This is the unformatted version of: Sun, W., Schmöcker, J.-D. and Nakamura, T. (2021). On the trade-off between sensitivity and specificity in bus bunching prediction. Journal of Intelligent Transportation Systems: Technology, Planning and Operations, 25(4), 384-400. https://doi.org/10.1080/15472450.2020.1725887

On the trade-off between sensitivity and specificity in bus bunching prediction

Wenzhe Sun, Jan-Dirk Schmöcker, Toshiyuki Nakamura

Wenzhe Sun
Department of Urban Management, Kyoto University, Kyoto, Japan
orcid: 0000-0002-7305-8671
Email: wz.sun@trans.kuciv.kyoto-u.ac.jp

Jan-Dirk Schmöcker, corresponding author
Department of Urban Management, Kyoto University, Kyoto, Japan
orcid: 0000-0003-2219-9447
Email: schmoecker@trans.kuciv.kyoto-u.ac.jp

Toshiyuki Nakamura
Institute of Innovation for Future Society, Nagoya University, Nagoya, Japan
orcid: 0000-0003-3093-4229
Email: tnakamura@mirai.nagoya-u.ac.jp

On the trade-off between sensitivity and specificity in bus bunching prediction

Abstract

Bus bunching resulting from initially small headway irregularities is a widely-known and studied problem. A variety of headway-prediction approaches, as well as corrective strategies, have been developed to identify and correct headway irregularity in real time. Instead of predicting an exact value for future headways, this study explores a probabilistic predictive methodology to forecast whether or not a bus will be bunched during its dwelling at a downstream stop, using a logistic regression model based on GPS records of buses at least k stops upstream to allow for sufficient time to implement control strategies. A case study is conducted on a circular bus route in Kyoto City. Compared to two headway-based prediction approaches using linear regression and support vector machine, the superior performance of the proposed tool in detecting bunching is illustrated by Receiver Operator Characteristic (ROC) analysis. The high reliability in long-term prediction gives adequate time for operators to employ countermeasures. Besides, the proposed method provides operators with trade-off options. We find that a bunching-averse operator can obtain 95% “sensitivity”, that is the ratio of correctly identified bunching events, at the cost of decreasing “specificity”, which is the ratio of correct non-bunching predictions over all events. This is true even if the prediction horizon is more than 10 stops.

Keywords: bus bunching prediction; logistic regression; sensitivity and specificity; bus GPS data; multiple-stop-ahead prediction

1. Introduction

Bus bunching is a frequently occurring undesired event. Generally it can be defined as the phenomenon of two successive bus runs of a single line arriving at a stop within significantly shorter headways than the designed one. Bunching involving more than two buses is also regularly observed. Bus bunching may be initiated by the arrival of one bus run being delayed at an upstream stop. More passengers are likely to accumulate for the delayed bus at that stop and the bus is thus further delayed. Conversely, the subsequent run has fewer passengers to pick up and departs earlier than scheduled. Accumulated delay to the first vehicle and increasingly earlier arrival of the second one result in obvious inequality in dwell times and on-board passenger numbers. As the inequality aggravates over a sequence of stops, the scheduled headway is significantly shortened or eventually offset and the leading bus among bunched bus is often overcrowded.

Accurate prediction on headway or bunching itself can help to spotlight the coming bunching and further assist the operator to eliminate bunching in real time. A useful prediction tool is expected to a) have a long enough prediction horizon to allow the operator's implementation of countermeasures and b) provide information on the reliability of the prediction. The latter point is important in order to account for different preferences among operators. A bunching-averse operator is willing to frequently control the service to avoid any possible bunching, whereas some other operators may hesitate to take control action that will negatively impact some passengers, they thus only correct the predicted bunching of high confidence level. Therefore, this paper suggests a probabilistic binary prediction method.

This study aims to extend the existing literature in two aspects. Firstly, this study builds a logistic regression (LOGR) model to predict the likelihood of bunching to occur using bus GPS data, and tests the prediction performance under a wide range of prediction horizons varying from 1-stop-ahead to 15-stop-ahead, with an emphasis on multi-stop-ahead prediction and understanding the regularity deterioration pattern. Secondly, this study tries to enhance the robustness and flexibility of existing prediction tools. To achieve this Receiver Operator Characteristic (ROC) curves are utilized. This method is widely used in evaluating the performance of binary classification models and in this study it is interpreted as the optimal front of the proposed LOGR. This study explains how to conduct the trade-off between "sensitivity" and "specificity" from an operator's perspective.

The paper is organized as follows. After this introduction, Section 2 conducts a literature review on the corrective and predictive models addressing the bus bunching problem. The predictive methodology using LOGR is elaborated in Section 3. We point out that LOGR might be biased when used for "rare events data" as is the case in our example and provide a correction method. Then two headway-predicting algorithms: linear regression (LR) and support vector machine (SVM) are taken as the two benchmark approaches in this study and are also briefly introduced in this section. In Section 4, the characteristics of the collected data are described, including data collection period, average stop-to-stop travel time, average scheduled headway, fluctuation patterns for headway, etc. Based on this, a proper prediction horizon and bunching threshold are determined. The case study is described in Sections 5-7. In Section 5, the prediction performance of the two headway-predicting algorithms is discussed. The prediction performance of the proposed LOGR is evaluated and compared with headway-based methods in Section 6. The trade-off functionality of LOGR is discussed in Section 7. Conclusions and further work can be found in Section 8.

2. Literature review

Most of the relevant existing literature can be cast into two categories according to their objective: bunching prediction and corrective strategies. A large body of literature discussed how to eliminate bus bunching using analytical or simulation methods following the seminal work by Newell and Potts (1964). Osuna and Newell (1972) and Newell (1974) tried to maintain the bus schedule by a single control point. On the other hand, advanced control methods such as dynamic holding control proposed by Eberlein, Wilson, and Bernstein (2001), Daganzo (2009), Xuan, Argote, and Daganzo (2011), Bartholdi and Eisenstein (2012), Zhang and Lo (2018) and velocity control developed by Daganzo and Pilachowski (2011) as well as stop skipping discussed by Sun and Hickman (2005) assume frequent and efficient communication between bus drivers and the control center. Berrebi et al. (2018) tested the control strategies proposed by Dagazo (2009), Xuan et al. (2011), Bartholdi and Eisenstein (2012), Daganzon and Pilachowski (2011), Berrebi, Watkins, and Laval (2015) on a bus route in Portland, Oregon. The experiment was based on bus automatic vehicle location (AVL) data, automatic passenger counter (APC) data and traffic signal data. The effectiveness of each strategy to stabilize bus headways was confirmed. Further, the effect of incorrect future headway prediction on each strategy was discussed. The variance of controlled headway was found rising significantly as the prediction errors increased. Instead of actively adjusting the headway, Schmöcker, Sun, Fonzone, and Liu (2016), Wu, Liu, and Jin (2017), Sun and Schmöcker (2018) discussed passive strategies such as passenger re-distribution and overtaking which are activated when bunching occurs. These strategies aim to equalize passenger boarding numbers for bunched buses through queue management.

Substantial development in data collection technology recently gives scholars access to massive bus operation data including AVL, APC and automatic fare collection (AFC) data, and has led to a large number of studies concerning real-time prediction of bus operational aspects. Rather than predicting bus bunching events, most existing literature focuses on bus arrival time and headway. Though closely related, this literature can again be grouped into three subcategories: bus trajectory, bus arrival time and headway prediction. Complete bus trajectory prediction is most challenging but also most informative. It provides predicted stop arrival and departure time, stop-to-stop travel time, as well as the headway between consecutive buses for bus operators and users. Hans, Chiabaut, Leclercq, and Bertini (2015) developed a sequential mesoscopic simulation that elaborately considered the stochastics generated during bus dwell time and link travel time. A bundle of possible future trajectories is simulated based on the distribution assumed for the time components in a bus trip and the associated parameters are calibrated with AVL, APC and traffic signal data. This method delivering robust prediction results to the operator. Distribution or range for future arrival time and headway can also be easily obtained. A shortcoming of this method is that the predicted range of arrival time or headway might be too wide to be conclusive for operators' decision making. Recent research by Dai, Ma, and Chen (2019) also modeled bus dwell time and link travel time in detail to reproduce the trip travel time variability for a bus line. They specifically considered the bus waiting time due to the interaction between buses at the stop intersected by multiple bus lines, which is also defined as common-line bunching in Schmöcker et al. (2016). They inferred the probabilities of the bus from a specific line queueing (bunching) after the other common lines at the stop from bus GPS data. Yu, Chen, Wu, Ma, and Wang (2016) conducted a solid literature review on the methods addressing bus arrival time prediction. They reviewed the implemented data source and algorithm of each relevant literature. SVM, Kalman filter (KF), k-nearest neighbor (KNN), artificial neural network (ANN) and regression-based methods are frequently used. Yu, Yang, and Yao (2006) made a successful attempt at predicting bus arrival time based on SVM method and AVL data. Yu, Lam, and Tam (2011) used SVM, ANN, KNN and LR to predict arrival time for a 0.7km common line section where more than 10 bus routes overlapped in Hong Kong. Kumar, Vanajakshi, and Subramanian (2018) combined KF and KNN to tackle the prediction of bus travel time and arrival time. In this hybrid model, KNN classifier is used to refine the model input of KF model.

