# Fast Scenario Reduction by Conditional Scenarios in Two-Stage Stochastic MILP Problems

C. Beltran-Royo*

July 31, 2020

**Abstract**

A common approach to model stochastic programming problems is based on scenarios. An option to manage the difficulty of these problems corresponds to reduce the original set of scenarios. In this paper we study a new fast scenario reduction method based on Conditional Scenarios (CS). We analyze the degree of similarity between the original large set of scenarios and the small set of conditional scenarios in terms of the first two moments. In our numerical experiment, based on the stochastic capacitated facility location problem, we compare two fast scenario reduction methods: the CS method and the Monte Carlo (MC) method. The empirical conclusion is twofold: On the one hand, the achieved expected costs obtained by the two approaches are similar, although the MC method obtains a better approximation to the original set of of scenarios in terms of the moment matching criterion. On the other hand, the CS approach outperforms the MC approach with the same number of scenarios in terms of solution time.

**Keywords:** Stochastic programming, scenario reduction, Monte Carlo sampling, conditional scenario, stochastic capacitated facility location problem.

## 1 Introduction

A common approach to model stochastic programming problems is based on scenarios [8]. One finds this approach in different fields and applications such as optimal design of energy systems [34], portfolio optimization [32], forestry planning [1], influence maximization in social networks [33] and trading in electricity markets [24], etc. Given the high dimension and complexity of stochastic programming problems, it is common to use specialized approaches to solve them. This is even more necessary in presence of integer variables, which make the problem size a critical issue. In this case one can use approaches such as Lagrangian relaxation [31], Benders decomposition [27], decomposition with branch-and-cut [28], parallel computing [23], among others. As an alternative, one can use some heuristic approach in order to obtain good, but suboptimal, solutions. See, for example, [2] for the branch-and-fix coordination heuristic, [21] for a greedy heuristic, etc.

A second option to manage the difficulty of stochastic programming problems corresponds to scenario reduction methods, which aim to trim down the number of scenarios whenever the stochastic programming problem results intractable or requires a very long solution time. The objective of scenario reduction is to balance two conflicting objectives, namely, to obtain an accurate model of the parameter uncertainty and to obtain a tractable problem. Different methods have been proposed for scenario reduction. The simplest approach corresponds to the Monte Carlo (MC) method, which selects, from the initial set of scenarios, a subset of scenarios by Monte Carlo sampling [25, 29]. Another well

*cesar.beltran@urjc.es, Computer Science, Statistics and Operations Research Department, King Juan Carlos University, Madrid, Spain.

known approach is the moment matching method, which selects a limited number of scenarios that satisfy specified statistical properties based on moments [18, 19]. In the scenario reduction method based on a given probability distance, one selects a subset of scenarios which is closest to the initial set of scenarios, in terms of a given probability distance. The choice of the probability distance is problem dependent [26]. Thus, in the case of a stochastic programming problem with continuous decision variables one can use, for example, the Kantorovich distance [11, 25]. However, in the presence of integer variables, one has to use a discrepancy-based probability distance [15].

The MC method is a fast scenario reduction method since it only requires selecting a subset of scenarios by sampling. In contrast, the probability distance method and the moment matching method cannot be considered fast scenario reduction methods, since they require solving an optimization problem to select the reduced set of scenarios. As pointed out in [14], the optimization problem associated to the probability distance method corresponds to a set-covering problem, which is NP-hard. On the other hand, the optimization problem associated to the moment matching method is nonlinear and generally not convex [19]. Even if one uses some heuristic method to approximately solve these difficult optimization problems, the corresponding solution times are much longer than the MC scenario reduction time. In this way, if the original set of scenarios is very large, only fast scenario reduction methods may result practical. This is the case of our numerical experiment (see Section 4) where we have an initial set of $10^5$ scenarios to be reduced (the largest scenario dimension there considered is 75). Thus, in this paper we focus on fast scenario reduction methods. Specifically, we compare the MC method to another fast scenario reduction method which is based on Conditional Scenarios (CS). In the CS method, one approximates the initial set of scenarios by a set of conditional expectations, which are called conditional scenarios. Notice that the CS method is fast since the computation of these conditional expectations is straightforward and can be done quickly (see Section 2).

The CS problem, which is based on conditional scenarios, was introduced in [4] as an effective approximation to the two-stage stochastic mixed-integer linear programming problem with recourse. For short, we will call it the recourse problem (RP). Some useful CS bounds for the optimal value of the RP problem were derived there. The definition of conditional scenario, introduced in [4], was basically suitable for the multivariate normal distribution. In [5] this definition was generalized in order to approximate any multivariate distribution (continuos or discrete). In contrast with the former definition, the new definition allows to approximate a potentially large set of scenarios by a small set of conditional scenarios.

In scenario based optimization one solves the RP problem, whereas in deterministic optimization one solves the so-called expected value problem [7]. The CS problem improves the ability of the expected value problem to deal with uncertainty by considering conditional scenarios instead of the expected scenario. On the other hand, the CS problem reduces the computational burden of the RP problem by considering a reduced number of conditional scenarios instead of a potentially large number of scenarios. In fact, the expected value problem is an approximation to the CS problem, which, in turn, is an approximation to the RP problem. Therefore, the CS solution is, in general, suboptimal for the RP problem but hopefully better than the expected value solution. For this reason, the CS approach should only be used in cases where the scenario approach results impractical regarding the solution time (Figure 1). Notice that in this paper we compare the CS problem to the MC problem (with a reduced number of scenarios) as *approximations* to the RP problem. With the MC method it would possible to solve the RP problem to optimality by taking a number of scenarios large enough. In contrast, as already pointed out, the CS method has only been designed to approximate the RP problem in order to obtain good, possibly suboptimal, solutions with low computational burden.

Depending on the scenario reduction method, we will have the CS problem or the MC problem. Although the CS problem has already been analyzed in [4, 5], the CS random vector, supported by the set of conditional scenarios, has not been analyzed yet. Therefore, the theoretical results of this paper complement the theoretical results of [4] and [5]. The objective of this paper is to study the first two moments of the CS random vector in the context of the RP problem. With this objective in mind, we consider the RP random vector, supported by a potentially large set of scenarios that models the parameter uncertainty of the RP problem, and the CS random vector that has been obtained by the CS

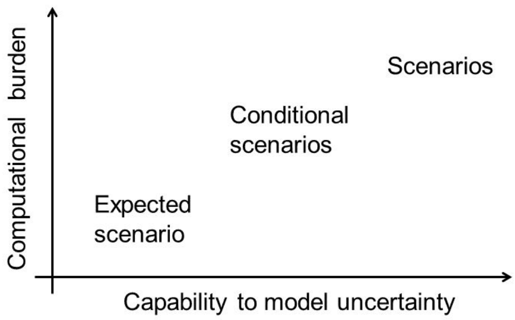


Figure 1: The 'Conditional scenario' approach represents an effective midpoint between the deterministic optimization (based on the 'Expected scenario') and the 'Scenario' based optimization, regarding the capability to model the uncertainty and the computational burden.

method. In this paper, we try to answer the following questions: *a) Which is the degree of similarity between the RP and CS random vector in terms of the first two moments? b) How does the CS problem compare with the MC problem as approximations to the RP problem?*

The contribution of this paper is the following:

- *Theoretical point of view:* Roughly speaking, we show that the RP and CS random vectors have the same expectation. We also show that they have covariance coefficients which are proportional, provided that the RP random vector is a discretization of a multivariate normal random vector (see Section 3). In this case, the CS method bears some similarity to the approaches preserving the first and second order moment information of the distribution, such as robust optimization [12] and stochastic optimization based on linear decision rules combined with semi-infinite programming [3].
- *Practical point of view:* In the numerical experiment, based on the stochastic capacitated facility location problem, we compare the MC and CS problems. The empirical conclusion is twofold: On the one hand, although the MC method obtains a better approximation to the original set of scenarios in terms of the moment matching criterion, the achieved expected costs obtained by the CS and MC approaches are similar. On the other hand, the CS problem outperforms the MC problem with the same number of scenarios in terms of solution time. One of the most interesting observations of this paper is the following conjecture: *in the context of MILP problems, the CS problem usually has a smaller LP gap than the corresponding MC problem with the same number of scenarios,* which would explain the faster performance of the CS approach (see Section 4).

Apart from the already mentioned main Sections 3 and 4, in Section 2 we review the MC and CS scenario reduction methods and in Section 5 the conclusions are drawn.

