



## Using fuzzy clustering of user perception to determine the number of level-of-service categories for bus rapid transit

Yueying Huo<sup>a,\*</sup>, Jinhua Zhao<sup>b</sup>, Xiaojuan Li<sup>a</sup>, Chen Guo<sup>a</sup>

<sup>a</sup> Inner Mongolia University, China

<sup>b</sup> Massachusetts Institute of Technology, USA



### ARTICLE INFO

#### Keywords:

Level of service  
User perception  
Level-of-service categories  
Bus rapid transit  
Fuzzy clustering  
Cluster validity indices

### ABSTRACT

The concept of level of service (LOS) is meant to reflect user perception of the quality of service provided by a transportation facility or service. Although the LOS of bus rapid transit (BRT) has received considerable attention, the number of levels of service of BRT that a user can perceive still remains unclear. Therefore, this paper addresses this issue using fuzzy clustering of user perception. User perception is a six-dimension vector including perceived arrival time, perceived waiting time, bus speed perception, passenger load perception, perceived departure time, and overall perception. The research team developed a smartphone-based transit travel survey system to conduct the user perception surveys in three BRT systems in China. Fuzzy C-means clustering, improved using a simulated annealing genetic algorithm, was adopted to partition user perception into 2–10 clusters. Seven cluster validity indices were used to determine the appropriate number of LOS categories. The results indicate that users can perceive two to four levels of service.

### Introduction

Level of service (LOS) is used to describe the traffic operating conditions of transportation facilities from intersections to highways in the *Highway Capacity Manual* (HCM) (Roess et al., 2010; TRB 2010). This important concept affects the planning, design, and operational aspects of transportation facilities as well as the allocation of limited financial resources among competing transportation projects (Choocharukul et al., 2004; Washburn and Kirschner, 2006).

The HCM states, “Quality of service describes how well a transportation facility or service operates from a traveler’s perspective,” and defines LOS as the quantitative stratification of a performance measure or measures that represent quality of service (TRB 2010). The definition of LOS indicates that it must reflect user perception of the quality of service provided by a transportation facility or service, which is also common in past research (Das and Pandit, 2013, 2015; Fang et al. 2003; Fang and Pecheux, 2009; Pecheux et al., 2000; Zhang and Prevedouros, 2011). Although the concept of LOS is meant to reflect user perception, its implementations—including determination of the number of LOS categories, selection of service measure(s), and identification of thresholds that separate one level from another—are based on the judgments of transportation professionals (Choocharukul et al., 2004; TRB 2010; Washburn and Kirschner, 2006).

Specifically, for the number of LOS categories, six levels of service are defined using the letters A through F and based on the judgments of transportation professionals in the HCM (TRB 2010). In addition, four, five, and nine or more levels of service have been proposed (Cameron, 1996; Das and Pandit, 2015; Huo et al. 2015; Maitra et al., 1999). Determining the number of LOS grades is part of the studies defining LOS criteria, which mostly assign the six levels conforming to the HCM and discuss deciding threshold values (Bhuyan and Mohapatra, 2014; Bhuyan and Rao, 2010, 2011). Other studies address the number of LOS categories at signalized intersections (Fang et al. 2003; Fang and Pecheux, 2009; Pecheux et al., 2000). These studies aim to test the appropriateness of the six levels of service in the HCM in terms of user perception rather than finding the optimal number of LOS categories. The LOS number is predetermined to be six and to date no special study has explored the issue of the number of LOS categories.

Bus rapid transit (BRT) combines the operational flexibility and low cost of conventional bus transit with the high efficiency and reliability of rail transit (Cervero and Kang, 2011). Currently, 168 cities worldwide have constructed BRT corridors, totaling 4998 kilometers (BRT Center of Excellence 2019). Although the topic of BRT level of service has received increasing attention, the number of BRT levels of service that users can perceive remains unclear. However, this is an important question because it helps to understand how passengers perceive LOS

\* Corresponding author.

E-mail address: [hyy@imu.edu.cn](mailto:hyy@imu.edu.cn) (Y Huo).

and guides the use of LOS to reflect user perception. In addition, determining the number of LOS users perceive will provide insights into developing strategies that meet passenger demand and effectively improve bus service.

This study therefore deals with the number of BRT levels of service that users can perceive (i.e., what is the appropriate number of BRT LOS categories from the users' perspective). Our objective was to provide a systematic methodology for defining the number of LOS categories for a transportation facility or service, propose recommendations regarding the appropriate number of BRT LOS categories, and spur further research by highlighting this concept.

The remainder of this paper is organized as follows. After a literature review, user perception is defined. Subsequently, data collection—including the development of the transit travel survey system, BRT user perception survey, and sample characteristics—is described. The methodology involving fuzzy C-means clustering is then explained, and this algorithm is improved by a simulated annealing genetic algorithm and cluster validity indices. Finally, the values of the cluster validity indices are analyzed, and an appropriate number of BRT LOS categories is recommended.

## Literature review

Based on the judgments of transportation professionals, six levels of service are defined in the HCM using the letters *A* through *F*. The HCM also provides LOS criteria with six levels for freeways, multilane highways, two-way highways, urban streets, and intersections (TRB 2010). Through collective professional judgments, the *Transit Capacity and Quality of Service Manual* (TCQSM) presents LOS criteria for service measures of availability, comfort, and convenience. The number of levels within the LOS criteria is not fixed to six in the TCQSM, allowing more or fewer levels depending on the service measure characteristics (TRB 2013).

Instead of relying on professional judgment, Das and Pandit (2015) and Huo et al. (2015) placed transit LOS on a solid academic research footing. Das and Pandit (2015) employed the law of successive interval scaling to define LOS criteria for bus transit service measures in India. LOS criteria for delay in total journey time, distance to bus stop, waiting time, on-time performance, and service hours were established. Respondents were shown to be comfortable rating their satisfaction on a five-point scale; thus, five levels of service from LOS A to LOS E were defined and there was no additional discussion on the number of categories. Huo et al. (2015) used fuzzy C-means clustering of delay to develop LOS criteria for BRT in China. By applying a delay estimation model and conducting field surveys, a delay dataset was formed, and then delay threshold values between adjacent grades were proposed. This paper predesignates the number of LOS grades to four following a tradition of LOS classification of transportation facilities in China and does not present any discussion on it.

