**Les Cahiers du GERAD** ISSN: 0711–2440

# Data-driven prioritization strategies for inventory rebalancing in bike-sharing systems

M.C.M. Silva, D. Aloise, S.D. Jena

G–2023–30

August 2023



La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2023
– Bibliothèque et Archives Canada, 2023

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2023
– Library and Archives Canada, 2023

**GERAD** HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

**Tél. : 514 340-6053**
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

# Data-driven prioritization strategies for inventory rebalancing in bike-sharing systems

**Maria Clara Martins Silva** [a, b]
**Daniel Aloise** [a, b]
**Sanjay Dominik Jena** [c, b]

[a] *Polytechnique Montréal, Montréal (Qc), Canada, H3T 1J4*

[b] *GERAD, Montréal (Qc), Canada, H3T 1J4*

[c] *Université du Québec à Montréal, Montréal (Qc), Canada, H2L 2C4*

maria-clara.martins-silva@polymtl.ca
daniel.aloise@polymtl.ca
jena.sanjay-dominik@uqam.ca

**August 2023**
**Les Cahiers du GERAD**
**G–2023–30**

**Abstract :** The popularity of bike-sharing systems has constantly increased throughout the last years. Most of such success can be attributed to their multiple benefits, such as user convenience, low usage costs, health benefits and their contribution to environmental relief. However, satisfying all user demands remains a challenge, given that the inventories of bike-sharing stations tend to be unbalanced over time. Bike-sharing system operators must therefore intervene to rebalance station inventories to provide both available bikes and empty docks to the commuters. Due to limited rebalancing resources, the number of stations to be rebalanced often exceeds the system's rebalancing capacity, especially close to peak hours. As a consequence, operators are forced to manually select stations and determine the appropriate quantities of bikes for rebalancing. In practice, such manual planning is likely to result in suboptimal system performance. In this paper, we propose four strategies to select the stations that should be prioritized for rebalancing, using features such as the predicted trip demand, as well as the inventory levels at the stations themselves and their surrounding stations. We evaluate the performance of these prioritization strategies by simulating real-world trips using data from 2019 and 2020, each of which exhibits distinct travel patterns given the restrictive measures implemented in 2020 to prevent the spread of COVID-19. One of these strategies significantly improves the system's performance by reducing the lost demand by up to 65%, while another strategy reduces the number of required rebalancing operations by up to 33%, when compared to the prioritization scheme currently used within our bike-sharing system use case. Finally, one prioritization strategy encourages the selection of stations that are geographically clustered, which may facilitate rebalancing operations afterwards.

**Keywords:** Bike-sharing, demand prediction, rebalancing, inventory management

**Acknowledgements:** The authors are grateful to BIXI Montreal who provided resources throughout the project. The authors also thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for its financial support.

# 1 Introduction

Demand for bike-sharing systems (BSS) has constantly increased over the recent years, as they continue to provide various advantages: they are typically simple to use and do not require previous reservation; they have been shown to be an environmentally friendly transportation mode by reducing the ever-increasing amount of cars in circulation (Wang and Zhou, 2017); and, they contribute to a healthy lifestyle (Pucher et al., 2010). Particularly throughout the COVID-19 pandemic, BSSs were considered a transportation alternative with a particularly low risk of user contamination (Pase et al., 2020; Basak et al., 2023).

In this paper, we focus on dock-based BSSs, in which stations are located in different parts of the city, and from which commuters may rent and return bikes. While dock-based systems have several advantages (e.g. users get used to the location of stations and bikes), a main issue is that the station inventories may quickly become unbalanced, i.e., either rental demand cannot be met, given that not a sufficient number of bikes is available, or return demand cannot be met, when the station has no empty docks. Such imbalances often occur during rush hours on weekdays, when commuters relocate from their residential areas to the areas they work in the morning and do the return trip in the afternoon (Mellou and Jaillet, 2019). Unmet user demand likely causes user dissatisfaction, which the system operators seek to avoid as best as possible, given that it ultimately reduces the user base as the system's reputation is damaged.

An effective way to fight station inventory imbalances is to redistribute bikes among stations, a process known as *rebalancing*. The literature distinguishes two main types of rebalancing: user-based rebalancing and operator-based rebalancing. The former consists of incentives given to the users in order to return bikes at stations before they become empty (Vallez et al., 2021). In contrast, operator-based rebalancing is carried out by the BSS operators themselves, typically by dispatching trucks that relocate bikes between the stations. In this work, we focus on operator-based rebalancing, which has shown to be effective to increase demand satisfaction (see, e.g. DeMaio, 2009) and is the common practice at major BSSs around the world, such as BIXI Montreal, Citi Bike in New York City and Ecobici in Mexico City. Such rebalancing is also a less expensive solution compared to installing more stations or adding more docks to already existing stations (Shu et al., 2013).

In most dock-based BSSs with operator-based rebalancing, the decision to actively rebalance a station depends on which stations are considered *unbalanced*. Depending on the BSS, the criteria may be different for a station to be categorized as such. For example, at NiceRide (Minneapolis, U.S), a station is considered to be unbalanced when it is either completely empty or completely full (Wang et al., 2018). The operators of Vélo'v (Lyon, France) classify a station as unbalanced if the absolute difference between the number of arrivals and departures is larger than the standard deviation of the distribution of these values over all the stations (Borgnat et al., 2011). BIXI Montreal uses *inventory intervals* that establish an acceptable quantity of bikes at each station. Inventory intervals are manually set by BIXI's dispatching team, based on their experience with the station location, intraday demand fluctuation, and the day of the week.

Nonetheless, the rebalancing process itself remains costly, as it accounts for gas, the maintenance of the truck fleet, drivers' salaries, etc. In addition, it reduces the favorable impact that bike sharing claims to have on the environment. All considered, having a fleet of trucks large enough to rebalance all unbalanced stations at every hour is not financially viable for most BSSs, especially during peak hours. According to JCDecaux, a company that offers self-service bikes to different cities around the world, the estimated cost in 2009 to relocate a single bike within a BSSs was about three dollars (DeMaio, 2009). A fleet of trucks available for rebalancing is therefore limited in size and cannot rebalance all unbalanced stations. It becomes imperative that the selection of stations to rebalance is carried out as effectively as possible.

Consider Figure 1, showing BIXI's (estimated on historical data) maximum hourly rebalancing capacity and the average daily number of unbalanced stations per hour for weekdays in July and

August of 2019 (left) and 2020 (right). Throughout this period, the number of unbalanced stations consistently exceeded BIXI's maximum rebalancing capacity of approximately 46 stations per hour in 2019 and 22 stations per hour in 2020. As a consequence, the operator must select a subset of these stations to be rebalanced.

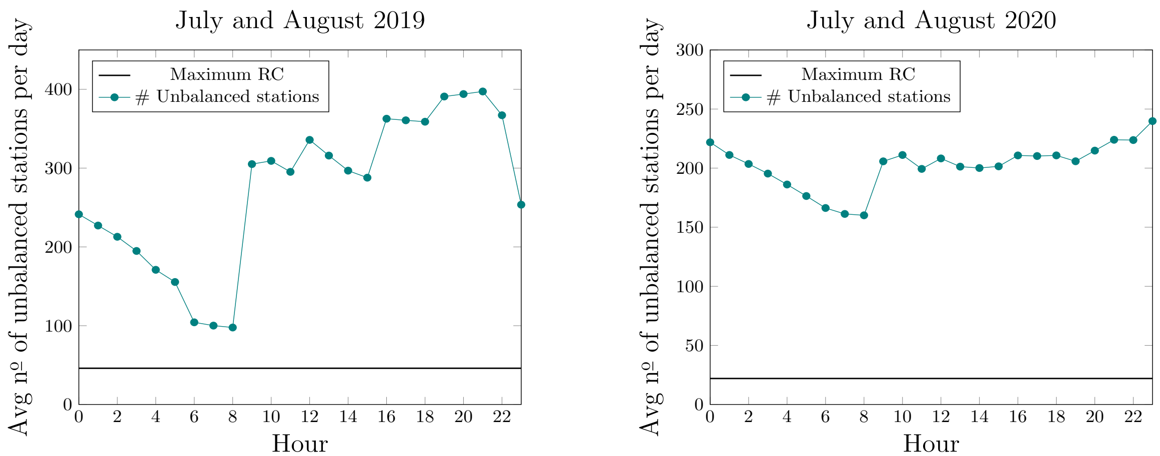



**Figure 1: Maximum hourly rebalancing capacity and average daily number of unbalanced stations per hour at BIXI BSS during weekdays in July and August 2019 and 2020.**

Ideally, the subset of unbalanced stations should be selected such that the number of served future demand requests is maximized, which requires an appropriate demand forecast. However, predicting the demand for a BSS is a complex task depending on several factors, such as the weather, the hour of the day, the day of the week, holidays, public events, etc.

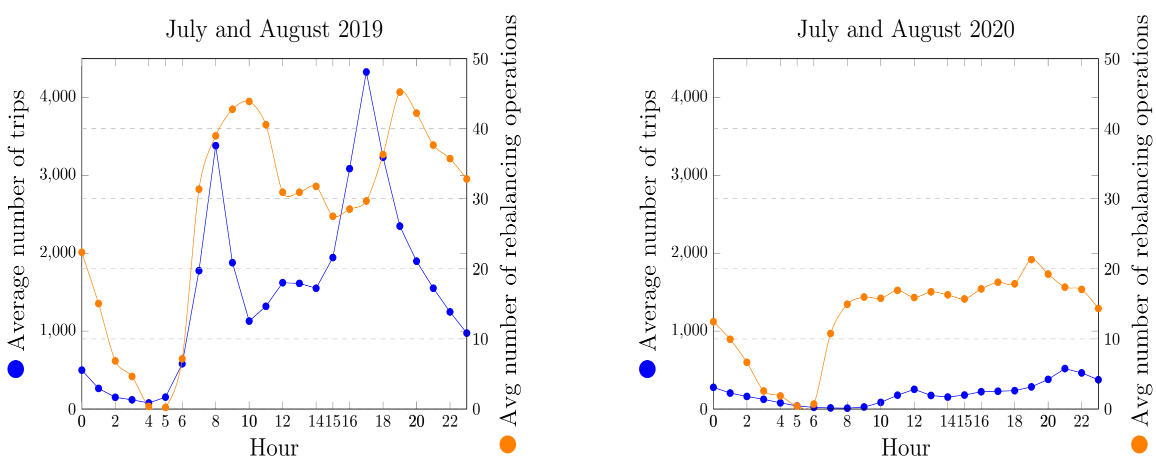



**Figure 2: Average number of trips and rebalancing operations for all stations per weekday for July & August 2019 (left) and 2020 (right).**

Demand prediction in BSS became particularly challenging in 2020 due to restrictive measures imposed by the governments in response to the COVID-19 pandemic, leading to a large part of the population working from home. Figure 2 shows the average number of trips and the average number of rebalancing operations as reported by BIXI during the weekdays in July and August 2019 (left) and 2020 (right). Not only did the number of trips in 2020 decreased considerably with respect to the same period in 2019, but the trip behavior also changed. In 2019, the peak hours occurred right before and right after the working hours, i.e., at 8 a.m. and 5 p.m., respectively. However, this pattern was no longer observed in 2020, resulting in a flatter demand along the day.

