

# Efficient Scenario Generation for Chance-Constrained Economic Dispatch Considering Ambient Wind Conditions

Qian Zhang <sup>✉</sup>, Graduate Student Member, IEEE, Apurv Shukla <sup>✉</sup>, Member, IEEE, and Le Xie <sup>✉</sup>, Fellow, IEEE

**Abstract**—Scenario approach is an effective data-driven method for solving chance-constrained optimization while ensuring desired risk guarantees with a finite number of samples. Crucial challenges in deploying this technique in the real world arise due to non-stationarity environments and the absence of appropriate risk-tuning models tailored for the desired application. In this article, we focus on designing efficient scenario generation schemes for economic dispatch in power systems. We propose a novel scenario generation method based on filtering scenarios using ambient wind conditions. These filtered scenarios are deployed incrementally in order to meet desired risk levels while using minimum resources. In order to study the performance of the proposed scheme, we illustrate the procedure on case studies performed for both 24-bus and 118-bus systems with real-world wind power forecasting data. Numerical results suggest that the proposed filter-and-increment scenario generation model leads to a precise and efficient solution for the chance-constrained economic dispatch problem.

**Index Terms**—Scenario generation, chance-constrained programming, economic dispatch, wind power forecasting error, scenario approach.

## I. INTRODUCTION

THE integration of intermittent renewable sources of generation into the existing power grids has posed significant challenge due to the inherent uncertainty associated with these sources. While both solar and wind generators introduce uncertainty, wind power, in particular, presents unique challenges. Unlike solar, wind energy lacks diurnal patterns, making it more difficult to predict. Furthermore, the scale of installed and consumed wind energy often surpasses that of solar power in many dispatch areas, leading to a higher level of uncertainty in the overall power system. For instance, wind power accounts for 25% of the total electricity consumed by the Electric Reliability

Manuscript received 28 June 2023; revised 26 October 2023 and 6 December 2023; accepted 23 December 2023. Date of publication 3 January 2024; date of current version 20 June 2024. This work was supported in part by NSF, in part by ECCS-2038963, in part by the U.S. Department of Energy (DoE) Office of Energy Efficiency and Renewable Energy (EERE) through the Solar Energy Technologies Office (SETO) under Award Number DEEE0009031, and in part by Texas A&M Energy Institute. Paper no. TPWRS-00998-2023. (Corresponding author: Le Xie.)

The authors are with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843 USA (e-mail: zhangqian-leo@tamu.edu; apurv.shukla@tamu.edu; le.xie@tamu.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TPWRS.2023.3349237>.

Digital Object Identifier 10.1109/TPWRS.2023.3349237

Council of Texas (ERCOT) region during 2022, compared to 6% for solar [1].

In response to these challenges, stochastic optimization (SO) and robust optimization (RO) are widely used methods for power engineers to deal with uncertainties [2]. SO utilizes probabilistic models to handle randomness [3], while RO employs set-based and deterministic uncertainty models [4]. We focus on chance-constrained optimization (CCO) which bridges the gap between probabilistic and deterministic approaches, providing explicit probabilistic guarantees on the feasibility of optimal solutions [5].

Over the past decade, many attempts have been made at reformulating CCO into a more computationally tractable form. Bienstock et al. [6] reformulate the chance-constrained DC-optimal power flow (DC-OPF) under affine control as a second-order cone program by a moment-based reformulation. Roald et al. [7] extend similar reformulations into the AC-OPF model with some efficient algorithms. Other reformulations have also been proposed considering the distributionally robust property, especially in economic dispatch problem [8], [9].

However, most of the above approaches require the explicit or approximated distribution function of uncertainty, which is hard to validate with streaming data. Data-driven optimization methods, unconstrained by specific underlying uncertainty distributions, have received substantial attention in recent years [5], especially the Sample Average Approximation (SAA) [10] and the Scenario Approach [11]. In the realm of power systems, applications of SAA are evident in day-ahead unit commitment [12], capacity planning [13], and other offline domains [14]. In real-time cases, the scenario approach proves advantageous due to its rapid computational speed, avoiding the long solving time caused by binary variables in SAA. Despite many recent works that have tried to apply the scenario approach in power systems [2], [5], [15], [16], [17], [18], [19], [20], there still exist two main limitations.

### A. Scenario Generation Limitation

Several studies validate scenario approach in the economic dispatch by using synthetic data created from representative distribution [15], [16]. These techniques do not utilize and exploit the fact that the scenario approach can be distribution-agnostic since the scenarios can be directly extracted from previous experience. In most real-world settings, empirical data is collected

from non-stationary environments, wherein the distribution of the random variable depends on environmental conditions [21]. Directly sampling from the past will incorporate all environmental factors from empirical data in scenario approach. For example, the wind power forecasting error addressed in this article, is the primary source of uncertainty in the economic dispatch process [22], [23], while the wind forecast error distribution varies based on forecasting techniques, power output levels, and ambient conditions, as noted in studies by [24], [25].

### B. Risk Tuning Limitation

After choosing the scenario space, the next step is to decide the number of scenarios. The conventional *sample and discard* approach [26] requires the decision-makers to first create an estimate of the sample size and calculate the exact risk level ex-ante. Then, it is decided whether scenarios need to be discarded to trade off risk and performance. To ensure meeting the risk requirements, a conservative sample size is used, which could be astronomical for problems with a large number of decision variables. This inefficient risk-tuning process consumes superfluous data and results in a long computation time.

The main contributions of this article are dealing with these two limitations, which can be summarized as follows:

- 1) To enhance scenario accuracy, the conditional distribution is considered without assuming any knowledge of the true distribution. Utilizing correlation analysis, wind forecast error scenarios are generated from empirical data similar to the present environment, ensuring a more precise representation of real-world conditions.
- 2) The incremental risk tuning method is introduced to meet the risk requirement with minimum data resources [27]. After declaring a desired risk level, scenarios are generated iteratively to eventually hit a desired level of risk.
- 3) Algorithms are designed to incorporate scenario generation with risk-tuning processes efficiently.

The remainder of this article is organized as follows: Section II formulates the chance-constrained economic dispatch problem and highlights the challenges associated with solving the problem with conditional wind power forecast error. Section III introduces our incremental scenario approach, and Section IV discusses the assumptions and limitations of the proposed method. We demonstrate the efficacy of the proposed approach on 24 and 118-bus systems in Section V.

## II. PROBLEM STATEMENTS

### A. Chance-Constrained Economic Dispatch

We consider the chance-constrained DC-OPF formulation in the presence of wind-forecasting uncertainty [6], [18]:

$$\min_{g, \eta} c(g) \quad (1a)$$

$$\text{s.t. } \mathbf{1}^\top g = \mathbf{1}^\top d - \mathbf{1}^\top \hat{w} \quad (1b)$$

$$\underline{g} \preceq g \preceq \bar{g} \quad (1c)$$

$$f(\hat{w}, \tilde{w}) = H_g(g - \mathbf{1}^\top \tilde{w}\eta) - H_d d + H_w(\hat{w} + \tilde{w}) \quad (1d)$$

$$\mathbb{P}_{\tilde{w}} \left( \begin{array}{l} f \preceq f(\hat{w}, \tilde{w}) \preceq \bar{f} \\ g \preceq g - \mathbf{1}^\top \tilde{w}\eta \preceq \bar{g} \\ R_d \preceq -\mathbf{1}^\top \tilde{w}\eta \preceq R_u \end{array} \right) \geq 1 - \epsilon \quad (1e)$$

$$\mathbf{1}^\top \eta = 1 \quad (1f)$$

The decision variables are generator output levels  $g \in \mathbf{R}^{n_g}$ , and an affine control policy  $\eta \in \mathbf{R}^{n_g}$  proportionally allocating total wind fluctuation  $\mathbf{1}^\top \tilde{w}$  to each generator<sup>1</sup>. The objective function is the total generations cost  $c(g)$ . The load level is  $d \in \mathbf{R}^{n_d}$ , and the wind generation  $w = \hat{w} + \tilde{w}$  consists of *deterministic* wind forecast value  $\hat{w} \in \mathbf{R}^{n_w}$  and the *uncertain* forecast error  $\tilde{w} \in \Delta$ , where  $\Delta \subseteq \mathbf{R}^{n_w}$  is the uncertainty set. Transmission line flows  $f \in \mathbf{R}^{n_f}$  are calculated using (3d), where  $H_g$ ,  $H_d$ , and  $H_w$  are the corresponding sub-matrix of the power transfer distribution factor (PTDF) matrix  $H$ . Constraints include transmission line flow limits  $[f, \bar{f}] \in \mathbf{R}^{n_f} \times \mathbf{R}^{n_f}$ , generation capacity limits  $[g, \bar{g}] \in \mathbf{R}^{n_g} \times \mathbf{R}^{n_g}$  and the ramp up(down) rate limits  $[R_d, \bar{R}_u] \in \mathbf{R}^{n_g} \times \mathbf{R}^{n_g}$  are modeled as a chance-constraint form under risk  $\epsilon$  in (1e).

