Public Transport (2018) 10:455–497
https://doi.org/10.1007/s12469-018-0187-1

ORIGINAL PAPER

A technology selection and design model of a semi-rapid transit line

Luigi Moccia[1,2] · **Duncan W. Allen**[3] · **Eric C. Bruun**[4]

Accepted: 26 August 2018 / Published online: 17 September 2018

Abstract

We present a new optimization model for technology selection and design of a semi-rapid transit line. With respect to previous studies, we improve the synthetic representation of the temporal and spatial variability of demand, and of several operational and design aspects. We apply the model to two scenarios offering comparable performance by commercially available technologies in terms of service, rather than assuming that service quality is strongly associated with technology. The model is validated by comparing some computed performance indices with best practices. We show that planning for a faster technology can be more important than the choice between bus and rail per se, except at very low demand density, and that differences of total cost, sum of passengers' time value and operator's cost, between the technologies are smaller than commonly held across a wide range of higher demands. At high demand density multiple-unit rail offers the most cost-effective way to achieve high capacities under many conditions. A scenario variation analysis shows the relevance of differences between value of time components, the bias of averaging vehicle load ratios when assessing the crowding disutility, the usefulness of a demand index abstracting from some specific parameter choices, and the high impact of the project discount rate.

Keywords Semi-rapid transit · Bus rapid transit · Light rail transit · Transit line optimization

✉ Luigi Moccia
moccia@icar.cnr.it

[1] Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni, via P. Bucci, Rende, Italy

[2] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT), Montreal, Canada

[3] IBI Group, Boston, USA

[4] Kyyti Group Ltd., Helsinki, Finland

1 Introduction

The technology selection and design model of Moccia and Laporte (2016) minimizes the sum of passengers and operator costs of a transit line. The passengers' cost is the sum of access, waiting, riding, and egress time values. The operator cost is the sum of fixed and variable costs of the transit line. We expand this model in five directions with the goal of improving the representation of semi-rapid transit (i.e. with higher quality of service attributes than local service). First, we improve the multi-period representation of the demand profile. Second, we propose a crowding penalty function that reduces the underestimation caused by Jensen's inequality when using the average vehicle occupancy rate. Third, we consider frequency-dependent travel time delays and platooning. Fourth, we define the maximum frequency as dependent on the longest dwell time. Fifth, we present a refined representation of the operator cost by differentiating between operating and commercial cycle times, and by modeling station construction and maintenance costs as variable with peak-period train length.

On the practical side, we provide a deep analysis of techno-economical parameters of two semi-rapid modes, namely bus rapid transit (BRT), and light rail transit (LRT). We use "mode" here to represent a relatively large set of specific implementations that we indicate as "technologies". We denote as "semi-rapid" the modes that require a right-of-way (RoW) with partial separation from other traffic. According to the classification scheme of Vuchic (2005), semi-rapid transit occurs in RoWs of category B, whereas properly defined rapid transit requires fully separated RoWs of category A, and local services take place in mixed traffic, RoW of category C (Vuchic et al. 2012).

We examine two scenarios offering comparable performance for each technology in terms of service, rather than assuming that service quality is strongly associated with technology. Two scenarios representing different levels of investment and performance are assessed by the proposed model. Scenario 1 represents operation in an urban roadway in RoW obtained by a simple change in designation of an existing lane, with relatively simple curb stops. Scenario 2 operates in an exclusive at grade RoW where a new laterally separated alignment is developed with more substantial stations and preferential signal priority. These two scenarios differ in performance levels, and, as a result, in productive capacity (Vuchic 2007). They have different operating and maintenance (O&M) and investment costs reflecting their design standards and technology assumptions but have identical values of time (VoTs) for the users.

The remainder of this paper is structured as follows. Section 2 reviews the relevant literature and outlines our approach. The model is presented in Sect. 3. Section 4 discusses the two scenarios and Sect. 5 describes the numerical results. Finally, we present some conclusions in Sect. 6.

2 Literature review

Moccia et al. (2017) establish an equivalence in the objective functions of fixed and elastic demand models for the optimization of a transit line when the elastic demand can be approximated by a linear function of travel time components and

fare. This finding reinforces the general result of Daganzo (2012) about the equivalence in the optimality conditions of fixed and elastic demand models for the optimization of public services. These results allow a decoupling between system design and demand forecast. The system modeling effort can therefore be directed toward a more refined representation of the operational details while the demand can be assumed as given. The fixed-demand transit line models of Moccia and Laporte (2016) integrate variable stop spacing and train length, a crowding penalty, and a multi-period generalization in the base model of Tirachini et al. (2010). The research question that we are dealing with is what functional extensions can increase the model's realism. In Sects. 2.1, 2.2 and 2.3 we discuss some literature on service variability, in-vehicle passenger crowding, and stop spacing, respectively. Our approach with respect to this literature is outlined in Sect. 2.4.

2.1 Travel time, speed, platooning and service variability

Fernandez and Planzer (2002) discuss the relevance of stops on transit capacity analysis given that stops can be more of a constraint on capacity than intersections or links. The authors show that simulation models of stop operations provide a better, and generally larger, estimation of stop capacity than analytical formulae. Tirachini and Hensher (2011) propose a total cost minimization model of a segregated bus line to choose the fare collection system and the infrastructure investment level that determines the bus running speed. To this latter aim, a linear relationship between infrastructure investment and running speed is assumed. High bus frequency induces delays at stops and at intersections. The delay at signaled intersections is estimated by the Akçelik and Rouphail (1993) formula and stops are assumed to be isolated from traffic lights. In the frequency range limited by stop capacity the overflow delay at signaled intersections is negligible. Queuing delay and dwell time at stops are modeled by regressions of simulations where frequency, types of fare payment, and boarding demand are key variables. Simulation results indicate rapidly increasing queuing delays when a threshold boarding demand and bus frequency are reached. These thresholds are higher with more efficient types of fare payment. The authors find that off-board fare payment is the most convenient option under many circumstances and that the optimal running speed grows with the logarithm of the demand.

Quality of service may be affected by bus bunching or platooning, the underlying process for which is well understood. Whenever a bus loses time or schedule it is likely to experience more passenger boardings than the trailing bus. This increases its dwell time at stops, and thus the process is self-reinforcing (Daganzo 2009). Negative consequences of platooning for the passengers are an increase of the average and the variance of the waiting times, crowding on the platforms and inside the vehicles, and discomfort when passengers are not able to board the first incoming bus and they have to relocate along the platform. Real-time information systems such as automatic vehicle location have allowed ambitious strategies to counteract platooning such as, for example, the self-equalizing headways of Bartholdi and Eisenstein (2012). In spite of these advancements, Verbich et al. (2016) report that in a state-of-the-art bus corridor in Portland, Oregon, platooning is still a frequent occurrence.

Service reliability, through its effects on waiting times, can have a significant effect on how users perceive the quality of service. Van Oort (2016) presents a framework to monetize service reliability, and estimates that 2/3 of the total benefits for a tram project in Utrecht, the Netherlands, can be attributed to improvements in service reliability.

2.2 Passenger in-vehicle crowding

Li and Hensher (2013) compare crowding levels as measured by some rail operators in Australia with results of passenger surveys. A significant gap exists between measured and expressed crowding levels. The authors observe that this gap may be caused by excessive aggregation of crowding measures with respect to the time and to the number of stations. Tirachini et al. (2016) estimate sitting and standing crowding multipliers by inferring passenger preferences from the smart card transaction database of the Singapore metro. Some passengers are willing to take a train in the opposite direction of their destination to secure a seat. Analyzing this behavior the authors find that the standing multiplier can be as much as 1.55 with a density of three standees per square meter. Klumpenhouwer and Wirasinghe (2016) define a single-period model of a LRT line where the frequency, the stop spacing, and the demand are given. The demand is centripetal, the passenger cost is related to the in-vehicle crowding disutility, and the operator cost is a function of the fleet size and of the platform length. The optimized variable is the train length. The model shows that crowding can significantly affect the optimal train length. Numerical experiments based on data from Calgary's C-Train yield train lengths similar to those in current and planned operation in the LRT network. Hörcher et al. (2017) analyze large scale automated demand and train location data from Hong Kong to estimate the user cost of crowding in a revealed preference route choice framework. A linear crowding multiplier model is validated, and this study found that at six passengers per square meter and no chance to find a seat, the value of time circa doubles. Haywood et al. (2017) report results of a survey on stated satisfaction of subway passengers in Paris. Crowding disutility grows linearly with in-vehicle density and it is sensitive to the passenger income. Three causes of crowding disutility are identified: standing probability, poorer use of time, and closeness to other passengers.

2.3 Stop spacing

Stop spacing in passenger transportation studies performed between 1915 and 1930 are reviewed in Vuchic and Newell (1968). These early works were the first to highlight the fact that an optimal stop spacing should be a function of the number of passengers on the vehicle and of those wishing to board it. Vuchic and Newell (1968) generalize their analysis to variable population densities, different access speeds, train kinematics, and competition for a single line where passengers commute to a terminus serving a central business district and the minimization of aggregate passenger travel time is sought. The proposed analytic model can be solved in closed form for the special case of a uniform population density. The optimal stop spacing

then follows an arithmetic progression and increases in the direction of cumulation. Optimal stop spacing is found to be highly sensitive and increasing with the ratio between passenger access speed and maximum train speed. Vuchic (1969) considers the maximization of patronage for a rapid transit line in an area with uniform population density, as well as other assumptions as in Vuchic and Newell (1968). This study allows an analytical comparison of the two objectives, namely maximum patronage vs minimum total travel time. In both cases, stop spacings increase in the direction of cumulation, but maximum patronage requires a considerably higher stop density. Kikuchi and Vuchic (1982) study the optimal number of stops and vehicle stopping policy of a transit line where boarding and alighting densities are constant along the line, an assumption that makes this study relevant to transit in central areas, as opposed to the cumulative boarding assumption more relevant to commuting lines, as in the studies previously reviewed. The authors compare two objective functions: minimization of passenger travel time, and minimization of total cost, defined as the sum of passenger and operator costs. For the large passenger volume case, the two objective functions converge to a similar number of optimal number of stops. In the context of transit technology selection, Moccia and Laporte (2016) analytically find that a faster technology requires longer stop spacing, but the ratio of the optimal stop spacings between two technologies follows a square root formula of their speed ratio, i.e. a sublinear relationship.

2.4 Proposed approach

Regarding travel time variability we chose to model a transit system where only a moderate level of this variability occurs at less than the maximum allowed frequency. Semi-rapid bus and rail systems can carry higher demand levels by using higher frequencies than those that we allow, but the resulting travel time variability is not likely to be acceptable for a new design. Regarding travel time speed we propose a scenario analysis because the alignment choice is often severely constrained and few options are available. Regarding platooning, we observe that this issue is tangential to our model's scope. However, since we consider frequencies that may generate some occurrences of these conditions, we conservatively model their impacts on the passengers by a reduction of the waiting time advantage of those frequencies, and on both the operator and the passengers by a travel time component as an increasing function of the frequency.

Regarding in-vehicle passenger crowding, both the recent survey of Haywood et al. (2017) and the revealed preference studies of Tirachini et al. (2016) and Hörcher et al. (2017) reaffirm the finding of the meta-analysis of Wardman and Whelan (2010) that the crowding multiplier of the base in-vehicle time with respect to the load factor is well approximated by a piecewise linear function which is equal to one up to the load factor corresponding to the seated capacity, and then starts to increase. The optimization model of Klumpenhouwer and Wirasinghe (2016) highlights the relevance of in-vehicle passenger crowding for station design and fleet sizing. Our contribution to this literature thread is that we overcome the inherent underestimation of the crowding disutility when applying directly a piecewise linear

function to the average vehicle load along a bidirectional line (we note that this issue does not pertain to the model of Klumpenhouwer and Wirasinghe (2016)). This is accomplished by a new crowding disutility function presented in the next section and derived in "Appendix 2".

Regarding stop spacing, in the computational experiments we discern between the effects of demand and trip length variations.

3 Model

This section is structured as follows. We first present the principal assumptions and the modeling of the cycle time in Sect. 3.1. The passengers' time value and the operator cost, which are treated separately and then combined into a total cost, are described in Sects. 3.2 and 3.3, respectively. The side constraints are described in Sect. 3.4, and the resulting optimization model is presented in Sect. 3.5.

Table 1 summarizes the main symbols. Dimensionless parameters are indicated by Greek letters. Additional symbols are derived as explained below. The subscripts *min* and *max* specify bounds of a parameter or of a variable. The subscript p refers to a specific period or portion of an operating day.

3.1 Main assumptions and cycle time

We assume that a bidirectional transit line operates on a route, where by "route" we indicate the physical alignment and the infrastructure, and by "line" we refer to the service operated. The route length is L and the line length is $2L$. The transit line serves passengers for $\hat{p}$ periods. The average bidirectional hourly demand (boardings) q_p in a period is equal to $q\gamma_p$, where p is the period index, γ_p is a positive parameter not larger than one, and q is the average hourly demand at the peak period. Without loss of generality we index the periods in decreasing order of demand, the peak period is indicated by $p = 1$ and thus $\gamma_1 = 1$. For each period the maximum demand is $q_p\tau_p$, where τ_p is a parameter larger than one. The ratio of a period's hours to the total service hours is denoted by χ_p. These demand parameters must reflect the total service hours in a year and they are usually derived by a procedure outlined in "Appendix 1". The average stop spacing is an optimized variable and is indicated by d. The frequency, f, is expressed as the number of transit units (TUs) per hour. The concept of transit unit, see Vuchic (2005), refers to a set of n physically linked vehicles traveling together. This number is usually called the "consist". For buses, n is equal to one, whereas for rail technology n can be an integer larger than one. TU is therefore a common concept for both single vehicles and trains used on a transit line. We posit a frequency for each period, namely f_p. The TU length for each period is indicated by n_p.

