Improved models for technology choice in a transit corridor with fixed demand

Luigi Moccia[a], Gilbert Laporte[b]

[a]*Istituto di Calcolo e Reti ad Alte Prestazioni, Consiglio Nazionale delle Ricerche, Rende, Italy*
[b]*HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7*

Abstract

We present three extensions to a base optimization model for a transit line which can be used to strategically evaluate technology choices. We add to the base model optimal stop spacing and train length, a crowding penalty, and a multi-period generalization. These extensions are analytically solvable by simple approximations and lead to meaningful insights. Their significance is illustrated by means of an example in which two road modes and two rail modes are defined by a set of techno-economical parameters. These parameters loaded in the base model yield dominance of road modes for all but the largest demand levels. We consistently keep this set of parameters for all models, and show how the break-even points between road and rail modes progressively recede toward lower demand levels when model refinements – not parameter changes – are applied. Scenario analyses of plausible parameter sets highlight the model's versatility, and caution on general conclusions regarding technology dominance.

Keywords: Transit optimization, bus rapid transit, light rail transit,

Email addresses: moccia@icar.cnr.it (Luigi Moccia), gilbert.laporte@cirrelt.ca (Gilbert Laporte)

1. Introduction

Rapid transit provides a key tool for sharply reducing the negative environmental externalities of transport in urban areas, and for improving access to employment and services to lower income groups (Fulton and Replogle, 2014). Rapid transit consists of bus rapid transit (BRT), light rail transit (LRT), metro, and commuter rail. In recent years, several models have been put forward for the strategic choice of transit technology (see e.g. Daganzo (2010), Tirachini et al. (2010b), Estrada et al. (2011), Sivakumaran et al. (2014)). Here we extend in several directions the model of Tirachini et al. (2010b) which is used as our base model. In it the optimized variable is the frequency, the objective function is the minimization of the sum of passenger and operator costs, and the demand is assumed to be fixed in a single period. This model can be solved analytically. We first extend the base model to account for optimal stop spacing. We then introduce optimal train length and a crowding penalty. Finally, we consider a two-period case and a multi-period generalization. In spite of some notational complexity, the proposed extensions can be solved by simple approximation schemes which provide some analytical insights into the structure of optimal solutions. In particular, we find that the ratio of optimal stop spacings among different modes follows a square root formula. A crowding penalty moves away the optimal frequency from the minimal values induced by the critical capacity. Road and rail modes handle crowding in different ways. Road modes aim to offer a higher frequency, whereas rail modes provide leverage on both

frequency and train lengths. The multi-period model further increases the model realism when comparing different technologies.

The remainder of this paper is organized as follows. Section 2 contains the literature review which motivates our work. Section 3 presents mathematical models and approximation schemes. Section 4 provides numerical analyses of a base parameter set, and Section 5 discusses plausible scenarios. Some conclusions are reported in Section 6.

2. Literature review

We first review the relevant literature on microeconomic models of transit systems in Section 2.1, and then the analytical models for transit network design in Section 2.2. Finally, we present our approach in Section 2.3.

2.1. Microeconomic models

Jara-Díaz and Gschwender (2003) review microeconomic models for the operation of public transport, propose extensions, and compare models' results. The authors extend the model of Jansson (1980), where demand is fixed, by including the effect of vehicle size on operating costs and the influence of crowding on the value of in-vehicle time. The authors show how a better characterization of user cost significantly increases optimal frequencies, and alters key design variables of a transit system. Bruun (2005) introduces a parametric cost model for BRT and LRT in order to compare these two technologies for base and peak service hours in trunk lines. Both the demand level and the shape of the demand profile determine operating costs. The author observes that rail technology can accommodate demand variations through the addition and removal of carriages from trains. According to cost

parameters representative of transit agencies in the United States, the model finds LRT to be increasingly attractive when demand is above 2000 spaces per hour. In recent years, the effect of crowding on passengers' value of time started to attract some attention. Tirachini et al. (2013) review different literature threads related to crowding in public transport such as psychometric methods, engineering, economics, etc. Additional references are found in Qin and Jia (2012), Qin (2014), and De Palma et al. (2015).

2.2. Analytical models for transit network design

The literature on structural transit analysis was initiated by Byrne (1975) for radial lines, Newell (1979) for a hub-and-spoke network, Vaughan (1986) for ring and radial routes, and Chien and Schonfeld (1998) for a rail trunk line with a bus service as a feeder. In recent years the continuous approximation literature has addressed the strategic evaluation of transit technologies. Daganzo (2010) studies structural characteristics of a transit system for a square shaped urban area. The main assumption is that the origin-destination flows are uniformly and independently distributed over the square. The author acknowledges that this assumption penalizes the transit system with respect to private car, but justifies this assumption since it sets a higher bar for transit success. The objective function minimizes the sum of user and agency costs under a fixed demand. The user cost is the sum of access, waiting, transfer, riding, and egress times for an average trip. The agency cost is the sum of fixed and variable costs of the transit system normalized per trip. The non-linear model can be easily solved by a grid search on the domains of the variables. The author compares road and rail rapid transit systems and concludes that BRT dominates LRT and metro. This result is not surprising

given the uniform demand assumption. In fact, the formula derived for the critical load results in relatively low values of transit occupancy, i.e. trains are not filled up. Estrada et al. (2011) extend the model of Daganzo (2010), and present a case study. Badia et al. (2014) further extend the model of Daganzo (2010) to cities with a radial street pattern. Sivakumaran et al. (2014) consider a hierarchical transit systems in which rail transit is the backbone of a bus feeder network.

Tirachini et al. (2010b) present models for a single transit line of fixed length where the number of stops is given. Under fixed demand, the goal of social welfare maximization is equivalent to the minimization of total user and operator cost. Under elastic demand, two goals are modeled: the maximization of the operator's profit and the maximization of social welfare. A distinctive characteristic of this model is that the critical load is defined by a specific parameter. Hence, this approach can synthetically model every type of demand distribution, centripetal or uniform. Passenger costs related to access, waiting, and in-vehicle times are finely represented. The authors also provide an extension to numerically account for crowding costs. This analysis is further expanded in Tirachini et al. (2010a) to a multi-period radial network in a circular service area.

2.3. Proposed approach

This literature review highlights the relevance of several issues. User values of time must be finely characterised (Jara-Díaz and Gschwender (2003), Tirachini et al. (2010b)), in particular crowding (Qin and Jia (2012), Tirachini et al. (2013), Qin (2014), De Palma et al. (2015), and references therein). The cost structure of different technologies is paramount (Bruun, 2005). As-

sumptions on demand distribution are crucial since rail transit systems arise when peak hour capacity is an issue (see e.g. Vuchic (2005), Vuchic et al. (2013)). Simple analytical models show broader applicability than numerical approaches (Daganzo (2010), Estrada et al. (2011)). Analytical models can be also useful to environmental assessments (Griswold et al. (2013), Griswold et al. (2014)). For these reasons, we extend the transit line model of Tirachini et al. (2010b) which can easily accomodate different assumptions on demand distribution and already provides a refined characterisation of user costs. The extensions proposed in the following section show how a transit line model can be made more realistic without sacrificing analytic tractability, by resorting to approximations.

3. Mathematical models and approximation schemes

This section presents mathematical models for the optimization of a transit line. Section 3.1 summarizes a classical model in which the optimized variable is the frequency, the objective function is the minimization of the sum of passenger and operator costs, and the demand is assumed to be fixed in a single period. This model, referred to as Model I, can be solved analytically and is our base model which is extended in Section 3.2 to account for optimal stop spacing. Section 3.3 introduces optimal train length, and a crowding penalty. Section 3.4 considers a two-period case, and a multi-period generalization. The new models are in general unsolvable by straightforward analytical procedures, and hence we propose analytically solvable approximation schemes for them. Formulae are reported with measurement units at their first mention. Table 1 summarizes the main symbols. Additional

symbols are derived as explained in the following. Subscripts may be used to specify relevant models. For example, t_{c3} indicates the cycle time of Model III introduced in Section 3.3, which differs from t_{c2}, the cycle time of Model II introduced in Section 3.2. The subscripts *min* and *max* specify bounds of a parameter or of a variable. The subscript *unc* refers to the optimal solution of an unconstrained objective function. A variable with a tilde is related to an approximation scheme. The superscripts (a, b) refer to the two cases for the waiting behaviour which depends on the frequency as detailed in Section 3.1. The superscripts (p, o) refer to the peak or off-peak periods.

Table 1: List of primary symbols, and units of measure used in the formulae

Symbol	Definition	Unit
$\bar{a}$	Average acceleration rate of a TU	km/h^2
$\bar{b}$	Average deceleration rate of a TU	km/h^2
B	Fleet size	TU
c_0	Fixed operator cost related to the infrastructure	\$/h
c_{0l}	Fixed operator cost related to the transit line	\$/h
c_{0s}	Fixed operator cost related to a stop	\$/h
c_1	Unit operator cost per TU-hour in Models I and II	\$/(TU-h)
c_{1t}	Unit operator cost per TU-hour in Models III and following	\$/(TU-h)
c_{1v}	Unit operator cost per vehicle-hour	\$/(veh-h)
c_2	Unit operator cost per TU-km	\$/(TU-km)
c_{2v}	Unit operator cost per veh-km	\$/(veh-km)
C_a	Access and egress time cost	\$/h
C_o	Operator cost	\$/h
C_p	Passenger cost	\$/h
C_{tot}	Total cost, sum of passenger and operator costs	\$/h
C_v	In-vehicle time cost	\$/h
C_w	Waiting time cost	\$/h
d	Distance between stops	km
f	Frequency	TU/h
$\bar{f}$	Threshold frequency for timetable behaviour	TU/h

Continued on next page

Table 1 – *Continued from previous page*

Symbol	Definition	Unit
K	Capacity of a TU	pax/TU
k_v	Capacity of a vehicle	pax/veh
l	Average trip length	km
L	Length of the transit line	km
n	Number of vehicles per TU	veh
P_a	Value of the access and egress time	\$/(pax-h)
P_v	Value of the in-vehicle time	\$/(pax-h)
P_w	Value of the waiting time	\$/(pax-h)
R	Running time	km/h
S	Average running speed	km/h
t_a	Average access and egress time of a passenger	h
T_a	Time loss caused by acceleration and deceleration phases	h
t_c	Cycle time	h
t_d	Time loss caused by opening and closing of doors	h
T_l	Time loss caused by acceleration, deceleration, and door operations	h
TU	Transit unit	-
t_v	Average in-vehicle time of a passenger	h
t_w	Average waiting time of a passenger	h
v	Walking speed	km/h
V	Commercial speed of the TU	km/h
V_{max}	Highest operational speed of the TU	km/h
w	Waiting time at a stop when $f < \bar{f}$	h
y	Fixed bidirectional demand	pax/h
α	Fraction of demand in the most loaded segment of the line	-
β	Boarding and alighting time per passenger and TU	h/(pax-TU)
β_v	Boarding and alighting time per passenger and vehicle	h/(pax-veh)
γ	Ratio of the off-peak demand to the peak demand	-
δ	Crowding penalty function	-
ϵ	Rate of the average waiting time to the headway	-
θ	Average occupancy rate	-
μ	Discount factor of the waiting time under timetable behaviour	-
ν	Spare capacity factor	-
ρ	Slope of the linear part of the crowding penalty function	-
$\chi^{(p,o)}$	Ratio of peak or off-peak hours to total service hours	-

3.1. Model I: Base model

We report the notation and the formulae of the Tirachini et al. (2010b) model with minor modifications to account for the extensions that we propose in the following sections. In this model a transit line of length L with a stop spacing equal to d serves a fixed bidirectional demand y. Passengers access and egress the line by walking to and from the nearest stop at speed v. Hence the average walking distance is $d/4$ at the origin and at the destination, the average total walking length is $d/2$, and the average access and egress time t_a of a passenger is

$$t_a = \frac{d}{2v} \qquad d[\text{km}], v[\text{km/h}]. \tag{1}$$

The cost of one unit of walking time is expressed by the parameter P_a, and the access and egress cost C_a borne by y passengers is

$$C_a = P_a t_a y \qquad P_a[\$/(\text{pax-h})], t_a[\text{h}], y[\text{pax/h}]. \tag{2}$$

Waiting time depends on the frequency f, expressed as the number of transit unit (TU) per hour. The concept of transit unit, see Vuchic (2005), defines a set of n vehicles traveling physically linked. For single-vehicle operations, such as for road modes, n is equal to one, whereas for rail modes n can be larger than one. Thus TU is the common concept for single vehicles and trains used on a transit line. Passenger behaviour differs for low and high frequencies. In the case of high frequencies, passengers arrive at stops at a constant rate and the average waiting time t_w can be modelled as a fraction $\epsilon \geq 1/2$ of the expected headway equal to $1/f$. Values of ϵ strictly larger than $1/2$ can model cases where the headways have a large variance. In

the case of low frequencies, passengers follow timetables and arrive at stops w minutes before the expected time of service. The waiting time saved by this behaviour still has a cost for the passenger but is discounted by a factor μ less than one. The threshold frequency for these two behaviour regimes is defined by $\bar{f}$, for example five transit units per hour, which results in a headway of 12 minutes. The average waiting time t_w of a passenger is

$$t_w = \begin{cases} t_w^a = w + \mu\dfrac{\epsilon}{f} & \text{if } \; f < \bar{f} \\ t_w^b = \dfrac{\epsilon}{f} & \text{if } \; f \geq \bar{f},\, f[\text{TU/h}], w[h] \end{cases}, \tag{3}$$

where we indicate by a the case with frequency lower than $\bar{f}$, and by b otherwise. The average cost of waiting C_w borne by y passengers is

$$C_w = P_w t_w y \qquad P_w[\$/(\text{pax-h})], t_w[\text{h}], y[\text{pax/h}], \tag{4}$$

where P_w is the cost of one waiting time unit.

The average in-vehicle time t_v of a passenger is modelled as a fraction of the cycle time t_c. The cycle time is the sum of the running time between stations, including acceleration and deceleration phases, and the dwell time for boarding and alighting. The running time, denoted R, is computed by assuming an average running speed S, and thus $R = 2L/S$. The dwell time depends on β, the boarding and alighting time per passenger of a TU, and the number of passengers using a TU, given by y/f. The boarding and alighting time of a TU depends on the number $\bar{n}$ of vehicles per TU, which is a fixed value in this model. We assume a boarding and alighting time per passenger of a vehicle as equal to β_v, thus $\beta = \beta_v/\bar{n}$.

