

Detection of Bus Bunching through the Analysis of Prevalent Public Transport Control Data

Leon Weinsziehr*¹, Antonios Tsakarestos², Frederik Bachmann², Klaus Bogenberger²

*Corresponding author: leon.weinsziehr@q-perior.com

¹ Consultant, Topic Chapter Travel Transport and Logistics, Q_PERIOR AG

² Chair of Traffic Engineering and Control, Technical University of Munich (TUM), Germany

SHORT SUMMARY

Bus bunching describes a phenomenon that is familiar to many public transport users. Two buses, running according to a scheduled frequency, arrive at a stop in immediate succession. In most cases, the leading vehicle is delayed. The delay causes an increasing number of waiting passengers at the stops. Through this higher number of boarding and alighting passengers, the dwell time of the leading bus lengthens and by that also its delay. This problem is made visible using freely available public transport control data of two routes from Sydney, Australia. To validate the bunching events captured from the bus control data, General Transit Feed Specification (GTFS) data is used. The buses' positioning logs are traced to determine the distance between bunched vehicles. Additionally, a direct association between late departures of buses induced by delay propagation from one direction and increased bunching occurrence in the opposite direction is observed.

Keywords: Big Data, Bus Bunching, Public Transport

1. INTRODUCTION

Bus bunching resembles a phenomenon that is frequently observed in urban bus operations. Whereas measuring or detecting its characteristics has not led to a thorough definition of the underlying problem, most literature determines bus bunching by the immediate queuing of two consecutive buses at one particular stop (ILIOPOULOU et al., 2020). Bus bunching resembles a phenomenon that is frequently observed in urban bus operations. Whereas measuring or detecting its characteristics has not led to a thorough definition of the underlying problem, most literature determines bus bunching by the immediate queuing of consecutive buses at one particular stop (MOREIRA-MATIAS et al., 2014; XIN et al., 2021; YU et al., 2016).

This paper has three objectives. First, to develop a methodology for identifying and analysing the spatiotemporal dimensions of bus bunching using publicly available data. Second, to investigate whether strict adherence to the timetable, as well as knock-on delays, also lead to bunching. Third, an overarching objective of this research is to ensure that potential results are obtained as accurately as possible, but with a minimum of effort for the operator.

Datasets revealing the punctuality of buses are already available to many transport companies worldwide. The common procedure is to analyse the actual arrival and departure times at bus stops and to calculate the deviation from the schedule. This paper shows how additional findings can be drawn from such analysis and headway deviation calculations. The methodology process

proves how common practices considering solely schedule adherence are easily extended to capture bus bunching. Service regularity gains importance compared to schedule adherence, especially in dense headways. It is noteworthy that these insights do not require supplementary data sources. Automatic Vehicle Location (AVL) is not necessarily available to all operators, such data is used in this paper as a means of validation of the methodology.

2. METHODOLOGY

The approach used for this project involves a total of four steps (see Table 1). Firstly, with the research questions concerning bus bunching identification and its spatiotemporal analysis are outlined. Secondly, the data acquisition determines the required level of detail to analyse bus bunching. Thirdly, the evaluation framework is described, which is crucial for clarifying which key performance indicators (KPI) will subsequently allow the interpretation of the results of the data analysis. The goal is to create a schedule adherence index. The index does not measure bus bunching directly but bus regularity, which is closely associated with bus bunching. Fourthly, several proposed solution approaches are compared regarding both the determined KPIs and the available data basis. A presented validation method helps to underline the results. In an additional step, the position and time of the bunching events are compared against a second data stream that tracks the actual coordinates of the vehicles.

Table 1: Methodological Steps for Bus Bunching Identification

| Step | Task | Question / Decision to be dealt with | Outcome |
|------|----------------------|--|---|
| 1 | Data Acquisition | How can the available data basis be evaluated in terms of the project's feasibility? | Level of detail of data basis and its feasibility for the intended analysis |
| 2 | Evaluation Framework | Which KPIs can be measured with the data basis? | Choice of KPIs |
| 3 | Choice of Method | Which solution approach from literature appears to be suitable? | Choice of Solution Approach (Algorithm) to capture chosen KPIs |
| 4 | Data Analysis | How can the desired results be obtained from the available data? | Specification of measurement tools and techniques |