We distinguish between operating and commercial cycle times. The *operating cycle time* is the sum of the running time between stations, including acceleration and deceleration, of the time lost at intersections, and of the dwell time for boarding and alighting. The *commercial cycle time* is determined as a function of the

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	m/s^2
$\bar{b}$	Average deceleration rate of a TU	m/s^2
B	Deployed fleet of TUs	TU
c_{0l}	Fixed operator cost related to the transit line	\$/h
c_{0s}	Fixed operator cost related to a stop	\$/h
c_{0sv}	Operator cost related to a stop per extra vehicle	\$/veh-h
c_{1t}	Unit operator cost per TU-hour	\$/TU-h
c_{1v}	Unit operator cost per vehicle-hour	\$/veh-h
c_{2v}	Unit operator cost per veh-km	\$/veh-km
C_a	Access and egress time value	\$/h
C_o	Operator cost	\$/h
C_u	Passengers' time value	\$/h
C_{tot}	Total cost, sum of C_u and C_o	\$/h
C_v	In-vehicle time value	\$/h
C_w	Waiting time value	\$/h
d	Average distance between stops	km
f	Frequency	TU/h
f_l	Threshold frequency for timetable behavior	TU/h
f_m	Threshold frequency for platooning	TU/h
$\dot{f}$	Threshold frequency for the high frequency penalty	TU/h
$\check{f}_{max}$	Cap on the maximum frequency	TU/h
H	Service hours per year	h/year
k	Capacity of a vehicle	pax/veh
K	Capacity of a TU	pax/TU
l	Average trip length	km
L	Length of the route	km
n	Number of vehicles per TU	veh
p	Index of a period	–
$\hat{p}$	Number of periods	–
q	Average bidirectional demand at the peak period	pax/h
R	Running time	h
s	Access and egress speed	km/h
S	Commercial speed of the TU	km/h
S_{max}	Maximum allowed speed of the TU	km/h
S_{run}	Running speed of the TU excluding stop service	km/h
t_a	Average access and egress time of a user	h
T_a	Time loss caused by acceleration and deceleration phases	h
t_b	Boarding and alighting time per user and vehicle	s/pax-veh
T_b	Boarding and alighting time per user and TU	s/pax-TU
t_c	Commercial cycle time	h
t_{oc}	Operating cycle time	h
t_d	Time loss caused by opening and closing of doors	s

Table 1 (continued)

Symbol	Definition	Unit
t_e	Average dwell time at stops	s
t_{e0}	Fixed stop clearance time	s
t_{ev}	Stop clearance time for an extra vehicle length	s/veh
T_l	Time loss caused by acceleration, deceleration, and door operations	h
t_{tf}	Fixed component of the terminal time	s
t_{tv}	Component of the terminal time variable with the consist	s
TU	Transit unit	–
t_u	Average time lost at intersections per unit distance	min/km
t_v	Average in-vehicle time of a user	h
t_w	Average waiting time of a user	h
V_a	Unit value of the access and egress time	\$/pax-h
V_v	Unit value of the in-vehicle time	\$/pax-h
V_w	Unit value of the waiting time	\$/pax-h
w	Waiting time at a stop when $f < f_l$	min
y	One-stage technical life	year
α	Fraction of the hourly demand in the most loaded segment of the line	–
β	Multiplicative factor of the operating cycle time	–
γ	Ratio of the period demand to the peak demand	–
δ	Crowding penalty function w.r.t. the instantaneous occupancy rate	–
Δ	Crowding penalty function w.r.t. the average occupancy rate	–
ϵ	Rate of the average waiting time to the headway	–
ζ	Spare capacity factor for the fleet	–
η	Fraction of the longest dwell time in the maximum frequency formula	–
θ	Vehicle occupancy rate	–
ι	Discount rate	–
λ	Ratio of the average trip length to the length of the route	–
μ	Discount factor of the waiting time under timetable behavior	–
ν	Spare capacity factor for the TU	–
ξ	Ratio of the residual value to the initial capital value	–
ρ	Slope of the linear part of the crowding penalty function δ	–
τ	Ratio of the maximum to the average period demand	–
ϕ	Ratio of the maximum to the average vehicle occupancy rate	–
ψ	Ratio of the maximum to the average dwell time	–
χ	Ratio of the period hours to the total service hours	–
$\omega_{(1,2)}$	Parameters of the high frequency penalty	–

operating cycle time plus some time components related to providing of variability in running times and operations at terminals.

More formally, we model the operating cycle time as the sum of four terms described in the following. We assume that a TU accelerates up to and decelerates from a speed S_{max} which is the maximum allowable speed. The TU loses an average t_u minutes per

km. This time loss occurs mainly at intersections and diminishes with higher investments in runningway improvements and traffic signal priority (TSP) systems. By "runningway" we intend the continuous infrastructure occupied by the TUs, including the lanes or tracks, any major structures required to support or accommodate the lanes or tracks, and any continuous systems such as electric traction power supply that is specifically required by a technology.

The average speed excluding user service at stops, S_{run}, is

$$S_{run} = \frac{1}{\frac{1}{S_{max}} + \frac{t_u}{60}} \quad S_{max}[\text{km/h}], t_u[\text{min/km}]. \tag{1}$$

The resulting running time, denoted R, is equal to $2L/S_{run}$ and is the first term of the operating cycle time. We assume that on average a TU leaves a stop accelerating up to S_{run}, 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

$$T_a = \frac{S_{run}}{25920}\left(\frac{1}{\bar{a}} + \frac{1}{\bar{b}}\right) \quad S_{run}[\text{km/h}], \bar{a}, \bar{b}[\text{m/s}^2], \tag{2}$$

(see e.g. Vuchic and Newell 1968). 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:

$$T_l = T_a + \frac{t_d}{3600} \quad T_a[\text{h}], t_d[\text{s}]. \tag{3}$$

Because the number of stops is equal to $2L/d$, the second term of the operating cycle time is $2LT_l/d$. The third term of the operating cycle time expresses the load-dependent dwell time which is related to T_b, the boarding and alighting time per user of a TU, and the number of passengers using a TU, given by q/f. The boarding and alighting time of a TU depends on the number n of vehicles per TU and their door configuration. We assume a boarding and alighting service time per user of a vehicle as equal to t_b, thus $T_b = t_b/n$. We introduce a fourth term of the operating cycle time accounting for extra delays at intersections, links, and stops under high frequencies, where for high frequencies we intend those exceeding a threshold frequency $\dot{f}$, for example equal to 25 TU/h. As this frequency is reached the design TSP may underperform, the interactions between signals and stops, as well as disturbances induced by trespassing, may become significant. These phenomena warrant a simulation approach for design purposes; here we propose a synthetic representation by a specific term. This term depends on the ratio of the frequency to this threshold frequency elevated to the power of ω_2, and is proportional by a coefficient ω_1 to the base operating cycle time loss at intersections, which is equal to $2t_uL$. The operating cycle time $t_{oc,p}$ for each period is

$$t_{oc,p}(f_p, d, n_p) = R + \frac{2L}{d} T_l + \frac{t_b}{3600} \frac{q_p}{n_p f_p} + \omega_1 \frac{2 t_u L}{60} \left(\frac{f_p}{\dot{f}} \right)^{\omega_2}$$
$$t_b[\text{s-pax/veh}], q_p[\text{pax/h}], T_l, R[\text{h}], t_u[\text{min/km}], d, L[\text{km}],$$
$$n_p[\text{veh/TU}], f_p, \dot{f}[\text{TU/h}]. \tag{4}$$

The commercial cycle time is obtained by multiplying the operating cycle time by β, a parameter larger than one that accounts for the schedule time recovery at the terminals, and by adding the following two terms. First, a fixed terminal time t_{tf} that accounts for minimum crew rest, securing vehicle keys, and in some cases navigating a terminal loop. Second, a terminal time proportional to t_{tv} and variable with the TU length that accounts for the crew walking distance between the TU ends, and safety checks. The commercial cycle time, in the following referred to as cycle time for brevity, is

$$t_{c,p}(f_p, d, n_p) = \beta t_{oc,p}(f_p, d, n_p) + \frac{t_{tf} + t_{tv} n_p}{3600}$$
$$t_{oc,p}[\text{h}], n_p[\text{veh/TU}], t_{tf}[\text{s}], t_{tv}[\text{s/veh}]. \tag{5}$$

We note that the accuracy of β as a linear factor goes down with shorter frequencies because of the rounding requirements within real-world schedules, but the accuracy of β is usually satisfactory at higher frequencies.

3.2 Passengers' time value

The passenger time value is a monetized value of time composed of three parts: access and egress, waiting, and in-vehicle time values.

Users access to and egress from the nearest stop at speed s. The average distance is d / 4 at the origin and at the destination, the average total access and egress length is d / 2, and the average access and egress time t_a of a user is

$$t_a(d) = \frac{d}{2s} \quad d[\text{km}], s[\text{km/h}]. \tag{6}$$

The value of one unit of access and egress time is expressed by the parameter V_a, and the average access and egress value C_a is

$$C_a(d) = V_a \frac{d}{2s} q \sum_p \gamma_p \chi_p. \tag{7}$$

Waiting time depends on the frequency and we distinguish between low, medium and high frequencies. In the case of high frequencies, defined as those above a

threshold frequency f_m, users arrive at stops at a constant rate but TU platooning starts to occur. We note that we use the term "platooning" to indicate the unorganized formation of clumps of vehicles due to the discrete red-green operation of traffic signals, not the organized "platoons" of vehicles dispatched together. Because of platooning, the additional TU capacity provided by frequencies larger than f_m do not yield a further reduction of the average headway with respect to $1/f_m$. Thus, the average waiting time t_w can be modeled as a fraction $\epsilon \geq 1/2$ of the expected headway. Values of ϵ strictly larger than 1 / 2 can model cases where the headways have a large variance. In the case of medium frequencies, users arrive at stops at a constant rate and the average waiting time is ϵ/f. In the case of low frequencies, users follow timetables and arrive at stops w minutes before the expected time of service. The waiting time saved by this behavior still has a cost for the user but is discounted by a factor μ less than one, for example $\mu = 1/3$ as in Tirachini et al. (2010). The threshold frequency for these two former behavior regimes is defined by f_l, for example six TU per hour, which results in a headway of 10 minutes. The average waiting time t_w of a user is

$$t_w(f) = \begin{cases} \frac{w}{60} + \mu\frac{\epsilon}{f} & \text{if } f < f_l \\ \frac{\epsilon}{f} & \text{if } f_l \leq f \leq f_m, \\ \frac{\epsilon}{f_m} & \text{if } f > f_m \quad f[\text{TU/h}], w[\text{min}] \end{cases} . \tag{8}$$

Figure 1 illustrates this new waiting time function and the difference with respect to a previous study.

The average value of waiting $c_w(f)$ borne by q users at the frequency f is

$$c_w(f) = V_w t_w(f) q \quad V_w[\$/\text{pax-h}], t_w[\text{h}], q[\text{pax/h}], \tag{9}$$

where V_w is the value of one waiting time unit. The average value of waiting for the $\hat{p}$ periods, C_w, is

$$C_w(\mathbf{f}) = \sum_p c_w(f_p)\gamma_p\chi_p \quad c_w[\$/\text{h}], \tag{10}$$

where we indicate by $\mathbf{f}$ the vector of $\hat{p}$ frequencies.

The average in-vehicle time t_v of a user is modeled as a fraction of the operating cycle time t_{oc}. This fraction is equal to the ratio of the average trip length l to the total distance $2L$ covered by a TU in a cycle. For notational compactness we indicate by λ the ratio of the average trip length to the route length, $\lambda = l/L$, and thus:

$$t_v = \frac{\lambda}{2} t_{oc} \quad t_{oc}[\text{h}]. \tag{11}$$

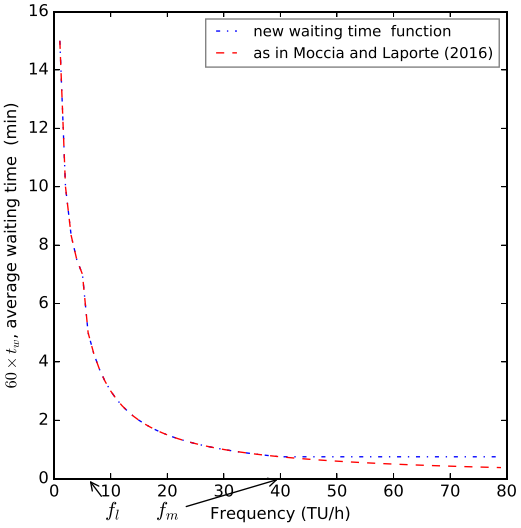

Fig. 1 An example of the new waiting time function and that of Moccia and Laporte (2016) with $f_l = 6, f_m = 40, \epsilon = 0.5, \mu = 0.33$, and $w = 5$

The value of the in-vehicle travel time is multiplied by the crowding penalty function Δ introduced in "Appendix 2". This penalty function aims at reducing underestimation of crowding when it is synthetically applied to the average of the vehicle occupancy rate $\bar{\theta}$. As explained in "Appendix 2", this improvement is accomplished by the introduction of a parameter ϕ belonging to the interval [1, 2] and such that $\phi\bar{\theta}$ is the maximum value of the vehicle occupancy rate. Figure 2 illustrates this new penalty function and the difference with respect to previous studies.

