



Comparison of dynamic control strategies for transit operations



Juan Carlos Muñoz^{a,*}, Cristián E. Cortés^b, Ricardo Giesen^a, Doris Sáez^c, Felipe Delgado^a, Francisco Valencia^b, Aldo Cipriano^d

^a Department of Transport Engineering and Logistics, Pontificia Universidad Católica de Chile, Chile

^b Civil Engineering Department, Universidad de Chile, Chile

^c Electrical Engineering Department, Universidad de Chile, Chile

^d Department of Electrical Engineering, Pontificia Universidad Católica de Chile, Chile

ARTICLE INFO

Article history:

Received 3 October 2011

Received in revised form 13 November 2012

Accepted 18 December 2012

Keywords:

Transit systems

Transit control

Performance comparison

Vehicle bunching

ABSTRACT

Real-time headway-based control is a key issue to reduce bus bunching in high frequency urban bus services where schedules are difficult to implement. Several mechanisms have been proposed in the literature, but very few performance comparisons are available. In this paper two different approaches are tested over eight different scenarios. Both methodologies solve the same problem, the former based on a deterministic optimization over a long-term rolling horizon, while the latter proposes a hybrid predictive approach considering a shorter horizon and a stochastic evolution of the system. The comparison is conducted through scenarios that include three different dimensions: (i) bus capacities which can be reached or not, (ii) service frequencies, considering high and medium frequency services and (iii) different load profiles along the corridor. The results show that the deterministic approach performs better under scenarios where bus capacity could be reached frequently along the route while the hybrid predictive control approach performs better in situations where this does not happen.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

The absence of a control system in a bus transit network tends to result in vehicle bunching due to the stochastic nature of traffic flows and passenger demand at bus stops. It also leads to an evident increase in bus headway variance and a consequent worsening of both the magnitude and variability of average waiting times. This in turn impacts heavily on the level of service as perceived by users given that their subjective valuation of this component of total trip time is higher than that of any other (access time, in-vehicle time) (Boardman et al., 2001).

In this study we address the problem of determining a bus control strategy for the various stops or stations in a transit system that will minimize the total time users must devote to making a trip. Historically, the literature in this field has evolved from the study of pre-planned fleet assignment and scheduling strategies, to the analysis of real-time control strategies, assuming that real-time information is available through on-vehicle equipment such as Automatic Passenger Counters (APCs) and Automatic Vehicle Location (AVL) devices. The first group of strategies works as a complement to a properly pre-planned bus schedule, since they are defined to deal with well-known demand imbalances at an aggregated level, in specific route sections and periods (for a deep description of these strategies, such as short turning, restricted zonal service and deadheading, among others, see Furth and Day, 1984). The second group of strategies has been designed to allow the operator to react dynamically to real-time system disturbances. Due to its simplicity and effectiveness, the most studied strategy of this type in recent years is the holding strategy, in which vehicles are held at specific stations for a certain time, in most cases oriented to keep the headway between successive buses as close as possible to a predefined value.

* Corresponding author. Tel.: +56 22 3547116; fax: +56 2 25530281.

E-mail address: jcm@ing.puc.cl (J.C. Muñoz).

In terms of the spatial configuration of the different control strategies, Eberlein (1995) classified them into three categories: (i) station control, including bus holding and stop skipping; (ii) interstation control, such as operating speed control, bus overtaking and traffic signal priority mechanisms; and (iii) other control measures such as adding vehicles.

This study will focus on strategies for vehicle holding that determine which buses are to be held where and for how long.

There are two types of bus holding, one based on schedules and the other on headways. The first type is used when the system has a predefined schedule, which is typical of services with long headways. The second type, on the other hand, is employed in the absence of a predefined schedule. This is commonly the case for services with short headways (e.g. less than 10 min).

As regards headway-based holding, the type that will concern us here, Barnett (1974) early research developed a simple holding model at a given control station, where the sum of total waiting time plus the extra delay of passengers on board deadheaded vehicles is minimized. Turnquist and Blume (1980) identified conditions under which holding results are attractive. The appearance of new information and communication technologies made possible the development of more complex models that determine the optimal holding time for each vehicle so as to minimize total passenger waiting time at stops, or a combination of this factor and in-vehicle delay of passengers due to holding. Fu and Yang (2002) present a formulation along these lines that minimizes average waiting time at stops but ignores both in-vehicle holding delay and bus capacity constraints. In such a case the results demonstrate that the optimal policy is to apply holding at a single stop. Sun and Hickman (2004) state the same problem in two dimensions, with holding of multiple buses but at a defined set of control stations. They show that if holding is applied at various stations, bus headways can be regularized and greater cost reductions thereby achieved compared to single station holding. The study of Hickman (2001) presented a stochastic holding model at a given control station. The author formulated a convex quadratic program in a single variable corresponding to the time lapse during which buses are held. More recent research has explored holding models that rely on real-time information, mainly referring to vehicle location (Eberlein, 1995; Eberlein et al., 2001). Zhao et al. (2003) propose a control model based on negotiation between two agents, one aboard the bus and the other at a stop. Simulation results indicate that the negotiation algorithm is robust for a variety of operating conditions. The vehicle capacity constraint approach is addressed in Zolfaghari et al. (2004), who formulate a problem in which the objective function minimizes the waiting times both of users who arrive at a stop and those who have to await more than one bus due to the activation of the capacity constraint. They do not, however, consider the extra waiting time endured by passengers held at a stop. Puong and Wilson (2004) extend this case by including the latter factor in their objective function in the context of interruptions in train service. They propose a non-linear mixed-integer model in which dwell time is assumed to be constant at any given station. The problem is solved in a reasonable amount of time using a branch-and-cut strategy. Delgado et al. (2012) address this problem using a model that considers holding at each stop and a capacity constrained heterogeneous fleet without using integer variables.

Yu and Yang (2007) present a dynamic holding strategy, in which the on-time performance of the early bus operation at the next stop is considered and the holding times of the held bus at the stop is optimized. A model based on support vector machine (SVM) for forecasting the early bus departure times from the next stop is also developed. Furthermore, in order to determine the optimal holding times, a model aiming to minimize the total user costs is developed. Genetic algorithms are proposed to optimize holding times.

Lately, an adaptive control scheme aiming to provide quasi-regular headways while maintaining as high a commercial speed as possible is proposed by Daganzo (2009) and Daganzo and Pilachowski (2011). In Daganzo (2009) the control dynamically determines bus holding times at a route's control points based on real-time information about the passage of the previous bus. Although the method proves to be efficient under small disturbances, under large perturbations it reduces performance. To overcome this problem, Daganzo and Pilachowski (2011) continuously adjust bus cruising speed based on a cooperative two way looking strategy based on the spacing of the front and back buses. The cooperative control shows to be effective in preventing bunching. Since in both cases they aim at maintaining regularity, their proposed approach is not well suited for operations in which vehicles reach their capacities. The cooperative control shows to be effective in preventing bunching. Bartholdi and Eisenstein (2012) abandon the idea of any a priori target headway, allowing the natural headway of the system to spontaneously emerge when disturbances appear by implementing a simple holding rule at a control point.

