J Grid Computing (2023) 21:15
https://doi.org/10.1007/s10723-023-09652-3

Interfering Spatiotemporal Features and Causes of Bus Bunching using Empirical GPS Trajectory Data

Xiaofeng Shan · Chishe Wang · Dongqin Zhou

Received: 12 October 2022 / Accepted: 12 February 2023 / Published online: 27 February 2023

Abstract Bus bunching refers to the phenomenon that several buses arrive at a station within a short period. It dramatically increases passengers' waiting time and reduces the quality of transit service. Evaluating the features of bus bunching and identifying the causes are important to developing countermeasures. The primary of this study was to analyze the temporal-spatial features of bus bunching by conducting an in-depth analysis of empirical bus GPS trajectory data obtained in Nanjing, China. The GPS data were inputted into the ArcGIS to track the spatial map's bus trajectories. A data processing procedure was proposed to analyze the data, including data cleaning, trip cutting, each station's arrival and departure time estimation, and time headway calculation. Then the spatiotemporal trajectory picture was drawn for the bus route where the bus bunching was identified. The study also analyzed the headway features of consecutive buses at the different stations and evaluated the variation of time headway, indicating the severity of bus bunching. The results showed that there are significant differences in the spatiotemporal features of bus bunching between bus stations. When the bus bunching occurred, it persisted on downstream stations for a long time. The bunching severity dramatically increased at downstream stations, reducing bus arrival reliability on the whole bus line. We also identified that the bus bunching was primarily caused by the overlong bus dwelling time at a station and the different travel times of buses between stations. The study fills the gap by developing the methodology to investigate the bus bunching features and causes with point-by-point empirical GPS trajectory data. Findings of the study can also support the real-time prediction and warning of bus bunching in practical applications.

Keywords GPS data · Bus bunching · Cause · Prediction

1 Introduction

Urban traffic has become more congested and unsafe in China due to the continuous increase in automobile ownership [1–5]. To reduce traffic jams on roads and improve travel quality, one of the most effective measures is to develop a sophisticated transport system at the network level [6–10]. Public travel modes such as public transport and the public bike could reduce the road space occupation per person [11–15]. Thus, the proper travel activity pattern could significantly improve traffic operation [16, 17]. The bus system has been considered an effective way to

X. Shan · C. Wang
Jinling Institute of Technology, 99 Hongjing Road, Nanjing 211169, China

D. Zhou (✉)
School of Transportation, Southeast University, 2 Si Pai Lou Road, Nanjing 211196, China
e-mail: 101101249@seu.edu.cn

reduce congestion, improve safety, and reduce emissions [18–21]. Service quality of transit is crucial to maintain passengers' satisfaction and promote transit usage. The bus schedule is usually designed in the planning and operation stage according to the transit demand.

In the real world, bus bunching is one major factor significantly affecting the service quality of public transport on bus routes. Bus bunching is a well-known phenomenon that refers to the fact that several buses arrive at a bus station within a short period. A bus bunching example is illustrated in Fig. 1. The scheduling frequency determines the initial headway between two buses on the initial departure station based on passenger demand at different times. However, during the duration of operation, the headway may decrease (or increase) gradually due to influences of some factors, which finally cause the occurrence of bus bunching at a station. Bus operation efficiency is affected because the two buses eventually pair up and travel as a single unit. More specifically, the red bunching bus in Fig. 1 will pick up much fewer passengers than scheduled, resulting in the waste of capacity. In addition, bunching is undesirable for passengers because it increases passengers' waiting time at the station after a bunching occurs.

The bus bunching effect was first described by Newell and Potts [22]. In the literature, many previous studies have been conducted to explore the potential causes of bus bunching at bus routes [23–25]. For example, some studies suggested that the primary reason for bus bunching is the random passenger arrivals at bus stations since the boarding and dwelling time affect the bus operating schedule [26]. Other studies showed that other factors, such as delay at intersections [27], delay due to traffic congestion, and uncertainties in boarding and dwelling time, also cause the bus bunching phenomenon. Based on the findings, several researchers have proposed control strategies for bus operation to relieve or prevent bus bunching. For example, Daganzo [28, 29] proposes an adaptive control scheme that adjusts a bus cruising speed in real-time to avoid bunching.

Most of the findings in the aforementioned studies were drawn from the theoretical analyses with assumptions of bus system operation. They generally lack the support of real-world data which could be over-idealistic. To overcome such issue in theoretical studies, some recent researchers estimated the bus bunching effects based on empirical data of transit systems [30–32]. For example, Yu et al.

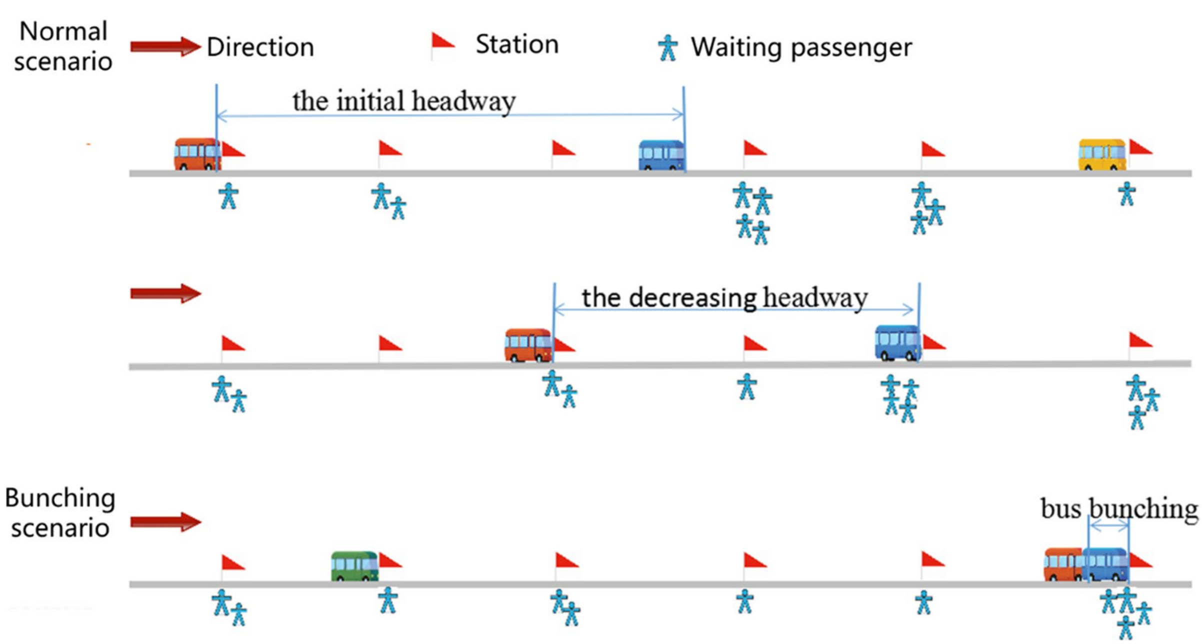


Fig. 1 Illustration of bus bunching at bus station

[33] proposed a predictive framework to capture the stop-level headway irregularity based on transit smart card data. The bus arrival time can be inferred from smart card data, which stores passengers' boarding time and location. However, smart card data do not record passengers' getting off time, so it is hard to estimate the bus departure time at stations accurately. Some other studies have analyzed bus bunching based on GPS trajectory data [34, 35]. However, those studies focused on the prediction of bus arrival time at stations with the GPS data, rather than evaluating the spatiotemporal features of bus bunching on actual bus routes. Until recently, it is still unknown how to analyze the bus bunching effect with the point-by-point empirical GPS data where significant uncertainties and outliers exist.