As mentioned in [18], the affine control policy  $\eta$  only focuses on the steady-state behavior of the *Automatic Generation Control* (AGC) action in dispatch time scale, i.e. 5 to 15 minutes, but not the 2 to 6 seconds fast time-scale regulation process. The system's active power deviation is allocated to generators based on  $\eta$ , which is also well known as *participation factors* in the conventional AGC scheme [28]. It is easy to confirm that constraints (3b) and (3f) imply the supply-demand balance in the presence of wind uncertainties:

$$\mathbf{1}^\top (g - \mathbf{1}^\top \tilde{w}\eta) = \mathbf{1}^\top d - \mathbf{1}^\top (\hat{w} + \tilde{w}) \quad (2)$$

Setting the constant affine control policy  $\eta$  prior to the next dispatch interval will unavoidably be economically inefficient if the net load's fluctuation or forecasting error is large. Changing the  $\eta$  more frequently within dispatch interval, or incorporating the optimization program into the AGC control policy may improve the economic efficiency [29], [30], but these methods ignore the network constraints and are hard to apply to the bulk power system due to the communication delay or solving time. In this article, we mainly concentrate on improving the dispatch performance in the chance-constrained problem. The description of affine control policy  $\eta$  in (1) is based on two assumptions: 1) All the traditional generators participate in the AGC actions; 2) The whole system is regarded as one control area.

### B. Conditional Wind Power Forecast Error

Due to the nonlinear wind power curve, the wind power forecast error is observed to vary with the level of its output [31], [32], while spatial and weather parameters also indirectly affect the forecasting quality [33], [34]. Based on these facts, the wind power forecast error measure  $\mathbb{P}_{\tilde{w}}$  in (1e) should be modeled as a conditional probability distribution from similar environments. Previous studies have focused on approximating the measure of

<sup>1</sup>In this article,  $\mathbf{1}^\top$  denotes a row vector of all ones, with its dimension adjustable to match the vector it multiplies.

conditional forecast error, eg., [31] generate the conditional error model based on the wind turbine power curve, [24] employ a fuzzy inference model to obtain conditional prediction intervals and [25] calculate the conditional forecasting error from joint distributed data using copula theory. All the above approximation methods for the distribution  $\mathbb{P}_{\tilde{w}}$  under similar conditions are incompatible with the data-driven approach in chance-constraint optimization, which uses distribution agnostic empirical data. Furthermore, the wind power output, temperature, and weather parameters are continuous variables, meaning the past scenarios' observations will be distinct with probability 1, which makes it impossible to generate scenarios in the data-driven scheme from an identical environment but similar environments. Scenario selection from similar environments presents a major engineering obstacle: a large search space results in conservative decision-making while a smaller search space results in limited data [21].

### III. METHOD

#### A. Scenario Approach

The scenario approach randomly extracts  $N$  independent and identically distributed (*i.i.d.*) scenarios to approximate the chance-constrained program. Supposing we have the random wind forecasting error scenarios set  $\mathcal{N} := \{\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_N\}$ , the chance-constrained inequalities (1e) in DC-OPF problem can be replaced by scenario-based inequalities (3e):

$$\min_{g, \eta} c(g) \quad (3a)$$

$$\text{s.t. } \mathbf{1}^\top g = \mathbf{1}^\top d - \mathbf{1}^\top \hat{w} \quad (3b)$$

$$\underline{g} \preceq g \preceq \bar{g} \quad (3c)$$

$$f(\hat{w}, \tilde{w}) = H_g(g - \mathbf{1}^\top \tilde{w} \eta) - H_d d + H_w(\hat{w} + \tilde{w}) \quad (3d)$$

$$\underline{f} \preceq f(\hat{w}, \tilde{w}_i) \preceq \bar{f}$$

$$\underline{g} \preceq g - \mathbf{1}^\top \tilde{w}_i \eta \preceq \bar{g}$$

$$R_d \preceq -\mathbf{1}^\top \tilde{w}_i \eta \preceq R_u \quad i = 1, 2, 3, \dots, N \quad (3e)$$

$$\mathbf{1}^\top \eta = 1 \quad (3f)$$

To distinguish from the original optimization problem (1), we name the above *scenario problem* as  $\text{SP}(\mathcal{N})$ .

**Definition 1 (Violation Probability):** The *violation probability* of a candidate solution  $(g^*, \eta^*)$  is defined as the probability that  $(g^*, \eta^*)$  is infeasible, i.e.,  $\mathbb{V}_{\tilde{w}}(g^*, \eta^*) := \mathbb{P}_{\tilde{w}}((g^*, \eta^*) \notin \mathcal{X}_{\tilde{w}})$ , where  $\mathcal{X}_{\tilde{w}}$  is the decision set generated by  $\text{SP}(\mathcal{N})$ .

**Definition 2 (Support Constraint):** The scenario-dependent constraint corresponding to sample  $\tilde{w}_s, s \in \{1, 2, \dots, S\}$ , is a

---

#### Algorithm 1: Searching Support Scenarios Using Dual Variables.

---

Solving the *scenario problem*  $\text{SP}(\mathcal{N})$

Generate the primal solution  $(g_N^*, \eta_N^*)$  and the constraints

(3e) related dual solution  $\mu_i^* (i = 1, 2, 3, \dots, N)$

Let  $\mathcal{M} = i \in \mathcal{N} : \|\mu_i^*\| > 0$ . Set  $\mathcal{S} \leftarrow \emptyset$ ;

**for**  $i \in \mathcal{M}$  **do**

Solve  $\text{SP}_{\mathcal{M}-i}$  and compute  $(g_{\mathcal{M}-i}^*, \eta_{\mathcal{M}-i}^*)$

**if**  $c(g_{\mathcal{M}-i}^*) < c(g_N^*)$  **then**

$\mathcal{S} \leftarrow \mathcal{S} + i$

**end if**

**end for**

**OUTPUT:** The support scenarios  $\mathcal{S}$

---

*support constraint* or *support scenario* if its removal improves the solution of  $\text{SP}(\mathcal{N})$ , i.e., if it decreases the optimal cost (3a).

**Definition 3 (Sample Complexity):** The number of support scenarios in  $\text{SP}(\mathcal{N})$  is defined as the sample complexity.

**Definition 4 (Helly's Dimension):** Helly's dimension of the scenario problem  $\text{SP}(\mathcal{N})$  is the smallest integer  $h$  that  $h \geq \text{ess sup}_{\mathcal{N} \subseteq \Delta^N} |\mathcal{S}(\mathcal{N})|$  holds for any finite  $N \geq 1$ , where  $|\mathcal{S}(\mathcal{N})|$  is the number of support constraints or sample complexity<sup>2</sup>.

The most important contribution of the scenario approach is the relationship between violation probability  $\mathbb{V}_{\tilde{w}}(g^*, \eta^*)$ , the number of scenarios  $N$  and the sample complexity.

**Theorem 1 (Exact Feasibility [11], [35]):** Under the assumptions of nondegeneracy and feasibility of the optimization problem, the deepest results show that the distribution of  $\mathbb{V}_{\tilde{w}}(g^*, \eta^*)$  is dominated by a Beta distribution, namely:

$$\mathbb{P}_{\tilde{w}}^N (\mathbb{V}_{\tilde{w}}(g^*, \eta^*) > \epsilon) \leq \sum_{i=0}^{h-1} \binom{N}{i} \epsilon^i (1-\epsilon)^{N-i} := \beta \quad (4)$$

where  $h$  is the Helly's dimension of  $\text{SP}(\mathcal{N})$ , and  $1 - \beta$  is defined as the confidence bound for the solution based on any  $N$  *i.i.d* scenarios.

