

## Research Article

# The Reliability Model for Bike-Sharing Dispatch Based on Hotspot Detection and Hypothesis Test: A Case Study in Beijing

Chao Sun  and Jian Lu 

*School of Transportation, Southeast University, Nanjing 210000, China*

Correspondence should be addressed to Jian Lu; [lujian\\_1972@seu.edu.cn](mailto:lujian_1972@seu.edu.cn)

Received 20 October 2021; Revised 26 December 2021; Accepted 11 January 2022; Published 28 January 2022

Academic Editor: Tingsong Wang

Copyright © 2022 Chao Sun and Jian Lu. 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.

In this article, a novel reliability model for bike-sharing dispatch is established using a hypothesis test. Based on the bike-sharing trajectory data from hotspot detection, we first perform the kernel density analysis to identify the dispatch points. As a result, a buffer area of 500 meters radius is designated as the studied dispatch area. From a systematic perspective, the reliability of the dispatch system is user-oriented during an ideal period when shared bikes constantly enter and leave the area. We propose the performance function of bike-sharing dispatch, in which the difference between origin and destination (OD) is defined as the main parameter of the failure probability of the system. By adopting different distribution forms, including Poisson distribution, Rayleigh distribution, exponential distribution, normal distribution, and gamma distribution, we examine the distribution characteristics of OD differences. The maximum likelihood estimation (MLE) technique is applied for model calibration, and chi-squared statistics are used to identify the acceptance of the null hypothesis. Finally, we take Beijing city as a case to verify this model. The results show that among many distribution models, the fitting goodness of normal distribution is the best. According to the properties and parameters of the distribution functions, we solve the dispatch scale for bike sharing at different confidence levels, allowing the dispatch strategy to be more flexible. Moreover, we find that the variation of dispatch quantity across different time periods and locations follows a systematic fluctuating trend.

## 1. Introduction

With the rise of sharing economy, bike sharing for public service has become a significant supplementary mode for urban transportation and thus has been revitalized in recent years [1]. Nowadays, it has been popularized in big cities, especially such as Beijing and New York [2]. Bike sharing produces no pollution and almost consumes no natural resources except during the period of manufacture. In addition, bike sharing is so flexible as to be the solution to the “last mile” problem, and it is very attractive as a bus alternative [3].

Bike sharing makes transportation convenient for residents but creates challenges in traffic management. For example, it is urgent to effectuate the bike-sharing dispatch. The undersupplied bikes will lead to the lower acceptance of potential users. However, if the supply of bikes is enhanced, it will lead to much higher dispatch costs and, possibly,

underutilization of resources. In general, the study of bike-sharing dispatch can be divided into two main aspects: location selection and determination of dispatch quantity. Zhou [4] proposed a feasible framework consisting of point location-based mean drift clustering, VRP model, genetic algorithm, and TOPSIS evaluation method. In terms of dispatch quantity, Yang et al. [5] employed a multiple regression approach to establish the relationship between bike-sharing trips, subway distance, and check-in for different types of points of interest (POI). Regue and Recker [6] established a dynamic bike-sharing scheduling model based on the bicycle inventory model and demand redistribution model at parking spots. In recent years, with the development of data acquisition and analysis technology, more and more algorithms have been employed for experiments with site selection or bike-sharing dispatch by GPS and mobile cellular: clustering progress [7–9], incentive mechanism [10–12], mixed-integer linear programming [13], and heuristic algorithm [14, 15].

Furthermore, these models can be assessed by comfort level [16], the total cost of management [11], walking time [3, 17], and accessibility [18, 19], which have been well published and applied. Before studying bike-sharing dispatch, a very important step is to predict the demand and analyze the travel behavior. After all, the transportation system is constantly changing. For example, Bao et al. [20] utilized deep learning methods to predict short-term demand for Free-Floating Bike Sharing (FFBS). Moreover, the cost-effective analysis [21], extensive analysis [22], and mathematical models [23] achieve a breakthrough for further demonstration and prediction of bike-sharing distribution specials, usage patterns, and exploration of quantitative relationships between land use and bike-sharing demand.

The existing parametric methods-based dispatch models give more attention to the inherent connection between bike sharing, traffic geography, and social elements and less attention to simpler nonparametric theory. Still, dispatch models tend to be derived from usage forecast, which lacks a full and robust understanding of inter-OD interactions during the movement of shared bikes. In other words, as the completely spatiotemporal-free vehicle, shared bikes have a certain self-sufficient capability, ignoring which will lead to a decrease in model accuracy and thus the possibility of overdispatch. However, most of the research only presents a strict and definite result (like literature [5–8]), which does not describe the fluctuation mechanism of the confidence level that the system can meet users' travel demand as the dispatch quantity changes. The confidence level is the effectiveness degree of dispatch strategies believed by bike-sharing managers, which is actually a measure of the rationality of dispatch from the perspective of probability. Only with the help of confidence level, we can determine the threshold of dispatch according to the goal to be achieved by the system so as to avoid overdispatch to the greatest extent and increase flexibility. To solve the challenging difficulties, this article aims to establish a flexible and reliable system based on hotspot detection to dispatch the shared bikes in a single area using hypothesis testing.

Our presented system can address the following questions:

- (1) How to identify the best dispatch points and areas according to the spatial and temporal distribution characteristics of the bike-sharing demand?
- (2) What kind of probability distribution does the bike-sharing trip generation and attraction follow? And how to investigate the probability distribution characteristics of a single region from a whole according to the sampling theorem?
- (3) For a certain area, how to determine the reliability of different dispatch quantities to meet the travel demand based on OD difference? Or equivalently, how to find out the threshold of dispatch quantity based on OD difference at given reliability?

By answering the above-mentioned questions, the research gaps and contributions in this article are mainly as follows. (1) According to our proposal, the hotspot detection

procedure-based trajectory data of bike sharing and kernel density analysis is employed to determine the dispatch points and dispatch areas. Compared with the traditional grid and traffic analysis area, our presented model solves modifiable area unit problems in a quicker and more direct manner. Then, the reliability model is deployed to appraise the failure probability of bike-sharing dispatch system. (2) We put forward the performance function and reliability model of bike-sharing dispatch based on OD differences, thus ensuring dynamic supply to be sufficient in an open system. (3) From the entire area to the local, we evaluate the probability distribution of bike-sharing demand, which can provide insights for future related research.

## 2. Methodology

*2.1. Methodology Framework.* In this article, we investigate the variation of dispatch location and dispatch quantity when reliability levels differ. The methodology framework is mainly constructed based on hotspot analysis, reliability modeling, probability distribution analysis, and case verification, respectively, as shown in Figure 1.

### *2.2. Model for Dispatch Location Based on Hotspot Detection*

*2.2.1. Mining Dispatch Points Based on Hotspot Detection.* "A hotspot for bike-sharing means an area where similar travel events take place at the same time and are geo-referenced on the map [24]. As spatial density analysis algorithms (including spatial relative analysis and clustering) make research progresses, it becomes readily easy to accurately grasp the location and influence the degree of hotspots. For example, the distribution may be responsible for the crime, disease, and traffic collision, and their evaluation is performed by hotspot detection [25]. Advancement in navigation technology enables us to receive extensive bike-sharing trajectory data from moving agents and gain insights into both individual behaviors and group mobility. The study of travel characteristics based on big data and urban computing is no longer limited to the traditional coarse-grained classification. In recent years, the thinning hotspot extraction model-based density field has become an important tool to analyze OD features and tracks [26, 27].

