Bus Holding Control of Running Buses in Time Windows

Dr Konstantinos Gkiotsalitis
University of Twente
Center for Transport Studies (CTS)
Department of Civil Engineering
P.O. Box 217
7500 AE Enschede
The Netherlands
Email: k.gkiotsalitis@utwente.nl

Dr Oded Cats
Delft University of Technology
Department of Transport and Planning
P.O. Box 5048
2600 GA Delft
The Netherlands
Email: o.cats@tudelft.nl

6247 words + 5 figures + 0 tables = 7497

Transportation Research Board, 98th Annual Meeting, Paper number: 19-00213

January 2019

ABSTRACT
This work proposes a periodic bus holding control method where the bus holding times of all running trips are computed simultaneously within each optimization time period; thus, increasing the coordination among running buses to avoid bus bunching. We consider the adverse effects of the bus holding control on the in-vehicle travel times of on-board passengers and formulate holistic bus holding decisions by modeling the bus holding problem as a discrete, nonlinear, constrained optimization problem. Given the computational complexity of the bus holding problem, an alternating minimization approach is introduced for computing the optimal holding times at each optimization instance. The performance of the periodic control method is evaluated against the performance of event-based control methods using 5-month automated vehicle location and automated passenger count data from bus line 1 in Stockholm demonstrating an improvement potential of 5% for the in-vehicle travel times and 11% for the service regularity.

Keywords: Dynamic Bus Holding; Nonlinear Programming; Periodic Control; Discrete Optimization; Bus Bunching

INTRODUCTION

Transport authorities have different metrics to evaluate the performance of bus services ranging from safety Key Performance Indicators (KPIs) to planning and operational KPIs. Focusing on the operational KPIs, transport authorities use punctuality-based KPIs to evaluate the performance of low frequency bus services and regularity-based KPIs to evaluate the performance of high frequency services (Gkiotsalitis and Maslekar (*1*), Gkiotsalitis and Stathopoulos (*2*)). In the case of high frequency bus services (i.e., bus lines with more than 5 buses per hour), the most common objective of the bus operations is to maintain an appropriate headway among buses in order to ensure that the waiting times of passengers at stops are close to the planned ones.

Not having a specific, punctuality-based arrival time target provides more flexibility to bus operators in applying control measures that can reduce the waiting time variation of passengers. Given the spatio-temporal variations of traffic and passenger demand, some buses are delayed at some specific segments of the line. These delays have two adverse effects: (i) they increase the headway of the delayed bus from its preceding one; (ii) the delayed bus arrives late at the next stop and, as a result, it needs to accommodate more passengers who are waiting for an extended period of time. This causes a "domino effect" that makes it difficult for a delayed bus to recover and leads to bus bunching.

Several metropolises with high-frequency bus services such as London, Singapore, Hong Kong, Barcelona (to name a few) have adopted specific KPIs to evaluate the regularity of the running services based on the waiting times of passengers (TfL (*3*), LTA (*4*)). For this purpose, real-time Automated Vehicle Location (AVL) information from the Fleet Telematic System (FTS) of each bus is collected and the recorded arrival times of buses at stops are used for evaluating the performance of the services. This constant monitoring of the bus operations has increased the pressure on bus operators to improve the regularity of their services and has led to the adoption of real-time control systems that can communicate actuation measures to bus drivers in near real-time. Nevertheless, most control options are currently based on ad-hoc rules (Cats et al. (*5*), Laskaris et al. (*6*)) which are triggered when a specific event occurs (i.e., a bus arrives late at a stop) and are not capable of finding an adequate trade-off between the in-vehicle travel times and the waiting times of passengers at stops.

RELATED STUDIES

In general, the daily operations of a bus line are formed on the basis of an extensive tactical planning process that involves the planning of bus frequencies, crew scheduling and several other constraints such as the minimum layover times between successive trips and the permitted dispatching headway ranges (Kepaptsoglou and Karlaftis (*7*), Farahani et al. (*8*), Gkiotsalitis and Cats (*9*), Gkiotsalitis and Cats (*10*)). During the tactical planning phase, the dispatching time of each bus trip is defined and the daily timetable is generated. Nevertheless, tactical planning assumes that the trip travel times remain stable during the actual operations or variate very little from their expected values (Gkiotsalitis and Kumar (*11*), Gkiotsalitis and Maslekar (*12*)). This is not the case though in real-world operations and several studies have proposed the introduction of real-time corrective actions, such as bus holding control, that can reduce the waiting time variation of passengers at stops.

The plurality of studies on the bus holding problem (Daganzo (*13*), Koehler et al. (*14*), Hickman (*15*), Eberlein et al. (*16*), Berrebi et al. (*17*), Sánchez-Martínez et al. (*18*), Gkiotsalitis and Maslekar (*19*)) focus on optimizing the bus holding times at a pre-determined sub-set of bus stops, known as control points, for reducing the waiting times of passengers. Such studies have a predefined target for the waiting time of passengers at each stop (i.e., scheduled waiting time) and employ heuristics for computing the optimal bus holding times at some bus stops for adhering to the target waiting times. The work of Bartholdi and Eisenstein (*20*) is an exception because its objective was to reduce the waiting time variance without adhering to a specific target waiting time.

In most of these works, the implications of the bus holding measures to the future trips are not considered and it is assumed that the slack times are long enough to ensure that delays are not propagated. Fu and Yang (*21*) tested two of the most common bus holding strategies: (i) the one-headway-based control where a bus is held at a control point stop if it is decided upon its arrival there that it is too close to its preceding bus and (ii) the two-headway-based control that estimates also the arrival time of the following bus at the same stop and considers the headways between its preceding and its following bus when computing a holding time. In addition, several works have also considered travel time variations when computing the bus holding times (Chen et al. (*22*), Daganzo (*13*)). For instance, Hickman (*15*) introduced stochasticity at the single holding problem by modeling it as a convex quadratic program in a single variable. Finally, Daganzo and Pilachowski (*23*) and Zhao et al. (*24*) proposed distributed control models where buses act as agents that communicate in real-time to achieve dynamic coordination.

To the best of the authors' knowledge, most of the existing studies on bus holding are based on the assumption that the slack times are long enough so that the bus holding measures do not impact future trips and the respective crew and vehicle schedules (Fu and Yang (*21*), Zhao et al. (*24*)). In addition, most studies focus on myopic, event-based rules that decide about the holding time of a bus trip based on its headway from the preceding and following buses when it arrives at a control point. Such myopic approaches might fail to capture the repercussions of the bus holding measure to the entirety of the running trips and can be therefore counterproductive (Fu and Yang (*21*)).

Our study is one of the first to examine how a set of holistic holding time decisions can be determined while ensuring that the potential repercussions due to the holding time decisions are taken into consideration. For making such holistic holding time decisions, we consider the effects of different holding time combinations and this results in a decision making process where all holding time options for all running buses are evaluated simultaneously. The main contributions of this work to the state-of-the-art are (i) the introduction of a time-window-based model that makes holistic decisions regarding holding times instead of performing a myopic optimization considering the status of neighboring buses; and (ii) the explicit consideration of domino effects due to bus holdings, such as the increase of the in-vehicle travel times and the delay of future trip departures (also known as "schedule sliding"), with the use of a multi-constrained modeling approach.