**Theorem 2 (Property for Convex Problem [11]):** Supposing all the constraints in  $\text{SP}(\mathcal{N})$  is convex for every instance of  $\tilde{w}$ , the sample complexity  $|\mathcal{S}(\mathcal{N})|$  for  $\text{SP}(\mathcal{N})$  is at most  $n$ . In other words,  $h \leq n$ , where  $n$  is the number of decision variables after eliminating the equality constraints [36].

For the convex problem, Helly's dimension  $h$  can be replaced by  $n$  to simplify the problem by applying Theorem 2, but it often causes extremely conservative results. To compute the lower bound of  $h$ , we suggest using the dual-based Algorithm 1 proposed in [17].

#### B. Sampling Scenarios From Parameter Space

Many environmental parameters, such as location, wind speed, temperature, wind direction, and relative humidity affect the wind power forecasting quality [34]. In this article,

<sup>2</sup>Because the randomness of sampling, the number of support constraints might be different especially when sample size is small. Here  $\text{ess sup}$  means the essential supremum to ignore some exceptional cases.

we suppose the wind generators are from the same area with the same forecasting algorithm, while the *deterministic* wind forecast value, the wind power changing rate, temperature, and relative humidity are selected as the four main parameters that affect the forecasting error.

*Remark:* The *deterministic* wind forecast value and the wind power ramp rate are parameters that are integrated with other environmental information, especially the wind speed and wind speed ramp rate, which are observed having a close relationship with the wind power forecasting quality.

*Definition 5 (Parameter Space for Wind Power Forecasting Error):* The parameter space  $\mathcal{V}_N$  is defined as the set of environmental parameters which the past scenarios  $\mathcal{N}$  are extracted from. For instance, the temperature between  $70^{\circ}\text{F}$  and  $80^{\circ}\text{F}$  is a temperature parameter space.

*Definition 6 (Probability Distribution Over Parameter Space):* Let  $\mathbb{P}_{\tilde{w}|\mathcal{V}_N}$  be a probability distribution over the parameter space  $\mathcal{V}_N$ .

*Remark:* Strictly speaking,  $\mathcal{V}_N$  should be identical with the environment parameter at the forecasting moment  $\mathcal{V}_{\text{Now}}$  to acquire more precise risk guarantee. However, the input data in the scenario approach is directly extracted from the empirical experience with continuous environment parameters. Finding the past scenarios under the same environment parameter as the future is impossible with probability 1. Even in some frontier probabilistic prediction methods, it is also hard to guarantee the accuracy of predicted conditional distribution but check how close it approximates to the real distribution based on testing data [37].

Based on the definitions above, the barrier of bringing the scenario approach to the real world is to find the parameter space  $\mathcal{V}_N$  which includes both the potential environment parameter in the future and enough number of empirical scenarios to meet the risk requirement in Theorem 1.

### C. Main Result I: Correlation-Based Scenario Generation

The Pearson correlation coefficient is used to quantify the relationship between each environmental parameter and the wind power forecasting error from the past forecasting data, for example, the past half-year data before the decision-making day. Let random variable  $P$  represent one of the environmental parameters, and random variable  $\tilde{W}$  denote the wind power forecasting error, then the Pearson correlation coefficient between these two variables is given by:

$$\rho_{P,\tilde{W}} = \frac{\mathbb{E}[(P - \mu_P)(\tilde{W} - \mu_{\tilde{W}})]}{\sigma_P \sigma_{\tilde{W}}} \quad (5)$$

where  $\sigma_P$  and  $\sigma_{\tilde{W}}$  are the standard deviation of  $P$  and  $\tilde{W}$ , while  $\mu_P$  and  $\mu_{\tilde{W}}$  are the mean value of  $P$  and  $\tilde{W}$ , respectively. After calculating the correlation coefficient of each parameter, we combine the penalized parameters as the *indicator vector* for estimating the difference in environmental conditions. Suppose  $\rho_1, \rho_2, \rho_3, \rho_4$  are the correlation coefficients of the four normalized parameters  $p_1, p_2, p_3, p_4$  that affect the past forecasting

error respectively, then the *indicator vector*  $v$  is constructed as:

$$v = [\rho_1 p_1, \rho_2 p_2, \rho_3 p_3, \rho_4 p_4]^T \quad (6)$$

Based on the *indicator vector*, we can define the wind power forecasting environment difference  $d_{ij}$  as the distance between two forecasting environments  $i$  and  $j$ , that is:

$$d_{ij} = \|v_i - v_j\| \quad (7)$$

where  $\|\cdot\|$  can be any norm distance and Euclidean norm is used in the case study part.

After calculating the environment difference  $d_{ij}$  between the now and the past, we can pick scenarios from the empirical data under more similar decision-making environments. The overview of the proposed scheme and the conventional scenario approach is drawn in Fig. 1.

Except for correlation analysis, many statistical or learning methods can be used as the similar environment filter represented in Fig. 1, such as the coefficient of determination ( $R^2$ ), and other learning-based ranking algorithms [38]. Furthermore, some generative models, such as generative adversarial networks (GAN), are another path to get the scenarios under a similar environment. In this article, we do not compare the accuracy of different filters but rather focus on illustrating this scheme and its performance on the economic dispatch problem.

### D. Main Result II: Incremental Scenario Optimization

To find the proper number of scenarios meeting the risk requirement, two main scenario-based algorithms exist for convex problems. The simple one is called *a-priori* approach [11], where Helly's dimension  $h$  is approximated by its upper bound: the number of decision variables  $n$ . In *a-priori* approach, the number of the needed scenarios can be directly given by Theorem 1 before solving the optimization program. The scenario sampling from a similar environment embedded in *a-priori* scenario optimization approach is summarized in the Algorithm 2.

The *a-priori* approach has less computational complexity and risk guarantee before solving the optimization problem. When the number of decision variables is small, the *a-priori* approach is an efficient choice, but as the size of the system increases it leads to extreme conservatism. For instance, the sample complexity of the look-ahead economic dispatch problem [16] is 3 – 5 while the decision variables  $n$  is 864.

Leveraging on support scenario searching algorithms, such as Algorithm 1, the *a-priori* approach can be improved if the risk level is updated after solving the optimization problem, called the *a-posteriori* approach [21]. In many practical engineering situations, the decision makers are interested in seeking the optimal solution given a certain risk level, implying that if the updated risk level is too conservative, the decision maker trades off risks for better performance by discarding some scenarios [26]. The computation complexity of this risk-tuning process can be reduced by using an incremental scenario optimization algorithm [27]. In this article, we propose an improved incremental scenario optimization algorithm based on sampling from a similar environment, see Algorithm 3. In practice, this algorithm can generate the optimal solution with the given risk

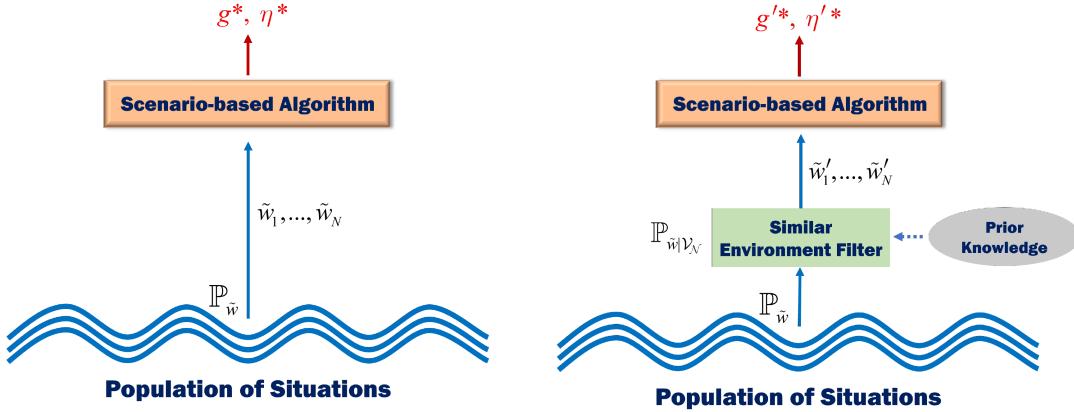


Fig. 1. Comparison of the conventional scenario approach (left) and the proposed scenario generation model (right).