The cycle time is

$$t_c = \frac{y}{f}\beta + R \qquad f[\text{TU/h}], R[\text{h}], \beta[\text{h/pax}], y[\text{pax/h}]. \tag{5}$$

The in-vehicle time is a fraction of the cycle time, equal to the ratio of the average trip length l, to the total distance $2L$ covered by a TU in a cycle:

$$t_v = \frac{l}{2L} t_c \qquad l, L[\text{km}], t_c[\text{h}]. \tag{6}$$

Let P_v be the cost of one unit of in-vehicle time, then the cost of in-vehicle time, C_v, is

$$C_v = P_v t_v y \qquad P_v[\$/(\text{pax-h})], t_v[\text{h}], y[\text{pax/h}]. \tag{7}$$

The total passenger cost C_p is then

$$C_p = C_a + C_w + C_v \qquad C_a, C_w, C_v[\$/\text{h}]. \tag{8}$$

The operator cost is divided into three components. The first comprises non-operational costs such as land and infrastructure capital costs. The second depends on the fleet size and reflects crew and TU capital costs. The third accounts for running costs such as fuel, tyres, lubricants, etc. Let c_0 be the fixed operator cost normalized for one hour, c_1 be the unit operator cost per TU-hour, and c_2 be the unit operator cost per TU-km. The fleet size B is the product of frequency and cycle time: $B = ft_c$. The amount of TU-km is the product of the commercial speed V and the fleet size. The commercial speed is obtained by dividing the total length $2L$ by the cycle time. Thus, the amount of TU-km is $V \times B = 2L/t_c \times ft_c = 2Lf$. The operator cost C_o is then

$$C_o = c_0 + c_1 f t_c + 2c_2 L f \qquad c_0[\$/\text{h}], c_1[\$/(\text{TU-h})], c_2[\$/(\text{TU-km})], \\ f[\text{TU/h}], t_c[\text{h}], L[\text{km}]. \tag{9}$$

The total cost C_{tot} is the sum of passenger and operator costs and it is a function of the frequency:

$$C_{tot}(f) = \begin{cases} C^a_{tot}(f) & f < \bar{f} \\ C^b_{tot}(f) & f \geq \bar{f} \end{cases}, \tag{10}$$

where $C^a_{tot}(f)$ and $C^b_{tot}(f)$ are defined as

$$C^{(a,b)}_{tot}(f) \quad = \quad C_a + P_w t^{(a,b)}_w y + C_v(f) + C_o(f). \tag{11}$$

The frequency is constrained to be less than or equal to a mode dependent value f_{max}, and equal to or larger than a value f_{min} which accounts for capacity as follows. We denote by k_v the capacity of a vehicle, and hence the capacity K of a TU is equal to $k_v \times \bar{n}$. Let αy be the largest load served by the line, where $\alpha \leq 1$, and ν be a spare capacity factor, for example $\nu = 0.9$, which accounts for random demand fluctuations. Thus, f_{min} and the feasible range for frequency are

$$f_{min} = \frac{\alpha y}{\nu K} \leq f \leq f_{max} \qquad K[\text{pax/TU}], y[\text{pax/h}]. \tag{12}$$

Model I minimizes (10) under constraints (12). To solve this problem, first observe that $C^a_{tot}(f)$ and $C^b_{tot}(f)$ are both convex functions, hence their unconstrained optimal solutions, f^a_{unc} and f^b_{unc}, can be stated in closed form:

$$f^{(a,b)}_{unc} = \sqrt{\frac{P_w \mu^{(a,b)} \epsilon y + P_v \frac{l\beta}{2L} y^2}{c_1 R + 2c_2 L}}, \tag{13}$$

where $\mu^a = \mu$, and $\mu^b = 1$.

Taking convexity into account, it is easy to compute the respective minima under bound constraints, and then the optimal frequency f^* of Model I is obtained by exploration of these minima.

3.2. Model II: Extension to optimal spacing

As reported in Section 4, a modal comparison through Model I is sensitive to the stop spacing d. Model I cannot be used directly to optimize over d since its definition of cycle time is independent from d. Hence the minimization of the total cost function (10) over d will lead to an unbounded problem. In the following, we propose an extension, Model II, where optimal stop spacing considers TU kinematics. We assume that a TU leaves a stop accelerating up to a speed V_{max}, travels at this speed, and then decelerates to halt at the next stop. Let $\bar{a}$ and $\bar{b}$ be the average acceleration and deceleration rates of a TU. The incremental time loss caused by the acceleration and deceleration phases is denoted by T_a, and is equal to

$$T_a = \frac{V_{max}}{2}\left(\frac{1}{\bar{a}} + \frac{1}{\bar{b}}\right) \qquad V_{max}[\text{km/h}], \bar{a}, \bar{b}[\text{km/h}^2] \tag{14}$$

(see e.g. Vuchic and Newell (1968)). Reaching the speed V_{max} requires a stop spacing larger than a threshold value d_{min}, which depends on acceleration and deceleration rates:

$$d_{min} = \frac{V_{max}^2}{2}\left(\frac{1}{\bar{a}} + \frac{1}{\bar{b}}\right) \qquad V_{max}[\text{km/h}], \bar{a}, \bar{b}[\text{km/h}^2]. \tag{15}$$

We add to the standing time a fixed component t_d, which accounts for the opening and closing of doors, and we denote by T_l the lost time for acceleration, deceleration, and door opening and closing:

$$T_l = T_a + t_d \qquad T_a[\text{h}], t_d[\text{h}]. \tag{16}$$

Since the number of stops is equal to $2L/d$, the new cycle time, t_{c2}, is

$$t_{c2} = \frac{y}{f}\beta + \frac{2L}{d}T_l + \frac{2L}{V_{max}} \qquad y[\text{pax/h}], f[\text{TU/h}], \beta[\text{h/pax}], \; d, L[\text{km}], T_l[\text{h}], V_{max}[\text{km/h}]. \tag{17}$$

The cycle time impacts on both passenger and operator costs. The passenger in-vehicle cost C_{v2} becomes

$$C_{v2}(f,d) = P_v \frac{l}{2L} t_{c2} y \qquad P_v[\$/(\text{pax-h})], l, L[\text{km}], t_{c2}[\text{h}], y[\text{pax/h}]. \quad (18)$$

The updated total passenger cost, $C_{p2}(f,d)$, is now a function of both f and d:

$$C_{p2}(f,d) = C_a(d) + C_w(f) + C_{v2}(f,d). \quad (19)$$

The varying stop spacing requires a disaggregation of the fixed infrastructure cost into two components: c_{0l}, the fixed hourly cost of the line, and c_{0s}, the fixed hourly cost of a stop. The terms c_{0l} and c_{0s} account for construction and maintenance costs of the line and of a stop, respectively. Thus, the updated operator cost C_{o2} is a function of both f and d:

$$C_{o2}(f,d) = c_{0l} + c_{0s}\frac{2L}{d} + c_1 t_{c2} f + 2c_2 L f \qquad c_{0l}[\$/\text{h}], c_{0s}[\$/(\text{stop-h})], \\ c_1[\$/(\text{TU-h})], c_2[\$/(\text{TU-km})], \\ d, L[\text{km}], f[\text{TU/h}], t_{c2}[\text{h}]. \quad (20)$$

To state the objective function in a more compact way, we introduce in Appendix A the coefficients $a_0^{(a,b)}, a_1, ..., a_5$. The total cost C_{tot2}, the sum of passenger and operator costs, is a function of frequency and stop spacing:

$$C_{tot2}(f,d) = \begin{cases} C^a_{tot2}(f,d) & f < \bar{f} \\ C^b_{tot2}(f,d) & f \geq \bar{f} \end{cases}, \quad (21)$$

where $C^a_{tot2}(f,d)$ and $C^b_{tot2}(f,d)$ are

$$C^{(a,b)}_{tot2}(f,d) = a_0^{(a,b)} + a_1 d + a_2 \frac{1}{d} + a_3 f + a_4^{(a,b)} \frac{1}{f} + a_5 \frac{f}{d}. \quad (22)$$

Model II follows:

$$\text{minimize}\, C_{tot2}(f,d) \tag{23}$$

subject to

$$\begin{aligned} d_{min} &\leq d, & (24)\\ \frac{\alpha y}{\nu K} &\leq f \leq f_{max}. & (25) \end{aligned}$$

Constraint (24) sets minimal stop spacing, and constraints (25) enforce minimal and maximal frequency.

The components of the objective function of Model II, unlike those of Model I, are not necessarily convex functions, see Appendix B. Hence, we propose $\tilde{C}_{tot2}$ as a lower convex envelope of the components of (21):

$$\tilde{C}_{tot2}(f,d) = \begin{cases} \tilde{C}^a_{tot2}(f,d) & f < \bar{f} \\ \tilde{C}^b_{tot2}(f,d) & f \geq \bar{f} \end{cases}, \tag{26}$$

where $\tilde{C}^a_{tot2}(f,d)$ and $\tilde{C}^b_{tot2}(f,d)$ are

$$\tilde{C}^{(a,b)}_{tot2}(f,d) = a_0^{(a,b)} + a_1 d + a_2 \frac{1}{d} + a_3 f + a_4^{(a,b)} \frac{1}{f} + a_5 \frac{f_{min}}{d}. \tag{27}$$

Lemma 1. *$\tilde{C}_{tot2}$ is a lower convex envelope of C_{tot2}.*

Proof — First, observe that the term f/d in (22) induces non-convexity. Second, note that the convex term f_{min}/d in (27) never exceeds f/d. □

Because of convexity, the unconstrained optimal solutions of (27) can be analytically determined. We denote by $\tilde{d}_{unc}$ the unconstrained optimal stop spacing of (27):

$$\tilde{d}_{unc} = \sqrt{\frac{a_2 + f_{min} a_5}{a_1}} = \sqrt{2vT_l \left(\frac{P_v l}{P_a} + \frac{2c_1 L f_{min}}{y P_a} + \frac{2c_{0s} L}{y P_a T_l} \right)}. \tag{28}$$

The unconstrained optimal frequencies of (27) are then

$$\tilde{f}_{unc}^{(a,b)} = \sqrt{\frac{a_4^{(a,b)}}{a_3}} = \sqrt{\frac{P_w \mu^{(a,b)} \epsilon y + P_v \frac{l\beta}{2L} y^2}{c_1 \frac{2L}{V_{max}} + 2c_2 L}}. \tag{29}$$

The optimal solution of the constrained lower convex envelope can be computed as follows. We observe that (27) is also separable, and whenever the unconstrained solution falls outside the box constraints, the optimal projection on the frontier is straightforward. The optimal solution $(\tilde{d}, \tilde{f})$ of the constrained lower convex envelope can be computed by examining at most two feasible solutions $(\tilde{d}, \tilde{f}^a)$ and $(\tilde{d}, \tilde{f}^b)$:

$$\tilde{d} = \begin{cases} d_{min} & d_{unc} < d_{min} \\ d_{unc} & d_{unc} \geq d_{min} \end{cases}, \tag{30}$$

$$\tilde{f}^{(a,b)} = \begin{cases} f_{min} & f_{unc}^{(a,b)} < f_{min} \\ f_{unc}^{(a,b)} & f_{min} \leq f_{unc}^{(a,b)} \leq f_{max}^{(a,b)} \\ f_{max}^{(a,b)} & f_{unc}^{(a,b)} > f_{max}^{(a,b)} \end{cases}, \tag{31}$$

where $f_{max}^a = \bar{f}$ and $f_{max}^b = f_{max}$.

A heuristic solution of Model II can be determined by applying to the starting solution $(\tilde{f}, \tilde{d})$ a quasi-Newton code for bound-constrained optimization. We denote by $(\hat{f}, \hat{d})$ the solution found by this procedure, and by (f^*, d^*) the unknown optimal solution. We then obtain a lower bound L_{b2} for Model II, and an easy-to-implement solution method:

$$L_{b2} = \tilde{C}_{tot2}(\tilde{f}, \tilde{d}) \leq C_{tot2}(f^*, d^*) \leq C_{tot2}(\hat{f}, \hat{d}) \leq C_{tot2}(\tilde{f}, \tilde{d}). \tag{32}$$

Proposition 2. *The optimality gap of the proposed procedure is*

$$\frac{C_{tot2}(\hat{f},\hat{d}) - L_{b2}}{C_{tot2}(\hat{f},\hat{d})}. \tag{33}$$

Hence (27) is a lower convex approximation of (22).

In Section 4 we show that the optimality gap is negligible on typical instances. Moreover, the gap of the approximate analytic solution $(\tilde{f},\tilde{d})$ with respect to the heuristic solution is insignificant. Hence the approximate analytic solution can also be used to discuss properties of the optimal frequency and stop spacing. Equation (28) indicates that the optimal stop spacing does not vary significantly with the level of demand beyond a low demand threshold. We note that the demand appears at the denominator of two terms with mode specific parameters at the numerator, and hence reduces the role of these parameters for all but the low demand levels. Thus, with the exception of T_l, the other mode specific parameters are toned down in (28) for medium and large demand levels. Let i and j be the indices of two modes to be compared. The ratio between their optimal stop spacings can be approximated as

$$\frac{d^{i*}}{d^{j*}} \approx \frac{\tilde{d}^i_{unc}}{\tilde{d}^j_{unc}} \approx \sqrt{\frac{T_l^i}{T_l^j}}. \tag{34}$$

Equation (34) expresses an interesting property of optimal stop spacing. We observe that the variation of the T_l of different modes is mainly related to the different maximal speeds V^i_{max} and V^j_{max}, see equations (14) and (16). As a consequence, the ratio between the optimal spacings of two modes can be further approximated as $\sqrt{V^i_{max}/V^j_{max}}$. A faster mode requires longer stop spacing, but the ratio of the optimal stop spacings among two modes

follows a square root formula of their speed ratio, i.e. a sublinear relationship. Exogenous values of stop spacing that do not reflect this property would violate the optimal structure of the model. We refer to Tirachini (2014) for further analysis on stop spacing, including the distribution of passenger arrivals at stops, queuing delays, fare collecting systems, and the case of multi-berth stops.

The unconstrained optimal frequencies of the approximation (29) are the same as the unconstrained optimal frequencies of Model I defined by (13). Hence Models I and II yield similar frequencies. A closer examination of (29) highlights a shortcoming of these two models. For moderate centripetal demand, and typical values of parameters, equations (29) yield frequencies lower than those required to satisfy typical passenger critical loads, i.e. $\tilde{f}_{unc} \leq f_{min}$. In these cases, Model I is redundant, frequency is implicitly set by the critical capacity, and the scope of Model II is limited to stop spacing. Moreover, the average occupancy rate θ is set to the highest value θ_{max}:

$$\theta(f_{min}) = \frac{ly}{2LKf_{min}} = \frac{l\nu}{2L\alpha} = \theta_{max}. \tag{35}$$

We observe that for some parameters, θ_{max} could be close to one and would likely result in overcrowding.

