

Research Article

A Bus-Service-Based Zone Division Approach for the Spatial Analysis of Public Bus: A Case Study in South China

Fanghao Fu^(b),^{1,2} Jiemin Xie^(b),^{1,2} Ming Cai^(b),^{1,2} and Wei Huang^(b),^{1,2}

¹Shenzhen Campus of Sun Yat-sen University, Shenzhen, China

²Guangdong Provincial Key Laboratory of Intelligent Transportation System, School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China

Correspondence should be addressed to Jiemin Xie; xiejm28@mail.sysu.edu.cn and Wei Huang; huangwei5@mail.sysu.edu.cn

Received 16 October 2023; Revised 11 February 2024; Accepted 12 February 2024; Published 28 February 2024

Academic Editor: Yajie Zou

Copyright © 2024 Fanghao Fu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Public bus is one of the major green transport modes in most cities, and a proper analysis of public bus can enhance its sustainable development. Hence, the analysis of public bus is always a hot research topic. Generally, to consider the heterogeneity in the studied area and reduce the computation difficulties, the study area is divided into multiple zones and then various analytical methods are applied. However, an improper zone division could lower the quality of the analysis results and then probably mislead the improvement strategy design. This study examines the influence of the two commonly used zone division methods (the grid-based method and the natural-boundary-based method) on the spatial analysis of public bus service using the analytical example. The result shows that the impact of zone division exists. Therefore, this study aims to identify the possible approach to reduce the negative impact and suggests a new zone division method, the bus-service-based (BS-based) method, for the spatial analysis of public bus service. The BS-based method is compared with the commonly used zone division methods using the real-world data, and the comparison results demonstrate that the BS-based method can be more efficient to support the spatial analysis of bus service.

1. Introduction

As a sustainable green travel mode, public bus is regarded as an effective tool to alleviate traffic congestion and land resource shortage caused by the motorized development. Generally, public bus in China heavily relies on the government subsidy to maintain the daily operation because the ticket fares are normally cheap so that the ticket revenue cannot cover the operation cost. However, the ticket revenue should be maximized, and thus the financial burden of the government can be relieved. Moreover, the public and the government expect the public bus to serve as many people as possible. But the loading rate of public bus is usually not high because of its relatively poor service quality compared with the private car or the customized bus [1] and the mismatch between public bus supply and residents' travel demand. The situation has become worse after the COVID-19. Thus, the current main task of the public bus operator is to improve

the service quality of public bus and meanwhile reduce bus operating costs [2–5]. To achieve this, the service characteristics of public bus should be analyzed.

As public bus services generally cover a wide city space, relevant analysis should consider the spatial dimension. Generally, the bus service analysis includes the zone division to consider the heterogeneity in the study area. For example, the levels of economic development and urbanization could vary in the study area. Zone division could be helpful in handling the heterogeneity during the analysis [6]. Furthermore, the management strategies could be specifically designed for different parts in the study area with consideration of their special characteristics [7, 8].

The main task of the zone division is to separate the whole area into smaller regions, and various zone division methods have been developed based on some specific rules to support a certain spatial analysis. For the spatial analysis, there are three general zone division methods as follows.

- The grid-based method adopts a regular shape (e.g., triangle, square, and hexagon) as the basic zone shape and then divides the study area into uniform grids [9]. The square shape is widely used because it is easily operated, and thus this study uses the square shape as the basic zone shape for the grid-based method.
- (2) The natural-boundary-based method uses natural boundaries (e.g., road and river) to generate zones with consideration of land use or resident information. This zone division method is widely used in the traditional traffic planning, and the new extension includes the cellular signaling data to improve the demand homogeneity within each zone [10]. The new extension is adopted in this study for the further tests.
- (3) The feature-based method generates zones with consideration of the difference and connectivity of local traffic characteristics, including traffic crash, land use, and resident trips. However, there are few comprehensive zone division research studies designed for exploring zonal bus service characteristics. Hence, this study suggests a new zone division method (bus-service-based method, BSbased method) which uses the bus service features (i.e., service coverage of bus) as the division criteria.

These three types of zone division methods separate a study area in different ways and thus influence the spatial analysis of the study area more or less. However, this influence is seldom analyzed, especially in the public bus network. For example, some zone division methods can result in the overestimation of the bus stop density which misleads the spatial analysis of the bus service quality. Therefore, this study aims at filling the gap, and the contributions of this study are listed as follows.

- (1) A new zone division method (BS-based method) is designed for the spatial analysis of public bus service. The new method divides the study area into smaller spatial units based on bus service coverage. The method heavily influences the division of service area of bus stops and then influences the other calculations based on bus stop. Finally, the new division method can improve the accuracy of the zonal evaluation of bus service, and support to explore the local features of bus services.
- (2) The three zone division methods (the grid-based method, the natural-boundary-based method, and the feature-based method) are tested in the analytical example of public bus to examine the impact of zone division on the spatial analysis. Moreover, the real-world data are used to compare the performance of these three zone division methods in the assessment of public bus services.

The remainder of the paper is organized as follows. Section 2 gives a literature review of relevant studies. Section 3 presents the BS-based method in detail. Section 4

compares the three methods using the analytical example and the real-world data. Section 5 concludes the work and gives possible future research directions.

2. Literature Review

This study wants to explore the influence of different zone division methods on the spatial analysis of bus service, so this review will firstly introduce the spatial analysis of bus service and then discuss the existing zone division methods.

The spatial analysis of current bus service is always a hot topic of research because the public bus system is an essential part of the modern urban transportation. One trend is to analyze the bus service in an aggregated manner. For example, Hawas et al. [11] evaluated the efficiency and effectiveness of the city line bus service for different scenarios. Esmailpour et al. [12] investigated the city-scale customer satisfaction in terms of public bus service. Paudel [13] evaluated the relationship between ridership and the reliability of bus service from the level of urban bus network. Sheng et al. [14] proposed a leader-follower game model to design quality incentive scheme to balance the cost and quality of public bus service in a large bus network. Deb et al. [15] explored the quality of the public bus service by a questionnaire survey in the city scale. The aforementioned studies do not consider the spatial heterogeneity in bus service, whereas the zone-based analysis can do so.

In the zone-based analysis, the shape of zones can be regular or irregular. For the regular shape, grids and hexagons are generally used. For example, grids are used to explore the influence of bus service on urban vitality [16], while hexagons are adopted to investigate the spatial information of opportunities to social resources [17] or explore the distribution of average crowding at bus stops in different time periods [18]. The regular shape is easily to operate, but the shape of the studied area might be irregular due to the geographic features. Therefore, irregular shape is also used, with consideration of census tracts [19], communities [20], and natural boundaries [21, 22]. Also, road network is a common factor which is used to divide zones, and divided zones can be used to analyze the travel behavior [23], the bus service quality [24], the relation between the bus service data and the spatial distribution of different income groups [25], the influence of bus stop on taxi demand [26], the impact of built environment on public transportation [27], etc. The abovementioned zone division methods could be divided into two types, the grid-based method to generate the regular-shape zones and the naturalboundary-based method to generate irregular-shape zones.

