Bike Sharing Systems: The Impact of Precise Trip Demand Forecasting on Operational Efficiency in Different City Structures

Selin Ataç*[1], Nikola Obrenović[2], and Michel Bierlaire[3]

[1]Institut Interdisciplinaire du Développement de l'Entreprise, Haute Ecole d'Ingénierie et de Gestion du Canton de Vaud, Switzerland
[2]BioSense Institute, University of Novi Sad, Serbia
[3]Transport and Mobility Laboratory, Ecole polytechnique fédérale de Lausanne, Switzerland

Short summary

Growing environmental concerns drive interest in sustainable solutions, with vehicle sharing systems addressing transportation needs. Existing studies focus on operational level challenges in one-way station-based bike sharing systems (BSSs), neglecting the added value of precise trip demand forecasting. This study assesses the worth of data collection and trip demand forecasting models. A simulation-optimization framework is created. Simulation module consists of a discrete-event simulator, representing a city BSS. Optimization module optimizes the relocation routes for rebalancing operations where clustering is used for computational efficiency. We experiment on extreme and intermediate scenarios using case studies from four city BSSs, different in location and size, that reveal varying impacts of trip demand forecasting on small- and large-scale systems. Findings emphasize the importance of demand forecasting in large-scale systems, offering insights for operators to enhance service levels, to optimize resource allocation, and to identify the maximum budget to allocate for trip demand forecasting.
Keywords: bike sharing systems, operational efficiency, operations research applications, shared mobility

1 Introduction

Increasing environmental awareness is driving industries to investigate eco-friendly alternatives. EPA (2021) indicates that transportation sector contributed to global greenhouse gas emissions by 28.7% in the US in 2019, when the end-use sectors are also considered. In response, sustainable solutions like shared mobility, including ride-hailing, ride-sharing, and vehicle sharing, have gained prominence, drawing attention from the research community due to the conflicting objectives of user convenience and operator profitability.

Bike sharing systems (BSSs) stand out as an integral component of shared mobility, providing an economically viable and health-conscious transportation mode (Bieliński et al., 2021). However, assumptions about the climate-friendliness of shared micro-mobility, including e-scooters and e-bikes, are challenged by studies revealing higher CO_2 emissions, particularly in Zurich, Switzerland, attributed to manufacturing and operational decisions (Reck et al., 2022). As a result, operational efficiency and effective BSS management emerge as critical factors.

The literature mainly focuses on operational challenges within BSSs, emphasizing trip demand forecasting, vehicle rebalancing operations, and station clustering (Ataç et al., 2021a). Surveys are conducted to understand the effects of factors such as socio-economic characteristics, weather conditions, temporal variables, and station attributes (Ashqar et al., 2019). Various modeling approaches use these data and orient themselves to predict trip demand as precisely as possible.

Rebalancing operations is one approach for addressing vehicle imbalance in the service area. Optimization models use diverse formulations, such as mixed integer linear programs (MILP; Dell'Amico et al., 2014), to solve the rebalancing problem while catering to operator- or user-oriented objectives. Computational efficiency is achieved through exact methods (Dell'Amico et al., 2014) and heuristics (Schuijbroek et al., 2017). Hierarchical clustering, k-means clustering, and MIP models aid in identifying station characteristics and solving optimization problems more efficiently (Lahoorpoor et al., 2019; Schuijbroek et al., 2017).

While the existing literature independently addresses demand forecasting and rebalancing operations in BSSs, there exists a gap in understanding the added value of constructing a precise demand model in station-based BSSs with static rebalancing operations. Thus, this paper aims to establish a connection between system characteristics and the significance of precise demand forecasting through multiple case studies, providing decision-makers with insights into the value of such forecasting in various city structures.

The contributions are as follows. We propose a simulation-optimization framework which consists of a tailored discrete-event simulator, rebalancing operations optimization model, and clustering. We enhance an existing rebalancing operations optimization model from Dell'Amico et al. (2014) and present two new datasets from online trip data. The study goes beyond experimental demonstrations by providing decision-makers with insights into the added value of demand forecasting specific to the city structure of interest. It emphasizes the need for unique city-specific management and optimization approaches for BSS operation, contributing to the field's comprehensive understanding of relationships among city characteristics, rebalancing operations, and trip demand forecasting within the context of BSSs.