3.3. Model III: Extension to crowding cost and optimal train length

We now consider the passenger cost of in-vehicle crowding. In-vehicle travel time is perceived by passengers as less convenient when the occupancy rate θ is larger than a threshold value θ_{min}. Usually, this threshold value is set in such a way that $\theta = \theta_{min}$ indicates saturation of seating capacity, and $\theta > \theta_{min}$ indicates that some passengers must stand. Road and rail modes can

manage crowding in different ways. Road modes can only attempt to offer a higher frequency, whereas rail modes can provide leverage on both frequency and the number of vehicles per TU. We introduce the integer variable n which indicates the number of vehicles per TU and ranges from n_{min} to n_{max}. Road modes are a special case with $n_{min} = n_{max} = 1$. We observe that different bus lengths, and hence capacities, could be analyzed for the bus services. This can be done by casting as different modes the services with different bus types. For example, a mode BRT1 with regular buses, a mode BRT2 with articulated buses, etc. Road services with different capacities can be modeled as different modes. Rail services can dynamically adjust capacities, and therefore it is useful to model the train length as a variable. Moreover, longer trains affects the boarding and alighting times since a higher number of doors becomes available. Let β_v be the boarding and alighting time per vehicle, the new cycle time t_{c3} is

$$t_{c3}(f,d,n) = \frac{y}{nf}\beta_v + \frac{2L}{d}T_l + \frac{2L}{V_{max}} \qquad \begin{array}{l} y[\text{pax/h}], f[\text{TU/h}], \beta_v[\text{h-veh/pax}], d, L[\text{km}], \\ T_l[\text{h}], V_{max}[\text{km/h}], n[\text{veh/TU}]. \end{array} \tag{36}$$

The value P_v of the in-vehicle travel time is multiplied by a crowding penalty which depends on the average occupancy rate θ:

$$\theta(f,n) = \frac{ly}{2Lk_v nf}. \tag{37}$$

We define a crowding penalty function δ as piecewise linear (see Wardman and Whelan (2010) for a review which validates this approach). There is no penalty up to an average occupancy rate of θ_{min}, e.g. it is equal to 0.3. For larger values of θ the penalty increases linearly with a slope value ρ, see Figure 1 for an example. The above definition of the average occupancy rate

assumes a uniform demand distribution and underestimates crowding on the segments affected by centripetal demand (see Li and Hensher (2013) for a critical appraisal of over-aggregation of crowding indices). We propose to adjust the parameter ρ to account for this, i.e. a higher value of ρ under centripetal demand should be used, other things being equal. Formally, the

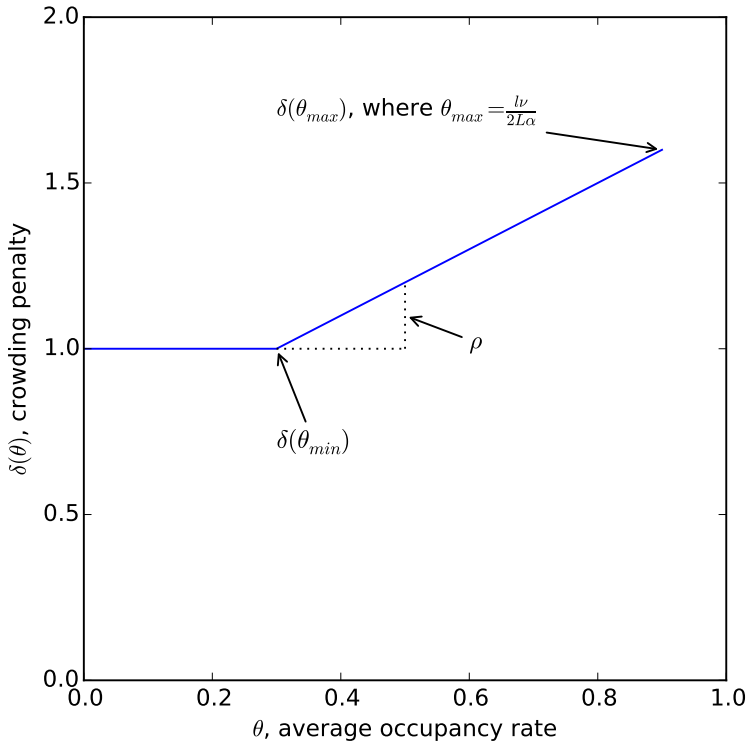

Figure 1: Example of the crowding penalty function

penalty function is

$$\delta(f,n) = \begin{cases} 1 + \rho\left(\theta - \theta_{min}\right) = \xi + \rho\dfrac{ly}{2Lk_v nf} & nf \leq \dfrac{ly}{2\theta_{min} Lk_v} \\ 1 & \text{otherwise} \end{cases}, \qquad (38)$$

where $\xi = 1 - \theta_{min}\rho$ is a parameter introduced for notational compactness.

In contrast to Tirachini et al. (2010b), we directly use the piecewise linear penalty function without resorting to an approximation by a quadratic function. In spite of this, we show that the complication of the crowding penalty function does not preclude an analytical solution, albeit approximate. Thus, the main differences of our approach with respect to that of Tirachini et al. (2010b) are twofold: we analytically assess crowding instead of resorting to a numerical appraisal, and we combine crowding with optimal stop spacing and train length. The passenger in-vehicle cost C_{v3} then becomes

$$C_{v3}(f, d, n) = P_v \delta(f, n) \frac{l}{2L} t_{c3}(f, d, n) y. \tag{39}$$

The updated total passenger cost $C_{p3}(f, d, n)$ is now a function of f, d, and n:

$$C_{p3}(f, d, n) = C_a(d) + C_w(f) + C_{v3}(f, d, n). \tag{40}$$

Varying the train length changes the structure of the operator's cost function. We have to distinguish between the TU costs and the vehicle costs. Let c_{1t} be the unit operator cost per TU-hour, and let c_{1v} be the unit operator cost per vehicle-hour. The former parameter accounts for the crew cost, and the latter for the capital cost of vehicles. Let c_{2v} be the unit operator cost per vehicle-km. The deployed fleet size B of TUs is, as in the previous models, the product of frequency and cycle time: $B = f t_{c3}$. The vehicle fleet size is equal to nB, and the amount of vehicle-km is $2nLf$. The operator cost C_{o3} is then

$$C_{o3}(f, d, n) = c_{0l} + c_{0s} \frac{2L}{d} + c_{1t} f t_{c3}(f, d, n) + c_{1v} n f t_{c3}(f, d, n) + 2 c_{2v} n L f. \tag{41}$$

In the case of road modes where $n = 1$, we have $c_{1t} + c_{1v} = c_1$, and $c_{2v} = c_2$, thus (41) is similar to (20).

The total cost C_{tot3}, the sum of passenger and operator costs, is a function of frequency, stop spacing, and number of vehicles per TU. Model III follows:

$$\text{minimize}\, C_{tot3}(f, d, n) \tag{42}$$

subject to

$$\begin{aligned} d_{min} &\leq d, & (43)\\ \frac{\alpha y}{\nu n k_v} &\leq f \leq f_{max}, & (44)\\ n_{min} &\leq n \leq n_{max},\, n \in \mathbb{N}. & (45) \end{aligned}$$

Constraint (43) sets minimal stop spacing. Constraints (44) enforce minimal and maximal frequency, and constraints (45) specify the feasible range of TU length.

We now define an approximation scheme for Model III. We first consider the domain where the variables n and f satisfy the following condition:

$$\theta \geq \theta_{min} \text{ i.e. } nf \leq \frac{ly}{2\theta_{min} L k_v}. \tag{46}$$

By so doing, we consider the linear part of the crowding penalty function δ. To state the objective function in a more compact way, we introduce in Appendix C the coefficients $a_{30}^{(a,b)}, a_{31}, ..., a_{311}$. We observe that outside the domain (46), the crowding penalty is not active, and the above coefficients must be computed by setting $\rho = 0$ and $\xi = 1$. The objective function of

Model III, C_{tot3}, can be restated as follows:

$$C_{tot3}(f,d,n) = \begin{cases} C^a_{tot3}(f,d,n) & f < \bar{f}, nf \leq \dfrac{ly}{2\theta_{min} L k_v} \\ C^b_{tot3}(f,d,n) & f \geq \bar{f}, nf \leq \dfrac{ly}{2\theta_{min} L k_v} \\ C^a_{tot3}(f,d,n)\Big|_{\rho=0} & f < \bar{f}, nf > \dfrac{ly}{2\theta_{min} L k_v} \\ C^b_{tot3}(f,d,n)\Big|_{\rho=0} & f \geq \bar{f}, nf > \dfrac{ly}{2\theta_{min} L k_v} \end{cases}, \tag{47}$$

where $C^a_{tot3}(f,d,n)$ and $C^b_{tot3}(f,d,n)$ are

$$\begin{aligned} C^{(a,b)}_{tot3r}(f,d,n) &= a^{(a,b)}_{30} + a_{31}d + a_{32}\frac{1}{d} + a_{33}f + a^{(a,b)}_{34}\frac{1}{f} + a_{35}\frac{f}{d} + a_{36}\frac{1}{n} + a_{37}nf + \\ &\quad + a_{38}\frac{1}{nf} + a_{39}\frac{nf}{d} + a_{310}\frac{1}{nfd} + a_{311}\frac{1}{(nf)^2}. \end{aligned} \tag{48}$$

Once the variable n is fixed to $\tilde{n}$, a lower convex envelope of (47) can be obtained as for Model II:

$$\tilde{C}_{tot3}(f,d,\tilde{n}) = \begin{cases} \tilde{C}^a_{tot3}(f,d,\tilde{n}) & f < \bar{f}, f \leq \dfrac{ly}{2\theta_{min} L k_v \tilde{n}} \\ \tilde{C}^b_{tot3}(f,d,\tilde{n}) & f \geq \bar{f}, f \leq \dfrac{ly}{2\theta_{min} L k_v \tilde{n}} \\ \tilde{C}^a_{tot3}(f,d,\tilde{n})\Big|_{\rho=0} & f < \bar{f}, f > \dfrac{ly}{2\theta_{min} L k_v \tilde{n}} \\ \tilde{C}^b_{tot3}(f,d,\tilde{n})\Big|_{\rho=0} & f \geq \bar{f}, f > \dfrac{ly}{2\theta_{min} L k_v \tilde{n}} \end{cases}, \tag{49}$$

where $\tilde{C}^a_{tot3}(f,d,\tilde{n})$ and $\tilde{C}^b_{tot3}(f,d,\tilde{n})$ are

$$\begin{aligned} \tilde{C}^{(a,b)}_{tot3}(f,d,\tilde{n}) &= \left(a^{(a,b)}_{30} + \frac{a_{36}}{\tilde{n}}\right) + a_{31}d + \left(a_{32} + a_{35}f_{min} + a_{39}\tilde{n}f_{min} + \frac{a_{310}}{\tilde{n}f_{max}}\right)\frac{1}{d} + \\ &\quad + \left(a_{33} + a_{37}\tilde{n}\right)f + \left(a^{(a,b)}_{34} + \frac{a_{38}}{\tilde{n}} + \frac{a_{311}}{(\tilde{n})^2 f_{max}}\right)\frac{1}{f}. \end{aligned} \tag{50}$$

Lemma 3. *$\tilde{C}_{tot3}(f,d,\tilde{n})$ is a separable lower convex envelope of $C_{tot3}(f,d,\tilde{n})$.*

Model III can be solved by a procedure similar to that devised for Model II for each feasible value of the integer variable n. For the approximation, the unconstrained optimal frequencies are denoted as $\tilde{f}^{(a,b)}_{unc3}(\tilde{n})$, and the unconstrained optimal stop spacing is denoted as $\tilde{d}_{unc3}$. The optimal solution of the constrained lower convex envelope can be easily computed by explorations of the relevant minima which are at most four for each feasible value of n. Let $\tilde{d}(\tilde{n})$ be the optimal spacing, and let $\tilde{f}(\tilde{n})$ be the optimal frequency of the constrained lower convex envelope for the TU length $\tilde{n}$. A heuristic solution of Model III can be determined by applying to each starting solution $(\tilde{f}, \tilde{d}, \tilde{n})$ a quasi-Newton code for bound-constrained optimization on the continuous variables f and d. We indicate by $(\hat{f}, \hat{d}, \tilde{n})$ the best solution found, and by $(f^*, d^*, \tilde{n})$ the unknown optimal solution for the value of TU length $\tilde{n}$. By so doing, we obtain a lower bound L_{b3} for Model III:

$$\begin{aligned} L_{b3} &= \min_{\tilde{n}\in(n_{min},\ldots,n_{max})} L_{b3}(\tilde{n}) = \tilde{C}_{tot3}(\tilde{f}, \tilde{d}, \tilde{n}) \leq C_{tot3}(f^*, d^*, \tilde{n}) \\ &\leq C_{tot3}(\hat{f}, \hat{d}, \tilde{n}) \leq C_{tot3}(\tilde{f}, \tilde{d}, \tilde{n}). \end{aligned} \tag{51}$$

The solution method is outlined in Algorithm 1 which iterates over n until the best solution $(\hat{f}, \hat{d}, \hat{n})$ is found. The optimality gap can then be computed.

Proposition 4. *The optimality gap of the proposed procedure is*

$$\frac{C_{tot3}(\hat{f}, \hat{d}, \hat{n}) - L_{b3}}{C_{tot3}(\hat{f}, \hat{d}, \hat{n})}. \tag{52}$$

Hence $\tilde{C}_{tot3}$ is a separable lower convex approximation of C_{tot3}.

In Section 4 we show that the optimality gap is very small on typical instances. Moreover, the gap of the approximate analytic solution $(\tilde{f}, \tilde{d}, \tilde{n})$

Algorithm 1 Solve Model III

$Best \leftarrow$ a very large number.

for $\tilde{n} \in (n_{min}, ..., n_{max})$ **do**

Compute $\tilde{f}^{(a,b)}_{unc3}(\tilde{n})$ and $\tilde{d}_{unc3}(\tilde{n})$ by (53) and (54), respectively.

Compute $\tilde{f}(\tilde{n})$ and $\tilde{d}(\tilde{n})$ by convexity arguments for separable functions.

Find $(\hat{f}, \hat{d})$ starting from $(\tilde{f}(\tilde{n}), \tilde{d}(\tilde{n}))$ by a bound constrained optimization solver.

if $C_{tot3}(\hat{f}, \hat{d}, \tilde{n}) < Best$ **then**

Update $Best$, store solution, compute and store gaps.

end if

end for

return best solution found, and gaps.

with respect to the heuristic solution is negligible. Hence the approximate analytic solution can also be used to derive properties of the optimal values. The unconstrained optimal frequencies of the approximation are

$$\tilde{f}^{(a,b)}_{unc3}(\tilde{n}) = \sqrt{\frac{P_w \mu^{(a,b)} \epsilon y + \frac{P_v l}{2L}\left(\frac{\xi \beta_v y^2}{\tilde{n}} + \rho \frac{l}{k_v \tilde{n}}\left(\frac{y^2}{V_{max}} + \frac{y^3 \beta_v}{2L\tilde{n} f_{max}}\right)\right)}{\frac{2L}{V_{max}}(c_{1t} + \tilde{n} c_{1v}) + 2c_{2v}\tilde{n}L}}. \quad (53)$$

The structure of (53) differs from that of (29) because of the two new terms under the square root related to the crowding penalty. One of these terms increases with the square of the demand, and the other increases with the cube of the demand, although it is multiplied by a very small number such as β_v. These terms have a relevant impact in increasing the optimal frequency with respect to formula (29). Modeling a crowding penalty moves away the

optimal frequency from the minimal values induced by the critical capacity. Rail modes can provide leverage on train length to counteract raising frequency since train length appears at the denominator in several terms. Overall, it is not anymore true that an increase in demand requires a less than proportional increase in frequency, the classic result of Mohring (1972). Faster modes with large and flexible capacity may yield economies of scale (observe the role played in (53) by V_{max}, k_v, and n).