As for strategies to increase bus operating speed, one of the most studied is stop-skipping Suh et al. (2002), Fu et al. (2003), Sun and Hickman (2005). Sáez et al. (2012) present a hybrid predictive controller in which holding and stop-skipping control actions are taken each time a bus arrives at a stop. The problem is solved using genetic algorithms in which holding time may only take some discrete values. On a similar combination of strategies, Delgado et al. (2009) present a continuous version of stop-skipping in which a certain fraction of passengers is prevented from boarding at each stop.

The major focus of this paper is to compare two different approaches for optimizing the real-time control for the operation of a bus system on a single-service corridor. One methodology based on Delgado et al. (2012) takes each single decision by solving a deterministic long-term prediction model while the other, based on Sáez et al. (2012), utilizes a hybrid predictive control methodology on a shorter horizon based on a stochastic evolution of the system conditions. Both approaches are embedded on a rolling horizon framework. Although both methodologies have been applied with a wider variety of control strategies, such as holding, station skipping, overtaking, and boarding limits, for comparison purposes, in this paper we focus only on holding, which is the most commonly implemented control strategy and widely studied in the literature. This comparison should allow us to detect the scenarios under which each of the two methods present advantages, highlighting their respective strengths for further construction of an integrated approach. We have made an effort in this document to present both approaches with a common nomenclature.

In Section 2 the problem formulation is described. In Sections 3 and 4, the deterministic and hybrid predictive control strategies are described respectively. Next, experiment simulations are shown in Section 5. Finally Section 6 presents the conclusion and further research.

2. Problem formulation

The system underlying our model is a high passenger demand one-way loop transit corridor with P stops operated by a single high-frequency service consisting of b homogeneous buses, with a limited capacity denoted by L_{max} (see Fig. 1). Vehicles start their run at a terminal defined as Stop 1, visiting all stops downstream (2, 3, ..., P) before returning to the same terminal ($P + 1$) where all remaining passengers must alight. The buses are numbered in strict order of advance along the corridor, bus 1 being furthest ahead and b furthest behind.

The proposed model incorporates the following assumptions:

- A homogeneous fleet of buses.
- Buses serve all stops, and overtaking is not permitted.
- Passenger arrival rates for each stop and travel times between stops are deterministic, known and fixed over the period of interest.
- Boarding time dominates alighting time at most stops. An estimate of boarding time based on predicted boarding passengers will therefore be used for dwell time. In all those stops where alighting time dominates boarding time, dwell time will be underestimated in both models when compared with the simulation.

3. Deterministic control strategy (DC)

The problem of vehicle holding with real time information assumed that at any time instant we have real-time information on the position of every bus and an estimation of the number of passengers aboard each bus as well as the number of passengers waiting at the various stops. The system is then completely determined by the following state variables:

- d_i distance between bus i and the last stop upstream from it (meters). (by “upstream” is meant a stop the bus has already visited).
- e_i stop immediately upstream from bus i . If bus i is at stop p then $e_i = p - 1$.
- B_{ip} number of passengers on bus i who boarded at stop p .
- I_p number of passengers waiting at stop p .

Regarding passenger demand, we assume, for each stop, that the average passenger arrival rate and the expected proportions vector that distributes the trips originating at that stop among all downstream ones by their respective destinations are known. Note that these two factors are specified separately for each stop that originates trips.

The following additional variables and parameters are used in the model:

- i index of buses, $i = 1, \dots, b$.
- p index of stops, $p = 1, \dots, P + 1$.
- t_k current time, instant when the control decision needs to be made.
- θ_z weight factors included in the objective function ($z = 1, 2, 3, 4$).
- L_{max} bus passenger capacity.
- λ_p average passenger arrival rate at stop p (passengers per minute).
- t_b passenger boarding time (minutes per passenger).
- r_p distance between stops p and $p + 1$ (meters).
- v_p bus operating speed between stops p and $p + 1$ while the bus is moving (meters per minute).
- Tv_{ip} travel time between stations p and $p + 1$ (minutes) ($Tv_{ip} = \frac{r_p}{v_p}$).
- \hat{F}_{iqp} expected fraction of passengers boarding bus i at stop q whose destination is stop p ($\forall q < p$).

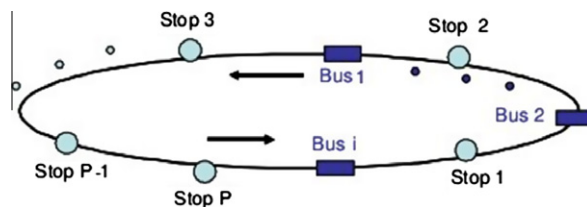


Fig. 1. Transit system model.

The following auxiliary variables regarding the evolution of the system during the rolling horizon can then be estimated:

\widehat{L}_{ip}	total number of passengers traveling in bus i when it reaches stop p .
$\widehat{L}m_{iqp}$	number of passengers boarding at stop q and traveling in bus i when it reaches stop p ($\forall q < p$). Notice that $\widehat{L}m_{iqe_i+1} = B_{ip}$.
\widehat{S}_{ip}	available capacity in bus i when it reaches stop p (number of passengers).
$\widehat{T}d_{ip}$	departure time of bus i from stop p .
\widehat{A}_{ip}	number of passengers who alight bus i at stop p .
$\widehat{B}p_{ip}^0$	number of passengers who board bus i at stop p ($p = e_i + 1 \dots P$).
$\widehat{B}p_{ip}$	number of passengers who board bus i at stop p ($p = 1 \dots e_i$).
\widehat{B}_{ip}	number of passengers who board bus i at stop p .
\widehat{W}_{ip}	number of passengers prevented to board bus i at stop p .
$\widehat{T}r_{ip}$	dwell time of bus i at stop p (minutes).
\widehat{H}_{ip}	length of time measured as the moment bus i departs stop p minus, the largest between t_k and the moment bus $i-1$ departs stop p .

We consider the following single set of decision variables for the control strategy:

h_{ip} holding time of bus i at stop p (minutes).

We now formulate a deterministic mathematical programming problem that determines the holding times of the buses at the various stops along the corridor during the rolling horizon.

The main objective of the controller is to minimize the total travel times of passengers from the moment they arrive at a stop to the moment they reach their destination. Since vehicle running times are assumed to be constant, the objective is to minimize both in-vehicle and at-stop waiting times. These components can be written in the following way:

$$\begin{aligned} & \theta_1 \cdot \sum_i \sum_p \widehat{\Gamma}_p \cdot \frac{\widehat{H}_{ip}}{2} + \theta_2 \cdot \left\{ \sum_{i=1}^b \sum_{p=1}^{P-1} \widehat{L}_{ip+1} \cdot h_{ip} + \sum_{i=1}^b \widehat{L}_{i1} \cdot h_{iP} \right\} + \theta_3 \cdot \left\{ \sum_{i=1}^{b-1} \sum_{p=e_{(i-1)+1}}^{e_i} \widehat{W}_{ip} \cdot \widehat{H}_{ip} + \sum_{p=e_b+1}^{e_1} \widehat{W}_{bp} \cdot \widehat{H}_{bp} \right\} \\ & + \theta_4 \cdot \left\{ \sum_{i=1}^b \sum_{p=1}^{P-1} \widehat{W}_{ip} \cdot \widehat{S}_{ip+1} + \sum_{i=1}^b \widehat{W}_{iP} \cdot \widehat{S}_{i1} \right\} \end{aligned} \quad (1)$$

