Sample average approximation for risk-averse problems: A virtual power plant scheduling application

Sample average approximation for risk-averse problems: A virtual power plant scheduling application

Ricardo M. Lima [a,*], Antonio J. Conejo [b], Loïc Giraldi [a,e], Olivier Le Maître [c], Ibrahim Hoteit [d], Omar M. Knio [a]

[a] *Computer, Electrical and Mathematical Sciences & Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*
[b] *Integrated Systems Engineering-Electrical and Computer Engineering, The Ohio State University, OH, USA*
[c] *Centre de Mathématiques Appliquées, CNRS, Inria, Ecole Polytechnique, Palaiseau, France*
[d] *Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Saudi Arabia*
[e] *CEA, DES, IRESNE, DEC, Cadarache F-13108 Saint-Paul-Lez-Durance, France*

A R T I C L E   I N F O

A B S T R A C T

In this paper, we address the decision-making problem of a virtual power plant (VPP) involving a self-scheduling and market involvement problem under uncertainty in the wind speed and electricity prices. The problem is modeled using a risk-neutral and two risk-averse two-stage stochastic programming formulations, where the conditional value at risk is used to represent risk. A sample average approximation methodology is integrated with an adapted L-Shaped solution method, which can solve risk-neutral and specific risk-averse problems. This methodology provides a framework to understand and quantify the impact of the sample size on the variability of the results. The numerical results include an analysis of the computational performance of the methodology for two case studies, estimators for the bounds of the true optimal solutions of the problems, and an assessment of the quality of the solutions obtained. In particular, numerical experiences indicate that when an adequate sample size is used, the solution obtained is close to the optimal one.

1. Introduction

The optimal operation and electricity market involvement of virtual power plants (VPPs) is currently an active field of research. VPP refers to an aggregation of distributed energy resources (DER) that interacts with the electricity market as a single entity (Awerbuch and Preston, 1997; Pudjianto et al., 2007). These DERs involve generators with small capacities or that face significant trading risks if operating by themselves. For example, a wind farm trading in an electricity market is subject to uncertain wind speeds, electricity prices, and imbalance costs, which may involve high risk. Aggregating multiple and diverse generators into a VPP creates an entity with a single power generation profile and a larger power capacity than the constituting units. A VPP has flexibility due to the complementarity of its diverse resources, and capacity to develop market intelligence to optimize its participation in the electricity market (through offers in the pool and by signing forwards contracts). However, to define a single power output profile, a self-scheduling problem considering all constituting generators, their interactions, and their generation constraints needs to be solved, which is more difficult than solving individual generator problems. The VPP concept is particularly relevant for the integration of distributed renewable energy resources in power systems to mitigate flexibility limitation and power output uncertainty. Additional details on the technical and commercial functions of VPPs can be found in Pudjianto et al. (2007), Morales et al. (2014), and Jansen et al. (2008). In Section 1.1, a review on VPP optimization is presented. Within the different aspects of a VPP operation, we are concerned with the optimal scheduling of the generation units and the interaction with the electricity market.

In the present work, we propose a sample average approximation (SAA) methodology (Kleywegt et al., 2001; Shapiro and Homem-De-Mello, 2000) to solve the risk-averse stochastic programming problem that describes the operation of a VPP. Three important aspects in the implementation of the SAA are studied: 1) the development of an efficient solution methodology; 2) the impact of the sample size on the performance of the methodology; and 3) the determination of point estimates and confidence intervals for the solutions.

---

* Corresponding author.
E-mail addresses: ricardo.lima@kaust.edu.sa (R.M. Lima), conejonavarro.1@osu.edu (A.J. Conejo), loic.giraldi@cea.fr (L. Giraldi), olivier.le-maitre@polytechnique.edu (O. Le Maître), ibrahim.hoteit@kaust.edu.sa (I. Hoteit), omar.knio@kaust.edu.sa (O.M. Knio).

1.1. Literature review: optimization of VPPs

In the literature, VPP models with distinct portfolios of generation units, different interactions with the electricity market, and, consequently, various sources of uncertainty can be found. In terms of interactions with the electricity market, the main features captured by VPP models include:

1. forward contracts (Lima et al., 2015; Pandžić et al., 2013a; Shabanzadeh et al., 2016);
2. offers to the day-ahead market (Baringo and Baringo, 2017; Dabbagh and Sheikh-El-Eslami, 2016; Kardakos et al., 2016; Moghaddam et al., 2013; Pandžić et al., 2013b; Rahmani-Dabbagh and Sheikh-El-Eslami, 2016; Shayegan-Rad et al., 2017; Tajeddini et al., 2014; Zamani et al., 2016);
3. offers/bids to the balancing market (Dabbagh and Sheikh-El-Eslami, 2016; Kardakos et al., 2016; Tajeddini et al., 2014);
4. offers to spinning reserves (Dabbagh and Sheikh-El-Eslami, 2016); and
5. strategic offering considering the market clearing and other market players' strategies (Kardakos et al., 2016).

The operational decision-making problem of VPPs involves decisions under uncertainty due to the nature of the renewable energy sources and electricity prices. Therefore, stochastic optimization approaches constitute a natural framework to address this type of problems. The most used approaches are based on stochastic programming and robust optimization. Some examples include the application of risk-neutral stochastic programming models (Lima et al., 2018; Pandžić et al., 2013a; 2013b), risk-averse stochastic programming (Dabbagh and Sheikh-El-Eslami, 2016; Kardakos et al., 2016; Lima et al., 2018; Moazeni et al., 2015; Moghaddam et al., 2013; Tajeddini et al., 2014), robust optimization (Lima et al., 2015; Rahimiyan and Baringo, 2016; Shabanzadeh et al., 2015), and recently, a hybrid approach based on stochastic programming and robust optimization (Baringo and Baringo, 2017).

When risk-neutral or risk-averse stochastic programming models are used, the problem is solved for a random uncertainty sample. The main assumption is that the sample used is representative of the full distribution of the uncertainties in the problem. In the works above, the sample size ranges from less than 100 elements to a maximum of 400, except in the work Lima et al. (2018) where samples of up to 25,000 elements were used. In some works, a reduced sample size was obtained using a scenario-reduction method (Dupačová et al., 2003). Typically, the size of the sample is limited by the capability to solve the problems using the corresponding extensive form and a branch and cut solver for mixed-integer linear programming (MILP) problems. Compared with the works cited above, our VPP model considers a time horizon of one week divided into periods of one hour, whereas most works above consider the 24 h of the next day. The time horizon of one week avoids the myopic view of a single-day horizon, which neglects the temporal coupling of some VPP constraints that condition next day operations. Therefore, this extended horizon enables considering constraints of thermal units, e.g., minimum up-time and minimum down-time, and water mass balances in the reservoir of hydro units across consecutive days, at the cost of a larger model.

We emphasize that risk-averse stochastic programming models are more complex to solve than risk-neutral stochastic programming; therefore, the impact of the sample size and the length of the time horizon on the computing time of risk-averse problems are comparatively more critical. Some works propose an out-of-sample analysis that provides an additional characterization of the solution resulting from the stochastic programming model (Baringo and Baringo, 2017). Compared to an out-of-sample analysis that performs a study for a single solution, the SAA methodology includes a first stage with multiple optimization replications using different samples to generate alternative solutions and to provide a point estimate of the upper bound on the true objective function value (for a maximization problem). The work

in Lima et al. (2018) does not use an SAA approach, but rather focuses on risk-averse problems and their solutions using the L-Shaped method (Van Slyke and Wets, 1969). In that work, 14 variants of the L-Shaped method using combinations of single and multiple optimality cuts and efficient parallel implementations are discussed, without considering SAA.

In this work, we propose a more elaborate approach that efficiently relies on multiple samples to assess solutions and determine confidence intervals on the relevant performance indices of the VPP operation. No such work is available for VPP applications.

1.2. Literature review: sample average approximation

The term sample average approximation is used in the literature to refer to a problem that approximates a stochastic optimization one (Shapiro and Homem-De-Mello, 2000), but it is also used to indicate a methodology where the solution of that approximate problem is performed multiple times and increasingly accurate solutions are obtained (Kleywegt et al., 2001). The repetition enables the calculation of confidence intervals of the optimal objective function value, and also the assessment of the solution.

The detailed characterization and statistical properties of the SAA methodology, namely consistency and rates of convergence to optimal objective function values and solution, have been studied by Mak et al. (1999), Shapiro and Homem-De-Mello (2000), Kleywegt et al. (2001), and reviewed in detail in Shapiro (2009) and Homem-de Mello and Bayraksan (2014). In these works, expressions for the convergence to the optimal value and solution are developed and expressions relating the probability of an optimal solution x_N^* to be equal to the true solution x^* with the sample size are established. However, as stated in Kleywegt et al. (2001) and Linderoth et al. (2006), it is impractical to determine some of the constants in those expressions. Furthermore, the expressions do not provide an estimate of the computing time required to obtain the approximated solutions and point estimates of the bounds. Therefore, computational experiments are essential to determine the trade-off between the computational resources and the accuracy obtained from the sample size. This type of analysis was performed, for example, in Linderoth et al. (2006) and Verweij et al. (2003).

Wang (2007) and Wang and Ahmed (2008) were the first to propose SAA methods for risk-averse stochastic programming problems. They focused on min-max problems with expected value objectives and problems with constraints involving expected values. For these problems, they proved that the results of their SAA problems converge exponentially fast to the true results of the stochastic problems as the sample size increases.

A related method to this work, which has been extended to integrate an SAA methodology with risk measures, is stochastic dual dynamic programming (SDDP) (Pereira and Pinto, 1991). In Philpott and de Matos (2012) and Shapiro et al. (2013) are developed risk-averse SDDP methods to address multistage linear stochastic programming models describing power systems planning problems. In risk-averse SDDP, the upper bound estimators are known to be weak (Shapiro, 2011), and in this regard, importance sampling was proposed in Kozmík and Morton (2015) to improve the quality of those estimators (for a minimization problem).

Regardless of the approach used, it is generally essential to develop an efficient SAA methodology that can handle a large number of replications as well as large sample sizes to calculate tight confidence intervals. Furthermore, an efficient methodology is paramount to study the impact of the sample size on the computational performance of the methodology. This is a central theme of the present work.

1.3. SAA applied to a risk-averse VPP optimization model

The application of SAA methodologies to the scheduling problems of VPPs has not been reported in the literature. To fill the gap, we propose

an SAA methodology to solve these problems that is based on methods from Linderoth et al. (2006). Specifically, a distinct class of risk-averse problems from those considered in Linderoth et al. (2006) is addressed, and thus, insights that are not found in Linderoth et al. (2006) and Verweij et al. (2003) are analyzed and discussed.

In this work, new avenues to solve the VPP problem are explored, focusing on the application of an SAA methodology that leads to comprehensive results for the VPP, including point estimates, confidence intervals, and the assessment of the solution obtained. The main contributions of this work are twofold: 1) providing new and efficient strategies to improve the performance of the SAA methodology; and 2) carrying out a comprehensive computational study applied to a VPP risk-averse two-stage stochastic programming problem. Specifically, an efficient initialization strategy to handle multiple optimization replications that reduces the computing time is proposed. In addition, a new method to reduce the number of times the bound estimation stage is ran is discussed. As a result, a detailed analysis of the effect of the sample size shows that, unexpectedly, increasing the sample size for the optimization leads to a reduction of the overall computing time in one of the procedures implemented. This reduction occurs because savings in the bound estimation stage compensate the additional time spent on the optimization with larger samples.

The paper is organized as follows. In Section 2, the problem statement and a description of the VPP studied are presented. The stochastic models with a focus on the objective functions are described in Section 3. The SAA approach and the implementation of two solution procedures are presented in Section 4, while the sampling techniques are shown in Section 5. The computational results are discussed in Section 6 and conclusions are summarized in Section 7. In a supplementary document, appendices with the deterministic model of the VPP, flow diagrams of the solution procedures used, and additional extensive computational results are made available.

2. Problem statement

In this work, we revisit the problem considered in Lima et al. (2018) but using an SAA methodology. The problem addresses the optimal operation and electricity market involvement of a VPP that consists of a thermal unit, a wind power farm, and a pumped-storage hydroelectric plant. This configuration provides the flexibility of two dispatchable units, two renewable sources, and storage to manage the energy available and to respond to the market. The VPP participates in the electricity market as a single entity by selling and buying electricity in the market pool and through weekly contracts. The time horizon is one week, divided into 168 hourly periods.

The VPP generates electricity through the three units and may consume electricity to pump back water to the upper reservoir of the hydro plant. Uncertainty in the wind speed and the electricity prices is considered, with a constant value for each hourly period.

The decision sequence of the VPP involves a two-stage framework. In the first-stage, the decisions are made before the beginning of the time horizon, where the VPP has to decide on the self-scheduling of the thermal unit and the electricity to buy or sell through forward contracts. The self-scheduling of the thermal unit involves only the commitment, defined as the periods where it is up or down. The VPP can choose to buy or sell electricity from two different weekly contracts, each one involving fixed prices and quantities. In the second-stage, the decisions for the full week are aggregated into one stage. These decisions are related to the dispatching of the thermal and hydro plants and to the additional electricity to buy or sell in the electricity pool (short-term electricity market). The dispatch involves the hourly determination of the power to generate by the thermal unit when it is up and the power generation or consumption by the pumped-storage hydro plant.

The objective of the VPP is to maximize the operational profit by determining an optimal solution concerning: a) the commitment, dispatch, and coordination between units in each period; b) the electricity bought

and sold in each period; and c) the electricity bought or sold through contracts. From the decision-maker perspective, the optimal first-stage variables are the most relevant. These decisions are implemented and fixed during the time horizon, whereas the optimal second-stage variables for the 168 periods represent recourse actions based on the realizations of the uncertain parameters. In practice, during the week, the VPP needs to solve additional decision problems to define the bidding strategy for the specific electricity market pool (day-ahead market, intra-day market, balancing market). These problems have a shorter horizon and use updated information for wind speed and electricity prices.

The thermal unit has lower and upper bounds on the power generation, minimum up-time and down-time, start-up and shutdown and power-up and power-down ramp rate limits. Regarding costs, this unit has fixed and variable generation costs, hot and cold start costs. The state of the thermal unit before the beginning of the time horizon is known. The hydro plant has lower and upper limits on the volume of water of the upper reservoir, and an upper bound on the pumped flow of water.