The unconstrained optimal stop spacing is

$$\tilde{d}_{unc3}(\tilde{n}) = \sqrt{2vT_l \left(\frac{P_v}{P_a} l\xi + \frac{2Lf_{min}}{yP_a} \left(c_{1t} + \tilde{n} c_{1v} \right) + \frac{2c_{0s}L}{yP_aT_l} + \rho \frac{l^2 yP_v}{2P_aLk_v \tilde{n} f_{max}} \right)}. \quad (54)$$

The structure of (54) differs from that of (28) because of the two new terms under the square root: a term related to the cost of a stop which decreases with the demand, and a term related to the crowding penalty which increases with the demand. However, both terms have typically small coefficients, hence the stop spacing of Model III is only slightly larger than that of Model II. Moreover, the structure of (54) confirms the approximation provided by (34).

3.4. Model IV: Extension to multiple periods

The single period assumption can be criticized on several counts. First, we observe that amortization factors play a relevant role when comparing technologies that differ in their infrastructure and vehicle costs. Hence the number of service hours per year is a critical parameter. Single period models must account for a number of "equivalent" service hours smaller than the effective hours. On the other hand, a single period has a higher number of "equivalent" peak hours, which helps spreading a higher fleet cost. Moreover,

as observed by Bruun (2005), both the demand level and the shape of the demand profile determine operating costs. Rail modes can accommodate demand variations through the addition and removal of carriages from trains. To disentangle these issues, a modal comparison between road and rail modes must represent a time-varying demand profile. We remove this limitation of the previous models by introducing a two-period model. We will show how formulae for multi-period cases can be obtained as simple extensions of the two-period case.

We assume a base demand in off-peak hours equal to γy, where γ is a positive parameter smaller than one. The ratio of peak hours to total service hours is denoted by χ^p, and $\chi^o = 1 - \chi^p$ is the ratio of off-peak hours to total total service hours. The components of passenger and operator costs are now described. The access cost C_{a4} is the weighted sum of two terms:

$$C_{a4} = \chi^p P_a \frac{d}{2v} y + \chi^o P_a \frac{d}{2v} \gamma y = P_a \frac{d}{2v} y(\chi^p + \gamma \chi^o). \tag{55}$$

The cost of waiting for a single period, defined by equation (4), is a function of frequency and it is directly proportional to the demand level. We now have two frequencies for the two periods, namely f^p for peak hours, and f^o for off-peak hours. The cost of waiting for the two periods C_{w4} is a weighted sum of the cost of waiting for the two periods:

$$C_{w4} = \chi^p C_w(f^p) + \chi^o \gamma C_w(f^o). \tag{56}$$

The cycle time has one of three terms dependent on the demand level. Hence we introduce the new cycle time $t_{c4}^{(p,o)}$, for peak and off-peak periods:

$$t_{c4}^{(p,o)}(f^{(p,o)}, d, n^{(p,o)}) = \frac{y^{(p,o)}}{n^{(p,o)} f^{(p,o)}} \beta_v + \frac{2L}{d} T_l + \frac{2L}{V_{max}}, \tag{57}$$

where $y^p = y$ and $y^o = \gamma y$. The average occupancy rate, defined by equation (37), is directly proportional to the demand level, hence requires only a scaling by a factor γ for the off-peak period. We denote by $\delta^{(p,o)}$ the resulting two penalty functions. The TU length for the peak period is n^p, and n^o denotes the TU length for the off-peak-period. The new in-vehicle cost C_{v4} is

$$C_{v4}(f^p, f^o, d, n^p, n^o) = P_v \frac{ly}{2L}(\chi^p \delta^p(f^p, n^p) t^p_{c4}(f^p, d, n^p) + \\ + \chi^o \gamma \delta^o(f^o, n^o) t^o_{c4}(f^o, d, n^o)), \tag{58}$$

and the operator cost C_{o4} is

$$\begin{aligned} C_{o4}(f^p, f^o, d, n^p, n^o) &= c_{0l} + c_{0s}\frac{2L}{d} + c_{1v} n^p f^p t^p_{c4}(f^p, d, n^p) + c_{1t}\chi^p f^p t^p_{c4}(f^p, d, n^p) + \\ &\quad + c_{1t}\chi^o f^o t^o_{c4}(f^o, d, n^o) + 2c_{2v} L(\chi^p n^p f^p + \chi^o n^o f^o). \end{aligned} \tag{59}$$

The total cost C_{tot4}, sum of passenger and operator costs, is a function of peak and off-peak frequencies, stop spacing, and peak and off-peak number of vehicles per TU. Model IV follows:

$$\text{minimize}\, C_{tot4}(f^p, f^o, d, n^p, n^o) \tag{60}$$

subject to

$$\begin{aligned} d_{min} &\leq d, && (61)\\ \frac{\alpha y^{(p,o)}}{\nu n^{(p,o)} k_v} &\leq f^{(p,o)} \leq f_{max}, && (62)\\ n_{min} &\leq n^{(p,o)} \leq n_{max},\ n^{(p,o)} \in \mathbb{N}, && (63)\\ n^p f^p t^p_{c4}(f^p, d, n^p) &\geq n^o f^o t^o_{c4}(f^o, d, n^o). && (64) \end{aligned}$$

Constraint (61) sets the minimal stop spacing. Constraints (62) enforce minimal and maximal values for peak and off-peak frequencies. Constraints (63)

specify the feasible range of TU lengths, and constraint (64) ensures that the maximal fleet is deployed at peak time.

The lower convex envelope and the solution algorithm for Model IV are a straightforward extension of those of Model III, and hence their description is omitted for the sake of brevity. However, in the case of Model IV we cannot always conclude optimality of the constrained lower convex envelope because of the non-linear constraint (64). The optimality gap cannot be computed whenever the unconstrained solution falls outside the domain of Model IV. In the following we report the optimal solution of the unconstrained lower convex envelope.

The unconstrained optimal peak frequency of the lower convex envelope is

$$\tilde{f}^{p(a,b)}_{unc4}(\tilde{n}^p) = \sqrt{\frac{P_w \mu^{(a,b)} \epsilon y + \frac{P_v l}{2L}\left(\frac{\xi \beta_v y^2}{\tilde{n}^p} + \rho \frac{l}{k_v \tilde{n}^p}\left(\frac{y^2}{V_{max}} + \frac{y^3 \beta_v}{2L\tilde{n}^p f_{max}}\right)\right)}{\frac{2L}{V_{max}}(c_{1t} + \frac{\tilde{n}^p c_{1v}}{\chi^p}) + 2c_{2v}\tilde{n}^p L}}. \quad (65)$$

Equation (65) differs from (53) in the denominator where the coefficient c_{1v}, which accounts for the fixed cost of vehicles, is divided by χ^p, a value smaller than one, typically equal to 0.25. Hence the fixed cost parameter is increased, which yields lower frequencies than in the single-period case.

The unconstrained optimal off-peak frequency of the lower convex envelope is

$$\tilde{f}^{o(a,b)}_{unc4}(\tilde{n}^o) = \sqrt{\frac{P_w \mu^{(a,b)} \epsilon \gamma y + \frac{P_v l}{2L}\left(\frac{\xi \beta_v (\gamma y)^2}{\tilde{n}^o} + \rho \frac{l}{k_v \tilde{n}^o}\left(\frac{(\gamma y)^2}{V_{max}} + \frac{(\gamma y)^3 \beta_v}{2L\tilde{n}^o f_{max}}\right)\right)}{c_{1t}\frac{2L}{V_{max}} + 2c_{2v}\tilde{n}^o L}}. \quad (66)$$

Equation (66) is similar to (53) applied to the demand level γy. The only difference lies in the denominator from which the term related to the fixed cost of vehicles is absent since the fixed cost of the fleet is accounted for by the peak hours in the equation (59). As a result, frequencies in off-peak hours are slightly larger than those at the same demand level in the single period case. An extension of these results to a multi-period case is simple. The formula for the peak frequency would not change, and the formulae for the off-peak periods must only update the relevant values of γ.

The unconstrained optimal stop spacing of the lower convex envelope is

$$\tilde{d}_{unc4}(\tilde{n}^p, \tilde{n}^o) = \sqrt{2vT_l \left(\frac{P_v}{P_a} l\xi + \frac{2Lf_{min}\left(c_{1t} + \tilde{n}^p c_{1v}\right)}{yP_a(\chi^p + \gamma\chi^o)} + \frac{2c_{0s}L}{yP_aT_l(\chi^p + \gamma\chi^o)} + \rho \frac{l^2 yP_v \left(\frac{\chi^p}{\tilde{n}^p} + \frac{\chi^o\gamma^2}{\tilde{n}^o} \right)}{2P_aLk_v\tilde{n}f_{max}(\chi^p + \gamma\chi^o)} \right)}. \quad (67)$$

Equation (67) mainly differs from the previous formula (54) by having two terms slightly increased because they are divided by $\chi^p + \gamma\chi^o$, a value smaller than one. However, the main structure of the formula remains stable, and confirms the approximation (34). In fact, the stop spacing resulting from (67) is close to that of (54) applied to an average demand level $y(\chi^p + \gamma\chi^o)$. Hence, the extension of this formula to the multi-period case is also straightforward. Moreover, since the stop spacing is not very sensitive to the demand level, the above formula hints at similar results with respect to the previous model.

4. Numerical analyses

This section shows how the proposed model refinements upend results of the base Model I. The parameters are such that Model I yields dominance

of road modes for all but the largest demand levels. We consistently keep this set of parameters for all models, and we show how the break-even points between road and rail modes progressively recede when model refinements are considered.

The section is structured as follows. We first presents the parameter set in Section 4.1, and some computational details in Section 4.2. Results of the four models are then discussed from Section 4.3 to Section 4.6.

4.1. Parameters

We derive the parameter set from the case study of Tirachini et al. (2010b) for a transit line in Australia. This case study considers a bidirectional line of 20 km patronized by passengers traveling on average 10 km per trip, and accessing the line by walking at a speed of 4 km/h. The critical load is equal to 35% of the total demand. This assumption models a moderate centripetal demand where, for example, 70% of the bidirectional demand uses the peak direction, and 50% of these passengers traverse the most loaded section. The demand is studied from 3000 to 60000 pax/h with a step of 500 pax/h, hence 94 demand levels are evaluated for each combination of mode and model. The set of parameters related to the passengers and to the transit line that are common to all models are listed in Table 2 where monetary figures are expressed in Australian dollars. Four modes are studied, namely buses in mixed traffic lanes (Bus), BRT, LRT, and heavy rail (HR). The main technical parameters of these four modes are listed in Table 3. These and the following parameters for the HR mode are derived from Tirachini et al. (2010b) by assuming a train length of three cars (recall that in that paper the train length is fixed, hence there is no differentiation between vehicle

Parameter	Definition	Unit	Value
$\bar{f}$	Threshold frequency for timetable behaviour	TU/h	5
l	Average trip length	km	10
L	Length of the transit line	km	20
P_a	Value of the access and egress time	\$/(pax-h)	12.5
P_v	Value of the in-vehicle time	\$/(pax-h)	10.0
P_w	Value of the waiting time	\$/(pax-h)	15.0
v	Walking speed	km/h	4
w	Waiting time at stops when $f < \bar{f}$	min	4
α	Fraction of demand in the most loaded segment of the line	-	0.35
ϵ	Rate of the average waiting time to the headway	-	0.5
μ	Discount factor of the waiting time under timetable behaviour	-	0.33
ν	Spare capacity factor	-	0.9

Table 2: Parameters related to the passengers and to the transit line that are common to all models

and TU parameters). The infrastructure and rolling stock costs of the four modes are presented in Table 4, and operational costs which include overhead are reported in Table 5. These data determine the cost parameters which vary with the model. Their derivation is described in the following. For infrastructure, land and rolling stock, the annual capital cost C is computed as

$$C = (P - R_v)\frac{i}{1-(1+i)^{-\tau}}, \tag{68}$$

where P is the purchase price, R_v is the residual value, i is the discount rate, and τ is the asset useful life in years. The used discount rate is 7%, the land cost is 9 million \$ per hectare, and a 5% residual value for rolling stock is assumed. Single period models consider an equivalent number of service hours per year equal to 2947. Two-period models use a value of 5500 hours per year. This difference yields different amortization factors, and hence different cost parameters between single and two-period models.

Table 6 lists the cost parameters for all models. The fixed hourly costs of the infrastructure, c_0 in Model I and c_{0l} in the following models, are computed by adding the amortized infrastructure and land costs from Table 4, and the hourly maintenance cost from Table 5. In order to maintain consistency among models, we conservatively consider the cost of stops as incremental with respect to that of the line, and the fixed hourly cost of a stop c_{0s} is the amortized building cost of a stop in Table 4. The crew cost of Table 5 determines the parameter c_{1t}, the hourly cost of a TU, used in models III, and IV. The amortized vehicle acquisition cost from Table 4 determines the hourly fixed cost of a vehicle, c_{1v}, in Models III, and IV. The hourly fixed cost of a TU in Model I and II, c_1, is obtained as the sum of c_{1t}, and c_{1v} multiplied by the number of fixed vehicles per TU. The cost per vehicle-km, c_{2v}, is directly derived from the operating cost of Table 5, and the cost per TU-km, c_2, is equal to c_{2v} multiplied by the number of fixed vehicles per TU. Some remarks on the limitations of the parameter set are presented in the following along the results.