Furthermore, other traffic analyses may adopt the third type method, the feature-based method. For example, Assunção et al. [28] and Wu et al. [29] divided zones using socioeconomic features, while Dong et al. [30] used the similar land-use function (e.g., residential function and commercial function). Some traffic safety studies suggested to consider the characteristic of traffic crashes [31, 32], and the analysis for the similarity of taxi travel characteristics also encouraged the development of the feature-based methods [33, 34]. The basic idea of the feature-based method is to maximize the homogeneity in zones. For example, Martinez et al. [35] designed the hierarchical heuristic algorithm to cluster grid cells using the origin-destination (OD) data to maximize the similarity of interzonal trips of each zone, while Xing et al. [36] combined the methods of Voronoi diagram and natural boundaries and clustered small units to zones, not only to maximize the similarity of interzonal trips of each zone but also to keep a similar level of population among zones. Besides, the K-means algorithm can be used to cluster grids to zones so that the zones have a similar level of population or demand for a transport mode, e.g., the taxi demand [33], while Cai et al. [10] suggested a genetic-algorithm-based clustering algorithm to cluster units into zones with consideration of the homogeneity of resident travel characteristics.

Furthermore, the optimization of bus service needs zone division. For instance, Ibeas et al. [7] divided the study area into zones considering regional socioeconomic difference, and based on the generated zones, the urban bus stop spacing optimization was established. Soares [37] considered natural boundaries to cluster origin zones and destination zones, and then the zone-based public bus route optimization was conducted. Lee et al. [38] adopted the grid-based method to generate zones which were used in an optimization model to maximize the profit of zonal-based flexible bus services.

In all, zone division is widely used in the traffic analysis. Researchers have developed different zone division methods for various analyses to achieve a better analysis quality. However, there are few zone division algorithms which are designed for bus service analysis, and the existing zone division algorithms could influence the separation result and then the spatial bus service analysis. This study fills this research gap and checks the possible impacts.

3. BS-Based Method

This section presents a new feature-based zone division method for the spatial analysis of bus service, the BS-based method. The core idea of the BS-based method is firstly to divide the study area into small grids, and then these small grids are clustered into zones with the consideration of service coverage of bus to ensure that the generated zones can serve the bus service analysis better. The generated zones are named as bus service analysis zone (BSAZ). The details will be illustrated in this section.

3.1. Definition. A bus runs based on the predesigned route and stops at the bus stops along the route to serve people. The area around a bus stop is called the designed bus service area. People whose origin and destination are within the designed bus service area are more likely to choose the public bus for their trips. It is generally believed that if a city has more designed bus service areas, it seems that the access time of the public bus is shorter, and the public bus can provide a wider service. Therefore, the designed bus service area is an important criterion of bus service, and this study chooses it as a basis factor when generating BSAZs. That is, the size of designed bus service area in each BSAZ is kept the same, which can simplify the bus service analysis. For example, when two BSAZs are compared, the bus service quality of the larger BSAZ is poorer than that of the smaller BSAZ because the larger BSAZ has a wider area which is not covered by the public bus, i.e., the larger BSAZ has a lower public transport coverage. More relative details will be discussed in Section 4.

To generate BSAZs, some concepts are clarified as follows.

- (1) *Bus Stop*. A bus site may include two stops, which serve the upstream and downstream vehicles of a bus line, respectively.
- (2) Designed Bus Service Area. The designed bus service area of each stop is defined as a circle shape, and its radius is set based on the Code for Transport Planning on Urban Road in China [39] which determined that a bus stop should serve people within a circle area with a radius of 300 meters. The value of the radius can be changed according to the local planning requirement. As shown in Figure 1, a yellow dot is a bus stop, and the circle area along a dot is the designed bus service area. The size of the circle area of bus stop *j* is marked as s_j . Take the bus stop *D* as an example. The size of red circle area surrounding bus stop *D* is s_D .
- (3) The overlapped area is counted once. For example, as shown in Figure 1, the designed bus service areas of bus stop A and bus stop B are overlapped, and the size of the overlapped area is s_{overlap}(A,B). The size of designed bus service area for bus stop A and bus stop B is s_A + s_B s_{overlap}(A,B), rather than s_A + s_B.
- (4) If the designed bus service area of a stop falls in more than one BSAZ, the designed bus service area is divided accordingly, and the size of designed bus service area of stop *j* in zone *z* is denoted as $s_{(j,z)}$. For example, as shown in Figure 1, the designed bus service area of bus stop C falls in BSAZ₁ and BSAZ₂, and it is divided into two parts, the left part in BSAZ₁ and the right part in BSAZ₂. The respective sizes are $s_{(C1,z)}$ and $s_{(C2,z)}$. The size of designed bus service area in BSAZ₁ is $s_A + s_B s_{overlap(A,B)} + s_{(C1,z)}$, and that in BSAZ₂ is $s_D + s_E + s_{(C2,z)}$.
- (5) The standard size of designed bus service area in a BSAZ, δ, is set as a critical criterion in the BS-based method, and it will be explained in the next section. The value of δ can be set based on the need of public bus operators with consideration of the local social and economic features.
- (6) The boundary of the studied area is considered when generating BSAZs.
- (7) The trips beginning at stop *j* would be separated, if the designed bus service area of stop *j* covers more than one BSAZ, such as stop C in Figure 1. This research assumes that the distribution of trip origins



FIGURE 1: Designed bus service area distribution to each zone.

is an even distribution, so the division of trips is based on the proportion of designed bus service area in each related zone. That is, the number of trips beginning at stop j in zone z is calculated as follows.

$$N_{(j,z)}^{\mathrm{TS}} = N_j^{\mathrm{TS}} \times \frac{s_{(j,z)}}{s_j},\tag{1}$$

where N_j^{TS} is the number of trips beginning at stop *j*. And the number of trips in zone *z* is calculated as follows.

$$N_z^{\text{TZ}} = \sum_j N_{(j,z)}^{\text{TS}}.$$
 (2)

3.2. BS-Based Method. The basic process of the BS-based method is to divide the studied area into small basic units and then cluster them. Therefore, the BS-based method has two steps, i.e., basic unit generation and clustering. This section will show the details of these two steps.

3.2.1. Basic Unit Generation. The basic unit could be in any shape. Researchers generally adopt the grid as a basic unit to generate zones [40–42] because the shape of grid is simple and grids are easily generated, searched, and clustered. Hence, this study uses a grid as the basic unit. The side length of a basic unit is α (unit: m). Furthermore, the grids located on the subdistrict boundary are reshaped based on the boundary as illustrated in Figure 2.

3.2.2. Clustering. After the basic units are generated, the basic units are clustered to zones, i.e., BSAZs. The clustering algorithm is given as follows (see Algorithm 1).

To better illustrate the clustering algorithm, an example area with 25 basic units is given in Figure 3. First, the basic units are numbered according to Step 1. Unit 1 is located at the southwest corner of the example area, and Unit 2 is located at the north of Unit 1. When all basic units of the first column are numbered (i.e., no more basic units can be found at the north of Unit 5), turn to the bottom of the second column (i.e., the southwest corner of the remaining basic units) and continue numbering. Repeat these steps until all basic units are numbered.

Second, the zone clustering starts following Steps 2–5. To cluster the first zone $Zone_1$, the unselected basic unit with the smallest ID (i.e., Unit 1) is selected as the starting point

and added into the set of basic units for Zone 1 ($\mathbf{B}_1^{\text{Selected}}$). The size of designed bus service area in Zone 1 is less than the standard size of designed bus service area in a BSAZ (δ) mentioned in Section 3.1, so the unselected basic units which surround Zone 1 (i.e., Units 2, 6, and 7) are selected and added to the set of basic units waiting for adding into a zone (\mathbf{B}^{Wait}). Unit 2 has the smallest ID and thus it is firstly selected to add, followed by Units 6 and 7, because the size of designed bus service area in Zone 1 is less than the standard size δ after adding Units 2, 6, and 7. Now, \mathbf{B}^{Wait} is empty, and then new basic units should be selected. That is, Units 3, 8, 11, 12, and 13 are selected. When Unit 3 is added into $\mathbf{B}_1^{\text{Selected}}$, the size of designed bus service area in Zone 1 is larger than the standard size δ , and the clustering for Zone 1 is completed.