$$\widehat{\Gamma}_p = \begin{cases} \lambda_p \cdot (\widehat{T}d_{ip} - t_k) + 2\Gamma_p & p = e_i + 1, \dots, e_{(i-1)}; \quad i = 1 \dots b \\ \lambda_p \cdot (\widehat{T}d_{ip} - \widehat{T}d_{i-1p}) & p = e_{(i-1)} + 1, \dots, e_i; \quad i = 2 \dots b \\ \lambda_p \cdot (\widehat{T}d_{1p} - \widehat{T}d_{bp}) & p = e_b + 1, \dots, e_1 \end{cases}$$

$$\widehat{H}_{ip} = \begin{cases} \widehat{T}d_{ip} - t_k & p = e_i + 1, \dots, e_{(i-1)}; \quad i = 1 \dots b \\ \widehat{T}d_{ip} - \widehat{T}d_{i-1p} & p = e_{(i-1)} + 1, \dots, e_i; \quad i = 2 \dots b \\ \widehat{T}d_{1p} - \widehat{T}d_{bp} & p = e_b + 1, \dots, e_1 \end{cases}$$

The first term in (1) refers to the at-stop waiting time (W_{first}) experienced by passengers as they wait for the first bus to arrive after t_k . It distinguishes between the waiting time for passengers at stops where bus i is the first one to arrive, and that for passengers at a stop p that has already been visited by a bus ($i-1$). It also distinguishes between passengers already waiting at the stop from those that have not arrived yet by t_k . The term also lends the objective function its non-linear nature given that total waiting time for all users is proportional to the square of bus headway. As for the second term in (1), it states the in-vehicle waiting time ($W_{\text{in-veh}}$) for passengers aboard a bus i being held at stop p . The third term, indicates the extra waiting time (W_{extra}) of passengers who are prevented from boarding bus i because it is at capacity. Finally, the fourth term represents a penalty (PE) for passengers prevented from boarding when there is still capacity in the bus (boarding limits). The weight associated to this term will be large enough for this never to be allowed. Each of the four terms are multiplied by a different weighting factor, θ_1 , θ_2 , θ_3 and θ_4 .

If we keep the objective function as expressed above (i.e. the total waiting time experienced by all passengers during the planning horizon), we encourage the model to leave more passengers at stops by the end of the planning horizon than those strictly needed to improve performance. This happens because the waiting times experienced by those passengers are not considered in the objective function. We prevent this problem by minimizing instead the average waiting time per passenger in the objective function. Therefore we divide the total waiting time by the total number of passengers involved which is given by:

$$Pax = \sum_{i=1}^b \sum_{p=e_i+1}^{e_{(i-1)}} \{ \lambda_p \cdot (\widehat{T}d_{ip} - t_k) + \Gamma_p \} + \sum_{i=2}^b \sum_{p=e_{(i-1)+1}}^{e_i} \{ \lambda_p \cdot (\widehat{T}d_{ip} - \widehat{T}d_{i-1p}) \} + \sum_{p=e_b+1}^{e_1} \{ \lambda_p \cdot (\widehat{T}d_{1p} - \widehat{T}d_{bp}) \}$$

Thus, we obtain the objective function for the proposed model:

$$\text{Min}_{h_{ip}} \frac{\theta_1 \cdot W_{\text{first}} + \theta_2 \cdot W_{\text{in-veh}} + \theta_3 \cdot W_{\text{extra}} + \theta_4 \cdot PE}{Pax}$$

The constraints are:

$$\widehat{T}d_{ip} = t_k + \frac{r_{p-1} - d_i}{v_{p-1}} + \widehat{T}r_{ip} + h_{ip} \quad \forall i; \quad p = e_i + 1 \tag{2}$$

$$\widehat{T}d_{ip} = \widehat{T}d_{ip-1} + T v_{ip-1} + \widehat{T}r_{ip} + h_{ip} \quad \forall i; \quad p \neq e_i + 1, 1 \tag{3}$$

$$\widehat{T}d_{i1} = \widehat{T}d_{ip} + T v_{ip} + \widehat{T}r_{i1} + h_{i1} \quad \forall i; \tag{4}$$

$$\widehat{L}_{iqp} = \widehat{B}_{iq}^0 \cdot \left(1 - \sum_{j=q+1}^{p-1} \widehat{F}_{ijq} \right) \quad \forall i; \quad p = e_i + 2, \dots, P; \quad q = 1, 2, \dots, p - 2 \tag{5}$$

$$\widehat{L}_{iq1} = \widehat{B}_{iq}^0 \cdot \left(1 - \sum_{j=q+1}^p \widehat{F}_{ijq} \right) \quad \forall i; \quad p = 1, 2, \dots, P \tag{6}$$

$$\widehat{L}_{iqp} = \widehat{B}_{iq} \cdot \left(1 - \sum_{j=q+1}^{p-1} \widehat{F}_{ijq} \right) \quad \forall i; \quad p = 2, \dots, e_i; \quad q = 1, 2, \dots, p - 2 \tag{7}$$

$$\widehat{L}_{ip-1p} = \widehat{B}_{ip-1} \quad \forall i; \quad p = 2, \dots, e_i \tag{8}$$

$$\widehat{L}_{ip} = \sum_{q=1}^{p-1} \widehat{L}_{iqp} \quad \forall i; \quad p = 2, \dots, P \tag{9}$$

$$\widehat{L}_{i1} = \sum_{q=1}^p \widehat{L}_{iq1} \quad \forall i; \tag{10}$$

$$\widehat{s}_{ip} = L_{\text{max}} - \widehat{L}_{ip} \quad \forall i, \quad p \tag{11}$$

$$\widehat{B}p_{ip} = \Gamma_p + \lambda_p \cdot (\widehat{T}d_{ip} - t_k) \quad \forall i; \quad p = e_i + 1, \dots, e_{(i-1)} \tag{12}$$

$$\widehat{B}p_{ip} = \lambda_p \cdot (\widehat{T}d_{ip} - \widehat{T}d_{i-1p}) + \widehat{W}_{i-1p} \quad \forall i \neq 1; \quad p = e_{(i-1)} + 1, \dots, e_i \tag{13}$$

$$\widehat{B}p_{1p} = \lambda_p \cdot (\widehat{T}d_{1p} - \widehat{T}d_{bp}) + \widehat{W}_{bp} \quad \forall i; \quad p = e_i + 1, \dots, e_1 \tag{14}$$

$$\widehat{A}_{ip} = \sum_{q=1}^{p-1} \widehat{B}_{iq}^0 \cdot \widehat{F}_{iqp} \quad \forall i; \quad p = e_i + 1, \dots, P \tag{15}$$

$$\widehat{A}_{ip} = \sum_{q=1}^{p-1} \widehat{B}_{iq} \cdot \widehat{F}_{iqp} \quad \forall i; \quad p = 2, \dots, e_i \tag{16}$$

$$\widehat{A}_{i1} = \sum_{q=1}^{p-1} \widehat{B}_{iq}^0 \cdot \widehat{F}_{iqp} \quad \forall i \tag{17}$$

$$\widehat{W}_{ip} \geq \widehat{B}p_{ip} - \widehat{s}_{ip} - \widehat{A}_{ip} \quad \forall i; \quad p \tag{18}$$