The Bhuyan group in India closely examined how to define urban street LOS criteria considering the country's highly heterogeneous traffic flow (Bhuyan and Mohapatra, 2014; Bhuyan and Rao, 2010, 2011). They selected the average travel speed as a service measure. Different clustering methods such as fuzzy C-mean, affinity propagation clustering, and self-organizing map were used to recognize the actual number of urban street classes and ascertain the speed ranges of six LOS categories. The problem of the number of LOS categories was not discussed in their studies and was specified as six following the HCM.

Choocharukul et al. (2004) examined user perception of a freeway LOS by conducting a multivariate statistical analysis of speed, density, vehicle headways, and the socioeconomic characteristics of participants. Freeway LOS criteria for perceived density were proposed. The research focused on comparing the density threshold values in the proposed and current HCM LOS criteria. The number of LOS categories was assigned to six to comply with the HCM and was not discussed. A

similar study specific to the rural freeway was conducted by Washburn and Kirschner (2006).

Arguing that the concept of LOS, although meant to reflect operational conditions as perceived by users is not explicitly based on studies of user perception in the HCM, Pecheux et al. (2000), Fang et al. (2003), and Fang and Pecheux (2009) conducted a series of studies to address "how users perceive quality of service and how many levels of service they are able to perceive" at signalized intersections. A total of 24 intersection approaches located in Pennsylvania were chosen to videotape the operations of actual intersections. In the laboratory experiment, 100 subjects were recruited to participate in experiments by watching the intersection approach videos, and they were required to estimate the time and assess the quality of service provided by the traffic signals. A database containing two types of data with user-estimated delays and quality-of-service ratings was developed and used as the basis on which the studies were performed.

Pecheux et al. (2000) performed rigid clustering of quality of service ratings to group intersection approaches. Different numbers of clusters ranging from six to three were tested. This study demonstrates that drivers do not perceive LOS as precisely as the existing HCM six levels, suggesting only three levels are perceived. Fang et al. (2003) stated that using the rigid LOS boundaries limited the subjective variations and uncertainties in user perception; thus, fuzzy C-means clustering was employed to partition delays or ratings into six to three clusters. They observed that six levels of service can be distinguished in the fuzzy domain for users' estimated delays; the current HCM LOS ranges for signalized intersections were acceptable after considering user perception.

Fang et al. (2003) did not examine whether the separate rating membership matched the delay membership and did not check whether users properly perceived each approach's LOS. Thus, Fang and Pecheux (2009) used fuzzy clustering again to revisit user perception of LOS at signalized intersections. After preprocessing the database, the delays and ratings were classified into seven clusters and mined to discover distinct clusters. The results were presented by analyzing delay membership, approach membership, and rating membership. The study concluded that users were able to perceive six levels of service but not the six defined in the HCM. A new set of six levels of service was proposed by merging HCM LOS A and LOS B and splitting HCM LOS F into two; fuzzy membership is effective at reflecting subjective perception. These studies used the HCM LOS criteria as guidelines for experimental design and examined the appropriateness of the current HCM LOS in terms of user perception. Their objective was not to determine the appropriate number of LOS categories, and they therefore did not discuss the issue of defining the LOS number.

Viewing that road users always consider multiple factors in LOS evaluation, Zhang and Prevedouros (2011) chose the factors of left-turn lane treatment, pavement markings, and delay to ascertain the signalized intersection LOS. To reflect the fuzziness of user perception, fuzzy logic was applied and rigid LOS thresholds were replaced with fuzzy ones. Level of service was classified into six levels to be consistent with the HCM, and there was no further discussion on the number of LOS categories.

In summary, no previous studies have addressed how to determine the number of LOS categories on a transportation facility or service. This study therefore addresses how to determine the actual number of LOS categories for BRT based on fuzzy clustering of user perception. We hope this work can lay the groundwork for methodology to determine the number of LOS categories for a transportation facility or service and encourage further exploration.

## User perception

User perception of a transportation facility or service is the user's personal experience of the operational conditions of a transportation facility or service. For BRT, the trip journey of an individual includes traveling from the origin to the boarding station, waiting for a bus at

the station, riding on the bus, and traveling from the alighting station to the destination. For arriving station perception and departing station perception, users are influenced by passenger arrival or departure time, availability of pedestrian overpasses or underpasses, width, isolation with bicycle lanes, and cleanliness of sidewalks. Waiting perception is affected by passenger waiting time, availability of real-time bus arrival information, availability of shelter and benches, availability of bus route map, and station width, height, and cleanliness. Riding perception is impacted by bus speed, passenger load, vehicle cleanliness, driver attitude, and comfort of seats.

Of the attributes mentioned above, the cleanliness of sidewalks, stations, and vehicles, driver attitude, and comfort of seats are difficult to quantify. The attributes of availability, such as the availability of pedestrian overpasses, involve a yes-or-no perception and do not involve a perception scale. On the other hand, people are very sensitive to time. The remaining attributes of facility condition (width, height, and isolation) and time-related attributes are more easily perceived by passengers. Crowding is also quite easily perceived by passengers. In other words, passengers intuitively perceive arrival time, waiting time, bus speed, passenger load, and departure time during the corresponding sub-journeys. Passengers also have an overall perception of the entire bus journey. User perception is therefore defined as a six-dimension vector including perceived arrival time, perceived waiting time, bus speed perception, passenger load perception, perceived departure time, and overall perception.

Perceived arrival time is a passenger's own perspective of travel time from the trip origin to the boarding station, and it differs from the actual arrival time. Perceived waiting time is a passenger's expected wait for a bus at the station, which differs from the actual waiting time. Bus speed perception is a passenger's experience rating on bus running speed generated during the bus ride, which is measured on a scale of 1–10 (1 = very poor, 10 = very good). Passenger load perception is a passenger's experience rating on crowding encountered during the bus ride, also measured on a scale of 1–10 (1 = very poor, 10 = very good). Perceived departure time is a passenger's perception of travel time from the alighting station to the destination, different from the actual departure time. Overall perception is a passenger's experience rating on the operational conditions of the entire bus travel from origin to destination, measured on a scale of 1–10 (1 = very poor, 10 = very good).