Third, the next zone clustering begins with unselected basic unit with the smallest ID (i.e., Unit 4), and then in order, Units 5, 8, 9, and 10 are selected to cluster Zone 2 (green grids). Each starting unit is marked with a fivepointed star in Figure 3, and Zone 1 (purple grids), Zone 2 (green grids), Zone 3 (yellow grids), and Zone 4 (blue grids) are clustered based on the given clustering algorithm.

During the clustering, two special situations may exist, and they need the special treatments (Steps 6-7).

(1) Special Situation A. As indicated in Step 4, the clustering for a zone may stop before its size of designed bus service area is larger than δ , when no unselected basic unit can be added into **B**^{Wait}. For example, as shown in Figure 4, after clustering Zone 4, Unit 22 is selected for a new clustering and its size of designed bus service area is less than δ , but there is no any unselected basic unit around it. Step 6 is given to handle such a situation. This study allows that the size of designed bus service area is less than δ , but the deviation cannot be too large. Hence, this study sets that if a zone's size of designed bus service area is less than $\mu\delta$, it will be combined with its neighbor zones. μ is a scale factor ranging from 0 to 1, and it can be set according to the analysis need. This study sets that $\mu = 0.5$.

(2) Special Situation B. The special situation B is that the ribbon-like zones (like Zone 3 in Figure 5(a)) exist. This special form reduces the zone compactness and may distort the influence of bus stops. To describe this situation, the compactness ratio [43] is adopted as follows.

$$_{i} = \frac{A_{i}}{A_{i}},$$
(3)

where A_i is the area of Zone_i (unit: km²) and A'_i is the minimum circumscribed circle area of Zone *i* (unit: km²). A larger τ_i means a higher BSAZ compactness. In Figure 5(a), Zone 3 has a large, circumscribed circle area marked by the green dot line, and it has a low compactness ratio. To solve the problem caused by the special situation B, Step 7 is performed in the clustering algorithm.

If a zone has a ribbon-like shape, the ribbon-like part is checked. For example, S'_2 in Zone 3 in Figure 5(b) is the ribbon-like part. x_w is the length of the short edge of the ribbon-like part, and x_l is the length of the long edge of the



FIGURE 2: Basic unit generation.

Input: basic units, information of the bus stops Output: BSAZs Step 1: Numbering. Number basic units from the southwest corner to the northeast corner and mark these numbers as the IDs of these basic units. Step 2: Initialization. The counter number *i* is set as 1, all basic units are put into a set **B**^{Unselected}, and the set of basic units waiting for adding into a zone, $\mathbf{B}^{\mathrm{Wait}}$, is empty. Step 3: Selection of the first basic unit. For a new zone z_i , select the basic unit with the smallest ID in **B**^{Unselected} as the starting point of the new clustering and add the basic unit into the set of basic units for z_i ($\mathbf{B}_i^{\text{Selected}}$) while it is removed from $\mathbf{B}^{\text{Unselected}}$ Step 4: Criteria check. If the designed bus service area in z_i is not less than the standard designed bus service area in a BSAZ (δ) then Step 4a: Empty B^{Wait}. else Step 4b: If \mathbf{B}^{Wait} is empty then Select the basic units which surround z_i and are in **B**^{Unselected}, and then add these selected basic units into **B**^{Wait}. else Turn to Step 5. Step 4c: If no unselected basic unit can be added into \mathbf{B}^{Wait} , then The clustering for z_i stops, and z_i is added into the set of zones in special situation A (\mathbf{Z}^A) because z_i may need further modification. else Turn to Step 5. Step 4d: If $B^{Unselected}$ is empty then Turn to Step 6. else i = i + 1 and return to Step 3. Step 5: Selection of basic units. Select the basic unit with the smallest ID in \mathbf{B}^{Wait} ; add it into $\mathbf{B}^{\text{Selected}}_{\text{Selected}}$ while removing it from \mathbf{B}^{Wait} and $\mathbf{B}^{\text{Unselected}}$; and then return to Step 4. Step 6: Special treatment A. For each zone in \mathbb{Z}^A , if its designed bus service area is less than $\mu\delta$, it is combined with the neighbor zone whose designed bus service area is less than δ and the least compared with those of other neighbor zones. If all neighbor zones' size of designed bus service area is not less than δ , the one with the smallest area is selected for the combination. Based on the combination result, put all formed zones into a set Z. Step 7: Special treatment B. Check whether the shape of each zone in Z needs a further modification (i.e., the zone is a ribbon-like zone). If yes, reshape the zone and its neighbor zones. Based on the modification, update Z. Step 8: Output. Output all zones in Z as BSAZs.

ALGORITHM 1: Clustering algorithm.

ribbon-like part. This study considers modifying Zone 3 if $x_w \le 2\alpha$ and $x_l \ge 10\alpha$. The condition can be changed based on the need of the analysis. The modification separates Zone 3 as two parts, S'_1 and S'_2 , and S'_2 is combined with the

neighboring zone which shares a long edge of Zone 3 and has the smallest area size. In this example, the selected neighboring zone is Zone 2, and we combine S'_2 with Zone 2. The modification is shown in Figure 5(c). During the



FIGURE 3: Clustering example.



FIGURE 4: Special situation A.



FIGURE 5: Special situation B. (a) Situation B. (b) The division of Zone 3. (c) Zones after the modification.

modification, A_1 , A'_1 , and A'_2 remain the same. A_2 becomes larger and thus τ_2 increases. Both A_3 and A'_3 become smaller, but the change of A'_3 is much larger. Therefore, τ_3 increases. That is, the zone compactness improves.

To test the BS-based method, we apply it in a real-world case in Section 4.

4. Analysis and Comparison

This section firstly examines the three types of zone division method, i.e., the grid-based method, the natural-boundarybased method, and the feature-based method, in an analytical example. The current zone division methods mainly include the grid-based method, the natural-boundary-based method, and the feature-based method. However, there are few feature-based methods for bus service analysis, as far as we know. Hence, the grid-based method [9] and the naturalboundary-based method [10] are used for comparison in this section, whereas the proposed BS-based method is selected to represent the feature-based method. Later, a real-world case is introduced to compare these three types of zone division methods in the spatial analysis of public bus.

4.1. Analytical Example. An analytical example is set to check the abilities of the three zone division methods in the bus service assessment. As shown in Figure 6, a study area has eight bus stops. If the grid-based method or the naturalboundary-based method is used, a possible zone division result is given in Figure 6(b). When the bus stop density is examined, it can be found that Zone 1 and Zone 3 have bus stops, whereas Zone 2 does not have a bus stop. That is, the bus stop density of Zone 2 is 0 bus stop per km². However, we cannot conclude that the bus service of the region covered by Zone 2 is poor, considering that bus stops in Zone 1 and Zone 3 may provide sufficient services if the size of Zone 2 is small. To avoid this misinterpretation, more efforts are required, such as ensuring a proper size of each zone or checking the neighbor zones. But the analysis of the bus stop density based on the BS-based method is much easier. As introduced in Section 3, the BS-based method can ensure that each zone contains at least one bus stop. As shown in Figure 6(c), the BS-based method divides the study area into Zone 1 and Zone 2 with a similar size of the designed bus service area. Zone 1 and Zone 2 have bus stops, and their bus stop densities are not zero. This feature of the BS-based method makes the generated zones more suitable for the bus stop density analysis and other bus service analyses which are illustrated in Section 4.2.