Data acquisition

The data used for this work is acquired from the Bus Opal Assignment Model (BOAM) hosted by Transport for New South Wales in Australia (TfNSW), which makes a wide variety of public transport (PT) related datasets publicly available. For buses, only actual arrival times at stops are recorded so that the actual departure times remain unknown. Figure 1 shows exemplarily those four columns that are relevant for the bunching identification and its spatiotemporal analysis:

Table 2: Excerpt of BOAM Daily Dataset Relevant for Bunching Identification

| Trip ID_Date | Stop | Scheduled Arrival | Actual Arrival |
|----------------------|------|-------------------|----------------|
| 179839616_2020-02-03 | 1 | 06:00:00 | 06:03:34 |
| 179839727_2020-02-03 | 1 | 06:10:00 | 06:10:58 |
| 179839617_2020-02-03 | 1 | 06:20:00 | 06:22:34 |
| 179839618_2020-02-03 | 1 | 06:30:00 | 06:30:18 |
| 179839730_2020-02-03 | 1 | 06:38:00 | 06:37:19 |
| 179839732_2020-02-03 | 1 | 06:45:00 | 06:45:09 |

From public transport control data to bus bunching analysis

The real-time extension of the General Transit Feed Specification (GTFS) differentiates trip updates, service alerts, and vehicle positions. The vehicle positions feed is of vital importance for the identification of bus bunching, it depicts the current location and movement parameters of vehicles (BARBEAU., 2018). AVL data is the primary of three forms of PT control data (alongside Automatic Fare Collection and Automated Passenger Counting) and typically involves information in three dimensions (latitude, longitude, time). Consequently, the identification of bus bunching which relies solely on AVL data can be regarded as a robust methodology (SUN, 2020).

The spatiotemporal analysis is designed as a retrospective evaluation of sample data to uncover patterns. Particularly, categorizing bus operations into predefined levels of service relies on the coefficient of variation of headway deviations. To obtain the required quotient, the standard deviation of headways is divided by the average headway (CAMPS AND ROMEU, 2016).

The calculation of the coefficient of variation c_{vh} follows equation (1):

$$c_{vh} = \frac{sd(h_A)}{\underline{h}_A} \quad (1)$$

sd : standard deviation
 h_A : actual headway
 \underline{h}_A : average actual headway

The further translation into Levels of Service (LOS) is applied according to the threshold ranges from the PT Capacity and Quality of Service Manual presented in Table 3 (TURNER et al., 2010).

Table 3: Levels of Service of Schedule Adherence based on Headway Deviations

| LOS | c_{vh} | $P(\text{abs}(h_i - h) > 0.5 * h)$ | Passenger and Operator Perspective |
|-----|-------------|------------------------------------|---------------------------------------|
| A | 0.00 - 0.21 | $\leq 2\%$ | Service provided like clockwork |
| B | 0.22 – 0.30 | $\leq 10\%$ | Vehicles slightly off headway |
| C | 0.31 – 0.39 | $\leq 20\%$ | Vehicles often off headway |
| D | 0.40 – 0.52 | $\leq 33\%$ | Irregular headway, with some bunching |
| E | 0.53 – 0.74 | $\leq 50\%$ | Frequent bunching |
| F | ≥ 0.75 | $> 50\%$ | Most vehicles bunched |

There are few studies in which schedule adherence is ascribed to minor importance. This is reasoned by the high utility of bus lines with short headways of less than ten minutes. Riders are assumed to travel spontaneously, meaning they do not check the upcoming departure times of their bus services (BARTHOLDI AND EISENSTEIN, 2012). However, a valuable contribution from the spatiotemporal analysis is to spot the locations at which bus bunching occurs regularly (LI et al., 2013).

Preliminary choice of method for data analysis

The prescribed methodology involves a suitable calculation method. Table 4 below describes the six steps carried out within that final methodological step. Because of the low traffic volumes during the night times, only the hour bands which are relevant to grasp the phenomenon are studied thoroughly. Data from weekdays in February are chosen as these do not interfere with public holidays or other strong seasonal influences.