Such drastic demand changes as observed from 2019 to 2020 severely affect the subset of stations that become unbalanced over time, complicating the demand forecast, and, in turn, affecting the choice of stations to be rebalanced. As a result, the pattern of BIXI's rebalancing operations changed drastically, reducing their activities by nearly half during the peak hours in 2020. Given that the manual planning is mostly based on the previous experience of the dispatching team, there is a significant risk that manually adjusted rebalancing strategies are ineffective in practice in a different environment of trip demand. This suggests that data-driven strategies that quickly adapt to changing demand hold certain benefits to assist BSS rebalancing operations.

The main contribution of this paper is to provide data-driven rebalancing recommendations to BSSs through the utilization of demand prediction models. These recommendations primarily focus on optimizing user satisfaction while simultaneously minimizing service costs. To this end, we first propose a model that automatically generates inventory intervals considering both rentals and returns forecast. Thereafter, we propose four strategies to prioritize the unbalanced stations according to their need for rebalancing. These strategies are based on the current inventory levels at the stations, the predicted demand for the next hours and, in the last strategy, the location of the stations. Each strategy aims to tackle a specific issue in order to improve the BSS performance. The first strategy focuses on prioritizing stations that are the most likely to generate lost demand. The second strategy aims to prioritize stations according to the amount of lost demand that can be avoided if they are immediately rebalanced. The third strategy prioritizes stations according to how unbalanced they are based on their inventory intervals. Finally, the fourth prioritization strategy further considers the proximity among unbalanced stations in order to favour the rebalancing of stations close to each other, leading to smaller operational routing costs for *a posteriori* truck route optimization. Throughout our prioritization strategies, we ensure that the set of stations selected for rebalancing results in a balanced number of bikes to be added and removed from them.

Our four prioritization strategies are compared to a systematic approach, which emulates the prioritization strategy currently employed at BIXI. More specifically, the comparison is conducted by a tailored discrete-time simulation that computes the estimated lost demand (i.e., demand that could not be satisfied), the total number of alerts raised each time a station becomes unbalanced, and the number of performed rebalancing operations in the system. Our computational experiments reveal that one of the proposed strategies is able to reduce the estimated lost demand by 35% for the 2019 season data, and by 65% for the 2020 season data, as compared to a baseline strategy that emulates the prioritization performed by the studied BSS. Besides, another prioritization strategy has shown effectiveness in selecting stations that are naturally grouped together while maintaining satisfactory performance measures. All proposed strategies are easy to implement and computationally cheap, therefore providing an attractive alternative to more elaborate BSS planning approaches, such as those based on optimization models.

The remainder of this paper is organized as follows. Section 2 reviews the most relevant literature in the area of rebalancing and prioritization strategies for bike-sharing systems. Section 3 describes how the inventory intervals are defined so as to serve as input to the prioritization strategies. Section 4 describes the different strategies proposed to score the rebalancing priorities of unbalanced stations. Section 5 presents and analyzes the computational experiments. Finally, Section 6 concludes the paper.

## 2 Literature review

Rebalancing in BSSs can be divided into two main steps: (a) tactical inventory management, and (b) operational bike repositioning. Step (a) aims to establish the number of bikes in each station to meet the predicted demand as best as possible. Step (b) focuses on the actual dispatching operations that are necessary to achieve the desired inventory levels at the stations in order to rebalance the system.

In order to define, in step (a), the optimal inventories that are likely to provide sufficient bikes and free docks to satisfy future demand, it is necessary to forecast the latter sufficiently well. Trip demand is influenced by numerous external factors, such as the weather, the day of the week, the time of the day, land use, the location of the stations, points of interest, and social-demographic characteristics (El-Assi et al., 2017; Hampshire and Marla, 2012). Most of the proposed approaches to predict trip demand are either based on machine learning (see, e.g. Feng et al., 2018; Hulot et al., 2018; Yin et al., 2012) or on statistical models (see, e.g. Borgnat et al., 2011; Chen et al., 2016; El-Assi et al., 2017; Gebhart and Noland, 2014). These models differ from each other in terms of the predicted time horizon (hourly, daily, or weekly), as well as on the geographic granularity of the predictions (station-level, cluster-level, or network-level). For instance, Yin et al. (2012) and Gebhart and Noland (2014) predict the total demand in the network for each observed hour. Feng et al. (2018) and Borgnat et al. (2011) predict the total demand for clusters of stations, while Chen et al. (2016) estimate the probability that stations in a cluster become either completely full or empty. El-Assi et al. (2017) propose a model that estimates the future demand for each station for five time periods along the day (morning, midday, afternoon, evening, and overnight). Hulot et al. (2018) predict the rentals and returns for each station per hour, using temporal (day, day of the week, holiday, etc.) and weather (temperature, humidity, rain, etc.) features. The authors also propose a reduction technique for the trip data, improving the computational execution time and erasing outliers from the dataset. For practical purposes, station-level demand predictions for shorter time-periods (such as one hour) seems preferable given that (i) the demand can drastically change from one hour to the next, and that (ii) the rebalancing process is actually planned and carried out at station-level.

Once the demand is properly predicted, optimal target inventory values can be determined. Schuijbroek et al. (2017) model the station inventory by means of a Poisson queuing system that estimates its optimal number of bikes while ensuring a given service level. Also using a queuing system, Huang et al. (2020) computes the target inventory values for the highest-demand stations, denoted central stations, considering their demand as well as those of their nearby stations. In the work of Liu et al. (2016), inventories are optimized in order to maximize the amount of time a station is considered balanced. In Raviv and Kolka (2013), they estimate the optimum inventory by minimizing a user dissatisfaction function with penalty variables to control the weight given to the rental or return dissatisfaction. Likewise, Hulot et al. (2018) introduces a hyperparameter to prioritize either the rental or the return service level when computing the target inventory value of a station. Datner et al. (2019) obtain target inventory values by minimizing the journey dissatisfaction, which is measured by the number of commuters who wait for missing bikes (for rentals) or docks (for returns), or by the number of times they change of station.

Regarding the operational decision-making step (b), the works in the literature can be categorized into two classes: those that assume that rebalancing operations are performed by the users of the system (under some incentive) and those that rebalance by means of a truck fleet coordinated by the operator. Chemla et al. (2013) propose a reward mechanism to encourage users to return bikes to certain stations in the BSS. In Fricker and Gast (2016), the authors conclude that encouraging the users to return the bikes to a non-saturated station does not significantly improve the system's performance. However, they also show that the performance can be improved by constantly stimulating users to return bikes to a nearby station with lower inventory.

In the case of dock-based BSSs, as the one here considered, operator-based rebalancing via trucks has typically been modeled via mixed-integer linear programming (see, e.g. Alvarez-Valdes et al., 2016; Brinkmann et al., 2016; Bulhões et al., 2018; Chemla et al., 2013; Contardo et al., 2012; Dell'Amico et al., 2014; Erdoğan et al., 2015; Lowalekar et al., 2017; Pal and Zhang, 2017; Papazek et al., 2013). These models generally aim to finding optimal truck routes to rebalance a set of stations, typically seeking to maximize customer satisfaction. The latter may be achieved by minimizing the total lost demand (see, e.g. Alvarez-Valdes et al., 2016; Contardo et al., 2012; Lowalekar et al., 2017), keeping station inventories close to their respective target inventory values (see, e.g. Brinkmann et al., 2016; Papazek et al., 2013), or even by optimizing several (possibly conflicting) objectives (see, e.g. Nunes et al., 2022). Unfortunately, the use of such models in practice is rather challenging, given that the resulting optimization models tend to be hard to solve. This typically limits their use to a small number of stations, given that intraday planning typically requires decisions within a matter of minutes. Schuijbroek et al. (2017) observe that their formulation becomes difficult to solve even for small instances with 50 stations and 3 trucks. The authors, therefore, propose a heuristic that clusters stations using a maximum spanning star and then rebalances among clusters. Likewise, Ghosh et al. (2017) cluster nearby stations and then rebalance among the clusters, where the capacity and the inventory of each cluster are given by the sum across its stations. While such an approach improves computational feasibility, it is based on the assumption that nearby stations have similar patterns and that rebalancing within each cluster is time feasible. Several other heuristic methods have been proposed (see, e.g. Lu et al., 2020; Papazek et al., 2014; Ren et al., 2020). In particular, Vergeylen et al. (2020) propose a large neighborhood search algorithm that optimizes the truck routes only for stations that have raised an alert to the system. An interesting characteristic of their optimization model is that such alerts have different priorities which are proportional to their importance in the objective function. The list of stations to rebalance (i.e., those that raised an alert) along with their associated priorities have to be provided as input to the optimization model.

Such approaches, based on alerts raised for stations that are susceptible to become unbalanced, are also easier to fit into existing practices at several BSSs, which often plan the dispatching operations based on such alerts. Nonetheless, because of limited resources in practice, planners are often required to choose a subset of the stations to rebalance. Indeed, a prior selection of stations can be very useful to scale optimization models to large BSSs by restricting the number of stations to be actually considered.

