Sample-Based Neural Approximation Approach for Probabilistic Constrained Programs

Xun Shen, *Member, IEEE*, Tinghui Ouyang, Nan Yang, and Jiancang Zhuang

***Abstract*— This article introduces a neural approximation-based method for solving continuous optimization problems with probabilistic constraints. After reformulating the probabilistic constraints as the quantile function, a sample-based neural network model is used to approximate the quantile function. The statistical guarantees of the neural approximation are discussed by showing the convergence and feasibility analysis. Then, by introducing the neural approximation, a simulated annealing-based algorithm is revised to solve the probabilistic constrained programs. An interval predictor model (IPM) of wind power is investigated to validate the proposed method.**

***Index Terms*— Neural network model, nonlinear optimization, probabilistic constraints, quantile function, sample average approximation.**

I. Introduction

PROBABILISTIC-constrained programs (PCPs) or chance-constrained programs (CCPs) are mathematical programs involved random variables in constraints, which is required to be satisfied in a given probability level [1]. In this study, the appellation of PCPs is adopted. PCPs have been used in various fields, such as decision-making and filtering in uncertain systems [2], motion planning with probabilistic constraints [5], engine combustion control [3], probabilistic bound estimation of uncertain state trajectories [4], and energy storage modeling toward distribution systems in electricity market [6].

PCPs have a great value for practical applications. However, it is NP-hard to solve the PCPs directly due to the existence of probabilistic constraints. Thus, a lot of research has been conducted to develop an approximation approach to solve the PCPs. For instance, Calariore and Campi [7] proposed the scenario approach in which the probabilistic constraints are replaced by deterministic constraints imposed for finite sets of independently extracted samples of random variables. The scenario approach ensures that the solution of the approximately formulated deterministic program satisfies the original probabilistic constraints with a determined bound of probability. The bound of probability is determined by the number of extracted samples. Afterward, the scenario approach has been improved to ensure more tight confidence bounds of probability by a sampling-and-discarding approach [8]. In the sampling-and-discarding scenario approach, a certain proportion of the extracted samples are used to define the deterministic constraints, while the rest ones are discarded. With a less sample number, the violation probabilities of the original probabilistic constraints are still preserved. However, the scenario approach has a fatal drawback. It gives a too conservative solution, which converges to the totally robust solution of the uncertain constraints rather than the original probabilistic constraints. Namely, as the sample number becomes infinity, the solution will satisfy the uncertain constraints with probability 1 rather than the given level of probability.

On the other hand, a sample average approach has been proposed in [9] and [10] to satisfy the probabilistic constraints strictly. The sample average program is formulated as an approximation of the PCP by replacing the probabilistic constraints with a measure to indicate the violation probability. Besides, an inner–outer approximate approach is proposed in [11] to get a tight sample average approximate program. A randomized optimization-based method is used to solve PCPs in [12].

This article extends the sample average approach by introducing neural network-based constraints to replace probabilistic constraints. After reformulating the probabilistic constraints as the quantile function, a sample-based neural network model is used to approximate the quantile function. The statistical guarantees of the neural approximation are discussed by showing the convergence and feasibility analysis. Then, by introducing the neural approximation, a simulated annealing-based algorithm is revised to solve the PCPs. An interval predictor model (IPM) of wind power is investigated to validate the proposed method.

This article is organized as follows. Section II gives the problem formulation. Section III demonstrates the proposed neural approximation approach and the proposed algorithm is introduced in Section IV. Section V presents how to use the proposed method to establish IPM of wind power. Finally, Section VI concludes this article.

Manuscript received May 29, 2020; revised January 3, 2021 and June 6, 2021; accepted July 31, 2021. This work was supported in part by the Graduate University for Advanced Studies, SOKENDAI. *(Corresponding author: Xun Shen.)*

Xun Shen is with the Department of Statistical Sciences, Graduate University for Advanced Studies, SOKENDAI, Tokyo 106-8569, Japan (e-mail: shenxun@ism.ac.jp).

Tinghui Ouyang is with the National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan (e-mail: ouyang.tinghui@aist.go.jp).

Nan Yang is with the Department of Hubei Provincial Collaborative Innovation Center for New Energy Microgrid, China Three Gorges University, Yichang 443002, China (e-mail: ynyyayy@ctgu.edu.cn).

Jiancang Zhuang is with The Institute of Statistical Mathematics, Tokyo 190-8562, Japan (e-mail: zhuangjc@ism.ac.jp).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3102323.

Digital Object Identifier 10.1109/TNNLS.2021.3102323

II. Problem Formulation

Consider the following optimization problem with probabilistic constraints:

$$\begin{aligned} &\min_{u\in\mathscr{U}} \ J(u) \\ &\text{s.t. } \Pr\{h(u,\delta)\le 0\}\ge 1-\alpha, \quad \delta\in\Delta,\ \alpha\in(0,1) \end{aligned} \tag{1}$$

where u represents the decision variable with a closed and compact feasible domain $\mathscr{U}\subseteq\mathbb{R}^{n_u}$, $J:\mathbb{R}^{n_u}\to\mathbb{R}$ is the objective function, δ represents an uncertain parameter vector with sample space $\Delta\subseteq\mathbb{R}^{n_\delta}$, probability measure is well defined on the $\mathcal{B}_\Delta$ (the sigma algebra of Δ) and the probability density of δ is denoted as $p_\delta(\delta)$, $h:\mathbb{R}^{n_u}\times\mathbb{R}^{n_\delta}\to\mathbb{R}^{n_h}$ is a vector-valued function, and α is a risk parameter. The notation of the probabilistic constraints is that the constraints represented by $h(u,\delta)\le 0$ should be satisfied with a probability larger than $1-\alpha$. Besides, this study focuses on problems in which J and h are continuous and differentiable $\forall u\in\mathcal{U}$ and $\forall\delta\in\Delta$. The uncertain variable $h(u,\delta)$ has a continuous cumulative distribution function (CDF) for all $u\in\mathcal{U}$.

There are several difficulties to address problem described by (1).

1) The structural properties of the feasible domain defined by $h(u,\delta)\le 0$ may not succeed to the domain defined by the constraints $\Pr\{h(u,\delta)\le 0\}\ge 1-\alpha$. For instance, even if h are all linear in u, the probabilistic constraint may not define a convex domain.
2) Instead of knowing the distribution of δ, only the samples of δ are available.
3) The tractable analytical function of probabilistic constraint does not exist even with the knowledge of the distribution of δ.