## 2 Fast scenario reduction for the recourse problem

In this paper we focus on fast scenario reduction methods for the two-stage stochastic mixed-integer linear programming problem with recourse (for short, the Recourse Problem (RP)). A thorough description of this problem, its applications and solution methods can be found in [7, 20, 25, 29], among others. The RP problem can be stated as follows:

$$
\begin{aligned}
\min_{x} \quad & z_{RP} = c_1 x_1 + \sum_{s \in \mathcal{S}_1} \tilde{p}^s \, \tilde{c}_2^s x_2^s && && (1)\\
\text{s.t.} \quad & A_1 x_1 = b_1 && && (2)\\
& \tilde{A}_2^s x_1 + \tilde{B}_2^s x_2^s = \tilde{b}_2^s && s \in \mathcal{S}_1 && (3)\\
& x_1 \geq 0 && && (4)\\
& x_2^s \geq 0 && s \in \mathcal{S}_1 && (5)\\
& x_{1j} \text{ integer} && j \in \mathcal{J}_1 && (6)\\
& x_{2j}^s \text{ integer} && s \in \mathcal{S}_1,\ j \in \mathcal{J}_2, && (7)
\end{aligned}
$$

where $\mathcal{J}_t$ is the index set for the integer variables at stage $t$ for all $t \in \mathcal{T} = \{1, 2\}$. In this context, scenario $\tilde{\xi}^s$ represents the $s$th realization of the random parameters of the problem, that is, $\tilde{\xi}^s = \text{vec}\big(\tilde{c}_2^s, \tilde{A}_2^s, \tilde{B}_2^s, \tilde{b}_2^s\big)$, where 'vec' is the operator that stacks vectors and matrix columns into a single vector. Then, $\tilde{\boldsymbol{\xi}}$ is the RP random vector, which is defined by the set of scenarios $\{\tilde{\xi}^s\}_{s \in \mathcal{S}_1}$ and the corresponding probability values $\{\tilde{p}^s\}_{s \in \mathcal{S}_1}$, such that $P\big(\tilde{\boldsymbol{\xi}} = \tilde{\xi}^s\big) = \tilde{p}^s$ for all $s \in \mathcal{S}_1 = \{1, \ldots, S_1\}$.

Given that the computational burden of this problem rapidly increases with the number of scenarios, a common approach is to reduce the initial set of scenarios to obtain a small set of representative scenarios. In this way, the new problem, formulated in terms of the reduced set of scenarios, results computationally tractable. As already pointed out in the introduction, in this paper we will focus on two fast scenario reduction methods, namely, the Monte Carlo method and the conditional scenario method.

## 2.1 Monte Carlo method

Scenario sampling by the Monte Carlo method is the simplest scenario reduction method. Given a potentially large set of scenarios $\{\tilde{\xi}^s\}_{s \in \mathcal{S}_1}$, the MC method randomly selects a subset of scenarios $\{\hat{\xi}^s\}_{s \in \mathcal{S}_2}$ of reduced cardinality indexed by $\mathcal{S}_2 = \{1, \ldots, S_2\}$. The MC scenario reduction is a fast method since it does not rely on any optimization problem [17]. The theoretical properties of the MC method can be found in [25] and the references therein. According to [16], presently the MC sampling method is the preferred approach for scenario generation and reduction in stochastic programming. A closely related stochastic programming approach is the Sample Average Approximation (SAA) method, where one formulates the RP problem in terms of a random sample of scenarios generated by Monte Carlo sampling techniques [29].

## 2.2 Conditional scenario method

In stochastic programming, scenarios and the expected scenario represent two extreme choices regarding computational burden and ability to model the parameter uncertainty. The conditional scenario (CS) concept was introduced in [4] as an effective midpoint between these two choices, since it showed a moderate computational burden and a reasonable ability to model the parameter uncertainty. Given a set of scenarios, computing the corresponding conditional scenarios can be seen as a scenario reduction method. In what follows we review the CS scenario reduction method.

**Method 1.** *(CS scenario reduction method [5])*

- *Objective:* To approximate a set of scenarios by a set of conditional scenarios.
- *Input:*
    a) A set of equiprobable scenarios $\{\tilde{\xi}^s\}_{s \in \mathcal{S}_1}$, such that $\tilde{\xi}^s = \big(\tilde{\xi}_1^s \ldots \tilde{\xi}_R^s\big)$, with probability value $\tilde{p}^s = 1/S_1$, for all $s \in \mathcal{S}_1$.

b) $E$, the number of conditional scenarios per each coordinate $r \in \mathcal{R} = \{1, \dots, R\}$. Then, $\mathcal{E} = \{1, \dots, E\}$ is the index set for the conditional scenarios for a given coordinate.

- *Output:* A set of conditional scenarios $\{\hat{\xi}^{re}\}_{re \in \mathcal{R}\times\mathcal{E}}$ and the corresponding probability values $\{\hat{p}^{re}\}_{re \in \mathcal{R}\times\mathcal{E}}$.
- *Steps:* For each $r \in \mathcal{R}$ :

1) Define the interval $\mathcal{I}_r = [a_r, b_r]$ such that

$$a_r = \min_{s\in\mathcal{S}_1}\{\tilde{\xi}_r^s\} \qquad\qquad b_r = \max_{s\in\mathcal{S}_1}\{\tilde{\xi}_r^s\}.$$

2) Partition $\mathcal{I}_r$ is into $E$ subintervals $\mathcal{I}_{re}$ such that $\mathcal{I}_r = \cup_{e\in\mathcal{E}}\mathcal{I}_{re}$, where $\mathcal{I}_{re} = [a_{re}, b_{re})$ for all $e \in \mathcal{E} \setminus \{E\}$ and $\mathcal{I}_{rE} = [a_{rE}, b_{rE}]$.
3) For all $e \in \mathcal{E}$ :

i) Classify the scenarios such that the index set $\mathcal{S}_{re}$ accounts for the scenarios that fulfill the condition $\tilde{\xi}_r^s \in \mathcal{I}_{re}$.
ii) Set $S_{re}$ as the cardinality of $\mathcal{S}_{re}$.
iii) Compute the corresponding conditional scenario and its probability value:

$$\hat{\xi}^{re} = \mathbb{E}[\,\tilde{\boldsymbol{\xi}} \mid \tilde{\boldsymbol{\xi}}_r \in \mathcal{I}_{re}\,] = \frac{1}{S_{re}} \sum_{s\in\mathcal{S}_{re}} \tilde{\xi}^s \tag{8}$$

$$\hat{p}^{re} = \frac{1}{R}\left(S_{re}/S_1\right).$$

As we can see, the CS scenario reduction is straightforward and does not rely on any optimization problem. In this method the input is the RP random vector $\tilde{\boldsymbol{\xi}}$, which has support $\{\tilde{\xi}^s\}_{s\in\mathcal{S}_1}$ and probability values $\{\tilde{p}^s\}_{s\in\mathcal{S}_1}$, and the output is the CS random vector $\hat{\boldsymbol{\xi}}^{\mathbf{r}}$, with support $\{\hat{\xi}^{re}\}_{re\in\mathcal{R}\times\mathcal{E}}$ and probability values $\{\hat{p}^{re}\}_{re\in\mathcal{R}\times\mathcal{E}}$. Notice that in the CS method, one computes each conditional scenario as a convex combination of a selected set of original scenarios (Equation (8)), in contract with the MC method where the scenarios are randomly selected from the original set of scenarios. Thus, in the MC method the structure of the original random process remains unchanged, while the CS method modifies this structure. In this respect, the CS method could be considered more a solution algorithm than a scenario reduction technique. However, in this paper we will use the term *CS scenario reduction method* to indicate that the initial large set of scenarios is approximated by a small set of conditional scenarios, that is, one reduces the original number of scenarios.

# 3 Moment properties of the randomized conditional expectation

According to [25], in very rare cases one can solve a stochastic programming problem formulated in terms of $\boldsymbol{\xi}$, a random vector with a general probability distribution that models the parameter uncertainty. For a numerical solution, the original problem is replaced by a simpler one, where $\boldsymbol{\xi}$ is replaced by a simpler random vector $\tilde{\boldsymbol{\xi}}$, which has a finite support given by a set of scenarios (we call it the RP random vector). For quite large sets of scenarios one may ignore the fact that the model is just an approximation since it produces accurate optimal solutions. However, in this case the approximate problem may become numerically intractable, and some scenario reduction method is needed to balance both numerical tractability and accuracy of the uncertainty model.

In this section we address the following question formulated in the introduction: *a) Which is the degree of similarity between the RP random vector $\tilde{\boldsymbol{\xi}}$ and the CS random vector $\hat{\boldsymbol{\xi}}^{\mathbf{r}}$ in terms of the first two moments?* With this objective in mind, we consider $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{\mathbf{r}}$, the Randomized Conditional Expectation (RCE) that approximates $\boldsymbol{\xi}$ by its conditional expectation [4]. At this point, it may be useful to briefly review the RCE concept. Given a random vector $\boldsymbol{\xi} = \left(\boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_R\right)$, the conditional expectation $\boldsymbol{\xi}^r := \mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]$, seen as a function of $\boldsymbol{\xi}_r$, is an optimal approximation to $\boldsymbol{\xi}$, for all $r \in \mathcal{R}$ [13]. Since it would not have sense to consider all these approximations simultaneously, they are considered one each time (randomly), that is, each $\boldsymbol{\xi}^r$ approximates $\boldsymbol{\xi}$ with probability $1/R$. For this reason, it is considered a random index $\mathbf{r}$ with support $\mathcal{R}$ and uniform probability $P\big(\,\mathbf{r} = r\,\big) = 1/R$ for all $r \in \mathcal{R}$ to define the RCE random vector $\boldsymbol{\xi}^{\mathbf{r}}$ (notice the $\mathbf{r}$, being a random variable, is written in boldface). Therefore, $\boldsymbol{\xi}^{\mathbf{r}}$ approximates the random vector $\boldsymbol{\xi}$ by a set of $R$ random vectors of dimension $R$, that is, $\{\boldsymbol{\xi}^r\}_{r\in\mathcal{R}}$, each one taken with probability $1/R$ (further details can be found in [4, 5]).