2 Methodology

We consider a one-way station-based BSS with static, operator-based rebalancing using relocation vehicles during the night. The vehicle and parking distribution across BSS stations is referred to as its configuration. The initial and final configurations represent the system state at the beginning and end of the day, respectively. The proposed framework consists of three main modules: trip demand forecasting, simulation, and optimization, as depicted in Figure 1.

The optimization module further divides into two components: agglomerative hierarchical clustering and rebalancing operations optimization, as illustrated in Figure 2. The former is responsible for creating the station clusters whereas the latter solves the rebalancing operations optimization for each cluster. In the simulation module, the trip demand for one day is simulated using a discrete-event simulator. The station imbalance is calculated for each station using the final configuration of the simulated day and the desired initial configuration for the following day. Then, the station imbalance of each station along with precomputed clustering information are passed to the optimization module that calculates the best routing for the rebalancing operations. The number of lost trip demand and the cost of rebalancing operations are recorded as performance measures and stored in the database. The dashed arrow in Figure 1 represents the iterative process of the framework.

Our objective is to investigate the added value of the trip demand forecasting module. We consider

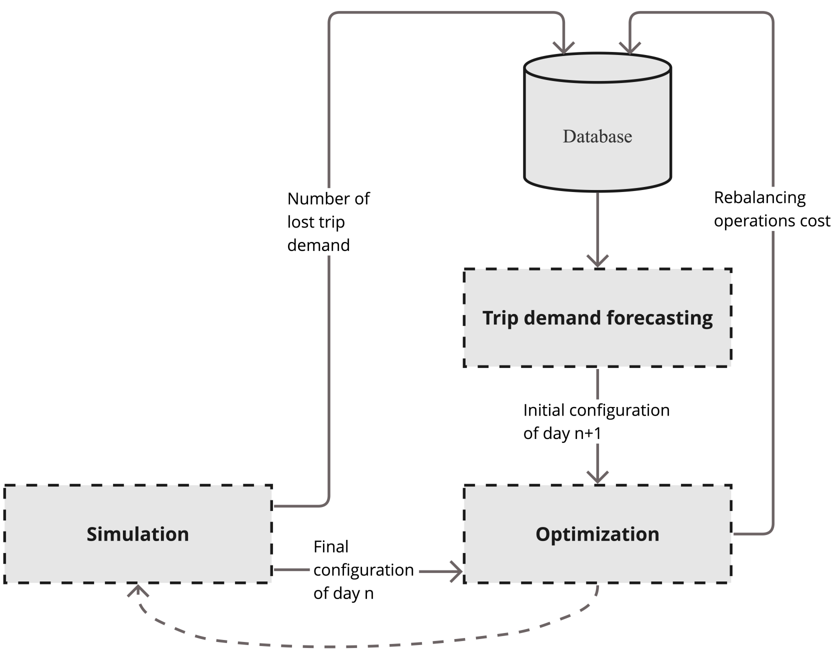

Figure 1: The framework for VSS simulation and rebalancing optimization

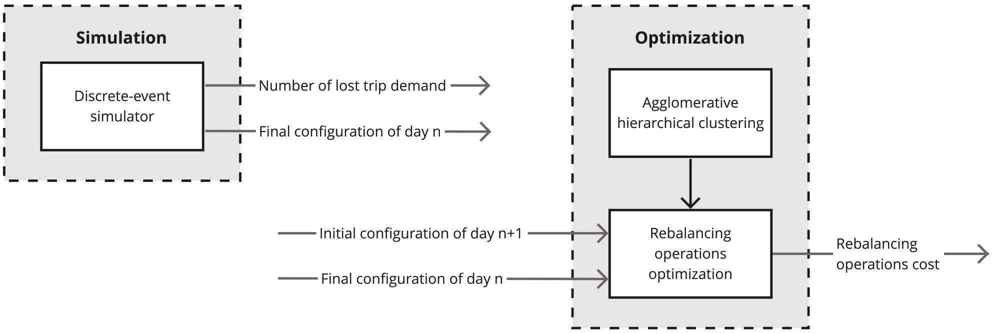

Figure 2: Simulation and optimization modules in detail

two extreme scenarios: (i) unknown demand and (ii) known demand. The first assumes that the trip demand information is not available to the operator. Therefore, the vehicles are rebalanced to the same initial configuration every day. The second assumes perfect knowledge about the future trip demand and the operator knows where and when a trip demand will occur. The desired initial configuration for the next day is determined based on this information. The number of vehicle deficiency or surplus throughout the day is calculated using perfect information on trip demand. Then, the maximum deficiency is assigned as the initial number of vehicles at the beginning of the day. If the number of vehicles is insufficient to assign the necessary number of vehicles to each station, we proportionally decrease the number of vehicles at each station. These two extreme scenarios are then relaxed and intermediate scenarios are explored, as presented in Section 3.