This study is to find out a tractable analytical function $H(u)$ to define a feasible domain $\tilde{\mathcal{U}}_f=\{u\in\mathcal{U}|H(u)\le 0\}$. The feasible domain $\tilde{\mathcal{U}}_f$ is equal to the feasible domain $\mathcal{U}_f=\{u\in\mathcal{U}|\Pr\{h(u,\delta)\le 0\}\ge 1-\alpha\}$ with probability as 1. Then, the problem with deterministic constraints

$$\begin{aligned} &\min_{u\in\mathscr{U}} \ J(u) \\ &\text{s.t. } H(u)\le 0 \end{aligned} \tag{2}$$

is the equivalent problem of (1). By solving (2), the optimal solution of (1) can be approximated.

III. Neural Approximation Approach

A. Quantile Function-Based Problem Reformulation

The cumulative probability function of a random variable X is denoted as

$$F(x)=\Pr\{X\le x\}. \tag{3}$$

While the $1-\alpha$ level quantile of a random variable X is defined as [13]

$$Q^{1-\alpha}(X)=\inf\{x\in\mathbb{R}|\Pr\{X\le x\}\ge 1-\alpha\}. \tag{4}$$

For the case $n_h=1$, $h:\mathbb{R}^{n_x}\times\mathbb{R}^{n_\delta}\to\mathbb{R}$ is a real-valued function and $h(u,\delta)$ is a scalar random variable. Thus, the cumulative probability function of $h(u,\delta)$ can be defined as

$$F(\gamma,u)=\Pr\{h(u,\delta)\le\gamma\}, \quad \gamma\in\mathbb{R}. \tag{5}$$

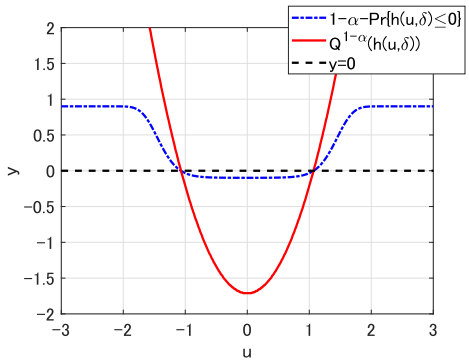

Fig. 1. Comparison of $1-\alpha-\Pr\{h(u,\delta)\le 0\}$ and $Q^{1-\alpha}(h(u,\delta))$, where $h(u,\delta)=1.5u^2-3+\delta$ and $\delta\sim N(0,1)$ ($\alpha=0.1$).

The quantile $Q^{1-\alpha}(h(u,\delta))\le 0$ is equivalent to $F(0,u)\ge 1-\alpha$. For the case $n_h>1$, $Q^{1-\alpha}(h(u,\delta)\le 0$ cannot be well defined since $h(u,\delta)$ is not a scalar random variable any more. Instead of using $h(u,\delta)$, $\bar{h}(u,\delta)=\max_{j=1,\dots,n_h}h_j(u,\delta)$ is used. Then, the cumulative probability function $F(\gamma,u)$ notes the same style as (5) only replacing $h(\cdot,\cdot)$ by $\bar{h}(\cdot,\cdot)$. Consequently, the quantile is written as $Q^{1-\alpha}(\bar{h}(u,\delta)\le 0$. Without losing the generality, $Q^{1-\alpha}(\bar{h}(u,\delta)\le 0$ can be used for $n_h=1$ as well. Then, the following reformulation of problem (1) is written as:

$$\begin{aligned} &\min_{u\in\mathscr{U}} \ J(u) \\ &\text{s.t. } Q^{1-\alpha}\big(\bar{h}(u,\delta)\big)\le 0, \quad \delta\in\Delta,\ \alpha\in(0,1). \end{aligned} \tag{6}$$

The advantage of using the quantile is that the quantile function is much less flat compared to the probability function $\Pr\{h(u,\delta)\le 0\}$. The comparison example of the quantile function and probability function is shown in Fig. 1. By the quantile function-based reformulation, the feasible region is measured in the image of $\bar{h}(u,\delta)$ instead of the bounded image $[0,1]$ of the probability function.

B. Neural Approximation of Quantile Function

The sample set $\Delta^N=\{\delta_1,\dots,\delta_N\}$ is obtained by extracting samples independently of Δ according to the identical distribution $p_\delta(\delta)$. For a given u, the empirical CDF of $\mathcal{H}^N=\{h(u,\delta_1),\dots,h(u,\delta_N)\}$ is written as

$$\tilde{F}^N(t,u)=\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\big(\bar{h}(u,\delta_i)\le t\big) \tag{7}$$

where $\mathbb{I}(h(u,\delta_i)\le t)$ denotes the indicator function written as

$$\mathbb{I}\big(\bar{h}(u,\delta_i)\le t\big)=\begin{cases}0, & \text{if } \bar{h}(u,\delta_i)>t\\ 1, & \text{if } \bar{h}(u,\delta_i)\le t.\end{cases} \tag{8}$$

The $(1-\alpha)$-empirical quantile at u can be obtained from a value t such that $\tilde{F}^N(t,u)\approx 1-\alpha$, which is defined as

$$\tilde{Q}^{1-\alpha}\big(\bar{h}(u,\delta)\big)=\inf\left\{y\,\middle|\,\frac{1}{N}\sum_{i=1}^{N}\mathbb{I}\big(\bar{h}(u,\delta_i)\le y\big)\ge 1-\alpha\right\} \tag{9}$$

equivalently written as

$$\tilde{Q}^{1-\alpha}\big(\bar{h}(u,\delta)\big)\big)=\bar{h}_{\lceil M\rceil}(u) \tag{10}$$

where $M = \lceil (1-\alpha)N \rceil$ and $\bar{h}_{\lceil M \rceil}(u)$ denotes the Mth smallest observation of the values $\mathcal{H}^N$ for a given u.

We can use the $(1-\alpha)-$empirical quantile as an approximation of the probabilistic constraints and form the following approximate problem:

$$\begin{aligned} &\min_{u \in \mathcal{U}} \ J(u) \\ &\text{s.t. } \tilde{Q}^{1-\alpha}\big(\bar{h}(u,\delta)\big) \le 0, \quad \delta \in \Delta, \ \alpha \in (0,1). \end{aligned} \tag{11}$$

Using the $(1-\alpha)$-empirical quantile as an approximation of the probabilistic constraints has two drawbacks.

1) $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta)))$ is usually not differentiable. M changes for different values of u. Thus, even if the constraint $h(u,\delta)$ is smooth for a fixed value of δ, $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta)))$ does not have to be smooth.
2) For small N, the uncertainty of $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta)))$ is large and the inward kinks of the feasible boundary will be very nonsmooth. Large N can bring some smoothness back, and however, this also increases the computation burden.