Density analysis is subject to a spatial relationship with the Earth, and it transmits data to the surface and thereby displays the distribution of spots. Nowadays, kernel density estimation is a widely used approach in geospatial analysis, and it is very suitable for density estimation of large-scale spatial point data [28, 29]. Figure 2 illustrates the working principle of kernel density estimation. In the research area  $R$ , the kernel density estimation model regards a random spot as the center (called kernel  $k$ ) and calculates the density value of the targeted spot in bandwidth  $r$  by determining the quantity and distance between matter spots within the bandwidth range. Kernel density analysis is used to calculate the density of points around each output raster cell. The density of each output raster cell is calculated as the sum of all the values covering the core surface of the raster cell.

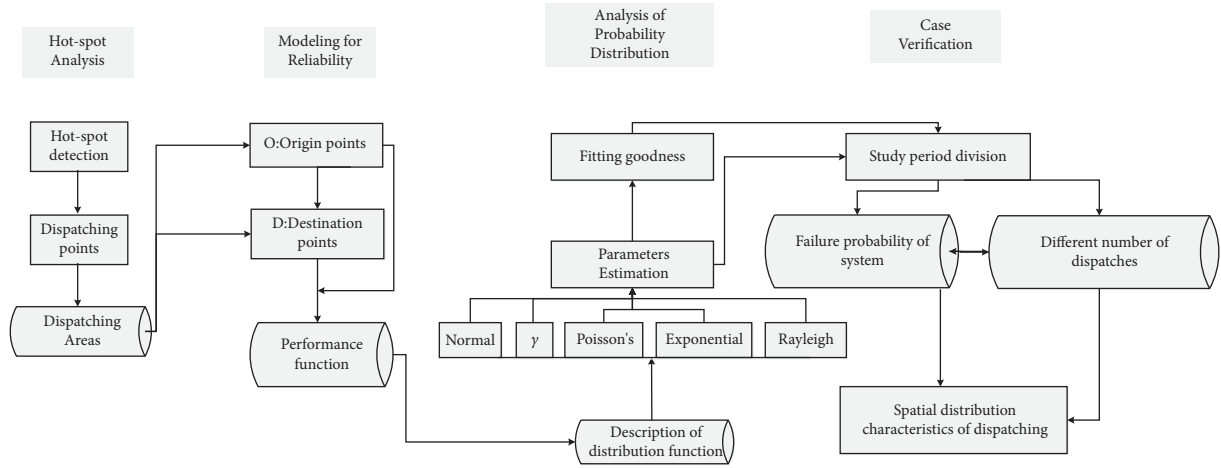


FIGURE 1: The flowchart of this article.

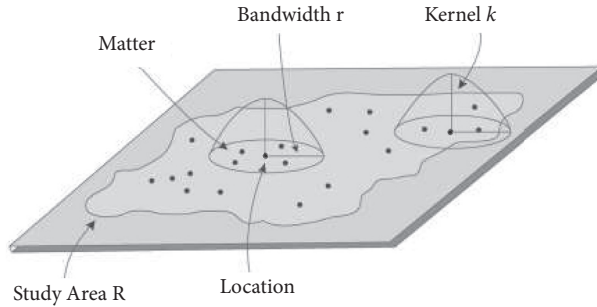


FIGURE 2: The model of kernel density estimation.

Figure 3 illustrates the flowchart of hotspot detection based on kernel density analysis. Data is firstly preprocessed by eliminating abnormal data, completing missing data, and extracting OD from the bike-sharing trajectory data. Then, OD data are input into ArcGIS, and kernel density estimation-based spatial analyst tools are applied to output raster cells with kernel density as the raster value. The tools including kernel density analysis, Window analysis, Minus, Reclassification, Raster to Polygon, and Turning Feature into Point are used in succession to obtain the travel hotspots of bike sharing and the corresponding raster values. The working principle of hotspot detection is demonstrated in Figure 4 (Sun C., Quan W. Evaluation of Bus Accessibility Based on Hotspot Detection and Matter-Element Analysis [J]. IEEE Access, 2020, PP (99):1-1).

**2.2.2. Establishment of Dispatch Areas.** A travel hotspot of bike sharing is defined as a dispatch point, which is deployed to buffer an area in the geographic database under the Geographic Information System so as to border the dispatch area. According to the evaluation system of the bus in the Transit Metropolis [30], the buffer radius  $r$  of the dispatch area is set as 500 meters in the article, as shown in Figure 5.

**2.3. Modeling for Reliability of Dispatch for the Bike-Sharing System.** Reliability, as an important index, is often defined as

the ability to complete a predetermined function within specific periods and conditions, including the safety, applicability, and durability of a product or system, which has been widely introduced in traffic safety, food safety, and so on [31]. In this article, the bike-sharing dispatch in an area is regarded as a system, and its reliability is user-oriented during an ideal period when shared bikes constantly enter and leave the area. To evaluate the reliability and failure probability, the performance function of the bike-sharing dispatch system ( $Z$ ) in a dispatch area is proposed for the first time, as follows:

$$Z = g(O, D) = (Q_D + x) - Q_O, \tag{1}$$

where  $Q_O$  denotes the quantity of shared bikes out of this area in unit time, i.e., quantity of  $O$ ;  $Q_D$  denotes the quantity of shared bikes into this area in unit time, i.e., quantity of  $D$ ;  $x$  denotes the dispatch quantity of shared bikes for this area. Therefore, the failure probability of the system can be calculated using the following equation, given that  $Z$  obeys a certain distribution form:

$$P_f = P(Z < 0). \tag{2}$$

Apparently, if the dispatch system is required to satisfy the demand at a certain probability level with the distribution characteristic of  $Z$  tested, the dispatch quantity of

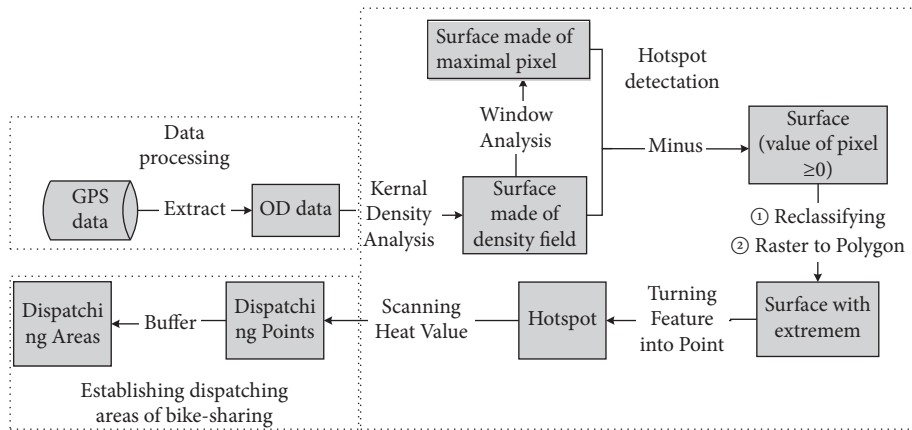


FIGURE 3: The flowchart of hotspot detection.

