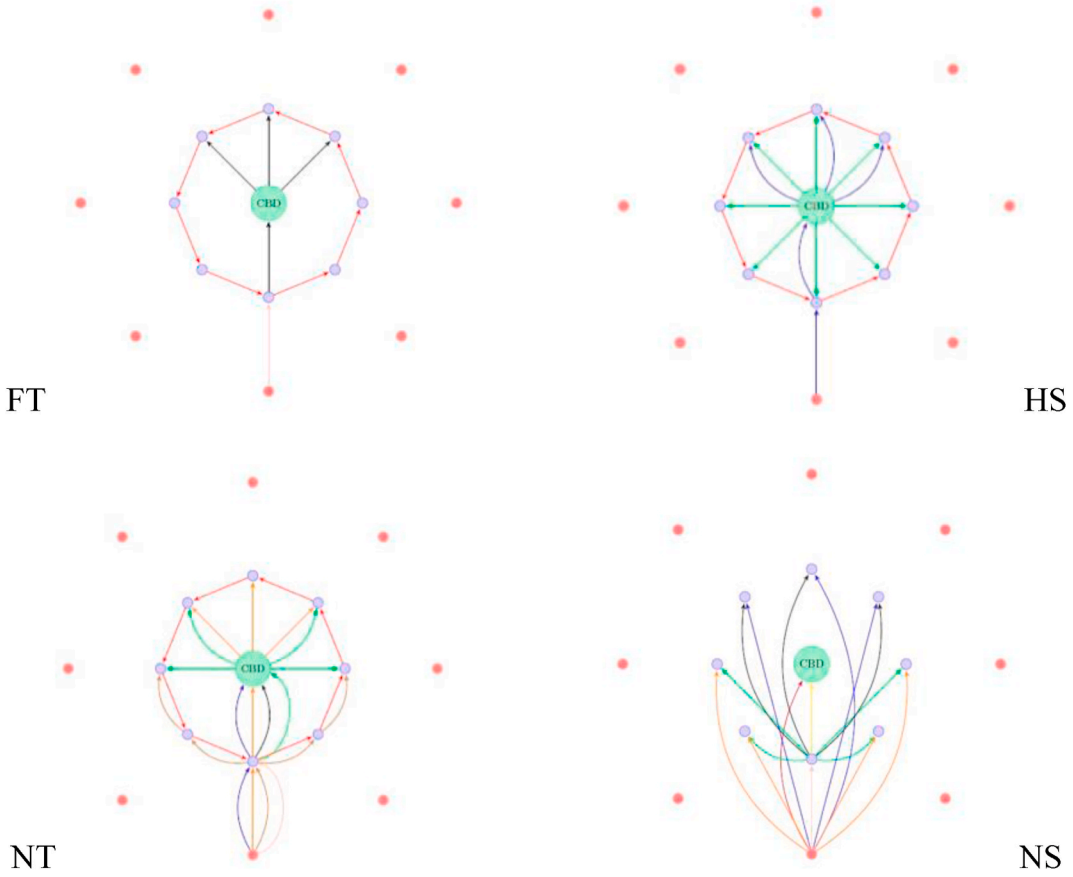

Fig. 4. Line structures in the parametric city.

- NT: each OD pair is connected by some line such that transfers are not needed. When optimizing frequencies, some lines might vanish (null frequency).
- NS: each OD pair is connected by an exclusive line that does not have intermediate stops.

In this approach, the design variables are the lines structure (i.e. the choice among FT, HS, NT and NS) and both frequency and bus size for each line within each structure.

Fielbaum et al. (2016) describe in detail each of these structures and explain how to obtain optimal frequencies for each line for all the four lines structures. The value of the resources consumed is now

$$VRC = \Sigma_L B_L(c_0 + c_1 K_L) + Y\left(p_v \bar{t_v} + p_w \bar{t_w} + p_a \bar{t_a} + p_{Tr} \bar{T}\, r\right) \quad (9)$$

As a brief synthesis, the number of buses B_L of each line depends on

its frequency and cycle time; bus size K_L depends on the most loaded segment of the respective routes; average waiting time $\bar{t}_w$ depends on the frequencies of each of the lines; average in-vehicle time $\bar{t}_v$ depends on the paths followed by each of the users and on the time spent at stops; and $\bar{T}\,r$ is the average number of transfers per trip; when transfers occur, the extra waiting time is included in $\bar{t}_w$, while the so-called pure transfer penalty (the discomfort due to the trip interruption) is captured by p_{Tr} (see García-Martinez et al., 2018, for a discussion and empirical estimation of this parameter). Note that the number of transfers and the length of the trips might depend on users' choices when they have more than one possible route. Fielbaum et al. (2016) use an iterative approach that assumes that all passengers choose their less costly route; fares are flat so travel time is all that matters. However, average access time $\bar{t}_a$ played no role in that formulation; this is exactly what we want to study in detail now.