As already mentioned, to reach tractability, $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{\mathbf{r}}$ are usually approximated by the finite support random vectors $\tilde{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^{\mathbf{r}}$, respectively. Thus, the degree of similarity between $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{\mathbf{r}}$ can be used as a proxy of the degree of similarity between $\tilde{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^{\mathbf{r}}$, provided that $\tilde{\boldsymbol{\xi}}$ and $\hat{\boldsymbol{\xi}}^{\mathbf{r}}$ are sufficiently close to $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{\mathbf{r}}$, respectively. The previous idea is summarized in the following graph:

$$
\begin{array}{ccc}
\boldsymbol{\xi} & \rightarrow \ \text{conditional expectation} \ \rightarrow & \boldsymbol{\xi}^{\mathbf{r}} \\
\downarrow & & \downarrow \\
\text{finite support} & & \text{finite support} \\
\text{approximation} & & \text{approximation} \\
\downarrow & & \downarrow \\
\tilde{\boldsymbol{\xi}} & \rightarrow \ \text{conditional expectation} \ \rightarrow & \hat{\boldsymbol{\xi}}^{\mathbf{r}}
\end{array}
$$

Therefore, we address question a) in an indirect way by analyzing the similarity between $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^{\mathbf{r}}$ in terms of the first two moments. Notice that the scenario reduction by the CS method corresponds to the bottom of the previous graph.

### 3.1 Expectation of the randomized conditional expectation

In this section we show that the RCE random vector and the original random vector $\boldsymbol{\xi}$, have the same expectation, that is, $\mathbb{E}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \mathbb{E}[\,\boldsymbol{\xi}\,]$. First, let us consider the following result, which specializes the *law of total expectation* ([9], Theorem 4.7.1) in the conditional scenario context and can be proved in a similar way.

**Proposition 1.** *Let us consider the random vector $\boldsymbol{\xi} = \big(\,\boldsymbol{\xi}_1 \ldots \boldsymbol{\xi}_R\,\big)$, such that $\boldsymbol{\xi}$ has finite expectation, then*

$$\mathbb{E}[\,\mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,] = \mathbb{E}[\,\boldsymbol{\xi}\,] \qquad \forall r \in \mathcal{R}.$$

In order to approximate the original stochastic programming problem, given the original random vector $\boldsymbol{\xi}$, the so-called expected value problem approximates it by its expectation, that is, $\bar{\xi} = \mathbb{E}[\,\boldsymbol{\xi}\,]$. If we consider $\bar{\xi}$ as a degenerate random vector then we have that $\mathbb{E}[\,\bar{\xi}\,] = \mathbb{E}[\,\boldsymbol{\xi}\,]$ which is a desirable property. By the previous proposition, this result is also true for the conditional expectation $\boldsymbol{\xi}^r$ since $\mathbb{E}[\,\boldsymbol{\xi}^r\,] = \mathbb{E}[\,\mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,] = \mathbb{E}[\,\boldsymbol{\xi}\,]$ for all $r \in \mathcal{R}$. Next proposition proves that this result is also true for the RCE random vector.

**Proposition 2.** *Let us consider the random vector $\boldsymbol{\xi}$ such that it has finite expectation (we do not assume any probability distribution) and let us also consider the randomized conditional expectation $\boldsymbol{\xi}^{\mathbf{r}} = \mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]$, where $\mathbf{r}$ is the discrete random variable with uniform distribution on $\mathcal{R} = \{1, \ldots, R\}$, such that, $P\big(\,\mathbf{r} = r\,\big) = 1/R$ for all $r \in \mathcal{R}$. Then*

$$\mathbb{E}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \mathbb{E}[\,\boldsymbol{\xi}\,].$$

*Proof.*

$$
\begin{aligned}
\mathbb{E}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \mathbb{E}[\,\mathbb{E}[\,\boldsymbol{\xi}^{\mathbf{r}} \mid \mathbf{r}\,]\,] &= \sum_{r\in\mathcal{R}} \frac{1}{R}\mathbb{E}[\,\boldsymbol{\xi}^{\mathbf{r}} \mid \mathbf{r} = r\,] \\
&= \sum_{r\in\mathcal{R}} \frac{1}{R}\mathbb{E}[\,\boldsymbol{\xi}^{r}\,] = \sum_{r\in\mathcal{R}} \frac{1}{R}\mathbb{E}[\,\boldsymbol{\xi}\,] = \mathbb{E}[\,\boldsymbol{\xi}\,].
\end{aligned}
$$

■

### 3.2 Covariance of the randomized conditional expectation

As is well known, the expected scenario $\bar{\xi} = \mathbb{E}[\,\boldsymbol{\xi}\,]$ ignores the variability of the original vector $\boldsymbol{\xi}$, and therefore it has a null covariance matrix, that is, $\mathbb{V}\mathrm{ar}[\,\bar{\xi}\,] = 0_{R\times R}$. In contrast, the RCE random vector takes into account the variability of the original vector and $\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,]$ shows some degree of 'similarity' with $\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}\,]$, as we will see in this section. This similarity is studied for general multivariate distributions in Proposition 4 and for the multivariate normal distribution in Propositions 5 and 6.

Let us consider the following result, which specializes the *law of total variance* ([9], Theorem 4.7.4) in the conditional scenario context and can be proved in a similar way.

**Proposition 3.** *Let us consider the arbitrary random vector $\boldsymbol{\xi} = \big(\boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_R\big)$ for which the necessary expectations and covariance matrices in equation (9) exist, then*

$$\mathbb{V}ar[\,\boldsymbol{\xi}\,] = \mathbb{E}[\,\mathbb{V}ar[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,] + \mathbb{V}ar[\,\mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,] \qquad \forall r \in \mathcal{R}. \tag{9}$$

**Proposition 4.** *Let us consider the random vector $\boldsymbol{\xi} = \big(\boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_R\big)$ for which the necessary expectations and covariance matrices in equation (10) exist (we do not assume any probability distribution) and let us also consider the randomized conditional expectation $\boldsymbol{\xi}^{\mathbf{r}} = \mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]$, where $\mathbf{r}$ is the discrete random variable with uniform distribution on $\mathcal{R}$. Then:*

$$\mathbb{V}ar[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \mathbb{V}ar[\,\boldsymbol{\xi}\,] - \frac{1}{R}\sum_{r\in\mathcal{R}} \mathbb{E}[\,\mathbb{V}ar[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,]. \tag{10}$$

*Proof.* By the definition of $\boldsymbol{\xi}^{\mathbf{r}}$, one has that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] &= \mathbb{V}\mathrm{ar}[\,\mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]\,] \\ &= \mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}\,] - \mathbb{E}[\,\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]\,],\end{aligned} \tag{11}$$

where the last equality is a consequence of the law of total variance (9). On the other hand by the law of total expectation

$$\begin{aligned}\mathbb{E}[\,\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]\,] &= \mathbb{E}[\,\mathbb{E}[\,\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]\,] \mid \mathbf{r}\,] \\ &= \sum_{r\in\mathcal{R}} \frac{1}{R}\mathbb{E}[\,\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,].\end{aligned}$$

Therefore Equation (11) can be written as

$$\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}\,] - \sum_{r\in\mathcal{R}} \frac{1}{R}\mathbb{E}[\,\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,].$$

■

The result in Proposition 4 applies for any distribution. In Propositions 5 and 6, we specialize this result for the multivariate normal distribution. As is well known, the multivariate normal distribution plays a central role in statistics because it can be viewed as an approximation and limit of many other distributions (the basis justification relies on the central limit theorem) . As a consequence, the multivariate normal distribution comes into play in many applications [13].