Thus, it is necessary to look for the smooth approximation of $(1-\alpha)$-empirical quantile.

Given the sample set Δ_N, the $(1-\alpha)$-empirical quantile $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta)))$ is essentially a function of u. $Q^{1-\alpha}(\bar{h}(u,\delta)))$ is also a function of u. By using $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta)))$ as an approximation of $Q^{1-\alpha}(\bar{h}(u,\delta)))$, a smooth function can be used to approximate the $(1-\alpha)$ quantile $Q^{1-\alpha}(\bar{h}(u,\delta)))$. In this study, single-layer neural network model is used to approximate the $(1-\alpha)$ quantile $Q^{1-\alpha}(\bar{h}(u,\delta)))$. Using N independently extracted samples of δ, the neural approximation of $Q^{1-\alpha}(\bar{h}(u,\delta)))$ with S hidden nodes and activation function $g(\cdot)$ is defined as

$$\hat{H}_S(u) = \sum_{i=1}^{S} \beta_i g(u, a_i, b_i) \tag{12}$$

where β_i denotes the weight vector connecting the ith hidden node and the output nodes, $a_i = [a_{i,1}, \ldots, a_{i,k}]$ represents the weight vector toward u, and b_i is the scalar threshold of the ith hidden node. Here, the activation function adopts the sigmoid function expressed as

$$g(u, a_i, b_i) = \frac{1}{1 + e^{-a_i^T u + b_i}}. \tag{13}$$

The purpose is to achieve

$$\hat{H}_S(u) \approx Q^{1-\alpha}\big(\bar{h}(u,\delta)\big)\big). \tag{14}$$

Then, $\hat{H}_S(\cdot)$ maps the decision variable $u \in \mathbb{R}^k$ into the image of $Q^{1-\alpha}(\bar{h}(u,\delta)))$ which is $\mathbb{R}$. In order to find solutions for (1) and (6), it is equivalent to solve the following problem:

$$\begin{aligned} &\min_{u \in \mathcal{U}} \ J(u) \\ &\text{s.t. } \hat{H}_S(u) \le 0. \end{aligned} \tag{15}$$

The above formulation is a sample-based neural approximation of the original PCP, which is obtained by a two-layer approximation: sample approximation and neural approximation. The convergence and feasibility of the two-layer approximation should be analyzed.

C. Convergence and Feasibility Analysis

We make the following assumptions.

Assumption 1: $\bar{h}(u,\delta)$ is a Carathéodory function. For every fixed $u \in \mathcal{U}$, $\bar{h}(u,\delta)$ is a continuous random variable, which is measurable and has a continuous distribution. For every fixed $\delta \in \Delta$, $\bar{h}(u,\delta)$ is a continuous function of u. Besides, for every $u \in \mathcal{U}$, $F_{\bar{h}}(\bar{h}(u,\delta))$, the CDF of $\bar{h}(u,\delta)$, is continuously differentiable and it has strictly positive derivative of $\bar{h}(u,\delta)$ (strictly monotonically increasing), $f_{\bar{h}}(\bar{h}(u,\delta))$, over the domain of $\bar{h}(u,\delta)$: $(-\infty, +\infty)$.

Remark 1: The CDF $F_{\bar{h}}(\bar{h}(u,\delta))$ is continuously and strictly monotonically increasing over $(-\infty, +\infty)$. Thus, the quantile function $Q(\bar{h}(u,\delta)) : [0,1] \to D_{\bar{h}(u,\delta)}$ is the inverse of $F_{\bar{h}}$ and thus continuous on $[0,1]$. Moreover, for a fixed $\delta \in \Delta$, $\bar{h}(\cdot,\delta)$ is continuous on $\mathcal{U}$. Thus, the values of the CDF $F_{\bar{h}}(\bar{h}(u,\delta))$ and the quantile function $Q(\bar{h}(u,\delta))$ vary continuously on $\mathcal{U}$ due to the fact that continuous functions' function composition leads to continuous function. Thus, Assumption 1 implies that the $1-\alpha$ level quantile $Q^{1-\alpha}(\bar{h}(u,\delta))$ is a continuous function of u on $\mathcal{U}$.

Assumption 2: The problem expressed by (1) has a globally optimal solution u^*, such that for any ϵ_u, there is $u \in \mathcal{U}$ such that $\|u - u^*\| \le \epsilon_u$ and $F(0,u) > 1-\alpha$ (or equivalently $Q^{1-\alpha}(\bar{h}(u,\delta)) < 0$).

Remark 2: Assumption 2 implies that there exists a sequence $\{u_k\}_{k=1}^{\infty} \subseteq \mathcal{U}$ that converges to an optimal solution u^* such that $F(0,u_k) > 1-\alpha$ (or equivalently $Q^{1-\alpha}(\bar{h}(u_k, \delta)) < 0$) for all $k \in \mathbb{N}$.

By applying [14, Corollary 21.5], we have the following lemma about the convergence of $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta))$.

Lemma 1: Suppose that Assumption 1 holds. For any fixed $\alpha \in (0,1)$ and any fixed $u \in \mathcal{U}$, if N samples of δ, δ_i, $i = 1, \ldots, N$, are independently extracted, then

$$\tilde{Q}^{1-\alpha} - Q^{1-\alpha} = -\frac{1}{N} \sum_{i=1}^{N} \frac{\mathbb{I}\big(\bar{h}_i \le Q^{1-\alpha}\big) - (1-\alpha)}{f\big(Q^{1-\alpha}\big)} + o(1) \tag{16}$$

where $\tilde{Q}^{1-\alpha}$, $Q^{1-\alpha}$, and $\bar{h}_i$ are shorts for $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta))$, $Q^{1-\alpha}(\bar{h}(u,\delta))$, and $\bar{h}(u,\delta_i)$, respectively.
Consequently, the sequence $\{\tilde{Q}^{1-\alpha} - Q^{1-\alpha}\}$ is asymptotically normal with mean 0 and variance $(\alpha(1-\alpha))/(N \cdot f^2(Q^{1-\alpha}))$. As N increases to ∞, the variance also vanishes to zeros.

For the convergence of $\hat{H}_S(u)$ to $Q^{1-\alpha}(\bar{h}(u,\delta))$, we have the following theorem.

Theorem 1: Suppose that Assumption 1 holds, as $N \to \infty$, $\forall \epsilon_H > 0$ there exists S, a, b, and δ such that

$$\sup_{u \in \mathcal{U}_c} \left\| \hat{H}_S(u) - Q^{1-\alpha}\big(\bar{h}(u,\delta)\big) \right\| < \epsilon_H, \quad \text{w.p.1} \tag{17}$$

where $\mathcal{U}_c \subseteq \mathcal{U}$ represents any compact set inside the feasible area.