BUS KINEMATIC MODEL FOR A PERIODIC OPTIMIZATION INSTANCE

During the daily bus operations, the running buses, $N = \{1, 2, ..., n, ..., |N|\}$, are expected to serve a number of predefined bus stops. Before proceeding further, it is important to present the main assumptions of this work:

Assumption 1: The number of trips of a bus line are defined during the tactical planning phase and must be fulfilled during the actual operations (i.e., no trip cancellations)

Assumption 2: Bus operators can apply bus holding control measures at control stops and cannot modify the scheduled dispatching times of buses except in the case of emergency (i.e., due to heavy traffic delays)

Assumption 3: Overtaking between buses of the same line is not allowed (a common assumption used in most related works Xuan et al. (*25*), Sun and Hickman (*26*), Chen et al. (*22*))

As it will be demonstrated later on, this work applies bus holdings measures that are not expected to postpone the departures of future trips operated by the same buses. This is a significant step towards ensuring that the optimization of the holding times of the running buses will not result in schedule sliding. In addition, this allows the discretization of the daily time horizon into a set of times periods $M = \{\tau, \tau+\delta, \tau+2\delta, ..., \tau+|M|\delta\}$ within which periodic optimizations of the daily operations are allowed to carry out.

The periodic, time-window-based optimization differs from the event-based headway optimization proposed in most studies (e.g., Yu and Yang (*27*), Fu and Yang (*21*)). In a periodic optimization holding control, it is assumed that within a time period δ, the link travel times of trips are available from a link travel time prediction model and the time period δ is short enough so that these travel times are relatively stable. In addition, the periodic optimization control is time-based and an optimization that computes all the holding times of the running buses is applied only at the beginning of the optimization time period $m \in M$.

Under the following assumptions, the expected arrival time of a running bus at one future bus stop, s', during the periodic optimization period $m = \{\tau+m\delta, \tau+(m+1)\delta\}$ is:

$$a_{m,n,s'}(\mathbf{x}) = (\tau+m\delta) + \xi_{\beta_{n,m},\rho_n} + \sum_{i=\rho_n}^{s'-1} t_{m,n,i} + \sum_{i=\rho_n}^{s'-1} d_{m,n,i} + \sum_{i=\rho_n}^{s'-1} x_{m,n,i} \tag{1}$$

where $a_{m,n,s'}$ is the expected arrival time of trip n at stop s' during the m^{th} optimization time period, ρ_n is the next bus stop of trip n given its current position at the beginning of the m^{th} optimization timeperiod and $\xi_{\beta_{n,m},\rho_n}$ is the expected travel time from the position of trip n at the beginning of the optimization, which is denoted as $\beta_{n,m}$, to its next bus stop ρ_n. The expected arrival of the bus at stop s' is affected also by the expected dwell time of trip n at each stop i which is denoted as $d_{m,n,i}$ and the holding times at each stop $x_{m,n,i}$.

Obviously, the expected arrival time of a bus trip at a stop s' is equal to the accumulation of the link travel times, the dwell times and the bus holding times from its current position, $\beta_{n,m}$, until stop s' plus the timestamp $\tau+m\delta$ that denotes the beginning of the optimization time period $m \in M$.

In Eq.1 the link travel times and the dwell times of the bus trip n are deterministic. However, in real-world operations these two parameters have some degree of randomness. Several works such as Shalaby and Farhan (*28*), Mazloumi et al. (*29*), Chien et al. (*30*) have focused on predicting the travel times of bus operations based on bus data with the use of several techniques spanning from time series analysis to artificial neural networks.

The randomness of the link travel times is explicitly considered in our analysis which studies the effects of travel time disturbances to the waiting times of passengers. By using the predicted link travel time and dwell time values at the time period m^{th}, the headway between a bus trip n and its preceding trip $n-1$ at stop s can be estimated as:

$$h_{m,n,s}(\mathbf{x}) = a_{m,n,s}(\mathbf{x}) - a_{m,n-1,s}(\mathbf{x}) \tag{2}$$

where its value depends on the applied bus holding measures $\mathbf{x}$. Finally, assuming random passenger arrivals at stops that follow the uniform distribution (Osuna and Newell (*31*)), the waiting time of passengers who are willing to board bus trip n is:

$$w_{m,n,s}(\mathbf{x}) = \frac{h_{m,n,s}(\mathbf{x})}{2} = \frac{a_{m,n,s}(\mathbf{x}) - a_{m,n-1,s}(\mathbf{x})}{2} \tag{3}$$

MODEL FORMULATION

In this section we formulate the mathematical model in order to solve the periodic bus holding optimization problem. The following nomenclature is used for developing the model.

Parameters

$S = \{1, 2, ..., s, ..., \lvert S\rvert\}$	set of bus stops
$N = \{1, ..., n, ..., \lvert N\rvert\}$	set of daily trips
J	set of bus stops where bus holdings are allowed
υ_n	scheduled dispatching time of each trip $n \in N$
u_n	scheduled arrival time of each trip $n \in N$ at the terminal $\lvert S\rvert$
k_n	embedded slack time for each trip $n \in N$
r_n	required resting time for each trip $n \in N$ before the dispatching of the next trip operated by the same bus
f_n	the next trip after trip n that is operated by the same bus
δ	time duration of each periodic optimization time period
$M = \{\tau, ..., \tau + \lvert M\rvert\delta\}$	set of daily periodic optimization time periods
m	identifier of a periodic optimization time period
$\beta_{n,m}$	location of bus trip n at the start of optimization period m
$b_{s,s+1}$	travel distance between bus stops s and $s+1$
$\alpha_{n,s}$	the recorded arrival time of trip n at stop s which is set equal to zero if such information is not available yet
$a_{m,n,s}$	expected arrival time of trip n at stop s during the m^{th} optimization time period without considering bus holdings
$a_{m,n,s}(\mathbf{x})$	expected arrival time of trip n at stop s during the m^{th} optimization time period as a function of the bus holding times
$h_{m,n,s}(\mathbf{x})$	headway between bus trips $n-1$ and n at stop s during the m^{th} optimization time period
$d_{m,n,s}$	expected dwell time of bus trip n at stop s during the m^{th} optimization time period
$t_{m,n,s}$	expected link travel time of trip n between consecutive bus stops s and $s+1$ during the m^{th} optimization time period
ζ_s	desired waiting time of passengers at stop s according to the planned schedule

Decision Variables

$\mathbf{x} = \{x_{m,n,s}\}$	holding times for each trip $n \in N$ at each stop $s \in S$ during the m^{th} optimization time period

Note that our aim is to find the optimal bus holding times, $\mathbf{x}$, for the running buses at each optimization time period. During the m^{th} periodic optimization period, the active and inactive decision variables can be defined by the following inequalities:

$$x_{m,n,s} = \begin{cases} x_{m,n,s} = 0 \text{ if } a_{m,n,s} < \tau + m\delta \text{ or } a_{m,n,s} > \tau + (m+1)\delta \text{ or } s \in J \\ x_{m,n,s} \in q \text{ if } \tau + m\delta \leq a_{m,n,s} \leq \tau + (m+1)\delta \text{ and } s \notin J \end{cases} \quad (4)$$

where J is the set of control point stops at which a but trip can be held at. Eq.4 includes several inequalities that determine whether a bus holding decision variable $x_{m,n,s}$ is equal to zero (inactive) or can take values from a candidate set q. From Eq.4 it is evident that if the expected arrival time of a bus trip n at one stop s is before the start of the m^{th} optimization window $\{\tau + m\delta\}$ or after the end of the m^{th} optimization window $\{\tau + (m+1)\delta\}$, then there is no need to make a decision for that holding time during this optimization time window.

