Stochastic Optimization Models for a Bike-Sharing Problem with Transshipment

Francesca Maggioni[a], Matteo Cagnolari[a], Luca Bertazzi[b], Stein W. Wallace[c]

[a] *Department of Management, Economics and Quantitative Methods, University of Bergamo, Italy, francesca.maggioni@unibg.it, matteo.cagnolari@unibg.it*
[b] *Department of Economics and Management, University of Brescia, Italy, luca.bertazzi@unibs.it*
[c] *Department of Business and Management Science, NHH Norwegian School of Economics, Bergen, Norway, stein.wallace@nhh.no*

Abstract

We study the problem faced by a bike-sharing service provider who needs to manage a fleet of bikes over a set of bike-stations, each with given capacity and time-varying stochastic demand. In particular, we focus on One-way bike sharing systems with transshipment in which: 1) The user can pick up a bike at a station and drop it off at a different station; 2) Transshipment of bikes among stations is performed at the end of the day, to have the optimal number of bikes at each station at the beginning of the service on the next day. For this problem, we propose two-stage and multistage stochastic optimization models, to determine the optimal number of bikes to assign to each station at the beginning of the service. Numerical results are provided for the bike-sharing service “LaBiGi” in Bergamo (Italy), from which managerial insights are drawn.

Keywords: Logistics, Bike-sharing, Stochastic programming, Transshipment

1. Introduction

Bike-sharing systems contribute towards obtaining a more sustainable mobility and decreasing traffic and pollution caused by motorized transportation. In Fishman et al. [14], bike-sharing programs are extensively analyzed, identifying sustainability challenges, bike-sharing schemes (e.g., one-way, two-way), mode substitution and impacts, usage rates, user motivation, preferences and purpose, safety concepts and re-balancing. The first bike-sharing system was installed in Amsterdam in 1965 and, since then, these systems have increased their popularity as a new transportation mode in a large number of cities worldwide, counting more than 1,000 operating systems and more than 300 planned or under construction.

Several interesting optimization problems can be studied to improve bike-sharing systems: we refer to Laporte et al. [23] for a survey on the main problems and methods arising in shared mobility systems. One of these is the re-balancing problem and another one is the fleet sizing problem. For the first problem, two types of repositioning can be distinguished: static repositioning and dynamic repositioning. In the static case, the repositioning occurs during the night when the traffic is low, the initial state of the system is considered known and fixed, and the aim of the repositioning is to get the system to a desired, predefined state. Most of the literature considers this type of repositioning. Among the contributions we mention Raviv et al. [31] and Dell'Amico et al. [9] which present mixed–integer linear programming models for the static re-balancing problem. Different exact, heuristic and metaheuristic algorithms for the this type of problem are proposed in Ho and Szeto [19], Erdoğan et al. [12], Forma et al. [15], Dell'Amico et al. [10], Alvarez-Valdes et al. [1] and Datner et al. [8]. Several variants of the static re-balancing problem have been also studied: in particular Li et al. [25] investigate the case in which multiple types of bikes are considered; Cruz et al. [7] study the case in which only one vehicle is available; Ho and Szeto [18] study the multi-vehicle bike-repositioning problem; Kloimüllner and Raidl [22] study the case in which only full vehicle loads are allowed among the rental stations; Szeto and Shui [37] study a bike repositioning problem that determines the routes of the repositioning vehicles and the loading/unloading quantities at each bike station to firstly minimize the positive deviation from the tolerance of total demand dissatisfaction and then service time; finally Wang and Szeto [38] formulate a mixed–integer linear programming model to determine the repositioning of both good and broken bikes, minimizing the total CO2 emissions. The effect of the randomness of user choices for the static repositioning problem have been investigated in Fricker and Gast [16] via stochastic optimization models while Erdoğan et al. [13] study the relocation problem with demand intervals, in which the inventory level at the stations is between given lower and upper bounds.

In the dynamic case, the repositioning system operates while the bike-sharing system is being used. Therefore, users are continuously taking and leaving bicycles at the stations, modifying their states. Among the contributions in this direction, Shu et al. [35] develop a network flow model to estimate the flow of bikes within the network and the number of trips supported, given an initial allocation of bikes at each station. Regue and Recker [32] propose a framework to solve the problem with a proactive approach while Zhang et al. [39] propose a mixed–integer linear programming model and a heuristic. Ghosh et al. [17] propose an optimization formulation to reposition bikes using vehicles, considering the routes for vehicles and future expected demand. Brinkmann et al. [4] study a stochastic-dynamic inventory routing problem for bike sharing systems. Çelebi et al. [5] propose an integrated approach to jointly considering location decisions and capacity allocation. A dynamic green repositioning problem that simultaneously minimizes the total unmet demand and the fuel and CO2 emission cost of the repositioning vehicle has been introduced in Shui and Szeto [36]. Legros [24] apply a Markov Decision Process to decide which station should be prioritized, and the number of bikes to add or remove at each station, in order to minimize the rate of arrival of unsatisfied users. Finally Schuijbroek et al. [34] design a cluster-first route-second heuristic to determine service level requirements at each station and routes to rebalance the inventory levels.

For the second problem (fleet sizing), Sayarshad et al. [33] presents an optimization formulation to design a bike-sharing system for travel inside small communities. Raviv and Kolka [30] study an inventory model for the management of bike rental stations. In Lu [27] the goal is to compute the number of bikes to assign to each bike-station on the basis of time-dependent bike flows in time-space networks. However, in these papers, re-balancing is not considered.

The first contribution of this paper is to propose two-stage and multistage stochastic optimization models, which integrate these two decisions with static repositioning. Multistage models allow us to split the day into a given number of shorter subperiods.

In particular, we study the problem faced by a bike-sharing service provider who needs to manage a fleet of bikes and a set of bike-stations with given capacity and stochastic demand. The corresponding probability distributions are not stationary and can be different for different bike stations. The studied bike-sharing system is a one-way rental system, which allows the user to pick a bike at one station and drop it off at a different bike-station. We formulate this bike-sharing problem as a variant of the *Newsvendor problem* *with transshipment* (see Dong and Rudi [11] and Zhang [40] for the application of transshipment in supply chain management) and investigate the role of the stochasticity. Then, we apply the models to instances inspired by the Bergamo bike-sharing service named “LaBiGi”, which started in May 2013 with 18 stations and then expanded in 2015 to a total of 22 stations located in different areas in the city center. The system is integrated with a mobile app, which allows the users to know in real time the number of available slots in all bike-stations throughout the city.

We carry out a computational experiment aiming at answering the following questions:

1. How does the optimal assignment of bikes to stations depend on the probability distribution of the rental demand?

2. How does the optimal expected total cost depend on the probability distribution of the rental demand?

3. What is the Value of the Stochastic Solution?

4. What is the quality of the expected value solution when solving the stochastic models?

5. What is the value of knowing in advance the probability distribution of the rental demand (Value of Ambiguity)?

6. How does the optimal assignment of bikes to stations and optimal total cost depend on the probability distribution of the rental demand?

7. What is the value of a solution obtained by a multistage optimization model with respect to the one from the two-stage model?

8. What is the effect of letting bike stations have different capacities?

9. How does the allocation plan implemented by the LaBiGi bike-sharing service perform compared to ones of the stochastic models?

The paper is organized as follows. Section 2 describes the bike-sharing problem. Section 3 provides the two-stage and multistage stochastic optimization models formulations. Section 4 shows the computational results that allow us to give an answer to the previous questions. Finally, managerial insights and conclusions are drawn in Section 5.

2. Problem Description

We consider the problem faced by a bike-sharing service provider who needs to manage a fleet of bikes over a set of bike-stations with given capacities, in order to serve the stochastic rental demand over space and time. A unit procurement cost is paid for each bike assigned to each station at the beginning of the service. The operational time frame is one day. The delivery of bikes to bike-stations is assumed to be instantaneous (lead time equal to zero), as this operation can be carried out before the start of the service. Backlogging is not allowed. A unit stock-out cost is paid if realized demand exceeds the number of bikes assigned to a station, and a unit transshipment cost is paid at the end of the rentals, when the bike-station inventory levels are rebalanced. Each bike rental demand is defined by an origin-destination pair, where the destination is unknown to the provider at the time of rental. A stochastic demand to each origin-destination pair is assigned. The rent must start at the user-defined time period or it is lost, determining a shortage for the provider and a reduction of the service level for the user. A shortage realizes when a rental demand arises in a bike-station, but no bikes are available: The user quits the service and looks towards an alternative transportation mode. A shortage causes a cost increase, a reduced service level and a reduced likelihood of future rental requests. The number of bikes that cannot be left in a bike-station as it is full when the user arrives at the station determines an overflow: The user cannot quit the service until the bike is redirected and positioned by the user in the nearest bike-station with available capacity. An overflow causes a waste of time for the user and a cost. Our aim is to determine the number of bikes to assign to each bike-station at the beginning of the service, in order to minimize the expected total costs, given by the sum of the procurement costs, the expected stock-out costs for unmet demand, the expected time-waste costs for overflow and the expected transshipment costs for repositioning bikes at the end of the service.