The simulation accounts for the capacity, number of bikes and available parking spots at each station. We denote the set of stations by N, and by V when the depot is also included. We define the set Γ which is a union of three subsets: Γ^{req} for trip requests, Γ^{des} for destinations, and Γ^{V} for station and depot locations. The elements of these sets are γ_{ω}^{req}, γ_{ω}^{des}, and γ_{k}^{V}, representing tuples of latitude and longitude values for the corresponding locations. The distance matrix Δ contains δ_{od}^{n}, the distances from origin o to destination d with mode n, where $o, d \in \Gamma$, and $n = \{$'walk', 'bike', 'drive'$\}$. A set of trips Ω occurs during the day. Each trip demand $\omega \in \Omega$ has both spatial and temporal dimensions, i.e., the request time, locations of the request and destination, τ_{ω}^{req}, γ_{ω}^{req}, and γ_{ω}^{des}, respectively. A trip demand ω is considered satisfied when there is an available vehicle at the time of pick-up at station i where $\delta_{\gamma_{\omega}^{req}\gamma_{i}^{V}}^{walk} \leq \phi$ and $\omega \in \Omega$. The simulated events include trip requests, bike pick-ups, bike drop-offs, and trip completions. A trip demand is considered lost when a user requests to use the system but cannot do so due to the unavailability of bikes or parking spots. The simulation outputs include the number of lost trip demands and the final configuration of the corresponding day. The former is calculated as the difference between the number of requested trips based on the perfect demand knowledge and the number of satisfied trips based on the simulation results.

Rebalancing operations occur instantaneously at the end of the day. We assume that there are m relocation vehicles, each with a capacity of Q. The cost of traveling from i to j where $i, j \in V$, is denoted by c_{ij} and is equal to δ_{ij}^{drive}. The imbalance of a station k, i.e., q_k, is computed as the difference between the number of vehicles left at the end of the simulated day and the desired number of vehicles at the beginning of the following day. Note that $q_k < 0$ if the station is a demand station, i.e., the number of pick-ups is more than the number of drop-offs, $q_k > 0$ if it is a supply station, i.e., the number of pick-ups is less than the number of drop-offs, and $q_k = 0$ if it is a self-balancing station, i.e., the number of pick-ups is equal to the number of drop-offs. Relocation vehicles visit station k if $q_k \neq 0$, and $\bar{q}$ is the number of stations to be visited. We utilize the $(F1)$ model from Dell'Amico et al. (2014) and enhance it as $(F1_M)$.

$(F1_M)$ is based on the Multiple Traveling Salesman Problem (m-TSP) and determines the optimal routing for the relocation vehicles, where x_{ij} is 1 if arc (i,j) is used by a relocation vehicle and 0 otherwise. θ_j represents the load of a vehicle after it leaves node j. The objective function (2) minimizes the routing cost. Constraints (3) and (4) ensure that every node, except the depot, is served exactly once. Constraints (5) and (6) guarantee that no more than m vehicles are used, and all used vehicles return to the depot at the end of their route.

The classical subtour elimination constraints used in $(F1)$ correspond to Dantzig-Fulkerson-Johnson (DFJ) formulation. These constraints result in $\mathcal{O}(2^N)$ complexity, making the model intractable for large instances. On the other hand, Miller-Tucker-Zemlin (MTZ) constraints utilize additional

decision variables and reduce the complexity to $\mathcal{O}(N^2)$. As we consider a cost function that is asymmetric, the model becomes a special case of asymmetric traveling salesman problem (ATSP). (Velednitsky, 2017) demonstrate that the DFJ polytope is contained in the MTZ polytope for the ATSP. On the other hand, Bazrafshan et al. (2021) argue that other criteria, such as number of constraints, number of variables, type of variables, time of solving, and differences between the optimum and the relaxed value, should also be considered. They show that when DFJ and MTZ formulations are compared, MTZ outperforms DFJ, especially for node size that is greater than 20. For this purpose, we introduce new decision variables, u_i, $\forall i \in V$ and replace DFJ constraints with (7)-(8). Constraints (9) are added to prevent the subtours to the same station.
Constraints (10)-(12) define the maximum and the initial load of a vehicle. Flow conservation is ensured by (13) and (14). The valid inequalities proposed by Dell'Amico et al. (2014), (15) and (16), exclude solutions that consecutively go through three nodes with a total station imbalance exceeding the capacity of the relocation vehicle. For this purpose, they define the set