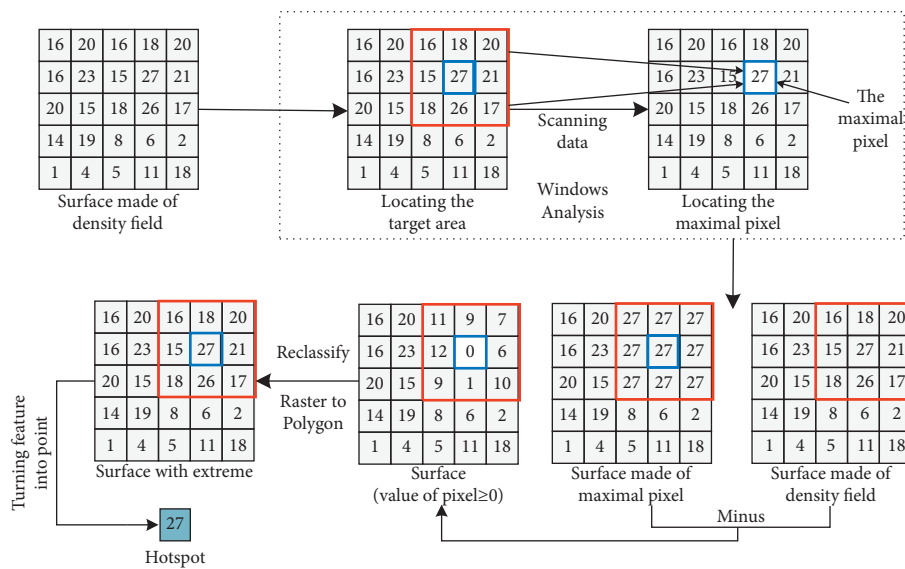


FIGURE 4: Demonstration of hotspot detection.

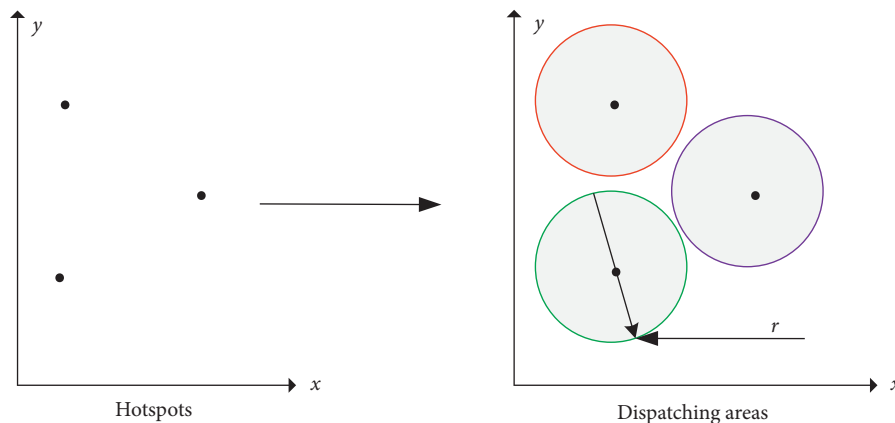


FIGURE 5: Establishment of dispatch area.

shared bikes in a single area can be obtained using equations (1) and (2). In order to further explain the principle of the model, we compare the dispatching system in this article to a water tank, as shown in Figure 6. Each dispatching area is

equivalent to a water tank, and the quantity of  $O$  per hour (veh/h) is equivalent to the water output (cubic meter/h), the quantity of  $D$  per hour (veh/h) is equivalent to the water input (cubic meter/h), and the dispatching quantity is

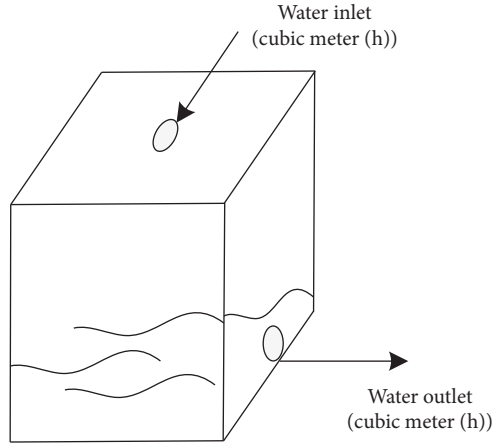


FIGURE 6: Water tank similar to the dispatching system.

equivalent to the initial water volume in the water tank. Therefore, the dispatch problem is equivalent to the following: when the water inlet and outlet all start to work continuously, what is the initial water volume in the water tank to ensure that the water tank will not dry? Therefore, to solve this problem, the most important thing is to determine the law of water inflow and outflow, that is, the law of bike-sharing OD time series.

Obviously, we take into account the interaction characteristics of bike-sharing ODs in each area, i.e., the demand-supply pattern of bike-sharing in the natural state. Not surprisingly, the shared bikes, which are constantly moving inside the city, are essentially a kind of nonhuman-intervention supply for each destination. Contrary to other dispatch models derived from usage forecast (e.g., literature [4–6, 8]), just the dispatch problem when “self-sufficiency” is ineffective needs to be solved through our model, thus being closer to the real state of the transportation system and considerably and greatly reducing the workload in practical applications.

## 2.4. The Analysis of Probability Distribution

### 2.4.1. The Statistical Distribution of Difference between O and D

(1) *The Description of Distribution Functions.* First, the statistical distribution of quantity difference between O and D in a single area should be detailed in order before assessment for the reliability of dispatch system as follows:

$$Z^* = Q_D - Q_O. \quad (3)$$

Next, according to statistics, the probability  $P(Z^* = d)$  where  $Z^*$  is equal to  $d$  can be calculated using equations (4)–(8) when  $Z^*$  follows the Poisson, Rayleigh, exponential, normal, and gamma distribution [32]:

$$P(Z^* = d) = \frac{\lambda_1^d e^{-\lambda_1}}{d!}, \quad (4)$$

where  $Z^*$  is the random variable representing the quantity difference between O and D,  $\lambda_1$  is the index of Poisson

distribution, and  $n$  is the integer (such as 1, 2, 3, ...). Alternatively, it is assumed that  $Z^*$  in a single area is continuous and follows the Rayleigh, exponential, normal, or gamma distribution, whereby the probability density of  $Z^* = d$  can be expressed as shown in the following equations:

$$f(Z^* = d) = \begin{cases} \frac{d}{\sigma_2^2} \exp\left(-\frac{d^2}{2\sigma_2^2}\right), & d \geq 0, \\ 0, & d < 0, \end{cases} \quad (5)$$

where  $f(Z^* = d)$  is the probability, where  $Z^*$  is equal to the determined value  $d$  and  $\sigma_3$  is the index of Rayleigh distribution.

$$f(Z^* = d) = \begin{cases} \lambda_3 e^{-\lambda_3 d}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (6)$$

where  $f(Z^* = d)$  is the probability, where  $Z^*$  is equal to the determined value  $d$  and  $\lambda_3$  is the index of exponential distribution.

$$f(Z^* = d) = \frac{1}{\sigma_4 \sqrt{2\pi}} e^{-(d-\mu_4)^2/2\sigma_4^2}, \quad (7)$$

where  $\sigma$  is the variance of normal distribution and  $\mu$  is the expectation of normal distribution.

$$\text{Gamma}(Z^* = d | \alpha_5, \beta_5) = \frac{\beta_5^{\alpha_5} e^{-\beta_5 d}}{\Gamma(\alpha_5)}, \quad (8)$$

where  $\alpha_5$  is the shape parameter of the gamma distribution, which mainly determines the shape of the curve, and  $\beta_5$  is the rate parameter or inverse scale parameter of the gamma distribution, which mainly determines how steep the curve is. The normalization constant of the gamma distribution (the point at which the gamma value is 1) is exactly gamma function at the point  $\alpha$   $\{\Gamma(\alpha)\}$ .