3. Stochastic Optimization Models

In this section, we present a two-stage and a multistage stochastic optimization formulation for the problem described in the previous section. We refer to Birge and Louveaux [3] and King and Wallace [21] for comprehensive books on Stochastic Programming.

3.1. A Two-Stage Stochastic Optimization Model

We first define the following notation:

Sets.

- $\mathcal{B}$: set of bike-stations, $\mathcal{B} = \{1, \ldots, B\}$;
- $\mathcal{S}$: set of scenarios, $\mathcal{S} = \{1, \ldots, S\}$ or finite set of possible realizations of the uncertainty.

Deterministic parameters.

- $c \in \mathbb{R}^+$: procurement cost per bike at each bike-station at the beginning of the service;
- $v_i \in \mathbb{R}^+$: stock-out cost per bike at bike-station $i \in \mathcal{B}$;
- $w_i \in \mathbb{R}^+$: time-waste cost per bike due to overflow at bike-station $i \in \mathcal{B}$;
- $t_{ij} \in \mathbb{R}^+$: unit transshipment cost per bike transshipped from bike-station i to bike-station j, $i, j \in \mathcal{B}$;
- $k_i \in \mathbb{Z}^+$: capacity of bike-station $i \in \mathcal{B}$.

Stochastic parameters. Let $(\Xi, \mathcal{A}, p)$ be a probability space with Ξ the set of outcomes, σ-algebra $\mathcal{A}$, probability p and $\xi \in \Xi$ a particular outcome representing the rental demand on each origin-destination pair of bike-stations. We define:

- $\xi_{ijs} \in \Xi \subset \mathbb{Z}^+$: rental demand from bike-station i to bike-station j in scenario s, $i, j \in \mathcal{B}$, $s \in \mathcal{S}$.
- $p_s \in [0, 1]$: probability of scenario $s \in \mathcal{S}$. Notice that $\sum_{s=1}^{S} p_s = 1$.

We now introduce the first-stage and second-stage variables. The first-stage variables are defined as follows:

- x_i: number of bikes to assign to bike-station $i \in \mathcal{B}$ at the beginning of the service. We denote with x_i^* an optimal number of bikes to assign to bike-station i and with $\boldsymbol{x} = [x_1, \ldots, x_B]^T$ the first-stage decision vector over all bike-stations. The decision must be taken before the realizations of the random rental demand ξ_{ijs}.

After the placement of the bikes, the stochastic demands ξ_{ijs} occur on each origin-destination pair $i, j \in \mathcal{B}$ and the minimum of the available and requested bikes is actually rented. Then, the surplus or shortage can be immediately computed at each bike-station. The second-stage decision variables are defined as follows:

- β_{ijs}: number of rented bikes from bike-station i to bike-station j in scenario s;
- I^+_{is} : realized surplus of bikes at bike-station i in scenario s. Note that the surplus does not involve any cost for the provider;
- I^-_{ijs}: realized shortage of bikes at origin-destination pair i, j in scenario s;
- ρ_{ijs}: number of redirected bikes from bike-station i to bike-station j in scenario s;
- O^+_{is}: residual capacity at bike-station i in scenario s;
- O^-_{is}: overflow at bike-station i in scenario s;
- τ_{ijs}: number of transshipped bikes from bike-station i to bike-station j in scenario s;
- T^+_{is}: excess of bikes at bike-station i in scenario s;
- T^-_{is} : lack of bikes at bike-station i in scenario s.

Figure 1 shows the sequence of operations for any scenario s.

The problem can be formulated as the following two-stage integer stochastic program:

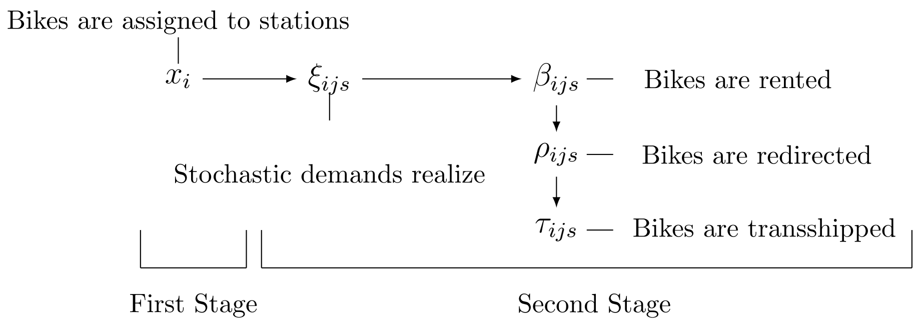


Figure 1: Sequence of operations for any scenario s.

$$
\begin{aligned}
&\min c \sum_{i=1}^{B} x_i + \sum_{s=1}^{S} p_s \sum_{i=1}^{B} \left[v_i \sum_{j=1}^{B} I^-_{ijs} + w_i O^-_{is} + \sum_{j=1}^{B} t_{ij} \tau_{ijs} \right] && && (1)\\
&\text{s.t.} \quad x_i \le k_i, && \forall i \in \mathcal{B}, && (2)\\
&\beta_{ijs} = \xi_{ijs} - I^-_{ijs}, && \forall i, j \in \mathcal{B}, \forall s \in \mathcal{S}, && (3)\\
&I^+_{is} - \sum_{j=1}^{B} I^-_{ijs} = x_i - \sum_{j=1}^{B} \xi_{ijs}, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (4)\\
&O^+_{is} - O^-_{is} = k_i - x_i + \sum_{j=1}^{B} \beta_{ijs} - \sum_{j=1}^{B} \beta_{jis}, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (5)\\
&\sum_{j=1}^{B} \rho_{ijs} = O^-_{is}, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (6)\\
&\sum_{j=1}^{B} \rho_{jis} \le O^+_{is}, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (7)\\
&T^+_{is} - T^-_{is} = k_i - O^+_{is} + \sum_{j=1}^{B} \rho_{jis} - x_i, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (8)\\
&\sum_{j=1}^{B} \tau_{ijs} = T^+_{is}, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (9)\\
&\sum_{j=1}^{B} \tau_{jis} = T^-_{is}, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (10)\\
&x_i, I^+_{is}, O^+_{is}, O^-_{is}, T^+_{is}, T^-_{is} \in \mathbb{Z}^+, && \forall i \in \mathcal{B}, \forall s \in \mathcal{S}, && (11)\\
&\tau_{ijs}, \beta_{ijs}, \rho_{ijs}, I^-_{ijs} \in \mathbb{Z}^+, && \forall i, j \in \mathcal{B}, \forall s \in \mathcal{S}. && (12)
\end{aligned}
$$

The objective function (1) is the minimization of the expected total cost, given by the sum of the procurement cost for the assigned bikes and the expected stock-out cost for shortage, the expected time-waste cost for overflow and the expected transshipment cost for repositioning bikes at the end of the service. Constraints (2) enforce the assigned bikes to be not greater than the bike-station capacity. Constraints (3) compute the number of rented bikes, given by the difference between the stochastic demand and the shortage. Notice that realized shortage has been defined as a variable. This makes the model optimistic. Constraints (4) ensure the balance between surplus and shortage, while constraints (5) ensure the balance between the residual quantity and the overflow. Constraints (6) define the sum of all the redirected bikes from a bike-station equal to its overflow, while constraints (7) guarantee that the sum of all the redirected bikes to a bike-station cannot exceed its residual capacity. Constraints (8) ensure the balance between exceeding and failure. Constraints (9) define the sum of all the transshipped bikes from a bike-station equal to its excess, while constraints (10) guarantee that the sum of all the transshipped bikes to a bike-station is equal to its failure. At last, constraints (11)-(12) define the integrality and non-negativity of first-stage and second-stage variables.

3.2. A Multistage Stochastic Optimization Model

We now formulate a multistage stochastic optimization model for the problem described in Section 2. The model aims at determining, for each bike-station, the optimal number of bikes to assign to each bike-station at the beginning of the service, to minimize the expected total cost, taking into account the dynamic nature of rental demands over the day. The main drawback of the two-stage formulation is given by the assumption that each bike is rented only once per day. In order to overcome this, we now split the daily rental demand of each origin-destination pair into three rental demand periods; morning, afternoon and evening, and we allow a bike to be rented in all three periods. The night is reserved for transshipment operations in order to re-balance the number of bikes in all the stations. The number of bikes available in a period in a station, is hence a function of initial distribution, rentals in the periods before, including re-directions and transshipment operations. So a bike can be rented up to three times in a day. Let us introduce the following notation, in addition to the one used in previous Section:

- $\mathcal{N} := \{n : n = 0, \ldots, N\}$: ordered set of nodes of the scenario tree structure;

- $0 \in \mathcal{N}$: root of the scenario tree;
- $\mathcal{F} := \{n : n = N - F + 1, \ldots, N\} \subset \mathcal{N}$: set of the leaves of the tree, that is the nodes in the last stage of the model; we have that the number of scenarios $S = F = |\mathcal{F}|$;
- pa(n): parent of node $n \in \mathcal{N} \setminus \{0\}$;
- x_{i0}: total number of bikes to assign to bike-station $i \in \mathcal{B}$ at the beginning of the service, represented by the root node $0 \in \mathcal{N}$;
- p_n: probability of node $n \in \mathcal{N}$, defined as:

$$p_n = \begin{cases} \frac{1}{|\mathcal{F}|} & \text{if } \ n \in \mathcal{F}, \\ \sum_{m \in \mathcal{N} \setminus \{0\}, pa(m)=n} p_m & \text{if } \ n \in \mathcal{N} \setminus \mathcal{F}. \end{cases}$$

All the recourse variables are defined as in the previous section by replacing the index s with n.