3. Risk-averse stochastic models

In this section, the main characteristics of the risk-averse stochastic programming formulations for the VPP problem are described. The deterministic version of the VPP optimization problem is described in Appendix A, where the objective function, constraints, variables, and input data are presented. Below, we focus on building a general risk-averse stochastic programming formulation that will be used in the SAA methodology outlined in Section 4.

3.1. VPP general stochastic programming formulation

The stochastic programming formulation of the VPP problem is an extension of the deterministic VPP problem, which results from considering the electricity prices and the wind power output as random variables. In the deterministic VPP problem, the aim is to maximize the profit defined as the difference between the revenues from selling electricity and the costs of operation and buying electricity. In the stochastic version, the profit is a random function, and thus the aim is to maximize a functional of the profit. We start by denoting the profit function $\overline{f}$ as

$$
\begin{aligned}
\overline{f}(x^+, x^-, u, y^+(\xi), y^-(\xi), s(\xi), \xi) := {}& (c^+)^\top x^+ + (c^-)^\top x^- + \overline{c}^{0\top} u \\
& + \tilde{c}^{0\top}(\xi) y^+(\xi) - \tilde{c}^{0\top}(\xi) y^-(\xi) + \underline{c}^\top s(\xi)
\end{aligned}
\tag{1}
$$

where $\xi : \Omega \to \mathbb{R}^r$ is a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$, with Ω being the set of all possible outcomes, $\mathcal{F}$ a σ-algebra and P a probability measure. The vectors c^+, c^-, $\overline{c}^0$, and $\underline{c} \in \mathbb{R}^{\overline{n}_1}$ denote deterministic vectors with known parameters, whereas $\tilde{c}^0(\xi)$ is a random vector. The vectors x^+, x^-, u, $y^+(\xi)$, $y^-(\xi)$, $s(\xi) \in \mathbb{R}^{\overline{n}_1}$ represent the variables of the VPP model.

The vectors x^+, x^-, and u correspond to the first-stage decisions. x^+, x^- capture the power to sell and buy through the contracts, respectively. The revenues related to the contracts are represented by $(c^+)^\top x^+$, while the costs by $(c^-)^\top x^-$. The vector u encapsulates the binary variables that define the operation mode of the thermal unit. Therefore, the term $\overline{c}^{0\top} u$ represents the startup costs, shutdown costs, and fixed costs of operation of the thermal unit.

The second-stage decisions are represented by the vectors $y^+(\xi)$, $y^-(\xi)$, $s(\xi)$, and $r(\xi)$. Note that $r(\xi)$ is not part of the objective function. The vectors $y^+(\xi)$, $y^-(\xi)$ denote the energy to sell or buy to the market, and the vector $\tilde{c}^0(\xi)$ denotes the uncertain hourly electricity prices. The vector $s(\xi)$ is the electricity generated by the thermal unit. Finally, the vector $r(\xi)$ captures the energy generated or consumed by the pumped-storage hydro unit and the volumes of water associated with this unit.

The stochastic model of the VPP aims at maximizing a functional of the profit subject to the constraints that define the region of operation of the thermal unit, pumped-storage hydro plant, and the two contracts

that are offered to the VPP:

$$\max_{x^+,x^-,v,u,y^+(\xi),y^-(\xi),s(\xi),r(\xi)} \psi\left[\overline{f}(x^+, x^-, u, y^+(\xi), y^-(\xi), s(\xi), \xi)\right]$$

$$\text{s.t.} \quad A^+ x^+ + A^- x^- + \overline{B}v \le \overline{b}$$

$$\overline{E}u \le \overline{g}$$

$$\overline{C}u + \overline{D}s(\xi) \le \overline{d}, \quad \text{a.s.}$$

$$s(\xi) - y^+(\xi) + y^-(\xi) + r(\xi) - x^+ + x^- \tag{2}$$

$$+ \overline{h}(\xi) = 0, \quad \text{a.s.}$$

$$Jr(\xi) \le j, \quad \text{a.s.}$$

$$x^+, x^- \in \mathbb{R}^{\overline{n}_1}_+, v, u \in \mathbb{B}^{\overline{n}_1}, y^+(\xi), y^-(\xi),$$

$$quad \, s(\xi) \in \mathbb{R}^{\overline{n}_1}_+, r(\xi) \in \mathbb{R}^{\overline{n}_1},$$

where $\mathbb{B} := \{0, 1\}$, v are the binary variables associated with the choice of blocks in the contracts, $\overline{h}(\xi) \in \mathbb{R}^{\overline{n}_1}$ defines the uncertain wind power, $A^+, A^- \in \mathbb{R}^{\overline{m}_1 \times \overline{n}_1}$, $\overline{B} \in \mathbb{R}^{\overline{m}_1 \times \overline{n}_1}$, $\overline{C} \in \mathbb{R}^{\overline{m}_2 \times \overline{n}_1}$, $\overline{D} \in \mathbb{R}^{\overline{m}_2 \times \overline{n}_1}$, $E \in \mathbb{R}^{\overline{m}_3 \times \overline{n}_1}$, $J \in \mathbb{R}^{\overline{m}_4 \times \overline{n}_1}$ are matrices with known parameters, and $\overline{b} \in \mathbb{R}^{\overline{m}_1}$, $\overline{d} \in \mathbb{R}^{\overline{m}_2}$, $g \in \mathbb{R}^{\overline{m}_3}$, $j \in \mathbb{R}^{\overline{m}_4}$ are vectors with known parameters.

In Problem (2), the first constraint covers the contracts. The second constraint represents the relationships among the binary variables associated with the commitment of the thermal unit. The third constraint captures the region of operation of the thermal unit, including minimum up and down times, and power generation limits. The fourth constraint represents the energy balance to the VPP. Finally, the last constraint captures the region of operation of the pumped-storage hydro plant. See Appendix A for details.

For the sake of clarity on the exposition of the SAA methodology, a compact formulation of the model with a simplified notation is introduced. We set $x := (x^+, x^-)$, $y := (y^+(\xi), y^-(\xi), s(\xi), r(\xi))$, $z := (u, v)$, $c := (c^+, c^-)$, $\overline{c} := \overline{c}^0$, $\tilde{c} := \tilde{c}^0$ with proper dimensions:

$$\max_{x,z,y(\xi)} \psi\left[f(x, z, y(\xi), \xi) := c^\top x + \overline{c}^\top z + \tilde{c}^\top(\xi) y(\xi)\right]$$

$$\text{s.t.} \quad Ax + Bz \le b$$

$$Cz + Dy(\xi) \le d, \quad \text{a.s.} \tag{3}$$

$$y(\xi) + x = h(\xi), \quad \text{a.s.}$$

$$x \in \mathbb{R}^{n_1}_+, z \in \mathbb{B}^{n_1}, y(\xi) \in \mathbb{R}^{n_1}_+,$$

where x, z denote first-stage variables and $y(\xi)$ second-stage variables, $h(\xi) \in \mathbb{R}^{n_1}$ and $\tilde{c}(\xi) \in \mathbb{R}^{n_1}$ are random vectors, c, $\overline{c}$, and $\tilde{c}$ are deterministic vectors with known parameters, $A \in \mathbb{R}^{m_1 \times n_1}$, $B \in \mathbb{R}^{m_1 \times n_1}$, $C \in \mathbb{R}^{m_2 \times n_1}$, and $D \in \mathbb{R}^{m_2 \times n_1}$ are matrices with known parameters, and $b \in \mathbb{R}^{m_1}$ and $d \in \mathbb{R}^{m_2}$ are vectors with known parameters.

3.2. Risk-averse VPP stochastic models

Based on (3), we consider a risk-averse formulation that involves the CVaR of the profit in the objective function. The CVaR is a coherent risk measure with relevant properties in terms of convexity (Artzner et al., 1999; Pflug, 2000; Rockafellar and Uryasev, 2000). Therefore, it has clear advantages over non-convex risk measures. Note that CVaR is not the only possibility to measure the risk; other measures have been suggested including the variance, value at risk, drawdown, or buffered probability of exceedance; see Pflug and Römisch (2007) and Shang et al. (2018). Also note that many works in the literature apply the CVaR to a loss function, whereas in this work the CVaR is applied to a profit one, which leads to different definitions below as compared with the works cited. The terms "average value at risk" and "expected shortfall" are also used to refer to the CVaR (Rockafellar, 2007).

The CVaR for the $(1 - \alpha)$ quantile of random variable f is defined as the conditional expectation of f for $f \le \text{VaR}_{1-\alpha}[f]$ (Rockafellar and Uryasev, 2000):

$$\text{CVaR}_{1-\alpha}[f] = \mathbb{E}\left[f | f \le \text{VaR}_{1-\alpha}[f]\right], \tag{4}$$

where $\text{VaR}_{1-\alpha}[f]$ is the value at risk defined as

$$\text{VaR}_{1-\alpha}[f] = \max\{w | F_f(w) \le 1 - \alpha\}, \tag{5}$$

and F_f is the cumulative distribution of f, $F_f(w) = P\{f \le w\}$.

From an optimization perspective, there are two additional definitions that are relevant. The first is due to Pflug (2000), which defines the CVaR through the optimization problem

$$\text{CVaR}_{1-\alpha}[f] = \max_\eta \left\{ \eta - \frac{1}{1-\alpha} \mathbb{E}\left[(\eta - f)^+\right] \right\}, \tag{6}$$

where $(\cdot)^+ = \max\{\cdot, 0\}$. In Pflug (2000), it is shown that the optimal value of η in (6) is the $\text{VaR}_{1-\alpha}[f]$. The second definition shows that the function that defines the CVaR of losses is a convex function (Rockafellar and Uryasev, 2000), which in our specific case leads to the equality

$$\max_{x,z,y(\xi)} \left\{ \text{CVaR}_{1-\alpha}[f(x, z, y(\xi), \xi)] \right\}$$

$$= \max_{x,z,y(\xi),\eta} \left\{ \eta - \frac{1}{1-\alpha} \mathbb{E}\left\{[\eta - f(x, z, y(\xi), \xi)]^+\right\} \right\}. \tag{7}$$

Based on the properties of the CVaR, we define an objective function representing the maximization of a combination of the expected profit and CVaR:

$$\max_{x,z,y(\xi)} \psi[f(x, z, y(\xi), \xi)] := \max_{x,z,y(\xi),\eta} \left\{ \mathbb{E}\left[(1 - \beta) f(x, z, y(\xi), \xi) \right. \right.$$

$$\left. \left. + \beta \left(\eta - \frac{1}{1-\alpha}[\eta - f(x, z, y(\xi), \xi)]^+ \right) \right] \right\}, \tag{8}$$

where $\beta \in [0, 1]$ is a parameter that defines the weights of the expectation and CVaR of f. We are particularly interested in the solution of Problem (3) with (8) defined with $\beta = (0, 1)$ and $\beta = 1$ (maximization of the CVaR of the profit) using an SAA approach.

Detailed analyses on the properties of the CVaR and VaR metrics and their integration into stochastic programming can be found in Pflug (2000), Rockafellar and Uryasev (2000), Rockafellar (2007), and Sarykalin et al. (2008). Risk metrics for multi-stage stochastic programming, such as the expected CVaR can be found in Pflug and Ruszczynski (2005), Shapiro et al. (2009), and Mello and Pagnoncelli (2016). Also, the work in Alonso-Ayuso et al. (2018) provides extensive comparisons on multi-period risk metrics.

4. Sample average approximation

For the generic problem formulation

$$w^* = \max_{x,y(\xi),z \in \mathcal{W}} \{\mathbb{E}_\xi[\phi(x, y(\xi), z, \xi)]\}, \tag{9}$$

where $\mathcal{W} := \{x, y(\xi), z | Ax + Bz \le b; Cz + Dy(\xi) \le d, \text{a.s.}; Fy(\xi) + Gx = h(\xi), \text{a.s.}; x \in \mathbb{R}^{n_1}_+, z \in \mathbb{B}^{n_1}, y(\xi) \in \mathbb{R}^{n_1}_+\}$, we define the sample average approximation problem as (Kleywegt et al., 2001; Shapiro and Homem-De-Mello, 2000):

$$w^*_N = \max_{x,y,z \in \mathcal{W}_N} \left\{ \frac{1}{N} \sum_{n=1}^N \phi(x, y, z, \xi^n) \right\}, \tag{10}$$

for a sample with independent and identically distributed elements ξ^n from the distribution of ξ, and where $\mathcal{W}_N := \{x, y(\xi^n), z | Ax + Bz \le b; Cz + Dy(\xi^n) \le d, \forall \xi^n \in \Xi; Fy(\xi^n) + Gx = h(\xi^n), \forall \xi^n \in \Xi; x \in \mathbb{R}^{n_1}_+, z \in \mathbb{B}^{n_2}, y(\xi^n) \in \mathbb{R}^{n_3}_+, n = 1, \ldots, N\}$, and ϕ stands for ϕ_β.

The optimal value and solution of Problem (9) are w^* and x^*, y^*, z^*, respectively, which are approximated by w^*_N and x^*_N, y^*_N, z^*_N obtained from (10). Note that w^*_N is a random variable that depends on the sample of ξ^n.

The objective function (8) involve the CVaR measure, which may not resemble the generic formulation in (9). However, note that (8) is reformulated as a function of the expectation operator.

Using the SAA methodology, two types of inference statistics for the solutions are calculated: 1) point estimates and confidence intervals for the upper and lower bounds on the optimal objective function value w^*; 2) a point estimates of the upper bound on the gap between w^* and w^*_N.

Table 1

Case 1. Optimization results for the formulation with $\beta = 0$. The results are ordered by the optimal objective function value ($\hat{w}_{N,m}$). $M = 30$, $N = 10$. **Bold** – set of distinct first-stage solutions.

