Data-driven optimization for rebalancing shared electric scooters

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Electronic
Research Archive

Research article

Data-driven optimization for rebalancing shared electric scooters

Yanxia Guan[1], Xuecheng Tian[1,*], Sheng Jin[2], Kun Gao[3], Wen Yi[4], Yong Jin[1], Xiaosong Hu[5] and Shuaian Wang[1]

[1] Faculty of Business, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

[2] College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China

[3] Department of Architecture and Civil Engineering, Chalmers University of Technology, Chalmers, Sweden

[4] Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

[5] State Key Laboratory of Mechanical Transmission/Automotive Collaborative Innovation Center, Chongqing University, Chongqing, China

* **Correspondence:** Email: xuecheng-simon.tian@connect.polyu.hk.

Abstract: Shared electric scooters have become a popular and flexible transportation mode in recent years. However, managing these systems, especially the rebalancing of scooters, poses significant challenges due to the unpredictable nature of user demand. To tackle this issue, we developed a stochastic optimization model (M0) aimed at minimizing transportation costs and penalties associated with unmet demand. To solve this model, we initially introduced a mean-value optimization model (M1), which uses average historical values for user demand. Subsequently, to capture the variability and uncertainty more accurately, we proposed a data-driven optimization model (M2) that uses the empirical distribution of historical data. Through computational experiments, we assessed both models' performance. The results consistently showed that M2 outperformed M1, effectively managing stochastic demand across various scenarios. Additionally, sensitivity analyses confirmed the adaptability of M2. Our findings offer practical insights for improving the efficiency of shared electric scooter systems under uncertain demand conditions.

Keywords: data-driven optimization; rebalancing problem; shared electric scooters; uncertain user demand

1. Introduction

As cities grow and the concept of sustainability strengthens, the sharing economy is receiving increasing attention [1]. Over the past decade, the sharing economy, particularly in the transport

sector [2], has received significant focus. A prominent area of study within the sharing economy in the transport sector focuses on vehicle-sharing systems. These systems provide shared vehicle services, such as bicycles and electric scooters, in public areas, including bus stops, metro stations, and other high-traffic locations. Users can utilize these services for commuting or travel and should return vehicles to a designated area upon completion of the rental.

The emergence of sharing systems is having an impact on transport patterns in cities [3–5]. This innovative transport mode reduces car usage, promotes the sharing economy, improves the quality of the urban environment [6–8], and enhances the health of the population [9]. An in-depth understanding of vehicle-sharing systems can provide meaningful insights for urban planning policymakers and companies providing vehicle-sharing services.

The success of vehicle-sharing systems hinges on both strategic design and effective operational strategies, particularly vehicle rebalancing. While infrastructure lays the groundwork, it is the operational strategies that enhance service efficiency [10, 11]. This study investigates the optimization strategies for rebalancing shared electric scooters. Unlike traditional shared bicycles, electric scooters utilize their electric drive systems that enable users to complete short-distance commuting without additional physical exertion and thus have a high degree of acceptance among users. In addition, electric scooters show higher adaptability when dealing with complex urban terrains and traffic environments. In particular, the electric drive system is able to respond quickly in scenarios with frequent starts and stops, improving operational flexibility and commute efficiency. Electric scooters offer distinct advantages in convenience and operational efficiency [12], making them an increasingly vital focus of research in modern shared transportation systems.

The rebalancing problem of the shared electric scooter system is mainly affected by the uncertainty of user demand and the shortage of charging station sites. User demand uncertainty introduces complexity to scheduling and allocation processes. Additionally, the constraints imposed by charging facilities affect the state of electric scooters. In our research, we assume that the uncertain demand for scooters each day follows a probability distribution; and accurate demand prediction and effective demand response strategies are vital for optimizing shared electric scooter systems. Moreover, high transportation and labor costs further complicate the rebalancing process.

To solve this uncertain operational problem, we propose a data-driven optimization model to obtain the operational scheduling scheme. Specifically, this data-driven approach fits stochastic user demand by the empirical demand distribution. Using this approach, we can better adjust the electric scooter configurations between stations to meet uncertain user demand. First, we construct a stochastic optimization model M0 to describe the rebalancing problem of electric scooter-sharing systems with uncertain user demand. Second, we use the mean value of historical user demand to fit the random user demand and construct a baseline mean-value model M1. Moreover, we use the empirical distribution of historical data to approximate the user demand, constituting an improved data-driven optimization model M2. Through different approaches, we approximate a stochastic optimization model M0 by the tractable optimization models M1 and M2, respectively. Finally, the performance of these models is evaluated through numerous computational experiments. The experimental results show that model M2 outperforms the benchmark model M1 in different scenarios, indicating that model M2 is better at dealing with stochastic user demand. Moreover, we conduct sensitivity analyses by altering model parameters to further validate the effectiveness of model M2. Consequently, our work has theoretical and practical significance for the development and enhancement of electric scooter-sharing systems.

The rest of the paper is structured as follows: Section 2 reviews the relevant literature. Section 3 constructs a mathematical model for the uncertain electric scooter scheduling problem. Section 4 presents solution approaches to solving the proposed stochastic model. Section 5 tests and validates the performance of our proposed models. Section 6 concludes the paper.

2. Related literature

There are lots of studies conducted on the sharing of bikes and cars. Yet, less research has been done about how to properly operate shared electric scooters [13]. Therefore, we review research on bike-sharing systems that are similar to the shared electric scooter systems.

