

Deep Sigma Point Processes-Assisted Chance-Constrained Power System Transient Stability Preventive Control

Tong Su, *Student Member, IEEE*, Junbo Zhao, *Senior Member, IEEE*, Xiao Chen

Abstract—This paper proposes a deep sigma point processes (DSPP)-assisted chance-constrained power system transient stability preventive control method to deal with uncertain renewable energy and loads-induced stability risk. The traditional transient stability-constrained preventive control is reformulated as a chance-constrained optimization problem. To deal with the computational bottleneck of the time-domain simulation-based probabilistic transient stability assessment, the DSPP is developed. DSPP is a parametric Bayesian approach that allows us to predict system transient stability with high computational efficiency while accurately quantifying the confidence intervals of the predictions that can be used to inform system instability risk. To this end, with a given preset confidence probability, we embed DSPP into the primal dual interior point method to help solve the chance-constrained preventive control problem, where the corresponding Jacobian and Hessian matrices are derived. Comparison results with other existing methods show that the proposed method can significantly speed up preventive control while maintaining high accuracy and convergence.

Index Terms—Deep sigma point processes, power system stability, probabilistic transient stability prediction, chance-constrained optimization, renewable energy.

I. INTRODUCTION

THE increased penetration of intermittent renewable energy has caused power system operation challenges. In particular, the forecasting uncertainties for renewable energy may yield an underestimate of power system transient stability. Therefore, probabilistic transient stability assessment has started to call attractions [1].

The traditional transient stability prediction is based on time-domain simulations by solving the nonlinear differential-algebraic equations, which is time-consuming and subject to scalability issues for large-scale systems [2]. To address that, many efficient alternatives have been proposed, including the transient energy function method [3], extended equal-area criterion [4], quasi-steady-state analysis [5], etc. These alternatives are model-based and thus require an accurate dynamic system model. With the advancement of artificial intelligence, learning-based methods have been introduced,

including the decision tree technique [6], support vector machine [7], artificial neural network [8], deep belief network [9], stacked denoising autoencoder (SDAE) [10], etc. Compared with the model-based methods, data-driven methods are much more computationally efficient [11]. However, these model-based and data-based methods do not consider the impacts of renewable energy and load uncertainties, which usually lead to either underestimated or overestimated transient stability prediction outcomes. To this end, probabilistic transient stability assessment approaches are proposed, which can be divided into two categories [12]. Some studies propose to approximate the true probability of the stability assessment outcome based on Monte Carlo simulations (MCS) and statistical analysis. In [10], a large number of possible operation scenarios are generated by the MCS to calculate the probability of transient stability and instability. In [13], MCS are used to identify the critical generators by counting the number of times each generator becomes unstable. Then, the statistical analysis is implemented to determine the thresholds for protection schemes. In [14], the historical statistics on the probabilistic states of load level factor, fault type, fault location, fault clearing, and automatic reclosing are used in MCS to generate samples and perform transient stability assessment. Other studies are based on analytical methods. In [15], the Kalman filter is used to estimate the system angle, and unscented transformation is applied to predict the distribution of the system transient stability margin. In [12], the analytical expression of the probabilistic transient stability index is derived, and three-point estimation and Cornish-Fisher expansion are used to deal with the wind farm uncertainties. However, MCS are computationally expensive to be used for operational planning, while the derivation of analytical expressions is challenging, especially when the accurate dynamical system model is difficult to obtain in practice [12].

On the other hand, the probabilistic transient stability assessment only provides the system risk but not the mitigation actions. Preventive control aims to prepare the system before the occurrence of credible contingencies by generator dispatch, wind curtailment, and load shedding [16]. Transient stability constrained optimal power flow (TSC-OPF) is the widely used method to find the optimal operating point under transient stability constraint. Because the time-domain simulation is very time-consuming, it is difficult to be applied in TSC-OPF for online applications. Several alternatives of time-domain simulations are combined with TSC-OPF to accelerate the calculation. In [17], deep learning-based transient stability

This work is supported by the National Science Foundation under ECCS 1917308 and the U.S. Department of Energy Advanced Grid Modernization program (*Corresponding Author*: Junbo Zhao).

Tong Su and Junbo Zhao are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT 06269, USA (e-mail: tongsu@uconn.edu; junbo@uconn.edu).

X. Chen is with the Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA (e-mail: chen73@llnl.gov).

constraint is used in TSC-OPF to speed up transient stability verification of preventive control strategies. However, [17] uses a deterministic prediction method, which is difficult to quantify the risk probability. In [18], probabilistic models of load injections, wind generations, and fault clearing time are constructed. Then, these probabilistic models are embedded into the TSC-OPF framework to construct probabilistic static inequality and transient stability constraints. The heuristic optimization approach, i.e., group search optimization is used that is not guaranteed to achieve optimal solution nor computationally efficient. In [19], a machine learning model is trained to predict transient stability. After that, the statistical stability probability of all scenarios is used in TSC-OPF as a transient stability constraint. The core of [19] is still based on MCS, which heavily depends on the accuracy of possible fault scenario generations.

In this paper, a computationally efficient deep sigma point processes (DSPP)-assisted chance-constrained power system transient stability preventive control framework is proposed. The main contributions are as follows:

- 1) A DSPP-assisted probabilistic transient stability prediction method is proposed that can quantify the predictive distribution of the transient stability index. DSPP takes the advantages of both Gaussian processes (GP) in uncertainty quantification and deep GP with scalability to yield accurate posterior approximations of stability index distribution and stability risk quantification.
- 2) The existing TSC-OPF formulation has been extended to the chance-constrained problem with the DSPP-assisted probabilistic transient stability prediction model informed risk. This allows us to better balance risk mitigation and system economics. It is worth noting that embedding the proposed probabilistic transient stability prediction model addresses the computational challenge of the traditional time-consuming simulations-based transient stability verification. Comparison results with existing approaches demonstrate the great improvements in the convergence and computing time without loss of generation dispatch accuracy, i.e., still maintaining good economics.

The rest of the paper is organized as follows. Section II presents the traditional TSC-OPF problem and shows the problem of interest. Section III shows the proposed DSPP-assisted chance-constrained transient stability preventive control. Results are presented and analyzed in Section IV and finally, Section V concludes the paper.

II. PROBLEM STATEMENT

For power system preventive control, TSC-OPF is widely used to find the optimal operating point under transient stability constraints, which is mathematically formulated as follows:

A. Objective Function

The objective function is to minimize the total cost, including generation cost, wind curtailment cost, and load shedding

cost.

$$\begin{aligned} \text{Minimize } & \sum_{i \in S_g} (a_{gi} P_{gi,t}^2 + b_{gi} P_{gi,t} + c_{gi}) + \\ & \sum_{j \in S_w} c_{wj} \Delta P_{wj,t} + \sum_{k \in S_d} c_{dk} \Delta P_{dk,t} \quad (1) \\ & \Delta P_{w,t} = P_{w,t}^{MPPT} - P_{w,t} \\ & \Delta P_{d,t} = P_{d,t}^0 - P_{d,t} \end{aligned}$$

where t denotes time instant; S_g is the synchronous generator set; a_{gi} , b_{gi} , and c_{gi} are fuel cost constants of i -th synchronous generator; $P_{gi,t}$ denotes the active power output of i -th synchronous generator; S_w is the wind farm set; c_{wj} is the wind curtailment cost constant of j -th wind farm; $\Delta P_{wj,t}$ is the curtailed active power of j -th wind farm; S_d is the load set; c_{dk} is the load shedding cost constant of k -th load; $\Delta P_{dk,t}$ denotes the active power shedding of k -th load; $P_{w,t}^{MPPT}$ is the maximum active power output of $P_{w,t}$ according to maximum power point tracking (MPPT); $P_{d,t}^0$ denotes the load demand; $P_{w,t}$ and $P_{d,t}$ denote active power of wind farms and loads according to preventive control strategy. It should be noted that both $\Delta P_{w,t}$ and $\Delta P_{d,t}$ are greater than 0, that is, they can only be reduced.

B. Power Balance Constraints

Suppose n_b and n_s are the numbers of buses and all generators (synchronous generators and wind farms), respectively. The power balance constraints are [20]

$$G_P(\Theta, V_m, P_g) = P_{\text{bus}}(\Theta, V_m) + P_d - C_s P_s = 0 \quad (2)$$

$$G_Q(\Theta, V_m, Q_g) = Q_{\text{bus}}(\Theta, V_m) + Q_d - C_s Q_s = 0 \quad (3)$$

where G_P and G_Q denote the active and reactive power balance equations; P_{bus} and Q_{bus} denote the $n_b \times 1$ vectors of bus active and reactive power injections; Θ and V_m denote the $n_b \times 1$ vectors of bus voltage angles and bus voltage magnitudes; P_s and Q_s denote the $n_s \times 1$ vectors of generators and wind farms active and reactive injections; P_d and Q_d denote the $n_b \times 1$ vectors of loads active and reactive demands; C_s denotes the $n_b \times n_s$ matrix of generators connection. If generator or wind farm s is in bus j , $C_{s,ij} = 1$.

C. Branch Flow Constraints

Suppose n_l denote the number of the branches. The branch flow inequality constraints are

$$|S_f(\Theta, V_m)| - S^{\max} \leq 0 \quad (4)$$

$$|S_t(\Theta, V_m)| - S^{\max} \leq 0 \quad (5)$$

where S_f and S_t denote the $n_l \times 1$ vectors of the apparent power flows at the start and end of the branches; $S^{\max}$ is the $n_l \times 1$ vector of max apparent power flow limits of the branches.

D. Variable Limits

Variable limits include reference bus voltage angle equality constraint, bus voltage magnitude inequality constraints, generator active and reactive power output constraints, i.e.,

$$\theta_i^{\text{ref}} \leq \theta_i \leq \theta_i^{\text{ref}}, i \in \zeta^{\text{ref}} \quad (6)$$

$$v_{mi}^{\min} \leq v_{mi} \leq v_{mi}^{\max}, i = 1, \dots, n_b \quad (7)$$

$$p_{si}^{\min} \leq p_{si} \leq p_{si}^{\max}, i = 1, \dots, n_s \quad (8)$$

$$q_{si}^{\min} \leq q_{si} \leq q_{si}^{\max}, i = 1, \dots, n_s \quad (9)$$

where ζ^{ref} denotes the reference bus; θ_i and θ_i^{ref} denote reference bus voltage angle and its rated value; v_{mi} denotes the i -th bus voltage magnitude; $v_{mi}^{\max}$ and $v_{mi}^{\min}$ are upper and lower limits of v_{mi} ; p_{si} denotes the i -th synchronous generator or wind farm active power output; $p_{si}^{\max}$ and $p_{si}^{\min}$ are upper and lower limits of p_{si} ; q_{si} denotes the i -th synchronous generator or wind farm reactive power output; $q_{si}^{\max}$ and $q_{si}^{\min}$ are upper and lower limits of q_{si} .

E. Transient Stability Constraint

Power system transient stability is constrained by differential and algebraic equations (DAEs) [1].

