This is the unformatted version of:

Fonzone, A., Schmöcker, J.-D. and Liu, R. (2015). A model of bus bunching under reliability-based passenger arrival patterns. Transportation Research C. Available from http://dx.doi.org/10.1016/j.trc.2015.05.020. Also available from Selected Proceedings of the 21th International Symposium on Transportation and Traffic Theory (ISTTT). Kobe, Japan, August 2015.

A model of bus bunching under reliability-based passenger arrival patterns

Achille Fonzone[a*], Jan-Dirk Schmöcker[b], Ronghui Liu[c]

[a] *Transport Research Institute, Edinburgh Napier University, Edinburgh, U.K.*
[b]*Kyoto University, Department of Urban Management, C1-2-436, Katsura Nishikyo-ku, Kyoto 615-8540, Japan*
[c] *Institute for Transport Studies, Leeds University, Leeds LS2 9JT, U.K.*

Abstract

If bus service departure times are not completely unknown to the passengers, non-uniform passenger arrival patterns can be expected. We propose that passengers decide their arrival time at stops based on a continuous logit model that considers the risk of missing services. Expected passenger waiting times are derived in a bus system that allows also for overtaking between bus services. We then propose an algorithm to derive the dwell time of subsequent buses serving a stop in order to illustrate when bus bunching might occur. We show that non-uniform arrival patterns can significantly influence the bus bunching process. With case studies we find that, even without exogenous delay, bunching can arise when the boarding rate is insufficient given the level of overall demand. Further, in case of exogenous delay, non-uniform arrivals can either worsen or improve the bunching conditions, depending on the level of delay. We conclude that therefore such effects should be considered when service control measures are discussed.

1. Introduction

Service irregularities increase passenger waiting times, decreasing the attractiveness of public transport. The more the passengers can trust the service schedule, the better they can time their arrival at stops. Whereas under completely random service arrivals the passengers can also do no better than "randomly" arrive at stops, in many cases at least some coherence of the actual arrivals with the service schedule might be expected. Therefore, even if the schedule might not be known to all passengers and uncertainties in the access time to the stop are considered, non-uniform passenger arrival patterns can be expected. With a few exceptions, the effect of such non-uniformity on bus loads has been largely ignored in the literature and is the topic of this contribution.

We propose a "mixed behaviour": Passengers consider the likely service departure times and leave some safety margins in order to ensure that they do not have to wait too long for a bus but also minimise the chances of missing a service. Such a behaviour seems reasonable for passengers in cities with fairly good bus services. As a motivating example familiar to the authors, consider the bus stop in front of Kyoto University. The stop is close to the office buildings and the most frequent service arrives around every 15 minutes during the evening hours. Some passengers, possibly those without knowledge of the schedule, will arrive randomly though the bulk of passengers will time their arrival to 2-3 minutes before scheduled service arrival. Some "risky" passengers, or those delayed by for example waiting times at the elevator, will arrive even closer to the scheduled departure.

[*] Corresponding author. Tel.: +44-131-455-2898.
E-mail address: a.fonzone@napier.ac.uk

In general, it is well known that passenger arrival is influenced by service characteristics, such as average value of headways and headway deviations. In particular, it is commonly accepted that passengers tend to arrive closer to the scheduled departure time (i.e., their arrivals are not uniform) when the headways are large. This behaviour is oftern termed as scheduled-based arrival behaviour. Bowman and Turnquist (1981) provide a model of passenger arrival behaviour, which links the arrival distribution to the characteristics of the service, including its reliability.

Therefore a good understanding of arrival patterns is the foundation to modelling boarding demand. Deriving bus loads are important to estimate potential capacity bottlenecks and possibly revenue splits among bus operators. Furthermore, bus loads and bus service dwell times at stops are closely correlated, and unexpected high loads can lead to the well-known "bus bunching" process. The seminal work of Newell and Potts (1964) presents a simplified model of the phenomenon, which casts light on some causes. However, their model does not provide a realistic representation of bunching as they neglect aspects such as en-route service perturbations, transport operator policies concerning holding and overtaking as well as complex network features such as the presence of "common lines" among which some of the passengers at a stop might choose. Some of these issues have been dealt with in later literature as reviewed in more detail in the next section.

Newell and Potts further assume uniform passenger arrival. In the above Kyoto bus stop example this might overestimate the bunching phenomena as only a few additional passengers arrive in the time interval between the scheduled and actual service departure and hence delayed buses have to board fewer additional passengers than predicted with uniform arrival. Furthermore, Newell and Potts do not capture the effect of severe bunching where buses might be overtaken.

The contributions of this paper are twofold. Firstly, a model of passenger arrival extending the approach of Bowman and Turnquist (1981) to allow for overtaking between buses at a stop. We refer to our model as the "reliability-based arrival pattern model" in line with the above example. Secondly, we include these passenger arrival patterns in a model of bus propagation, highlighting causes of bunching which are not identified by Newell and Potts.

In line with above discussion, our model will be mainly applicable to situations in which passengers consider timetables in deciding their arrival at stops. It is conventionally accepted that timetables influence passengers decisions for services with expected headways of more than 10 minutes and that, instead, if service headways are shorter, uniform passenger arrival patterns can be expected. Actually the threshold between schedule-dependent and uniform passenger arrival can be lower than the conventional one. A review of existing studies on the relation between service headway and passenger arrival at stops is provided by Luethi et al. (2006). Interestingly, this study finds that passengers consult schedules even when the headway is 5 minutes. We consider the topic discussed in this paper especially of topical importance due to the increasing presence of service schedule information to passengers before arrival at a stop even for passengers unfamiliar with the network due to online availability of journey planners. More and more cities now provide real-time information (RTI) for passengers. RTI changes the "visibility of the network" and hence passenger behaviour. For instance, it is reasonable to expect that ubiquitous RTI on departure time (accessed by internet and/or mobile phone apps) induces non-uniform passenger arrivals also for short headways and irregular services.

The reminder of this paper is organized as follows: Section 2 provides a more detailed review of the two key references for this paper, Newell and Potts (1964) and Bowman and Turnquist (1981) as well as further related and newer literature. Section 3 then introduces the notation that is utilized in later sections. Section 4 describes the passenger arrival model and Section 5 the bus propagation model. Section 6 illustrates both models through case study applications before Section 7 concludes this paper.

2. Literature Review

Bus bunching is generally defined as the effect of two successive services of a single line arriving at stops with shorter than scheduled headways. The effect occurs by the first service being delayed at previous stops due to unplanned long boarding times, or being delayed en-route by unforeseen traffic congestion. The subsequent service then has to pick up fewer passengers at that stop and departs earlier than scheduled. At downstream stops, the effect is further emphasised as the initial delay to the first vehicle and the early arrival of the subsequent service result in increasingly longer dwell times for the first bus and increasingly shorter dwell times for the second bus.

Bus bunching has a direct negative impact on the passengers as it leads to, on average, longer dwell-times. Quarmby (1967) already found that transit passengers value their time waiting two to three times more than their time on board travelling. Subsequent literature has confirmed this, sometimes reporting even higher disutility associated with waiting. Using a stated-preference survey, Hollander and Liu (2008) found that bus passengers value service reliability four times higher than they do to mean travel time. Hollander et al. (2007) further demonstrate that bus unreliability has a significant impact on passengers' response in their departure-times.

Bus bunching is a common feature in urban public transport, and a long-standing problem facing the bus/transit service providers and academic researchers alike. The bunching effect on a single line has been first analytically described by Newell and Potts (1964). Assuming that travel times between stops are identical and that passenger loads are constant, Newell and Potts show that if the passenger arrival rate at a stop is larger than half the loading rate of buses the bunching effect occurs for small perturbations in the original schedule. If the ratio (referred to below as ρ-ratio) is smaller, the system can recover from perturbations. Furthermore, bus bunching is more noted in high frequency services, where the headway between buses is small and the delay to headway ratio is more likely exceeding the threshold for a localised bus bunching to amplify (rather than being damped) further down the route. Recently, Schmöcker et al. (2015) extended the bunching research to a route section served by two lines with high frequency so that some passengers might board "whichever line comes first". They show therefore how bunching can spread to initially unaffected lines but also how a line serving similar stops can help with service recovery on the initially affected line.

There has been a significant body of research designing operational strategies to address the bus bunching problem. In particular, holding strategies of early buses as well as strategies to keep minimum distances between subsequent services have been analysed and shown to be successfully applied in literature. The holding strategies are implemented through building slacks in the schedule at key timing points and holding buses at these points to keep them to schedule (e.g. Osuna and Newell, 1971; Newell, 1974; Hickman, 2001; Xuan et al., 2011; Cats et al., 2012).

Due to the complexity of the problem, most of the analytical studies involved having just one controlled timing point. Hickman (2001) developed a simulation approach to solve the bus bunching problem through optimal holding points. These are static holding solutions, which do not respond to dynamical changes in the actual bus performances on the day. Eberlein et al. (2001) developed a model for dynamical bus holding which take real-time information on bus headways into consideration and achieves to minimise passenger waiting time. The model assumes deterministic passenger demand and bus travel times between stops. Daganzo (2009) developed a more systematic approach for dynamical holding analysis with real-time bus performance information. Daganzo's method is able to consider holding at multiple timing points, therefore providing opportunity for return to schedule for long bus routes. In addition, the model takes into account random effects, such as the random variations in bus travel time, bus dwell time and passenger demand, making it resemble more realistically to real-life situations. Daganzo and Pilachowski (2011) proposed an adaptive bus control scheme based on a two-way bus-to-bus cooperation, where a bus adjusts its speed to both its front and rear headways. They show that the scheme yields significant improvements in bus headways and bus travel time. Moving away from the traditional ideal of schedule and a prior target headway, Bartholdi and Eisenstein (2012) proposed a self-coordinating method to equalise bus headway, while Pilachowski (2009) proposed to use GPS data to counteract the cause of the bunching directly by allowing the buses to cooperate with each other and to determine their speed based on relative position. Using a dynamic simulation approach to model the stochastic nature of the transport networks, Hadas and Ceder (2010) examined a range of operational tactics to improve bus service reliability. Muñoz et al. (2013) compared the performance of dynamic bus control strategies under different scenarios of bus capacity, service frequency and passenger load profiles. Most of the existing studies are concerned with a simplified bus system, notably with a single line, fixed service frequency, constant passenger flows, and no bus overtaking. Newell and Potts (1964), for instance, assume fixed frequency, constant dwell times, equal-distance stops and equal-travel time between stops, and that buses cannot overtake. In real-life situations these simplifying assumptions obviously might not hold. Boyd (1983) presented empirical evidence that demonstrated the impact of variability in bus journey time on bunching. Nagatani (2001) shows a strong relationship between bus delay and the passenger number on bus, and proposes skipping a bus stop as a way of keeping to schedule. Liu et al (2013) studied the bus stop-skipping scheme considering random travel time. Recently, Nesheli and Ceder (2014) introduced the concept of skip-a-segment in addition to skip a stop. In general, bus corridors operating multiple lines that interact through sharing passengers and bus stops are considered complex scenarios for most of the existing analytical studies (Daganzo and Pilachowski, 2011).