Table 4: Six-step Heuristic as Final Methodological Step

| Step | Description |
|--|--|
| 1) Data cleaning | Elimination of faulty (e.g. double) or missing records. |
| 2) Data sorting | Sort records by scheduled / actual arrival time for each stop for scheduled and actual headway calculation, respectively. |
| 3) Headway calculation and bus identification | Headway can be easily obtained by subtracting two consecutive arrival times for each stop |
| 4) Bus bunching identification | The set headway threshold for bus bunching identification is set to be $0.25 * h_{sched}$ (scheduled headway). |
| 5) Bus bunching distribution and further KPI calculation | Count the number of identified bus bunching records for each stop in each hour for all days of the same type of day (weekday). Calculate the coefficient of variation for each stop in each hour and return the corresponding LOS. |
| 6) Data aggregation and plotting | Aggregate data sets of same day type and plot results. |

Supplementary analysis of delay propagation

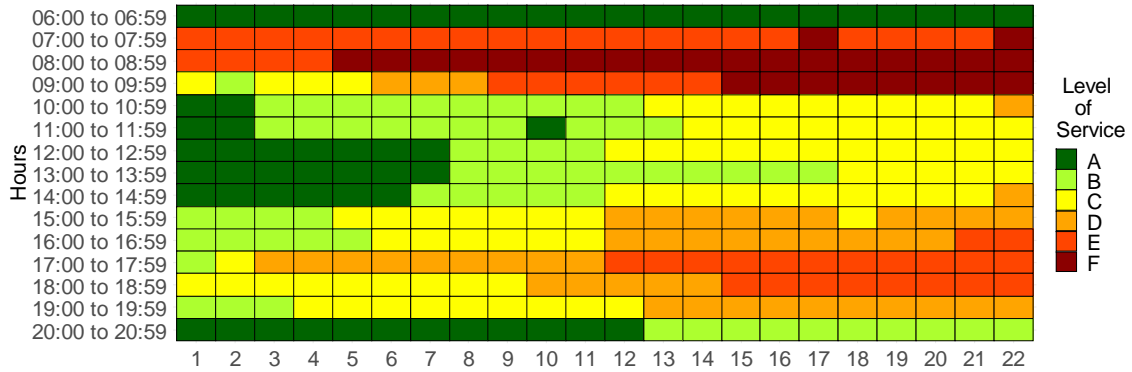
In classic scheduled services, buffer and turnaround times are scheduled at the terminals of scheduled routes. However, despite these preventive measures, it may happen that delayed buses from one direction do not re-enter the line route in the opposite direction on time. This delay propagation is closely linked to the phenomenon of bus bunching, as poor schedule adherence applies in both cases. The six-step heuristic for bus bunching analysis is perpetuated to conclude a direct association between late departures of buses induced by delay propagation from one direction and increased bunching occurrence in the opposite direction. By doing so, each trip is marked by a flag regarding its deviation from the scheduled departure at the start-stop. Thus, trips are divided into three categories - trips that depart more than a minute before their scheduled departure, trips that depart between a minute early and a minute late (one-minute tolerance), and trips that are more than a minute late. The latter serve as an indicator of the relationship between delay propagation and the occurrence of bus bunching events. Subsequently, for the three aforementioned categories of trips, bunching events are identified and additionally, the number of bunching events per trip is determined. The stop on the route at which the bunching event occurs has secondary importance – nevertheless, a trip can inherit more than one bunching event (PARK, 2020).

3. RESULTS AND DISCUSSION

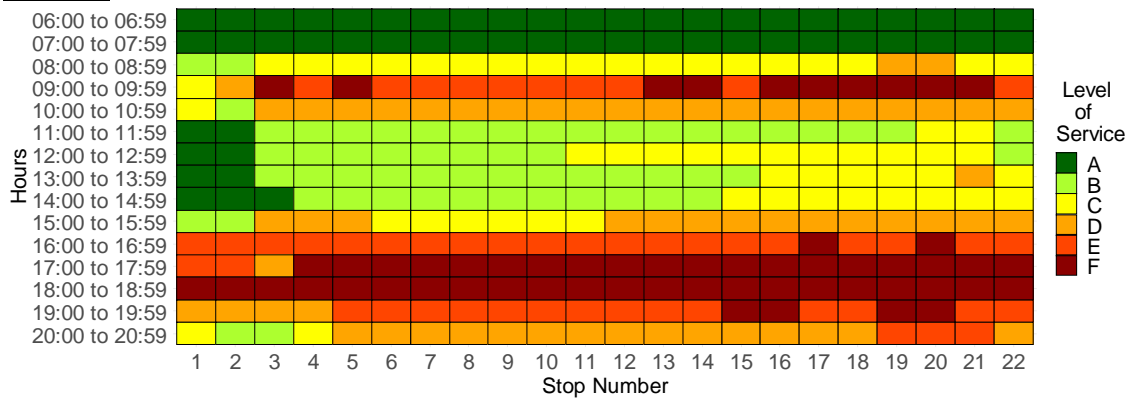
Low LOS resulting from high coefficients of variation of the calculated headway deviations indicate bus bunching. The following figure shows the LOS according to Table 3 for each stop of line 304 in both directions. For instance, the examined line 304 runs in the north-south direction in and out of Sydney's highly demanded central business district (CBD) on a ten-to-twelve-minute frequency during normal weekday hours. During peak hours, its headway is shortened to six minutes for the major commuting direction. On parts of the route close to the CBD, the headway is even lowered to three minutes. In contrast to the inbound results, Line 304 in the outbound direction reveals lower LOS in the afternoon.

