

# Demand Prediction and Optimal Allocation of Shared Bikes Around Urban Rail Transit Stations

Liang Yu<sup>1</sup> · Tao Feng<sup>2</sup>  · Tie Li<sup>3</sup> · Lei Cheng<sup>1</sup>

Received: 30 August 2022 / Revised: 9 September 2022 / Accepted: 26 October 2022 / Published online: 13 December 2022  
© The Author(s) 2022

**Abstract** The imbalance between the supply and demand of shared bikes is prominent in many urban rail transit stations, which urgently requires an efficient vehicle deployment strategy. In this paper, we propose an integrated model to optimize the deployment of shared bikes around urban rail transit stations, incorporating a seasonal autoregressive integrated moving average with long short-term memory (SARIMA-LSTM) hybrid model that is used to predict the heterogeneous demand for shared bikes in space and time. The shared bike deployment strategy was formulated based on the actual deployment process and under the principle of cost minimization involving labor and transportation. The model is applied using the big data of shared bikes in Xicheng District, Beijing. Results show

that the SARIMA-LSTM hybrid model has great advantages in predicting the demand for shared bikes. The proposed allocation strategy provides a new way to solve the imbalance challenge between the supply and demand of shared bikes and contributes to the development of a sustainable transportation system.

**Keywords** Bike sharing · Demand prediction · Bike-sharing deployment · Hybrid model · Big data

## 1 Introduction

Bike sharing has become a popular transport option in many cities because it is green and convenient and an essential feeder for other transportation modes such as buses and metro. In the emerging trend of mobility-as-a-service (MaaS) system deployment, which involves multimodal transportation, bike sharing may be promoted even further. As an important transport connector, shared bikes could increase the attractiveness of urban transit and thus potentially encourage transit use and contribute to sustainable transportation. Bicycle sharing complements and enhances public transportation [1]. As manifested in many cities, shared bikes are often used in combination with transit, shortening the travel time induced by the access/egress stages of a complete journey by transit. Especially for urban rail transportation, bike sharing has become the leading way to connect the last mile of rail transportation [2].

Shared bikes are traditionally dock-based because of the management and payment requirement and have also recently emerged as dockless. While the latter is considered more convenient because of the parking convenience, many problems remain, such as disordered parking,

---

✉ Tao Feng  
taofeng@hiroshima-u.ac.jp

Liang Yu  
20121214@bjtu.edu.cn

Tie Li  
120821904@qq.com

Lei Cheng  
18121186@bjtu.edu.cn

<sup>1</sup> School of Civil Engineering, Beijing Jiaotong University, No.3 Shangyuncun, Haidian District, Beijing 100044, People's Republic of China

<sup>2</sup> Urban and Data Science, Graduate School of Advanced Science and Engineering, Hiroshima University, 1-5-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8529, Japan

<sup>3</sup> Beijing Metro Network Administration Co.,Ltd, Beijing Jingtou Building, No.6 Xiaoying North Road, Chaoyang District, Beijing, People's Republic of China

abandoned bikes, and illegal space occupation. Rail transit stations, as hot spots for shared bicycles, make this contradiction unprecedentedly acute. On the one hand, the excessive influx of shared bicycles is not conducive to the organization and evacuation of passengers. On the other hand, insufficient bicycle storage will lead to passengers being forced to choose other ways to complete their journeys, reducing passenger comfort and leading to a decrease in traffic accessibility [3]. Both stations and bike-sharing operators have made efforts to this end. For example, planners can enhance resilience of urban transport networks by fully considering the capacity and the usage of bike-sharing docks, as well as the coherence of metro stations and bike-sharing docks, in distributing and rebalancing activities [4]. Some scholars have also proposed the joint operation of bike sharing and metro mainly to serve rail stations, and the results show that the total social cost can be reduced to a great extent [5]. However, the imbalance issue between demand and supply in space and time underlying the use of shared bikes is still prominent in the sense that the use rate of shared bikes at a specific location could be highly time-dependent, e.g., commuters pick up bikes in the morning peak, and varied in the different built environment around the station, e.g., central business districts (CBD) and residential neighborhoods. Therefore, it remains critical for shared bike operators and urban management bureaus to better coordinate the deployment of shared bikes in different places and at different times of the day.

Many studies have addressed the issues related to shared bikes from different perspectives, which provides a theoretical basis for the better planning of shared bikes around urban rail transit stations. Five main aspects were discussed in the literature, namely, the factors affecting the use of shared bikes [6], the characteristics of shared bike users [7], the spatial and temporal distribution of bike use [8, 9], the demand prediction of shared bikes [10, 11], the allocation of shared bikes [12], and the integration of shared bikes with transit [13]. These studies offer valuable insights into shared bikes planning, but there are still various remaining problems, especially in the demand forecasting and allocation strategy.

The prediction of demand for shared bikes is essential for planning bike-sharing deployment, because the optimal allocation strategy can only be obtained when the predicted bike-sharing demand is close to the actual deployment needs. Existing research has developed different models by focusing on either the demand prediction or allocation strategy using actual data. There is a lack of integrated models which combine the demand prediction and bike allocation. We argue that the separate models may be more useful for a specific purpose. In contrast, models integrating the two are more useful to effectively realize the

balance between supply and demand and have more practical significance.

In this study, therefore, we propose an improved bike-sharing allocation strategy to optimize the bike allocation around urban rail transit stations. A hybrid model based on time series and deep learning prediction theory is proposed, combining a seasonal autoregressive integrated moving average (SARIMA) model and a long short-term memory (LSTM) model to predict the usage of shared bikes around urban rail transit stations. Based on production and sales balance theory, a model to optimize bike allocation while minimizing the total dispatch cost is developed. Using 3-month big data of shared bikes in Beijing, the dynamic deployment strategy around urban rail transit stations is obtained by combining the actual dispatch process for shared bikes to meet the demand for shared bicycles during peak periods.

The remainder of this paper is organized as follows: Section 2 reviews the related literature on bike-sharing demand forecasting and deployment planning. Section 3 introduces the methodology employed in this study. Section 4 describes a case study of bike sharing in Beijing. The last section concludes this paper.

## 2 Literature Review

The purpose of the literature review is twofold: to discuss the relevant work on demand prediction and the allocation of shared bikes. These two aspects of research have been primarily conducted separately in the existing literature. The purpose of this review is not to offer a systematic review of the research on shared bikes but rather to present the content-wise research and the methods related to the two perspectives.

The bike-sharing demand prediction has been discussed in different spatial scales, namely the city level, station clustering level, and single station level. City-level forecasting aims to predict coarser-grained bike usage at all stations throughout the city. Borgnat et al. [14], based on the data generated by the bike-sharing system in Leon City, developed a combination model to predict the hourly demand for shared bikes within a city. The results help provide an understanding of the travel characteristics of bike-sharing users. Based on the assumption that geographically adjacent stations have similar characteristics of time demand, Chen et al. [15] and Feng et al. [16] classified stations into different clusters and predicted station-based demand for the different categories. Lin et al. [17] and Wang et al. [18] showed that, although the prediction with station clustering can better capture the local demand than city-level results, the local differences between different bike-sharing stations are not addressed.

In the case of the station-based prediction models, Sohrabi et al. [19] proposed a two-step prediction method based on historical traffic and spatiotemporal characteristics to predict the demand at bike-sharing stations dynamically. Wu et al. [20] found that the current site-level bike-sharing demand prediction did not fully mine the information in the ordering data and ignored the potential relationship between different sites/stations. Machine learning models were used to predict the time-by-time demand for shared bikes in different sites.

Among the studies on demand prediction, time-series regression models have most typically been applied. Kaltenbrunner et al. [21] used the classic autoregressive moving average model (ARMA) to predict the number of shared bikes. Realizing that the ARMA model requires stationarity of the time series, Yoon et al. [22] used the autoregressive integrated moving average (ARIMA) model to predict the demand of each station for shared bikes, such that the stationarity problem of the time series in the number of shared bikes is solved by setting the time cycle. Apart from the statistical models, recent efforts have also involved the use of machine learning methods. For example, Liu et al. [23] developed a demand prediction model for shared bikes using an artificial neural network.