Another significant simplification in the existing studies that was mentioned in the introduction is the assumption of a uniform passenger demand distribution over time and space. This hypothesis of random passenger arrivals is common in models of public transport. When the randomness of the bus arrival at stops is taken into consideration, a model of passenger arrivals is needed to compute the waiting time and so the overall journey time. In networks with frequent transit services and no ubiquitous real time information, models based on the random incidence assumption can be justified: in fact passengers may not consider schedules if they know that their waiting time for the next service will be low in any case. As shown for instance in Larson and Odoni (1981), under the assumption of passenger random arrival, the waiting time expectation can be easily derived from the first and second moment of the bus arrival distribution. Random incidence is assumed in most of the literature on frequency-based route choice and assignment, notably in the seminal work of Spiess and Florian (1989) when deriving “optimal strategies”. The assumption can be still retained when passengers have access to real time information but only once they have reached a stop, as considered by Gentile et al. (2005). Conversely, passenger arrival distributions can be expected not to be uniform when real time information is available to passengers even before they go to a stop. With the increasing diffusion of internet-based information and smart phones, this scenario is becoming more and more common in many cities around the world. It can be anticipated that passengers who know the predicted bus departure times in advanced will try to coordinate their arrival with that of the next attractive bus. Watkins et al. (2011) observed a reduction of the actual waiting time from 11 to 9 minutes for passengers of King County Metro using information disseminated by internet. Clearly real time information is not perfectly reliable: bus arrival prediction normally entails information about the current position of transit vehicles and the use of models to predict the travel time to downstream stops. Both these elements are affected by uncertainty (see for instance Crout, 2007).

Liu and Sinha (2007) have further confirmed the non-uniformity of demand. They collected data on bus travel time, dwell time, and passenger boarding and alighting along a commuter bus route in the city of York, in England. Significant variations of demand (boarding and alighting) were detected both across bus stops and over time. Building their observed data into a microsimulation model of the bus corridor, Sorratini et al. (2008) showed that the variability of the passenger demand distribution has the most significant impact on bus reliability measures.

Based on this empirical evidence therefore it appears reasonable to argue that passenger decision making contemplates the overall system reliability deriving from the combination of the transport service reliability and the information system accuracy. A model in which some “schedule-aware” passengers choose their arrival considering system reliability was put forward by Bowman and Turnquist (1981). In their model, the passenger arrival pattern at a bus stop is described by a continuous logit function. The utility function depends on the expected waiting time calculated by averaging waiting times over all possible bus arrival times for a specific passenger arrival time, i.e.

$$E\big(W(t)\big) = [1 - P(t)]W(t) + P(t)W'(t) \tag{1}$$

where $E\big(W(t)\big)$ is the expected waiting time, $P(t)$ the probability that the intended bus is missed by a passenger arriving at t, $W(t)$ the expected waiting time in case the intended bus arrives after t and $W'(t)$ the expected waiting time in case the bus has departed before t. Bowman and Turnquist provided no details regarding the calculation of the probabilities and expected values. Considering the context of their paper, it seems that they do not consider overtaking. This assumption can be valid for stops served by a single line, when severe bunching is not likely or when overtaking is prohibited by the operator or not feasible due to road space. It has to be relaxed though when bus bunching is such that the presence of more than one bus at the same stop can be an issue, and/or when more attractive lines serve the same stop and their services can overlap. Bowman and Turnquist showed that their model predicts the actual arrival patterns very well, and that their model can cover both the headway ranges normally modelled with a frequency-based approach and those modelled by a scheduled-based one. Furthermore, they demonstrated that passengers are more sensitive to schedule reliability than to service frequency. A limitation of their model, stressed by the authors themselves, is the use of the continuous logit model. GEV models like the logit one are commonly used for discrete choices within the framework of random utility maximization. The application of the logit model is constrained by the underpinning hypothesis of the independence of errors across alternatives. The validity of this assumption becomes particularly debatable when the decision dimension is continuous. The problem has been addressed in the literature concerning the topic of the departure-time choice of travellers using private transport. Choices in a continuous setting instead have been often described by models based on hazard functions; but such an approach is criticized because it lacks a proper behavioural support. Recently Lemp et al. (2010) suggested a continuous cross-nested logit model, which retains the behavioural justification of the logit models and allows for correlation across alternatives.

In summary, most literature related to bus bunching since Newell and Potts (1964) has focused on deriving control strategies to minimize the bunching effect. There is a separate set of literature discussing arrival patterns and the value of waiting times but, to the best of our knowledge, the results of the latter set of literature have not yet been included in bus bunching models. We propose that considering more advanced passenger arrival time models seems necessary to build a realistic model of bus bunching, especially in the era of ubiquitous (real-time) information.

3. Notation

The following notation is used throughout the paper.

Table 1 Notation

i	Stop number. Stop 0 is the bus depot, where no boarding is allowed.
n	Bus index or bus run (concisely, "bus") designated by index n (the meaning will be clarified by the context)
$\tau_{in}^{a}, \tau_{in}^{d}$	Arrival, departure time of bus n from stop i. Superscript d is omitted when redundant.
$f_{in};\ F_{in}$	Perceived probability distribution function of τ_{in}; perceived cumulative distribution function of τ_{in}. These and all the remaining probabilities mentioned in the paper have to be considered as describing the perception of the travellers rather than the actual characteristics of the service.
T_{in}	Interval of perceived possible departure times of bus m from stop i, i.e. $\mathrm{T}_{in} = \{\tau_{in} \vert f_{in}(\tau_{in}) > 0\}$
B_i	Set of buses serving stop i
$B_i(t)$	Set of buses available at stop i after t, i.e. $B_i(t) = \{\mathrm{n} \in B_i \vert \sup \mathrm{T}_{in} \geq t\}$
$A_{in}(t)$	Set of buses alternative to the focal bus serving stop i after t, i.e. $A_{in}(t) = \{m \in B_i(t), m \neq n\}$
$A_{in}^{-}(t)$	Set of alternative buses which can depart from stop i before t, i.e. $A_{in}^{-}(t) = \{m \in A_{in}(t) \vert \inf \mathrm{T}_{im} \leq t\}$
$\mathcal{P}\left(A_{in}^{-}(t)\right)$	Power set of $A_{in}^{-}(t)$, i.e. is set of all subsets of $A_{in}^{-}(t)$, including the empty set and $A_{in}^{-}(t)$ itself.
$P_{in}(t);\ \bar{P}_{in}(t)$	Generic element of $\mathcal{P}\left(A_{in}^{-}(t)\right)$; set of elements of $A_{in}(t)$ not included in $P_{in}(t)$, i.e. $\bar{P}_{in}(t) = \{S \in A_{in}(t) \vert S \notin P_{in}(t)\}$
$\varphi_i(t, \tau_{in}, S)$	Probability distribution function of the event "bus n departs from stop i at $\tau_{in} > t$ and before any other bus in the set S".
$\Phi_{in}(t, S)$	Probability that n departs from stop i after t and before any bus in the set S
$\Xi_i(t, S)$	Probability that all buses in the set S have departed from stop i before t
$\pi_{in}(t, \tau)$	Probability distribution function of the event "passenger arriving at stop i at t and boarding bus n at τ"
$\Pi_{in}(t)$	Probability that a passenger arriving at stop i at t boards bus n
$E\left(w_i(t)\right)$	Expected waiting time of a passenger arriving at stop i at t
$E\left(w_{in}(t)\right)$	Expected waiting time for bus n of a passenger arriving at stop i at t
$W(t)$	Anticipated risk-averse waiting time of a passenger arriving at stop i at t
α, β	Parameters measuring aversion of passengers to wait
$q_i(t)$	Passenger arrival rate at stop i
$b_i(t), b$	Time-dependent boarding rate at stop i; time and stop independent boarding rate
ρ	Ratio between arrival and boarding rate, referred to later as the saturation rate
v_{in}	Travel time of bus n from stop i to stop $i+1$
D_{in}	Dwell time of bus n at stop i
δ_{in}	Exogenous delay of bus n at stop i

4. Reliability-based passenger arrival

4.1. Assumptions and model limitations

To explore the impact of non-uniform passenger arrivals on bus bunching, in this section we derive a model of passenger arrival at bus stops based on the assumption that travellers aim to reduce their expected waiting time considering that they may miss some or all of their intended buses because of service irregularity. Our passenger utility function is based on an anticipated risk-averse waiting time. By "anticipated" we mean that passengers decide their arrival time based on their perceived probability of bus departure times in a non-perfectly reliable system, i.e. in a system in which the actual departure times can differ from the scheduled ones. By "risk-averse" we mean that passengers attach a cost to the possibility of missing their last service. Our utility function is based on a behavioural model similar to that considered by Bowman and Turnquist, but we extend their approach by allowing for the existence of an indefinite number of buses and for the possibility of overtaking among buses. Different specifications of the utility functions are discussed below. The model presumes that passengers have a priori information on the scheduled departure time of buses and on the accuracy of such information. The first hypothesis holds in the case of relatively low-frequency services for which a timetable is published and known by the travellers. In the case of services for which schedules are not published, it may still be valid if the day-by-day service regularity is high. The accuracy of the information considered in the decision-making process depends both on the punctuality of the service and on the passenger's knowledge of it: passengers who have used a given transit system only few times, might not trust the information they have and hence consider a large variability of departure times even when the system is in fact very reliable. In the following, by system reliability we mean the reliability accounted for in the decision-making process, as a consequence of the perceived information accuracy.

Both the expected waiting times and the derived utility function are deterministic. However, we consider that the perception of the value of each arrival time varies among users. In particular, following Bowman and Turnquist, we assume that the passenger arrival distribution can be described by a continuous logit model. Not much work has been done concerning passenger arrival patterns after the contribution of Bowman and Turnquist (1981). Since our focus here is on the derivation of the utility function, we use the continuous logit model as it has been proven in the Bowman and Turnquist paper to be able to describe the actual behaviour of public transport users for their scenario assumptions despite its theoretical limitations. In our model, passengers always get to a bus stop at their intended arrival time. In reality, this might not happen because passengers cannot always anticipate their walking time exactly. This phenomenon exceeds the scope of this study.

It is assumed that passengers board the bus that departs first after their arrival. Such an assumption is realistic if we assume that only one bus line serves the stop or that all bus lines travel the same route downstream from the boarding point. In these cases passengers do not face the "common line issue" where the choice set might depend on the time spent waiting at the stop already (see Noekel and Weckek (2009) for elaboration of such cases). The "board the first departing bus" behaviour is further justified by the empirical findings that passengers put more cost on wait time than in-vehicle travel time (e.g. Quarmby, 1967). However, we acknowledge that there are cases in which "board the first bus" strategy is not realistic, for instance when crowding on buses affects the decision/possibility to board, when passengers might anticipate variations in downstream travel times between buses, or when passengers prefer express services. Our model does not cover these cases. Statistical independence is assumed for both the departure times of different buses at the same stop and the departure times of the same bus at consecutive stops. This assumption is clearly an over-simplified representation of reality as departures of buses from the same stop can be correlated. Our assumption does not hold if, for example, particular operational policies exist, such as a FIFO rule for buses departing from the stop, or, if the stop layout does not allow more than one bus to board passengers at a time.