4.2. Real-World Case Study. A case study is given based on the real-world data collected in Shunde District (WIKI Shunde District: https://en.wikipedia.org/wiki/Shunde_ District (Accessed on September 26, 2022)) of Foshan (WIKI Foshan: https://en.wikipedia.org/wiki/Foshan (Accessed on September 26, 2022)), China. Foshan is a city located in the Pearl River Delta which is known as a center of advanced manufacturing and modern service industries (WIKI Pearl River Delta Economic Zone: https:// en.wikipedia.org/wiki/Pearl_River_Delta_Economic_Zone (Accessed on September 26, 2022)). The dataset used in this study includes the following data.

- Subdistrict boundary information. Shunde District has 10 subdistricts, including Chencun, Lecong, Beijiao, Longjiang, Leliu, Lunjiao, Daliang, Xingtan, Ronggui, and Jun'an, as shown in Figure 7.
- (2) Bus stop locations. The locations of 2520 bus stops in Shunde District were collected.
- (3) Bus route information. The information of 207 bus routes operating in Shunde District was collected.
- (4) Population. Data of the population were collected in November 2020.
- (5) Point of interest (POI). The POI data of Shunde District were collected in 2018. POIs include shopping centers, scenic spots, factories, automobile maintenance store, etc.
- (6) Smartcard data (SCD). When a passenger uses a smartcard to pay for a public bus service, a record of SCD is automatically generated and mainly includes the unique card number, boarding stop, boarding time, and bus line number. The SCD of Shunde District recorded in five working days (November 4–8, 2019) was given. The dataset of the studied period contains more than 1 million records. A record can be referred to one trip, and thus according to SCD and the zone division result, the average number of trips in each zone per day can be calculated as mentioned in Section 3.1 and used to assess the bus flow of each zone.

4.2.1. Zone Division Results

(1) BS-Based Method. After the sensitivity test (given in the Appendix), this case study set that $\delta = 0.9 \text{ km}^2$ and $\alpha = 60 \text{ m}$. These parameters can be changed based on the needs of the analysis. After the division, there were 364 BSAZs in which 28 BSAZs were modified to increase the average compactness ratio by 4.5%, and the result of BSAZs is shown in Figure 7. As δ was set to 0.9 km², the designed bus service area of these 364 BSAZs ranged from 0.49 km² to 1.34 km², while 348 BSAZs' designed bus service area sizes were within 0.8 km^2 -1.0 km², as shown in Figure 8(a). As shown in Figure 8(b), a few BSAZs with different-level sizes of designed bus service area were located at the subdistrict boundaries because the subdistrict boundaries could affect the efficiency of the BS-based method to uniform the designed bus service area in each BSAZ. However, most BSAZs had a similar size of designed bus service area and were thus marked in the same color, indicating that the BSbased method can ensure the high similarity of the generated BSAZs in terms of the size of designed bus service area. That is, the proposed BS-based method can meet the need of keeping the size of designed bus service area in each BSAZ in a similar level, thus simplifying the bus service analysis, as discussed in Section 3.



FIGURE 6: Analytical example for the analysis of bus stop density. (a) Study area. (b) The result by inappropriate zone division method. (c) The result by BS-based method.



FIGURE 7: The BSAZs generated for Shunde District.

(2) Grid-Based Method and Natural-Boundary-Based Method. For the grid-based method, the zone size of each grid was set to 2.25 km^2 ($1.5 \text{ km} \times 1.5 \text{ km}$), which was close to the mean zone size of BSAZs (2.230 km^2). The grid-based method divided Shunde District into 599 zones, as shown in Figure 9(a). The natural-boundary-based method divided Shunde District into 520 zones, as shown in Figure 9(b).

4.2.2. Comparison When Bus Stop Density Is Analyzed. The bus stop density in a zone is calculated as follows.

Bus stop density in a zone =
$$\frac{\text{Number of bus stop in a zone}}{\text{Area size of a zone}}$$
. (4)

As the basic logic of the BS-based method suggests, the divided zones have a similar size of the designed bus service area, but the size of each zone is different, that is, a BSAZ with a smaller zone size indicates that the area covered by the BSAZ has a relatively high density of bus stop. The BSAZs were ordered in the ascending order in terms of the zone size, and the zone size and the bus stop density are plotted in

Figure 10. When the zone size increased, the bus stop density had shown an obvious decreasing trend, which is consistent with the basic logic of the BS-based method. That is, the result of the proposed BS-based method can be used to reflect the geographical distribution of the bus stops. As shown in Figure 11, the BSAZs with a smaller zone size were marked in a lighter green. For example, the northwest part of the Daliang Subdistrict had the darker green BSAZs where the bus stop density was in a relatively high level, whereas the southeast part of the Jun'an Subdistrict had the lighter green BSAZs where the bus stop density was in a relatively low level. These zones with unusually higher or lower bus stop densities are identified and the bus operator should check the demand and supply in these zones to ensure the proper use of public bus resources. This illustrates how the proposed method can be used to identify zones with unnormal bus characteristics.

In this case study, the grid-based method and the natural-boundary-based method generated 214 and 102 zones without a bus stop, respectively, whereas all zones generated by the BS-based method had at least a bus stop. Therefore, the BS-based method had a lower discrete degree of the



FIGURE 8: Distribution of the size of designed bus service area. (a) Numerical distribution. (b) Spatial distribution.



FIGURE 9: Division results given by two zone division methods. (a) Grid-based method. (b) Natural-boundary-based method.

number of bus stops (calculated by equation (5)) and a lower discrete degree of the designed bus service area (calculated by equation (6)), compared with the other two methods, as shown in Table 1. Except zones with zero density of bus stop, it was found that there were zones with extremely high density of bus stop generated by the grid-based method and the natural-boundary-based method. In all, different zone division methods can result in different distributions of the bus stop density and then may influence the relative analysis.



FIGURE 10: The relation between zone size and bus stop density.



FIGURE 11: Zone size distribution of BSAZs.

TABLE 1	•	Indicators	of	discrete	degree	and	hus	ston	density	7
INDLL I	٠.	malcutors	O1	anserete	uegree	unu	Uu3	Stop	actionty	

Method	Discrete degree of the number of bus stops in a zone	Discrete degree of the designed bus service area in a zone	Average bus stop density in the study area (unit: stops/km ²)	Range of bus stop density in the study area (unit: stops/km ²)
BS-based	0.404	0.065	4.535	(0.342, 19.947)
Grid-based	1.297	1.031	2.439	(0, 62.015)
Natural-boundary-based	1.302	1.196	4.465	(0, 38.579)

Discrete degree of the number of bus stops in a zone =
$$\frac{\text{Standard deviation of the number of bus stops in a zone}}{\text{Average number of bus stops in a zone}}$$
, (5)
Discrete degree of the designed bus service area = $\frac{\text{Standard deviation of the designed bus service area in a zone}}{\text{Average designed bus service area in a zone}}$, (6)
Average bus stop density in the study area = $\frac{\sum \text{Bus stop density in a zone}}{\text{Number of zones in the study area}}$. (7)

In this case, two relevant analyses have been examined, including (a) the analysis of the relation between bus stop density and the bus flow (i.e., the average number of trips in a zone per day) and (b) the analysis of the relation between bus stop density and the bus popularity which represents the willingness of residents in different regions to choose public bus travel. The bus popularity is calculated by the following equation.

$$Bus popularity = \frac{Bus flow}{Population in a zone}.$$
 (8)

First, the relation between bus stop density and bus flow was investigated as shown in Figure 12.