Future headway is the difference between the predicted arrival times of two consecutive buses and can be obtained by the arrival time prediction model. There are also some studies directly focusing on the prediction of headway itself. Yu, Wu, Chen, and Ma (2017) proposed a probabilistic prediction approach using RVM (Relevance Vector Machine) to attach a confidence interval for each predicted headway for 2- and 3-stop-ahead. Outperformance with respect to robustness was concluded by comparing the results with the deterministic single values derived by SVM, KF, KNN and ANN algorithms. Andres and Nair (2017) integrated headway prediction and bus holding control strategies. Regression, ANN and autoregressive models are used in their work to predict future headways with 5min and 10min prediction horizons. The prediction results are applied as input to an analytical model extending Daganzo (2009).

Although headway prediction methods have made great advancement, it remains a challenging work to precisely identify coming bunching events in multiple-stop-ahead prediction. The accuracy of bunching prediction is heavily dependent on the reliability of headway prediction whose results deteriorate gradually as the prediction horizon extends. Yu, Chen, Wu, Ma, and Wang (2016) used several well-developed algorithms to predict headway first then convert the result to binary bunching occurrence. 2min RMSE is obtained for headway and 99% sensitivity is realized for bunching in 2-

stop-ahead prediction, but the performance deteriorates to 6min RMSE and 73% sensitivity for 5-stop-ahead prediction. Moreira-Matias, Cats, Gama, Mendes-Moreira, and De Sousa (2016) built a regression-based model to predict the headway for a downstream stop and calculate the likelihood of bus bunching to occur for all the further downstream stops. The focus of their study was to propose a proactive control framework in which every suspicious event triggers a bunching alarm. The effect of bunching likelihood thresholds was not investigated. It should be noted that Moreira-Matias et al. (2016), Andres and Nair (2017), Berrebi et al. (2018) combined predictive and corrective models, and tested the feasibility and benefit of putting control strategies into practice. Instead of bunching prediction, Arriagada et al (2019) used bus GPS data and smartcard data to investigate the causes of bus bunching, with an emphasis on the planning side. Scheduled frequency, stop location and configuration (number of the berths), traffic signal and bus lane design are found influential. This research provides insight into bunching prevention in the planning stage.

3. Methodology

3.1 The identification of bus bunching event

As a bunching event involves two buses we refer to these as front bus and back bus respectively. Let a binary variable b_m^n denote whether bus run *m* is caught in bunching as the back bus during its dwelling at stop *n*. a_m^n and d_m^n denote the arrival and departure time of bus run *m* at stop *n* respectively. At stop *n*, for each bus run m ($m \geq 2$) we can obtain $\Delta_{m-1,m}^n$ which is the time interval between the arrival time of bus *m* and the departure time of bus *m*-1 in Eq. (1). Bus run *m* is considered bunched with bus run *m*-1 at the stop when $\Delta_{m-1,m}^n$ is below a threshold Δ_0. The threshold can be determined by the operator. Yu et al. (2016) and Moreira-Matias et al. (2016) used 1/4 of the scheduled headway. $\Delta_{m-1,m}^n$ is defined as the departure-to-arrival headway in this study. Different from arrival-to-arrival or departure-to-departure headway, $\Delta_{m-1,m}^n$ is negative when two buses overlap at the stop. As overtaking is not allowed, for each stop *n*, bus *m*-1 always arrives and departs earlier than bus *m*, and accordingly time interval $\Delta_{m-1,m}^n$ can always be obtained before the departure of bus *m*.

$$\Delta_{m-1,m}^n = a_m^n - d_{m-1}^n \tag{1}$$

For each bus *m* ($m \geq 2$), the binary bunching status b_m^n can be derived by Eq. (2)

$$b_m^n = \begin{cases} 1, \Delta_{m-1,m}^n \leq \Delta_0 \\ 0, \Delta_{m-1,m}^n > \Delta_0 \end{cases} \tag{2}$$

3.2 Variable selection

Following afore reviewed literature, the continuous $\Delta_{m-1,m}^n$ can be used as the dependent variable for headway-prediction approaches. For bunching prediction then an additional step is required judging whether the predicted headway is below a prior defined bunching threshold or not. Instead, in this study, b_m^n is used as the dependent variable of the logistic regression to directly predict the binary bunching status and bunching probabilities.

Gradually accumulated or suddenly significant inequality in dwell time and travel time might lead two successive buses to be bunched. The back bus in a bunching event tends to have a shorter forward-looking headway, negative deviation from timetable (ahead of schedule), less on-board passengers and shorter dwell time than those of front buses in a bunching event or of non-bunched buses (Degeler Heydenrijk-Ottens, Luo, Oort, & Lint, 2018). Yu et al. (2016) used boarding and alighting numbers of two successive buses, link travel time and headway at an upstream stop as the input to their headway-based prediction approach. As only bus GPS data is used in this study, information regarding boarding, alighting as well as on-board passengers are not available. Instead dwell time is included in the variable in addition to headway. Deviation from the timetable is excluded here, as bus dispatching is not based on the timetable in some cities and the data for this variable might not be available. To conclude, dwell time of two successive buses and their headway at an upstream stop *n-k* are used as the main leading indicators of a coming bunching event in the *k*-step-ahead prediction. The detailed notation is as follows:

k	prediction horizon in terms of number of stops, k = 1,2,3,…*N*-1 and *N* denote the last stop of the bus route
t_m^{n-k}	dwell time of bus run *m* at stop *n-k*
t_{m-1}^{n-k}	dwell time of bus run *m-1* at stop *n-k*
$\Delta_{m-1,m}^{n-k}$	time interval between the arrival time of bus *m* and the departure time of bus *m-1* at stop *n-k*

We always have $n>k$, so that the *k*-stop-ahead prediction cannot be carried out until bus run *m* passes the initial *k* bus stops, e.g. the prediction starts from stop 6 in the 5-stop-ahead prediction by using the data at stop 1. Also note that $m \geq 2$ and that the first bus has zero probability to be bunched as the back bus.

3.3 Logistic regression

Logistic regression (LOGR) modeling is widely used in classification problems. In binary classification, it not only helps to categorize observations into positive or negative class but also interprets the causality by producing the significance of each independent variable. Moreover, it computes the probability of each observation to be in the positive or negative class. The binary bunching status from the perspective of the back bus b_m^n (m≥2) is taken as the dependent variable. t_m^{n-k}, t_{m-1}^{n-k}, and $\Delta_{m-1,m}^{n-k}$ are the independent variables. Let $\boldsymbol{X}_m^n = [t_m^{n-k},\ t_{m-1}^{n-k},\ \Delta_{m-1,m}^{n-k}]$, then the probability of bus run m being bunched at stop n as a back bus can be derived as

$$Pr(b_m^n = 1|X_m^n) = \frac{1}{1 + e^{-\boldsymbol{\beta} \boldsymbol{X}_m^n}} \tag{3}$$

With parameters $\boldsymbol{\beta} = [\beta_0, \beta_1, \beta_2, \beta_3]$ estimated by fitting the model with real data, $Pr(b_m^n = 1|\boldsymbol{X}_m^n)$ for each bus run m (m≥2) at any stop n (n>k) can be computed k-stop ahead in the prediction stage. b_m^n is predicted to be positive (one-event) if $Pr(b_m^n = 1|\boldsymbol{X}_m^n)$ exceeds a probability threshold Pr_x which is also known as the cut-off point, otherwise, negative (zero-event), as in Eq. (4).

$$b_m^n = \begin{cases} 1, Pr(b_m^n = 1|\boldsymbol{X}_m^n) > Pr_x \\ 0, Pr(b_m^n = 1|\boldsymbol{X}_m^n) \leq Pr_x \end{cases} \tag{4}$$

3.4 Rare events bias

Irregular arrivals are common in bus transit operation, however few of them turn into severe bunching. The prior bunching probability which is the ratio of bunching occurrence to the total number of dwelling in Yu et al. (2016) varies from 3% to 17%, from 0.15% to 7.17% in Moreira-Matias et al. (2016), and from 3% to 9% in our 5-day testing data. Bunching is hence a "rare" event in the dataset.

"Rare events data" refer to large datasets in which it is significantly less likely that the binary dependent variables take one than zero. King and Zeng (2001) considered events such as wars, natural disasters or epidemiological infections within long term time series data. They found logistic regression underestimates the probability of rare events because they tend to be biased towards the majority class, which is the less important class in most cases. This can be explained as follows:

The dependent variable Y_i follows a Bernoulli probability distribution that can take the values of one and zero with probabilities π_i and $1 - \pi_i$ respectively. The probability function can be written as

$$Pr(Y_i|\pi_i) = {\pi_i}^{Y_i}(1 - \pi_i)^{1-Y_i} \tag{5}$$

It is easy to derive the expectation and variance of Y_i as

$$E(Y_i) = \pi_i \tag{6}$$

$$V(Y_i) = \pi_i(1 - \pi_i) \tag{7}$$

If the regression model has some explanatory power, the variance in the dependent variable has to be large enough. The variance becomes larger as π_i increases and reaches its maximum if π_i = 0.5, which indicates that it is favorable to involve an equal number of ones and zeros in the dataset. Cosslett (1981) and Imbens (1992) also showed that equally sampling the two classes is optimal.