(2) *Parameters Estimation.* The MLE is employed to estimate the parameters of the above-mentioned distribution functions. The probability density of entirety is  $f(w; \theta_1, \theta_2 \dots \theta_m)$ ,  $w$  is sample value, and  $\theta_1, \theta_2 \dots \theta_m$  are unknown parameters, based on which the likelihood function is shown in the following equation:

$$L(w; \theta_1, \theta_2 \dots \theta_m) = \prod_{i=1}^N f(w; \theta_1, \theta_2 \dots \theta_m). \quad (9)$$

We can get the following equation:  $\hat{\theta}_j = \hat{\theta}_j(w_1; w_2, \dots w_N)$ . When  $\hat{\theta}_j$  satisfies equation (10),  $\hat{\theta}_j = \hat{\theta}_j(w_1; w_2, \dots w_N)$  is named after MLE of  $\theta_j$ , as follows:

$$L(\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_m) = \max_{\theta_1, \theta_2 \dots \theta_m} L(w; \theta_1, \theta_2 \dots \theta_m). \quad (10)$$

Therefore, according to the MLE, the parameter estimation results of the above-distribution models are shown in Table 1.  $\lambda$  is the expectation,  $a$  is the reciprocal of overdiscrete parameter, and  $S^2$  is the sample variance.



TABLE 1: The results of estimation for parameters.

No.	Distribution function	Method	Result of estimation
①	Poisson distribution	Maximum likelihood estimation	$\lambda_1 = 1/N \sum_{i=1}^N X_i$
②	Rayleigh distribution	Maximum likelihood estimation	$\sigma_2 = \sqrt{1/2N \sum_{i=1}^N X_i^2}$
③	Exponential distribution	Maximum likelihood estimation	$\lambda_3 = 1/N \sum_{i=1}^N X_i$
④	Gaussian distribution	Maximum likelihood estimation	$\mu_4 = 1/N \sum_{i=1}^N X_i \quad \sigma_4^2 = (1/N)S^2$
⑤	Gamma distribution	Quasi-maximum likelihood estimation	$\alpha_5 = [1/N - 1 \sum_{i=1}^N (X_i/\bar{X} - 1) \ln X_i/N\bar{X}]^{-1} \quad \beta_5 = \bar{X}/\alpha$

2.4.2. *Test of Fitting Goodness.* The chi-square test of fitting goodness for a probability distribution is applied to verify the conformity of OD difference with a certain distribution. Followings are the specific steps of hypothesis testing:

Step 1: we make the assumptions according to hypothesis testing.  $H_0$ : OD difference follows a certain distribution  $F_0(x; \theta_1, \theta_2, \dots, \theta_m)$ ;  $H_1$ : OD difference does not follow a certain distribution.

Step 2: the maximum likelihood estimation is employed to calculate the parameters  $(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m)$ .

Step 3: the data are divided into  $K$  disjoint sets of the real numbers.  $X$  is identified as one of the  $A_i (i=1, 2, \dots, K)$  and  $p_i = P(X \in A_i) (i=1, 2, \dots, k)$ . Therefore, we regard the original hypothesis ( $H_0$ ) as  $p_i = p_{i0} (i=1, 2, \dots, k)$

Step 4: the chi-squared statistic is calculated using the following equation:

$$\begin{aligned} \chi_p^2 &= \sum_{i=1}^k \frac{(Q_i - E_i)^2}{E_i} \\ &= \sum_{i=1}^k \frac{(N_i - N p_{i0})^2}{N p_{i0}}, \end{aligned} \quad (11)$$

where  $Q_i$  and  $E_i$  are the distributions of actual and theoretical values of samples on the interval  $A_i$ , respectively, and  $N_i$  is the number of samples  $(x_1, x_2, \dots, x_N)$  falling into the interval  $A_i$  and  $v = k - 1 - a$ .

Step 5: according to  $P$  value and test level ( $\alpha = 0.05$ ), we can judge the acceptable level of assumptions. If  $P \geq 0.05$ , then the null hypothesis would be accepted.

### 3. Case Study

3.1. *Study Area and Data.* This article takes Beijing as a case study. As one of the first cities to introduce shared bikes, Beijing has invested heavily in constructing a bike-sharing system in recent years, with a high coverage rate and the utilization rate among the highest in China. Therefore, by analyzing the usage characteristics of bike sharing in Beijing, this article verifies the corresponding statistical distribution and proposes the dispatch strategy, including location and quantity.

This article takes Beijing bike-sharing data provided by open data sources as research data. The main fields of the dataset include ordered and bikeid (as shown in Table 2).

TABLE 2: The description of data.

Name of fields	The description of fields
Orderid	The order reference
Bikeid	The vehicle serial number
start_time	The initial time of the order
start_location_x	The longitude of the starting position of the order
start_location_y	The latitude of the starting position of the order
end_time	The ending time of the order
end_location_x	The longitude of the ending position of the order
end_location_y	The latitude of the ending position of the order

After data preprocessing (deletion of invalid data and redundant and abnormal data), it is input into ArcGIS software. By utilizing the tool of cutting, OD data of shared bikes in the research area are obtained. To sum up, we only need to obtain the amount of OD within each dispatch area without the need for assistance from other multisource spatial data (like spatial accessibility, social networking data, simulation data, and so on [4–17]), which can be easily achieved through GIS software. More importantly, the method in this article does not need to predict the travel demand like the existing research (such as literature [4–6, 8]), which reduces the complexity and time consumption in practical application.

3.2. *The Division of Study Period.* According to the principle of unified division, the whole day is divided into equal periods according to a fixed length of time, regardless of whether the demand for the bikes-sharing wobbles or rises. Taking every hour in the whole city as the analysis unit, a single day is divided into 24 time periods. As for the study period of seven days, there are 168 periods. During the whole study period, the total bike-sharing usage is 1830100.

As shown in Figure 7, hourly bike-sharing usage ( $b_u$ ) in the entire city shows a clear bimodal distribution during the day. In the morning peak and evening peak hours, the maximum usage in bike sharing exceeds 25,000 times per hour, and the effects of the two are almost the same. In addition, there tend to be zero shared bikes between 0:00 a.m. and 6:00 a.m.

3.3. *Evaluation for Fitting Goodness.* According to the equations in Table 1 based on the maximum likelihood method, parameters of the Poisson, Rayleigh, exponential, normal, and gamma distribution are estimated using hourly bike-sharing usage out of 168 periods in the entire city (as shown in Table 3).

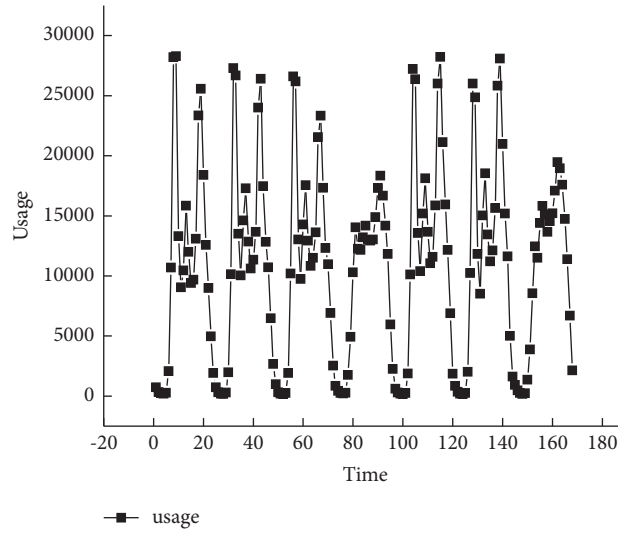


FIGURE 7: The variation of bike-sharing usage between different hours.

TABLE 3: The results of parameter estimation.