The optimization objectives of some studies focus on the strategic design of bike-sharing systems, such as the arrangement of station locations. García et al. [14] used the geographic information system of Madrid city to design the location-allocation model. Raviv et al. [15] constructed a new inventory routing model to standardize emerging bike-sharing systems. Fricker and Gast [16] evaluated the impact of station capacity on the optimization of the bike-sharing system. Liu and Tian [17] optimized companies' decision making for bike-sharing services by constructing virtual sites to promote bike use based on the k-means method. Chen et al. [18] proposed an improved repositioning model using trajectory data from Shenzhen to optimize the system design solution.

More infrastructure development is not enough to improve the service of bike-sharing systems. How to effectively rebalance bikes between stations from an operational perspective has obtained the attention of many researchers [10, 11]. For example, Erdoğan et al. [19] presented the first exact algorithm for the bike rebalancing problem. Kadri et al. [20] used the branch-and-bound algorithm to solve the proposed bike rebalancing problem that optimizes the stations' operational costs. However, existing studies [21–24] are more biased towards the static sharing scheduling problem; that is, they focus more on situations where user demand is constant or has only minor changes, which restricts the system's flexibility and responsiveness.

To better improve the system's performance, some studies have considered analyzing the uncertain user demand. Hulot et al. [25] constructed a station-level user demand prediction model based on two years of real data from BIXI Montréal (a local bus company in Montréal) for rebalancing problems. VE and Cho [26] predicted user demand by analyzing multi-dimensional data containing information such as weather. However, these models have high requirements for data quality and quantity.

In summary, existing literature has extensively explored strategic design and rebalancing strategies for shared bicycle systems. However, compared to shared bicycles, shared electric scooters have significant advantages in terms of storage space requirements and operational flexibility, which enable them to show higher adaptability in urban environments. The shared electric scooters reduce physical effort for users, making scooters preferable for short commutes. Additionally, they are better suited to complex urban environments, with faster responses in stop-and-go conditions, improving both flexibility and commuting efficiency. Nevertheless, the scheduling and rebalancing problems faced by electric scooters, especially those limited by charging equipment, have not been adequately studied. Therefore, this study aims to investigate how shared electric scooter systems can achieve effective rebalancing in the face of user demand uncertainty under charging facility constraints.

3. Problem description

Take one bus route in an area that stops at multiple stations as an example for a company that offers shared electric scooter services. In the electric scooter scheduling optimization problem, the initial state of electric scooters is influenced by the placement of charging hubs. In order to ensure that the electric scooter can be properly managed and used with the support of charging facilities, we assume that all the electric scooters will return to their designated stations for recharging at the end of each day's operation. On this basis, our model focuses on adjusting the distribution of electric scooters in the face of stochastic user demand, thereby optimizing the scheduling strategy to meet user demand more efficiently.

The rebalancing decision not only determines the operating costs, but also affects user satisfaction. More importantly, this decision is made in the face of the unknown user demand. Therefore, our research attempts to formulate a model that provides the optimal solution to this scheduling problem considering uncertain user demand. This model is constructed based on the following assumptions: i) Each site has an unconstrained capacity for electric scooters; and ii) The total number of electric scooters owned in the shared electric scooter system is fixed.

Our research primarily focuses on implementing operational rebalancing decisions before users begin using the scooters every day. The research aims to achieve effective rebalancing of the electric scooter system on a daily basis, with the objective of minimizing transportation costs associated with transferring scooters between stations and minimizing the penalties resulting from unsatisfied user demand. To handle uncertain user demand, we assume that historical demand information is available.

To formulate the problem discussed above, we consider a region with a set of transit stations, represented as P, where passengers may choose electric scooters for commuting to work upon arriving at the destination station by transit. The region also comprises a set of electric scooter storage stations, denoted as K, where the number of electric scooters available at the station k is q_k, $\forall k \in K$. Here, the transit station set P is a subset of the storage station set K. Additionally, the number of users utilizing electric scooters from station p to work is represented by the stochastic parameter $\tilde{\xi}_p$. Although the underlying distribution of $\tilde{\xi}_p$ is unknown, we are accessible to M pieces of historical user demand at each station p, $p \in P$ over a historical period, denoted as $\zeta = \{\boldsymbol{\xi}^1, \cdots, \boldsymbol{\xi}^m, \cdots, \boldsymbol{\xi}^M\}$; here, $\boldsymbol{\xi}^m = (\xi_1^m, \cdots, \xi_{|P|}^m)$ denotes the vector of user demand in the mth instance, where ξ_p^m means the user demand in the mth instance for station p, $\forall m \in \{1, 2, \cdots, M\}$, $\forall p \in P$. To handle uncertain user demand, we assume that the distribution of the stochastic user demand $(\tilde{\xi}_1, \cdots, \tilde{\xi}_{|P|})$ is consistent with the historical data, which can be characterized by $\mathbb{P}(\tilde{\xi}_p = \xi_p^m) = 1/M$, where $\forall p \in P$, $\forall m \in \{1, \cdots, M\}$.