Identifying the causes for bus bunching is also a crucial research topic. Previously, most researchers tried to identify the factors which could affect bus bunching based on statistical methods. For example, Feng and Figliozzi [35] investigated the causes of bus bunching using Automatic Vehicle Location (AVL) and Automatic Passenger Count (APC) data, unveiling the influence of late departures from stops. Fonzone et al. [24] created a continuous logit model for emulating passenger behavior at the bus stop. Moreira-Matias et al. [36] found that the occurrence of bus bunching amplifies in the next stops by time-series analysis. Recently, a binary logistic regression and survival analysis was conducted to identify some significant factors, including stop location, departure headways, and the vehicle. Recently, a binary logistic regression and survival analysis was conducted to determine essential factors, including stop location, departure headways, and vehicle types [37]. Later, some research focused on the bus bunching characteristics in the service network. For example, Verbich et al. [38] found that bus bunching increased dwell time and bus travel time in a network. Later, Diab et al. [39] investigated the influence of increased headway delays and service times caused by overlapping lines on bus bunching. Schmöcker et al. [40] also explored the impact of overlapping lines on service regularity. Later, Wu et al. [41] conducted the simulation analysis to corroborate the positive effect of bus overtaking and distribute passenger boardings on service regularity. They underline that the proposed benefits would be higher for high-frequency services.

The literature review show that most previous studies focused on the factors influencing bus bunching and the links between bunching and networks. At the same time, few researchers explored the spatiotemporal features and deep causes of bus bunching. The previous methodologies are mostly theoretical analysis and simulation analysis. They cannot be directly used to identify the causing factors for bus bunching with the empirical GPS trajectory data. A study should be conducted to propose the data mining approach framework that can be followed to explore the spatiotemporal features and causes of bus bunching with the real GPS data collected on actual bus routes.

The primary objective of this study is to analyze the temporal-spatial features of bus bunching by conducting an in-depth analysis of empirical bus GPS trajectory data obtained in Nanjing, China. A data processing procedure was followed to estimate each bus's arrival and departure times at stations. Then the spatiotemporal trajectory picture was drawn for the bus route. The study also analyzed the headway features of consecutive buses at the different stations. The findings of the study are helpful to transit management agencies to understand the trend and severity of bus bunching in the city and develop countermeasures to reduce bus bunching and improve transit service quality. The rest of the study is organized as follows. Section II introduces the data used for analysis. Section III presents the data processing methodologies for bus bunching analysis. Section IV shows the major results and findings. The paper ends with concluding remarks in section V.

2 Data Description

The GPS trajectory data used for our analysis is obtained from the Nanjing Bus Company. Bus route 100 was selected for study, as shown in Fig. 2. The bus stations are given in Table 1. The reason for choosing this route is that it crosses the city's business center with significant uncertainties on the bus route. In addition, a high passenger demand requires a high service quality.

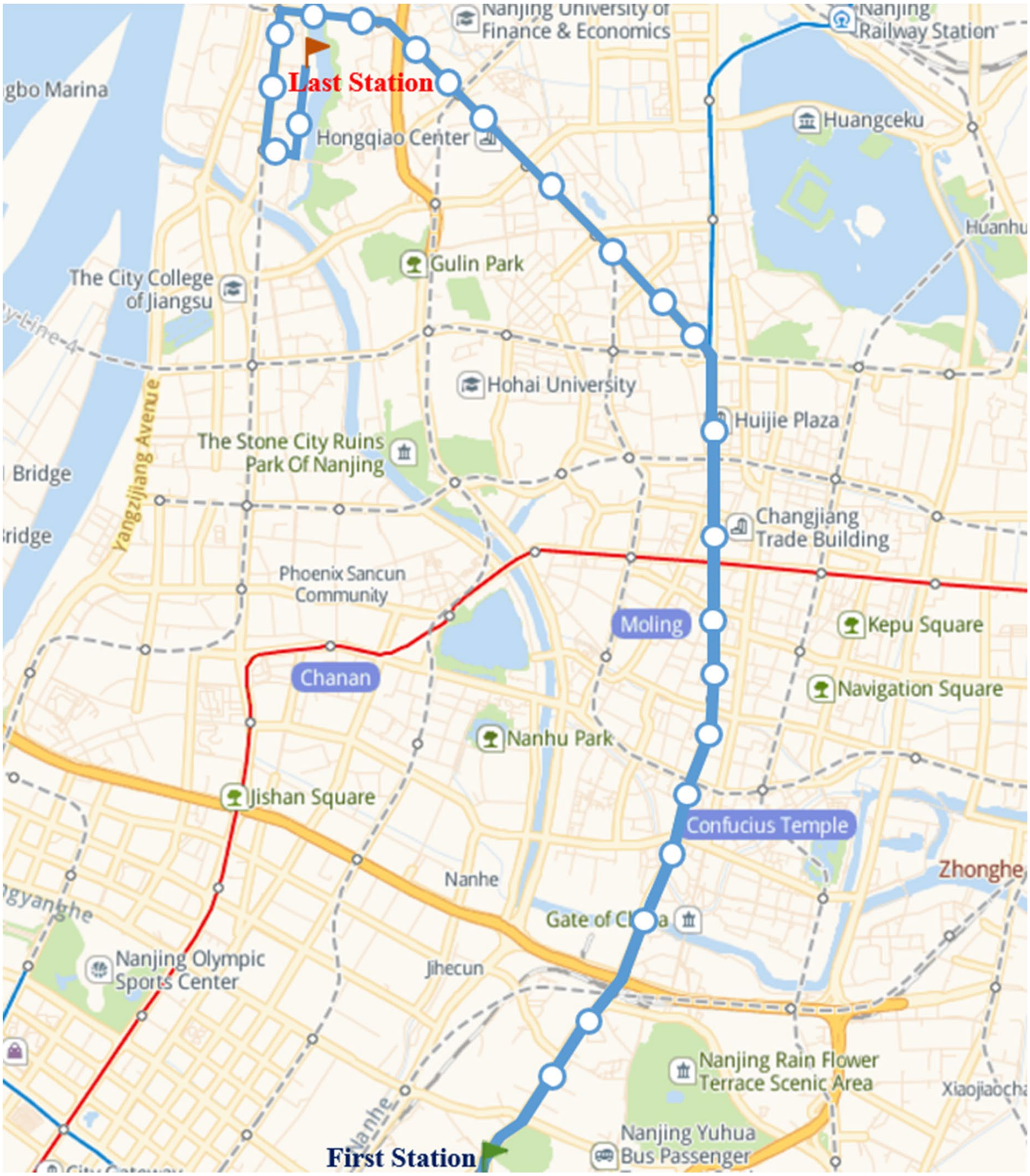


Fig. 2 Illustration of study bus route in Nanjing

The company's data system keeps GPS records updated per 10 s. We reconstructed the GPS data by a linear interpolation approach to estimate the values between two sampling timestamps. In total, about 690,000 GPS records were generated for the bus route. The samples of GPS trajectory data are shown in

Table 1 Stations on the Study Bus Route

Sequence	Longitude	Latitude	Station
1	118.7576922	31.99584545	Andemen
2	118.7632571	32.00177399	Nengrenli
3	118.7665624	32.00603138	Yuhua West Road
4	118.7723052	32.01488849	Yaowan Street
5	118.7743442	32.01867141	Diaoyutai
6	118.7755583	32.02235154	Xinqiao
7	118.7776964	32.0279242	Shengzhou Road
8	118.7789635	32.03431713	Sanyuan Lane
9	118.7789481	32.03837703	Xinjiekou South
10	118.7789244	32.04574066	Xinjiekou North
11	118.7789237	32.05344038	Zhujiang Road North
12	118.7761272	32.0633884	Gulou
13	118.7737512	32.06543433	Dafang Lane
14	118.7690538	32.06953178	Shanxi Road
15	118.7635787	32.07427056	Hongqiao
16	118.7564478	32.08043994	Sanpailou
17	118.752448	32.08389528	Sajiawan
18	118.7507084	32.08539018	South Medial Hospital Eastern
19	118.7471137	32.08804923	Yancang Bridge
20	118.7422919	32.08860244	Yijiang Door
21	118.7369385	32.08764637	Rehenan Road
22	118.7364501	32.08371669	South Medial Hospital
23	118.7366726	32.07725623	Jiangjiawei
24	118.7386088	32.07836592	Jiangjiayuan South
25	118.7393941	32.08363186	South Medial Hospital Terminal