In addition, scholars have increasingly begun to combine different types of methods to predict the demand for shared bicycles. Mehdizadeh et al. [24] used a hybrid convolutional neural network (CNN)-LSTM model to predict bicycle demand during the COVID-19 pandemic, and the results showed higher accuracy than ARIMA. A model combining LSTM and gated recurrent units (GRU) was proposed by Boonjubut et al. [25] to predict the demand for shared bicycles, which can improve the effectiveness and accuracy of a single recurrent neural network (RNN) model. Similarly, Ma et al. [26] proposed a spatial-temporal graph attentional (STGA)-LSTM neural network framework to predict short-term bike-sharing demand using a multi-source data set. Most of these models combine LSTM, improve the prediction accuracy of a single model to some extent, and provide a new method for demand prediction of shared bikes. However, for strongly periodic demand forecasting, there are still inadequate studies in giving full play to the performance of LSTM.

Based on the predicted demand for shared bikes at different locations, optimal bike deployment has been an important subject for operators to balance the demand and supply in space and time. Developing an effective vehicle deployment planning model has been discussed extensively by scholars. For instance, Forma et al. [27] proposed a bike static reset deployment method and designed a three-step heuristic algorithm to make real-time responses for bike demand and vacant piles at sites. Angeloudis et al. [28] redistributed the allocation lines by developing a dynamic

bike-sharing allocation algorithm and demonstrated the model's effectiveness through numerical examples. Jost et al. [29] proposed a dynamic pricing strategy to improve the imbalance between the supply and demand of shared bikes based on the idea of self-distribution balance in public bike spaces. However, results showed that although the pricing strategy could improve the imbalance issue, it would reduce travelers' willingness to use shared bikes. Jiménez et al. [30] combined the real-time changes in shared bike demand by optimizing the deployment paths. A station-based hybrid deployment model was proposed to effectively reduce the deployment cost. The optimal allocation path based on the minimum cost has been a major focus in many studies. Furthermore, Raviv et al. [31] focused on the stochastic and dynamic changes in public bike demand to propose an optimization model based on the static bike relocation problem. The model uses a time-dependent objective function to optimize vehicle paths, which can effectively assist path and inventory decisions.

Reviewing the literature on the allocation strategy of shared bikes shows that most of the existing allocation schemes are based on simple manual allocation, and the allocation schemes do not cater to the hotspot usage areas of shared bicycles from a station perspective. In other words, they did not consider the demand for shared bikes from stations at different times, resulting in more significant problems in the management and operation of shared bikes in the surrounding rail transit stations. On the other hand, the existing prediction models have been developed independently, and their contribution to the optimal allocation of shared bikes has not been examined. Given this shortcoming, we propose an optimal shared bike allocation model incorporating the demand heterogeneity of shared bikes. More specifically, we propose a hybrid model to predict the demand for shared bikes by giving full play to the advantages of SARIMA and LSTM models and a deployment model to meet station-wide bike-sharing demand under the premise of cost minimization.

## 3 Methodology

### 3.1 Bike-Sharing Demand Forecasting Model

The bike-sharing order data has an obvious time pattern, and its generation and ending have clear time points, so the time-by-time order data derived from the statistics belong to the classical time-series data. Classical models in time-series research include ARMA, ARIMA, and SARIMA. These regression-based models can effectively deal with the linear part of time-series data but are not sensitive to the nonlinear characteristics.

In order to overcome the disadvantage of linear models in prediction with nonlinear features, recent studies have increasingly applied machine learning methods that are considered to effectively deal with the nonlinear association in the data, including traditional neural networks, deep learning, models such as recurrent neural network and LSTM, to enhance the model predictability. However, these machine learning methods depend highly on data and model selection. The results obtained also lack a certain degree of interpretability. Thus, recent activities regarding the synergy between theory-driven and data-driven methods have emerged [32].

In this study, we contend that the time-dependencies in the time-series data can be best captured using the time-series data-based models, while the residual (random component) in the linear model may be best captured with a highly nonlinear structure in deep learning. In other words, we propose a model that integrates SARIMA and LSTM models, a SARIMA-LSTM hybrid model, such that the ridership of shared bikes can be predicted using advanced machine learning models while simultaneously maintaining the statistical features of the time-series data. In the following sections, we will first present the prediction models such as the SARIMA and LSTM models and focus on the hybrid version of the two models. Because the prediction model generates input data to optimize bike deployment, we also present the optimization model and explain how the two models are integrated.

### 3.1.1 SARIMA Model

SARIMA is an autoregressive model for seasonal sequences developed based on the ARIMA model. By decomposing the time series into a regular time series and seasonal cycles, the model can eliminate the interference of periodic changes and improve its prediction accuracy. The model is generally denoted as SARIMA( $p, d, q$ )( $P, D, Q$ ) $s$  and the expression is shown in Eq. (1).

$$\phi_p(K)\varphi_P(K)(1-K)^d(1-K^s)^D Y_t = \theta_q(K)\omega_Q(K)\varepsilon_t \quad (1)$$

where  $Y_t$  is the time series;  $K$  represents the lag operator, which serves to regress the sequence by one or more cycles;  $s$  represents cycles, such as seasonal cycles  $s = 4$ , monthly cycles  $s = 12$ , and so on;  $\varepsilon_t$  is the white noise sequence conforming to normal distribution;  $d$  and  $D$  represent the number of nonseasonal and seasonal differences, respectively;  $\phi_p(K)$  and  $\varphi_P(K)$  are functions describing the relationship between adjacent moments of time series  $Y_t$ , as shown in Eqs. (2) and (3), which  $p$  and  $P$  mean the order of the nonseasonal and seasonal autoregressive terms, respectively.

$$\phi_p(K) = 1 - \alpha_1 K - \alpha_2 K^2 - \dots - \alpha_p K^p \quad (2)$$

$$\varphi_P(K) = 1 - \beta_1 K^s - \beta_2 K^{2s} - \dots - \beta_P K^{Ps} \quad (3)$$

where  $\theta_q(K)$  and  $\omega_Q(K)$  are relational functions describing the seasonal period of time series  $Y_t$ . The expressions are shown in Eqs. (4) and (5), where  $q$  and  $Q$  represent the nonseasonal and seasonal moving average term orders, respectively.

$$\theta_q(K) = 1 + \lambda_1 K + \lambda_2 K^2 + \dots + \lambda_q K^q \quad (4)$$

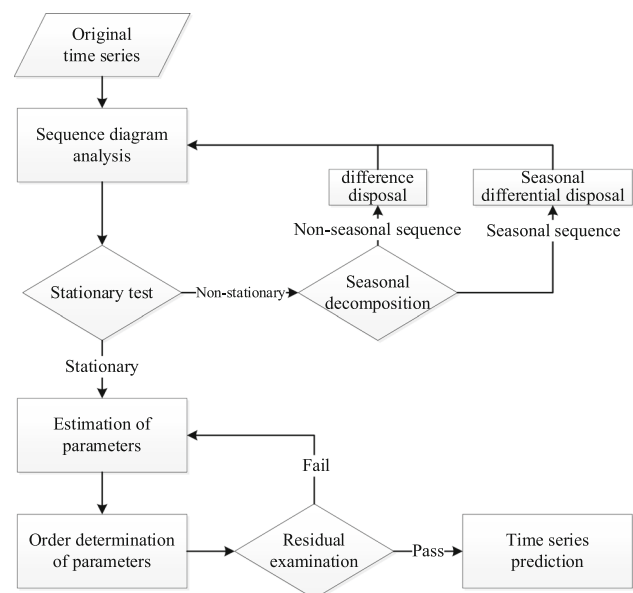
$$\omega_Q(K) = 1 + \lambda_1 K + \lambda_2 K^{2s} + \dots + \lambda_Q K^{Qs} \quad (5)$$

The process of implementing a SARIMA model is depicted in Fig. 1. Firstly, the time-series data are analyzed by time-series plot to determine the nonstationary series. Then the nonstationary series is differenced and seasonally differenced to obtain the minimum values of  $d$  and  $D$ . Next, the model' parameters are estimated and determined, and the residuals of the model are tested. Finally, numerical predictions are made by the eligible SARIMA( $p, d, q$ )( $P, D, Q$ ) $s$ .