$$S(i,j) = \{h \in N, h \neq i, h \neq j : |q_i + q_j + q_h| > Q\}. \tag{1}$$

Finally, (17) impose binary restrictions on x_{ij}'s. The validity of the model is ensured as it is based on Dell'Amico et al. (2014) and the subtour elimination constraints are replaced with a set of constraints proven to be equivalent to the existing ones.

$$
\begin{aligned}
(F1_M)\ \min \quad & \sum_{i \in V}\sum_{j \in V} c_{ij} x_{ij} && && (2)\\
\text{s.to} \quad & \sum_{i \in V} x_{ik} = 1 && \forall k \in N && (3)\\
& \sum_{i \in V} x_{ki} = 1 && \forall k \in N && (4)\\
& \sum_{j \in V} x_{0j} \leq m && && (5)\\
& \sum_{k \in N} x_{0j} - \sum_{k \in N} x_{k0} = 0 && && (6)\\
& u_k - u_l + |N| * x_{kl} \leq |N| - 1 && \forall k, l \in N && (7)\\
& 1 \leq u_i \leq |N| - \bar{q} && \forall i \in V && (8)\\
& x_{ii} = 0 && \forall i \in V && (9)\\
& \theta_j \geq \max\{0, q_j\} && \forall j \in V && (10)\\
& \theta_j \leq \min\{Q, Q + q_j\} && \forall j \in V && (11)\\
& \theta_0 = 0 && && (12)\\
& \theta_k - \theta_i + M(1 - x_{ik}) \geq q_k && \forall i \in V, k \in N && (13)\\
& \theta_k - \theta_j + M(1 - x_{kj}) \geq -q_j && \forall k \in N, j \in V && (14)\\
& x_{kl} + \sum_{h \in S(k,l)} x_{lh} \leq 1 && \forall k, l \in N, h \in S(k,l) && (15)\\
& \sum_{h \in S(k,l)} x_{hk} + x_{kl} \leq 1 && \forall k, l \in N, h \in S(k,l) && (16)\\
& x_{ij} \in \{0, 1\} && \forall i, j \in V && (17)
\end{aligned}
$$

We employ agglomerative hierarchical clustering to group stations. The rebalancing operations optimization model is solved for each cluster in which the total bike imbalance is minimized. After incorporating travels between clusters, the final route for relocation vehicles is determined. For a detailed discussion on the simulation-optimization framework, the reader is kindly referred to Ataç et al. (2020).

3 Results and discussion

We conduct computational experiments on four real-life case studies: nextbike Sarajevo, nextbike Berlin, Divvy Chicago, and Citi Bike New York City. These systems operate with 21, 298, 681, and 1361 stations and approximately 120, 3000, 6000, and 22000 bikes, respectively. The data for the first two systems, which are available on request, are obtained from their online system