Table 2 Samples of GPS Trajectory Data

Bus ID	Latitude	Longitude	velocity	Local Time	Mileage
2,017,030,708	32.005241	118.765823	28.02	170,307,073,309.00	3,242,240
2,017,030,708	32.00573	118.766273	18.797	170,307,073,319.00	3,249,149
2,017,030,708	32.005844	118.766357	0.0185	170,307,073,329.00	3,250,830
2,017,030,708	32.005844	118.76635	0.0185	170,307,073,339.00	3,250,908
2,017,030,708	32.006077	118.766502	16.816	170,307,073,349.00	3,253,846
2,017,030,708	32.006645	118.76696	30.169	170,307,073,359.00	3,261,506
2,017,030,708	32.007004	118.767578	4.1855	170,307,073,409.00	3,268,589
2,017,030,708	32.00713	118.767761	13.852	170,307,073,420.00	3,270,972
2,017,030,708	32.007317	118.767998	7.4265	170,307,073,430.00	3,274,010
…	…	…	…	…	…

Table 2. The Bus ID was used to identify bus vehicles. The latitude and longitude information and the velocity at each time stamp were recorded for each bus. The total mileage was also provided for data cleaning and processing. The characteristics of the bus line and the speed information of buses are shown in Table 3.

Table 3 Characteristics and speed information in the bus line

Sequence	Number of intersections	Distance of bus station (m)	Maximum number of lanes	Average speed (m/s)	Max/min speed (m/s)	Standard deviation of speed
1	2	849	2	12.7	13.4/12.0	0.56
2	3	568	2	12.2	12.9/11.8	0.64
3	3	1200	2	12.0	12.8/11.5	0.43
4	3	457	2	11.7	12.4/11.4	0.65
5	1	411	2	11.8	12.3/11.4	0.55
6	3	738	3	11.6	12.3/11.0	0.63
7	3	608	2	11.3	12.2/10.5	0.46
8	2	472	2	10.9	11.6/10.0	0.50
9	3	835	2	11.2	11.9/10.2	0.48
10	2	849	2	10.2	10.8/9.4	0.42
11	4	1000	2	10.3	10.9/9.1	0.47
12	1	313	2	9.5	10.3/8.2	0.43
13	1	204	2	9.6	10.2/8.2	0.40
14	1	700	3	10.4	10.9/9.2	0.48
15	2	921	3	10.7	11.4/9.7	0.38
16	1	566	2	11.5	12.1/10.8	0.35
17	1	240	2	11.7	12.9/10.2	0.45
18	2	475	2	11.2	12.8/10.5	0.38
19	2	433	2	10.9	11.9/9.9	0.53
20	2	585	2	11.7	12.7/10.9	0.43
21	3	294	2	12.2	12.8/11.2	0.36
22	4	741	2	12.4	12.9/11.5	0.52
23	2	215	2	12.2	12.7/11.7	0.42
24	1	245	2	12.4	13.2/11.8	0.55

3 Methodology

3.1 Bus Bunching Measurement

The headway regularity can measure the occurrence of bus bunching. "Time headway" in traffic flow is the elapsed time between the instant of time that one car begins passing a fixed location and the moment that the following vehicle begins to pass that location (see Fig. 2). The traffic conditions and the passenger demand affect headway regularity on a bus route. As can be seen in Fig. 3, the headway fluctuation is apparent. The original headway is the same when the bus departs from the initiating station. The bus bunching occurs when the two tracts intersect (see the red point in Fig. 2). The uneven time headway may cause the bus bunching. The variation of headway indicates the instability of bus operation.

In our study, the mean time headway between consecutive buses and the standard deviation of time headway were considered crucial indicators for bus bunching. The two indicators are calculated as follows:

$$h_{in} = t_{in} - t_{(i-1)n} \tag{1}$$

where

h_{in}: headway of the n^{th} bus at the i^{th} station

t_{in}: the arrival time of the n^{th} bus at the i^{th} station

$t_{(i-1)n}$: the arrival time of the n^{th} bus at the i-1^{th} station

$$h_{m(i)} = \frac{1}{N}\left(h_{i1} + h_{i2} + \ldots + h_{iN}\right) \tag{2}$$

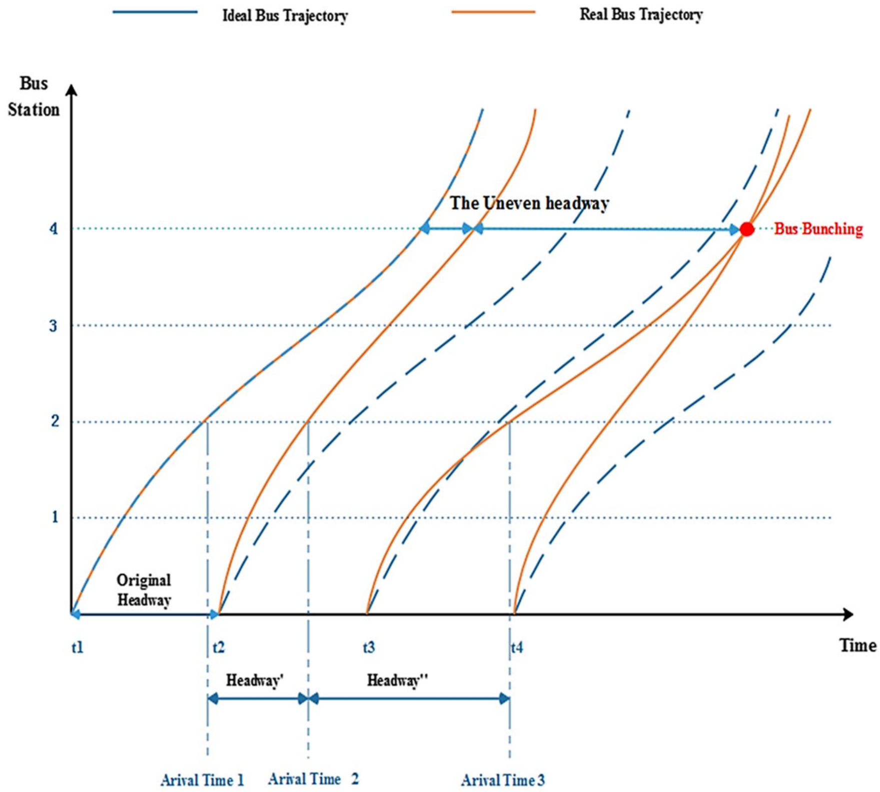


Fig. 3 Bus bunching measurement in bus trajectories

$$h_{std(i)} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(h_{in} - h_{m(i)})^2} \tag{3}$$

where

N: the total number of buses

$h_{m(i)}$:the mean headway of all buses at the i^{th} station

$h_{std(i)}$: the standard deviation of headway of all buses at the i^{th} station

According to Eq. 1, the time headway of the n^{th} bus at the i^{th} station is the time gap that the n^{th} bus passes two consecutive bus stations, i-1^{th} and i^{th}. $h_{m(i)}$ denotes the mean headway of all bus at i^{th} station and $h_{std(i)}$ denotes the standard deviation of time headway. According to Eq. 2 and Eq. 3, the two indicators can be calculated accordingly.

As for their meaning, the two indicators are mean time headway and standard deviation of time headway for all buses passing a specific bus station. In terms of the differences, the mean time headway reflects the level of the time headway at a specific bus station. In contrast, the standard deviation reflects the differences between different buses' headway at the same bus station.