## Data collection

### Transit travel survey system

The research team developed a smartphone-based transit travel survey system, comprised of an app, a server, and a web interface. The app, shown in Fig. 1, interacts with users and collects and uploads data to the server. The server stores data and prepares the data for the web interface. The web interface is the output end of the system, from which the collected data are downloaded.

The data downloaded from the web interface (i.e., the collected data by this system) include personal demographics, the moments users present at each travel node, the times users spend in each sub-journey, user perception, and other travel information. Specifically, personal demographics include gender, age group, highest completed educational level, occupation, monthly income, car ownership status, most common travel mode, BRT use frequency, and city of residence. The moments include when a user sets out from the origin, arrives at the boarding station, boards a bus, alights from the bus, arrives at the transfer station, boards a transfer bus, alights from the transfer bus, and arrives at the destination. The times include those spent by a user arriving at a boarding station, waiting for a bus, riding on the bus, arriving at the transfer station, waiting for a transfer bus, transferring a bus, and departing from the station. User perception (i.e., perceived arrival time, perceived waiting time, bus speed perception, passenger load perception, perceived transfer time, perceived departure time, and overall perception) is gathered. Other

travel information such as travel purpose, bus route number, ride distance, and bus speed are collected as well.

### BRT user perception survey

The Guangzhou BRT, Changzhou BRT, and Yichang BRT are the most representative systems in Asia and were selected to conduct the user perception surveys using the transit travel survey system. The basic survey process included recruiting participants to use the app in daily travel for data collection. Recruiting participants is crucial and difficult. The research team tried several recruiting methods, such as using an investigation firm, acquaintance recommendations, and WeChat marketing. Ultimately, face-to-face communication with passengers at the station proved the most effective way of recruiting, specifically at stations with large passenger volumes. For Guangzhou BRT, we entered the Gangding, Normal University & Jinan University, Shipai Qiao, and Tangxiacun stations, which are illustrated in Fig. 2(a). Four surveyors worked November 4–6, 2017, and 65 participants were gained. For Changzhou BRT, the stations of Wanfu Qiao, Huaide Road & Yanling Road, Lanling Road & Renmin Road, and Qinggong College were entered, as shown in Fig. 2(b). A total of 100 participants were recruited during November 26–29, 2017. For Yichang BRT, the working stations included Wuyi Square, Pagoda River, Liujiadayan, and Lvluo Road, as shown in Fig. 2(c). A total of 100 participants were obtained January 16–19, 2018.

After finding participants interested in the survey, data collection began through participants using the app in daily BRT travel. One data record was received per participant successfully using the app in one BRT trip. The data collection continued for one month in each city. Throughout the duration of the survey, participants were able to choose to take part in it as they wished, although surveyors frequently encouraged them to use the app when taking BRT.

Participation in the survey took place as follows:

1. Surveyors invited passengers to participate in the BRT user perception survey.
2. Passengers accepted invitations and became participants.
3. Participants downloaded and installed apps available for an Android-based smartphone and registered.
4. Prior to the formal survey, participants took a pilot survey to become familiar with the app.
5. In daily BRT travel, participants used the app during the entire trip. They ran the app and logged in before setting out from the trip origin. They operated the app following prompts at each travel node. The app was placed in the backstage when no operation was performed.
6. After finishing the BRT travel (i.e., app use ended), participants took a screenshot of the final app interface and sent the screenshot to the surveyors.
7. Surveyors judged whether participants used the app truly and correctly (i.e., whether the collected data were valid) according to the final interface.
8. After completing a valid survey (i.e., correctly using the app in one BRT travel), participants received a monetary reward. Participants could choose to end or continue their participation.

### Sample characteristics

BRT user perception surveys produced 1304 valid records. Fifty-five percent of the sample is male. Approximately 95% of the sample belongs to both the 19–30 and 31–50 age groups, and 83% have received an upper-middle education of technical college and bachelor's degree. Basically, the sample equally covers different occupations, with 65% having a monthly income lower than \$780 and 63% not owning a car. Most of the sample includes frequent transit users, with 98.24% of the most common travel mode being transit and 68% using BRT every day.



Fig. 1. App interface (partial).

For travel purpose, 67% of participants listed commuting and 25% listed recreation. In general, the sample is, perhaps not surprisingly, composed heavily of youth and middle-aged people, upper-middle education level riders, particularly low-income riders, riders without a car, and frequent transit users.

The traffic management departments in Guangzhou, Changzhou, and Yichang often conduct BRT passenger satisfaction surveys and they have a deep understanding of passenger characteristics. The research team asked the traffic management departments about the socio-economic characteristics of BRT passengers before data collection. The