Inbound

Calculation based on observations per Hourband aggregated from 20 weekday days
 (Indication of Bus Bunching using Levels of Service relying on Coefficients of Variation of Headway Deviations)



Outbound



Data acquired from: TfNSW's Bus Opal Assignment Model

Figure 1: Line 304 - Schedule Adherence Index in February 2020

During the week, most bunching events occur in the rush hour between 8 and 9 a.m. Weekends feature most bunching events around midday. According to the major commuting direction, the evening peak is significantly more affected by bunching than the morning, which is evident not only during the week but also on weekends as shown in Table 5 below.

Table 5: Line 304 – Number of Bus Bunching Events in February 2020

| <u>Inbound</u> | | | | | | | | | | |
|----------------|------------|------------|------------|------------|------------|------------|-------------|--------------|--------------|---------------|
| daytype | mon | tue | wed | thu | fri | sat | sun | weekdays | weekends | total |
| 06:00 to 06:59 | 1 | 11 | 9 | 0 | 15 | 0 | 0 | 36 | 0 | 36 |
| 07:00 to 07:59 | 120 | 117 | 138 | 103 | 91 | 0 | 0 | 569 | 0 | 569 |
| 08:00 to 08:59 | 146 | 110 | 111 | 156 | 128 | 0 | 0 | 651 | 0 | 651 |
| 09:00 to 09:59 | 88 | 88 | 70 | 104 | 48 | 0 | 0 | 398 | 0 | 398 |
| 10:00 to 10:59 | 13 | 26 | 33 | 48 | 1 | 0 | 0 | 121 | 0 | 121 |
| 11:00 to 11:59 | 11 | 5 | 16 | 6 | 19 | 6 | 0 | 57 | 6 | 63 |
| 12:00 to 12:59 | 14 | 9 | 3 | 11 | 9 | 23 | 9 | 46 | 32 | 78 |
| 13:00 to 13:59 | 4 | 34 | 8 | 21 | 12 | 44 | 1 | 79 | 45 | 124 |
| 14:00 to 14:59 | 19 | 7 | 9 | 12 | 6 | 26 | 2 | 53 | 28 | 81 |
| 15:00 to 15:59 | 43 | 20 | 53 | 34 | 50 | 7 | 5 | 200 | 12 | 212 |
| 16:00 to 16:59 | 71 | 41 | 32 | 51 | 40 | 2 | 0 | 235 | 2 | 237 |
| 17:00 to 17:59 | 19 | 6 | 29 | 20 | 35 | 12 | 2 | 109 | 14 | 123 |
| 18:00 to 18:59 | 29 | 61 | 22 | 78 | 47 | 17 | 2 | 237 | 19 | 256 |
| 19:00 to 19:59 | 0 | 19 | 17 | 15 | 38 | 3 | 0 | 89 | 3 | 92 |
| 20:00 to 20:59 | 0 | 1 | 0 | 20 | 5 | 0 | 0 | 26 | 0 | 26 |
| total | 578 | 555 | 550 | 679 | 544 | 140 | 21 | 2906 | 161 | 3067 |
| mean | 145 | 139 | 138 | 170 | 136 | 28 | 5,25 | 145,3 | 17,89 | 105,76 |