Parameter	Unit	Bus	BRT	LRT	HR
f_{max}	TU/h	200	150	80	40
k_v	pax/veh	64	101	190	250
S	km/h	20	30	35	40
β_v	s/veh	4	0.33	0.33	0.33

Table 3: Main technical parameters of the four studied modes

4.2. Computational details

The algorithms were implemented in Python 2.7 with the L-BFGS-B solver, a quasi-Newton code for bound-constrained optimization, see Zhu

Parameter	Unit	Bus	BRT	LRT	HR
Infrastructure building cost	million $/km	0	10	20	35
Width required by the infrastructure	m	0	10	10	15
Stop building cost	million $/stop	0	0.125	0.25	0.5
Infrastructure technical life	year	-	50	100	100
Vehicle acquisition cost	million $/veh	0.4	0.62	3.49	2.77
Vehicle technical life	year	20	20	35	35

Table 4: Parameters for the fixed costs related to the infrastructure and to the rolling stock

Parameter	Unit	Bus	BRT	LRT	HR
Crew	$/(TU-h)	42	42	73	130
Distance related vehicle cost	$/(veh-km)	1.13	1.42	1.83	1.11
Infrastructure maintenance	$/h	0	295	1078	1851

Table 5: Operating costs, including overhead

Parameter	Unit	Model	Bus	BRT	LRT	HR
c_0, c_{0l}	$/h	I, II, III	0	9638	14871	24918
c_{0l}	$/h	IV	0	5301	8468	14211
c_{0s}	$/stop-h	II, III	0	3.1	5.9	11.9
c_{0s}	$/stop-h	IV	0	1.6	3.2	6.4
c_1	$/TU-h	I, II	54.2	60.9	159.9	336.9
c_{1t}	$/TU-h	III, IV	42	42	73	130
c_{1v}	$/veh-h	III	12.17	18.87	86.89	68.97
c_{1v}	$/veh-h	IV	6.52	10.11	46.56	36.95
c_2	$/TU-km	I, II	1.13	1.42	1.83	3.32
c_{2v}	$/veh-km	III, IV	1.13	1.42	1.83	1.11

Table 6: Cost parameters used in Models I, II, III, and IV

et al. (1997). The algorithms compute optimality, approximation, and heuristic gaps as defined in Section 3. The algorithm for Model IV does not always return an optimality gap, hence only the heuristic gap is reported. Table 7 lists the median gaps which are negligible for all modes. Because of this, in the following we refer to the results as optimal, although they are in gen-

eral near-optimal for the models solved by the approximation scheme coupled with the optimization solver. The code is available XXXXXXXXXXXXXXXXXX.

	Model II		Model III		Model IV
Mode	Median optimality gap (%)	Median approx. gap (%)	Median optimality gap (%)	Median approx. gap (%)	Median heuristic gap (%)
Bus	0.0	0.0	0.6	0.0	0.0
BRT	0.0	0.0	0.6	0.0	0.0
LRT	0.0	0.0	1.0	0.0	0.0
HR	0.4	0.1	0.9	0.0	0.0

Table 7: Median optimality, approximation, and heuristic gaps

4.3. Results of Model I

Model I assumes a fixed TU length equal to $\bar{n}$ which determines the TU capacities, and the boarding and alighting times per passenger. The mode specific parameters used for Model I are listed in Table 8.

Parameter	Unit	Bus	BRT	LRT	HR
d	km	0.4	0.8	1.0	1.2
K	pax/TU	64	101	190	750
$\bar{n}$	veh/TU	1	1	1	3
β	s/pax	4	0.33	0.33	0.11

Table 8: Mode specific parameters for Model I

The results of Model I are similar to those presented in Tirachini et al. (2010b), and therefore we present in Figure 2 only the average total cost of the four studied modes. As in Tirachini et al. (2010b), road modes dominate rail modes. LRT is never the cheapest option, and HR becomes competitive only when BRT capacity is not sufficient. In fact, Model I is sensitive to stop spacing. Figure 3 reports the average total cost when the stop spacing for

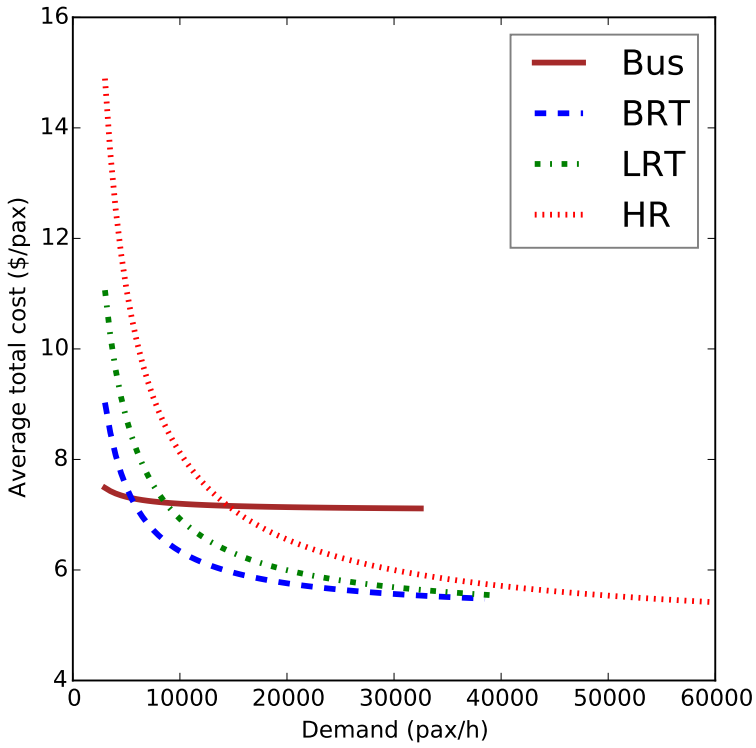

Figure 2: Model I, average total cost, C_{tot}/y

the tree rapid transit modes is set equal to 0.8 km. As can be observed, this modification yields break-even points between BRT and LRT, and between LRT and HR. Hence exogenous stop spacing is critical, and justifies the development of Model II.

4.4. Results of Model II

Model II requires new parameters, maximal speed, average acceleration and deceleration rates, and time lost for door operations, in order to compute the parameter T_l, fixed time lost for a stop. The mode specific parameters

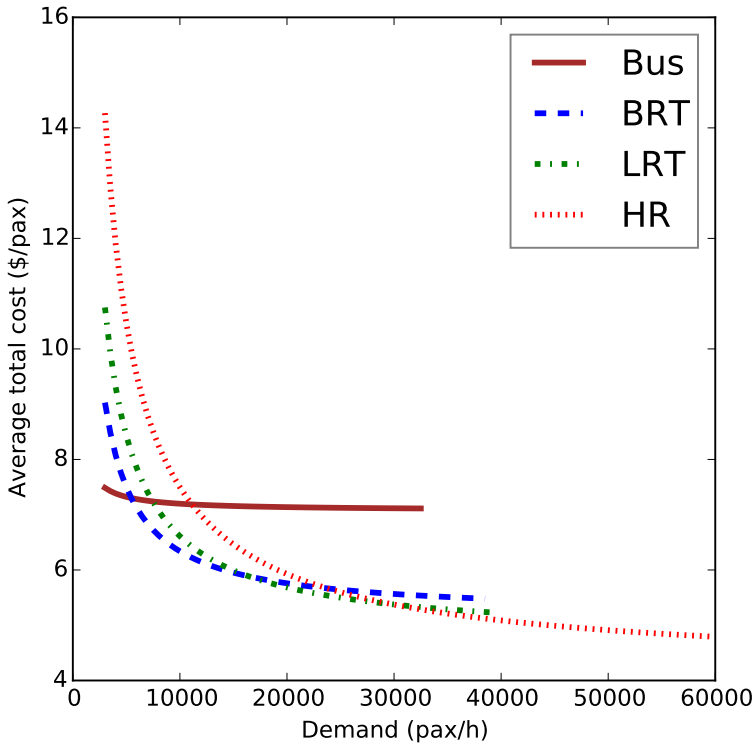

Figure 3: Model I, average total cost, C_{tot}/y, when the stop spacing is set equal to 0.8 km for the three rapid transit modes

used for Model II are listed in Table 9, and are derived from Vuchic (2005). The parameter V_{max} is set in such a way that the resulting average running speed is close to the value of S in Model I. Figure 4 illustrates the average total cost. The break-even point between BRT and LRT occurs at *circa* 18000 pax/h. The optimal stop spacing is reported in Figure 5. As predicted by the analytical results, the modes with larger T_l yield longer stop spacing. However, the ratio of stop spacings between modes follows the square root formula (34), hence the difference between modes is not as large as previously

assumed. Figure 6 reports the average running speeds which are sufficiently close to those exogenously fixed in Model I. Figure 7 illustrates the optimal frequencies with respect to the minimal frequencies. With the exception of the high capacity HR, all modes are such that the optimal frequency is set by the constraint on the minimal frequency. Recall that the optimal frequency of Model II is similar to that of Model I, and this shortcoming justifies the development of Model III.

Parameter	Unit	Bus	BRT	LRT	HR
V_{max}	km/h	22.0	36.0	44.5	55.0
$\bar{a}$	m/s^2	1.2	1.2	1.4	1.4
$\bar{b}$	m/s^2	1.4	1.4	1.2	1.1
t_d	s	2	2	2	3
T_l	h	0.0017	0.0023	0.0026	0.0033
T_l	s	6	8	10	12

Table 9: Mode specific parameters for Model II

4.5. Results of Model III

Model III requires additional parameters for the crowding penalty function. We assume, as in Tirachini et al. (2010b), a critical average occupancy rate θ_{min} equal to 0.3, and a slope ρ equal to one. The other mode specific parameters used for Model III are listed in Table 10. Figure 8 illustrates the average total cost. The break-even point between BRT and LRT occurs at *circa* 13000 pax/h, and that between LRT and HR occurs at *circa* 19000 pax/h. The crowding penalty significantly changes the optimal total cost for road modes which now exhibit a diseconomy of scale. The maximal frequency is reached earlier than in the previous model (compare Figures 7 and 9) and when this occurs serving increasing demand yields a higher total cost

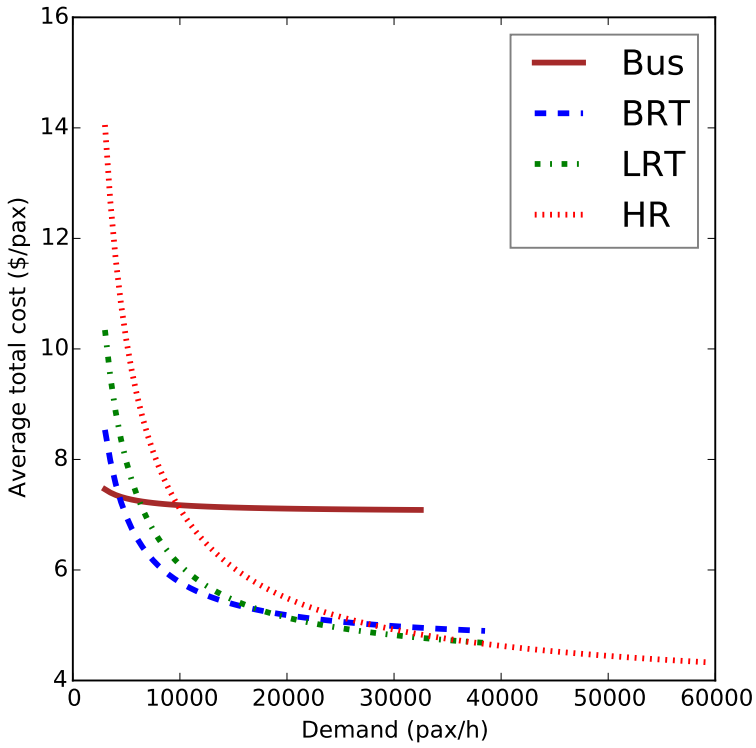

Figure 4: Model II, average total cost, C_{tot2}/y

because of passenger inconvenience. This can be appreciated by disagregating the total cost in passenger cost (Figure 10) and operator cost (Figure 11). The former strongly increases with demand when maximal frequency is reached, and the smaller decrease in average operator cost is not sufficient to compensate this effect. This fact highlights the crucial role of the parameter f_{max}.

Figure 12 compares the optimal frequency with the frequency where the crowding penalty starts, i.e. when $\theta = \theta_{min}$. In the following we refer to

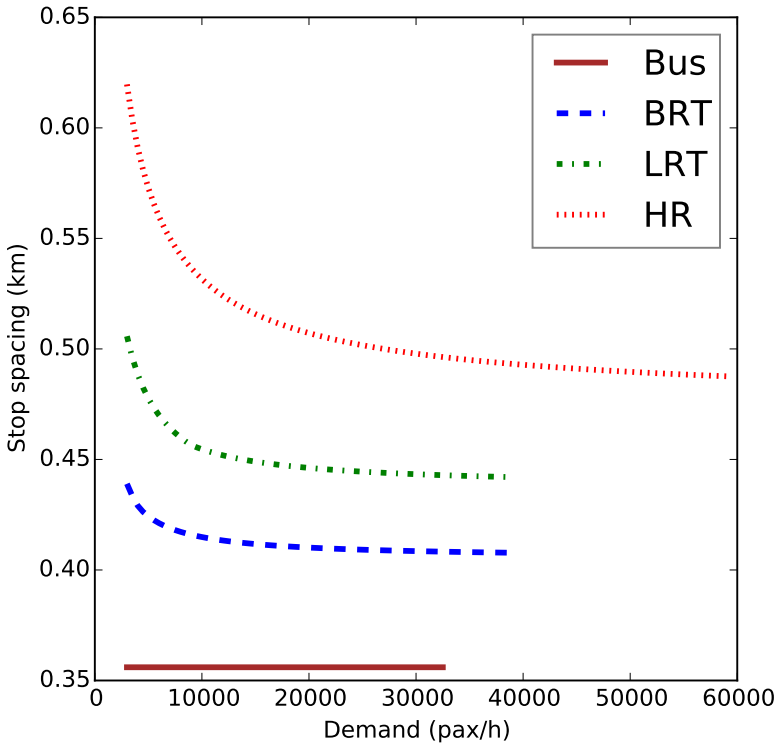

Figure 5: Model II, optimal stop spacing

this frequency as *critical*. The optimal frequency is almost always set by this critical frequency, highlighting the relevance of the crowding penalty.

The optimal stop spacing is reported in Figure 13. Here the curves show some local oddities. We observe that there are many near-optimal solutions, and hence, because of the presence of many local minima, some variations in the curve of a significant but still secondary variable such as the stop spacing can occur. As predicted by the analytical results, Model III yields slightly longer stop spacings with respect to Model II. For road modes, which have smaller capacities than rail modes, the crowding penalty increases stop

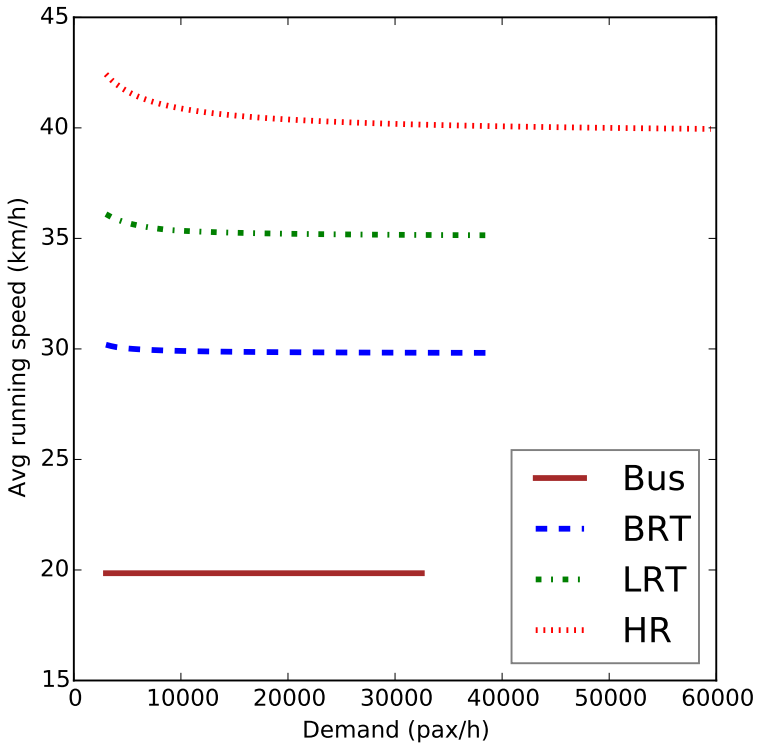

Figure 6: Model II, average running speed

spacing at higher demand levels. However, the ratio of stop spacings between modes still follows the square root formula (34) which confirms its generality. The optimal spacings for HR are lower than current practice. This may be caused by the assumption of walking as access mode. In fact, the Paris metro network, which was built to provide a dense coverage suitable for access by walking, has an average inter-station distance of 562 meters, a value close to what we find in our example.