					First-stage variables (aggregated)				
m	$\hat{w}_{N,m}$ (€)	ITER	GAP (%)	T (s)	UT (%)	SUP	SD	SELLC (MW)	BUYC (MW)
23	2,437,900	2	0.0000	1	**100**	**0**	**0**	**155**	**0**
22	2,754,781	2	0.0000	1	**100**	**0**	**0**	**50**	**55**
27	2,791,618	2	0.0000	1	100	0	0	50	55
18	2,807,663	2	0.0000	1	100	0	0	50	55
17	2,808,236	2	0.0000	1	100	0	0	50	55
2	2,828,660	2	0.0000	1	100	0	0	50	55
19	2,881,600	2	0.0000	1	**100**	**0**	**0**	**50**	**110**
20	2,950,385	2	0.0000	1	100	0	0	50	110
3	2,987,171	2	0.0000	1	**100**	**0**	**0**	**0**	**160**
9	3,051,161	2	0.0000	1	100	0	0	0	160
21	3,075,317	2	0.0000	1	100	0	0	0	160
28	3,086,196	2	0.0000	1	100	0	0	0	160
26	3,098,970	2	0.0000	1	100	0	0	0	160
13	3,112,520	2	0.0000	1	100	0	0	0	160
7	3,141,942	2	0.0000	1	100	0	0	0	160
4	3,182,338	2	0.0000	1	100	0	0	0	160
14	3,225,259	2	0.0000	1	100	0	0	0	160
12	3,241,059	2	0.0000	1	100	0	0	0	160
15	3,245,519	2	0.0000	1	100	0	0	0	160
25	3,251,400	2	0.0000	1	100	0	0	0	160
10	3,273,009	2	0.0000	1	100	0	0	0	160
8	3,311,875	2	0.0000	1	100	0	0	0	160
16	3,326,421	2	0.0000	1	100	0	0	0	160
29	3,331,775	2	0.0000	1	100	0	0	0	160
11	3,371,868	2	0.0000	1	**100**	**0**	**0**	**0**	**215**
6	3,525,814	2	0.0000	1	**100**	**0**	**0**	**0**	**265**
30	3,543,351	2	0.0000	1	100	0	0	0	265
5	3,682,847	2	0.0000	1	100	0	0	0	265
1	3,738,467	2	0.0000	1	100	0	0	0	265
24	3,743,889	2	0.0000	1	100	0	0	0	265

m – Optimization replication index, $\hat{w}_{N,m}$ - optimal objective function value for optimization replication m, ITER – number of iterations of the L-Shaped method, GAP – gap between the upper and lower bound within the L-Shaped method, T – elapsed wall-clock time, UT – percentage of up-time of the thermal unit, SUP/SD – number of startups/shutdowns of the thermal unit, SELLC/BUYC – power sold/bought through contracts.

Besides, for the specific application, the solutions are complemented with the corresponding point estimates and confidence intervals for the expected profit and the CVaR of the profit.

To evaluate the inference statistics on the bounds of the optimal objective function value w^*, the following notation is used: a) for the upper bound, N denotes the sample size, n is the index of an element of the sample, M is the number of replications, and m is the replication index; b) for the lower bound, N' denotes the sample size, n' is the index of an element of the sample, T and T' are the number of replications, and t and t' are replication indices. For the upper bound on the gap between w^* and w_N^*, N, n, M, and m are used with the same purpose as in a). Note that the symbols t, t', and T are not related to the symbols t and T used in Appendix A to refer to time periods and time horizon.

4.1. Motivating example

In this section, we illustrate the variance of the optimal objective function values and first-stage solutions obtained using different samples having the same size. The samples and results used in this example are part of the computational experiments described in Section 6. Table 1 outlines the results of 30 optimization replications, each with a sample of size $N = 10$, for the maximization of the expected profit.

The variability of these results can be highlighted with the results of two samples. The first sample ($m = 1$) has optimal objective function value of 3,738,467 and first-stage solution $\{100, 0, 0, 0, 265\}$, whereas for $m = 2$, the optimal objective function value is 2,828,660 with first-stage solution $\{100, 0, 0, 50, 55\}$. This table also shows that the first-stage solution $\{100, 0, 0, 0, 265\}$ corresponds to the higher profits, and the solution with $\{100, 0, 0, 0, 160\}$ has the higher number of occur-

rences. These results do not support a clear-cut decision on the best solution to choose. Obviously, the variance of these results is due to the size of the sample used for the optimization. However, with a small sample and the methodologies presented in this work, it is possible to provide a set of inference statistics and to assess the solutions. This information can help the selection of first-stage solutions.

The impact of the sample size N and the number of replications M on the probability of obtaining an optimal solution to the true problem from the SAA problem is discussed in Kleywegt et al. (2001). In that work, it is claimed that the relation between N and this probability is problem specific, and that beyond a given M, it would be better to increase N.

4.2. Upper bound on the optimal objective function value w^*

The approximation of w^* is done by calculating point estimates of lower and upper bounds on w^*. An upper bound on w^* is defined by the well-known relation $\mathbb{E}[w_N^*] \geq w^*$ (Mak et al., 1999). The value of the $\mathbb{E}[w_N^*]$ is estimated by the statistical estimator $UB_{N,M}$ that is determined by solving M optimization problems (10)

$$\hat{w}_{N,m} = \max_{x,y,z \in \mathcal{W}_{N,m}} \left\{ \frac{1}{N} \sum_{n=1}^{N} \phi(x, y, z, \xi^{n,m}) \right\}, \quad \forall m \in M, \tag{11}$$

and by using $\hat{w}_{N,m}$ to estimate the $UB_{N,M}$ through

$$UB_{N,M} = \frac{1}{M} \sum_{m=1}^{M} \hat{w}_{N,m}. \tag{12}$$

The sample variance estimator is

$$(s_{N,M}^1)^2 = \frac{1}{M(M-1)} \sum_{m=1}^{M} \left(\hat{w}_{N,m} - UB_{N,M} \right)^2. \tag{13}$$

Thus, $UB_{N,M}$ is a statistical estimator of $\mathbb{E}[\hat{w}_N]$ with the $100(1-\alpha')\%$ confidence interval

$$\left[LUB_{N,M}, UUB_{N,M} \right] := \left[UB_{N,M} - t_{M-1,\alpha'} s_{N,M}, UB_{N,M} + t_{M-1,\alpha'} s_{N,M} \right], \tag{14}$$

where $t_{M-1,\alpha'}$ is the critical value from the *t*-distribution.

4.3. Lower bound on the optimal objective function value w^*

We define $LB_{N',m,T}$ as an estimator of a lower bound on w^*, which is associated with a first-stage solution from the optimization replication m denoted by $(\hat{x}_{N,m}, \hat{z}_{N,m})$. The calculation of $LB_{N',m,T}$ is done using T independent samples of size N', as follows

$$\hat{w}_{N',m,t} = \max_{y \in \mathcal{W}_{N',m,t}} \left\{ \frac{1}{N'} \sum_{n'=1}^{N'} \phi(\hat{x}_{N,m}, y, \hat{z}_{N,m}, \xi^{n',t}) \right\}, \quad \forall m \in M, \forall t \in T, \tag{15}$$

$$LB_{N',m,T} = \frac{1}{T} \sum_{t=1}^{T} \hat{w}_{N',m,t}, \quad \forall m \in M. \tag{16}$$

Note that the estimator $LB_{N',m,T}$ is associated with the solution of the optimization replication with index m. To generate a single estimator of the lower bound (from the M bounds $LB_{N',m,T}$), the first-stage solution that corresponds to the maximum value of $LB_{N',m,T}$ is selected and for this solution a new estimator is calculated using T' independent samples of size N':

$$\hat{w}_{N',t'} = \max_{y \in \mathcal{W}_{N',t}} \left\{ \frac{1}{N'} \sum_{n'=1}^{N'} \phi(\hat{x}_N, y, \hat{z}_N, \xi^{n',t}) \right\}, \quad \forall t' \in T', \tag{17}$$

where $(\hat{x}_N, y, \hat{z}_N)$ is the first-stage solution selected. For this first-stage solution, the estimator of the lower bound is given by

$$LB_{N',T'} = \frac{1}{T'} \sum_{t'=1}^{T'} \hat{w}_{N',t'}, \tag{18}$$

and the sample variance of $LB_{N',T'}$ through

$$(s_{N',T'}^2)^2 = \frac{1}{T(T-1)} \sum_{m=1}^{M} \left(\hat{w}_{N',t'} - LB_{N',T'} \right)^2, \tag{19}$$

with the confidence interval defined as

$$\left[LLB_{N',T'}, ULB_{N',T'} \right] := \left[LB_{N',T'} - t_{T-1,\alpha'} s_{N',T'}, LB_{N',T'} + t_{T-1,\alpha'} s_{N',T'} \right]. \tag{20}$$

This approach to calculate the point estimate of the lower bound is based on Linderoth et al. (2006).

4.4. Upper bound on the gap between w^* and $\mathbb{E}_\xi[\phi(\hat{x}_N, y, \hat{z}_N, \xi)]$

In Mak et al. (1999), it is derived a method for assessing the quality of a solution $(\hat{x}_N, \hat{z}_N)$, which defines an estimator of an upper bound on the following gap:

$$gap(\hat{x}_N, \hat{z}_N) := w^* - \mathbb{E}_\xi[\phi(\hat{x}_N, y, \hat{z}_N, \xi)]. \tag{21}$$

This gap represents the difference between the true optimal solution w^* and the true optimal solution $\mathbb{E}_\xi[\phi(\hat{x}_N, y, \hat{z}_N, \xi)]$ for $(\hat{x}_N, \hat{z}_N)$. The upper bound, $\mathcal{G}$, on the $gap(\hat{x}_N, \hat{z}_N)$ is defined as (Mak et al., 1999):

$$\mathcal{G} := \mathbb{E}\left[\max_{x,y,z \in \mathcal{W}_N} \left\{ \frac{1}{N} \sum_{n=1}^{N} \phi(x, y, z, \xi^n) \right\} - \frac{1}{N} \sum_{n=1}^{N} \phi(\hat{x}_N, y, \hat{z}_N, \xi^n) \right]$$

$$\geq gap(\hat{x}_N, \hat{z}_N). \tag{22}$$

$\mathcal{G}$ is calculated using $\mathbb{E}[w_N^*]$ as a valid statistical upper bound on w^*, $\mathbb{E}[w_N^*] \geq w^*$, and $\mathbb{E}[\hat{w}_N]$ as an estimator of $\mathbb{E}_\xi[\phi(\hat{x}_N, y, \hat{z}_N, \xi)]$. In Mak et al. (1999), it is proposed to use the upper limit of the confidence interval of the estimator of $\mathcal{G}$ as an upper bound on the $gap(\hat{x}_N, z_N)$. This upper limit calculation is presented in Procedure 2. Note that this procedure uses the same sample to calculate the two terms on the left hand side of the inequality in (22), and the upper bound on the gap is always positive. Additional details regarding variants of evaluation procedures and convergence analysis of the gap estimator can be found in Bayraksan and Morton (2006) and Mak et al. (1999).

4.5. Implementation

The practical computation of the statistical point estimates and confidence intervals described in Sections 4.2–4.4 is presented in Procedures 1 and 2, and the corresponding diagrams in Appendix B. These two procedures involve two stages denoted as optimization stage and bound estimation stage. In line 3 of Procedures 1 and 2, a risk-neutral or a risk-averse stochastic programming problem is solved. Each of these problems can be solved directly using an MILP solver or by decomposition. We adopt the latter and use the L-Shaped method proposed in the seminal work of Van Slyke and Wets (1969). This choice is supported by a previous successful application of this method in solving large scale problems (Lima et al., 2018). The L-Shaped method is an extension of Benders decomposition (Benders, 1962) for two-stage stochastic programming problems, which is straightforward to describe using problem (23):

$$\max_{x,z,\eta} \quad c^\top x + \bar{c}^\top z + (1-\beta) \sum_{n=1}^{N} \left[p_n Q(x, z, \xi^n) \right]$$

$$+ \beta \left\{ \eta - \frac{1}{1-\alpha} \sum_{n=1}^{N} \left[p_n(\eta - Q(x, z, \xi^n))^+ \right] \right\} \tag{23}$$

$$s.t. \quad Ax + Bz \leq b$$
$$x \in \mathbb{R}_+^{n_1}, z \in \mathbb{B}^{n_2}, \eta \in \mathbb{R},$$

where $Q(x, z, \xi^n)$ is defined as

$$Q(x, z, \xi^n) := \left. \begin{array}{ll} \max_{y_n} & \tilde{c}_n^\top y_n \\ s.t. & Dy_n \leq d - Cz \\ & y_n = h_n - x \\ & y_n \in \mathbb{R}_+^{n_3} \end{array} \right\} \quad \forall n = 1, \dots, N, \tag{24}$$

where the first problem is an MILP problem followed by a collection of LP subproblems that constitutes a sample element-wise decomposition, meaning one subproblem for each element $n = 1, \dots, N$ of the sample. This formulation moves the variable y_n to the subproblems. The L-Shaped method is an iterative method, where in each iteration the following steps are performed:

1. A relaxation of problem (23), called master problem, is solved. This master problem is constructed by outer-approximating the terms involving $Q(x, z, \xi^n)$, and thus, its objective function value provides an upper bound on the objective function value of problem (23). The solution of this master problem is the first-stage variables: (x, z).
2. The subproblems (24) are solved independently for fixed values of the first-stage variables obtained from the master problem.
3. The VaR and CVaR of the second stage profit are evaluated based on the distribution of the second stage profits (the objective function values of the subproblems), and they are used in the next step.
4. The upper and lower bounds are calculated and the method terminates if the gap between these bounds reaches a specified threshold.
5. The dual variables solution of the subproblems are used to build the outer-approximations to be added to the master problem in the next iteration. In the next iteration, these outer-approximations "cut" the solution of the master problem obtained in step 1.

Procedure 1 Proposed implementation of an SAA method. Input: M samples of size N: $\xi^{n,m}$, $n = 1, \dots, N$, $m = 1, \dots, M$; T samples of size N': $\xi^{n',t}$, $n' = 1, \dots, N'$, $t = 1, \dots, T$; T' samples of size N': $\xi^{n',t'}$, $n' = 1, \dots, N'$, $t' = 1, \dots, T'$; and confidence level. Output: first-stage solution, point estimates of bounds on the optimal objective function value w^*, and confidence intervals.