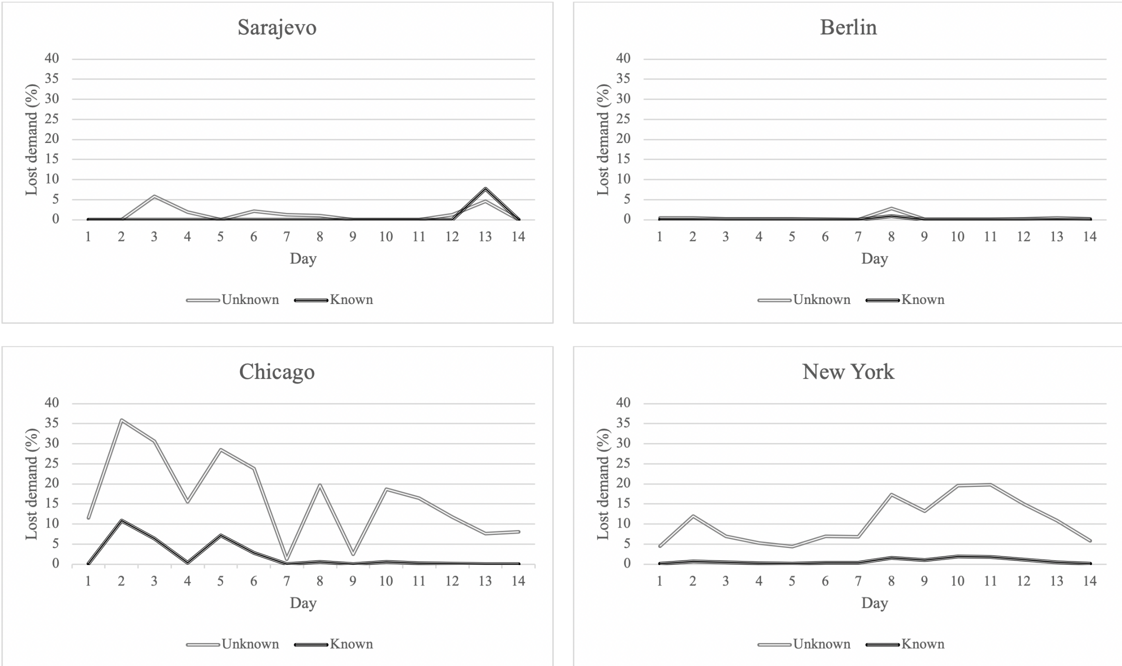

Figure 3: Unknown vs known demand scenarios

in real-time and then processed to obtain O-D trip information. The data for the last two are publicly available (Bike, 2024; Divvy, 2024).

In Figure 3, we present results for 14 consecutive days. The horizontal axis represents the days, and the vertical axis shows the ratio of lost and total trip demand. As expected, the unknown demand scenario results in more lost trip demand compared to the known scenario. However, perfect trip demand knowledge does not guarantee the satisfaction of all trip demand, influenced by fleet size and station availability. We assume that users are rational and deterministic in terms of station choice, deciding based on proximity. We observe that the benefit of trip demand forecasting in larger-scale instances, i.e., Chicago and New York City, is clearer and more significant than in smaller-scale ones, i.e., Sarajevo and Berlin. This suggests that smaller-scale systems may not necessarily require precise trip demand forecasting.

We also explore intermediate scenarios, i.e., X-percent scenarios, wherein we assume that we per-