$$\widehat{W}_{ip} \geq 0 \quad \forall i; \quad p \tag{19}$$

$$\widehat{B}_{ip}^0 = \widehat{B}p_{ip} - \widehat{W}_{ip} \quad \forall i; \quad p = e_i + 1, \dots, P \tag{20}$$

$$\widehat{B}_{ip} = \widehat{B}p_{ip} - \widehat{W}_{ip} \quad \forall i; \quad p = 1, \dots, e_i \tag{21}$$

$$\widehat{T}r_{ip} = \widehat{B}_{ip}^0 \cdot t_b \quad \forall i; \quad p = e_i + 1, \dots, P \tag{22}$$

$$\widehat{T}r_{ip} = \widehat{B}_{ip} \cdot t_b \quad \forall i; \quad p = 1, \dots, e_i \tag{23}$$

$$\widehat{T}d_{ip} - \widehat{T}d_{i-1p} \geq 0 \quad \forall i \neq 1; \quad p = e_{(i-1)} + 1, \dots, e_i \tag{24}$$

$$\widehat{T}d_{1p} - \widehat{T}d_{bp} \geq 0 \quad \forall p = e_b + 1, \dots, e_1 \tag{25}$$

$$\widehat{T}d_{i-1p} - \widehat{T}d_{ip} \geq 0 \quad \forall i \neq 1; \quad p = e_i + 1, \dots, e_{(i-1)} \tag{26}$$

$$\widehat{T}d_{bp} - \widehat{T}d_{1p} \geq 0 \quad \forall p = e_i + 1, \dots, e_{(i-1)} \tag{27}$$

Constraints (2)–(4) determine the departure times from downstream stops for each bus, the former associated to the stop immediately downstream from the bus's current location while (3) and (4) to the rest of the stops.

Constraints (5)–(8) establish the number of passengers traveling in bus i before arriving at stop p who boarded at some previous stop q . In (9) and (10) the total number of passengers in bus i before it arrives at stop p is stated simply as the sum of those who boarded at previous stops, while (11) relates the available capacity of a bus before arriving at a stop to the total number aboard the bus and its capacity.

Constraints (12)–(14) establish potential passenger demand at a given stop for a particular bus. Constraint (15)–(17) relates the total number of passengers who alight from a bus at a given stop to the estimated probability each of them gets off at that stop given the stop they boarded at.

Constraints (18) and (19) address the capacity constraint. They indicate that the quantity of passengers prevented from boarding at a given stop p must be equal to or greater than the number prevented from boarding bus i because it was at capacity.

In (20) and (21), the quantity of passengers who are allowed to board bus i at stop p is given as the difference between potential demand and the number prevented from boarding there. In (22) and (23), dwell time for a bus i at a stop p is the sum of boarding times for all passengers allowed to board. And finally, constraints (24)–(27) establish that buses cannot overtake each other.

For a detailed explanation of the constraints of this model and its implementation see Delgado et al. (2012).

The objective function in (1) is quadratic in h_{ip} and not convex, whereas the model's constraints are linear.

4. Hybrid predictive control strategy (HPC)

The objective of this section is to summarize the HPC approach proposed in Sáez et al. (2012) and Cortés et al. (2010) for a real-time bus system optimization. To be efficient in terms of computation time, the HPC framework is written in discrete time. The problem is then formulated as a hybrid predictive system, where events are triggered by specific actions. Unlike traditional HPC formulations written for a fixed step-size, this scheme is based on the occurrence of relevant events (corresponding to the instants at which control actions have to be taken); in this case the formulation results in a variable step-size as a proxy for expected bus arrival times at bus stops.

Specifically, the events are triggered when a bus arrives at a bus stop, which determines a variable time-step. Hereafter, the following variables and parameters are used in the state space model:

t	continuous time.
k	discrete event associated with the arrival of a specific bus to a specific bus stop.
t_k	continuous time at which event k occurs.
$x_i(t)$	position of the bus i at any continuous instant t .
i	index of buses, $i = 1, \dots, b$
$\hat{T}_i(t)$	expected remaining time for the bus to reach the next stop.
$h_i(k)$	holding time of bus i at event k .
$\hat{L}_i(k+1)$	estimated passenger load once the bus departs from its current stop, associated with the bus i that triggered event k .
$\hat{Td}_i(k+1)$	estimated departure time once the bus departs from its current stop, associated with the bus i that triggered event k .
$\hat{B}_i(k)$	expected number of passenger that will board bus i while it is at the stop.
$\hat{A}_i(k)$	estimated number of passenger alighting from bus i at event k .
v_i	bus instantaneous speed as a function of the continuous time.
$\hat{T}r_i(k)$	estimated passenger transference delay (maximum between the boarding and alighting times).
$Tv_i(k)$	travel time between two consecutive stations.
$\Gamma_p(k)$	estimated bus stop load.
p	index of stops, $p = 1, \dots, P+1$
θ_z	weight factors included in the objective function ($z = 1, 2, 3$).
N_p	prediction horizon.
$\hat{H}_i(k+\ell)$	headway of bus i that triggers the event $k+\ell-1$.
\bar{H}	desired headway (set-point).

The analytical expressions for such a dynamic model associated with bus i that triggered event k can be summarized as presented in Sáez et al. (2012) and Cortés et al. (2010):

$$x_i(t) = x_i(t_k) + \int_{t_k}^t v_i(\vartheta) d\vartheta \quad (28)$$

$$\hat{T}_i(t) = t_k + h_i(k) + \hat{T}r_i(k) + Tv_i(k) - t \quad t_k \leq t \leq t_{k+d} \quad (29)$$

$$\widehat{L}_i(k+1) = \begin{cases} \min\{\bar{L}, L_i(k) + \widehat{B}_i(k) - \widehat{A}_i(k)\} & \text{if bus } i \text{ triggered event } k \\ L_i(k) & \text{otherwise} \end{cases} \quad (30)$$

$$\widehat{Td}_i(k+1) = \begin{cases} t_k + h_i(k) + \widehat{T}r_i(k) & \text{if bus } i \text{ triggered event } k \\ Td_i(k) & \text{otherwise} \end{cases} \quad (31)$$

The demand estimator work with statistical analysis of data collected from sensors that should be located at stops and buses. In our approach, these estimations are obtained from data of both a set of previous similar days (offline historical data) and dynamic information occurring the same day (online data).

Based on offline data, we are able to estimate $\widehat{A}_i(k)$ using the most frequent destination patterns from previous days over the same period; then, those estimations are corrected with online destination data obtained from observed preferences from passengers already in the system. $\widehat{B}_i(k)$ is computed based on both the estimated bus stop load $\Gamma_p(k)$ at instant k and the bus capacity; it is estimated considering autoregressive moving average models for the arrival time of passengers at stops.