2.2. Lines spacing as a new design variable

How can we adapt the model explained above in order to also consider lines spacing? To do so, we will consider that each former line l is now a "super-line" containing D parallel lines per unit width, each one with the same frequency f_l. This design variable D represents the spatial density of lines. The frequency of the super-line F_l is then given by

$$F_l = f_l D \tag{10}$$

The introduction of D has two effects on the users' cost that have to be taken into account: each user has to walk to the closest line inducing an access time, and buses distribute on more lines diminishing perceived frequency, increasing waiting time. Passengers walk in average $\bar{t}_a = \frac{1}{v_a}\frac{Y}{4D}$ minutes to access the nearest line, where v_a is walking speed. We assume that whenever transfers are required the bus stops coincide in space such that walking is negligible (but additional waiting and the interruption of the trip - the pure transfer penalty - are indeed considered). The design exercise was made using the four strategic lines structures explained in 3.1, including D as a design variable in addition to frequency and bus size for every line.

For a given lines structure, *VRC* depends on the lines frequencies and D only. In order to obtain the first order conditions, *VRC* can be written in a very compact way because derivatives require a local analysis (in a neighborhood of a given point) such that assignment can be assumed constant.[8] Following Fielbaum et al. (2016), fleets, (optimal) bus sizes, in-vehicle times and transfers can be expressed as functions of frequencies only. In this case, and with a fixed passenger's assignment, these functions depend on the frequencies of the super-lines:

$$B_l = B_l(F_1,\ldots,F_L),\ K_l = K_l(F_1,\ldots,F_L),\ \bar{t}_v = \bar{t}_v(F_1,\ldots,F_L),\ \bar{T}\,r = \text{constant} \tag{11}$$

The intuition behind these relations is straightforward. Fleets depend on the total number of buses running per unit time; buses should be large enough to carry the maximum load on each line, which again depends on the total number of buses per hour; in-vehicle time depends on total time in-motion (which is constant) and on time spent at bus stops, which depends on the number of passengers that board a specific bus, which in turn depends on the total number of buses; and finally, the number of transfers usually depends on the lines structure and on passengers assignment, both fixed in our case.

[8] Varying D might induce changes in passenger assignment because it affects waiting times of the different routes according to the frequencies of the corresponding lines. To be clear, if D doubles average waiting time of a route with 4 min increases to 8 min, while it increases to 20 in a route that exhibited an average waiting time of 10 min. Then the second route is more likely to lose passengers.

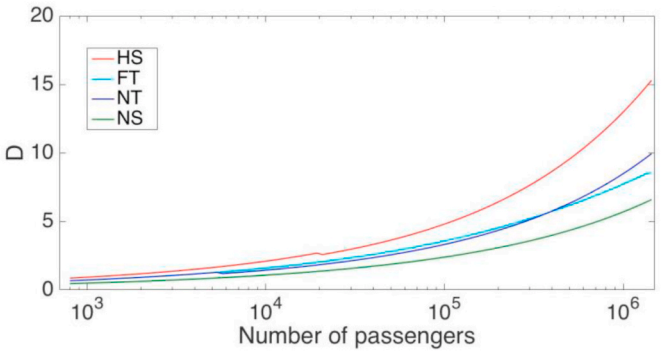

Fig. 5. Optimal density for each structure as a function of total demand.

Considering relations (11) and equation (10), we can rewrite *VRC* as a function of D and frequencies. If no common lines exist, then

$$VRC = G(f_1 D,\ldots,f_L D) + \frac{\theta_a}{D} + \sum_l \frac{\theta_l}{f_l} \tag{12}$$

G is a differentiable function that encompasses all terms in equation (9) but waiting and access users' cost. θ_l contains all the information related to waiting costs (like the number of passengers and their assignment, among others) and $\theta_a = \frac{p_a}{v_a}\frac{YP}{4}$.

If common lines are present, the third term in equation (12) becomes more complex, as some passengers' routes can use several lines such that the observed frequency is the sum over all common lines. In this case, total waiting time can be better expressed adding over OD-pairs w rather than lines:

$$VRC = G(f_1 D,\ldots,f_L D) + \frac{\theta_a}{D} + \sum_{w\in OD}\sum_{q=1}^{M_w} \frac{\theta_{wq}}{f_1\varepsilon_{1wq} + \ldots. + f_L\varepsilon_{Lwq}} \tag{13}$$

Trips on the OD-pair w are composed by M_w stages ($M_w - 1$ transfers), and ε_{lwq} is a binary variable whose value is 1 if passengers on that OD-pair use line l at stage q, and 0 otherwise. Note that equation (13) can be written as (14), with only one $\varepsilon_{lwq} = 1$for each w, q, and

$$\theta_l = \theta_{wq}\varepsilon_{lwq} \tag{14}$$

Now we can prove the following

Proposition 1. *Total access costs equal total waiting costs in this scheme.*

Proof:

Making the derivative with respect to f_l in (13) yields:

$$D\partial_l G - \sum_{w,q}\frac{\varepsilon_{lwq}\theta_{wq}}{\left(f_1\varepsilon_{1wq} + \ldots + f_L\varepsilon_{Lwq}\right)^2} = 0 \Rightarrow Df_l\partial_l G = \sum_{w,q}\frac{\varepsilon_{lwq}\theta_{wq}f_l}{\left(f_1\varepsilon_{1wq} + \ldots + f_L\varepsilon_{Lwq}\right)^2} \tag{15}$$

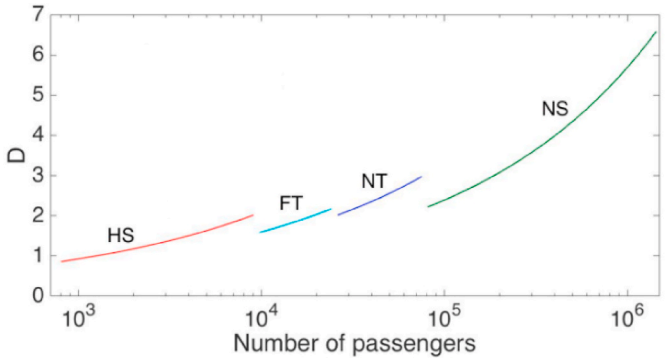

Fig. 6. Optimal density as a function of total demand.