Now, we denote by decision variable $x_{kp} \in \mathbb{Z}_+$ the number of electric scooters transported from site k to site p, $k \in K$, $p \in P$. Additionally, the parameter c_{kp} signifies the transport cost per unit of electric scooter from site k to site p. Therefore, we quantify the transportation costs in the whole scheduling process by $\sum_{k \in K} \sum_{p \in P} c_{kp} x_{kp}$. Moreover, the decision variable y_p denotes the number of electric scooters at site p after the scheduling process, and it can be calculated through Eq (3.1) as follows:

$$y_p = q_p + \sum_{k \in K} x_{kp} - \sum_{p' \in P} x_{pp'}. \tag{3.1}$$

Similarly, the expected decision loss stemming from the inadequate inventory level y_p at site p is expressed as $d_p \mathbb{E}\left[\max\{0, \tilde{\xi}_p - y_p\}\right]$, where d_p characterizes the loss incurred due to the lack of an

electric scooter at site p. When the final number of electric scooters y_p could meet the users' demand $\tilde{\xi}_p$ at station p, the decision loss at station p is zero; otherwise, it is $d_p(\tilde{\xi}_p - y_p)$. Before we present the mathematical programming model for our research problem, we present all of the notions as follows.

Sets and Indices	
P	Set of transit stations
K	Set of electric scooter storage stations, $P \subset K$
ξ^m	Set of mth piece of user demand data, $m = 1, 2, \cdots, M$, where M is the total number of historical pieces
ζ	Set of user data, which contains a total of M pieces of historical user demand data, $\zeta = \{\xi^1, \cdots, \xi^M\}$
p	Index of transit stations, $p = 1, 2, 3, \cdots, |P|$
k	Index of electric scooter storage stations, $k = 1, 2, 3, \cdots, |K|$
m	Index of historical data pieces, $m = 1, 2, 3, \cdots, M$

Deterministic parameters	
ξ_p^m	The user demand of station p, $p \in P$, in the mth piece of historical data, $m = 1, 2, \cdots, M$ and $\xi^m = (\xi_1^m, \cdots, \xi_{|P|}^m)$
q_k	The number of electric scooters available at the station k, $\forall k \in K$
c_{kp}	The transporting cost per unit of electric scooter from station k to station p, $\forall k \in K, \forall p \in P$
d_p	The loss incurred due to the lack of an electric scooter at station p, $\forall p \in P$

Stochastic parameter	
$\tilde{\xi}_p$	The stochastic demand of electric scooters at station p, $\forall p \in P$

Decision variables	
x_{kp}	The number of electric scooters transported from station k to station p, $\forall k \in K, \forall p \in P$
y_p	The number of electric scooters at station p after the rebalancing process, $\forall p \in P$

The objective of this study is to optimize the allocation of available electric scooters to meet users' uncertain demand. The uncertain scheduling problem, termed model M0, can be mathematically modeled as follows:

$$\min \sum_{k \in K} \sum_{p \in P} c_{kp} x_{kp} + \sum_{p \in P} d_p \mathbb{E}\left[\max\left\{0, \tilde{\xi}_p - y_p\right\}\right] \tag{3.2a}$$

$$\text{s.t.} \quad y_p = q_p + \sum_{k \in K} x_{kp} - \sum_{p' \in P} x_{pp'}, \quad \forall p \in P \tag{3.2b}$$

$$\sum_{p \in P} x_{kp} \leq q_k, \quad \forall k \in K \tag{3.2c}$$

$$x_{kp} \in \mathbb{Z}_+, \quad \forall k \in K, \quad \forall p \in P \tag{3.2d}$$

$$y_p \in \mathbb{Z}_+, \quad \forall p \in P. \tag{3.2e}$$

Objective function (3.2a) minimizes the sum of transportation cost and decision losses. Constraints (3.2b) compute the number of electric scooters owned by each site after the scheduling is completed.

Constraints (3.2c) ensure that the number of electric scooters transported from station k does not exceed its storage capacity q_k. Constraints (3.2d) and (3.2e) ensure the domains of the decision variables.

Proposition 1. *The integer constraints* (3.2e) *can be relaxed to the continuous constraints* $y_p \in \mathbb{R}_+, \forall p \in P$, *where* $\mathbb{R}_+$ *denotes the set of non-negative real numbers.*

Proof. The value of y_p is formed through integer additions and subtractions in the calculation displayed in (3.2b). Relaxing the constraints of y_p does not affect the consistency of the solution. □

4. Solution methods

4.1. Mean-value optimization model M1

In model M0, objective function (3.2a) incorporates the random user demand $\tilde{\xi}_p$ at station p. Solving it requires the knowledge of the uncertain demand distribution. To address this challenge, we initially follow the common approach by using the average values of user demand at transit stations from the historical data $\{\xi^m\}_{m=1}^{M}$ to approximate the random user demand. Incorporating the average values into model M0, the objective function of model M0 is approximated as follows:

$$\min \sum_{k \in K} \sum_{p \in P} c_{kp} x_{kp} + \sum_{p \in P} d_p \max \left\{ 0, \frac{1}{M} \left(\sum_{m=1}^{M} \xi_p^m \right) - y_p \right\}. \tag{4.1}$$

The mean-value method approximates the stochastic objective function (3.2a) in model M0 by a deterministic objective function (4.1). However, the objective function (4.1) contains the max operator, which is a nonlinear term. In order to transform the nonlinear term into a linear term, we introduce the auxiliary decision variable w_p to represent the number of unmet electric scooters for the users at station p, $\forall p \in P$, and it can be calculated as follows:

$$w_p = \begin{cases} \dfrac{1}{M} \displaystyle\sum_{m=1}^{M} \xi_p^m - y_p & \text{if } \dfrac{1}{M} \displaystyle\sum_{m=1}^{M} \xi_p^m - y_p > 0 \\ 0 & \text{if } \dfrac{1}{M} \displaystyle\sum_{m=1}^{M} \xi_p^m - y_p \leq 0. \end{cases} \tag{4.2}$$