At each period, the average occupancy rate $\bar{\theta}$ depends on the frequency and on the TU length. For notational compactness in the following two equations we have dropped the period subscript. The average occupancy rate is

$$\bar{\theta}(f, n) = \frac{lq}{2Lknf}, \tag{12}$$

where we denote by k the passenger capacity of a vehicle, and therefore the capacity K of a multi-unit TU is equal to $k \times n$. The penalty function Δ is here expressed as a function of the frequency and of the TU length:

$$\Delta(f, n) = \begin{cases} 1 + \rho\left(\dfrac{lq}{2Lknf} - \theta_{min}\right) & nf \leq \dfrac{lq}{2Lk\theta_{min}}(2 - \phi) \\ 1 + \rho\left(\phi\dfrac{lq}{2Lknf} - \theta_{min}\right)^2 \dfrac{Lknf}{2lq(\phi - 1)} & \dfrac{lq}{2Lk\theta_{min}}(2 - \phi) < nf < \phi\dfrac{lq}{2Lk\theta_{min}} \\ 1 & nf \geq \phi\dfrac{lq}{2Lk\theta_{min}} \end{cases}. \tag{13}$$

We denote by Δ_p the resulting $\hat{p}$ penalty functions.

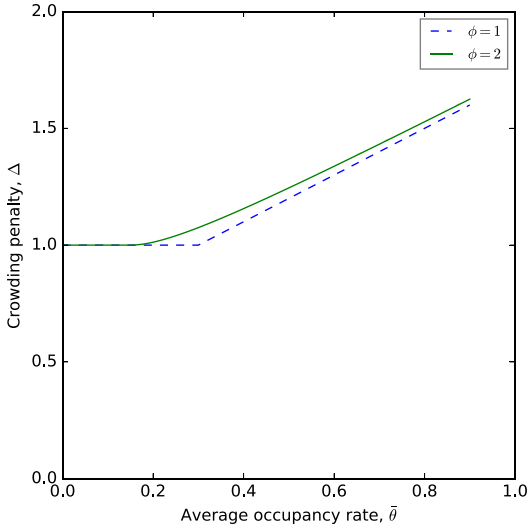

Fig. 2 An example of the crowding penalty function. With $\phi = 1$ the occupancy rate is constant and the crowding penalty is the same as that of Moccia and Laporte (2016). With $\phi = 2$ a maximum correction is applied

Let V_v be the value of one unit of in-vehicle time, then the value of in-vehicle time, C_v, is

$$C_v(\mathbf{f}, d, \mathbf{n}) = V_v \frac{\lambda}{2} q \sum_p \chi_p \gamma_p \Delta_p(f_p, n_p) t_{oc,p}(f_p, d, n_p), \tag{14}$$

where we indicate by $\mathbf{n}$ the vector of $\hat{p}$ TU lengths. The total passengers' time value C_u is then

$$C_u = C_a + C_w + C_v \quad C_a, C_w, C_v[\$/\text{h}]. \tag{15}$$

3.3 Operator cost

The operator cost has five components. The first is the construction and maintenance of the route and is indicated by c_{0l}. The second includes the construction and maintenance costs of the stations. For each station this term has a fixed part c_{0s}, and a variable part, c_{0sv}, which depends on the TU length at peak hours. The third depends on the fleet size and reflects vehicle capital and administrative costs. Let c_{1v} be the unit operator cost per vehicle-hour which accounts for the capital and administrative costs. The deployed fleet size B of TUs is the product of frequency and cycle time: $B = ft_c$. The vehicle fleet size is equal to ζnB, where $\zeta > 1$ provides for O&M spares. The fourth expresses the crew costs and depends on c_{1t}, the unit operator cost per TU-hour. The fifth accounts for running costs such as energy, tires, lubricants, etc. Let c_{2v} be the unit operator cost per vehicle-km. The amount of vehicle-km is the product of the commercial speed S and the fleet size. The commercial speed is obtained by dividing the total length $2L$ by the cycle time. Thus, the amount of vehicle-km is $S \times nB = 2L/t_c \times nft_c = 2Lnf$. The operator cost C_o is then

$$\begin{aligned} C_o(\mathbf{f}, d, \mathbf{n}) =& c_{0l} + (c_{0s} + c_{0sv}(n_1 - 1))\frac{2L}{d} + c_{1v}\zeta n_1 f_1 t_{c,1}(f_1, d, n_1) \\ &+ c_{1t} \sum_p \chi_p f_p t_{c,p}(f_p, d, n_p) + 2c_{2v} L \sum_p \chi_p n_p f_p. \end{aligned} \tag{16}$$

3.4 Side constraints

The frequency is constrained to be equal to or larger than f_{min}, and less than or equal to f_{max}. The value f_{min} can be set by a "policy headway" rationale, i.e. there is a minimum guaranteed frequency f_{pol}, or can account for capacity as follows. Let $\alpha q \tau$ be the largest load served by the line in a generic period, where $\alpha \leq 1$, and ν be a spare capacity design factor. For example, a value of ν smaller than one accounts for random demand fluctuations and represents a safety margin, whereas ν larger than one allows crush loading. Thus, f_{min} is

$$f_{min} = \max\left(f_{pol}, \frac{\alpha q \tau}{\nu k n}\right) \quad f_{pol}[\text{TU/h}], q[\text{pax/h}], k[\text{pax/veh}], n[\text{veh/TU}]. \tag{17}$$

The f_{max} is defined according to general principles from the Transit Capacity and Quality of Service Manual (TCQSM 2013). We set a cap $\check{f}_{max}$ on the value of f_{max} to reflect rail vehicle operation with drivers responsible for maintaining safe separation. For buses $\check{f}_{max}$ can attain a larger value, twice as high as the one for rail. TCQSM (2013) effectively defines f_{max} as the maximum number of transit vehicles that can pass a given location in a given time period, and is inversely proportional to a minimum achievable headway under which acceptable service can be maintained. For BRT and LRT operations within or parallel to urban arterial roadways the minimum achievable headway depends on the longest dwell time. We assume that the longest dwell time is ψ times the average dwell time, where ψ is a parameter larger than one. The average dwell time t_e is a function of frequency, stop spacing, and TU length

$$t_e(f, d, n) = t_d + t_b \frac{qd}{2nfL} \qquad \begin{array}{l} t_d[\text{s}], t_b[\text{s-pax/veh}], q[\text{pax/h}], \\ d, L[\text{km}], n[\text{veh/TU}], f[\text{TU/h}]. \end{array} \tag{18}$$

We simplify the f_{max} formula for the bus and light rail modes to share the following common form:

$$f_{max}(f, d, n) = \min\left(\check{f}_{max}, \frac{3600}{t_{e0} + t_{ev}(n-1) + \eta\psi t_e}\right) \qquad \begin{array}{l} \check{f}_{max}[\text{TU/h}], \\ t_{e0}, t_{ev}, t_e[\text{s}]. \end{array} \tag{19}$$

The values of t_{e0}, the fixed stop clearance time, t_{ev}, the stop clearance time for an extra vehicle length, and η, the fraction of the longest dwell time, are set in accordance with the vehicle characteristics we define for each scenario, and with the assumed station and runningway characteristics for which capital cost estimates are established.

Finally, we define side constraints on the average stop spacing. Reaching the speed S_{max} requires a stop spacing larger than a threshold value d_{min}, which depends on acceleration and deceleration rates

$$d_{min} = \frac{S_{max}^2}{25920}\left(\frac{1}{\bar{a}} + \frac{1}{\bar{b}}\right) \quad S_{max}[\text{km/h}], \bar{a}, \bar{b}[\text{m/s}^2], \tag{20}$$

(see e.g. Vuchic and Newell 1968). Values of d less than this minimum distance will not be allowed. We observe that d_{min} will be different for types of vehicles and speed limits, and is a constraint that is rarely binding. An upper bound d_{max} is defined as well. We remark that the optimal d is an average stop spacing that does not impede specific implementations where the stop spacing varies along the route.

3.5 Optimization model

The total cost C_{tot}, sum of passengers' time value and operator cost, is a function of frequencies, stop spacing, and consists. The model follows:

$$\text{minimize } C_{tot}(\mathbf{f}, d, \mathbf{n}) \tag{21}$$

subject to

$$d_{min} \leq d \leq d_{max} \tag{22}$$

$$\max\left(f_{pol,p}, \frac{\alpha q_p \tau_p}{\nu n_p k}\right) \leq f_p \leq f_{max,p}(f_p, d, n_p), \quad \forall p \in \{1, \ldots, \hat{p}\} \tag{23}$$

$$n_{min} \leq n_p \leq n_{max},\ n_p \in \mathbb{N}, \quad \forall p \in \{1, \ldots, \hat{p}\} \tag{24}$$

$$n_p f_p t_{c,p}(f_p, d, n_p) \leq n_1 f_1 t_{c,1}(f_1, d, n_1), \quad \forall p \in \{2, \ldots, \hat{p}\}. \tag{25}$$

Constraints (22) set minimum and maximum values for the stop spacing. Constraints (23) enforce minimum and maximum values for the frequencies. Constraints (24) specify the feasible range of TU lengths, and constraint (25) ensures that the maximum fleet is deployed at peak times, where for "maximum deployed fleet" we refer to the vehicles needed for the scheduled service (the reserve that bears a capital cost is accounted in the operator cost function). The model is solved by an updated version of the algorithm presented in Moccia and Laporte (2016) where the new nonlinear constraint, the right side of (23), is imposed iteratively, and the lower convex approximation scheme, reported in "Appendix 3", accounts for the new definitions (4), (5), (13), and (16).

4 Scenarios

We consider two scenarios that capture operational aspects dependent upon the interaction between service frequency and crowding, with infrastructure for each designed to offer comparable performance for both bus and rail until capacity is approached. These scenarios are based on an assessment of the literature (Vuchic 2005; Vuchic et al. 2012; Casello et al. 2014; Bruun et al. 2018), and on IBI Group's internal database of transit projects. These scenarios represent design values that planners of new systems in developed countries with high labor costs are likely to use and thus represent best practices. A preliminary version of these scenarios for a two-period case was presented in Moccia et al. (2016). Because some introduced model features were not present in Moccia et al. (2016) the following results are new.

Scenario 1 consists of a dedicated arterial RoW obtained by a simple change in use of an existing lane. We assume an average of six intersections per km, four of which are signalized. TSP is assumed to be present in the relatively modest form prevalent in North America. Stations are part of the curb or sidewalk environment, with only modest passenger shelters and amenities, but allow level boarding. Fares are collected on board, so no fare vending occurs at stations. Most stations are not multi-modal transfer points; users access and egress stations by walking at a speed of $s = 4$ km/h. Rail vehicles are single-unit low-floor trams of three different lengths ranging from 34 to 56 m. A 24 m tram was also considered in Moccia et al. (2016) but we have omitted it because it is indicated to be dominated in both the current and previous models. For buses we consider a single-articulated bus of 18 m, and a double-articulated bus of 24 m, the latter of which would generally require special authorization in mixed rights of way. Platooning is assumed to occur for buses at frequencies in the range between 30 and 80 TU/h, whereas it does not occur for trams because their frequency is capped at 40 TU/h. For each technology, the two terminals have a loop that is not part of the service route length and vehicles are in a single-ended configuration. The maximum authorized speed is assumed to be the same for each technology (50 km/h). In the case of rail technology, this might commonly be referred to as an enhanced streetcar or tram. In the case of bus technology, the labels "enhanced bus service", or "BRT lite" might apply.

Scenario 2 represents an exclusive at grade RoW when a new laterally separate alignment is developed. Intersections are assumed to be identical to the above scenario, but TSP is provided in a more robust form prevalent in many European cities. Stations provide for off-vehicle fare collection, ample passenger amenities including weather protection and lighting, off-board fare collection, level boarding, a distinct architectural treatment and branding appropriate to the streetscape, and, in some cases, revisions to the streetscape as may be warranted to accommodate the stations without interfering with sidewalks. Stations are multi-modal transfer points so that access and egress occur at an average speed $s = 12$ km/h, i.e. a bike speed is representative of the system average between the walking speed and the speed of motorized transport. Rail TUs may operate as multiple units with a consist ranging from one to four cars, and are served by stub terminals, i.e. they are in a double-ended

configuration. Capital and O&M cost figures for rail in this scenario consider the higher complexity of variable consists. Bus vehicles are the same as for Scenario 1 and change directions at the terminals via a loop that is not part of the service route length. Platooning is assumed to occur for buses at frequencies in the range between 40 and 80 TU/h, whereas it does not occur for rail TUs because their frequency is capped at 40 TU/h. The maximum authorized speed is assumed to be the same for each technology (75 km/h). These might commonly be referred to as LRT and BRT, in the case of rail and bus technology, respectively.