### 3.1.2 LSTM Model

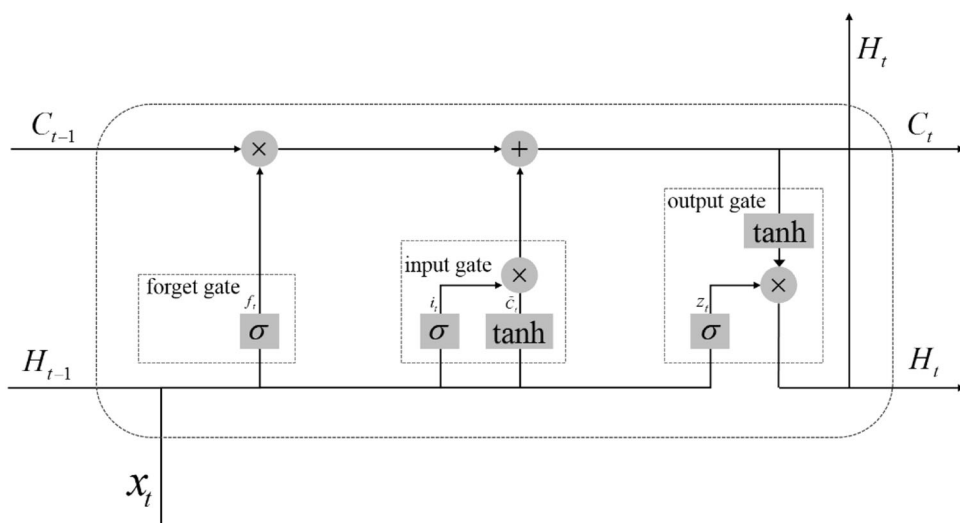
LSTM is used to solve the long-term dependency and gradient problems, and its basic unit is shown in Fig. 2.

According to Fig. 2, LSTM mainly consists of three gates: input gate, forget gate and output gate. The process of LSTM is to pass the input sequence  $\{x_1, x_2, \dots, x_t\}$  through an encoder-decoder and finally output the predicted sequence  $\{y_1, y_2, \dots, y_t\}$ .  $x_t$ ,  $C_t$  and  $H_t$  means the



**Fig. 1** The implementation process of the SARIMA model

**Fig. 2** LSTM standard cell structure diagram



input value, the regenerative cell state value and the output cell hidden state value at time  $t$ , respectively.

First, the forget gate is passed, and the forget gate operation expression is shown in Eq. (6), where  $H_{t-1}$  is the hidden state value of the cell at the previous moment and  $x_t$  is the input value at this moment, and  $\sigma$  function is used to output a weight  $f_t$  between 0 and 1, which is taken as the probability value of the node state at the time before the neural unit forgets.

$$f_t = \sigma(W_f[H_{t-1}, x_t] + b_f) \tag{6}$$

In the input gate, the state information of the candidate cell  $\tilde{C}_t$  can be generated from the  $\tanh$  function, and  $\sigma$  function generates a weight  $i_t$  to determine the size of the state information of the candidate cell.

$$\tilde{C}_t = \tanh(W_c[H_{t-1}, x_t] + b_c) \tag{7}$$

$$i_t = \sigma(W_i[H_{t-1}, x_t] + b_i) \tag{8}$$

Combining with the cell state at the previous moment  $C_{t-1}$ , the regenerative cell state at this moment  $C_t$  is shown in Eq. (9).

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{9}$$

Finally, the value of the output gate  $z_t$  can be calculated from Eq. (10), and the current hidden state  $H_t$  can be calculated by combining the current state  $C_t$  with Eq. (11). By processing  $H_t$  through the decoder, the predicted sequence  $\{y_1, y_2, \dots, y_t\}$  is obtained.

$$z_t = \sigma(W_z[H_{t-1}, x_t] + b_z) \tag{10}$$

$$H_t = z_t \cdot \tanh(C_t) \tag{11}$$

In the above equations,  $W_f, W_c, W_i, W_z$  are the weight matrixes,  $b_f, b_c, b_i, b_z$  are bias vectors. These are exactly the eight sets of parameters that machine learning is designed to learn to obtain better predictions.

### 3.1.3 SARIMA-LSTM Hybrid Model

The SARIMA-LSTM hybrid model combines the time-series prediction model and the neural network prediction model to predict the time-series volume of shared bikes,  $\{Y_t\} (t = 1, 2, 3, \dots, n)$ . According to the characteristics of different models,  $\{Y_t\}$  can be divided into linear structure  $\{L_t\}$  and nonlinear structure  $\{e_t\}$ , which are expressed in Eq. (12).

$$Y_t = L_t + e_t \tag{12}$$

Because the hybrid model involves parameter estimation of SARIMA and the training of LSTM, a joint model estimation process is needed. The structure diagram of the hybrid model is shown in Fig. 3, and the three core steps are described below.

- (1) Extract linear features of the data.  
The data set was divided into a training set and a test set, which were built with a SARIMA model using SPSS. Then the linear fitting value  $\hat{L}_t$  of shared bike time-series data and linear value  $\hat{L}_{t+1}, \hat{L}_{t+2}, \dots, \hat{L}_{t+n}$  of future period  $n$  are obtained by the SARIMA prediction.

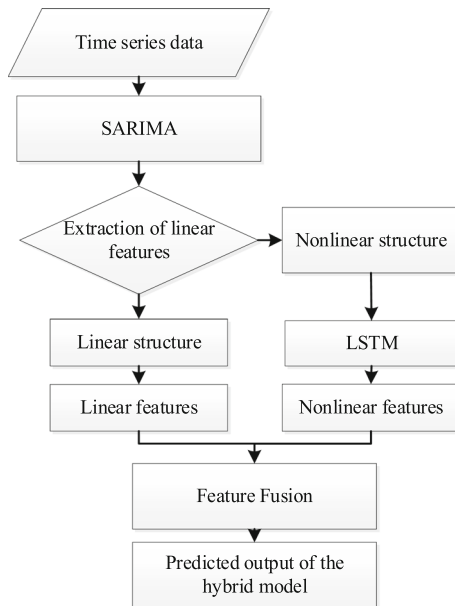
- (2) Analyze the nonlinear characteristics of data.

After SARIMA modeling and prediction, the residual sequence reflecting the nonlinear structure of the data can be obtained, as shown in Eq. (13).

$$e_t = Y_t - \hat{L}_t \tag{13}$$

Then, the Keras deep learning library in Python and the TensorFlow package are used to build the LSTM network model. The residual sequence is used as the input value for training. After the parameters are stabilized, the model is





**Fig. 3** Structure of hybrid prediction model.

used to predict the nonlinear value  $\hat{e}_{t+1}, \hat{e}_{t+2}, \dots, \hat{e}_{t+n}$  in future period  $n$ .

(3) The predicted value of future period  $n$  is obtained by using the hybrid model prediction.

$$\hat{Y}_{t+i} = \hat{L}_{t+i} + \hat{e}_{t+i} \quad (i = 1, 2, \dots, n) \quad (14)$$

### 3.1.4 Evaluation Metrics

To assess the predictive performance of the models, mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) are adopted. MAE and RMSE reflect the absolute error in the predicted and actual values, while MAPE reflects the relative percentage error. Smaller values of MAE and RMSE indicate a smaller fitting deviation of the model, thus better accuracy in prediction. In MAPE cases, a value lower than 10% has been considered empirically to be satisfactory. The formulas of the three indicators are shown below:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (16)$$

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \times 100 \right| \quad (17)$$

where  $\hat{y}_t$  represents the predicted value of the model,  $y_t$  represents the actual value, and  $n$  represents the number of samples.

## 3.2 Bike-Sharing Allocation Strategy

By physically moving the bikes from one location to another, the demand in target areas can be matched, while the over-allocation in original areas can be released. Before redeployment begins, a redeployment plan will be developed to achieve a better balance. The plan usually includes transportation volumes and routes, with transportation costs being the essential basis for its development.

### 3.2.1 Bike-Sharing Allocation Process

The deployment of bike sharing is a system project. Preparing a deployment plan based on order data also requires the full cooperation of deployment command points, deployment locations, and deployment experts. In order to standardize the operational process of deployment, we summarize the management methods of different companies deploying shared bikes and propose the following deployment strategies.

