2019

Observational Data-Based Quality Assessment of Scenario Generation for Stochastic Programs

Didem Sari Ay
Sarah M. Ryan

Observational Data-Based Quality Assessment of Scenario Generation for Stochastic Programs

Didem Sarı Ay

Industrial Engineering

Bartın University

Bartın, Turkey, 74100

Sarah M. Ryan*

Industrial & Manufacturing Systems Engineering

Iowa State University

Ames, IA 50011-2164

*Corresponding Author: smryan@iastate.edu, Phone (+1) 515-294-4347 ORCID 0000-0001-5903-1432

Abstract

In minimization problems with uncertain parameters, cost savings can be achieved by solving stochastic programming (SP) formulations instead of using expected parameter values in a deterministic formulation. To obtain such savings, it is crucial to employ scenarios of high quality. An appealing way to assess the quality of scenarios produced by a given method is to conduct a re-enactment of historical instances in which the scenarios produced are used when solving the SP problem and the costs are assessed under the observed values of the uncertain parameters. Such studies are computationally very demanding. We propose two approaches for assessment of scenario generation methods using past instances that do not require solving SP instances. Instead of comparing scenarios to observations directly, these approaches consider the impact of each scenario in the SP problem. The methods are tested in simulation studies of server location and unit commitment, and then demonstrated in a case study of unit commitment with uncertain variable renewable energy generation.

Keywords: Stochastic programming, Scenario generation method assessment, Scenario quality

1 Introduction

The quality of a solution obtained from solving a stochastic program (SP) depends strongly on the quality of the set of scenarios employed to represent the joint distribution of the uncertain parameters. The solution quality is judged conceptually by comparing the expected cost of the solution with the minimal expected cost, where both expectations are taken with respect to the underlying “true” parameter distribution. In practice, this true distribution may not be known exactly and, even if it were, the optimization problem to find the minimal expected cost is usually not solvable because the true distribution is continuous or has too many supporting points for computational tractability. Nevertheless, this concept of solution quality has led to several rigorous and useful approaches to assessing solutions and, by extension, the scenario sets used to obtain them. These include statistical methods based on sampling from the true distribution (Bayraksan and Morton 2006) as well as heuristics for testing in-sample and out-of-sample solution stability (Kaut and Wallace 2007).

The concept of solution stability also motivated the development of widely-used scenario reduction methods based on probability metrics (Dupacova et al. 2003; Heitsch and Römisch 2003). These methods posit the existence and knowledge of a true discrete distribution, which typically has a high-cardinality supporting set, and view a reduced scenario set as an approximate distribution having fewer supporting points. They rely on results linking the differences among expected costs of solutions with the distances among discrete distributions for the parameters. Loosely, for a given cardinality of the approximate supporting set, an upper bound on the distance from the optimal expected cost to the expected cost of a solution to the approximate problem is minimized by optimizing a certain distance metric between the approximate and the true distribution. Thus, the quality of a scenario set is often formulated in terms of how well it approximates a (conceptual) true distribution. Because the computational burden of solving the SP is related to the number of scenarios employed, scenario reduction and solution assessment methods focus on identifying a minimal set of scenarios that yield a high quality solution.

Two recent trends have inspired efforts to explore alternative approaches for generating, reducing and evaluating scenario sets. One is the impetus to allow observational data to drive optimization modeling directly (Feng and Ryan 2016). The second is a recognition or intuition that, mathematical stability theorems notwithstanding, the choice of scenarios to include in a stochastic program should somehow explicitly account for the impact of those scenarios on the solutions obtained (Papavasiliou and Oren 2013) or the costs they incur (Bruninx et al. 2014; Morales et al. 2009). Our goal in this paper is to combine both of these notions in computationally efficient methods to assess the quality of scenario generation methods. For simplicity, we focus on two-stage stochastic programs. To employ observational data, we focus on applications in which multiple similar instances of a problem are to be solved and data are available for a collection of past instances. These data include values of fixed parameters, information that could have been used to generate scenarios for parameters that were uncertain at the time of solution, and observed values of those parameters that were revealed later. An emphasis on stochastic mixed-integer programs motivates the need for computational efficiency.

Stochastic unit commitment in the electric power industry motivated this work. Here, a system operator would generate scenario time series for load and variable renewable generation on the day ahead of the target day for use in optimizing a daily on-off schedule of thermal generators, and then dispatch those generators during the target day for the realized load and variable generation amounts. A set of historical days supplies the past instances. We also test our methods on stochastic server location, where server locations are to be chosen before knowing which potential clients will materialize and the past instances could represent various geographical regions. In our simulation studies, we employ a "true" distribution as a source from which to sample observed values of the uncertain parameters. But in our case study on stochastic unit commitment, we purposely avoid defining such a distribution and simply rely on actual past observations.

We define a scenario generation method (SGM) as any combination of stochastic process modeling, approximation, sampling and reduction techniques that results in a set of probabilistic scenarios based on the information available at the time the SP is to be solved. We assume that the application of a given SGM can be re-enacted over our collection of past instances (Staid et al. 2017). In this way, a SGM is identified interchangeably with the sets of scenarios it would have produced, one set for each instance. When comparing SGMs we assume each that, for each instance, each SGM produces a scenario set with the same specified cardinality. One attractive characteristic of a SGM is *reliability*, defined as correspondence between the scenario probabilities and the relative frequencies with which corresponding parameter values are observed. Reliability roughly corresponds to a small distance between the scenario distribution and the empirical distribution of observations. To reflect this correspondence, we use a statistical metric for reliability constructed from mass transportation distances (Sari et al. 2016; Sari and Ryan 2016).

The more important characteristic we aim to assess is *quality*, which describes how well the generated scenario sets perform in our collection of instances. We claim that high quality is reflected in low average cost incurred by, repeatedly over our collection of instances, (i) applying the SGM, (ii) solving the resulting stochastic program, (iii) implementing the first stage decision, and (iv) taking optimal recourse to the observed values. Because step (ii) of the above process may be time-consuming, we aim to develop SGM quality assessment methods that circumvent it and rely, instead, on solving single-scenario sub-instances. In this paper we propose a generic approach wherein for each instance, a single-scenario version of the SP is solved to find a candidate first-stage solution. Then, for each scenario as well as the observation, the second-stage solution is optimized assuming the candidate solution has been implemented, and the total cost for the scenario is computed. Reliability assessment is then applied to these costs. Variants of this approach differ according to whether the expected value (EV) scenario, perfect information (PI, i.e., the observation), or a randomly selected (RS) scenario is used to find the candidate solution.

Simulation studies demonstrate that reliability of SGMs can be assessed accurately by the EV-based method. The stochastic unit commitment case study indicates that the PI- and RS-based methods can be used to distinguish between higher and lower quality SGMs, as have been identified by re-enactment (Sari and Ryan 2017).

In Section 2 we place this work in the context of a brief literature review. Section 3 provides more detail on re-enactment and reviews our motivating applications. In Section 4 we present our proposed generic approach for SGM quality assessment. The EV- and PI-based variants are presented along with simulation studies in Sections 5 and 6, respectively. In Section 7 we describe a stochastic unit commitment case study showing the results of quality assessment for wind power scenarios generated by two different SGMs, including variants within each method. Finally, we conclude in Section 8 with a brief summary and discussion of further research directions.

2 Literature Review

Kaut and Wallace (2007) discussed and formulated important properties that a scenario generation method should possess to be usable for a given decision model. They defined the optimality gap as the difference between the objective function values, assessed using the true distribution, at the optimal solutions of the true and the approximated problems. They observed that it is impossible to test the optimality gap in most practical problems because it requires solving the optimization problem with the true distribution, which may be unknown and/or intractable. As proxies, they defined two stability measures. In-sample stability exists if different approximate distributions (i.e., scenario sets) produce solutions with similar expected costs with respect to the approximate distributions. Out-of-sample stability requires the true expected cost of the solutions produced by the alternative scenario sets to be similar. While out-of-sample stability assessment requires knowledge of the true distribution, it does not require optimizing with it.