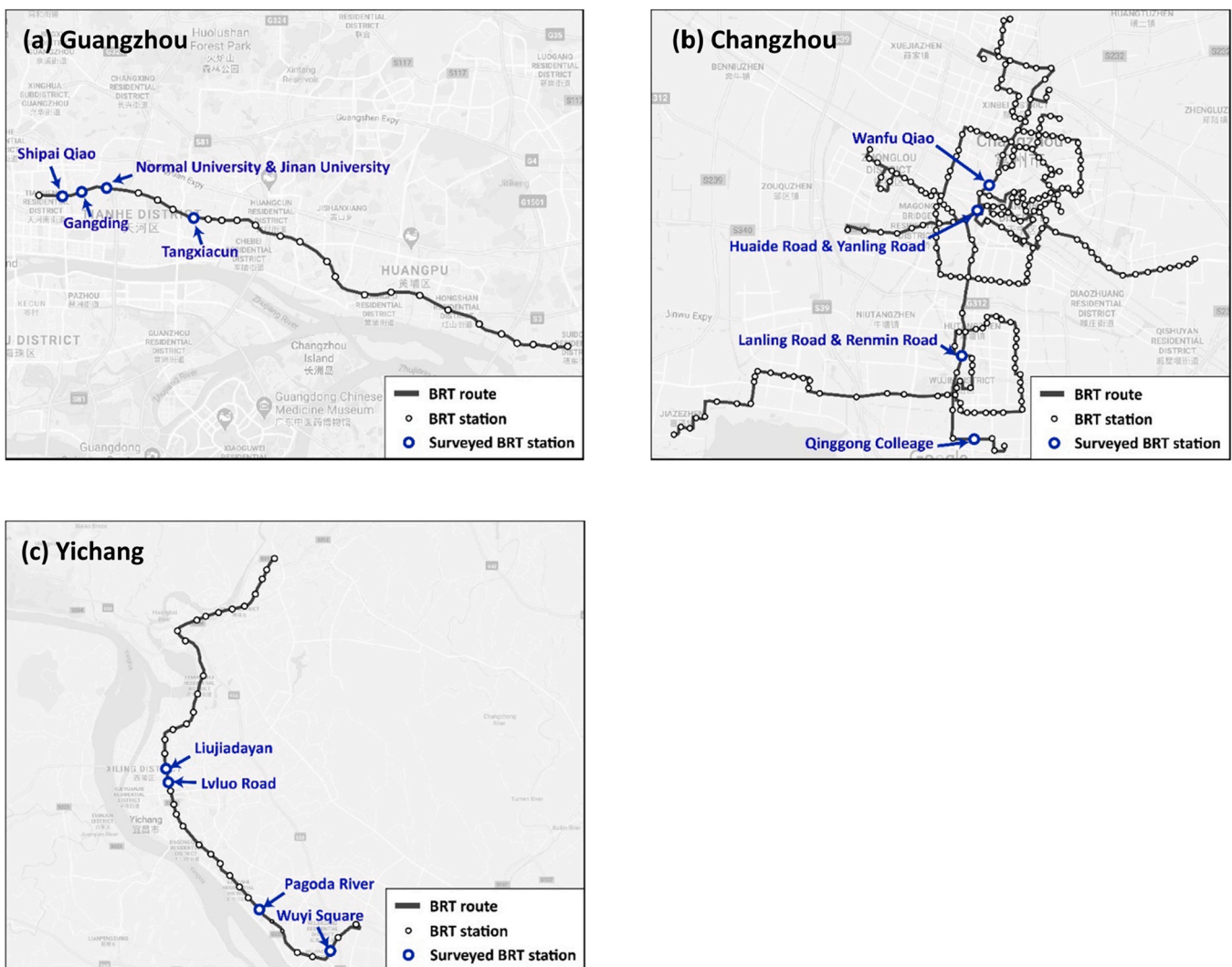


Fig. 2. BRT routes in Guangzhou, Changzhou, and Yichang.

sample collected fits the BRT passenger characteristics provided by the traffic management departments.

### Methodology

Clustering, a data-mining technology, is an unsupervised learning method always used to discover the potential knowledge hidden behind data (Fang and Pecheux, 2009). Clustering refers to the procedure of partitioning a finite dataset of objects into a certain number of clusters such that objects in the same clusters are homogeneous, whereas objects in different clusters are heterogeneous (Fang and Pecheux, 2009). Defining LOS is basically a clustering problem (Bhuyan and Rao, 2010). Hence, a clustering method was used for this study. Since user perception is fundamentally subjective and inexact (Fang et al. 2003), fuzzy C-means clustering (FCM) was adopted. Because FCM may be trapped in a local optimum, it was improved by a simulated annealing genetic algorithm (SAGA-FCM), thereby making FCM more effective by fast convergence to a global optimal solution (Zhang et al. 2017).

The objective of this study was to identify the appropriate number of LOS categories for BRT. The angle of clustering was used to discover the optimal number of clusters for a given dataset. Clustering is seen as an unsupervised learning method because there is no concrete answer to what is the optimal number of clusters for a given dataset (Kim et al., 2018), which requires developing a methodology to evaluate clustering results. This methodology is cluster validity indices (CVIs); a decision

on the optimal number of clusters for a given dataset is always judged from CVIs (Askari et al., 2017). The following CVIs were used herein to determine the appropriate number of LOS categories: Calinski–Harabasz index (CH), Dunn’s index (D), partition coefficient (PC), partition entropy (PE), Fukuyama–Sugeno index (FS), Xie–Beni index (XB), and Pakhira–Bandyopadhyay index (PB) (Kim et al., 2003; Pakhira et al., 2004; Zhao and Fränti 2014).

The user perception dataset contains a total of 1304 records that were used to study the number of LOS categories for BRT. First, the potential number of LOS categories was assumed to be 2–10. Subsequently, user perception was partitioned into 2–10 clusters using FCM and SAGA-FCM. Next, the values of multiple CVIs were computed for 2–10 clusters based on cluster results and user perception. Finally, the optimal number of BRT LOS categories was identified by analyzing the CVIs.

### Fuzzy C-means clustering (FCM)

FCM is a clustering algorithm that groups the data points in multi-dimensional space into a specific number of clusters with every data point in the dataset belonging to every cluster for a certain membership (Zhang et al. 2017). The purpose of FCM is to determine the cluster center  $V = (v_1, v_2, \dots, v_c)$  and associated membership matrix  $U = \{u_{ij}\}$ . It is based on the minimization of the following objective function:

$$J_m = \sum_{j=1}^n \sum_{i=1}^c \mu_{ij}^m \|x_j - v_i\|^2 \quad (1)$$

where  $x_j$  is the  $j$ th  $d$ -dimensional measured data point,  $n$  is the number of data points,  $v_i$  is the  $i$ th  $d$ -dimensional cluster center,  $c$  is the number of clusters,  $\mu_{ij}$  is the membership of  $x_j$  in cluster  $i$ ,  $m$  is a fuzziness index, and the default value is 2.

Fuzzy partitioning is performed through an iterative optimization of the objective function in Eq. (1) with adjustment of membership  $\mu_{ij}$  and cluster center  $v_i$  as follows:

$$\mu_{ij} = 1 / \sum_{q=1}^c \left( \frac{\|x_j - v_i\|}{\|x_j - v_q\|} \right)^{2/m-1} \quad 1 \leq i \leq c, 1 \leq j \leq n \quad (2)$$

$$v_i = \sum_{j=1}^n \mu_{ij}^m x_j / \sum_{j=1}^n \mu_{ij}^m \quad 1 \leq i \leq c \quad (3)$$

This iteration will stop when

$$\max_{ij} (|\mu_{ij}^{(k+1)} - \mu_{ij}^{(k)}|) < \varepsilon \quad (4)$$

where  $\varepsilon$  is a termination criterion.