The strategies proposed in the next sections seek to recommend the best subset of stations to be rebalanced over a prespecified period of time (e.g. one hour). Given that these prioritization strategies are easy to implement and computed within a matter of seconds, they provide an attractive alternative to computationally expensive optimization models and can easily complement systems that perform rebalancing planning based on raised station alerts.

# 3 Inventory intervals

BSS operators (such as BIXI Montreal) often use intervals of acceptable inventory levels, referred to as inventory intervals. They are composed of a lower and an upper bound, as well as a target inventory value, which refers to the ideal inventory for that station and falls within the lower and upper bounds. Typically, each station has its specific inventory interval defined for a specific time period, and its values may change depending on the hour and day. When the inventory of a station falls outside of its specified inventory interval, the station is classified as unbalanced, which in turn triggers a rebalancing *alert*.

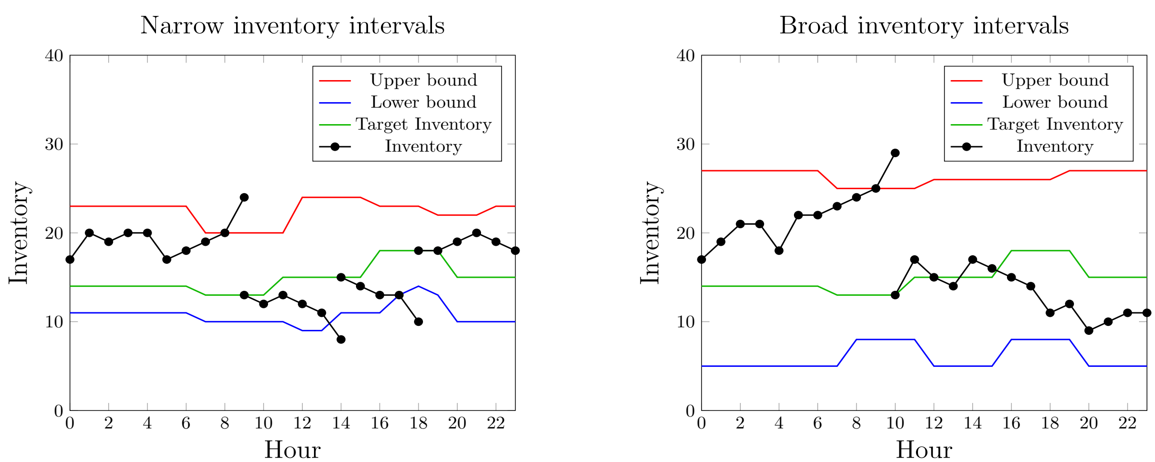



**Figure 3: Trade-off between the number of alerts and lost demand when defining inventory intervals.**

Figure 3 illustrates the evolution of the inventory of two stations with different inventory intervals and target values. Once the inventory is outside the interval, the station is assumed to be rebalanced such that, at the next hour, the inventory is set back to the target value. The width and the level of the inventory intervals have a major impact on the quantity of lost demand, on the number of raised alerts, and ultimately, on the total number of rebalancing operations. For example, narrow intervals generate more alerts, but tend to keep the inventory closer to its target value, which in turn, decreases the likelihood of lost demand (see Figure 3, left). In contrast, wide inventory intervals create fewer alerts, but at the expense of keeping the inventory farther from its target values, which may increase the amount of lost demand (see Figure 3, right). Defining inventory intervals that make this delicate trade-off in order to maximize system performance over the entire day is therefore a sensitive challenge.

The model proposed here computes the inventory intervals and the target values based on the *service level* – which is defined as the proportion of satisfied trips at a given station (Schuijbroek et al., 2017). In the literature, the metric used to estimate user satisfaction can be computed considering either rentals or returns individually (see, e.g. Schuijbroek et al., 2017; Hulot et al., 2018; Huang et al., 2020), or, as in our case, in a combined manner in which both rentals and returns are considered (see, e.g. Raviv and Kolka, 2013; Datner et al., 2019).

The inventory of the stations is typically modelled as an M/M/1/K queue, whose parameters are the time between the rentals, the time between the returns, the number of servers (here, a single station), and the maximum capacity of the server (i.e., the total number of docks $C_s$). As often assumed in the literature, we describe the trips by a Poisson distribution (Schuijbroek et al., 2017; Hulot et al., 2018; Ghosh et al., 2017; Shu et al., 2013; Kabra et al., 2020; Raviv et al., 2013; George and Xia, 2011). Hence, the times between rentals and returns follow a exponential distribution. To estimate the rentals and returns rate, we used the model presented by Hulot et al. (2018) which is able to predict the hourly trips at a station of the system by training a gradient-boosted tree (GBT) model and using a Singular Value Decomposition (SVD) method to reduce the dimension size and exclude outliers.

Given a station $s$ with initial inventory $f$, a time period $[0, T]$, and the parameters mentioned above, we propose to compute the joint service level of $s$ as:

$$SL_s(f, T) = \frac{\int_0^T \mu_s(t)(1 - p_s(f, 0, t)) + \lambda_s(t)(1 - p_s(f, C_s, t))dt}{\int_0^T \mu_s(t) + \lambda_s(t)dt}, \tag{1}$$

where $p_s(f, N, t)$ is the probability that the station $s$ has $N$ bikes at time $t$, assuming that it had $f$ bikes at time 0, and $\mu_s(t)$ (resp. $\lambda_s(t)$) is the expected demand at time $t$ at station $s$ for renting (resp. returning) bikes. The numerator of Equation (1) represents the expected number of satisfied trips at station $s$ for the referred period and initial inventory value. After normalization, $SL_s(f, T) \in [0, 1]$.

Thus, we can compute the minimum and maximum service levels for a station $s$ in a time period $[0, T]$ depending on the initial inventory at time 0 as follows:

$$SL_s^{min}(T) = \min_{f \in \{0, \dots, C_s\}} SL_s(f, T), \qquad \text{and} \qquad SL_s^{max}(T) = \max_{f \in \{0, \dots, C_s\}} SL_s(f, T). \tag{2}$$

Next, we establish a threshold $\Omega_s$ for the acceptable service level for a station $s$ during time period $[0, T]$, defined as:

$$\Omega_s(T) = SL_s^{min}(T) + \beta(SL_s^{max}(T) - SL_s^{min}(T)), \tag{3}$$

where the hyperparameter $\beta \in [0, 1]$ controls how exigent the operator is about the network service. A small value of $\beta$ approximates the threshold to the minimum service level, while a large $\beta$ brings the threshold closer to the maximum service level. Here, $\Omega_s(T) \in [0, 1]$, given that each individual service level $SL_s \in [0, 1]$.

The inventory interval for station $s$ for time period $[0,T]$ compatible with threshold $\Omega_s(T)$ is then defined as

$$\mathcal{I}_s(T) = \{f \in \{0,...,C_s\} | \mathcal{L}_s(T) \leq f \leq \mathcal{U}_s(T)\}, \tag{4}$$

where

$$\mathcal{U}_s(T) = \max\{f \in \{0,...,C_s\} | SL_s(f,T) \geq \Omega_s(T)\}, \text{ and} \tag{5}$$

$$\mathcal{L}_s(T) = \min\{f \in \{0,...,C_s\} | SL_s(f,T) \geq \Omega_s(T)\}. \tag{6}$$

Finally, the target inventory value $\mathcal{T}_s(T)$ for station $s$ for time period $[0,T]$ is given by the inventory value that provides the highest service level:

$$\mathcal{T}_s(T) = \underset{f \in \{0,...,C_s\}}{\arg\max} \{SL_s(f,T)\}. \tag{7}$$

Based on the above defined inventory intervals and the target inventory value, we next propose new strategies to effectively prioritize unbalanced stations that should be considered for rebalancing.

# 4 Prioritizing strategies for unbalanced stations

Operators typically follow a systematic approach to prioritize stations for rebalancing among those that have raised an alert. In particular, BIXI Montreal segments the unbalanced stations into three priority groups, each with specific criteria[1]. The first group, denoted $K_1$, contains stations classified as critical, i.e., stations that are either completely empty or completely full, and for which all their neighboring stations within a radius of 600 meters are also completely empty (or completely full)[2]. The second group, denoted $K_2$, includes unbalanced stations that are not part of $K_1$, but are located within a 600-meter radius of a metro station. This criterion recognizes the fact that stations located near metro stations tend to have a higher demand, since it is common for both modes of transportation to complement each other in the urban environment. The third group, $K_3$, comprises stations that are not part of the previous groups but are within 600 meters of any station in $K_1$ or $K_2$. Any unbalanced station not classified under $K_1, K_2$ or $K_3$ is not considered for rebalancing in the analyzed time period. The priority of stations in $K_1$ is always higher than that of stations in $K_2$, and the priority of stations in $K_2$ is higher than that of stations in $K_3$. Finally, within each group, stations are sorted based on their proximity to the nearest metro station. This means that stations closer to the metro have a greater likelihood of being selected for rebalancing.

In the following, we present our proposed prioritization strategies, that provide priority scores to the unbalanced stations according to different criteria.

## 4.1 Prioritization strategy based on inventory forecasting

The first prioritization strategy, denoted $Pa_1$, selects stations for rebalancing taking into consideration the current stations inventories, provided as input, and the expected demand, which is obtained by the gradient-boosted tree (GBT) regression model of Hulot et al. (2018). We predict the inventory of each station $s$ in the network at the next hour $t+1$ as follows:

$$\overline{f}_s^1(t+1) = f_s(t) + \mu_s(t) - \lambda_s(t), \tag{8}$$

[1]This procedure was conceived after several exchanges with BIXI's planners. As such, it is not an official representation of BIXI's decision-making process.