In this case, we will pursue the minimization of expression (32) next, which comprises three components oriented to the improvement of the passengers' level of service by means of waiting time and penalty due to control actions. In this expression the first two terms are identical as the three first terms in expression (1). Analytically,

$$J = \sum_{\ell=1}^{N_p} \left[\theta_1 \cdot \widehat{\Gamma}(k+\ell) \frac{\widehat{H}_i(k+\ell)}{2} + \theta_2 \cdot \widehat{L}_i(k+\ell)h_i(k+\ell-1) + \theta_3 \cdot \left(\widehat{H}_i(k+\ell) - \bar{H} \right)^2 \right] \Bigg|_{i=i(k+\ell-1)} \quad (32)$$

where $\{u(k), \dots, u(k+Np-1)\}$ is the control–action sequence with $u(k+\ell-1) = h_i(k+\ell-1)$ when bus i triggers event $k+\ell-1$. N_p is the prediction horizon and b is the number of buses in the fleet.

Note that $i = i(k+\ell-1) \in \{1, \dots, b\}$, if we consider that the future event $k+\ell-1$ is triggered by one bus $i(k+\ell-1)$ arriving to a specific station downstream. In expression (32), θ_j are weighting parameters, and have to be tuned depending on the specific problem to be treated and on the physical interpretation of the different components as well. Moreover, these parameters allow the modeller to give different importance to the specific terms of J .

Note that expression (32) depends on the predicted headway between consecutive buses. By using the prediction of the departure time as detailed in (29), it is possible to predict the headway $\widehat{H}_i(k+\ell)$ of bus i that triggers the event $k+\ell-1$, with respect to its precedent bus $i-1$ when it reaches the same stop, which corresponds to event $k+\ell-z_{i-1}$. Analytically:

$$\widehat{H}_i(k+\ell) = \widehat{Td}_i(k+\ell) - \widehat{Td}_{i-1}(k+\ell-z_{i-1}) \quad (33)$$

where $\widehat{Td}_i(k+\ell)$ is associated with the bus i that triggers the event $k+\ell-1$, and $\widehat{Td}_{i-1}(k+\ell-z_{i-1})$ represents the predicted departure time of precedent bus $i-1$ that triggers the event $k+\ell-z_{i-1}-1$, at the same stop. The variable z_{i-1} represents the number of events between the arrival of the previous bus $i-1$ and the bus i , both reaching the same stop.

In expression (32), \bar{H} is designed for servicing the system demand during a certain time period. Normally, the desired headway \bar{H} is related to the design frequency that directly depends on the segment loads, and can be determined as the minimum required for moving the passengers on the most loaded segment along the bus route. The first term in (32) quantifies the total passenger waiting time at stops, and depends on the predicted headway along with the bus-stop load, which at the same time quantifies the level of service. The second term in (32) measures the delay associated with passengers on-board a vehicle when they are held at a control station due to the application of the holding strategy. The last term captures the regularization of bus headways, and that intends to maintain the headway as close as possible to the desired headway.

The predictive model of the public transport system must satisfy some physical and operational constraints.

The first constraint corresponds to the capacity constraint. This is a physical constraint in the sense that the bus cannot transport more passengers than its maximum capacity. We can also apply a service policy by setting such a capacity differently in order to avoid overcrowding.

Both the precedence constraint and the demand consistency are relevant, because every passenger has a specific origin and destination. Precedence constraints avoid passengers getting off before they get on any bus. With regard to the demand, it is assumed that there are no transfer nodes, and therefore, once a passenger is on board a bus, he (she) will alight from the same bus at his (her) destination stop. Also, once a passenger arrives at their destination, he (she) will always get off the bus there (passengers want to minimize their travel time, so we assume that passengers do not stay on buses in loops).

Regarding bus operation, if the model determines a holding action at a certain stop, which is not physically appropriated for such an operation, then the bus just stops during a lapse required for a normal passenger transfer operation.

Each bus is identified by a unique internal label. However, the model allows the indices to be updated when a bus arrives at its next stop, sorted in such a way that bus $i-1$ always precedes bus i . One important issue is that overtaking is allowed in the model as the indices associated with buses (i and $i-1$ for two consecutive buses) are set each time an event occurs and a control action is applied; in such cases the indices are properly updated and sorted.

With regard to the implementation of the HPC methodology, one difficulty inherent to the approach is in theory the amount of information required to test the potential future scenarios in the prediction process.

If the objective is to evaluate the approach on a realistic transit system, with many buses and bus stops, the chance that a near future event will happen close to a specific stop is low (for example, the one where the current event occurs); therefore, if we predict to n steps ahead, the future events really affected by the current decision will be very few, reducing the predictive power of the approach. The premise is that the events analyzed in the future will most likely be distributed all along the system which prevents us from focusing our analysis to the real impact of the current decision at the given stop. This problem could be avoided by increasing the future steps n considerably to make sure that enough events that are really affected by our decision are included; however this approach is not computationally feasible for real implementations.

To avoid the previous problem, we implemented a local analysis of the impact of the control actions to be compared, avoiding a detailed picture of actions that will occur far from the current decisions to be made. Thus, the controller must only predict the events that happen in a vicinity of bus stops around the one where the current event occurs. As a consequence, we dissociate a future step for the controller from a future step in the simulator. Then, between the occurrences of two events within the vicinity, many more events occur outside its limits, which are not really significant in the selection of the best action to be applied. The conditions of the buses in the future outside the vicinity are approximated to average values, mainly to avoid losing track of the position of the buses in the future due to this approximation.

5. Experiments

The proposed models are now applied, to a simulated transit corridor of 10 km of length, with 30 bus stops evenly spaced, where the terminal is denoted by stops 1 and 31. Vehicle operating speed between stops for all of the buses is assumed to be 26 km/h, while boarding and alighting time per passenger is set at 2.5 and 1.5 s respectively. All passenger arrivals are governed by Poisson processes.

5.1. Scenarios

In order to evaluate and compare the proposed model under different operational conditions, two different load bus profiles are tested: (i) bus loads are concentrated in the center of the corridor; (ii) bus loads are concentrated at the end of the corridor. On both scenarios, we also distinguish scenarios where: (i) bus capacities are reached; (ii) bus capacities are never reached; and (iii) scenarios with high frequency services, i.e. short headways; and (iv) scenarios with medium frequency services. Therefore, we considered eight scenarios as shown in Table 1.

5.2. Simulated control strategies

The objective function of the modeled scenarios is solved using three different control strategies. The first (*without control*, WC) is a benchmark for purposes of comparison; the second is the deterministic control strategy, DC, proposed in Delgado et al. (2012) with only holding in which high weights on the penalty for passengers left behind are implemented if there is available capacity so that boarding limits are never implemented, see Eq. (1); and the last one HPC proposed in Sáez et al. (2012) but using just holding strategy. In all scenarios, all the components of travel times involved in objective functions (1) and (32) are weighted equally; i.e. $\theta_1 = \theta_2 = \theta_3 = 1$. The DC strategy was coded in AMPL and solved using MINOS, while the HPC strategy was coded in Matlab with optimization toolbox.

Summarizing, we compare the three following control strategies:

- WC, Without control. That is the spontaneous evolution of the system, where buses are dispatched from the terminal at a designed headway, without taking any control action along the route, therefore some holding is expected to occur at terminal in which few passengers are affected.
- DC, Deterministic control strategy.
- HPC, Hybrid predictive control with 3 discrete holding values.

Table 1
Simulation scenarios.