#### FCM Based on Simulated Annealing Genetic Algorithm (SAGA-FCM)

The basic steps of SAGA-FCM are as follows (Zhang et al. 2017):

- 1) Initiate the controlling parameters, population size  $sizepop$ , maximum generation  $MAXGEN$ , crossover probability  $P_c$ , mutation probability  $P_m$ , initial annealing temperature  $T_0$ , temperature cooling coefficient  $q$ , and ending temperature  $T_{end}$ .
- 2) Initiate random  $c$  clustering centers, create the initial population  $Chrom$ , calculate the membership of individual samples for each clustering center using Eq. (2), and calculate the fitness value  $f_i$  of each sample, where  $i = 1, 2, \dots, sizepop$ .
- 3) Let the cycle count parameter  $gen$  be 0.
- 4) For the group  $Chrom$ , perform genetic operations including selection, crossover, and mutation. Eqs. (2) and (3) are used to calculate the membership for each cluster center and the fitness value  $f'_i$  of each individual. If  $f'_i > f_i$ , replace the old individual with the new one. Otherwise, accept the new individual with probability  $P = \exp((f'_i - f_i)T)$  and abandon the old one, where required.
- 5) If  $gen < MAXGEN$ , then  $gen = gen + 1$ . Go to step 4. Otherwise, go to step 6.
- 6) If  $T_i < T_{end}$ , the algorithm ends successfully and returns the globally optimal solution. Otherwise, perform the cooling operation  $T_{i+1} = qT_i$  and proceed to step 3.

#### Cluster validity indices

The CVIs of CH, D, PC, PE, FS, XB, and PB were used to determine the appropriate number of LOS categories. Table 1 provides brief explanations of these indices.

#### Results and analysis

We wrote two main MATLAB programs. One invokes FCM to classify user perception data into 2–10 clusters and calculate the values of the seven CVIs for 2–10 clusters. The other executes SAGA-FCM and computes the values of CVIs over each cluster. The values of the SAGA-FCM parameters are  $sizepop = 10$ ,  $MAXGEN = 100$ ,  $P_c = 0.7$ ,  $P_m = 0.01$ ,  $T_0 = 100$ ,  $q = 0.8$ , and  $T_{end} = 20$ . The outputs of the two programs are CVI values, membership matrices, and cluster centers for a cluster number of 2–10.

The CVIs of CH, D, and PC are designed to be maximized with the optimal number of clusters and others are designed to be minimized.

Thus, CH, D, and PC are changed to  $1/CH$ ,  $1/D$ , and  $1/PC$  to make analysis convenient. The value ranges of the seven CVIs differ greatly, so they are standardized using  $CVI_{si} = \frac{CVI_i - CVI_i^{\min}}{CVI_i^{\max} - CVI_i^{\min}}$ ,  $CVI_{si}$  is the  $i$ th standardized cluster validity index.

Fig. 3 displays the values of the standardized CVIs with FCM partitioning user perception into 2–10 clusters. The values of  $1/CH$  decline in a fluctuating manner as the cluster number increases from 2 to 10. The minimum value of  $1/CH$  is attained as the cluster number is eight. Thus, the appropriate cluster number of user perception is eight according to CH. That is, users can perceive eight categories; therefore, BRT LOS should be partitioned into eight categories according to CH from the perspective of users. Since  $1/D$  shows a declining trend with user perception grouped into 2–10 clusters and obtains the minimum value at 5 and 6 clusters, the most desirable partition of user perception is 5 or 6 clusters based on D. Correspondingly, users can distinguish five or six LOS categories, and the appropriate number of BRT LOS categories is five or six according to D. As the cluster number of user perception varies from 2 to 10,  $1/PC$  and PE gradually increase. The minimum values of  $1/PC$  and PE are reached at cluster two, meaning an optimal partition of user perception is two clusters. That is, users can differentiate two LOS categories and BRT LOS should be divided into two categories according to  $1/PC$  and PE. Since FS fluctuates over 2–10 clusters and the minimum value appears at 6 and 10 clusters, users can distinguish 6 or 10 LOS categories. According to FS, BRT level of service should be partitioned into 6 or 10 categories. XB indicates a slow increase as the cluster number increases from two to nine and then a sharp increase. The minimum value of XB appears as user perception, which is classified into two clusters. Therefore, users can perceive two LOS categories and the appropriate number of LOS categories is two based on XB. PB also displays a slow increase as the cluster number increases from 2 to 6, but then it sharply increases from 6 to 10. The minimum value of PB appears when user perception is grouped into two clusters, while the values for three and four clusters are nearly equal to the minimum value. Hence, the proper cluster number of user perception is two, three, or four based on PB, and accordingly the appropriate number of BRT LOS categories is two, three, or four.

Fig. 4 shows the values of standardized CVIs using SAGA-FCM, classifying user perception into 2–10 clusters. For the CVIs of D, PC, PE, XB, and PB, the results using SAGA-FCM are the same as those of FCM, as shown in Table 2. The minimum value of  $1/CH$  using SAGA-FCM and grouping user perception into 2–10 clusters appears at 7 clusters. Therefore, the correct number of BRT LOS categories is seven according to CH. The minimum value of FS is achieved as user perception and classified into 10 clusters using SAGA-FCM, and the values for 4, 8, and 9 clusters are nearly equal to the minimum value. Thus, the desirable partition of user perception is 4, 8, 9, or 10 clusters. Correspondingly, the appropriate number of BRT LOS categories is 4, 8, 9, or 10 according to FS.