[2]The radius of 600 meters is defined by BIXI based on the fact that an average person may walk this distance within 10-15 minutes, which is the time that BIXI considers acceptable for a commuter to walk seeking to be served.

where $\mu_s(t)$ and $\lambda_s(t)$ are, respectively, the expected demand for the number of rentals and returns at station $s$ during time period $[t, t+1]$, and $f_s(t)$ refers to the actual number of bikes available at station $s$ at the beginning of hour $t$. Note that (8) can also be extended to predict inventories for subsequent hours in advance. In this case, it suffices to predict bike demands for the hours of the considered time period.

Then, for each station $s$, the strategy computes two indices

$$\mathcal{B}_s^{1-}(t+1) = \max\{0, -\overline{f}_s(t+1)\} \qquad \text{and} \qquad \mathcal{B}_s^{1+}(t+1) = \max\{0, \overline{f}_s(t+1) - C_s\}, \tag{9}$$

which represent, respectively, the predicted inventory shortfall and surplus at hour $t+1$ at station $s$. By definition, only one of the two indices $\mathcal{B}_s^-(t)$ and $\mathcal{B}_s^+(t)$ can assume a value greater than 0 at the same time $t$. Finally, a prioritization score is computed for station $s$ at hour $t$ proportional to its predicted inventory shortfall or surplus, as follows:

$$\mathcal{P}_s^1(t) = \max\{\mathcal{B}_s^{1-}(t+1), \mathcal{B}_s^{1+}(t+1)\}. \tag{10}$$

As a result, strategy $Pa_1$ ranks and prioritizes stations in non-increasing order of their predicted inventory shortfall or surplus for the next hour(s).

## 4.2 Prioritization strategy based on inventory forecasting with immediate rebalancing

The second proposed strategy, denoted $Pa_2$, sorts unbalanced stations according to the amount of lost demand avoided by rebalancing operations. It assumes that a rebalancing operation at the beginning of hour $t$ sets the inventory of a station to its target inventory value. Thus, the predicted inventory for each station $s$ at the hour $t+1$ is computed as:

$$\overline{f}_s^2(t+1) = \mathcal{T}_s(t) + \mu_s(t) - \lambda_s(t), \tag{11}$$

The predicted shortfall and surplus of bikes at station $s$ for the next hour are then computed as:

$$\mathcal{B}_s^{2-}(t+1) = \max\{0, -\overline{f}_s^2(t+1)\} \qquad \text{and} \qquad \mathcal{B}_s^{2+}(t+1) = \max\{0, \overline{f}_s^2(t+1) - C_s\}. \tag{12}$$

Finally, the prioritization score of station $s$ at hour $t$ is given by strategy $Pa_2$ as:

$$\mathcal{P}_s^2(t) = \mathcal{P}_s^1(t) - \max\{\mathcal{B}_s^{2-}(t+1), \mathcal{B}_s^{2+}(t+1)\}. \tag{13}$$

The first term of (13), i.e., $\mathcal{P}_s^1(t)$ represents the lost demand if station $s$ is not rebalanced at hour $t$, whereas the second part, i.e., $\max\{\mathcal{B}_s^{2-}(t+1), \mathcal{B}_s^{2+}(t+1)\}$ refers to the lost demand if that station has its inventory set to its target value $\mathcal{T}_s(t)$ at the beginning of hour $t$. Note that, theoretically, $\mathcal{P}_s^2(t)$ may be negative, i.e., rebalancing a station may increase the lost demand. Throughout our computational experiments with real-world data, however, we have never observed negative values for $\mathcal{P}_s^2(t)$. Finally, the strategy outputs the stations in non-increasing order of their prioritization scores.

## 4.3 Prioritization strategy using inventory intervals

The third prioritization strategy we propose, denoted $Pa_3$, is more conservative than the strategies above. Instead of giving a high prioritization score to stations whose inventories are predicted below zero or above their dock capacity (i.e., yielding lost demand), $Pa_3$ prioritizes unbalanced stations according to the predicted deviation of the station's inventory from its inventory interval bounds (eqs. (5–6)). Thus, stations might be rebalanced before they actually start to generate lost demand (i.e., before they are completely full or empty). Strategy $Pa_3$ computes two indices:

$$\mathcal{B}_s^{3-}(t+1) = \max\{0, \mathcal{L}_s(t) - \overline{f}_s^1(t+1)\} \quad \text{and} \quad \mathcal{B}_s^{3+}(t+1) = \max\{0, \overline{f}_s^1(t+1) - \mathcal{U}_s(t)\}, \tag{14}$$

where $\mathcal{L}_s(t)$ and $\mathcal{U}_s(t)$ correspond, respectively, to the lower and upper bounds of the inventory interval of station $s$ at hour $t$. We also recall that $\overline{f}_s^1(t+1)$ corresponds to the predicted station inventory at hour $t+1$ if no rebalancing is carried out at station $s$ (see Section 4.1). The priority score for strategy $Pa_3$ is then computed as:

$$\mathcal{P}_s^3(t) = \max\{\mathcal{B}_s^{3-}(t+1), \mathcal{B}_s^{3+}(t+1)\}. \tag{15}$$

As a result, $Pa_3$ prioritizes stations that are expected to remain unbalanced in the next hour if the operator does not rebalance them. These stations are then sorted and prioritized in non-increasing order of the deviation between their predicted inventories and their inventory interval bounds.

## 4.4 Prioritization strategy based on neighbourhood

The fourth prioritization strategy proposed here, denoted $Pa_4$, diverts from the previous ones in the sense that it additionally takes into consideration geographical information to rank unbalanced stations. By favoring the rebalancing of neighbouring stations, $Pa_4$ is expected to obtain more compact and thus less costly (and faster) dispatching routes that group together nearby stations.

Let us define $\mathcal{N}_s$ as the index set of all stations within a radius of 600 meters from station $s$. The priority score computed by $Pa_4$ is given by:

$$\mathcal{P}_s^4(t) = \begin{cases} \gamma \times \mathcal{P}_s^*(t) + (1-\gamma) \times \frac{\sum_{s' \in \mathcal{N}_s, s' \neq s} \mathcal{P}_{s'}^*(t)}{|\mathcal{N}_s|}, & \text{if } \mathcal{P}_s^*(t) > 0 \\ \mathcal{P}_s^*(t), & \text{otherwise.} \end{cases} \tag{16}$$

Here, the priority score $\mathcal{P}_s^*(t)$ refers to any of the three priority scores $\mathcal{P}_s^1(t), \mathcal{P}_s^2(t)$ or $\mathcal{P}_s^3(t)$ presented above. The hyperparameter $\gamma \in [0,1]$ controls the weight of the neighbourhood inventory information on the computation of the priority score. By using $\gamma < 1$, the priority score computed by $Pa_4$ for a station $s$ also takes into consideration the priority scores of its neighboring stations. As $\gamma$ approaches 0, those scores tend to prevail in (16), prioritizing stations surrounded by others that raised an alert. As before, the stations are then sorted in non-increasing order of their priority scores.

For all of the proposed prioritization strategies, stations with a priority score of 0 or lower are simply discarded. Among stations with the same score, priority is given to those closer to a metro station.

## 4.5 Post-processing to ensure a balanced selection of pick-ups and drop-offs

Our prioritization strategies are responsible for sorting the stations according to their rebalancing priority, but they do not take into consideration practical constraints such as the hourly rebalancing capacity of the operator or the parity between the amount of bikes added and removed from the BSS. Indeed, the latter is an important criterion for the rebalancing team in order to ensure that the number of bikes picked up is approximately the same as the number of bikes dropped off.

In view of that, we developed a post-processing procedure that follows the application of any of the previous prioritization strategies.This procedure has two steps and it is illustrated in Figure 4. In the first step, the procedure receives the list of sorted unbalanced stations and splits it into two sorted sub-lists of stations, namely: (i) a list $\mathcal{S}^{add}$ of stations at which bikes are added (i.e., dropped off), and (ii) a list $\mathcal{S}^{rem}$ of stations from which bikes should be removed (i.e., picked up). The amount of bikes added or removed from the stations is exemplified in the figure and corresponds to the number of required bikes to restore the inventory of the stations to their respective target values. In the second step, the stations are selected for rebalancing according to the value of the *accumulator* variable, which is initialized with 0. If *accumulator* $\geq 0$, the procedure selects the highest priority station from $\mathcal{S}^{rem}$ to be rebalanced, and updates the accumulator accordingly. Otherwise, the procedure selects for rebalancing the highest priority station from $\mathcal{S}^{add}$, and the accumulator variable is increased. Each time a station is selected in step 2, it is removed from its corresponding sub-list. The post-processing procedure is halted when either $\mathcal{S}^{add}$ or $\mathcal{S}^{rem}$ are empty, or when the number of rebalancing operations equals the BSS rebalancing capacity (i.e., the maximum number of rebalancing operations the operator is likely to be able to carry out). At the end, the procedure outputs the set of stations to be rebalanced in the simulation.

# 5 Computational experiments

We now report on computational experiments that have been carried out to evaluate the performance of the proposed prioritization strategies. We first present the details of the dataset used in our study. In the sequel, we present the simulation used to estimate the inventory response to each of the applied rebalancing strategies. Specifically, it computes three performance metrics: lost demand, the number of raised alerts, and the number of performed rebalancing operations. These metrics are then compared for the different prioritization strategies. Finally, in a specific subsection, we assess $Pa_4$ with respect to the clustering property of its prioritized stations.