Scenario	Capacity of buses reach	Design headway	Scenario load	Scenario characteristics
1	Yes	Short	Center	Headway = 140 s; Bus capacity = 100 passenger
2	No	Short	Center	Headway = 120 s; Bus capacity = 100 passenger
3	Yes	Medium	Center	Headway = 290 s; Bus capacity = 150 passenger
4	No	Medium	Center	Headway = 250 s; Bus capacity = 150 passenger
5	Yes	Short	End	Headway = 140 s; Bus capacity = 100 passenger
6	No	Short	End	Headway = 120 s; Bus capacity = 100 passenger
7	Yes	Medium	End	Headway = 290 s; Bus capacity = 150 passenger
8	No	Medium	End	Headway = 250 s; Bus capacity = 150 passenger

Table 2
Results for eight scenarios.

	WC	DC	Benefit %	HPC	Benefit %
<i>Scenario 1. Capacity reached - Centered load - Headway = 140 s - Cap = 100 pax/bus</i>					
W_{first}	0.65	0.15	76.75	0.24	62.81
(Std)	(0.06)	(0.05)		(0.05)	
W_{extra}	0.09	0.02	80.66	0.06	33.33
(Std)	(0.04)	(0.02)		(0.03)	
W_{total}	0.74	0.17	77.22	0.30	59.24
(Std)	(0.07)	(0.07)		(0.07)	
R_{total}	10.01	9.62	3.96	9.70	3.10
(Std)	(0.17)	(0.16)		(0.15)	
<i>Scenario 2. No Capacity reached - Centered load - Headway = 120 s - Cap = 100 pax/bus</i>					
W_{first}	0.31	0.07	78.69	0.13	57.28
(Std)	(0.04)	(0.02)		(0.02)	
W_{extra}	0.00	0.00	–	0.00	–
(Std)	(0.00)	(0.00)		(0.00)	
W_{total}	0.31	0.07	78.69	0.13	57.28
(Std)	(0.04)	(0.02)		(0.02)	
R_{total}	8.73	8.62	1.24	8.85	–1.40
(Std)	(0.10)	(0.10)		(0.91)	
<i>Scenario 3. Capacity reached - Centered load - Headway = 290 s - Cap = 150 pax/bus</i>					
W_{first}	0.90	0.12	86.26	0.67	25.86
(Std)	(0.40)	(0.09)		(0.33)	
W_{extra}	0.11	0.00	96.96	0.14	–27.60
(Std)	(0.14)	(0.00)		(0.12)	
W_{total}	1.01	0.13	87.42	0.81	20.04
(Std)	(0.54)	(0.09)		(0.45)	
R_{total}	10.88	10.27	5.60	11.17	–2.70
(Std)	(0.61)	(0.23)		(0.58)	
<i>Scenario 4. No Capacity reached - Centered load - Headway = 250 s - Cap = 150 pax/bus</i>					
W_{first}	0.41	0.05	86.92	0.10	75.92
(Std)	(0.29)	(0.04)		(0.13)	
W_{extra}	0.00	0.00	–	0.00	–
(Std)	(0.00)	(0.00)		(0.00)	
W_{total}	0.41	0.05	86.92	0.10	75.92
(Std)	(0.29)	(0.04)		(0.08)	
R_{total}	9.21	9.02	2.04	9.00	2.28
(Std)	(0.36)	(0.17)		(0.15)	
<i>Scenario 5. Capacity reached - Ended load - Headway = 140 s - Cap = 100 pax/bus</i>					
W_{first}	0.58	0.11	80.77	0.14	75.93
(Std)	(0.06)	(0.03)		(0.07)	
W_{extra}	0.11	0.01	86.67	0.04	63.64
(Std)	(0.05)	(0.01)		(0.02)	
W_{total}	0.69	0.13	81.71	0.18	73.97
(Std)	(0.09)	(0.04)		(0.03)	
R_{total}	10.48	9.96	4.98	10.13	3.34
(Std)	(0.20)	(0.18)		(0.09)	
<i>Scenario 6. No Capacity reached - Ended load - Headway = 120 s - Cap = 100 pax/bus</i>					
W_{first}	0.37	0.08	77.11	0.08	78.43
(Std)	(0.06)	(0.07)		(0.05)	
W_{extra}	0.00	0.00	–	0.00	–
(Std)	(0.00)	(0.00)		(0.00)	
W_{total}	0.37	0.08	77.11	0.08	78.43
(Std)	(0.06)	(0.07)		(0.04)	
R_{total}	9.41	9.21	2.12	9.14	2.87
(Std)	(0.16)	(0.14)		(0.12)	
<i>Scenario 7. Capacity reached - Ended load - Headway = 290 s - Cap = 150 pax/bus</i>					
W_{first}	0.55	0.06	88.46	0.16	71.50
(Std)	(0.26)	(0.06)		(0.01)	
W_{extra}	0.18	0.02	89.64	0.11	38.89
(Std)	(0.14)	(0.02)		(0.08)	
W_{total}	0.73	0.08	88.75	0.27	63.41
(Std)	(0.39)	(0.09)		(0.17)	
R_{total}	11.48	11.06	3.66	11.31	1.48
(Std)	(0.43)	(0.26)		(0.29)	

(continued on next page)

Table 2 (continued)

	WC	DC	Benefit %	HPC	Benefit %
<i>Scenario 8. No Capacity reached - Ended load - Headway = 250 s - Cap = 150 pax/bus</i>					
W_{first}	0.43	0.04	90.78	0.12	72.59
(Std)	(0.20)	(0.04)		(0.15)	
W_{extra}	0.00	0.00	–	0.00	–
(Std)	(0.00)	(0.00)		(0.00)	
W_{total}	0.43	0.04	90.78	0.12	72.59
(Std)	(0.20)	(0.04)		(0.09)	
R_{total}	10.11	9.73	3.70	9.81	2.97
(Std)	(0.30)	(0.16)		(0.21)	