The multistage model can be formulated as follows:

$$\min c\sum_{i=1}^{B} x_{i0} + \sum_{n=1}^{N} p_n \left[\sum_{i=1}^{B} \left[v_i \sum_{j=1}^{B} I_{ijn}^{-} + w_i O_{in}^{-} \right] \right] + \sum_{n=N-F+1}^{N} p_n \left[\sum_{i=1}^{B} \sum_{j=1}^{B} t_{ij} \tau_{ijn} \right] \tag{13}$$

$$\begin{aligned}
\text{s.t.} \quad & x_{i0} \leq k_i, && \forall i \in \mathcal{B}, && (14)\\
& \beta_{ijn} = \xi_{ijn} - I_{ijn}^{-}, && \forall i,j \in \mathcal{B}, \forall n \in \mathcal{N} \setminus \{0\}, && (15)\\
& I_{in}^{+} - \sum_{j=1}^{B} I_{ijn}^{-} = x_{i0} - \sum_{j=1}^{B} \xi_{ijn}, && \forall i \in \mathcal{B}, \forall n \in \mathcal{N} \setminus \{0\}, && (16)\\
& O_{in}^{+} - O_{in}^{-} = k_i - x_{i0} + \sum_{j=1}^{B} \beta_{ijn} - \sum_{j=1}^{B} \beta_{jin}, && \forall i \in \mathcal{B}, \forall n \in \mathcal{N} \setminus \{0\}, && (17)\\
& \sum_{j=1}^{B} \rho_{ijn} = O_{in}^{-}, && \forall i \in \mathcal{B}, \forall n \in \mathcal{N} \setminus \{0\}, && (18)\\
& \sum_{j=1}^{B} \rho_{jin} \leq O_{in}^{+}, && \forall i \in \mathcal{B}, \forall n \in \mathcal{N} \setminus \{0\}, && (19)\\
& T_{in}^{+} - T_{in}^{-} = k_i - O_{in}^{+} + \sum_{j=1}^{B} \rho_{jin} - x_{i0}, && \forall i \in \mathcal{B}, \forall n \in \mathcal{F}, && (20)\\
& \sum_{j=1}^{B} \tau_{ijn} = T_{in}^{+}, && \forall i \in \mathcal{B}, \forall n \in \mathcal{F}, && (21)\\
& \sum_{j=1}^{B} \tau_{jin} = T_{in}^{-}, && \forall i \in \mathcal{B}, \forall n \in \mathcal{F}, && (22)\\
& x_{i0} \in \mathbb{Z}^{+}, && \forall i \in \mathcal{B}, && (23)\\
& \tau_{ijn}, T_{in}^{+}, T_{in}^{-} \in \mathbb{Z}^{+}, && \forall i,j \in \mathcal{B}, \forall n \in \mathcal{F}, && (24)\\
& \beta_{ijn}, \rho_{ijn}, I_{in}^{+}, I_{ijn}^{-}, O_{in}^{+}, O_{in}^{-} \in \mathbb{Z}^{+}, && \forall i,j \in \mathcal{B}, \forall n \in \mathcal{N} \setminus \{0\}. && (25)
\end{aligned}$$

The objective function (13) is the minimization of the expected total cost, given by the sum of the procurement cost for the assignment of bikes to stations at the beginning of the service, the expected stock-out cost for the shortage, the expected time-waste cost for the overflow and the expected transshipment cost for the transshipped bikes in the last time stage. Constraints (14) enforce the assigned bikes at the beginning of the day to be not greater than the bike-station capacity. Constraints (15) define the rented quantities, given by the difference between the stochastic demand and the shortage. Constraints (16) ensure the balance between surplus and shortage, while constraints (17) ensure the balance between the residual quantity and the overflow. Constraints (18) define the sum of all the redirected bikes from a bike-station equal to its overflow, while constraints (19) guarantee that the sum of all the redirected bikes to a bike-station cannot exceed its residual capacity. Constraints (20) ensure the balance between exceeding and failure. Constraints (21) define the sum of all the transshipped bikes from a bike-station equal to its exceed, while constraints (22) guarantee that the sum of all the transshipped bikes to a bike-station is equal to its failure, at the end of last time period. Finally, constraints (23)-(24)-(25) define the integrality and non-negativity of the first-stage and recourse variables, respectively.

4. Numerical Results

In this section, we provide numerical results to give an answer to the questions we stated in the Introduction.

We use AMPL and CPLEX 12.8.0 to formulate and solve the stochastic programming models. All the computations were run on a 64-bit machine with 12 GB of RAM and a 3.2 gigahertz processor and all the models presented are solved to optimality. Summary statistics on the CPU time and number of MIP iterations required by CPLEX to find an optimal solution will be reported later.

4.1. Data Analysis and Scenario Generation for the Two-stage Model

We generate a set of instances inspired by the real bike-sharing service "LaBiGi" in Bergamo (Italy). This service started in May 2013 with 18 bike-stations. Then, it was expanded in 2015 and now has a total of 22 operational bike-stations. Let $\mathcal{B} = \{\mathscr{A}, \mathscr{B}, \ldots, \mathscr{V}\}$ be the set of bike-stations. The mobile app `https://www.atb.bergamo.it/it/mobilita/bike-sharing` allows the user to know in real-time, the number of available slots in all bike-stations throughout the city. Notice that the number of re-directed bikes ρ_{ijs} modelled in Section 3, can be interpreted as an approximation of what the app would do in reality. Since the models optimize the system, while the app is myopic, the models are optimistic. We first focus on the following instance:

- Unit procurement cost $c = 2$ (Euros);

- Unit stock-out cost $v_i = 4$ (Euros), $i \in \mathcal{B}$;

- Unit time-waste cost $w_i = 8$ (Euros), $i \in \mathcal{B}$;

- Unit transshipment cost $t_{ij} = 1$ (Euros), $i, j \in \mathcal{B}$;

- Bike-station capacity k_i, $i \in \mathcal{B}$;

where the unit procurement cost c is and the bike-station capacities k_i are set based on the Bergamo case, while the rest are our best guesses. Notice that the system has recently been updated with a different capacity k_i at each station i. Results based on this case will be discussed in Section 4.10. However, in order to understand how the probability distribution of rental demands affect the optimal assignment of bikes and the total cost, we now keep the capacities of the stations constant and equal to 30, as they were originally in Bergamo. We generally allow stock-out and time-waste costs to depend on bike-station (to reflect potential geographical preferences), but in our example they are all bike-station independent. We deliberately choose a large value of the time-waste cost w, since we want to penalize the overflow and consequently the time-costly redirection of bikes carried out by the user to find the nearest bike-station with an available slot. We assume that the random variable rental demand ξ_{ij} between each pair of bike-stations i and j has a finite number of scenarios (as discussed in Section 3), and the outcomes are integer and exogenous to the problem. Consequently, its probability distribution is not influenced by decisions. Several methods are adopted in the literature for discretizing distributions and generating scenarios. Among the most common we list: Conditional sampling, sampling from specified marginals and correlations, moment matching, path-based methods and optimal discretization. For a short overview of scenario-generation methods, see Chapter 4 of King and Wallace [21].