Proof (Theorem 1): Due to Assumption 1 and Remark 1, $Q^{1-\alpha}(\bar{h}(u,\delta))$ is the continuous function of u. Then, according to the universal approximation theorem [15], [16], $\forall \epsilon_H > 0$, $\forall Q^{1-\alpha}(\bar{h}(u,\delta))$, $\exists S \in \mathbb{N}^+, \beta, b \in \mathbb{R}^S$, and $a \in \mathbb{R}^{S \times k}$ such that

$$\left\| \hat{H}_S(u) - Q^{1-\alpha}\big(\bar{h}(u,\delta)\big) \right\| < \epsilon_H \tag{18}$$

for all $u \in \mathcal{U}$. However, $Q^{1-\alpha}(\bar{h}(u,\delta))$ is not used directly for fitting the quantile function. The used one is the approximate value $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta))$, which has bias compared to $Q^{1-\alpha}(\bar{h}(u,\delta))$. The bias is defined by (16). Denote $dQ^{1-\alpha} = \tilde{Q}^{1-\alpha} - Q^{1-\alpha}$. According to Lemma 1, $dQ^{1-\alpha}$ is normal with mean 0 and variance $(\alpha(1-\alpha))/(N \cdot f^2(Q^{1-\alpha}))$. As $N \to \infty$, $(\alpha(1-\alpha))/(N \cdot f^2(Q^{1-\alpha})) \to 0$. Namely, $\|dQ^{1-\alpha}\| \to 0$ with probability 1.

Since we can only obtain $\tilde{Q}^{1-\alpha}(\bar{h}(u,\delta))$, instead of (18) which needs $Q^{1-\alpha}(\bar{h}(u,\delta))$, the following inequality:

$$\left\|\hat{H}_S(u) - \tilde{Q}^{1-\alpha}\left(\bar{h}(u,\delta)\right)\right\| < \epsilon_H \tag{19}$$

can be obtained $\forall \epsilon_H > 0, \forall Q^{1-\alpha}(\bar{h}(u,\delta)), \exists S \in \mathbb{N}^+, \beta, b \in \mathbb{R}^S$, and $a \in \mathbb{R}^{S\times k}$. Since $\|\hat{H}_S(u) - Q^{1-\alpha}(\bar{h}(u,\delta))\| = \|\hat{H}_S(u) - \tilde{Q}^{1-\alpha}(\bar{h}(u,\delta)) + \tilde{Q}^{1-\alpha}(\bar{h}(u,\delta)) - Q^{1-\alpha}(\bar{h}(u,\delta))\| \le \|\hat{H}_S(u) - \tilde{Q}^{1-\alpha}(\bar{h}(u,\delta))\| + \|dQ^{1-\alpha}\|$, we have

$$\left\|\hat{H}_S(u) - Q^{1-\alpha}\left(\bar{h}(u,\delta)\right)\right\| < \epsilon_H + \left\|dQ^{1-\alpha}\right\|. \tag{20}$$

Namely, for any $\epsilon_H > 0$, we have parameters to satisfy (20). Notice that $\|dQ^{1-\alpha}\|$ becomes 0 with probability 1 as $N \to \infty$. Besides, for any N, there exist S, a, b, and δ to satisfy (20) for any $\epsilon_H > 0$. Thus, (17) is proven. □

Let A and A_N^S denote the sets of optimal solutions for problems (1) and (15), respectively. Notice that problem (6) shares the same set of optimal solutions as problem (1). Besides, denote J^* and J_N^S the optimal values of cost functions in problems (1) and (15), respectively. The convergences of J_N^S and A_N^S as N and S increase are summarized in the following.

Theorem 2: Suppose that $\mathcal{U}$ is compact, the function $J(\cdot)$ is continuous, and Assumptions 1 and 2 hold. Then, $J_N^S \to J^*$ and $\mathbb{D}(A_N^S, A) \to 0$ w.p.1.

Proof (Theorem 2): Due to Assumption 2, the set A is nonempty and there exists $u \in \mathcal{U}$ such that $Q^{1-\alpha}(\bar{h}(u,\delta)) < 0$. According to Theorem 1, $\hat{H}_S(u)$ converges to $Q^{1-\alpha}(\bar{h}(u,\delta)) < 0$, and thus, we can find S_0, a_0, b_0, β_0, and N_0 such that $\hat{H}_S(u) - \epsilon_H - \|dQ^{1-\alpha}\| \le 0$. ($\epsilon_H$ can be any small value to 0 and $\|dQ^{1-\alpha}\|$ decreases to 0 with probability 1 as $N \to \infty$.) Because $\hat{H}_S(u)$ is continuous in u and $\mathcal{U}$ is compact, the feasible set of the approximation problem (15) is compact as well. Besides, A_N^S is not empty w.p.1 for all $N \le N_0$ and all S, a, b, β, and N for smaller error bound (with higher probability) $\epsilon_H + \|dQ^{1-\alpha}\|$ of (20). Here, we denote $\hat{H}_{S,0}(u)$ for $\hat{H}_S(u)$ with parameters $\{S_0, a_0, b_0, \beta_0, N_0\}$ and input u. Analogously, $\hat{H}_{S,k}(u_k)$ is for $\hat{H}_S(u)$ with parameters $\{S_k, a_k, b_k, \beta_k, N_k\}$ and input u_k. With parameters $\{S_k, a_k, b_k, \beta_k, N_k\}$, the corresponding notations for the optimal solution set and optimal cost value are generally defined as $A_{N,k}^{S,k}$ and $J_{N,k}^{S,k}$.