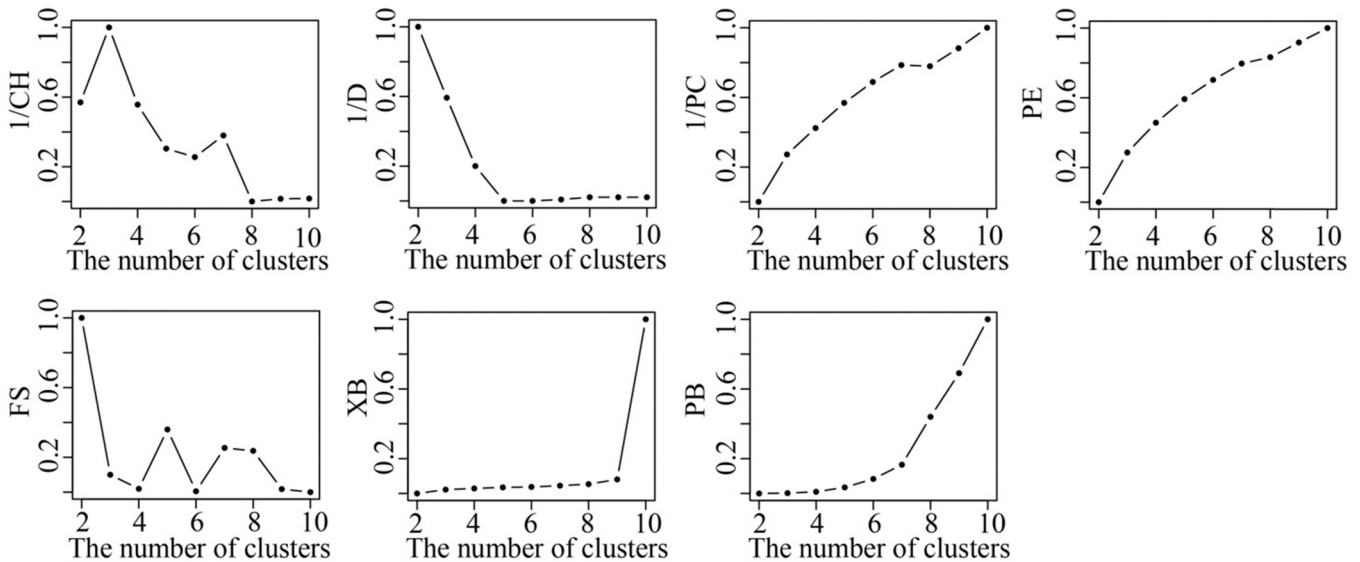
The CVIs of CH and D only consider the geometric structure of the dataset, and they are more suitable for evaluating hard clustering. PC and PE are designed to reflect the membership of fuzzy clustering. FS, XB, and PB are validity indices that simultaneously consider the geometric structure of the dataset and fuzzy membership. Some new indices proposed in the latest research on cluster validity evaluation are defined to report both data structure and membership (Askari et al., 2017; Kim et al., 2018, 2003; Pakhira et al., 2004; Zhao and Fränti 2014). Therefore, it is reasonable to state that FS, XB, and PB are more comprehensive and effective indices and that they influence the partition evaluation results.

The indices FS, XB, and PB focus on two properties of fuzzy partition: intra-cluster compactness and inter-cluster separation. FS is the difference between compactness and separation, XB is compactness divided by separation, and PB is the product of three factors, the minimization of which ensures the formation of a small number of compact clusters with large separation between clusters (Pakhira et al.,

**Table 1**  
Definitions of Cluster Validity Indices.

| Index                       | Abb. | Definition   | Optimal Value |
|-----------------------------|------|--|---------------|
| Calinski–Harabasz Index     | CH   | $\frac{SS_B}{SS_W} \times \frac{n-c}{c-1}$ $SS_B = \sum_{i=1}^c n_i \ v_i - \bar{X}\ ^2$ $SS_W = \sum_{i=1}^c \sum_{x \in C_i} \ x - v_i\ ^2$ $\bar{X} = \frac{1}{n} \sum_{j=1}^n x_j$       | max           |
| Dunn's Index                | D    | $\frac{\min_{i=1}^c \min_{j=i+1}^c d(C_i, C_j)}{\max_{k=1}^c \text{diam}(C_k)}$ $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} \ x - y\ ^2$ $\text{diam}(C_k) = \max_{x, y \in C_k} \ x - y\ ^2$ | max           |
| Partition Coefficient       | PC   | $\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m$   | max           |
| Partition Entropy           | PE   | $-\frac{1}{n} \sum_{i=1}^c \sum_{j=1}^n [\mu_{ij} \log_a(\mu_{ij})]$   | min           |
| Fukuyama–Sugeno Index       | FS   | $\sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \ x_j - v_i\ ^2 - \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \ v_i - \bar{v}\ ^2$  | min           |
| Xie–Beni Index              | XB   | $\frac{\sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \ x_j - v_i\ ^2}{n \min_{i \neq k} \ v_i - v_k\ ^2}$   | min           |
| Pakhira–Bandyopadhyay Index | PB   | $1 / \left( \frac{1}{c} \frac{E_1}{J} D \right)^2$ $E_1 = \sum_{j=1}^n \ x_j - v\ $ $J = \sum_{j=1}^n \sum_{i=1}^c \mu_{ij}^m \ x_j - v_i\ ^2$ $D = \max_{i, j=1}^c \ v_i - v_j\ $           | min           |

The above-mentioned variables are as follows:  $x_j$  is the  $j$ th  $d$ -dimensional measured data point;  $n$  is the number of data points;  $v_i$  is the  $i$ th  $d$ -dimensional cluster center;  $c$  is the number of clusters, cluster center  $V = (v_1, v_2, \dots, v_c)$ ;  $\mu_{ij}$  is the membership of  $x_j$  in cluster  $i$ ;  $m$  is a fuzziness index;  $C_i$  is the  $i$ th cluster;  $\bar{v}$  is the center of dataset.



**Fig. 3.** Values of standardized CVIs based on FCM of user perception.

2004).

Kim et al. (2003) conducted performance reliability comparisons across eight CVIs on seven datasets. These seven datasets—BENSAID, SYNTHETIC, STARFIELD, X30, BUTTERFLY, IRIS, and COLOR—are commonly used to test the effectiveness of the clustering method and validity indices. The eight CVIs contain XB and FS. The authors observed that XB is one of the most reliable indices because it correctly recognizes the number of clusters for six of the seven datasets, and FS is more unreliable than the other indices. In addition, XB is more frequently used than FS in current cluster validity evaluation research (Kim et al., 2018). Hence, XB demonstrates more reliable performance than FS in finding the best partition.

Pakhira et al. (2004) chose four real-life and four artificial datasets to compare the performance of PB and XB. The four artificial datasets are CIRCULAR\_5\_2, CIRCULAR\_6\_2, ELLIPTICAL\_10\_2, and

SPHERICAL\_4\_3, and the four real-life datasets are IRIS, CRUDE\_OIL, CANCER, and KALAZAAR, which are well-known. These authors discovered that PB can identify the right cluster number for all datasets except KALAZAAR, while XB recognizes the right cluster number for two of the eight datasets. Thus, based on the findings of this study, PB outperforms XB in searching for optimal clustering.