The computational time available for solving SP models might force the use of a smaller number of scenarios. Thus, scenario reduction techniques, aimed at keeping most of the stochastic information embedded in the scenarios, are frequently used to trim the number of scenarios included. Scenario reduction concepts are discussed extensively by Dupacova et al. (2003) and Heitsch and Römisch (2003). A stability approach led to the commonly used forward selection and backward reduction heuristics. An upper bound on the distance between the optimal value of the problem with the reduced scenario set and the optimal value of the solution to the original problem is minimized if the scenario sets are sufficiently close in terms of the probability distance (Dupacova et al. 2003). The most common probability distance used for stochastic optimization problems is the Monge-Kantorovich (mass transportation) distance (Rachev 1991). For two-stage problems, Kantorovich distances are used to derive several heuristics for scenario reduction, including forward selection and backward reduction. A reduced number of scenarios that best retains the essential features of a given original scenario set according to a probability metric can be obtained with these algorithms (Heitsch and Römisch 2003; Heitsch and Römisch 2007). Reduction is based on the norm of the difference between pairs of random vectors. The effect of scenarios on optimal solutions is not addressed directly.

In a recent study on scenario assessment in the context of power system planning, Pinson and Girard (2012) discussed statistical metrics for assessing the reliability of equally likely wind power scenarios; however, they did not examine the scenarios' performance in a SP problem. Sari et al. (2016) modified these statistical evaluation metrics for assessing unequally likely wind power scenarios for use in stochastic unit commitment (SUC), including a mass transportation distance (MTD) rank histogram. Other metrics proposed in that study were specific to the SUC problem.

In recent power system planning studies, researchers have devised scenario reduction techniques based on the optimal objective function values of single-scenario problems. In these approaches, forward selection or backward reduction heuristics are modified to account for the impact of each single scenario realization on the objective function of the stochastic problem. Some numerical evidence indicates that the new scenario reduction procedures outperform the traditional ones. Morales et al. (2009) applied such a scenario reduction technique and compared the results with forward selection. The reduced set of scenarios that was obtained by the proposed technique gives more similar results to those of the original set of scenarios in the SP than does the reduced set of scenarios that was obtained by the existing scenario reduction approach. The superiority of the new approach was illustrated by two different two-stage stochastic problems in the electricity market solved by the producer and the retailer. A scenario reduction method of Bruninx et al. (2014) depends on the objective value of the single scenario equivalent of the stochastic problem. Their approach is similar to that of Morales et al.; however, Bruninx et al. do not fix the first stage decision variables whereas Morales et al. compute the cost of the single-scenario equivalent problem with first-stage decision variables fixed to values obtained by solving the expected value problem. Similarly, a heuristic scenario reduction method that selects scenarios based on their cost and reliability impacts is presented by Feng and Ryan (2016). In SUC, they found that fewer load imbalances result from the proposed reduction technique, which clusters scenarios according to their impact on solutions and then applies the fast forward selection heuristic. These approaches are reminiscent of importance sampling, which inspired a scenario selection procedure developed by Papavasiliou and Oren (2013). They select uncertain scenarios for SUC on the basis of their likelihood of occurrence and the severity of their impact on operating costs.

Our proposed scenario assessment approach is inspired by the recent scenario reduction techniques. To evaluate the scenarios we employ a reliability metric that is motivated by mass transportation distances (Sari et al. 2016). Our methodology also accounts for the impact of each single scenario realization while the assessment relies on MTD rank histograms. We demonstrate the proposed methodology in the context of unit commitment and server location problems. Because the interest in stochastic optimization-based unit commitment grown rapidly in the past several years due to deepening penetration of renewable energy (Bakirtzis et al. 2014; Bruninx et al. 2016a; Bruninx et al. 2016b; Du et al. 2018; Wu and Shahidehpour 2014; Zheng et al. 2015), we focus on the SUC problem as our case study.

3 Scenario quality assessment by historical re-enactment

A good scenario generation method (SGM) should result in low costs in historical simulation over a long sequence of instances. The formalization of scenario quality assessment by historical simulation and the proposed approach will be explained step by step on an abstract form of the general two-stage SP with fixed recourse. A generic two-stage SP is formulated as (Birge and Louveaux 1997):

$$(\mathbb{P}) \qquad \min_{x} \quad c^T x + \mathrm{E}_{\xi} Q\left(x, \xi(\omega)\right) \tag{1}$$

$$\text{s.t.} \quad Ax = b, \tag{2}$$

$$x \in X, \tag{3}$$

where

$$Q\left(x,\xi(\omega)\right)=\min_{y}\ \left\{q(\omega)^{T}\ y \mid \mathrm{T}(\omega)x+Wy=h(\omega),\ y\in Y\right\} \tag{4}$$

The first stage decisions, x, must be taken without full information on random events, $\omega \in \Omega$. The random vector, ξ, composes the parameters of the second-stage problem, $\xi(\omega)=\left(q(\omega),h(\omega),\mathrm{T}(\omega)\right)$. The second stage decisions, denoted by $\boldsymbol{y}$, are taken after a realization of $\xi(\omega)$ becomes known. Either of the feasible sets X and $\boldsymbol{Y}$ may include integer restrictions. At the first stage, optimization is achieved by minimizing the cost of the first-stage decisions, $\boldsymbol{c}^{T}\boldsymbol{x}$, plus the expected cost of optimal second-stage decisions. When the uncertain data are revealed, optimal second-stage costs are obtained by minimizing $Q\left(x,\xi(\omega)\right)$ with respect to $\boldsymbol{y}$. We restrict attention to fixed recourse models with deterministic W.

Using a set of past instances, an appealing way to assess scenario quality is as follows: For each instance in the past set, generate scenarios using historical data available up to that time and employ them in the SP problem. Simulate the implementation of the first-stage decisions, followed by the second-stage decisions optimized according to the observational data for that instance. The historical instance of $\mathbb{P}$ for $d\in\{1,2,...,D\}$ is:

$$\begin{aligned}
\left(\mathbb{P}^{d}\right) \qquad & \min_{x} \quad \left(c^{d}\right)^{T} x+\mathrm{E}_{\boldsymbol{\xi}^{d}}Q\left(x,\xi^{d}(\omega)\right)\\
& \text{s.t.} \quad A^{d}x=b^{d},\\
& \qquad\quad x\in X
\end{aligned}$$

We assume we have a corresponding set of historical observations $\left\{\left(q_{o}^{d},h_{o}^{d},T_{o}^{d}\right)\right\}_{d=1}^{D}$ and scenario sets $\left\{\left(q_{s}^{dk},h_{s}^{dk},T_{s}^{dk}\right),\ s\in \mathrm{S}_{d}\right\}_{d=1}^{D}$ generated by SGM k under assessment, along with the set of corresponding probabilities, $\left\{p_s^{dk}, s \in \mathrm{S}_{dk}\right\}_{d=1}^{D}$ where $0 \le p_s^{dk} \le 1, \sum_{s \in S_{dk}} p_s^{dk} = 1$ for each $k = 1, 2, ..., K$. This produces a collection of extensive forms generated by SGM k :

$$\begin{aligned}
\left(\mathbb{P}^{dk}\right) \qquad \min_x \quad & \left(c^d\right)^T x + \sum_{s \in S_{dk}} p_s^{dk} Q^{dk}(x, s) \\
\text{s.t.} \quad & A^d x = b^d, \\
& x \in X,
\end{aligned}$$

where

$$Q^{dk}(x, s) = \min_y \ \left\{ q_s^{dk} y \mid W^d y = h_s^{dk} - T_s^{dk} x,\ y \in Y \right\}$$

Let x^{dk} be an optimal solution to $\mathbb{P}^{dk}$. For each $k = 1, 2, ..., K$, we conduct the historical re-enactment as follows:

For each $d \in \{1, 2, ..., D\}$,

solve $\mathbb{P}^{dk}$ for x^{dk}

solve $Q^{do}\left(x^{dk}\right) = \min_y \ \left\{ q_o^d y \mid W^d y = h_o^d - T_o^d x^{dk},\ y \in Y \right\}$

set $z_o^{dk} = \left(c^d\right)^T x^{dk} + Q^{do}\left(x^{dk}\right)$

Compute $c^k = \frac{1}{|D|} \sum_{d \in D} z_o^{dk}$

We claim that SGM i has a higher quality than SGM j if $c^i < c^j$. Solving $\mathbb{P}^{dk}$ for each $d \in \{1, 2, ..., D\}$ and $k \in \{1, 2, ..., K\}$ may be difficult due to the challenging computational complexity of the SP, especially in the mixed-integer case. Thus, we seek to replace this process with a computationally easier method.