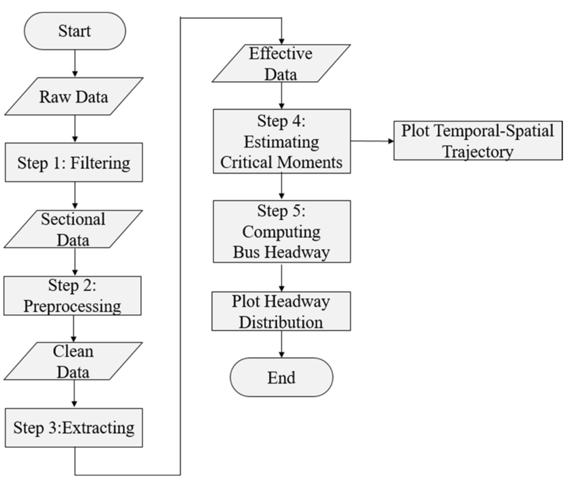


Fig. 4 Bus GPS data processing procedure

3.2 Data Processing

The primary purpose of the data processing procedure is to extract critical information from massive GPS records for identifying bus bunching measurements (see Fig. 3). The overall procedure of data processing is illustrated in Fig. 4. First, data filtering was adopted to eliminate unreasonable and erroneous data. A method was then put forward to extract practical trips during the studied time. Subsequently, an algorithm was proposed to estimate the bus arrival and departure time at every station. The bus bunching measurements were calculated for multiple buses at different stations based on such information.

The details of the data processing step are given as follows:

Step 1: Filtering all bus GPS data between 06:30:00 and 09:30:00 AM and eliminating duplicate records. The study period was from 6:30 am to 9:30 am, covering both off-peak and peak hours. By analyzing the data in the three hours, typical characteristics of the bus bunching period could be identified.

Step 2: Data cleaning was performed to eliminate some records that are unreasonable. After a closer inspection, some outliers have been observed and should be eliminated. We first removed the GPS data that was located outside of the road area. Then, we filled the vacant data with the estimated one by averaging the data point before and after it. Specifically, some trips have absurdly scarce records; some trips have missing records in a short duration of a complete trip, and some buses are just wandering near the shunting yard. These situations mentioned above were then eliminated.

Step 3: Extract bus GPS records of all valid upper stream trips generated during the time above duration. As shown in Fig. 5, we first specify segmental records whose recordings are continuous. The segment is the road region between two adjacent intersections. Hence, a bus recording may include several segments. When a segment represents incomplete bus recordings, another segment that follows is then searched and combined. When a segment contains all records of a bus, we then check whether it is reasonable via its traveled mileage. With all adequate buses collected, all upper stream trips from Stop 1 to Stop 25 are separated from the buses, giving us 31 effective trips. To ensure the accuracy of the bus GPS data and reduce the measurement errors, some odometer readings of some buses were collected for cross-validation purposes. Odometer positioning was achieved by tracing the distance measured by the odometer along the bus route road, which can be considered the ground truth data. We paired the two kinds of data to validate the accuracy of the GPS data. Results showed that the bus GPS data measurement error was within an acceptable range. We further conducted data cleaning work to reduce the outliers and improve the data quality in the data processing.

Step 4: Estimating every station's vehicle arrival/departure time with the following procedures.

(1) We first defined a station velocity v_s as the average travel velocity during arriving and leaving the station. Then we calculated the statistical distribution of the speed data and removed those beyond the two times of standard deviation. Finally, we run the moving average algorithm to smooth the data and reduce outliers' impact. Two respective moments, t_1, t_2, together with according velocity v_1, v_2 were then specified as the beginning/ending moments of the vehicle closest to the bus stop.

(2) For intermediate stations, we extrapolated the records around $t_1(t_2)$ to find a recording whose instant travel velocity was more significant than the velocity threshold, and the recording gives us $t_1^{'}(t_2^{'})$ and $v_1^{'}(v_2^{'})$. Station velocity can now be computed as (take arriving as an example)

$$v_s = \begin{cases} \frac{v_1 + v_1^{'}}{2}, & if\ v_1 > velocity\ threshold \\ \frac{v_1^{'}}{2}, & else \end{cases} \tag{4}$$

(3) Station travel time Δ_t was defined as the travel time in the vicinity of a bus stop. This study was the time spent at a certain distance (concretely, set Distance = 30 m) with station velocity. The arriving moment as estimated as

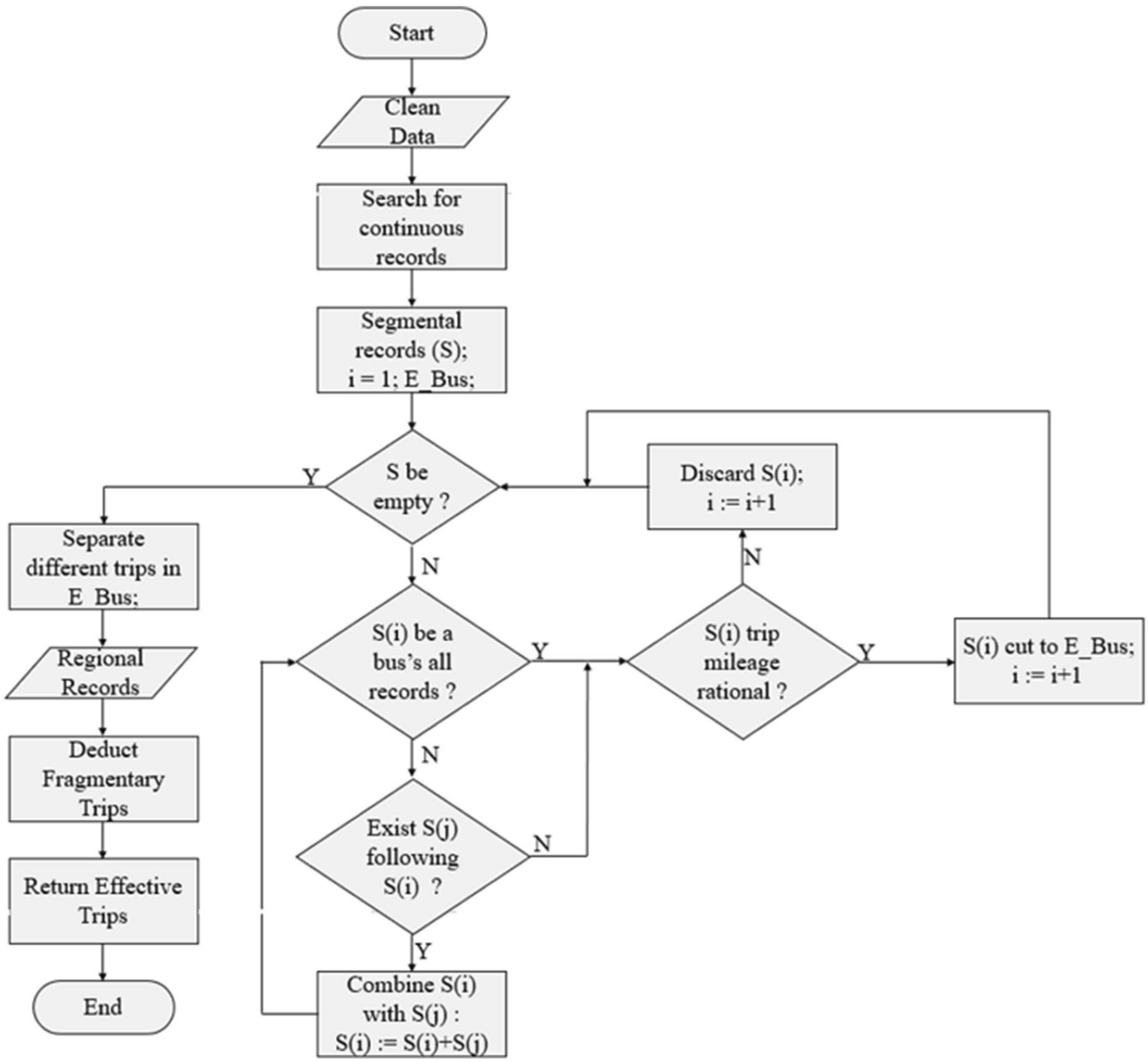


Fig. 5 Procedure to extract effective trips in step 3

$$T_a = \begin{cases} t_1 - \frac{Distance}{v_s}, & if\ v_1 > velocity\ threshold \\ t_1' - \frac{Distance}{v_s}, & else \end{cases} \tag{5}$$

(4) The departure moment at a stop is estimated similarly.