In general, PB performs better in determining the optimal number of clusters among FS, XB, and PB. This paper adopted the result of the clustering evaluation of PB, meaning users can perceive two, three, or four levels of service; accordingly, BRT LOS should be divided into two, three, or four categories. Pecheux et al. (2000) addressed the number of levels of service individuals could perceive at signalized intersections and found that two and perhaps three levels of service are generally perceived. The research findings in this study are similar to the findings in their study.

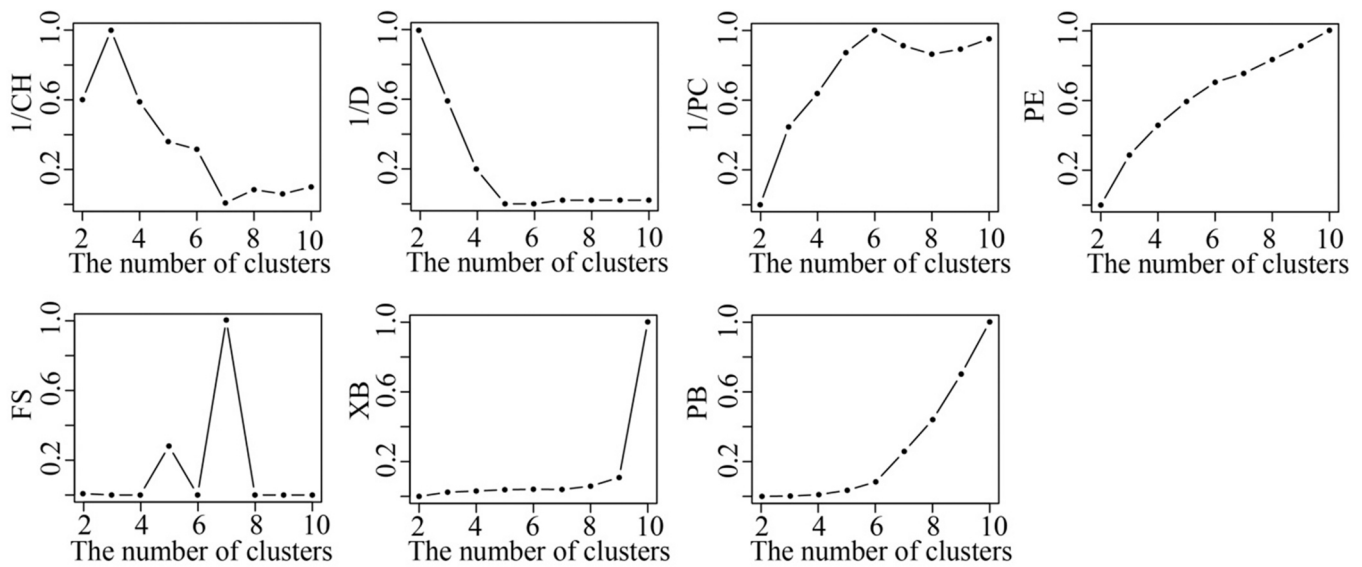


Fig. 4. Values of standardized CVIs based on SAGA-FCM of user perception.

Table 2  
Result Summary of BRT LOS Categories Based on User Perception.

| CVIs | Appropriate Number of BRT LOS Categories Based on FCM | Appropriate Number of BRT LOS Categories Based on SAGA-FCM |
|------|---|--|
| CH   | 8   | 7  |
| D    | 5, 6  | 5, 6   |
| PC   | 2   | 2  |
| PE   | 2   | 2  |
| FS   | 6, 10   | 4, 8, 9, 10  |
| XB   | 2   | 2  |
| PB   | 2, 3, 4   | 2, 3, 4  |

Conclusions

This paper presents a framework to determine the appropriate number of BRT LOS categories based on fuzzy clustering of user perception (i.e., the number of BRT levels of service that users can perceive). The research team developed a smartphone-based transit travel survey system to collect personal demographics, travel characteristics, and user perception. The surveys were conducted in Guangzhou BRT, Changzhou BRT, and Yichang BRT, which produced a user perception dataset containing 1304 records. To deal with the ambiguity in user perception, FCM and SAGA-FCM were adopted to partition user perception into 2–10 clusters. Seven CVIs of CH, D, PC, PE, FS, XB, and PB were used to search for the appropriate number of LOS categories. By analyzing the values of CVIs, we observed that users could perceive two to four levels of service, and accordingly BRT LOS should be divided into two to four categories, thereby supporting the conclusion of Pecheux et al. (2000).

User perception was defined to represent the entire bus journey and each sub-journey, enabling determination of the number of LOS categories to rely on the comprehensive perception of bus transit passengers. The development of the transit travel survey system collected data in real time instead of depending on respondents’ recall. In contrast to video laboratory studies, field surveys were conducted to collect data. To classify user perception, SAGA-FCM was used, which improved the shortcoming of FCM becoming trapped in local optima. The idea of cluster validity evaluation was applied to determine the appropriate number of LOS categories. These are the special features of this study.

Several studies have shown that the value of waiting time is higher than that of in-vehicle time. Indeed, passengers feel more dissatisfied with their waiting time than in-vehicle travel time. For example, Lu

et al. (2018) found that for the London-specific waiting time, the in-vehicle time multiplier was 2.0. This means that on average London bus travelers value changes in their waiting time two times more than changes in their in-vehicle time. In this paper, we considered perceived waiting time to study the number of LOS categories. We did not consider time values from the user’s viewpoint. However, it would be interesting for future research to consider time values to address the number of LOS categories for public transit.

The perceived in-vehicle crowding levels (i.e., passenger load perception) differ for seated and standing passengers. User perception in this paper is defined as perceived arrival time, perceived waiting time, bus speed perception, passenger load perception, perceived departure time, and overall perception. The passenger load perception is a component of user perception, but for other components, there is no need to distinguish between seated and standing passengers. Therefore, we did not make a distinction between the level of discomfort experienced by seated and standing passengers. In the future, it would be beneficial to make this distinction when developing BRT level of service criteria for passenger load.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China under grant number 51668048, 52062039, and the authors thank them for their support. The authors also thank Dr. Yuan Zhu from Inner Mongolia University for his suggestions and the anonymous reviewers for their constructive comments.