By using the average values of historical data and auxiliary variables w_p, $\forall p \in P$, we approximate the stochastic model M0 by a deterministic mean-value model M1, shown below:

$$\min \sum_{k \in K} \sum_{p \in P} c_{kp} x_{kp} + \sum_{p \in P} d_p w_p \tag{4.3a}$$

$$\text{s.t.} \quad y_p = q_p + \sum_{k \in K} x_{kp} - \sum_{p' \in P} x_{pp'}, \quad \forall p \in P \tag{4.3b}$$

$$w_p \geq \frac{1}{M} \sum_{m=1}^{M} \xi_p^m - y_p, \quad \forall p \in P \tag{4.3c}$$

$$\sum_{p \in P} x_{kp} \leq q_k, \quad \forall k \in K \tag{4.3d}$$

$$x_{kp} \in \mathbb{Z}_+, \quad \forall k \in K, \quad \forall p \in P \tag{4.3e}$$

$$w_p \in \mathbb{R}_+, \quad \forall p \in P \tag{4.3f}$$

$$y_p \in \mathbb{R}_+, \quad \forall p \in P. \tag{4.3g}$$

Objective function (4.3a) aims to minimize the sum of transportation costs and decision losses. Constraints (4.3b) calculate the number of scooters owned by each site after the dispatching has been completed. Constraints (4.3c) calculate the number of users whose demand for scooters is unmet. Constraints (4.3d) guarantee that the number of electric scooters transported from site k does not exceed its original storage capacity q_k. Constraints (4.3e)–(4.3g) ensure the domains of the decision variables.

4.2. Data-driven optimization model M2

Simple average values may not adequately reflect the real-world uncertainties, thereby reducing the model's effectiveness. Therefore, we approximate the distribution of the random user demand $(\tilde{\xi}_1, \cdots, \tilde{\xi}_{|P|})$ by using the empirical dataset $\zeta = \{\boldsymbol{\xi}^m\}_{m=1}^M$. Specifically, we assume $\mathbb{P}(\tilde{\xi}_p = \xi_p^m) = 1/M$, $\forall p \in P$, $\forall m \in \{1, \cdots, M\}$. This method approximates the calculation of $\mathbb{E}[\max\{0, \tilde{\xi}_p - y_p\}]$ by $(M)^{-1} \sum_{m=1}^M \max\{0, \xi_p^m - y_p\}$. Consequently, the objective function of model M0 is approximated by:

$$\min \sum_{k \in K} \sum_{p \in P} c_{kp} x_{kp} + \sum_{p \in P} d_p \left(\frac{1}{M} \sum_{m=1}^M \max\{0, \xi_p^m - y_p\} \right). \tag{4.4}$$

Likewise, this approach approximates the original stochastic objective function (3.2a) by the deterministic objective function (4.4), which is tractable. Since Objective function (4.4) contains the max operation, it is nonlinear. Therefore, we introduce the auxiliary decision variable z_p^m to represent the number of users whose demand for scooters is unmet at site p in the mth instance. Specifically, $z_p^m = 0$ denotes the successful fulfillment of user needs when the actual inventory level y_p equals or exceeds the user demand ξ_p^m at station p in the mth instance; otherwise, z_p^m means the shortage, indicating the discrepancy between the actual inventory level y_p and the user demand ξ_p^m. z_p^m is calculated as follows:

$$z_p^m = \begin{cases} \xi_p^m - y_p & \text{if } \xi_p^m - y_p > 0, \\ 0 & \text{if } \xi_p^m - y_p \leq 0, \end{cases} \tag{4.5}$$

By doing so, objective function (4.4) is expressed as follows:

$$\min \sum_{k \in K} \sum_{p \in P} c_{kp} x_{kp} + \sum_{p \in P} d_p \left(\frac{1}{M} \sum_{m=1}^M z_p^m \right). \tag{4.6}$$

Meanwhile, to simplify the model representation, we introduce the decision variable z_p to represent the average number of users whose demand for scooters is unmet at site p, which can be calculated as follows:

$$z_p = \frac{1}{M} \sum_{m=1}^M z_p^m. \tag{4.7}$$

By utilizing historical data ζ to approximate the distribution of stochastic user demand $(\tilde{\xi}_1, \cdots, \tilde{\xi}_{|P|})$ and introducing auxiliary decision variables z_p, $\forall p \in P$, the complex stochastic optimization model M0

is approximated by the mixed-integer programming model M2, shown as follows:

$$\min \sum_{k \in K} \sum_{p \in P} c_{kp} x_{kp} + \sum_{p \in P} d_p z_p \tag{4.8a}$$

$$\text{s.t.} \quad y_p = q_p + \sum_{k \in K} x_{kp} - \sum_{p' \in P} x_{pp'}, \quad \forall p \in P \tag{4.8b}$$

$$z_p^m \geq \xi_p^m - y_p, \quad \forall p \in P, \quad \forall m \in \{1, \cdots, M\} \tag{4.8c}$$

$$z_p = \frac{1}{M} \sum_{m=1}^{M} z_p^m, \quad \forall p \in P \tag{4.8d}$$

$$\sum_{p \in P} x_{kp} \leq q_k, \quad \forall k \in K \tag{4.8e}$$

$$x_{kp} \in \mathbb{Z}_+, \quad \forall k \in K, \quad \forall p \in P \tag{4.8f}$$

$$z_p \in \mathbb{R}_+, \quad \forall p \in P \tag{4.8g}$$

$$y_p \in \mathbb{R}_+, \quad \forall p \in P \tag{4.8h}$$

$$z_p^m \in \mathbb{Z}_+, \quad \forall p \in P, \quad \forall m \in \{1, \cdots, M\}. \tag{4.8i}$$