(5) As to both ends of the route, it's only necessary to estimate the moment of leaving the origin and arriving at the terminus. The estimating procedure is also shown in Fig. 6.

Step 5: Computing the time headway of adjacent vehicles at each station. Time headway in this study is computed as the time difference between arriving moments of adjoining vehicles.

4 Results

After the data processing, the spatiotemporal trajectories of 31 buses were drawn based on each station's arrival and departure time. The bus trajectories are

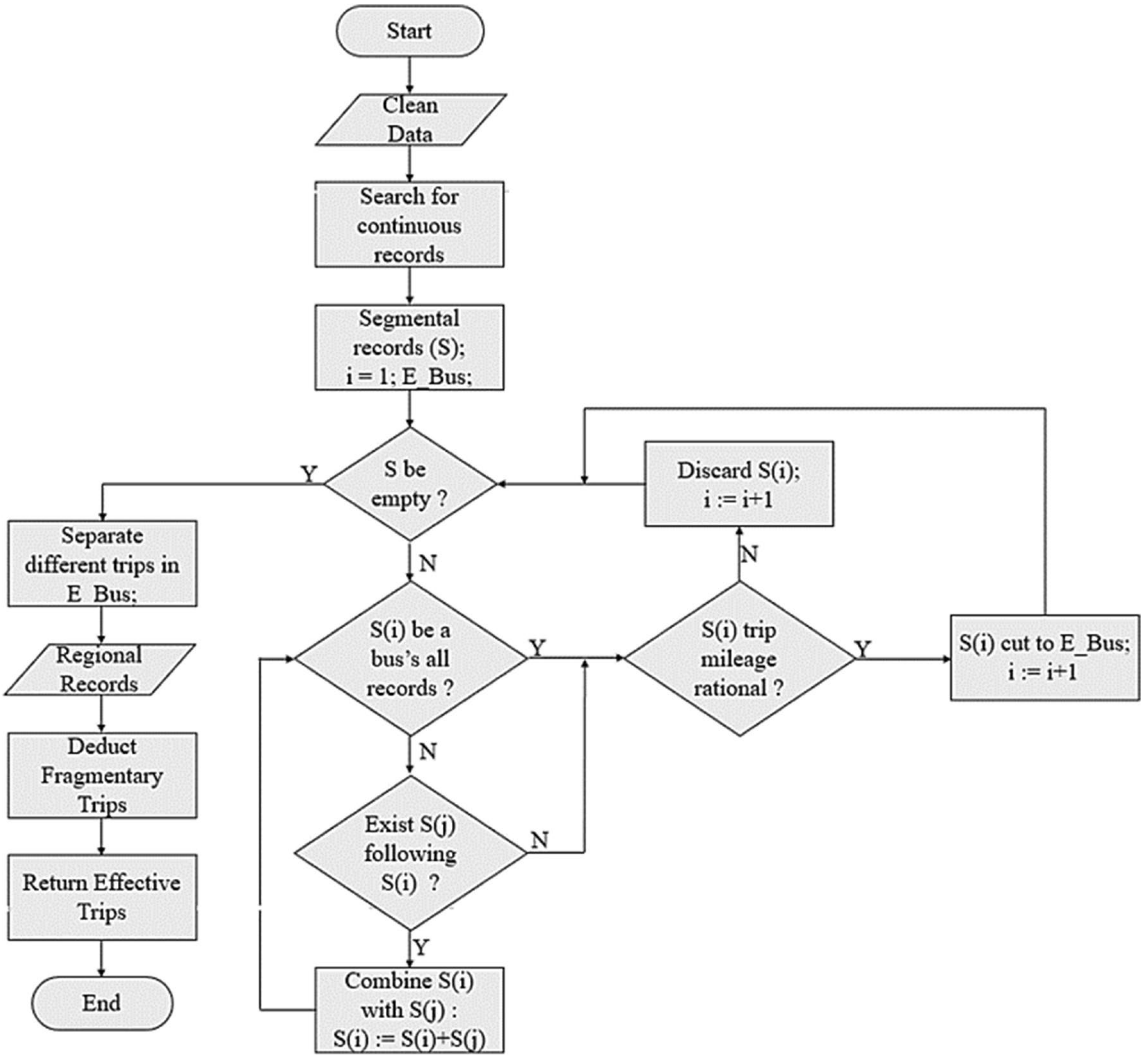


Fig. 6 Procedure to estimate arrival and departure time in step 4

shown in Fig. 7. We assumed a linear link between the departure time at a station and the arrival time at its next station. This link is not linear in actual trajectory because a bus may stop at intersections. The assumption was for better illustration instead of statistical computation since we don't care about how buses run between two stations.

In Fig. 7, the bus bunching can be identified in the spatiotemporal trajectory picture. Closer spacing between two trajectories at a station indicates a more significant probability of bus bunching. Consequently, more intensive trajectories in a spatiotemporal region indicate a more severe bus bunching phenomenon. And a significant distance between two bus trajectories at a station suggests that the waiting time of passengers at the station is very long. It should be noted that the bus departures at the first station at the beginning of the study period are already unbalanced, probably because bus operation has already been disturbed before the study period.

As reported in previous studies, multiple factors can trigger the occurrence of bus bunching. Several reasons for bus bunching can be identified in the spatiotemporal trajectory figure. As shown in Fig. 8, a part of the bus trajectory encircled in a red circle in Fig. 8 suggests that bus bunching occurs at stop eight