Figure 14 shows how the optimal number of vehicles per transit unit varies for rail modes. The flexible capacity is fully exploited in these modes.

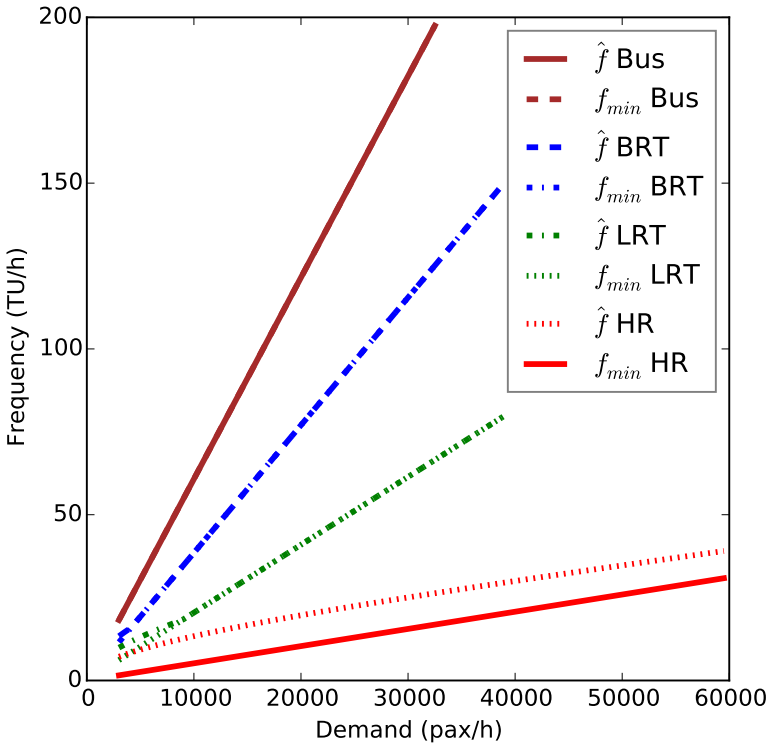

Figure 7: Model II, optimal and minimal frequencies

Hence, riding comfort and flexible capacity of rail modes prove to be related and crucial issues.

4.6. Results of Model IV

Model IV introduces peak and off-peak periods. We assume an off-peak demand level equal to γy, where γ is equal to 0.5, i.e. the peak demand is twice that of the off-peak period, as in Bruun (2005). The peak hours represent one quarter of the total service hours. All other parameters are equal to those listed previously. Figure 15 illustrates the average total cost.

Parameter	Unit	Bus	BRT	LRT	HR
n_{min}	veh	1	1	1	2
n_{max}	veh	1	1	2	5

Table 10: Mode specific parameters for Model III

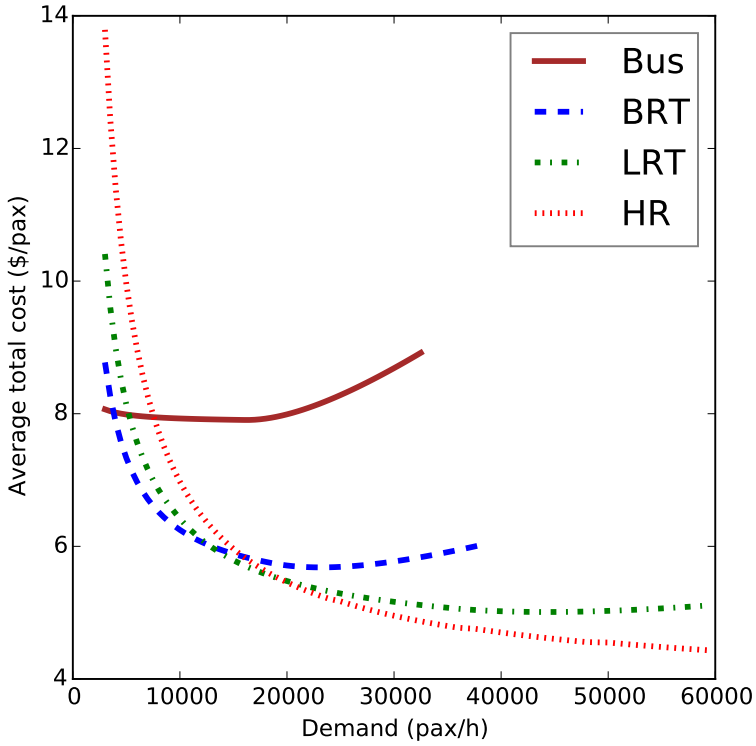

Figure 8: Model III, average total cost, C_{tot3}/y

The break-even point between BRT and LRT occurs at *circa* 13000 pax/h, and that between LRT and HR occurs at *circa* 20000 pax/h. The optimal stop spacing is not reported since the results are very similar to those of the previous model, as analytically observed in Section 3.4. Figure 16 illustrates the optimal peak frequencies with respect to the minimal frequencies. As

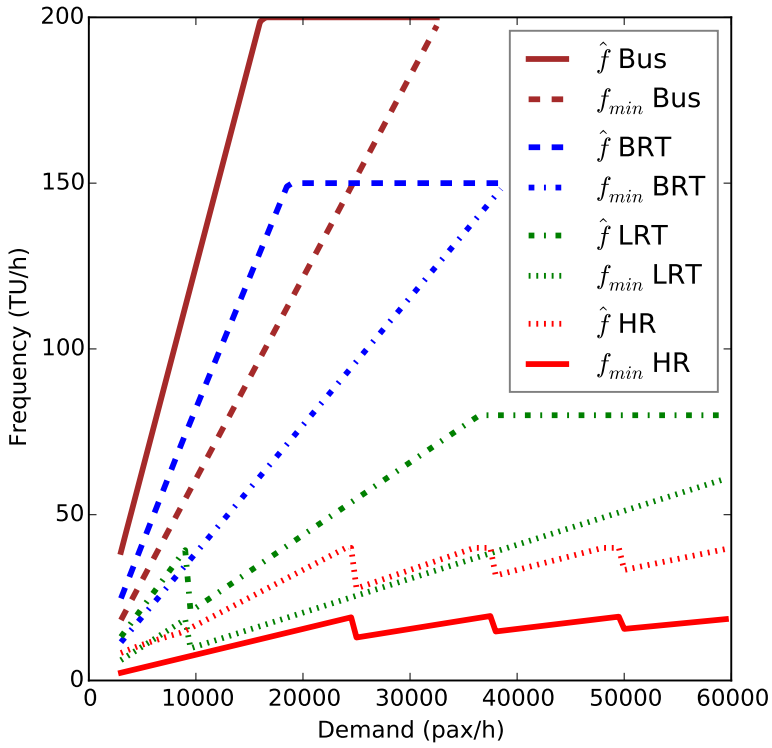

Figure 9: Model III, optimal and minimal frequencies

in Model III, the peak frequency is significantly larger than the minimal frequency. However, as can be observed in Figure 17, the optimal peak frequency does not reach the critical frequency. This happens because in Model IV the peak frequency is more expensive that the uniform frequency of Model III. The reverse is true for the off-peak frequency which benefits from the removal of fleet acquisition cost. Consequently, the optimal off-peak frequency is set by the critical frequency (see Figure 18). Figure 19 shows how the optimal number of vehicles per transit unit varies between peak and off-peak periods. These results further highlight how flexible capacity is

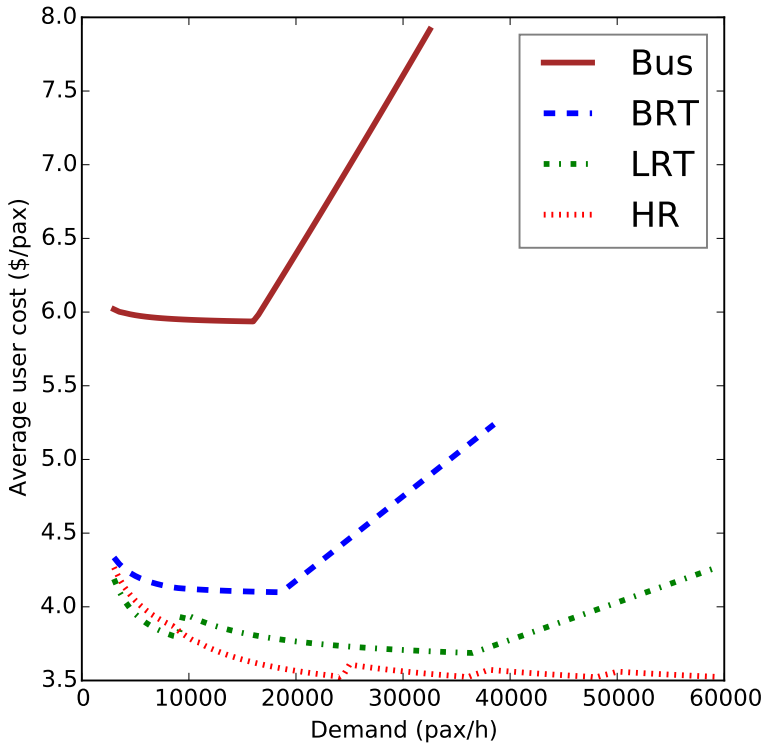

Figure 10: Model III, average passenger cost, C_{p3}/y

exploited in rail modes.

5. Scenario analyses

Because optimal solutions are well approximated by square root formulae, the proposed models are in general robust to small perturbations of the parameters. However, significantly different parameter sets could arise in real-world applications. We discuss in the following economic, technical, and environmental issues.

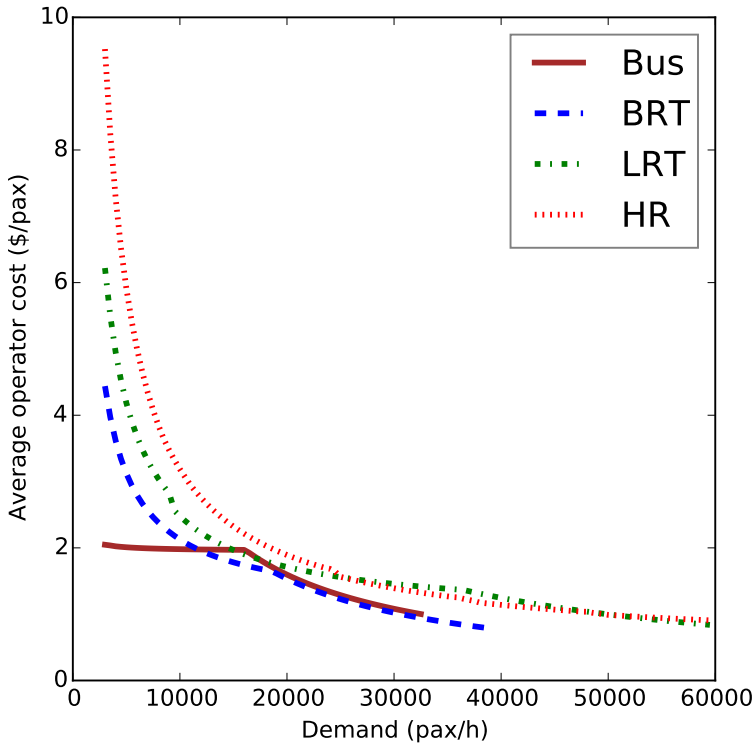

Figure 11: Model III, average operator cost, C_{o3}/y

5.1. Economic issues

A crucial parameter, the discount rate, is not directly used in the objective functions. It determines, in a non-linear way, see equation (68), the cost parameters related to the capital expenditures, infrastructure and vehicles. These cost parameters are part of the square root formulae, *and* are coefficients of the objective function. Therefore, different results must be expected by varying the discount rate. Assessing investments that differ in their capital expenditure profiles over time is a context-specific exercise. In a

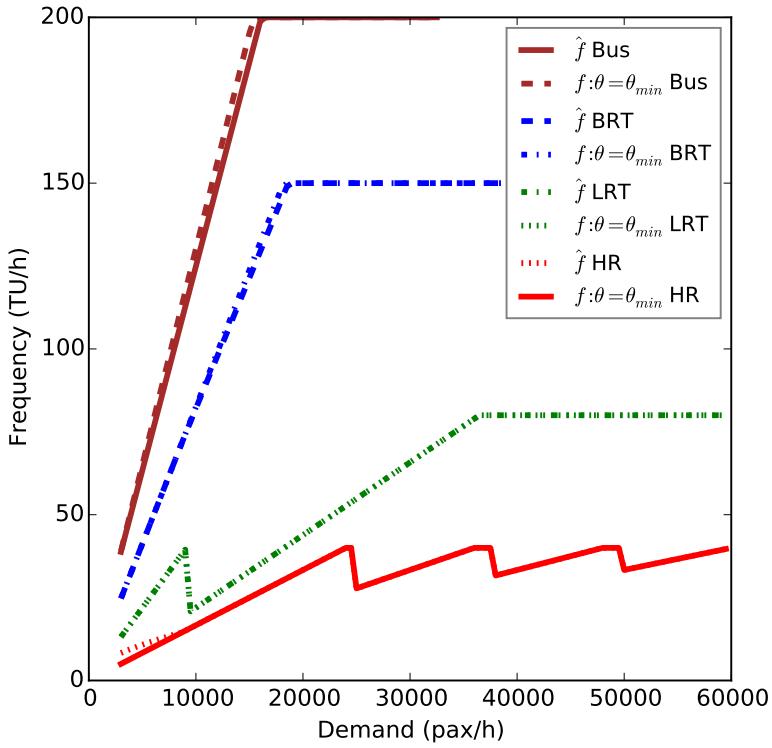

Figure 12: Model III, optimal frequency and critical frequency for crowding penalty

developing country capital is requested for many competing allocations, and therefore a high discount rate must be considered. The reverse may occur in a developed country depending on the phase of the economic cycle. Figure 20 illustrates how the break-even demand between rapid transit modes vary with the discount rate in Model IV. The parameters that do not depend on the discount rate are the same as in the base scenario of the previous section. The sensitivity to the discount rate is substantial for the comparison of BRT vs LRT and of LRT vs HR, but is larger for the former.