**Proposition 5.** *Let us consider the multivariate normal random vector $\boldsymbol{\xi} = \big(\boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_R\big) \sim N_R(\mu, \Sigma)$ and the randomized conditional expectation $\boldsymbol{\xi}^{\mathbf{r}} = \mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]$, such that $\mathbf{r}$ is the discrete random variable with uniform distribution on $\mathcal{R}$. Then*

$$\mathbb{V}ar[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \frac{1}{R}\sum_{r\in\mathcal{R}} \frac{1}{\sigma_r^2}\,\Sigma_{*r}\Sigma_{r*},$$

*where $\Sigma_{*r}$ and $\Sigma_{r*}$ are the rth column and the rth row of the covariance matrix $\Sigma$, respectively.*

*Proof.* It is straightforward to see that the conditional covariance matrix of $\boldsymbol{\xi}$ given $\boldsymbol{\xi}_r$ can be computed as follows:

$$\mathbb{V}\text{ar}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,] = \Sigma - \frac{1}{\sigma_r^2}\,\Sigma_{*r}\Sigma_{r*} \qquad \text{(by Theorem 5.3 in [13]).}$$

Therefore:

$$\begin{aligned}
\mathbb{V}\text{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] &= \mathbb{V}\text{ar}[\,\boldsymbol{\xi}\,] - \frac{1}{R}\sum_{r\in\mathcal{R}} \mathbb{E}[\,\mathbb{V}\text{ar}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_r\,]\,] \qquad \text{(by Proposition 4)}\\
&= \Sigma - \frac{1}{R}\sum_{r\in\mathcal{R}} \mathbb{E}\Big[\Sigma - \frac{1}{\sigma_r^2}\,\Sigma_{*r}\Sigma_{r*}\Big]\\
&= \Sigma - \frac{1}{R}\sum_{r\in\mathcal{R}} \Sigma + \frac{1}{R}\sum_{r\in\mathcal{R}} \frac{1}{\sigma_r^2}\,\Sigma_{*r}\Sigma_{r*}\\
&= \frac{1}{R}\sum_{r\in\mathcal{R}} \frac{1}{\sigma_r^2}\,\Sigma_{*r}\Sigma_{r*}.
\end{aligned}$$

■

**Proposition 6.** *Let us consider the multivariate normal random vector $\boldsymbol{\xi} \sim N_R(\mu, \Sigma)$, such that $\Sigma = (\sigma_{ij})$, and the randomized conditional expectation $\boldsymbol{\xi}^{\mathbf{r}} = \mathbb{E}[\,\boldsymbol{\xi} \mid \boldsymbol{\xi}_{\mathbf{r}}\,]$. Furthermore, let us also consider $\rho_{ij}$, the correlation between the components $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$, for all $i, j \in \mathcal{R}$ :*

1. *If the components of $\boldsymbol{\xi}$ are pairwise uncorrelated ($\rho_{ij} = 0$ for all $i, j \in \mathcal{R}$ such that $i \neq j$) then*

$$\mathbb{V}ar[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \frac{1}{R}\mathbb{V}ar[\,\boldsymbol{\xi}\,].$$

   *Notice that $\mathbb{V}ar[\,\boldsymbol{\xi}\,]$ is a diagonal matrix (therefore, $\mathbb{V}ar[\,\boldsymbol{\xi}^{\mathbf{r}}\,]$ is also diagonal).*

2. *If the components of $\boldsymbol{\xi}$ are pairwise correlated ($\rho_{ij} \neq 0$ for all $i, j \in \mathcal{R}$) then*

$$\mathbb{V}ar[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \Big(\alpha_{ij}\sigma_{ij}\Big)_{ij\in\mathcal{R}\times\mathcal{R}},$$

   *where*

$$\alpha_{ij} = \frac{1}{R}\sum_{r\in\mathcal{R}} \frac{\rho_{ir}\rho_{rj}}{\rho_{ij}}.$$

*Proof.* 1. By Proposition 5 we have

$$\begin{aligned}
\mathbb{V}\text{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] &= \frac{1}{R}\sum_{r\in\mathcal{R}} \frac{1}{\sigma_r^2}\,\Sigma_{*r}\Sigma_{r*}\\
&= \frac{1}{R}\sum_{r\in\mathcal{R}} \frac{1}{\sigma_r^2}\begin{pmatrix}\sigma_{1r}\\ \vdots \\ \sigma_{Rr}\end{pmatrix}\Big(\sigma_{r1}\ldots\sigma_{rR}\Big)\\
&= \frac{1}{R}\sum_{r\in\mathcal{R}} \Big(\frac{\sigma_{ir}\sigma_{rj}}{\sigma_r^2}\Big)_{ij\in\mathcal{R}\times\mathcal{R}}\\
&= \frac{1}{R}\sum_{r\in\mathcal{R}} \Big(\frac{\rho_{ir}\sigma_i\sigma_r\rho_{rj}\sigma_r\sigma_j}{\sigma_r^2}\Big)_{ij\in\mathcal{R}\times\mathcal{R}}\\
&= \frac{1}{R}\Big(\sum_{r\in\mathcal{R}} \rho_{ir}\rho_{rj}\sigma_i\sigma_j\Big)_{ij\in\mathcal{R}\times\mathcal{R}}. \qquad (12)
\end{aligned}$$

By hypothesis we have that $\rho_{ir} = 0$ if $i \neq r$ and, as usual, $\rho_{ir} = 1$ if $i = r$, for all $i, r \in \mathcal{R}$. In this case equation (12) becomes

$$\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] = \frac{1}{R}\Big(\,\delta_{ij}\sigma_i\sigma_j\,\Big)_{ij\in\mathcal{R}\times\mathcal{R}} = \frac{1}{R}\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}\,],$$

where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. We observe that $\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,]$ is a diagonal matrix.

2. To prove the second statement we use equation (12):

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\,\boldsymbol{\xi}^{\mathbf{r}}\,] &= \frac{1}{R}\Big(\,\sum_{r\in\mathcal{R}}\rho_{ir}\rho_{rj}\sigma_i\sigma_j\,\Big)_{ij\in\mathcal{R}\times\mathcal{R}}\\ &= \frac{1}{R}\Big(\,\sum_{r\in\mathcal{R}}\rho_{ir}\rho_{rj}\,\frac{\sigma_{ij}}{\rho_{ij}}\,\Big)_{ij\in\mathcal{R}\times\mathcal{R}}\\ &= \Big(\,\alpha_{ij}\sigma_{ij}\,\Big)_{ij\in\mathcal{R}\times\mathcal{R}},\end{aligned}$$

where

$$\alpha_{ij} = \frac{1}{R}\sum_{r\in\mathcal{R}}\frac{\rho_{ir}\rho_{rj}}{\rho_{ij}} \qquad\qquad ij \in \mathcal{R}\times\mathcal{R}.$$

■

# 4 Numerical experiment

In this section we address the second question formulated in the introduction: *b) How does the CS problem, based on conditional scenarios, compare with the MC problem, based on scenarios, as approximations to the RP problem?* To answer this question we use the Capacitated Facility Location (CFL) problem with uncertain demand [6, 30].

In the CFL problem we have a set of candidate facilities indexed by $\mathcal{I} = \{1, \ldots, I\}$ and a set of clients indexed by $\mathcal{J} = \{1, \ldots, J\}$. In the stochastic CFL problem here solved future demand is modelled by a set of equiprobable scenarios $\{\tilde{d}^s\}_{s\in\mathcal{S}_1}$. The goal is to choose a subset of facilities in order to satisfy the demand of all the clients at the minimum expected cost, which accounts for the fixed costs $f_i$ and for the supplying costs $c_{ij}$ (see Table 1). Since demands are unknown and facilities have a limited capacity $g_i$, the facilities opened in the first stage may be insufficient to satisfy all of the demands in the second stage. In this case, a penalty term for unmet demand, proportional to $q_j$, is added to the objective function of the deterministic CFL formulation [10]

Notice that solving the CFL problem with recourse, formulated in terms of a set of scenarios generated by the MC method, can be considered as a simple version of the Sample Average Approximation (SAA) method [22]. For this reason we could name it the SAA capacitated facility location problem. However, to maintain the acronyms used in previous sections, we will call it the MC capacitated facility location problem or, for short, the MC problem. The MC problem is a mixed-integer linear programming problem which has been solved by CPLEX 12.6 with default parameter values. Computations have been conducted on a PC under Windows 7 (64 bits), with an Intel Core i5 processor, 2.67GHz and 8 GB of RAM.