In both scenarios, stops and stations are configured to allow in-line double-berthing by buses, and offer all-door level boarding. In both cases, higher frequencies cause increased delays at stops and stations due to queuing and higher demand causing longer dwell times.

We consider a route of 20 km with users traveling on average a fraction $\lambda = 0.45$ of this length, i.e. 9 km per trip. Scenario variants with different values of λ are discussed in Sect. 5.4.

The flow at the peak direction and at the most loaded section is equal to 40% of the total bidirectional demand. This assumption represents a moderate centripetal demand where, for example, 60% of the bidirectional demand moves in the heavy or peak direction, and 67% of these users traverse the most loaded section. Scenario variants of this assumption are presented in Sect. 5.4.

The passenger time values were derived as follows. The in-vehicle value of time is assumed to be the base VoT. A wide defendable range for this parameter can be assumed as discussed in Litman (2017). We decided for the US-DOT (2011) estimate of the average surface transportation VoT equal to 12.5 in year 2009 US dollars, updated for inflation to 13.37 in year 2012 US dollars. The waiting and the access VoTs were obtained by multiplying the base VoT for 1.25 and 1.5, respectively, as in Tirachini et al. (2010) and Moccia and Laporte (2016). A scenario variation analysis on the VoT assumptions is carried out in Sect. 5.4.

A technology is indicated by three letters, "BRT" or "LRT", to distinguish between bus and rail, respectively, followed by a number, one or two, referring to the scenario, and, in the case of a single-unit technology, a number specifying the length of the vehicle in meters. For example, "BRT1_18" refers to the BRT of Scenario 1 with a bus of 18 m, and "LRT2" indicates the multi-unit rail technology of Scenario 2. The set of parameters related to the users and to the transit line that are common to all scenarios is listed in Table 2. Table 3 reports parameters that are scenario-specific and Table 4 lists those that are mode-specific. Parameters that are technology-specific are reported for Scenario 1 and 2 in Tables 5 and 6, respectively. Monetary figures are expressed in year 2012 US dollars.

The demand profile is derived from turnstile data of the Boston subway network presented in "Appendix 1". The parameters of the demand profile are listed in Table 7.

Capital costs are annualized assuming a discount rate of 3% over a one-stage technical life for the vehicles and the infrastructure. The discount rate is indicated by the symbol ι, and the one-stage technical life by y. The one-stage technical life is lower than a typical service lifetime because it expresses the equivalent years including cost of a mid-life rebuild at the prevalent discount rate. For example, a bus may have a useful

Table 2 Parameters related to the users and to the transit system that are common to both scenarios and to all technologies

Parameter	Symbol	Unit	Value
Unit operator cost per TU-hour	c_{1t}	\$/TU-h	60
Threshold frequency for timetable behavior	f_l	TU/h	6.0
Threshold frequency for the high frequency penalty	$\dot{f}$	TU/h	25
Number of service hours per year	H	h/year	5940
Route length	L	km	20.0
Unit value of access time	V_a	\$/h	20.05
Unit value of waiting time	V_w	\$/h	16.71
Unit value of in-vehicle time	V_v	\$/h	13.37
Waiting time at a stop when $f < f_l$	w	min	5
One-stage infrastructure technical life (route and stations)	y	year	40
Fraction of demand in the most loaded segment of the line	α	–	0.40
Multiplicative factor of the operating cycle time	β	–	1.07
Rate of the average waiting time to the headway	ϵ	–	0.5
Spare capacity factor for the fleet	ζ	–	1.20
Average occupancy rate up to $\delta = 1$	θ_{min}	–	0.3
Discount rate	$\imath$	–	0.03
Ratio of the average trip length to the length of the route	λ	–	0.45
Discount factor of the waiting time under timetable behavior	μ	–	0.33
Spare capacity factor for the TU	ν	–	0.95
Ratio of the residual value to the initial value of the rolling stock	ξ	–	0.05
Ratio of the residual value to the initial value of the infrastructure	ξ	–	0.00
Slope of the linear part of δ	ρ	–	1.0
Ratio of the maximum to the average occupancy rate	ϕ	–	1.8
Ratio of the maximum to the average dwell time	ψ	–	2.0

Table 3 Scenario-specific parameters

Parameter	Symbol	Unit	Scenario 1	Scenario 2
Maximum allowed speed	S_{max}	km/h	50	75
Land acquisition unitary cost	–	m\$/hectare	1.0	10.7

service life of 15 years with a mid-life rebuild, but because that overhaul costs 25% of the initial vehicle cost, the equivalent one-stage bus technical life is assumed to be 12 years, which fits the one-stage capital accounting in the relevant range of discount rates (1–10%). Similarly, runningway and station infrastructure were annualized on the basis of a 40-year one-stage technical life, determined from the inclusion of the periodic re-investments of different elements on specific timeframes. A 5% residual value is factored into the annuity computation for the rolling stock by setting a parameter ξ to 0.05.

Table 4 Mode-specific parameters

Parameter	Symbol	Unit	BRT	LRT
Route width (two-ways)		m	10	9
One-stage vehicle technical life	y	year	12	25
Average acceleration rate	$\bar{a}$	m/s^2	1.00	1.15
Average deceleration rate	$\bar{b}$	m/s^2	1.15	1.00
Fraction of the longest dwell time in the maximum frequency formula	η	–	0.571	1.000
Cap on the maximum frequency	$\check{f}_{max}$	TU/h	80	40
Parameter of the high frequency penalty	ω_1	–	0.075	0.135
Exponent of the high frequency penalty	ω_2	–	1.25	1.40

No residual value is assumed for the infrastructure, $\xi = 0$. In general, for infrastructure and rolling stock, the capital amortization per hour of service is computed as

$$\frac{P(1-\xi)\imath}{H(1-(1+\imath)^{-y})} \tag{26}$$

where P is the purchase price and H is the number of service hours in a year. Land capital cost is annualized by multiplying it by the discount rate, i.e. assuming an infinite service life.

Vehicle capacities are determined such that all vehicles for both technologies have the same ratio of seats to total users. This ratio is equal to 0.3, equal to θ_{min}, the threshold load ratio of the δ penalty function up to which there is no penalty. The total user capacities for rail are computed assuming an industry-standard 4 standees per available square meter, whereas for bus we use 3.6 standees per square meter to account for the higher vertical accelerations and less controlled lateral accelerations experienced on buses which necessitates that passengers use more space to brace themselves.

5 Results

Optimal results of the decision variables and of other relevant indices will be presented as functions of the density of demand, a key input variable for any system designer. This latter index, referred to as passenger travel density (PTD), expresses the amount of traveled distance by passengers per unit of route length. More formally, PTD is

$$\frac{Hql\sum_p \chi_p\gamma_p}{L} = Hq\lambda\sum_p \chi_p\gamma_p. \tag{27}$$

Table 5 Scenario 1, technology-specific parameters

Parameter	Symbol	Unit	BRT1_18	BRT1_24	LRT1_34	LRT1_45	LRT1_56
Route capital cost		m$/km	7.33	7.33	11.63	11.63	11.63
Route maintenance cost		m$/km-year	0.006	0.006	0.028	0.028	0.028
Stop capital cost, one-way		m$/stop	0.33	0.35	0.36	0.38	0.41
Stop maintenance cost, one-way		$/stop-year	11,587	14,302	18,827	22,900	28,782
Vehicle capital cost		m$/veh	0.75	1.0	3.2	3.9	4.6
Vehicle capacity	k	pax/veh	114	154	227	317	400
Unit operator cost per vehicle-km	c_{2v}	$/veh-km	1.83	2.47	3.51	4.46	5.70
Vehicle administrative cost	c_{1vb}	$/veh-year	44,230	45,437	56,129	73,211	90,294
Time lost at intersections per average km	t_u	min/km	1.05	1.05	0.99	0.99	0.99
Fixed time lost for a stop (doors and others fixed times)	t_d	s	6.0	7.0	9.0	9.0	10.0
Vehicle boarding and alighting time per user	t_b	s-veh/pax	4.5	3.0	1.5	1.1	0.9
Fixed terminal time	t_{tf}	s	360	360	360	360	360
Fixed stop clearance time	t_{e0}	s	32	34	57	58	59

Table 6 Scenario 2, technology-specific parameters

Parameter	Symbol	Unit	BRT2_18	BRT2_24	LRT2
Route capital cost	–	m$/km	8.90	8.90	15.58
Route maintenance cost	–	m$/km-year	0.013	0.013	0.042
Stop capital cost, one-way	–	m$/stop	1.36	1.46	1.42
Stop maintenance cost, one-way	–	$/stop-year	37640	45276	35236
Incremental stop capital cost per additional vehicle, one-way	–	m$/stop-veh	0.00	0.00	0.73
Incremental stop maintenance cost per additional vehicle, one-way	–	$/stop-veh-year	0	0	24184
Vehicle capital cost	–	m$/veh	0.75	1.00	2.9
Vehicle capacity	k	pax/veh	114	154	191
Unit operator cost per vehicle-km	c_{2v}	$/veh-km	1.71	2.30	1.96
Vehicle administrative cost	c_{1vb}	$/veh-year	44230	45437	54908
Time lost at intersections per average km	t_u	min/km	0.74	0.74	0.62
Fixed time lost for a stop (doors and others fixed times)	t_d	s	6.0	7.0	7.0
Vehicle boarding and alighting time per user	t_b	s-veh/pax	2.1	1.4	1.4
Fixed terminal time	t_{tf}	s	360	360	210
Variable terminal time	t_{tv}	s	0	0	80
Minimum number of vehicles per transit unit	n_{min}	veh/TU	1	1	1
Maximum number of vehicles per transit unit	n_{max}	veh/TU	1	1	4
Fraction of the longest dwell time in the maximum frequency formula	η	–	0.571	0.571	1.000
Fixed stop clearance time	t_{e0}	s	32	34	57
Stop clearance time for an extra vehicle length	t_{ev}	s/veh	0	0	2

Table 7 Parameters related to the demand profile common to both scenarios and to all technologies

Ratio	Symbol	$p = 1$	$p = 2$	$p = 3$
Period average demand to average peak demand	γ_p	1.00	0.51	0.17
Maximum to average peak period demand	τ_p	1.38	1.44	2.28
Period hours to total service hours	χ_p	0.12	0.22	0.66

We indicate the PTD unit of measure as "pax-km/year-km" because the "km" at the numerator has a different meaning of that to the denominator—the former refers to the unitary distance traveled by a passenger while the latter expresses the unitary route length. The rationale for PTD as the index for the amount of service provided is that it incorporates the ratio between the average trip length and the route length, and the number of service hours in a year. This index allows a degree of abstraction from some of our parameter choices as we illustrate in the following by a scenario variation analysis. The PTDs are large numbers and therefore we use a logarithm scale with the exception of the frequency and consist figures where a linear scale is preferred to highlight the proportionality of these variables to the demand. For the frequency figures we also use the following secondary index: the users per peak hour in the peak direction at the maximum load segment (MLS)—an index often used in practice.

The remaining of this section is structured as follows. Sections 5.1 and 5.2 report numerical results for the Scenarios 1 and 2, respectively. Section 5.3 presents a model's validation by comparing computed O&M costs with those of light rail systems in North America. Sensitivity analysis and scenario variants are discussed in Sect. 5.4.

5.1 Results for Scenario 1

Figure 3 illustrates the average total cost of the studied technologies in Scenario 1. The breakeven points between the technologies occur at PTDs between 6.3 and 8.6×10^6. However, we note that there is a wider range where cost differences are small, and therefore other factors beside passengers' time value and operator cost could be decisive.

Figure 4 compares the BRT1_18 optimal and minimum frequencies for all periods. At the peak period, $p = 1$, the optimal frequency aligns with the minimum feasible value. This means that even from a total cost perspective providing extra frequent service (i.e. a frequency higher than the minimum) at the peak hour is suboptimal. At the medium demand period, $p = 2$, extra-service is provided with a reduced difference at high frequencies when platooning starts to occur. At the low demand period, $p = 3$, the minimum frequency again becomes the dominating factor. These observations also apply to the BRT1_24 optimal frequencies, and hence such a figure is omitted for brevity. Figure 5 compares the LRT1_34 optimal and minimum frequencies for all periods. At the peak period, the tram optimal frequency behavior is the same as that of buses, i.e the minimum frequency rules. In contrast,

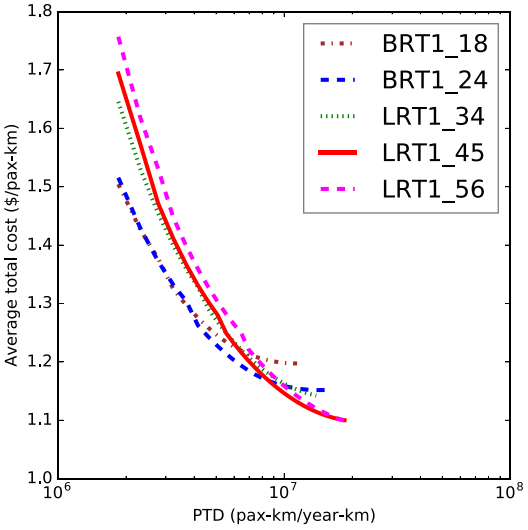

Fig. 3 Scenario 1, average total cost, $C_{tot}/(ql\sum_p \chi_p\gamma_p)$

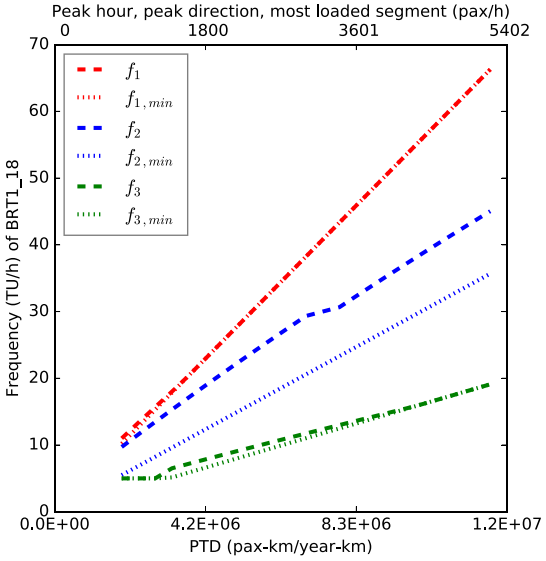

Fig. 4 Scenario 1, mode BRT1_18, optimal and minimum frequencies for all periods

at the medium and low demand periods significantly more frequent service is provided with respect to it being strictly necessary. This results from the lower operating expenses of rail vs. bus when the fleet acquisition cost is considered as a sunk cost. Similar observations hold for the remaining Scenario 1 rail modes, and hence figures are omitted for brevity.