- (1) After the morning peak, the enterprise allocation center calculates the inflow and outflow of shared bikes at various rail transit stations and deployment sites based on the order data uploaded to the system.
- (2) According to the combined prediction model, input the inflow and outflow of shared bikes in the morning peak hours, predict the inflow and outflow of shared bikes in the peak hours and evening peak hours of all stations, and determine the export and import volume of the target area around the urban rail transit station and the upper limit of the number of shared bikes that can be accommodated and supplemented by the allocation station.
- (3) Aiming at the company's minimum deployment cost, establish a shared bike deployment model based on the transportation problem of production and sales balance, and solve a specific shared bike allocation plan.

The allocation center sends the allocation plan to the specialists around each site. The allocation specialists complete the allocation of shared bikes in the allocation skylight according to the allocation instructions and report them in time.

### 3.2.2 Deployment Areas

Analysis of the allocation strategy shows that the division of bike-sharing allocation areas is the basis for allocating work. In other words, we need to know which station needs to move bikes out and which one needs to replenish before we can optimize.

As there are no fixed stations for bike sharing, users can rely on intelligent location devices such as onboard GPS to park and pick up their bikes anytime. That poses a challenge for deployment. The shared bike drop-off area is divided into different deployment units to simplify the deployment of shared bikes. In the actual scheduling process, the deployment center often determines the areas to be assigned based on historical experience, big data, analysis and dispatcher’s instructions—in other words, without long transport distances for cost savings in the deployment. According to the analysis of shared bike data, 77.4% of shared bikes are manually assigned a straight-line distance of 3 km or less. Therefore, the straight-line distance between stations within the setup dispatch unit does not exceed 3 km.

Based on the deployment units, the deployment sites of shared bicycles can be further divided. The areas around the urban rail transit station were considered, in general, a hot spot for the use of shared bikes; another study showed that shared bikes going to or leaving a station were distributed within 2 km of the station [33].

To disperse or supplement shared bicycles around stations, it is necessary to set up additional deployment sites. We filtered the latitude and longitude at the starting and ending points of bike-sharing orders within 2 km of all stations. We used a *k*-means clustering algorithm to identify hot spots of bike-sharing usage. A 500-meter radius of the cluster center was used as the deployment site for the shared bikes.

### 3.2.3 Amount of Demand for Deployment

Here, we analyzed the number of shared bikes used around stations (Fig. 4).

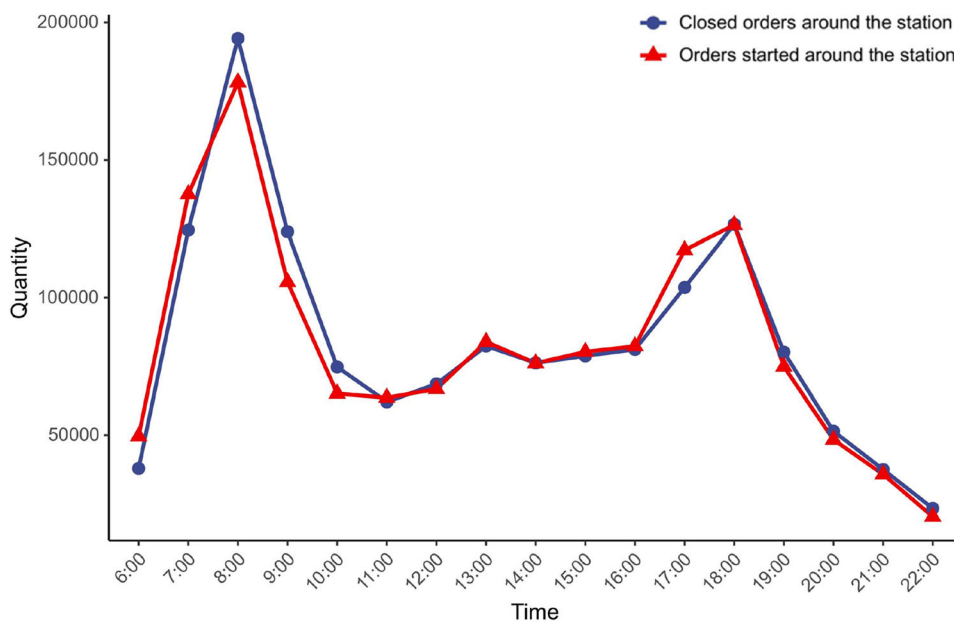
Figure 4 shows that there are two distinct peaks in the volume of bike-sharing usage around rail stations. After calculation, the number of bicycles used during the peak period accounts for 40% of the whole day. Combined with the actual operation strategies of the operating companies, dynamic deployment with a high degree of refinement would result in excessive operating costs. Therefore, this paper still adopts the deployment scheme based on the skylight time and determines the skylight time as from 9:00 to 17:00.

The allocation demand in each area is calculated according to the flow difference, which refers to the difference between the inflow and outflow of regional shared bikes in the same period. The deployment requirement for each place can be calculated by Eq. (18).

$$\Delta D_k = \sum_t (\text{Flow\_in}_t - \text{Flow\_out}_t) \tag{18}$$

The inflow and outflow of shared bikes are predicted according to the SARIMA-LSTM combination model. Then, we calculated the allocation demand in the stations and the allocation site areas, respectively. For the urban rail transit stations, when  $\Delta D_k > 0$ , the shared bikes flow into this station, and the shared bikes need to be transferred out to the surrounding allocation stations, which is called the export urban rail station; when  $\Delta D_k < 0$ , the shared bikes flow out of this station, and the shared bikes need to be

**Fig. 4** Usage of shared bicycles around urban rail transit stations



transferred into this station from the surrounding allocation stations, which is called the import urban rail station.

For the deployment sites, when  $\Delta D_k > 0$ , the continuous inflow of shared bikes in this station can provide a certain number of shared bikes for outflow urban rail transit stations with insufficient shared bikes nearby, which are supplementary deployment sites; when  $\Delta D_k < 0$ , the shared bikes in this site continue to flow out, so there is a certain capacity for shared bikes, which can accommodate the excess shared bikes of nearby inflow urban rail transit stations, which are the accommodation deployment sites.

### 3.2.4 Bike-Sharing Allocation Model

By physically moving the bikes from one location to another, the demand in target areas can be matched while the over-allocation in original areas can be released. Before redeployment begins, a redeployment plan will be developed to achieve a better balance. The plan usually includes transportation volumes and routes, with transportation costs being the essential basis for its development. The allocation task is to allocate the shared bikes to be transferred from all export urban rail stations to import urban rail stations and accommodation deployment sites. At the same time, the shared bikes required by all import urban rail stations can be transferred in from export urban rail stations and supplementary deployment sites. The critical point is that we need to minimize the costs incurred in shipping. Therefore, based on the balanced transportation model, we allocate shared bikes with the minimized cost as the objective function, as shown in Eq. (19).

$$\min \left( \sum_{i=1}^m \sum_{j=1}^n \gamma_{ij} + \sum_{i=1}^m \sum_{j=1}^N \gamma_{ij} + \sum_{i=1}^M \sum_{j=1}^n \gamma_{ij} \right)$$

$$s.t. \begin{cases} \sum_{j=1}^n x_{ij} = a_i, i = 1, 2, 3, \dots, m & (1) \\ \sum_{j=1}^n x_{ij} \leq p_i, i = 1, 2, 3, \dots, M & (2) \\ \sum_{i=1}^m x_{ij} = b_j, j = 1, 2, 3, \dots, n & (3) \\ \sum_{i=1}^m x_{ij} \leq q_j, j = 1, 2, 3, \dots, N & (4) \\ x_{ij} = 0, i = 1, 2, 3, \dots, M; j = 1, 2, 3, \dots, N & (5) \\ x_{ij} \geq 0, i, j \text{ are not all deployment stations} & (6) \end{cases} \quad (19)$$

The definitions of the parameters in the formula are shown in Table 1.

The effect of constraints ① and ③ is to transport the excess bikes to the stations that need resupply. Constraint conditions ② and ④ mean that the number of shared bikes allocated at the supplementary and accommodation allocation stations should not exceed its upper limit. The last two constraints mean that the allocation of shared bikes cannot occur between the allocation stations.