Objective function (4.8a) aims to minimize the sum of transportation costs and decision losses. Constraints (4.8b) calculate the number of scooters available at a site after the dispatching process. Constraints (4.8c) determine the number of unsatisfied users in any station of any instance. Constraints (4.8d) calculate the average number of users whose demand for scooters is unmet among all of the instances. Constraints (4.8e) ensure that the number of electric scooters transported from site k does not exceed q_k. Constraints (4.8f)–(4.8i) define the domains of the decision variables.

5. Computational experiments

5.1. Experiment settings

Based on the historical data and the introduced auxiliary variables, we approximate the original optimization model M0 by the mixed-integer programming models M1 and M2. This section conducts computational experiments to validate their performance.

In our computational experiments, we run programs using an M3 Pro chip with an 11-core configuration. Simultaneously, we utilize the CBC solver from Python to solve optimization models, an open-source tool specifically designed for integer and mixed-integer linear programming problems.

5.1.1. Parameter settings

Firstly, we determine the values of the parameters to be adopted. The respective considerations and parameter values are listed below:

1). The number of transit stations $|P|$ is set to 20.
2). The number of electric scooter storage stations $|K|$ is set to 40.
3). The number of instances of historic data M is set to 100.
4). The number of electric scooters q_k available at the station k, $k \in K$, is set to 100.

5). As for the number of users needing electric scooters, ξ_p^m, in the site p of the mth instance, we take an arbitrary integer from a discrete uniform distribution over the interval $[0, 5q_k]$, $\forall k \in K$, $\forall p \in P$, $\forall m \in \{1, \cdots, M\}$.

6). The transport cost c_{kp} per unit of electric scooter from site k to site p is associated with the distance d_{kp} (km) between site k and site p. We first set $c_{kp} = 0.5d_{kp}$ ($). Here, d_{kp} is a randomly generated positive value less than 100. The distances between sites all satisfy the principle of triangulation which means that $d_{kp'} + d_{p'p} > d_{kp}$, $\forall k \in K$, $\forall p, p' \in P$. This distance design approach intends to simplify the complexity of our models.

7). For the loss, d_p, incurred due to the lack of an electric scooter at station p, $p \in P$, we set d_p to be the value of 10 ($).

5.1.2. Experimental design

In this section, we describe our experimental design, as illustrated in Algorithm 1. The goal of this algorithm is to evaluate the models' performances based on available data ζ. We divide the historical user demand data ζ into the training dataset ζ_{train} and the test dataset ζ_{test} in the ratio of $9:1$. Here, the training dataset ζ_{train} can be seen as the historical dataset, and the test dataset ζ_{test} is used to measure the out-of-sample performances of our proposed models. First, we obtain the initial optimal solution X_1^* for the first instance of the test dataset from the model (model M1 or model M2) using the training dataset ζ_{train}. In the optimal solution X_1^*, the element $X_1^*[i, j]$ represents the optimal number of electric scooters transported from station i to station j, $\forall i \in K$ and $\forall j \in P$. In addition, q_k is the initial inventory level of the first instance. And then we update the number of electric scooters q_k^1 at station k based on X_1^* through Eq (5.1), which is the end inventory level of the first instance, shown as follows:

$$q_k^1 = q_k + \left(\sum_{i=1}^{|K|} X_1^*[i, k] \right) - \left(\sum_{i=1}^{|P|} X_1^*[k, i] \right). \tag{5.1}$$

For easier evaluation of the model performance, we assume that users will return electric scooters to their original stations after use. Thus, q_k^1 is also the original inventory level of the second test instance.

Next, we use ζ_{test} to evaluate the performance of the model based on q_k^1. We define ξ^i as the ith instance of ζ_{test}, $\forall i \in \{1, 2, \cdots, |\zeta_{test}|\}$. Specifically, ξ_p^i means the actual user demand in the station p in the ith instance of ζ_{test} and here $\xi^i = (\xi_1^i, \cdots, \xi_{|P|}^i)$. As for the test instance ξ^i, we concatenate the known data ζ_{train} and $\{\xi^j\}_{j=1}^{j=i-1}$ to obtain the optimal solution X_i^* and q_k^i, $k \in K$. Then, we evaluate the performance res_i of ξ^i, $\forall i \in \{1, 2, \cdots, |\zeta_{test}|\}$ by

$$res_i = \sum_{k \in K} \sum_{p \in P} c_{kp} X_i^*[k, p] + \sum_{p \in P} d_p \max\{0, \xi_p^i - q_p^i\}. \tag{5.2}$$