---

**Algorithm 2:** A-Priori Scenarios Approach Sampling from Similar Environment.

---

**INPUT1:** The past forecasting error data with environment parameters

**INPUT2:** The environment parameter at decision-making time

**INPUT3:** The risk and confidence level  $\epsilon$  and  $\beta$

```

 $N \leftarrow (4)$ 
if  $N >$  empirical data size then
    Reset risk and confidence level
else sampling the scenarios  $\mathcal{N}$  under similar environments
     $d_{ij} \leftarrow (5)(6)(7)$ 
     $\mathcal{N} \leftarrow$  the scenarios with the  $N$  smallest  $d_{ij}$ 
end if
Solve the scenario-based optimization problem  $SP(\mathcal{N})$  with optimal solution  $(g^*, \eta^*)$ .
```

**OUTPUT:** The optimal solution  $(g^*, \eta^*)$

---

guarantee and provide some higher-risk solutions with better economic performance to meet the risk-tuning needs of system operators.

In the previous sample and discard risk tuning method [26], the sample complexity is first supposed to be its upper bound, i.e. the number of decision variables, which means a large sample size may be used to solve the problem. After calculating the true sample complexity and risk level based on the solution, the decision-makers can trade off the risk and performance by gradually discarding some support scenarios. Compared to the incremental optimization method, the sample and discard method tunes the risk in a decremented way, which is inefficient when the true sample complexity is much smaller than the number of decision variables.

#### IV. DISCUSSION

The most important assumption in the scenario approach is the samples are identical independent distributed (*i.i.d.*) random

---

**Algorithm 3:** Incremental Scenario Optimization Sampling from Similar Environment.

---

**INPUT1:** The past forecasting error data with environment parameters

**INPUT2:** The environment parameter at decision-making time

**INPUT3:** The risk and confidence level  $\epsilon$  and  $\beta$

- 1: Set  $j := 1$  and  $N_0 = 0$
- 2: Suppose the problem has  $j$  support scenarios, i.e.  $h = j$ , and calculate the number of needed scenarios  $N_j$  from (4).
- 3: Collect a sample of scenarios  $\tilde{w}_{N_{j-1}+1}, \tilde{w}_{N_{j-1}+2}, \dots, \tilde{w}_{N_j}$  with  $N_j - N_{j-1}$  smallest environment difference  $d$  without replacement.
- 4: Solve the scenario-based optimization problem  $SP(\mathcal{N}_j)$  with optimal solution  $(g^*, \eta^*)_{N_j}$ .
- 5: Compute the exact sample complexity  $h_j$  of the solution  $(g^*, \eta^*)_{N_j}$  based on Algorithm 1.
- 6: (Optional) Compute the exact risk level  $\epsilon_j$  corresponding to  $(g^*, \eta^*)_{N_j}$  after updating  $h = h_j$  in (4).
- 7: **if**  $h_j \leq j$  **then**
  - 8: halt the algorithm and **return**  $(g^*, \eta^*) := (g^*, \eta^*)_{N_j}$
  - 9: **else** set  $j := j + 1$  and GOTO step 2
- 10: **end if**

**OUTPUT1:** The optimal solution  $(g^*, \eta^*)$ .

**OUTPUT2:** (Optional) The solution  $(g^*, \eta^*)_{N_j}$  with higher risk level  $\epsilon_j$ .

---

variables. In this section, we will discuss how the proposed method meets this *i.i.d.* property and our weakness.

##### A. Identical Distribution

In the chance-constrained problem, the uncertainties are modeled as random variables sampled from the identical probability distribution. For example, the uncertainty of wind forecasting error in chance-constrained DC-OPF problem (1) is sampled from the fixed probability distribution  $\mathbb{P}_{\tilde{w}}$ . Because each forecasting happens in a complex and varying environment, the precise

description of uncertainty should be under the environment condition, i.e.  $\mathbb{P}_{\tilde{w}|\mathcal{V}_{\text{Now}}}$ . For data-driven optimization methods, if we apply the whole empirical data to describe uncertainty, the identical probability distribution  $\mathbb{P}_{\tilde{w}}$  should be regarded as the marginal distribution over the (past) environment parameters.

Existing papers focusing on modeling the gap between the solution under empirical distribution  $\mathbb{P}_{\tilde{w}}$  and exact distribution  $\mathbb{P}_{\tilde{w}|\mathcal{V}_{\text{Now}}}$  require prior knowledge of the distance or mean value between two probability measures, which is hard to calculate in the wind forecasting scenario. Some machine learning approaches [39] may be useful to approximate this exact distribution, but their results are not compatible with the direct data-driven program, especially the scenario approach.

In this article, instead of quantifying the gap between empirical distribution and exact distribution, we seek to filter empirical data through the similar environment parameter space  $\mathcal{V}_{\mathcal{N}}$ . The proposed solution's risk guarantee is based on the conditional distribution  $\mathbb{P}_{\tilde{w}|\mathcal{V}_{\mathcal{N}}}$ , but the simulations in the next section show that the testing results also meet the setting risk threshold even in the real-world data.

### B. Independent Random Variables

The purpose of sampling scenarios from parameter space is to make the conditional distribution  $\mathbb{P}_{\tilde{w}|\mathcal{V}_{\mathcal{N}}}$  much closer to the exact one  $\mathbb{P}_{\tilde{w}|\mathcal{V}_{\text{Now}}}$ , which do not affect the independent property of scenarios under the two assumptions. The first is that the forecasting algorithm does not use the previous forecasting error data, which is true for most updated forecasting methods [34]. The second assumption is that the dispatch decision itself will not affect future forecasting errors, which is also justifiable because the dispatch decision does not interact with the whole weather system in the short term.

## V. CASE STUDY

The scenario approach formulation (4) has been validated in the chance-constrained economic dispatch many times [5], [15], [16]. The scenarios in these previous papers were all obtained by sampling from some particular distribution, such as normal and beta distribution, which didn't utilize and exploit the distribution-free advantage of the original scenario theory.

In our simulation, the knowledge about uncertainty is acquired directly through experience, i.e. the past recorded data. We focus on the economic dispatch under the 5-min unit, where the 5-min ahead wind forecast uncertainty plays an important part. The peak hours 16:00–18:00 in August 2022 is selected as the testing period with 744 dispatch intervals. For each testing interval, the empirical forecasting error scenarios are generated from the past three or six months, i.e. a dataset with 25920 or 51840 5-min ahead forecasting errors with their corresponding environment parameters. All the data is acquired from the ERCOT website based on the five wind forecasting regions in Fig. 2 [1], while the weather data is from [40].

All the problems are solved using 64 GB RAM on the Intel XEON-10885 M CPU (2.4 GHz). The mathematical models were formulated using YALMIP on Matlab R2023a and solved using Gurobi v9.5.

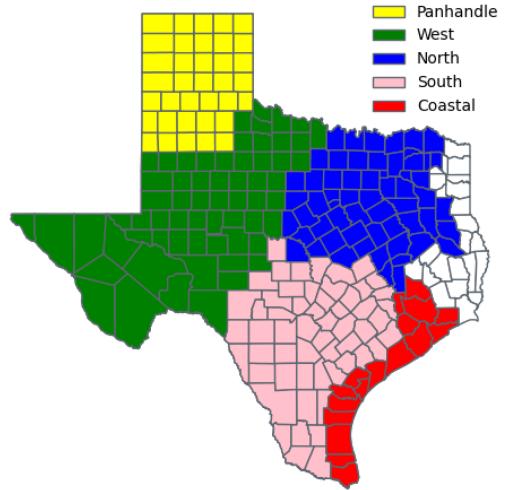


Fig. 2. Wind forecasting region in ERCOT.