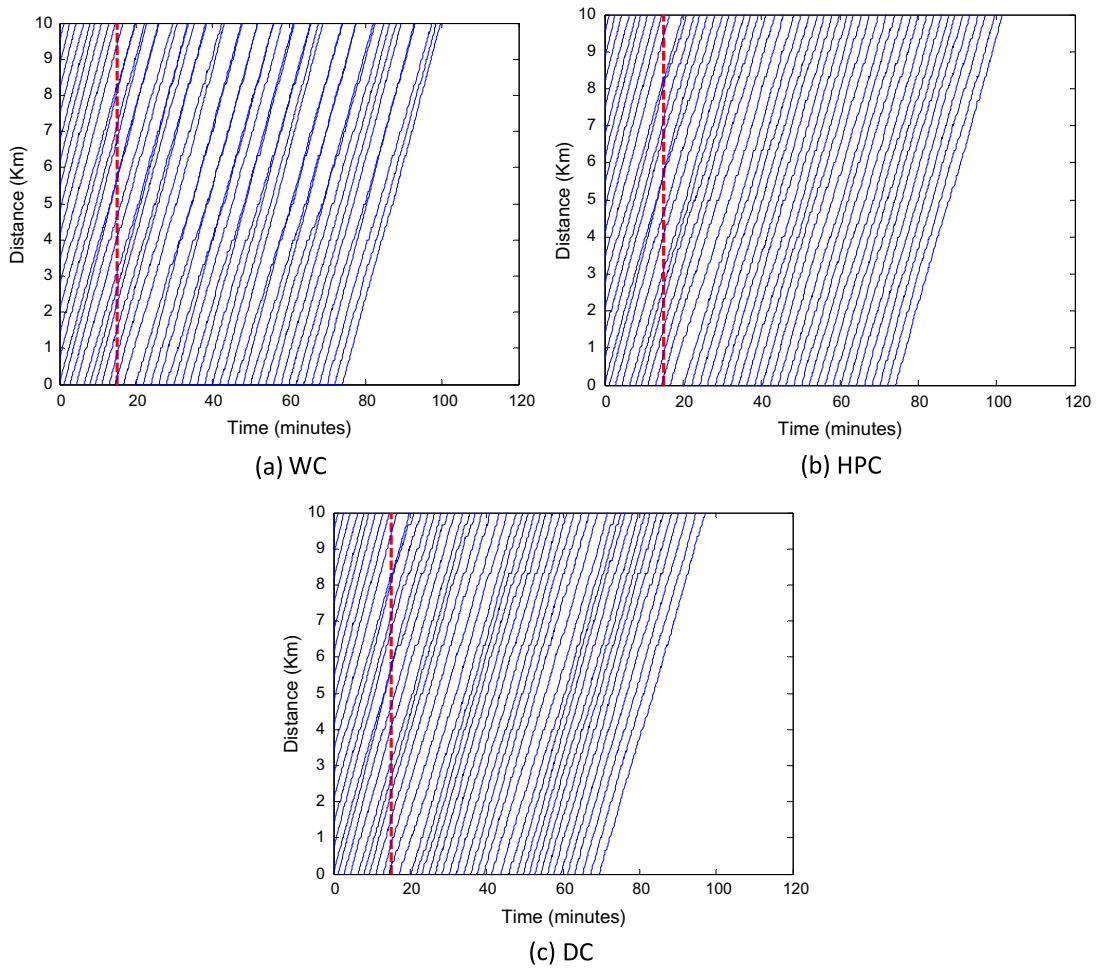


Fig. 2. Trajectories of buses for the different strategies: (a) WC; (b) HPC; (c) DC.

5.3. Simulation results

For every combination of strategies and scenarios, simulations were carried for 30 replication runs, each of them representing 2 h of bus operations. The system was simulated, using and adaptation of the simulator developed by Cortés et al. (2010), using common random numbers and the same initial conditions, corresponding to buses without passengers aboard and evenly spaced along the corridor. A warm-up period of 15 min is considered for all scenarios, before any control strategy is applied.

For each simulation run, the following measures of performance were examined:

- W_{first} waiting time experienced by passengers at stops as they wait for the first bus to arrive, prorated by all passenger. Since there is a minimum waiting time for a given frequency which is unavoidable (i.e. half of the average headway), Table 2 only presents the difference between the total waiting time until the arrival of the first bus and this lower bound.
- W_{extra} extra waiting time of passengers that cannot board the first bus, prorated by all passenger.
- W_{total} W_{first} plus W_{extra} .
- R_{total} W_{total} plus the individual travel time experienced by passenger on the bus (including holding).

6. Analysis of results and conclusions

From the analysis of the previous results shown in Table 2, we can obtain some conclusions with regard to the conditions under which one methodology can perform better than the other and vice versa. We can also devise a combined method over long periods in cases where the conditions change over time and therefore could utilize the benefits of either one or the other approach depending on the better option at each time. From Table 2, the following considerations and conclusions seem to be relevant:

- In all cases, independently of the scenario to be modeled and the method to be used, we observe a significant reduction in the following indicators: W_{first} , W_{total} . For the DC approach we also see a significant reduction of W_{extra} and R_{total} for all scenarios. For the HPC approach, W_{extra} and R_{total} improves in all scenarios with the exception of scenario 3 for the former and scenarios 2 and 3 for the latter. This shows the benefits of using control in operational decisions of real transit systems.

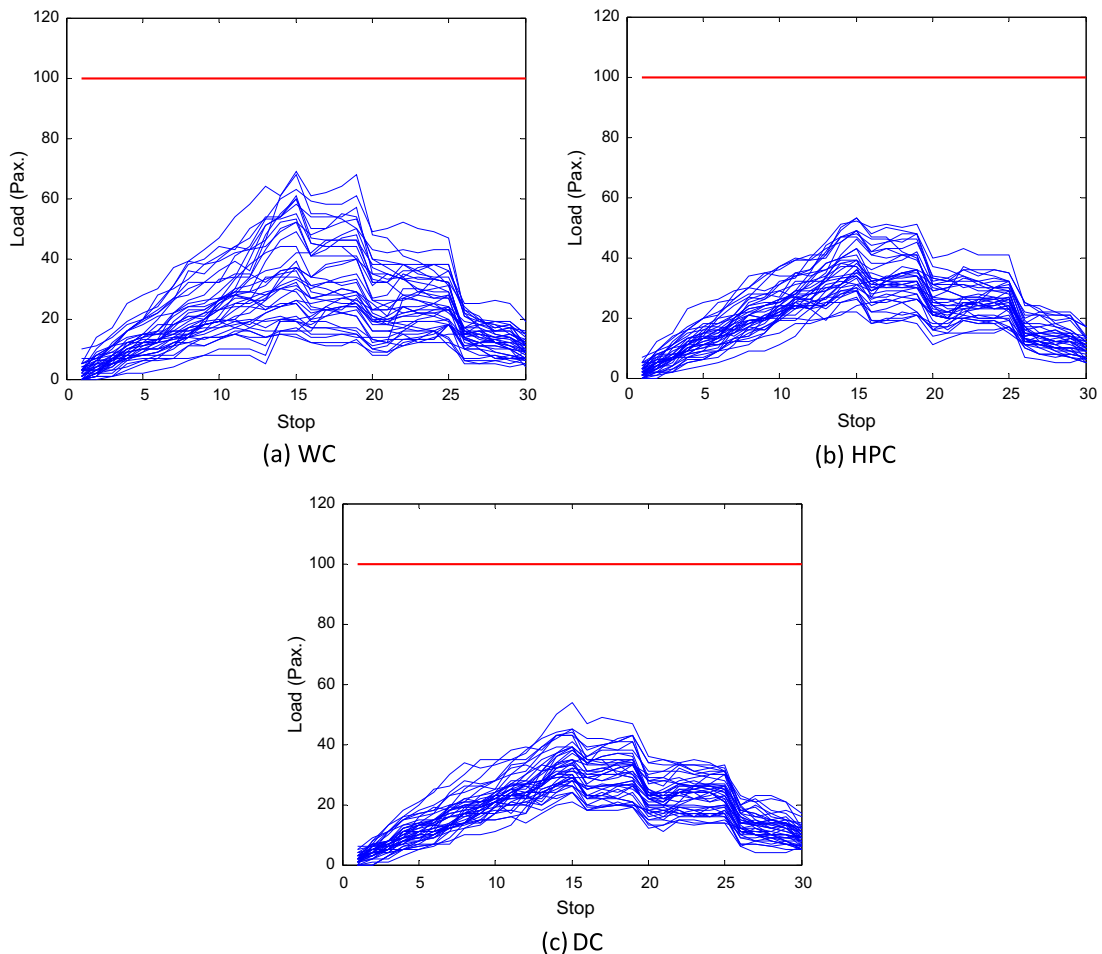


Fig. 3. Bus load at different stops for the different strategies: (a) WC; (b) HPC; (c) DC.