We further assume that the passenger arrival distributions at each stop are independent from what happens at other stops and therefore in this section the subscript representing the stop is redundant and will be omitted for the sake of simplicity. We add the subscript in Section 5 when considering the (obviously correlated) departures of the same bus from subsequent stops. This stop-independence assumption in the behavioural model implicitly assumes that real time information are not available but that passengers only know the scheduled information (plus their perception of the service reliability and the variance in their stop access time) when deciding their stop arrival time.

We hypothesize that a bus leaves a stop as soon as all waiting passengers have boarded, i.e. we do not consider any holding policy. This assumption implies that we consider bus stops as on the main road, rather than in a bus bay. Meng and Qu (2013) show that bus dwell time at bus bays possesses a high degree of uncertainty caused by the fact that buses having to merge onto the main road and that the merging process is dependent on the level of traffic on the shoulder lane. In our model, as in most works in literature, the dwell time depends only on the boarding process. This is realistic when the boarding time is longer than the alighting time, which can occur because (a) either passengers who board are generally less than those who alight, and/or (b) the boarding time per passenger is considerably longer than the alighting one, due for instance to the fact that tickets are issued when boarding. In addition, no influence between boarding and alighting is considered. This happens when buses are not crowded, and/or boarding and alighting take place at different doors. Finally, we note that we also do not include capacity effects in this model; thus assume all waiting passengers can board the first departing bus.

4.2. Perceived probability of boarding a bus

Clearly if $B(t) = \{n\}$ (i.e. if bus n is the only departing after time t), the probability that a passenger who arrives at the stop at time t boards bus n is equal to the probability that n departs after t. Consider now that the passenger arrives at t but $B(t) = \{n, n+1\}$ and $\inf \mathrm{T}_{n+1} \le t$. Still the passenger can only board n if it departs after t. In this case he will actually board n if $n+1$ has already departed or it has not departed yet and actually leaves after n. In the remaining scenario, in which $n+1$ has not departed yet but it leaves before n, the passenger boards $n+1$. In other words, the overall probability of boarding n is equal to the sum of the probability of boarding it in each of two cases of $A_n^-(t) \in \mathcal{P}\big(A_n^-(t)\big) = \big\{\emptyset, \{n+1\}\big\}$:

$$\Pi_n(t) = \Pr(\tau_n > t\ \cap \tau_{n+1} < t) + \Pr(\tau_n > t\ \cap \tau_{n+1} > \tau_n) \tag{2}$$

In the general case in which $c = \big|\mathcal{P}\big(A_n^-(t)\big)\big| \ge 2$, i.e. considering the possibility of more than one alternative to n which can depart before t, n is boarded in all the cases in which some of the alternatives have departed before t and all others depart after n. The number of the scenarios to consider in the calculation of $\Pi_n(t)$ is equal to the number of combinations of buses in $A_{in}^-(t)$, i.e.

$$\Pi_n(t) = \sum_{P_n(t) \in \mathcal{P}\left(A_n^-(t)\right)} \Pr\Big((\tau_n > t)\ \cap \big(\tau_l < t\ \forall l \in P_n(t)\big) \cap \big(\tau_m > \tau_n\ \forall m \in \bar{P}_n(t)\big)\Big) \tag{3}$$

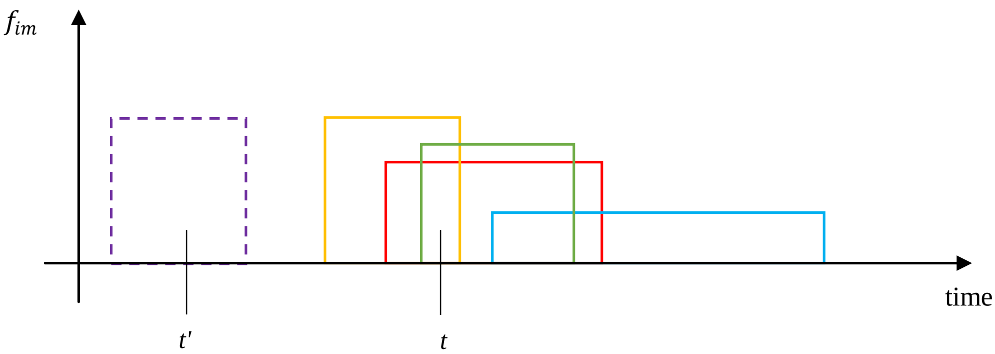

Fig. 1. Example of bus arrival probability distributions functions at a stop

To illustrate (3), consider the case in Fig. 1, in which lines of different colours represent the perceived probability distribution functions of different buses (in the explanation of the example, buses are indicated with the initial letter of their colour). Following the definitions in Section 3, $B = \{Purple, Yellow, Red, Green, Cyan\}$, $B(t) = \{Y, R, G, C\}$, $A_R(t) = \{Y, G, C\}$, $A_R^-(t) = \{Y, G\}$, $\mathcal{P}\big(A_P^-(t)\big) = \big\{\emptyset, \{Y\}, \{G\}, \{Y, G\}\big\}$. Bus P has surely departed at t, therefore $P \notin A_R(t)$ and it does not influence the probability of boarding n. Note that it does influence such probability if the passenger arrives at t'. There are four possible circumstances (note that $\big|\mathcal{P}\big(A_n^-(t)\big)\big| = 4$) under which R is boarded by a passenger arriving at t:

- No bus has departed before t, and R departs before all alternative buses: $P_R(t) = \emptyset, \bar{P}_R(t) = A_R(t) = \{Y, G, C\}$
- Y has departed before t, and R has not departed yet and departs before G and P: $P_R(t) = \{Y\}, \bar{P}_R(t) = \{G, C\}$
- G has departed before t, and R has not departed yet and departs before Y and P: $P_R(t) = \{G\}, \bar{P}_R(t) = \{Y, C\}$
- G and Y have departed before t, and R has not departed yet and departs before P: $P_R(t) = \{G, Y\}, \bar{P}_R(t) = \{C\}$

The four cases are mutually exclusive; therefore, the overall probability of boarding R is equal to the sum of the probability of each case, as indicated by (3).

To calculate the probabilities in the summation in (3) we need to derive the probabilities that a bus leaves before or after a set of other buses. As to the former, consider a set of buses S alternative to n and assume without loss of generality $S = \{1, \dots, k\}$. In general, the probability that n departs after t and before any other bus in S can be calculated from the joint distribution of departure time probability as

$$\Phi_n(t,S) = \int\limits_{\tau_n \geq t; \tau_j > \tau_n, j=1,\dots,k} f(\tau_n, \tau_1, \dots, \tau_k)\, d\tau_n d\tau_1 \dots d\tau_k \tag{4}$$

Under the assumption that passengers perceive the departure times of different buses as statistically independent, (4) becomes[1]

$$\Phi_n(t,S) = \int_t^{+\infty} f_n(\tau_n) \cdot \prod_{m\in S} \big(1 - F_m(\tau_n)\big)\, d\tau_n \tag{5}$$

The integrand function used to calculate $\Phi_n(t,S)$ can be interpreted as the pdf of the event "bus n departs at $\tau_n > t$ and before any other bus in the set S", i.e.

$$\varphi_n(t, \tau_n, \mathrm{S}) = f_n(\tau_n) \cdot \prod_{m\in S} \big(1 - F_m(\tau_n)\big) \tag{6}$$

Note that, as expected, the probability of bus n departing before any other bus in the set S decreases when t increases. In fact

Proposition 1: $\Phi_n(t,S)$ is a non-increasing function of t

Proof: $\varphi_n(t, \tau_n, \mathrm{S}) \geq 0$ because it is a product of pdf and CDF. Applying the Leibniz's rule for integral differentiation

$$\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx = \int_{a(\theta)}^{b(\theta)} \partial_\theta f(x,\theta)dx + f(b(\theta),\theta)b'(\theta) - f(a(\theta),\theta)a'(\theta)$$

it follows

$$\frac{\partial \Phi_n(t,S)}{\partial t} = \int_t^{\bar{t}>t} \partial_t f_n(\tau_n) \cdot \prod_{m\in S} \big(1 - F_m(\tau_n)\big)\, d\tau_n + \varphi_n(\bar{\mathrm{t}}, \tau_n, \mathrm{S})\cdot 0 - \varphi_n(\mathrm{t}, \tau_n, \mathrm{S})\cdot 1 = -\varphi_n(\mathrm{t}, \tau_n, \mathrm{S}) \leq 0$$

which shows that $\Phi_n(t,S)$ is a non-increasing function of t. QED.

In general the probability that all buses in the set S have departed earlier than t is

$$\Xi(t,S) = \int\limits_{\tau_j < t, j=1,\dots,k} f(\tau_n, \tau_1, \dots, \tau_k)\, d\tau_n d\tau_1 \dots d\tau_k \tag{7}$$

Under the assumption of independence of the departures

$$\Xi(t,S) = \int\limits_{\tau_j < t, j=1,\dots,k} f(\tau_1, \dots, \tau_k)\, d\tau_1 \dots d\tau_k = \int_{-\infty}^{t} f(\tau_1) \dots \int_{-\infty}^{t} f(\tau_k)\, d\tau_k \dots d\tau_1 = \prod_{m\in S} F_m(t) \tag{8}$$

Intuitively such probability has to increase in time and in fact

Proposition 2: $\Xi(t,S)$ is a non-decreasing function of t

Proof: The proposition follows immediately by the fact that $F_m(t)$ are non-decreasing functions of t.

Let E and L denote two disjoint sets of buses not including n. Using (5) and (8) the joint probability that all buses in E depart before t, and n departs after t and before all buses in L is

$$\Xi(t,E) \cdot \Phi_n(t,L) \tag{9}$$

Using this result, (3) becomes

[1] In the following, we stick to the usual convention of considering an empty summation, i.e. a summation with no addend, equal to 0 and an empty product, i.e. a product with no factor, equal to 1.

$$\Pi_n(t) = \sum_{P_n(t)\in\mathcal{P}(A_n^-(t))} \Xi\big(t, P_n(t)\big)\cdot\Phi_n\big(t,\bar{P}_n(t)\big) \tag{10}$$

The pdf of the event "a passenger arriving at t departs at τ with bus n" is

$$\pi_n(t,\tau) = \sum_{P_n(t)\in\mathcal{P}(A_n^-(t))} \Xi\big(t, P_n(t)\big)\cdot\varphi_n\big(t,\tau,\bar{P}_n(t)\big) \tag{11}$$

Proposition 3: The probability of boarding a bus n is not influenced by the presence of services which can depart only after bus n has surely left the stop.

Proof: It is sufficient to prove that $\pi_n(t,\tau)$ is not influenced by a bus m that surely departs after n. In fact, $\Pi_n(t)$ is the integral of $\pi_n(t,\tau)$ over τ.Therefore if $\pi_n(t,\tau)$ is not affected by m, neither is $\Pi_n(t)$. We show now that the presence of m affects neither $\Xi\big(t,P_n(t)\big)$ nor $\varphi_n\big(t,\bar{P}_n(t)\big)$.

Case 1: $t > \max T_n$. Since $t > \max T_n \Leftrightarrow f_n(\tau_n|\tau_n \geq t) = 0$, it turns out that

$$\varphi_n(t,\mathrm{S}) = \int_t^{+\infty} f_n(\tau_n)\prod_{k\in S}\big(1-F_k(\tau_n)\big)\,d\tau_n = \int_t^{+\infty} 0\cdot\prod_{k\in S}\big(1-F_k(\tau_n)\big)\,d\tau_n = 0$$

Independent from $F_m(\tau_n)$.