1: **for** $m = 1$ **to** M **do**
2:     Initialize the L-Shaped method with $(\hat{x}_{N,m}^{best}, \hat{z}_{N,m}^{best})$, if this solution is available
3:     Solve using the L-Shaped method each MILP problem with the sample with size N of i.i.d. $\xi^{n,m}$:
$$\hat{w}_{N,m} = \max_{x,y,z \in \mathcal{W}_{N,m}} \left\{ \frac{1}{N} \sum_{n=1}^{N} \phi(x, y, z, \xi^{n,m}) \right\} \quad (25)$$
4:     Let $(\hat{x}_{N,m}, \hat{y}_{N,m}^n, \hat{z}_{N,m})$ be the optimal solution from (25)
5:     **if** $\hat{w}_{N,m} > \hat{w}_{N,m-1}$ **then**
6:         $(\hat{x}_{N,m}^{best}, \hat{z}_{N,m}^{best}) \leftarrow (\hat{x}_{N,m}, \hat{z}_{N,m})$
7:     **end if**
8:     **if** $(\hat{x}_{N,m}, \hat{z}_{N,m})$ is a new solution **then**
9:         **for** $t = 1$ **to** T **do** {Lower bound estimation}
10:             Solve each LP subproblem with the sample with size N' of i.i.d. $\xi^{n',t}$:
$$\hat{w}_{N',m,t} = \max_{y \in \mathcal{W}_{N',m,t}} \left\{ \frac{1}{N'} \sum_{n'=1}^{N'} \phi(\hat{x}_{N,m}, y, \hat{z}_{N,m}, \xi^{n',t}) \right\} \quad (26)$$
11:         **end for**
12:         Evaluate $LB_{N',m,T}$ using (16)
13:         **if** $m = 1$ **then**
14:             $(\hat{x}_N, \hat{z}_N) \leftarrow (\hat{x}_{N,m}, \hat{z}_{N,m})$
15:         **else if** $m > 1$ **and** $LB_{N',m,T} > LB_{N',m-1,T}$ **then**
16:             $(\hat{x}_N, \hat{z}_N) \leftarrow (\hat{x}_{N,m}, \hat{z}_{N,m})$
17:         **end if**
18:     **end if**
19: **end for**
20: Evaluate $UB_{N,M}$ using (12) and the confidence interval using (14)
21: **for** $t' = 1$ **to** T' **do** {Lower bound estimation}
22:     Solve each LP subproblem with the sample with size N' of i.i.d. $\xi^{n',t'}$:
$$\hat{w}_{N',t'} = \max_{y \in \mathcal{W}_{N',t'}} \left\{ \frac{1}{N'} \sum_{n'=1}^{N'} \phi(\hat{x}_N, y, \hat{z}_N, \xi^{n',t'}) \right\} \quad (27)$$
23: **end for**
24: Evaluate $LB_{N',T'}$ using (18) and the confidence interval using (20)
25: **return** $(\hat{x}_N, \hat{z}_N)$, $UB_{N,M}$, $LB_{N',T'}$, and confidence intervals

---

Procedure 2 Proposed implementation of the upper bound on the gap between w^* and $\mathbb{E}_\xi[\phi(\hat{x}_N, y, \hat{z}_N, \xi)]$ (Bayraksan and Morton, 2006; Mak et al., 1999). Input: $(\hat{x}_N, \hat{z}_N)$ from Procedure 1; M samples of size N: $\xi^{n,m}$, $n = 1, \dots, N$, $m = 1, \dots, M$; and confidence level. Output: upper bound on the $gap(\hat{x}_N, \hat{z}_N) = w^* - \mathbb{E}_\xi[\phi(\hat{x}_N, y, \hat{z}_N, \xi]$

1: **for** $m = 1$ **to** M **do**
2:     Initialize the L-Shaped method with $(\hat{x}_N, \hat{z}_N)$ **or** if available, with $(\hat{x}_{N,m}^{best}, \hat{z}_{N,m}^{best})$
3:     Solve using the L-Shaped method each MILP problem with the sample with size N of i.i.d. $\xi^{n,m}$:
$$\hat{w}_{N,m} = \max_{x,y,z \in \mathcal{W}_{N,m}} \left\{ \frac{1}{N} \sum_{n=1}^{N} \phi(x, y, z, \xi^{n,m}) \right\} \quad (28)$$
4:     Let $(\hat{x}_{N,m}, \hat{y}_{N,m}^n, \hat{z}_{N,m})$ be the optimal solution of (28)
5:     **if** $\hat{w}_{N,m} > \hat{w}_{N,m-1}$ **then**
6:         $(\hat{x}_{N,m}^{best}, \hat{z}_{N,m}^{best}) \leftarrow (\hat{x}_{N,m}, \hat{z}_{N,m})$
7:     **end if**
8:     Solve each LP subproblem with the sample with size N of i.i.d. $\xi^{n,m}$:
$$\max_{y \in \mathcal{W}_{N,m}} \left\{ \frac{1}{N} \sum_{n=1}^{N} \phi(\hat{x}_N, y, \hat{z}_N, \xi^{n,m}) \right\} \quad (29)$$
9:     Evaluate
$$\mathcal{G}_{N,m} = \frac{1}{N} \sum_{n=1}^{N} \left[\phi(\hat{x}_{N,m}, \hat{y}_{N,m}^n, \hat{z}_{N,m}, \xi^{n,m}) - \phi(\hat{x}_N, \hat{y}_N^n, \hat{z}_N, \xi^{n,m}) \right] \quad (30)$$
10: **end for**
11: Evaluate the statistical point estimate for the expected gap and sample variance
$$\overline{\mathcal{G}}_{N,M} = \frac{1}{M} \sum_{m=1}^{M} \mathcal{G}_{N,m} \quad (31)$$
$$(s_M)^2 = \frac{1}{M(M-1)} \sum_{m=1}^{M} \left(\mathcal{G}_{N,m} - \overline{\mathcal{G}}_{N,M} \right)^2 \quad (32)$$
12: **return** upper bound $UG_{N,M} = \overline{\mathcal{G}}_{N,M} + t_{M-1,\alpha} s_M$

The outer-approximations are known as optimality cuts. These cuts require that all subproblems in step 2 are feasible, otherwise, a feasibility subproblem needs to be solved for each infeasible subproblem. The optimal values of the dual variables of these feasibility subproblems are then used to build cuts, known as feasibility cuts, to be added to the master problem in the next iteration. In the problems studied in this work, the subproblems are always feasible for all values of the first-stage variables. This is due to the possibility to buy electricity from the pool, which is implemented in the subproblems to satisfy contracts decided in the master problem. An algorithmic description, specific problem formulations, expressions for bounds, and details on handling the variable η and the CVaR calculation can be found in Lima et al. (2018), and the outer-approximations derivations and convergence analysis in Van Slyke and Wets (1969), Birge and Louveaux (2011), and Kall and Mayer (2011).

Clearly, the solution of multiple replications of risk-averse problems in Procedures 1 and 2 is a computationally demanding process, especially for MILP problems. Therefore, implementations of Procedures 1 and 2 should include strategies to reduce their overall computing time. An attractive approach is to parallelize the implementation of these procedures taking advantage of the modeling and hardware ca-

pabilities available. In this work, the parallelization is exploited in the solution of the independent LP subproblems in two places: 1) within the L-Shaped method; and 2) at the lower bound estimation, where for each first-stage solution, N' LP subproblems are solved.

In Procedures 1 and 2, we implement an initialization step that uses the best first-stage solution obtained in a replication $m' < m$ to initialize the L-Shaped method in replication m. This initialization is indicated in line 2 of Procedure 1, where $(\hat{x}_{N,m}^{best}, \hat{z}_{N,m}^{best})$ is the best first-stage solution. We note that this first-stage solution is not available in the first optimization replication ($m = 1$). In contrast, in Procedure 2, by construction, the solution $(\hat{x}_N, \hat{z}_N)$ is available at the beginning of the procedure, thus, the initialization mentioned in line 2 is active from the first replication. The initialization step entails that in the first iteration of the L-Shaped method, the best first-stage solution is assigned to the first-stage solution, instead of finding a solution using the master problem. With this initialization step, the initial lower bound of the L-Shaped method, calculated after solving the LP subproblems, is generally better than the one calculated using the solution of the master problem in the first iteration. Note that the corresponding computational gain is multiplied by the number of optimization replications.

In the bound estimation stage, the bounds described in Sections 4.2–4.4 are evaluated using the first-stage solutions and the objective function values obtained from the optimization stage. In Procedure 1, the lower bound estimation is computationally more demanding than the upper bound estimation. Therefore, we implement another important time saving strategy: the lower bound estimation is only carried out for new distinct first-stage solutions. Note that an alternative would be to perform the lower bound estimation only for ϵ-different first-stage solutions. This alternative would avoid the computing time for the lower bound estimation for slightly different first-stage solutions. In the dis-

cussion of the results, the impact of avoiding the lower bound estimation for repeated solutions on the computing time is analyzed and situations where it is relevant are identified.

The risk-averse formulations involving the CVaR measure add an extra burden to the L-Shaped method and lower bound estimation. The evaluation of the CVaR measure requires the calculation of the conditional average of the $Q(x, z, \xi^n)$ (defined in (24)). To reduce this evaluation time, an efficient approach in the lower bound estimation is implemented, where the evaluation of the CVaR for $\alpha > \alpha'$ re-uses the information of the distribution of the previous CVaR evaluation for α'.

With this overall approach, we provide a full characterization of the solution of our problem, which encompasses the following information:

1. approximation of the optimal solution, in terms of the first-stage variables;
2. point estimate and confidence interval for the upper bound on the true optimal objective function value;
3. point estimate and confidence interval for the lower bound on the true optimal objective function value;
4. point estimate on the upper bound on the gap between the true optimal objective function value and the expected optimal objective function value associated with a first-stage solution;

Therefore, this information provides inference statistics on the solutions that are more informative than the single optimal objective function value and corresponding first-stage variables of a deterministic problem.

In the next section, we describe the sampling procedure to generate the samples of the wind and electricity prices.

5. Sampling

5.1. Sampling the electricity prices

We generate the samples for the electricity prices from an ARIMA model that is fitted to an electricity price time series. Specifically, the samples are drawn by sampling the error term of the ARIMA model; see Conejo et al. (2010) for a discussion on this type of approach. Specific details for the data used are given in Section 6.1.1.

5.2. Sampling the wind speed with a truncated Karhunen–Loève expansion

In this section, we briefly introduce the KLE, and in Section 6, we provide additional information regarding the wind data used. See Maître and Knio (2010, Section 2.1) for a presentation on KLE.

5.2.1. Wind speed model
Assuming that the wind speed $h(t)$ is always strictly positive, we adopt here a log-normal model where $X(t) = \log(h(t))$ is a Gaussian process with mean function $\mu(t)$ and covariance function $C(t, t')$, which are approximated in our application using a sample $(x^i(t))_{i=1}^N$ and the empirical estimation

$$\mu(t) \approx \frac{1}{N} \sum_{i=1}^{N} x^i(t) \tag{33}$$

and $\quad C(t, t') \approx \frac{1}{N-1} \sum_{i=1}^{N} (x^i(t) - \mu(t))(x^i(t') - \mu(t')). \tag{34}$

Given a time interval uniformly discretized $(t_i)_{i=1}^n$ with a resolution $\Delta t = 1$ h, the rest of Section 5.2 is dedicated to the KLE-based approximation of the wind speed model under the form

$$h(t_i) \approx \exp\left(\mu(t_i) + \sum_{i=j}^{r} \sqrt{\lambda_j} W_{ji} \xi_j \right), \tag{35}$$

where $(\xi_i)_{i=1}^r$ are independent random variables drawn according to the standard normal distribution.

5.2.2. Continuous formulation
Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space, with Ω being the sample space, $\mathcal{F}$ a σ-algebra and $\mathcal{P}$ a probability measure. Let $I = [a, b]$ be an interval of $\mathbb{R}$. X is a square-integrable stochastic process that is assumed to be defined on I. The Karhunen–Loève decomposition of the process can be expressed as

$$X(t) = \mu(t) + \sum_{i=1}^{\infty} \sqrt{\lambda_i} V_i(t)\xi_i, \tag{36}$$

where $(V_i)_{i=1}^\infty$ is a Hilbert basis of $L^2(I)$, the random variables $(\xi_i)_{i=1}^\infty$ are mutually independent, with zero mean and unit variance, and $(\lambda_i)_{i=1}^\infty$ are positive constants in a decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$). Under sufficient regularity of the covariance C, the pairs $(\lambda_i, V_i)_{i=1}^\infty$ are solutions of the following Fredholm integral equation of the second kind

$$\int_{t' \in I} C(t, t')V_i(t')\mathrm{d}t' = \lambda_i V_i(t). \tag{37}$$

The random variable ξ_i is defined by

$$\xi_i = \frac{1}{\sqrt{\lambda_i}} \int_{t \in I} V_i(t)(X(t) - \mu(t))\mathrm{d}t. \tag{38}$$

Note that if X is a Gaussian process, then the random variables $(\xi_i)_{i=1}^\infty$ follow the standard Gaussian distribution.

5.2.3. Discretization
Let $(s_i)_{i=0}^n$ be a regular discretization of the interval I such that

$$a = s_0 \leq s_1 \leq \cdots \leq s_n = b, \quad \text{and} \quad s_{i+1} - s_i = \Delta t, \tag{39}$$

and let $(t_i)_{i=1}^n$ be the midpoints $(s_i - s_{i-1})/2$. Evaluating the integral in Eq. (37) with a midpoint-quadrature formula at every t_j yields

$$\sum_{j=1}^{n} C(t_i, t_j)V_i(t_j)\Delta t = \lambda_i V_i(t_i). \tag{40}$$

Let the matrix $K \in \mathbb{R}^{n \times n}$ and the vector $W_i \in \mathbb{R}^n$ be defined by

$$K_{ij} = \Delta t C(t_i, t_j), \quad \text{and} \quad W_{ij} = V_i(t_j). \tag{41}$$

The discretized Fredholm integral equation leads to the following algebraic eigenvalue problem

$$KW_i = \lambda_i W_i. \tag{42}$$

The time discretization followed by a truncation of the sum yield therefore the approximations

$$X(t_i) \approx \mu(t_i) + \sum_{j=1}^{n} \sqrt{\lambda_j} W_{ji} \xi_j \approx \mu(t_i) + \sum_{j=1}^{r} \sqrt{\lambda_j} W_{ji} \xi_j. \tag{43}$$

Given a tolerance ϵ, the constant r is chosen such that

$$\left(\frac{\sum_{i=r+1}^{n} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \right)^{\frac{1}{2}} \leq \epsilon, \tag{44}$$

ensuring that the relative L^2 error between the n-terms and the r-terms approximations is less than ϵ.