As shown in Figure 12, not matter which zone division method was used, a similar pattern between bus flow and stop density could be found, i.e., the higher stop density generally had a larger bus flow. However, there were some obvious outliers when the grid-based method and the naturalboundary-based method were used. Generally, outliers are expected to indicate a problem, but these outliers in Figures 12(b) and 12(c) might not be related to a poor bus service. For example, outlier A in Figure 12(b) had a high density of bus stop but its bus flow was low, and Figure 12(d) shows the respective zone (Zone A). Zone A covered a very small area due to the subdistrict boundaries, and thus its density of bus stop was quite high. However, when the neighborhood area was examined, this area should not be regarded as an area with a high density of bus stop. In contrast, the BS-based method can divide the same area into a zone with a more reasonable size as shown in Figure 12(e) and then give a more proper value in terms of the density of bus stop. Similarly, the natural-boundary-based method suffers the same problem.

Second, the relation between bus stop density and bus popularity was examined as shown in Figures 13–16.

When Figures 13–15 are compared, more outliers are found in Figures 14 and 15. However, these outliers might be misleading and cause extra analysis efforts. Take the outlier B in Figure 15 as an example. The respective Zone B did not have bus stops as shown in Figure 16(a) and then its density of bus stop was zero, although there were bus stops very close to Zone B. In contrast, the BS-based method can achieve a better division result as shown in Figure 16(b).

In all, compared with the BS-based method, the gridbased method and natural-boundary-based method are less likely to be suitable for such bus service analysis which involves the bus stop density because they could generate misleading results and require more efforts to verify the analysis result. 4.2.3. Comparison When the Relation between POI and Bus Service Characteristic Is Analyzed. Except the bus stop density analysis, the analysis of the relation between POI and bus service characteristic is preformed to check the influence of the three zone division methods. In this comparison, the geographic weighted regression (GWR) [44] analysis was adopted to estimate the effects of land use on the bus system, i.e., the spatial relationship between POI and bus flow/bus popularity. The GWR model is an extension of the general linear regression model which considers the local effects of spatial objects and its functional form can be expressed as follows.

$$y_j = \beta_0(u_j, v_j) + \sum_{i=1}^p \beta_i(u_j, v_j) x_{ij} + \varepsilon_j, \qquad (9)$$

where y_j is the dependent variable at zone j, x_{ij} (i = 1, 2, ..., p) is the independent variable at zone j, (u_j, v_j) represents the coordinate of zone centroid, β_0 and β_i ($i \in [1, p]$) are calculated according to the distance between the zone i and nearby zones, and ε_j is the random error of the independent distribution.

As for the input of the GWR model, there were three sets of bus flows which were calculated by the BS-based method, the grid-based method, and the natural-boundary-based method, respectively. The regression result of the naturalboundary-based method was not presented in this paper because its R^2 was extremely low.

After carefully checking POI-related factors, four POIrelated factors were selected based on the test performance, including the number of scenic spots in a zone, the number of factories in a zone, the number of automobile maintenance shops in a zone, and the number of shops in a zone. They represent cultural, economic, infrastructural, and social dimensions, respectively, which can reflect the impact of different aspects on spatial public bus service and the city's level of development. There are two sets of zone division results generated by the BS-based method and the grid-based method, and for each zone division result, the spatial distributions of four factors were calculated and the collinearity of the four factors was checked. The correlation coefficients among four factors were below 0.7 no matter which zone division method was used, indicating that these four factors could be statistically independent. To further check the collinearity, the variance inflation factor (VIF), an indicator of the severity of multicollinearity, was adopted to test the independency. The VIFs have been computed by the method suggested by Belsley et al. [45] which uses the diagonal of the inverse of the correlation coefficient matrix of the exogenous



FIGURE 12: The relation between bus flow and stop density using three zone division methods. (a) By BS-based method. (b) By grid-based method. (c) By natural-boundary-based method. (d) Grid-based method's result for an extreme case (outlier A). (e) BS-based method's result for comparison.

24



21 18 Bus popularity 15 12 9 6 3 0 0 8 16 24 32 40 48 56 64 Stop density (stops/km2) • Bus popularity

FIGURE 13: The relation between bus popularity and stop density when the BS-based method is used.

variables. If variables with VIF values are greater than 10, they are considered to be multicollinear variables [46].

The calculation result is shown in Table 2, and all VIF values were less than 10, which illustrated that four factors were independent no matter which zone division result was used. Hence, the four POI-related factors were included in the GWR model.

The calibration result is presented in Table 3. R^2 was 0.789 when the BS-based method was used, which was more than that for grid-based method (0.746). Moreover, the AIC

FIGURE 14: The relation between bus popularity and stop density when the grid-based method is used.

and residual sum of squares were significantly better when the BS-based method was used. Hence, adopting the BSbased method could help to improve the calibration results of the GWR model for the bus service analysis.

As the results for the BS-based method in Table 3 show, the mean coefficient for the number of shops presents a positive correlation with bus flow, which is consistent with the general phenomenon that high-consumption areas attract people and induce the public bus trips. The mean coefficient for the number of scenic spots is also positive,



• Bus popularity

FIGURE 15: The relation between bus popularity and stop density when the natural-boundary-based method is used.



FIGURE 16: Zone division results of the natural-boundary-based method and the BS-based method for a special case. (a) Natural-boundarybased method's result for an extreme case (outlier B). (b) BS-based method's result for comparison.

which means that the increasing number of scenic spots results in the increase of tourists who travel by the public bus. However, the mean coefficient for the number of factories demonstrates a negative correlation. The possible reason is that a factory generally provides shuttle bus service for its employees, and employees are more likely to live in or near the factory. It shows that the construction of public transport facilities needs to take the cluster of surrounding factory areas into account.

Additionally, Figure 17 displays the hierarchy of the coefficient between the number of shops and the bus flow calculated by the BS-based method and the grid-based method. No matter which method was used, the coefficient spatial distribution presented a similar pattern in which the values in the west part were generally smaller than those in the east part in terms of the relation between the number of shops and the bus flow. This pattern

illustrated that the number of shops had a larger impact on the bus flow in the east part than that in the west part. The east subdistricts (Daliang and Ronggui Subdistricts) are adjacent to Guangzhou (one of China's three largest cities) (Guangzhou: https://en.wikipedia.org/wiki/Guangzhou (Accessed on September 26, 2022)), and thus the east subdistricts have many active commercial activities which induce trips. But the BS-based method can help to achieve a better fitness, so that the respective regional coefficients have higher credibility when analyzing the influencing factors of bus demand.

4.3. Results of the BS-Based Method. To further illustrate the performance of the BS-based method, this section conducts more spatial analysis of the bus service based on the zone division result of the BS-based method.

Variables	VIFs calculated based on the zone division result of the BS-based method	VIFs calculated based on the zone division result of the grid-based method
Number of shops	1.34	1.44
Number of scenic spots	1.24	1.24
Number of factories	1.11	1.13
Number of automobile maintenance shops	1.16	1.17

TABLE 2: Variance inflation factor (VIF) of explanatory variables.

Variables	Min	Max	Mean	p value
Bus flow calculated by the BS-based method				
Intercept	-0.383	0.432	-0.018	
Number of shops	0.136	0.945	0.517	≤0.001
Number of scenic spots	-0.094	0.695	0.137	≤0.001
Number of factories	-0.753	0.052	-0.211	≤0.001
Number of automobile maintenance shops	-0.271	0.415	0.145	≤0.001
AIC	587.662			
R^2	0.789			
Residual sum of squares	76.882			
Bus flow calculated by the grid-based method				
Intercept	-0.209	0.430	0.013	
Number of shops	0.065	1.050	0.611	≤0.001
Number of scenic spots	-0.052	0.377	0.105	≤0.001
Number of factories	-0.421	0.106	-0.096	≤0.001
Number of automobile maintenance shops	-0.377	0.874	0.183	≤0.001
AIC	942.358			
R^2	0.746			
Residual sum of squares	139.793			





FIGURE 17: Spatial distribution of shopping service coefficient. (a) By BS-based method. (b) By grid-based method.