Optimal stop spacings, illustrated in Fig. 6, vary in a small range across the studied PTDs. Differences between the technologies are small and follow the approximate formula (34) of Moccia and Laporte (2016) which holds as well for the model of this paper. We further observe that PTD is proportional to λ, the ratio of the

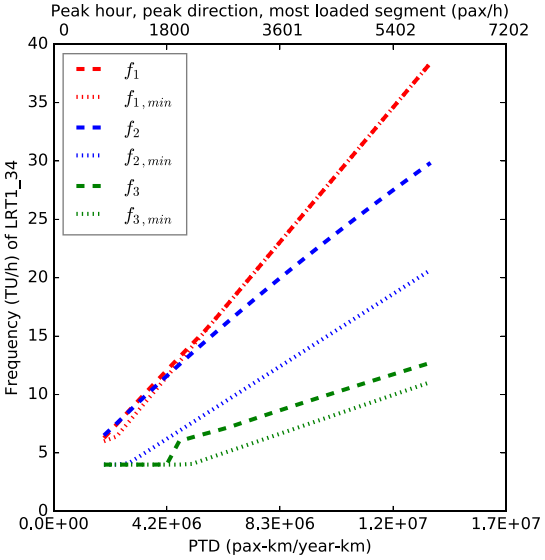

Fig. 5 Scenario 1, technology LRT1_34, optimal and minimum frequencies for all periods

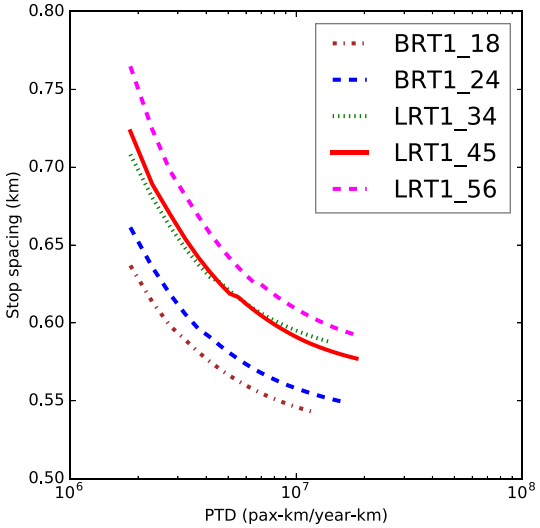

Fig. 6 Scenario 1, optimal stop spacings

average trip length to the route length, but the reported stop spacing figures are computed for a fixed λ. In those figures, PTD varies because q, the average demand at the peak period, varies. Formula (42), the unconstrained optimal stop spacing of the approximation, explains the observed decrease of the stop spacing with an increasing q: the component related to the stop fixed costs is divided by q. This formula also indicates that, *ceteris paribus*, the stop spacing would increase with λ as we will show in Sect. 5.4.

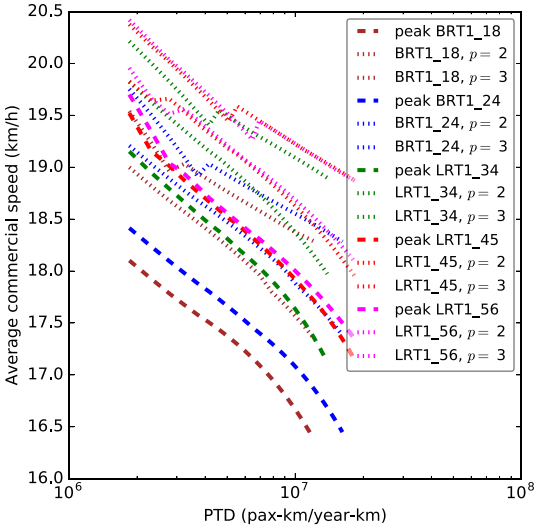

Fig. 7 Scenario 1, average commercial speed for all periods

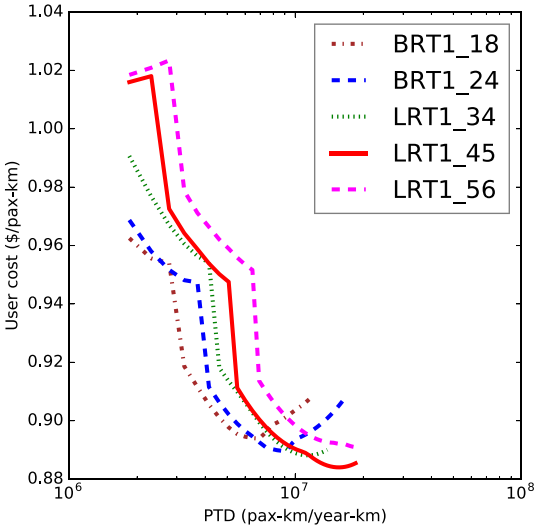

Fig. 8 Scenario 1, average passenger time value, $C_u/(ql\sum_p \chi_p\gamma_p)$

At the high frequencies required by higher PTDs, user inconvenience occurs both because of crowding and of longer operating cycle times. The degradation of cycle time can be observed in Fig. 7 which depicts the average commercial speeds. This phenomenon can be further appreciated by disaggregating the total cost into average passenger time value (for brevity in the following indicated by "user cost", see Fig. 8) and average operator cost (Fig. 9). The former increases with demand density when the optimal frequency deviates from that guaranteeing uncrowded or moderately crowded conditions. For buses at high PTDs the decrease in average operator cost is not sufficient to compensate for the increase in user cost which rail technologies can avoid.

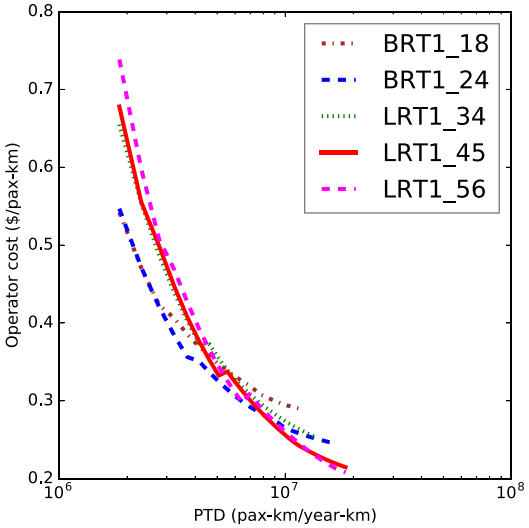

Fig. 9 Scenario 1, average operator cost, $C_o/(ql\sum_p \chi_p\gamma_p)$

We further observe in this scenario that high frequency penalty does not impact breakeven points, which occur at frequencies where this penalty is not significant. The same holds for the platooning effect. The combination of the high frequency penalty and of the self-adjusting maximum frequency constraint are significant in limiting the attainable capacities of the BRTs and increasing the average passenger time value for large demand levels.

This dynamic highlights the crucial role of the high frequency penalty and of the maximum frequency constraint which are two of the main extensions proposed in this paper. These combine to limit operations to capacity in the sense defined by the TCQSM, e.g. "without unreasonable delay and with reasonable certainty". Many semi-rapid bus and rail systems carry throughputs on their maximum load sections that are higher than capacity in this sense, at low operating speeds, severe overcrowding, or other less-than-ideal circumstances which are not likely to be acceptable for a new design. This model does not attempt to represent such overcapacity operations.

We have also evaluated the operational impact of an exclusive operator cost minimization versus a total cost minimization. In this former case the stop distance is set to the optimal value as in total cost minimization. Figures are not presented for brevity, and the results can be synthesized as follows. The resulting optimal frequencies align to the f_{min} policy. The breakeven points between bus and rail recede toward lower PTDs because the waiting time advantage of buses is not included in the single objective of operator cost minimization.

5.2 Results for Scenario 2

Figure 10 illustrates the average total cost of the studied technologies in Scenario 2. The breakeven points between the technologies occur at PTDs between 7.7 and 11.2 $\times 10^6$. Comparing total cost between scenarios, Figs. 3 and 10, it may be concluded

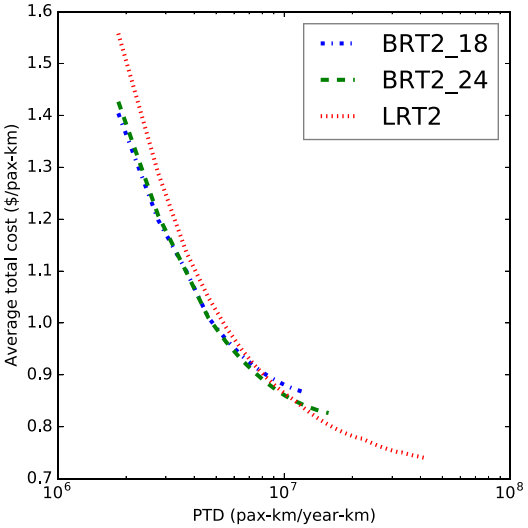

Fig. 10 Scenario 2, average total cost, $C_{tot}/(ql\sum_p \chi_p\gamma_p)$

that the faster and more capital-intensive technologies of Scenario 2 dominate those of Scenario 1 with the only exception of the LRT2 vs BRT1 at the lowest demand density studied.

For buses, the optimal frequencies compared to the minimum frequencies behave similarly to those of Scenario 1 (see Fig. 11 for BRT2_18). Figure 12 shows how the optimal number of rail vehicles per Transit Unit varies between peak and off-peak periods. Figure 13 compares the LRT2 optimal and minimum frequencies for all periods. As in Scenario 1, rail provides more extra frequent service at non-peak periods with respect to buses. The sawtooth pattern of the frequencies highlights

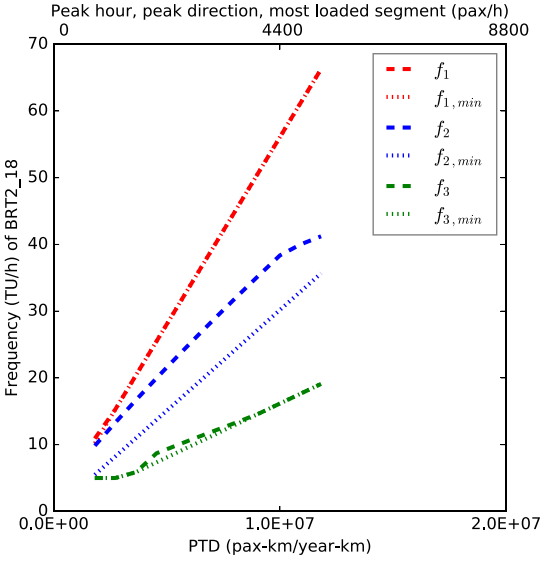

Fig. 11 Scenario 2, technology BRT2_18, optimal and minimum frequencies for all periods

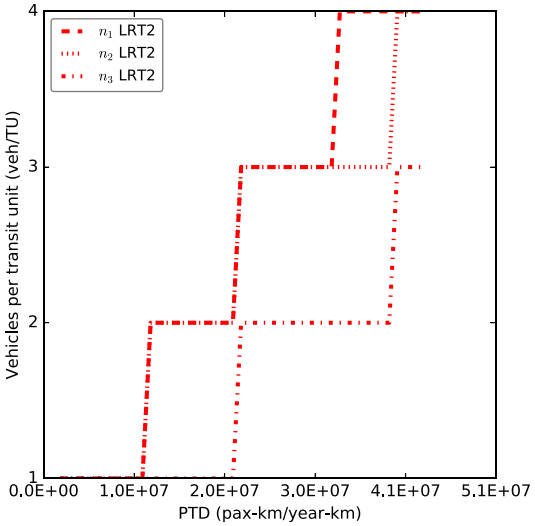

Fig. 12 Scenario 2, technology LRT2, optimal number of vehicles per Transit Unit for all periods

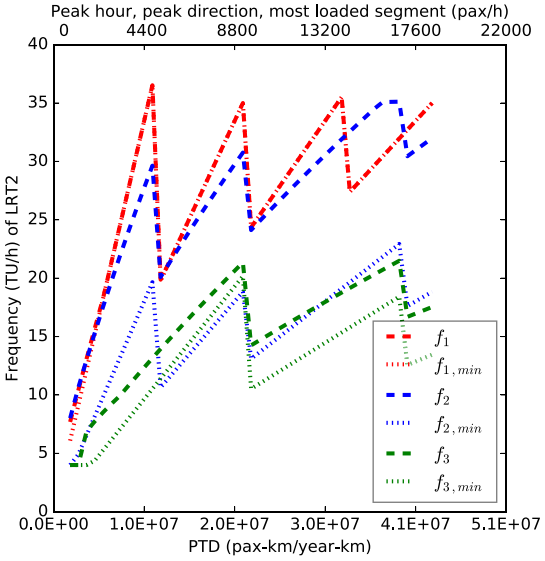

Fig. 13 Scenario 2, technology LRT2, optimal and minimum frequencies for all periods

how flexible capacity can be offered by multiple-unit operation light rail. Stop spacings are illustrated in Fig. 14. We illustrate the degradation of the cycle times in Figure 15, its effect on the average passenger time value, Fig. 16, and the average operator cost, Fig. 17.