Many studies such as Lee et al. (*32*) have shown that a driver has a minimum response time of $\simeq$ 1.5 second and if a driver is requested to hold a bus at one stop for a very small amount of time, he/she will not be able to implement the holding time request with high accuracy. Several works such as Cortés et al. (*33*), Gkiotsalitis and Maslekar (*34*) have proposed to discretize the holding time instructions to 30-second periods (i.e., 0, 30, 60, ...) seconds. The same 30-second discretization for the holding time decisions is used in commercial software as it is described in Cats et al. (*5*). In this study, we propose a 10-second discretization scheme for the holding time options. In such case, the minimum driver response time of $\simeq$ 1.5sec. can only affect the holding time suggestion by 15% at most.

Slack Time Constraints

During the daily scheduling of the bus operations a slack time is allocated to each trip in order to ensure that the time delay of a bus trip due to traffic will not result in a knock-down effect for the departure of the next trip operated by the same bus. If a bus trip, n, is expected to arrive at terminal $|S|$ at time $a_{m,n,|S|}(\mathbf{x})$, then the slack time $k_n \in \mathbb{Z}_+$ is the extra time that the bus is allowed to remain at the terminal before starting its next trip.

The expected arrival time of a running trip n at terminal $|S|$, $a_{m,n,|S|}(\mathbf{x})$, can be directly linked to the slack time k_n if we consider the scheduled arrival time of trip n at the terminal, u_n. Satisfying the slack time constraints can be expressed by the following set of inequalities that consider the scheduled arrival time of a trip at the terminal and the associated slack time:

$$a_{m,n,|S|}(\mathbf{x}) \leq u_n + k_n \quad \forall n \in R \quad (5)$$

where R is the set of running trips during the m^{th} optimization period. The inequalities of Eq.5 denote that a bus trip n should be anticipated to arrive at the last stop before its scheduled arrival time at that stop plus the slack time. Note that if, due to heavy traffic, any of these inequality constraints cannot be satisfied (i.e., the examined trip is delayed beyond the slack time buffer); then, Eq.5 does not allow to hold this trip at control point stops.

Travel Time Limits for On-Board Passengers

Excessive holding for regularizing the headways of the running buses is inconvenient for the on-board passengers. The obvious side effects are (i) the inflicted increases on the on-board passenger travel times due to holdings, and (ii) the inconvenience of on-board passengers that are held at several stops beyond the required dwell time for the boardings/alightings.

Let assume that g_n denotes the total bus holding time that is permitted for a trip n. This cumulative limit for the holding times, g_n, can be used for controlling the maximum amount of holding times that are allowed for one trip and can ensure that the on-board passengers will not have a travel time increase of more than g_n due to holdings. Depending on the passenger demand of each trip $n \in N$, the g_n values can vary (i.e., trips with higher demand might be allowed to operate without significant holdings ($g_n \simeq 0$), whereas trips with lower demand might be more flexible and allow more time for holdings). The cumulative holding time limits, g_n, per each trip can be provided from the transport authority or derived from historical data analysis regarding the passenger demand levels imposing the following inequality constraints:

$$\sum_{s=1}^{|S|} \sum_{m=1}^{|M|} x_{m,n,s} \leq g_n \quad \forall n \in N \tag{6}$$

One should note that Eq.6 accumulates all the holding times that are calculated during different periodic optimization time periods because the state-space model of this work is implemented in a rolling-horizon.

Bounding the Upper Limits of Bus Holdings

As described in Eq.4, each holding decision is allowed to take a value from the discrete set q that contains all potential holding time options with a 10-second difference for improving the probability that a bus driver will be able to implement accurately this holding option. However, each holding time should have an upper limit because one bus cannot be held at one stop for a very long time since (i) this increases the on-board passengers' and driver inconvenience; (ii) other buses from the same or different lines should use the same stop for boardings/alightings.

Several studies such as Cortés et al. (*33*), Gkiotsalitis (*35*) and Cats et al. (*5*) have proposed to impose an upper limit to the bus holding time which is in the range of 60-90 seconds depending on the characteristics of the control point stop. Adopting such upper limit for each holding time, the admissible set of holding times is modified to:

$$q = \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90\} \text{ seconds} \tag{7}$$

Reducing the Variance of the Actual Passenger Waiting Times from the Planned Waiting Times

Bus operators strive to improve the service reliability by minimizing the passenger waiting time variance from the planned waiting time, ζ_s, at each stop $s \in S$ (Daganzo (*13*), LTA (*4*)). If the scheduled waiting time ζ_s at one stop is different than the actual waiting time of passengers, then the service reliability deteriorates and the bus operators should take corrective actions.

We illustrate this using an example where the trips $R = \{n, n+1, n+2, n+3\}$ are the running trips at the start of the m^{th} periodic optimization instance as they are presented in Fig.1. In addition, let $J = \{s = 3, s = 5, s = 7, s = 9\}$ be the set of the control point stops of the bus service in Fig.1.

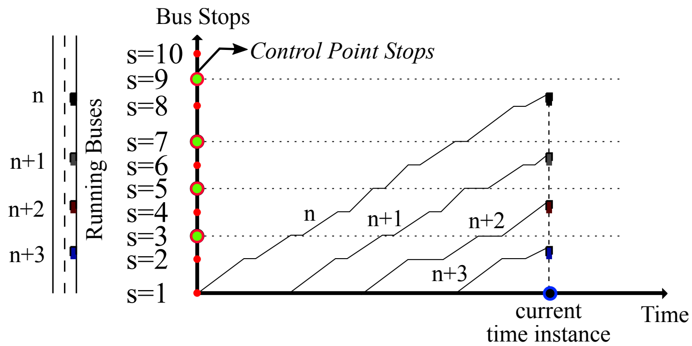

FIGURE 1 : Illustration of the running buses at the beginning of the m^{th} periodic optimization instance

If trips $R = \{n, n+1, n+2, n+3\}$ are the running trips at the m^{th} optimization instance, then trips $R_{prior} = \{1, 2, ..., n-1\}$ have already completed their service prior to the start of the m^{th} optimization time period.

In this way, we can split the daily trips $N = \{1, ..., n, ..., |N|\}$ into three sets during the m^{th} optimization instance: (i) the trips that are already completed $R_{prior} = \{1, 2, ..., n-1\}$; (ii) the running trips that are expected to operate during the m^{th} time period $R = \{n, n+1, n+2, n+3\}$; and (iii) the remaining trips, $R_{future} = \{n+4, n+5, ..., |N|\}$, until the end of the day.

Having defined the set of running trips, we proceed one step further by identifying the set of trips that are expected to arrive at each bus stop $s \in S$ during the m^{th} optimization time period.