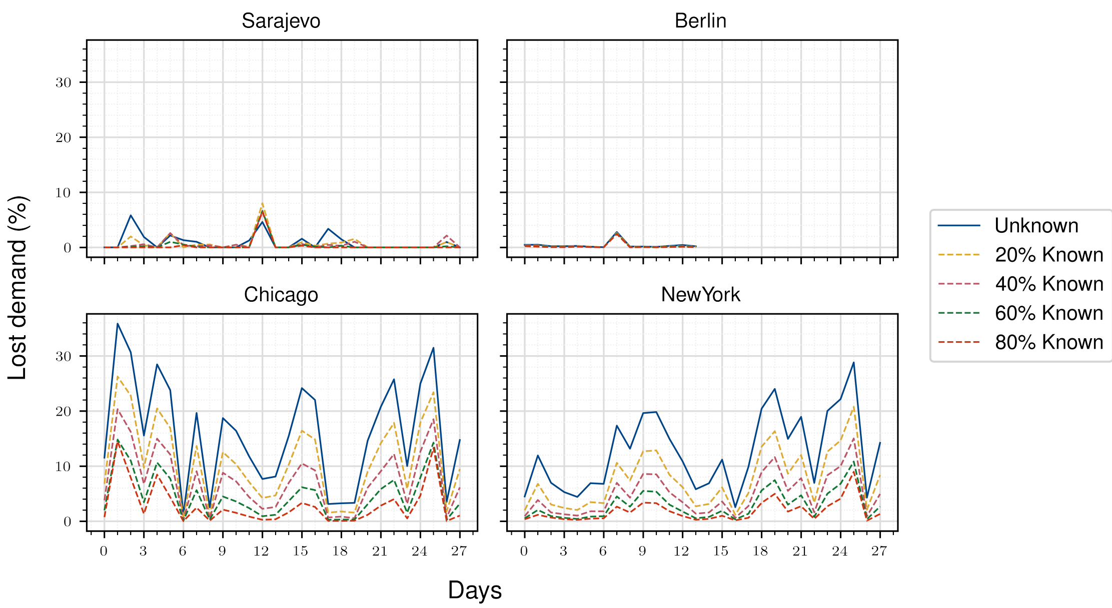

Figure 4: Lost demand for the unknown and X-percent scenarios

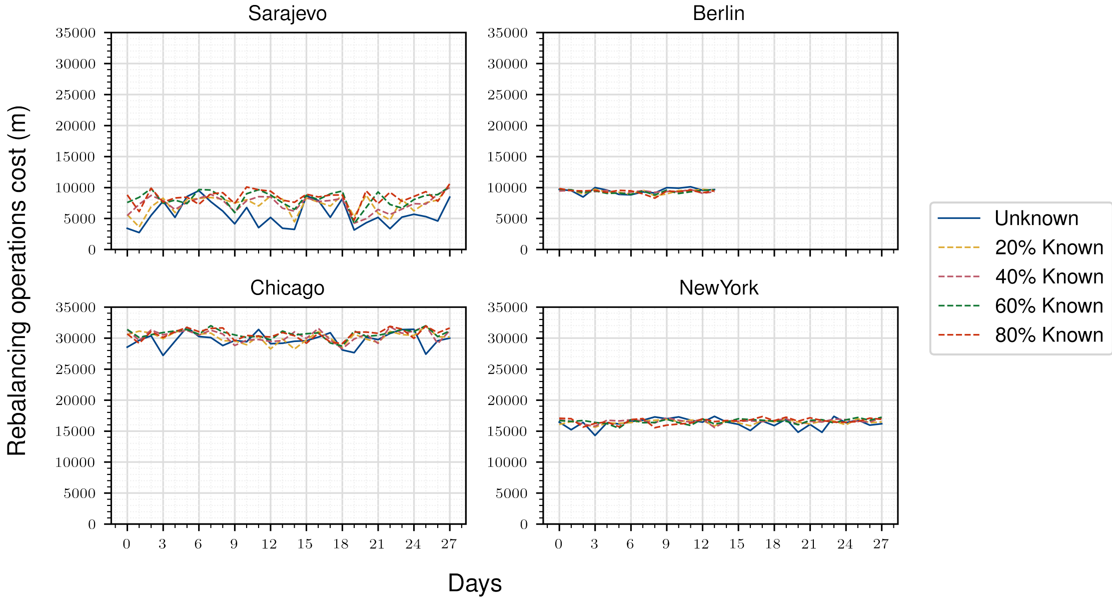

Figure 5: Rebalancing operations cost for the unknown and X-percent scenarios

fectly forecast $X\%$ of the trips while the remaining trips cannot be identified. X-percent scenarios can be associated with reservation-based systems. This also serves the evaluation of whether promoting reservations in the system would enhance its profitability. For the X-percent scenarios, we experiment with four levels of knowledge, i.e., 20%, 40%, 60%, and 80%. Each scenario is executed 100 times with different seeds to account for the variability depending on the sample, and the averages over 100 repetitions are presented in Figures 4 and 5.

Figure 4 consolidates each case study for a clearer observation of trends. The vertical axes are adjusted among the case studies for easier comparison. Each X-percent scenario is represented by one dashed line. Interestingly, we see fluctuations in the Sarajevo case study for different levels of knowledge; these fluctuations diminish for the Berlin case study; and for the other two case studies exhibit a consistent trend. In other words, as the system size increases, we notice a significant decrease in lost trip demand with an increase in the knowledge of trip demand data.

These results support the observation that small- and large-scale systems react differently to changes in forecasting methods. Small-scale systems do not exhibit a clear trend as the numbers of stations and trips are significantly less compared to large-scale systems. In large-scale systems, the higher number of trips stabilizes the system, hence the lost demand is mitigated with increased knowledge involved in the forecasting process. In essence, the results show that including reservations in the system is more beneficial for larger-scale systems than smaller ones.

In Figure 5, we present findings on the rebalancing operations cost. The horizontal axis represents the days, and the vertical axis shows the rebalancing operations cost in meters. Notably, the rebalancing operations cost does not exhibit significant changes across different X-percent scenarios. This indicates that the use of trip demand forecasting does not considerably impact the rebalancing cost as the routes tend to remain similar. This finding is important, suggesting that forecasting demand allows for a better service in large-scale case studies without necessarily incurring higher costs. Similarly, small-scale case studies do not significantly benefit from forecasting and the rebalancing operations cost remains consistent, supporting the observation that precise trip demand forecasting is not essential for small-scale systems.