6. Numerical experiments

The results presented in this section demonstrate the new features implemented in the proposed procedures and provide a complete set of statistical results to better support the decision-maker. The objectives of the experiments are fourfold:

1. to demonstrate the positive impact of the proposed strategies on the computing time;
2. to analyze the performance of the optimization stage for risk-averse approaches;
3. to compute high-quality solutions for the VPP problem;
4. to assess those solutions using statistic estimators and confidence intervals.

Table 2

Summary of the specifications of the thermal unit in Cases 1 and 2.

Thermal unit specifications	Case 1	Case 2
Initial state	On	Off
Maximum power (MW)	455	55
Fixed generation cost (€ /h)	1000	660
Variable generation cost (€ /MWh)	16.19	25.92
Minimum up-time (h)	8	1
Minimum down-time (h)	8	1

Table 3

Size of the extensive form and problems within the L-Shaped method.

Problem	N	β	NCNST	NVAR	0-1 NVAR
Extensive	10	0.0	21,346	14,298	508
Extensive	5000	0.0	10,081,186	6,720,858	508
Extensive	5000	1.0	10,086,186	6,725,859	508
Master†	10	0.0	1187	859	508
Master†	5000	0.0	1187	859	508
Master†	5000	0.5	6188	5861	508
Master†	5000	1.0	6187	5860	508
Subproblem*	–	–	1345	2353	–

† – Sizes refer to the second iteration of the L-Shaped method. * – Size refers to the dual of (24). NCNST – number of constraints plus objective function; NVAR – number of total variables; 0–1 NVAR – number of binary variables.

The first and second objectives above are related to the first contribution of this work and the third and fourth objectives to the second contribution. In terms of the organization of the results, we discuss the computational performance of Procedure 1 in Section in 6.2, the objective function values in Section 6.3, the first-stage solutions in Section 6.4, and the maximum gaps obtained in Section 6.5.

More specifically, in Procedure 1, we analyze the impact of a) the sample size N on the computing time of the L-Shaped method and the variance of the solutions; b) the sample size N' on the computing time and variance of the solutions; c) the number of replications T and T' on the lower bound estimation. In Procedure 2, we study the influence of the sample size N on the upper bound estimation.

In a supplementary document, detailed experimental results that support the ones presented in this section are provided.

6.1. Description of the case studies

We consider two cases pertaining to the VPP described in Section 2, which we refer to as Cases 1 and 2. In both cases a similar wind farm and a similar pumped-storage hydro unit are considered. However, the specifications of the thermal unit in each case are different in terms of initial state, operational constraints, and costs. Table 2 specifies the differences in terms of capacity and costs between the thermal unit in Cases 1 and 2.

For the same samples of the wind speed and electricity prices, these differences induce a distinct reaction of the VPP. Furthermore, the higher generation cost of the thermal unit in Case 2 leads to a situation where the gap between the electricity price and generation costs is smaller than the gap in Case 1. The results presented in this section show that the two cases are sufficiently distinct to give a broad view on the performance of the methods used in this work.

The solutions of interest for the VPP problem are the optimal first-stage solutions and the performance indices expected profit and the CVaR of the profit for different quantiles. Given the number of elements in each sample, and the fact that in practice, the second-stage variables have to be adjusted to the future realization of the random variables, these variables are less meaningful and thus their values are not reported.

6.1.1. Wind speed and electricity prices data

We consider a time horizon of 168 h, which corresponds to the week of August 25–31, 2014. The ARIMA model uses the same structure and fit of the model proposed in Lima et al. (2018). The electricity price time series has 12 weeks before the week studied, from the Iberian Peninsula electricity market (Iberian Electricity Market, 2015).

The raw data for the wind speed for a specific wind farm location consists of a wind speed ensemble with 51 members obtained from the European Centre for Medium Range Weather Forecasts (ECMWF). We use the wind speed ensemble and the KLE to generate additional samples of the wind speed for the SAA methodology. The electricity prices are considered independent of the wind speed available at the location of the considered VPP, due to the small capacity of the VPP in each case study, which has no market power to influence the electricity prices.

6.1.2. Setup of the parameters of the procedures

The risk-neutral and risk-averse stochastic programming problems are solved with an L-Shaped based method with single optimality cuts. For this method, the stop criteria are a maximum wall-clock time of 10,800 s, a maximum gap between the bounds of $1\times10^{-4}\%$, and a maximum number of iterations of 5000. Note that this gap is related to the bounds on the objective function value within the L-Shaped method for a given sample. These bounds are not related to the bounds described in Sections 4.2 and 4.3.

We perform a sensitivity analysis with Procedure 1 combining the following parameters $M = 30$ (number of optimization replications), $N \in \{10, 50, 100, 500, 5000\}$ (size of the samples used in each optimization replication), $T \in \{10, 30\}$, $T' \in \{10, 30\}$ (number of replications in the lower bound estimation), $N' \in \{5000, 25,000\}$ (size of the samples used in the lower bound estimation).

We consider four combinations of $(\beta, \alpha) \in \{(0, -), (0.5, 0.9), (1, 0.9), (1, (N - 1)/N)\}$ for the optimization replication. The case with $\alpha = (N - 1)/N$ corresponds to the maximization of the worst profit consistently with the definition of the CVaR. In the lower bound estimation, for each first-stage solution we evaluate point estimates and confidence intervals of the CVaR of the profit for $\alpha \in \{0.9, 0.95, (N' - 1)/N'\}$.

For Procedure 2, we set $M = 30$ (number of replications of the optimization) and $N \in \{500, 5000\}$ (size of the samples). All confidence intervals presented correspond to 95%.

A workstation with 40 Intel Xeon CPU E5-2680 v2 @ 2.80 GHz processors, and 125.8 Gb of RAM was used and the solution of the LP subproblems was distributed among the 40 CPUs. The MILP and LP problems were solved with CPLEX 12.7.1.0 using the GAMS/GRID/GUSS capabilities to distribute the solution of the LP subproblems.

6.2. Computational performance

In this subsection, we first present the dimensions of the problems considered and discuss the choice of the L-Shaped method over the direct solution of the extensive form. Next, the benefits of the proposed initialization of the L-Shaped method are shown and then analysis on the overall performance of Procedure 1 is provided.

In Table 3, we show that the size of the extensive form reaches millions of constraints and variables for samples with $N = 5000$, and that the L-Shaped method decomposes the extensive form into one smaller master problem and N smaller subproblems.

We adopt the L-Shaped variant that provided the best performance in Lima et al. (2018), i.e., Algorithm 1, which corresponds to using a single optimality cut for each of the expectation and CVaR operators and calculating VaR based on the solution of the subproblems, rather than using VaR as a first-stage variable. In that work it is shown that for $\beta = 0$, the L-Shaped method with an efficient parallelization solution of the subproblems is one order of magnitude faster than the direct solution of the extensive form. Whereas for $\beta \in \{0.1, 0.5, 0.9, 1.0\}$, $\alpha = 0.9$, sample
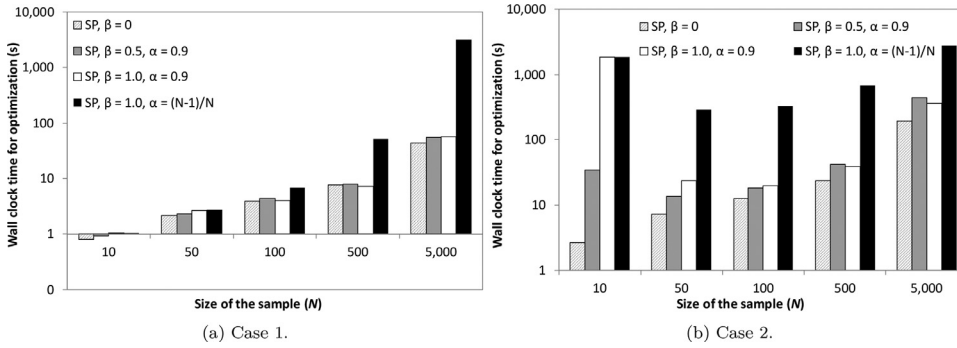
Fig. 1. Average wall-clock time for 30 optimization replications for the combinations of sample sizes and risk parameters studied.

(a) Case 1.   (b) Case 2.

Table 4

Ratio between the total wall-clock time for 30 optimization replications using an initial solution (Tinit) and without using an initial solution (T0) in the L-Shaped method. $M = 30$.

β	α	$N = 500$		$N = 5000$	
		Case 1 Tinit/T0	Case 2 Tinit/T0	Case 1 Tinit/T0	Case 2 Tinit/T0
0	–	0.68	0.82	0.63	0.82
0.5	0.9	0.52	0.75	0.56	0.80
1	0.9	0.56	0.78	0.61	0.81
1	$(N-1)/N$	0.96	1.19	1.06	1.09

sizes of 5100 elements, and maximum wall clock time of 7200 s, the MILP solver applied to the extensive form did not find a feasible solution for four case studies similar to the two considered in this work. We refer the reader to that work for detailed computational results showing the superior performance of the L-Shaped method over solving directly the extensive form.

6.2.1. Initialization of the optimization stage

Table 4 shows the ratios between the total wall-clock time with and without initialization for 30 optimization replications.

These results cover four combinations of risk parameters using two sample sizes – 500 and 5000 elements – over 30 replications. A ratio smaller than one means that the initialization is effective in reducing the total computing time. Overall, the results suggest that the initialization reduces the required computing time. The exception is the formulation with $\beta = 1$ and $\alpha = (N-1)/N$, (last row of Table 4). This peculiar behavior occurs because convergence is not reached for the first-stage solution over the replications, and thus, one solution from one replication is not necessarily a good initial solution for another replication. For the remaining risk metrics, there is convergence for the first-stage variables, and therefore, the initialization has a positive impact on computing time. This is further discussed in Section 6.2.4.

6.2.2. Performance of the optimization stage

Fig. 1 a and b shows for Cases 1 and 2, respectively, the average wall-clock time for 30 optimization replications as a function of the risk parameters and size of the sample. These results show that with a sample with 10 elements, the average wall-clock time for Case 1 is approximately 1 s, independently of the risk parameters. However, this computing time increases with the sample size, with the formulation with $\beta = 1$, $\alpha = (N-1)/N$ exhibiting a steeper increase for an average of 3146 s. For Case 2 the average wall-clock time for the formulations with the CVaR is independent of the size of the sample. The next sub-section focuses on the results of the formulations with $\beta = 1$, $\alpha = (N-1)/N$.

6.2.3. Performance of the optimization stage for $\beta = 1$, $\alpha = (N-1)/N$

The formulation with $\beta = 1$, $\alpha = (N-1)/N$ exhibits some relevant results that are worth discussing. First of all, for the formulations with

the CVaR, the L-Shaped method evaluates the CVaR in each iteration, which requires resources, however, for $\beta = 1$, $\alpha = (N-1)/N$ this evaluation is straightforward. In general, comparing the formulations with $\beta = 0$ and $\beta > 0$, $\alpha < (N-1)/N$ in terms of computing time, we can see that the CVaR specific calculation is not computationally the most demanding. For example, in Fig. 1a for $N = 500$, the computing times for $\beta = 0$ and $\beta > 0$, $\alpha < (N-1)/N$ are similar. On the other hand, for $\beta = 1$, $\alpha = (N-1)/N$ the average number of iterations is 5.7, whereas for $\beta = 1$, $\alpha = 0.9$ it is 2.0. Therefore, the higher computing times obtained with $\beta = 1$, $\alpha = (N-1)/N$ do not arise from the CVaR evaluation, but rather from the number of iterations executed by the L-Shaped method, as discussed below.

For Case 1, for $N = 10$, the L-Shaped method requires on average 2.1 iterations to meet the stopping criteria, while for $N = 5000$ it requires on average 8.6 iterations. Fig. 2 presents the profiles of the bounds in the L-Shaped method for the optimization replications with the lower and higher number of iterations within each optimization replication with $N = 10$ and $N = 5000$. The figures for Case 1 show that for $N = 10$ the number of iterations ranges from 2 to 4, while for $N = 5000$ it ranges from 2 to 21 iterations. However, it is not the size of the sample that demands extra iterations, since for $\beta = 0$ and $N = 5000$, the L-Shaped method requires on average 2.0 iterations to meet the stopping criteria. It is rather the elements within large samples that have an impact on the performance of the L-Shaped method if applied to the maximization of the CVaR of the profit with $\alpha = (N-1)/N$.

For Case 2, Fig. 1b shows that the average wall-clock times for the formulations with the CVaR are independent of the size of the sample. In the specific case of $\beta = 1$, $\alpha = (N-1)/N$, for $N = 10$, the L-Shaped method requires on average 101.2 iterations to meet the stopping criteria, whereas for $N = 5000$ it requires on average 11.4 iterations. These results combined with the results in Fig. 1b indicate that 1) Case 2 forces the L-Shaped method to perform more iterations than Case 1, for the same samples, which is explained by the higher generation costs of the thermal unit in Case 2; and 2) consequently, for the formulations that maximize the worst profit observation, the L-Shaped method requires additional iterations than the other formulations. Fig. 2c and d shows the profiles of the optimization replications with the lower and higher number of iterations obtained with $N = 10$ and $N = 5000$. From these figures, it is also clear that in some replications the L-Shaped method requires few iterations. For example, 10 and 6 for $N = 10$ and $N = 5000$, respectively, while in others it may require a large number – 417 and 25 for $N = 10$ and $N = 5000$, respectively.