Labor productivity is also relevant. In the parameter set of our illustrative

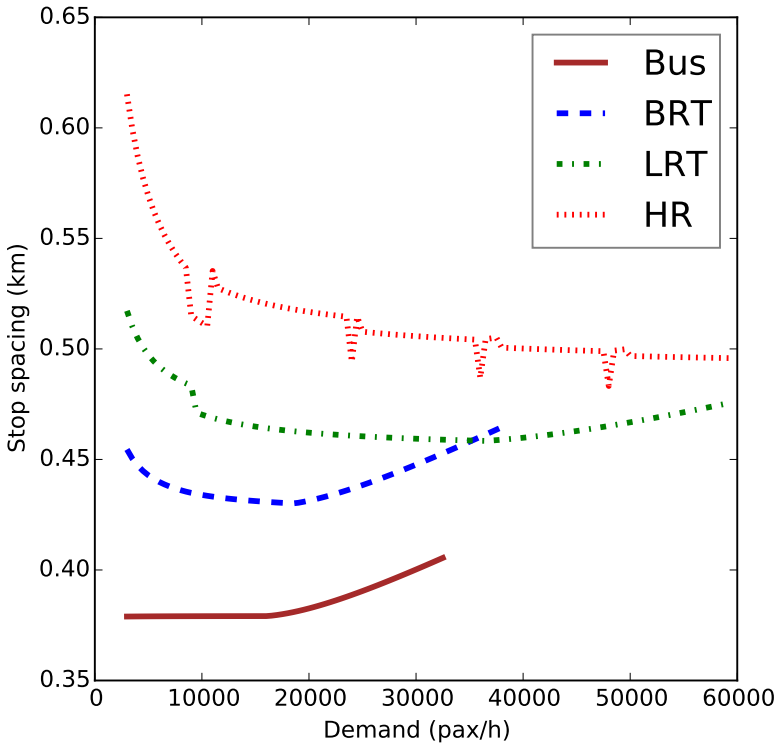

Figure 13: Model III, optimal stop spacing

example (derived in Tirachini et al. (2010b) from an Australian case study), the labor productivity of the rail mode is relatively low. Hence, the effect described in Bruun (2005) of a higher convenience of rail modes under a time-varying demand profile cannot be observed within our base parameter set. In a scenario of high labor productivity in LRT, a similar crew cost for BRT and LRT may be assumed. This could be assessed by setting the same value for the parameter c_{1t} for these modes. Under this assumption, the break-even demand between these two modes decreases to 10000 pax/h from 13000 pax/h of Model IV loaded with the base parameter set. Combining

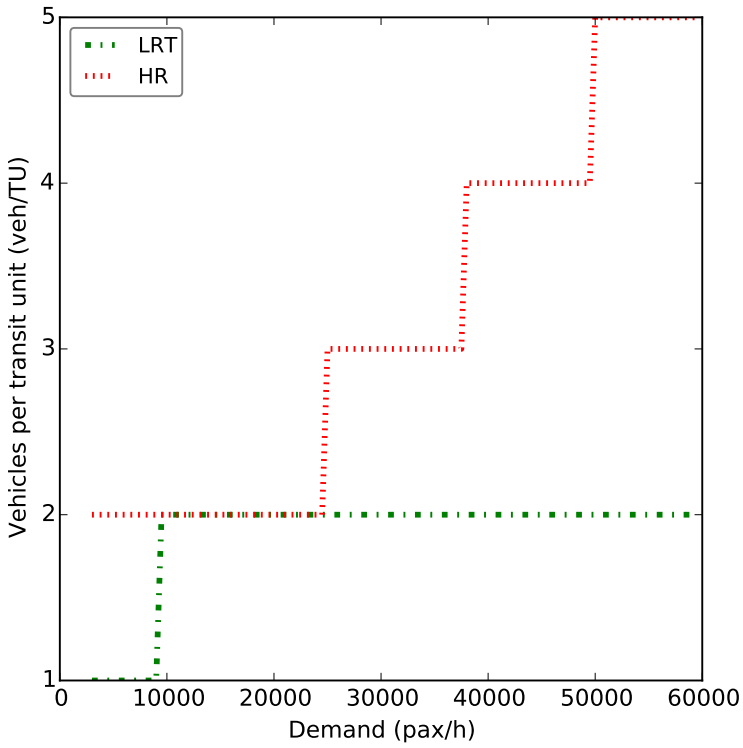

Figure 14: Model III, optimal number of vehicles per transit unit

in a scenario a low discount rate (e.g. 3%), and the previous assumption of high labor productivity, allows to reduce the break-even demand between LRT and BRT to *circa* 5000 pax/h.

We further note that income levels differ significantly between countries. Hence, both the operator cost parameters related to wages, and the user time values are context-specific. For example, by halving the user time values and the crew costs the break-even demand between LRT and BRT increases to 22000 pax/h.

In synthesis, economic issues are relevant for technology choice. Fulton

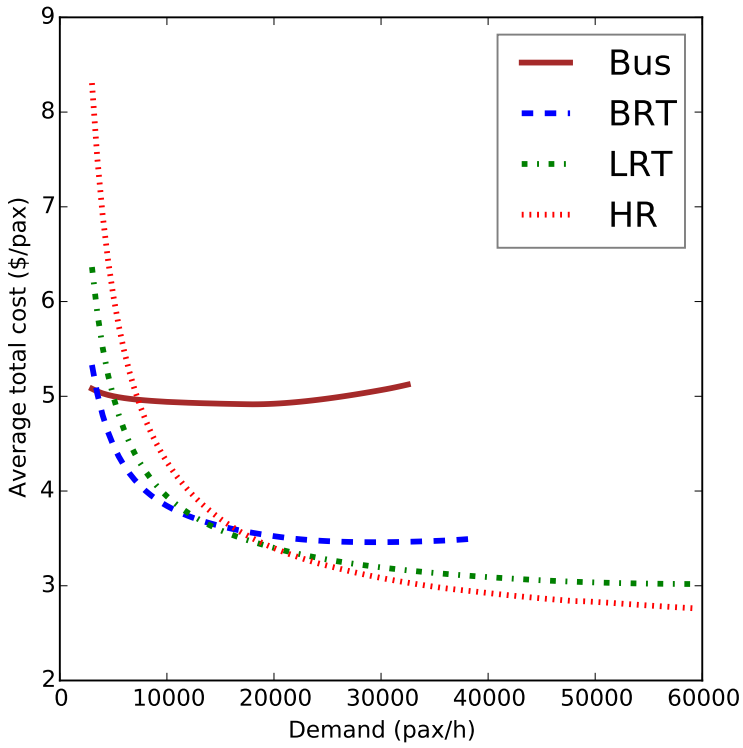

Figure 15: Model IV, average total cost, C_{tot4}/y

and Replogle (2014) discuss a 2050 scenario where the ratio of rapid transit kilometers per million urban residents is boosted in emerging economies closer to to the levels found today in developed economies, and increased in wealthy countries where it falls short of current global best practice. In this scenario, rail modes are more often preferred in OECD countries, whereas BRT is more often featured in non-OECD countries, altough all regions develop both systems to some extent. These findings are consistent with the above discussion on economic parameter variations which can be assessed by

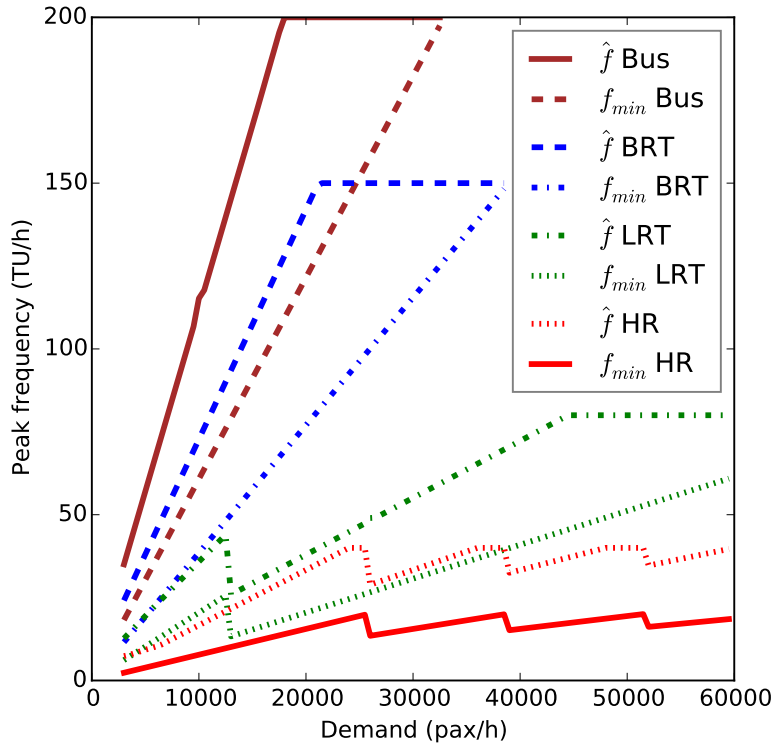

Figure 16: Model IV, optimal and minimal peak frequencies

our models.

5.2. Technical issues

The maximal speed, as well as the acceleration and deceleration rates determine the average running speed. Our experiments indicate model robustness, and the ratio of the maximal speeds is a good proxy of the ratio of the average running speed, which in turn determines relative competitiveness. A significant change in the maximal speed ratio influences results. For example, by setting the same maximal speed for both BRT and LRT, a ratio

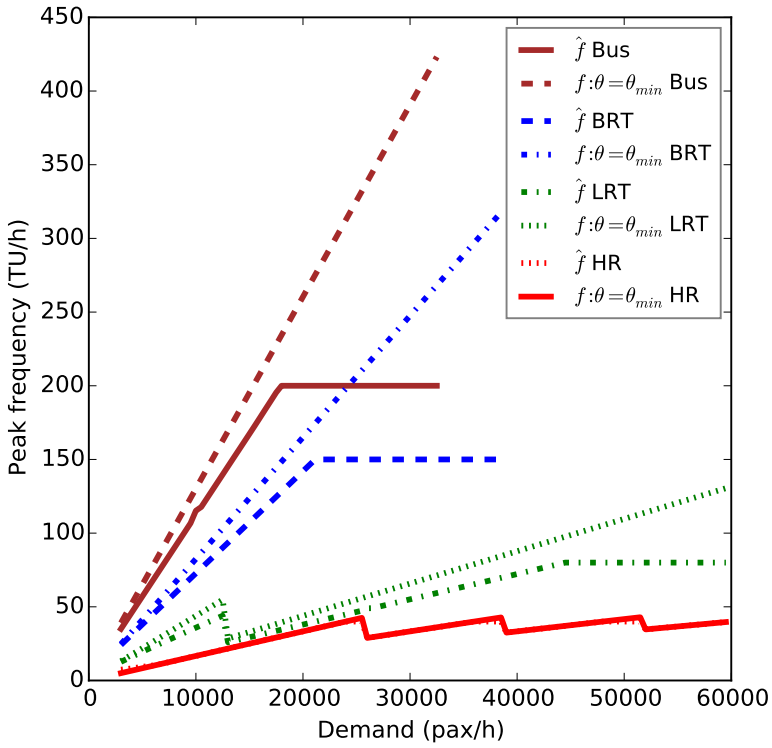

Figure 17: Model IV, optimal peak frequency and critical frequency for crowding penalty

of one, BRT is always cheaper than LRT. A more capital intensive mode requires a speed advantage.

Increasing access speed to account for access by a mix of modes faster than walking (e.g. cycling, park-and-ride) does not significantly change the break-even points. For example, by setting the average access speed to 12 km/h results in a break-even demand of 12000 pax/h between BRT and LRT, and of 18000 pax/h between LRT and HR. However, this assumption significantly changes the optimal stop distances, which for HR become larger than 800 meters, as often found in current practice.

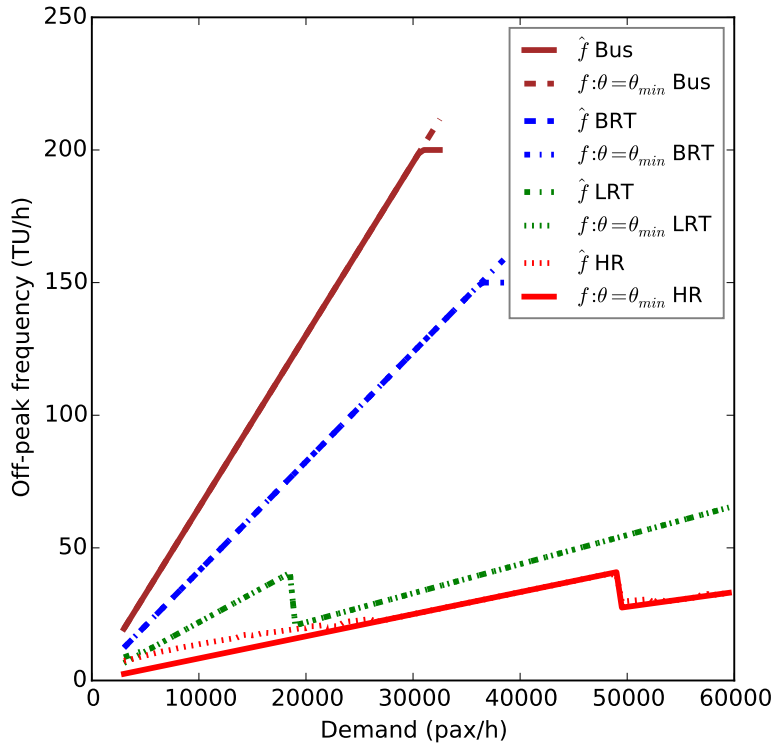

Figure 18: Model IV, optimal off-peak frequency and critical frequency for crowding penalty

The high values of f_{max} for bus, BRT, and LRT in our parameter set are technically feasible, but they could induce negative externalities in terms of urban segregation. The vehicular and pedestrian traffic flows crossing the transit line would be hampered with high transit frequencies. The interested reader is referred to Vuchic et al. (2013) for a critical appraisal of the maximal frequency in BRT and LRT. We further observe that f_{max} depends on the capacity of stops, and this parameter is context-specific. In view of this, a scenario could include lower values of f_{max}, for example $f_{max} = 60$ for both

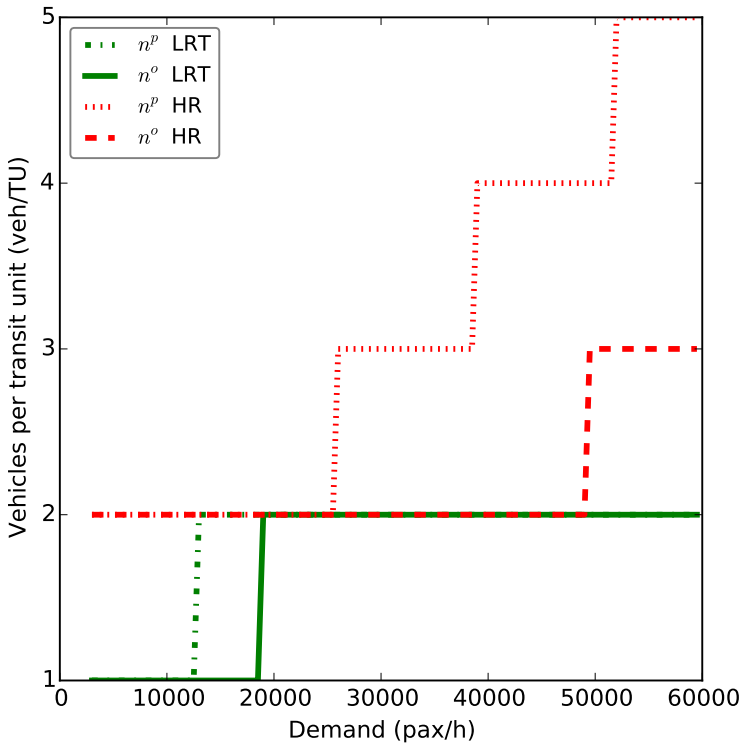

Figure 19: Model IV, optimal number of vehicles per transit unit

BRT and LRT. Coupled with a mostly centripetal demand, e.g. by setting the critical load α to 0.5, BRT is capacity constrained at *circa* 10000 pax/h.