In order to understand why HPC performs worse than DC in most cases, and for some exceptional scenarios HPC apparently seems to be worse than the WC strategy in some indicators, Fig. 2, shows the bus trajectories for scenario 2 for these three strategies for a typical simulation run. While in the WC strategy buses tend to bunch up, increasing the waiting time for passengers, the application of any of the proposed approaches improves headway regularity. If we compare the HPC with the DC strategy we can observe that the former achieve a better headway regularity between buses than the latter, which is reasonable as keeping regular headways is one of the objectives of HPC strategy. However, to increase regularity, too much holding needs to be applied which explained the increase in travel time for passengers inside the bus. In any case, we should recognize that users perceive in-vehicle travel time as significantly less relevant than waiting time. Thus, even though in this scenario WC yielded a lower R_{total} than HPC, the HPC trajectories are much more appealing than WC's.

The better headway regularity achieved by the two proposed strategy is also reflected in bus loads showing more uniform patterns, compared with the WC situation, as shown in Fig. 3.

- Moreover, in all these cases the reduction in standard deviation of the indicators is even more noticeable. This is especially relevant since the focus in regularity is aimed at reducing the fraction of passengers experiencing long waiting times.
- By comparing the two approaches (DC and HPC), we can see that HPC results in a better performance in terms of total waiting time (holding included) in scenarios 4, and 6, while DC performs better under scenarios 1, 2, 3, 5, 7 and 8. This conclusion shows that DC is convenient in cases where bus capacity could be reached frequently along the route while HPC takes advantage of situations where this does not happen, with the exception of scenario 2 explained above. We see two possible explanations for this.
- DC has a longer prediction horizon of future operational conditions than HPC; however, in DC the prediction is deterministic. This longer prediction horizon allows for preventive actions avoiding vehicles from reaching capacity downstream in the corridor to be taken. On the other hand, HPC works with considerably less steps ahead, but with more precision than DC and incorporates a term that looks for regular bus intervals (third term in expression (32)). These two elements appear to allow HPC to achieve better results when bus capacity is never reached. However, this last term in the objective function could also negatively affect its performance when capacity is reached, since in these cases regular headways may not be the best control strategy. The results also seem to show that the DC's explicit treatment of capacity constraints in the rolling horizon optimization model pays off in scenarios when these constraints are binding.

Acknowledgments

We thank the financial support of the ACT-32 Project “Real-Time Intelligent Control for Integrated Transit Systems, the Millennium Institute Complex Engineering Systems (ICM: P-05-004-F, CONICYT: FBO16) and FONDECYT projects 1100239 and 1120993. This research was also supported by the Across Latitudes and Cultures- Bus Rapid Transit Centre of Excellence funded by the Volvo Research and Educational Foundations (VREF).

References

- Barnett, A., 1974. On controlling randomness in transit operations. *Transportation Science* 8 (2), 102–116.
- Bartholdi III, J.J., Eisenstein, D.D., 2012. A self-coordinating bus route to resist bus bunching. *Transportation Research Part B* 46 (4), 481–491.
- Boardman, A., Greenberg, D., Vining, A., y Weimer, D., 2001. *Cost-Benefit Analysis*. Prentice Hall, E.E.U.U.
- Cortés, C.E., Sáez, D., Milla, F., Riquelme, M., Núñez, A., 2010. Hybrid predictive control for real-time optimization of public transport system' operations based on evolutionary multiobjective optimization. *Transportation Research Part C-Emerging Technologies* 18 (5), 757–769.
- Daganzo, C.F., 2009. A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B* 43 (10), 913–921.
- Daganzo, C.F., Pilachowski, J., 2011. Reducing bunching with bus-to-bus cooperation. *Transportation Research Part B* 45 (1), 267–277.
- Delgado, F., Muñoz, J.C., Giesen, R., Cipriano, A., 2009. Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. *Transportation Research Record* 2090, 59–67.
- Delgado, F., Muñoz, J.C., Giesen, R., 2012. How much can holding and/or limiting boarding improve transit performance? *Transportation Research Part B* 46 (9), 1202–1217.
- Eberlein, X.-J., 1995. *Real Time Control Strategies in Transit Operations: Models and Analysis*. PhD thesis. Massachusetts Institute of Technology, Cambridge, MA.
- Eberlein, X.J., Wilson, N.H.M., Bernstein, D., 2001. The holding problem with real-time information available. *Transportation Science* 35 (1), 1–18.
- Fu, L., Yang, X., 2002. Design and implementation of bus-holding control strategies with real-time information. *Transportation Research Record* 1791, 6–12.
- Fu, L., Liu, Q., Calamai, P., 2003. Real-time optimization model for dynamic scheduling of transit operations. *Transportation Research Record* 1857, 48–55.
- Furth, P.G., Day, F.B., 1984. Transit routing and scheduling strategies for heavy demand corridors. *Transportation Research Record* 1011, 23–26.
- Hickman, M., 2001. An analytic stochastic model for the transit vehicle holding problem. *Transportation Science* 35 (3), 215–237.
- Puong A., Wilson, N.H.M., 2004. A train holding model for urban rail transit systems. In: 9th International Conference on Computer-Aided Scheduling of Public Transport (CASPT) San Diego, California. <<http://fugazi.engr.arizona.edu/caspt/puong.pdf>> (accessed 15.03.08).
- Sáez, D., Cortés, C., Riquelme, M., Núñez, A., Milla, F., Tirachini, A., 2012. Hybrid predictive control strategy for a public transport system with uncertain demand. *Transportmetrica* 8 (1), 61–86.
- Suh, W., Chon, K., Rhee, S., 2002. Effect of skip-stop policy on a Korean subway system. *Transportation Research Record* 1793, 33–39.
- Sun, A., Hickman, M., 2004. The holding problem at multiple holding stations. In: 9th International Conference on Computer-Aided Scheduling of Public Transport (CASPT) San Diego, California. <<http://fugazi.engr.arizona.edu/caspt/sun.pdf>> (accessed 15.03.08).
- Sun, A., Hickman, M., 2005. The real-time stop-skipping problem. *Journal of Intelligent Transportation Systems*, 91–109, <[http://www.informaworld.com/smp/title~content=t713398522~db=all~tab=issueslist~branches=9-v99\(2\)>](http://www.informaworld.com/smp/title~content=t713398522~db=all~tab=issueslist~branches=9-v99(2)>).

- Turnquist, M., Blume, S., 1980. Evaluating potential effectiveness of headway control strategies for transit systems. *Transportation Research Record* 746, 25–29.
- Yu, B., Yang, Z., 2007. A dynamic holding strategy in public transit systems with real-time information. *Applied Intelligence* 31 (1), 69–80.
- Zhao, J., Bukkapatnam, S., Dessouky, M., 2003. Distributed architecture for real-time coordination of bus holding in transit networks. *IEEE Transactions on Intelligent Transportation Systems* 4 (1), 43–51.
- Zolfaghari, S., Azizi, N., Jaber, M., 2004. A model for holding strategy in public transit systems with real time information. *International Journal of Transport Management* 2, 99–110.