The costs incurred for deployment in this paper consist of labor and transportation costs, which are calculated in Eqs. (20, 21, 22 and 23). The definitions of the parameters in the formula are shown in Table 2.

$$\gamma_{ij} = \gamma_{H,ij} + \gamma_{T,ij} \quad (20)$$

$$\gamma_{H,ij} = x_{ij} \cdot L_{ij} \cdot v \quad (21)$$

$$v = \frac{\text{monthly salary}}{\text{distance} \cdot \text{number of bikes}} \quad (22)$$

$$\gamma_{T,ij} = 2 \frac{x_{ij}}{n} s \cdot L_{ij} \quad (23)$$

The model can be solved by the table on the operating method. We implemented such an optimization algorithm via MATLAB, in which the initial feasible solution uses the Vogel method, and the optimal solution is determined by closed-loop adjustment.

## 4 Empirical Results

In this section, we introduce the empirical analysis conducted based on the data in Beijing, the capital of China. First, we describe the data and fields being analyzed. Then the effectiveness of the hybrid model is verified based on the data. Taking the Xicheng District of Beijing as the basic dispatching unit, the allocation process for shared bikes around urban rail transit stations within the dispatching unit is elaborated in detail, and the deployment scheme for shared bikes around urban rail transit stations in the Xicheng District is formulated to verify the applicability of the proposed model.

### 4.1 Data and Research Area

The Xicheng District is located in the west of the central city area. Some stations have a higher number of shared bikes than other areas during morning and evening peak hours, which is more likely to have the imbalance problem between supply and demand of shared bikes. The  $k$ -means was used to filter the possible deployment sites in the Xicheng District. According to the silhouette coefficient and within-groups sum-of-squares errors (SSE), the clustering results show the best performance when  $k = 932$ . The results show that there are 22 known urban rail transit



**Table 1** Model parameters and definitions

Parameters	Definition
$\gamma_{ij}$	The deployment cost from station $i$ to $j$
$x_{ij}$	the volume of bikes transported from $i$ to $j$ , and $x_{ij}$ is an integer
$m$	The number of export urban rail station
$M$	The number of supplementary deployment sites
$n$	The number of import urban rail station
$N$	The number of accommodation deployment sites
$a_i$	The number of bikes that need to be called out for the $i$ th exit urban rail station
$b_i$	Demand quantity for bicycles of the $j$ th import urban rail station
$p_i$	The number of bikes that need to be called out for the $i$ th supplementary deployment sites
$q_i$	Demand quantity for bicycles of the $j$ th accommodation deployment sites

**Table 2** Parameters and definitions of deployment costs

Parameters	Definition
$\gamma_{H,ij}$	The allocated labor cost, which is calculated based on interviews with staff
$L_{ij}$	The shortest road network distance from location $i$ to location $j$ , which is obtained by Baidu Maps API
$v$	The labor cost per km of shared bikes allocation, in this paper $v = 1$ according to our interview results
$\gamma_{T,ij}$	The allocated transportation cost, which is mainly influenced by the fuel consumption and maintenance cost of the transport vehicle
$n$	The maximum number of shared bikes that can be accommodated by each transport vehicle, in this paper, $n = 60$
$s$	The driving cost per kilometer of the transport vehicle, we supposed $s = 2$ yuan/km according to our interview results

stations and 28 qualified allocation sites were selected for deployment.

The primary data set used in this study is the payment transaction records of Mobike in Beijing from April to June 2018. Mobike is one of the first bike-sharing enterprises to enter the Chinese market and has recorded a huge number of rental trips over the years. Through vehicle smart locks and GPS positioning devices embedded in the bikes, location data are recorded as well as the registration data of users, including user ID, bike ID, and latitude and longitude information of start and end locations. Mobike generates about 900,000 orders daily, and some of the order data contain abnormal data due to various disturbing situations, such as smart lock failure, unstable GPS signal, and human error. These data will interfere with the accuracy of model results. Thus, before the analysis, a data cleaning process is implemented to screen out standard order data. Abnormal data in this study are identified based on the following five criteria:

- 1) Records having empty values in crucial fields such as order time, latitude, and longitude
- 2) Abnormal data repeatedly uploaded to the system in the same order
- 3) Abnormal data where the end time is earlier than the start time and multiple bikes are parked at the same location

- 4) Data on starting or ending locations (longitude and latitude) outside the administrative geographical scope of Beijing
- 5) Abnormal data with usage time less than 2 minutes or longer than 3 hours and driving distance less than 150 m or more than 10 km

In addition to Mobike data, we use the data for Beijing rail transit stations, road networks, and administrative areas. These data are downloaded from OpenStreetMap. We use ArcGIS software to calculate the inflow and outflow of shared bikes around the Xicheng District. Moreover, we also call the Baidu Maps API to generate the shortest path data, which is used for the cost calculation of the optimization model.

#### 4.2 Prediction Results of the Hybrid Model

To verify the effectiveness of the hybrid model, we take the surroundings of Fuchengmen station as an example and use SARIMA, LSTM, and SARIMA-LSTM models for prediction. There were 35 valid data sets on orders for shared bikes around the Fuchengmen subway station. The time interval for the analysis of each data set was from 6:00 to 21:00 with a granularity of 1 hour. In this paper, the first 34 sets of data were used as the training set for the prediction model, and the last set of data was predicted.

The training set data were first analyzed using the SARIMA model. Based on the results of the data processed by SARIMA modeling, when the model performance was optimal and the error was minimal, the parameter of SARIMA( $p, d, q$ ) ( $P, D, Q$ ) $s =$  SARIMA (1, 0, 2) (0, 1, 1) 16 was used for prediction.

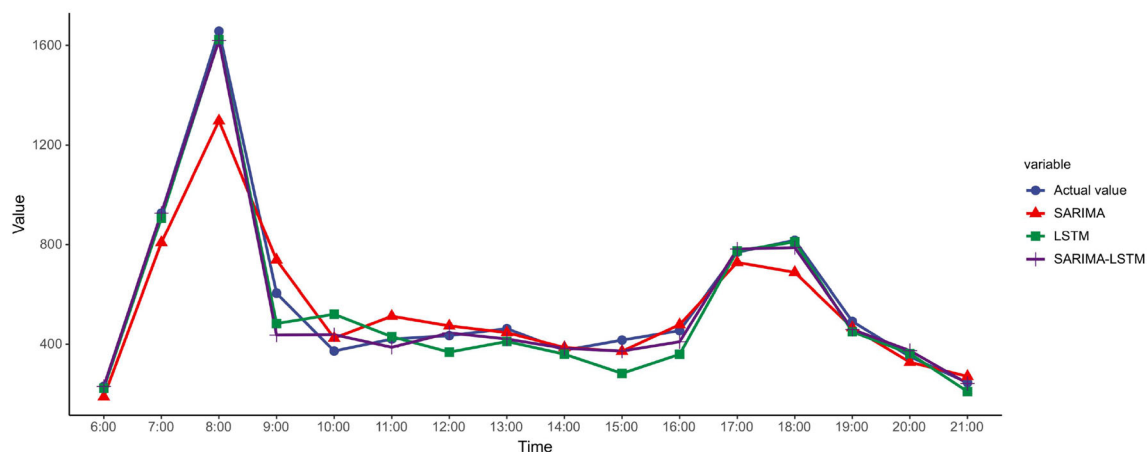
An LSTM model was then used to perform the same prediction work. The LSTM network model framework in this paper was built using the Keras deep learning library in Python and the TensorFlow package. When using the LSTM model, a set of hyperparameters (as listed in Table 1) needs to be determined because different settings affect the prediction results. To avoid difficult convergence of the model or excessive computational work, the number of hidden layers `lstm_size` was set to 1. because of the small amount of data, `time_step = 15`, `batch_size = 2`, and `epochs = 100` were set to ensure accuracy. MAE was used as the loss, which can better characterize the distribution of standard data. The model used Adaptive Moment Estimation (Adam) as the optimizer during the training process because it has a tremendous advantage over the optimizer in terms of learning speed and convergence speed (Table 3).