## 4.1 The RP capacitated facility location problem

The RP capacitated facility location problem, for short, the RP problem, corresponds to solve the stochastic CFL problem with recourse formulated in terms of a set of scenarios. The deterministic parameters, random parameters and decision variables of the RP problem can be found in Tables 1, 2

Table 1: Indexes and parameters of the RP problem.

| Parameter | Value | Unit | Description |
|---|---|---|---|
| $I$ | 10 | - | Number of (candidate) facilities |
| $\mathcal{I}$ | $\{1, \ldots, I\}$ | - | Index set for facilities |
| $i$ | - | - | Index for facilities |
| $J$ | 30 | - | Number of clients |
| $\mathcal{J}$ | $\{1, \ldots, J\}$ | - | Index set for clients |
| $j$ | - | - | Index for clients |
| $\mathcal{IJ}$ | $\mathcal{I} \times \mathcal{J}$ | - | Index set for pairs $ij$ |
| $f_i$ | $1000 + 300i$ | euros | Fixed cost of facility $i$ |
| $c_{ij}$ | $0.3IJ - 0.01\big(I(j-1) + i\big)$ | euros / unit | Cost of supplying client $j$ by facility $i$ |
| $q_j$ | $3\sum_{i \in \mathcal{I}} c_{ij}/I$ | euros / unit | Penalty for unmet demand |
| $\bar{d}_j$ | $100 + 10j$ | units | Expected demand of client $j$ |
| $K$ | $\sum_{j \in \mathcal{J}} \bar{d}_j/I$ | units | Ratio: |
| | | | 'Total expected demand / Number of facilities' |
| $g_i$ | $4\big(0.5K + (K/I)i\big)$ | units | Capacity of facility $i$ |

Table 2: Random demand of the RP problem.

| Parameter | Value | Unit | Description |
|---|---|---|---|
| $\tilde{\mathbf{d}}$ | $\big(\tilde{\mathbf{d}}_1 \ldots \tilde{\mathbf{d}}_J\big)$ | units | Random demands modeled by a random sample |
| | | | of $S_1 = 10^5$ equiprobable scenarios $\{\tilde{d}^s\}_{s \in \mathcal{S}_1}$ drawn |
| | | | from the multivariate normal distribution $\mathbf{d} \sim N_J(\mu, \Sigma)$ |
| $\mu_j$ | $\bar{d}_j$ | units | Expected demand of client $j$ |
| $\sigma_j$ | $0.2\,\bar{d}_j$ | units | Standard deviation of $\mathbf{d}_j$ |
| $\rho_{j_1,j_2}$ | 0.7 | - | Correlation between $\mathbf{d}_{j_1}$ and $\mathbf{d}_{j_2}, \quad j_1 \neq j_2$ |
| $\sigma_{j_1,j_2}$ | $\rho_{j_1,j_2}\,\sigma_{j_1}\sigma_{j_2}$ | units$^2$ | Covariance between $\mathbf{d}_{j_1}$ and $\mathbf{d}_{j_2}, \quad j_1 \neq j_2$ |
| $\tilde{d}^s$ | - | units | Realization of $\tilde{\mathbf{d}}$ (scenario) |
| $R$ | 30 | - | Number of random parameters |
| $E_r$ | 8 | - | Number of conditional scenarios |
| | | | corresponding to the $r$th component of $\tilde{\mathbf{d}}$ |
| $S_1$ | $10^5$ | - | Initial number of scenarios |
| $\mathcal{S}_1$ | $\{1, \ldots, S_1\}$ | - | Index set for the initial set of scenarios |
| $s$ | - | - | Index for scenarios |

Table 3: Decision variables of the RP problem.

| Decision | Unit | Description |
|---|---|---|
| $u_i$ | - | $u_i = 1$, if facility $i$ is opened |
| | | $u_i = 0$, otherwise |
| $x_{ij}^s$ | % | Fraction of the demand of client $j$ supplied by facility $i$ |
| $y_j^s$ | % | Fraction of the demand of client $j$ unmet |

and 3, respectively. Then, the RP problem can be written as follows:

$$\min_{u,x,y} \quad z_{RP} = \sum_{i\in\mathcal{I}} f_i\, u_i + \frac{1}{S_1} \sum_{s\in\mathcal{S}_1} \Big( \sum_{ij\in\mathcal{IJ}} c_{ij} \tilde{d}_j^s\, x_{ij}^s + \sum_{j\in\mathcal{J}} q_j \tilde{d}_j^s\, y_j^s \Big) \tag{13}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{I}} x_{ij}^s + y_j^s = 1 \qquad s\in\mathcal{S}_1,\ j\in\mathcal{J} \tag{14}$$

$$\sum_{j\in\mathcal{J}} \tilde{d}_j^s\, x_{ij}^s \le g_i\, u_i \qquad s\in\mathcal{S}_1,\ i\in\mathcal{I} \tag{15}$$

$$u_i \in \{1, 0\} \qquad i\in\mathcal{I} \tag{16}$$

$$x_{ij}^s \ge 0,\ y_j^s \ge 0 \qquad s\in\mathcal{S}_1,\ ij\in\mathcal{IJ}. \tag{17}$$

This problem is a well known example of the two-stage stochastic MILP problem introduced in (1)–(7). Notice that we have not included the constraints $x_{ij}^s \le u_i$ for all $s \in \mathcal{S}_1, ij \in \mathcal{IJ}$. These constrains are normally used in facility location problems to strengthen the MILP formulation, however in a preliminary computational test we have observed that they considerably increase the problem dimension resulting in longer solution times compared with the current formulation (13)–(17).

## 4.2 Fast scenario reduction

The above RP problem, formulated in terms of the initial set of $S_1 = 10^5$ scenarios that model the random demand, is computationally intractable. In order to attain tractability we will approximate the initial set of scenarios by three methods: the Expected Value (EV) method, the MC method and the CS method. After these scenario reduction processes, one can formulate the corresponding EV, MC and CS problems, respectively, which have the same structure as the RP problem (13)-(17) but with a reduced size. The EV method approximates the initial set of scenarios by the expected scenario. In the CS method (see Method 1) we set $E_r = 8$ for all $r \in \mathcal{R}$ which, combined with $R = 30$ random parameters, gives $R{\cdot}E = 240$ conditional scenarios. We arbitrarily choose $E_r = 8$ for two reasons. On the one hand, to keep the computational burden of the corresponding MILP problem low. On the other hand, since we know that the scenarios, componentwise, come from the normal distribution which has probability almost one along the interval of length $8\sigma$ centered at $\mu$, we consider that one discretization point per each subinterval of length $\sigma$ is reasonable. Although this is a heuristic approach, taking $E_r = 8$ represents a good balance between the computational burden and the quality of the approximated solution, as we will see in Section 4.4. In the MC method we consider $S_2 = 240$ scenarios randomly sampled from the initial set of scenarios, in order to match the number of conditional scenarios.

In Table 4 we summarize the results of the three scenario reductions. In this table, $\tilde{\mathbf{d}}$ is the random vector associated to the inial set of $10^5$ scenarios. $\hat{\mathbf{d}}$ is the random vector associated to the scenario reduction method: EV, MC or CS. $\|\|_2$ and $\|\|_F$ are the Euclidean and the Frobenius norms, respectively. We have obtained the following results. *Scenario reduction time:* we observe that the three methods require less than 2 seconds. *Relative expectation error:* the three methods approximate well the original expectation vector, since all of them have a relative expectation error under or equal to $1\%$. *Relative covariance error:* obviously, the relative covariance error for the EV method is the maximum (100%) since in this case $\mathbb{V}\text{ar}[\,\hat{\mathbf{d}}\,] = 0$ ($\hat{\mathbf{d}} = \bar{d}$ is considered a degenerate random vector). For the MC and CS methods this error is 7% and 37%, respectively. Otherwise said, the quality of the approximation to the original covariance matrix is 0%, 93% and 63% by the EV, MC and CS methods, respectively (measured by the Frobenius norm). Notice that, although moment similarity is a desired property, it does not guarantee the similarity of two probability distributions, as pointed out in [25].

Table 4: Fast scenario reduction.

| | EV | MC | CS |
|---|---|---|---|
| Scenario reduction time (s) | <1 | <1 | <2 |
| Relative expectation error (%) | 0 | 1 | 0 |
| $\Vert\, \mathbb{E}[\,\tilde{\mathbf{d}}\,] - \mathbb{E}[\,\hat{\mathbf{d}}\,] \,\Vert_2 \;/\; \Vert\, \mathbb{E}[\,\tilde{\mathbf{d}}\,] \,\Vert_2$ | | | |
| Relative covariance error (%) | 100 | 7 | 37 |
| $\Vert\, \mathbb{V}\mathrm{ar}[\,\tilde{\mathbf{d}}\,] - \mathbb{V}\mathrm{ar}[\,\hat{\mathbf{d}}\,] \,\Vert_F \;/\; \Vert\, \mathbb{V}\mathrm{ar}[\,\tilde{\mathbf{d}}\,] \,\Vert_F$ | | | |
| Relative covariance approximation (%) | 0 | 93 | 63 |
| 100 – Relative covariance error | | | |

In order to improve the quality of the scenario reduction, one could augment the number of scenarios considered (at the price of worsening the computational tractability of the resulting optimization problem). In this way the MC random vector could get arbitrarily close to the original continuous random vector $\mathbf{d}$ (see Section 3). However, this would not be possible with the CS method. Even by considering a very large number of conditional scenarios, the CS random vector $\hat{\mathbf{d}}^{\mathbf{r}}$ could get arbitrarily close to the RCE random vector $\mathbf{d}^{\mathbf{r}}$ but not to the original continuous random vector $\mathbf{d}$. Thus, the relative covariance error of the CS method in Table 4 (37%) has a theoretical lower bound which, by Proposition 6.2, can be computed as follows (see the parameter values in Table 2):

$$\frac{\Vert\, \mathbb{V}\mathrm{ar}[\,\mathbf{d}\,] - \mathbb{V}\mathrm{ar}[\,\mathbf{d}^{\mathbf{r}}\,] \,\Vert_F}{\Vert\, \mathbb{V}\mathrm{ar}[\,\mathbf{d}\,] \,\Vert_F} = \frac{1}{\Vert\, \Sigma \,\Vert_F} \left\Vert \Big( \sigma_{ij} - \alpha_{ij}\sigma_{ij} \Big)_{ij \in \mathcal{R}\times\mathcal{R}} \right\Vert_F = 31.52\%. \tag{18}$$

This bound could be attained by considering a very large number of conditional scenarios. Notice that in this paper we compare the CS problem to the MC problem (with a reduced number of scenarios) as *approximations* to the RP problem. With the MC method it would possible to solve the RP problem to optimality by taking a number of scenarios large enough. In contrast, the CS method has been designed only to approximate the RP problem in order to obtain good, possibly suboptimal, solutions with low computational burden.