As motivating examples, we consider two challenging stochastic mixed integer programming problems. The stochastic server location problem (SSLP) and the SUC problem are briefly introduced in this section.

The SSLP is to choose locations of servers from potential locations and allocate clients to the chosen servers to maximize the total expected net revenue subject to the given constraints (Ntaimo and Sen 2005). Network design for electric power, internet services, telecommunications, and water distribution are some of its applications. This problem is formulated as a two-stage SP model. Binary first stage decisions determine whether or not to invest in a server at each of the potential locations. Second stage decisions, which are also binary, assign clients to each server. There are constraints on the total number of servers that can be installed and the server capacity. Moreover, each available client can be served by at most one server. The uncertainty occurs in the availability of clients. Thus, the scenarios can be represented as binary vectors where a value of 1 denotes that the corresponding client materializes. The first-stage cost, which is the investment cost of server siting, is denoted by $c^T x$ in (1). The expected second stage cost, as the negative of the revenue obtained by serving material customers, is denoted by $\mathrm{E}_{\xi} Q\left(x, \xi(\omega)\right)$ in (1). The constraint on the total number of servers that can be installed is expressed by (2). Binary restrictions on the first-stage decision variables are expressed by (3). Unserved demand due to the limitations of server capacity (which results in a loss of revenue), the requirement that each available client is served by at most one server, and binary restrictions on the second-stage decision variables are summarized in the feasible region described by (4).

Unit commitment is an important short-term planning problem for electric power generation in which a commitment schedule is identified for each thermal generating unit over a planned time period (Takriti et al. 1996). In our application, we consider a two-stage SUC formulation where the binary commitment decisions are made in the first stage and the dispatch decisions for the committed units are made in the second stage (Feng et al. 2015). The objective, represented by equation (1), is to minimize the total cost which includes the start-up and shutdown costs in the first stage and the expected generation costs along with heavy penalties on load mismatch in the second stage subject to the operational constraints considering all scenarios. In our application, scenarios represent probabilistic time series for wind energy over the target day. Operational constraints include minimum up and down time constraints represented by equation (2), along with energy balance, ramp rate limits and generation level limitations that are summarized in the feasible region described by (4). The uncertain parameters appear in the net load; i.e., the load less the wind energy, on the right-hand-side of the energy balance constraint for each time period. If the total amount of dispatched energy from the committed units is less than the net load then a shortage occurs or, inversely, if it is greater than an excess occurs.

4 Proposed generic approach for assessment of scenario generation methods

The proposed scenario generation assessment approach accounts for the impact of each single scenario realization on the optimal cost. A rank histogram is employed to assess the reliability of scenario sets, where the ranks are computed based on the MTD (Sari and Ryan 2016). Although the MTD generally is found by solving a linear program, in our application it is the minimum cost of transporting the probability from the group to the individual and can be found in a single greedy step. The notations below relate to computing the MTD and constructing the MTD rank histogram:

u : characteristic of scenario or observation

u_s^d : the value of u in scenario s for instance d

u_o^d : the observed value of u in instance d

$\delta(u',u)$: distance metric

Given observations, $\boldsymbol{u_o^d}$, and scenario-probability pairs, $\boldsymbol{V^d} = \left\{ \left(\boldsymbol{u_s^d}, \boldsymbol{p_s^d} \right) \right\}_{\boldsymbol{s=1}}^{|\mathbf{S}_d|}$, for a given set of instances along with a pre-rank function $f\left(u', V; \delta\right) = \sum_{(u,p)\in V} \delta\left(u', u\right) p,$ the MTD rank histogram is constructed as follows (Sari et al. 2016):

1. For each $d = 1, 2, ..., D$, find l_o^d , the distance from the scenarios to the observation:

$$l_o^d \equiv f\left(u_o^d, V^d; \delta\right)$$

2. For each $s = 1, ..., \left|S_d\right|$ compute l_s^d as the distance from the set $S_d \cup \{o\} \setminus s$ to the scenario *s*, where probability p_s^d is assigned to u_o^d :

$$l_s^d \equiv f\left(u_s^d, V^d \setminus \left(u_s^d, p_s^d\right) \cup \left\{\left(u_o^d, p_s^d\right)\right\}\right)$$

3. Find the rank of l_o^d, denoted $\boldsymbol{r^d}$, when $\left\{l_o^d\right\} \cup \left\{l_s^d\right\}_{s=1}^{|S_d|}$ are ordered from largest to smallest.

4. Construct the histogram of $\left\{\boldsymbol{r^d}\right\}_{\boldsymbol{d=1}}^{\boldsymbol{D}}$.

The MTD rank histogram is able to distinguish between sets of scenarios that are more or less reliable according to their bias, variability and autocorrelation when applied to scenarios directly. MTD rank histograms display a downward slope from left to right for an under-dispersed ensemble of scenarios and an upward slope for an over-dispersed ensemble. Bias overpopulates the small ranks similarly as under-dispersion. For scenarios with a higher (lower) autocorrelation level than the observation, a sloping downward (upward) shape is observed. A hill-shaped MTD rank histogram is observed for scenarios with heterogeneous autocorrelation levels. Flat histograms result when the scenarios are reliable (Sari et al. 2016).

While in (Sari et al. 2016) the distances among realizations are measured directly, in this paper we modify the definition of *u* and its metric, δ. In our proposed method, distances among scenarios and the observed value are measured by fixing first-stage decisions to a common value and computing the differences among second-stage objective values obtained by solving the single-scenario deterministic sub-problems of the SP problem. Thus, the impact of each scenario in the SP problem is considered.

Our generic approach for SGM assessment is explained as follows: As in Section 3, we have observational data $\left\{\left(q_o^d, h_o^d, T_o^d\right)\right\}_{d=1}^{D}$ and scenario sets $\left\{\left(q_s^d, h_s^d, T_s^d\right), s \in \mathrm{S}_d\right\}_{d=1}^{D}$ generated by the method under assessment. To compute distances among scenarios and the observation for each instance d, we solve a single-scenario problem and obtain the optimal first-stage decision variables; i.e., a candidate solution. Then, the second-stage problem is solved for each scenario as well as the observation with the first-stage decision variables fixed to the candidate solution and the second-stage cost is recorded. The distances among the second stage costs are used as the function δ to construct the MTD rank histogram. The steps are formalized as follows:

For each $d \in \{1, ..., D\}$,

Step 1: Solve a single-scenario (deterministic) version of the SP problem with parameters $\left(\hat{q}^d, \hat{h}^d, \hat{T}^d\right)$:

$$\left(\hat{x}^d, \hat{y}^d\right) = \arg\min_{x,y} \quad c^T x + \hat{q}^d y \tag{5}$$

$$\text{s.t.} \quad Ax = b, \tag{6}$$

$$\hat{T}^d x + Wy = \hat{h}^d \tag{7}$$

$$x, y \geq 0 \tag{8}$$

Step 2: For each $g \in G_d = S_d \cup \{o\}$, solve a single-scenario version of the second stage of the SP by fixing the first stage decision variables to the values, $\hat{x}^d$, that are obtained from *Step 1*:

$$u_g^d = \min_y \quad q_g^d y \tag{9}$$

$$\text{s.t.} \quad Wy = h_g^d - T_g^d \hat{x}^d \tag{10}$$

$$y \geq 0 \tag{11}$$

Step 3: Construct the MTD rank histogram using u_o^d, $V^d = \left\{\left(u_s^d, p_s^d\right)\right\}_{s=1}^{|S_d|}$ and $\delta(u', u) = |u' - u|$.