For each bus stop $s \in S$ we initialize a set $R_{m,s}$ that contains all the bus trips that are expected to arrive at that stop during the m^{th} optimization time period. As a result, at the start of the periodic optimization each trip $n \in N$ is assigned to a set $R_{m,s}$ according to the following inequalities:

$$n = \begin{cases} n \in R_{m,s} \text{ if } \alpha_{n,|S|} = 0 \text{ and } \tau + m\delta \leq a_{m,n,s} \leq \tau + (m+1)\delta \text{ and } n \in R \\ n \notin R_{m,s} \text{ otherwise} \end{cases} \tag{8}$$

where $\alpha_{n,|S|} = 0$ denotes that trip n has not been completed at the start of the m^{th} optimization time period and $\tau + m\delta \leq a_{m,n,s} \leq \tau + (m+1)\delta$ denotes that the expected arrival time of trip n at stop s lies within the range of the m^{th} optimization time period where:

$$a_{m,n,s} = (\tau + m\delta) + \xi_{\beta_{n,m},\rho_n} + \sum_{i=\rho_n}^{s-1} t_{m,n,i} + \sum_{i=\rho_n}^{s-1} d_{m,n,i} \quad \forall n \in R \tag{9}$$

In Eq.9, $a_{m,n,s}$ is the expected arrival time of trip n at stop s during the m^{th} optimization time period without considering bus holding measures, ρ_n is the next bus stop of trip n given its current position at the beginning of the m^{th} optimization timeperiod and $\xi_{\beta_{n,m},\rho_n}$ is the expected travel time from the position of trip n at the beginning of the optimization, which is denoted as $\beta_{n,m}$, to its next bus stop ρ_n.

After defining the set of trips $R_{m,s}$ that are expected to arrive at stops during the m^{th} optimization time period, the actual waiting time variance from the scheduled waiting times at stops, ζ_s, should be minimized. Assuming random passenger arrivals at stops, those arrivals follow the uniform distribution and the waiting times of passengers are half the headway between two consecutive bus trips (Osuna and Newell (*31*)) resulting in the following objective function:

$$\min_x f(x) = \sum_{s \in S} \sum_{n \in R_{m,s}} \Big(\frac{a_{m,n,s}(\mathbf{x}) - a_{m,n-1,s}(\mathbf{x})}{2} - \zeta_s \Big)^2 \tag{10}$$

where

$$a_{m,n,s}(\mathbf{x}) = (\tau + m\delta) + \xi_{\beta_{n,m},\rho_n} + \sum_{i=\rho_n}^{s-1} t_{m,n,i} + \sum_{i=\rho_n}^{s-1} d_{m,n,i} + \sum_{i=\rho_n}^{s-1} x_{m,n,i} \quad \forall n \in R \tag{11}$$

From the above descriptions, a mathematical model that (i) minimizes the passenger waiting time variation at stops, (ii) satisfies the slack time constraints and (iii) limits the travel times of on-board passengers according to the on-board passenger demand levels is formulated by combining equations 10, 5, 6 for the optimization of the m^{th} rolling horizon:

$$\begin{aligned}
\min_{\mathbf{x}} \quad & f(\mathbf{x}) = \sum_{s \in S} \sum_{n \in R_{m,s}} \Big(\frac{a_{m,n,s}(\mathbf{x}) - a_{m,n-1,s}(\mathbf{x})}{2} - \zeta_s \Big)^2 \\
\text{subject to} \quad & a_{m,n,|S|}(\mathbf{x}) \leq u_n + k_n, \quad \forall n \in R \\
& \sum_{s=1}^{|S|} \sum_{m=1}^{|M|} x_{m,n,s} \leq g_n, \; n = 1, \dots, |N|. \\
& x_{m,n,s} = \begin{cases} 0 \text{ if } a_{m,n,s} < \tau + m\delta \text{ or } a_{m,n,s} > \tau + (m+1)\delta \text{ or } s \in J \\ x_{m,n,s} \in q \text{ if } \tau + m\delta \leq a_{m,n,s} \leq \tau + (m+1)\delta \text{ and } s \notin J \end{cases} \\
& q = \{0, 10, 20, ..., 90\} \text{ seconds}
\end{aligned} \tag{12}$$

Modeling the Dwell Times

The dwell time $d_{m,n,s}$ of a bus trip n at stop s during the m^{th} optimization time period can be associated with the number of boardings and alightings. Detailed studies such as Levinson (*36*), Guenthner and Sinha (*37*) and Bertini and El-Geneidy (*38*) showed that there is a linear relationship between the dwell time and the number of boardings where the dwell time can be equal to a minimum value p_0 that can range from 2-5 seconds in case of no boardings and it increases by p_1=1.5-4.5 seconds for each extra passenger boarding depending on the fare payment structure.

In accordance with the above studies, the underlying relationship between the observed dwell time and the number of boardings can be expressed by a linear equation:

$$d_{m,n,s} = p_0 + p_1 B_{n,s} \tag{13}$$

where p_0 is expressed in seconds, p_1 in seconds per passenger and $B_{n,s}$ is the expected number of passengers that will board on trip n at stop s.

If we assume a passenger arrival rate B_s that expresses the number of passengers per second that are expected to arrive at stop s, the number of passenger boardings, $B_{n,s}$, on one trip n at stop s is associated with the headway between trip n and its preceding trip $n-1$ at that stop. Consequently, the boarding time of passengers $d_{m,n,s}$ is:

$$d_{m,n,s} = p_0 + p_1 B_s (a_{m,n,s} - a_{m,n-1,s}) \tag{14}$$

Eq.14 denotes that the dwell time at one stop is proportional to the headway between two consecutive bus trips $n-1, n$ multiplied by the passenger arrival rate at that stop. As a result, holding a running bus $n-1$ at stop s impacts the dwell time of trip n at stop s:

$$d_{m,n,s}(\mathbf{x}) = p_0 + p_1 B_s (a_{m,n,s}(\mathbf{x}) - a_{m,n-1,s}(\mathbf{x})) \tag{15}$$

Plugging Eq.11 and Eq.15 into the model of Eq.12 yields:

$$\begin{aligned}
\min_{\mathbf{x}} \quad & f(\mathbf{x}) = \sum_{s \in S} \sum_{n \in R_{m,s}} \Big(\frac{1}{2} \big((\tau + m\delta) + \xi_{\beta_{n,m},\rho_n} + \sum_{i=\rho_n}^{s-1} t_{m,n,i} + \sum_{i=\rho_n}^{s-1} d_{m,n,i}(\mathbf{x}) + \sum_{i=\rho_n}^{s-1} x_{m,n,i} \big) \\
& \quad - \frac{1}{2} \big((\tau + m\delta) + \xi_{\beta_{n-1,m},\rho_{n-1}} + \sum_{i=\rho_{n-1}}^{s-1} t_{m,n-1,i} + \sum_{i=\rho_{n-1}}^{s-1} d_{m,n-1,i}(\mathbf{x}) + \sum_{i=\rho_{n-1}}^{s-1} x_{m,n-1,i} \big) \\
& \quad - \zeta_s \Big)^2 \\
\text{s.t.} \quad & a_{m,n,|S|}(\mathbf{x}) \leq u_n + k_n, \quad \forall n \in R \\
& \sum_{s=1}^{|S|} \sum_{m=1}^{|M|} x_{m,n,s} \leq g_n, \; n = 1, \dots, |N|. \\
& x_{m,n,s} = \begin{cases} 0 \text{ if } a_{m,n,s} < \tau + m\delta \text{ or } a_{m,n,s} > \tau + (m+1)\delta \text{ or } s \in J \\ x_{m,n,s} \in q \text{ if } \tau + m\delta \leq a_{m,n,s} \leq \tau + (m+1)\delta \text{ and } s \notin J \end{cases} \\
& d_{m,n,s}(\mathbf{x}) = p_0 + p_1 B_s (a_{m,n,s}(\mathbf{x}) - a_{m,n-1,s}(\mathbf{x})) \\
& q = \{0, 10, 20, \dots, 90\} \text{ seconds}
\end{aligned} \tag{16}$$

Remark: From the model of the m^{th} periodic optimization described in Eq.16, one can note that the dwell time of a trip at a stop is affected by the holding times. This results in an open-loop relationship between the arrival times of trips at stops and the dwell times because the dwell time modifications impact the arrival times of trips at stops according to Eq.15 and the arrival time modifications are modifying again the dwell times expressed in Eq.11 leading to a recursive optimization problem with a non-convex objective function.