Fig. 7 Spatiotemporal trajectory picture for bus route

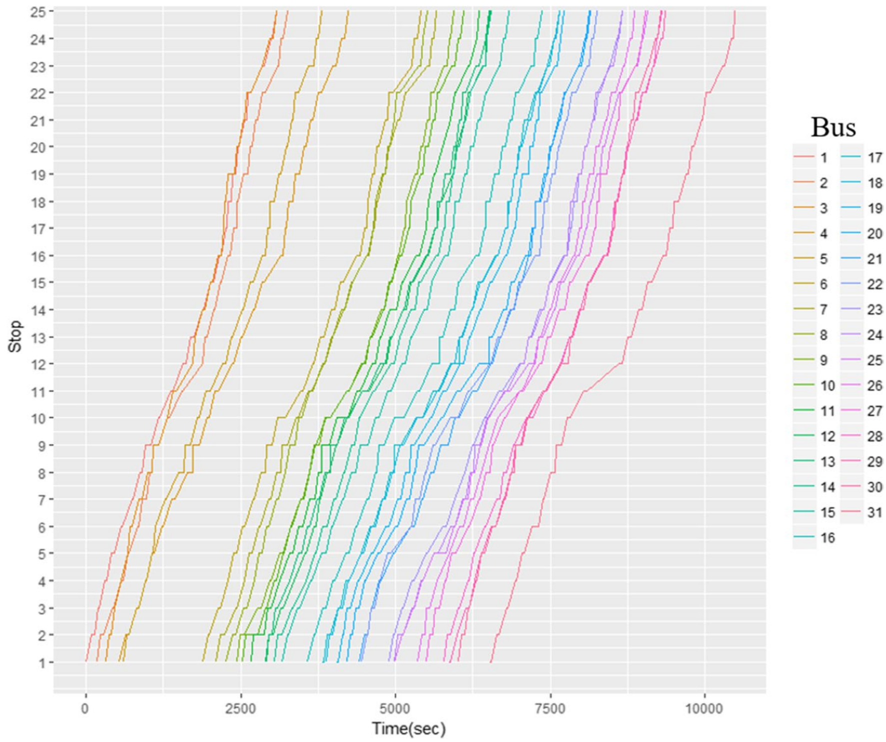
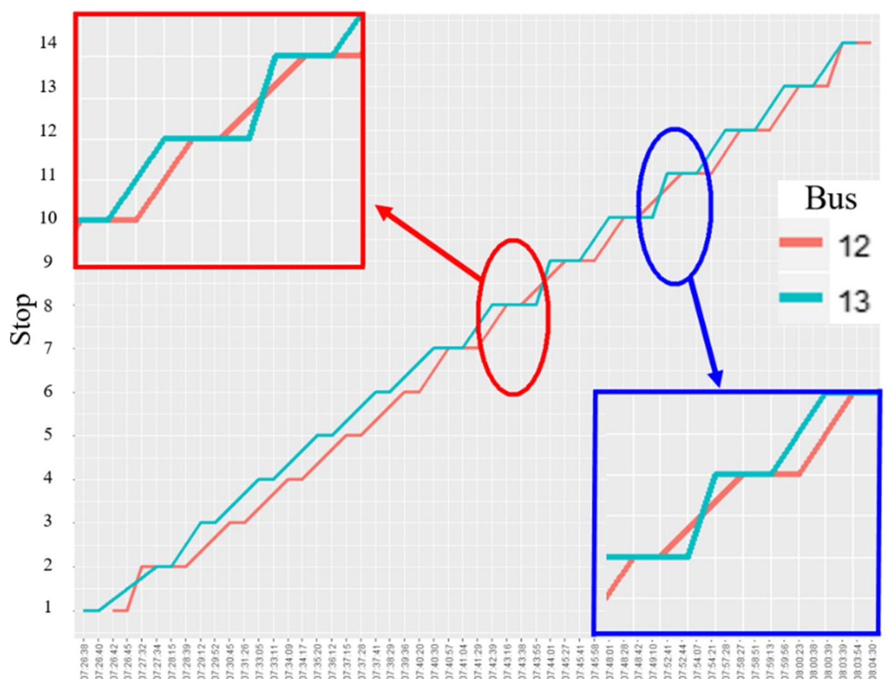


Fig. 8 Dwelling time as a cause for bus bunching



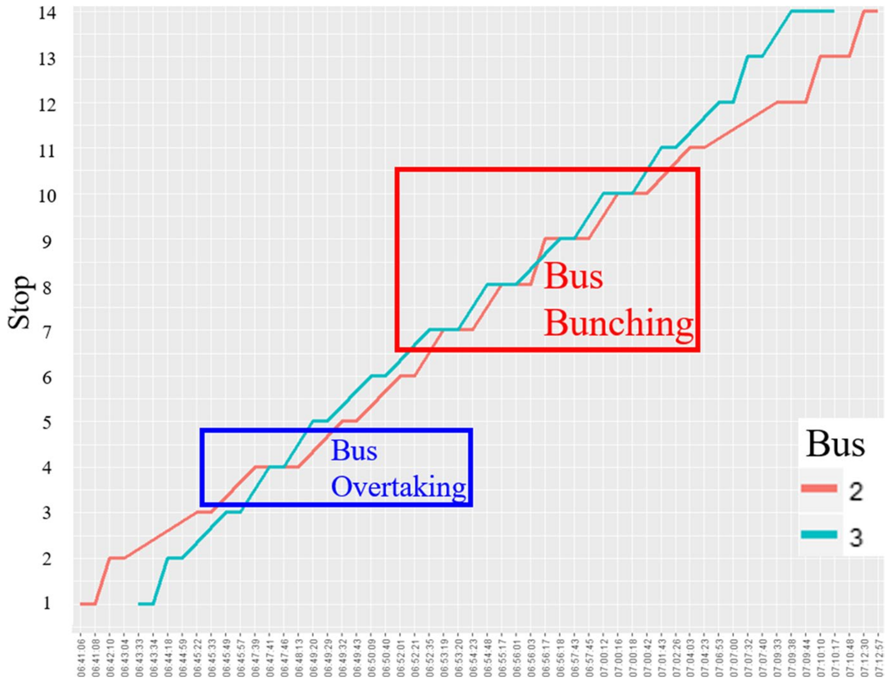


Fig. 9 Trajectories showing bus bunching during operation

as a consequence of the excessive dwelling time of bus 13. Bus 12 stays at stop 8 for a long time, and bus 13 quickly catches up, contributing to the bus bunching. Similar pattern can also be identified in the trajectory figure.

The trajectories in the blue circle in Fig. 8 illustrate that a bus bunching was caused by the difference in travel time between the two stations. As can be identified, bus 13 ran faster than bus 12 and overtook bus 12 (probably because it was temporally blocked by other vehicles on the lane) during running between two stations. As a result, bus 13 arrived the station 11 first but had to wait longer for the dwelling, which caused another bunching occurrence at station 11. Figure 8 also shows that once a bus bunching initially occurred on the route, it could persist for a long time in the downstream stations.

The spatiotemporal trajectory files in Fig. 9 show how the bus overtaking causes bus bunching during the operation. In station 1 buses 2 and 3 departed as scheduled with a reasonable headway. The headway was kept when they arrived stop 2. However, between stop 2 and 3, bus 2 runs slower than bus 3; consequently, the headway between the two buses was much smaller than the scheduled value. At stop 4, bus 2 stayed longer than bus 3 due to the passenger dwelling, and bus 3 overtook bus 2 at the stop. A reasonable headway was kept for a while before the difference in bus travel time between stop 6 and 7 caused another bunching occurring at stop 7. The two buses tangled, resulting in bus bunching at the following several stops. Finally, the headway between the buses became larger due to the difference in dwelling time and travel time, and the bunching did not occur in the last several stations. The headway between the two buses is also shown in Fig. 9. A bus overtaking may not always result in a bus bunching. The conditional rule is that the headway between the two buses does not keep at a reasonable range. If the headway is too small, a bunching will likely occur at the following stations. If the headway is kept reasonably large (around the designed value), the bunching will not likely occur. However, if the headway is too large, it indicates that the bus waiting time at one or more stations is too long, and a bunching will occur at those stations.

According to empirical records, the bunching time increased with decreased bus speed, also the bunching length. The reason is that when buses run in a congested region, their velocities decrease, and they even queue up at bus stations. In addition, the bunching time increased downstream, especially in the city's central region with busy streets.