No.	Distribution function	Result of estimation
①	Poisson distribution	$\lambda_1 = 10875.02$
②	Rayleigh distribution	$\sigma_2 = 9620.67$
③	Exponential distribution	$\lambda_3 = 10875.02$
④	Gaussian distribution	$\mu_4 = 10875.02 \sigma_4^2 = 8200.54$
⑤	Gamma distribution	$\alpha_5 = 0.9428 \beta_5 = 11534.68$

In turn, the probability values calculated by the above-mentioned chi-square test (Section 2.4) show that based on other fitting indicators, we can identify whether the bike-sharing usage follows a certain probability distribution (as shown in Table 4).

$P$  is the probability of accepting the hypothesis. The closer  $P$  is to 0, the more likely it is to reject the null hypothesis. LSTAT is the value of the test statistic, and CV is the cut-off value for rejecting the null hypothesis. According to Table 4, for all the selected distributions, there are large gaps between the actual usage and distribution value for bike sharing. Combining the fitting results in the  $R$  language, we reject the null hypothesis because values of  $P$  are all less than 0.05. Therefore, bike-sharing usage is not expected to follow the normal distribution,  $\gamma$  distribution, Poisson distribution, exponential distribution, or Rayleigh distribution. However, we find that within seven days, the average usage of bike sharing per hour is lower than 2000 from 0:00 a.m. to 6:00 a.m., while the average value is larger than 15000 from 6:00 a.m. to 24:00 p.m. The small usage during the early morning hours quite differs from other periods. Therefore, we divide a day into the early morning time period (0:00–6:00) and the nonearly morning time period (6:00–24:00). In this article, we just focus on the bike-sharing dispatch during the nonearly morning time period.

Based on the above-mentioned division principle, the hourly data during nonearly morning time period are input into the distribution model for refitting. The results show that, compared with other models, the data for this period are more suitable for the normal distribution model, which

meets the significance requirements of accepting the null hypothesis ( $H_0$ ). According to MLE, the parameters of the normal distribution have changed into  $\lambda = 14296$  and  $\sigma^2 = 42532474$ . In addition to the chi-square test, the Kolmogorov–Smirnov method is also used to test whether the data follow the normal distribution in  $R$  language. By comparing the frequency distribution  $F(x)$  with the theoretical distribution  $G(x)$  or the distribution of two observed values, whether the data are consistent with the test hypothesis can be judged. The null hypothesis ( $H_0$ ) indicates that the data follow the normal distribution with the rejection area of  $DS = \max |f(x) - g(x)|$ . If the actual observed value  $DS > DS(n, \alpha)$ ,  $H_0$  would be rejected. Otherwise,  $H_0$  is accepted. According to the results of fitting,  $DS = 0.148 < DS(n, \alpha)$ ,  $P = 0.0512 > 0.05$ ,  $LSTAT = 0.1192 > CV = 0.0046$ . Hence, we can accept  $H_0$ , indicating that the bike-sharing usage amount in each hour is normally distributed. The probability density function is shown as follows:

$$f(b_u) = \frac{1}{6521\sqrt{2\pi}} e^{-\frac{(b_u - 14296)^2}{85046882}}. \tag{12}$$

Equation (12) is integrated on the frequency interval so as to obtain the probability value corresponding to the frequency interval. The product of probability and the total quantity of samples is the fitting value of normal distribution for bike-sharing usage amount. For the entire city, the quantities at  $O$  and  $D$  are equal to the bike-sharing usage. Therefore, both  $O$  and  $D$  follow the same normal distribution, as shown in equation (12). Besides, according to the sampling theorem, the distribution characteristics of a part

TABLE 4: The fitting results of different distribution models.

Type of distribution	$H$	$P$	LSTAT	CV
Poisson distribution	1	$6.31399e-42$	0.5288	$6.31399e-42$
Rayleigh distribution	1	$5.40742e-10$	0.2639	$5.40742e-10$
Exponential distribution	1	$2.48725e-07$	0.2157	$2.48725e-07$
Normal distribution	1	0.0046	0.1334	0.0046
$\gamma$ distribution	1	$2.27069e-08$	0.2314	$2.27069e-08$

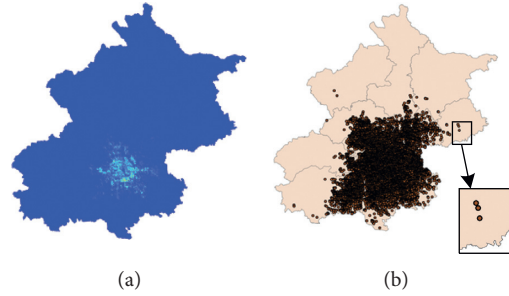


FIGURE 8: (a) Heatmaps and (b) travel hotspots of bike sharing.

extracted from the whole should be the same as that of the entirety. In other words, when a small sample area is selected from the entire city, the bike-sharing OD points in this area should also follow the normal distribution.

**3.4. The Mining Results of Dispatch Point and Area.** After coordinate conversion and data cleaning for data pre-processing, OD points of bike-sharing in urban Beijing are collected and displayed on the coordinate system of WGS1984. Based on the kernel density analysis algorithm, the heatmap of bike-sharing trips within Beijing is obtained by capturing OD data, as shown in Figure 8(a). Next, bike-sharing dispatch points are mined under hotspot detection and dispatch areas are established to identify the dispatch quantity. According to the methodology described in Section 2, all of the hotspot detections are performed with tools in ArcGIS, including 8334 travel hotspots of bike sharing calculated by raster value, as shown in Figure 8(b). Next, based on travel hotspots, 8334 buffers are constructed as dispatch areas, and the search radius is 500 meters, which are input into the geographic database. Especially, 50 hotspots out of all are selected to demonstrate the morphological characteristics of dispatch areas and to illustrate the modeling process for identifying the dispatch quantity, as shown in Figure 9.

Apparently, most studies have been keen on the cross-time and cross-regional management of bike sharing, such as literature [4, 7, 8, 12–15], but there are few good methods for single small-scale regional allocation or optimization. The hot travel zones enable the modeling problem of bike-sharing dispatch to be solved in a single area, even in a community or a school. Compared with others, by applying the dispatch model to areas with a large usage amount, the balance between supply and demand of shared bikes can be realized more effectively, which is also one of the difficulties in bike-sharing management. Moreover, unlike other

complicated multiredistribution models such as literature [11, 14], it is necessary to dispatch bike sharing within each small area only once a day in this article, and the cost of manpower and material resources will be greatly saved.

**3.5. The Identification for Dispatch Quantity.** The dispatch areas obtained by clustering algorithms and hotspot detection are essentially samples randomly subject to sampling within the entirety. This means that the difference between  $O$  and  $D$  during the nonearly morning time period in every dispatch area also conforms to the normal distribution, which is also consistent with the test for probability distribution and fitting goodness according to Section 2.4. Therefore, based on the distribution function, we can expect the bike-sharing dispatch quantity in each dispatch area to be at different reliability levels. For example, for the No. 27 dispatch area selected in this article, the mean and variance of  $O$  and  $D$  for shared bikes during the nonearly morning time period are calculated as follows:  $\lambda_3^{(O)} = 34$ ,  $\sigma_{(O)}^2 = 358.97$ ,  $\lambda_3^{(D)} = 37$ , and  $\sigma_{(D)}^2 = 545.68$ . Therefore, the normal distribution models of the two are shown in the following equations, respectively.