The data provided by the bike-sharing system on the stochastic rental demand ξ_{ij} between each pair of bike-stations i and j, are by the month for the years 2013, 2014 and 2015, where we assume summer from April to September and winter from October to March. However, since the time scale of the model is a day, daily rental demands are then estimated by normalizing monthly data by the number of days in a month and rounding to the nearest integer below. We think this is good enough for our tests, but are aware that our approximation underestimates the variation. Moreover, due to the limited size of the available historical data, we compute for each origin-destination pair (i, j), the minimum m_{ij} and maximum demand M_{ij}, its expected value $\mathbb{E}[\xi_{ij}] = \bar{\xi}_{ij}$, its standard deviation σ_{ij} and we assume $\xi_{ij} \in [m_{ij}, M_{ij}]$. Since the inverse of the Kaplan-Meier estimate of the cumulative distribution function (also known as the empirical cdf) of the real rental demand cannot be built, due to the limited amount of available data, we assume that there are four possible probability distributions: Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ and scenarios are sampled according to a Monte Carlo sampling procedure for each of them. Consequently we suppose that the four distributions have the same mean $\bar{\xi}_{ij} := \sum_{s=1}^{S} \frac{\xi_{ijs}}{S}$. For the standard deviation we suppose that, in the case of normal distribution, its standard deviation $\sigma_{ij}^{\mathcal{N}}$ is equal to the one derived from historical data $\sigma_{ij} := \sqrt{\frac{\sum_{s=1}^{S}(\xi_{ijs} - \bar{\xi}_{ij})^2}{S-1}}$, while for the other probability distributions we apply the corresponding definitions as follows:

- $\sigma_{ij}^{\mathcal{U}} = \sqrt{\frac{1}{12}(M_{ij} - m_{ij})^2}$: standard deviation of the Uniform distribution $\mathcal{U}$;
- $\sigma_{ij}^{\mathcal{E}} = \bar{\xi}_{ij}$: standard deviation of the Exponential distribution $\mathcal{E}$;
- $\zeta_{ij}^{\mathcal{L}} = \ln(\bar{\xi}_{ij}) - \frac{1}{2}\ln(1 + \mathrm{v}_{ij}^2)$: location parameter of the Log-Normal distribution $\mathcal{L}$;
- $\eta_{ij}^{\mathcal{L}} = \ln(1 + \mathrm{v}_{ij}^2)^{\frac{1}{2}}$: scale parameter of the Log-Normal distribution $\mathcal{L}$,

with v_{ij} the variation coefficient Papoulis [29]. Note that we do not require the same standard deviation, otherwise we a-priori exclude some probability distributions. We think such a requirement would be too strong for the problem we study, especially for new bike-stations for which no data are available. In the following, we will consider the four aforementioned probability distributions, even if the analysis can be done with any possible distribution which match the mean $\bar{\xi}_{ij}$ from the data. In section 4.7 we will analyze the value of knowing in advance the probability distribution of rental demand.

Scenario trees, with increasing cardinality $S = 100, 200, \ldots, 1000$, are then generated using the four probability distributions listed above according to Monte Carlo sampling. A discussion on the influence of the size of the scenario tree with tests on the quality of the solution of the optimization model is presented in the next section.

In the following, we will focus our analysis on the winter months only, as similar results were obtained for the summer months.

4.2. The Number of Scenarios for the Two-stage Model

We first perform an in-sample analysis (see Kaut and Wallace [20]) to establish the number of scenarios to use in our computational study. The analysis is carried out by optimally solving the two-stage stochastic program with scenario trees of increasing size.

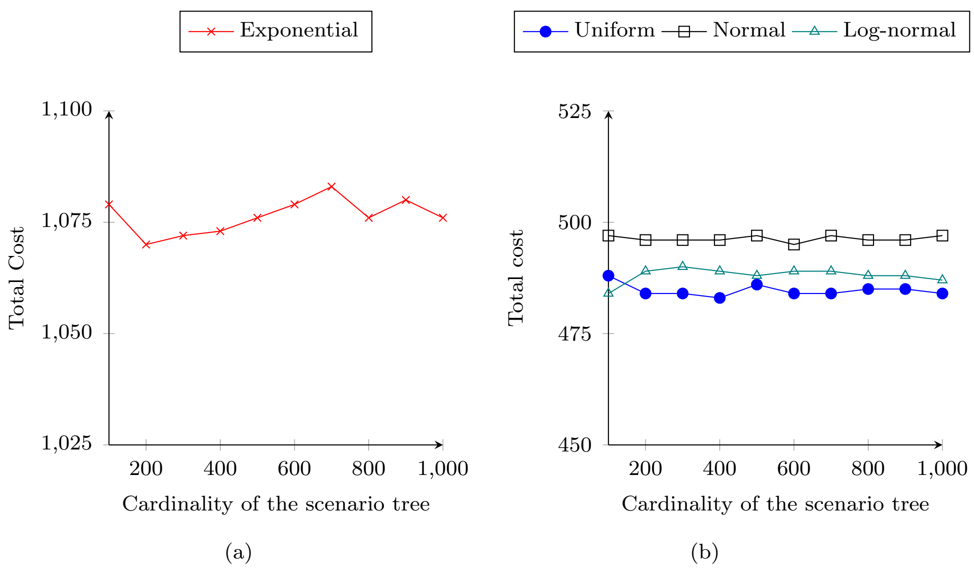


Figure 2: In-sample stability for the expected total cost under the Exponential distribution (a) and Uniform, Normal and Log-normal distributions (b) of the rental demand.

The results, illustrated in Figure 2, show that the objective function value is quite stable over the four probability distributions for an increasing number of scenarios. In the following we use as benchmark scenario trees with 500 scenarios.

We now provide statistics concerning the performance of a state-of-the-art solver (CPLEX 12.8.0) to find an optimal solution over 20 instances. In particular, Table 1 reports the CPU time (in seconds) and the number of MIP iterations required by CPLEX to find an optimal solution of the corresponding two-stage stochastic programming model with its number of binary, integer, continuous variables and constraints (equalities and inequalities) where the integrality on the recourse variables $\rho_{ijs}, I^+_{is}, I^-_{ijs}, O^+_{is}, O^-_{is}, T^+_{is}, T^-_{is},\ i, j \in$

$\mathcal{B}$, $s \in \mathcal{S}$, have been relaxed because of the integer nature of random demand. Results are reported for an increasing number of scenarios S and refer to the adjusted problem after presolve has been applied to the original problem.

Table 1: Summary statistics.

Number of Scenarios	100	300	500	700	900	1000
CPU time (s)	3.53	6.71	93.29	311.672	541.125	757.938
MIP iter.	9251	8115	52253	97350	115570	155553
Binary V.	10952	32805	54472	76197	98091	108784
Integer V.	3323	10137	16912	23460	30353	33757
Continuous V.	122053	366320	610362	854235	1098622	1220519
Eq. Constr.	27453	82520	137362	192035	247222	274519
Ineq. Constr.	2200	6600	11000	15400	19800	22000

4.3. Optimal Assignment of Bikes to Stations

In this section, we aim at answering our first question: How does the optimal assignment of bikes to stations depend on the probability distribution of the rental demand?

Table 2 shows the optimal number of bikes x_i^*, i.e. the number of bikes assigned to each bike-station $i \in \mathcal{B}$ at the beginning of the service, under the four considered probability distributions, using 500 scenarios.

Table 2: Optimal number of bikes assigned to each bike-station $i \in \mathcal{B}$ for the four different probability distributions.

x_i^*	Station i										
	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
$\mathcal{U}$	30	0	2	3	3	11	28	3	6	28	11
$\mathcal{E}$	15	3	10	6	14	26	14	18	18	23	20
$\mathcal{N}$	30	0	2	3	3	12	28	3	6	28	11
$\mathcal{L}$	30	0	2	3	3	11	28	3	6	28	11

x_i^*	Station i											
	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total
$\mathcal{U}$	0	2	5	7	16	2	9	12	6	13	6	203
$\mathcal{E}$	0	8	14	19	23	9	23	27	22	25	18	355
$\mathcal{N}$	0	2	5	7	16	3	9	12	6	13	6	205
$\mathcal{L}$	0	2	5	7	16	2	8	12	5	13	5	200

The results show that the optimal solutions under the assumption of Uniform, Normal and Log-normal distributions are similar, with a total number ranging from 200 to 205 (see last column), while a different behavior is obtained with the Exponential distribution: When we place more under the former distributions, we place less under the latter, and vice versa. This is mainly due to the different shape of the Exponential distribution. The average optimal number of bikes over all the bike-sharing stations is 9 under the assumptions of Uniform, Normal and Log-normal distributions, while it is 16 bikes under the assumption of the Exponential distribution. Interestingly, the optimal stochastic solution suggests a fleet size, as the total number of bikes initially allocated to the stations, to be less than 1/3 of the $660 = 30 \cdot 22$ available slots, for Uniform, Normal and Log-normal distributions while more than 1/2 for the Exponential. This low "utilization rate" of slots is there to avoid too high overflow, stock-out and transshipping costs.

4.4. Optimal Expected Total Cost

In this section, we aim at answering our second question: How does the optimal expected total cost depend on the probability distribution of the rental demand?

From Figure 2 we can see that the highest expected total costs are obtained under the assumption of the Exponential distribution, while the lowest is by the Uniform distribution. This is justified by the different values of the standard deviations of the two distributions and the different shape of the Exponential distribution.

We now carry out a systematic computational experiment to test how the optimal solution of the two-stage model depends on the unit costs. The obtained results are reported in Tables 3 and 4, which respectively show the optimal total cost and, for the sake of completeness, the optimal first-stage solution for the bike-stations $\mathscr{A}$, $\mathscr{B}$, $\mathscr{J}$ and $\mathscr{P}$ (the ones with the highest rental demand), under the four probability distributions.

The results show that the optimal first-stage decisions and the corresponding expected total cost vary depending on the ratio between the costs c, v, w and t, on the demand distribution and the capacity of the bike-stations. In particular, for instance 3, the optimal inventory levels are set to the minimum realization of the demand, while instance 8 sets them to the available capacity (see Table 4).

4.5. Analyzing the Value of the Stochastic Solution

In this section, we aim at answering to our third question: What is the value of the stochastic solution? This is done by considering the well known

Table 3: Expected total cost for a given combination of unit cost parameters under the assumption of Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ probability distributions.