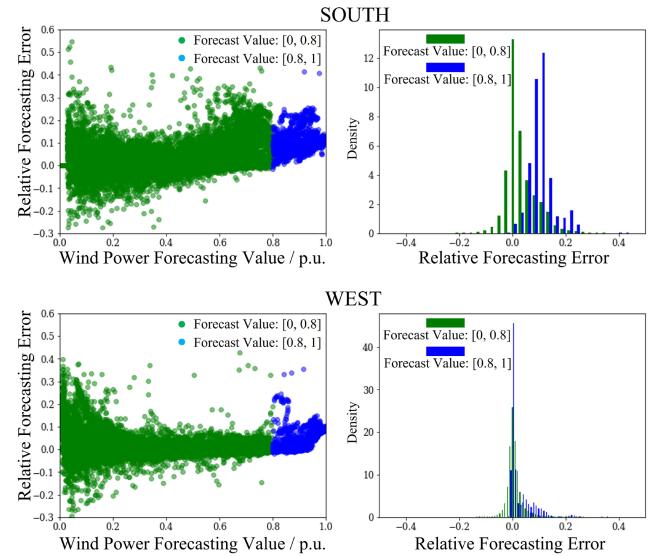


Fig. 3. Scatter plot (left) and density function (right) of forecasting error under different wind power forecasting value.

### A. Conditional Wind Forecasting Error

In this section, we mainly focus on the wind forecasting error affected by the *deterministic* wind forecast value, the wind power ramp rate, temperature, and relative humidity. To illustrate the wind forecasting error under different wind forecast values, the empirical forecasting error data is extracted from June to August 2022 in a similar geographic region in Texas. Fig. 3. shows the wind forecasting error density function under different normalized wind forecasting levels. It is clear that the empirical distribution under high wind output level ( $[0.8, 1]$ , blue) is right-skewed to the lower wind output level ( $[0, 0.8]$ , green), which means directly using the whole past scenario may make the decision aggressive under high wind weather. Similar results are also found in [25], [41].

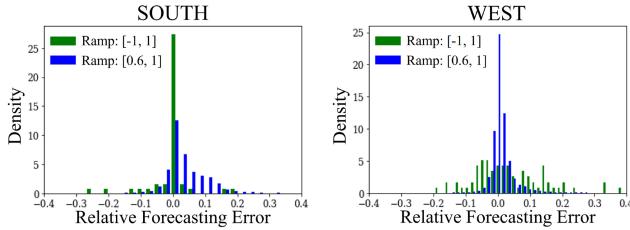


Fig. 4. Density function of forecasting error under different wind power ramping rate.

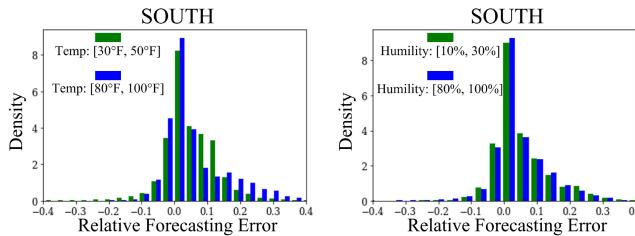


Fig. 5. Density function of forecasting error under different temperatures and relative humidity.

TABLE I  
CORRELATION COEFFICIENT BETWEEN WIND POWER FORECASTING ERROR  
AND SOME ENVIRONMENTAL PARAMETERS

Correlation between forecasting error and	Coefficient
Wind Power Forecasting Value	0.50
(Absolute) Wind Power Changing Rate	-0.17
Temperature	-0.07
Relative Humidity	-0.03

Similarly, under different wind power ramp rates, the forecasting error has different patterns. The ramping rate of wind power is an aggregated parameter reflecting the stationary level of the environment. Typically, the high absolute wind power ramp rate means the weather data used for wind power forecasting is less precise than the stable environment, which results in different forecasting error distributions. After normalizing the ramping rate to  $[-1, 1]$  interval, the distribution of forecasting error conditions on a high ramp-up rate ( $[0.6, 1]$ ) from June to August 2022 compared with its marginal distribution is shown in Fig. 4.

Unlike the wind power forecast value and ramp rate, the temperature and relative humidity have less influence on the wind power forecasting error. The distribution of wind power forecasting errors in southern Texas during hot and cold, dry and wet days of 2022 are compared in Fig. 5. It is obvious that the forecasting error under different temperatures or humidity shares a very similar distribution.

To quantify the correlation level between the above four parameters and the wind power forecasting error, we calculate the Pearson correlation coefficient between each parameter and the forecasting error in southern Texas based on the past half year's data, i.e. February to July 2022. As shown in Table I, the wind power forecasting value and changing rate have more

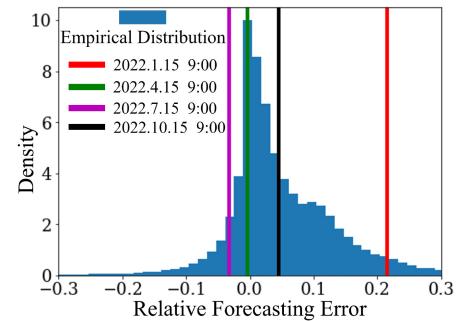


Fig. 6. Future error values and the scenarios' empirical distribution (conventional method).

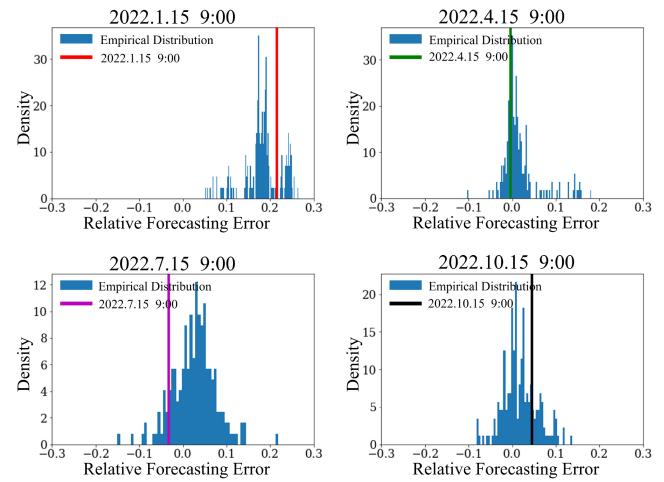


Fig. 7. Future error values and the scenarios' empirical distribution (proposed method).

correlation with the forecasting error than the temperature and relative humidity.

*Remark:* Due to confidentiality concerns, the chosen four parameters are publicly available information accessible online. Decision-makers also have the option to incorporate locally measured data, including wind speed and direction, air pressure, and freezing level, to enhance the resolution of their analysis. By combining these diverse meteorological parameters, it becomes possible to delineate specific scenarios for various extreme weather events based on their respective thresholds.

After the above calculation, the scenarios under similar environments can be generated by Algorithm 2 or 3, where the scenarios are selected from a smaller parameter space rather than the conventional method whose scenarios are randomly extracted from the whole past data. In other words, the distribution of empirical scenarios is adaptive changing with the environment, while it is fixed in the conventional method.

Supposing 200 scenarios are needed to meet the risk requirement, Fig. 6 shows some real *future* error values and the scenarios' empirical distribution in the whole last year, while the proposed method gives the adaptive empirical distribution for each decision-making time in Fig. 7.

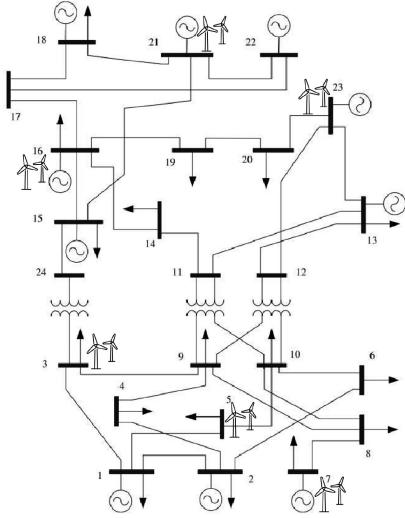


Fig. 8. Modified 24-bus power system integrated with wind.

The advantages of sampling from similar parameter space are obvious from the above simulation. First, the distribution of scenarios will be compressed into narrower intervals giving a higher-resolution description of the future uncertainty. Second, the *future* error values are also covered by this empirical distribution, which somewhat validates our method's correctness.