$$f(O) = \frac{1}{18.95\sqrt{2\pi}} e^{-(O-34)^2/717.94}, \quad (13)$$

$$f(D) = \frac{1}{23.36\sqrt{2\pi}} e^{-(D-37)^2/1091.36}. \quad (14)$$

Further, when both  $O$  and  $D$  follow the normal distribution, the dispatch quantity  $x$  is equivalent to a constant. According to the nature of normal distribution and equation (1), the OD difference of bike sharing ( $Z^*$ ) and the performance function of bike-sharing dispatch system ( $Z$ ) should also both follow the normal distribution, and then the distribution function of  $Z$  is as shown in the following equation:



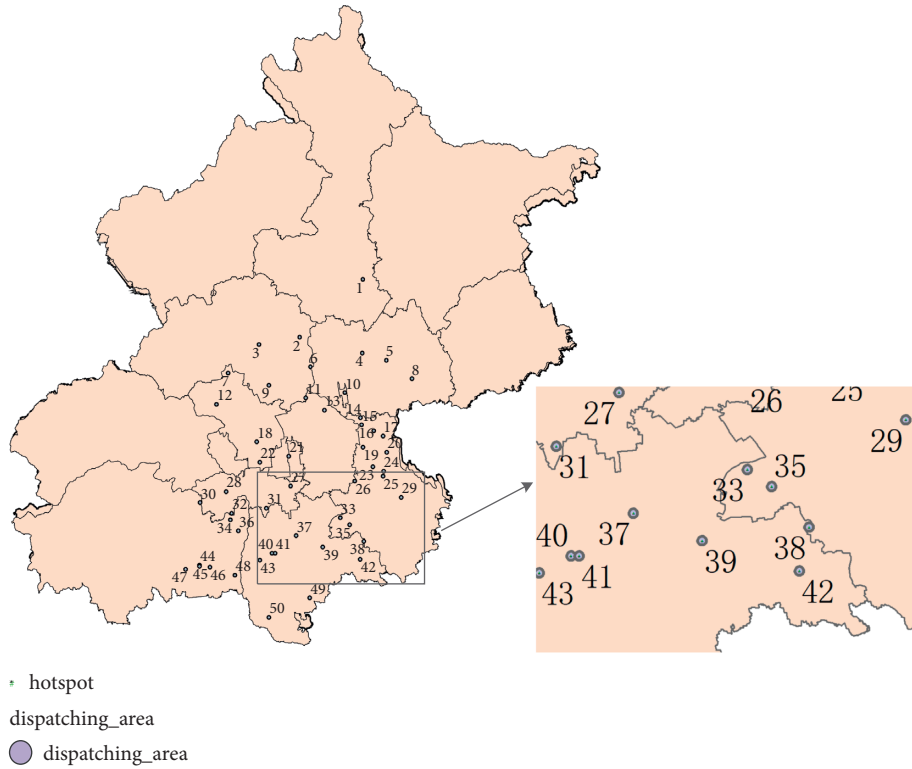


FIGURE 9: Selected dispatch points and areas for demonstrating the modeling process.

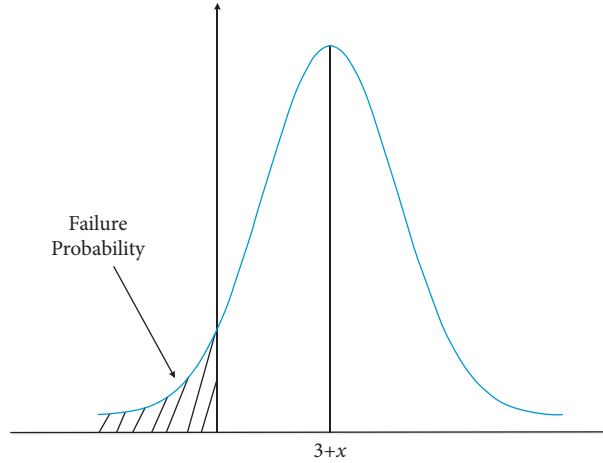


FIGURE 10: The probability distribution of the difference between  $O$  and  $D$ .

$$f(Z) = \frac{1}{30.07\sqrt{2\pi}} e^{-[Z-(3+x)]^2/1809.30}. \quad (15)$$

According to (2), the system would fail when  $Z < 0$ , which is the area corresponding to the shaded section in Figure 10. The larger the initial quantity of dispatch ( $x$ ) is, the smaller the failure probability of the system will be as the distribution function moves to the right, or in other words, the higher the system reliability of bike-sharing dispatch will be. According to the image of the distribution function, the failure probability of the system under different input states can be obtained, as shown in Table 5. It is obvious that for the

dispatch model in this article, we just need to calculate the mean and variance of OD within each dispatch area, which are the most basic metrics in statistics. Compared to other complex algorithms, including Demand-Forecast-Based Self-adaptive Model [6], Nonlinear Autoregressive Neural Network Forecasting [8], Mixed-Integer Linear Program [13], and 3-Step Math Heuristic [15], the model proposed in this article does not require a specially designed solution algorithm and has a low threshold for application in different cities, making it more practical.

According to Table 5, despite the initial dispatch quantity being zero, the probability that shared bikes from other areas

TABLE 5: The failure probability of system failure at different dispatch quantities within the 27th dispatch area.

State number	The initial quantity of dispatch	The probability of failure (%)
1	0	46.00
2	1	44.69
3	3	42.07
4	5	39.49
5	7	36.95
6	9	34.47
7	11	32.05
8	13	29.71
9	15	27.44
10	17	25.27
11	19	23.19
12	21	21.21
13	23	19.33
14	25	17.56
15	27	15.99
16	29	14.33
17	31	13.88
18	33	11.53
19	35	10.28
20	37	9.14
21	39	8.19
22	41	7.13
23	43	6.27
24	45	5.48
25	47	4.78
26	49	4.15
27	51	3.59
28	55	2.65
29	70	0.72

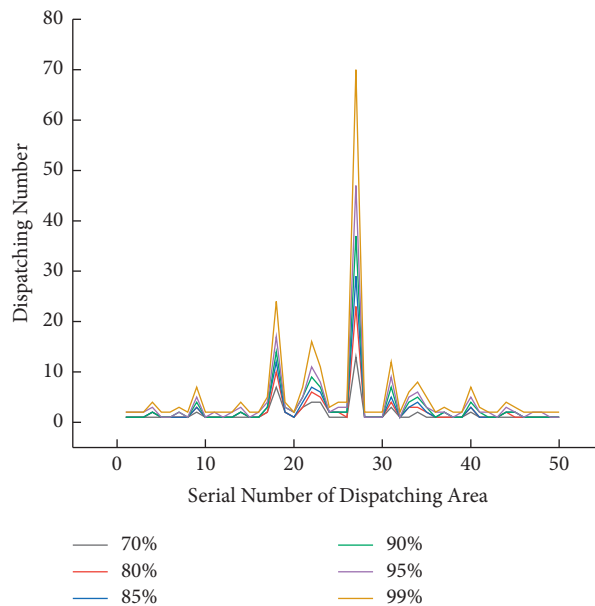


FIGURE 11: The dispatch quantities at different system reliability degrees within the different dispatch areas.

can meet the travel demands within the 27th dispatch area is larger than 50%. With the increase of initial dispatch quantity, the system reliability to meet travel demands in the dispatch area is gradually improved. Moreover, the system reliability is over 80% after the initial dispatch quantity

reaches up to 25. While the initial quantity of bike-sharing dispatch exceeds 47, the system reliability is over 95%. More importantly, the failure probability corresponding to a dispatch quantity of about 70 is 0. Based on the reliability principle, the above process solves the dispatch model within