		Total cost			
Instance	Parameters	$\mathcal{U}$	$\mathcal{E}$	$\mathcal{N}$	$\mathcal{L}$
1	$c = v = w = t = 5$	1057	2015	1068	1036
2	$c = w = t = 5, v = 5c$	1400	3544	1436	1424
3	$c = v = t = 5, w = 5c$	1057	2015	1068	1037
4	$c = v = w = 5, t = 5c$	1057	2015	1068	1037
5	$c = 5, v = 2c, w = 3c, t = \frac{1}{5}c$	1207	2637	1228	1198
6	$c = 5, v = 3c, w = 2c, t = \frac{1}{5}c$	1270	2953	1298	1274
7	$t = 5, v = 2t, w = 3t, c = \frac{1}{5}t$	382	1215	395	396
8	$t = 5, v = 3t, w = 2t, c = \frac{1}{5}t$	412	1398	425	432

Table 4: Optimal number of bikes to assign to bike-stations $\mathscr{A}$, $\mathscr{B}$, $\mathscr{J}$ and $\mathscr{P}$ for a given combination of unit cost parameters under the assumption of Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ probability distributions.

	x^*															
	$\mathcal{U}$				$\mathcal{E}$				$\mathcal{N}$				$\mathcal{L}$			
Instance	$\mathscr{A}$	$\mathscr{G}$	$\mathscr{J}$	$\mathscr{P}$	$\mathscr{A}$	$\mathscr{G}$	$\mathscr{J}$	$\mathscr{P}$	$\mathscr{A}$	$\mathscr{G}$	$\mathscr{J}$	$\mathscr{P}$	$\mathscr{A}$	$\mathscr{G}$	$\mathscr{J}$	$\mathscr{P}$
1	18	17	17	8	4	3	6	6	16	17	14	9	4	4	3	4
2	30	30	30	18	19	20	30	30	30	30	30	19	30	30	30	18
3	10	11	12	9	4	3	6	4	8	6	6	4	10	13	14	9
4	12	15	15	9	4	3	6	6	11	13	14	9	12	18	16	9
5	30	28	29	17	16	15	24	25	30	28	28	17	30	28	28	16
6	30	28	30	18	18	17	28	28	30	29	29	18	30	29	29	17
7	30	29	30	19	19	18	28	28	30	30	29	19	30	29	29	18
8	30	30	30	19	21	22	30	30	30	30	30	20	30	30	30	19

the *Value of the Stochastic Solution* (VSS) (see Birge and Louveaux [3]). VSS indicates the expected gain from solving the stochastic recourse problem (RP) rather than its deterministic counterpart (in which the random parameters are replaced with their expected values) and it is defined as follows:

$$VSS = EEV - RP, \tag{26}$$

where EEV denotes the solution value of the RP model, having the first-stage decision variables fixed at the optimal values obtained by setting the coefficients equal to their expected values.

Table 5 shows the optimal number of bikes assigned to the bike-stations under the expected value demand $\bar{\xi}_{ij}$, $i, j \in \mathcal{B}$. We get 105 bikes (see last column), which is not enough to deal with the high demand scenarios, yielding a loss of approximately 30% of the total cost in a stochastic setting (see VSS values in Table 6), with the largest loss (37.7%) in the case of the Exponential distribution.

Table 5: Expected Value Solution $\bar{x}_i(\bar{\xi})$, $i \in \mathcal{B}$.

	Station i										
	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
$\bar{x}_i(\xi)$	21	0	1	2	1	2	22	1	0	19	4

	Station i											
	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total
$\bar{x}_i(\xi)$	0	1	2	2	10	0	3	5	1	6	2	105

Table 6: VSS and the corresponding losses, under the assumption of Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ probability distributions of the rental demand.

	Distribution			
	$\mathcal{U}$	$\mathcal{E}$	$\mathcal{N}$	$\mathcal{L}$
VSS	145	406	144	140
$\frac{VSS}{RP} \cdot 100$	29.8	37.7	29.0	28.7

4.6. Analyzing the Quality of the Expected Value Solution

In this section, we aim at answering our fourth question: What is the *quality* of the expected value solution in the stochastic setting? An approach to answer this question has been proposed in Maggioni and Wallace [28] and extended in Crainic et al. [6]. Starting from the solution of the expected value problem, it assesses whether 1) the deterministic model produced the right non-zero variables, but possibly was off on the *values* of these variables; and 2) the deterministic solution is upgradable to become good (if not optimal) in the stochastic setting. The resulting measures, called *Loss Using the Skeleton Solution* ($LUSS$) and the *Loss of Upgrading the Deterministic Solution* ($LUDS$) are obtained by restricting the values of the first stage variables based on the solution of the expected-value problem. The former is defined as the difference between the *Expected Skeleton Solution Value* $ESSV$, obtained by fixing, in the stochastic program, the non-basic variables of the first stage expected value solution, and the objective value of the stochastic problem RP, given by

$$LUSS = ESSV - RP. \tag{27}$$

The latter measure $LUDS$, is defined as the difference of the *Expected Input Value* EIV obtained by solving a restricted stochastic model by using as a lower bound on all variables their corresponding values in the expected value solution, and the objective value of the stochastic problem RP, given by

$$LUDS = EIV - RP. \tag{28}$$

Table 7 shows, for the four probability distributions, the value of $LUSS$ and of $LUDS$, and the corresponding percentage losses. The results show

Table 7: $LUSS$, $LUDS$, and the corresponding losses, under the assumption of Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ probability distributions.

	Distribution			
	$\mathcal{U}$	$\mathcal{E}$	$\mathcal{N}$	$\mathcal{L}$
$LUSS$	16	56	14	66
$\frac{LUSS}{RP} \cdot 100$	3.3	5.2	2.8	3.3
$LUDS$	0	12	0	1
$\frac{LUDS}{RP} \cdot 100$	0.0	1.0	0.0	0.0

that $LUSS$ is positive under all the four considered distributions: This means that the deterministic solution has a somewhat bad structure (skeleton solution), allocating bikes in a lower number of stations than in the here-and-now solution. More specifically, the expected value solution does not considers $\mathscr{B}$, $\mathscr{I}$, $\mathscr{L}$, $\mathscr{Q}$ (see Table 5), while the stochastic one only disregards stations $\mathscr{B}$ (with exception of the Exponential distribution) and $\mathscr{L}$ (see Table 2). This means that the expected value solution is bad in a stochastic setting, as it chooses the wrong stations where to allocate bikes and, at the same time, it allocates the wrong number of bikes in the selected stations.

From the last two lines of Table 7, we finally observe that the Expected Value Solution is upgradeable to become good in the stochastic setting only for the Uniform and Normal distributions, since in these two cases we have $LUDS = 0$. This means that, for these two distributions, the number of bikes allocated by the deterministic solution in all the stations, can be considered as a lower bound for the optimal stochastic solution.

4.7. Analyzing the Value of Ambiguity

In this section, we aim at answering our fifth question: What is the value of knowing in advance the probability distribution of the rental demand (Value of Ambiguity)?

A critical assumption of the stochastic programming approach, is that the probability distribution of uncertain parameters, in this case the rental demand, is known. However, such an assumption can be too strong in the problem we study, especially for new bike-stations for which no data are available. In this case, the decision-maker has to choose one distribution, without having information on its true shape, to solve the model. Therefore, the optimal solution obtained on the basis of this guessed probability distribution can perform poorly under the true demand distribution, when it will be revealed. To handle this case, in Section 4.1, we have assumed to know just the support and the first moment of the probability distribution. After selecting one probability distribution $\mathcal{G}$ (guessed), we optimally solve the model using $\mathcal{G}$ with the parameters we know. Let $\boldsymbol{x}^*_{\mathcal{G}}$ be the corresponding optimal value of the first-stage variables. Then, we compute the total cost of $\boldsymbol{x}^*_{\mathcal{G}}$ by using the probability distribution $\mathcal{R}$ (the right one) that will actually happen. We call this value *Out-of-Distribution value* (OD), while $RP_{\mathcal{R}}$ is the optimal cost of solving the model with the right probability distribution (for the definition and more details see Bertazzi et al. [2]).

Table 8 reports the OD values and the percentage losses of the *Value of the Right Distribution* (VRD), where VRD is computed as follows:

$$VRD = OD - RP_{\mathcal{R}},$$

which is the loss from solving the two-stage stochastic program using a solution obtained assuming a guessed probability distribution $\mathcal{G}$ instead of the right one $\mathcal{R}$. VRD is a measure of the importance of ambiguity in the problem considered. Results show that the ambiguity plays an important role only when the Exponential distribution is considered as possible candidate to be both right and guessed. The VRD is negligible for all the other possible mismatches between distributions, since the corresponding optimal first-stage solutions are similar (see Table 2). We therefore see that the uncertainty of the future, plays a crucial role for determining the Value of the Right Distribution.

4.8. Testing the role of the standard deviation

In this section we aim to answer to our sixth question: How does the op-

Table 8: *Out-of-Distribution* (OD) values and *Value-of-the-Right-Distribution* (VRD), under the assumption of Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ probability distributions.