Finally, we ran the SARIMA-LSTM hybrid model with the same parameters as the separate predictions. The prediction results of the three models are summarized in Fig. 5, and the effect evaluation is shown in Table 4. The ‘original’ in Fig. 5 meant the actual sequence data of this day. It can be seen from Fig. 5 that the SARIMA-LSTM model, which is a combination of two different models, not only accurately captured the linear features in the data but also better grasped the remaining trend of the data and had a high fit with the actual data. The hybrid model was able to maximize the closeness to the actual data for most of the time, especially during the peak hours, and it could also be observed that the LSTM also performed relatively well. In other periods, the LSTM showed significant deviations from the actual results compared to the other models. In contrast, the hybrid model effectively took advantage of the two separate models, and the predictions did not deviate much.

Table 4 shows that the hybrid model had smaller MAE, RMSE, and MAPE than the other models. It is worth mentioning that the MAPE of the hybrid model was less than 10%. The results demonstrated that the model with a hybrid of SARIMA and LSTM is more accurate than when used separately in predicting bike-sharing time-series data.

**Table 3** The main parameters in the LSTM model

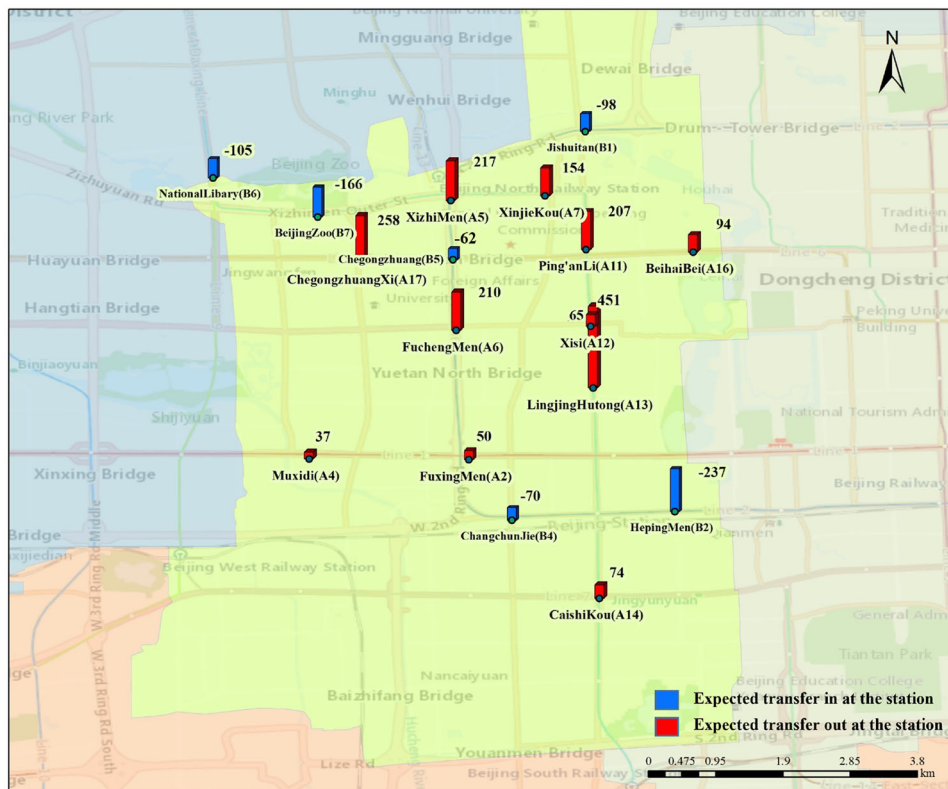
Name of the parameter	Meaning
<code>Lstm_size</code>	Number of hidden layers
<code>Time_step</code>	Number of blocks of the expanded LSTM
<code>Batch_size</code>	Number of rows in each batch feeding into the LSTM
<code>Epochs</code>	Number of iterations to train the model
<code>Optimizers</code>	Method for adjusting the weight of each node
<code>Loss</code>	Method for calculating the difference between the output of a neural network and the sample markers



**Fig. 5** Comparison of prediction results of different models

**Table 4** Effect evaluation of different prediction models

Models	Evaluation indexes		
	MAE	RMSE	MAPE (%)
SARIMA	63.085	101.136	22.486
LSTM	58.480	99.772	10.642
SARIMA-LSTM	43.513	54.467	9.765



**Fig. 6** Participating deployment sites of urban rail transit stations and deployment volume

### 4.3 Bike-Sharing Allocation

With the forecast data, an effective and dynamic deployment plan can be developed in advance to meet the demand of residents for bike-sharing usage. Because of the low demand for shared bikes at some sites, when the demand at a site is less than half the maximum capacity of a dispatched bike, i.e., 30 bikes, it is considered that the site will not participate in the deployment. After the statistical screening, the Xicheng District participated in the deployment of 17 urban rail transit stations, including 11 exit stations with 1817 bikes and six import stations with a total of 738 bikes. These stations and the demand for shared bikes are plotted using GIS, as shown in Fig. 6. There are 21 participating deployment sites, including nine

supplemental sites for up to 1512 supplemental bikes and 12 accommodating sites for up to 1474 bikes. We show these sites in Fig. 7.

Then the planning was done according to the bike-sharing allocation model. With a total of 3329 bikes to be shipped out and total demand of 2212 bikes, it is necessary to add a virtual allocation center as the 13th accommodation allocation station to accommodate an extra 1117 bikes. Based on the above consideration, the unknown quantities in Eq. (18) were determined where  $m = 11, N = 12, M = 9, n = 6$ . The optimal deployment scheme was obtained after 20 iterations of running the algorithm in MATLAB. The iterative process is shown in Fig. 8, and using this option yields a minimum total cost of 8353.6 yuan.