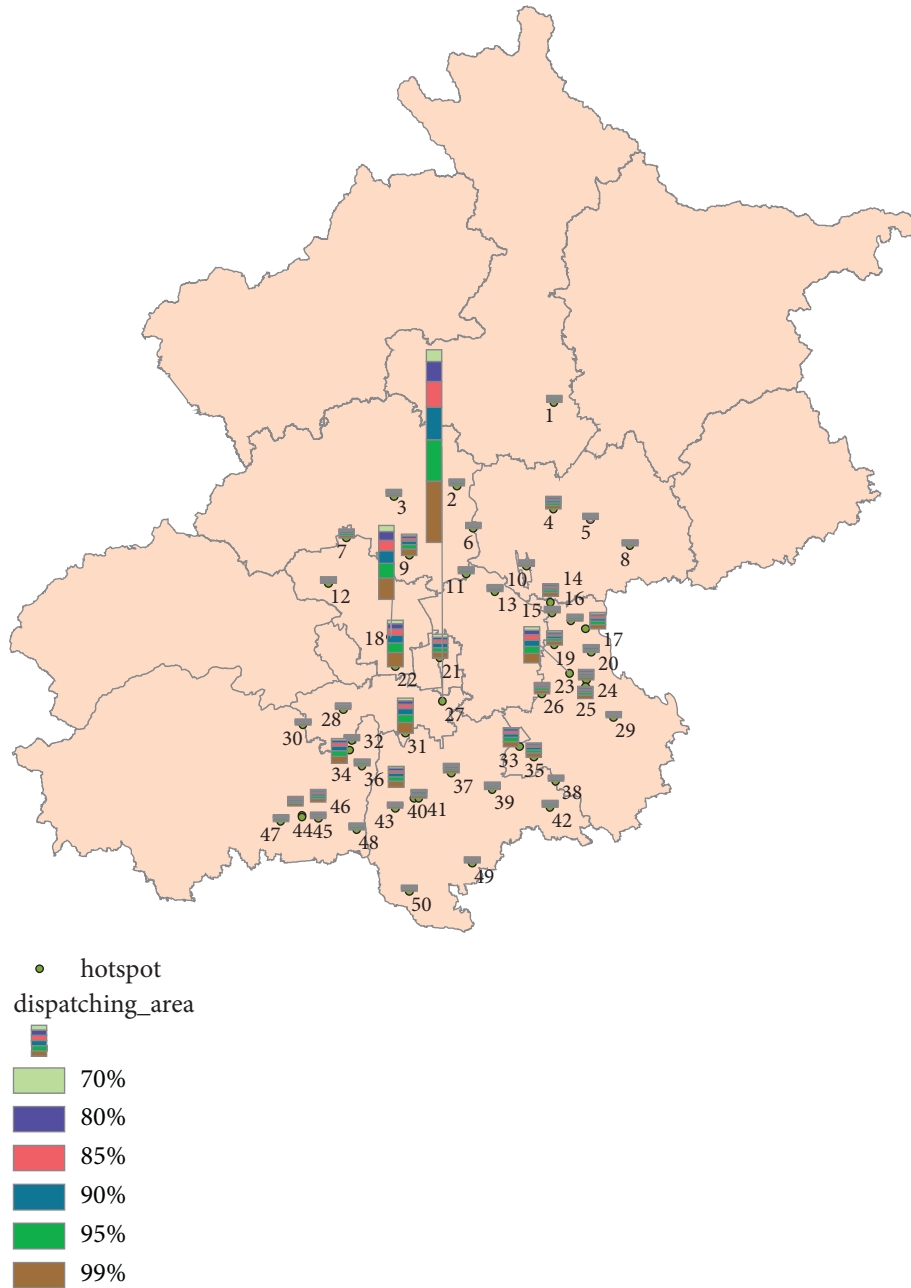


FIGURE 12: Spatial distribution characteristics of dispatch quantity.

all 8334 dispatch areas so as to obtain the dispatch quantities for bike sharing at different reliability degrees. Taking selected dispatch points and dispatch areas (as shown in Figure 9) as examples, Figure 11 shows that the optimal dispatch quantities of shared bikes during the nonearly morning time period in each area are at 70%, 80%, 85%, 90%, 95%, and 99% reliability degrees, respectively.

This article shows the spatial variation pattern of bike-sharing dispatch during the nonearly morning time period with selected 50 dispatch areas as an example. As can be seen in Figure 12, the initial dispatch quantities at the downtown area required to meet the system reliability are larger. Especially, the dispatch quantities are more significant when

consistent with the 99% reliability. Some dispatch areas require more than ten initial dispatch quantities, such as 18th, 22nd, 27th, and 31st, while the initial dispatch quantities required in the outlying district are obviously smaller. Even in some areas, the initial dispatch quantities meeting the reliability requirements are close to zero. These patterns are consistent with the intracity commuting characteristics and provide a reference for the next step of traffic planning and control. In addition, all of the dispatch results and variation patterns described above are derived from data for the nonearly morning time period. However, the usage amount of bike sharing in the early morning time period is very small, and the average hourly usage amount of

each dispatch area is close to 0. This indicates that when the dispatch quantities during the nonearly morning time period meet the reliability requirements of the system, they should also meet the travel demand during the early time morning, so there is no need for further discussion.

#### 4. Discussion and Conclusion

From the perspective of OD differences, this article explored a new model for the bike-sharing dispatch based on the reliability and hypothesis testing, which realized analysis of the distribution characteristics and the establishment of relatively flexible dispatch strategies at different failure probabilities. (1) The article completed dispatch site selection through the hotspot detection method. The kernel density analysis and bike-sharing trajectory data were carried out to explore the optimum dispatch location and establish the dispatch areas with a higher heat value corresponding to the demand of bike-sharing travel, which could be used to dynamically monitor and track the variations of bike-sharing demand temporally and spatially. (2) This article implemented the definition and modeling for the reliability of the bike-sharing dispatch system. Bike sharing was of public nature, whose demand and supply in a single area were regarded as a system. The reliability of the dispatch system was for the first time defined as the ability to satisfy the demand of potential users during ideal periods when shared bikes constantly went in and out of a certain area. The reliability model of the system to meet the travel demand was derived from the performance function of the bike-sharing dispatch based on the OD difference. (3) The probability of bike-sharing usage distribution has been evaluated. A test of fitting goodness based on chi-squared statistics was selected to demonstrate the applicability of some distribution models, including the Poisson, Rayleigh, exponential, normal, and gamma distribution, the principle of which was introduced in this article. Furthermore, the fatal defectives of the Bayesian method were overcome using MLE to avoid information of the prior probability in statistical inference. (4) Beijing was taken as the sample in the case study for methodology verification. The whole day was divided into 24 time windows and 2 study periods. The bike-sharing usage within 168 hours in a week served to analyze probability distribution and hypothesis test. The results showed that both the hourly bike-sharing usage and the OD difference in a single dispatch area conformed to the normal distribution. Additionally, according to the properties and parameter values of the distribution function, the dispatch quantities for bike-sharing within all the 8334 dispatch areas at different confidence levels were assessed based on reliability. Moreover, the spatial variation patterns of dispatch quantities for bike-sharing were consistent with the intracity commuting characteristics, which provided a reference for the next step of traffic planning and control.

Further discussions are still necessary: (1) the dispatch quantity is only demonstrated by the probability distribution of bike-sharing usage. Compared with machine learning models and time series-based track characteristics of bike sharing, this article never gives sufficient consideration to

other spatial geographic and social elements. (2) We mainly focus on the dispatch of bike sharing without the connections to other modes of transportation, such as subway and bus, which have been developing fast in many cities. (3) OD data of bike-sharing within one week is provided as a model for the reliability of matching between dispatch and demand, which lacks considerations in weather and sudden events.

#### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

#### Conflicts of Interest

The authors declare that there are no conflicts of interest.

#### Acknowledgments

The authors also thank TopEdit (<https://www.topeditsci.com>) for its linguistic assistance during the preparation of this manuscript. This work was supported by the National Natural Science Foundation of China under Grant 51478110.

#### References

- [1] A. Soltani, T. Mátrai, R. Camporeale, and A. Allan, "Exploring shared-bike travel patterns using big data: evidence in chicao and budapest," in *Computational Urban Planning and Management for Smart Cities. CUPUM 2019. Lecture Notes in Geoinformation and Cartography*, S. Geertman, Q. Zhan, A. Allan, and C. Pettit, Eds., Springer, Manhattan, NY, USA, 2019.
- [2] S. Shen, C. Lv, X. Xu, and X. Liu, "A bicycle-borne sensor node for monitoring air pollution based on NB-IoT," in *Machine Learning and Intelligent Communications. MLICOM 2019*, X. Zhai, B. Chen, and K. Zhu, Eds., Springer, Manhattan, NY, USA, 2019.
- [3] A. A. Campbell, C. R. Cherry, M. S. Ryerson, and X. Yang, "Factors influencing the choice of shared bicycles and shared electric bikes in Beijing," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 399–414, 2016.
- [4] Q. Zhou, J. Zhou, and C. J. Wong, "A balanced strategy for FFBS operator integrating dispatch area, route and depot based on multi-model technologies," *Journal of Advanced Transportation*, vol. 2021, Article ID 6637251, 16 pages, 2021.
- [5] F. Yang, F. Ding, X. Qu, and B. Ran, "Estimating urban shared-bike trips with location-based social networking data," *Sustainability*, vol. 11, no. 11, p. 3220, 2019.
- [6] R. Regue and W. Recker, "Proactive vehicle routing with inferred demand to solve the bikesharing rebalancing problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 72, no. C, pp. 192–209, 2014.
- [7] L. Mingyuan and W. Guan, "A study of parking space of shared bikes near rail transit stations, Chengdu," *Planners*, vol. 35, no. 18, pp. 54–61, 2019, [http://en.cnki.com.cn/Article\\_en/CJFDTotal-GHSI201918009.htm](http://en.cnki.com.cn/Article_en/CJFDTotal-GHSI201918009.htm).
- [8] L. Caggiani, R. Camporeale, M. Ottomanelli, and W. Y. Szeto, "A modeling framework for the dynamic management of free-floating bike-sharing systems," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 159–182, 2018.

- [9] W. Xia, S. Shu-Hua, T. Jun, Y.-G. Liu, and Y. University, "Rationality of shared bicycles parking location based on hybrid cluster analysis," *Digital Technology Applied*, 2019.
- [10] L. Gao, Y. Ji, X. Yan, Y. Fan, and W. Guo, "Incentive measures to avoid the illegal parking of dockless shared bikes: the relationships among incentive forms, intensity and policy compliance," *Transportation*, vol. 48, pp. 1–28, 2020.
- [11] Y. Luo, W. Dou, H. Yan, L. Liu, and S. Lu, "An automated planning and dispatch method of shared bikes based on reward and punishment mechanism," in *Proceedings of the 2018 17th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, Wuxi, China, October 2018.
- [12] J. Zhang, M. Meng, and D. Wang, "A dynamic pricing scheme with negative prices in dockless bike sharing systems," *Transportation Research Part B: Methodological*, vol. 127, 2019.
- [13] A. Pal and Y. Zhang, "Free-floating bike sharing: solving real-life large-scale static rebalancing problems," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 92–116, 2017.
- [14] S. C. Ho and W. Y. Szeto, "Solving a static repositioning problem in bike-sharing systems using iterated tabu search," *Transportation Research Part E: Logistics and Transportation Review*, vol. 69, pp. 180–198, 2014.
- [15] I. A. Forma, T. Raviv, and M. Tzur, "A 3-step math heuristic for the static repositioning problem in bike-sharing systems," *Transportation Research Part B: Methodological*, vol. 71, pp. 230–247, 2015.
- [16] C. Yuan, Y. Sun, J. Lv, and A. Lusk, "Cycle tracks and parking environments in China: learning from college students at Peking university," *International Journal of Environmental Research and Public Health*, vol. 14, no. 8, p. 930, 2017.
- [17] D. Luo, O. Cats, H. van Lint, and G. Currie, "Integrating network science and public transport accessibility analysis for comparative assessment," *Journal of Transport Geography*, vol. 80, Article ID 102505, 2019.
- [18] M. Song, K. Wang, Y. Zhang, M. Li, H. Qi, and Y. Zhang, "Impact evaluation of bike-sharing on bicycling accessibility," *Sustainability*, vol. 12, 2020.
- [19] L. J. Jia and C. W. Ting, "Bike-sharing systems and destination satisfaction in overseas tourists," *Asian Transport Studies*, vol. 5, 2018.
- [20] J. Bao, H. Yu, and J. Wu, "Short-term FFBS demand prediction with multi-source data in a hybrid deep learning framework," *IET Intelligent Transport Systems*, vol. 13, no. 9, pp. 1340–1347, 2019.
- [21] K. Kim, "Investigation on the effects of weather and calendar events on bike-sharing according to the trip patterns of bike rentals of stations," *Journal of Transport Geography*, vol. 66, pp. 309–320, 2018.
- [22] J. Zhang, X. Pan, M. Li, and P. S. Yu, "Bicycle-sharing system Analysis and trip prediction," in *Proceedings of the 2016 17th IEEE International Conference on Mobile Data Management (MDM)*, pp. 174–179, Porto, Portugal, June 2016.
- [23] C. Rudloff and B. Lackner, "Modeling demand for bikesharing systems," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2430, no. 1, pp. 1–11, 2014.
- [24] S. E. Wilson, A. Bunko, S. Johnson, J. Murray, and Y. Li, "The geographic distribution of un-immunized children in Ontario, Canada: hotspot detection using Bayesian spatial analysis," *Vaccine*, vol. 39, no. 1, 2021.
- [25] Y. Zhang, J. Min, C. Liu, and Y. Li, "Hotspot detection and spatiotemporal evolution of catering service grade in mountainous cities from the perspective of geo-information tupu," *International Journal of Geo-Information*, vol. 10, no. 287, p. 287, 2021.
- [26] W. Yu, "Discovering frequent movement paths from taxi trajectory data using spatially embedded networks and association rules," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 855–866, 2019.
- [27] R. Bandyopadhyaya and S. Mitra, "Fuzzy cluster-based method of hotspot detection with limited information," *Journal of Transportation Safety & Security*, vol. 7, no. 4, pp. 307–323, 2015.
- [28] A. Gramacki, "FFT-based algorithms for kernel density estimation and bandwidth selection," *[Studies in Big Data] Nonparametric Kernel Density Estimation and its Computational Aspects*, Studies in Big Data, 2018.
- [29] J. Aitchison and I. J. Lauder, "Kernel density estimation for compositional data," *Journal of Applied Statistics*, vol. 34, no. 2, pp. 129–137, 2018.
- [30] R. Cervero, *Transit Metropolis*, Island Press, Washington, D.C., USA, 1998.
- [31] P. E. Shrout and J. L. Fleiss, "Intraclass correlations: uses in assessing rater reliability," *Psychological Bulletin*, vol. 86, no. 2, pp. 420–428, 1979.
- [32] P. L. Gatti, *Probability Theory and Mathematical Statistics for Engineers*, Elsevier Science, Amsterdam, Netherlands, 1984.