As for Scenario 1, we have also evaluated the exclusive operator cost minimization. The breakeven points are similar to those of total cost minimization. The optimal frequencies under operator cost minimization align to a f_{min} policy. However, there is a relevant difference for LRT2 under the two objective functions, total and operator cost minimizations. Operator cost minimization leads to minimum

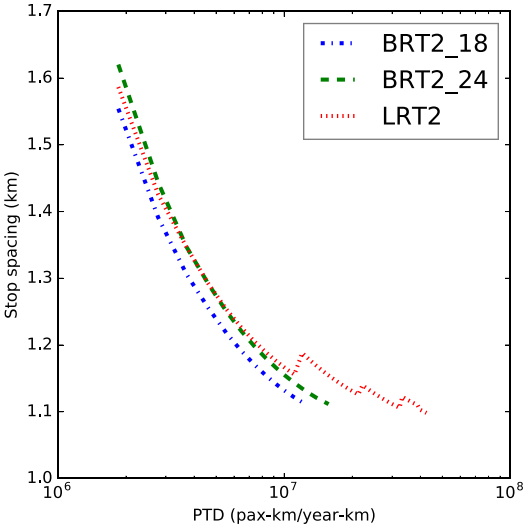

Fig. 14 Scenario 2, optimal stop spacings

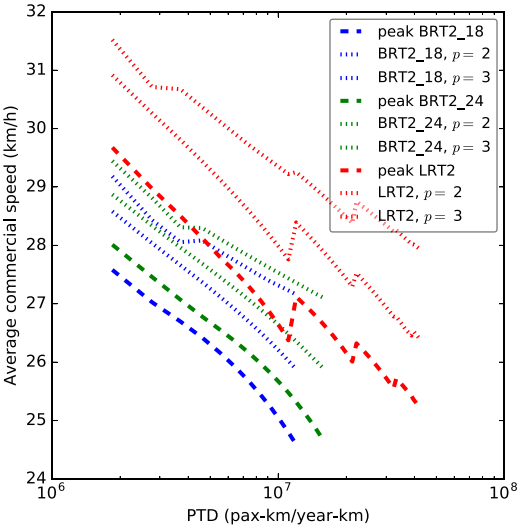

Fig. 15 Scenario 2, average commercial speed for all periods

frequencies with longer consists and thus less frequent service than under total cost minimization.

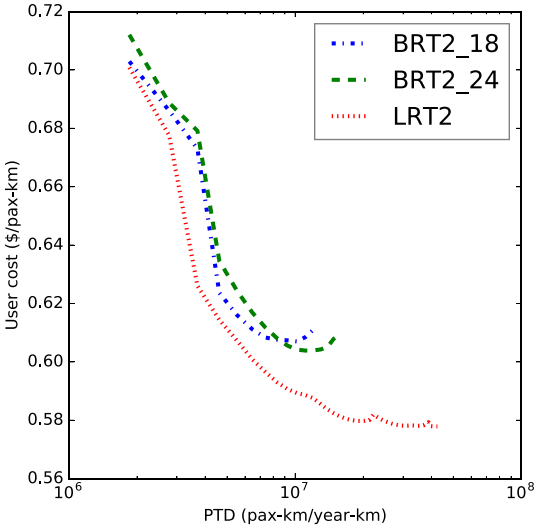

Fig. 16 Scenario 2, average passenger time value, $C_u/(ql\sum_p \chi_p\gamma_p)$

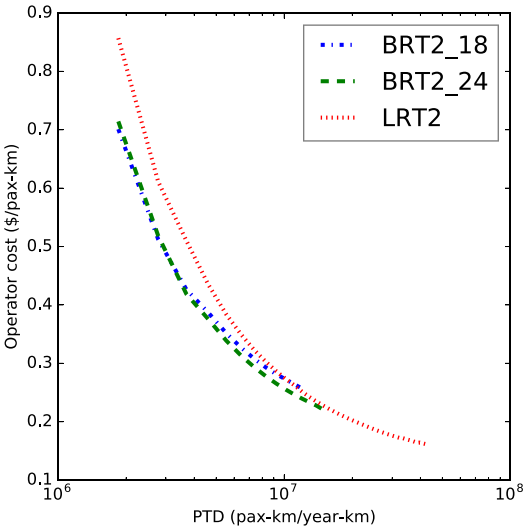

Fig. 17 Scenario 2, average operator cost, $C_o/(ql\sum_p \chi_p\gamma_p)$

5.3 Model validation

The previously presented figures of computed stop spacings, commercial speeds, and consists align with observed best practices in Europe, North America and Australia. Because of space limits, we do not include additional figures of other performance indices, but we report that computed passenger in-vehicle occupancy, labor percentage of O&M cost, fleet size, service vehicle hours and kilometres per year are in good accord with the above mentioned best practices. As a synthetic illustration of this, Fig. 18 compares the computed O&M costs of LRT2 with those of light rail systems in North America. The data points of the

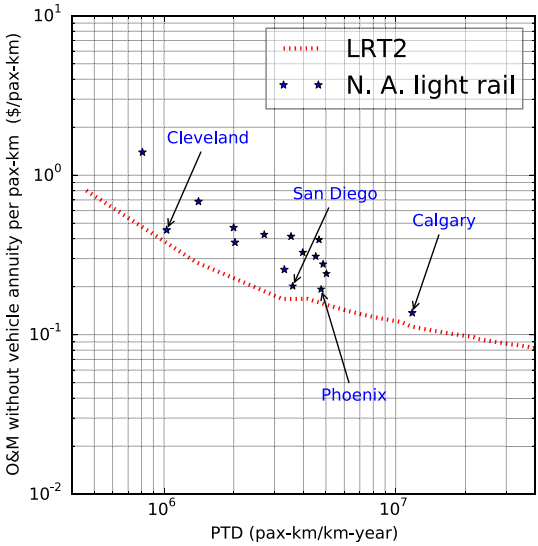

Fig. 18 O&M costs of North America light rail systems *versus* those of the computed LRT2

systems that are widely recognized as some of the best-in-class are well enveloped by the computed values.

5.4 Sensitivity analysis and scenario variants

This section discusses the model's sensitivity and the dependence of our results to VoT assumptions, parameters influencing demand concentration and density, discount rate, and maximum vehicle loading.

The presented model is robust in the strict sense of optimization theory where sensitivity analysis deals with very small input changes. This robustness is a consequence of the fact that the optimal solution is well approximated by square root formulae ("Appendix 3"). However, significantly different assumptions, hence significantly different parameters, may yield different results as discussed in the following.

Table 8 summarizes the breakeven points between the bus and rail technologies in the two base case scenarios and in eleven variants of them. We list breakeven points for the user, operator and total cost curves, but, unless stated otherwise, in the following discussion we refer to the total cost curves when we use the term "breakeven points". Nine scenario variants are indicated by an alphabetical letter from "A" to "I". Two scenario variants are obtained by combining some of the previous nine variants and are labelled by the corresponding set of letters.

Variant A halves all VoTs with respect to the base case. Because of the large VoT reduction, the total cost curves are significantly lower than those of our base scenarios. However, the total cost breakeven points between bus and rail technologies do not change dramatically, see the second data row of Table 8, although there are differences in the user and operator cost breakeven points. This allows us to conclude that our results are robust to uniform small and medium VoT variations.

Table 8 Breakeven PTDs between bus and rail technologies in the base case and in eleven scenario variants

Variant	Experiment	Scenario 1			Scenario 2		
		User cost	Operator cost	Total cost	User cost	Operator cost	Total cost
	Base case	7.6–9.5	4.4–8.0	6.3–8.6	Rail	10.4–15.0	7.7–11.2
A	VoTs halved	10.5–11.0	4.4–6.1	5.6–9.4	n.a.	9.3–15.0	8.5–12.6
B	$V_a = V_w = V_v$	n.a.	4.4–5.8	5.3–7.0	Rail	9.5–15.0	7.0–10.0
C	$\rho = 0$	12.0–13.2	5.0–6.1	7.8–10.6	n.a.	10.2–15.0	9.2–13.7
D	$\phi = 1$	10.6–11.3	5.0–7.7	7.8–10.5	n.a.	9.1–15.0	8.6–12.9
E	Rail V_v cut by 5%	n.a.	5.6–7.9	4.6–5.7	Rail	10.5–15.0	5.7–6.3
F	$\alpha = 0.2$	6.4–7.8	4.6–9.3	6.4–8.6	Rail	24.0	8.0–11.5
G	$\alpha = 0.6$	Bus	3.4–4.2	5.2–7.3	n.a.	6.0–10.2	6.0–8.9
H	$\iota = 6\%$	8.3–9.8	6.9–12.4	7.7–10.9	2.5	15.0	11.3–15.0
I	$\iota = 9\%$	7.5–9.9	10.2–15.0	9.3–13.9	2.9	15.0	15.0
C+G		9.9–10.1	6.6–12.6	9.6–12.6	n.a	15.0	14.3–23.2
B+F		n.a.	3.4–4.2	4.1–5.6	n.a.	6.0–10.3	4.9–6.1

Data are reported in millions of pax-km/year-km. When there is a rail or a bus technology dominance we indicate it by "rail" and "bus", respectively. When the cost curves present multiple crossing points we indicate this by "n.a."

Variant B sets access and waiting VoTs equal to the base in-vehicle VoT. This experiment assesses the influence of the access and waiting time penalties. In fact, because of the wide adoption of info-mobility in recent years and the station quality features embedded in our scenarios' assumptions, it is unlikely that passengers will experience additional discomfort of waiting. This should also be true for the access time given the current emphasis and the popular uptake of active transport (Khreis et al. 2016). We further note that our access value of time formula (7), by assuming a uniform spatial demand distribution around stations, provides an overestimation of access time when the origins and the destinations are instead clustered and centered at stations—as often is the case. We observe that VoT equality among the three passenger time components penalizes BRT, because at the medium and low demand densities this assumption reduces the advantage of a higher frequency and a reduced access distance provided by bus vs. rail technologies (with the waiting component playing the major role). Consequently, the breakeven points shift toward lower demand levels.

Variant C tests the impact of the crowding penalty by setting $\rho = 0$. Removing the crowding penalty does not recognize the capacity advantage of rail TU, and therefore the breakeven points significantly move toward higher demand levels. A similar effect can be obtained by disabling the newly introduced correction for the average vehicle occupancy by setting $\phi = 1$ as done in Variant D.

Variant E reduces by 5% the in-vehicle VoT for rail modes with respect to the base case while it keeps the in-vehicle VoT for bus modes as in the base case. These VoT variations attempt to represent a rail ride quality advantage in terms of both

lateral and vertical accelerations. Even this modest in-vehicle VoT reduction for rail yields significantly lower breakeven points.

We now present results obtained by varying α, the parameter referring to the degree of demand concentration along the line. Variant F sets a low value for this parameter, $\alpha = 0.2$, which expresses a very low demand concentration. Breakeven points do not change significantly with respect to our base case, but we note that bus technologies can serve higher PTDs than the base case before reaching maximum capacity. Variant G posits a high value of α, equal to 0.6, which indicates an unusual concentration of traffic in a relatively short part of the line. Breakeven points shift toward lower demand levels with respect to our base case because the large capacity of rail TU show its usefulness under this variant.

We explored the dependence to a project discount rate, which had a default value of 3%. The defensible range for this assumption can be considerable, ranging from a "fiscal" discount rate reflecting only the cost for the project sponsor of borrowing money to a "social" discount rate reflecting the urgency of the problem or competing demands for funds, more typical of developing countries. Table 8 reports two variants, H and I, with a discount rate of 6 and 9%, respectively. Further experiments on the discount rate for Scenario 2 are synthesized in Fig. 22. These results indicate that the breakeven points are strongly related to the discount rate.

Variant C+G presents a best case for bus technologies, where crowding is not considered, $\rho = 0$, and the demand concentration along the line is low, $\alpha = 0.2$. An opposite combined scenario variant is B+F where there is VoT equality among the three passenger time components and a high demand concentration along the line, $\alpha = 0.6$. These opposite scenario variants illustrate how particular and legitimate assumptions can yield very different results.