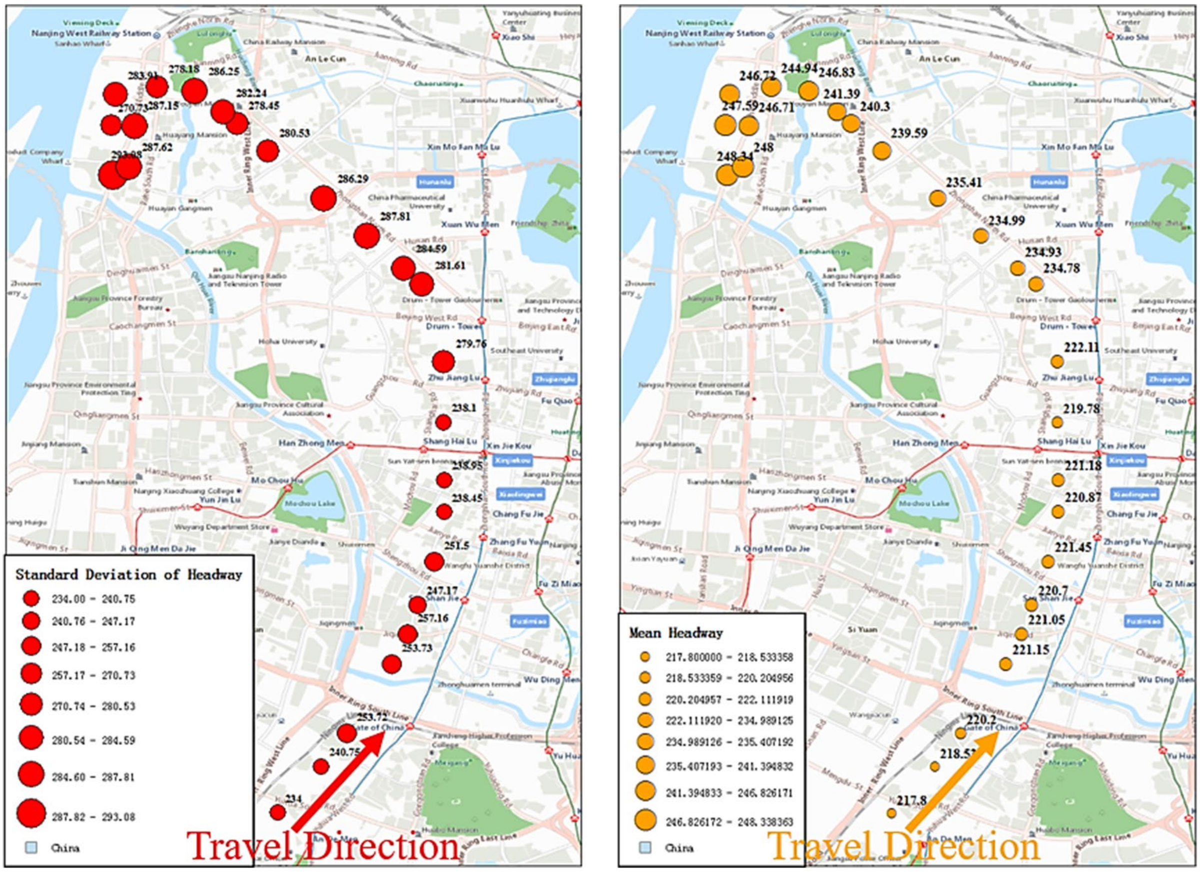


Fig. 10 Headways features at different stations

The mean headway and the standard deviation at different stations on the bus route are shown in Fig. 10. Mean headway reflects the difference in bus arrival time, and the standard deviation of headway indicates fluctuation intensity. Comparing vehicle headway at different stations can assist in determining the likelihood of a bunching phenomenon. From the figure below, the mean headway slightly increased in the several initial stations and then grew quickly after passing station 11. In our case, station 11 locates within the city center, so both passenger arrival rate and bus travel time are highly likely to be affected by uncertain factors. The standard deviation of headway also experienced a significant fluctuation with a similar pattern. The results suggested that bus bunching is more severe at the downstream stations on the bus route. The probable reason is that once disturbed, the bus operation could become more severe due to the commutative effect of uncertainties.

5 Conclusions

This study developed a data processing procedure to identify recurrent bus bunching incidents at stations based on the GPS trajectory data. The spatiotemporal features of bus bunching on a bus route were analyzed by in-depth analysis of the empirical bus GPS trajectory data obtained in Nanjing, China. A procedure was followed to estimate each station's arrival and departure time, including data cleaning, trip cutting, arrival and departure time estimation, and time headway calculation. Then the spatiotemporal trajectory picture was drawn for the bus route. The study also analyzed the headway features of consecutive buses at the different stations.

The data analysis results showed that there are significant differences in the spatiotemporal features of bus bunching between bus stations. When the bus bunching first took place, it persisted on downstream stations for a long time. The bunching severity was found to dramatically increase at downstream stations, reducing bus arrival reliability on the whole bus line. We also identified that the bus bunching was primarily caused by the overlong bus dwelling time at a station and the different travel times of buses between stations. The study fills the gap by developing the methodology to investigate the bus bunching features and causes with point-by-point empirical GPS trajectory data.

This study's findings helped analyze the characteristic of bus bunching using GPS trajectory data. Knowing the features and the probable causes for bus bunching can help alleviate or even eliminate the occurrence of bus bunching. The data processing procedure utilized in this study can be easily transferable to other city bus routes. The study can help transit management agencies develop countermeasures to improve transit service quality. Firstly, this paper proposed a novel data processing procedure to estimate each station's arrival and departure time. Secondly, this study explored the spatiotemporal evolution of the bus bunching phenomenon. The government can draft effective policies by adjusting the bunching distribution. Lastly, it was found that the time headway and bus dwelling time were the crucial factors causing bus bunching time. Future bus bunching control strategies can be proposed based on the prediction of the bunching occurrence. For example, each bus can install a headway control device that suggests the optimal headway between the bus and its preceding and following ones. The device can guide the bus driving at different speeds to keep the optimal headway. In addition, the bus can be directed to jump over some stations to prevent bunching. Such countermeasures should be tested in both simulation and practical applications.

This study can be further enhanced by considering the following future research directions. First, GPS trajectory data from more bus routes on more cities can be used to validate the methodology proposed in this study. Findings from more data can also be validated and compared. Second, other advanced prediction methods, such as the Bayesian probabilistic model [42], could be tested in future to improve the prediction performance of bus bunching. Last but not least, more factors should be considered in the bus bunching cause analysis, such as weather conditions, traffic conditions, bus route features, etc., to make the research more solid and practical.

6 Abbreviations

Not applicable

Authors' Contributions Xiaofeng Shan: conception or design of the work, data collection, data analysis and interpretation, drafting the article, critical revision of the article.

Chishe Wang: critical revision of the article.

Dongqin Zhou: data collection, data analysis and interpretation.

Data Availability Not applicable

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication This is to state that the authors give full permission for the publication of the article.

Competing Interests The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

References

1. Li, M., Li, Z., Xu, C., Liu, T.: Short-term prediction of safety and operation impacts of lane changes in oscillations with empirical vehicle trajectories. Accid. Anal. Prev. **1**(135), 105345 (2020)
2. Wang, C., Xu, C., Dai, Y.: A crash prediction method based on bivariate extreme value theory and video-based vehicle trajectory data. Accid. Anal. Prev. **1**(123), 365–373 (2019)
3. Gu, X., Abdel-Aty, M., Xiang, Q., Cai, Q., Yuan, J.: Utilizing UAV video data for in-depth analysis of drivers' crash risk at interchange merging areas. Accid. Anal. Prev. **1**(123), 159–169 (2019)
4. Guo, Y., Li, Z., Liu, P., Wu, Y.: Modeling correlation and heterogeneity in crash rates by collision types using full Bayesian random parameters multivariate Tobit model. Accid. Anal. Prev. **1**(128), 164–174 (2019)
5. Wang, C., Xu, C., Xia, J., Qian, Z., Lu, L.: A combined use of microscopic traffic simulation and extreme value methods for traffic safety evaluation. Transp. Res. C Emerg. Technol. **1**(90), 281–291 (2018)
6. Liu, Z., Liu, Y., Meng, Q., Cheng, Q.: A tailored machine learning approach for urban transport network flow