Variants of the assessment approach differ in how we choose $\left(\hat{q}^d, \hat{h}^d, \hat{T}^d\right)$ in (5) - (8) and are explained in detail in Sections 5 and 6. Fig. 1 summarizes the generic scenario assessment for reference in explaining the variants of the approach. Note that only input and output quantities differ among the variants.

For each $d \in \{1, \ldots, D\}$,

	input	output
Step 1: solve $(5)-(8)$	$\left(\hat{q}^d, \hat{h}^d, \hat{T}^d\right)$	$\hat{x}^d$
Step 2: for each $g \in G_d$	$G_d = S_d \cup \{o\}$	
solve $(9)-(11)$	$x = \hat{x}^d$	u_g^d
Step 3: construct MTD	$\left\{u_o^d, V^d\right\}_{d=1}^{D}, \delta(u', u)$	MTD rank histogram

Fig. 1 Generic scenario assessment

5 Expected value based scenario assessment

Recently developed scenario reduction techniques take the impact of each scenario on the SP problem into account. Similar to (Morales et al. 2009), we judge a scenario set based on the results of each single scenario by fixing the first-stage variables to the values that are obtained from solving the expected-value problem, as a first approximation of the optimal values of the decision variables. As summarized in Fig. 2, in this variant we set $\left(\hat{q}^d, \hat{h}^d, \hat{T}^d\right) = \left(\overline{q}^d, \overline{h}^d, \overline{T}^d\right)$, the expected value scenario.

	input	output
Step 1:	$\left(\overline{q}^d, \overline{h}^d, \overline{T}^d\right)$	$\overline{x}^d$
Step 2:	$\mathrm{G}_d = S_d \cup \{o\}$	
	$x = \overline{x}^d$	$\overline{u}_g^d$
Step 3:	$\left\{\overline{u}_o^d, V^d\right\}_{d=1}^{D}, \delta(u', u)$	MTD rank histogram

Fig. 2 EV-based scenario assessment

Because a flat MTD rank histogram based on distances among scenarios results from reliable scenarios (Sari et al. 2016), the histogram constructed here might be hypothesized to be flat if the scenarios are of high quality. To investigate this hypothesis, we simulated the application of this approach when applied to SSLP. By systematically varying the parameters of scenarios and simulated observations, we observed the results of EV-based scenario assessment when scenarios have defects and when scenarios are reliable.

In the test instances used for the EV-based scenario verification simulation study, there are 5 potential locations for servers and 50 potential clients. Scenario-independent instance data are obtained which specifies the set of potential server and customer locations, server capacities, installation costs, and revenues (Ahmed et al. 2015). Scenario-dependent instance data additionally specifies the set of customers that are actually realized in that specific scenario. Each simulation study consists of 1000 instances of ten randomly-generated scenarios, where the binary existence of each client follows a Bernoulli distribution.

In the first simulation, the client availabilities were independent and identically distributed. In the observed data, the Bernoulli parameter was $p_{obs} = 0.5$. To test whether the proposed approach detects bias, we set the Bernoulli parameter used to generate scenarios, p_{scen}, to 0.1, 0.3, 0.4, 0.5, 0.6, 0.7, and 0.9. Fig. 3 shows the resulting MTD rank histograms.

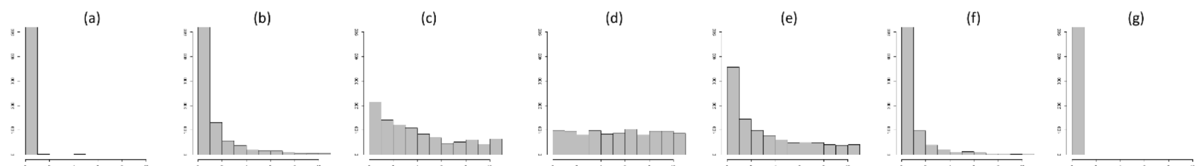

Fig. 3 MTD rank histograms obtained by EV-based scenario assessment for various values of parameter $p_{scen} = (\mathrm{a})\ 0.1\ (\mathrm{b})\ 0.3\ (\mathrm{c})\ 0.4\ (\mathrm{d})\ 0.5\ (\mathrm{e})\ 0.6\ (\mathrm{f})\ 0.7\ (\mathrm{g})\ 0.9$ with $p_{obs} = 0.5$.

A downward slope appears in all the MTD rank histograms except in panel (d) where $p_{obs} = p_{scen}$. The magnitude of the slope of the MTD rank histograms increases with the difference between p_{scen} and p_{obs} due to the increasing bias in the second stage cost results. When the parameters are equal for the scenarios and observational data we observe a flat histogram (d).

Fig. 4 shows the results of a simulation when there is no bias; i.e., $p_{obs} = p_{scen} = 0.5$ but there are correlation inconsistencies between observations and scenarios. Correlated binary variates are generated by using the exchangeable correlation structure method of Lunn and Davies

(1998). For observational data the pairwise correlation of availability among clients is $\rho_{obs} = 0.01$, whereas for scenario data, we vary the corresponding parameter, ρ_{scen}.

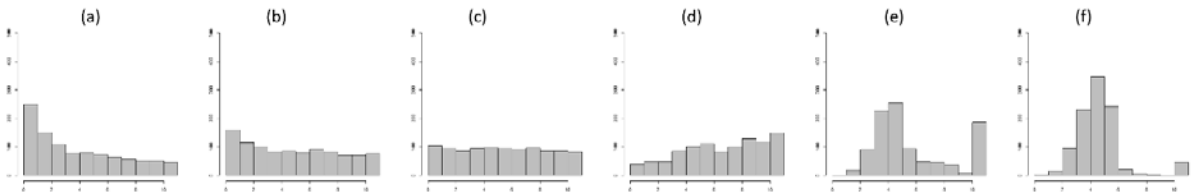

Fig. 4 MTD rank histograms obtained by EV-based scenario assessment for various values of parameter $\rho_{scen} = (\mathrm{a})\ 0.0001\ (\mathrm{b})\ 0.0025\ (\mathrm{c})\ 0.01\ (\mathrm{d})\ 0.04\ (\mathrm{e})\ 0.36\ (\mathrm{f})\ 0.81$ when $\rho_{obs} = 0.01$.

When scenarios have lower correlation than the observation as in panels (a) and (b), the histogram has a downward slope that is steeper when the difference correlation is greater. When scenarios and the observation have the same correlation, the MTD rank histogram appears to be flat in panel (c), indicating the scenarios are of good quality. As the correlation of scenarios is increased to 0.04 the histogram takes on an upward slope in (d) which indicates that the second stage costs of the scenarios are over-dispersed. Larger values of the scenario correlation relative to the observation result in hill-shaped rank histograms observed in (e) and (f). These indicate that the range of the MTDs among the costs of scenarios is wide, so the MTD from scenarios to the observation falls in the middle frequently.

From this study, EV-based scenario assessment appears to be a useful approach for assessing scenario quality, insofar as it corresponds to reliability. The results are easy to interpret. For good quality scenarios we expect a flat rank histogram because the second stage cost of observation should be indistinguishable among the second stage costs of scenarios. Well-known goodness-of-fit tests can be used to test uniformity. However, this approach may not be satisfactory for comparing SGMs. The scenarios produced by different methods may have different expected values. For example, the same empirical distribution may be used for either moment-matching or sampling to obtain competing scenario sets. But commonly-used scenario reduction heuristics do not necessarily preserve the mean of the set of sampled scenarios. To compare the results of distinct SGMs, we should evaluate them in Steps 2 and 3 of our procedure using the same first-stage decisions. With the perfect information based scenario assessment described next, first-stage decision variables are obtained through the hindsight available in a historical re-enactment.