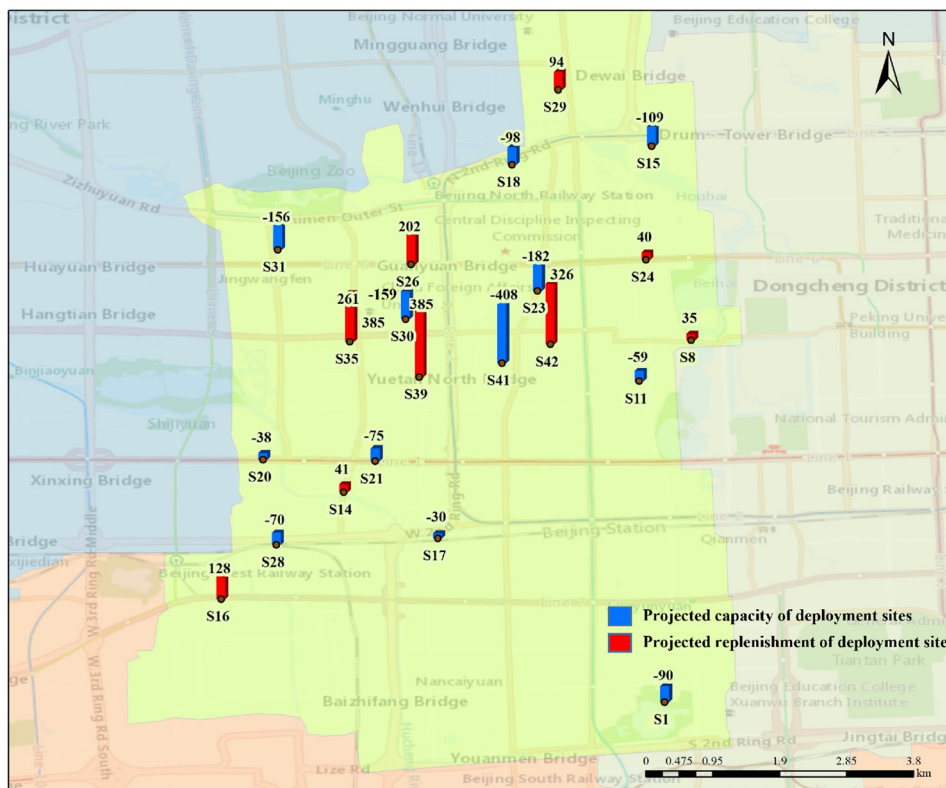


Fig. 7 Participating deployment sites and deployment volume

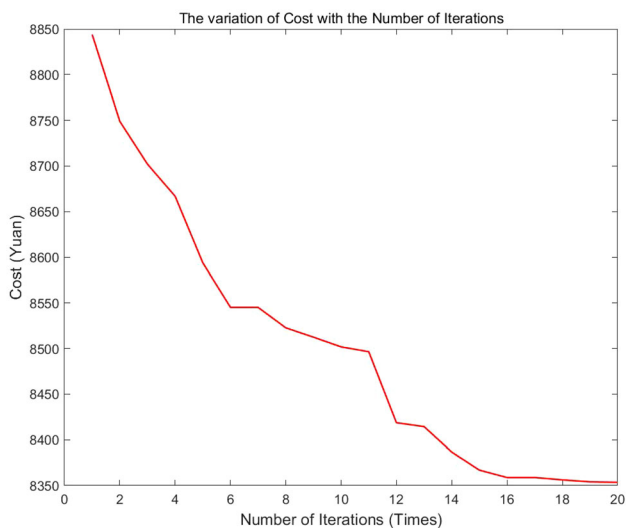


Fig. 8 The variation in cost with the number of iterations

A more visible representation of the deployment plan is shown in Figs. 9 and 10. In particular, Fig. 9 shows which stations need to transport the excess bikes to deployment sites. Figure 10 shows which stations need to replenish bikes from nearby deployment sites. The amount of transportation was marked. Finally, the allocation process was executed according to the method proposed in Sect. 3.2.

Furthermore, a sensitivity analysis is conducted for the labor cost. The result indicates that the current solution is still optimal when  $v \in [0.8, 1.0]$ . This result can provide a meaningful reference for managers in pricing and programming.

### 5 Conclusion

As a sustainable transportation mode with the apparent features of green, flexible, and easy access, bike sharing can be used to connect urban rail transit, which contributes to solving not only the first and last mile problems but also transportation decarbonization as a whole. However, the imbalance between the supply and demand of shared bikes in space and time has remained prominent in many cities, resulting in various social problems. For both planners and the convenience of users, an efficient vehicle deployment strategy based on accurate demand prediction is seen as essential to ensure the sustainable benefits of shared bikes in urban transportation systems.

In line with this concern, in this paper, we proposed a hybrid model which focuses on the dynamic deployment of bike sharing around urban rail transit stations based on a machine learning-enabled time-series prediction model. The purpose is to achieve the best deployment plan which

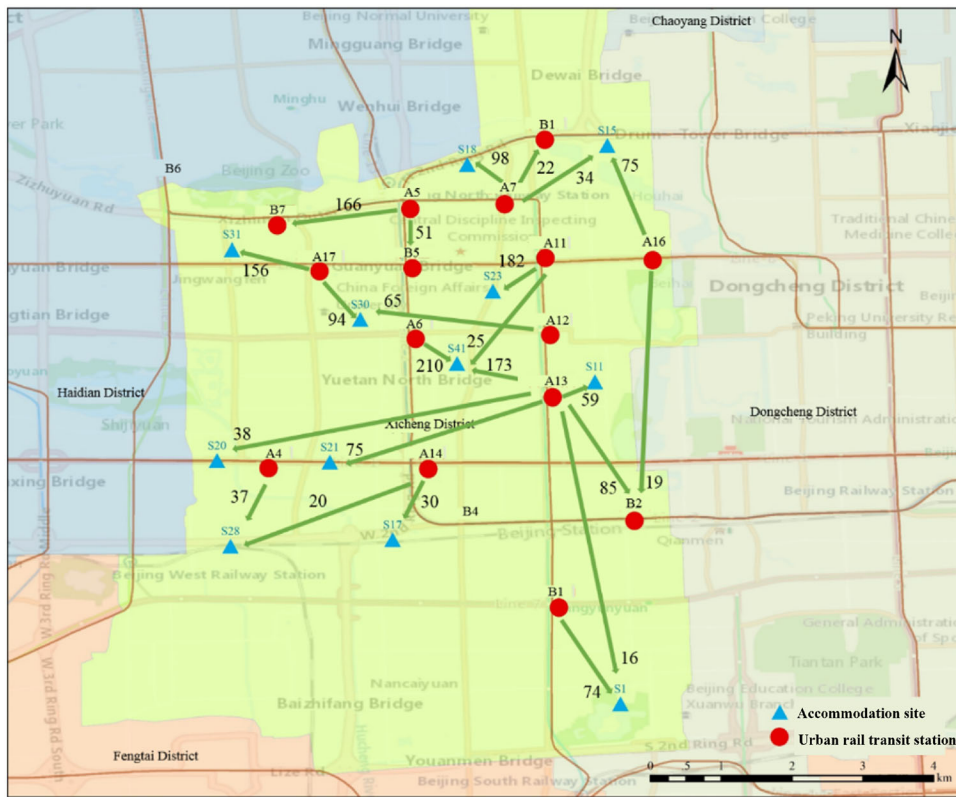


Fig. 9 Deployment schematic diagram of export stations

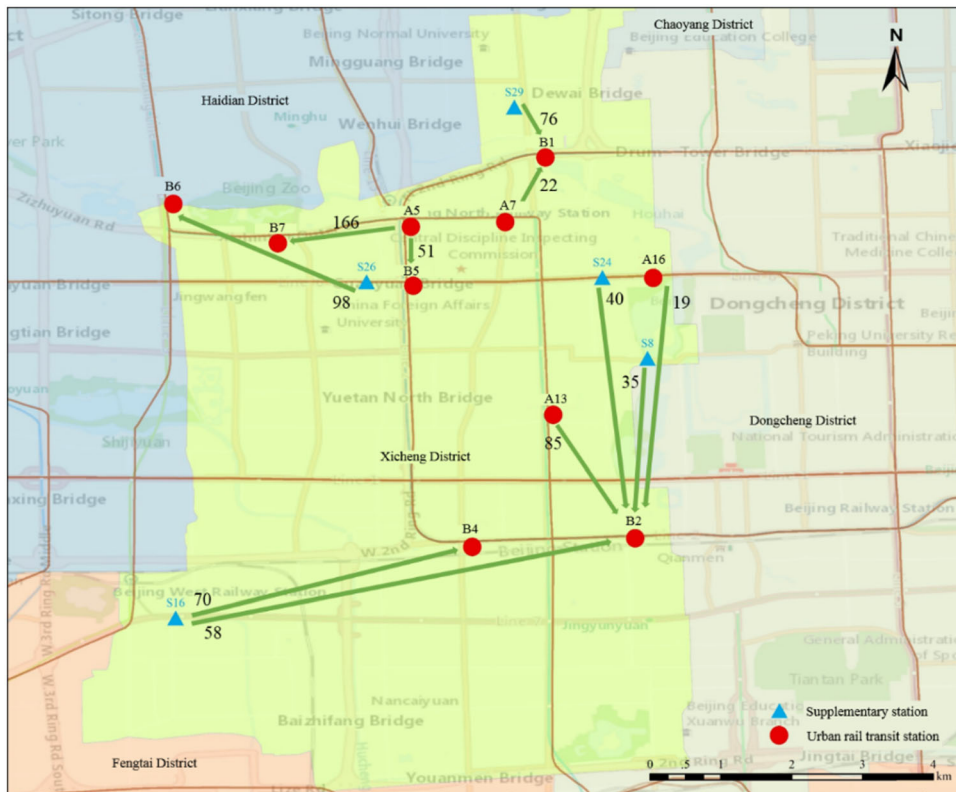


Fig. 10 Deployment schematic diagram of import stations



minimizes the total cost by balancing the supply and predicted demand at different stations. The proposed hybrid model incorporates the advantages of both SARIMA and LSTM to predict the use of shared bikes.

Using the actual data for the Xicheng District of Beijing, we compared the predicted values of a single model and the hybrid model. Results show that the hybrid model has higher prediction efficiency. This enriches the method for demand prediction of shared bikes. Apart from that, the bike-sharing deployment model around urban rail transit stations was developed to minimize transportation costs based on forecasting data. The resulting bike deployment plan would advise on the bike-sharing imbalance around urban rail transit stations.