Case 2: $t \leq \max T_n$. The hypothesis that a bus m can depart only after bus n has surely left the stop implies that $\min T_m < \max T_n$. It follows that $F_m(\tau_m|\tau_m \leq \min T_m < \max T_n) = 0$. Therefore

$$\varphi_n(t,\mathrm{S}) = \int_t^{+\infty} f_n(\tau_n)\prod_{k\in S}\big(1-F_m(\tau_n)\big)\,d\tau_n$$

$$= \int_t^{\max \mathrm{T}_n} f_n(\tau_n)\prod_{k\in S}\big(1-F_k(\tau_n)\big)\,d\tau_n = \int_t^{\max \mathrm{T}_n} f_n(\tau_n)\prod_{k\in S\setminus m}\big(1-F_k(\tau_n)\big)\,d\tau_n$$

The previous holds for any S so also for any $\bar{P}_n(t)$. Since m cannot be part of $P_n(t)$ for any $t \leq \max T_n$, the presence of m does not affect $\Xi\big(t,P_n(t)\big)$. Since m does not influence $\varphi_n\big(t,\bar{P}_n(t)\big)$ and $\Xi\big(t,P_n(t)\big)$, it does not influence their product $\pi_n(t,\tau)$. QED.

4.3. Expected waiting time

The waiting time of a passenger arriving at t and departing at τ is $w = \tau - t$. Since boarding different buses are mutually exclusive events, the probability that a passenger arriving at t departs at τ is equal to the sum of the probabilities that he departs with any of the available buses. Therefore the expected waiting time for a passenger arriving at t is

$$E\big(w(t)\big) = \int_t^{+\infty} (\tau - t)\sum_{\mathrm{n}\in B(t)} \pi_n(t,\tau)\,d\tau \tag{12}$$

and the expected waiting time corresponding to bus n

$$E\big(w_n(t)\big) = \int_t^{+\infty} (\tau - t)\,\pi_n(t,\tau)d\tau \tag{13}$$

Proposition 4: The overall expected waiting time is equal to the sum of the expected waiting times of all available services

Proof:

$$E\big(w(t)\big) = \int_t^{+\infty} (\tau - t)\sum_{\mathrm{n}\in B(t)} \pi_n(t,\tau)\,d\tau = \sum_{\mathrm{n}\in B(t)}\int_t^{+\infty} (\tau - t)\,\pi_n(t,\tau)\,d\tau = \sum_{\mathrm{n}\in B(t)} E\big(w_n(t)\big)$$

QED.

4.4. Passenger arrival distribution

The expected waiting time obtained from (12) cannot be used to specify the utility function of passengers deciding their arrival time at a stop as we need to consider the possibility of missing all buses after a certain $\bar{t}$, i.e. if $\exists \bar{t} | B(t) = \emptyset \forall t \geq t$ or, more commonly, the possibility that the passenger aims to board before a time $\bar{t}$ and attaches higher disutilities to later departures. To solve this shortcoming we consider

$$W(t) = E\big(w(t)\big) + \left(1 - \sum_{n \in B(t)} \Pi_n(t)\right) \mathrm{c} \tag{14}$$

where c is a constant representing the aversion to the possibility of missing the last attractive boarding opportunity. We further assume that the utility function has the form

$$U(t) = \alpha\big(W(t)\big)^{\beta} \tag{15}$$

and that the arrival time choices can be described by the continuous logit model

$$q(t) = -\frac{e^{U(t)}}{\int_{t_1}^{t_2} e^{U(t)} dt} \tag{16}$$

Where $[t_1; t_2]$ is an interval spanning all the possible departure times of the buses considered by passengers. We note that different specifications of the anticipated risk-averse expected value are possible. For instance, an optimistic passenger may choose the arrival time with

$$W_{opt}(t) = \min\big\{E_n\big(w(t)\big) | 1 - F_n(t) > 0\big\} \tag{17}$$

i.e. considering the bus with the minimum waiting time among the buses that can depart after t. Instead, the utility function of a pessimist could be based on the maximum waiting time among those buses that can be surely boarded, i.e.

$$W_{pes}(t) = \max\big\{E_n\big(w(t)\big) | 1 - F_n(t) > 0\big\} \tag{18}$$

Our reliability-based passenger arrival model is illustrated in Fig. 2 for the case of three buses with the triangular pdf as in (a). We assume that the modal value is the scheduled departure time of each bus (in general, the modal value may differ from the scheduled departure time, for instance when passengers think that the buses are systematically delayed). Note that perceived departure times earlier than the scheduled ones, although normally unlikely, cannot be completely ruled out, for instance because bus drivers may not comply with the schedule, or because of discrepancies between the time of the passenger and the time of the bus system.

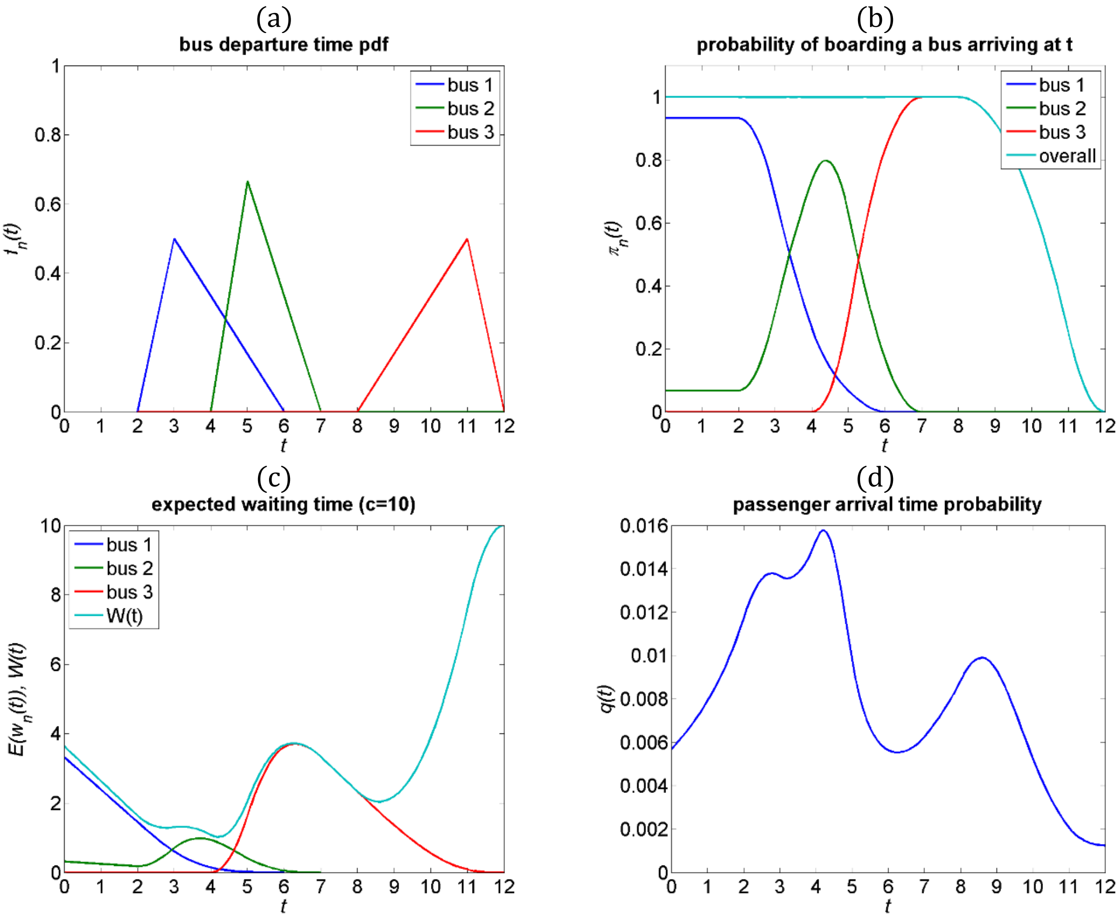

Fig. 2. Illustration of the reliability-based passenger arrival model

In (a) it can be seen that bus 1 is scheduled to depart before bus 2, but it can be overtaken by bus 2. Instead bus 3 departs surely after buses 1 and 2. Fig. 2(b) shows that, because of the possibility that bus 1 is later than bus 2, the perceived probability of boarding bus 1 (and analogously that of bus 2) is always smaller than 1. The probability of boarding 1 and 2 are constant while the pdf of bus 1 is nil, i.e. until the first possible departure of bus 1. After that, the chances of missing bus 1 increase, so the probability of boarding it decreases and instead the probability of boarding bus 2 increases. Analogously, the perceived probability of boarding bus 3 is nil until the moment when the passenger is sure to be able to board at least bus 2, then it starts increasing. Different from buses 1 and 2, the probability of boarding bus 3 reaches 1, because between times 7 and 8 it has not departed for sure and it cannot be surpassed by any other bus. The overall probability of boarding at least one bus is equal 1 until the possibility of missing bus 3 arises. Fig. 2(c) shows the expected waiting times of each bus and the corresponding overall expected waiting time. After time 7, the probability of catching a bus coincides with the probability of boarding bus 3 and therefore the overall waiting time is equal to the expected waiting time of bus 3. Such expected waiting time decreases when the probability of boarding bus 3 decreases. Hence, without the correction suggested in (14), the perceived expected waiting time of a passenger arriving at $t > 8$ would decrease with the paradoxical consequence that the probability of arrival would increase when the chances of missing the last service increase. In Fig. 2(d) it is evident that the arrivals when service reliability and risk aversion are considered are far from being uniform as commonly assumed for bus load models.

5. Bus propagation model

In the following, we develop a bus bunching model consistent with the non-uniform arrival process. Consider a common scenario of a route served by buses of different lines which serve the same destinations. To avoid the issue of having to consider different passenger destinations and hence different attractive sets, we assume that the buses might originate from different terminals, then merge in the city centre and all are destined for the same terminal. As in Newell and Potts (1964), the undisturbed trajectories of the buses are described by the following bus propagation equations

$$\begin{cases} \tau_{in}^{a} = \tau_{i-1n}^{d} + v_{i-1n} \\ \tau_{in}^{d} = \tau_{in}^{a} + D_{in} + \delta_{in} \end{cases} \tag{19}$$

Assuming that v_{in} and the scheduled departure times of the buses from the first stop (τ_{0n}^{d}) are known, (19) can be solved recursively if a formulation of D_{in} is available as a function of τ_{jl}^{d} and τ_{kl}^{a}, with $j < i; k \leq i; l \leq n$. Remember that δ_{in} is an exogenously determined delay, and so it is considered known in (19).

5.1. Newell & Potts model

In their bus propagation model, Newell and Potts assume that passenger arrival and boarding rate are time-independent, and that the boarding process follows the law

$$D_{in} = \rho\left(\tau_{in}^{d} - \tau_{in-1}^{d}\right) \tag{20}$$

where ρ is the (time-independent) ratio between the passenger arrival and boarding rates. That is, they consider boarding as a stationary unsaturated deterministic queuing process with constant degree of saturation. They also assume that the bus departs as soon as the queue of waiting passengers vanishes, independent of any schedule or holding policy.