#### B. 24-Bus System

The 24-bus System is modified from the IEEE Reliability Test System (RTS-24) [42] with an additional six wind farms to mimic the high renewable energy penetration as illustrated in Fig. 8. The detailed information, including generator parameters, reactance and capacity of transmission lines, and the load profile can be found in [43]. The forecasting profile of wind generators located at bus 3, 5 and 7 are directly scaled from the west Texas region wind forecasting results, while the wind farms at bus 16, 21 and 23 are from the south Texas region. Each wind farm is assumed with 400MW capacity and a low marginal price (3 \$/MWh).

As suggested in [44], the capacity on the transmission lines connecting the node pairs (15, 21), (14, 16), and (13, 23) is reduced to 400MW, 250MW, and 250MW, respectively. This is done to introduce bottlenecks or congestion in this high wind penetration system, which shares a similar situation when comes to the real power grids [45].

The number of decision variables  $n$  in the 24-bus system is 22 after eliminating the equation constraints. In this case, a rough approximation of Helly's dimension  $h$  by  $n$  will increase the needed number of scenarios to meet the risk threshold, which results in great conservatism of the final decision. To illustrate this, Fig. 9 shows the relationship between the needed number of scenarios and sample complexity under the risk  $\epsilon = 0.05$  and confidence parameter  $\beta = 10^{-3}$ . It can be seen that when the sample complexity is 22, the needed number of scenarios is

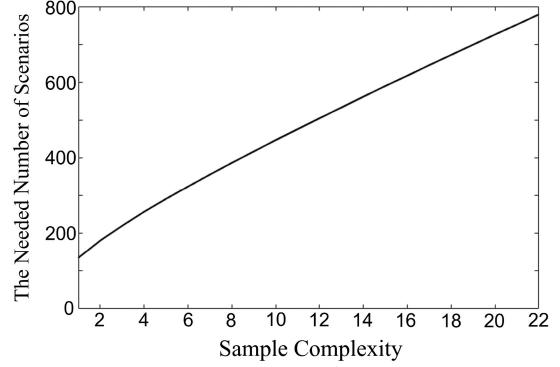


Fig. 9. Relationship between the needed number of scenarios and sample complexity.

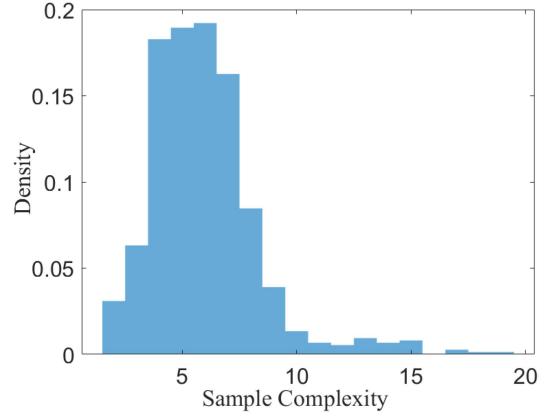


Fig. 10. Density function of the sample complexity over the 744 dispatch intervals.

779, which is much larger than the 324 needed scenarios under 6 sample complexity.

Instead of using the a-prior approach, we apply Algorithm 3 incrementally tuning risk for finding the exact sample complexity and the needed number of scenarios. Because the wind output and demand are changing, the sample complexity needs to be updated for each dispatch interval. Fig. 10 shows the empirical distribution of the sample complexity of the 744 dispatch intervals during August 16:00-18:00 we studied, where the average sample complexity (5.8) is much smaller than the number of decision variables (22). This evidence implies that in the previous *sample and discard* risk-tuning method [16], where the sample complexity is first supposed to be the number of decision variables 22, the decision makers may need to discard some redundant scenarios to achieve the same performance as our incremental method.

Table II compares the initial and final input sample size of these two risk-tuning methods when sample complexity is 6. Compared with *sample and discard*, incremental optimization guarantees the minimum input sample size with fewer iteration times, which is beneficial in the situation of limited high-accuracy data. Fig. 11 shows more details about the risk-tuning process of these two methods. The incremental optimization

TABLE II  
COMPARISON OF THE RISK TUNING PROCESS

Risk Tuning Method	Sample and Discard [26]	Incremental Optimization
Initial Input Sample Size	779	135
Initial Risk Level	0.021	0.117
Final Input Sample Size	771	324
Iteration Times	8	6

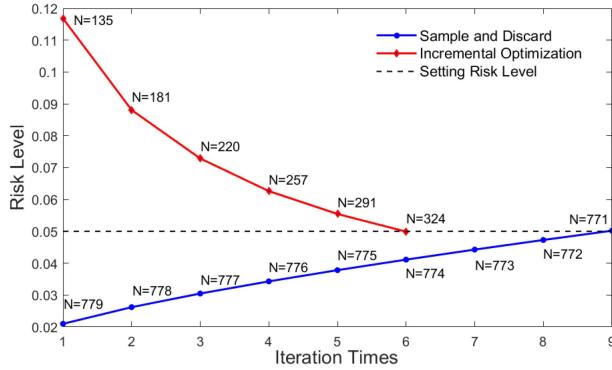


Fig. 11. Risk tuning details of two different methods.

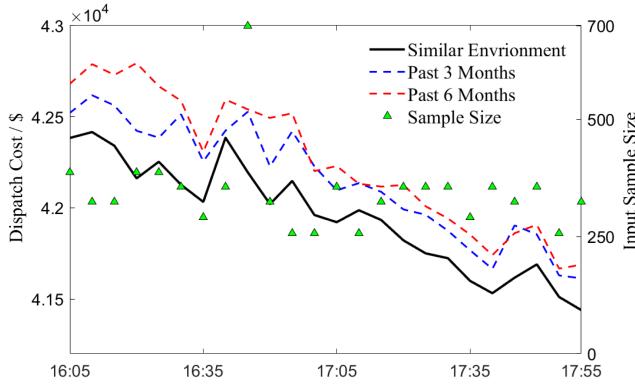


Fig. 12. Dispatch cost by sampling from different parameter space and input sample size between 16:00 and 18:00 on August 2nd, 2022 (24-bus system, high wind day).

method's efficiency results in fewer iteration times and less sample size in each iteration.

*Remark:* The final input sample size in the sample and discard method is much larger than the incremental method because only support scenarios are discarded in [26] but not randomly, and this risk bound is proved to be not tight in recent research [46].

So far, the simulation results have provided insights into the specifics of scenario generation and risk-tuning processes. Subsequent simulations will focus on examining the cost and actual violation outcomes following the scenario-based optimization model (3).

After setting the same risk and confidence parameter as Fig. 9, we can input *different* size of scenarios to meet the risk requirement based on the sample complexity of each dispatch interval, which is more efficient than the traditional sample and discarding method. Fig. 12 illustrates the adaptive input sample size (green triangle) between 16:00 and 18:00 on August 2nd, 2022, where

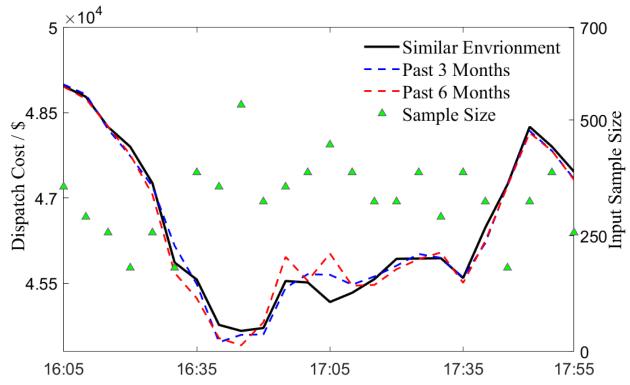


Fig. 13. Dispatch cost by sampling from different parameter space and input sample size between 16:00 and 18:00 on August 30th, 2022 (24-bus system, low wind day).

the dispatch cost of sampling from different parameter spaces is also compared.<sup>3</sup> During these two hours, the average wind power output is 23% of the total wind generation capacity, which is a relatively high wind period during August rush hours. It is clear that sampling from a similar environment space results in lower costs in each dispatch interval than others.