Here we are using ∂_l to represent the partial derivative with respect to the *l-th* variable in G; the arguments of G have been omitted to simplify notation. Making the derivative with respect to D yields:

$$\sum_l f_l \partial_l G - \frac{\theta_a}{D^2} = 0 \Rightarrow \sum_l D f_l \partial_l G = \frac{\theta_a}{D} \tag{16}$$

Introducing (15) into the second equality in (16):

$$\frac{\theta_a}{D} = \sum_l \sum_{w,\, q} \frac{\varepsilon_{lwq}\theta_{wq} f_l}{\left(f_1\varepsilon_{1wq} + \ldots + f_L\varepsilon_{Lwq}\right)^2} = \sum_{w \in OD} \frac{\theta_{wq}}{\left(f_1\varepsilon_{1wq} + \ldots + f_L\varepsilon_{Lwq}\right)^2} \sum_l \varepsilon_{lwq} f_l = \sum_{w \in OD} \sum_{q=1}^{M_w} \frac{\theta_{wq}}{f_1\varepsilon_{1wq} + \ldots + f_L\varepsilon_{Lwq}} \tag{17}$$

2.2.1. Q.E.D

Proposition 1 proves that the property obtained in our simple parallel lines model in Section 2 (Property 2) remains valid: at the optimum design, average waiting and access costs are equal. The intuition behind this result is interesting, as it is not the usual microeconomic equality between some marginal costs. For any given fleet, optimizing D means deciding on how many parallel lines the buses should be distributed. If access cost is larger than waiting cost, then splitting the buses into more lines will induce savings in access costs that will outbalance the losses in waiting costs (analogously in the opposite situation). This happens because access and waiting costs are proportional to $\frac{1}{D}$ and $\frac{1}{f_l}$ respectively and, therefore, decreasing the largest of them in some proportion is more relevant than increasing the other one in the same proportion. For example, if there is a single line that needs walking 14 min and waiting 2 min on average, splitting it into two lines would increase waiting time to 4 min but reduce walking times to 7 min, which is more efficient.

Proposition 1 extends to a network our generalization in Section 2 of what had been obtained by Hurdle (1973), Schonfeld (1981), Kocur and Hendrickson (1982) and Chang and Schonfeld (1991). It is worth noting as well that our result resembles a property shown by Kraus (2008), that states that in a cost-minimizing network "the degree of local economies of scale in the cost function for the network's outputs is the same along all margins for adjusting capacity". In this case, the "margins for adjusting capacity" are frequency and density; however, as discussed by Fielbaum et al. (2020), Kraus' result relies on users' assignment to routes under a system optimizing rule which is not the case when users decide their routes according to individual preferences (as in this parametric city).

2.3. Numerical analysis

Numerical analysis was done using Y as the variable, with Santiago-like parameters $\alpha = 0.22$, $\beta = 0.25$ and $\gamma = 0.53$ (Fielbaum et al., 2017). The rest of the parameters (also used in Fielbaum et al., 2016) are found in the appendix. The procedure to find the best lines structure for each Y has two steps: first, for a given structure, the optimal (social cost minimizing) frequencies are found for each line together with the optimal D. Second, the best lines structure is found as the one that exhibits the minimum *VRC* across structures for each Y. Results are shown for a wide range of passenger volumes which makes the logarithmic scale preferable in order to facilitate the analysis for lower values of Y.

The results of step one are shown in Fig. 5; the optimal value of D increases with $Y$for each of the four structures, which fits intuition and is consistent with results in section 2. Note that structures that are less direct, i.e. those whose routes are longer with more transfers and stops (Fielbaum et al., 2020), present in general larger D; in other words D increases from NS (bottom-green curve) to HS (upper-red curve). This is because less direct structures involve super-lines that collect passengers from many OD pairs; the resulting large volume of passengers on the super-lines increases D.

In Fielbaum et al. (2020) we define a *threshold point* that, simply stated, is the value of Y where a slight increase in the number of trips makes a different lines structure the one that minimizes total cost. This is a discrete change because each lines structure is defined by a specific set of lines, which induces "jumps" (discrete changes) in frequencies, lines density, etc. In our example, as the number of passengers increases, the best lines structure (i.e. the one that minimizes *VRC*) evolves from HS to FT, then NT, and finally to NS, increasing directness. The corresponding D evolves as shown in Fig. 6. Within each structure lines density increases with Y, but it decreases locally at those threshold points where the line structure changes; this happens because, as directness increases, it is necessary to compensate for the fact that less passengers are being collected. Overall, however, D increases with Y as lines structure evolves from HS (left-red) to NS (right-green).