The calculation of res_i is the sum of transport costs $\sum_{k \in K} \sum_{p \in P} c_{kp} X_i^*[k, p]$ and the penalties caused by insufficient electric scooters $\sum_{p \in P} d_p \max\{0, \xi_p^i - q_p^i\}$. Finally, res is calculated by summing the individual values res_i for all of the instances in the test dataset ζ_{test}. The final value res is used to evaluate the performance of the model, calculated by

$$res = \sum_{i=1}^{|\zeta_{test}|} res_i. \tag{5.3}$$

Algorithm 1 Experimental setup for training and testing

1: **Input**: The historical user demand data ζ, training ratio $\alpha = 0.9$,
   parameter tuple $\theta = (P, K, c_{kp}, d_p, q_k, \forall k \in K, \forall p \in P)$
2: **Output**: The transport cost and the penalties caused by inadequate scooters in the test dataset
3: Split ζ into a training set ζ_{train} and a test set ζ_{test}, based on ratio α
4: Solve the model by dataset ζ_{train} and obtain the initial optimal solution X_1^*
5: $q_k^1 = q_k + (\sum_{i=1}^{|K|} X_1^*[i, k] - \sum_{i=1}^{|P|} X_1^*[k, i]), \forall k \in K$
6: **for** $i = 1, 2, \cdots, |\zeta_{test}|$ **do**
7:    Calculate res_i through $res_i = \sum_{k \in K} \sum_{p \in P} c_{kp} X_i^*[k, p] + \sum_{p \in P} d_p \max\{0, \xi_p^i - q_p^i\}, \forall \xi^i \in \zeta_{test}$
8:    Solve the model based on new $\theta' = (P, K, c_{kp}, d_p, q_k^i, \forall k \in K, \forall p \in P)$ and $\zeta_{train} = \zeta_{train} \cup \xi^i$
9:    Obtain the optimal solution X_{i+1}^*
10:   $q_k^{i+1} = q_k^i + (\sum_{j=1}^{|K|} X_{i+1}^*[j, k] - \sum_{j=1}^{|P|} X_{i+1}^*[k, j]), \forall k \in K$
11: **end for**
12: The final objective function value res through $res = \sum_{i=1}^{|\zeta_{test}|} res_i$
13: **return** res

5.2. Evaluation of models

5.2.1. Model performance

In the previous sections, we presented various models for solving electric scooter scheduling problems. We used the *res* calculated through the test dataset ζ_{test} as the evaluation metric for comparing model performances; here, a larger value of *res* indicates a poorer ability of the model to minimize transport costs and penalties caused by inadequate electric scooters. We use Algorithm 1 for assessing the models' performances.

Table 1 presents the results (*res*) under varying conditions, particularly differences in the number of transit stations ($|P|$). An analysis of these results reveals that while the performance gap between the two models fluctuates with changes in $|P|$, model M2 consistently outperforms the baseline model M1 across all scenarios. Notably, model M2 achieves its greatest efficacy at $|P| = 15$, showing an improvement of approximately 13.12% over model M1. The smallest observed improvement, at $|P| = 16$, is approximately 3.61%. These results underscore the enhanced capability of model M2 in managing the operational challenges of electric scooter rebalancing in response to uncertain user demand, confirming its superior performance overall.

Table 1. *res* values of models in different scenarios.

$|S|, |P|$	M1	M2
$|S| = 40, |P| = 15$	77, 824.59	67, 615.38
$|S| = 40, |P| = 16$	85, 813.63	82, 717.11
$|S| = 40, |P| = 17$	102, 597.44	96, 771.46
$|S| = 40, |P| = 18$	123, 794.27	115, 172.27
$|S| = 40, |P| = 19$	138, 456.39	131, 222.93
$|S| = 40, |P| = 20$	160, 383.08	149, 227.68

5.2.2. Sensitivity analyses

To analyze the electric scooter scheduling problem on a specific bus route, the parameters P and K are generally assumed to be fixed. In contrast, the parameters d_p and q_k often fluctuate dynamically. For instance, in colder winter weather, the likelihood of choosing electric scooters decreases, leading to a downward adjustment in q_k at the station. Considering the dynamic nature of these parameters in real-world scenarios, it is crucial to perform sensitivity analyses to investigate their impact on the scheduling decisions.

To account for the volatility observed in real-world scenarios, this study conducts sensitivity analyses by considering a range of values for the penalty coefficient d_p and the storage quantity q_k. Specifically, when conducting sensitivity analyses for d_p, we assume that $|S| = 40$, $|P| = 20$, and $q_k = 100$, allowing d_p to vary with an interval of 1 over the integer range from 0 to 10. Similarly, when performing sensitivity analyses for q_k, we assume that $|S| = 40$, $|P| = 20$, and $d_p = 10$, allowing q_k to vary with an interval of 10 over the integer range from 50 to 150. At the same time, we calculate the improvement Δ between models M1 and M2 using Eq (5.4). Here, Δ_{d_p} corresponds to parameter d_p, and similarly, Δ_{q_k} is associated with parameter q_k. The results of these calculations are presented in Table 2.

$$\Delta = \frac{res_{M1} - res_{M2}}{res_{M1}}. \tag{5.4}$$

Table 2. Results of sensitivity analyses of d_p and q_k.