Overall, these results show that:

1. the maximization of the CVaR of the profit with $\alpha = (N-1)/N$ forces the L-Shaped method to perform more iterations, by comparison with $\beta = 0$. The reason is that in the former problem, the search is driven by the worst profit, which forces the L-Shaped method to search for the optimal commitment of the thermal unit or contract selection to respond to the low electricity prices that induce the worst case profit. By contrast, the problem with $\beta = 0$ is driven by the max-
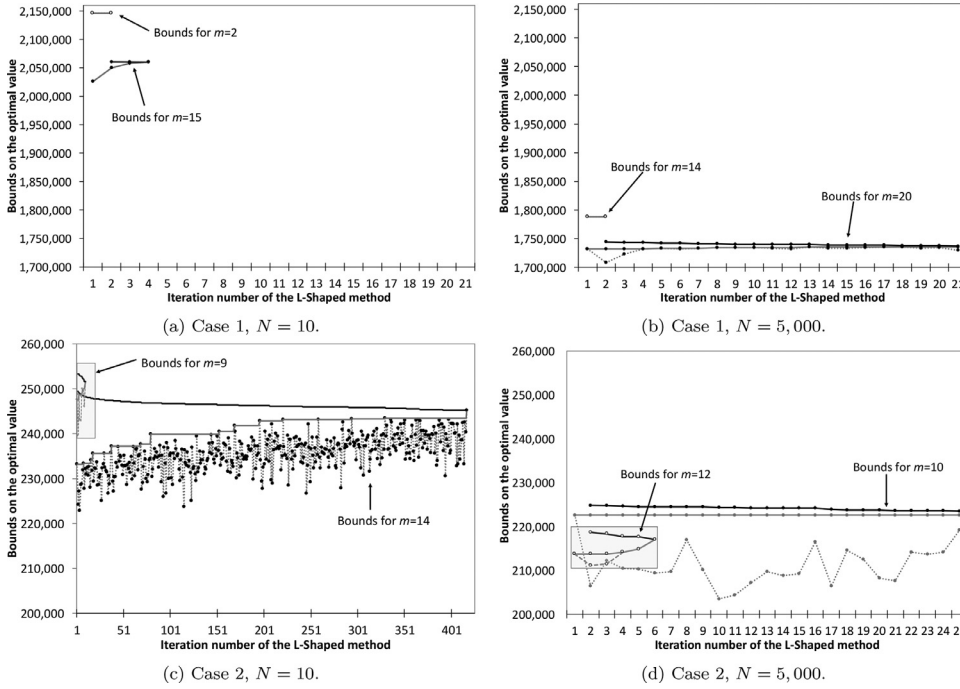
(a) Case 1, $N = 10$.

(b) Case 1, $N = 5,000$.

(c) Case 2, $N = 10$.

(d) Case 2, $N = 5,000$.

Fig. 2. Upper and lower bounds (continuous lines) from the L-Shaped method. The dashed line is the objective function for the first-stage solution. In each sub-figure, there are the two optimization replications with the lower and higher number of iterations. For $M = 30$, $\beta = 1$, and $\alpha = (N-1)/N$.

imization of the expected profit that is less sensitive to low electricity price observations;

2. the size of the samples have an impact on the performance of the L-Shaped method because increasing the sample size increases the likelihood of observations with lower electricity prices;

3. the thermal plant with higher generation costs induces more iterations of the L-Shaped method by comparison with the thermal plant with lower generation costs; and

4. the variability in the number of iterations required by the L-Shaped method in each optimization replication is induced by the different characteristics of the observations with lower electricity prices among the samples. By different characteristics, we mean that across the samples, the lower electricity prices occur on different hours of the week and distinct minimum values.

6.2.4. Combined performance of the optimization and bound estimation stages

Fig. 3 presents the total wall-clock time required by Procedure 1 and by its two main stages: optimization and bound estimation. This figure compares the computational performance obtained with $\beta = 0$ and $\beta = 1$, $\alpha = (N-1)/N$. It also shows the impact on the computing time of increasing the number of replications and the sample size in the lower bound estimation, from $\{T, T'\} = 10$ and $N' = 5000$ to $\{T, T'\} = 30$ and $N' = 25,000$. Using $N' = 25,000$ and $\{T, T'\} = 30$ the overall time increases, by comparison with $N' = 5000$ and $\{T, T'\} = 10$, but there is no influence on the optimization time.

These results provide relevant insights regarding the trade-off between the sample size and the overall time required. We can observe that increasing the sample size in the optimization, the overall time does not monotonically increase. In fact, for $\beta = 0$, there is a minimum overall time for $N = 500$, instead of $N = 10$; see Fig. 3a and c. This behavior is explained by noting that: 1) the lower bound estimation is only performed when a new solution is obtained from an optimization replication; and 2) for $N = 10$, multiple distinct solutions are obtained, whereas for $N = 500$, only one distinct solution is obtained. In Fig. 3, as N increases the time required by the optimization stage increases, but not the lower bound estimation time. For $\beta = 1$, $\alpha = (N-1)/N$, there is also a minimum time that corresponds to $N = 10$ for $N' = 5000$ and

$\{T, T'\} = 10$, and $N = 50$ for $N' = 25,000$ and $\{T, T'\} = 30$; see Fig. 3b and d.

We complement these results by presenting in Table 5 the number of distinct solutions obtained as a function of N and the risk parameters for Case 1.

This table shows that increasing the sample size, the problem converges to one distinct solution for all combinations of risk parameters; except for $\beta = 1$, $\alpha = (N-1)/N$. For $\beta = 1$, $\alpha = (N-1)/N$, increasing the sample size increases the number of distinct solutions. In this specific situation, $\beta = 1$, $\alpha = (N-1)/N$, the objective is to maximize the single worst observation of the profit. The primary difference among the first-stage solutions for $\beta = 1$, $\alpha = (N-1)/N$ is the hourly commitment of the thermal plant; see also additional results available in the supplementary document. This difference suggests that the thermal plant is prone to shutdown and startup to follow the lowest electricity price observation. The samples with more elements have a higher likelihood of lower electricity prices, which from one sample to the other may occur at different hours of the week. Therefore, distinct first-stage solutions emerge among the optimization replications.

The number of distinct solutions obtained with $\beta = 1$, $\alpha = (N-1)/N$ for $N = 500$ also justifies the worst performance of the initialization implemented in the L-Shaped method. Thus, one solution from one optimization replication may not be a good starting point for an optimization replication with a different sample.
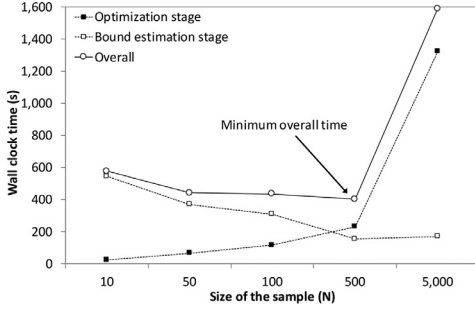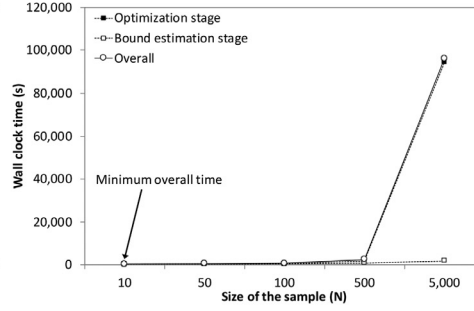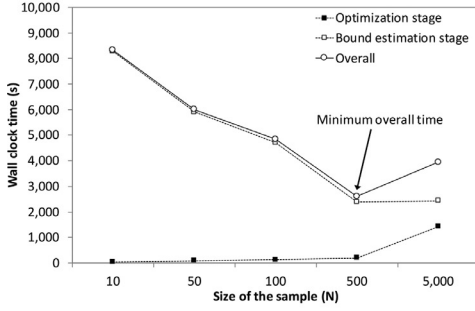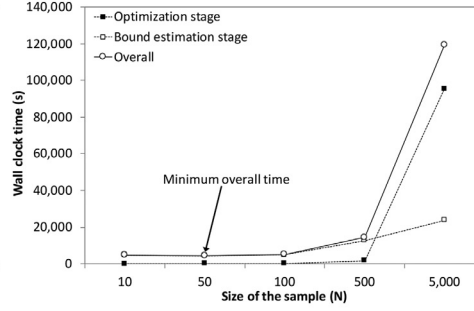
Table 5

Case 1. Number of successful optimization replications and distinct solutions. $M = 30$.

	$\beta = 0$		$\beta = 0.5, \alpha = 0.9$		$\beta = 1, \alpha = 0.9$		$\beta = 1, \alpha = (N-1)/N$	
N	SR	DS	SR	DS	SR	DS	SR	DS
10	30	6	30	6	30	3	30	3
50	30	4	30	4	30	2	30	3
100	30	3	30	3	30	2	30	3
500	30	1	30	1	30	2	30	10
5,000	30	1	30	1	30	1	25	18

SR – number of successful optimization replications that meet the stop criteria (out of 30); DS – number of distinct solutions.

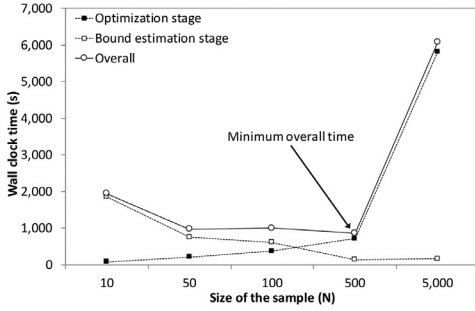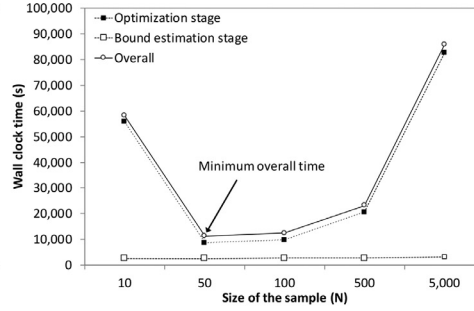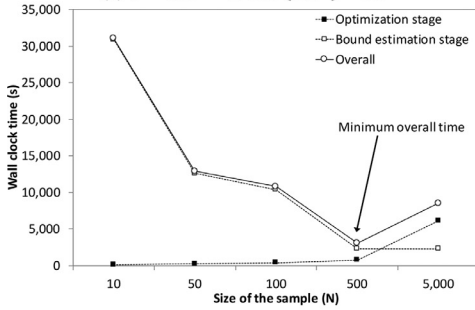**Fig. 3.** Case 1. Wall-clock time as a function of the size of the sample for the optimization. $M = 30$.

(a) $\beta = 0$, $N' = 5,000$, $\{T, T'\} = 10$.

(b) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 5,000$, $\{T, T'\} = 10$.

(c) $\beta = 0$, $N' = 25,000$, $\{T, T'\} = 30$.

(d) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 25,000$, $\{T, T'\} = 30$.

Fig. 4. Case 2. Wall-clock time as a function of the size of the sample for the optimization. $M = 30$.
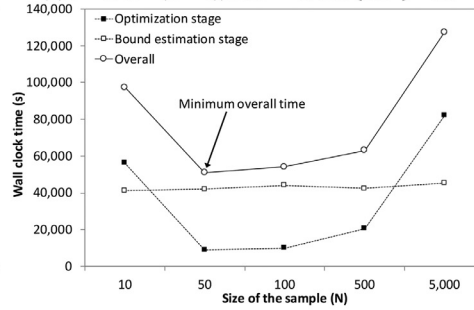
(a) $\beta = 0$, $N' = 5,000$, $\{T, T'\} = 10$.

(b) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 5,000$, $\{T, T'\} = 10$.

(c) $\beta = 0$, $N' = 25,000$, $\{T, T'\} = 30$.

(d) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 25,000$, $\{T, T'\} = 30$.

For Case 2 with $\beta = 0$, the results show that increasing the sample size for the optimization, the number of distinct solutions decreases, which means a decrease in the time for the lower bound estimation; see Fig. 4a and c. However, for $\beta = 1$ with $\alpha \in \{0.9, (N-1)/N\}$, the number of distinct solutions is equal to the number of optimization replications, which means that there is no convergence to a distinct solution; see Table 6. Therefore, the computing time for the lower bound estimation is independent of the sample size for the optimization; see Fig. 4b and d.

6.3. Objective function values and its bounds

This section presents the bounds on the optimal objective function value w^* for the different formulations as a function of the samples size

and the number of replications. Figs. 5 and 6 show the limits of the confidence intervals for the upper and lower bounds on the true optimal objective function value w^* obtained for Case 1 and Case 2, respectively. These figures illustrate the impact of the sample size and number of replications used in the lower bound estimation on the confidence intervals for $\beta = 0$ and $\beta = 1$, $\alpha = (N-1)/N$. In the supplementary results document, the bounds for additional risk parameters are presented.

A relevant result shows that for the formulation that maximizes the expected profit ($\beta = 0$), the point estimate of the upper bound on the true optimal objective function value w^* falls within the confidence interval of the upper bound for all N tested. Increasing the sample size reduces the confidence interval, converging approximately to the center of the confidence intervals. For the formulation with $\beta = 1$, $\alpha = (N-1)/N$, the
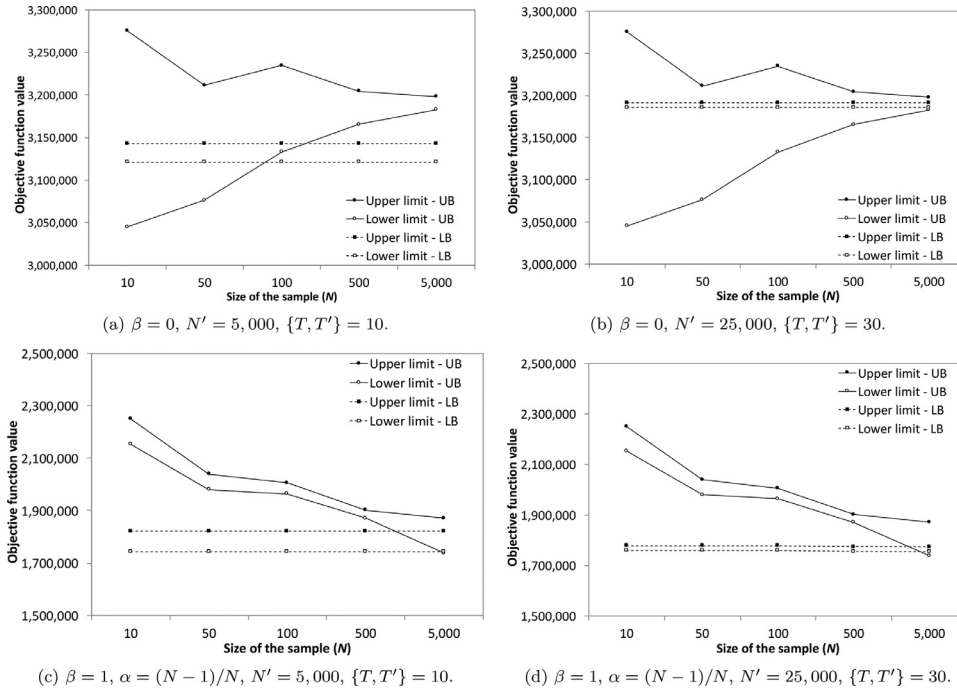
Fig. 5. Case 1. Limits of the confidence intervals on the point estimate of the lower and upper bounds ($LB_{N',T'}$ and $UB_{N,M}$) on the true optimal objective function value (w^*) as a function of risk parameters, N', T, and T'. $M = 30$.