It needs to be clarified that because of the stochastic property, sampling from a similar environment does not ensure a lower cost in each dispatch interval, especially on a low wind day. For example, the average wind power output is only 13% of the total wind generation capacity between 16:00 and 18:00 on August 30th, which makes the economic benefits of sampling from a similar environment not comparable with the others (Fig. 13).

However, sampling from a similar environment builds a more precise model of the uncertainty variables, which results in a more trustworthy risk guarantee of the final solution. Table III compares the testing risk and the average cost of each 5-min dispatch interval by sampling from different parameter spaces during the rush hours 16:00 to 18:00 in August, where sampling from similar environment parameter space results in both lower testing violations and less dispatch cost. Meanwhile, when randomly sampling from the past half year, the testing violation 0.054 (the bold value in Table III) even exceeds the setting risk 0.05, which invalidates the risk-guarantee property of the scenario approach.

### C. 118-Bus System

We apply our method to a larger system in this section, based on the test case *c118swf.m* in MATPOWER [47]. This system includes 118 nodes, 210 lines, and 52 generators, 11 of which are modeled as wind farms. To address the influence of wind uncertainty on the economic dispatch, we replace the 4 storage units with the same capacity wind farms. Meanwhile, the transmission line capacities are set to be 60% of the original value to introduce more congestion in the system.

<sup>3</sup>The dispatch cost is calculated after each dispatch with true wind output data, and the searching space of a similar environment is set to the past 3 months.

TABLE III  
VIOLATION AND AVERAGE DISPATCH COST OF SAMPLING FROM DIFFERENT PARAMETER SPACES (24-BUS SYSTEM)

	Sampling Space	Past Half Year	Past Three Months	Similar Environment
Whole August	<b>Setting Violation</b>	0.05	0.05	0.05
	<b>Actual Violation</b>	<b>0.054</b>	0.040	0.036
	Average Cost( $10^4$ \$)	4.7522	4.7512	4.7444
	Average Solving Time(s)	0.181	0.166	0.165
	Average Sampling Time(s)	0.015	0.014	0.020
2nd August (high wind)	Average Cost( $10^4$ \$)	4.2257	4.2154	4.1956
30th August (low wind)	Average Cost( $10^4$ \$)	4.6414	4.6430	4.6433

TABLE IV  
VIOLATION AND AVERAGE DISPATCH COST OF SAMPLING FROM DIFFERENT PARAMETER SPACES (118-BUS SYSTEM)

	Sampling Space	Past Half Year	Past Three Months	Similar Environment
Whole August	<b>Setting Violation</b>	0.05	0.05	0.05
	<b>Actual Violation</b>	0.042	0.035	0.035
	Average Cost( $10^5$ \$)	1.7445	1.7454	1.7414
	Average Solving Time(s)	13.66	13.41	13.30
	Average Sampling Time(s)	0.022	0.021	0.028
2nd August (high wind)	Average Cost( $10^5$ \$)	1.5294	1.5435	1.5173
30th August (low wind)	Average Cost( $10^5$ \$)	1.7251	1.7255	1.7271

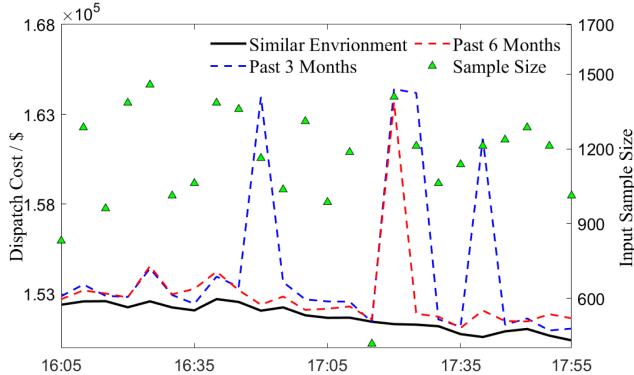


Fig. 14. Dispatch cost by sampling from different parameter space and input sample size between 16:00 and 18:00 on August 2nd, 2022 (118-bus system, high wind day).

Similar to the simulation in the 24-bus system, the wind power forecasting and real value are directly modeled by the data from ERCOT with 7 wind generators from south Texas and 8 wind generators from west Texas. Fig. 14 shows the adaptive input sample size and the dispatch cost of sampling from different parameter spaces between 16:00 and 18:00 on August 2nd, 2022. In this period, the average wind power output accounts for 24% of the total wind generation capacity, a relatively high wind output level during August peak hours.

Benefiting from the smaller uncertainty space (see Figs. 6, 7), sampling from a similar environment helps the decision maker avoid some odd scenarios when applying scenario approach in the real world. These odd scenarios typically make the final decision more conservative, which can be illustrated by the cost spikes in Fig. 14 when we directly sample scenarios from the past.

Table IV lists the testing risk and the average cost of 5-min dispatch during August peak hours (16:00–18:00), where the result has a similar pattern with the 24-bus system. The 5-min short-term wind power forecasting error is relatively lower than other long-term forecasting errors, but a more efficient scenarios generation model will make the chance-constrained dispatch solution more trustworthy.

Although there's no theory guarantee for a lower cost solution by sampling from a similar environment, the simulation shows the economic advantage of the proposed efficient scenarios generation model, especially during high wind situations. To quantify the relationship between the accuracy of the scenario generation model and economic benefits will be one possible direction of future research.

Regardless of the selection of sample space, the number of needed scenarios is the same, which results in a similar problem-solving time of three different sample spaces in both Tables III and IV. Meanwhile, computing *indicator vector* (7) of similar environments increases the sampling time, but it's far less than the solving time by one or more orders of magnitude.

## VI. CONCLUSION

This article studies the main two barriers to applying the scenario approach to the economic dispatch with high penetration of renewable resources, i.e. lack of accurate scenario generation models and inefficient risk tuning process. Leveraging correlation analysis, we generate scenarios via an environment filter with empirical distribution closer to the true probability measure. After embedding this scenario generation model with the incremental scenario optimization algorithm, we propose an efficient risk-tuning scheme, which can solve the optimal solution meeting risk requirement with minimum data size and provide other higher-risk solutions to system operators in meanwhile. Case studies based on real-world wind data and modified

IEEE benchmark systems show the effectiveness and advantages of our methods.

Directly generating scenarios from past experience may be a naive approach, but it works well in practice when the needed data size is much smaller than the size of the past data pool. Future work includes (1) comparing the results of using other scenario generation methods, such as the generative model in the field of machine learning; (2) extending the economic dispatch problem to a multi-stage framework, i.e. look-ahead economic dispatch, where varying risk tolerance levels are allowed across different time horizons to enhance the adaptability; and (3) applying the proposed scheme to unit commitment and other non-convex decision-making processes in electric power systems, which requires efficient handling of discrete variables under incremental scenarios to ensure timely solutions.