In conclusion, the results suggest that the added value of demand forecasting increases with the system size. Although larger-scale systems allocate larger budgets for their operation and are more likely to consider rebalancing operations, this may not be the case for small-scale and local systems. Therefore, we believe that these findings offer valuable insights for decision-makers.

4 Conclusions

We introduced a simulation-optimization framework to assess the necessity for precise trip demand forecasting in BSSs. This framework includes (i) a discrete-event simulator representing trip de-

mand in a BSS for one day and (ii) an optimization module consisting of clustering to identify groups of stations to reduce the computational complexity and an enhanced rebalancing operations optimization model. With this framework, we investigate both lost demand and the cost of rebalancing operations, allowing us to determine the trade-off between them. I.e., our framework assists decision-makers by providing insights into the upper limit of the budget for demand forecasting tasks, such as data collection and the development of demand models. The results from the four case studies, representing different city sizes and types reveal valuable insights.
The numerical experiments involve two small-scale systems with 21 and 298 stations and two large-scale systems with 681 and 1361 stations. We present results for two extreme scenarios, namely known and unknown demand scenarios, along with intermediate scenarios, i.e., X-percent scenarios. The key finding suggests that, for large-scale systems, planning rebalancing operations based on demand forecasts is crucial. It allows improving the level of service (the percentage of trip requests served) without any significant increase in rebalancing costs. We have also reported that, for small-scale systems, there is no significant benefit from a precise trip demand forecasting model and the benefit obtained from including reservations in the system increases as the scale of the system increases.

References

Ashqar, H. I., Elhenawy, M., & Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. *Case Studies on Transport Policy*, *7*, 261 - 268.

Ataç, S., Obrenović, N., & Bierlaire, M. (2020). Vehicle sharing systems: Does demand forecasting yield a better service? In *20th Swiss Transport Research Conference (STRC 2020).*

Ataç, S., Obrenović, N., & Bierlaire, M. (2021a). Vehicle sharing systems: A review and a holistic management framework. *EURO Journal on Transportation and Logistics*, *10*, 100033.

Bazrafshan, R., Hashemkhani Zolfani, S., & Al-e hashem, S. M. J. M. (2021). Comparison of the sub-tour elimination methods for the asymmetric traveling salesman problem applying the SECA method. *Axioms*, *10*(1).

Bieliński, T., Kwapisz, A., & Ważna, A. (2021). Electric bike-sharing services mode substitution for driving, public transit, and cycling. *Transportation Research Part D: Transport and Environment*, *96*, 102883.

Bike, C. (2024). *Citi Bike System Data: Citi Bike NYC.* `https://ride.citibikenyc.com/system-data`. ([Online; accessed 18 January 2024])

Dell'Amico, M., Hadjicostantinou, E., Iori, M., & Novellani, S. (2014). The bike sharing rebalancing problem: Mathematical formulations and benchmark instances. *Omega*, *45*, 7 - 19.

Divvy. (2024). *Divvy System Data: Divvy Bikes.* `https://divvybikes.com/system-data`. ([Online; accessed 18 January 2024])

EPA. (2021). *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2019.* `https://www.epa.gov/sites/default/files/2021-04/documents/us-ghg-inventory-2021-main-text.pdf?VersionId=yu89kg1O2qP754CdR8Qmyn4RRWc5iodZ`.

Lahoorpoor, B., Faroqi, H., Sadeghi-Niaraki, A., & Choi, S.-M. (2019). Spatial cluster-based model for static rebalancing bike sharing problem. *Sustainability*, *11*(11).

Reck, D. J., Martin, H., & Axhausen, K. W. (2022). Mode choice, substitution patterns and environmental impacts of shared and personal micro-mobility. *Transportation Research Part D: Transport and Environment*, *102*, 103134.

Schuijbroek, J., Hampshire, R. C., & van Hoeve, W.-J. (2017). Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research*, *257*, 992 - 1004.

Velednitsky, M. (2017). Short combinatorial proof that the DFJ polytope is contained in the MTZ polytope for the asymmetric traveling salesman problem. *Operations Research Letters*, *45*(4), 323-324.