King and Zeng (2001) further discuss that selective data collection strategies instead of sampling all available events could save data collection costs and correct the bias. Maalouf and Trafalis (2011) implemented kernel logistic regression to rare events data, making use of a fast and robust adaptation of kernel logistic regression and taking the weight of rare events into account. In this paper, selective sampling and corresponding prior correction are used to reduce bias induced by rare events.

3.4.1 Sampling

Since bus GPS records are plentiful and easy to filter, efficient sampling thus can be achieved by creating a balanced dataset in which all bunching events are included and part of the non-bunching events are excluded. A balanced selection to include ones (bunching) and an equal number of zeros (non-bunching) is applied.

3.4.2 Prior correction

Following King and Zeng (2001), prior correction is to correct the estimates according to the fraction of ones in the population, denoted by τ, and the observed fraction of ones in the sample, denoted by $\bar{y}$, since the probability of events to be predicted as ones is overestimated in the sample. The correction is applied to the intercept β_0 as

$$\widehat{\beta_0} = \beta_0 - \ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right] \tag{8}$$

3.5 LR and SVM as benchmark solutions

We now turn to two headway prediction methods that we consider as benchmarks compared to the afore introduced direct bunching prediction method. Firstly, we consider linear regression (LR) which is a basic tool in addressing prediction problems. To make LR comparable with LOGR, the same set of independent variables $\boldsymbol{X}_m^n = [t_m^{n-k},\ t_{m-1}^{n-k},\ \Delta_{m-1,m}^{n-k}]$ is applied. With $\boldsymbol{\beta}' = [\beta_0', \beta_1', \beta_2', \beta_3']$ the relationship between the headway at stop *n* and the set of the independent variables containing information *k*-stop-ahead is modeled as

$$\Delta_{m-1,m}^{n} = \boldsymbol{\beta}' \boldsymbol{X}_{m}^{n} \tag{9}$$

Secondly, support vector machine (SVM) can map a non-linear relationship for model input and output, and is tested by a number of studies in predicting bus headway or arrival time (B., Yu et al., 2006; 2011; H., Yu et al., 2016). The same independent variables and dependent variable are applied to the SVM regression, and a RBF (Radial Basis Function) kernel is selected because it is found both efficient for bus arrival time prediction (Yu et al., 2011) and for bus headway prediction (Yu et al., 2016).

4. Data description and case study settings

Buses are the main mode of public transport in Kyoto, Japan with more than 100 lines being served by several operators. Bus GPS data of two primary bus operators has been obtained for a period of six months in 2016. The data is collected every 8 seconds and provides the geographic coordinates of bus location in real-time as well as associated bus line and vehicle number. Due to the lack of stop-based information, it is essential to identify arrival and departure times for each bus run at each stop. Using bus stop coordinates the distances of a bus from previous and next stops can be computed for every GPS record. Considering that bunching and traffic congestion might make it difficult for the bus driver to stop the bus at the exact bus stop coordinates as well as inaccuracy of GPS records, the bus is regarded arriving at the stop once it approaches the bus stop within 30m. In the same way the departure time is obtained when the GPS records indicate that the bus has moved 30m from the bus stop.

The data collection period includes the months of April and November. During these months, Kyoto City experiences vast numbers of domestic and foreign visitors who come to enjoy the cherry blossoms (April) and red leaves (November) in various sites around the city. The bus operators thus encounter a huge challenge during these seasons to deliver a reliable service.

A circular bus line, Kyoto City Bus No. 205, which connects the city center, railway station and several famous tourist attractions (Figure 1(middle)) is selected for the case study. There are 53 stops on this bus line in total. To exclude the effect of dispatching at the terminal and factors for which we do not have data (e.g. crew shedule, departure time adjustments), the 2nd stop of the line is taken as the initial stop and the 52nd stop as the last one so that each bus run passes 51 bus stops. Data of five weekdays in April 2016 are used as the training dataset and data of another five weekdays in the same month are used for testing the model.

The scheduled headway varies from hour to hour, and the mean scheduled headway at the initial stop is 6.97min from 6 am to 8 pm. The shortest scheduled headway is 3min at 7 am. Based on this, 1min is used for the bunching threshold as larger threshold can include headway variance that does not lead to bunching.

Adequate time is required to project a successful correction, in particular, if the control strategy is based on manual communication between the dispatcher and the bus drivers. In this study, the proposed approach is tested under a long prediction horizon of 10 stops or more which gives the operator more than 15min to react since the mean stop-to-stop travel time is 1.77min.

Figure 2 illustrates the bus runs departing from the initial stop between 8 am and 10 am. Bunching occurs frequently along the bus line. Bus runs that are involved in bunching as the back bus of two or more buses at least once are denoted in red, and the front buses of a bunching sequence are denoted in blue. Buses in green are not involved in any bunching. The headway fluctuation patterns of seven red trajectories are demonstrated in Figure 3. Because of the bunching effect, the forward-looking headway of back buses fluctuate within a small range, but always below one minute, once bunching has been occurring giving further support to our threshold choice of one minute.

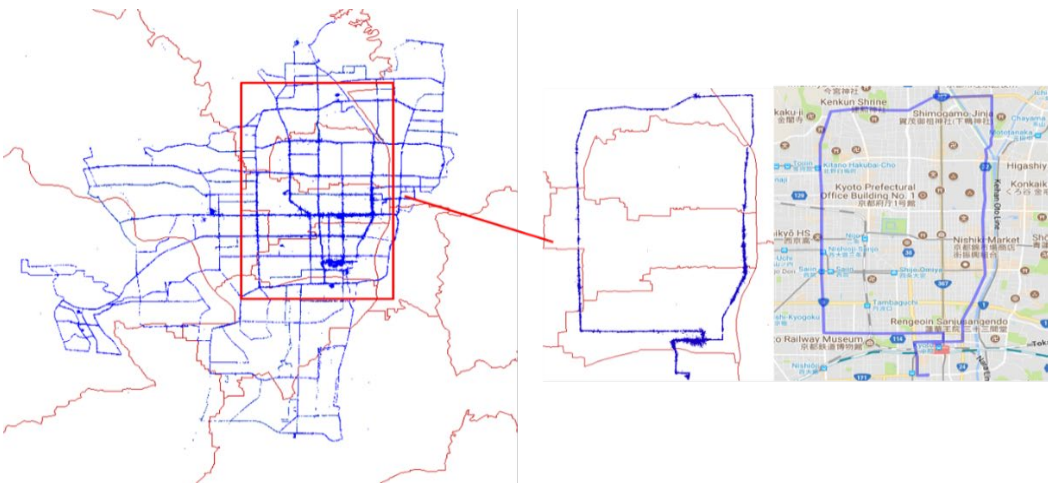


Figure 1. Data collected (left), data of Kyoto City Bus No. 205 (middle) and its configuration on real map (right).

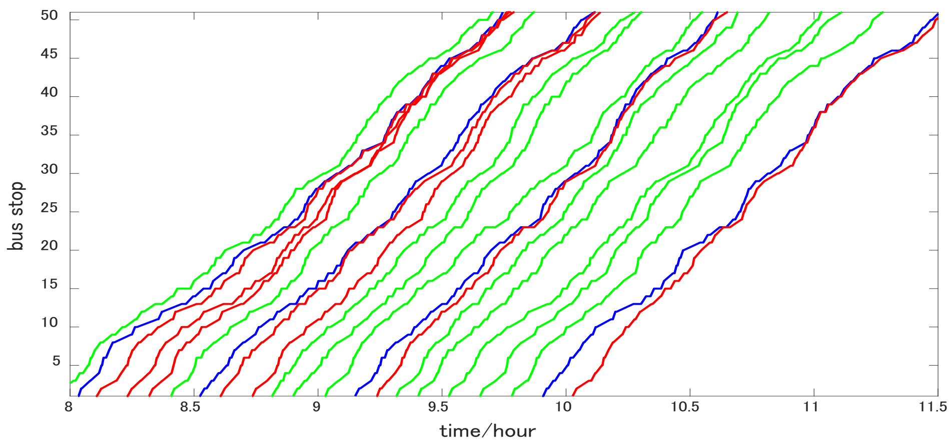


Figure 2. Trajectories of Kyoto City Bus No. 205 in one day of April 2016

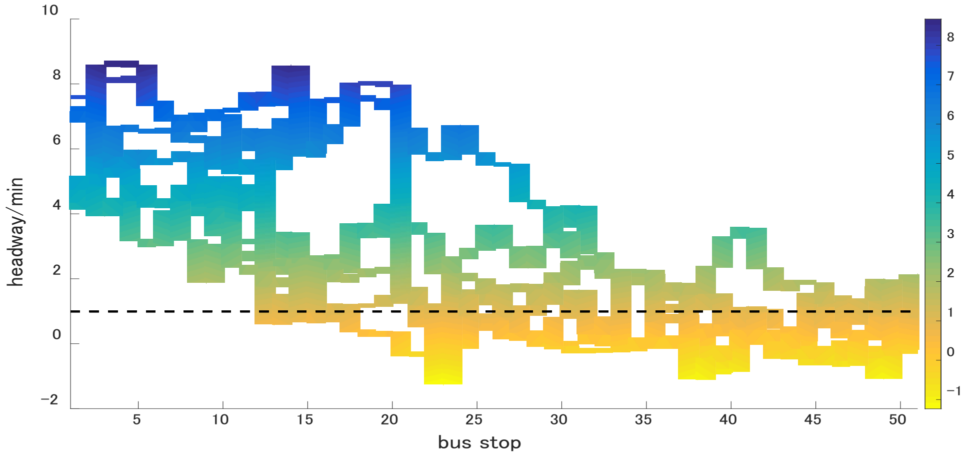


Figure 3. Headway fluctuation along the line for bunched buses

5 Headway prediction

In the following case study, the headway prediction results derived by LR and SVM are discussed at first including a comparison of these results. In Section 6 then the focus is on the bunching prediction using these two methods as well as the newly proposed LOGR model. In the third part of our case study we compare the measures using ROC curves.