## REFERENCES

- [1] ERCOT, "Information on generation resources," 2023. [Online]. Available: <https://www.ercot.com/gridinfo/generation>
- [2] L. A. Roald, D. Pozo, A. Papavasiliou, D. K. Molzahn, J. Kazempour, and A. Conejo, "Power systems optimization under uncertainty: A review of methods and applications," *Electric Power Syst. Res.*, vol. 214, 2023, Art. no. 108725.
- [3] A. Prékopa, *Stochastic Programming*. Berlin, Germany: Springer Science & Business Media, 2013, vol. 324.
- [4] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust Optimization*, vol. 28. Princeton, NJ, USA: Princeton University Press, 2009.
- [5] X. Geng and L. Xie, "Data-driven decision making in power systems with probabilistic guarantees: Theory and applications of chance-constrained optimization," *Annu. Rev. Control*, vol. 47, pp. 341–363, 2019.
- [6] D. Bienstock, M. Chertkov, and S. Harnett, "Chance-constrained optimal power flow: Risk-aware network control under uncertainty," *SIAM Rev.*, vol. 56, no. 3, pp. 461–495, 2014.
- [7] L. Roald and G. Andersson, "Chance-constrained AC optimal power flow: Reformulations and efficient algorithms," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2906–2918, May 2018.
- [8] Y. Zhang, S. Shen, and J. L. Mathieu, "Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1378–1388, Mar. 2017.
- [9] B. K. Poolla, A. R. Hota, S. Bolognani, D. S. Callaway, and A. Cherukuri, "Wasserstein distributionally robust look-ahead economic dispatch," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2010–2022, May 2021.
- [10] J. Luedtke and S. Ahmed, "A sample approximation approach for optimization with probabilistic constraints," *SIAM J. Optim.*, vol. 19, no. 2, pp. 674–699, 2008.
- [11] M. C. Campi and S. Garatti, "The exact feasibility of randomized solutions of uncertain convex programs," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1211–1230, 2008.
- [12] A. Bagheri, C. Zhao, and Y. Guo, "Data-driven chance-constrained stochastic unit commitment under wind power uncertainty," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2017, pp. 1–5.
- [13] A. N. Madavan, N. Dahlin, S. Bose, and L. Tong, "Risk-based hosting capacity analysis in distribution systems," *IEEE Trans. Power Syst.*, vol. 39, no. 1, pp. 355–365, Jan. 2024.
- [14] P. Jirutitijaroen and C. Singh, "Reliability constrained multi-area adequacy planning using stochastic programming with sample-average approximations," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 504–513, May 2008.
- [15] H. Ming, L. Xie, M. C. Campi, S. Garatti, and P. Kumar, "Scenario-based economic dispatch with uncertain demand response," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1858–1868, Mar. 2019.
- [16] M. S. Modarresi et al., "Scenario-based economic dispatch with tunable risk levels in high-renewable power systems," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 5103–5114, Nov. 2019.
- [17] X. Geng, L. Xie, and M. S. Modarresi, "Computing essential sets for convex and nonconvex scenario problems: Theory and application," *IEEE Trans. Control Netw. Syst.*, vol. 9, no. 1, pp. 269–281, Mar. 2022.
- [18] M. Vrakopoulou, K. Margellos, J. Lygeros, and G. Andersson, "A probabilistic framework for reserve scheduling and N-1 security assessment of systems with high wind power penetration," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3885–3896, Nov. 2013.
- [19] S. Yan, F. Parise, and E. Bitar, "Data-driven approximations of chance constrained programs in nonstationary environments," *IEEE Contr. Syst. Lett.*, vol. 6, pp. 2671–2676, 2022.
- [20] X. Geng and L. Xie, "Chance-constrained unit commitment via the scenario approach," in *Proc. North Amer. Power Symp.*, 2019, pp. 1–6.
- [21] M. C. Campi, A. Carè, and S. Garatti, "The scenario approach: A tool at the service of data-driven decision making," *Annu. Rev. Control*, vol. 52, pp. 1–17, 2021.
- [22] H. Bludszuweit, J. A. Domínguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 983–991, Aug. 2008.
- [23] B.-M. Hodge, A. Florita, K. Orwig, D. Lew, and M. Milligan, "Comparison of wind power and load forecasting error distributions," National Renewable Energy Lab. (NREL), Golden, CO, USA, Tech. Rep. NREL/CP-5500-54384, 2012.
- [24] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845–1856, Nov. 2010.
- [25] N. Zhang, C. Kang, Q. Xia, and J. Liang, "Modeling conditional forecast error for wind power in generation scheduling," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1316–1324, May 2014.
- [26] M. C. Campi and S. Garatti, "A sampling-and-discriminating approach to chance-constrained optimization: Feasibility and optimality," *J. Optim. Theory Appl.*, vol. 148, no. 2, pp. 257–280, 2011.
- [27] S. Garatti, A. Carè, and M. C. Campi, "Complexity is an effective observable to tune early stopping in scenario optimization," *IEEE Trans. Autom. Control*, vol. 68, no. 2, pp. 928–942, Feb. 2023.
- [28] P. S. Kundur and O. P. Malik, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill Education, 2022.
- [29] Q. Liu and M. D. Ilić, "Enhanced automatic generation control (E-AGC) for future electric energy systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2012, pp. 1–8.
- [30] N. Li, C. Zhao, and L. Chen, "Connecting automatic generation control and economic dispatch from an optimization view," *IEEE Trans. Control Netw. Syst.*, vol. 3, no. 3, pp. 254–264, Sep. 2016.
- [31] M. Lange, "On the uncertainty of wind power predictions—analysis of the forecast accuracy and statistical distribution of errors," *J. Sol. Energy Eng.*, vol. 127, no. 2, pp. 177–184, 2005.
- [32] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, and Z. Yan, "A review on the forecasting of wind speed and generated power," *Renewable Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, 2009.
- [33] J. Miettinen, H. Holttilin, and B.-M. Hodge, "Simulating wind power forecast error distributions for spatially aggregated wind power plants," *Wind Energy*, vol. 23, no. 1, pp. 45–62, 2020.
- [34] S. Hanifi, X. Liu, Z. Lin, and S. Lotfian, "A critical review of wind power forecasting methods—Past, present and future," *Energies*, vol. 13, no. 15, 2020, Art. no. 3764.
- [35] M. C. Campi, S. Garatti, and F. A. Ramponi, "A general scenario theory for nonconvex optimization and decision making," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4067–4078, Dec. 2018.
- [36] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
- [37] V. Vovk, "Conditional validity of inductive conformal predictors," in *Proc. Asian Conf. Mach. Learn.*, 2012, pp. 475–490.
- [38] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, "Learning to rank: From pairwise approach to listwise approach," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 129–136.
- [39] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.
- [40] Iowa State University, "Iowa environmental mesonet: Asos-awos-metar data download," 2022. [Online]. Available: <https://mesonet.agron.iastate.edu/request/download.phtml>
- [41] B. Mauch, J. Apt, P. M. Carvalho, and P. Jaramillo, "What day-ahead reserves are needed in electric grids with high levels of wind power?," *Environ. Res. Lett.*, vol. 8, no. 3, 2013, Art. no. 034013.

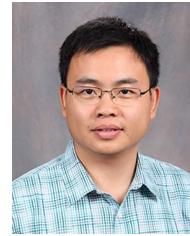
- [42] C. Grigg et al., "The IEEE reliability test system-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.
- [43] C. Ordoudis, P. Pinson, J. M. Morales, and M. Zugno, "An updated version of the IEEE RTS 24-bus system for electricity market and power system operation studies," Tech. Univ. Denmark, Lyngby, Denmark, 2016.
- [44] M. Zugno and A. J. Conejo, "A robust optimization approach to energy and reserve dispatch in electricity markets," *Eur. J. Oper. Res.*, vol. 247, no. 2, pp. 659–671, 2015.
- [45] D. Millstein, R. Wiser, A. D. Mills, M. Bolinger, J. Seel, and S. Jeong, "Solar and wind grid system value in the United States: The effect of transmission congestion, generation profiles, and curtailment," *Joule*, vol. 5, no. 7, pp. 1749–1775, 2021.
- [46] L. Romao, A. Papachristodoulou, and K. Margellos, "On the exact feasibility of convex scenario programs with discarded constraints," *IEEE Trans. Autom. Control*, vol. 68, no. 4, pp. 1986–2001, Apr. 2023.
- [47] C. E. Murillo-Sánchez, R. D. Zimmerman, C. L. Anderson, and R. J. Thomas, "Secure planning and operations of systems with stochastic sources, energy storage, and active demand," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 2220–2229, Dec. 2013.



**Qian Zhang** (Graduate Student Member, IEEE) received the B.E. and M.S. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2019 and 2022, respectively. He is currently working toward the Ph.D. degree with Texas A&M University, College Station, TX, USA. His research interests include machine learning, optimization in the electricity market, and power system stability and control.



**Apurv Shukla** (Member, IEEE) received the bachelor's degree in manufacturing science and engineering from Indian Institute of Technology Kharagpur, Kharagpur, India, in 2016, and the Ph.D. degree in operations research from Columbia University, New York, NY, USA, in 2022. He is currently a Postdoctoral Associate with Texas A&M University, College Station, TX, USA. His research focuses on systems and control.



**Le Xie** (Fellow, IEEE) received the B.E. degree in electrical engineering from Tsinghua University, Beijing, China, in 2004, the M.S. degree in engineering sciences from Harvard University, Cambridge, MA, USA, in 2005, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA, in 2009. He is currently a Professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA. His research interests include modeling and control of large-scale complex systems, smart grid applications with renewable energy resources, and electricity markets.