(a) $\beta = 0$, $N' = 5,000$, $\{T, T'\} = 10$.

(b) $\beta = 0$, $N' = 25,000$, $\{T, T'\} = 30$.

(c) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 5,000$, $\{T, T'\} = 10$.

(d) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 25,000$, $\{T, T'\} = 30$.
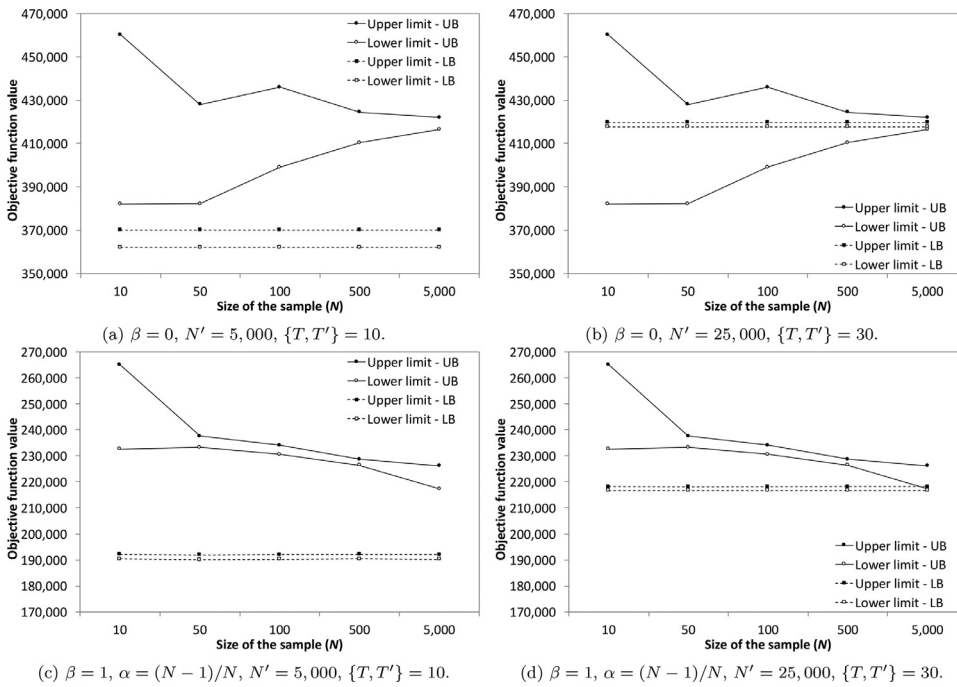

Fig. 6. Case 2. Limits of the confidence intervals on the point estimate of the lower and upper bounds ($LB_{N',T'}$ and $UB_{N,M}$) on the true optimal objective function value (w^*) as a function of risk parameters, N', T, and T'. $M = 30$.

(a) $\beta = 0$, $N' = 5,000$, $\{T, T'\} = 10$.

(b) $\beta = 0$, $N' = 25,000$, $\{T, T'\} = 30$.

(c) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 5,000$, $\{T, T'\} = 10$.

(d) $\beta = 1$, $\alpha = (N-1)/N$, $N' = 25,000$, $\{T, T'\} = 30$.

Table 6

Case 2. Number of successful optimization replications and distinct solutions. $M = 30$.

N	$\beta = 0$		$\beta = 0.5$, $\alpha = 0.9$		$\beta = 1$, $\alpha = 0.9$		$\beta = 1$, $\alpha = (N-1)/N$	
	SR	DS	SR	DS	SR	DS	SR	DS
10	30	26	30	17	28	30	28	30
50	30	10	30	30	30	30	30	30
100	30	8	30	30	30	30	30	30
500	30	1	30	23	30	30	30	30
5,000	30	1	30	30	30	30	27	30

SR – number of successful optimization replications that meet the stop criteria (out of 30); DS – number of distinct solutions.

point estimate of the upper bound overestimates the true optimal objective function value, and the point estimate is not included in all the confidence intervals (for different N). Thereby, increasing the sample size for the optimization decreases the value of the point estimate and range of the confidence intervals.

Figs. 5 and 6 indicate that the confidence interval for the lower bound on the true optimal objective function value is underestimated with the samples with $N' = 5000$, $\{T, T'\} = 10$. For example, this can be seen comparing in Fig. 5a and b the confidence interval for the lower bound represented by the dashed lines. These results indicate that for $T' = 10$ there is a bias in the estimation of the lower bound on the true optimal objective function value. However, as N' is increased to 25,000

Table 7
Case 1. First-stage solutions and point estimates of the lower bound on the true optimal objective function value ($LB_{N',T'}$) with confidence intervals. $N' = 25,000$, $\{M, T, T'\} = 30$.

β	α	N	UT (%)	SUP	SD	SELLC (MW)	BUYC (MW)	$LB_{N',T'}$ (€)	LLB (€)	ULB (€)
									CI, [LLB,ULB]	
			First-stage solutions (aggregated)							
0	–	10	100	0	0	0	160	3,188,599	3,185,869	3,191,329
0	–	50	100	0	0	0	160	3,188,599	3,185,869	3,191,329
0	–	100	100	0	0	0	160	3,188,599	3,185,869	3,191,329
0	–	500	100	0	0	0	160	3,188,599	3,185,869	3,191,329
0	–	5000	100	0	0	0	160	3,188,599	3,185,869	3,191,329
0.5	0.9	10	100	0	0	155	0	2,454,870	2,453,669	2,456,071
0.5	0.9	50	100	0	0	155	0	2,454,870	2,453,669	2,456,071
0.5	0.9	100	100	0	0	155	0	2,454,870	2,453,669	2,456,071
0.5	0.9	500	100	0	0	155	0	2,454,870	2,453,669	2,456,071
0.5	0.9	5000	100	0	0	155	0	2,454,870	2,453,669	2,456,071
1	0.9	10	100	0	0	315	0	2,100,859	2,100,156	2,101,562
1	0.9	50	100	0	0	315	0	2,100,859	2,100,156	2,101,562
1	0.9	100	100	0	0	315	0	2,100,859	2,100,156	2,101,562
1	0.9	500	100	0	0	315	0	2,100,859	2,100,156	2,101,562
1	0.9	5000	100	0	0	315	0	2,100,859	2,100,156	2,101,562
1	$(N-1)/N$	10	100	0	0	315	0	1,768,962	1,759,967	1,777,956
1	$(N-1)/N$	50	100	0	0	315	0	1,768,962	1,759,967	1,777,956
1	$(N-1)/N$	100	100	0	0	315	0	1,768,962	1,759,967	1,777,956
1	$(N-1)/N$	500	95.2	1	1	315	0	1,764,836	1,755,568	1,774,104
1	$(N-1)/N$	5000	95.2	1	1	315	0	1,764,554	1,755,197	1,773,911

CI – Confidence interval, LLB and ULB as in (20), UT – percentage of up-time of the thermal unit, SUP/SD – number of startups/shutdowns of the thermal unit, SELLC/BUYC – power sold/bought through contracts.

and $\{T, T'\}$ to 30, the confidence interval of the lower bound converges with the confidence interval of the upper bound.

Note that for the optimization replications where the L-Shaped method did not meet the stopping criteria, the upper bound from the L-Shaped method is used to evaluate the point estimate on the upper bound on the true optimal solution. This replacement justifies the larger range of the confidence interval for the upper bound for Case 1 with $\beta = 1$, $\alpha = (N-1)/N$, $N = 5000$ in Fig. 5c and d, and for Case 2 with $\beta = 1$, $\alpha = (N-1)/N$, $N \in \{10, 5000\}$ in Fig. 6c and d.

Overall, the convergence of the bounds for $N' = 25,000$, $\{T, T'\} = 30$ indicate that high-quality solutions are obtained, which are further assessed in the next sub-section.

6.4. First-stage solutions

Tables 7 and 8 provide the best first-stage solution and inference statistics resulting from each set of risk parameters and sample size.

The first stage solutions are given in aggregated form to simplify their presentation. Therefore, to distinguish between equal aggregated solutions that correspond to distinct disaggregated solutions, the point estimates of their lower bounds are provided.

For Case 1, for each set of risk parameters $(\beta, \alpha) \in \{(0, -), (0.5, 0.9), (1, 0.9)\}$ there is only one distinct solution, independently of the sample size for the optimization. However, $\beta = 1$, $\alpha = (N-1)/N$ has three distinct solutions, which are identified by the three distinct values of $LB_{N',T'}$, see Table 7.

For Case 2 with $\beta = 0$, the same first-stage solution is obtained independently of the sample size for optimization, while for the other set of risk parameters, within each set there are slight variations on the power bought through contracts. The slight variations in this variable lead to small variations in the lower bound and confidence interval for the point estimate of the lower bound on the true optimal objective function value.

These results explain why the confidence interval on the point estimate for the lower bounds in Figs. 5 and 6 are represented by (almost) horizontal dashed lines.

6.5. Gap between w^* and $\mathbb{E}_\xi\big[\phi(\hat{x}_N, y, \hat{z}_N, \xi)\big]$

We further evaluate each distinct solution using a point estimate on the upper bound on the gap between w^* and $\mathbb{E}_\xi\big[\phi(\hat{x}_N, y, \hat{z}_N, \xi)\big]$, as described in Section 4.4. Tables 9 and 10 show this point estimate for $N = 500$ and $N = 5000$, for Cases 1 and 2, respectively.

For Case 1, a gap of zero is obtained for the solutions with $\beta = 0$ and $\beta = 0.5$, independently of N, and for $N = 5000$ for $\beta = 1$, $\alpha = 0.9$. For the other risk parameters, the maximum relative gap is 1.5%.

For Case 2, the solution obtained for $\beta = 0$ has a gap of zero for both values of N. The remaining gaps are relatively small, as the maximum relative gap is 1.1%.

We also present the wall-clock time required, which consists of the time for the optimization and lower bound estimation. It is clear that the formulations with $\beta = 1$, $\alpha = (N-1)/N$ are more demanding, in particular for $N = 5000$. The limiting step in this procedure is the optimization stage, which in Case 1 with $\beta = 1$, $\alpha = (N-1)/N$ and in Case 2 is much more demanding than the lower bound estimation, as it can be noticed in Figs. 3a, b and 4 a, b. For example, in Case 2 with $N = 500$, the optimization stage requires on average 1216 s and 20,998 s for $\alpha < (N-1)/N$ and $\alpha = (N-1)/N$, respectively, while the lower bound estimation on average only requires 73 s.

For the assessment of multiple solutions of the same problem, Procedure 2 can be improved by noting that the optimization stage is independent of the solution to assess. Therefore, to assess multiple solutions, it is necessary one optimization stage (valid for all solutions) plus one lower bound estimation stage per solution. In this way, and given the computing times for the optimization stage mentioned in the previous paragraph, the computing time required by Procedure 2 could be significantly reduced.

Conceptually, Procedure 1 is oriented to the generation of solutions and specific inference statistics, whereas the objective of Procedure 2 is to assess the solution quality of a given solution. The optimization stage is similar in both procedures, however, the lower bound estimation in Procedure 1 is made for each new solution, whereas in Procedure 2 it is only for the single input solution of the procedure. The common parts between the procedures suggest that they can be merged to take advantage of these similarities to reduce the overall computing time. A merging would require Procedure 1 to be adapted so that during the

Table 8
Case 2. First-stage solutions and point estimates of the lower bound on the true optimal objective function value ($LB_{N',T'}$) with confidence intervals. $N' = 25,000$, $\{M, T, T'\} = 30$.

| | | | First-stage solutions (aggregated) | | | | | | CI, [LLB,ULB] | |
| | | | | | | | | | | |
β	α	N	UT (%)	SUP	SD	SELLC (MW)	BUYC (MW)	$LB_{N',T'}$ (€)	LLB (€)	ULB (€)
0	–	10	100	1	0	0	160	418,652	417,656	419,648
0	–	50	100	1	0	0	160	418,652	417,656	419,648
0	–	100	100	1	0	0	160	418,652	417,656	419,648
0	–	500	100	1	0	0	160	418,652	417,656	419,648
0	–	5000	100	1	0	0	160	418,652	417,656	419,648
0.5	0.9	10	100	1	0	100	50.89	260,050	259,967	260,132
0.5	0.9	50	100	1	0	100	50.81	260,048	259,966	260,130
0.5	0.9	100	100	1	0	100	51.15	260,055	259,971	260,138
0.5	0.9	500	100	1	0	100	52.00	260,061	259,975	260,146
0.5	0.9	5000	100	1	0	100	50.99	260,052	259,969	260,134
1	0.9	10	100	1	0	100	38.24	239,852	239,795	239,910
1	0.9	50	100	1	0	100	38.68	239,869	239,811	239,927
1	0.9	100	100	1	0	100	38.84	239,865	239,807	239,923
1	0.9	500	100	1	0	100	38.64	239,869	239,811	239,927
1	0.9	5000	100	1	0	100	38.64	239,869	239,811	239,927
1	$(N-1)/N$	10	100	1	0	100	38.22	217,371	216,616	218,127
1	$(N-1)/N$	50	100	1	0	100	38.31	217,361	216,610	218,113
1	$(N-1)/N$	100	100	1	0	100	38.00	217,389	216,617	218,161
1	$(N-1)/N$	500	100	1	0	100	38.04	217,387	216,619	218,156
1	$(N-1)/N$	5000	100	1	0	100	38.05	217,387	216,619	218,155

CI – Confidence interval, LLB and ULB as in (20), UT – percentage of up-time of the thermal unit, SUP/SD – number of startups/shutdowns of the thermal unit, SELLC/BUYC – power sold/bought through contracts.

Table 9
Case 1. Distinct first-stage solutions and the upper bound on the absolute gap ($UG_{N,M}$) obtained from Procedure 2. $M = 30$.

| | | First-stage variables (aggregated) | | | | | $N = 500$ | | $N = 5000$ | |
β	α	UT (%)	SUP	SD	SELLC (MW)	BUYC (MW)	$UG_{N,M}$	Time (s)	$UG_{N,M}$	Time (s)
0	–	100	0	0	0	160	0	254	0	3304
0.5	0.9	100	0	0	155	0	0	277	0	3312
1	0.9	100	0	0	315	0	563	268	0	3312
1	$(N-1)/N$	100	0	0	315	0	4921	1665	10,618	93,279
1	$(N-1)/N$	95.2	1	1	315	0	27,188	1680	16,127	92,788
1	$(N-1)/N$	95.2	1	1	315	0	27,977	1667	16,646	92,012

UT – percentage of up-time of the thermal unit, SUP/SD – number of startups/shutdowns of the thermal unit, SELLC/BUYC – power sold/bought through contracts.