## 4.3 Comparing the CS, MC and EV solutions

The simplest approximation of the RP problem corresponds to the EV problem, which approximates the random vector of demands by its expectation $\bar{d} = \mathbb{E}[\,\tilde{\mathbf{d}}\,]$. After solving the EV problem we have obtained:

$$\begin{aligned} z^*_{EV} &= \;\; 681,264 \text{ euros} \\ u^*_{EV} &= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \end{aligned}$$

To formulate the MC problem we consider $S_2 = 240$ scenarios randomly sampled from the initial set of $S_1 = 10^5$ scenarios that model the random demand. As already pointed out, we consider 240 scenarios in order to match the number of conditional scenarios that we will consider in the CS method. After solving the MC problem we have obtained:

$$\begin{aligned} z^*_{MC} &= \;\; 681,627 \text{ euros} \\ u^*_{MC} &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Table 5: Comparing the EV, MC and CS solutions.

| | Predicted expected cost | | | Achieved expected cost | | |
|---|---|---|---|---|---|---|
| | EV | MC | CS | E-EV | E-MC | E-CS |
| Cost (euros) | 681,264 | 681,627 | 683,319 | 751,414 | 687,422 | 687,347 |
| Solution time (s) | $< 1$ | 29 | 18 | 32 | 32 | 32 |

To formulate the CS problem we consider 240 conditional scenarios obtained as described in Section 4.2. After solving the CS problem we have obtained:

$$\begin{aligned} z^*_{CS} = &\quad 683,319 \text{ euros} \\ u^*_{CS} = &\quad \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Notice that the MC and the CS problems have the same structure and size. The only difference between them is the way it is used to approximate the random vector $\tilde{\mathbf{d}}$ (scenarios $\tilde{d}^s$ versus conditional scenarios $\hat{d}^{re}$). Therefore, the resulting MILP problems have the same dimensions for the two approaches.

Let us call EV the optimal value obtained by solving the EV problem, that is EV $= z^*_{EV}$. As pointed out in [7], the expected cost *predicted* by EV is, in general, different from the expected cost *achieved* by the EV solution. The achieved expected cost is known in the literature as the 'Expected result of using the EV solution' (E-EV) and can be computed by solving the E-EV problem, which is nothing but the RP problem (13)–(17) with the additional constraint $u = u^*_{EV}$, that is, fixing the first stage decision to the first stage EV solution.

This remark is also valid for the CS and MC problems and, therefore, one can define and compute the values CS, E-CS, MC and E-MC in an analogous way. To set the E-EV, E-CS and E-MC problems we have used a set of scenarios $\mathcal{S}_3$ randomly sampled from the initial set of scenarios $\mathcal{S}_1$, such that $|\mathcal{S}_3| = S_3 = 10^4$. In this case one has $\tilde{p}^s = 1/S_3$ for all $s \in \mathcal{S}_3$. In order to compare the MC, EV and CS solutions, we have collected all the above mentioned values in Table 5. We observe that the the achieved expected cost obtained by the CS and MC approaches are similar. We also observe that for the EV problem the achieved expected cost (E-EV=751,414 euros) is more than 9% worse than the MC and CS counterparts.

### 4.4 How many conditional scenarios?

In this section we analyze the relationship between the number of conditional scenarios and the quality of the approximated solution. To this end we solve the CS problem by considering 120, 240 and 480 conditional scenarios, which corresponds to $E_r = 4$, 8 and 16, respectively. In Table 6, we observe that, as the number of conditional scenarios increases, the 'Relative covariance error' decreases and asymptotically approaches 31.52%, the theoretical lower bound computed in (18). In all the cases, the 'Relative expectation error' has been equal to zero and the CS scenario reduction time has been under 2 seconds.

Table 6: Covariance approximation quality.

| | Conditional scenarios | | |
|---|---|---|---|
| | 120 | 240 | 480 |
| Relative covariance error (%) | 49 | 37 | 32 |
| $\parallel \mathbb{V}\mathrm{ar}[\,\tilde{\mathbf{d}}\,] - \mathbb{V}\mathrm{ar}[\,\hat{\mathbf{d}}\,] \parallel_F \,/\, \parallel \mathbb{V}\mathrm{ar}[\,\tilde{\mathbf{d}}\,] \parallel_F$ | | | |
| Relative covariance approximation (%) | 51 | 63 | 68 |
| 100 – Relative covariance error | | | |

In Table 7 the ‘Predicted expected cost’ increases monotonically with the number of conditional scenarios. However, the ‘Achieved expected cost’ does not necessarily improve as the number of conditional scenarios increases (although they are similar). On the other hand, the ‘Time to solve the CS problem’ increases quickly with the number of conditional scenarios. Therefore, in this context and considering Tables 6 and 7, taking 240 conditional scenarios ($E_r$=8) seems a good balance between the computational burden and the quality of the approximated solution. In the remaining of the paper we will use $E_r$=8. In this section the E-CS problem has been solved with $S_3 = 10^4$ in all the cases.

Table 7: Quality of the approximated solution.

| | Conditional scenarios | | |
|---|---|---|---|
| | 120 | 240 | 480 |
| Predicted expected cost (euros) | 683,254 | 683,319 | 683,444 |
| Time to solve the CS problem (s) | 3 | 18 | 45 |
| Achieved expected cost (euros) | 687,539 | 687,347 | 687,828 |
| Time to solve the E-CS problem (s) | 32 | 32 | 32 |

## 4.5 Comparing the MC and CS performances

In Section 4.3 we have observed that the MC and CS approaches obtain solutions with a significantly lower achieved expected cost compared to the EV approach (see Table 5). In this section we compare the MC and CS performances by solving the stochastic CFL problem for different number of facilities and clients. Table 8 reports the size of the CFL instances, that is, number of facilities and clients, the reduced number of scenarios (the initial number of scenarios is $10^5$ for the ten instances) and the number of rows and columns of the constraint matrix of the corresponding MILP instance. Notice that the CS and the MC instances have the same size.

Table 8: Size of the stochastic CFL instances.

| Instance | Facilities | Clients | Scenarios | Rows | Columns |
|---|---|---|---|---|---|
| 1 | 16 | 48 | 384 | 24,576 | 313,360 |
| 2 | 17 | 51 | 408 | 27,744 | 374,561 |
| 3 | 18 | 54 | 432 | 31,104 | 443,250 |
| 4 | 19 | 57 | 456 | 34,656 | 519,859 |
| 5 | 20 | 60 | 480 | 38,400 | 604,820 |
| 6 | 21 | 63 | 504 | 42,336 | 698,565 |
| 7 | 22 | 66 | 528 | 46,464 | 801,526 |
| 8 | 23 | 69 | 552 | 50,784 | 914,135 |
| 9 | 24 | 72 | 576 | 55,296 | 1,036,824 |
| 10 | 25 | 75 | 600 | 60,000 | 1,170,025 |

Table 9 reports the value of the relative covariance approximation introduced in Table 4, whose value is, on average, 92.15% and 63.42% for the MC and CS methods, respectively. Thus, in order to reduce the number of scenarios, the MC method will be better than the CS method in terms of the second moment matching criterion. However, in the context of stochastic MILP problems, it can be useful to use the CS method in cases where the MC solution times are too long. As we will see below, in this context the CS method may obtain solutions of the same quality with shorter solution times compared to the MC method. Notice that, in this table, the scenario reduction time for the MC method is not reported since it is under one second for all the instances. The relative expectation error, introduced in Table 4, is also not reported since it is 0% and under 1% for the CS and MC approaches, respectively.