SOLUTION METHOD

Decoupling the Arrival and Dwell Times with Alternating Minimization

The authors utilize the alternating minimization method for non-convex optimization by solving iteratively the model of Eq.18 where at each iteration the dwell time values are assumed to be deterministic resulting in a simpler sub-problem with a convex objective function. The alternating minimization method initializes the problem by assuming that the optimal bus holding times are equal to zero. For this initial solution, $\mathbf{x}^{\kappa}$, where $x^{\kappa}_{m,n,s} = 0, \forall x^{\kappa}_{m,n,s} \in \mathbf{x}^{\kappa}$, the resulting dwell times are computed:

$$d_{m,n,s} = p_0 + p_1 B_s(a_{m,n,s}(\mathbf{x}^{\kappa}) - a_{m,n-1,s}(\mathbf{x}^{\kappa})) \tag{17}$$

Then, Eq.18 computes the new optimal solution $\mathbf{x}^{\kappa+1}$ under the assumption of fixed dwell time values which were computed by Eq.17. After that, the dwell time values are updated for the bus holding solution $\mathbf{x}^{\kappa+1}$ and the model of Eq.18 is solved iteratively until a termination criterion is satisfied.

$$\begin{aligned}
\min_{\mathbf{x}^{\kappa+1}} \quad & f(\mathbf{x}^{\kappa+1}) = \sum_{s\in S}\sum_{n\in R_{m,s}} \Big(\frac{1}{2}\big((\tau+m\delta)+\xi_{\beta_{n,m},\rho_n}+\sum_{i=\rho_n}^{s-1} t_{m,n,i}+\sum_{i=\rho_n}^{s-1} d_{m,n,i}+\sum_{i=\rho_n}^{s-1} x^{\kappa+1}_{m,n,i}\big) \\
& -\frac{1}{2}\big((\tau+m\delta)+\xi_{\beta_{n-1,m},\rho_{n-1}}+\sum_{i=\rho_{n-1}}^{s-1} t_{m,n-1,i}+\sum_{i=\rho_{n-1}}^{s-1} d_{m,n-1,i}+\sum_{i=\rho_{n-1}}^{s-1} x^{\kappa+1}_{m,n-1,i}\big) \\
& -\zeta_s\Big)^2 \\
\text{s.t.} \quad & a_{m,n,|S|}(\mathbf{x}^{\kappa+1}) \leq u_n + k_n, \quad \forall n \in R \\
& \sum_{s=1}^{|S|}\sum_{m=1}^{|M|} x^{\kappa+1}_{m,n,s} \leq g_n,\ n=1,\dots,|N|. \\
& x^{\kappa+1}_{m,n,s} = \begin{cases} 0 \text{ if } a_{m,n,s} < \tau+m\delta \text{ or } a_{m,n,s} > \tau+(m+1)\delta \text{ or } s\in J \\ x^{\kappa+1}_{m,n,s} \in q \text{ if } \tau+m\delta \leq a_{m,n,s} \leq \tau+(m+1)\delta \text{ and } s \notin J \end{cases} \\
& d_{m,n,s} = p_0 + p_1 B_s(a_{m,n,s}(\mathbf{x}^{\kappa}) - a_{m,n-1,s}(\mathbf{x}^{\kappa})) \\
& q = \{0,10,20,...,90\} \text{ seconds}
\end{aligned} \tag{18}$$

The alternating minimization approach is formed in such a way that, even if the main problem is non-convex, each alternating step in isolation is tractable. For solving this problem, one can deploy the branch-and-bound (B&B) method that is proposed by Land and Doig (*39*) and is one of the most common approaches for solving discrete, NP-Hard problems.

The B&B method explores branches of a dynamically generated tree which represent subsets of the solution set of discrete options. In contrary to the brute-force algorithm that explores the solution space exhaustively, the B&B method uses upper and lower bounds for pruning the set of candidate solutions. In such way, if the estimation of the upper and lower bounds is effective, the exploration effort is significantly reduced.

In order to define the lower bound (LB) of our problem, we allow the holding times to take any real value from 0 to 90 seconds instead of values from the discrete set $q = \{0,10,20,...,90\}$. Then, after expanding the problem of Eq.18 by plugging Eq.11 into Eq.18, it becomes a continuous Quadratic Programming (QP) problem that can be solved in polynomial time with an exact optimization algorithm such as the Active Set method (for more information regarding the Active Set method, one can refer to Murty and Yu (*40*) or Nocedal and Wright (*41*)).

By solving the continuous QP, the lower bound of the problem is established because any discrete solution has inferior performance compared to the global optimum solution of the continuous

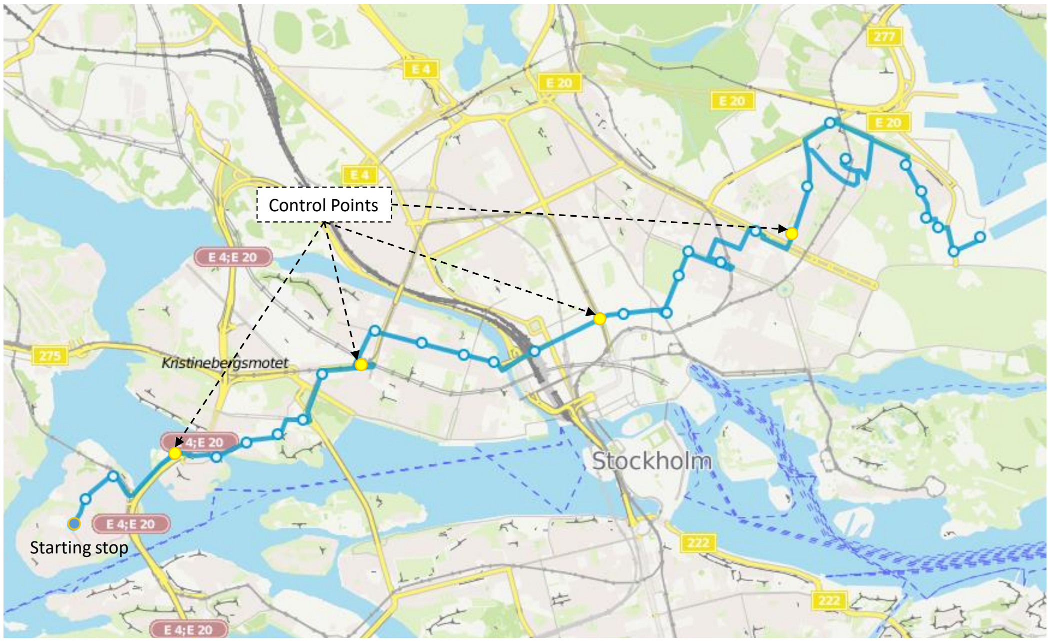

FIGURE 2 : Eastbound direction of Bus Line 1 in Stockholm that serves 33 bus stops from which 4 are control point stops

problem and the steps of the B&B algorithm can be performed[1].