Let $\{N_k\}_{k=1}^{\infty} \ge N_0$ and $\{S_k, a_k, b_k, \beta_k, N_k\}$ be two sequences such that the corresponding $\epsilon_{s,k}$ as in (20) decreases to 0. Let $\hat{u}_k \in A_{N,k}^{S,k}$, which means that $\hat{u}_k \in \mathcal{U}$, $\hat{H}_{S,k}(\hat{u}_k) \le 0$, and $J_{N,k}^{S,k} = J(\hat{u}_k)$. Let $\hat{u} \in \mathcal{U}$ be any cluster point of $\{\hat{u}_k\}_{k=1}^{\infty}$. Define $\{\hat{u}_l\}_{l=1}^{\infty}$ a subsequence converging to $\hat{u}$. Since $\hat{H}_{S,l}(u)$ defined by $\{S_l, a_l, b_l, \beta_l, N_l\}$ is continuous and converges uniformly to $Q^{1-\alpha}(\bar{h}(u,\delta))$ on $\mathcal{U}$ w.p.1, we have that $Q^{1-\alpha}(\bar{h}(\hat{u},\delta)) = \lim_{l\to\infty} \hat{H}_{S,l}(\hat{u}_l)$ w.p.1. Therefore, $Q^{1-\alpha}(\bar{h}(\hat{u},\delta)) \le 0$ and $\hat{u}$ is feasible for the true problem, and $J(\hat{u}) \ge J^*$. Moreover, $J(\hat{u}_l) \to J(\hat{u})$ w.p.1, which means that $\lim_{l\to\infty} J_{N,l}^{S,l} \ge J^*$. The above is true for any cluster point of $\{\hat{u}_k\}_{k=1}^{\infty}$ in the compact set $\mathcal{U}$, and we have

$$\lim_{k\to\infty} \inf J_{N,k}^{S,k} \ge J^*, \quad \text{w.p.1.} \tag{21}$$

Now, by Assumption 2 and Remark 2, there exists an optimal solution u^* and a sequence $\{\hat{u}_l\}_{l=1}^{\infty}$ converging to u^* with $Q^{1-\alpha}(\bar{h}(\hat{u}_l,\delta)) < 0$. Note that $\hat{H}_{S,l}(\hat{u}_l)$ converges to $Q^{1-\alpha}(\bar{h}(\hat{u}_l,\delta))$ w.p.1, and thus, there exist $K(l)$ such that $\hat{H}_{S,k}(\hat{u}_l) \le 0$ for every $k \ge K(l)$ and every l, w.p.1. Assume that $K(l) < K(l+1)$ for every l without loss of generality and define the sequence $\{\bar{u}_k\}_{k=K(1)}^{\infty}$ by setting $\bar{u}_k = u_l$ for all k and l with $K(l) \le k < K(l+1)$. We then have $\hat{H}_{S,k}(\bar{u}_k) \le 0$, which implies that $J_{N,k}^{S,k} \le J(\bar{u}_k)$ for all $k \ge K(1)$. Since f is continuous and $\bar{u}_k$ also converges to u^*, we have

$$\lim_{k\to\infty} \sup J_{N,k}^{S,k} \le J\left(u^*\right) = J^*, \quad \text{w.p.1.} \tag{22}$$

Thus, $J_{N,k}^{S,k} \to J(u^*)$ w.p.1 when $k \to \infty$. Namely, $J_N^S \to J^*$.

For the proof of $\mathbb{D}(A_N^S, A) \to 0$ w.p.1, it can refer to [18, Th. 5.3]. □

For the finite sample feasibility analysis of the approximation problem solutions, we will make use of Hoeffding's inequality [18], [19]:

Theorem 3: Denote $Z_1, \ldots, Z_N$ for independent random variables with bounded sample spaces, namely $\Pr\{Z_i \in [z_{i,\min}, z_{i,\max}]\} = 1, \forall i \in \{1, \ldots, N\}$. Then, if $s > 0$

$$\Pr\left\{\sum_{i=1}^{N}(Z_i - \mathbb{E}\{Z_i\}) \ge sN\right\} \le e^{-\frac{2N^2s^2}{\sum_{i=1}^{N}\left(z_{i,\max}-z_{i,\min}\right)^2}}. \tag{23}$$

Based on Hoeffding's inequality, probabilistic feasibility guarantee of the sample approximation method is proven in [9] and summarized here as:

Theorem 4: Let $u \in \mathcal{U}$ be such that $u \notin \mathcal{U}_f$. Then,

$$\Pr\left\{\tilde{F}^N(-\gamma, u) \ge 1 - \beta\right\} \le e^{-2N\tau_u^2} \tag{24}$$

where $\gamma > 0$, $\beta \in [0, \alpha]$, and $\tau_u > 0$ is written as

$$\tau_u = F(0,u) - F(-\gamma,u) + (\alpha - \beta). \tag{25}$$

Theorem 4 shows an important property of approximating the cumulative probability function by samples. Set $\gamma \approx 0$ and $\beta \approx \alpha$, if the sample number goes to ∞ and $\Pr\{\tilde{F}^N(-\gamma, u) \ge 1 - \beta\}$ goes to 0. Namely, the sample-based neural approximation problem has the solution, which satisfies the original probabilistic constraint in a higher probability with more sample numbers.

IV. Proposed Algorithms for PCPs

Two algorithms are used to solve PCPs. First, sample-based algorithm is summarized to train the deterministic constraints, which has a neural network form. Then, a simulated annealing algorithm is summarized to solve the deterministic program obtained by the neural approximation approach.

A. Algorithm for Training $\hat{H}_S(u)$

Due to the discussion in Section III, the algorithm for training $\hat{H}_S(u)$ is summarized as follows.

1) Generate the sample set of uncertain parameter vector $\Delta^N = \{\delta_1, \ldots, \delta_N\}$ by extracting samples independently of sample space Δ according to the identical distribution $p_\delta(\delta)$.
2) Generate the sample set of decision variable $\mathcal{U}^K = \{u_1, \ldots, u_K\}$ by extracting samples independently of feasible domain $\mathcal{U}$ according to uniform distribution.
3) For all $u_k \in \mathcal{U}^K$, calculate the $(1-\alpha)$-empirical quantile at u_k, $\tilde{Q}^{1-\alpha}(\bar{h}(u_k, \delta))$, using (9) and obtain the output sequence $Y^K = \{\tilde{Q}^{1-\alpha}(\bar{h}(u_1, \delta)), \ldots, \tilde{Q}^{1-\alpha}(\bar{h}(u_K, \delta))\}$.
4) Train β, a, b, and c in (12) based on $\mathcal{U}^K$ and Y^K by the extreme learning machine (ELM) algorithm that is introduced in [20].

B. Simulated Annealing Algorithm for Approximate Program

Simulated annealing algorithm has been widely used to address global optimizations with constraints in many fields [21]–[25] and has also been proved to be practical for nonconvex optimization [26], [27]. After using neural approximation approach, we obtain the original problem with deterministic constraints. The deterministic constraints have the neural network formulation. The original edition of simulated annealing algorithm is not presented here, which can be found in [22] and [26]. After minor revision on the original algorithm, it can be used to solve the deterministic constrained problem. The algorithm is summarized as follows.