Table 9: Scenario reduction: Relative covariance approximation.

| Instance | Relative covariance approximation (%) | | Scenario reduction time (s) |
|---|---|---|---|
| | MC | CS | CS |
| 1 | 92.85 | 63.37 | 7 |
| 2 | 94.94 | 63.38 | 8 |
| 3 | 87.49 | 63.41 | 8 |
| 4 | 89.92 | 63.40 | 9 |
| 5 | 95.17 | 63.43 | 10 |
| 6 | 88.45 | 63.43 | 12 |
| 7 | 90.88 | 63.44 | 12 |
| 8 | 91.30 | 63.44 | 13 |
| 9 | 96.28 | 63.45 | 14 |
| 10 | 94.20 | 63.44 | 16 |
| Average | 92.15 | 63.42 | 11 |

The CS and MC performances are compared in Figure 2, regarding the 'Achieved expected cost', in Figure 3, regarding the 'Solution time', and in Figure 4, regarding the 'LP gap' (the relative gap between the optimal solution of the MILP problem and the optimal solution of its LP relaxation). The figures used to plot these graphs can be found in Tables 10, 11 and 12, respectively. In these figures we observe that: 1) The achieved expected cost obtained by the CS and MC approaches are similar. 2) On average, the CS approach has been 2.5 times faster than the MC approach (the solution time, plotted in Figure 3, accounts for the scenario reduction time plus the time to solve the corresponding stochastic CFL instance). 3) A possible reason for the faster performance of the CS approach is its smaller LP gap (the relative gap between the optimal solution of the MILP problem and the optimal solution of its LP relaxation).

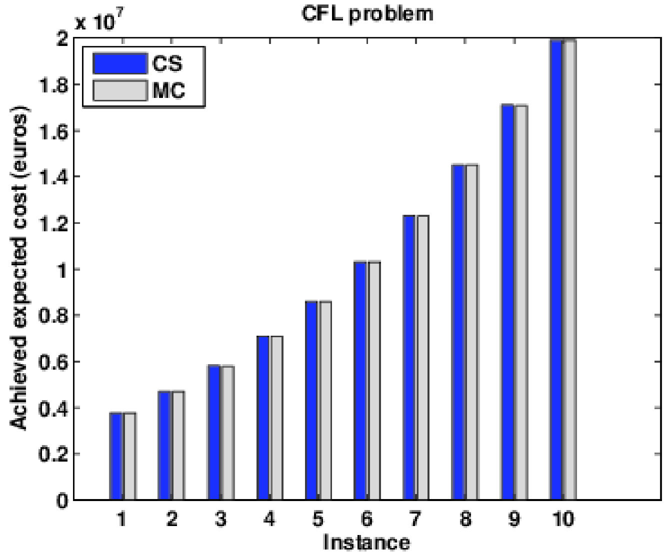


Figure 2: The achieved expected cost obtained by the CS and MC approaches are similar (see Table 10).

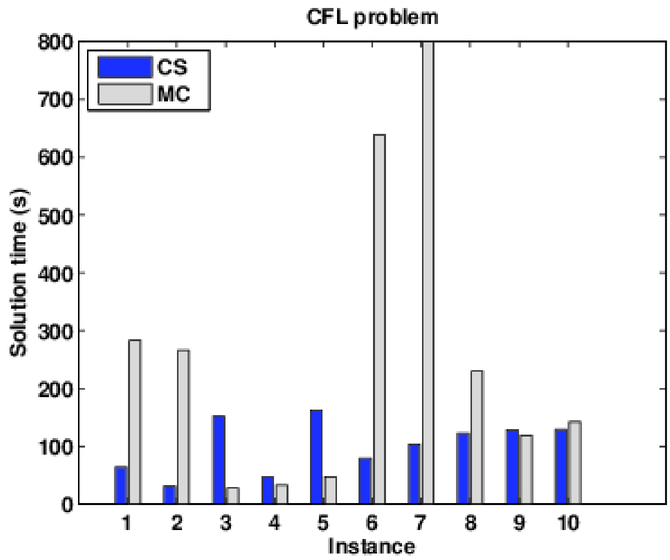


Figure 3: On average, the CS approach has been 2.5 times faster than the MC approach (see Table 11).

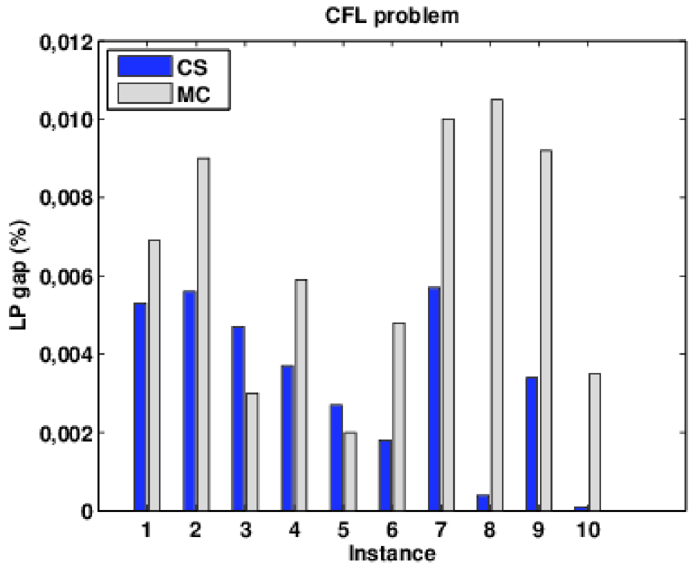


Figure 4: On average, the LP gap of the CS instances has been almost 2 times smaller than the LP gap of the MC instances (see Table 12).

Finally, let us have a look to the following three tables, which have been used to plot Figures 2, 3 and 4, respectively. In Table 10 we observe that, on average, the CS achieved expected cost has been 0.05% higher than the MC counterpart (CS:10,410,564 versus MC:10,405,241 euros). In Table 11 we have used the following notation: $\text{CS}_1$ = 'Time to compute the conditional scenarios', $\text{CS}_2$ = 'Time to solve the CS problem', Total = $\text{CS}_1 + \text{CS}_2$. We do not report the time to sample the MC scenarios since it has been under one second for all the instances. On average, the CS solution time has been 2.5 times faster (CS:102 versus MC:259 seconds). In Table 12 the LP gap corresponds to the relative gap between the optimal solution of the MILP problem and the optimal solution of its LP relaxation. On average, the LP gap of the CS approach has been almost 2 times smaller (CS:0.0033% versus MC:0.0065%).

Table 10: Comparing the E-CS and E-MC values.

| Instance | E-CS (euros) | E-MC (euros) |
|---|---|---|
| 1 | 3,759,329 | 3,758,544 |
| 2 | 4,698,521 | 4,698,246 |
| 3 | 5,806,134 | 5,801,227 |
| 4 | 7,089,521 | 7,089,521 |
| 5 | 8,602,819 | 8,594,859 |
| 6 | 10,305,452 | 10,304,597 |
| 7 | 12,316,605 | 12,301,684 |
| 8 | 14,518,430 | 14,518,665 |
| 9 | 17,108,169 | 17,084,409 |
| 10 | 19,900,656 | 19,900,656 |
| Average | 10,410,564 | 10,405,241 |

Table 11: Comparing the CS and MC solution times.

| Instance | CS problem | | | MC problem |
|---|---|---|---|---|
| | $CS_1$ | $CS_2$ | Total (s) | Total (s) |
| 1 | 7 | 57 | 64 | 283 |
| 2 | 8 | 23 | 31 | 266 |
| 3 | 8 | 144 | 152 | 28 |
| 4 | 9 | 37 | 46 | 33 |
| 5 | 10 | 152 | 162 | 47 |
| 6 | 12 | 67 | 79 | 638 |
| 7 | 12 | 91 | 103 | 800 |
| 8 | 13 | 110 | 123 | 230 |
| 9 | 14 | 114 | 128 | 119 |
| 10 | 16 | 113 | 129 | 142 |
| Average | 11 | 91 | 102 | 259 |

Table 12: Comparing the LP gap of the MC and CS problems.

| Instance | CS problem | | | MC problem | | |
|---|---|---|---|---|---|---|
| | LP bound (euros) | $z^*_{CS}$ (euros) | LP gap (%) | LP bound (euros) | $z^*_{MC}$ (euros) | LP gap (%) |
| 1 | 3,755,710 | 3,755,909 | 0.0053 | 3,729,246 | 3,729,505 | 0.0069 |
| 2 | 4,697,833 | 4,698,096 | 0.0056 | 4,640,258 | 4,640,675 | 0.0090 |
| 3 | 5,806,075 | 5,806,346 | 0.0047 | 5,837,008 | 5,837,185 | 0.0030 |
| 4 | 7,098,450 | 7,098,711 | 0.0037 | 7,155,997 | 7,156,420 | 0.0059 |
| 5 | 8,594,403 | 8,594,634 | 0.0027 | 8,688,482 | 8,688,658 | 0.0020 |
| 6 | 10,313,915 | 10,314,097 | 0.0018 | 10,384,489 | 10,384,986 | 0.0048 |
| 7 | 12,278,587 | 12,279,282 | 0.0057 | 12,163,593 | 12,164,811 | 0.0100 |
| 8 | 14,510,677 | 14,510,728 | 0.0004 | 14,565,506 | 14,567,034 | 0.0105 |
| 9 | 17,032,515 | 17,033,095 | 0.0034 | 17,055,797 | 17,057,359 | 0.0092 |
| 10 | 19,868,189 | 19,868,204 | 0.0001 | 19,698,378 | 19,699,063 | 0.0035 |
| Average | 10,395,635 | 10,395,910 | 0.0033 | 10,391,875 | 10,392,570 | 0.0065 |

## 5 Conclusions

The Conditional Scenario (CS) problem was introduced in [4] as an effective approximation to the two-stage stochastic mixed-integer linear programming problem with recourse (for short, the RP problem). In this context, the scenario based random vector that models the uncertain parameters of the RP problem is called the the RP random vector. From a theoretical point of view, we have analyzed the CS random vector, obtained by the CS scenario reduction method, as an approximation to the RP random vector. In the introduction we have raised question a): *Which is the degree of similarity between the RP and CS random vectors in terms of the first two moments?* Roughly speaking, we have shown that the RP and CS random vectors have the same expectation. We have also shown that they have covariance coefficients which are proportional, provided that the RP random vector is a discretization of a multivariate normal random vector.