We have tested a wide interval for the parameter λ. Breakeven points presented in terms of PTD do not change significantly for $\lambda \in [0.35, 0.6]$, whereas the breakeven points expressed as average hourly peak demand would vary misleadingly. This

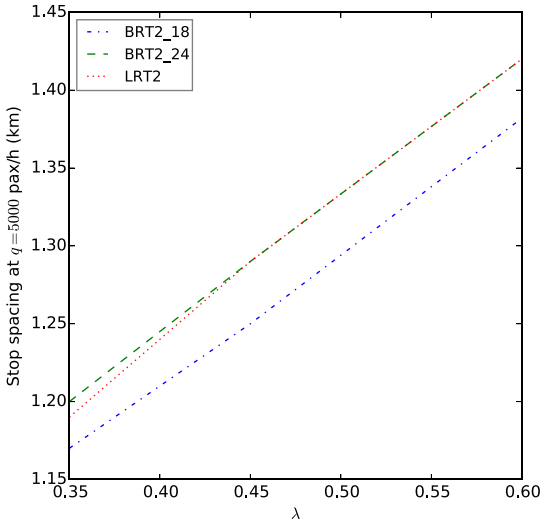

Fig. 19 Scenario 2, stop spacing as a function of λ at $q = 5000$ pax/h

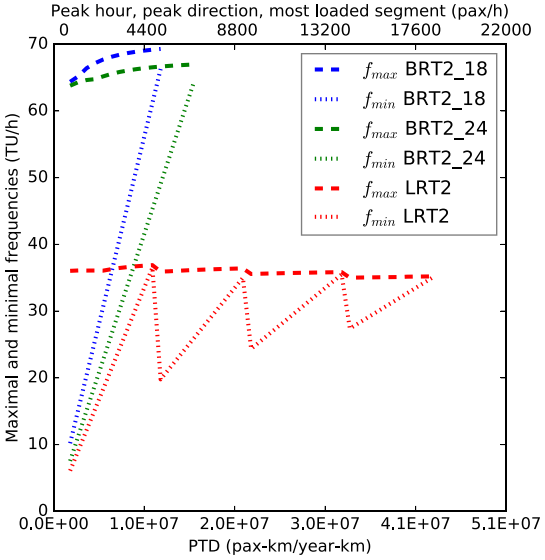

Fig. 20 Scenario 2, maximum and minimum frequencies with $\nu = 0.95$

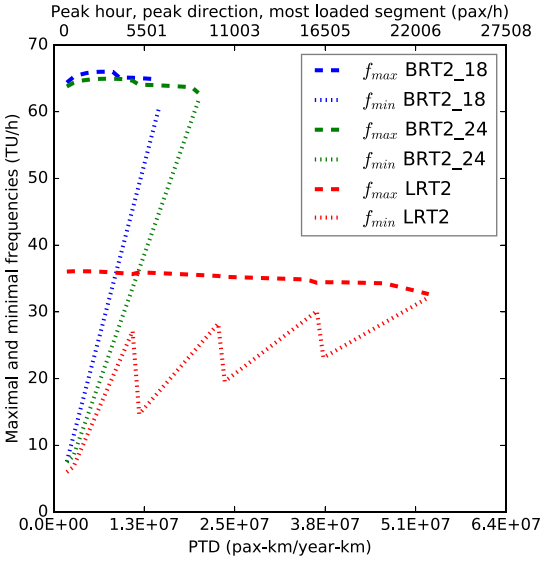

Fig. 21 Scenario 2, maximum and minimum frequencies with $\nu = 1.28$

highlights the advantage of the PTD index for the demand level when comparing technologies. As discussed in Sect. 5.1, λ does have an impact on the stop spacings. As an example, we report in Fig. 19 the stop spacings in Scenario 2 at $q = 5000$ pax/h.

We now explore the dependence of the model results to maximum vehicle loading. Such loading can be represented by setting the spare capacity factor ν larger than one. In the base case we have $\nu = 0.95$ indicating a safety margin of 5%. In our vehicle capacity parameter set, $\nu = 1.28$ allows loads up to six standees per

square meter, commonly referred to as a "crush" load. Allowing this crowding level does not affect the breakeven points in our two base scenarios. This crowding level increases capacities but by less than the $1.28/0.95 = 1.35$ factor that one would expect. This occurs because the higher load ratios induce longer dwell times, and hence smaller maximum frequencies due lo longer operating times. For example, compare the maximum and minimum frequencies in the Scenario 2 with $\nu = 0.95$, Fig. 20, and those with $\nu = 1.28$, Fig. 21.

6 Conclusions

We believe that the results of the model to date support the following conclusions. The model is validated by a good alignment of its results with performance indices of best practices in North America, and to a lesser extent Europe and Australia. For a particular corridor and specific cost conditions, a breakeven point between rail and bus technologies can be identified from both an operator perspective and a total (operator and user) perspective. However, we note that there is a wide range where the cost differences between the technologies are small, and therefore other factors beside passengers' time value and operator cost, both capital and O&M, could be decisive. It may be concluded that faster, although more capital-intensive, best commonly available technologies, as represented in our choice of coefficients and parameters, dominate the slower ones with the only exception of the multiple-unit rail vs. BRT at the lowest demand density studied. Planning for a faster technology is then more relevant than the choice between bus and rail per se, except at very low demand density. At high demand density multiple-unit rail as in Scenario 2 offers the most cost-effective way to achieve high capacities (in the TCQSM sense) under many conditions. A scenario variation analysis shows that differences between waiting and in-vehicle VoTs are more relevant to the technology choice than the exact estimate of the VoT. Assessing the crowding disutility by reducing the bias of averaging vehicle load ratios is found to be significant as well. Even a small reduction of the rail in-vehicle VoT with respect to that of bus, reflecting a difference in riding comfort, yields lower breakeven points than in the conservative scenarios were in-vehicle VoTs are equal for all modes. The choice of a density demand index allows a degree of abstraction with respect to a straightforward comparison by average peak demand level. Allowing crush loading does not modify breakeven points and extends capacities in a sub-linear way. Breakeven points are very sensitive to the project discount rate. Thus, developing countries will tend to have a higher breakeven point for choosing rail, as the added capital costs of rail do not offset labor costs as quickly.

With respect to the impact of the new model's features, the computational experiments show the following. The reduction of the crowding disutility underestimation has a significant effect on the breakeven points between bus and rail technologies that otherwise would move toward higher demand levels, because the full comfort advantage of higher capacity rail TU would not be recognized. The high frequency penalty, the reduction of the waiting advantage in case of

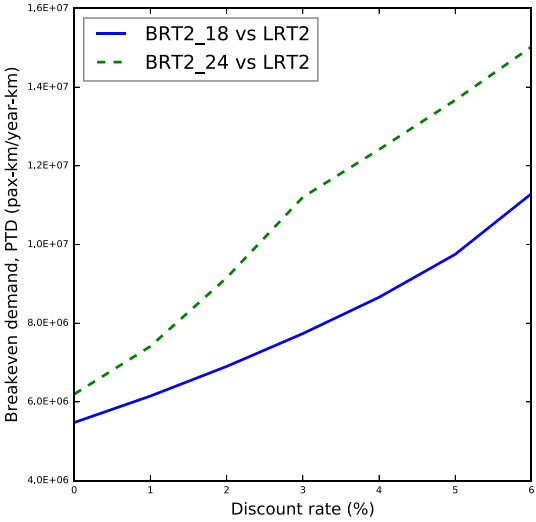

Fig. 22 Breakeven demand as a function of the discount rate for Scenario 2

platooning, and the maximum frequency constraint do not impact the breakeven points, hence the technology selection issue, but are instrumental in aligning computed attainable capacities, and TU speeds to those that can be empirically observed in good quality semi-rapid transit lines. These results are relevant from a design perspective, and attain our research objective of functional extensions that increase the model's realism. We also have developed and offer techno-economic parameters that reflect the types of construction characteristic of the two scenarios we present. As discussed in Bruun et al. (2018) these costs can exhibit a significant range even within a class of construction because of local circumstances including terrain and availability of right-of-way. These parameters represent an empirical mid-point of the observed range of basic infrastructure elements within the applicable classes of construction. With respect to innovation management in transit, our model can be used to sharpen the value proposition for the users and the operator in broader assessment frameworks as that of Newman et al. (2018) without limiting the technology spectrum to rail.

We observe that we have compared technologies that may have significantly different pollution profiles both in terms of local air quality and greenhouse gas emissions, noise levels, visual aesthetics, and degree of local manufacturing content, and that the model does not represent these aspects. Specific planning requirements, for example gradients or a transit section in a pedestrian street, can tilt the choice between modes. See Vuchic et al. (2012) for an extensive discussion of several transit planning issues, and Vuchic (1984) on the limits of total cost appraisal in transportation. Moreover, total cost minimization does not address equity concerns (Perugia et al. 2011; Gutiérrez-Jarpa et al. 2017). Future work may address these particulars.

Acknowledgements This work was partly supported by CNR (Italy) under project "Smart data and models". This support is gratefully acknowledged. Luigi Moccia and Eric Bruun also show their appreciation

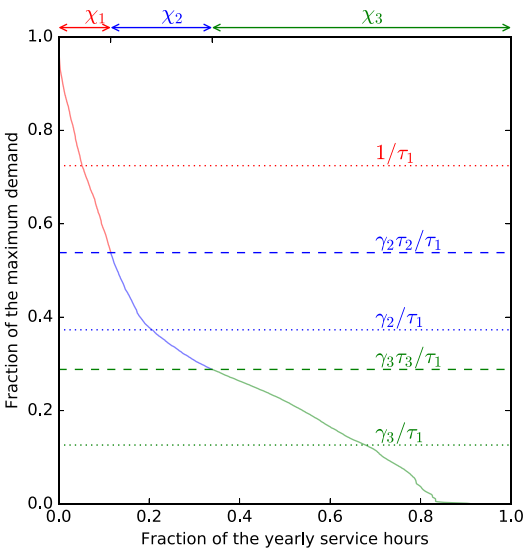

Fig. 23 An example of the disaggregation of a demand profile in three periods

to IBI Group for permitting Duncan Allen to share his dataset and expertise. Thanks are also due to the reviewers for their valuable comments.

Appendix 1: Demand profile

Figure 23 shows an example of a demand profile scaled to the peak maximum demand and subdivided in three periods. This demand profile is derived from 1 month of turnstile data of the Boston subway network. These data are extended to 1 year of service applying average month-specific corrective factors retrieved at MBTA (2017). In this type of chart, the average peak demand is equal to $1/\tau_1$, and the other parameters (γ, τ, and χ) can be derived by a graphical inspection as explained in the following. The vertical axis reports the demand levels as fractions with respect to the maximum demand in a year. The bottom horizontal axis shows the fraction of service hours in a year, and the top horizontal axis indicates the fraction of the period service hours with respect to the total service hours, i.e. the parameters χ. The thick line indicates the demand fractions, and different shades of gray (different colors in the pdf version) represent the different periods. The remaining horizontal axes illustrate the average and maximum demand fractions for each period. The positions of these axes determine the parameters γ and τ.

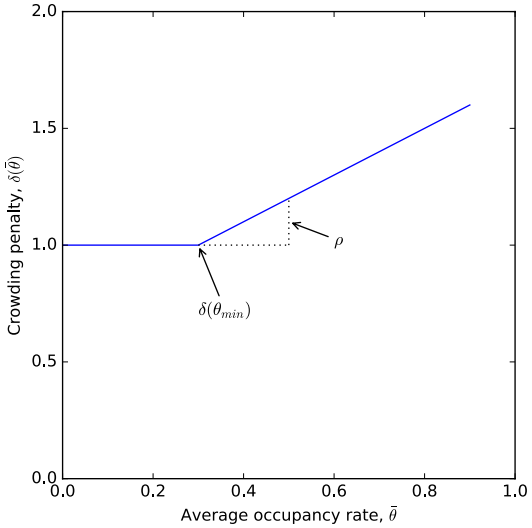

Fig. 24 Example of the crowding penalty function of Moccia and Laporte (2016)

Appendix 2: Approximating the crowding penalty under non-uniform vehicle occupancy rate

As discussed in the literature review, there is a wide consensus that the value of the in-vehicle travel time should be multiplied by a crowding penalty function. Moccia and Laporte (2016), as well as previous studies, use a piecewise linear function δ with respect to the average vehicle occupancy rate $\bar{\theta}$. By such a function, there is no penalty up to an average occupancy rate θ_{min} that indicates the filling of available seats. For larger values of $\bar{\theta}$ the penalty increases linearly with a slope value ρ, see Fig. 24 for an example.