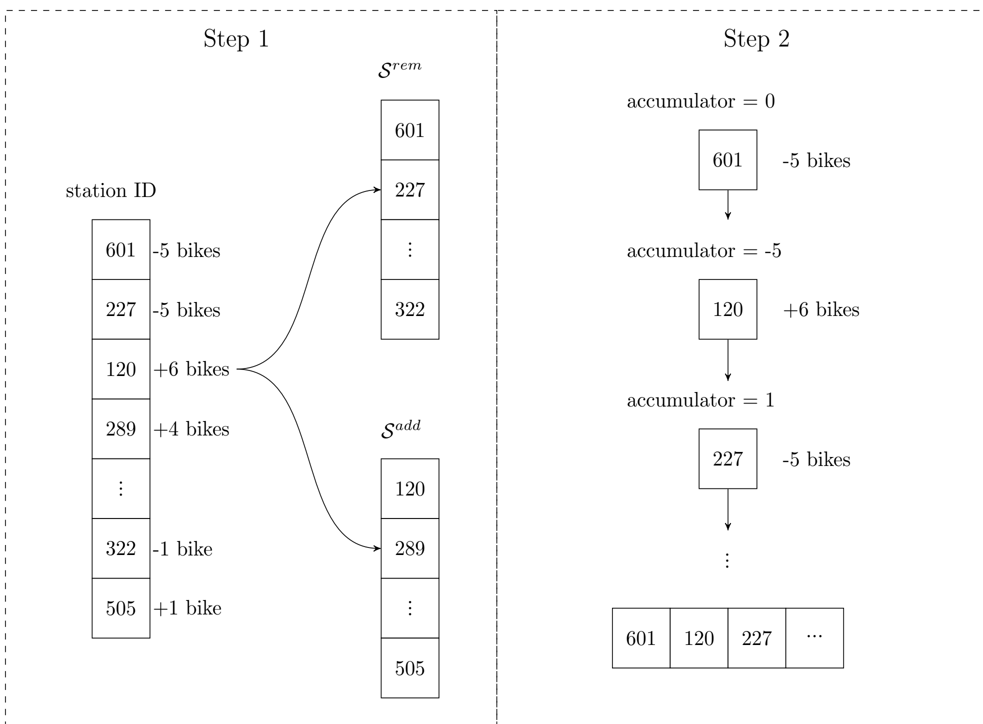



**Figure 4: Illustration of the post-processing procedure is divided into two steps. Note that the original sorting of the stations provided by the prioritization strategy is preserved in the sub-lists.**

## 5.1 Data set

We consider real-world data from BIXI Montreal for the 2019 and 2020 seasons. The dataset used in the experiments contains hourly information for time, weather, trips, and stations. Time features hold temporal data such as the hour, day, whether it is a holiday, and the day of the week. Weather features store information such as temperature, wind speed and relative humidity. Trip features are composed of the number of bike rentals and returns, observed at each station of the network. Finally, station features contain information regarding the geographical location of each station, as well as their corresponding number of docks. Time and weather features were collected from `https://climate.weather.gc.ca` (except for the holiday feature that was manually imputed), while trip and station features were both provided by BIXI Montreal. More details about the importance of the different features for the GBT used here can be found in Hulot (2018).

The here considered BSS network from BIXI Montreal contains a total of 620 stations in 2019 and 641 stations in 2020. Each dataset was split into training, validation, and test data, and they each contain all stations from the corresponding year. The training and validation data are used, respectively, to fit the machine learning model parameters and tune its hyperparameters. Finally, the test dataset was used to provide an unbiased performance evaluation of the different prioritization strategies.

Because of a large observed discrepancy in the frequency and the behavior of trips in 2020, caused by the COVID-19 pandemic, we used different strategies for selecting the dataset for each simulated season. For the 2019 tests, the training dataset contains data from April 2018 to June 2019, minus the months during which BIXI is out of service (i.e., December, January, February, and March). The validation dataset is composed of data from the first 15 days of July and August 2019, and the test dataset uses data from the remaining days of July and August 2019. We opted to divide both July and August into validation and test datasets so that the model is less sensitive to demand changes observed between consecutive months.

Note that the physical network of BSSs typically changes over time, which makes it often difficult to use data linked to a specific station ID over longer periods of time. Therefore, we do not use data prior to the 2018 season. We have also observed that, for the 2020 season, adding training data from previous years deteriorates traffic prediction. Consequently, we used training data from April 2020 to June 2020. In that case, the validation dataset contains the first 15 days of July and August 2020, while the test dataset contains the remaining days of July and August 2020. Note that we focus our experiments on the months of July and August because of their high demand and importance throughout the season.

## 5.2 Simulation

Our simulation aims to emulate inventory fluctuations in the BSS according to the trip demand and the performed rebalancing operations. Thus, we can evaluate the proposed prioritization rebalancing strategies regarding the measured performance metrics.

The simulation starts by initializing the inventory of all stations with their respective target values. By doing so, we guarantee that the experiment starts with an optimal number of bikes in the system. Then, at every simulated hour, the inventory of all stations is modified according to the historical rentals and returns. We assume that all trips at each hour take place simultaneously to stress the system, thus capturing its possible failures. In the sequel, the simulation launches the rebalancing process by gathering the stations that raised alerts. The set of unbalanced stations is then forwarded for prioritization and, after being sorted, to the post-processing procedure, which ensures that the total number of bike pick-ups and bike drop-offs is sufficiently close to each other. The latter then returns a subset of the stations to be rebalanced. The number of stations within this subset is limited by the total rebalancing capacity, which is given as input to the simulator and represents the total number of stations the operator is capable of rebalancing at each hour. After the prioritized stations have been rebalanced by setting their inventories to their target values, the simulation proceeds to the next simulated hour until all hours are iterated.

After execution, the simulator returns three metrics that are important to the BSS operator: (i) the number of raised alerts, indicating how often a station has been classified as unbalanced, thus representing how stressed the system was; (ii) the number of rebalancing operations, allowing for an estimation of the operational rebalancing costs; and (iii) the amount of rental and return requests that could not be satisfied, directly affecting the customer satisfaction.

The pseudo-code of our simulator, as well as its execution pipeline and description can be found at `github.com/clara91/Data-driven-prioritization-strategies-for-BSSs`.

## 5.3 Simulation results

We now report on the simulation results of BIXI's data for the 2019 and 2020 seasons. We here consider that the maximum rebalancing capacity per hour was 46 for the 2019 season and 22 for the 2020 season. These numbers correspond to the average number of rebalancing operations performed by BIXI during the peak rebalancing hours in the two observed years.

All prioritization strategies have been fed with the same inventory intervals in order to ensure a fair comparison of their performance. Even though the inventory intervals used by BIXI were available to us, incorporating them into the analysis may have compromised the validity of our conclusions.

As shown in Section 3, the value of parameter $\beta$ plays an important role in the definition of the inventory intervals and directly impacts the performance measures. Large values of $\beta$ result in narrow inventory intervals, and consequently, in a larger number of alerts, while potentially improving demand satisfaction. In contrast, small $\beta$ values lead to wide intervals, decreasing the number of raised alerts at the stations, but possibly decreasing demand satisfaction. We, therefore, evaluated the prioritization strategies using three values for $\beta$, specifically $\beta = 0.75$ (narrow intervals), 0.50 (medium intervals), and 0.25 (wide intervals).

### 5.3.1 Evaluation of $Pa_1$, $Pa_2$ and $Pa_3$

We first focus on the comparison of the first three prioritization strategies $Pa_1$, $Pa_2$ and $Pa_3$, as well as the emulated strategy employed by BIXI Montreal. Table 1 provides an overview of the total lost demand (expressed as a percentage of the total demand, i.e., rental plus return demand) calculated across the entire network for the 768 simulated hours from the test dataset. The table also includes the hourly average of raised alerts and the hourly average number of rebalancing operations performed in the system over the same 768 simulated hours. The presented results were collected from the simulation using the proposed prioritization strategies $Pa_1, Pa_2$, and $Pa_3$, as well as the emulation of BIXI's prioritization strategy. Finally, for each of the presented metrics, we report the relative difference (column $\Delta(\%)$) calculated with respect to the values obtained using BIXI's prioritization strategy.

Table 1 suggests the following conclusions:

1. **General impact on lost demand.** The three proposed strategies consistently resulted in lower levels of lost demand compared to BIXI's prioritization strategy across all simulated scenarios. This suggests that the proposed strategies excel at prioritizing stations for rebalancing.
2. **Best strategy to reduce lost demand.** Among the proposed prioritization strategies, $Pa_3$ is the most effective strategy to reduce lost demand and the number of raised alerts, achieving reductions of ≈ 35% in lost demand for 2019, and ≈ 65% in 2020. However, it's important to consider the tradeoff involved when using $Pa_3$ as it requires a higher number of rebalancing operations compared to BIXI's prioritization strategy. Specifically, even though the performed rebalancing respects the maximum rebalancing capacity, more rebalancing operations also translate into higher operational costs.
3. **Best strategies with moderate rebalancing.** $Pa_1$ and $Pa_2$ reduce the lost demand without necessarily increasing the number of rebalancing operations. In fact, they often led to fewer rebalancing operations in the system – for all $\beta$ values in 2019, and 2 out of 3 in 2020. This is explained by the fact that both $Pa_1$ and $Pa_2$ assign a positive score to unbalanced stations only