4.3.1. Moran's I Test for Spatial Autocorrelation. This study investigated the spatial autocorrelation among BSAZs in terms of (a) the bus flow and (b) the bus popularity. Global Moran's I and Local Moran's I [47]

were adopted to explore these spatial autocorrelation, and Global Moran's I (I) and Local Moran's I (I_i) for BSAZ *i* can be calculated by equations (10) and (11), respectively.

$$I = \frac{n}{\sum_{i}\sum_{j}w_{ij}} \frac{\sum_{i}\sum_{j}w_{ij}(y_{i}-\overline{y})(y_{j}-\overline{y})}{\sum_{i}(y_{i}-\overline{y})^{2}},$$
(10)

$$I_{i} = \left(\frac{y_{i} - \overline{y}}{\sum_{i} (y_{i} - \overline{y})^{2}}\right) \sum_{j} w_{ij} (y_{j} - \overline{y}), \qquad (11)$$

where *n* is the total number of BSAZs, w_{ij} is the spatial weight between the centroids of two BSAZs *i* and *j*, y_i is the one of the characteristics (bus flow or bus popularity) of BSAZ *i*, and \overline{y} is the mean value of the respective characteristic. Moran's I's statistical significance is based on the *p* value. Moran's I value ranges from -1 (perfect dispersion) to +1 (perfect clustering).

The results of Global Moran's I for bus flow and bus popularity are shown in Table 4, showing that both characteristics indicated a statistically significantly positive spatial correlation (p values for both characteristics were significant at 0.01 level). That is, the bus flow and bus popularity of a zone were similar to those of its nearby zones. Hence, there was a widespread existence of spatial agglomeration in terms of residents' public bus travel pattern in Shunde District.

As shown in Figures 18(a) and 18(b) ((a) and (b) can take the value "High" or "Low" to represent that the characteristic value is high or low, respectively), clusters represent the units with A surrounded by units with B. For example, Low-High clusters represent the units with low values surrounded by units with high values.

For both characteristics, the distributions of high-high clusters were mostly in the southeast and the northwest (i.e., the central and western regions in Daliang Subdistrict and the northeastern regions in Lecong Subdistrict). It indicated that these regions had a high homogeneity in terms of bus travel pattern in which the region enjoyed the high volume of bus passengers and the high willingness of residents to travel by public bus. Daliang Subdistrict has a high GDP (gross domestic product) of around 546,000,000 RMB per year and the largest population of 265,500 people in Shunde District (Shunde District Government Affairs Data Service Network: https://data.shunde.gov.cn/ (Accessed on September 26, 2022)). Business centers are mainly concentrated in the central and western regions in Daliang Subdistrict. The northwestern regions of Lecong Subdistrict are considered the important area connecting Shunde District and Chancheng District which is the economic center of Foshan (Chancheng District: https://en.wikipedia.org/wiki/ Chancheng District (Accessed on September 26, 2022)). Moreover, most clusters of the southwest of Shunde District were Low-Low in terms of bus flow as shown in Figure 18(a), which indicated that the daily bus flow in the southwest of Shunde District seemed to be generally low. However, such a spatial correlation did not seem to exist when the bus popularity (i.e., the number of daily bus trips per resident) was examined, since the Local Moran's I for the bus popularity was not significant as shown in Figure 18(b). It illustrated that the individual preference to bus trips might randomly vary in the southwest of Shunde District, although there was a homogeneous pattern in the aggregated level. It

TABLE 4: Global Moran's I of bus flow and bus popularity.

Characteristic	Global Moran's I	<i>p</i> value
Bus flow	0.470	≤0.001
Bus popularity	0.236	≤0.001

means that a more detailed analysis for these zones is needed and it might not be suitable to treat these zones in a similar manner.

4.3.2. Spatial Relation between the Number of Bus Routes and Bus Popularity. The spatial relation between the number of bus routes and bus popularity was investigated. As shown in Figure 19(a), bus popularity tended to increase with the increase of the number of bus routes. However, the outliers existed and were detected by K Nearest Neighbors (KNN) algorithm [48]. For each zone, the probability of being an outlier was calculated, and the BSAZs with the first 5% highest probabilities were marked as outliers and plotted as red points in Figure 19(a).

These outliers' locations are presented in Figure 19(b) and these outliers should attract the operators' attention. For example, there were no outliers in Lunjiao Subdistrict, indicating that for the bus service offered in Lunjiao Subdistrict, its influence on bus popularity was similar to those in most parts in Shunde District, whereas the outliers in Daliang Subdistrict illustrated that the bus service's influence on bus popularity in the respective BSAZs had a relatively different pattern, and these BSAZs might need different regional bus service management strategies. Three examples are given in this case.

First, the BSAZ 346 had a high bus popularity but with relatively less bus routes. Checking the POI in BSAZ 346, we found that BSAZ 346 had a wide industrial area in which commuting needs would be high. It might explain that the bus popularity was high. However, for such a level of bus popularity, the number of bus routes in BSAZ 346 was much lower than that in the general case of Shunde District, since BSAZ 346 was identified as an outlier. For this case, the bus operators should carefully examine the bus service quality in BSAZ 346 and consider whether more bus routes should be added because people in BSAZ 346 seem to be more willing to use public bus for their daily trips.

Second, the situation of another outlier, BSAZ 162, was different, i.e., BSAZ 162 had a low bus popularity but with relatively more bus routes. BSAZ 162 has a high-speed railway station, so the bus operators set 19 bus routes. However, the high-speed railway passengers seemed to use other transport modes rather than bus as indicated by the low bus popularity. Therefore, the bus operators should consider the optimization of bus setting in BSAZ 162.

Third, the outliers BSAZs 115, 127, and 213 had a low bus popularity but with relatively more bus routes. These BSAZs were industrial areas where some factories operated their own shuttle buses to serve their employees and some offered the accommodation. Hence, the need of bus was affected (as



FIGURE 18: Local Moran's I for bus flow and bus popularity at the BSAZ level. (a) Local Moran's I for bus flow. (b) Local Moran's I for bus popularity.



FIGURE 19: Outlier distribution of BSAZs. (a) Bus popularity and the number of bus routes in BSAZs. (b) Outlier distribution in map.

α (m)	Size of average designed bus service area (km ²)	Deviation (km ²)	Average compactness	Computation time (min)
30	0.906	0.006	0.474	164
60	0.906	0.006	0.479	29
90	0.908	0.008	0.474	11
120	0.906	0.006	0.471	6
150	0.913	0.013	0.467	4

TABLE 5: The sensitivity analysis for α .

TABLE 6: The sensitivity analysis for δ .

δ (km ²)	Deviation (km ²)	Average zone size (km ²)	Average number of bus stops
0.80	0.002	1.976	6.131
0.85	0.007	2.109	6.545
0.90	0.006	2.230	6.923
0.95	0.006	2.353	7.304
1.00	0.008	2.483	7.706

shown in Table 3, the number of factories has a negative impact on the bus flow), and the bus operators might need to reorganize the bus services in BSAZs 115, 127, and 213.

5. Conclusions

This study has suggested a new feature-based zone division method called BS-based method. The division result was analyzed, and the analysis result showed that the BS-based method could keep the size of designed bus service area in each zone in a similar level and support different spatial analyses of bus service.

Furthermore, in this study, the influence of the different zone division strategies has been checked in the analytical example and the case study which used the real-world data to conduct a series of analyses that investigated bus service characteristics (e.g., bus stop density, bus flow, and bus popularity) and the relation between land use and bus service pattern. The results have illustrated that a proper zone division strategy can give a good support to transport planning and management. That is, with the consideration of designed bus service area, the feature-based method (e.g., the BS-based method) can offer a better zone division result to increase the efficiency of the bus-stop-related analyses, compared with the grid-based method and the naturalboundary-based method.