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{u}, \mathbf{p}, \boldsymbol{\tau}) \quad (10)$$

$$\mathbf{0} = \mathbf{g}(\mathbf{x}(t), \mathbf{y}(t), \mathbf{u}, \mathbf{p}, \boldsymbol{\tau}) \quad (11)$$

where t denotes the time instant during the transient period; $\mathbf{x}(t)$ and $\mathbf{y}(t)$ denote vectors of state and algebraic variables, respectively; $\mathbf{u}$ is the input vector; $\mathbf{p}$ denotes the model parameters; $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are nonlinear differential equations and algebraic equations, respectively. They are solved by time domain simulation, which is time-consuming and subject to scalability issues for large-scale systems; $\boldsymbol{\tau}$ denotes the uncertainties that can change the system operating point, such as the rapid wind power fluctuations by speed change [19].

Due to the presence of DAEs within the TSC-OPF problem, there is a huge computational challenge for most existing approaches, especially when the uncertainties from renewable energy-induced transient stability risk are considered. This paper focuses on preventive control, where renewable energy uncertainty-induced transient stability should be accurately taken into account while yielding optimal control actions.

III. DSPP-ASSISTED CHANCE-CONSTRAINED TRANSIENT STABILITY PREVENTIVE CONTROL

To address the TSC-OPF computational efficiency issue as well as mitigate the renewable energy uncertain induced stability risk, we extend the TSC-OPF to the chance-constrained formulation and develop the DSPP-assisted probabilistic transient stability preventive control.

A. Chance-Constrained TSC-OPF

Traditional TSC-OPF is based on the deterministic transient stability constraint, which cannot quantify the stability risk of different uncertain power generation and demand scenarios. Probabilistic optimization or chance-constrained optimization enables a constraint to be satisfied with a preset probability [21]. In this paper, the transient stability constraint is reformulated as chance constraint, and the chance-constrained TSC-OPF is expressed as follows [18], [19]

$$\text{Minimize (1)} \quad (12)$$

$$\text{s.t. (2) } \sim (9) \quad (13)$$

$$\mathbb{P}(\chi(\mathbf{x}(t), \mathbf{y}(t), \mathbf{u}, \mathbf{p}, \boldsymbol{\tau}) > \alpha) \geq \varepsilon \quad (14)$$

where χ denotes the stability index of interest, which can be selected according to different studies while this paper focuses on transient stability; α denotes the threshold of the pre-defined stability index. $\mathbb{P}(\cdot)$ denotes the probability; ε is the confidence probability of the chance constraint. The condition that the transient stability constraint is satisfied means that the probability of $\chi > \alpha$ is not less than ε .

In this paper, the transient stability index (TSI) is used to describe the system stability, which reflects the maximum rotor angle difference of all generators during the transient period [9]. Formally, we have

$$\text{TSI} = \frac{360^\circ - \Delta^{\max}}{360^\circ + \Delta^{\max}} \times 100 \quad (15)$$

$$\Delta^{\max} = \max_{t \in T} (|\Delta_{i,t} - \Delta_{j,t}|), \forall i, j \in S_g \quad (16)$$

where T is the whole time-domain simulation period; $\Delta_{i,t}$ and $\Delta_{j,t}$ denote the rotor angle of generator i and j at time t ; $\Delta^{\max}$ denotes the maximum rotor angle difference of any two generators. The increase of TSI means the improvement of system transient stability. Generally speaking, $\Delta^{\max} = 360^\circ$, i.e. TSI = 0, is considered as the boundary between transient stable and unstable [22]. However, in some conservative transient stability prediction studies, this boundary is considered as $\Delta^{\max} = 180^\circ$, i.e. TSI = 33.3 [23]. Based on DSPP, the transient stability chance constraint can be expressed as [24]

$$\mathbb{P}(\text{TSI} \geq \text{TSI}_B) \geq \varepsilon \quad (17)$$

where TSI_B denotes the boundary TSI for distinguishing transient stable and unstable; (17) denotes the probability of $\text{TSI} \geq \text{TSI}_B$ must be no less than the pre-defined threshold ε , which can be selected according to the engineering experiences of the operator. It should be noted that if MCS-based methods are used to approximate probability in (17), the computational burden will be further increased. To this end, we propose the DSPP-based probability transient stability prediction method to speed up the calculation.

B. DSPP Approach for Chance-Constraints Modeling

Gaussian Processes (GP) is a machine learning model based on Bayesian theory for regression and classification. GP can qualify the uncertainty of the predicted target and obtain its predictive distribution, which is meaningful in security operation and dispatch [25]. A GP is represented by mean function $\mu(\mathbf{X})$ and covariance function $k(\mathbf{X}, \mathbf{X}')$ [27], i.e.,

$$f(\mathbf{X}) \sim \mathcal{GP}(\mu(\mathbf{X}), k(\mathbf{X}, \mathbf{X}')) \quad (18)$$

where $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ denotes the $D \times N$ matrix of the training inputs (D is the dimension of $\mathbf{x}_i$, N is the dimension of input feature space); $f(\mathbf{X}) = \mathbf{f} = \{\mathbf{f}_i\}_{i=1}^N$ is the latent function values matrix. The joint density of GP is

$$p(\mathbf{y}, \mathbf{f} | \mathbf{X}) = p(\mathbf{y} | \mathbf{f}, \sigma_{\text{obs}}^2) p(\mathbf{f} | \mathbf{X}) \quad (19)$$

where $p(\mathbf{y} | \mathbf{X})$ denotes the multivariate Normal distribution; $\mathbf{y} = \{\mathbf{y}_i\}_{i=1}^N$ is the target matrix; σ_{obs}^2 denotes the variance; $p(\mathbf{y} | \cdot)$ is the likelihood function. The marginal likelihood is

$$p(\mathbf{y} | \mathbf{X}) = \int d\mathbf{f} p(\mathbf{y} | \mathbf{f}, \sigma_{\text{obs}}^2) p(\mathbf{f} | \mathbf{X}) \quad (20)$$

The above form is time-consuming in the presence of large samples, i.e., large N . The time complexity of GP is $O(N^3)$. For this reason, the inducing point method-based sparse GP

(SGP) is developed [28]. In particular, the inducing points $\mathbf{Z} = \{z_i\}_{i=1}^M$ is introduced as inputs, where $M \ll N$, and the corresponding outputs are $\mathbf{u} = \{u_i\}_{i=1}^M$. The GP prior varies with $\mathbf{u}$ as

$$p(\mathbf{f} | \mathbf{X}) \rightarrow p(\mathbf{f} | \mathbf{u}, \mathbf{X}, \mathbf{Z})p(\mathbf{u} | \mathbf{Z}) \quad (21)$$

Using the Jensen's inequality to reduce the log joint density of $\mathbf{y}$ and $\mathbf{u}$, yielding

$$\begin{aligned} \log p(\mathbf{y}, \mathbf{u} | \mathbf{X}, \mathbf{Z}) &= \log \int d\mathbf{f} p(\mathbf{y} | \mathbf{f}) p(\mathbf{f} | \mathbf{u}) p(\mathbf{u}) \\ &\geq \mathbb{E}_{p(\mathbf{f} | \mathbf{u})} [\log p(\mathbf{y} | \mathbf{f}) + \log p(\mathbf{u})] \\ &= \sum_{i=1}^N \log \mathcal{N}(y_i | \mathbf{k}_i^T \mathbf{K}_{MM}^{-1} \mathbf{u}, \sigma_{\text{obs}}^2) \\ &\quad - 1/2\sigma_{\text{obs}}^2 \text{Tr} \tilde{\mathbf{K}}_{NN} + \log p(\mathbf{u}) \end{aligned} \quad (22)$$

where $\mathbb{E}$ denotes the mean value function; Tr denotes the trace of the matrix $\tilde{\mathbf{K}}_{NN}$; $\tilde{\mathbf{K}}_{NN} = \mathbf{K}_{NN} - \mathbf{K}_{NM} \mathbf{K}_{MM}^{-1} \mathbf{K}_{MN}$; $\mathbf{K}_{MM} = k(\mathbf{Z}, \mathbf{Z})$; $\mathbf{k}_i = k(x_i, \mathbf{Z})$; $\mathbf{K}_{NM} = \mathbf{K}_{MN}^T = k(\mathbf{X}, \mathbf{Z})$. SGP can then be obtained by applying variational inference to the lower bound of (22). The evidence lower bound (ELBO) of SGP is

$$\begin{aligned} \mathcal{L}_{\text{SGP}} &= \mathbb{E}_{q(\mathbf{u})} [\log p(\mathbf{y}, \mathbf{u} | \mathbf{X}, \mathbf{Z})] + H[q(\mathbf{u})] \\ &= \sum_{i=1}^N \left\{ \log \mathcal{N}(y_i | \mu_f(x_i), \sigma_{\text{obs}}^2) - \frac{1}{2} \frac{\sigma_f(x_i)^2}{\sigma_{\text{obs}}^2} \right\} \\ &\quad - \text{KL}(q(\mathbf{u}) | p(\mathbf{u})) \end{aligned} \quad (23)$$

where KL is the Kullback-Leibler divergence; $q(\mathbf{u}) = \mathcal{N}(\mathbf{m}, \mathbf{S})$ denotes Normal variational distribution; $H[q(\mathbf{u})]$ denotes the entropy term; $\mu_f(x_i) = \mathbf{k}_i^T \mathbf{K}_{MM}^{-1} \mathbf{m}$ denotes the predictive mean function; $\sigma_f(x_i)^2 = \tilde{\mathbf{K}}_{ii} + \mathbf{k}_i^T \mathbf{K}_{MM}^{-1} \mathbf{S} \mathbf{K}_{MM}^{-1} \mathbf{k}_i$ is the latent function variance. The complete variational distribution of SGP is

$$q(\mathbf{f}, \mathbf{u}) = p(\mathbf{f} | \mathbf{u}, \mathbf{X}) q(\mathbf{u}) \quad (24)$$

The marginal distribution is

$$q(\mathbf{f}) = \int d\mathbf{u} q(\mathbf{f}, \mathbf{u}) = \mathcal{N}(\mathbf{f} | \mu_f(\mathbf{X}), \Sigma_f(\mathbf{X})) \quad (25)$$

where $\Sigma_f(\mathbf{X}) = \tilde{\mathbf{K}}_{NN} + \mathbf{K}_{NM}^T \mathbf{K}_{MM}^{-1} \mathbf{S} \mathbf{K}_{MM}^{-1} \mathbf{K}_{MN}$ is the $N \times N$ covariance matrix. The final optimization objective of SGP is to maximize $\mathcal{L}_{\text{SGP}}$.