**Table 1: Performance metrics for the 2019 and 2020 seasons.**

| Season | Prioritization Strategy | $\beta$ | Lost Demand total % | Lost Demand $\Delta$(%) | Alerts per hour | Alerts $\Delta$(%) | Rebalancing per hour | Rebalancing $\Delta$(%) |
|---|---|---|---|---|---|---|---|---|
| 2019 | BIXI | 0.25 | 5.22 | | 69.86 | | 23.86 | |
| | | 0.5 | 4.96 | | 95.64 | | 29.48 | |
| | | 0.75 | 5.13 | | 142.41 | | 35.87 | |
| | $Pa_1$ | 0.25 | 4.19 | ▼ -19.72 | 69.34 | ▼ -0.74 | 23.35 | ▼ -2.14 |
| | | 0.5 | 4.05 | ▼ -18.23 | 109.83 | ▲ 14.84 | 23.83 | ▼ -19.17 |
| | | 0.75 | 4.00 | ▼ -22.03 | 170.38 | ▲ 19.64 | 24.02 | ▼ -33.04 |
| | $Pa_2$ | 0.25 | 4.16 | ▼ -20.24 | 69.03 | ▼ -1.19 | 23.20 | ▼ -2.77 |
| | | 0.5 | 4.05 | ▼ -18.35 | 110.01 | ▲ 15.03 | 23.63 | ▼ -19.84 |
| | | 0.75 | 3.95 | ▼ -22.95 | 169.87 | ▲ 19.28 | 23.89 | ▼ -33.40 |
| | $Pa_3$ | 0.25 | 3.99 | ▼ -23.46 | 53.54 | ▼ -23.36 | 27.19 | ▲ 13.96 |
| | | 0.5 | 3.58 | ▼ -27.71 | 74.80 | ▼ -21.79 | 31.51 | ▲ 6.89 |
| | | 0.75 | 3.33 | ▼ -35.13 | 116.63 | ▼ -18.10 | 36.03 | ▲ 0.45 |
| 2020 | BIXI | 0.25 | 2.47 | | 43.60 | | 9.55 | |
| | | 0.5 | 2.13 | | 57.22 | | 11.42 | |
| | | 0.75 | 2.10 | | 88.21 | | 14.59 | |
| | $Pa_1$ | 0.25 | 1.57 | ▼ -36.44 | 31.74 | ▼ -27.20 | 10.93 | ▲ 14.45 |
| | | 0.5 | 1.53 | ▼ -28.17 | 51.98 | ▼ -9.16 | 10.97 | ▼ -3.94 |
| | | 0.75 | 1.53 | ▼ -27.14 | 92.29 | ▲ 4.63 | 10.99 | ▼ -24.67 |
| | $Pa_2$ | 0.25 | 1.55 | ▼ -37.25 | 32.02 | ▼ -26.56 | 10.92 | ▲ 14.35 |
| | | 0.5 | 1.53 | ▼ -28.17 | 51.98 | ▼ -9.16 | 10.97 | ▼ -3.94 |
| | | 0.75 | 1.52 | ▼ -27.62 | 92.65 | ▲ 5.03 | 11.06 | ▼ -24.19 |
| | $Pa_3$ | 0.25 | 1.15 | ▼ -53.44 | 20.64 | ▼ -52.66 | 12.53 | ▲ 31.20 |
| | | 0.5 | 0.88 | ▼ -58.69 | 28.26 | ▼ -50.61 | 14.07 | ▲ 23.20 |
| | | 0.75 | 0.72 | ▼ -65.71 | 48.87 | ▼ -44.60 | 16.47 | ▲ 12.89 |

if they are predicted to generate lost demand in the next hour, while $Pa_3$ bases its prioritization score on the violation of their inventory intervals. As a result, $Pa_3$ prioritizes a larger number of stations with scores greater than 0, leading to an increased frequency of rebalancing operations compared to $Pa_1$ or $Pa_2$.

4. **Impact of inventory interval size.** For our prioritization strategies, the relative reduction of the lost demand was more substantial when narrow inventory intervals ($\beta = 0.75$) were employed, generally leading to a higher number of alerted stations and slightly more rebalancing operations. In contrast, this behavior has not been observed during the 2019 season for BIXI's prioritization strategy. Here, narrow inventory intervals ($\beta = 0.75$) resulted in more alerts and rebalancing operations, but this did not translate into a lower lost demand. We analyze this curious behaviour below in Section 5.3.2.
5. **Pattern shift from 2019 to 2020.** As a result of the lockdown measures applied by the Canadian authorities in response to the COVID-19 pandemic, user behavior drastically changed. Most likely attributable due to the increased work-from-home, the demand peaks shifted and the total number of trips reduced by 85% (e.g., from about 4000 trips to about 600 trips at the respective peak hours). As a result, the relative lost demand also dropped by at least 50% from 2019 to 2020, as stations were substantially less stressed and users were less often faced with empty or full stations.
6. **Benefits of data-driven methods.** Even though the relative lost demand reduced from 2019 to 2020, this was expected (see previous point) and it is questionable whether the BIXI's rebalancing strategy has adapted to the new demand pattern in an ideal manner. Our prioritization strategies demonstrated a substantially higher reduction of the relative lost demand over BIXI's strategy in 2020. Such an improvement, despite the fact the total demand was much lower, is remarkable. It is, in fact, much more difficult to further reduce lost demand when the total number of trips is low. For example, 5% of lost demand at a peak-hour in 2020 may refer to about 15 out of

600 trips, while it refers to 200 out of 4000 trips in 2019. Reducing an abundant lost demand is easier than a sparse one, since the latter requires a very precise identification of the stations at which lost demand can be further reduced. It is therefore even more impressive to see the lost demand in 2019 reducing to about 0.72% for strategy $Pa_3$. This does not only highlight that our prioritization strategies are effective, it illustrates their ability to adapt to new demand patterns and, as such, highlights the importance of data-driven strategies in general that can adapt to changes in demand patterns much faster than manual adjustments.

### 5.3.2 Trade-off between the number of alerts and the rebalancing capacity

BIXI's prioritization strategy presenting higher lost demand with tighter inventory intervals ($\beta = 0.75$) is unexpected, given that our prioritization strategies consistently reduced lost demand as the inventory intervals got tighter. It turns out that the trade-off between the number of alerts and the rebalancing capacity is an important one.

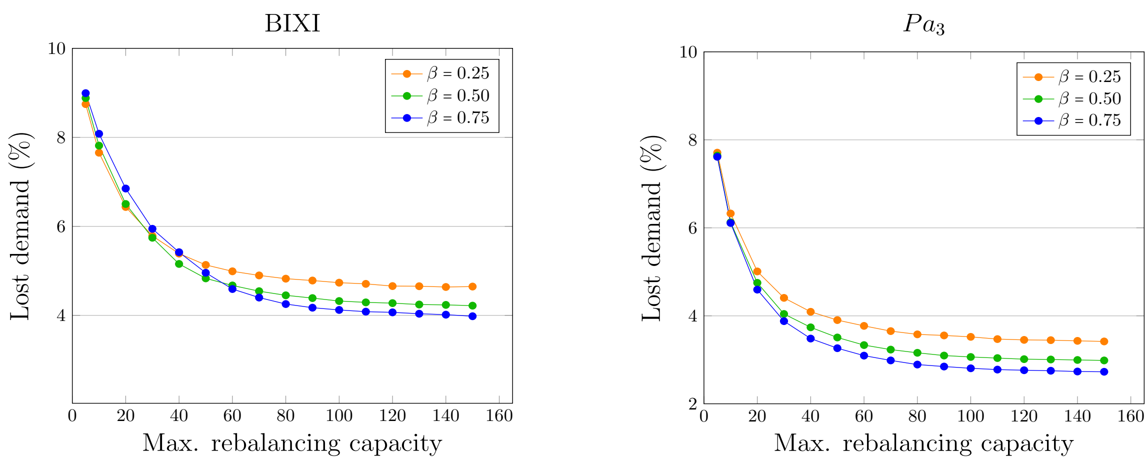



**Figure 5: Percentage of lost demand as a function of the rebalancing capacity of the system on the 2019 test dataset.**

Figure 5 presents the lost demand as a function of the BSS rebalancing capacity for both BIXI's prioritization strategy (left) and $Pa_3$ (right) throughout the 2019 test dataset. For strategy $Pa_3$, more narrow intervals (i.e., higher $\beta$ values) consistently lead to lower lost demand. In contrast, for BIXI's prioritization strategy, narrow intervals only perform well when the rebalancing capacity is sufficiently high (more than 50) to deal with the large amount of raised alerts. When stations are stressed (i.e., there is a high number of raised alerts), BIXI's prioritization strategy has difficulties identifying the most critical stations, which leads to higher lost demand when the rebalancing capacity is low (here, less than 40). Here, wider inventory intervals raise alerts only for the most unbalanced stations, which is more aligned with the limited rebalancing capacity. Indeed, the rebalancing capacity used in our case-study in Table 1 assumes a rebalancing capacity of 46, which lies at the threshold where more narrow intervals are beneficial when using BIXI's prioritization strategy.

This analysis allows for deriving two key insights for BSS operators, particularly those with a limited rebalancing capacity (which is an economical key concern for operators). First, it highlights the importance of an effective prioritization strategy, particularly when the number of alerts is high. Prioritizing the wrong stations is as ineffective as raising the wrong alerts. Second, being capable of adjusting the number of alerts (by smartly selecting the intervals and choosing an appropriate interval size) is crucial, since the operator may want to keep the number of alerts aligned with the rebalancing capacity, particularly when the subsequent prioritization process is not robust to changes in the demand pattern. The prioritization strategies here presented improve over BIXI's prioritization strategies in both concerns, which makes them an attractive tool to integrate while working towards more effective and automated rebalancing strategies.

### 5.3.3 Evaluation of $Pa_4$

Recall that all proposed prioritization strategies (i.e., $Pa_1$–$Pa_4$) respect the total rebalancing capacity prescribed by the operator (i.e., the maximum number of stations that can be selected for rebalancing). Further, we ensure in the post-processing procedure (see Section 4.5) that the number of bikes to be picked up and dropped off at the selected stations is approximately the same. While this may facilitate the planning of feasible rebalancing routes of the vehicles, stations may still be quite far from each other. Selecting a subset of stations that is naturally more clustered may therefore facilitate the generation of vehicle routes that rebalance the selected stations efficiently.

To this end, in this section, we turn our analysis to the prioritization strategy $Pa_4$, which is designed to favour the rebalancing of proximal unbalanced stations as $\gamma$ tends to zero. Thus, it is expected that less lengthy, and consequently, less expensive routes can be derived for the trucks that are in charge of the system rebalancing. We report in this section results of strategy $Pa_4$ with $\mathcal{P}_s^3(t)$ in eq. (16). The results of $Pa_4$ obtained with the other prioritization algorithms are similar to the ones reported here in this section.

We first introduce an additional metric to measure the compactness of a set of stations prioritized for rebalancing. Let $G = (N, E)$ be a graph for which there is a node $n_i \in N$ corresponding to each station selected for rebalancing at a given hour. Set $E$ contains all edges $e_{ij}$ such that $n_i \in N$ and $n_j \in N$, and the distance between the stations represented by $n_i$ and $n_j$ is not more than 600m. Let us also denote $S_i \subseteq N$ as the set of nodes connected to $n_i \in N$. We then compute the *Watts–Strogatz clustering coefficient* (Watts and Strogatz, 1998) of $G$ as:

$$\eta_G = \frac{1}{|N|} \sum_{n_i \in N} \eta_i, \quad \text{where } \eta_i \text{ is computed as} \quad \eta_i = \frac{|\{e_{jk} \in E : n_j \in S_i, n_k \in S_i\}|}{\binom{|S_i|}{2}}. \tag{17}$$

The Watts–Strogatz clustering coefficient measures the inherent tendency of a graph to form clusters. In fact, the value $\eta_i$ of a node $n_i \in N$ measures how close its neighbours are to forming a clique.