In Fig. 7 we show the impact of including D as a new design variable. The dotted lines represent the average cost curves of the best structures when $D = 1$ (fixed), while the solid lines show the result when D is optimized (i.e. the ones that correspond to Fig. 6). In both cases the evolution is towards those structures that are more direct, i.e. reducing transfers, stops and traveled distances. Different models (that usually consider only some of these three aspects) have shown that, as the number of passengers grows, structures that are increasingly direct become superior (Jara-Díaz & Gschwender, 2003a; Daganzo, 2010; Badia et al., 2014; Fielbaum et al., 2016, 2020). The novelty here is that introducing D not only reduces average cost but also postpones the emergence of those structures with increasing degrees of directness as Y increases: HS, FT, NT, and NS. This can be graphically seen by noting the lower levels of patronage at which a change in lines structure occurs along the dotted lines when compared with the solid lines in Fig. 7. Note that the difference in average cost between the best structures considering D or not increases with Y from nearly zero to nearly 24%.

For synthesis, less direct structures become more competitive when spacing is taken into account. This happens because when the number of passengers is low, reducing waiting times is the most relevant target, which is achieved through lines structures that collect passengers (at the transferring stations), making them share the same intermediate

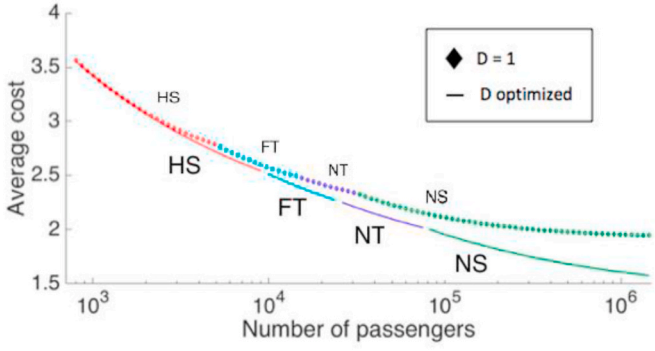

Fig. 7. Average total cost as a function of total demand.

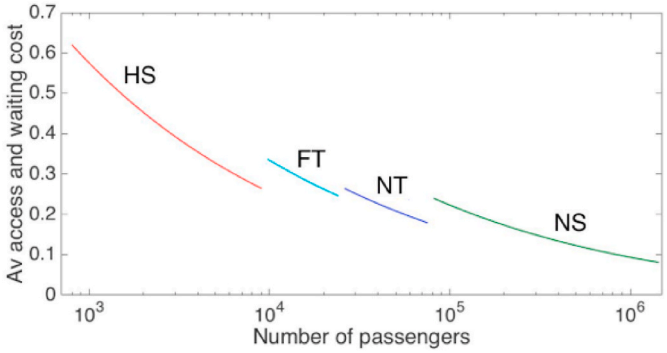

Fig. 8. Average access and waiting cost as a function of total demand.

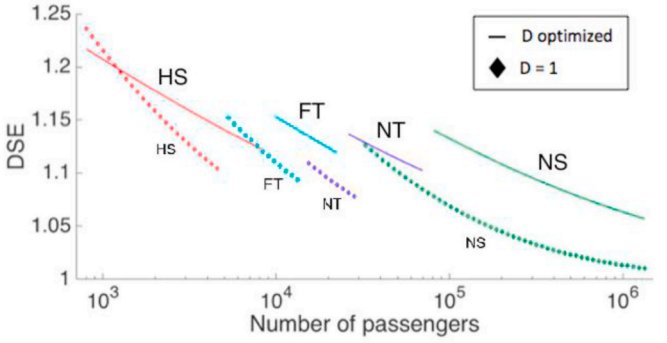

Fig. 9. Effect of spacing on the Degree of Scale Economies as a function of total demand.

destinations; the change towards more direct structures takes place only when the number of passengers is high enough to make waiting times less relevant (Fielbaum et al., 2020). When spacing is included as a design variable, frequencies decrease (because each line is split into many) such that the number of passengers that make direct line structures more convenient, increases.

We have shown that under an optimal design average access and waiting costs are going to be equal. Their (joint) evolution as Y increase is shown in Fig. 8, where one can see that for any given lines structure these costs decrease (the Mohring effect operates). But when the lines structure changes, this curve jumps upwards (just as D decreases locally). This is consistent with the results of Fielbaum et al. (2020) when studying the impact of lines structure changes on waiting times.

In section 2 we showed the effect of considering lines spacing on scale economies. The same analysis can be replicated here, by comparing the *DSE* obtained with the model just presented (solid lines in Fig. 9) against the *DSE* obtained if access time is considered but with $D = 1$ (dotted lines in Fig. 9). In both cases at the points in which lines structure changes, the *DSE* jumps discreetly upwards (as shown in Fielbaum et al., 2020)[9] and then continues decreasing asymptotically to 1. Overall, solid lines lay above dotted lines which means that optimizing spacing is yet another source of scale economies: increasing the number of users makes lines density increase, reducing the average walking time; this can be interpreted as the spatial counterpart of the so-called "Mohring effect" and the reduction of waiting time after an increase in patronage due to the associated increase in frequency. As advanced in Fig. 7, the jumps occur first (i.e. for lower levels of Y) in the dotted lines, which explains the small anomaly when HS turns into FT for the $D = 1$ model. The only true exception occurs at the beginning of the graph, where the optimal D is lower than 1: frequencies are large in this zone, making the Mohring effect less relevant.