Deep Gaussian Processes (DGP) is the multi-layer generalization of GP and forms a hierarchical model [29]. For a 2-layer DGP, the joint likelihood for $(\mathbf{y}, \mathbf{X})$ is

$$p(\mathbf{y}, \mathbf{f}, \mathbf{G} | \mathbf{X}) = p(\mathbf{y} | \mathbf{f}, \sigma_{\text{obs}}^2) p(\mathbf{f} | \mathbf{G}) p(\mathbf{G} | \mathbf{X}) \quad (26)$$

where $\mathbf{G} = \{\mathbf{g}_i\}_{i=1}^N$ denotes the $N \times W$ matrix of output of the first DPs (W is the dimension of $\mathbf{g}_i$); $p(\mathbf{G} | \mathbf{X}) = \prod_{w=1}^W p(\mathbf{g}_w | \mathbf{X})$ denotes prior of $\mathbf{G}$.

The analytical form of DGP is difficult to obtain and therefore, the doubly stochastic variational inference (DSVI) is developed to achieve approximation. Inducing points and variational distribution $Q(\mathbf{f}, \mathbf{u}_f, \dots, \mathbf{g}_W, \mathbf{u}_{gW})$ are used. The ELBO is

$$\mathcal{L}_{\text{DSVI}} = \mathbb{E}_Q [\log p(\mathbf{y} | \mathbf{f}, \sigma_{\text{obs}}^2)] - \sum \text{KL} \quad (27)$$

where $\sum \text{KL}$ is the sum of all KL divergences for the inducing variables $\{\mathbf{u}_f, \dots, \mathbf{g}_W, \mathbf{u}_{gW}\}$. $\mathbf{f}$ at top layer can be analytically integrated, while $\mathbf{G}$ at other layers must be sampled using

the reparameterization trick. The predictive distribution is

$$\mathbb{E}_{\prod_{n=1}^W q(\mathbf{g}_{z_N} | \mathbf{x}_i)} \left[\mathcal{N}(y_* | \mu_f(\mathbf{g}_*), \sigma_f(\mathbf{g}_*)^2 + \sigma_{\text{obs}}^2) \right] \quad (28)$$

where subscript $*$ is the test set; $\mu_f(\mathbf{g}_*)$ is the predictive mean; $\sigma_f(\mathbf{g}_*)^2$ is the predictive variance. The calculation of (28) is intractable. In DSPP, a parametric finite mixture method is developed to simplify this, where a sigma point approximation method is applied to (28) [26]. Suppose $W = 2$ for the first layer of DSPP, the marginal likelihood is

$$\begin{aligned} p_{\text{DSPP}}(y_i | \mathbf{x}_i) &= \int d\mathbf{g}_i \mathcal{N}(y_i | \mu_f(\mathbf{g}_i), \sigma_f(\mathbf{g}_i)^2 + \sigma_{\text{obs}}^2) \prod_{w=1}^2 q(\mathbf{g}_{iw} | \mathbf{x}_i) \end{aligned} \quad (29)$$

Through Gauss-Hermite quadrature, $q(\mathbf{g}_{i1} | \mathbf{x}_i)$ in (29) can be approximated by S mixtures of Dirac delta distributions

$$q(\mathbf{g}_{i1} | \mathbf{x}_i) = \sum_{s_1=1}^S \omega_1^{(s_1)} \delta\left(\mathbf{g}_{i1} - \left(\mu_{g_1}(\mathbf{x}_i) + \xi_1^{(s_1)} \sigma_{g_1}(\mathbf{x}_i)\right)\right) \quad (30)$$

where $q(\mathbf{g}_{i2} | \mathbf{x}_i)$ has a similar equation to $q(\mathbf{g}_{i1} | \mathbf{x}_i)$; $\omega_1^{(s_1)}$ denotes the weights; δ is the Dirac delta function; $\xi_1^{(s_1)}$ represents the quadrature points. Replace (30) into (29), a mixture with S^2 components is obtained

$$\begin{aligned} p_{\text{DSPP}}(y_i | \mathbf{x}_i) &= \sum_{s_1=1}^S \sum_{s_2=1}^S \omega_1^{(s_1)} \omega_2^{(s_2)} \times \mathcal{N}(y_i | \\ &\mu_f\left(\mu_{g_1}(\mathbf{x}_i) + \xi_1^{(s_1)} \sigma_{g_1}(\mathbf{x}_i), \mu_{g_2}(\mathbf{x}_i) + \xi_1^{(s_2)} \sigma_{g_2}(\mathbf{x}_i)\right), \\ &\sigma_f\left(\mu_{g_1}(\mathbf{x}_i) + \xi_1^{(s_1)} \sigma_{g_1}(\mathbf{x}_i), \mu_{g_2}(\mathbf{x}_i) + \xi_1^{(s_2)} \sigma_{g_2}(\mathbf{x}_i)\right)) \end{aligned} \quad (31)$$

In (31), the predictive distribution of a 2-layer DSPP has S^W Normal distribution mixtures, which grow exponentially. To solve this problem, a more flexible alternative is to 'line-up' the quadrature points across a different $\mathbf{g}_i$

$$\begin{aligned} \prod_{w=1}^W q(\mathbf{g}_{iw} | \mathbf{x}_i) &\rightarrow \sum_{s=1}^S \omega^{(s)} \prod_{w=1}^W \delta\left(\mathbf{g}_{iw} - \left(\mu_{g_w}(\mathbf{x}_i) + \xi_w^{(s)} \sigma_{g_w}(\mathbf{x}_i)\right)\right) \end{aligned} \quad (32)$$

The objective function of DSPP is

$$\mathcal{L}_{\text{DSPP}} = \sum_{i=1}^N \log p_{\text{DSPP}}(y_i | \mathbf{x}_i) - \beta_{\text{reg}} \sum \text{KL} \quad (33)$$

where $\beta_{\text{reg}} > 0$ denotes optional regularization constant. The optimization parameter of $\mathcal{L}_{\text{DSPP}}$ is σ_{obs} , $\mathbf{m}$, $\mathbf{S}$, $\mathbf{Z}$ and kernel hyperparameters for each layer, which can be optimized by stochastic gradient methods and subsampling.

In this paper, DSPP is used for probabilistic transient prediction. The input variables are synchronous generator active power outputs $\mathbf{P}_g$, wind farm active power outputs $\mathbf{P}_w$, synchronous generator bus voltage magnitudes $\mathbf{V}_g$, wind farms bus voltage magnitude $\mathbf{V}_w$, active power demands of loads $\mathbf{P}_d$ and reactive power demands of loads $\mathbf{Q}_d$. The input matrix of DSPP is $\mathcal{I} = [\mathbf{P}_g, \mathbf{P}_w, \mathbf{V}_g, \mathbf{V}_w, \mathbf{P}_d, \mathbf{Q}_d]^T$. The output variables are mean of TSI μ_{TSI} and standard deviation of TSI σ_{TSI} . Because the sum of the probability density of the Normal distribution is 1, the transient stability chance constraint is rewritten as

$$\mathbb{C}(\mu_{\text{TSI}}, \sigma_{\text{TSI}}, \lambda) \geq \varepsilon \quad (34)$$

where $\mathbb{C}$ is the cumulative probability density of the Normal distribution in the interval $[\mu_{\text{TSI}} - \lambda\sigma_{\text{TSI}}, \mu_{\text{TSI}} + \lambda\sigma_{\text{TSI}}]$; λ is a coefficient that depends on ε . The calculation of λ is

based on the integral of the probability density function of the Normal distribution.

$$\frac{1}{\sqrt{2\pi}\sigma_{\text{TSI}}} \int_{\mu_{\text{TSI}} - \lambda\sigma_{\text{TSI}}}^{\mu_{\text{TSI}} + \lambda\sigma_{\text{TSI}}} \exp\left(-\frac{(x - \mu_{\text{TSI}})^2}{2\sigma_{\text{TSI}}^2}\right) dx = \varepsilon \quad (35)$$

For a preset ε and corresponding λ , the specific expression of (34) is

$$\mu_{\text{TSI}} - \lambda\sigma_{\text{TSI}} > 0 \quad (36)$$

The meaning of (36) is that the probability of $\text{TSI} > 0$ is more than ε . At the same time, if $\mu_{\text{TSI}} - \lambda\sigma_{\text{TSI}} > 0$ is satisfied, $\mu_{\text{TSI}} + \lambda\sigma_{\text{TSI}} > 0$ is true.

The training samples of DSPP is generated through time-domain transient stability simulations. According to the parameters of the test system, we set the variation range of the variables in $\mathcal{I}$. Then, the Latin hypercube sampling (LHS) [9] is applied to generate N input samples, which correspond to N operation scenarios. For each scenario, we perform three-phase short-circuit faults for credible contingencies in the test system. Suppose there are L credible contingencies in the test system, L TSIs are generated through L time-domain simulations. We choose the minimum TSI in L TSIs as the target, which corresponds to the most severe contingencies in this scenario. It is worth noting that this paper focuses on preventive control and thus the credible contingency list is known to system operators. This list is typically determined by engineering practice. By performing the above process for all N scenarios, we get N sample pairs for training DSPP.

C. Chance-Constrained Transient Stability Preventive Control Framework and Solution

The DSPP approach allows us to achieve a computationally efficient assessment of the probabilistic transient stability as compared to Monte Carlo and its variants. It also enables modeling the probabilistic transient stability a chance constraint that can be effectively used for OPF. To this end, the chance-constrained transient stability preventive control framework is established. This is a nonlinear programming problem. The primal dual interior point method (PDIPM) is one of the most widely used algorithms for solving such problems [30]. In this paper, PDIPM can be used to solve this chance-constrained optimal power flow problem, where the objective function is (1) with constraints (2)-(9) and (36). In each iteration of PDIPM, the required μ_{TSI} and λ_{TSI} for chance constraint (36) can be calculated from DSPP.