1) Initialize a temperature T_0 and a decision value u^* and calculate the cost value as

$$E^* = J(u^*) + C(\hat{H}_S(u^*)) \tag{26}$$

where

$$\mathbb{C}(u^*) = \begin{cases} 0, & \text{if } \hat{H}_S(u^*) < 0 \\ \bar{C}, & \text{if } \hat{H}_S(u^*) \le 0 \end{cases} \tag{27}$$

where $\bar{C}$ should be chosen as a very larger value.

2) For iteration $m = 1, 2, \ldots, M$, do the following step iteratively.
   a) Randomly select another point $u_m \in \mathcal{V}_r(u^*) = \{u_m \in \mathcal{U} | \|u_m - u^*\| \le r\}$. $\mathcal{V}_r(u^*)$ a neighborhood of the previous point and r is the radius. Calculate the corresponding cost value E_m similarly as (27).
   b) Calculate

$$\nabla E_m = E_m - E^*. \tag{28}$$

   c) Move to the new point by setting $u_m = u^*$ if a random variable μ distributed uniformly over (0,1) satisfies

$$\mu \le e^{-\frac{\nabla E_m}{T_{m-1}}} \tag{29}$$

   or equivalently

$$\nabla E_m \le -T_{m-1} \log \mu. \tag{30}$$

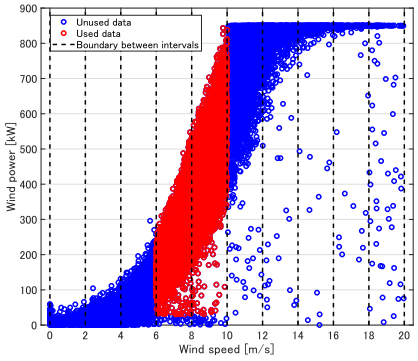

Fig. 2. Power curve of an industrial wind turbine.

   d) Update the temperature as

$$T_m = \begin{cases} T_{m-1}, & \text{if (29) holds} \\ \rho \cdot T_{m-1}, & \text{if (29) doesn't hold} \end{cases} \tag{31}$$

   where $\rho \in (T_{\min}/T_0, 1)$.
   e) Terminate the algorithm if $T_m < T_{\min}$, where $T_{\min}$ is the lower boundary for the temperature or $m = M$.

V. Application to IPM of Wind Power

A. Wind-Turbine Power Curve and Dataset

The predictor model of wind power is used to predict the wind power based on wind speed measurement [28]. Fig. 2 shows the power curve constructed from the industrial data. The data come from a large wind farm located in Jiugongshan, Hubei, China. The dataset was collected at turbines at a sampling interval of 10 min. In total, 56618 data points were collected from March 17, 2009, to April 17, 2009. The unit of the active wind power is kW, and the value of power is normalized for air density of 1.18 kg/m^3.

Piecewise models are often used to improve the prediction accuracy [29], [30]. The wind speed range is discretized into intervals, and the corresponding wind speed and power data make the partitions. Supposing that the cut-out speed is v_{co}, the speed range $[0, v_{\text{co}}]$ is divided into N_w equal length intervals. The data points in each interval are defined as

$$D_i = \{(v, p(v)) \in \mathbb{R}^2 | v \in [v_i, v_{i+1})\}, \quad i = 1, \ldots, N_w \tag{32}$$

where D_i represents the points set of the ith partition, v represents the wind speed, $p(v)$ is the corresponding wind power, and $v_i = ((i-1)/N_w)v_{\text{co}}$ is the demarcation speed between the ith and $(i-1)$th partitions. In this study, v_{co} is chosen as 20, and N_w is 10, as shown in Fig. 2. Besides, we do not establish models for all partitions in this study. As an example to demonstrate the proposed method, the data in partitions 4 and 5 are used as a unity, which are marked red circles in Fig. 2. Notice that some under-power points or stopping points are wiped out by the data preprocess method introduced in [28, Sec. III]. For convenience, denote the used dataset as D.

B. IPM of Wind Power

Consider a system with the mechanism or model from the input to output written as

$$y = f(\varphi, \delta_y) \tag{33}$$

where $y \in \mathbb{R}$ denotes the system output and $\varphi \in \mathbb{R}^{n_\varphi}$ is a regression vector containing input variables $u \in \mathbb{R}^{n_u}$ or decision variables on which the system output y depends. Besides, δ_y denotes the uncertain parameter vector and $\delta_y \in \Delta_y \subseteq \mathbb{R}^{n_{\delta_y}}$. Besides, assume that probability measure is well defined on the sigma algebra of Δ_y, $\mathcal{B}_{\Delta_y}$. Due to the existence of δ_y, for a fixed φ, y is supposed to be located in a sample space with well-defined sigma algebra and probability measure. Let $\mathcal{Y}(\varphi)$ be the sample space and $p_y(y|\varphi)$ be the probability density function conditioned on φ. Standard predictor models can only give a specific value of y for φ. However, the probabilistic interval of y conditioned on φ is important for various applications.

Differently from standard predictor models, IPMs return a prediction interval instead of a single prediction value [31]. As defined in [31], an IPM is a set-valued map

$$I : \varphi \to I(\varphi) \subseteq Y \subset \mathbb{R} \tag{34}$$

where $\varphi \in \mathbb{R}^{n_\varphi}$ is a regression vector containing input variables $u \in \mathbb{R}^{n_u}$ or decision variables on which the system output y depends. Given an observed φ, $I(\varphi)$ is an informative interval or the prediction interval, containing y with a given guaranteed probability $1-\alpha, \alpha \in [0, 1)$. Output intervals are obtained by considering the span of parametric families of functions. Here, we consider the family of linear regression functions defined by

$$\mathcal{M} = \{y = \theta^T \varphi + e, \theta \in \Theta \subseteq \mathbb{R}^{n_\theta}, \|e\| \le \varepsilon \in \mathbb{R}\}. \tag{35}$$

Then, a parametric IPM is obtained by associating to each φ the set of all possible outputs given by $\theta^T\varphi + e$ as θ varies over Θ and e varies over $\mathcal{E} = \{e \in \mathbb{R} | \|e\| \le \varepsilon\}$, which is defined as

$$I(\varphi) = \{y : y = \theta^T \varphi + e \ \forall \theta \in \Theta \ \forall e \in \mathcal{E}\}. \tag{36}$$

A possible choice for the set Θ is a ball with center c and radius r

$$\Theta = \mathcal{B}_{c,r} = \{\theta \in \mathbb{R}^{n_\theta} : \|\theta - c\| \le r\}. \tag{37}$$

Then, the interval output of the IPM can be explicitly computed as

$$I(\varphi) = [c^T\varphi - (r\|\varphi\| + \varepsilon), c^T\varphi + (r\|\varphi\| + \varepsilon)]. \tag{38}$$

For the wind power prediction problem, the input is wind speed and output is wind power as $u = v$ and $y = p$. Besides, φ is chosen as $\varphi = [u \ u^2]^T$. Then, $\varphi = [v \ v^2]^T$ in the wind power prediction problem. Identifying the IPM is essentially identifying (c, r, ε). Without loss of the generality, we use $I_{(c,r,\varepsilon)}(v)$ for $I(\varphi)$ in the latter part for IPM identification problem of wind power prediction.