5.1 Linear regression

Table 1 shows the estimation results of the fitted LR model. For all the prediction horizons, the headway between the target bus and its front bus $\Delta_{m-1,m}^{n-k}$ and the dwell time of the target bus t_m^{n-k} are always significant at 0.1% level and have positive signs. For short-term prediction the coefficient of $\Delta_{m-1,m}^{n-k}$ is close to 1 and it begins to deviate from 1 as the prediction horizon increases. Meanwhile the coefficient of t_m^{n-k} increases gradually as the prediction horizon extends. t_{m-1}^{n-k} is insignificant in some cases, but it is still considered an important variable indicating at-stop activities and passenger loads. Long t_{m-1}^{n-k} may shorten the headway, but it sometimes results from in-vehicle crowding as well as high boarding demand which may cause boarding failures that lead the following bus to dwell longer and increase the headway, thus the sign of t_{m-1}^{n-k} is inconclusive but mostly negative.

Table 1. Coefficients of the independent variables in the LR model

Prediction horizon	Intercept	$\Delta_{m-1,m}^{n-k}$	t_m^{n-k}	t_{m-1}^{n-k}	Adjusted R^2
1-stop-ahead	-0.3604***	1.0009***	0.7084***	0.1247***	0.9681
2	-0.3689***	1.0026***	0.7735***	0.0381*	0.9431
3	-0.4890***	1.0037***	0.9786***	0.0845***	0.9173
4	-0.4664***	1.0051***	0.9611***	0.0203	0.8922
5	-0.5081***	1.0065***	1.0315***	0.0115	0.8673
6	-0.4676***	1.0075***	0.9974***	-0.0751*	0.8424
7	-0.5320***	1.0092***	1.0664***	-0.0247	0.8183
8	-0.5070***	1.0108***	1.0177***	-0.0738*	0.7942
9	-0.5458***	1.0111***	1.1231***	-0.0931*	0.7710
10	-0.5613***	1.0123***	1.1489***	-0.1072**	0.7474
11	-0.5937***	1.0136***	1.1834***	-0.0932*	0.7245
12	-0.6405***	1.0160***	1.2057***	-0.0459	0.7012
13	-0.6283***	1.0186***	1.1944***	-0.0935	0.6790
14	-0.6675***	1.0212***	1.2323***	-0.0886	0.6564
15	-0.6325***	1.0235***	1.2236***	-0.1681***	0.6349

*** <= 0.001, ** <= 0.01, * <= 0.05

5.2 Support vector machine

The RBF function has two tuning parameters (C, γ) to enhance the predicting power of the SVM model. C is the cost parameter to penalize the misclassifying of a sample. C thus controls the complexity of the classifier; a high C may greatly bend the "prediction hyperplane" to avoid any misclassifying (Cherkassky and Ma, 2004). γ is the inverse of the radius of influence by the samples selected as the support vectors of the model. γ determines the influence of a single sample, a high γ thus may reduce the radius and limit the generalization performance of the model. According to the findings on bus arrival time prediction in Yu et al. (2011), $C \in [2^{-5}, 2^5]$, $\gamma \in [0.1, 0.3]$ are recommended for the two parameters. In this paper, $(2^2, 1)$ is set for the two parameters after a grid search in which $\gamma = 1$ performs better in our dataset.

5.3 Performance evaluation index

MAPE (Mean Absolute Percentage Errors) and RMSE (Root Mean Square Errors) are commonly used to evaluate the prediction performance regarding exact value arrival time or headway prediction. Let M and N denote the total number of bus runs and stops for a bus line, $\Delta_{m-1,m}^{n-k}$ and $\hat{\Delta}_{m-1,m}^{n-k}$ denote the actual value and predicted value for headway, MAPE and RMSE are obtained respectively in Eq.

(10) and Eq. (11). In order to prevent the denominator being close to zero, we follow the method of Yu et al. (2016) to calculate MAPE and use the mean of actual headways $\bar{\Delta}$ instead of $\Delta_{m-1,m}^{n-k}$

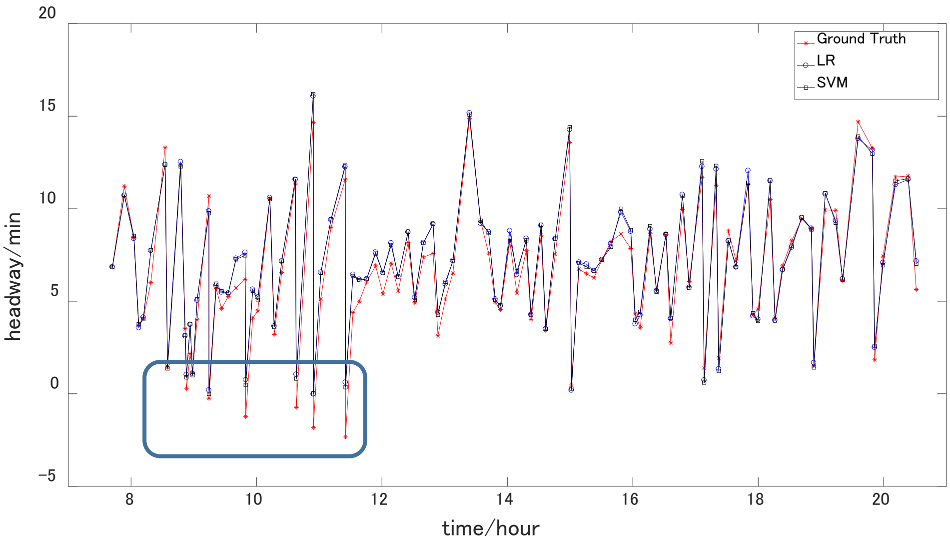
$$MAPE = \frac{1}{(M-1)(N-k)} \sum_{n=k+1}^{N} \sum_{m=2}^{M} \left| \frac{\Delta_{m-1,m}^{n-k} - \hat{\Delta}_{m-1,m}^{n-k}}{\bar{\Delta}} \right| \times 100\% \tag{10}$$

$$RMSE = \sqrt{\frac{1}{(M-1)(N-k)} \sum_{n=k+1}^{N} \sum_{m=2}^{M} (\Delta_{m-1,m}^{n-k} - \hat{\Delta}_{m-1,m}^{n-k})^2} \tag{11}$$

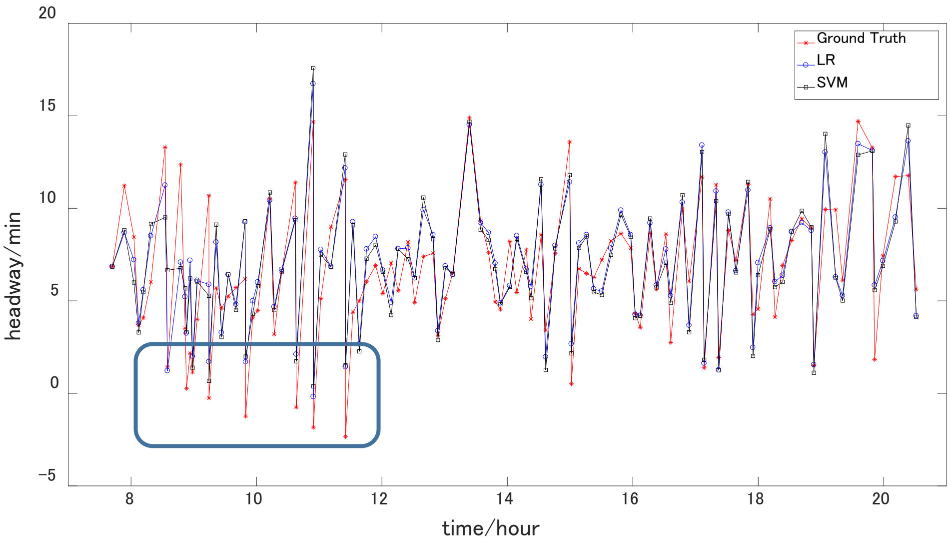
5.4 Performance comparison

Headway prediction results at Stop 23 "Kinkaku Temple", one of the most frequented sightseeing spots in Kyoto, is used to illustrate the performance of the aforementioned two methods. The results of 1-stop-ahead and 10-stop-ahead predictions are illustrated in Figure 4 and evaluated in Figure 5.