It is worth noting that for the PDIPM approach, the Jacobian and Hessian matrices associated with the constraints are required. The Jacobian and Hessian matrices of (1)-(9) can be easily calculated by following [31]. However, for those associated with (36), they are derived as follows:

$$\begin{aligned} \mathbf{J}_{\text{CC}} &= \mathbf{J}(\mu_{\text{TSI}} - \lambda\sigma_{\text{TSI}}) = \mathbf{J}(\mu_{\text{TSI}}) - \lambda\mathbf{J}(\sigma_{\text{TSI}}) \\ &= \frac{\partial\mu_{\text{TSI}}}{\partial\mathcal{I}} - \lambda\frac{\partial\sigma_{\text{TSI}}}{\partial\mathcal{I}} \\ &= \begin{bmatrix} \frac{\partial\mu_{\text{TSI}}}{\partial\mathcal{I}_1} & \cdots & \frac{\partial\mu_{\text{TSI}}}{\partial\mathcal{I}_K} \end{bmatrix}^T \\ &\quad - \lambda \begin{bmatrix} \frac{\partial\sigma_{\text{TSI}}}{\partial\mathcal{I}_1} & \cdots & \frac{\partial\sigma_{\text{TSI}}}{\partial\mathcal{I}_K} \end{bmatrix}^T \end{aligned} \quad (37)$$

$$\begin{aligned} \mathbf{H}_{\text{CC}} &= \mathbf{H}(\mu_{\text{TSI}} - \lambda\sigma_{\text{TSI}}) = \mathbf{H}(\mu_{\text{TSI}}) - \lambda\mathbf{H}(\sigma_{\text{TSI}}) \\ &= \frac{\partial^2\mu_{\text{TSI}}}{\partial\mathcal{I}^2} - \lambda\frac{\partial^2\sigma_{\text{TSI}}}{\partial\mathcal{I}^2} \\ &= \begin{bmatrix} \frac{\partial^2\mu_{\text{TSI}}}{\partial\mathcal{I}_1\partial\mathcal{I}_1} & \cdots & \frac{\partial^2\mu_{\text{TSI}}}{\partial\mathcal{I}_1\partial\mathcal{I}_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\mu_{\text{TSI}}}{\partial\mathcal{I}_K\partial\mathcal{I}_1} & \cdots & \frac{\partial^2\mu_{\text{TSI}}}{\partial\mathcal{I}_K\partial\mathcal{I}_K} \end{bmatrix} \\ &\quad - \lambda \begin{bmatrix} \frac{\partial^2\sigma_{\text{TSI}}}{\partial\mathcal{I}_1\partial\mathcal{I}_1} & \cdots & \frac{\partial^2\sigma_{\text{TSI}}}{\partial\mathcal{I}_1\partial\mathcal{I}_K} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2\sigma_{\text{TSI}}}{\partial\mathcal{I}_K\partial\mathcal{I}_1} & \cdots & \frac{\partial^2\sigma_{\text{TSI}}}{\partial\mathcal{I}_K\partial\mathcal{I}_K} \end{bmatrix} \end{aligned} \quad (38)$$

where $\mathbf{J}_{\text{CC}}$ and $\mathbf{H}_{\text{CC}}$ denote the Jacobian and Hessian matrices of the chance constraint, respectively; K denotes the dimension of $\mathcal{I}$. The essence of (37) and (38) is to calculate the first and second order partial derivatives of μ_{TSI} and σ_{TSI} to each variable in $\mathcal{I}$.

In this paper, DSPP is implemented in the GPyTorch library which is based on PyTorch framework [32], [33]. Before training DSPP, the training samples should be normalized via

$$\mathcal{I}^{\text{nor}} = 2 \times \frac{\mathcal{I} - \mathcal{I}^{\min}}{(\mathcal{I} - \mathcal{I}^{\min})^{\max}} - 1 \quad (39)$$

$$\text{TSI} = \frac{\text{TSI} - \mu(\text{TSI})}{\sigma(\text{TSI} - \mu(\text{TSI}))} \quad (40)$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are used to calculate mean and standard deviation. Then, we train DSPP with normalized samples, and $\mathbf{J}(\mu_{\text{TSI}})$, $\mathbf{J}(\sigma_{\text{TSI}})$, $\mathbf{H}(\mu_{\text{TSI}})$ and $\mathbf{H}(\sigma_{\text{TSI}})$ can be calculated through an automatic differentiation function in PyTorch [34]. For example, $\mathbf{J}(\mu_{\text{TSI}})$ and $\mathbf{H}(\sigma_{\text{TSI}})$ can be calculated by

$$\begin{aligned} \mathbf{J}(\mu_{\text{TSI}})^{\text{nor}} &= \text{torch.autograd.functional.jacobian} \\ &\quad \left(\sum_{s=1}^S \exp(w^{(s)}) \times \mu_{\text{TSI}}, \mathcal{I}^{\text{nor}} \right) \end{aligned} \quad (41)$$

$$\begin{aligned} \mathbf{H}(\mu_{\text{TSI}})^{\text{nor}} &= \text{torch.autograd.functional.jacobian} \\ &\quad (\mathbf{J}(\mu_{\text{TSI}})^{\text{nor}}, \mathcal{I}^{\text{nor}}) \end{aligned} \quad (42)$$

$$\mathbf{J}(\mu_{\text{TSI}}) = \frac{\mathbf{J}(\mu_{\text{TSI}})^{\text{nor}} \times \mu(\text{TSI})}{2 \times (\mathcal{I} - \mathcal{I}^{\min})^{\max}} \quad (43)$$

$$\mathbf{H}(\mu_{\text{TSI}}) = \frac{\mathbf{H}(\mu_{\text{TSI}})^{\text{nor}} \times \mu(\text{TSI}) \times 4}{(\mathcal{I} - \mathcal{I}^{\min})^{\max, T} \otimes (\mathcal{I} - \mathcal{I}^{\min})^{\max}} \quad (44)$$

where $\otimes$ denotes the matrix multiplication. For $\mathbf{J}(\sigma_{\text{TSI}})$ and $\mathbf{H}(\sigma_{\text{TSI}})$, they have the same calculation process as $\mathbf{J}(\mu_{\text{TSI}})$ and $\mathbf{H}(\mu_{\text{TSI}})$. Then, we substitute $\mathbf{J}$ and $\mathbf{H}$ of μ_{TSI} and σ_{TSI} in (37) and (38) to get $\mathbf{J}_{\text{CC}}$ and $\mathbf{H}_{\text{CC}}$. It should be noted that in each iteration of PDIPM, $\mathbf{J}_{\text{CC}}$ and $\mathbf{H}_{\text{CC}}$ need to be updated, so the use of GPU can greatly speed up the iteration of PDIPM. The detailed steps of the chance-constrained preventive control framework proposed in this paper are shown in Fig. 1, including sample generation, DSPP training and preventive control. The sample generation and DSPP training are offline steps, and only PDIPM-based preventive control is implemented online. The detailed iteration process of DSPP training and preventive control are shown in Algorithm 1

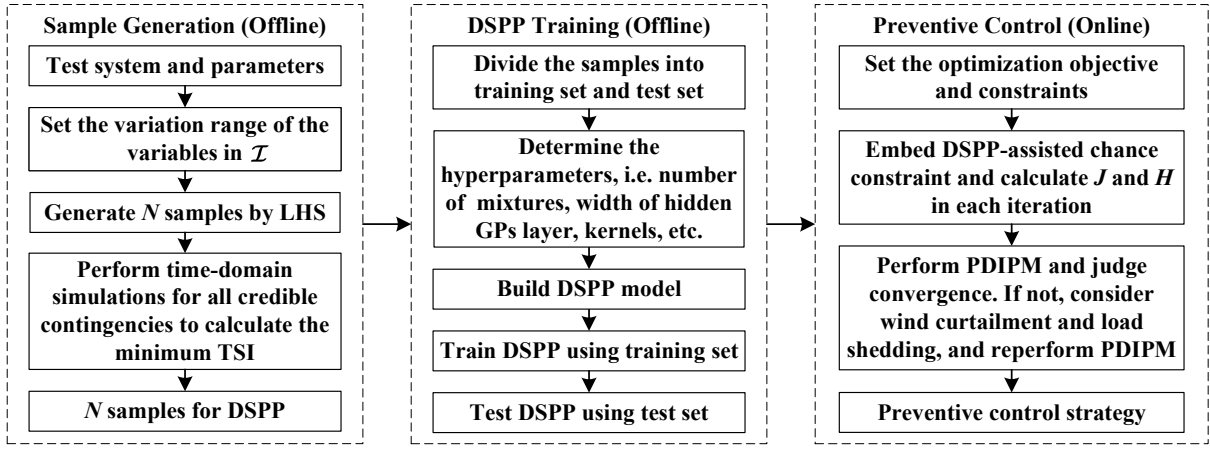


Fig. 1. Proposed DSPP assisted chance-constrained power system transient stability preventive control framework.

and Algorithm 2, respectively. The proposed method can be applied in other stability issues, following the steps similar to Fig. 1, but beyond the scope of this paper.

Algorithm 1: Data generation and DSPP training

Input: Parameters of systems; PyTorch

Output: Well-trained DSPP

- 1 Set variation range of P_g, P_w, V_g, V_w, P_d and Q_d
- 2 Generate N random discrete samples using LHS, where P_g, P_w, V_g, V_w, P_d and Q_d are limited to their variation range
- 3 **for** $i = 1 : N$ **do**
- 4 **for** $j = 1 : C(\text{number of credible contingencies})$ **do**
- 5 Perform time-domain simulation to sample i for contingencies j , and get $\text{TSI}_{i,j}$
- 6 **end**
- 7 $\text{TSI}_i = \min(\text{TSI}_{i,1:C})$
- 8 **end**
- 9 Generate N samples, where P_g, P_w, V_g, V_w, P_d and Q_d are input, and $\text{TSI}_{1:N}$ are output
- 10 Divide samples into training set and test set
- 11 Build DSPP in PyTorch
- 12 Train DSPP using training set
- 13 Test DSPP using test set

IV. NUMERICAL RESULTS

The proposed method is tested on the modified IEEE 39-bus system with two wind farms located on buses 17 and 21, both of which have a maximum active power output of 500 MW and belong to the type III wind generator, i.e., DFIG. The variation range of synchronous generator active power outputs P_g and active power demands of loads P_d are 80% ~ 120%. The variation range of wind farm active power outputs P_w is 0% ~ 100%. The variation range of generators' bus voltage magnitudes V_g and wind farms' bus voltage magnitudes V_w are 1.0 ~ 1.05 pu. Reactive power demands of loads Q_d changes with active power demands of loads P_d . Based on the above variation range, 10000 samples are generated by using

Algorithm 2: Iteration process of preventive control

Input: $P_g, P_w, V_g, V_w, P_d, Q_d$, and other system parameters; well-trained DSPP

Output: Optimal control strategy

- 1 Set variation range of P_g, P_w, V_g, V_w, P_d and Q_d , and use the middle value of that range as the initial iteration value of PDIPM
- 2 Substitute transient stability constraint with (36)
- 3 **while** not converge or not reach the maximum iteration number **do**
- 4 Use DSPP to calculate the value, Jacobian and Hessian matrices of (36) by calculating $\mu_{\text{TSI}}, \sigma_{\text{TSI}}, \mathbf{J}(\mu_{\text{TSI}}), \mathbf{J}(\sigma_{\text{TSI}}), \mathbf{H}(\mu_{\text{TSI}})$ and $\mathbf{H}(\sigma_{\text{TSI}})$ according to (37) and (38)
- 5 The value, Jacobian and Hessian matrices are applied in PDIPM for iteration [30]
- 6 Calculate value, Jacobian and Hessian matrices for cost function and other static stability constraints
- 7 **end**

LHS, of which 8000 samples are used as training set and 2000 samples are used as the test set. The corresponding minimum TSI is obtained by performing time-domain simulations on credible contingencies. The combination of $\mathcal{I}$ and minimum TSI represent the generated samples. A 64-Bit computer with Intel(R) Core(TM) i9-12900KF 3.19GHz CPU, 128GB RAM and NVIDIA GeForce RTX 3090 24GB GPU is used to perform numerical results.