d_p	M1	M2	Δ_{d_p}	q_k	M1	M2	Δ_{q_k}
$d_p = 0$	0	0	0.00%	$q_k = 50$	302,800.66	294,019.64	2.90%
$d_p = 1$	27,383.0	27,598.73	−0.79%	$q_k = 60$	272,505.80	263,476.50	3.31%
$d_p = 2$	36,351.88	38,908.07	−7.03%	$q_k = 70$	241,615.10	233,188.19	3.49%
$d_p = 3$	49,801.44	48,549.93	2.51%	$q_k = 80$	210,953.11	203,972.29	3.31%
$d_p = 4$	65,600.94	62,232.04	5.14%	$q_k = 90$	183,706.23	176,605.01	3.87%
$d_p = 5$	81,399.44	76,880.99	5.55%	$q_k = 100$	160,383.08	149,227.68	6.95%
$d_p = 6$	97,201.44	91,390.25	5.98%	$q_k = 110$	134,635.92	126,016.66	6.40%
$d_p = 7$	113,012.02	105,917.89	6.28%	$q_k = 120$	108,877.22	106,244.49	2.42%
$d_p = 8$	128,809.08	120,088.38	6.77%	$q_k = 130$	99,543.95	87,298.08	12.30%
$d_p = 9$	144,602.08	134,727.28	6.83%	$q_k = 140$	98,992.93	71,183.24	28.10%
$d_p = 10$	160,383.08	149,227.68	6.95%	$q_k = 150$	98,616.64	56,279.06	42.93%

Table 2 compares the performance of baseline models M1 and M2 under varying values of d_p and q_k. The data show that model M2 generally surpasses model M1 in performance. However, exceptions are noted at lower d_p values ($d_p = 0, 1, 2$), where the models perform similarly or where M1 may exhibit comparative advantages.

In detail, in terms of the penalty coefficient d_p, $\forall p \in P$, when $d_p = 0$, both models yield 0 transport costs, which means that no electric scooter is dispatched between sites when there is no penalty for the shortage of electric scooters at a station. However, model M2 fails to improve performance when $d_p = 1$ and $d_p = 2$. This lack of enhancement may be attributed to the low penalty values assigned for scooter shortages, which are insufficient to significantly influence operational decisions under user demand uncertainty. Consequently, the advantages of model M2, which are designed to optimize around

variability, are not effectively leveraged. Model M2 shows a significant performance improvement compared to model M1 when $d_p \geq 3$, and the Δ_{d_p} gradually increases along with the increase of d_p, with the highest Δ_{d_p} at $d_p = 10$, which is about 6.95%. This trend underscores model M2's adaptability in scenarios where higher penalties compel more strategic rebalancing efforts to accommodate uncertain demand fluctuations.

As for q_k, $\forall k \in K$, an increase in the initial inventory q_k means that each site has more electric scooters at the start of the operation. As a result, the number of electric scooters that need to be dispatched from the storage sites to meet user demand decreases, which directly reduces transport costs. Additionally, it also reduces the number of electric scooters in shortage to meet user demand, reducing the corresponding penalties. Thus, the total cost *res*, consisting of the sum of transport costs and the penalties value due to the lack of electric scooters, decreases gradually as the initial inventory q_k increases.

However, according to Table 2, the variation of q_k does not change the superior position of model M2, verifying its significant advantages. These findings underscore the effectiveness of the data-driven model M2 in managing variations in d_p and q_k, making it more applicable in practical scenarios where these parameters may fluctuate. The overall superior performance of the model M2 suggests its capability to provide more effective solutions to electric scooter scheduling problems.

6. Conclusions

This study addresses the operational rebalancing challenges in shared electric scooter systems by developing and comparing several optimization models. Initially, we construct a stochastic optimization model (M0) to capture the complexity of unknown user demand. To operationalize this model, we introduce the mean-value optimization model (M1), which estimates demand using historical averages. Further enhancing our approach, we develop a data-driven optimization model (M2) that uses the empirical distribution of user demand for greater accuracy.

Through extensive computational experiments, we demonstrate that model M2 consistently outperforms model M1 in various scenarios. These findings underscore the superiority of the data-driven model M2 in effectively managing the complexities of scooter rebalancing. The performance of M2 across different test conditions confirms its practical applicability and potential to significantly improve operational efficiencies in real-world electric scooter sharing systems.

In our research, we have utilized stochastic optimization to cope with user demand uncertainty, and future research can further explore methods such as robust optimization and distributionally robust optimization. These methods provide alternative strategies to cope with worst-case scenarios and limited distributional information, which can help to enhance the robustness of the model. Therefore, future work can be centered around these methods to extend and enrich the existing research results to provide more comprehensive solutions for practical applications.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare there is no conflicts of interest.

References

1. S. K. Curtis, O. Mont, Sharing economy business models for sustainability, *J. Cleaner Prod.*, **266** (2020), 121519. https://doi.org/10.1016/j.jclepro.2020.121519

2. S. Castellanos, S. Grant-Muller, K. Wright, Technology, transport, and the sharing economy: towards a working taxonomy for shared mobility, *Transport Rev.*, **42** (2022), 318–336. https://doi.org/10.1080/01441647.2021.1968976

3. D. Fuller, L. Gauvin, Y. Kestens, P. Morency, L. Drouin, The potential modal shift and health benefits of implementing a public bicycle share program in Montreal, Canada, *Int. J. Behav. Nutr. Phys. Act.*, **10** (2013), 1–6. https://doi.org/10.1186/1479-5868-10-66

4. E. W. Martin, S. A. Shaheen, Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two US cities, *J. Transp. Geogr.*, **41** (2014), 315–324. https://doi.org/10.1016/j.jtrangeo.2014.06.026

5. C. Hsu, J. J. Liou, H. Lo, Y. Wang, Using a hybrid method for evaluating and improving the service quality of public bike-sharing systems, *J. Cleaner Prod.*, **202** (2018), 1131–1144. https://doi.org/10.1016/j.jclepro.2018.08.193