CASE STUDY: BUS LINE 1 IN STOCKHOLM

Application

Our application uses data from the eastbound direction of bus line 1 in Stockholm. Line 1 connects the main eastern harbor to a residential area in the western part of the city through the commercial center. In this work we processed detailed 5-month datasets of Automatic Vehicle Location (AVL) and Automatic Passenger Counting (APC) data for this line that serves 33 bus stops and has 4 control point stops as presented in Fig.2.

The 5-month AVL data is used for estimating the link travel times of the 118 daily trips that operate in the eastbound direction (direction 1) from 07:00 until 19:00. Note that the time-period 07:00-19:00 is selected as the main focus area of the daily operations since trips prior to 07:00 and after 19:00 do not experience significant spatio-temporal link travel time variations due to road traffic.

In this work, we perform a time-window-based bus holding control on the daily operations where the entire day is split into 10-minute periods within which the optimization is performed. For applying the computed bus holding measures, we build the Stockholm city network in the open source Simulation of Urban MObility (SUMO) using information from OpenStreetMaps. The daily AVL data from the actual bus operations was used for replicating the daily operations in SUMO. For performing such action, we updated automatically the parameters of the traffic scenarios with the use of the Constrained Optimization BY Linear Approximation (COBYLA) algorithm from the SciPy library in Python that computes the optimal parameter values of traffic

[1] since the B&B algorithm is a well-known solution method, we will not provide here its detailed steps. Instead, we refer the interested reader to the work of Land and Doig (*39*))

flows and traffic signal cycles which minimize the root-mean-square error (RMSE) between the actual and simulated arrival times of buses at stops.

The vehicle parameters of the running buses (i.e., acceleration, max. speed, deceleration, vehicle length) are specified in SUMO using an .xml file that follows the guidelines of SUMO (*42*). After the calibration stage, the acceleration, deceleration and maximum speed values are set as 2.61 m/s^2, 4.68 m/s^2 and 100.0 km/h, respectively. In addition, the daily boarding rates, expressed as number of boarding passengers per hour, are derived from the APC data. For demonstration purposes, the hourly boardings from 08:00 until 12:00 are presented in Fig.3.

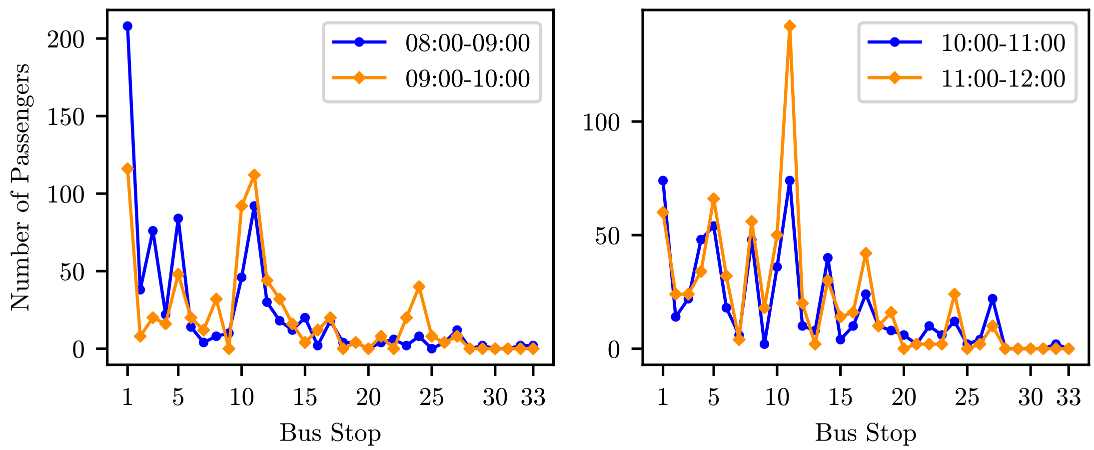

FIGURE 3 : Hourly passenger boardings rates at each bus stop (from 08:00 until 12:00)

The results of the bus holding control for the eastbound direction of line 1 in Stockholm are presented in Fig.4 where the computed bus holding measures at every 10-minute optimization horizon are implemented in SUMO (from 07:00 until 19:00 are performed $|M| = 72$ periodic optimizations and the respective bus holding times are applied in SUMO). In the same figure, the results of the one-headway-based control method (Fu et al. (*43*)) are presented where each bus trip that arrives at one control point stop is held there according to its headway from its preceding trip and the variance of the passenger waiting times from their target values is evaluated every 10 minutes. In this way, the proposed periodic bus holding optimization approach is tested against the one-headway-based control method which is one of the most used event-based approaches.

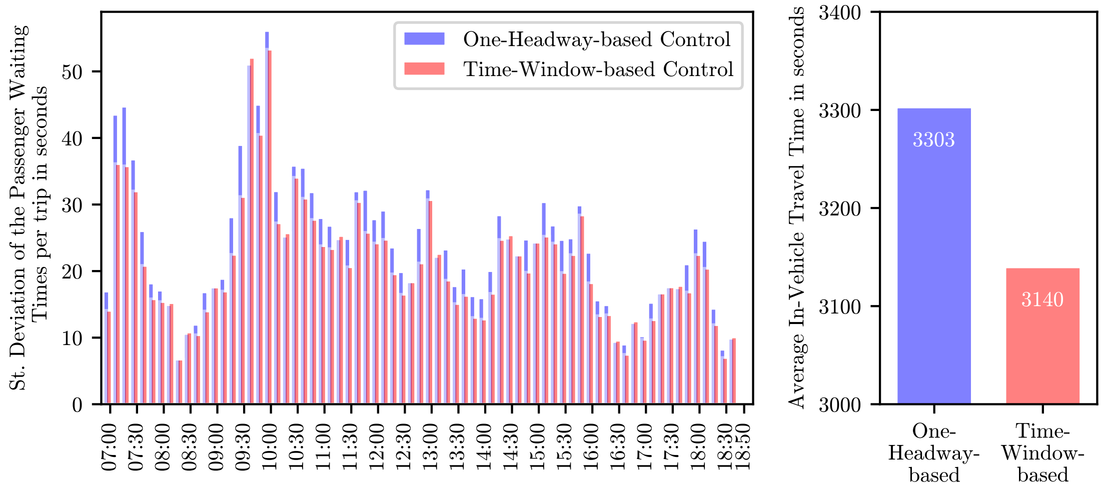

FIGURE 4 : Results of the Bus Holding control in the eastbound direction of line 1 in Stockholm

Sensitivity to Travel Time Variation

Finally, Fig.5 presents different instances of the results of the one-headway-based bus holding measures and the periodic optimization bus holding measures when the actual link travel times exhibit noise levels of 10%, 20%, 50% and 80%. An important observation from Fig.5 is that the bus holding measures for noise levels of 10% maintain a passenger waiting time variance close to the desired passenger waiting times whereas when the noise levels are in the range of 80% the waiting times variate significantly from their target values due to bus bunching.