A. DSPP Training and Performance Evaluation

We build a DSPP, whose size of minibatch is 500, the number of mixtures of the outputs is 8, the width of the hidden GP layer is 7, the number of epochs of training is 400, the number of inducing points in each hidden layer is 300; the optimizer is Adam, the kernel is Matérn kernel, the initial learning rate is 0.01. In addition, the learning rate piecewise constant decay method is used to accelerate the convergence speed and improve the fitting accuracy. Then, we use training

samples to train DSPP, and verified by test set samples. The performance of DSPP can be estimated by transient stable and unstable classification accuracy, mean absolute error (MAE) between TSI target and μ_{TSI} , false alarm rate (FA) and miss detection rate (MD). The equations of classification accuracy, MAE, FA and MD are

$$A_P = \frac{1}{N_T} \sum_{i=1}^{N_T} (\mu_{TSI} \times TSI_{Ti} > 0) \quad (45)$$

$$MAE = \frac{1}{N_T} \sum_{i=1}^{N_T} |\mu_{TSI} - TSI_{Ti}| \quad (46)$$

$$FA = \frac{1}{N_T} \sum_{i=1}^{N_T} (\mu_{TSI} < 0 \text{ and } TSI_{Ti} > 0) \quad (47)$$

$$MD = \frac{1}{N_T} \sum_{i=1}^{N_T} (\mu_{TSI} > 0 \text{ and } TSI_{Ti} < 0) \quad (48)$$

where A_P is the classification accuracy; N_T is the number of test set samples; TSI_{Ti} denotes the TSI target of the test set samples.

We also compare DSPP with several other probabilistic and deterministic prediction models. For probabilistic prediction models, the GP, DGP, and deep kernel learning (DKL) [35] are compared. For deterministic models, the four-layer stacked autoencoder (SAE), the three-layer artificial neural network (ANN), and the four-layer deep belief network (DBN) are compared. The hyperparameters of the above models are tuned to have optimal performances. The accuracy comparison results of above models are shown in Table I.

TABLE I
COMPARISON OF PERFORMANCE OF DIFFERENT PREDICTION MODELS
FOR 39-BUS SYSTEM

Type	Model	A_P	MAE	FA	MD
Probabilistic model	DSPP	97.4%	6.1	1.50%	1.10%
	GP	86.4%	35.7	5.95%	7.65%
	SGP	88.1%	33.2	7.10%	4.80%
	DGP	94.1%	22.3	3.55%	2.35%
	DKL	94.0%	12.6	2.90%	3.10%
Deterministic model	SAE	96.1%	9.2	2.25%	1.65%
	ANN	96.3%	12.4	1.40%	2.30%
	DBN	97.5%	8.6	1.05%	1.45%

From Table I, we can see that DSPP has the best performance compared with other probabilistic prediction models according to classification accuracy and MAE. By contrast, in the deterministic prediction models, DBN has the best performance. The classification accuracy of DSPP is 97.4%, which is almost the same as DBN, while the MAE of DSPP is smaller than DBN (6.1 and 8.6 respectively). In addition, DSPP can give the distribution of the predicted TSI, which is crucial in transient stability risk assessment. It should be noted that FA will carry out unnecessary control and increase costs, while MD will generate potential security risks. In addition, the accuracy, FA and MD all take $TSI = 0$ as the boundary of stability. If μ_{TSI} increases and moves away from $TSI = 0$, prediction reliability will increase because FA and MD mostly occur close to $TSI = 0$. The offline training time of DSPP is

337s and the prediction times for all models are less than 0.01s, which meets the requirements of online applications. In addition, we test the single calculation times of prediction, Jacobian and Hessian matrices, which are 0.008s, 0.008s, and 0.2s respectively.

In preventive control, different control reliability probabilities correspond to different control costs. In this paper, we set four confidence probabilities, including 95%, 90%, 80%, and 70%. According to (35), the corresponded λ is 1.960, 1.645, 1.282 and 1.036 respectively. We randomly select 50 samples in the test set and plot confidence intervals of four different probabilities, as shown in Fig. 2 (The discrete samples are arranged in ascending order of TSI true labels).

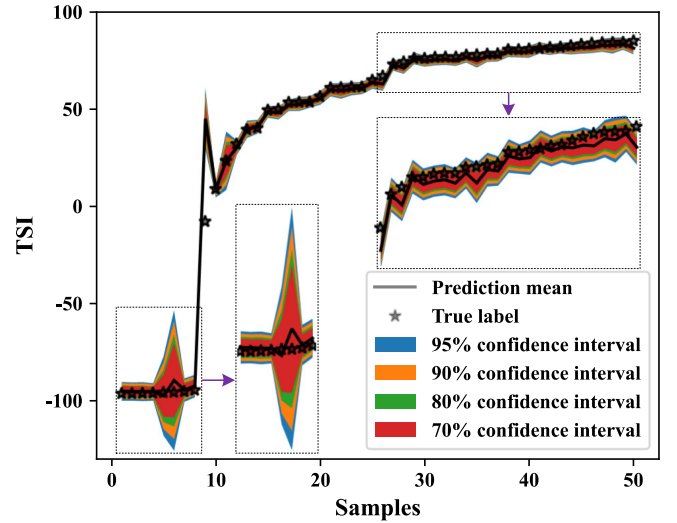


Fig. 2. The visualization of the prediction performance of DSPP.

It can be seen from Fig. 2 that the prediction mean and four confidence intervals are accurately distributed along the true labels. The condition of (36) is satisfied when the lower bound of the confidence interval of the preset confidence probability is greater than 0. In addition, transient stability margins vary with different operating scenarios, which can better increase the control success rate and reduce the control costs. In the preventive control later, we set the confidence probability to 95% to ensure the reliability of the control strategy.

B. Chance-Constrained Power System Preventive Control

Based on the DSPP, the preventive control can be implemented according to Fig. 1. In real-time power system operation, we perform look-ahead transient stability predictions at every time interval (e.g., 15 minutes) for all credible contingencies. These predictions can be quickly achieved through DSPP, taking less than 1 second per prediction according to Table I. If a transient instability risk is detected, the PDIPM-based preventive control is activated.

In this section, we first show the generator redispatch as the control method while ignoring the wind curtailment and load shedding to ensure full generation of renewable energy and power supply reliability. This curtailment and load shedding is tested in the later sections. For a scenario with unstable risk,

TABLE II
ITERATIVE PROCESS OF FOUR INDICES

Index	1	2	3	4	5	6	7	Iteration	8	9	10	11	12	13	14	15
objective	41500	41671	42716	43028	43165	43062	42726	41730	41385	41257	41226	41215	41207	41205	41204	
step size	10.92	18.56	3.45	2.17	0.93	0.30	0.46	1.92	1.36	0.65	1.11	1.18	1.15	0.27	0.35	
fcond	3.1e-3	2.5e-3	3.0e-4	1.3e-4	9.5e-6	5.6e-7	8.8e-7	1.7e-5	9.1e-6	1.3e-6	5.5e-6	1.1e-5	4.9e-6	1.7e-6	9.3e-7	
gcond	1.44	1.57	1.31	1.13	0.99	0.96	0.86	0.66	0.11	3.8e-2	2.0e-2	2.4e-2	6.9e-3	2.1e-3	4.1e-4	
ccond	4835	498	218	122	48.9	4.8	1.2	0.55	7.2e-2	7.6e-3	2.7e-3	1.1e-3	3.5e-4	1.3e-4	2.2e-5	
ocond	3.6e-2	1.5e-2	2.0e-2	5.9e-3	2.6e-3	1.9e-3	6.3e-3	1.9e-3	6.7e-3	2.5e-3	6.1e-4	2.1e-4	1.6e-4	3.5e-5	1.8e-5	

we perform the preventive control using PDIPM with 15 iterations. According to [30], four indices are defined to judge the convergence of PDIPM, including fcond (termination tolerance for feasibility), gcond (termination tolerance for gradient), ccond (termination tolerance for complementary condition) and ocond (termination tolerance for cost condition). The iterative process of the objective function, step size and four indices are shown in Table II. From Table II, we can see that the optimization objective, i.e., the fuel cost of generators, has changed from 41500 to 41204 with a reduction of 296. The step size of iteration is changed from the initial 10.92 to 0.35 to accelerate convergence in the early stage of iteration and improve control accuracy in the later stage. In the last iteration, the four convergences indices are all less than $1e-3$, which is the tolerance of PDIPM convergence. The generation redispatch results, cost constants, and total cost are shown in Table III.

TABLE III
GENERATOR REDISPATCH RESULTS FOR 39-BUS SYSTEM

Gen.	Before control/MW	After control/MW	a_g	b_g	c_g	Gen. cost/\$
1	246.3	329.8	0.013	0.32	0.20	1520
2	565.6	561.1	0.012	0.31	0.20	3952
3	510.9	607.5	0.011	0.30	0.22	4242
4	724.8	588.1	0.011	0.31	0.20	3987
5	447.2	539.2	0.012	0.32	0.20	3661
6	753.4	649.6	0.010	0.30	0.20	4415
7	593.6	645.9	0.010	0.29	0.20	4360
8	625.5	637.1	0.010	0.30	0.21	4250
9	743.3	707.6	0.009	0.29	0.20	4711
10	860.2	856.3	0.008	0.28	0.20	6106

The total cost is \$41204.

In Table III, the fuel cost constants a_g , b_g and c_g are randomly generated according to a reference value in [20], where $a_g = 0.01$, $b_g = 0.3$ and $c_g = 0.2$ for all generators. For three fuel cost constants, a_g is the dominant cost constant, which accounts for about 93% ~ 96% of the total cost. Generator 1, which has the highest a_g , has the minimum active power output after control, whereas generators 9 and 10 have the maximum power output for the lower a_g . From Tables II and III, it can be concluded that the PDIPM-based preventive control has achieved good generator redispatch. Then, we test the reliability of PDIPM-based strategy through time-domain simulations for credible contingencies. For this scenario, three transmission lines are at risk of transient instability when in the presence of a three-phase short-circuit fault. The rotor angle trajectories of the generators before and after the preventive

control are shown in Fig. 3 (the trajectories before the control are on the left, and the corresponding trajectories after the control are on the right). From Fig. 3, we can see that before the preventive control, the trajectory is divergent, meaning transiently unstable, but after the preventive control, the generator power angle trajectory fluctuates within a small range, indicating that the system is transient stable. The running time of PDIPM is only 8.4s, which is fast enough for online application.