From a computational point of view, we have compared the CS random vector to the Monte Carlo (MC) random vector, obtained by sampling a reduced set of scenarios. We have performed a numerical experiment where we have solved the capacitated facility location problem with uncertain demand. We have observed that the MC method obtains a better approximation to the original probability distribution than the CS method in terms of the moment matching criterion. Furthermore, we have used this experiment to answer se second question raised in the introduction: *b) How does the CS problem compare with the MC problem as approximations to the RP problem?* Regarding the solution quality, we have observed that the two methods obtain solutions with a very similar achieved expected cost. Regarding performance, the CS method has resulted, on average, 2.5 times faster than the MC method. Here we compare the total time which accounts for the scenario reduction time plus the time to solve the corresponding MILP problem. One possible explanation for this observation is that the CS instances show an LP gap which is, on average, 2 times smaller than the LP gap of the corresponding MC instances (with the same number of scenarios).

Of course, with the empirical results reported in this paper we can only formulate the following conjecture: *The CS problem usually has a smaller LP gap than the corresponding MC problem with the same number of scenarios.* This conjecture is a matter of further research. Notice that this potential computational advantage of the CS method over the MC method only applies for MILP problems. In the case of LP problems, this potential computational advantage disappears and one should prefer the MC method, considering its better approximation to the original set of scenarios in terms of the moment matching criterion.

Therefore, this paper could be useful for Operations Research practitioners in order to answer the following question: ”Which method could one use to efficiently model and solve two-stage stochastic optimization problems?”. The answer of this paper is as follows: If the problem is continuous, the MC approach is a good choice. However, if the problem is mixed-integer and the MC approach takes too long, one could try the CS approach with the same number of scenarios (possibly, it will produce smaller LP gaps and, thus, a faster solution of the problem).

**Acknowledgments:** We are grateful to the editor, associate editor and reviewers for their constructive and stimulating suggestions. We are also grateful to the Spanish Ministry of Economy, Industry and Competitiveness (grants MTM2015-63710-P and RTI2018-094269-B-I00).

## References


[1] A. Alonso-Ayuso, L. F. Escudero, M. Guignard, and A. Weintraub. Risk management for forestry planning under uncertainty in demand and prices. *European Journal of Operational Research*, 267:1051–1074, 2018.

[2] A. Alonso-Ayuso, L.F. Escudero, and M.T. Ortuño. BFC, a branch-and-fix coordination algorithmic framework for solving some types of stochastic pure and mixed 0-1 programs. *European Journal of Operational Research*, 151:503–519, 2003.

[3] J. Ang, F. Meng, and J. Sun. Two-stage stochastic linear programs with incomplete information on uncertainty. *European Journal of Operational Research*, 233(1):16–22, 2014.

[4] C. Beltran-Royo. Two-stage stochastic mixed-integer linear programming: The conditional scenario approach. *Omega - The International Journal of Management Science*, 70:31–42, 2017.

[5] C. Beltran-Royo. Two-stage stochastic mixed-integer linear programming: from scenarios to conditional scenarios. *Optimization Online*, March, 2018.

[6] M. Bieniek. A note on the facility location problem with stochastic demands. *Omega, The International Journal of Management Science*, 55:53–60, 2015.

[7] J. R. Birge and F. V. Louveaux. *Introduction to Stochastic Programming (Second Edition)*. Springer (New York), 2011.

[8] M. Chen, S. Mehrotra, and D. Papp. Scenario generation for stochastic optimization problems via the sparse grid method. *Computational Optimization and Applications*, 62(3):669–692, 2015.

[9] M. H. DeGroot and M. J. Schervish. *Probability and Statistics (Fourth Edition)*. Addison Wesley (Boston, USA), 2012.

[10] J. A. Diaz and E. Fernandez. A branch-and-price algorithm for the single source capacitated plant location problem. *Journal of the Operational Research Society*, 53(7):728–740, 2002.

[11] J. Dupacova, N. Growe-Kuska, and W. Romisch. Scenario reduction in stochastic programming: An approach using probability metrics. *Mathematical Programming*, 95(3):493–511, 2003.

[12] V. Gabrel, C. Murat, and A. Thiele. Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3):471–483, 2014.

[13] W. Hardle and L. Simar. *Applied Multivariate Statistical Analysis (Third Edition)*. Springer, Berlin, 2012.

[14] H. Heitsch and W. Romisch. Scenario reduction algorithms in stochastic programming. *Computational Optimization and Applications*, 24(2-3):187–206, 2003.

[15] R. Henrion, C. Kuchler, and W. Romisch. Discrepancy distances and scenario reduction in twostage stochastic mixed-integer programming. *Journal of Industrial and Management Optimization*, 4:363–384, 2008.

[16] R. Henrion and W. Romisch. Optimal scenario generation and reduction in stochastic programming, humboldt-universitat zu berlin, institut fur mathematik. Technical report, 2017.

[17] T. Homem-de-Mello and G. Bayraksan. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1):56–85, 2014.

[18] K. Hoyland, M. Kaut, and S. W. Wallace. A heuristic for moment-matching scenario generation. *Computational Optimization and Applications*, 24(2-3):169–185, 2003.

[19] K. Hoyland and S. W. Wallace. Generating scenario trees for multistage decision problems. *Management Science*, 47(2):295–307, 2001.

[20] P. Kall and J. Mayer. *Stochastic Linear Programming*. Kluwer academic publishers (New York, USA), 2005.

[21] S. Karademir, N. Kong, and O. Aa Prokopyev. On greedy approximation algorithms for a class of two-stage stochastic assignment problems. *Optimization Methods and Software*, 29(1):42–67, 2014.

[22] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502, 2002.

[23] M. Lubin, C. G. Petra, and M. Anitescu. The parallel solution of dense saddle-point linear systems arising in stochastic programming. *Optimization Methods and Software*, 27(4-5):845–864, 2012.

[24] J. M. Morales, S. Pineda, A. J. Conejo, and M. Carrion. Scenario reduction for futures market trading in electricity markets. *IEEE Transactions on Power Systems*, 24(2):878–888, 2009.

[25] G. Ch. Pflug and A. Pichler. *Multistage Stochastic Optimization*. Springer, Cham (Switzerland), 2014.

[26] W. Romisch. Stability of stochastic programming problems, in: Stochastic programming (A. Ruszczynski and A. Shapiro eds.). *Handbooks in operations research and management science*, 10:483–554, Elsevier, Amsterdam, 2003.

[27] T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro. A stochastic programming approach for supply chain network design under uncertainty. *European Journal of Operational Research*, 167(1):96–115, 2005.

[28] S. Sen and H. D. Sherali. Decomposition with branch-and-cut approaches for two-stage stochastic mixed-integer programming. *Mathematical Programming*, 106(2):203–223, 2006.

[29] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9. Society for Industrial Mathematics, Philadelphia, 2009.

[30] L. V. Snyder. Facility location under uncertainty: a review. *IIE Transactions*, 38(7):547–564, 2006.

[31] S. Takriti and J. R. Birge. Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs. *Operations Research*, 48(1):91–98, 2000.

[32] X. Tong, F. Wu, and L. Qi. Worst-case CVaR based portfolio optimization models with applications to scenario planning. *Optimization Methods and Software*, 24(6):933–958, 2009.

[33] H. H. Wu and S. Kucukyavuz. A two-stage stochastic programming approach for influence maximization in social networks. *Computational Optimization and Applications*, 69(3):563–595, 2018.

[34] Z. Zhou, J. Zhang, P. Liu, Z. Li, M. C. Georgiadis, and E. N. Pistikopoulos. A two-stage stochastic programming model for the optimal design of distributed energy systems. *Applied Energy*, 103:135–144, 2013.