5.2. Our model

We retain the hypotheses that boarding can be represented as a stationary unsaturated deterministic queue, and that buses depart immediately after boarding all waiting passengers. In line with Section 4 we relax though the assumption that the degree of saturation is constant. Eq. (20) can be regarded as a particular case of the general relationship (21) regulating an unsaturated deterministic boarding process with time-dependent arrival and boarding rates:

$$\int_{\tau_{in-1}^{d}}^{\tau_{in}^{d}} [q_i(t) - b(t)]dt = \int_{\tau_{in-1}^{d}}^{\tau_{in}^{d}} q_i(t)dt - \int_{\tau_{in}^{a}}^{\tau_{in}^{d}} b_i(t)dt = 0 \tag{21}$$

Note that evidently $b_i(t) = 0$ between τ_{in-1}^{d} and τ_{in}^{a}. We consider that passengers do not arrive uniformly but they follow the distribution derived from (16). We further assume that the boarding rate is constant which appears to be a reasonable assumption since the rate is mainly linked to vehicle characteristics and the ticketing system. Under our hypotheses, (21) simplifies to

$$\int_{\tau_{in-1}^{d}}^{\tau_{in}^{a}+D_{in}} q_i(t)dt - bD_{in} = 0 \tag{22}$$

In our formulation, $q_i(t)$ does not depend on the actual bus arrival and departure times but only on passenger perception of the pdf of the bus departure times. Therefore it is exogenous to the bus propagation model, i.e. it is known in (22). If τ_{n}^{a} and τ_{n-1}^{d} are known, (22) can be easily solved with the iterative approach described in Table 2. Note that the algorithm works also if the boarding rate is not constant but it can be known once τ_{jl}^{d} and τ_{kl}^{a}, with $j < i; k \leq i; l \leq n$ are known. Since D_{in} in (22) is a function of τ_{in-1}^{d} and τ_{in}^{a}, (19) and (22) can be easily used to derive the bus trajectories from τ_{0n}^{d}.

Table 2 Algorithm to calculate D_{im}

Initialisation	$\delta t \leftarrow$ small time step $m \leftarrow 1$	
Main	$Q \leftarrow \int_{\tau_{in-1}^{d}}^{\tau_{in}^{a}+n\,\delta t} q_i(t)dt$	Obtain number of passengers who arrived between τ_{in-1}^{d} and $\tau_{in}^{a} + m\delta t$
	$B \leftarrow \int_{\tau_{in-1}^{d}}^{\tau_{in}^{a}+m\,\delta t} b_i(t)dt = bm\delta t$	Obtain number of passengers who boarded between τ_{in-1}^{d} and $\tau_{in}^{a} + m\delta t$
	If $Q \leq B$ then $D_{in} \leftarrow m\delta t$ otherwise $m \leftarrow m + 1$ and repeat *Main*	

6. Bus bunching with passenger reliability-based behaviour

To illustrate the effects of reliability-based passenger arrivals on bunching, we compare the solutions provided by our model of bus propagation to those given by the Newell and Potts model in an analogous situation. The exemplification is carried out by varying some parameters of the basic scenario described in Table 3.

Table 3 Case study, base parameter settings

Bus service

- 6 buses, 10 stops (including the depot)
- Scheduled headway: $\tau_{0n+1}^{d} - \tau_{0n}^{d} = 10\text{min}, n = 1, \ldots, 5$
- Travel time: $v_{in} = 3\text{min}, i = 0, \ldots, 8, n = 1, \ldots, 6$
- $\delta_{in} = 0\text{min}, i = 0, \ldots, 9, n = 1, \ldots, 6$

Reliability-based arrival model

- Perceived pdf of departure time at all stops: triangular function with $\min \mathrm{T}_n = \bar{\tau}_n - 1\text{min}, \max \mathrm{T}_n = \bar{\tau}_n + 2\text{min}$ where $\bar{\tau}_n$ is the scheduled departure time. The scheduled departure time coincides with the departure time predicted by the Newell and Potts model when no delay occurs.
- To avoid the influence of boundary conditions and to have periodic demand, the considered six buses do not include the first and the last service along the route. The passenger arrival probability function in the basic scenario is shown in Fig. 3.
- $\alpha = -1, \beta = 0.55$ as in Bowman and Turnquist (1981)
- At each stop
    - Passengers start arriving one scheduled headway (i.e. 10min) before the scheduled departure of the first bus from the stop (so that there are passengers waiting for the first bus at the stop);
    - All passengers have arrived before the schedule departure of the last bus;
    - Therefore passenger arrival simulation time: $\mathcal{T} = \text{number of buses} * \text{scheduled headway} = 60\text{min}$
- Total demand: $Q = 100\text{pax}$ when not specified differently

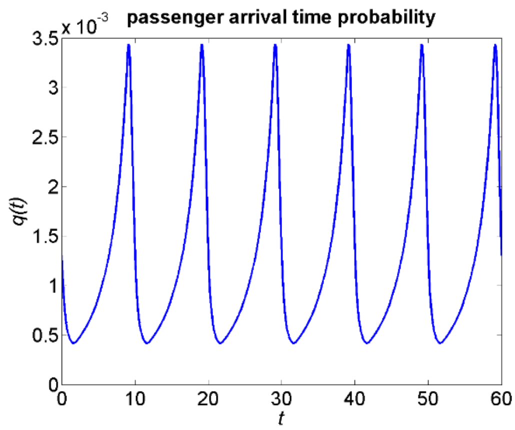

Fig. 3. Arrival patterns under basic demand conditions

In the following, to allow for easy comparison of our model with Newell and Potts model, in the latter we assume that the Q passengers arrive uniformly during the simulation period $\mathcal{T}$ and so we set

$$\rho = {q^{NP}}/{b} = {\left({Q}/{\mathcal{T}}\right)}/{b}$$

In the figures showing bus trajectories, red lines represent the results of the Newell and Potts model (with "uniform" arrival of passengers, blue lines the results of our model (the arrival distribution is indicated in the legend of each figure).

6.1. Effect of boarding rate

In Fig. 4 we present the results provided by the two bunching models with different boarding rate values, namely b=11.11 [2] and b=2.78 pax/min.

Since no random delay is considered, the Newell and Potts model predicts no bunching. When a variable passenger arrival rate is considered, a regular service is provided for high values of boarding rates (Fig. 4(a)). But when the boarding rate is low, the service can be severely disrupted also without exogenous delays (Fig. 4(b)). Note that our bus propagation model, as well as the Newell and Potts one, is not adjusted to deal with cases in which overtaking occurs. In case of low saturation rates ρ (as mostly expected in reality) this does not lead to large errors as the overtaking and overtaken buses will remain bunched, however, for large ρ this might lead to some more significant errors in the trajectories of, both, overtaken and overtaking buses.

To avoid the issue and exclude overtaking, one could introduce a condition where buses do not depart prior to the departure of the previous bus in the main loop in Table 2. This adjustment can be easily implemented. To explicitly deal with overtaking, further adjustments are needed in (21) and the main loop in Table 2 In the present paper, we decided to avoid both adjustments to allow for comparability with the Newell and Potts model and to remain the focus of this paper.

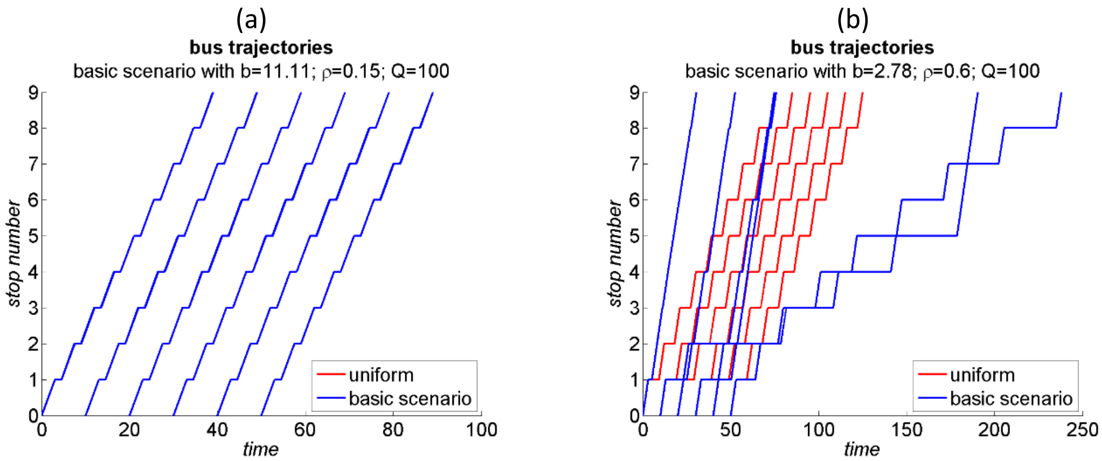

Fig. 4. Effect of boarding rate; no exogenous delay

The trajectories in Fig. 4(b) can be explained by looking at the cumulative passenger arrival and boarding functions. Fig. 5 shows cumulative arrivals and boarding for two different values of boarding rates. The two scenarios share the same scheduled headway (10 min). Since we assume the schedule equal to that predicted by the N&P model with no delay, in Fig. 5(b) the first scheduled arrival occurs at minute 4 and the following ones every 10 minutes after; each bus leaves the stop after 6 minutes dwell time (corresponding to ρ=0.6). In (b), with ρ=0.15, the first scheduled arrival is at minute 8.5 and the following ones every 10 minutes after; the dwell time of each bus is 1.5 minutes.

Both the Newell and Potts model and our model assume that a bus leaves the stop as soon as all the waiting passengers have boarded, i.e. when the cumulative boarding function intersects the cumulative arrival function. As illustrated in both Fig. 5(a) and 5(b), in the case of uniform passenger arrival the first departure happens at time

[2] We developed our case study by setting the value of ρ, and then deriving the boarding rate corresponding to the assumed total demand. This explains the "unusual" value of b. Although it has not been possible to retrieve literature on the boarding rates, 11.11 pax/min seems realistic for passengers using travel cards.

10 and then every 10 minutes (when the red line crosses the blue one. In Fig. 5(a) the cyan and red lines overlap), i.e. the service operates according to schedule. Conversely, if passengers arrive considering the schedule as well as system unreliability (basic scenario) and the boarding rate is low (Fig. 5(b)), at each instant the cumulative number of passenger arrivals is smaller than in the case of uniform arrivals. Therefore, when the first bus arrives at time 4, it has to board fewer passengers and it leaves earlier, in the figure at the time in which the cyan line crosses the green one. This earlier departure triggers bunching.