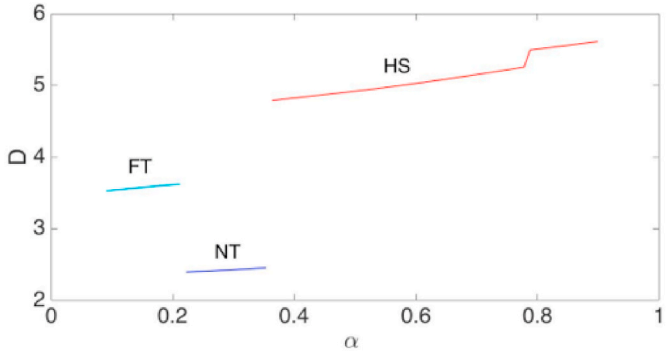

Fig. 10. Effect of monocentricity on the optimal lines density.

As advanced by Daganzo (2010), Gschwender et al. (2016) or Fielbaum et al. (2016), the internal distribution of the trips has an impact over public transport design for a given level of total trips. To have an idea of this effect, in Fig. 10 we show the evolution of the optimal density as the degree of monocentrism - represented by α - increases from 0.1 to 0.9 keeping $\beta = \gamma$, for $Y = 90,000$ passengers per hour. Here the lines structure evolves from FT to NT to HS; as α grows within a given lines structure, lines density increases a little (when trips are more concentrated, lines can split up to reduce access times), but it varies notably when the lines structure changes. When most trips go from the peripheries to the subcenters - a non-monocentric city - FT dominates with the circular line playing an important role. As α increases such that $\alpha \cong \beta = \gamma$, the CBD, the own subcenter and the other subcenters become equally important such that direct lines are superior using all arcs and density drops. When α begins to dominate (mildly monocentric) it is more efficient to begin using the CBD as a hub combined with the circular line (see Fig. 4) until the city becomes very centralized (α>0.8) such that it is better to let the few trips that go to other subcenters transfer at the CBD, which explains the slight increase of D within HS because the circular line disappears, i.e. its frequency becomes zero.

Fig. 10 admits an intuitive explanation, keeping in mind that $\beta = \gamma$ and that only α varies. The intuition behind the dominant lines structures is quite clear, as discussed in detail in Fielbaum et al. (2016): when α is small, trips to all subcenters are very important and the own subcenter is the dominant one for each periphery, such that the feeder-trunk structure is obviously the more convenient; when α is large it is the CBD the destination that dominates, making an HS structure more convenient with an obvious hub at the CBD; when α is around 0.33 the CBD, the own subcenter and the other subcenters have the same importance which makes direct lines the "balanced" choice. Regarding lines density, the two extremes are quite intuitive because, compared against HS, FT dominates when trips to the subcenters are larger (i.e. trips are more distributed in space) such that waiting times are larger than those occurring in the HS zone; as waiting time values have to equal access time values, density is smaller for FT than for HS. The small density of the direct lines occurs because they have no transfers and exhibit the largest waiting (and therefore also access) times, which means large spacing. Regarding the (slight) increase of density for a given lines structure, the explanation is simple: trips to the CBD increase with α so that optimal frequencies increase and waiting time diminishes along with access time (density increases).

Finally, D has been treated as a continuous variable: the number of lines per unit width. Nevertheless, as streets are exogenous to the model, flexibility to adjust this variable is somewhat limited. This is why it is worth analyzing if results previously found are still valid if D is modeled as discrete. The results presented in Fig. 11 reveals that they tend to hold. Fig. 11a shows the optimal discrete D as a function of the number of passengers (similar to Fig. 6); Fig. 11b shows the differences in average costs (similar to Fig. 7); and Fig. 11c represents the ratio between access and waiting costs (always 1 in the continuous case). This

[9] At each threshold point average costs are equal but marginal cost of the emerging structure is lower, increasing *DSE*.