| <u>Outbound</u> | | | | | | | | | | |
|-----------------|------------|------------|------------|------------|------------|------------|-------------|---------------|--------------|---------------|
| daytype | mon | tue | wed | thu | fri | sat | sun | weekdays | weekends | total |
| 06:00 to 06:59 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 07:00 to 07:59 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 3 | 0 | 3 |
| 08:00 to 08:59 | 14 | 7 | 9 | 32 | 1 | 0 | 0 | 63 | 0 | 63 |
| 09:00 to 09:59 | 55 | 39 | 38 | 55 | 24 | 0 | 0 | 211 | 0 | 211 |
| 10:00 to 10:59 | 49 | 49 | 49 | 83 | 21 | 0 | 0 | 251 | 0 | 251 |
| 11:00 to 11:59 | 10 | 11 | 0 | 23 | 0 | 1 | 0 | 44 | 1 | 45 |
| 12:00 to 12:59 | 7 | 14 | 3 | 4 | 20 | 34 | 0 | 48 | 34 | 82 |
| 13:00 to 13:59 | 18 | 18 | 7 | 29 | 1 | 22 | 0 | 73 | 22 | 95 |
| 14:00 to 14:59 | 7 | 23 | 7 | 42 | 11 | 54 | 1 | 90 | 55 | 145 |
| 15:00 to 15:59 | 53 | 23 | 6 | 28 | 50 | 22 | 0 | 160 | 22 | 182 |
| 16:00 to 16:59 | 67 | 57 | 110 | 53 | 139 | 20 | 0 | 426 | 20 | 446 |
| 17:00 to 17:59 | 106 | 97 | 135 | 136 | 108 | 1 | 0 | 582 | 1 | 583 |
| 18:00 to 18:59 | 99 | 133 | 160 | 110 | 121 | 47 | 0 | 623 | 47 | 670 |
| 19:00 to 19:59 | 17 | 50 | 65 | 103 | 180 | 24 | 0 | 415 | 24 | 439 |
| 20:00 to 20:59 | 0 | 46 | 9 | 55 | 92 | 0 | 0 | 202 | 0 | 202 |
| total | 503 | 567 | 600 | 753 | 768 | 225 | 1 | 3191 | 226 | 3417 |
| mean | 126 | 142 | 150 | 188 | 192 | 45 | 0,25 | 159,55 | 25,11 | 117,83 |

To conclude this section, it can be said that in both directions of line 304, the counted bunching events fit the heat maps of the schedule adherence index (Figure 1). Both the peak hours, as well as the major commuting direction, are apparent.

Validating identified bunching events using records from GTFS-real-time feed

Table 6: Identified Bunching Event for Exemplary Validation

| Route | Trip | Stop | Sched- uled headway | Scheduled arrival time | Actual headway | Actual arrival time | Schedule deviation |
|-------|------|------|---------------------------|---------------------------|-------------------|------------------------|-----------------------|
| 304 | 21 | 8 | 540 s | 18:14:00 | 807 s | 18:14:04 | -4 s |
| 304 | 22 | 8 | 180 s | 18:17:00 | 25 s | 18:14:29 | 151 s |

Table 6 depicts an identified bunching event as two consecutive arrivals of line 304 at stop 8 (trips 21 and 22) are recorded only 25 seconds after another. To validate whether a bus bunching event has occurred, the matching GTFS-real-time feed records at 18:14 needs to be considered (see Table 7). For the validation, the two trip IDs are to be checked for bunching at stop 8 (despite an insignificant five-second delay of the following vehicle's record).

Table 7: Matching GTFS records for Exemplary Validation

| Route | Trip | Start time | Lati- tude | Longi- tude | Timestamp | Vehicle | Direction |
|-------|------|---------------|---------------|----------------|-----------|---------|-----------|
| 304 | 21 | 17:59:00 | -33.885 | 151.214 | 18:14:40 | 1339858 | Inbound |
| 304 | 22 | 18:02:00 | -33.881 | 151.214 | 18:14:45 | 1340083 | Inbound |

Finally, a comparison of the latitude and longitude coordinates using the statistical software R computes the distance of the allegedly bunching buses. R yields a distance of only 515 meters, which unambiguously indicates bus bunching.

Dependency of bunching occurrence from delay propagation

Beyond the presented suitability of the methodology to study bus bunching and the spatiotemporal dimensions of the phenomenon, the relation between delay propagation on trips and bunching occurrences in the opposite direction is discovered. Table 8 notes bunching events according to one of three predefined categories concerning the start delay of the respective trips. Although the number of trips per category varies greatly, it appears that the dispersion of bunching events in the case of trips that suffer from delay propagation of more than one minute reaches higher value ranges.