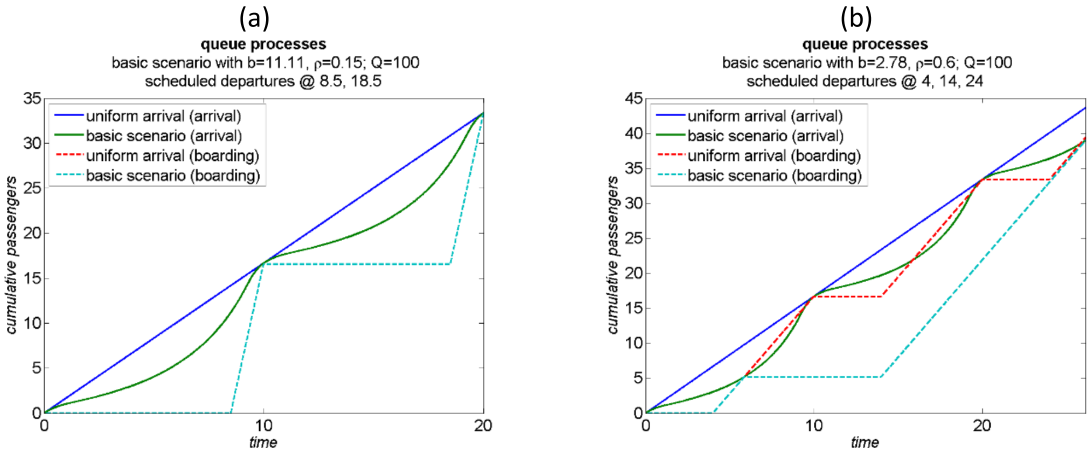

Fig. 5. Queue processes under different scenarios

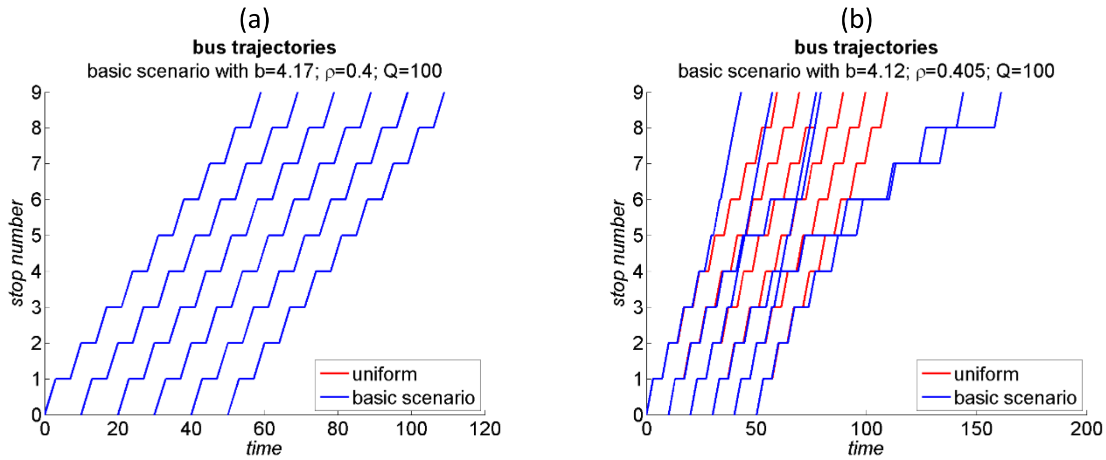

Fig. 6. Effect of small positive (a) and negative (b) deviations from b_{crit}; no exogenous delay

Note that the case in Fig. 5(b) is a case of serious bunching. In fact, assuming that all waiting passengers board the bus arriving at time 14, this bus would leave at 26, i.e. after the third bus is due at the stop. Of course, the possibility for passengers to choose between two buses can alter the queuing process shown in Fig. 5(b) but this clearly exceeds the scope of the analysis here. Fig. 5(a) shows also that bunching is not triggered by non-uniform arrival if b is high enough, In other words, there is a critical value b_{crit} such that non-uniform passenger arrivals generate instability for $b \leq b_{crit}$ whereas they do not compromise stability for $b > b_{crit}$. Even a small negative deviation from b_{crit} can induce large perturbations: For example in Fig. 6 the difference between b in (a) and (b) is only 0.05pax/min. The existence of the critical value is clearly linked to the shape of the passenger arrival function: suffice it to notice that if the green line was above the blue one in Fig. 5(b), the cyan line would not intersect it before time 10 and bunching would not arise.

6.2. Effect of total demand

In the Newell and Potts model the boarding rate does not appear explicitly but is linked to the arrival rate via the degree of saturation ρ. It follows that an increase of the total demand Q results in a lower saturation rate, assuming that the boarding rate stays constant. Therefore higher total demand does not lead to bunching in the Newell and Potts model unless the saturation rate is assumed to be bus specific. In our model, the boarding and arrival rates are represented explicitly and enter the model separately. The consequence is that, as illustrated in Fig. 7, bus bunching can arise in case the demand is different from the one planned for by the operator, even if the boarding rate remains constant (b=11.11pax/min) and there is no exogenous delays. Note that bunching arises for values of total demand both lower (Q=75pax, Fig. 7(a)) and higher (Q=125pax, Fig. 7(b)) than the "design" one i.e. that used to derive the scheduled departure times. In the former case, bunching is triggered by an early departure of the first bus, in the latter by a delay of the same bus. Again the phenomenon can be explained by considering the cumulative arrival and boarding functions. Increasing the total demand while the boarding rate stays constant means that, in Fig. 5(a), the cyan line does not change whereas at each instant the green line is below (if the demand is smaller) or above (if the demand is greater) that in the figure. Therefore they can intersect before (in case of smaller demand) or after (in the opposite case) the first scheduled departure time.

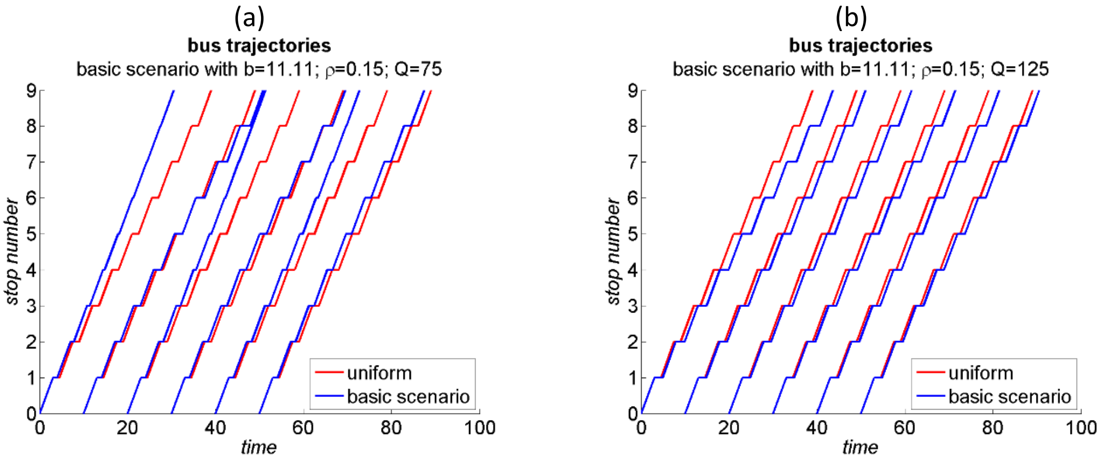

Fig. 7. Effect of total demand; no exogenous delay

6.3. Effect of exogenous delay

In the Newell and Potts formulation, bunching arises only when an exogenous delay occurs at any stop. In the following we study what happens when the second bus is delayed at stop 1, with b=11.11pax/min and Q=100pax (Fig. 8). Fig. 8(c) to (f) show system performance measures for two values of delay. The bars in Fig. 8(c) and 8(d) represent the mean of the absolute deviations of the headways from the scheduled ones calculated over all buses at each stop. The bars in Fig. 8(e) and 8(f) instead show the mean of the absolute headway deviations from the schedule calculated over all stops for pairs of subsequent buses. Crosses hatch the bars when not all buses (stops) are used in the calculations because the models predict overtaking and so the results from the moment of the overtaking on are not fully comparable. For the sake of comparison, the same subsets of buses (stops) are used to calculate the means for both passenger arrival processes. It is evident that larger initial delays give rise to larger deviations, and that bunching becomes a more significant problem at stops further from the depot. The performance of the system is more or less similar under the two arrival processes at the initial stops. In fact, the schedule deviations are slightly lower if passengers consider service reliability. However, the headway deviations in case of reliability-based passenger arrivals are greater than those occurring in case of uniform arrival for stops further downstream. The differences between the two processes diminish for larger initial delay. In Fig. 8(e) and 8(f) it can be observed that irregularities at the beginning increase but then tend to reduce. In both cases, the system performs better under reliability-based passenger arrivals for the first runs. The better performance of non-uniform passenger arrivals in the case of the first delayed bus can be easily explained by considering the cumulative number of passengers arriving at bus stops shown in Fig. 5. At each instant the cumulative passenger number in the basic scenario is smaller than that under the uniform arrival assumption, except at time 0 and time 100. Therefore, when bus 2 experiences a delay at stop 1, the additional passengers waiting to board at stop 2 will be fewer in the case of reliability-based arrivals than in the case of uniform arrivals, and so bus 2 is less delayed along its route. The difference of performance for later runs between the two arrival processes is attenuated for the larger delay in this case as well. For later runs, the headways tend to be less different from the scheduled ones if passengers arrive uniformly. Note that deviation peaks for intermediate runs, in particular, in the case of $\delta_{12} = 1.5$, the maximum headway deviation is between the second and the third bus, when the performance with non-uniform arrivals is still better than with uniform ones. In the latter scenario, the deviation between bus 2 and 3 is so large that overtaking occurs, i.e. the system experiences a very serious distortion at stop 8. In contrast, there is no overtaking if the arrivals are non-uniform. This shows that despite the system underperforming on average with non-uniform arrivals, consideration of system reliability in passenger arrival times avoids the occurrence of extreme bunching conditions in the presence of large delays.