Figure 6 illustrates four graphs constructed from the stations prioritized by $Pa_4$ in our simulation run for July 19, 2019 at 11 am, using $\beta = 0.75$ and $\gamma = 0.25, 0.5, 0.75$ and 1. Nodes in $N$ are drawn as red dots, whereas edges in $E$ are indicated as yellow lines. We note from the figure that small $\gamma$ values yield graphs for which the selected stations can be more naturally clustered, yielding higher Watts–Strogatz clustering coefficient values.

Next, Figure 7 illustrates the impact of the hyperparameter $\gamma$ on the Watts-Strogatz clustering coefficient, as well as on the three performance measures used in our study. Moreover, we include in Figure 7 the best metric values obtained by the BIXI's prioritization strategy among those obtained with $\beta = 0.25, 0.5, 0.75$, which can be found in Table 1.

In particular, for each value of $\gamma \in \{0, 0.05, 0.10 \ldots, 1\}$, Figure 7a presents the average Watts-Strogatz clustering coefficient of the graphs built from the stations selected by $Pa_4$ around the demand peak hours of the 2019 BIXI's season, i.e., between 7-11 am and 4-8 pm. We can notice that $\eta_G$ reaches its maximum for $\gamma = 0.15$, and not at $\gamma = 0$. Since the stations available for prioritization might vary over time due to previous rebalancing operations, being too greedy towards the prioritization of neighbouring stations might cause an absence of clusters of unbalanced stations for subsequent simulated hours. For reference, $Pa_4$ yields an $\eta_G$ superior to that obtained by means of the BIXI's prioritization strategy for $\gamma \leq 0.5$, regardless of the tested value of $\beta$. Moreover, as expected, we can observe in Figures 7b–7d that $Pa_4$ improves its performance with respect to the other three metrics as $\gamma$ approaches 1, as the influence of the station's neighborhood in the prioritization procedure is decreased. Note that $Pa_4$ is equivalent to $Pa_3$ when $\gamma = 1$.

Although the number of rebalancing operations observed with the use of $Pa_4$ is larger for smaller $\gamma$, such operations are potentially less costly since the selected stations are closer to each other. Such tradeoff between the clustering coefficient and the cost of the system's rebalancing operations indeed merits further investigation. However, it must be conducted in conjunction with routing optimization algorithms, which are out of the scope of the present paper. We leave this to future research.

# 6 Concluding remarks

In this paper, we proposed a series of strategies to select the unbalanced stations that should be prioritized for rebalancing in dock-based BSSs. This is an important issue for BSS operators, given that it is neither economically feasible nor ecologically desirable to provide a sufficiently large truck fleet in order to rebalance all unbalanced stations at the same time.

Our prioritization strategies make use of a wide range of indicators to select stations for rebalancing, including the forecast of future station inventories, the impact of potential rebalancing on station inventory, and the predicted deviation of the station inventories from their inventory intervals (the latter being automatically computed using demand prediction based on historical trips, weather, and temporal data). As inventory intervals are already currently used in practice by BSS operators, the implementation of our prioritization strategies into the practical information- and decision process is straightforward. Once the predictive model is trained, inventory intervals and prioritization strategies are computed in a matter of seconds, which is a practical requirement in the ongoing effort to rebalance stations throughout the day.

By comparing the proposed prioritization strategies, we observe that all of them are capable of reducing the occurrences of lost demand. This shows that a judicious choice of stations to be rebalanced has a considerable impact on the service level and, hence, customer satisfaction. Among the proposed strategies, $Pa_1$ and $Pa_2$ considerably reduce the number of performed rebalancing operations in most scenarios, whereas $Pa_3$ is the most effective strategy to improve the service level at the stations. Hence, the choice of the prioritization strategy to implement should depend on the objective pursued by the operator. The fourth strategy, $Pa_4$, also proved to be an effective alternative to improve service level while rebalancing proximal stations, with the potential to decrease operational routing costs.

Our strategies also seem particularly useful during periods in which the demand changes its patterns. The proposed prioritization strategies seem to adapt better to the demand changes in 2020 in comparison to BIXI's prioritization strategy. This outcome is due to the fact that the first is based on demand prediction while the former is based mostly on the location of the stations.

In future research, our strategies may be used along with routing optimization models, such that the entire rebalancing process is optimized in an integrated manner, taking into consideration the capacity of the dispatched trucks, as well as the loading/unloading of bikes among the stations.

# References

M. Wang, X. Zhou, Bike-sharing systems and congestion: Evidence from us cities, Journal of transport geography 65 (2017) 147–154.

J. Pucher, R. Buehler, D. R. Bassett, A. L. Dannenberg, Walking and cycling to health: a comparative analysis of city, state, and international data, American journal of public health 100 (2010) 1986–1992.

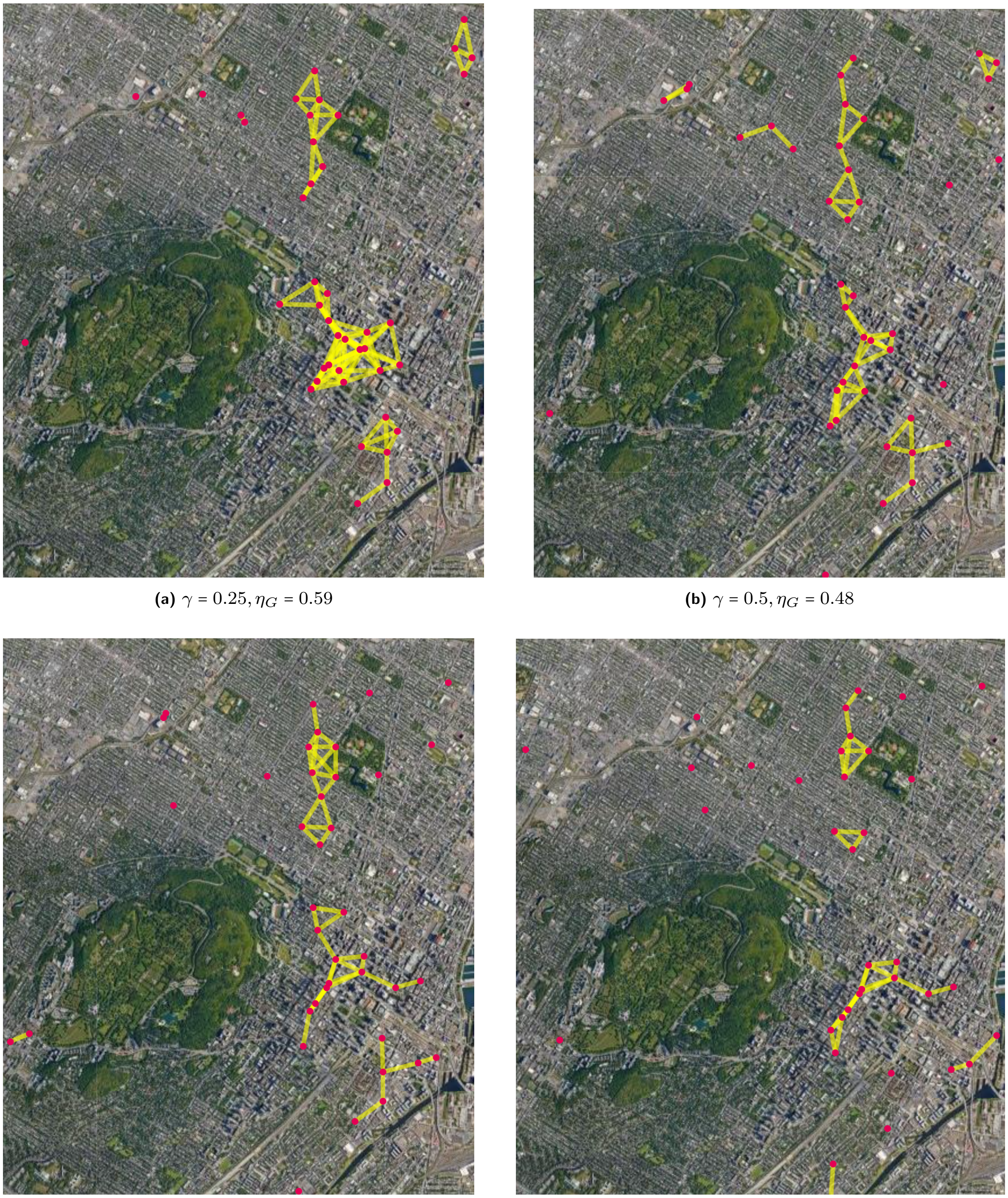
F. Pase, F. Chiariotti, A. Zanella, M. Zorzi, Bike sharing and urban mobility in a post-pandemic world, IEEE Access 8 (2020) 187291–187306.

E. Basak, R. Al Balawi, S. Fatemi, A. Tafti, When crisis hits: Bike-sharing platforms amid the covid-19 pandemic, Plos one 18 (2023) e0283603.

K. Mellou, P. Jaillet, Dynamic resource redistribution and demand estimation: An application to bike sharing systems, SSRN Electron (2019) 1–58.

C. M. Vallez, M. Castro, D. Contreras, Challenges and opportunities in dock-based bike-sharing rebalancing: a systematic review, Sustainability 13 (2021) 1829.

P. DeMaio, Bike-sharing: History, impacts, models of provision, and future, Journal of public transportation 12 (2009) 3.