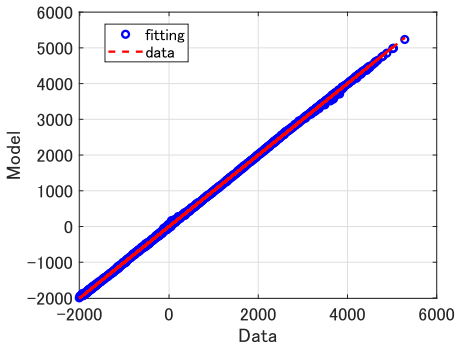

Fig. 3. Estimation results of quantile function by using neural networks.

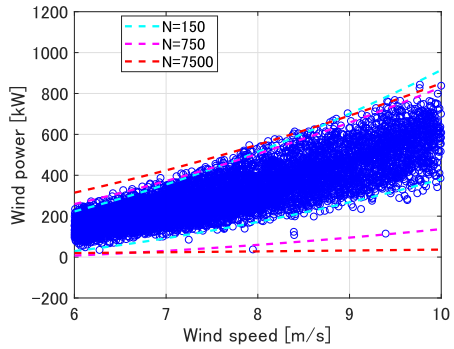

Fig. 4. Results of interval predictions by scenario approach using different choices of sample number.

Due to the above discussions, the IPM identification problem of wind power prediction is to find IPM $I_{(c,r,\varepsilon)}(v)$ (or find (c, r, ε)) such that

$$\min_{(c,r,\varepsilon)} \ wr + \varepsilon \tag{39}$$

$$\text{subject to } r, \varepsilon \ge 0$$

$$\Pr\{p \in I_{c,r,\varepsilon}(v)\} \ge 1 - \alpha, \quad \alpha \in [0, 1). \tag{40}$$

Notice that w is a weight and often chosen as $w = \mathbb{E}\{\|\varphi\|\}$ [31]. In this study, w is chosen as 20.

C. Results and Discussion

Fig. 3 shows one example of the estimation results of quantile function by using neural networks; 5000 different combinations of (c, r, ε) are considered. The dataset was divided into a train set and a test set. The blue circles show the quantile function value given by the trained fitting model. The red dotted line shows the result of the empirical quantile function value calculated by using the data of the test set. The mean value of error is -0.0421 and the mean of the abstract value of mean is 12.5882.

We compare the proposed method with scenario approach, which is proposed in [7]. The required α is set as 0.05. Figs. 4 and 5 show the results of interval predictions by scenario approach and the proposed method, respectively. The number of used data samples is 150, 750, or 7500. As the number of samples increases, the scenario approach gives more conservative results, while the proposed method performs with better robust on sample numbers.

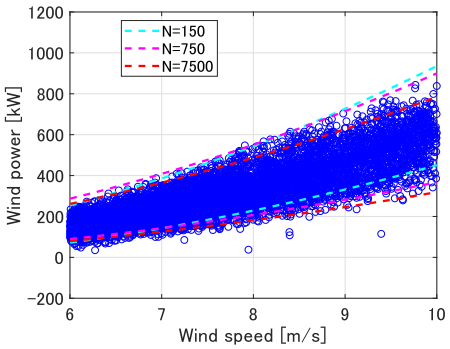

Fig. 5. Results of interval predictions by neural approximation approach using different choices of sample number. The limitation of violation probability is $\alpha = 0.05$.

TABLE I

PROBABILITY THAT CONSTRAINT FAILURE PROBABILITY $\alpha > 0.05$: $\Pr\{\alpha > 0.05\}$

N	150	375	750	1500	3750	7500
SA	0.224	0.004	0	0	0	0
Proposed	0.156	0.07	0.006	0	0	0

TABLE II

MEAN VALUE OF COST FUNCTION

N	150	375	750	1500	3750	7500
SA	61.56	70.47	74.72	77.58	79.95	80.93
Proposed	50.25	50.45	50.69	50.97	51.33	51.72

A more comprehensive validation was conducted through the posterior Monte Carlo analysis. In the validation, the number of data sample was increased from 150 to 7500. For each number of data samples, 500 times sampling and calculation of c, r, ε processes were done. For instance, if we set the number of data samples as 150. Then, 500 times of extracting 150 samples of (v, p) from the training set were done. For each extracted 150 samples, the corresponding c, r, ε was calculated by the scenario approach or proposed method. The statistical analysis results are summarized in Tables I and II. Apparently, the proposed method can achieve a better tradeoff between the probability constraint and the cost value.

VI. CONCLUSION

In this article, a neural approximation-based method for solving PCP has been proposed. The statistical guarantees of the proposed method are discussed. Furthermore, IPM of wind power is investigated with experimental data to validate the proposed method. In the validation, the proposed method is compared with the scenario approach. The results show that the proposed method performs better in the tradeoff between satisfying the probabilistic constraints and minimizing the cost.

REFERENCES

[1] A. Charnes and W. W. Cooper, "Chance constrained programming," *Manage. Sci.*, vol. 6, no. 1, pp. 73–79, 1959.

[2] M. C. Campi and S. Garatti, *Introduction to the Scenario Approach*. Philadelphia, PA, USA: MOS-SIAM Series on Optimization, 2019.

[3] X. Shen, Y. Zhang, T. Shen, and C. Khajorntraidet, "Spark advance self-optimization with knock probability threshold for lean-burn operation mode of Si engine," *Energy*, vol. 122, pp. 1–10, Mar. 2017.

[4] X. Shen, T. Ouyang, Y. Zhang, and X. Zhang, "Computing probabilistic bounds on state trajectories for uncertain systems," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Sep. 3, 2020, doi: 10.1109/TETCI.2020.3019040.

[5] M. Guo and M. M. Zavlanos, "Probabilistic motion planning under temporal tasks and soft constraints," *IEEE Trans. Autom. Control*, vol. 63, no. 12, pp. 4051–4066, Dec. 2018.

[6] P. Gautam, R. Karki, and P. Piya, "Probabilistic modeling of energy storage to quantify market constrained reliability value to active distribution systems," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 1043–1053, Apr. 2020.