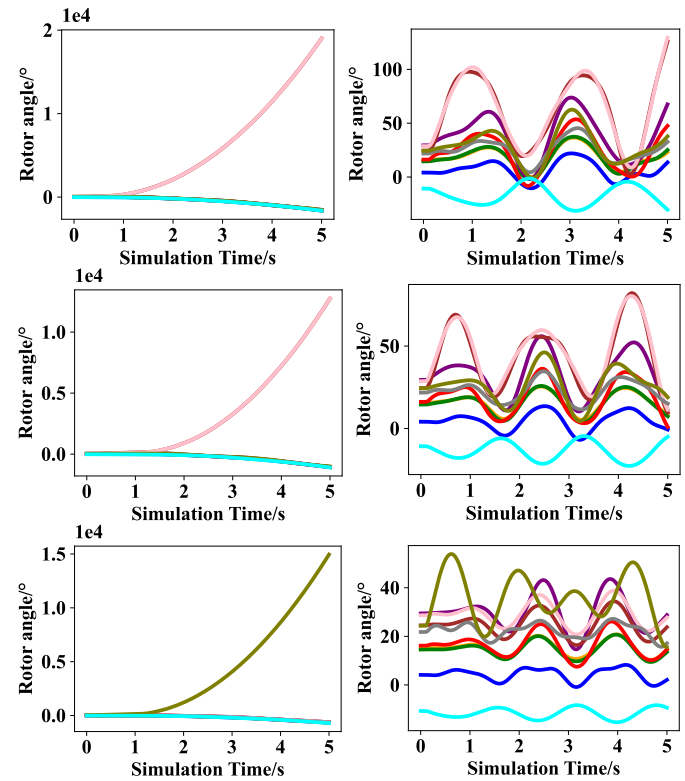


Fig. 3. Rotor angle trajectories of three faults, where the trajectories before the control are on the left, and the corresponding trajectories after the control are on the right (TSI from -96.6 to 38.5).

Then, we randomly select 50 scenarios with instability risk to verify (36), as shown in Fig. 4. According to [22], all scenarios on the purple line in Fig. 4 are greater than 0, which is considered transient stable. In this paper, the chance constraint is $\mu_{TSI} - \lambda\sigma_{TSI} > 0$, which is satisfied when the orange line is greater than 0. The orange line partially overlaps with the zero line, i.e., $\mu_{TSI} - \lambda\sigma_{TSI} = 0$. This is because for these scenarios, when PDIPM calculates the

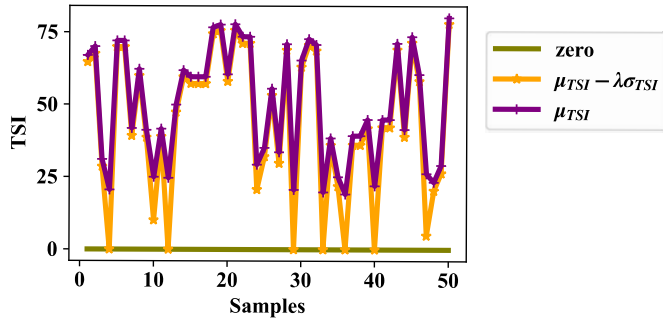


Fig. 4. Visualization of chance constraints in the preventive control.

optimal generator redispatch, the transient stability chance constraint is the dominant constraint, so it converges to 0 to minimize the cost of power generation. However, other scenarios are mainly limited by other constraints, such as the line congestion, so the chance constraint is greater than 0.

C. Preventive Control with Wind Curtailment and Load Shedding

For some severe scenarios, using only generator redispatch cannot find a preventive control strategy that satisfies all constraints. At this time, wind curtailment and/or load shedding are required to identify the right control strategy. For a scenario with instability risk in Fig. 5, when three-phase short circuit fault occurs on lines 2-25 or 16-21, the generator redispatch based PDIPM diverges. Therefore, we consider wind curtailment and load shedding in the preventive control and re-execute PDIPM. The preventive control strategy with wind curtailment and load shedding is shown in Fig. 5.

From Fig. 5, we can see that some loads shed part of the active power demand to make PDIPM converge, i.e., ensuring system stability, and other loads maintain full power supply, such as loads 2, 6, 8, 13, and 18. Due to load shedding, most generators reduce their active power output to maintain active power balance. It is worth noting that the absolute values of the Jacobian and Hessian matrices of the wind farms are small, which means that the wind curtailment has little effect on transient stability improvement. Therefore, the active power output of the wind farms remains unchanged. We also verify the preventive control strategy through time-domain simulations. According to simulation results, TSI is changed from -35.3 to 81.2, indicating the effectiveness of the proposed method for ensuring system stability.

D. Scalability to Larger-Scale System

To further demonstrate its scalability to larger-scale systems, the proposed method is tested on the South Carolina 500-bus system, which serves 21 counties and about 2.6 million people and built on the PowerWorld [40], [41]. In the 500-bus system, 200 loads are supplied by 51 synchronous generators and 9 wind farms (type III wind generator, i.e., DFIG). The wind farms, with a total capacity of 1372 MW, account for 21.3% of the total system capacity (1372 MW/6456 MW). The one-line diagram of the 500-bus system is shown in Fig.

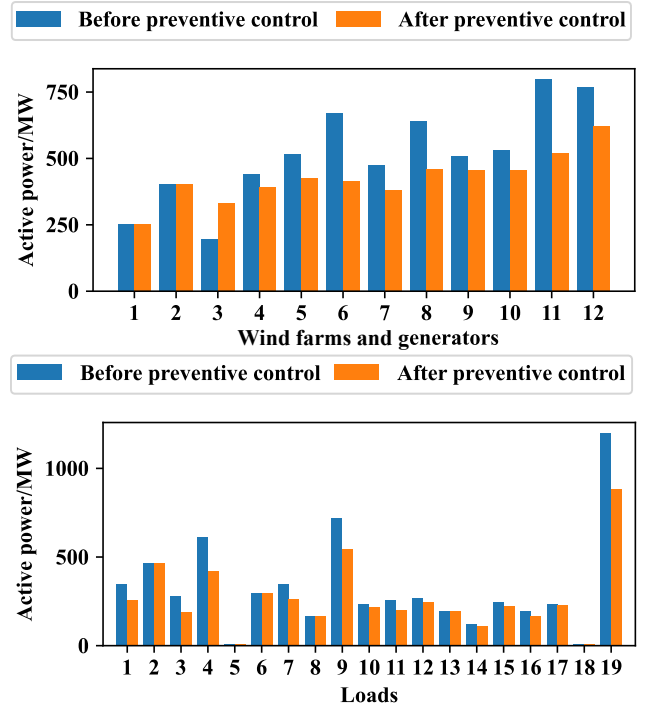


Fig. 5. Preventive control strategy of generators and loads (generators 1 and 2 are wind farms while the others are synchronous generators).

6. The variation range of synchronous generator active power outputs P_g and active power demands of loads P_d are 55% ~ 105%, respectively. The variation range of wind farm active power P_w is 0% ~ 100%. Similar to 39-bus system, Q_d varies with P_d . Based on the aforementioned variation range, 20000 samples are generated using LHS, of which 16000 samples are utilized as the training set and 4000 samples as the test set. The corresponding minimum TSI is obtained by performing time-domain simulations for all credible contingencies.

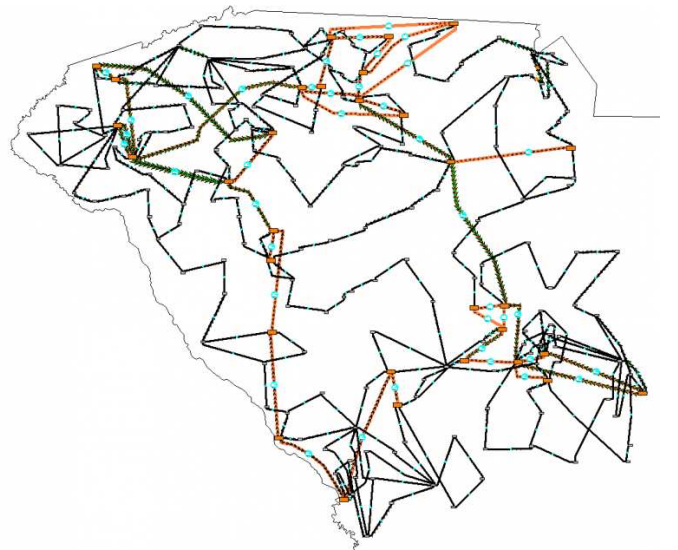


Fig. 6. One-line diagram of the South Carolina 500-bus system [40], [41].

Similar to the 39-bus system, DSPP is built in PyTorch, in which the size of the minibatch is 500, the number of mixtures

of the outputs is 8, the width of the hidden GP layer is 6, the number of training epochs is 400, the number of inducing points in each hidden layer is 300; the optimizer is Adam, the kernel is Matérn kernel, the initial learning rate is 0.01, the learning rate piecewise constant decay method is also used. The performance of the DSPP is compared to that of other probabilistic and deterministic models in Table IV.

TABLE IV
COMPARISON OF PERFORMANCE OF DIFFERENT PREDICTION MODELS FOR 500-BUS SYSTEM

Type	Model	A_P	MAE	FA	MD
Probabilistic model	DSPP	99.0%	3.4	0.950%	0.050%
	GP	/	/	/	/
	SGP	/	/	/	/
	DGP	/	/	/	/
Deterministic model	DKL	96.8%	7.8	2.750%	0.450%
	SAE	97.1%	4.2	2.050%	0.850%
	ANN	99.2%	2.5	0.075%	0.725%
	DBN	99.4%	2.3	0.375%	0.225%

In Table IV, we evaluate a variety of hyperparameters for GP, SGP, and DGP, but the results do not converge, demonstrating the limitations of these models for large-scale systems. DSPP has the highest A_P (99.0%) and minimal MAE (3.4) in five probabilistic models. This demonstrates that DSPP is the most accurate probabilistic model for this problem. Compared with deterministic models, DSPP outperforms SAE, but is slightly worse than ANN and DBN. However, DSPP's accuracy is sufficient and it can generate the predicted TSI distribution for evaluating the risk of transient stability. According to MD of DSPP, the probability of potential risks is only 0.050%. The offline training time of DSPP is 659s and the prediction times for all models are less than 0.03s, which meets the requirements of online applications. In addition, we test the single calculation times of prediction, Jacobian and Hessian matrices, which are 0.02s, 0.009s and 1.5s respectively.

Based on the DSPP, the preventive control is implemented according to Fig. 1 for a random unstable scenario, where the control strategy can be found via generator dispatch. The control strategy is shown in Fig. 7 and the corresponding cost calculation is shown in Table V. In Table V, the leading cost coefficient is b_g , which affects about 60% ~ 90% of the total cost. Generators 9, 11, 15 and 30, which have the lower b_g , all increase the active power output to decrease the total cost, whereas generators 1, 2 and 42 decrease the power output for the higher b_g . According to the PDIPM results, the DSPP-based TSI is 36.4. To verify this control strategy, the time-domain simulation-based TSI is calculated, which equals 42.4 and satisfies the requirement for transient stability. The corresponding rotor angle trajectories before and after preventive control are shown in Fig. 8. The running time of the proposed method is 113s, which meets the requirement of 15 minutes or 1-hour preventive control interval.