Reliable prediction results (MAPE = 7.42% and RMSE = 0.71min by LR, MAPE = 7.45% and RMSE = 0.71min by SVM) are produced for 1-stop-ahead prediction. For 10-stop-ahead prediction, the results obviously deteriorate (MAPE = 21.64% and RMSE = 1.93min by LR, MAPE = 21.51% and RMSE = 1.92min by SVM). We suggest they can still provide insights into expected fluctuation patterns downstream, but the exact value is not reliable. Furthermore, neither in 1- nor 10-stop-ahead prediction can these two methods perform favorably under the circumstance that the actual headway becomes extremely short and bunching is going to happen, as is highlighted by the blue box in Figure 4. Furthermore, Figure 5 illustrates that in terms of MAPE and RMSE, both methods produce close prediction accuracy and deteriorate similarly. Instead of significant increases in prediction errors, evaluation metrics deteriorate gradually as the prediction horizon extends.

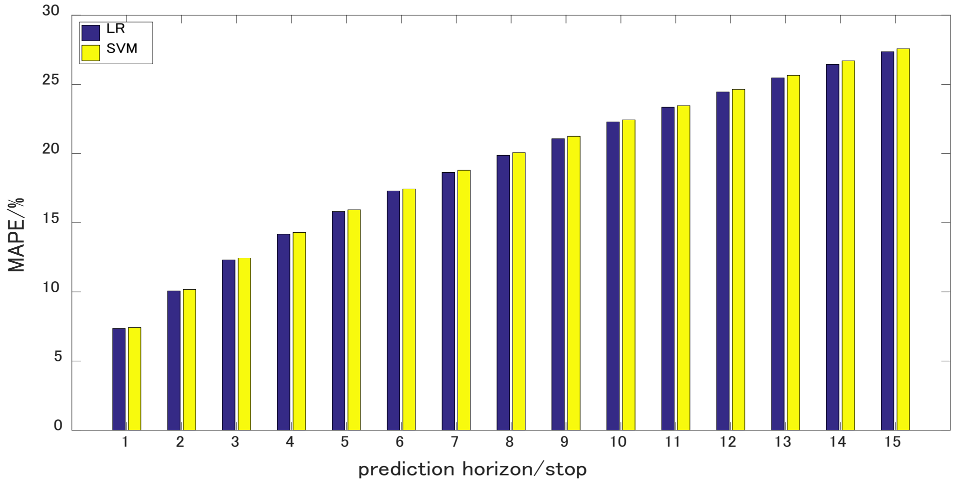


(a) 1-stop-ahead headway prediction

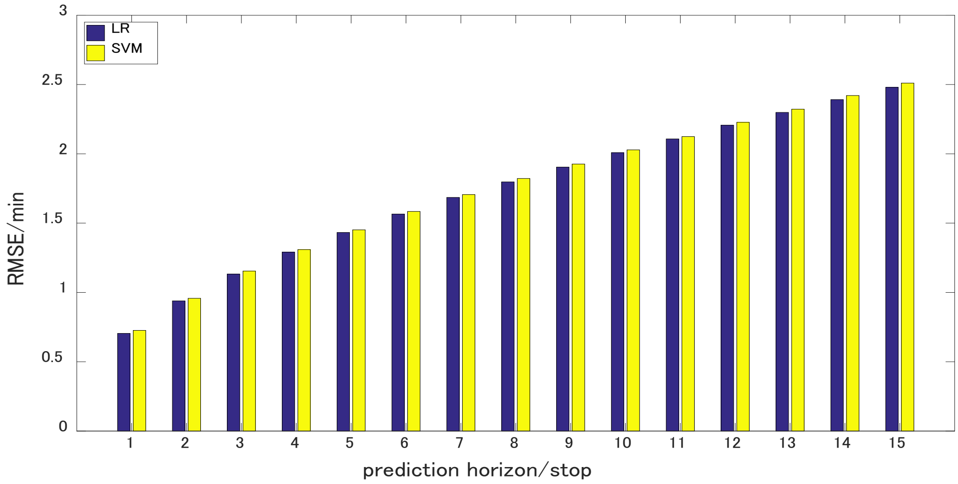


(b) 10-stop-ahead headway prediction

Figure 4. Performance comparison in terms of exact headway value



(a) Deterioration in MAPE as the prediction horizon extends



(b) Deterioration in RMSE as the prediction horizon extends

Figure 5. Performance comparison in terms of RMSE and MAPE under various prediction horizons

6 Bunching prediction

6.1 Logistic regression

We now focus on bunching prediction, firstly with logistic regression. Estimation results with and without rare events bias correction are shown in Table 2. Adjusted McFadden's R^2 obtained by Eq. (12) is selected to measure the overall goodness of fit for the logistic regression model.

$$R_{MCF}^2 = 1 - \frac{lnL_{full} - K}{lnL_{null}} \tag{12}$$

where L_{full} is the likelihood derived by the fitted model, and L_{null} is the likelihood of a null model with intercept as the only predictor. K is the number of independent variables in the proposed model. Due to the randomness generated by drawing non-bunching observations from the 5-day dataset to correct the rare event bias, we run the model for 100 times and report the mean values for the coefficients and adjusted McFadden's R^2. The significance is not based on any specific run but on all the 100 runs, and for each variable the p-value is obtained by one sample t-test on the 100 estimated coefficients. Bunching probability is negatively correlated with the value of headway, thus the coefficients of the variables in the fitted LOGR have a reversed sign compared to those in the LR model. The correction is proven effective as the adjusted R_{MCF}^2 is increased by at least 0.05 for each prediction horizon.

Table 2. Coefficients of the independent variables in the LOGR model

Prediction horizon	Intercept	$\Delta_{m-1,m}^{n-k}$	t_m^{n-k}	t_{m-1}^{n-k}	Adjusted R_{MCF}^2
Without correction					
1-stop-ahead	3.0842***	-2.2341***	-1.4015***	-0.3552***	0.7508
2	2.3653***	-1.7302***	-1.0123***	-0.1913*	0.6899
3	2.0826***	-1.4265***	-1.1577***	-0.1057	0.6357
4	1.8370***	-1.2634***	-1.0737***	0.0149	0.6016
5	1.6296***	-1.1234***	-0.9013***	-0.0036	0.5651
6	1.5044***	-1.0363***	-0.9266***	0.1282	0.5385
7	1.4301***	-0.9596***	-0.8913***	0.1007	0.5097
8	1.2780***	-0.8957***	-0.7406***	0.1680*	0.4838
9	1.2385***	-0.8352***	-0.8607***	0.1960**	0.4558
10	1.1162***	-0.7811***	-0.8076***	0.2664***	0.4289
11	1.1189***	-0.7361***	-0.7486***	0.0807	0.4028
12	1.0504***	-0.7003***	-0.7425***	0.1446*	0.3819
13	0.9763***	-0.6679***	-0.6612***	0.1593*	0.3615
14	0.9184***	-0.6403***	-0.5862***	0.1666**	0.3431
15	0.8498***	-0.6137***	-0.5155***	0.1807**	0.3246
With correction (mean values of 100 runs reported with significance also based on all runs)					
1-stop-ahead	2.3647***	-2.0343***	-1.1165***	0.0619***	0.8214
2	1.8730***	-1.5919***	-0.8407***	0.0385*	0.7732
3	1.7571***	-1.3429***	-1.1465***	0.0985***	0.7267
4	1.6035***	-1.2226***	-0.9988***	0.1627***	0.6971
5	1.4231***	-1.0874***	-0.9060***	0.1598***	0.6568
6	1.2997***	-1.0042***	-0.8988***	0.2585***	0.6248
7	1.2399***	-0.9326***	-0.7756***	0.1574***	0.5938
8	1.1046***	-0.8738***	-0.7128***	0.2740***	0.5641
9	1.1111***	-0.8146***	-0.8855***	0.2514***	0.5311
10	1.0183***	-0.7663***	-0.8584***	0.3289***	0.5023
11	0.9848***	-0.7188***	-0.7544***	0.1642***	0.4709
12	0.9108***	-0.6817***	-0.7792***	0.2616***	0.4461
13	0.8268***	-0.6467***	-0.7202***	0.2776***	0.4195
14	0.8020***	-0.6183***	-0.6655***	0.2246***	0.3970
15	0.6966***	-0.5887***	-0.5887***	0.2788***	0.3739

*** <= 0.001, ** <= 0.01, * <= 0.05

6.2 Performance evaluation index

We define an actual bunching as "observed positive" and a predicted bunching as "predicted positive". Similarly, for non-bunching we define "observed negative" and "predicted negative". All the prediction results can be cast into four categories as is shown in Table 3, e.g. it is a true positive if an observed bunching is correctly labeled one in the prediction outcomes. Four indexes can be obtained from Eqs. (13) to (16). A binary classifier with high true positive rate and high true negative rate is desired. The former is commonly referred to as "sensitivity" and the latter as "specificity". Sensitivity, specificity and accuracy which is an index computed with Eq. (17) to indicate overall prediction performance, are applied to evaluate the binary classification performance of the three algorithms.