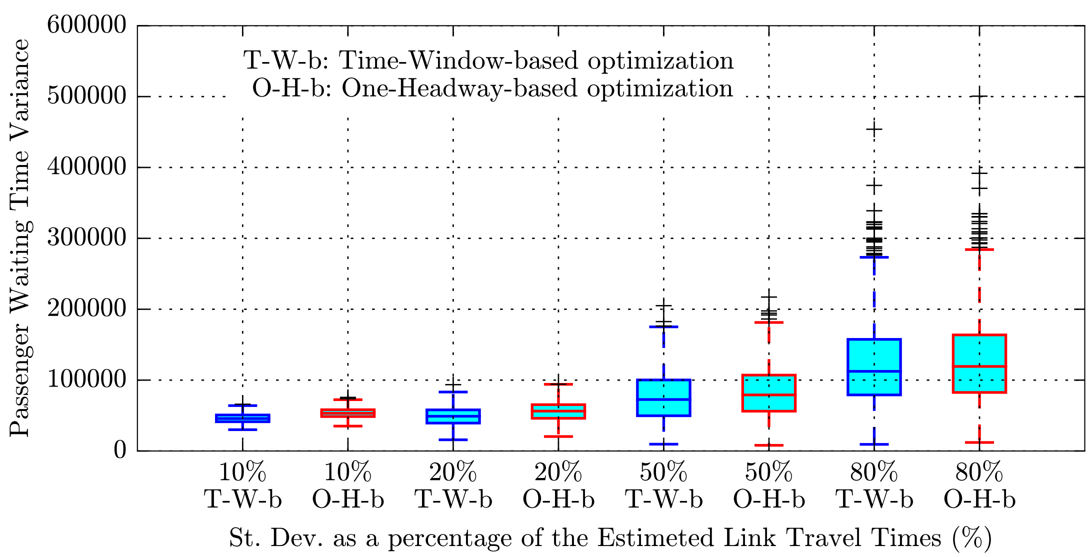

FIGURE 5 : Comparative Analysis of the periodic optimization and the one-headway based approach in the presence of travel time noise. The overall daily variance of the actual passenger waiting times from the expected ones is measured in seconds

Discussion

Our work replicates the daily operations of the eastbound direction of bus line 1 in Stockholm using the AVL and APC data from the daily operations for building a realistic simulation scenario in SUMO by minimizing the root-mean-square error (RMSE) between the actual and simulated arrival times of buses at stops. This calibration enabled us to apply the periodically optimized bus holding times to a realistic simulation scenario that is close to the real operations.

In this simulation scenario, this work evaluated the performance of the bus holding control measures computed by the proposed time-window-based optimization method. The one-headway-based bus holding method was also applied in the same scenario resulting in higher waiting time variance from the desired passenger waiting times compared to the time-window-based optimization method. The main reason behind the improved performance of the time-window-based method compared to the event-based methods is that the proposed time-window-based optimization method evaluates the bus holding options *holistically* within each time window; thus, enabling the coordination of the running buses.

The improvement with the use of the time-window-based optimization method is in the range of $\simeq 0-17\%$ as it is presented in Fig.4 (an improvement of 11% on average). The average waiting time difference of passengers at stops from the desired waiting time values was only 3 seconds in some cases whereas in others increased up to 58 seconds (i.e., during the time period 10:00-10:10). The main findings of the analysis are:

1. the consideration of the slack time and in-vehicle travel time constraints in the optimization process of the time-window-based method can improve the waiting time variance of passengers (by up to $\simeq 17\%$) while improving also the average in-vehicle travel times (by $\simeq 5\%$);

2. the performances of the time-window-based control and the one-headway-based control deteriorate following a similar trend when the external disturbance is significant and the estimated link travel time values are not close to the actual ones (travel time noise levels of 80%).

CONCLUSION

This study proposed a time-window-based optimization method for improving the waiting time variance of passengers from the planned waiting time values with the use of bus holding control. Past research works Cats et al. (*5*), Fu and Yang (*21*) have shown that event-based bus holding control methods and bus holding methods that hold the bus trips at more than two control point stops can be effective in reducing the passenger waiting time variance only up to a certain level. For this reason, this study investigated whether a time-window-based approach that computes the bus holding times of all running trips within a time period in a holistic manner can reduce further the passenger waiting time variance from the desired waiting time values.

The proposed time-window-based bus holding method was evaluated against the one-headway-based method and exhibited an improved performance with improvements of the passengers' waiting time variance of up to 17% in some specific time periods for bus line 1 in Stockholm where the control measures computed within each 10-minute optimization period were applied to a simulation scenario in SUMO. This simulation scenario replicated the actual operations by using the AVL and APC data from the daily operations for its calibration and the application of the bus holding measures computed by the time-window-based optimization demonstrated an improvement potential of $\simeq 5\%$ for the in-vehicle travel times and 11% (on average) for the waiting time variance of passengers at stops.

Although the effect of the travel time noise was considered in the validation phase, the developed time-window-based model does not consider the travel time stochasticity as a problem variable when computing the bus holding measures for improving the robustness of the solution. Works in event-based bus holding control such as Hickman (*15*) have considered the stochasticity of travel times in simple holding models that hold a bus at a single control point stop (the departure stop) because of the computational complexity. This challenge however is a very interesting topic for future investigation in order to determine the improvement potential of robust time-window-based control.

Author Contribution Statement

The authors confirm contribution to the paper as follows: study conception and design: K. Gkiotsalitis, O. Cats; data collection: O. Cats; analysis and interpretation of results: K. Gkiotsalitis, O. Cats; draft manuscript preparation: K. Gkiotsalitis, O. Cats.
All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] Gkiotsalitis, K. and N. Maslekar, Towards transfer synchronization of regularity-based bus operations with sequential hill-climbing. *Public transport*, Vol. 10, No. 2, 2018, pp. 335–361.

[2] Gkiotsalitis, K. and A. Stathopoulos, Demand-responsive public transportation re-scheduling for adjusting to the joint leisure activity demand. *International Journal of Transportation Science and Technology*, Vol. 5, No. 2, 2016, pp. 68–82.

[3] TfL, *Transport for London: Bus Routes and Borough Reports*. https://www.tfl.gov.uk/forms/14144.aspx, 2016, accessed: 2016-04-21.

[4] LTA, *Land Transport Authority: Bus Service Reliability Framework*. http://www.lta.gov.sg/data/apps/news/press/2014/20140124_BSRF(final2)-Annex.pdf, 2016, accessed: 2016-04-21.

[5] Cats, O., A. Larijani, Á. Ólafsdóttir, W. Burghout, I. Andreasson, and H. Koutsopoulos, Bus-holding control strategies: simulation-based evaluation and guidelines for implementation. *Transportation Research Record: Journal of the Transportation Research Board*, , No. 2274, 2012, pp. 100–108.

[6] Laskaris, G., O. Cats, E. Jenelius, and F. Viti, A real-time holding decision rule accounting for passenger travel cost. In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, IEEE, 2016, pp. 2410–2415.