estimation. Transp. Res. Emerg. Technol. **1**(108), 130–150 (2019)
7. Li, D., Jin, C.J., Yang, M., Chen, A.: Incorporating multi-level taste heterogeneity in route choice modeling: From disaggregated behavior analysis to aggregated network loading. Travel Behav. Soc. **1**(19), 36–44 (2020)
8. Sun, C., Cheng, L., Zhu, S., Han, F., Chu, Z.: Multi-criteria user equilibrium model considering travel time, travel time reliability and distance. Transp. Res. Part D: Transp. Environ. **1**(66), 3–12 (2019)
9. Liu, Z., Wang, S., Zhou, B., Cheng, Q.: Robust optimization of distance-based tolls in a network considering stochastic day to day dynamics. Transp. Res. C Emerg. Technol. **1**(79), 58–72 (2017)
10. Liu, Z., Chen, X., Meng, Q., Kim, I.: Remote park-and-ride network equilibrium model and its applications. Transp. Res. B Methodol. **1**(117), 37–62 (2018)
11. Ji, Y., Ma, X., He, M., Jin, Y., Yuan, Y.: Comparison of usage regularity and its determinants between docked and dockless bike-sharing systems: a case study in Nanjing, China. J. Clean. Prod. **11**, 120110 (2020)
12. Ji, Y., Fan, Y., Ermagun, A., Cao, X., Wang, W., Das, K.: Public bicycle as a feeder mode to rail transit in China: The role of gender, age, income, trip purpose, and bicycle theft experience. Int. J. Sustain. Transp. **11**(4), 308–317 (2017)
13. Guo, Y., Li, Z., Wu, Y., Xu, C.: Exploring unobserved heterogeneity in bicyclists' red-light running behaviors at different crossing facilities. Accid. Anal. Prev. **1**(115), 118–127 (2018)
14. Guo, Y., Li, Z., Wu, Y., Xu, C.: Evaluating factors affecting electric bike users' registration of license plate in China using Bayesian approach. Transport. Res. F: Traffic Psychol. Behav. **1**(59), 212–221 (2018)
15. Bai, L., Liu, P., Chan, C.Y., Li, Z.: Estimating level of service of mid-block bicycle lanes considering mixed traffic flow. Transp. Res A Policy Pract. **1**(101), 203–217 (2017)
16. Cheng, L., Chen, X., Yang, S., Cao, Z., De Vos, J., Witlox, F.: Active travel for active ageing in China: The role of built environment. J. Transp. Geogr. **1**(76), 142–152 (2019)
17. Fu, X., Lam, W.H.: Modelling joint activity-travel pattern scheduling problem in multi-modal transit networks. Transportation **45**(1), 23–49 (2018)
18. Wang, C., Sun, Z., Ye, Z.: On-Road Bus Emission Comparison for Diverse Locations and Fuel Types in Real-World Operation Conditions. Sustainability. **12**(5), 1798 (2020)
19. Pan, Y., Chen, S., Qiao, F., Ukkusuri, S.V., Tang, K.: Estimation of real-driving emissions for buses fueled with liquefied natural gas based on gradient boosted regression trees. Sci. Total Environ. **10**(660), 741–750 (2019)
20. Liu, Y., Liu, Z., Jia, R.: DeepPF: A deep learning based architecture for metro passenger flow prediction. Transp. Res C Emerg. Technol. **1**(101), 18–34 (2019)
21. Xu, C., Zhao, J., Liu, P.: A geographically weighted regression approach to investigate the effects of traffic conditions and road characteristics on air pollutant emissions. J. Clean. Prod. **1**(239), 118084 (2019)
22. Newell, GF, Renfrey, B.P.: Maintaining a bus schedule. Australian Road Research Board (ARRB) Conference, 2nd, Melbourne. **2**(1) (1964)
23. Figliozzi, M.A., Wu-chi F., Gerardo L., Wei F.: A study of headway maintenance for bus routes: causes and effects of "Bus Bunching" in extensive and congested service areas. OTREC-RR-12-09. Portland, OR: Transportation Research and Education Center (TREC), (2012). https://doi.org/10.15760/trec.107
24. Fonzone, A., Schmocker, J., Liu, R.: A Model of Bus Bunching under Reliability-based Passenger Arrival Patterns. Transp. Res. Procedia **7**, 276–299 (2015)
25. Feng, W., Figliozzi, M.A.: Empirical analysis of bus bunching characteristics based on bus AVL/APC Data. Civil and environmental engineering faculty publications and presentations. 315, (2015). http://archives.pdx.edu/ds/psu/16620
26. Lüthi, M., Weidmann, U., Nash, A.: Passenger arrival rates at public transport stations. In: Transportation research board board 86th annual meeting, 07-0635 (2007)
27. Fu, L., Hellinga, B.: Delay Variability at Signalized Intersections. Transp. Res. Rec. **1710**(1710), 215–221 (2000)
28. Daganzo, C.F.: A headway-based approach to eliminate bus bunching: Systematic analysis and comparisons. Transp. Res. B Methodol. **43**(10), 913–921 (2009)
29. Daganzo, C.F., Pilachowski, J.: Reducing bunching with bus-to-bus cooperation. Transp. Res. B Methodol. **45**(1), 267–277 (2011)
30. M. Chen, S. Chien, X. Liu, and J. Brickey, "Application of APC/AVL archived data support system," In: TRB 82nd Annual Meeting (CD-ROM), Washington, D.C., 2003.
31. Chien, S.I., Ding, Y., Wei, C.: Dynamic bus arrival time prediction with artificial neural networks. J. Transp. Eng. ASCE **128**(5), 429–438 (2002)
32. Jeong, Hee, R.: "The prediction of bus arrival time using Automatic Vehicle Location Systems data," Doctoral dissertation, Texas A&M University. vol. 65–12, Section: B, pp 6530 2004.
33. Yu, H., Chen, D., Wu, Z., Ma, X., Wang, Y.: Headway-based bus bunching prediction using transit smart card data. Transp. Res. C Emerg. Technol. **72**, 45–59 (2016)
34. Padmanaban, R.P.S., Divakar, K., Vanajakshi, L., Subramanian, S.C.: Development of a real-time bus arrival prediction system for Indian traffic conditions. IET Intel. Transport Syst. **4**(3), 189–200 (2010)
35. Lin, Y., Yang, X., Zou, N., Jia, L.: Real-Time Bus Arrival Time Prediction: Case Study for Jinan, China. J. Transp. Eng. ASCE **139**(11), 1133–1140 (2013)
36. Feng, W., Figliozzi, M: Empirical findings of bus bunching distributions and attributes using archived AVL/APC bus data. Proceedings of the 11th international conference of chinese transportation professionals: towards sustainable transportation systems. 4330–4341 (2011)
37. Moreira-Matias, L., Ferreira, C., Gama, J., Mendes-Moreira, J., de Sousa, J. F.: Bus bunching detection by mining sequences of headway deviations. In advances in data mining. applications and theoretical aspects: 12th industrial conference, industrial conference on data mining, **12**, 77–91. Springer, Berlin, Heidelberg (2012)

38. Nguyen, P., Diab, E., Shalaby, A.: Understanding the factors that influence the probability and time to streetcar bunching incidents. Public Transp. **11**(2), 299–320 (2019)
39. Verbich, D., Diab, E., El-Geneidy, A.: Have they bunched yet? An exploratory study of the impacts of bus bunching on dwell and running times. Public Transp. **8**(2), 225–242 (2016)
40. Diab, E., Bertini, R., & El-Geneidy, A.: Bus transit service reliability: Understanding the impacts of overlapping bus service on headway delays and determinants of bus bunching. In: 95th Annual Meeting of the Transportation Research Board, Washington DC, USA (2016)
41. Schmöcker, J.D., Sun, W., Fonzone, A., Liu, R.: Bus bunching along a corridor served by two lines. Transp. Res. B Methodol. **93**, 300–317 (2016)
42. Wu, W., Liu, R., Jin, W.: Modelling bus bunching and holding control with vehicle overtaking and distributed passenger boarding behaviour. Transp. Res. B Methodol. **104**, 175–197 (2017)