	$\mathcal{G}$			
OD	$\mathcal{U}$	$\mathcal{E}$	$\mathcal{N}$	$\mathcal{L}$
$(\mathcal{R} = \mathcal{U})$	-	877	486	487
$(\mathcal{R} = \mathcal{E})$	1348	-	1344	1355
$(\mathcal{R} = \mathcal{N})$	495	877	-	496
$(\mathcal{R} = \mathcal{L})$	489	879	490	-

	$\mathcal{G}$			
$\frac{VRD}{RP} \cdot 100$	$\mathcal{U}$	$\mathcal{E}$	$\mathcal{N}$	$\mathcal{L}$
$(\mathcal{R} = \mathcal{U})$	-	80.45	0	0.20
$(\mathcal{R} = \mathcal{E})$	25.27	-	24.90	25.92
$(\mathcal{R} = \mathcal{N})$	0	76.45	-	0
$(\mathcal{R} = \mathcal{L})$	0.20	80.12	0.40	-

timal assignment of bikes to stations and total costs depend on the standard deviation of the rental demand? We do not consider changing the standard deviation of the Exponential and Uniform distributions since this would imply changing also the mean of the Exponential distribution and the support of the Uniform distribution. As example we report in Table 9 the optimal number of bikes x_i^*, i.e., the number of bikes assigned to each bike-station $i \in \mathcal{B}$ at the beginning of the service, under the Normal probability distribution having standard deviation modified by $\Delta\sigma_{ij}^{\mathcal{N}} = +20\%$ and $\Delta\sigma_{ij}^{\mathcal{N}} = -20\%$ relative to the one derived from the historical data, σ_{ij}. Results show that increasing the standard deviation by 20% implies assigning only one bike more in station $\mathscr{E}$ with an increased cost of 2.81% (511.45 instead of 497), while decreasing the standard deviation by 20% implies assigning only one bike less in station $\mathscr{H}$ with a decreased cost of 2.75% (483.30 instead of 497). From the test we can conclude that if we know the support and the probability distribution of the rental demand, it is not so crucial to have an exact estimate of the standard deviation to find a good fleet size.

Table 9: Optimal number of bikes assigned to each bike-station $i \in \mathcal{B}$ for the Normal probability distribution and standard deviation modified of $\Delta\sigma_{ij}^{\mathcal{N}} = +20\%$ and $\Delta\sigma_{ij}^{\mathcal{N}} = -20\%$ from σ_{ij}.

x_i^*	Station i										
$\Delta\sigma_{ij}^{\mathcal{N}}$	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
+20%	30	0	2	3	4	12	28	3	6	28	11
−20%	30	0	2	3	3	12	28	2	6	28	11

x_i^*	Station i												Total Cost
$\Delta\sigma_{ij}^{\mathcal{N}}$	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total	
+20%	0	2	5	7	16	3	9	12	6	13	6	206	511.45
−20%	0	2	5	7	16	3	9	12	6	13	6	204	483.30

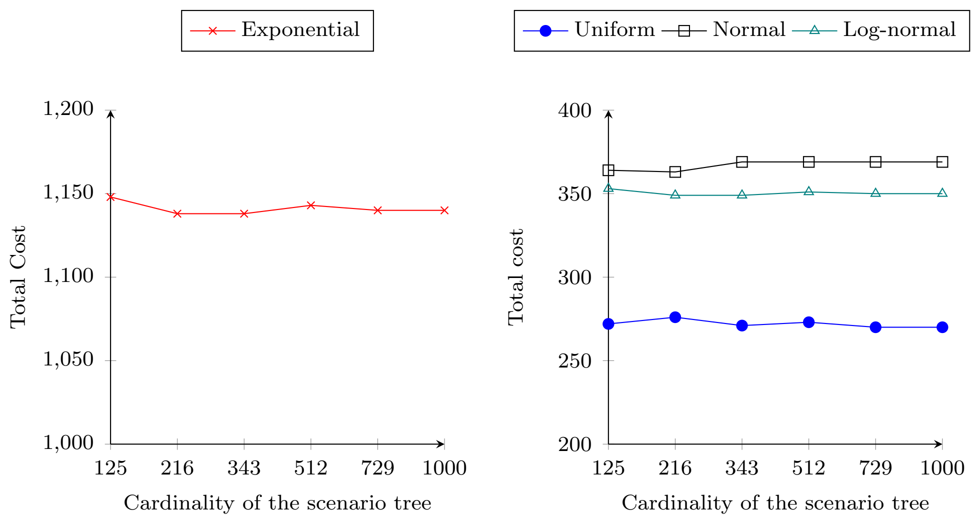
4.9. Solving the Multistage Stochastic Programming Model

In this section we aim to answer our seventh question: What is the value of the solution obtained by the two-stage stochastic optimization model with respect to the one from the multistage one? In other words, we test the effect of adding several stages (i.e., stages with branching) on the quality and cost of the first-stage solution.

In the multistage setting, the day is split into three time periods: From 6 am to 12 noon, from 12 noon to 6 pm and from 6 pm to midnight, while transshipment operations take place from midnight to 6 am. The demand in the second period is assumed to be larger than in the first and in the third, but the sum of the three corresponds to the daily demand of the corresponding scenario considered for the two-stage model. Four-stages scenario trees with increasing and constant branching factors (5 to 10) among periods have been built, for a total number of scenarios from $S = 5 \times 5 \times 5 = 125$ to $S = 10 \times 10 \times 10 = 1000$. As for the two-stage case, in-sample stability has been tested, identifying the minimum number of scenarios needed in order to get stable results in terms of total costs. We ended up with $S = 512$ scenarios (see Figure 3). We declare the four-stage scenario tree with $S = 512 = 8 \times 8 \times 8$ scenarios and $N = 585$ nodes, to be the true representation of the real world and use it as a benchmark to compare the solution obtained using the two-stage scenario tree described in Section 4.2.

As for the two-stage case we now provide statistics concerning the performance of CPLEX 12.8.0 to find an optimal solution over 20 instances. Table 10 reports the CPU time (in seconds) and the number of MIP iterations required by CPLEX to find an optimal solution of the corresponding multistage stochastic programming model with its number of binary, integer, continuous variables and constraints. Results refer to the adjusted problem after presolve has been applied to the original problem. As we see CPU times are not significant in any of the cases.

Comparing Tables 2 and 11, we observe that the optimal number of bikes to assign to each bike-station in the multistage setting (see Table 11), are approximately 2/3 of what we obtained in the two-stage setting (see Table 2) in the case of Uniform, Normal and Log-normal distributions, while a



(a) (b)

Figure 3: In-sample stability for the expected total cost of the four-stage problem under the assumption of (a) Exponential and (b) Uniform, Normal and Log-normal probability distributions of rental demand.

Table 10: Summary statistics.

Number of Scenarios	125	216	343	512
CPU time (s)	1.96	3.57	6.48	9.67
MIP iter.	3553	5059	6931	9700
Binary V.	6171	9812	14639	20842
Integer V.	136443	230973	361498	533942
Continuous V.	22802	37879	58413	85370
Eq. Constr.	25552	42631	65959	96634
Ineq. Constr.	3410	5676	8778	12848

slightly larger number of bikes is obtained in case of Exponential distribution (in total 384 instead of 355). The results can be justified by the fact the multistage model allows for more flexibility having more stages and consequently more recourse decisions: Instead of assigning a large and costly (in terms of procurement cost) number of bikes at the beginning of the service, uncertain demand can be satisfied through renting and redirecting bikes over the day, and finally transshipping them over the night, without incurring in the risk of paying larger stock-out costs. The same considerations apply to the total costs of the multistage model which is 2/3 of the two-stage one for the former distributions, while slightly larger for the Exponential one. However, we have to remember that these are in-sample objective values and

Table 11: Optimal number x^*_{i0} of bikes to place in each bike-station $i \in \mathcal{B}$ at the beginning of the service, as solution of the multistage stochastic model.

x^*_{i0}	Station i										
	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
$\mathcal{U}$	13	0	1	2	2	8	10	2	3	12	7
$\mathcal{E}$	25	2	12	7	15	27	18	20	19	28	22
$\mathcal{N}$	16	0	2	2	4	9	14	3	6	16	9
$\mathcal{L}$	14	0	2	2	3	9	13	2	6	15	7

x^*_{i0}	Station i											
	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total
$\mathcal{U}$	0	1	3	5	9	2	6	9	5	8	4	112
$\mathcal{E}$	0	8	15	19	25	9	25	28	22	20	18	384
$\mathcal{N}$	0	2	5	7	12	3	7	9	5	10	6	147
$\mathcal{L}$	0	2	5	7	11	2	7	10	4	9	6	136

the true costs of the solutions, the out-of-sample objective values over the four considered distributions, are likely to be higher. To see how much, we can solve the multistage stochastic model and the first-stage variables fixed to the solution of the two-stage model. We indicate with *VMS* the *Value of the Multistage Solution*, which represents the expected gain from solving the stochastic multistage problem (MRP) instead of the two-stage one, and it is defined as follows:

$$VMS := E2RP - MRP, \tag{29}$$

where $E2RP$ denotes the solution value of the MRP model, having the first-stage decision variables fixed at the optimal values obtained using the two-stage model. Table 12 shows that the large number of assigned bikes by the two-stage model yields a loss in a multistage setting, of approx 30% in the case of Normal and Log-normal distribution, and of 57.57% for the Uniform distribution, paying for a larger and unnecessary procurement cost. On the other hand, in case of an Exponential distribution, the optimal two-stage allocation performs well in a multistage setting, being able to face the high demand standard deviation, because of the large number of bikes allocated at the beginning of the day.