E. Comparison with Sensitivity-based Method

In this section, we compare the proposed method with the sensitivity-based method, where the sensitivity index is calculated based on the full-system model and time-domain

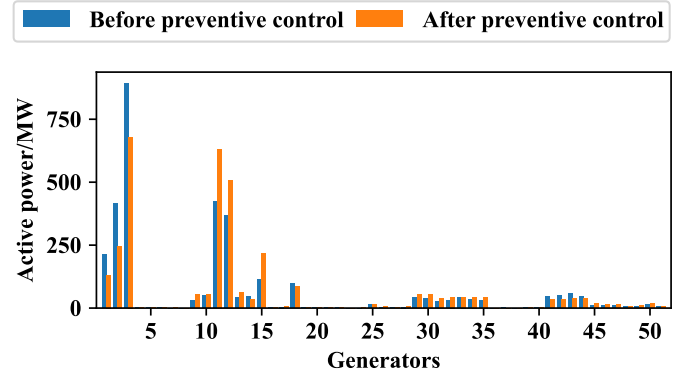


Fig. 7. Preventive control strategy of generators of 500-bus system.

TABLE V
TSC-OPF RESULTS AND TOTAL COST FOR 500-BUS SYSTEM

Gen.	Before control/MW	After control/MW	C_g/C_r			Gen. cost/\$
			a_g	b_g	c_g	
1	214.0	131.4	0.0015	26	1241	4683
2	416.3	244.5	0.0021	26	306	6787
3	892.6	679.2	0.0022	21	296	15574
9	29.9	52.6	0.0010	15	424	1215
11	424.8	629.6	0.0011	16	1046	11555
12	368.6	507.8	0.0022	19	620	10836
13	41.3	62.8	0.0017	20	808	2070
14	47.4	32.9	0.0023	24	1292	2084
15	113.6	216.6	0.0016	17	651	4408
18	98.8	86.9	0.0028	25	239	2432
30	38.1	54.6	0.0017	15	1043	1867
42	50.6	33.5	0.0016	26	1150	2022
43	56.8	37.5	0.0027	24	755	1658
Others	458.6	533.2	/	/	/	35491

Others denote other synchronous generators; The total cost is \$102683.

simulation. We replace the intractable non-convex nonlinear transient stability constraints with sensitivity-based constraint [36], [37]. This optimization problem can be expressed as:

$$\text{Minimize (1)} \quad (49)$$

$$\text{s.t. (2) ~ (9)} \quad (50)$$

$$\text{TSI} > \text{TSI}_B \mid \mathbf{J} = \frac{\partial \text{TSI}}{\partial \mathbf{x}}, \mathbf{H} = 0 \quad (51)$$

where the objective is to minimize the total cost; the transient stability constraint is the time-domain simulation-based TSI. According to [23], TSI is set to be 33.3; $\mathbf{x}$ denotes the controllable variable, including P_g , P_w and P_d ; $\frac{\partial \text{TSI}}{\partial \mathbf{x}}$ denotes the sensitivity of TSI to $\mathbf{x}$; $\mathbf{J}$ of PDIPM is updated every iteration; Since sensitivity-based constraint is a linear

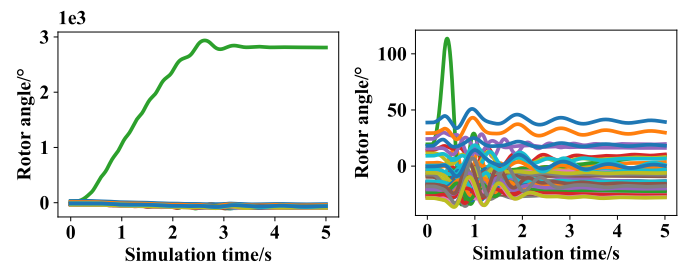


Fig. 8. Rotor angle trajectories of 500-bus system.

constraint, $\mathbf{H}$ is constant at 0. The main difference between the sensitivity-based method and the proposed method is that the sensitivity is calculated based on the time-domain simulation, and the Hessian matrix of the transient stability constraint is not considered. For the 39-bus system, we randomly select 10 scenarios with instability risk to compare their performances. The results are shown in Table VI.

TABLE VI
COMPARISON WITH SENSITIVITY-BASED METHOD FOR 39-BUS SYSTEM

Scenario	Control method		Total cost/\$		Time	
	A	B	A	B	A	B($\approx$)
1	G	G	37224	37224	12s	15min
2	G	S	/	/	10s	70min
3	G	S	/	/	14s	70min
4	G	N	/	/	11s	/
5	G	G	36892	36892	13s	55min
6	G	S	/	/	11s	130min
7	G	N	/	/	10s	/
8	G	N	/	/	11s	/
9	G	S	/	/	9s	170min
10	G	S	/	/	9s	90min

A: proposed method; B: sensitivity-based method; G: generator redispatch; S: generator redispatch, wind curtailment and load shedding; N: not converge.

In Table VI, we only compare the total cost of using the same control method and calculate the iteration time of PDIPM. It can be found that our proposed method has better convergence and can find preventive control strategies that satisfy all constraints by generator redispatch in these 10 scenarios, while the sensitivity-based method requires consideration of wind curtailment and load shedding and may not converge. Scenarios 1 and 5 have the same control method and the same total cost because they get the same preventive control strategy, indicating the DSPP in the proposed method can achieve the same accuracy as the time-domain simulation. There is a significant difference in the iteration time between the two methods. The proposed method can iteratively converge within 15s, while the sensitivity-based method requires time-domain simulations for all credible contingencies in each iteration, which is time-consuming and difficult to meet the requirements of online applications. In the sensitivity-based method, when only generator redispatch is used, it is only necessary to calculate the sensitivities of $\mathbf{P}_g$, and the iteration time is tens of minutes, such as 15min and 55min. When considering wind curtailment and load shedding, the sensitivities of $\mathbf{P}_w$ and $\mathbf{P}_d$ also need to be calculated, and the iteration time can reach up to 170min.

For the 500-bus system, we also choose 10 random unstable scenarios to compare the performances, see Table VII. Similar to the 39-bus system, only scenarios with the same control method compare the total cost in Table VII. The sensitivity-based method can only converge for three of the ten scenarios (2 G and 1 S), while the proposed method can converge for all ten (9 G and 1 S). The total cost of the two methods in scenarios 2 and 3 is the same, but the calculation times differ by around 50 times. For scenario 9, due to the wind curtailment and load shedding, the dimension of Jacobian matrices increases from 51 (the number of $\mathbf{P}_g$) to 460 (the number

of $\mathbf{P}_g$, $\mathbf{P}_w$, $\mathbf{P}_d$ and $\mathbf{Q}_d$). As a result, the calculation time increases 370 times from 97s to 10h. From the comparison results, it can be concluded that the proposed method has a great improvement in the convergence and computational efficiency to meet practical applications.

TABLE VII
COMPARISON WITH SENSITIVITY-BASED METHOD FOR 500-BUS SYSTEM

Scenario	Control method		Total cost/\$		Time	
	A	B	A	B	A	B($\approx$)
1	G	N	/	/	143s	/
2	G	G	104007	104007	107s	120min
3	G	G	103911	103911	112s	100min
4	G	N	/	/	102s	/
5	G	N	/	/	97s	/
6	G	N	/	/	102s	/
7	G	N	/	/	92s	/
8	G	N	/	/	107s	/
9	G	S	/	/	97s	10h
10	S	N	/	/	159s	/

A, B, G, S and N are defined in Table VI.

F. Comparison with Heuristic Algorithm

Heuristic algorithms are widely employed in engineering control. NSGA-II and NSGA-III are two advanced multi-objective heuristic algorithms and applied in transient stability control of power systems [38], [39]. In this section, the proposed method is compared with the DBN-based NSGA-II and NSGA-III, and time-domain simulation-based NSGA-II and NSGA-III. The comparison results of the two systems are shown in Table VIII. It should be emphasized that just the generator redispatch is used, and the total cost of generators and TSI are objectives to construct a multi-objective optimization problem in NSGA-II and NSGA-III. Similar to [38], the population size and number of iterations are both set to 200.

TABLE VIII
COMPARISON WITH HEURISTIC ALGORITHM

Method	39-bus system			500-bus system		
	Cost	TSI	Time	Cost	TSI	Time
TDS-NSGA-II	41624	68.9	$\approx 105h$	100638	51.2	$\approx 100h$
TDS-NSGA-III	41285	60.2	$\approx 105h$	100642	50.8	$\approx 100h$
DBN-NSGA-II	41580	40.3	145s	100764	51.1	164s
DBN-NSGA-III	41299	66.9	3.5s	100639	50.8	7.0s
PM	41204	38.5	8.4s	100635	42.4	113s

TDS-NSGA-II/III: time-domain simulation-based NSGA-II/III; DBN-NSGA-II/III: DBN-based NSGA-II/III; PM: proposed method.

From Table VIII, we can see that the proposed method performs well in terms of cost, while the TSI is lower than the heuristic algorithms. This is because the cost is the only optimization objective in the proposed method, and TSI is the constraint and not maximized. In addition, the proposed method and DBN-based NSGA-II/III can achieve preventive control within 200s and meet the requirement of online application, but time-domain simulation-based NSGA-II/III is very time-consuming ($> 100h$). Although the heuristic algorithm can obtain results close to those of the proposed method, other static stability constraints, such as bus voltage

magnitude and branch flow constraints, are not considered in its iteration process. Therefore, it is necessary to verify the results because they may not be accurate and reliable. By contrast, the proposed method fully takes into account all constraints during the iteration. In addition, the chance-constrained preventive control is another main advantage of the proposed method.

G. Computational Complexity Analysis

The computational complexity of DSPP-based preventive control is shown in Table IX. For each iteration, one prediction should be performed to obtain μ_{TST} and σ_{TST} , two calculations of Jacobian matrices should be performed to obtain $\mathbf{J}(\mu_{\text{TST}})$ and $\mathbf{J}(\sigma_{\text{TST}})$, and two calculation of Hessian matrices should be performed to obtain $\mathbf{H}(\mu_{\text{TST}})$ and $\mathbf{H}(\sigma_{\text{TST}})$. For the 39-bus and 500-bus systems, the total calculation times related to transient stability constraint for each iteration are 0.424s (0.008s + 2*0.008s+2*0.2s) and 3.038s (0.02s + 2*0.009s+2*1.5s), respectively. Note that these calculation times are appropriate for both G (generator redispatch) and S (generator redispatch, wind curtailment and load shedding) in Tables VI and VII. This is because DSPP can generate Jacobian and Hessian matrices for all $\mathbf{P}_g$, $\mathbf{P}_w$, $\mathbf{V}_g$, $\mathbf{V}_w$, $\mathbf{P}_d$ and $\mathbf{Q}_d$ at one time.

TABLE IX
COMPUTATIONAL COMPLEXITY OF DSPP-BASED METHOD

System	Control method	Prediction	Time Jacobian	Hessian
39-bus system	G/S	0.008s	0.008s	0.2s
500-bus system	G/S	0.02s	0.009s	1.5s

G and S are defined in Table VI.