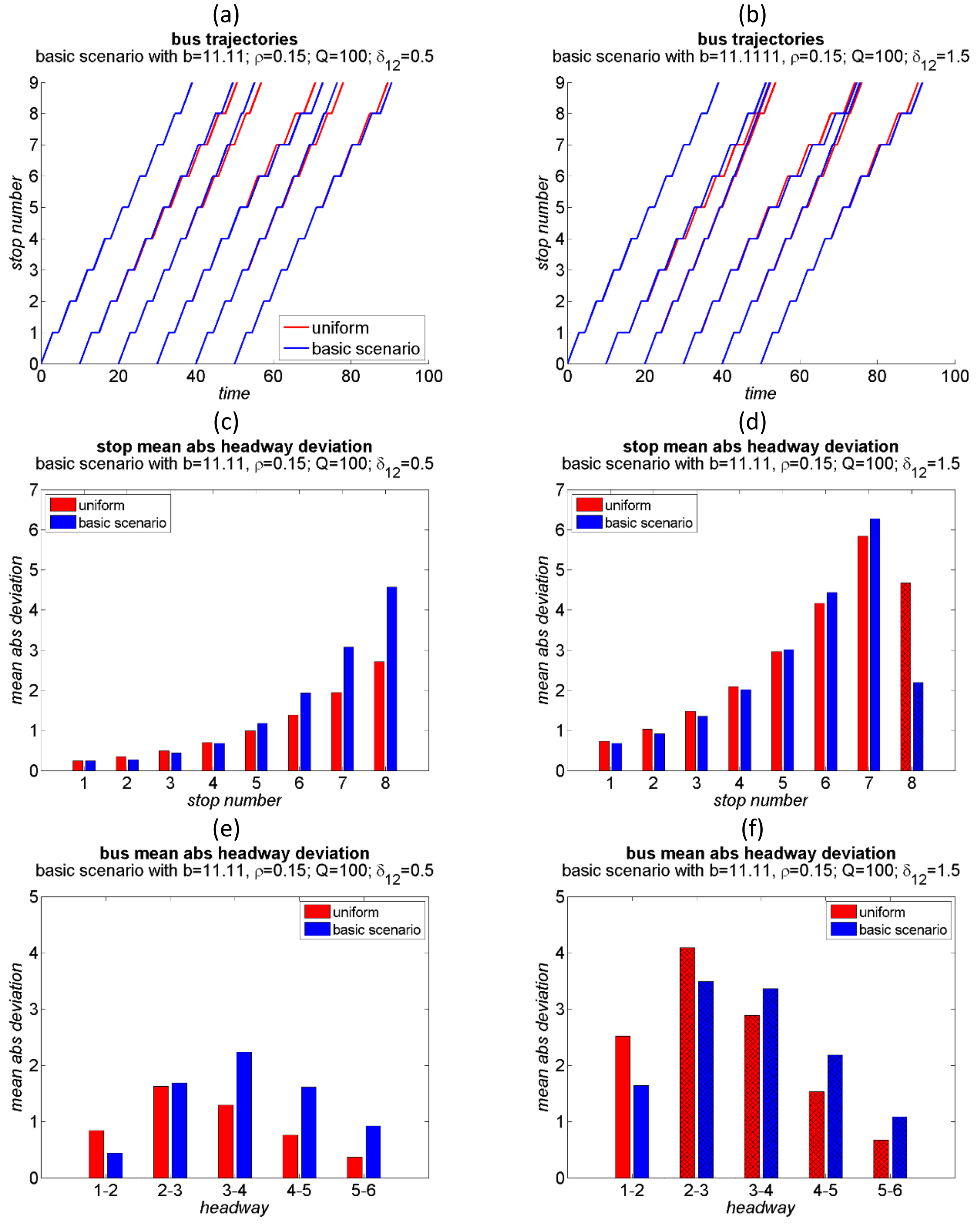

Fig. 8. Effect of exogenous delay

6.4. Effect of passenger perceptions and preferences

The passenger arrival distribution depends on the perceived pdf of the bus departure times, linked to the system reliability, and to the passenger risk-aversion, measured by parameters α and β.

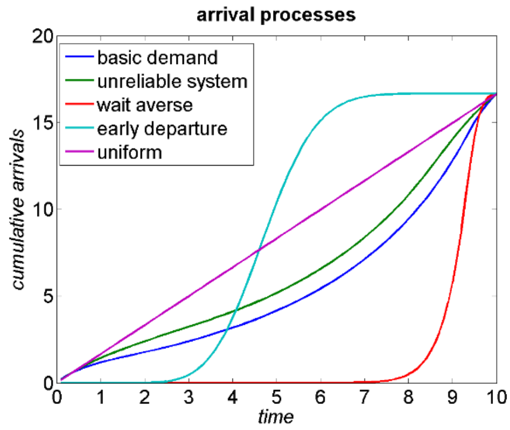

Fig. 9. Arrival processes for different values of system reliability and risk-aversion

In the following we analyse the effects of three different arrival profiles (the cumulative arrival distributions under the different scenarios are represented in Fig. 9 for one scheduled headway), namely

- Unreliable system: The perceived unreliability is higher than in the basic scenario, i.e. triangular pdf functions are considered by the passengers with $\min \mathrm{T}_n = \bar{\tau}_n - 2, \max \mathrm{T}_n = \bar{\tau}_n + 4$. Therefore passengers anticipate that bus arrivals can occur along a wider interval, therefore their arrivals tend to be more spread as well and the distribution is more similar to that predicted under the uniform arrival assumption.
- Wait averse: The aversion to wait ($\alpha = 1.55$) is higher than in the basic scenario. High expected wait times have higher disutilities under this scenario than under the basic one. The opposite holds for low expected wait times. Therefore passengers tend to avoid both very early arrivals – because they might give rise to long waits for the intended buses – and very late arrivals – because the might wait long for the next bus, in case the intended one is missed. The consequence is that passenger arrivals tend to be concentrated around a specific time.
- Early departures: In this scenario we consider passengers who fear the possibility of very early departures but not of late departures, so that we set $\min \mathrm{T}_n = \bar{\tau}_n - 8, \max \mathrm{T}_n = \bar{\tau}_n$. Because of high aversion to long wait, arrivals tend to be concentrated as in “wait averse”. Given the pdf of the bus departure times, the arrival time is earlier than in “wait averse”. As a result, there are times in which the cumulative arrivals are higher than under uniform arrivals. This scenario is unlikely in reality, where normally buses cannot leave long before the scheduled departure time. We add this scenario mainly in order to illustrate the effect of a partly concave cumulative arrival function.

Fig. 10 shows the bus trajectories arising from the three profiles assuming the same exogenous delays of bus 2 at stop 1 (and same b and Q) considered in Fig. 8. Since the Newell and Potts results depend only on the value of the exogenous delay, the red trajectories are the same in (a), (c), (e) and in (b), (d), (f), and they coincide with those in Fig. 8(a) and (b) respectively. Under the “unreliable system” scenario the behaviour of the system ((a) and (b)) is more similar to that with uniform arrivals than that illustrated in Fig. 8, as it can be anticipated by the fact the arrival distributions tend to be more uniform.

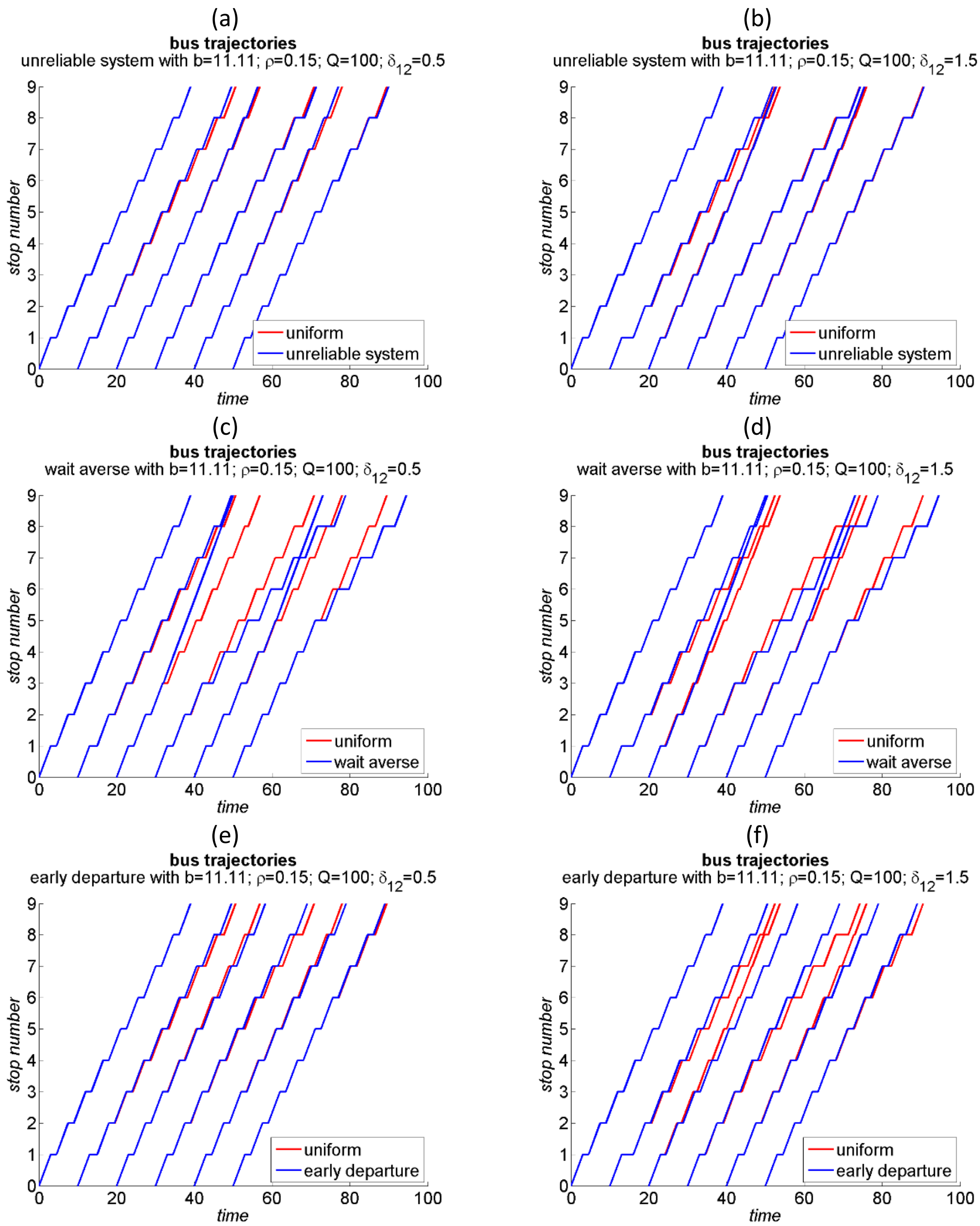

Fig. 10. Effect of passenger perceptions and preferences

The worst performance of the system occurs under “wait averse” arrival patterns, i.e. when passenger arrivals are concentrated near the scheduled departure time. In this case, even a small delay gives rise to large headway deviation (Fig. 11(a)).

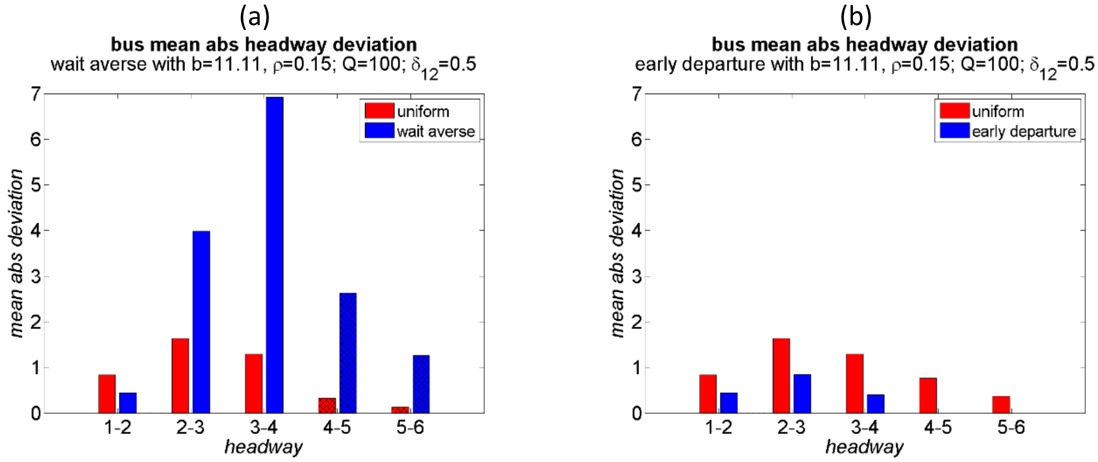

Fig. 11. System performance under different values of system reliability and risk-aversion

When real time information on bus arrival/departure is available, one can expect that most passengers using the information arrive at stops very close to the advised time. Given the result obtained for "wait averse", the impact of this behaviour in the presence of unpredicted bus delays should be carefully examined. Under "early departure" the system is only little disturbed, because in the first instants of the scheduled headway there is almost no passenger arrival and therefore a delayed bus faces a boarding situation very similar to the case with no delay. In this case, the system performs even better than with uniform passenger arrivals (Fig. 11 (b)).