6 Perfect information based scenario assessment

In this variant, we solve a single scenario version of the SP problem with the perfect information (PI) parameter values $\left(q_o^d, h_o^d, T_o^d\right)$ to obtain optimal values of the first- and second-stage decision variables, $\breve{x}_o^d, \breve{u}_o^d$. This approach is summarized in Fig. 5.

	input	output
$STEP\ 1:$	$\left(q_o^d, h_o^d, T_o^d\right)$	$x_o^d, \breve{u}_o^d$
$STEP\ 2:$	$G_d = S_d$	
	$x = x_o^d$	$\breve{u}_g^d$
$STEP\ 3:$	$\left\{\breve{u}_o^d, V^d\right\}_{d=1}^{D}, \delta\left(u', u\right)$	MTD rank histogram

Fig. 5 PI-based scenario assessment

6.1 Server location simulation study

The results of PI-based scenario assessment when applied to our server location instances are shown in Figs. 6 and 7, based on the same sets of scenarios and observations as Section 5. The results of PI-based scenario verification are similar to the results of EV-based scenario assessment. When scenarios and observations are drawn from the same distribution, we observe flat MTD rank histograms. When scenarios are biased or have the client availability correlation from the observations, downward slopes or hill shapes are observed.

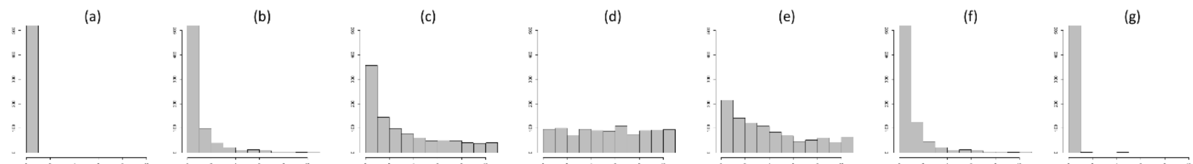

Fig. 6 MTD rank histograms obtained by PI-based scenario assessment for various values of parameter $p_{scen} = (\mathrm{a})\ 0.1\ (\mathrm{b})\ 0.3\ (\mathrm{c})\ 0.4\ (\mathrm{d})\ 0.5\ (\mathrm{e})\ 0.6\ (\mathrm{f})\ 0.7\ (\mathrm{g})\ 0.9$ when $p_{obs} = 0.5$.

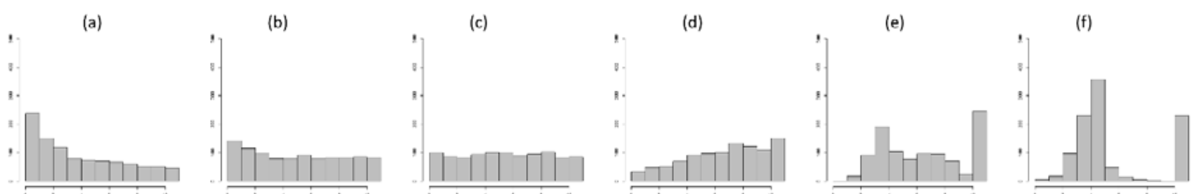

Fig. 7 MTD rank histograms obtained by PI-based scenario assessment for various values of parameter $\rho_{scen} = (\mathrm{a})\ 0.0001\quad (\mathrm{b})\ 0.0025\quad (\mathrm{c})\ 0.01\quad (\mathrm{d})\ 0.04\quad (\mathrm{e})\ 0.36\quad (\mathrm{f})\ 0.81$ when $\rho_{obs} = 0.01$.

However, when the optimal objective value of the SP is highly sensitive to the first-stage solution, the PI-based MTD rank histogram may not expected to be flat. Because we evaluate the results of employing scenarios by fixing the first stage decisions to the optimal first stage decision variables obtained from the observation, the second stage costs of scenarios are expected almost always to exceed the corresponding costs for the observation. Moreover, the range of the distances among the second stage costs of scenarios might be large. We would expect a downward slope in the rank histogram if the bias in second stage costs of scenarios dominates or a hill-shaped rank histogram if the wide range of the distances among second stage costs dominates. This phenomenon is illustrated by the SUC problem as explored in a simulation study.

6.2 Unit commitment simulation study

Simulation studies were designed to explore the results of PI-based scenario assessment approach when scenarios have defects (bias, under/over-dispersion, bias that is hidden by variation, and autocorrelation inconsistencies) and when scenarios are reliable. For the simulation studies, we generated simulated wind scenarios and observations from AR(1) distribution controlling the values of the mean $\left(\mu_{obs} \text{ and } \mu_{scen}\right)$, standard deviation $\left(\sigma_{obs} \text{ and } \sigma_{scen}\right)$ and autocorrelation $\left(\rho_{obs} \text{ and } \rho_{scen}\right)$. Each panel represents 1000 instances, with 10 scenarios for instance consisting of 24 hourly values. We use the concrete stochastic single-bus unit commitment model described in (Feng and Ryan 2016). The deterministic unit commitment instance, having 20 thermal generators used for this simulation study is described in detail in (Sari and Ryan 2017).

Fig. 8 shows the results of reliable scenarios, where the mean, variation, and autocorrelation of scenarios are equal to those of the observations.

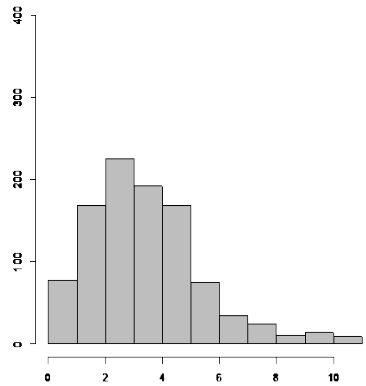

Fig. 8 MTD rank histogram obtained by PI-based scenario assessment for $\mu_{obs} = \mu_{scen} = 2500,\ \sigma_{obs} = \sigma_{scen} = 100,$ and $\rho_{obs} = \rho_{scen} = 0.70$.

The MTD rank histogram in Fig. 8 is not flat for reliable scenarios when applied to unit commitment problem. The middle ranks are overpopulated because the second stage cost of observational data (for which the first stage decisions are optimized) is lower than the second stage cost under most scenarios and the range of differences among scenario costs is large. This causes the MTD from the scenario costs to the observation's cost to tend toward a middle rank.

For reliable scenarios, we may also observe downward slopes, hill shapes, or a right-skewed hill shape under different parameter settings. Some possible outcomes are illustrated in Fig. 9. The second stage cost for the observation is almost always lower than the second stage costs in the scenario subproblems because the first-stage decisions are optimized for the observation. If the resulting bias in the second-stage costs for scenarios is pronounced, we see a downward slope in the rank histogram. Alternatively, the range of the distances among the second stage costs of scenarios might be large which causes a hill shape. Some combinations of different parameters can result in hill-shaped rank histograms that are right-skewed.