In the future study, the applications of the BS-based method will be explored to assist in the improvement of bus service quality. Moreover, the BS-based method will be further developed. Considering more factors could enhance our proposed division method, such as the varied actual bus service area which is affected by the different number of routes through a bus stop [49] and the time headway of bus routes through a bus stop, and the Bayesian Model Averaging techniques can be adopted to analyze the bus characteristics [50]. The future study will consider this extension, so that the divided zones can support different spatial analyses of bus service and help to deliver a more realistic result.

Appendix

A. Sensitivity Analysis of the BS-Based Method

The BS-based method has two parameters, the standard size of designed bus service area in a BSAZ (δ) and the side length of a basic unit (α). This appendix performs a sensitivity analysis for δ and α to examine the BS-based method using a personal computer (Intel(R) Xeon(R) Gold 6230 CPU @ 2.10 GHz). Xingtan Subdistrict in Shunde District was selected for the sensitivity analysis.

First, α varied from 30 m to 150 m with a step of 30 m while $\delta = 0.9 \text{ km}^2$. The value of δ is firstly set to be approximately triple the size of designed bus service area of a stop (0.3 km²), and the sensitivity test for δ will be performed later. The value of δ can be set based on the need of public bus operators. The result is shown in Table 5. The second column gives the size of average designed bus service area of BSAZs. The third column presents the deviation of designed bus service area in each BSAZ. The fourth column gives the average compactness of BSAZs. The fifth column presents the computation time.

The result showed that the average size of designed bus service area and the average compactness changed a little when α varied from 30 m to 150 m, but a general trend was that with the increase of α , the deviation of designed bus service area (the difference between the average size of designed bus service area and δ) increased while the average compactness decreased. A more accurate spatial analysis requires that BSAZs should have a similar designed bus service area and the shape of the BSAZs should be more compact. However, a smaller α requires a longer computation time, and when α increases, the decrease rate of computation time changes dramatically. Therefore, both the efficiency and the accuracy should be considered when selecting a proper α . When $\alpha = 60$ m, the computation time was short, the average size of designed bus service area was close to the standard size, and the average compactness was the highest, and thus this study set α to be 60 m.

Second, δ varied from 0.80 km² to 1.00 km² with a step of 0.05 km^2 while $\alpha = 60 \text{ m}$. The result is shown in Table 6. The second column gives the absolute difference between the average designed bus service area of BSAZs and δ , and the third and fourth columns show the average zone size and the average number of bus stops, respectively. Because a bus site may have two bus stops for the upstream and downstream buses (according to the bus network data in Foshan, a bus site averagely includes 1.4 bus stops) and these two bus stops are generally close to each other, the average number of bus stops ranged from 6.1 to 7.7 when δ varied from 0.80 km² to 1.00 km². With the increase of δ , the average zone size and the average number of bus stops increased. Moreover, when δ increased, the deviation tended to remain a small value, indicating that the BS-based method has a relatively stable performance when it varies from 0.80 km² to 1.00 km². Hence, the medium value (0.90 km^2) is used.

Data Availability

The data that support the findings of this study are provided by Foshan Transport Institute, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Fanghao Fu was responsible for methodology, validation, visualization, and original draft preparation. Jiemin Xie was responsible for conceptualization, methodology, validation, review and editing, funding acquisition, and supervision. Ming Cai was responsible for conceptualization, methodology, project administration, and review and editing. Wei Huang was responsible for funding acquisition, supervision, and review and editing. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

This research was supported by the Shenzhen Science and Technology Program (grant nos. 202206193000001 and 20220817201129001), Guangdong Basic and Applied Basic Research Foundation (grant nos. 2022A1515110235 and 2023A1515012542), and Science and Technology Planning Project of Guangdong Province (grant no. 2023B1212060029).

References

- T. Liu, A. A. Ceder, R. Bologna, and B. Cabantous, "Commuting by customized bus: a comparative analysis with private car and conventional public transport in two cities," *Journal of Public Transportation*, vol. 19, no. 2, pp. 55–74, 2016.
- [2] M. G. Badami and M. Haider, "An analysis of public bus transit performance in Indian cities," *Transportation Research Part A: Policy and Practice*, vol. 41, no. 10, pp. 961–981, 2007.

- [3] N. Guschinsky, M. Y. Kovalyov, E. Pesch, and B. Rozin, "Cost minimizing decisions on equipment and charging schedule for electric buses in a single depot," *Transportation Research Part E: Logistics and Transportation Review*, vol. 180, 2023.
- [4] Y. H. Kuo, J. M. Leung, and Y. Yan, "Public transport for smart cities: recent innovations and future challenges," *European Journal of Operational Research*, vol. 306, no. 3, pp. 1001–1026, 2023.
- [5] A. Nikolaidou, A. Kopsacheilis, G. Georgiadis, T. Noutsias, I. Politis, and I. Fyrogenis, "Factors affecting public transport performance due to the COVID-19 outbreak: a worldwide analysis," *Cities*, vol. 134, 2023.
- [6] S. Openshaw, "Optimal zoning systems for spatial interaction models," *Environment & Planning A: Economy and Space*, vol. 9, no. 2, pp. 169–184, 1977.
- [7] Á. Ibeas, L. dell'Olio, B. Alonso, and O. Sainz, "Optimizing bus stop spacing in urban areas," *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 3, pp. 446–458, 2010.
- [8] C. Saifei, F. Hui, W. Naiqi, W. Yefei, and Q. Yan, "Passengeroriented traffic management integrating perimeter control and regional bus service frequency setting using 3D-pMFD," *Transportation Research Part C: Emerging Technologies*, vol. 135, 2022.
- [9] E. Lee, X. Cen, and H. K. Lo, "Zonal-based flexible bus service under elastic stochastic demand," *Transportation Research Part E: Logistics and Transportation Review*, vol. 152, 2021.
- [10] M. Cai, L. Hong, and C. Xiong, "Data-driven traffic zone division in smart city: framework and technology," Sustainable Energy Technologies and Assessments, vol. 52, 2022.
- [11] Y. E. Hawas, M. Khan, and N. Basu, "Evaluating and enhancing the operational performance of public bus systems using GIS-based data envelopment analysis," *Journal of Public Transportation*, vol. 15, no. 2, pp. 19–44, 2012.
- [12] J. Esmailpour, K. Aghabayk, M. Abrari Vajari, and C. De Gruyter, "Importance–Performance Analysis (IPA) of bus service attributes: a case study in a developing country," *Transportation Research Part A: Policy and Practice*, vol. 142, pp. 129–150, 2020.
- [13] J. Paudel, "Bus ridership and service reliability: the case of public transportation in Western Massachusetts," *Transport Policy*, vol. 100, pp. 98–107, 2021.
- [14] D. Sheng, Q. Meng, and Z. C. Li, "Optimal quality incentive scheme design in contracting out public bus services," *Transportation Research Part C: Emerging Technologies*, vol. 133, 2021.
- [15] S. Deb, M. Ali Ahmed, and D. Das, "Service quality estimation and improvement plan of bus Service: a perception and expectation based analysis," *Case Studies on Transport Policy*, vol. 10, no. 3, pp. 1775–1789, 2022.
- [16] Y. Chen, B. Yu, B. Shu, L. Yang, and R. Wang, "Exploring the spatiotemporal patterns and correlates of urban vitality: temporal and spatial heterogeneity," *Sustainable Cities and Society*, vol. 91, 2023.
- [17] F. Basso, J. Frez, L. Martínez, R. Pezoa, and M. Varas, "Accessibility to opportunities based on public transport gpsmonitored data: the case of Santiago, Chile," *Travel Behaviour and Society*, vol. 21, pp. 140–153, 2020.
- [18] F. Basso, J. Frez, H. Hernández, V. Leiva, R. Pezoa, and M. Varas, "Crowding on public transport using smart card data during the COVID-19 pandemic: new methodology and case study in Chile," *Sustainable Cities and Society*, vol. 96, 2023.
- [19] Y. Chen, A. Bouferguene, Y. Shen, and M. Al-Hussein, "Assessing accessibility-based service effectiveness (ABSEV) and social equity for urban bus transit: a sustainability perspective," *Sustainable Cities and Society*, vol. 44, pp. 499–510, 2019.