7. Conclusions

In this paper we discussed the relationship between passenger bus stop arrival patterns and dwell times of buses that might lead to bunching effects. We advanced the Bowman and Turnquist (1981) arrival model by developing a "reliability-based arrival pattern model" in which possibly overlapping perceived departure time probability density functions for several buses are considered. Passengers hence consider the schedule as well as possible early or late departures in order to minimize their expected waiting time. We then implemented these behavioural assumptions into a bus propagation model similar to that of Newell and Potts (1964) but in which passenger arrival and boarding rates enter the model separately. In our case study we compared our model to the assumption of uniform arrival as commonly done in the literature on bus bunching.

Throughout the paper we assume that after the release of the buses from the depot the buses are uncontrolled. That is, buses leave stops whenever there are no more waiting passengers. Holding points or headway equalizing strategies are not considered. This allows us to illustrate how a mismatch between the operators' perception of service demand and actual demand can lead to bunching effects even without exogenous delays. Remarkably bunching occurs also for values of demand lower than the planned one. Bunching generated by unexpected demand cannot be derived with previous models, which assumed uniform demand. As all holding related control measures introduce inefficiency for the operator into the service we therefore suggest that our model shows the importance for operators to correctly assess the demand and resulting dwell time at stops.

Another main finding is that non-uniform arrival patterns can lead to more severe bunching effects over time. This highlights the need for control measures. We believe therefore that the model developed here could be used to test the effectiveness of various control strategies, i.e. that it is worth revisiting literature discussed in our review section with consideration of the behavioural issues discussed in this paper. An alternative helpful implication from this research for operators might be to aim to smoothen the arrival pattern of passengers if on-time arrivals cannot be guaranteed. That is, if exogenous delays are feared and boarding rates are substantial it might be worth for operators to consider making notes to schedules that slightly early or late departures are possible in order to encourage a more spread arrival pattern of passengers.

Our approach is based on the assumption that passengers are aware of a "static" service schedule information. In the introduction we already mentioned the importance of real time information (RTI) to our topic as it will encourage even more peaked arrival time patterns also for services with short headways. Modelling the effect of upstream delays on downstream arrival patterns we leave though as further work. Not only due to the modeling challenges (consideration of RTI requires a dynamic modelling framework) but also since it is not clear in how far RTI is (and should be) trusted by passengers from available literature. Possibly, a simple Bayesian approach could be investigated to model behaviour once passengers obtain information on the next departure(s).

In this paper, we have considered an unsaturated boarding process, assuming that buses can always accommodate all passengers waiting to board. This might not be true in very busy bus systems where capacity constraints affect system operations. On the supply side, consideration of bus capacity is equivalent to constrain the dwell time. In addition, the interaction between boarding and on-board passengers may reduce the boarding speed at high levels of bus occupancy. Therefore, system performance forecast requires knowledge about origin-destination matrices, so that the number of on-board passengers can be derived in each bus trip leg. On the demand side, one can expect that at least some categories of passengers will try to avoid demand peaks; hence, risk aversion may change in time in the presence of time-dependent demand. Finally, the utility of a given arrival time should include the level of comfort, which may vary considerably as a function of crowdedness.

We note further the effect of electronic ticketing. Through smart card technologies, boarding has become faster which possibly reduces the bunching effect (see a review in Pelletier et al., 2011). At the same time though more efforts are made by bus operators to cater also for the needs of population groups with special needs such as wheelchair users. These require additional boarding times, leading to more variability in the boarding rates. Therefore, one possible further work direction is to consider non-constant boarding rates b where boarding per passenger in off-peak times might take longer when less passengers with travel cards and proportionally more passengers with special needs travel (Khoo, 2013). More generally, one might consider different passenger classes with different boarding rates as also considered in Bowman and Turnquist (1981) who distinguish a group of randomly arriving passenger and a group of optimizing passengers.

There are various other directions in which this work could be extended, partially already mentioned within previous sections but we believe worth summarising here. Firstly, developing a bus propagation model that, different from Newell and Potts (1964), gives a realistic representation of the system operations after once overtaking between buses has taken place. Secondly, more complex network structures with common lines should be considered. In combination with the work presented here, this could lead to conclusions in how far network structure and information to passengers could avoid the need for or change the characteristics of holding strategies. Thirdly, better understanding of passenger behaviour is needed. We take the parameters α and β in our model from Bowman and Turnquist (1981) and showed that these values affect the service quality which might justify further calibration efforts. Furthermore, currently we assume that passengers consider all the possible departures from the stop, but some passenger groups might be interested only in a subset of the services as indicated in Schmöcker et al. (2013). Related to this, the value of the cost of missing the last desired service has been simply set to a large value in our model. Furthermore, also the shape of the perceived departure probabilities needs calibration. Finally, obviously the bus choice model could be made more complex by considering crowdedness on buses and expectations on residual travel times for different lines (similar to Gentile et al., 2005).

Acknowledgements

This paper is partly supported by the UK Rail Safety and Standards Board Project RSSB-T1071, DITTO. In addition, we would like to thank the editor and reviewers for their critical comments and constructive suggestions.

References

Bartholdi, J.J., Eisenstein, D.D., 2012. A self-coordinating bus route to resist bus bunching. Transportation Research, Part B, 46, 481–491.

Bowman, L.A., Turnquist, M.A., 1981. Service frequency, schedule reliability and passenger wait times at transit stops. Transportation Research, Part A, 15(6), 465-471.

Boyd, C.W., 1987. Notes on the theoretical dynamics of intermittent public passenger transportation systems. Transportation Research, Part A, 17(5), 347-354.

Cats, O., Nabavi, A., Larijani, Burghout, W., Koutsopoulos, H.N., 2012. Impacts of holding control strategies on transit performance: A bus simulation model analysis. Transportation Research Record, 2584, 51-58.

Crout, D.T., 2007. Accuracy and precision of TriMet's Transit Tracker system. Transportation research board 86th annual meeting, Washington, DC, January 2007.

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, 267–277.

Eberlein, X.J., Wilson, N.H.M., Bernstein, D., 2001. The holding problem with real-time information available. Transp. Sci., 35(1), 1-18.

Gao, Z, Sun, H, Shan, L.L., 2004. A continuous equilibrium network design model and algorithm for transit systems. Transportation Research, Part B, 38, 235-250.

Gentile, G., Nguyen, S., Pallotino, S., 2005. Route Choice on Transit Networks with Online Information at Stops. Transportation Science 29(3), 289-297.

Gibson, J., Baeza, I., Willumsen, L.G., 1989. Bus-stops, congestion and congested bus-stops. Traffic Eng. & Control, 30(6), 291-302.

Hadas, Y., Ceder, A. 2010 Optimal coordination of public-transit vehicles using operational tactics examined by simulation. Transportation Research, Part C, 18, 879-895.

Hickman, M.D., 2001. An analytic stochastic model for the transit vehicle holding problem. Transportation. Science, 35(3), 215-237.

Hickman, M.D., Wilson, N.H.M., 1995. Passenger travel time and path choice implications of real-time transit information. Transportation Research, Part C, 3(4), 211-226.

Hollander, Y., Liu, R., Mackie, P., 2007. Application of a departure time choice model in a cost-benefit analysis. Proceedings of the European Transport Conference. Leiden, October 2007.

Hollander, Y., Liu, R., 2008. Estimation of the distribution of travel times by repeated simulation. Transportation Research, Part C, 16, 212-231.

Khoo, H.L., 2013. Statistical Modeling of Bus Dwell Time at Stops. Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 9.

Larson, R.C., Odoni A.R., 1981. Urban operations research. Prentice-Hall, NJ.

Lemp, J.D., Kockelman, K.M., Damien, P., 2010.The continuous cross-nested logit model: Formulation and application for departure time choice. Transportation Research. Part B, 44, 646–661.

Liu, R., Sinha, S., 2007. Modelling urban bus service and passenger reliability. Proceedings of the International Symposium on Transportation Network Reliability. The Hague, July 2007.

Liu, Z., Yan, Y., Qu, X., Zhang, Y. 2013. Bus stop-skipping scheme with random travel time. Transportation Research, Part C, 35, 46-56.

Luethi, M., Weidmann, U., Nash, A., 2007. Passenger arrival rates at public transport stations. Transportation Research Board 86th annual meeting (No. 07-0635), Washington, DC, January 2007.

Meng, Q., Qu., X. 2013. Bus dwell time estimation at bus bays: A probabilistic approach. Transportation Research, Part C, 36, 61-71.

Muñoz, J.C., Cartés, C.E., Giesen, R., Sáez, D., Delgado, F., Valencia, F., Cipriano, A., 2013. Comparison of dynamic control strategies for transit operations. Transportation Research, Part C, 28, 101-113.

Nagatani, T., 2001. Interaction between buses and passengers on a bus route. Physica A, 296, 320-330.

Nesheli, M.M., Ceder, A. 2014. Optimal combinations of selected tactics for public-transport transfer synchronization. Transportation Research, Part C, 48, 491-502.

Newell, G.F., 1974. Control of pairing of vehicles on a public transportation route, two vehicles, one control point. Transportation Science, 8(3), 248-264.

Newell, G.F. Potts, R.B., 1964. Maintaining a bus schedule. In: Proceedings of the 2nd Australian Road Research Board, 2, 388-393.

Noekel, K., Wekeck, S., 2009. Boarding and Alighting in Frequency-Based Transit Assignment. Transportation Research Record, 2011

Osuna, E.E. Newell, G.F., 1972. Control strategies for an idealized bus system. Transportation Science, 6(1), 52-71.

Pilachowski, J.M., 2009. An Approach to Reducing Bus Bunching. PhD Thesis, University of California, Berkeley.

Pelletier, M-P., Trépanier, M., Morency, C. 2011. Smart card data use in public transit: A literature review. Transportation Research, Part C, 19(4), 557-568.

Powell, W.B., Sheffi, Y., 1983. A probabilistic model of bus route performance. Transportation Science, 17(4), 376-404.

Quarmby, D.A., 1967. Choice of Travel Mode for the Journey to Work: Some Findings. Journal of Transport Economics and Policy, 1(3), 273-314.

Schmöcker, J.-D., Kurauchi, F., Shimamto, H., 2013. Generation and Calibration of Transit Hyperpaths. Transportation Research, Part C, 36, 406-418.

Schmöcker, J.-D., Sun, W., Liu, R., Fonzone, A., 2015. Bus Bunching Along a Corridor Served by Two Lines. Submitted to the 6th International Symposium on Transportation Network Reliability (INSTR), August 2015, Nara, Japan.

Sorratini, J., Liu, R., Sinhan, S., 2008. Assessing bus transport reliability using microsimulation. Transport Planning & Technology, 31(3), 303-324.

Spiess, H., Florian, M., 1989. Optimal strategies: A new assignment model for transit networks. Transportation Research, Part B, 23, 83–102.

Watkins, K.E., Ferris, B., Borning, A., Rutherford, G.S., Layton, D., 2011. Where Is My Bus? Impact of mobile real-time information on the perceived and actual wait time of transit riders. Transportation Research, Part A. 45, 839–848.

Xuan, Y., Argote, J. Daganzo, C.F., 2011. Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. Transportation Research, Part B, 45, 1831-1845.