TABLE X
COMPUTATIONAL COMPLEXITY OF TIME-DOMAIN SIMULATION-BASED METHOD FOR EACH ITERATION

System	Control method	Prediction	Time Jacobian	Hessian
39-bus system	G	6.5s	38s	/
	S	6.5s	222s	/
500-bus system	G	5.8s	280s	/
	S	5.8s	1950s	/

G and S are defined in Table VI.

For time-domain simulation-based preventive control, the computational complexity is shown in Table X. For example, when the 39-bus system has 7 credible contingencies, each control iteration requires 7 time-domain simulations for prediction and 70 (7 * 10, for G) and 434 (7 * 62, for S) times for Jacobian matrices, respectively. Similarly, if the 500-bus system has 5 credible contingencies, each control iteration requires 5 time-domain simulations for prediction and 255 (5 * 51, for G) and 2300 (5 * 460, for S) times for Jacobian matrices, respectively. From the analysis above, it is clear that as the number of credible contingencies rises, the calculation time also increases proportionally. The calculations of Hessian matrices are very time-consuming, so they are not solved in a sensitivity-based method.

H. Impact and Selection of λ

For the 39-bus system, we test the impact of different λ for an unstable scenario, which is shown in Table XI. In Table XI, six different λ are set, corresponding to a probability range of 60.0% to 99.7%. With the increase of λ and probability, the increase in μ_{TST} denotes a rise in control reliability. As a result, the cost also goes up. $\mu_{\text{TST}} - \lambda\sigma_{\text{TST}} = 0$ means that the control strategy is always solved on the boundary of transient stability constraint in this unstable scenario. The choice of λ is the balance of cost and reliability. Generally, it is recommended to choose 2.0 or 3.0, which corresponds to the probability of 95.4% and 99.7%. If we want to decrease FA and MD and increase the prediction reliability, λ should be increased.

TABLE XI
IMPACT OF DIFFERENT λ IN PREVENTIVE CONTROL

λ	Probability	μ_{TST}	$\mu_{\text{TST}} - \lambda\sigma_{\text{TST}}$	Cost/\$
0.8416	60.0%	9.3	0	42697
1.0364	70.0%	11.1	0	42698
1.2815	80.0%	13.0	0	42700
1.6448	90.0%	15.7	0	42702
2.0	95.4%	17.9	0	42704
3.0	99.7%	23.2	0	42710

Note that when PDIPM's iteration is limited by other constraints (e.g., line congestion) as shown in Fig. 4, an increase in λ , i.e., the confidence probability, may not immediately lead to an increase in μ_{TST} . This is because the transient stability constraint is not the dominant constraint until it surpasses the line congestion constraint.

V. CONCLUSION

This paper proposes a new chance-constrained power system transient stability preventive control framework, which can limit the transient security risk under a preset confidence probability. The key idea is to embed the DSPP enabled probabilistic prediction model into the TSC-OPF as a transient stability chance constraint and accelerate the calculation. DSPP can fit the transient stability index offline and generate probability distribution of the predicted value online in less than 0.03s. DSPP-assisted TSC-OPF uses PDIPM to analytically calculate the optimal preventive control strategies that satisfy the chance constraint through the generator redispatch, wind curtailment and load shedding. Numerical results show that the proposed method can quickly and reliably generate preventive control strategies and has a great improvement in convergence and iteration time compared with the sensitivity-based method. In our future work, the preventive control framework will be extended to coordinate different control modes from inverter-based resources with synchronous generators. We will also extend the approach to deal with other stability problems with high penetration of inverter-based resources, such as converter-driven stability control.

REFERENCES

- [1] J. Zhao, *et al.*, "Power system dynamic state estimation: motivations, definitions, methodologies, and future work," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3188–3198, Jul. 2019.

- [2] X. Wang and H.-D. Chiang, "Application of pseudo-transient continuation method in dynamic stability analysis," in *IEEE PES General Meeting*, National Harbor, MD, USA, Jul. 2014, pp. 1–5.
- [3] P. Bhui and N. Senroy, "Real time prediction and control of transient stability using transient energy function," *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 923–934, March 2017.
- [4] H. Yuan and Y. Xu, "Preventive-corrective coordinated transient stability dispatch of power systems with uncertain wind power," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 3616–3626, Sep. 2020.
- [5] Q. Wang and V. Ajjarapu, "PA novel approach to implement generic load restoration in continuation-based quasi-steady-state analysis," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 516–518, Feb. 2005.
- [6] T. Amraee and S. Ranjbar, "Transient instability prediction using decision tree technique," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3028–3037, Aug. 2013.
- [7] F. R. Gomez, A. D. Rajapakse, U. D. Annakkage, and I. T. Fernando, "Support vector machine-based algorithm for post-fault transient stability status prediction using synchronized measurements," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1474–1483, Aug. 2011.
- [8] S. A. Siddiqui, K. Verma, K. R. Niazi, and M. Fozdar, "Real-time monitoring of post-fault scenario for determining generator coherency and transient stability through ANN," *IEEE Trans. Ind. Appl.*, vol. 54, no. 1, pp. 685–692, Jan. 2018.
- [9] T. Su, Y. Liu, J. Zhao, and J. Liu, "Deep belief network enabled surrogate modeling for fast preventive control of power system transient stability," *IEEE Trans. Ind. Inform.*, vol. 18, no. 1, pp. 315–326, Jan. 2022.
- [10] T. Su, Y. Liu, J. Zhao, and J. Liu, "Probabilistic stacked denoising autoencoder for power system transient stability prediction with wind farms," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3786–3789, Jul. 2021.
- [11] B. Wang, B. Fang, Y. Wang, H. Liu, and Y. Liu, "Power system transient stability assessment based on big data and the core vector machine," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2561–2570, Sep. 2016.
- [12] Z. Yue, Y. Liu, Y. Yu, and J. Zhao, "Probabilistic transient stability assessment of power system considering wind power uncertainties and correlations," *Int. J. Electr. Power Energy Syst.*, vol. 117, p. 105649, May 2020.
- [13] P. N. Papadopoulos and J. V. Milanovic, "Probabilistic framework for transient stability assessment of power systems with high penetration of renewable generation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3078–3088, Jul. 2017.
- [14] E. Vaahedi, W. Li, T. Chia, and H. Dommel, "Large scale probabilistic transient stability assessment using bc hydro's on-line tool," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 661–667, May 2000.
- [15] K. Hua, Y. Mishra, and G. Ledwich, "Fast unscented transformation-based transient stability margin estimation incorporating uncertainty of wind generation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1254–1262, Oct. 2015.
- [16] H. Yuan, Y. Xu, and C. Zhang, "Robustly coordinated generation dispatch and load shedding for power systems against transient instability under uncertain wind power," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1032–1043, Mar. 2022.
- [17] G. Qiu *et al.*, "Analytic deep learning-based surrogate model for operational planning with dynamic TTC constraints," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3507–3519, Jul. 2021.
- [18] S. Xia, X. Luo, K. W. Chan, M. Zhou, and G. Li, "Probabilistic transient stability constrained optimal power flow for power systems with multiple correlated uncertain wind generations," *IEEE Trans. Sustain. Energy*, vol. 7, no. 3, pp. 1133–1144, Jul. 2016.
- [19] Y. Chen, S. M. Mazhari, C. Y. Chung, and S. O. Faried, "A preventive dispatching method for high wind power-integrated electrical systems considering probabilistic transient stability constraints," *IEEE Open Access J. Power Energy*, vol. 8, pp. 472–483, 2021.
- [20] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb. 2011.
- [21] D. Pozo and J. Contreras, "A chance-constrained unit commitment with an n-k security criterion and significant wind generation," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 2842–2851, Aug. 2013.
- [22] A. Gupta, G. Gurralla, and P. S. Sastry, "An online power system stability monitoring system using convolutional neural networks," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 864–872, Mar. 2019.
- [23] A. Shamisa, B. Majidi, and J. C. Patra, "Sliding-window-based real-time model order reduction for stability prediction in smart grid," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 326–337, Jan. 2019.
- [24] 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.
- [25] D. Cao *et al.*, "Robust deep gaussian process-based probabilistic electrical load forecasting against anomalous events," *IEEE Trans. Ind. Inform.*, vol. 18, no. 2, pp. 1142–1153, Feb. 2022.
- [26] M. Jankowiak, G. Pleiss, and J. R. Gardner, "Deep sigma point processes," in *Conference on Uncertainty in Artificial Intelligence*, 2020, pp. 789–798.
- [27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. Cambridge, MA: The MIT Press, 2005.
- [28] J. Hensman, N. Fusi, and N. Lawrence, "Gaussian processes for big data," in *Uncertainty in Artificial Intelligence*, Sep. 2013, pp. 282–290.
- [29] H. Salimbeni and M. Deisenroth, "Doubly stochastic variational inference for deep gaussian processes," in *Advances in neural information processing systems*, Nov. 2017, pp. 4588–4599.
- [30] H. Wang, C. E. Murillo-Sanchez, R. D. Zimmerman, and R. J. Thomas, "On computational issues of market-based optimal power flow," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1185–1193, Aug. 2007.
- [31] R. D. Zimmerman, "AC power flows, generalized opf costs and their derivatives using complex matrix notation," Matpower Technical Note 2. Feb. 2010 [Online]. Available: <http://www.pserc.cornell.edu/matpower/TN2-OPF-Derivatives.pdf>.
- [32] J. Gardner, G. Pleiss, K. Q. Weinberger, D. Bindel, and A. G. Wilson, "GPYtorch: blackbox matrix-matrix gaussian process inference with gpu acceleration," in *Advances in Neural Information Processing Systems*, 2018, vol. 31.
- [33] A. Paszke *et al.*, "PyTorch: an imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.
- [34] A. Paszke *et al.*, "Automatic differentiation in pytorch," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS) Workshop*, 2017.
- [35] A. G. Wilson, Z. Hu, R. Salakhutdinov, and E. P. Xing, "Deep kernel learning," in *Artificial intelligence and statistics*, 2016, pp. 370–378.
- [36] T. B. Nguyen and M. A. Pai, "Dynamic security-constrained rescheduling of power systems using trajectory sensitivities," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 848–854, May 2003.
- [37] M. C. Passaro, A. P. A. da Silva, and A. C. S. Lima, "Preventive control stability via neural network sensitivity," *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2846–2853, Nov. 2014.
- [38] T. Han, Y. Chen, J. Ma, Y. Zhao and Y. -y. Chi, "Surrogate modeling-based multi-objective dynamic VAR planning considering short-term voltage stability and transient stability," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 622–633, Jan. 2018.
- [39] T. Han, Y. Chen, and J. Ma, "Multi-objective robust dynamic VAR planning in power transmission grids for improving short-term voltage stability under uncertainties," *IET Gener. Transmiss. Distrib.*, vol. 12, no. 8, pp. 1929–1940, Apr. 2018.
- [40] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *Proc. Bulk Power Syst. Dynamics Control Symp., Portugal*, pp. 1–6, 2017.
- [41] "Electric grid test case repository - synthetic electric grid cases," [Online]. Available: <https://electricgrids.engr.tamu.edu/>.