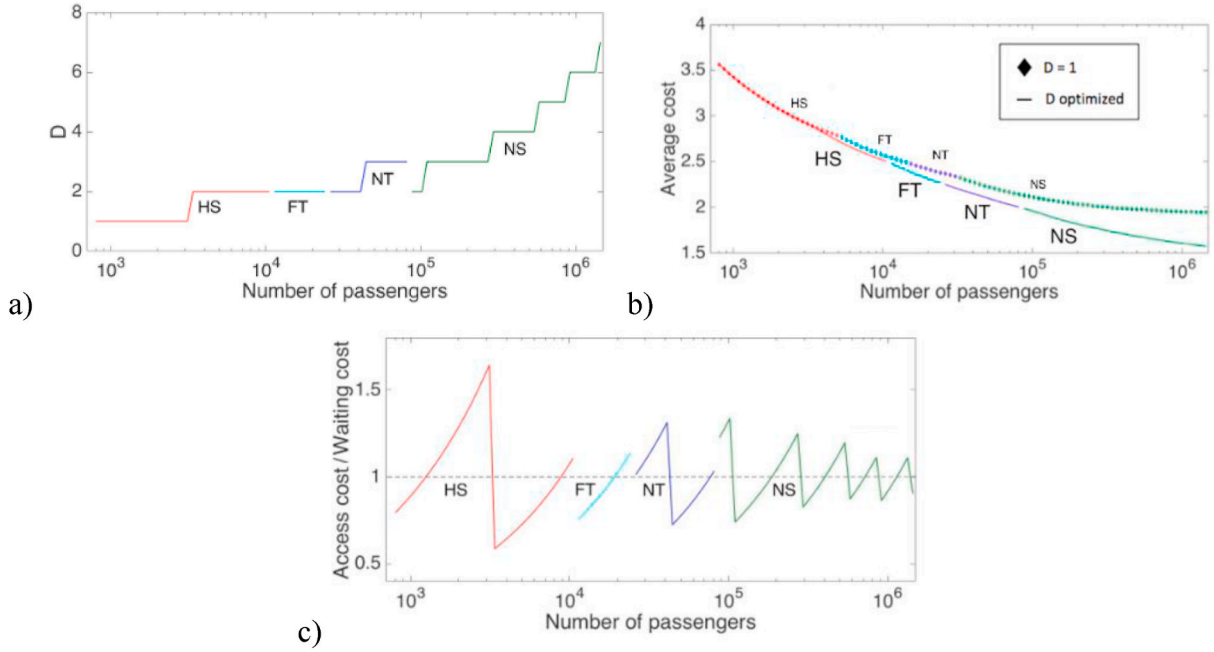

Fig. 11. Optimal density (a), average costs (b) and access/waiting cost ratio (c) as a function of patronage for the discrete density case.

last Figure is quite interesting, as these costs are not exactly equal anymore because the discreteness of D prevents full adjustment between density and frequencies; nevertheless this ratio oscillates around 1 and approaches this value as Y grows.

3. Synthesis, conclusions, and further research

In this paper the role of spacing as a design variable emerges very clearly. From a strategic viewpoint, the optimal design of transit systems involves decisions that somehow belong to a rank. In very simple settings (networks), different combinations of fleet and vehicle sizes can provide the same total capacity. To generate the same number of seats, the choice of an optimal combination depends on what the designer wants to achieve: a small fleet of large buses would be the cheapest (with a low frequency) while a large fleet of small vehicles would be the best for the users (but very expensive), a trade-off that emerges indeed when a budget constraint or self-financing policies are imposed (Jara-Díaz & Gschwender, 2009). In not so simple networks this type of choice remains as an important one, but another dimension has to be considered as well: the design of the transit network - a lines structure -, i.e. a set of transit lines that can be thought of in many ways, from lines organized involving a certain amount of transfers (e.g. hub-and-spoke or feeder-trunk systems), to a set of OD specific lines with neither transfers nor stops, which begins to be attractive as demand gets large enough. Here we have shown that these dimensions do not exhaust all design possibilities, however: lines can be replicated in space, i.e. decreasing lines spacing could be an adequate response to demand growth.

We first developed a version of a single line model adding spacing between parallel lines as a design variable, including boarding and alighting time and the role of bus size on operators' cost, which yielded some interesting properties: first, frequency and bus size for each of the parallel lines replicate the relations found in the single-line model; second, as total demand grows, optimal spacing decreases (inducing less walking) while frequency increases (inducing less waiting), and this happens in such a way that total waiting and access times remain equal, showing that spacing is yet another source of scale economies. This latter property remains valid in the context of the design of a transit network - a lines structure - in a city described around the role of peripheries, sub-centers, and a center (Fielbaum et al., 2017), using trip distribution parameters that admit the representation of monocentric, polycentric or dispersed scenarios. In this case density -the inverse of spacing-also grows with total demand except at the points where a change in lines structure takes place, making access and waiting costs increase locally. In this context, a new property emerges, namely that lines spacing is relevant not only when total demand is large but also plays a role that is linked with what we have called "directness" of a lines structure (Fielbaum et al., 2020), which is said to increase as routes become shorter and transfers and stops diminish. When lines spacing is considered as a design variable, it begins to grow at relatively low values of total demand such that the nice properties of more direct routes require larger flows to emerge. For short, spacing exhibits some degree of substitution with both frequency and directness.

From the point of view of the introduction of spacing in the parametric representation of a city, we have to recall that this model is conceived as a spatial setting based on the structure of centers that is simple enough to find the most appropriate strategic transit lines structure as a basis for a detailed design. There are parameters that represent the basic network - its centers and arcs - and those that represent the demand structure that corresponds to different city types - the proportion of trips to subcenters and to the CBD. When spacing is allowed, polycentrism and dispersion favor a feeder-trunk structure with the sub-centers as transfer points while mostly monocentric cities favor hub-and-spoke with a hub at the CBD.

In a model aimed at the strategic design of a transit network, there are many elements that are either simplified or neglected. This is due not only to the aggregated approach inherent to a strategic view but also to the trade-off that exists in public transport models between the number of variables and details included in the model and the possibility of deriving analytical results. There are several directions in which our models can move in order to incorporate some elements that have not been considered.