**(a)** $\gamma = 0.25, \eta_G = 0.59$

**(b)** $\gamma = 0.5, \eta_G = 0.48$

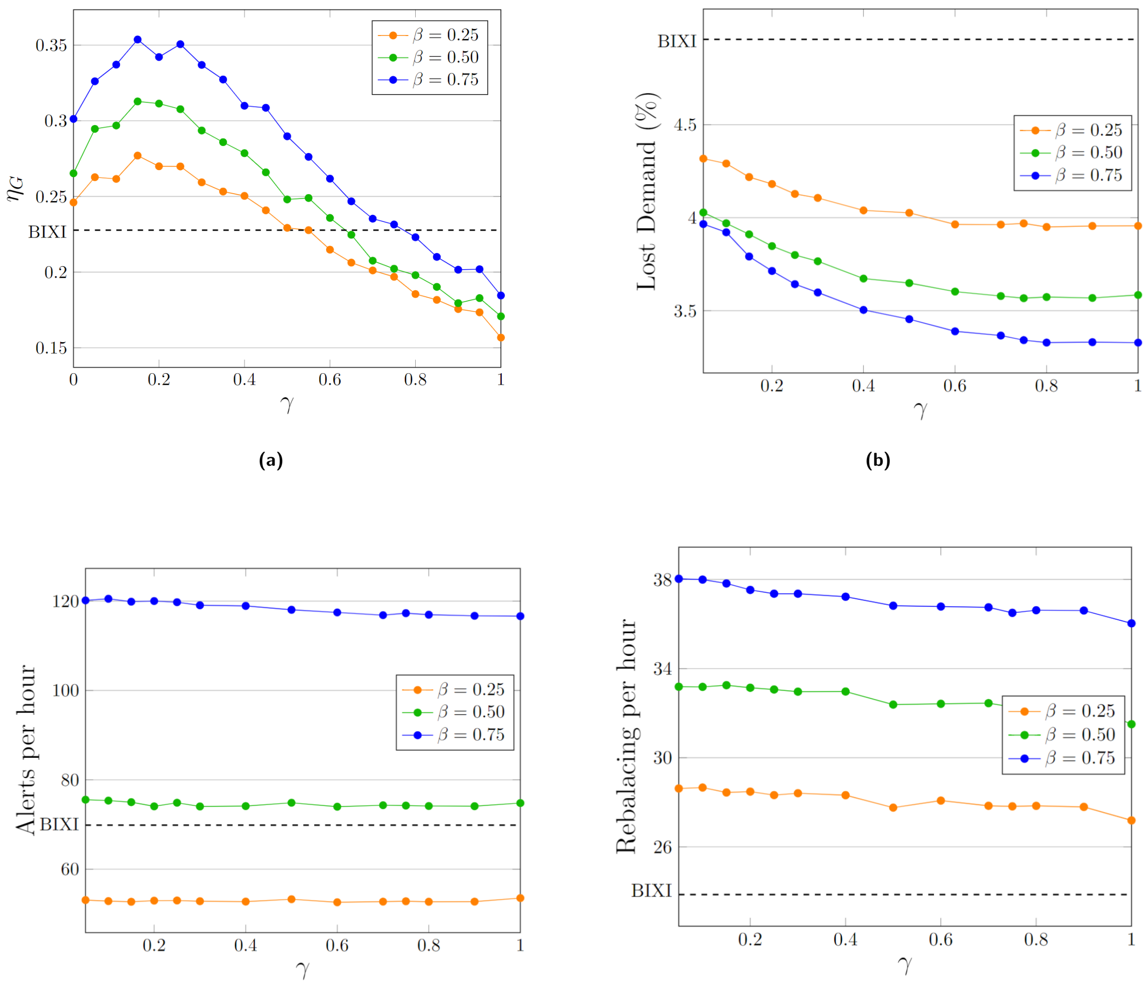
**(c)** $\gamma = 0.75, \eta_G = 0.29$

**(d)** $\gamma = 1, \eta_G = 0.27$

**Figure 6: Stations prioritized by $Pa_4$ on July 19, 2019 at 11 a.m. with different values of $\gamma$.**

J. Shu, M. C. Chou, Q. Liu, C.-P. Teo, I.-L. Wang, Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems, Operations Research 61 (2013) 1346–1359.

S. Wang, T. He, D. Zhang, Y. Shu, Y. Liu, Y. Gu, C. Liu, H. Lee, S. H. Son, Bravo: Improving the rebalancing operation in bike sharing with rebalancing range prediction, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2 (2018) 1–22.




(a) (b)

(c) (d)

**Figure 7:** $Pa_4$ **performance measures.**

P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, E. Fleury, Shared bicycles in a city: A signal processing and data analysis perspective, Advances in Complex Systems 14 (2011) 415–438.

W. El-Assi, M. Salah Mahmoud, K. Nurul Habib, Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in toronto, Transportation 44 (2017) 589–613.

R. C. Hampshire, L. Marla, An analysis of bike sharing usage: Explaining trip generation and attraction from observed demand, in: 91st Annual meeting of the transportation research board, Washington, DC, 2012, pp. 12–2099.

S. Feng, H. Chen, C. Du, J. Li, N. Jing, A hierarchical demand prediction method with station clustering for bike sharing system, in: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), IEEE, 2018, pp. 829–836.

P. Hulot, 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, pp. 378–386.

Y.-C. Yin, C.-S. Lee, Y.-P. Wong, Demand prediction of bicycle sharing systems (2012). [Online].

L. Chen, D. Zhang, L. Wang, D. Yang, X. Ma, S. Li, Z. Wu, G. Pan, T.-M.-T. Nguyen, J. Jakubowicz, Dynamic cluster-based over-demand prediction in bike sharing systems, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2016, pp. 841–852.

K. Gebhart, R. B. Noland, The impact of weather conditions on bikeshare trips in washington, dc, Transportation 41 (2014) 1205–1225.

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

J. Huang, H. Sun, H. Li, L. Huang, A. Li, X. Wang, Central station-based demand prediction for determining target inventory in a bike-sharing system, 2020.

J. Liu, L. Sun, W. Chen, H. Xiong, Rebalancing bike sharing systems: A multi-source data smart optimization, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1005–1014.

T. Raviv, O. Kolka, Optimal inventory management of a bike-sharing station, IIE Transactions 45 (2013) 1077–1093.

S. Datner, T. Raviv, M. Tzur, D. Chemla, Setting inventory levels in a bike sharing network, Transportation Science 53 (2019) 62–76.

D. Chemla, F. Meunier, T. Pradeau, R. W. Calvo, H. Yahiaoui, Self-service bike sharing systems: simulation, repositioning, pricing, Technical Report hal-00824078f, Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique - CERMICS, 2013.

C. Fricker, N. Gast, Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity, Euro journal on transportation and logistics 5 (2016) 261–291.

R. Alvarez-Valdes, J. M. Belenguer, E. Benavent, J. D. Bermudez, F. Muñoz, E. Vercher, F. Verdejo, Optimizing the level of service quality of a bike-sharing system, Omega 62 (2016) 163–175.

J. Brinkmann, M. W. Ulmer, D. C. Mattfeld, Inventory routing for bike sharing systems, Transportation research procedia 19 (2016) 316–327.

T. Bulhões, A. Subramanian, G. Erdoğan, G. Laporte, The static bike relocation problem with multiple vehicles and visits, European Journal of Operational Research 264 (2018) 508–523.

D. Chemla, F. Meunier, R. W. Calvo, Bike sharing systems: Solving the static rebalancing problem, Discrete Optimization 10 (2013) 120–146.

C. Contardo, C. Morency, L.-M. Rousseau, Balancing a dynamic public bike-sharing system, volume 4, Cirrelt Montreal, 2012.

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

G. Erdoğan, M. Battarra, R. W. Calvo, An exact algorithm for the static rebalancing problem arising in bicycle sharing systems, European Journal of Operational Research 245 (2015) 667–679.

M. Lowalekar, P. Varakantham, S. Ghosh, S. D. Jena, P. Jaillet, Online repositioning in bike sharing systems, in: Twenty-seventh international conference on automated planning and scheduling, 2017.

A. Pal, Y. Zhang, Free-floating bike sharing: Solving real-life large-scale static rebalancing problems, Transportation Research Part C: Emerging Technologies 80 (2017) 92–116.

P. Papazek, G. R. Raidl, M. Rainer-Harbach, B. Hu, A pilot/vnd/grasp hybrid for the static balancing of public bicycle sharing systems, in: International Conference on Computer Aided Systems Theory, Springer, 2013, pp. 372–379.

P. Nunes, A. Moura, J. Santos, Solving the multi-objective bike routing problem by meta-heuristic algorithms, International Transactions in Operational Research (2022).

S. Ghosh, P. Varakantham, Y. Adulyasak, P. Jaillet, Dynamic repositioning to reduce lost demand in bike sharing systems, Journal of Artificial Intelligence Research 58 (2017) 387–430.

Y. Lu, U. Benlic, Q. Wu, An effective memetic algorithm for the generalized bike-sharing rebalancing problem, Engineering Applications of Artificial Intelligence 95 (2020) 103890.

P. Papazek, C. Kloimüllner, B. Hu, G. R. Raidl, Balancing bicycle sharing systems: an analysis of path relinking and recombination within a grasp hybrid, in: International Conference on Parallel Problem Solving from Nature, Springer, 2014, pp. 792–801.

Y. Ren, L. Meng, F. Zhao, C. Zhang, H. Guo, Y. Tian, W. Tong, J. W. Sutherland, An improved general variable neighborhood search for a static bike-sharing rebalancing problem considering the depot inventory, Expert Systems with Applications 160 (2020) 113752.

N. Vergeylen, K. Sörensen, P. Vansteenwegen, Large neighborhood search for the bike request scheduling problem, International Transactions in Operational Research 27 (2020) 2695–2714.

A. Kabra, E. Belavina, K. Girotra, Bike-share systems: Accessibility and availability, Management Science 66 (2020) 3803–3824.

T. Raviv, M. Tzur, I. A. Forma, Static repositioning in a bike-sharing system: models and solution approaches, EURO Journal on Transportation and Logistics 2 (2013) 187–229.

D. K. George, C. H. Xia, Fleet-sizing and service availability for a vehicle rental system via closed queueing networks, European journal of operational research 211 (2011) 198–207.

P. Hulot, Towards station-level demand prediction for effective rebalancing in bike-sharing systems, Master's thesis, École Polytechnique de Montréal, 2018.

D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world'networks, Nature 393 (1998) 440–442.