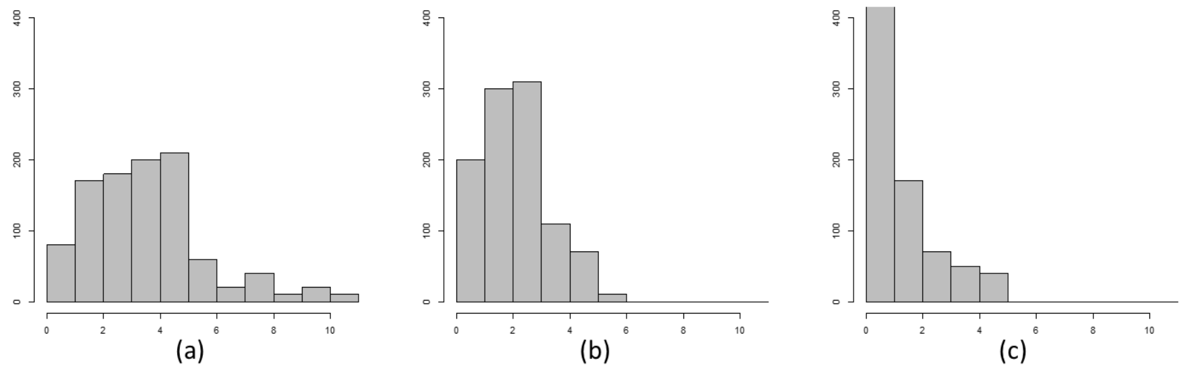

Fig. 9 MTD rank histograms obtained by PI-based scenario assessment in unit commitment with wind power scenarios and observations generated from AR(1) processes with $\mu_{obs} = \mu_{scen} = \mu,\ \sigma_{obs} = \sigma_{scen} = \sigma,$ and $\rho_{obs} = \rho_{scen} = \rho$ and (μ, σ, ρ) = (a) $(70, 2500, 0.75)$; (b) $(300, 1500, 0.50)$; (c) $(800, 800, 0.50)$

These results indicate that a flat rank histogram cannot be expected from PI-based assessment even when the scenarios are reliable. We conjecture that sensitivity of the objective value to small deviations from the optimal solution causes the behaviors observed in Figs. 8 and 9. As a result, we cannot apply goodness-of-fit testing for uniformity to assess scenario quality. Instead, in the next section we propose an evaluation method based on the notion that reliable scenarios are statistically indistinguishable from the observations.

6.3 Evaluation of the MTD rank histogram from PI-based scenario assessment

We conjecture that if the observed data have similar characteristics to the scenario sets, a MTD rank histogram very similar to the PI-based one would result from ignoring the observed data and treating one of the scenarios as if it were the observation. Thus, for each historical instance we randomly select a scenario and assign it the role of the observation in PI-based assessment to obtain the random selection (RS)-based rank histogram. As summarized in Fig. 10, for each instance the single-scenario problem in Step 1 is solved for a scenario, $\left(q_{s'}^d, h_{s'}^d, T_{s'}^d\right)$, randomly selected for that instance. A SGM for which the RS-based histogram is similar to the PI-based histogram would be expected to perform well in the application.

	input	output
Step 1:	$\left(q_{s'}^d, h_{s'}^d, T_{s'}^d\right)$	$x_{s'}^d, \tilde{u}_{s'}^d$
Step 2:	$G_d = S_d \setminus s'$	
	$x = x_{s'}^d$	$\tilde{u}_g^d$
Step 3:	$\left\{V^d\right\}_{d=1}^D, \delta\left(u', u\right)$	MTD rank histogram

Fig. 10 Random scenario (RS) based scenario assessment

When this method is applied using the same set of reliable scenarios as in Fig. 8, which is reproduced for convenience in Fig. 11(a), we obtain the rank histogram shown in Fig. 11(b). The similarity of the two panels confirms that the scenarios are of high quality.

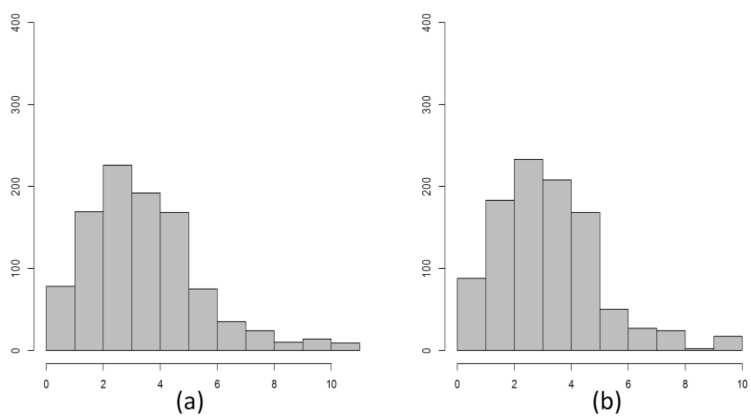

Fig. 11 MTD rank histogram obtained by a RS scenario (a) to evaluate the MTD rank histogram shown in Fig. 8, reproduced in (b).

Fig. 12 depicts the results of RS-based rank histograms when applied to the reliable scenarios used in Fig. 9. The similarity of corresponding panels between the two figures confirms the scenario quality.

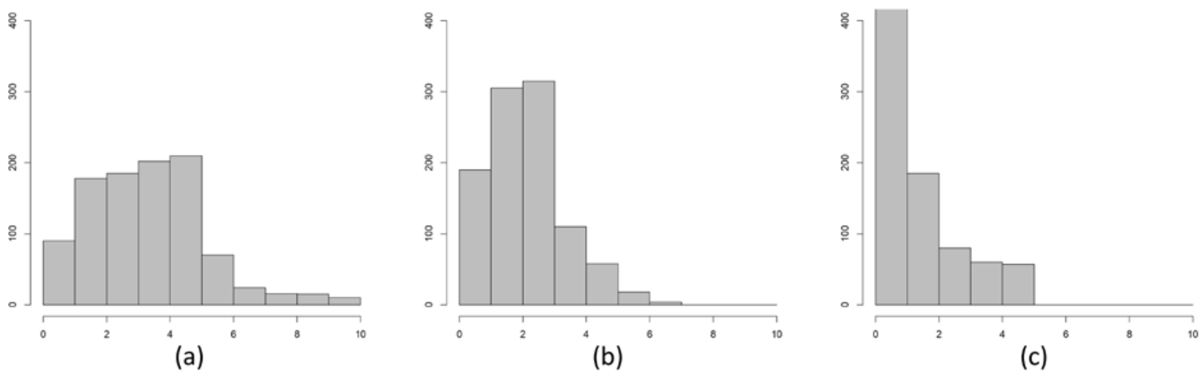

Fig. 12 MTD rank histogram obtained by a RS scenario to evaluate the MTD rank histograms shown in Fig. 9

7 Unit commitment case study

Wind power scenarios are generated from the day-ahead wind forecast and observation data from the Bonneville Power Administration from 2012/10/01 to 2013/09/31 using the quantile regression with Gaussian copula approach (QR) (Pinson et al. 2009) and epi-spline approximation approach (EPI) (Rios et al. 2015; Staid et al. 2017). We test two variants of each approach labeled as QR, QRnew, EPIwide, and EPInarrow. We obtain the load data from Independent System Operator of New England (ISO-NE). The details of scenario generation methods and how the input data are obtained are documented in (Sari et al. 2016). We assumed a 20% wind penetration and omitted the reserve requirements in the SUC. The deterministic unit commitment instance used for the case study is described in (Sari and Ryan 2017). For illustration, in Fig. 13, we plot the scaled wind power scenarios that are generated for the same day.