- [20] L. Yang, K. W. Chau, W. Y. Szeto, X. Cui, and X. Wang, "Accessibility to transit, by transit, and property prices: spatially varying relationships," *Transportation Research Part D: Transport and Environment*, vol. 85, 2020.
- [21] J. Cai, B. Huang, and Y. Song, "Using multi-source geospatial big data to identify the structure of polycentric cities," *Remote Sensing of Environment*, vol. 202, pp. 210–221, 2017.
- [22] F. Shi, "Research on accessibility and equity of urban transport based on multisource big data," *Journal of Advanced Transportation*, vol. 2021, Article ID 1103331, 18 pages, 2021.
- [23] L. Sun and K. W. Axhausen, "Understanding urban mobility patterns with a probabilistic tensor factorization framework," *Transportation Research Part B: Methodological*, vol. 91, pp. 511–524, 2016.
- [24] S. Mortazavi and M. Akbarzadeh, "A framework for measuring the spatial equity in the distribution of public transportation benefits," *Journal of Public Transportation*, vol. 20, no. 1, pp. 44–62, 2017.
- [25] Q. L. Gao, Q. Q. Li, Y. Yue, Y. Zhuang, Z. P. Chen, and H. Kong, "Exploring changes in the spatial distribution of the low-to-moderate income group using transit smart card data," *Computers, Environment and Urban Systems*, vol. 72, pp. 68–77, 2018.
- [26] Q. Liu, C. Ding, and P. Chen, "A panel analysis of the effect of the urban environment on the spatiotemporal pattern of taxi demand," *Travel Behaviour and Society*, vol. 18, pp. 29–36, 2020.
- [27] R. An, Z. Wu, Z. Tong, S. Qin, Y. Zhu, and Y. Liu, "How the built environment promotes public transportation in Wuhan: a multiscale geographically weighted regression analysis," *Travel Behaviour and Society*, vol. 29, pp. 186–199, 2022.
- [28] R. M. Assunção, M. C. Neves, G. Câmara, and C. da Costa Freitas, "Efficient regionalization techniques for socioeconomic geographical units using minimum spanning trees," *International Journal of Geographical Information Science*, vol. 20, no. 7, pp. 797–811, 2006.
- [29] S. Wu, Y. Pei, and G. Cheng, "Method of traffic zone division based on spectral graph theory," *Computer Modelling and New Technologies*, vol. 18, no. 2, pp. 186–191, 2014.
- [30] H. Dong, M. Wu, X. Ding et al., "Traffic zone division based on big data from mobile phone base stations," *Transportation Research Part C: Emerging Technologies*, vol. 58, pp. 278–291, 2015.
- [31] C. Siddiqui, M. Abdel-Aty, and H. Huang, "Aggregate nonparametric safety analysis of traffic zones," *Accident Analysis* & *Prevention*, vol. 45, pp. 317–325, 2012.
- [32] P. Xu, H. Huang, N. Dong, and M. Abdel-Aty, "Sensitivity analysis in the context of regional safety modeling: identifying and assessing the modifiable areal unit problem," *Accident Analysis & Prevention*, vol. 70, pp. 110–120, 2014.
- [33] Y. Zheng, G. Zhao, and J. Liu, "A novel grid based k-means cluster method for traffic zone division," in *Second International Conference on Cloud Computing and Big Data in Asia*, pp. 165–178, Springer, Cham, 2015.
- [34] D. Xia, B. Wang, Y. Li, Z. Rong, and Z. Zhang, "An efficient MapReduce-based parallel clustering algorithm for distributed traffic subarea division," *Discrete Dynamics in Nature and Society*, vol. 2015, Article ID 793010, 18 pages, 2015.
- [35] L. M. Martínez, J. M. Viegas, and E. A. Silva, "A traffic analysis zone definition: a new methodology and algorithm," *Transportation*, vol. 36, no. 5, pp. 581–599, 2009.
- [36] X. Xing, W. Huang, G. Song, and K. Xie, "Traffic zone division using mobile billing data," in *Proceedings of the 2014 11th International Conference on Fuzzy Systems and Knowledge*

Discovery (FSKD), pp. 692–697, IEEE, Xiamen, China, August 2014.

- [37] P. H. Soares, "Zone-based public transport route optimisation in an urban network," *Public Transport*, vol. 13, no. 1, pp. 197–231, 2021.
- [38] E. Lee, X. Cen, and H. K. Lo, "Scheduling zonal-based flexible bus service under dynamic stochastic demand and Time-dependent travel time," *Transportation Research Part E: Logistics and Transportation Review*, vol. 168, 2022.
- [39] Ministry of Construction of the People's Republic of China, Code for Transport Planning on Urban Road, Ministry of Construction of the People's Republic of China, Beijing, China, 1995.
- [40] X. Liu, L. Gong, Y. Gong, and Y. Liu, "Revealing travel patterns and city structure with taxi trip data," *Journal of Transport Geography*, vol. 43, pp. 78–90, 2015.
- [41] Z. Cai, D. Wang, and X. M. Chen, "A novel trip coverage index for transit accessibility assessment using mobile phone data," *Journal of Advanced Transportation*, vol. 2017, Article ID 9754508, 14 pages, 2017.
- [42] K. Kim, "Identifying the structure of cities by clustering using a new similarity measure based on smart card data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 2002–2011, 2020.
- [43] R. S. Montero and E. Bribiesca, "State of the art of compactness and circularity measures," *International Mathematical Forum*, vol. 4, no. 27, pp. 1305–1335, 2009.
- [44] C. Brunsdon, A. S. Fotheringham, and M. E. Charlton, "Geographically weighted regression: a method for exploring spatial nonstationarity," *Geographical Analysis*, vol. 28, no. 4, pp. 281–298, 1996.
- [45] D. A. Belsley, E. Kuh, and R. E. Welsch, Regression Diagnostics, Wiley, New York, NY, USA, 1980.
- [46] X. Qian and S. V. Ukkusuri, "Spatial variation of the urban taxi ridership using GPS data," *Applied Geography*, vol. 59, pp. 31–42, 2015.
- [47] G. Sharma and G. R. Patil, "Public transit accessibility approach to understand the equity for public healthcare services: a case study of Greater Mumbai," *Journal of Transport Geography*, vol. 94, 2021.
- [48] K. Fukunaga and P. M. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Transactions on Computers*, vol. 7, pp. 750–753, 1975.
- [49] A. M. El-Geneidy, P. Tetreault, and J. Surprenant-Legault, Pedestrian Access to Transit: Identifying Redundancies and Gaps Using a Variable Service Area Analysis, National Academics, Washington, DC, USA, 2010.
- [50] S. Wu, Y. Zou, L. Wu, and Y. Zhang, "Application of Bayesian model averaging for modeling time headway distribution," *Physica A: Statistical Mechanics and its Applications*, vol. 620, 2023.