4.10. Solving the case of different capacity at different stations

In this section we aim at answering to our eighth question: What is the effect of letting bike stations have different capacities? Table 13 shows the number of available docks at the bike-sharing service "LaBiGi" after a recent considerable reduction from a total number of 660 to only 283 available docks.

Table 12: VMS and the corresponding losses, under the assumption of Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ probability distributions of the rental demand.

	Distribution			
	$\mathcal{U}$	$\mathcal{E}$	$\mathcal{N}$	$\mathcal{L}$
VMS	157.28	29.19	108.85	111.54
$\frac{VMS}{RP} \cdot 100$	57.57	2.55	29.46	31.76

Table 13: Capacity k_i of each bike-station $i \in \mathcal{B}$.

	Station i										
	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
k_i	22	10	10	17	19	12	20	8	8	20	10

	Station i											
	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total
k_i	19	8	10	10	10	18	10	8	12	12	10	283

Optimal number of bikes at the beginning of the service, under the four considered probability distributions, in the two-stage and the multistage setting are shown in Tables 14 and 15 respectively.

From Table 14 we observe that the optimal two-stage solutions in the cases of Normal, Uniform and Log-Normal distributions, suggest to saturate the capacities of stations $\mathscr{A}$, $\mathscr{G}$, $\mathscr{J}$, $\mathscr{K}$, $\mathscr{P}$, $\mathscr{S}$, $\mathscr{U}$, while for all other stations almost the same solutions as in Table 2 are obtained (the case with $k_i = 30$). This is not the case for the Exponential distribution where the capacities are saturated only in stations $\mathscr{J}$, $\mathscr{K}$, $\mathscr{P}$, $\mathscr{S}$, $\mathscr{U}$, while for all the other stations a lower number of bikes are allocated compared to the case of equal capacities. The results show that the optimal solutions under the assumption of Uniform, Normal and Log-normal distributions are similar, with a total number ranging from 163 to 168, while 208 with the Exponential distribution. As observed before in the case of constant capacity, this is mainly due to the different shape of the Exponential distribution. The average optimal number of bikes over all the bike-sharing stations is 7 under the assumptions of Uniform, Normal and Log-normal distributions, while it is 9 bikes under the assumption of the Exponential distribution. The optimal stochastic solution suggests a fleet size, to be around 60% of the 283 available slots, for Uniform, Normal and Log-normal distributions while 73% for the Exponential. This avoids too high overflow, stock-out and transshipping costs. Besides, the reduction in the number of available docks from 660 to 283 implies an increase in costs of 10% for Uniform, Normal and Log-normal distributions and 15% for Exponential because of larger overflow, stock-out and transshipping costs.

The optimal multistage solutions (see Table 15) suggest not to saturate any available station and to allocate a lower number of bikes compared to the case of constant capacity (see Table 11). Comparing Tables 14 and 15, we observe that the optimal number of bikes to assign to each bike-station in the multistage setting are approximately $60-80\%$ of what we obtained in the two-stage setting for the case of Uniform, Normal and Log-normal distributions, while approximately 97.5% in the case of the Exponential distribution (in total 203 instead of 208). The results can be again justified by the fact the multistage model allows us for more flexibility having more stages and consequently more recourse decisions: Instead of assigning a large and costly (in terms of procurement cost) number of bikes at the beginning of the service, uncertain demand can be satisfied through renting and redirecting bikes over the day, and finally transshipping them over the night, without incurring in the risk of paying larger stock-out costs. The same considerations apply to the total costs of the multistage model, which is $50-70\%$ of the two-stage one for the former distributions, while higher for the Exponential one.

Table 14: Optimal number of bikes assigned to each bike-station $i \in \mathcal{B}$ for the four different probability distributions obtained solving the two-stage stochastic program.

x_i^*	Station i										
	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
$\mathcal{U}$	22	0	2	3	3	11	20	3	6	20	10
$\mathcal{E}$	14	2	9	5	12	12	12	8	8	20	10
$\mathcal{N}$	22	0	2	3	3	11	20	3	6	20	10
$\mathcal{L}$	22	0	2	3	3	11	20	3	6	20	10

x_i^*	Station i												Total Cost
	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total	
$\mathcal{U}$	0	2	5	7	10	2	9	8	6	12	6	167	535.70
$\mathcal{E}$	0	6	10	10	10	8	10	8	12	12	10	208	1238.762
$\mathcal{N}$	0	2	4	7	10	2	8	8	5	12	5	163	549.56
$\mathcal{L}$	0	2	5	7	10	3	9	8	6	12	6	168	536.34

4.11. A comparison with the implemented system

We now compare how the plan prescribed by the stochastic program performs relative to what was the initial allocation plan in the actual system.

Table 15: Optimal number of bikes assigned to each bike-station $i \in \mathcal{B}$ for the four different probability distributions obtained solving the four-stage stochastic program.

x^*_{i0}	STATION i										
	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
$\mathcal{U}$	13	0	1	2	2	7	11	2	3	13	7
$\mathcal{E}$	20	2	10	7	14	10	14	7	8	20	9
$\mathcal{N}$	16	0	2	2	3	8	14	3	6	16	7
$\mathcal{L}$	14	0	2	2	3	9	13	2	6	15	7

x^*_{i0}	STATION i												Total Cost
	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total	
$\mathcal{U}$	0	1	3	5	8	3	6	7	5	7	4	110	273.625
$\mathcal{E}$	0	7	8	7	9	9	7	8	12	5	10	203	1429.74
$\mathcal{N}$	0	2	5	8	8	3	7	6	5	9	6	136	385.15
$\mathcal{L}$	0	2	5	7	9	2	6	7	4	9	6	130	362.14

We derived the initial allocation of bikes to stations by collecting the number of bikes at each station at 4 a.m., for each day of the considered months. Table 16 shows what we found.

Table 16: Real initial allocation of bikes x_i^{REAL} implemented by the "LaBiGi" bike-sharing service.

	STATION i										
	$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	$\mathscr{D}$	$\mathscr{E}$	$\mathscr{F}$	$\mathscr{G}$	$\mathscr{H}$	$\mathscr{I}$	$\mathscr{J}$	$\mathscr{K}$
x_i^{REAL}	11	8	10	9	8	8	14	6	4	9	8

	STATION i											
	$\mathscr{L}$	$\mathscr{M}$	$\mathscr{N}$	$\mathscr{O}$	$\mathscr{P}$	$\mathscr{Q}$	$\mathscr{R}$	$\mathscr{S}$	$\mathscr{T}$	$\mathscr{U}$	$\mathscr{V}$	Total
x_i^{REAL}	14	8	8	7	7	13	9	6	8	7	8	190

Table 17: Losses under the assumption of Uniform $\mathcal{U}$, Exponential $\mathcal{E}$, Normal $\mathcal{N}$ and Log-normal $\mathcal{L}$ probability distributions of the rental demand, using the initial allocation of bikes $\mathbf{x}^{REAL}$ implemented by the "LaBiGi" bike-sharing service.

Problem	Loss	DISTRIBUTION			
		$\mathcal{U}$	$\mathcal{E}$	$\mathcal{N}$	$\mathcal{L}$
Two-stage	$\frac{\text{Cost}(\mathbf{x}^{REAL})-RP}{RP} \cdot 100$	38.02	10.03	35.76	38.25
Multistage	$\frac{\text{Cost}(\mathbf{x}^{REAL})-RP}{RP} \cdot 100$	58.57	8.01	39.30	40.16

The performance of the allocation plan implemented by "LaBiGi" bike-sharing service has been evaluated by computing its total cost in an out-of-sample fashion substituting it in our two-stage and multistage models (see Table 17 for the four probability distributions). We see that the two-stage stochastic program allocates fewer bikes (15% fewer) in case of Uniform, Normal and Log-Normal distributions compared to the implemented system, yielding a saving of approximately 38% (see Table 17), while a little more bikes (8% more) in case of Exponential with a saving of 10%. This is due to lower overflow, stock-out and transshipping costs of our solution than the "LaBiGi" allocation plan. The multistage stochastic model allocates 72% less bikes in the case of the Uniform distribution with a saving of approximately 58% (see Table 17), while 40% less in case of the Normal and Log-Normal distributions with a saving of 40% relative to the actual "LaBiGi" service. This is again due to the higher flexibility of the multistage model having more recourse decisions over the day avoiding to incur in the risk of paying larger stock-out costs. As in the two-stage case, under Exponential distribution, the optimal multistage stochastic programming solution allocates a little more bikes (6% more) with a saving of 8% to the one in the real system. This can be explained, as observed before, by the different shape of the Exponential distribution having larger standard deviation which consequently allocates a larger number of bikes at the beginning of the day.