First, the heterogeneity of trip length across users has been proven to have relevant impacts over an optimal design (Hörcher & Graham, 2018; Dakic et al., 2020). This could be included in the simple model studied in Section 2, and radial asymmetries could make the model studied in Section 3 more general (although trips of different length are indeed present in that model). Second, the models considered here are static, not accounting for intra-day traffic dynamics, a factor whose impact on public transport design has been studied by Jara-Díaz et al. (2017) regarding frequencies across periods; it is noteworthy that

whereas lines' frequencies can easily be changed throughout a day, this is not the case for either the lines structure and its spatial density.

Third, the models we studied in this paper consider many elements in a deterministic way, but there are many factors of uncertainty in public transport: traveling times might change due to congestion, the arrival of passengers might not be uniform, or buses might face bunching, among others. All these random facts affect the optimal design, as the system needs to stay reliable and robust when unexpected changes happen. In contrast to the said sources of uncertainty, access times are quite stable, so spacing might not be directly affected when randomness is included in the model.

Fourth, users might be heterogeneous, which could be included in the model either by considering different time values in different zones within the parametric city (representing inter-zonal differences in income, for example) or by assuming that there is some random noise when passengers choose their routes, leading to logit-type models.

Fifth, users' experience on the bus could be modeled in more detail. So far we have assumed that buses can be filled up to a certain capacity that is reached in the most loaded segment of the line. It has been shown, though, that a more crowded bus is usually perceived as less comfortable by users (Jara-Díaz & Gschwender, 2003b; Tirachini et al., 2013; Hörcher & Graham, 2018), which has an impact on scale analysis, as increasing the load factor (a measure of crowding) of a bus is a negative externality. In a similar note, the fare collection system (Tirachini & Hensher, 2011), as well as the boarding-alighting technology (Jara-Díaz & Tirachini, 2013) might also be designed optimally; the effect of these on spacing and scale is yet unknown.

There are two other elements that correspond to another level of analysis. One is optimal pricing, which depends on short-run long-run considerations. If everything can be varied in an optimal way, increasing the demand volume increases frequencies, vehicle sizes, directness and density, reducing waiting, in-vehicle and access times; these are scale economies that translate into (long-run) optimal prices and revenues that fall short of operators' costs, inducing optimal subsidies. When prices are not the optimal ones, the design is affected indeed, as shown by Jara-Díaz and Gschwender (2009) regarding lower than optimal frequencies and larger than optimal vehicles.

Optimal prices (and subsidies), as well as level-of-service variables, might affect demand, which calls for an expanded general equilibrium approach including not only mode choice but also long-run effects in location, land use, and time use which links with the second element: transit network design interacts with the development of the city in which it is immersed. This can affect the general urban form (Anas & Moses, 1979; Brueckner, 2005; Basso et al. 2020), the internal distribution of people (Glaeser et al., 2008), and the evolution of zones close to transit stations (Bertolini, 1999; Diao et al. 2017), all of which depend on whether the city is considered as closed (fixed boundaries) or open (Brueckner, 1987; Kono & Joshi, 2012). Relating this research to lines' density is a promising research direction, as space and time play a vital role in all these phenomena.

Finally, density is a design variable that has an upper limit given by the local size of the underlying network, i.e. the number of parallel lines that fit into each of the (aggregated) arcs. This number, in turn, would depend on the technology conceived for the transit system. Admitting more than one technology would make this an important avenue for future research.

CRediT authorship contribution statement

Andrés Fielbaum: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Sergio Jara-Diaz:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Visualization, Writing - original draft, Writing - review & editing. **Antonio Gschwender:** Conceptualization, Formal analysis, Investigation, Methodology, 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 two unknown referees that helped us improve the exposition of the paper. Remaining errors are our responsibility.

References

Anas, A., & Moses, L. N. (1979). Mode choice, transport structure and urban land use. *Journal of Urban Economics, 6*(2), 228–246.

Badia, H., Estrada, M., & Robuste, F. (2014). Competitive transit network design in cities with radial street patterns. *Transportation Research Part B: Methodological, 59*, 161–181.

Basso, L., Navarro, M., & Silva, H. (2020). *Public transport and urban structure*. Documento de trabajo Instituto de Economía PUC, 549.

Bertolini, L. (1999). Spatial development patterns and public transport: The application of an analytical model in The Netherlands. *Planning Practice and Research, 14*(2), 199–210.

Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth-Mills model. *Handbook of Regional and Urban Economics, 2*(20), 821–845.

Brueckner, J. K. (2005). Transport subsidies, system choice, and urban sprawl. *Regional Science and Urban Economics, 35*(6), 715–733.

Byrne, B. (1976). Cost minimizing positions, lengths and headways for parallel public transit lines having different speeds. *Transportation Research, 10*(3), 209–214.

Chang, S., & Schonfeld, P. (1991). Multiple period optimization of bus transit systems. *Transportation Research Part B: Methodological, 25*(6), 453–478.

Chen, H., Gu, W., Cassidy, M., & Daganzo, C. (2015). Optimal transit service atop ring-radial and grid street networks: A continuum approximation design method and comparisons. *Transportation Research Part B: Methodological, 81*, 755–774.

Daganzo, C. (2010). Structure of competitive transit networks. *Transportation Research Part B: Methodological, 44*(4), 434–446.