5.3. Environmental issues

One of the main motivations for rapid transit is the reduction of the environmental impacts of private cars. However, different transit technologies may have significantly different environmental impacts. Hence, a full technological appraisal cannot disregard them. There is a growing literature attempting to quantify the life cycle environmental impacts of transit, see

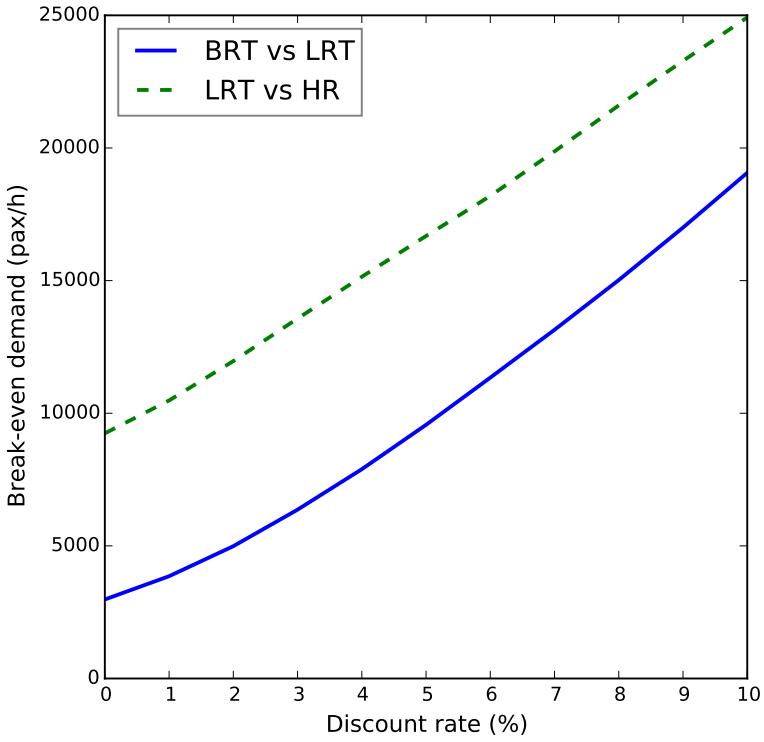

Figure 20: Model IV, Break-even demand between modes depending on the discount rate

e.g. Chester et al. (2013), and the bibliography therein. Monetized environmental impacts could also be handled by our models by expanding the scope of the fixed and variable cost components. However, the intricacies of such analyses are so high that an adequate appraisal would fall outside the scope of this paper. For example, there are vastly different environmental impacts for rail modes depending on the energy source, diesel or electricity, and in the latter case there are several sub-scenarios depending on the power source, with unabated coal and fossil-free sources being at the extremes of a wide spectrum. Development in batteries are bringing these variations to

buses too, with the added intricacies of the technology choice for batteries, and their different life cycle environmental impacts.

6. Conclusions

We have presented mathematical models amenable to approximate closed form solutions for the optimization of a transit line with fixed demand. A literature review has motivated the choice of the model of Tirachini et al. (2010b) as our base model which has been extended in several directions. Namely, we have added to the base model variable stop spacing and train length, a crowding penalty, and a multi-period generalization. Since the new models are in general unsolvable by straightforward analytical procedures, we have proposed analytically solvable approximation schemes for them. The significance of the proposed extensions has been discussed both through analytical results and an illustrative example. Faster modes require longer stop spacing, but the ratio of optimal stop spacings among different modes follows a square root formula. A crowding penalty moves away the optimal frequency from the minimal values induced by the critical capacity. Moreover, road and rail modes manage crowding in different ways. Road modes try to offer a higher frequency, whereas rail modes leverage on both frequency and train lengths. A multi-period model further increases the model realism when comparing different technologies.

We have applied the proposed models to an illustrative example where two road modes and two rail modes are defined by a set of techno-economical parameters taken from the literature. These parameters loaded in the base model yield dominance of road modes for all but the largest demand levels.

We have consistently kept this set of parameters for all models, and we have shown how the break-even points between road and rail modes progressively recede toward lower demand levels when the proposed model refinements are applied. Model refinements, not parameter changes, have brought the break-even point between BRT and LRT at *circa* 13000 pax/h, and that between LRT and HR at *circa* 20000 pax/h.

While the models are robust to small perturbations of the coefficients, we have discussed plausible scenarios and additional issues that could yield significantly different parameter sets. This is why, beside methodological transparency reasons, we have released as open source the computer code of the proposed models. Practitioners and transportation researchers could then propose their assessment of such a complex question.

Acknowledgements

Luigi Moccia was partially supported by the CNR (Italy) under the STM-2015 grant. Gilbert Laporte was funded by the Canadian Natural Sciences and Engineering Research Council under grant 2015-06189. These supports are gratefully acknowledged. We tank Manlio Gaudioso and Giovanni Giallombardo for fruitful discussions. Thanks are also due to the reviewers for their valuable comments.

Appendix A. Coefficients of the objective function of Model II

To state Model II in a more compact way, we introduce the following coefficients:

$$
\begin{aligned}
a_0^a &= P_w w y + P_v \frac{l}{V_{max}} y + c_{0l} + c_1 \beta y && \text{(A.1)}\\
a_0^b &= P_v \frac{l}{V_{max}} y + c_{0l} + c_1 \beta y && \text{(A.2)}\\
a_1 &= P_a \frac{y}{2v} && \text{(A.3)}\\
a_2 &= P_v l T_l y + 2 c_{0s} L && \text{(A.4)}\\
a_3 &= c_1 \frac{2L}{V_{max}} + 2 c_2 L && \text{(A.5)}\\
a_4^{(a,b)} &= P_w \mu^{(a,b)} \epsilon y + P_v \frac{l\beta}{2L} y^2 && \text{(A.6)}\\
a_5 &= 2 c_1 L T_l. && \text{(A.7)}\\
& && \text{(A.8)}
\end{aligned}
$$

Appendix B. Properties of the objective function of Model II

In the following we report the gradient, the Hessian, and the determinant of the Hessian of the two components a and b of (21):

$$
\begin{aligned}
\nabla C_{tot2}^{(a,b)} &= \left(a_3 - \frac{a_4^{(a,b)}}{f^2} + \frac{a_5}{d}, a_1 - \frac{a_2}{d^2} - \frac{a_5 f}{d^2} \right), && \text{(B.1)}\\
H^{(a,b)} &= \begin{pmatrix} \frac{2\, a_4^{(a,b)}}{f^3}, & -\frac{a_5}{d^2} \\ -\frac{a_5}{d^2}, & \frac{2\, a_5 f}{d^3} + \frac{2\, a_2}{d^3} \end{pmatrix}, && \text{(B.2)}\\
\det(H^{(a,b)}) &= \frac{4\, a_4^{(a,b)} \left(\frac{a_5 f}{d^3} + \frac{a_2}{d^3} \right)}{f^3} - \frac{a_5^2}{d^4}. && \text{(B.3)}
\end{aligned}
$$

Convexity would require the satisfaction of the following condition, which is not guaranteed in the common ranges of variables and parameters:

$$\det(H^{(a,b)}) > 0 \text{ if and only if } 4a_4^{(a,b)} d\left(1 + \frac{a_2}{a_5 f}\right) > a_5 f^2. \tag{B.4}$$

Appendix C. Coefficients of the objective function of Model III

To state Model III in a more compact way, we introduce the following coefficients:

$$\begin{aligned}
a_{30}^{a} &= P_w w y + P_v \frac{l\xi}{V_{max}} y + c_{0l} + c_{1v}\beta_v y && \text{(C.1)}\\
a_{30}^{b} &= P_v \frac{l\xi}{V_{max}} y + c_{0l} + c_{1v}\beta_v y && \text{(C.2)}\\
a_{31} &= P_a \frac{y}{2v} && \text{(C.3)}\\
a_{32} &= P_v l T_l \xi y + 2c_{0s} L && \text{(C.4)}\\
a_{33} &= c_{1t} \frac{2L}{V_{max}} && \text{(C.5)}\\
a_{34}^{(a,b)} &= P_w \mu^{(a,b)} \epsilon y && \text{(C.6)}\\
a_{35} &= 2c_{1t} L T_l && \text{(C.7)}\\
a_{36} &= c_{1t}\beta_v y && \text{(C.8)}\\
a_{37} &= c_{1v} \frac{2L}{V_{max}} + 2c_{2v} L && \text{(C.9)}\\
a_{38} &= P_v l y^2 \left(\frac{\xi\beta_v}{2L} + \frac{\rho l}{2Lk_v V_{max}}\right) && \text{(C.10)}\\
a_{39} &= 2c_{1v} L T_l && \text{(C.11)}\\
a_{310} &= P_v \rho \frac{l^2 y^2 T_l}{2Lk_v} && \text{(C.12)}\\
a_{311} &= P_v \rho \frac{l^2 y^3 \beta_v}{4L^2 k_v}. && \text{(C.13)}
\end{aligned}$$

References

Badia, H., Estrada, M., and Robusté, F. (2014). Competitive transit network design in cities with radial street patterns. *Transportation Research Part B: Methodological*, 59(0):161–181.

Bruun, E. (2005). Bus rapid transit and light rail: Comparing operating costs with a parametric cost model. *Transportation Research Record: Journal of the Transportation Research Board*, 1927(1):11–21.

Byrne, B. F. (1975). Public transportation line positions and headways for minimum user and system cost in a radial case. *Transportation Research*, 9(2):97–102

Chester, M. V., Pincetl, S., Elizabeth, Z., Eisenstein, W., and Matute, J. (2013). Infrastructure and automobile shifts: positioning transit to reduce life-cycle environmental impacts for urban sustainability goals. *Environmental Research Letters*, 8(1):015041+10.

Chien, S. and Schonfeld, P. (1998). Joint optimization of a rail transit line and its feeder bus system. *Journal of Advanced Transportation*, 32(3):253–284.

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

De Palma, A., Kilani, M., and Proost, S. (2015). Discomfort in mass transit and its implication for scheduling and pricing. *Transportation Research Part B: Methodological*, 71(1):1–18.

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., and Daganzo, C. F. (2011). Design and implementation of efficient transit networks: Procedure, case study and validity test. *Transportation Research Part A: Policy and Practice*, 45(9):935–950.

Fulton, L. M. and Replogle, M. A. (2014). A global high shift scenario: Impacts and potential for more public transport, walking, and cycling with lower car use. Technical report, University of California, Davis, US.

Griswold, J. B., Cheng, H., Madanat, S., and Horvath, A. (2014). Unintended greenhouse gas consequences of lowering level of service in urban transit systems. *Environmental Research Letters*, 9(12):124001+11.

Griswold, J. B., Madanat, S., and Horvath, A. (2013). Tradeoffs between costs and greenhouse gas emissions in the design of urban transit systems. *Environmental Research Letters*, 8(4):044046+14.

Jansson, J. O. (1980). A simple bus line model for optimisation of service frequency and bus size. *Journal of Transport Economics and Policy*, 14(1):53–80.

Jara-Díaz, S. and Gschwender, A. (2003). Towards a general microeconomic model for the operation of public transport. *Transport Reviews*, 23(4):453–469.

Li, Z. and Hensher, D. A. (2013). Crowding in public transport: a review of objective and subjective measures. *Journal of Public Transportation*, 16(2):107–134.

Mohring, H. (1972). Optimization and scale economies in urban bus transportation. *The American Economic Review*, 62(4):591–604.

Newell, G. F. (1979). Some issues relating to the optimal design of bus routes. *Transportation Science*, 13(1):20–35.

Qin, F. (2014). Investigating the in-vehicle crowding cost functions for public transit modes. *Mathematical Problems in Engineering*, 2014:1–13.

Qin, F. and Jia, H. (2012). Remodeling in-vehicle crowding cost functions for public transit. In *Transportation Research Board 91st Annual Meeting*, number 12-2393.

Sivakumaran, K., Li, Y., Cassidy, M., and Madanat, S. (2014). Access and the choice of transit technology. *Transportation Research Part A: Policy and Practice*, 59(0):204–221.

Tirachini, A. (2014). The economics and engineering of bus stops: Spacing, design and congestion. *Transportation Research Part A: Policy and Practice*, 59(0):37–57.

Tirachini, A., Hensher, D. A., and Jara-Díaz, S. R. (2010a). 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., and Jara-Díaz, S. R. (2010b). Restating modal investment priority with an improved model for public transport analysis. *Transportation Research Part E: Logistics and Transportation Review*, 46(6):1148–1168.

Tirachini, A., Hensher, D. A., and 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(0):36–52.

Vaughan, R. (1986). Optimum polar networks for an urban bus system with a many-to-many travel demand. *Transportation Research Part B: Methodological*, 20(3):215–224.

Vuchic, V. R. (2005). *Urban Transit: Operations, Planning, and Economics*. Wiley, Hoboken, New Jersey.

Vuchic, V. R. and Newell, G. F. (1968). Rapid transit interstation spacings for minimum travel time. *Transportation Science*, 2(4):303–339.

Vuchic, V. R., Stanger, R. M., and Bruun, E. C. (2013). Bus rapid bus rapid transit (BRT) versus light rail transit light rail transit (LRT): Service quality, economic, environmental and planning aspects. In *Transportation Technologies for Sustainability*, pages 256–291. Springer, Berlin.

Wardman, M. and Whelan, G. (2010). Twenty years of rail crowding valuation studies: Evidence and lessons from British experience. *Transport Reviews*, 31(3):379–398.

Zhu, C., Byrd, R. H., Lu, P., and Nocedal, J. (1997). Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–560.