Table 3. Four categories for binary classification results

	Observed positive (OP)	Observed negative (ON)
Predicted positive (PP)	True positive (TP)	False positive (FP)
Predicted negative (PN)	False negative (FN)	True negative (TN)

$$\text{True positive rate (TPR, sensitivity, SES)} = \frac{\sum \mathrm{TP}}{\sum \mathrm{OP}} \tag{13}$$

$$\text{False positive rate (FPR)} = \frac{\sum \mathrm{FP}}{\sum \mathrm{ON}} \tag{14}$$

$$\text{True negative rate (TNR, specificity, SPC)} = \frac{\sum \mathrm{TN}}{\sum \mathrm{ON}} \tag{15}$$

$$\text{False negative rate (FNR)} = \frac{\sum \mathrm{FN}}{\sum \mathrm{OP}} \tag{16}$$

$$\text{Accuracy (ACC)} = \frac{\sum \mathrm{TP} + \sum \mathrm{TN}}{\sum \mathrm{OP} + \sum \mathrm{ON}} \tag{17}$$

For headway-based methods, only one combination of sensitivity and specificity is derived, as headway prediction produces an exact value for each headway, resulting in deterministic true positive and negative outcomes. Instead, by using logistic regression different combinations are obtained depending on the cut-off point applied to the predicted probability. The cut-off point is the threshold to determine the predicted positive. The event is judged as positive if its predicted probability exceeds the cut-off point. A high cut-off point tends to only identify events presenting convincingly high probability as positives, and consequently, it thus might misclassify observed positives as negative. Vice versa, a low cut-off point will lead to more false positives. Therefore the cut-off point choice should depend on the operator's attitude towards bunching. Two scenarios are assumed here to represent operators with different weights to false negative errors (missing actual bunching). Moreira-Matias et al (2016) employed a large weight of 10:1 for false negative compared to false positive for aggressive control purposes. We consider more moderate weights of 1:1 and 3:1.

Scenario 1 (LOGR-N): the operator is bunching-neutral, and gives equal weight to false positive and false negative.

Scenario 2 (LOGR-A): the operator is bunching-averse, and gives a 3:1 weight to false negative over false positive predictions.

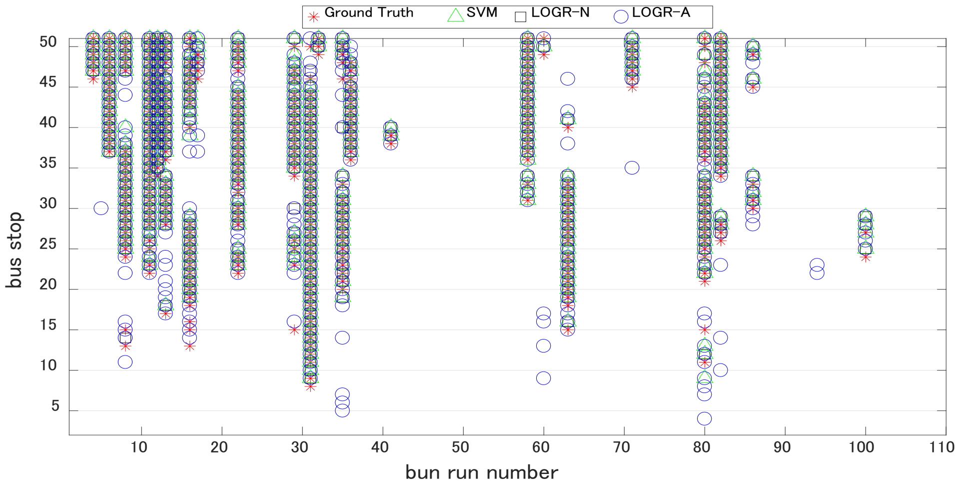
The cost function in Eq. (18) computes the total weighted errors given a cut-off point. For LOGR-N, $w_{FP} = w_{FN} = 1$, and for LOGR-A, $w_{FP} = 1$, $w_{FN} = 3$. The cut-off point generating the lowest cost is taken as the optimal one. Based on the scenario-specific predicted positives and negatives, the combination of sensitivity and specificity is determined.

$$c = w_{FP} \sum \mathrm{FP} + w_{FN} \sum \mathrm{FN} \tag{18}$$

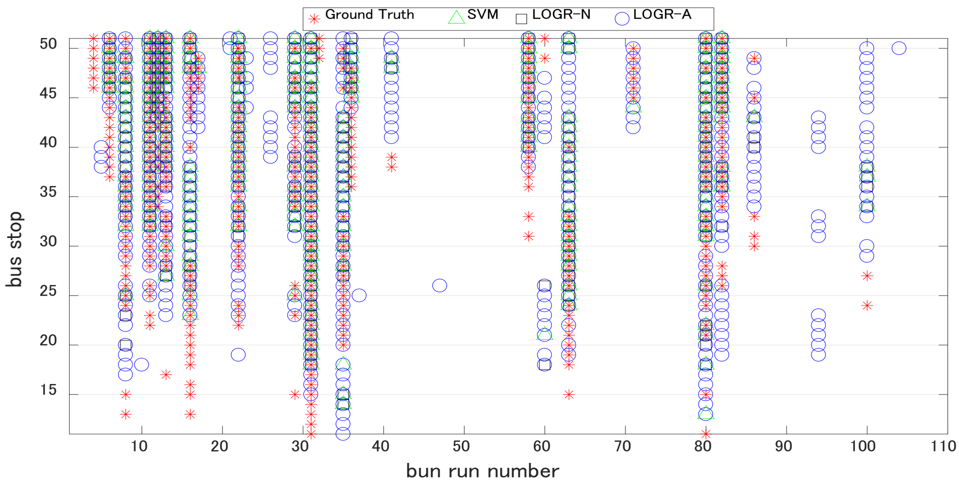
6.3 Performance comparison

Considering that the results derived by LR and SVM are similar, the comparison here is among SVM and two distinguished scenarios based on LOGR. As is presented in Figure 6(a), most bunching events can be detected 1-stop in advance by all three methods, and LOGR-A produces several false positives because it applies a more aggressive strategy to potential bunching events. However, LOGR-A significantly outperforms in 10-stop-ahead prediction, as is illustrated in Figure 6(b). LOGR-A captures a number of observed positives that are misclassified by SVM and LOGR-N although it generates a few more false positives.

A further comparison among two headway-based approaches and two scenarios of logistic regression is demonstrated in Figure 7. Sensitivity, specificity and accuracy for the four methods under various prediction horizons are presented. LOGR-A shows remarkable robustness in terms of sensitivity. On the contrary to the obvious deterioration of the other three methods, the sensitivity of LOGR-A keeps above 65% under all the prediction horizons. Besides, it only slightly underperforms the other three methods in terms of specificity, indicating an acceptable trade-off cost. Non-bunching events overwhelm bunching events in the daily operation, and a slight underperformance in specificity might introduce a large number of false positive. The exact numbers of true positives, false positives, true negatives, false negatives derived in the 5-day testing data are listed in Table 4. LOGR-N always generates the least total errors (highest accuracy). LOGR-A always correctly detects most bunching events (highest sensitivity) at the cost of most total errors (lowest accuracy). The notable advantage of LOGR over the other two methods is its trade-off functionality. It can achieve highest overall accuracy and can outperform the other methods in terms of sensitivity, although it cannot realize both objectives simultaneously.