This study was conducted by setting the hot spot within 500 m of a station. However, as a flexible transportation mode, the bike-sharing system has many types of hot spots in the city. Therefore, the hot areas for using different types of shared bikes may also be combined to conduct further research on different travel structures and the global deployment optimization of shared bikes. In addition, as this paper focuses on meeting the demand for peak bike-sharing usage, the temporal granularity of the prediction and the establishment of the allocation model were simplified based on the actual deployment of the operating companies. Further research can be conducted to cover dynamic requirements on a time-by-time basis.

**Acknowledgements** This work was supported by the Beijing Municipal Natural Science Foundation under Grant No. 9202012.

#### Declarations

**Conflict of 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.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

#### References

- Lu K, Han B, Zhou X (2018) Smart urban transit systems: from integrated framework to interdisciplinary perspective. *Urban Rail Transit* 4(2):49–67. <https://doi.org/10.1007/s40864-018-0080-x>
- Bai Y, Liu X, Huang Y, Wan X (2018) The influence of shared bicycle on urban rail transit operation and management in Beijing. *Urban Rapid Rail Transit* 31(03):119–124
- Zuo T, Wei H, Chen N, Zhang C (2020) First-and-last mile solution via bicycling to improving transit accessibility and advancing transportation equity. *Cities* 99:102614. <https://doi.org/10.1016/j.cities.2020.102614>
- Cheng L, Mi ZF, Coffman D, Meng J, Liu DN, Chang DF (2021) The role of bike sharing in promoting transport resilience. *Netw Spat Econ*. <https://doi.org/10.1007/s11067-021-09518-9>
- Zhang F, Liu W (2021) An economic analysis of integrating bike sharing service with metro systems. *Transp Res Part D Transp Environ* 99:103008. <https://doi.org/10.1016/j.trd.2021.103008>
- Hu B, Gao Y, Yan J, Sun Y, Ding Y, Bian J, Dong X, Sun H (2021) Understanding the operational efficiency of bicycle-sharing based on the influencing factor analyses: a case study in Nanjing China. *J Adv Transp* 2021:1–14. <https://doi.org/10.1155/2021/8818548>
- Molinillo S, Ruiz-Montañez M, Liébana-Cabanillas F (2020) User characteristics influencing use of a bicycle-sharing system integrated into an intermodal transport network in Spain. *Int J Sustain Transp* 14(7):513–524. <https://doi.org/10.1080/15568318.2019.1576812>
- Lin P, Weng J, Hu S, Alivianistos D, Li X, Yin B (2020) Revealing spatio-temporal patterns and influencing factors of dockless bike sharing demand. *IEEE Access* 8:66139–66149. <https://doi.org/10.1109/ACCESS.2020.2985329>
- Zeng P, Wei M, Liu X (2020) Investigating the spatiotemporal dynamics of urban vitality using bicycle-sharing data. *Sustainability* 12(5):1714. <https://doi.org/10.3390/su12051714>
- Xu M, Liu H, Yang H (2020) A deep learning based multi-block hybrid model for bike-sharing supply-demand prediction. *IEEE Access* 8:85826–85838. <https://doi.org/10.1109/ACCESS.2020.2987934>
- Xu T, Han G, Qi X, Du J, Lin C, Shu L (2020) A hybrid machine learning model for demand prediction of edge-computing-based bike-sharing system using internet of things. *IEEE Internet Things J* 7(8):7345–7356. <https://doi.org/10.1109/JIOT.2020.2983089>
- Chen J, Li K, Li K, Yu PS, Zeng Z (2021) Dynamic Bicycle Dispatching of Dockless Public Bicycle-sharing Systems using Multi-objective Reinforcement Learning. *arXiv preprint arXiv:2101.07437*
- Ma T, Chao L, Erdogan S (2015) Bicycle sharing and public transit: does capital bikeshare affect metrorail ridership in Washington, D.C.? *Transp Res Record J Transp Res Board* 2534(1):1–9
- Borgnat P, Abry P, Flandrin P, Robardet C, Rouquier J, Fleury E (2011) Shared bicycles in a city: a signal processing and data analysis perspective. *Adv Complex Syst* 14(3):415–438. <https://doi.org/10.1142/S0219525911002950>
- Chen L, Zhang D, Wang L, Yang D, Ma X, Li S, Wu Z, Pan G, Thi-Mai-Trang N, Jakubowicz J (2016) Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems. In: *UBI-COMP'16: Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* (2016) 841–852. <https://doi.org/10.1145/2971648.2971652>
- Feng S, Hao C, Du C, Li J, Ning J (2018) A Hierarchical Demand Prediction Method with Station Clustering for Bike Sharing System. In: *2018 IEEE Third International Conference on Data Science in Cyberspace (DSC)*, 2018
- Lin L, He Z, Peeta S (2018) Predicting station-level hourly demand in a large-scale bike sharing network: a graph convolutional neural network approach. *Transp Res Part C Emerging Technol* 97:258–276. <https://doi.org/10.1016/j.trc.2018.10.011>

18. Wang B, Kim I (2018) Short-term prediction for bike-sharing service using machine learning. *Transp Res Procedia* 34:171–178
19. Sohrabi S, Ermagun A (2021) Dynamic bike sharing traffic prediction using spatiotemporal pattern detection. *Transp Res Part D Transp Environ* 90:102647. <https://doi.org/10.1016/j.trd.2020.102647>
20. Wu X, Lyu C, Wang Z, Liu Z (2019) Station-level hourly bike demand prediction for dynamic repositioning in bike sharing systems. Springer, Berlin, pp 19–27
21. Kaltenbrunner A, Meza R, Grivolla J, Codina J, Banchs R (2010) Urban cycles and mobility patterns: exploring and predicting trends in a bicycle-based public transport system. *Pervasive Mob Comput* 6(4):455–466
22. Yoon JW, Pinelli F, Calabrese F (2012) Cityride: A Predictive Bike Sharing Journey Advisor. In: *IEEE International Conference on Mobile Data Management*, 2012, pp. 306–311.
23. Liu J, Qiao L, Meng Q, Chen W, Fu Y (2016) Station Site Optimization in Bike Sharing Systems. In: *2015 IEEE International Conference on Data Mining (ICDM)*, 2016
24. Mehdizadeh Dastjerdi A, Morency C (2022) Bike-sharing demand prediction at community level under COVID-19 using deep learning. *Sensors* 22(3):1060. <https://doi.org/10.3390/s22031060>
25. Boonjubut K, Hasegawa H (2020) Multivariate time series analysis using recurrent neural network to predict bike-sharing demand. Springer, Singapore, pp 69–77
26. Ma X, Yin Y, Jin Y, He M, Zhu M (2022) Short-term prediction of bike-sharing demand using multi-source data: a spatial-temporal graph attentional LSTM approach. *Appl Sci Basel* 12(3):1161. <https://doi.org/10.3390/app12031161>.
27. Forma IA, Raviv T, Tzur M (2015) A 3-step math heuristic for the static repositioning problem in bike-sharing systems. *Transp Res Part B Methodolog* 71:230–247. <https://doi.org/10.1016/j.trb.2014.10.003>
28. Angeloudis P, Hu J, Bell MGH (2014) A strategic repositioning algorithm for bicycle-sharing schemes. *Transp A Transp Sci* 10(8):759–774. <https://doi.org/10.1080/23249935.2014.884184>
29. Jost JD, Herr T, Lecaplain C, Brasch V, Pfeiffer M, Kippenberg TJ (2014) Counting the cycles of light using a self-referenced optical microresonator. *Optica* 2(8):706
30. Jiménez E, Soriguera F (2020) A new dynamic repositioning approach for bike sharing systems. *Transp Res Procedia* 47:227–234
31. Raviv T, Tzur M, Forma IA (2013) Static repositioning in a bike-sharing system: models and solution approaches. *EURO J Transp Logist* 2(3):187–229
32. Feng T, Arentze TA, Timmermans HJ (2011) Assessing the relative importance of input variables for route choice modeling: a neural network approach. *J East Asia Soc Transp Stud* 9:341–353. <https://doi.org/10.11175/easts.9.341>
33. Wang Z, Cheng L, Li Y, Li Z (2020) Spatiotemporal characteristics of bike-sharing usage around rail transit stations: evidence from Beijing, China. *Sustainability* 12(4):1299. <https://doi.org/10.3390/su12041299>