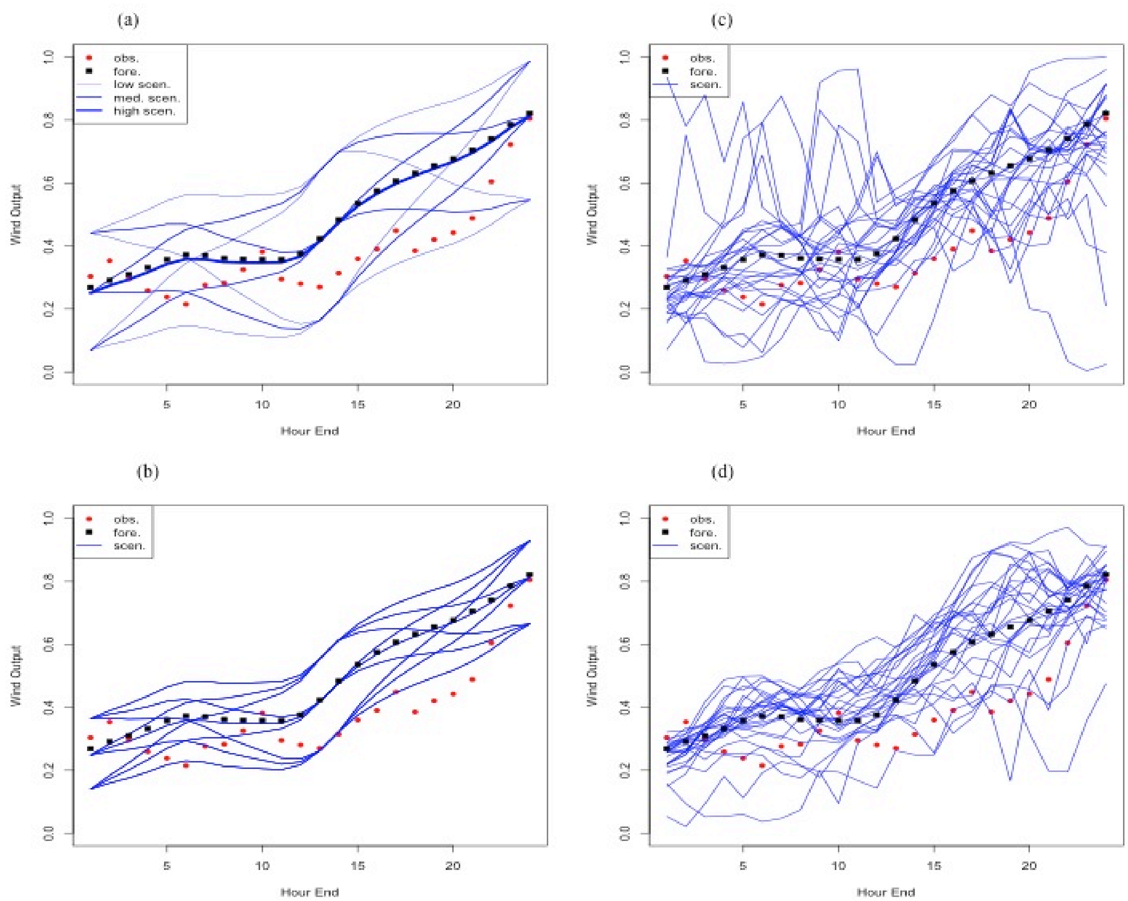

Fig. 13 Wind power scenarios generated for 2012/10/19 (a) EPIwide, (b) EPInarrow, (c) QR, (d) QRnew

Fig. 14 shows the results of EV-based scenario assessment when applied to UC problem with scenarios generated by each of these methods for 343 days.

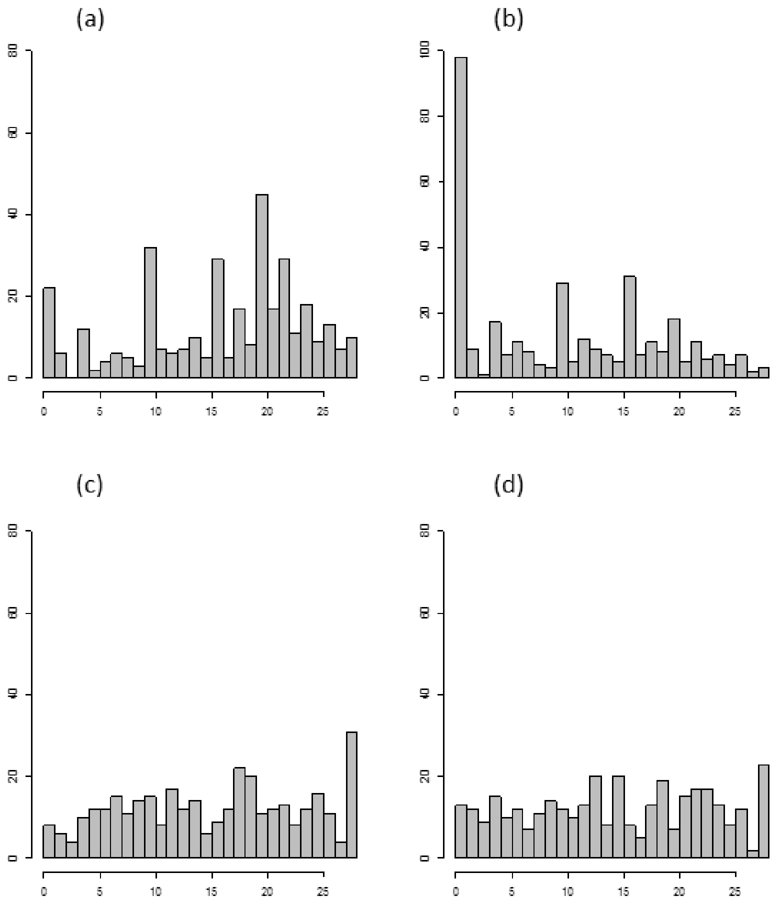

Fig. 14 MTD rank histograms obtained by EV-based scenario assessment when applied to UC case study with scenarios generated by approaches (a) EPIwide, (b) EPInarrow, (c) QR and (d) QRnew.

The MTD rank histogram constructed by EPInarrow scenarios display a downward slope. The smallest rank is over-populated. This is a result of high bias and/or under-dispersion in the results of second stage costs. There is no obvious pattern in the MTD rank histogram constructed based on EPIwide scenarios. According to the MTD rank histograms, we expect to achieve higher cost savings in unit commitment and dispatch problem with EPIwide scenarios than with EPInarrow scenarios. The QR method results in MTD rank histograms with an upward slope because the resulting second stage costs of QR scenarios are over-dispersed. QRnew scenarios result in a flatter rank histogram than QR. Similarly, we expect that solutions obtained with QRnew scenarios will incur lower costs than those found with QR scenarios.

A high quality scenario set should result in a relatively flat histogram in the EV-based assessment, which means the results of observational data are indistinguishable among the results of scenario data. With this approach, we can verify a scenario set to be of high quality for the related SP problem and compare the variants of each scenario generation method. In this case study, we can eliminate the scenarios generated by EPInarrow over EPIwide and QR over QRnew. However, the EV-based assessment may not differentiate among distinct scenario generation approaches. In order to make a comparison between quantile regression and epi-spline scenario, we must evaluate them on the same basis. To do so, we apply PI-based scenario assessment where the values of first stage decision variables are fixed across the scenario generation methods. Fig. 15 displays the hill-shaped rank histograms obtained for both EPIwide and QRnew scenarios.

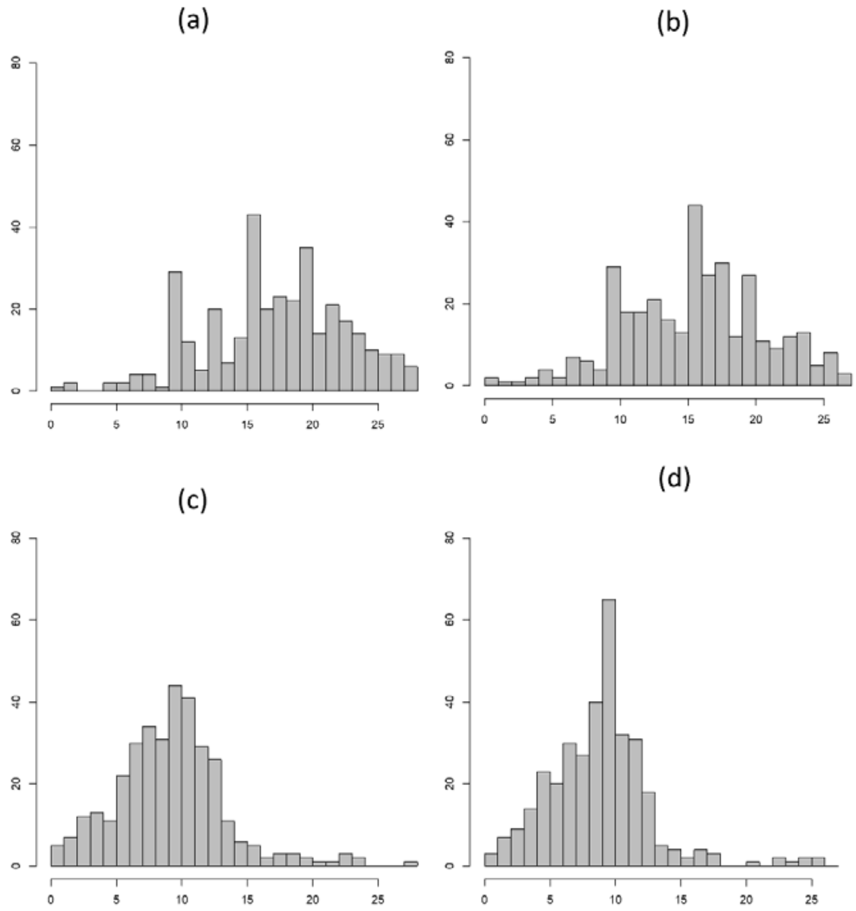

Fig. 15 MTD rank histograms obtained with scenarios generated by approaches EPIwide by (a) PI-based scenario assessment, (b) RS-based scenario assessment and QRnew by (c) PI-based scenario assessment, (d) RS-based scenario assessment.