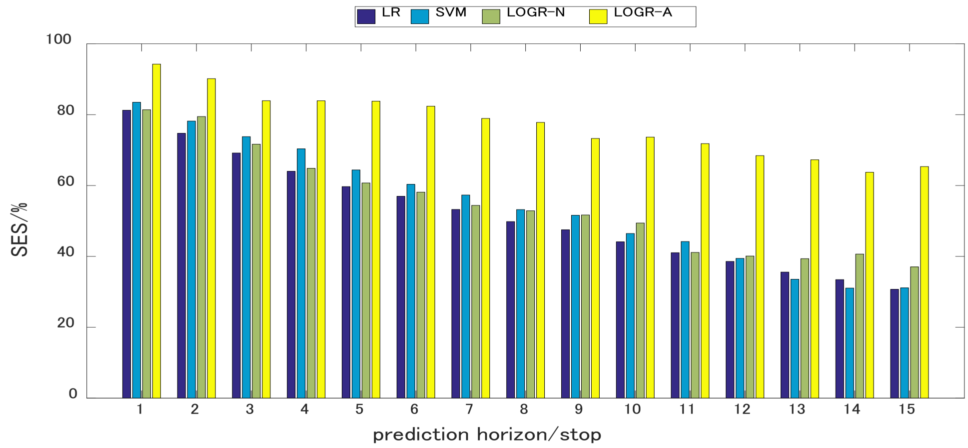


(a) 1-stop-ahead bunching prediction

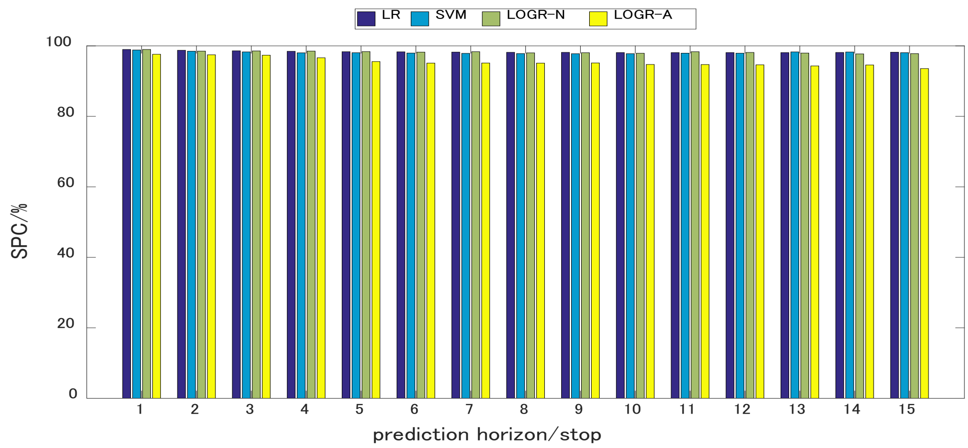


(b) 10-stop-ahead bunching prediction

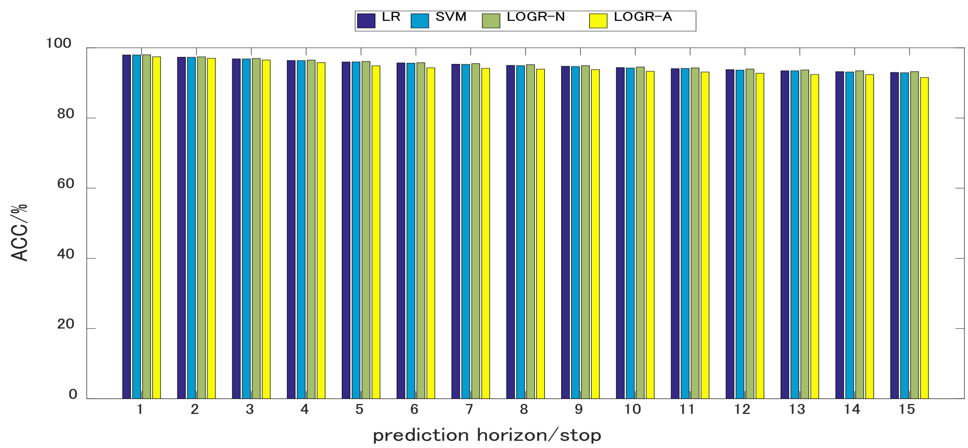
Figure 6. Performance comparison in terms of binary bunching identification



(a) Deterioration in SES as the prediction horizon extends



(b) Deterioration in SPC as the prediction horizon extends



(c) Deterioration in ACC as the prediction horizon extends

Figure 7. Performance comparison in terms of SES, SPC and ACC under various prediction horizons

Table 4. Performance comparison for 10-stop-ahead bunching prediction

	Size	OP	PP		PN		SES	SPC	ACC
			TP	FP	TN	FN	(%)	(%)	(%)
					Day 1				
LR	4223	344	161	77	3802	183	46.80	**98.01**	93.84
SVM	4223	344	171	89	3790	173	49.71	97.71	93.80
LOGR-N	4223	344	208	118	3761	136	60.47	96.96	**93.99**
LOGR-A	4223	344	262	200	3679	82	**76.16**	94.84	93.32
					Day 2				
LR	4182	384	158	118	3680	226	41.15	96.89	91.77
SVM	4182	384	166	138	3660	218	43.23	96.37	91.49
LOGR-N	4182	384	143	94	3704	241	37.24	**97.53**	**91.99**
LOGR-A	4182	384	266	345	3453	118	**69.27**	90.92	88.93
					Day 3				
LR	4264	355	175	62	3847	180	49.30	**98.41**	94.32
SVM	4264	355	187	73	3836	168	52.68	98.13	94.35
LOGR-N	4264	355	201	82	3827	154	56.62	97.90	**94.47**
LOGR-A	4264	355	263	202	3707	92	**74.08**	94.83	93.11
					Day 4				
LR	4182	146	51	42	3994	95	34.93	98.96	96.72
SVM	4182	146	57	51	3985	89	39.04	98.74	96.65
LOGR-N	4182	146	50	37	3999	96	34.25	**99.08**	**96.82**
LOGR-A	4182	146	100	112	3924	46	**68.49**	97.22	96.22
					Day 5				
LR	4182	254	123	67	3861	131	48.43	**98.29**	95.27
SVM	4182	254	121	79	3849	133	47.64	97.99	94.93
LOGR-N	4182	254	176	115	3813	78	69.29	97.07	**95.38**
LOGR-A	4182	254	204	164	3764	50	**80.31**	95.82	94.88

7 Discussion on the trade-off between sensitivity and specificity

ROC curves created by plotting (1-SPC, SES) for given cut-off points are commonly used to evaluate the classification performance. AUC (Area Under the Curve) being close to one indicates good classification power. ROC curves under various prediction horizons are presented in Figure 8. Furthermore, the four combinations of sensitivity and specificity derived by the four methods discussed in the previous section are indicated on each curve.

For each horizon, the corresponding curve can be considered the optimal front derived by LOGR. If an algorithm outperforms LOGR, the point it represents should appear above the curve with a higher SES and lower 1-SPC. It can be observed that the two headway-based methods (LR and SVM) generally fall below and sometimes on the LOGR curve, although the downward deviation from the curve is not significant.

It is easy to conduct the trade-off between sensitivity and specificity on the LOGR curve. The LOGR curve contains all combinations of prediction performance given continuous cut-off points where each cut-off point can be considered as optimal. A bunching-averse operator who is aggressive to eliminate bunching might desire to detect 99% of the positives regardless of the cost to increase false positive rate. This trade-off functionality significantly enhances the flexibility and robustness of existing bunching prediction approaches, especially for putting the predictive methodology into real practice. The curves provide a robust benchmark and insights for future algorithms that address bunching prediction problem. Deterministic methods can only produce one combination of prediction performance which greatly limits its contribution to the real application unless its sensitivity and specificity simultaneously achieve a highly reliable level. Other probabilistic methods generating a curve having higher AUC than LOGR or deterministic methods producing points of substantial upward deviation from the curve under various prediction horizons should be further promising extensions.

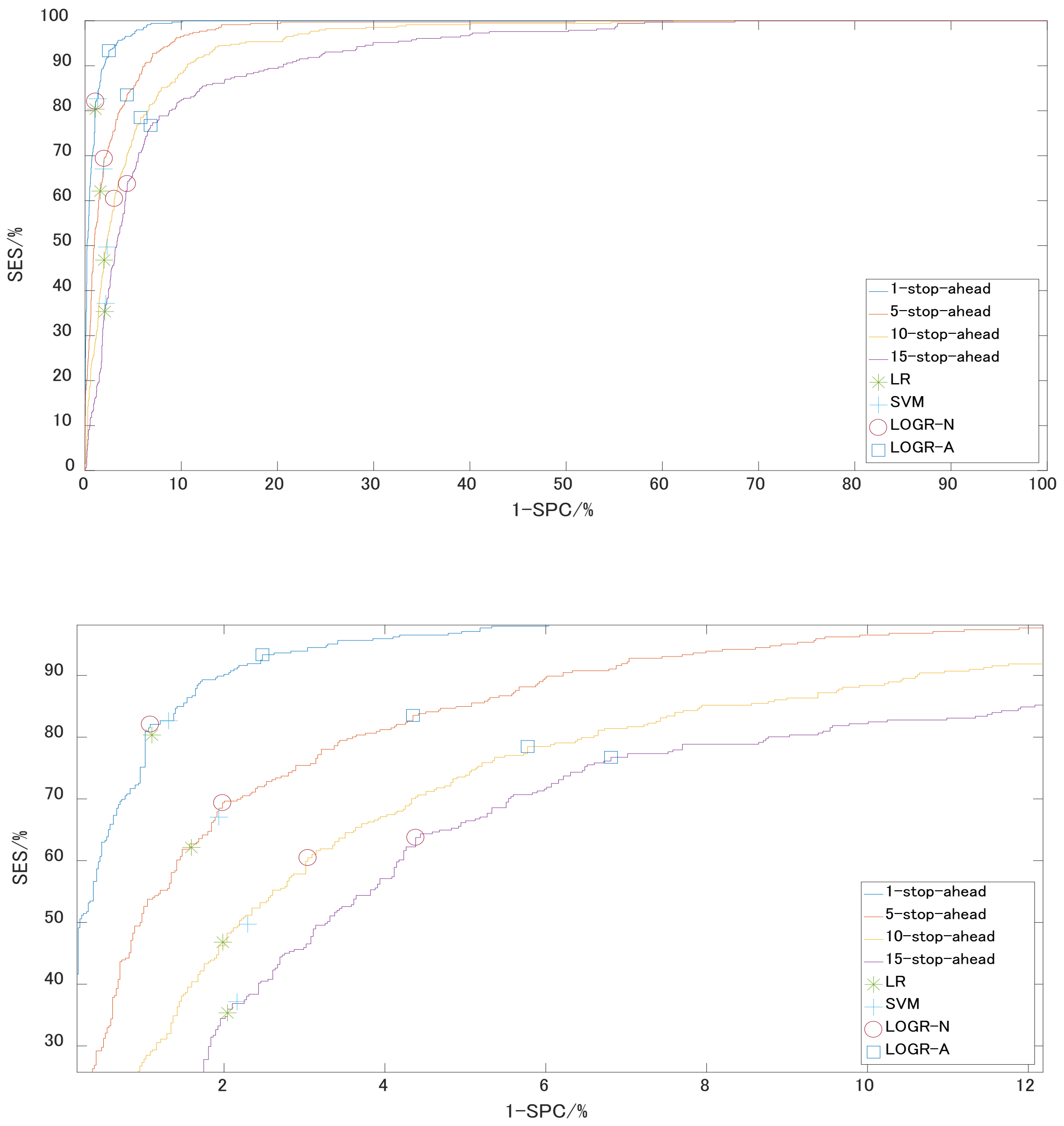


Figure 8. ROC curves under various prediction horizons (1-stop, 5-stop, 10-stop and 15-stop ahead)