Table 8: Bunching Events per Trip categorised by Schedule Adherence at Stop 1 (Line 304 Inbound –Weekdays in February 2020)

| Schedule adherence at first stop (Category) | Number of trips | Total Bunching Events | Mean (Bunching events / trip) |
|--|--------------------|--------------------------|----------------------------------|
| Earlier than 1 min before schedule | 24 | 74 | 3.08 |
| Within 1 min deviation from the schedule | 1239 | 1744 | 1.41 |
| More than 1 min late | 490 | 1002 | 2.04 |
| Total | 1753 | 2820 | 1.61 |

The 1239 trips recorded between one minute before and one minute after the schedule at the start-stop show a considerably lower mean of only 1.41 bunching events per trip (1744 bunching events

recorded). At the same time, indicating the negative effect of the delay propagation, 1002 bunching events are counted among 490 trips that are recorded for the category of more than one minute late at the start-stop. The corresponding average of 2.04 is clearly above the overall average (1.61) of all trips. The category of trips that start more than one minute early on the line route is very rare (only 24 occasions) and the high average value of 3.08 bunching events per trip most likely results from considerable irregularities in the operation of the vehicles.

4. CONCLUSIONS

Although various paradigms and algorithms have already sufficiently addressed the topic of spatiotemporal analysis of bus bunching, the selected measuring instruments allow a transparent view of this phenomenon. The uniqueness of the methodology is the type of data used. The data's prevalence as well as its scope and format are globally distinctive, which caters to a high transferability of the methodology. It shows that bus bunching can be analysed with publicly available PT control data. Typically, punctuality is the focus of analysis, but on-time performance is often not influenceable due to prevailing external factors. However, service regularity is a more promising indicator to assess the service quality of a line. PT agencies that record actual values of buses' arrivals or departures along the route can use the methodology presented here to better understand the occurrence of bus bunching in their network.

Following this work, the influence of short turns or buffer times on bus bunching events gives room for further investigation. These are the simplest tool for transport operators and can mitigate the proven delay propagation and associated bunching occurrence. Further influencing factors such as weather conditions, and temporal dimensions like the day of the week, time of day, and season could be additionally differentiated. Overall, bunching analysis and drawing the right conclusions from it could bridge the time until automated mitigation actions might be implemented in the onboard computers of buses.

REFERENCES

- Barbeau, S. (2018): Quality control-lessons learned from the deployment and evaluation of GTFS-realtime feeds. *97th Annual Meeting of the Transportation Research Board, Washington, DC*. 2018.
- Bartholdi, J. J.; Eisenstein, D. D. (2012): A self-coordinating bus route to resist bus bunching, *Transportation Research Part B: Methodological*. vol. 46, no. 4.
- Camps, J. M.; Romeu, M. E. (2016): Headway Adherence. Detection and Reduction of the Bus Bunching Effect. *European Transport Conference 2016*. Association for European Transport (AET).
- Iliopoulou, C.; Vlahogianni, E. I.; Kepaptsoglou, K. (2020): Understanding the factors that affect the bus bunching events' duration. *IEEE, 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1-6.
- Li, F.; Yang, D.; Ma, K. (2013): Bus Rapid Transit (BRT) Bunching Analysis with Massive GPS Data, *American Society of Civil Engineers. ICTE 2013: Safety, Speediness, Intelligence, Low-Carbon, Innovation*. 2013. 959-965.
- Moreira-Matias, L.; Gama, J.; Mendes-Moreira, J.; Freire de Sousa, J. (2014): An Incremental Probabilistic Model to Predict Bus Bunching in Real-Time *Advances in Intelligent Data Analysis*

XIII: 13th International Symposium, IDA 2014, Leuven, Belgium, October 30–November 1, 2014. Proceedings 13. Springer International Publishing, 2014.

Park, Y.; Mount, J.; Liu, L.; Xiao, N.; Miller, H. J. (2020): Assessing public transit performance using real-time data: spatiotemporal patterns of bus operation delays in Columbus, Ohio, USA, *International Journal of Geographical Information Science*.

Sun, W. (2020): *Bus Bunching Prediction and Transit Route Demand Estimation Using Automatic Vehicle Location Data*. Dissertation

Turner, S.; Robinson, B.; Koorey, G. (2010): Technical Note - Transportation Research Board (TRB) Annual Meeting (Conference) 2009 & 2010: Highlights, University of Canterbury. Civil and Natural Resources Engineering.

Xin, Q.; Fu, R.; Yu, S.; Ukkusuri, S.; Jiang, R. (2021): Modeling bus bunching and anti-bunching control accounting for signal control and passenger swapping behavior, University of South Florida Libraries, *Journal of Public Transportation*. vol. 23, no. 1.

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: Emerging Technologies*. vol. 72, pp. 45–59.