The MTD rank histograms resulting from PI-based and RS-based assessment appear similar for both SGMs. To evaluate their similarity quantitatively we compute the MTD between empirical distributions of ranks in each bin. The MTDs between rank histograms of PI-based and RS-based approaches for EPIwide and QRnew scenarios are 1.2976 and 1.8479, respectively. According to this metric, the rank histograms of EPIwide scenarios are more similar than those of QRnew scenarios. Thus, the EPIwide variant of the EPI scenario generation method is the overall choice. According to detailed re-enactment using the method described in Section 3, this SGM resulted in the lowest cost of the four scenario generation methods (Sari and Ryan 2017).

8 Conclusions

High quality scenarios are very important for achieving costs savings by solving SP problems rather than deterministic approximations. We proposed EV-based and PI-based scenario assessment approaches aiming to assess the quality of scenarios quickly. We applied them to server location and unit commitment problems with the simulated scenarios to show the results of approaches when scenarios are reliable or unreliable and discussed how to interpret the results. Two different scenario generation methods, along with two variants, are tested with the proposed approaches in a unit commitment case study where uncertainty occurs in wind energy production.

The EV-based scenario assessment is expected to produce a flat histogram for a high quality scenario set. It is a useful approach when comparing the variants of a scenario generation method. If distinct scenario generation methods that may result in different expected values are under evaluation, we suggest using PI-based and RS-based scenario assessment and evaluating the similarity of the resulting rank histograms. With the proposed approaches, the scenario generation methods that are expected to lead to low costs in SP problem can be identified quickly.

REFERENCES

Ahmed S, Garcia R, Kong N, Ntaimo L, Parija G, Qiu F, Sen S (2015) SIPLIB: A Stochastic Integer Programming Test Problem Library. http://www.isye.gatech.edu/~sahmed/siplib. Accessed 20 February 2018

Bakirtzis EA, Biskas PN, Labridis DP, Bakirtzis AG (2014) Multiple time resolution unit commitment for short-term operations scheduling under high renewable penetration. IEEE T Power Syst 29:149-159

Bayraksan G, Morton DP (2006) Assessing solution quality in stochastic programs. Math Program 108:495-514 doi:10.1007/s10107-006-0720-x

Birge JR, Louveaux F (1997) Introduction to stochastic programming. Springer series in operations research. Springer, New York

Bruninx K, Bergh KVd, Delarue E, D'haeseleer W (2016a) Optimization and allocation of spinning reserves in a low-carbon framework. IEEE T Power Syst 31:872-882 doi:10.1109/TPWRS.2015.2430282

Bruninx K, Delarue E, D'haeseleer W (2014) A practical approach on scenario generation & reduction algorithms for wind power forecast error scenarios. Retrieved from https://www.mech.kuleuven.be/en/tme/research/energy_environment/Pdf/wp2014-15b.pdf,

Bruninx K, Dvorkin Y, Delarue E, Pandžić H, D'haeseleer W, Kirschen DS (2016b) Coupling pumped hydro energy storage with unit commitment. IEEE T Sustain Energ 7:786-796 doi:10.1109/TSTE.2015.2498555

Du E, Zhang N, Kang C, Xia Q (2018) Scenario map-based stochastic unit commitment. IEEE T Power Syst PP:1-1 doi:10.1109/TPWRS.2018.2799954

Dupacova J, Gröwe-Kuska N, Römisch W (2003) Scenario reduction in stochastic programming: An approach using probability metrics. Math Program 95:493-511 doi:10.1007/s10107-002-0331-0

Feng Y, Ryan SM (2016) Solution sensitivity-based scenario reduction for stochastic unit commitment. Computational Management Science 13:29-62 doi:10.1007/s10287-014-0220-z

Feng YH, Rios I, Ryan SM, Spurkel K, Watson JP, Wets RJB, Woodruff DL (2015) Toward scalable stochastic unit commitment. Part 1: load scenario generation. Energy Syst 6:309-329 doi:10.1007/s12667-015-0146-8

Heitsch H, Römisch W (2003) Scenario reduction algorithms in stochastic programming. Computational Optimization and Applications 24:187-206 doi:10.1023/a:1021805924152

Heitsch H, Römisch W (2007) A note on scenario reduction for two-stage stochastic programs. Oper Res Lett 35:731-738 doi:10.1016/j.orl.2006.12.008

Kaut M, Wallace SW (2007) Evaluation of scenario-generation methods for stochastic programming. Pacific Journal of Optimization 3:257-271

Lunn AD, Davies SJ (1998) A note on generating correlated binary variables. Biometrika 85:487-490 doi:10.1093/biomet/85.2.487

Morales JM, Pineda S, Conejo AJ, Carrion M (2009) Scenario reduction for futures market trading in electricity markets. IEEE T Power Syst 24:878-888 doi:10.1109/TPWRS.2009.2016072

Ntaimo L, Sen S (2005) The million-variable "march" for stochastic combinatorial optimization. Journal of Global Optimization 32:385-400 doi:10.1007/s10898-004-5910-6

Papavasiliou A, Oren SS (2013) Multiarea stochastic unit commitment for high wind penetration in a transmission constrained network. Oper Res 61:578-592

Pinson P, Girard R (2012) Evaluating the quality of scenarios of short-term wind power generation. Appl Energ 96:12-20 doi:http://dx.doi.org/10.1016/j.apenergy.2011.11.004

Pinson P, Madsen H, Nielsen HA, Papaefthymiou G, Klockl B (2009) From probabilistic forecasts to statistical scenarios of short-term wind power production. Wind Energy 12:51-62

Rachev ST (1991) Probability metrics and the stability of stochastic models. Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, Chichester

Rios I, Wets RJB, Woodruff DL (2015) Multi-period forecasting and scenario generation with limited data. Computational Management Science 12:267-295 doi:10.1007/s10287-015-0230-5

Sari D, Lee Y, Ryan S, Woodruff D (2016) Statistical metrics for assessing the quality of wind power scenarios for stochastic unit commitment. Wind Energy 19:873-893

Sari D, Ryan S (2016) MTDrh: Mass transportation distance rank histogram. https://cran.r-project.org/web/packages/MTDrh/index.html

Sari D, Ryan S (2017) Statistical reliability of wind power scenarios and stochastic unit commitment cost. Energy Systems doi:10.1007/s12667-017-0255-7

Staid A, Watson J-P, Wets RJB, Woodruff DL (2017) Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators. Wind Energy 20:1911-1925 doi:10.1002/we.2129

Takriti S, Birge JR, Long E (1996) A stochastic model for the unit commitment problem. IEEE T Power Syst 11:1497-1508 doi:10.1109/59.535691

Wu H, Shahidehpour M (2014) Stochastic SCUC solution with variable wind energy using constrained ordinal optimization. IEEE T Sustain Energ 5:379-388 doi:10.1109/TSTE.2013.2289853

Zheng QPP, Wang JH, Liu AL (2015) Stochastic optimization for unit commitment-A review. IEEE T Power Syst 30:1913-1924