[7] Kepaptsoglou, K. and M. Karlaftis, Transit route network design problem. *Journal of transportation engineering*, Vol. 135, No. 8, 2009, pp. 491–505.

[8] Farahani, R. Z., E. Miandoabchi, W. Y. Szeto, and H. Rashidi, A review of urban transportation network design problems. *European Journal of Operational Research*, Vol. 229, No. 2, 2013, pp. 281–302.

[9] Gkiotsalitis, K. and O. Cats, Exact optimization of Bus Frequency Settings considering Demand and Trip time variations. In *Transportation Research Board 96th Annual Meeting*, TRB, 2017.

[10] Gkiotsalitis, K. and O. Cats, Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways. *Transportation Research Part C: Emerging Technologies*, Vol. 88, 2018, pp. 187–207.

[11] Gkiotsalitis, K. and R. Kumar, Bus Operations Scheduling Subject to Resource Constraints Using Evolutionary Optimization. In *Informatics*, Multidisciplinary Digital Publishing Institute, 2018, Vol. 5, p. 9.

[12] Gkiotsalitis, K. and N. Maslekar, Multiconstrained Timetable Optimization and Performance Evaluation in the Presence of Travel Time Noise. *Journal of Transportation Engineering, Part A: Systems*, Vol. 144, No. 9, 2018, p. 04018058.

[13] Daganzo, C. F., A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B: Methodological*, Vol. 43, No. 10, 2009, pp. 913–921.

[14] Koehler, L. A., W. Kraus, and E. Camponogara, Iterative quadratic optimization for the bus holding control problem. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 4, 2011, pp. 1568–1575.

[15] Hickman, M. D., An analytic stochastic model for the transit vehicle holding problem. *Transportation Science*, Vol. 35, No. 3, 2001, pp. 215–237.

[16] Eberlein, X. J., N. H. Wilson, and D. Bernstein, The holding problem with real–time information available. *Transportation science*, Vol. 35, No. 1, 2001, pp. 1–18.

[17] Berrebi, S. J., E. Hans, N. Chiabaut, J. A. Laval, L. Leclercq, and K. E. Watkins, Comparing bus holding methods with and without real-time predictions. *Transportation Research Part C: Emerging Technologies*, Vol. 87, 2018, pp. 197–211.

[18] Sánchez-Martínez, G., H. Koutsopoulos, and N. Wilson, Real-time holding control for high-frequency transit with dynamics. *Transportation Research Part B: Methodological*, Vol. 83, 2016, pp. 1–19.

[19] Gkiotsalitis, K. and N. Maslekar, Dynamic Bus Operations Optimization with REFLEX. *NEC Tech. J*, Vol. 11, No. 1, 2016.

[20] Bartholdi, J. J. and D. D. Eisenstein, A self-coördinating bus route to resist bus bunching. *Transportation Research Part B: Methodological*, Vol. 46, No. 4, 2012, pp. 481–491.

[21] Fu, L. and X. Yang, Design and implementation of bus-holding control strategies with real-time information. *Transportation Research Record: Journal of the Transportation Research Board*, , No. 1791, 2002, pp. 6–12.

[22] Chen, Q., E. Adida, and J. Lin, Implementation of an iterative headway-based bus holding strategy with real-time information. *Public Transport*, Vol. 4, No. 3, 2013, pp. 165–186.

[23] Daganzo, C. F. and J. Pilachowski, Reducing bunching with bus-to-bus cooperation. *Transportation Research Part B: Methodological*, Vol. 45, No. 1, 2011, pp. 267–277.

[24] Zhao, J., S. Bukkapatnam, and M. M. Dessouky, Distributed architecture for real-time coordination of bus holding in transit networks. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4, No. 1, 2003, pp. 43–51.

[25] Xuan, Y., J. Argote, and C. F. Daganzo, Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. *Transportation Research Part B: Methodological*, Vol. 45, No. 10, 2011, pp. 1831–1845.

[26] Sun, A. and M. Hickman, The holding problem at multiple holding stations. In *Computer-aided systems in public transport*, Springer, 2008, pp. 339–359.

[27] Yu, B. and Z. Yang, A dynamic holding strategy in public transit systems with real-time information. *Applied Intelligence*, Vol. 31, No. 1, 2009, pp. 69–80.

[28] Shalaby, A. and A. Farhan, Bus travel time prediction model for dynamic operations control and passenger information systems. *Transportation Research Board*, Vol. 2, 2003.

[29] Mazloumi, E., G. Rose, G. Currie, and S. Moridpour, Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. *Engineering Applications of Artificial Intelligence*, Vol. 24, No. 3, 2011, pp. 534–542.

[30] Chien, S. I.-J., Y. Ding, and C. Wei, Dynamic bus arrival time prediction with artificial neural networks. *Journal of Transportation Engineering*, Vol. 128, No. 5, 2002, pp. 429–438.

[31] Osuna, E. and G. Newell, Control strategies for an idealized public transportation system. *Transportation Science*, Vol. 6, No. 1, 1972, pp. 52–72.

[32] Lee, J. D., D. V. McGehee, T. L. Brown, and M. L. Reyes, Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human factors*, Vol. 44, No. 2, 2002, pp. 314–334.

[33] Cortés, C. E., D. Sáez, F. Milla, A. Núñez, and M. Riquelme, Hybrid predictive control for real-time optimization of public transport systems' operations based on evolutionary multi-objective optimization. *Transportation Research Part C: Emerging Technologies*, Vol. 18, No. 5, 2010, pp. 757–769.

[34] Gkiotsalitis, K. and N. Maslekar, Improving Bus Service Reliability with Stochastic Optimization. In *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*, IEEE, 2015, pp. 2794–2799.

[35] Gkiotsalitis, K., Rolling-horizon optimization of the Bus Holding problem with Sequential Hill Climbing. In *97th Transportation Research Board Annual Meeting*, TRB, 2018.

[36] Levinson, H. S., Analyzing transit travel time performance. *Transportation Research Record*, , No. 915, 1983.

[37] Guenthner, R. P. and K. C. Sinha, *Modeling bus delays due to passenger boardings and alightings*. 915, 1983.

[38] Bertini, R. L. and A. M. El-Geneidy, Modeling transit trip time using archived bus dispatch system data. *Journal of transportation engineering*, Vol. 130, No. 1, 2004, pp. 56–67.

[39] Land, A. H. and A. G. Doig, An automatic method of solving discrete programming problems. *Econometrica: Journal of the Econometric Society*, 1960, pp. 497–520.

[40] Murty, K. G. and F.-T. Yu, *Linear complementarity, linear and nonlinear programming*, Vol. 3. Heldermann Berlin, 1988.

[41] Nocedal, J. and S. J. Wright, *Sequential quadratic programming*. Springer, 2006.

[42] SUMO, *SUMO: Definition of Vehicles, Vehicle Types, and Routes*. http://sumo.dlr.de/wiki/Definition_of_Vehicles,_Vehicle_Types,_and_Routes, 2017, accessed: 2017-12-04.

[43] Fu, L., Q. Liu, and P. Calamai, Real-time optimization model for dynamic scheduling of transit operations. *Transportation Research Record: Journal of the Transportation Research Board*, , No. 1857, 2003, pp. 48–55.