Dakic, I., Leclercq, L., & Menendez, M. (2020). *On the optimization of the bus network design: An analytical approach based on a bi-modal Macroscopic Fundamental Diagram*. SVT Working Papers. ETH Zürich.

Diao, M., Zhu, Y., & Zhu, J. (2017). Intra-city access to inter-city transport nodes: The implications of high-speed-rail station locations for the urban development of Chinese cities. *Urban Studies, 54*(10), 2249–2267.

Díaz, G., Gómez-Lobo, A., & Velasco, A. (2002). *Micros en Santiago: Hacia la licitación del 2003*. Kennedy School of Government, Harvard University. https://sites.hks.harvard.edu/cid/archive/events/chile/diaz_gomez-lobo_velasco.pdf. (Accessed 17 September 2019).

Fawaz, M., & Newell, G. (1976). Optimal spacings for a rectangular grid transportation network—I: A hierarchy structure. *Transportation Research, 10*(2), 111–119.

Fielbaum, A., Jara-Diaz, S., & Gschwender, A. (2016). Optimal public transport networks for a general urban structure. *Transportation Research Part B: Methodological, 94*, 298–313.

Fielbaum, A., Jara-Diaz, S., & Gschwender, A. (2017). A parametric description of cities for the normative analysis of transport systems. *Networks and Spatial Economics, 17*(2), 343–365.

Fielbaum, A., Jara-Diaz, S., & Gschwender, A. (2020). Beyond the Mohring effect: Scale economies induced by transit lines structures design. *Economics of Transportation, 22*, Article 100163.

García-Martínez, A., Cascajo, R., Jara-Díaz, S. R., Chowdhury, S., & Monzón, A. (2018). Transfer penalties in multimodal public transport networks. *Transportation Research Part A, 114*, 52–66.

Appendix: Parameters used in the simulations.

Parameter	α	β	c_0	c_1	T	g	t		
Value	0.25	0.22	10.65 [$US/h]	0.204 [$US/h]	30 [min]	1/3	2.5 [s]		
Parameter	p_v	p_w	p_a	p_R	n	a	P		v_a
Value	1.48 [US$/h]	4.44 [US$/h]	5.33 [US$/h]	0.37 [$US]	8	0.8	2 [km]	5 [km/h]	

Glaeser, E. L., Kahn, M. E., & Rappaport, J. (2008). Why do the poor live in cities? The role of public transportation. *Journal of Urban Economics, 63*(1), 1–24.
Gschwender, A., Jara-Díaz, S., & Bravo, C. (2016). Feeder-trunk or direct lines? Economies of density, transfer costs and transit structure in an urban context. *Transportation Research Part A: Policy and Practice, 88*, 209–222.
Hörcher, D., & Graham, D. J. (2018). Demand imbalances and multi-period public transport supply. *Transportation Research Part B: Methodological, 108*, 106–126.
Hurdle, V. F. (1973). Minimum cost locations for parallel public transit lines. *Transportation Science, 7*(4), 340–350.
Jansson, J. O. (1980). A simple bus line model for optimisation of service frequency and bus size. *Journal of Transport Economics and Policy*, 53–80.
Jara-Díaz, S., Fielbaum, A., & Gschwender, A. (2017). Optimal fleet size, frequencies and vehicle capacities considering peak and off-peak periods in public transport. *Transportation Research Part A: Policy and Practice, 106*, 65–74.
Jara-Díaz, S., & Gschwender, A. (2003a). From the single line model to the spatial structure of transit services: Corridors or direct? *Journal of Transport Economics and Policy, 37*(2), 261–277.
Jara-Díaz, S., & Gschwender, A. (2003b). 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., & Tirachini, A. (2013). Urban bus transport: Open all doors for boarding. *Journal of Transport Economics and Policy, 47*(1), 91–106.
Kocur, G., & Hendrickson, C. (1982). Design of local bus service with demand equilibration. *Transportation Science, 16*(2), 149–170.
Kono, T., & Joshi, K. (2012). A new interpretation on the optimal density regulations: Closed and open city. *Journal of Housing Economics, 21*(3), 223–234.
Kraus, M. (2008). Economies of scale in networks. *Journal of Urban Economics, 64*(1), 171–177.
Lovett, A., Munden, G., Saat, M., & Barkan, C. (2013). High-speed rail network design and station location: Model and sensitivity analysis. *Transportation Research Record, 2374*(1), 1–8.
Marin, Á., & García-Ródenas, R. (2009). Location of infrastructure in urban railway networks. *Computers & Operations Research, 36*(5), 1461–1477.
Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *The American Economic Review, 62*(4), 591–604.
Quadrifoglio, L., & Li, X. (2009). A methodology to derive the critical demand density for designing and operating feeder transit services. *Transportation Research Part B: Methodological, 43*(10), 922–935.
Schonfeld, P. (1981). *Minimum cost transit and paratransit services*. College Park: Transportation Studies Center Report, Department of Civil Engineering, University of Maryland.
Tirachini, A., & Hensher, D. A. (2011). Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. *Transportation Research Part B: Methodological, 45*(5), 828–844.
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. *Research in Transportation Economics, 29*(1), 231–242.
Tirachini, A., Hensher, D. A., & Rose, J. M. (2013). Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice, 53*, 36–52.