6. C. Beckx, S. Broekx, B. Degraeuwe, B. Beusen, L. I. Panis, Limits to active transport substitution of short car trips, *Transp. Res. Part D Transp. Environ.*, **22** (2013), 10–13. https://doi.org/10.1016/j.trd.2013.03.001

7. P. S. Cerutti, R. D. Martins, J. Macke, J. A. R. Sarate, "Green, but not as green as that": An analysis of a Brazilian bike-sharing system, *J. Cleaner Prod.*, **217** (2019), 185–193. https://doi.org/10.1016/j.jclepro.2019.01.240

8. A. Li, K. Gao, P. Zhao, P. Qu, K. Axhausen, High-resolution assessment of environmental benefits of dockless bike-sharing systems based on transaction data, *J. Cleaner Prod.*, **296** (2021), 126423. https://doi.org/10.1016/j.jclepro.2021.126423

9. J. Lee, S. He, D. W. Sohn, Potential of converting short car trips to active trips: The role of the built environment in tour-based travel, *J. Transp. Health*, **7** (2017), 134–148. https://doi.org/10.1016/j.jth.2017.08.008

10. J. Sultan, G. Ben-Haim, J. Haunert, S. Dalyot, Extracting spatial patterns in bicycle routes from crowdsourced data, *Trans. GIS*, **21** (2017), 1321–1340. https://doi.org/10.1111/tgis.12280

11. Y. Can, G. Gidófalvi, Mining and visual exploration of closed contiguous sequential patterns in trajectories, *Int. J. Geogr. Inf. Sci.*, **32** (2018), 1282–1304. https://doi.org/10.1080/13658816.2017.1393542

12. K. Wang, X. Qian, D. Fitch, Y. Lee, J. Malik, G. Circella, What travel modes do shared e-scooters displace? A review of recent research findings, *Transport Rev.*, **43** (2023), 5–31. https://doi.org/10.1080/01441647.2021.2015639

13. S. Kim, G. Lee, S. Choo, Optimal rebalancing strategy for shared e-scooter using genetic algorithm, *J. Adv. Transp.*, **2023** (2023). https://doi.org/10.1016/j.ejor.2015.03.043

14. P. García, C. Juan, J. Gutiérrez, M. Latorre, Optimizing the location of stations in bike-sharing programs: A GIS approach, *Appl. Geogr.*, **35** (2012), 235–246. https://doi.org/10.1016/j.apgeog.2012.07.002

15. T. Raviv, C. Juan, M. Tzur, I. Forma, Static repositioning in a bike-sharing system: models and solution approaches, *EURO J. Transp. Logist.*, **2** (2013), 187–229. https://doi.org/10.1007/s13676-012-0017-6

16. C. Fricker, N. Gast, Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity, *EURO J. Transp. Logist.*, **5** (2016), 261–291. https://doi.org/10.1007/s13676-014-0053-5

17. Y. Liu, L. Tian, A graded cluster system to mine virtual stations in free-floating bike-sharing system on multi-scale geographic view, *J. Cleaner Prod.*, **281** (2021), 124692. https://doi.org/10.1016/j.jclepro.2020.124692

18. Q. Chen, X. Pan, F. Liu, Y. Xiong, Z. Liu, J. Tang, Reposition optimization in free-floating bike-sharing system: A case study in Shenzhen City, *Physica A*, **593** (2022), 126925. https://doi.org/10.1016/j.physa.2022.126925

19. G. Erdoğan, M. Battarra, R. Calvo, An exact algorithm for the static rebalancing problem arising in bicycle sharing systems, *Eur. J. Oper. Res.*, **245** (2015), 667–679. https://doi.org/10.1016/j.ejor.2015.03.043

20. A. Kadri, I. Kacem, K. Labadi, A branch-and-bound algorithm for solving the static re-balancing problem in bicycle-sharing systems, *Comput. Ind. Eng.*, **95** (2016), 41–52. https://doi.org/10.1016/j.cie.2016.02.002

21. Z. Zhang, X. Zhang, Shared bikes scheduling under users' travel uncertainty, *IEEE Access*, **8** (2019), 3123–3143. https://doi.org/10.1109/ACCESS.2019.2961628

22. B. Vishkaei, I. Mahdavi, N. Mahdavi-Amiri, E. Khorram, Balancing public bicycle sharing system using inventory critical levels in queuing network, *Comput. Ind. Eng.*, **141** (2020), 106277. https://doi.org/10.1016/j.cie.2020.106277

23. Y. Wang, W. Szeto, The dynamic bike repositioning problem with battery electric vehicles and multiple charging technologies, *Transp. Res. Part C Emerging Technol.*, **131** (2021), 103327. https://doi.org/10.1016/j.trc.2021.103327

24. Y. Li, Y. Liu, The static bike rebalancing problem with optimal user incentives, *Transp. Res. Part E Logist. Transp. Rev.*, **146** (2021), 102216. https://doi.org/10.1016/j.tre.2020.102216

25. H. Pierre, D. Aloise, S. D. Jena, Towards station-level demand prediction for effective rebalancing in bike-sharing systems, in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (2018), 378–386. https://doi.org/10.1145/3219819.3219873

26. S. VE, Y. Cho, A rule-based model for seoul bike sharing demand prediction using weather data, *Eur. J. Remote Sens.*, **53** (2020), 166–183. https://doi.org/10.1080/22797254.2020.1725789

AIMS Press