[7] G. Calafiore and M. Campi, "The scenario approach to robust control design," *IEEE Trans. Autom. Control*, vol. 51, no. 5, pp. 742–753, May 2006.

[8] M. C. Campi and S. Garatti, "A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality," *J. Optim. Theory Appl.*, vol. 148, no. 2, pp. 257–280, 2011.

[9] J. Luedtke and S. Ahmed, "A sample approximation approach for optimization with probabilistic constraints," *SIAM J. Optim.*, vol. 19, no. 2, pp. 674–699, Jul. 2008.

[10] A. Peña-Ordieres, J. R. Luedtke, and A. Wächter, "Solving chance-constrained problems via a smooth sample-based nonlinear approximation," *SIAM J. Optim.*, vol. 30, no. 3, pp. 2221–2250, Jan. 2020.

[11] A. Geletu, A. Hoffmann, M. Kloppel, and P. Li, "An inner-outer approximation approach to chance constrained optimization," *SIAM J. Optim.*, vol. 27, no. 3, pp. 1834–1857, 2017.

[12] X. Shen, J. Zhuang, and X. Zhang, "Approximate uncertain program," *IEEE Access*, vol. 7, pp. 182357–182365, 2019.

[13] G. Steinbrecher and W. T. Shaw, "Quantile mechanics," *Eur. J. Appl. Math.*, vol. 19, no. 2, pp. 87–112, 2008.

[14] A. W. van der Vaart, *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[15] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control, Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.

[16] M. Hassoun, *Fundamentals of Artificial Neural Networks*. Cambridge, MA, USA: MIT Press, 1995.

[17] R. R. Selmic and F. L. Lewis, "Neural-network approximation of piecewise continuous functions: Application to friction compensation," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 745–751, May 2002.

[18] A. Shapiro, D. Dentcheva, and A. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory*. 2nd ed. Philadelphia, PA, USA: SIAM, 2014.

[19] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.

[20] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.

[21] S. Kirkpatrick, C. D. Gelatt, and M. Vecchi, "Optimization BT simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[22] S. P. Brooks and B. J. T. Morgan, "Optimization using simulated annealing," *J. Roy. Stat. Soc., Ser. D, Statistician*, vol. 44, no. 2, pp. 241–257, 1995.

[23] A. S. Azad, M. Islam, and S. Chakraborty, "A heuristic initialized stochastic memetic algorithm for MDPVRP with interdependent depot operations," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4302–4315, Dec. 2017.

[24] A. Albanna and H. Yousefi'Zadeh, "Congestion minimization of LTE networks: A deep learning approach," *IEEE/ACM Trans. Netw.*, vol. 28, no. 1, pp. 347–359, Feb. 2020.

[25] A. Dekkers and E. Aarts, "Global optimization and simulated annealing," *Math. Program.*, vol. 50, pp. 367–393, Mar. 1991.

[26] H. Szu and R. Hartley, "Fast simulated annealing," *Phys. Lett. A*, vol. 122, no. 3, pp. 157–162, 1987.

[27] L. Ingber, "Simulated annealing: Practice versus theory," *Math. Comput. Model.*, vol. 18, no. 11, pp. 29–57, 1993.

[28] T. Ouyang, A. Kusiak, and Y. He, "Modeling wind-turbine power curve: A data partitioning and mining approach," *Renew. Energy*, vol. 102, pp. 1–8, Mar. 2017.

[29] M. Lydia, A. I. Selvakumar, S. S. Kumar, and G. E. P. Kumar, "Advanced algorithms for wind turbine power curve modeling," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 827–835, Jul. 2013.

[30] M. G. Khalfallah and A. M. Koliub, "Suggestions for improving wind turbines power curves," *Desalination*, vol. 209, nos. 1–3, pp. 221–229, Apr. 2007.
[31] M. C. Campi, G. Calafiore, and S. Garatti, "Interval predictor models: Identification and reliability," *Automatica*, vol. 45, no. 2, pp. 382–392, Feb. 2009.

Xun Shen (Member, IEEE) was born in Yueyang, China. He received the B.S. degree in electrical engineering from Wuhan University, Wuhan, China, in 2012, and the Ph.D. degree in mechanical engineering from Sophia University, Tokyo, Japan, in 2018. He is currently pursuing the Ph.D. degree with the Department of Statistical Sciences, Graduate University for Advanced Studies, Tokyo.

He has been a Post-Doctoral Research Associate with the Department of Engineering and Applied Sciences, Sophia University, since April 2018. Since June 2019, he has been an Assistant Professor with the Department of Mechanical Systems Engineering, Tokyo University of Agriculture and Technology, Fuchu, Japan. His current research interests include chance constrained optimization, point process, and their industrial applications.

Tinghui Ouyang received the B.S. and Ph.D. degrees in electrical engineering and automation from Wuhan University, Wuhan, Hubei, China, in 2012 and 2017, respectively.

He is currently a Research Scientist at the Artificial Intelligence Research Center (AIRC), National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan. Before joining AIST, he was a Research Fellow with Nanyang Technological University (NTU), Singapore, from 2018 to 2019, and a Post-Doctoral Fellow at the University of Alberta, Edmonton, AB, Canada, from 2017 to 2018. His major research interests include computational intelligence, data mining, machine learning, and deep learning.

Nan Yang was born in Xiangyang, China. He received the B.S. degree in electrical engineering from the College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, China, in 2009, and the Ph.D. degree in electrical engineering from Wuhan University, Wuhan, China, in 2014.

He is currently an Associate Professor with the Department of Hubei Provincial Collaborative Innovation Center for New Energy Microgrid, China Three Gorges University, Yichang, China. His major research interests include optimal operation of power system, planning of power system, uncertainty modeling, and artificial intelligence technology.

Jiancang Zhuang received the Ph.D. degree in statistics from the Department of Statistical Science, Graduate University for Advanced Studies, Tokyo, Japan, in 2003.

From 2003 to 2004, he was a Lecturer with The Institute of Statistical Mathematics, Tokyo. From 2004 to 2006, he was a Post-Doctoral Fellow at The Institute of Statistical Mathematics, funded by the Japanese Science Promotion Society. From 2006 to 2007, he was a Post-Doctoral Fellow at the Department of Earth and Space Sciences, University of California at Los Angeles, Los Angeles, CA, USA. In 2007, he returned to The Institute of Statistical Mathematics, where he had been an Assistant Professor until 2011. Since 2011, he has been an Associate Professor with The Institute of Statistical Mathematics and the Department of Statistical Sciences, Graduate University for Advanced Studies. His research interests include statistics, probability, and modeling.