It would have been useful to run a simulation on the actual Bergamo system. But that would require access to data that has not been collected and stored, namely actual demand data over shorter time periods, not just added up over a month. Such data could of course be collected in the Bergamo system, given the properties of the app, and it would allow for a much more detailed analysis. Also, ambiguities, such as the type of distributions we are facing, could be removed or reduced. We would recommend that this is done in the future. But even that would suffer from the classical problem that occurs almost whenever demand is an issue; What is measured is sales (satisfied demand) and not actual demand.

It is clear that variation (if not randomness) is higher than what we have used, given that we distributed the monthly demand evenly across days. We also lack daily time patterns and dependencies in the demand, resulting from for example people going into the city center in the morning and back to the suburbs in the afternoon. Such patterns also increase variation (even if not randomness) in actual demand. It is therefore fairly reasonable to say that the real system is more noisy than the one we analyzed.

Even so, there are some conclusions we can draw. One is that we do believe that our multistage model, using better data as just indicated, combined with simulation to verify, would help in Bergamo. The optimization model can pick up patterns that are hard to utilize manually. We can certainly verify that the reduction in the station capacities was wise, and that the capacities now are much closer to what they should be. We have also shown that a deterministic model would definitely not be very useful.

5. Managerial Insights and Conclusions

In this paper we proposed two-stage and multistage stochastic optimization models for a bike-sharing problem and tested it on the bike-sharing system in Bergamo, named "LaBiGi". The problem we studied allows us to re-balance the number of bikes at each bike-station at the end of the service, by transshipping bikes among bike-stations. Doing that, the number of bikes at the beginning of the service on the next day is equal to the optimal number of bikes assigned to the bike-stations at the beginning of the service on the current day. Optimal solutions for the two-stage and multistage models have been provided under different assumptions of probability distributions for the rental demand for each origin-destination pair. An analysis of the optimal solutions obtained by varying the unit cost parameters is performed.

There are some general conclusions that can be drawn from this diverse set of tests on our case from Bergamo. The first, and probably most important, is that it is not safe to solve a deterministic model. A stochastic model will have a larger set of bikes in the system in order to handle variation (more the higher is the variance) and will also tend to use too few stations. The first issue is rather classical (see for example Lium et al. [26] for a case from network design), while the latter will be more problem dependent. Concerning the fleet size, as the total number of bikes to allocate to the stations, in any case it should be much less than the available slots for holding bikes; this to prevent overflow, stock-out and transshipping costs.

We also learn that it might be important to use a multistage model as that provides a more realistic picture of the reuse of bikes during the day. Just looking at a two-stage model (and even more a one-stage deterministic one) loses the dynamics of re-rentals. We also expect that if there are daily patterns in the rental (to school in the morning, home in the afternoon), this is even more important. Analyzing this is for later work.

Future works will take into account correlations and dependencies for rental demand, based on a larger set of real data, as we are still collecting.

Acknowledgments

The authors would like to thank the Municipality of Bergamo, in particular Stefano Zenoni, Councilor for Mobility and Silvio Cavati, for the description of the problem and the historical data provided. F. Maggioni acknowledges *Gruppo Nazionale per il Calcolo Scientifico* (GNCS-INdAM) for the financial support. We are also indebted to the referees who provided many good suggestions.

References

References

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

[2] Bertazzi, L., Cagnolari, M., Maggioni, F., 2018. The value of the right distribution in stochastic programming and its application to a lot sizing problem. Under Evaluation in Computational Management Science .

[3] Birge, J.R., Louveaux, F., 2011. Introduction to stochastic programming. Springer Science & Business Media.

[4] Brinkmann, J., Ulmer, M., Mattfeld, D., 2018. Dynamic lookahead policies for stochastic-dynamic inventory routing in bike sharing systems. Computers & Operations Research .

[5] Çelebi, D., Yörüsün, A., Işık, H., 2018. Bicycle sharing system design with capacity allocations. Transportation Research Part B: Methodological 114, 86–98.

[6] Crainic, T.G., Maggioni, F., Perboli, G., Rei, W., 2018. Reduced cost-based variable fixing in two-stage stochastic programming. Annals of Operations Research URL: `https://doi.org/10.1007/s10479-018-2942-8`, doi:`10.1007/s10479-018-2942-8`.

[7] Cruz, F., Subramanian, A., Bruck, B., Iori, M., 2017. A heuristic algorithm for a single vehicle static bike sharing rebalancing problem. Computers & Operations Research 79, 19–33.

[8] Datner, S., Raviv, T., Tzur, M., Chemla, D., 2017. Setting inventory levels in a bike sharing network. Transportation Science .

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

[10] Dell'Amico, M., Iori, M., Novellani, S., Stützle, T., et al., 2016. A destroy and repair algorithm for the bike sharing rebalancing problem. Computers & Operations Research 71, 149–162.

[11] Dong, L., Rudi, N., 2004. Who benefits from transshipment? exogenous vs. endogenous wholesale prices. Management Science 50, 645–657.

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

[13] Erdoğan, G., Laporte, G., Calvo Wolfler, R., 2014. The static bicycle relocation problem with demand intervals. European Journal of Operational Research 238, 451–457.

[14] Fishman, E., Washington, S., Haworth, N., 2013. Bike share: a synthesis of the literature. Transport reviews 33, 148–165.

[15] Forma, I.A., Raviv, T., Tzur, M., 2015. A 3-step math heuristic for the static repositioning problem in bike-sharing systems. Transportation research part B: methodological 71, 230–247.

[16] Fricker, C., Gast, N., 2016. Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. Euro Journal on Transportation and Logistics 5, 261–291.

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

[18] Ho, S., Szeto, W., 2017. A hybrid large neighborhood search for the static multi-vehicle bike-repositioning problem. Transportation Research Part B: Methodological 95, 340–363.

[19] Ho, S.C., Szeto, W., 2014. Solving a static repositioning problem in bike-sharing systems using iterated tabu search. Transportation Research Part E: Logistics and Transportation Review 69, 180–198.

[20] Kaut, M., Wallace, S.W., 2007. Evaluation of scenario-generation methods for stochastic programming. Pacific Journal of Optimization 3 (2), 257-271 .

[21] King, A.J., Wallace, S.W., 2012. Modeling with stochastic programming. Springer Science & Business Media.

[22] Kloimüllner, C., Raidl, G., 2017. Full-load route planning for balancing bike sharing systems by logic-based benders decomposition. Networks 69, 270–289.

[23] Laporte, G., Meunier, F., Calvo Wolfler, R., 2018. Shared mobility systems: an updated survey. Annals of Operations Research , 1–22.

[24] Legros, B., 2019. Dynamic repositioning strategy in a bike-sharing system; how to prioritize and how to rebalance a bike station. European Journal of Operational Research 272, 740–753.

[25] Li, Y., Szeto, W., Long, J., Shui, C., 2016. A multiple type bike repositioning problem. Transportation Research Part B: Methodological 90, 263–278.

[26] Lium, A.G., Crainic, T.G., Wallace, S.W., 2009. A study of demand stochasticity in stochastic network design. Transportation Science 43, 144–157.

[27] Lu, C.C., 2016. Robust multi-period fleet allocation models for bike-sharing systems. Networks and Spatial Economics 16, 61–82.

[28] Maggioni, F., Wallace, S.W., 2012. Analyzing the quality of the expected value solution in stochastic programming. Annals of Operations Research 200, 37–54.

[29] Papoulis, A., 1965. Probability, random variables, and stochastic processes. McGraw-Hill.

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

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

[32] Regue, R., Recker, W., 2014. Proactive vehicle routing with inferred demand to solve the bikesharing rebalancing problem. Transportation Research Part E: Logistics and Transportation Review 72, 192–209.

[33] Sayarshad, H., Tavassoli, S., Zhao, F., 2012. A multi-periodic optimization formulation for bike planning and bike utilization. Applied Mathematical Modelling 36, 4944–4951.

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

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

[36] Shui, C., Szeto, W., 2018. Dynamic green bike repositioning problem–a hybrid rolling horizon artificial bee colony algorithm approach. Transportation Research Part D: Transport and Environment 60, 119–136.

[37] Szeto, W., Shui, C., 2018. Exact loading and unloading strategies for the static multi-vehicle bike repositioning problem. Transportation Research Part B: Methodological 109, 176–211.

[38] Wang, Y., Szeto, W., 2018. Static green repositioning in bike sharing systems with broken bikes. Transportation Research Part D: Transport and Environment 65, 438–457.

[39] Zhang, D., Yu, C., Desai, J., Lau, H., Srivathsan, S., 2017. A time-space network flow approach to dynamic repositioning in bicycle sharing systems. Transportation research part B: methodological 103, 188–207.

[40] Zhang, J., 2005. Transshipment and its impact on supply chain members performance. Management Science 51, 1534–1539.