Table 5. Supplementary information for ROC curves shown in Figure 8

Prediction horizon	AUC (area under the curve)	Cut-off point of LOGR-N (%)	Cut-off point of LOGR-A (%)
1-stop-ahead	0.9922	90.31	73.50
5	0.9763	87.19	77.53
10	0.9546	87.56	79.74
15	0.9279	81.60	71.50

8. Conclusion and further work

In this study, the potential of logistic regression to predict bus bunching is discussed. We consider the "rare event" nature of our problem which leads logistic regression to lose prediction power due to being biased to the majority in the dataset where positive events are by far outnumbered by negative events. Thus a selective sampling method and intercept correction is applied. We then compare this method with existing approaches that predict headways and then utilize the headway prediction for bunching prediction. Clearly headway prediction can be used for a larger range of purposes and deeper understanding of the service regularity developments as well as control strategies. However, bunching prediction in itself is important as it can be considered a distinctive state. This paper and other literature illustrate that headways fluctuate, but that, once bunching is reached, this state mostly continues along the line with far less headway fluctuation. We illustrate that when it comes to predicting bunching itself the newly proposed method has the potential to outperform headway-based methods such as LR and SVM in several aspects.

Firstly, LOGR provides superior prediction results under a long prediction horizon. It outperforms LR and SVM by 28% in sensitivity and maintains the same level of specificity in 10-stop-ahead prediction. It also shows improved resistance against deterioration in prediction performance as the prediction horizon extends.

Secondly, robustness and flexibility are significantly enhanced. LOGR provides robust prediction results that contain various sets of bunching outcomes under different cut-off points. This enables the operator to apply weights that are in accordance with their attitude towards bunching and operation budget. Some operators with limited possibility or willingness to apply corrective measures can use SVM or LOGR with neutral cut-off point setting. On the contrary, operators who desire to eliminate any possible bunching might be unwilling to choose headway-based methods which omit a considerable number of bunching in the long-term prediction cases. In this case LOGR-A becomes a much-preferred option. To conclude, LOGR provides operators with a wide range of options that can be tailored by their attitudes towards unexpected system disturbances.

We find that the headway-predicting approaches deviate slightly downward from the optimal front and we discuss that their shortcomings are inadequate robustness and flexibility from the operator's perspective. We note that it is also feasible to form a curve in terms of sensitivity and specificity for probabilistic headway prediction methods with confidence intervals to realize the trade-off on the curve discussed in this paper. Hans et al. (2015) developed a simulation-based prediction tool, Yu et al. (2017) tested RVM algorithm on headway prediction problem. Both methods could be extended to compute the probability of a headway falling below 1min and then to construct the ROC curves. By doing so and comparing the different ROC curves more insights might be obtained.

Finally, other extensions that potentially strengthen the predictive power of the models presented in this study should be noted. The model itself has space for improvement by including variables such as weather, traffic signals and passenger demand that are not incorporated due to missing data. Furthermore, the study could be extended to simultaneously predict bunching for several lines, in which case common line effects such as the interaction between buses of different lines at a common stop need to be considered.

Acknowledgements

We acknowledge the contribution of Hiroshi Shimamoto who provided valuable insights in the early stage of this research. We also acknowledge the comments from three anonymous reviewers that helped us improve this paper substantially. Besides, we gratefully acknowledge ASTEM (Advanced Science, Technology and Management Research Institute of Kyoto), notably Hideyuki Yamauchi, for his support in providing us with the bus GPS data.

References

Andres, M., & Nair, R. (2017) A predictive-control framework to address bus bunching. *Transportation Research Part B*, 104, 123-148.

Arriagada J., Gschwender, A., Munizaga, M., & Trepanier, M. (2019). Modeling bus bunching using massive location and fare collection data. *Journal of Intelligent Transportation Systems*, 23(4), 332-344.

Bartholdi, J. J., & Eisenstein, D.D. (2012). A self-coordinating bus route to resist bus bunching. *Transportation Research, Part B*, 46(4), 481-491.

Berrebi, S. J., Watkins, K. E., & Laval, J. A. (2015). A real-time bus dispatching policy to minimize passenger wait on a high frequency route. *Transportation Research Part B*, 81, 377-389.

Berrebi, S. J., Hans, E., Chiabaut, N., Laval, J. A., Leclercq, L., & Watkins, K. E. (2018). Comparing bus holding methods with and without real-time predictions. *Transportation Research Part C*, 87, 197-211.

Cherkassky, V., & Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks* 17(1), 113-126.

Cosslett, S. R. (1981). Maximum Likelihood Estimator for Choice-Based Samples. *Econometrica*, 49(5), 1289-1316.

Daganzo, C.F. (2009). A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. *Transportation Research Part B*, 43(10), 913-921.

Daganzo, C.F., & Pilachowski, J. (2011). Reducing bunching with bus-to-bus cooperation. *Transportation Research Part B*, 45(1), 267-277.

Dai, Z., Ma, X., & Chen, X. (2019). Bus travel time modelling using GPS probe and smart card data: A probabilistic approach considering link travel time and station dwell time. *Journal of Intelligent Transportation Systems*, 23(2), 175-190.

Degeler, V., Heydenrijk-Ottens, L., Luo, D., Oort, N., & Lint, H. (2018, July). *Unsupervised approach to public transport bunching swings formations phenomenon analysis*. Paper presented at the 14th International Conference on Advanced Systems in Public Transport (CASPT), Brisbane, Australia.

Eberlein, X.J., Wilson, M.H.M., & Bernstein, D. (2001). The holding problem with real-time information available. *Transportation Science*, 35(1), 1-18.

Hans, E., Chiabaut, N., Leclercq, L., & Bertini, R. L. (2015). Real-time bus route state forecasting using partical filter and mesoscopic modeling. *Transportation Research Part C*, 61, 121-140.

Imbens, G. (1992). An Efficient Method of Moments Estimator for Discrete Choice Models with Choice-Based Sampling. *Econometrica*, 60(5), 1187-1214.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9(2), 137-163.

Kumar, B. A., Vanajakshi, L., & Subramanian, S. C. (2018). A hybrid model based method for bus travel time estimation. *Journal of Intelligent Transportation Systems*, 22(5), 390-406.

Maalouf, M., & Trafalis, T.B. (2011). Robust weighted kernel logistic regression in imbalanced and rare events data. *Computational Statistics & Data Analysis*, 55(1), 168-183.

Moreira-Matias, L., Cats, O., Gama, J., Mendes-Moreira, J., & De Sousa, J.F. (2016). An online learning approach to eliminate Bus Bunching in real-time. *Applied Soft Computing*, 47, 460-482.

Newell, G.F., & Potts, R.B. (1964). Maintaining a bus schedule. *Proceedings of 2nd Australian Road Research Board*, 2, 388-393.

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.

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

Schmöcker, J. D., Sun, W., Fonzone, A., & Liu, R. (2016). Bus bunching along a corridor served by two lines. *Transportation Research Part B*, 93, 300-317.

Sun, A., & Hickman, M. (2005). The real-time stop-skipping problem. *Journal of Intelligent Transportation Systems*, 9(2), 91-109.

Sun, W., & Schmöcker, J. D. (2018). Considering passenger choices and overtaking in the bus bunching problem. *Transportmetrica B*, 6(2), 151-168.

Wu, W., Liu, R., & Jin, W. (2017). Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour. *Transportation Research Part B*, 104, 175-197.

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(10), 1831-1845.

Yu, B., Yang, Z., & Yao, B. (2006). Bus arrival time prediction using support vector machines. *Journal of Intelligent Transportation Systems*, 10(4), 151-158.

Yu, B., Lam, W. H., & Tam, M. L. (2011). Bus arrival time prediction at bus stop with multiple routes. *Transportation Research Part C*, 19(6), 1157-1170.

Yu, H., Chen, D., Wu, Z., Ma, X., & Wang, Y. (2016). Headway-based bus bunching prediction using transit smart card data. *Transportation Research Part C*, 72, 45-59.

Yu, H., Wu, Z., Chen, D., & Ma, X. (2017). Probabilistic prediction of bus headway using relevance vector machine regression. *IEEE Transactions on Intelligent Transportation Systems*, 18(7), 1772-1781.

Zhang, S., & Lo, H. (2018). Two-way-looking self-equalizing headway control for bus operations. *Transportation Research Part B*, 110, 280-301.