References

Askari, S., Montazerin, N., Zarandi, M.H.F., 2017. Generalized possibilistic fuzzy C-means with novel cluster validity indices for clustering noisy data. *Appl. Soft Comput.* 53, 262–283.  
 Bhuyan, P.K., Mohapatra, S.S., 2014. Affinity propagation clustering in defining level of service criteria of urban streets. *Transport* 29, 401–411.  
 Bhuyan, P.K., Rao, K.V.K., 2010. FCM clustering using GPS data for defining level of service criteria of urban streets in Indian context. *Transport Problems* 5, 105–113.  
 Bhuyan, P.K., Rao, K.V.K., 2011. Defining level of service criteria of urban streets in Indian context. *Eur. Transp.* 49, 38–52.  
 Bus Rapid Transit-Centre of Excellence. <http://www.brt.cl/>. Accessed June 26, 2020.



- Cameron, R., 1996. "G3F7-an expanded LOS gradation system. ITE J. 66, 40–41.
- Cervero, R., Kang, C.D., 2011. Bus rapid transit impacts on land uses and land values in Seoul, Korea. *Transp. Policy* 18, 102–116.
- Choocharukul, K., Sinha, K.C., Mannering, F.L., 2004. User perceptions and engineering definitions of highway level of service: an exploratory statistical comparison. *Transp. Res. Part A* 38, 677–689.
- Das, S., Pandit, D., 2013. Importance of user perception in evaluating level of service for bus transit for a developing country Like India: a review. *Transp. Rev.* 33, 402–420.
- Das, S., Pandit, D., 2015. Determination of level-of-service scale values for quantitative bus transit service attributes based on user perception. *Transp. A. Transp. Sci.* 11, 1–21.
- Fang, C.F., Elefteriadou, L., Pecheux, K.K., Pietrucha, M.T., 2003. Using fuzzy clustering of user perception to define levels of service at signalized intersections. *J. Transp. Eng.* 129, 657–663.
- Fang, C.F., Pecheux, K.K., 2009. Fuzzy data mining approach for quantifying signalized intersection level of services based on user perceptions. *J. Transp. Eng.* 135, 349–358.
- Huo, Y.Y., Zhao, J.H., Zhang, J., Qiu, F., 2015. Development of level-of-service criteria based on a single measure for BRT in China. *J. Public Transp.* 18, 20–33.
- Kim, B., Lee, H., Kang, P., 2018. Integrating cluster validity indices based on data envelopment analysis. *Appl. Soft Comput.* 64, 94–108.
- Kim, D.W., Lee, K.H., Lee, D., 2003. Fuzzy cluster validation index based on inter-cluster proximity. *Pattern Recog. Lett.* 24, 2561–2574.
- Lu, H., Burge, P., Heywood, C., et al., 2018. The impact of real-time information on passengers' value of bus waiting time. *Transp. Res. Procedia* 31, 18–34.
- Maitra, B., Sikdar, P.K., Dhingra, S.L., 1999. Modeling congestion on urban roads and assessing level of service. *J. Transp. Eng.* 125, 508–514.
- Pakhira, M.K., Bandyopadhyay, S., Maulik, U., 2004. Validity index for crisp and fuzzy clusters. *Pattern Recog.* 37, 487–501.
- Pecheux, K.K., Pietrucha, M.T., Jovanis, P.P., User perception of level of service at signalized intersections Presentation at 79th Annual Meeting of the Transportation Research Board, Washington, DC 2000.
- Roess, R.P., Vandehey, M.A., Kittelson, W., 2010. Level of service 2010 and beyond transportation research record. *J. Transp. Res. Board* 2173, 20–27.
- TRB (Transportation Research Board), 2010. Highway capacity manual. Washington, DC: TRB. Nat. Res. Council.
- TRB (Transportation Research Board), 2013. TCRP Report 165: Transit Capacity and Quality of Service Manual, third ed.,. TRB, National Research Council.,, Washington, DC.
- Washburn, S.S., Kirschner, D.S., Rural Freeway Level of Service Based upon Traveler Perception Presentation at 85th Annual Meeting of the Transportation Research Board, Washington, DC 2006.
- Zhang, J.H., Liu, Y.Q., Infield, D., Ma, Y.C., Cao, Q.S., Tian, D., 2017. Optimal power dispatch within wind farm based on two approaches to wind turbine classification. *Renew. Energy* 102, 487–501.
- Zhang, L., Prevedouros, P.D., 2011. User perceptions of signalized intersection level of service using fuzzy logic. *Transportmetrica* 7, 279–296.
- Zhao, Q.P., Fränti, P., 2014. WB-. Data and Knowledge Engineering 92. a Sum-of-Squares Based Index for Cluster Validity, Index, pp. 77–89.
- Yueying Huo** ([hyy@imu.edu.cn](mailto:hyy@imu.edu.cn)) is a professor at the Transportation Institute of Inner Mongolia University. She finished her PhD in School of Transportation at Southeast University. She studied at the University of British Columbia from 2011 to 2012. Her research interests include public transit planning, capacity, and service quality.
- Jinhua Zhao** ([jinhua@mit.edu](mailto:jinhua@mit.edu)) is an associate professor of city and transportation planning at the Massachusetts Institute of Technology. His research interests include urban development and planning, transportation economics, and public transportation. He brings behavioral science and transportation technology together to shape travel behavior, design mobility systems, and reform urban policies. He has developed methods to sense, predict, judge, and regulate travel behavior and design multimodal mobility systems that integrate autonomous vehicles, shared mobility, and public transport.
- Xiaojuan Li** ([311987431@imu.edu.cn](mailto:311987431@imu.edu.cn)) is an associate professor at the Transportation Institute of Inner Mongolia University. She finished her PhD at the School of Transportation in Beijing Jiaotong University. Her research interests include public transit planning, capacity, and service quality.
- Chen Guo** ([guochen@chd.edu.cn](mailto:guochen@chd.edu.cn)) is a lecturer at the Transportation Institute of Inner Mongolia University. She finished her master's degree at the School of Transportation in Shandong University. Her research interests include public transit planning, capacity, and service quality.