Table 10
Case 2. Distinct first-stage solutions and the upper bound on the absolute gap ($UG_{N,M}$) obtained from Procedure 2. $M = 30$.

| | | First-stage variables (aggregated) | | | | | $N = 500$ | | $N = 5000$ | |
β	α	UT (%)	SUP	SD	SELLC (MW)	BUYC (MW)	$UG_{N,M}$	Time (s)	$UG_{N,M}$	Time (s)
0	–	100	1	0	0	160	0	784	0	7488
0.5	0.9	100	1	0	100	50.89	160	1341	43	15,587
0.5	0.9	100	1	0	100	50.81	160	1381	45	15,033
0.5	0.9	100	1	0	100	51.15	160	1365	35	15,165
0.5	0.9	100	1	0	100	52.00	182	1447	24	15,595
0.5	0.9	100	1	0	100	50.99	160	1444	40	15,332
1	0.9	100	1	0	100	38.24	79	1368	35	13,320
1	0.9	100	1	0	100	38.68	56	1333	13	13,286
1	0.9	100	1	0	100	38.84	64	1306	18	13,543
1	0.9	100	1	0	100	38.64	55	1225	13	13,075
1	0.9	100	1	0	100	38.64	55	1198	13	13,000
1	$(N-1)/N$	100	1	0	100	38.22	2412	20,984	1365	82,972
1	$(N-1)/N$	100	1	0	100	38.31	2434	21,362	1351	84,861
1	$(N-1)/N$	100	1	0	100	38.00	2371	21,144	1422	81,950
1	$(N-1)/N$	100	1	0	100	38.04	2377	20,910	1407	82,132
1	$(N-1)/N$	100	1	0	100	38.05	2379	20,945	1403	82,635

UT – percentage of up-time of the thermal unit, SUP/SD – number of startups/shutdowns of the thermal unit, SELLC/BUYC – power sold/bought through contracts.

lower bound estimation, the required information can be extracted to evaluate the upper bound on the gap defined in (22).

7. Conclusions

We presented and applied a methodology based on SAA to generate first-stage solutions and inference statistics for the optimal operation of a VPP. To cope with the computational complexity of the problems addressed, new strategies were developed to reduce the computing time of the optimization replications and innovations on managing the repetition of sampling, optimization, and evaluation of first-stage solutions.

We performed a detailed characterization of the overall solution by providing point estimates and confidence intervals for the main quantities of interest, including a) upper and lower bounds for the true optimal objective function value; and b) an upper bound on the gap between the true optimal objective function value and the optimal objective function value for a given first-stage solution. This methodology is applied to formulations involving a parameterized combination of the expected profit and the CVaR of the profit. The results and discussions focus on the extremes of the parameterization – the risk-neutral and risk-averse solutions.

For specific conditions, we identified a relevant trade-off between the size of the sample used in the optimization and the time spent on the lower bound estimation; specifically, increasing the size of the sample for the optimization reduces the time spent in the lower bound estimation and the overall required time. Two reasons justify this behavior: 1) the lower bound estimation is only performed for new first-stage solutions; and 2) increasing the sample size for the optimization reduces the number of distinct first-stage solutions. There is one exception to this behavior, which occurs for specific instances of the maximization of the CVaR with $\alpha = (N-1)/N$, where each optimization replication provides one distinct first-stage solution.

For the present setup, computational experiments indicated that with the number of replications performed, the five sample sizes tested converge to the same solution or a neighboring solution. These results indicate that the replications with samples of ten elements can generate first-stage solutions that are the same or close to the best solutions generated from samples of 500 elements. However, the variance of the solutions obtained with samples of 10 elements makes it difficult to identify whether the best solution has been obtained.

The inference statistics indicate that some of the optimal first-stage solutions of the SAA problem are relatively close to the true optimal solution.

Although the definition of the sample size and the number of replications that guarantee specific bounds or convergence is problem dependent, some practical insights are in order. As general guidelines, given limited time and computational resources, the first issue to address is the sample size for the optimization. On the one hand, if the optimization replications are relatively efficient, then larger samples should be used to reduce the variance of the solutions, eventually avoiding the lower bound estimation for repeated solutions. On the other hand, more difficult optimization problems may require smaller samples for the optimization, which increase the variance. However, our results indicate that among the solutions found, there is a high probability of identifying an excellent first-stage solution that can be further evaluated with the lower bound estimation.

Future work can evolve in two directions: 1) a comparison between risk-averse stochastic programming and robust optimization; and 2) the study of alternative risk metrics. Regarding the comparison, the first-stage solutions and point estimates of the expected profit and the CVaR of the profit obtained in this work can be contrasted with the solutions from a robust optimization approach. Additionally, it would be relevant to explore alternative risk metrics, such as stochastic dominance (Escudero and Monge, 2018), and to assess the quality of the solutions and the overall SAA methodology performance with alternative risk metrics.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejco.2021.100005.

References

Alonso-Ayuso, A., Escudero, L.F., Guignard, M., Weintraub, A., 2018. Risk management for forestry planning under uncertainty in demand and prices. Eur. J. Oper. Res. 267, 1051–1074.
Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent measures of risk. Math. Finance 9, 203–228.
, 1997. In: Awerbuch, S., Preston, A. (Eds.), The Virtual Utility: Accounting, Technology & Competitive Aspects of the Emerging Industry. Springer US.
Baringo, A., Baringo, L., 2017. A stochastic adaptive robust optimization approach for the offering strategy of a virtual power plant. IEEE Trans. Power Syst. 32, 3492–3504.
Bayraksan, G., Morton, D.P., 2006. Assessing solution quality in stochastic programs. Math. Program. 108, 495–514. Ser. B
Benders, J.F., 1962. Partitioning procedures for solving mixed-variables programming problems. Numer. Math. 4, 238–252. doi:10.1007/BF01386316.
Birge, J.R., Louveaux, F., 2011. Introduction to Stochastic Programming. Springer.
Conejo, A.J., Carrión, M., Morales, J.M., 2010. Decision Making Under Uncertainty in Electricity Markets. Springer US.
Dabbagh, S.R., Sheikh-El-Eslami, M.K., 2016. Risk assessment of virtual power plants offering in energy and reserve markets. IEEE Trans. Power Syst. 31, 3572–3582.
Dupačová, J., Gröwe-Kuska, N., Römisch, W., 2003. Scenario reduction in stochastic programming – an approach using probability metrics. Math. Program. 95, 493–511.
Escudero, L.F., Monge, J.F., 2018. On capacity expansion planning under strategic and operational uncertainties based on stochastic dominance risk averse management. Comput. Manag. Sci. 15, 479–500.
Iberian Electricity Market, 2015. www.omie.es.
Jansen, J.C., van der Welle, A., Nieuwenhout, F., 2008. Deliverable D3.2.4: The Virtual Power Plant Concept From an Economic perspective: Updated Final Report. Technical Report. Energy Research Centre of the Netherlands (ECN). http://www.fenix-project.org
Kall, P., Mayer, J., 2011. Stochastic Linear Programming. Springer US.
Kardakos, E.G., Simoglou, C.K., Bakirtzis, A.G., 2016. Optimal offering strategy of a virtual power plant: a stochastic bi-level approach. IEEE Trans. Smart Grid 7, 794–806.
Kleywegt, A., Shapiro, A., Homem-De-Mello, T., 2001. The sample average approximation method for stochastic discrete optimization. SIAM J. Optim. 12, 479–502.
Kozmík, V., Morton, D.P., 2015. Evaluating policies in risk-averse multi-stage stochastic programming. Math. Program. 152, 275–300.
Lima, R.M., Conejo, A.J., Langodan, S., Hoteit, I., Knio, O.M., 2018. Risk-averse formulations and methods for a virtual power plant. Comput. Oper. Res. 96, 350–373.
Lima, R.M., Novais, A.Q., Conejo, A.J., 2015. Weekly self-scheduling, forward contracting, and pool involvement for an electricity producer. An adaptive robust optimization approach. Eur. J. Oper. Res. 240, 457–475.
Linderoth, J., Shapiro, A., Wright, S., 2006. The empirical behavior of sampling methods for stochastic programming. Ann. Oper. Res. 142, 215–241.
Maître, O.L., Knio, O.M., 2010. Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics. Springer Science & Business Media.
Mak, W., Morton, D., Wood, R., 1999. Monte carlo bounding techniques for determining solution quality in stochastic programs. Oper. Res. Lett. 24, 47–56.
Homem-de Mello, T., Bayraksan, G., 2014. Monte Carlo sampling-based methods for stochastic optimization. Surv. Oper. Res. Manag. Sci. 19, 56–85.
Mello, T.H.-d., Pagnoncelli, B.K., 2016. Risk aversion in multistage stochastic programming: a modeling and algorithmic perspective. Eur. J. Oper. Res. 249, 188–199.
Moazeni, S., Powell, W.B., Hajimiragha, A.H., 2015. Mean-conditional value-at-risk optimal energy storage operation in the presence of transaction costs. IEEE Trans. Power Syst. 30, 1222–1232.
Moghaddam, I.G., Nick, M., Fallahi, F., Sanei, M., Mortazavi, S., 2013. Risk-averse profit-based optimal operation strategy of a combined wind farm-cascade hydro system in an electricity market. Renew. Energy 55, 252–259.
Morales, J.M., Conejo, A.J., Madsen, H., Pinson, P., Zugno, M., 2014. Integrating Renewables in Electricity Markets. Springer US.

Pandžić, H., Kuzle, I., Capuder, T., 2013. Virtual power plant mid-term dispatch optimization. Appl. Energy 101, 134–141.

Pandžić, H., Morales, J.M., Conejo, A.J., Kuzle, I., 2013. Offering model for a virtual power plant based on stochastic programming. Appl. Energy 105, 282–292.

Pereira, M., Pinto, L., 1991. Multistage stochastic optimization applied to energy planning. Math. Program. 52, 359–375.

Pflug, C., 2000. Probabilistic Constrained Optimization: Methodology and Applications. Springer US, pp. 272–281.

Pflug, G., Ruszczynski, A., 2005. Measuring risk for income streams. Comput. Optim. Appl. 32, 161–178.

Pflug, G.C., Römisch, W., 2007. Modeling, Measuring and Managing Risk. World Scientific Publishing Co. Pte. Ltd..

Philpott, A.B., de Matos, V.L., 2012. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. Eur. J. Oper. Res. 218, 470–483.

Pudjianto, D., Ramsay, C., Strbac, G., 2007. Virtual power plant and system integration of distributed energy resources. IET Renew. Power Gen. 1, 10–16.

Rahimiyan, M., Baringo, L., 2016. Strategic bidding for a virtual power plant in the day-ahead and real-time markets: a price-taker robust optimization approach. IEEE Trans. Power Syst. 31, 2676–2687.

Rahmani-Dabbagh, S., Sheikh-El-Eslami, M.K., 2016. A profit sharing scheme for distributed energy resources integrated into a virtual power plant. Appl. Energy 184, 313–328.

Rockafellar, R.T., 2007. Coherent approaches to risk in optimization under uncertainty. INFORMS 38–61.

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. J. Risk 2, 21–41.

Sarykalin, S., Serraino, G., Uryasev, S., 2008. Value-at-risk vs. conditional value-at-risk in risk management and optimization. INFORMS 270–294.

Shabanzadeh, M., Sheikh-El-Eslami, M.-K., Haghifam, M.R., 2015. The design of a risk-hedging tool for virtual power plants via robust optimization approach. Appl. Energy 155, 766–777.

Shabanzadeh, M., Sheikh-El-Eslami, M.-K., Haghifam, M.R., 2016. A medium-term coalition-forming model of heterogeneous DERs for a commercial virtual power plant. Appl. Energy 169, 663–681.

Shang, D., Kuzmenko, V., Uryasev, S., 2018. Cash flow matching with risks controlled by buffered probability of exceedance and conditional value-at-risk. Ann. Oper. Res. 260, 501–514.

Shapiro, A., 2009. On a time consistency concept in risk averse multistage stochastic programming. Oper. Res. Lett. 37, 143–147.

Shapiro, A., 2011. Analysis of stochastic dual dynamic programming method. Eur. J. Oper. Res. 209, 63–72.

Shapiro, A., Dentcheva, D., Ruszczyński, A., 2009. Lectures on Stochastic Programming: Modeling and Theory. In: MPS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, pp. 253–332.

Shapiro, A., Homem-De-Mello, T., 2000. On the rate of convergence of optimal solutions of monte carlo approximations of stochastic programs. SIAM J. Optim. 11, 70–86.

Shapiro, A., Tekaya, W., Costa, J.P.d., Soares, M.P., 2013. Risk neutral and risk averse stochastic dual dynamic programming method. Eur. J. Oper. Res. 224, 375–391.

Shayegan-Rad, A., Badri, A., Zangeneh, A., 2017. Day-ahead scheduling of virtual power plant in joint energy and regulation reserve markets under uncertainties. Energy 121, 114–125.

Tajeddini, M.A., Rahimi-Kian, A., Soroudi, A., 2014. Risk averse optimal operation of a virtual power plant using two stage stochastic programming. Energy 73, 958–967.

Van Slyke, R., Wets, R., 1969. L-shaped linear programs with applications to optimal control and stochastic programming. SIAM J. Appl. Math. 17, 638–663.

Verweij, B., Ahmed, S., Kleywegt, A.J., Nemhauser, G., Shapiro, A., 2003. The sample average approximation method applied to stochastic routing problems: a computational study. Comput. Optim. Appl. 24, 289–333.

Wang, W., 2007. Sample Average Approximation of Risk-Averse Stochastic Programs. H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology Ph.D. thesis.

Wang, W., Ahmed, S., 2008. Sample average approximation of expected value constrained stochastic programs. Oper. Res. Lett. 36, 515–519.

Zamani, A.G., Zakariazadeh, A., Jadid, S., 2016. Day-ahead resource scheduling of a renewable energy based virtual power plant. Appl. Energy 169, 324–340.