Formally, this penalty function is

$$\delta(\bar{\theta}) = \begin{cases} 1 + \rho\left(\bar{\theta} - \theta_{min}\right) & \bar{\theta} \geq \theta_{min} \\ 1 & \text{otherwise} \end{cases}. \tag{28}$$

A *caveat* of this approach is that crowding is underestimated when the vehicle occupancy rate significantly varies along the cycle time. This underestimation derives from Jensen's inequality. In the following we indicate by t the time step belonging to the cycle time, and we consider for notational brevity a cycle time of unitary length, i.e. $t \in [0, 1]$. Assuming that the piecewise linear function δ accurately describes users' preferences with respect to the instantaneous vehicle occupancy rate θ, then, because of the convexity of δ, we have the following Jensen's inequality:

$$\delta(\bar{\theta}) = \delta\left(\int_0^1 \theta(t)\mathrm{d}t\right) \leq \int_0^1 \delta(\theta(t))\mathrm{d}t = \Delta, \tag{29}$$

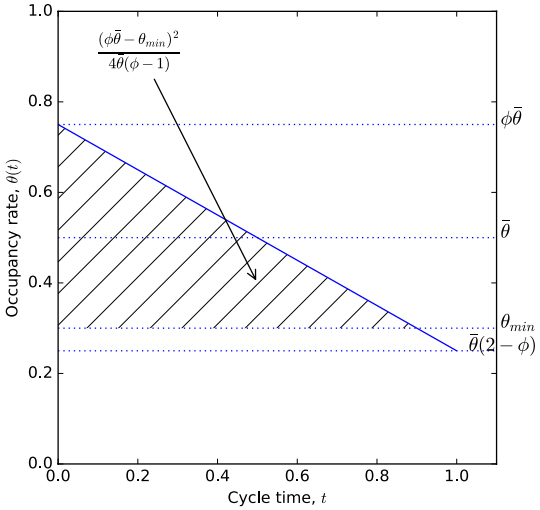

Fig. 25 Example of an occupancy rate function such that $\theta_{min} \geq \bar{\theta}(2-\phi)$. The area corresponding to the users' time under crowding is depicted as hatched and can be computed by the formula reported in the upper left corner of the figure

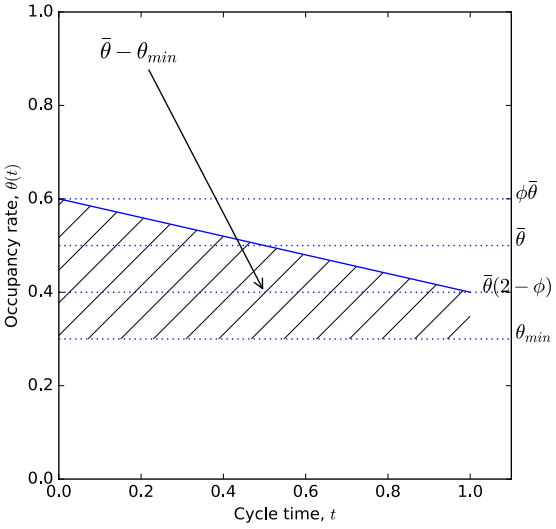

Fig. 26 Example of an occupancy rate function such that $\theta_{min} \leq \bar{\theta}(2-\phi)$. The area corresponding to the users' time under crowding is depicted as hatched and can be computed by the formula reported in the upper left corner of the figure. This formula is the same as in the case of a constant occupancy rate

where Δ is the penalty function that we want to estimate. Observe that $\delta(\bar{\theta}) = \Delta$ whenever $\theta(t)$ is a constant. In this case we have that $\delta(\bar{\theta})$ exactly gauges the users' time at an occupancy rate larger than θ_{min}, if any.

For the purpose of this paper, strategic appraisal, we want a synthetic representation of non-uniform occupancy rate $\theta(t)$. We assume that the occupancy rate can be approximated by a linear function. Because the users' time in the unitary cycle time is an invariant, i.e. $\int_0^1 \theta(t)\mathrm{d}t = \bar{\theta}$, such a linear function can be described by only one parameter, ϕ, which indicates the ratio of the maximum to the average occupancy rate. Observe that for any practical use ϕ is constrained in the interval [1, 2], because values larger than two would imply that a fraction of the cycle time occurs with zero

occupancy rate. Simple geometric analyses (see Figs. 25, 26 for the two relevant cases) allow to define the function Δ for the above defined occupancy rate:

$$\Delta(\bar{\theta}) = \begin{cases} 1 + \rho\left(\bar{\theta} - \theta_{min}\right) & \theta_{min} \leq \bar{\theta}(2-\phi) \\ 1 + \rho\dfrac{\left(\phi\bar{\theta} - \theta_{min}\right)^2}{4\bar{\theta}(\phi - 1)} & \bar{\theta}(2-\phi) < \theta_{min} < \phi\bar{\theta} \\ 1 & \theta_{min} \geq \phi\bar{\theta} \end{cases} \cdot \tag{30}$$

Appendix 3: Lower convex approximation scheme

We construct a separable lower convex envelope of the objective function C_{tot} in the feasible frequency range as follows. First, we use the minimum feasible value of the crowding penalty (13). Second, we assume a discounted waiting time by the factor μ whenever the minimum frequency is smaller than f_l. Third, we compute the intersection delay in the cycle time at the minimum frequency. Fourth, we remove non-convex terms by fixing some variables to proper bounds.

Thus, we can define a separable lower convex envelope of the objective function for a given vector of consists $\bar{\mathbf{n}}$ as

$$\tilde{C}_{tot}(\mathbf{f}, d, \bar{\mathbf{n}}) = a_0 + a_1 d + \frac{a_2}{d} + \sum_p \left(a_{3,p} + \frac{a_{5,p}}{d_{max}}\right) f_p + \sum_p \frac{a_{4,p}}{f_p}. \tag{31}$$

The coefficients of the previous equation are defined as follows

$$\begin{aligned} a_0 = {} & c_{0l} + c_{1v}\beta\zeta\frac{t_b q}{3600} + c_{1t}\beta\frac{t_b q}{3600}\sum_p \frac{\gamma_p \chi_p}{\bar{n}_p} \\ & + V_v \frac{\lambda}{2} q \sum_p \chi_p \gamma_p \bar{\Delta}_p (R + t_{\omega,p}) + c_{1v}\beta\zeta f_{1,min}\bar{n}_1 t_{\omega,1} \\ & + c_{1t}\beta \sum_p \chi_p f_{min,p} t_{\omega,p} \end{aligned} \tag{32}$$

$$a_1 = \frac{q V_a}{2s} \sum_p \gamma_p \chi_p \tag{33}$$

$$a_2 = 2L(c_{0s} + c_{0sv}(\bar{n}_1 - 1)) + T_l V_v \frac{\lambda}{2} q \sum_p \chi_p \gamma_p \bar{\Delta}_p \tag{34}$$

$$a_{3,p} = c_{1t} \chi_p R'_p + 2c_{2v} L \chi_p \bar{n}_p + {}_{\text{if } p=1}\, c_{1v} \beta \zeta \bar{n}_1 R'_1 \tag{35}$$

$$a_{4,p} = V_w q \gamma_p \chi_p \epsilon \mu_p + V_v l (q\gamma_p)^2 \chi_p \bar{\Delta}_p \frac{t_b}{3600 \bar{n}_p} \tag{36}$$

$$a_{5,p} = 2c_{1t} \beta L T_l \chi_p + {}_{\text{if } p=1}\, 2c_{1v} \beta \zeta \bar{n}_1 L T_l, \tag{37}$$

where, for notational compactness, we have used the following notation:

$$R'_p = \beta R + \frac{t_{tf} + t_{tv} \bar{n}_p}{3600} \tag{38}$$

$$\mu_p = \begin{cases} \mu \text{ if } f_{min,p} < f_l \\ 1 \text{ if } f_{min,p} \geq f_l \end{cases} \tag{39}$$

$$\bar{\Delta}_p = \Delta(f_{max,p}, \bar{n}_p) \tag{40}$$

$$t_{\omega,p} = \omega_1 \frac{2t_u L}{60} \left(\frac{f_{min,p}}{\dot{f}} \right)^{\omega_2}. \tag{41}$$

By calculus, as done in Moccia and Laporte (2016), the unconstrained optimal stop spacing of the approximation $\tilde{d}_{unc}$ is

$$\tilde{d}_{unc} = \sqrt{\frac{2s(2L(c_{0s} + c_{0sv}(\bar{n}_1 - 1)) + T_l V_v \frac{\lambda}{2} q \sum_p \chi_p \gamma_p \bar{\Delta}_p)}{q V_a \sum_p \gamma_p \chi_p}}. \tag{42}$$

The unconstrained optimal frequencies $\tilde{f}_{unc,p}$ are

$$\tilde{f}_{unc,p} = \sqrt{\frac{a_{4,p}}{a_{3,p} + \frac{a_{5,p}}{d_{max}}}}. \tag{43}$$

References

Akçelik R, Rouphail NM (1993) Estimation of delays at traffic signals for variable demand conditions. Transp Res Part B Methodol 27(2):109–131

Bartholdi JJ, Eisenstein DD (2012) A self-coördinating bus route to resist bus bunching. Transp Res Part B Methodol 46(4):481–491

Bruun EC, Allen DW, Givoni M (2018) Choosing the right public transport solution based on performance of components. Transport (Forthcoming)

Casello JM, Lewis GM, Yeung K, Santiago-Rodríguez D (2014) A transit technology selection model. J Public Transp 17(4):50–75

Daganzo CF (2009) A headway-based approach to eliminate bus bunching: systematic analysis and comparisons. Transp Res Part B Methodol 43(10):913–921

Daganzo CF (2012) On the design of public infrastructure systems with elastic demand. Transp Res Part B Methodol 46(9):1288–1293

Fernandez R, Planzer R (2002) On the capacity of bus transit systems. Transp Rev 22(3):267–293

Gutiérrez-Jarpa G, Laporte G, Marianov V, Moccia L (2017) Multi-objective rapid transit network design with modal competition: the case of Concepción, Chile. Comput Oper Res 78:27–43

Haywood L, Koning M, Monchambert G (2017) Crowding in public transport: who cares and why? Transp Res Part A Policy Pract 100:215–227

Hörcher D, Graham DJ, Anderson RJ (2017) Crowding cost estimation with large scale smart card and vehicle location data. Transp Res Part B Methodol 95:105–125

Khreis H, Warsow KM, Verlinghieri E, Guzman A, Pellecuer L, Ferreira A, Jones I, Heinen E, Rojas-Rueda D, Mueller N, Schepers P, Lucas K, Nieuwenhuijsen M (2016) The health impacts of traffic-related exposures in urban areas: understanding real effects, underlying driving forces and co-producing future directions. J Transp Health 3(3):249–267

Kikuchi S, Vuchic VR (1982) Transit vehicle stopping regimes and spacings. Transp Sci 16(3):311–331

Klumpenhouwer W, Wirasinghe SC (2016) Cost-of-crowding model for light rail train and platform length. Public Transp 8(1):85–101

Li Z, Hensher DA (2013) Crowding in public transport: a review of objective and subjective measures. J Public Transp 16(2):107–134

Litman T (2017) Valuing transit service quality improvements. Technical report, Victoria Transport Policy Institute

MBTA (2017). http://www.mbtabackontrack.com. Accessed 1 June 2017

Moccia L, Allen DW, Bruun EC (2016) New results of a technology choice model for a transit corridor. In: European Transport Conference 2016. Association for European Transport (AET)

Moccia L, Giallombardo G, Laporte G (2017) Models for technology choice in a transit corridor with elastic demand. Transp Res Part B Methodol 104:733–756

Moccia L, Laporte G (2016) Improved models for technology choice in a transit corridor with fixed demand. Transp Res Part B Methodol 83:245–270

Newman P, Davies-Slate S, Jones E (2018) The entrepreneur rail model: funding urban rail through majority private investment in urban regeneration. Res Transp Econ 67:19–28

Van Oort N (2016) Incorporating enhanced service reliability of public transport in cost-benefit analyses. Public Transp 8(1):143–160

Perugia A, Cordeau J-F, Laporte G, Moccia L (2011) Designing a home-to-work bus service in a metropolitan area. Transp Res Part B Methodol 45(10):1710–1726

TCQSM (2013) Transit capacity and quality of service manual. Transportation Research Board, Washington

Tirachini A, Hensher DA (2011) Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. Transp Res Part B Methodol 45(5):828–844

Tirachini A, Hensher DA, Jara-Díaz SR (2010) Restating modal investment priority with an improved model for public transport analysis. Transp Res Part E Logist Transp Rev 46(6):1148–1168

Tirachini A, Sun L, Erath A, Chakirov A (2016) Valuation of sitting and standing in metro trains using revealed preferences. Transp Policy 47:94–104

US-DOT (2011) The value of travel time savings: departmental guidance for conducting economic evaluations. Technical report, Department of Transportation, US

Verbich D, Diab E, El-Geneidy A (2016) Have they bunched yet? An exploratory study of the impacts of bus bunching on dwell and running times. Public Transp 8(2):225–242

Vuchic VR (1969) Rapid transit interstation spacings for maximum number of passengers. Transp Sci 3(3):214–232
Vuchic VR (1984) The auto versus transit controversy: toward a rational synthesis for urban transportation policy. Transp Res Part A 18(2):125–133
Vuchic VR (2005) Urban transit: operations, planning, and economics. Wiley, Hoboken
Vuchic VR (2007) Urban transit systems and technology, chapter 4 transit system performance: capacity, productivity, efficiency, and utilization. Wiley, Oxford, pp 149–201
Vuchic VR, Newell GF (1968) Rapid transit interstation spacings for minimum travel time. Transp Sci 2(4):303–339
Vuchic VR, Stanger RM, Bruun EC (2012) Bus rapid transit (BRT) versus light rail transit (LRT): service quality, economic, environmental and planning aspects. Transp Technol Sustain. Springer, Berlin, pp 256–291
Wardman M, Whelan G (2010) Twenty years of rail crowding valuation studies: evidence and lessons from British experience. Transp Rev 31(3):379–398