

Interpretable Data-Driven Probabilistic Power System Load Margin Assessment with Uncertain Renewable Energy and Loads

Bendong Tan, *Student Member, IEEE*, Junbo Zhao, *Senior Member, IEEE*, Weijia Liu, Nan Duan

Abstract—The increasing uncertainties caused by the high-penetration of stochastic renewable generation resources poses a significant threat to the power system voltage stability. To address this issue, this paper proposes a probabilistic deep kernel learning enabled surrogate model to extract the hidden relationship between uncertain sources, i.e., wind power and loads, and load margin for probabilistic load margin assessment (PLMA). Unlike other deep learning approaches, a kernel SHAP provides the sensitivity analysis as well as interpretability of the inputs to outputs influences. This allows identifying the critical factors that affect load margin so that corrective control can be initiated for stability enhancement. Numerical results carried out on the IEEE 118-bus power system demonstrate the accuracy and efficiency of the proposed data-driven PLMA scheme.

Index Terms—Voltage stability, deep kernel learning, surrogate model, interpretability, uncertainty quantification.

I. INTRODUCTION

With the increased penetrations of stochastic renewable generations and flexible loads induced uncertainties, there is an increased concern for quantifying system stability under uncertainties. If these uncertainties are ignored, inappropriate actions may be taken, leading to power system instability, i.e., static voltage stability [1]. Static voltage stability, also known as the load margin, refers to the capability of maintaining at a stable voltage operating point after a small disturbance. Once static voltage stability occurs, power system voltage will collapse. Thus, it is critical to assess the static voltage stability (or load margin) under the strong uncertainties from renewable generations and loads.

To quantify the uncertainties for load margin assessment, many model-based probabilistic load margin assessment (PLMA) approaches have been proposed. The widely utilized method is the Monte Carlo sampling (MCS)-based technique [2], which obtains the load margin distribution via a large number of continuation power flow (CPF) evaluations [3]. However, as each CPF takes a rather long time for large-scale systems, MCS involves tens of thousands of CPF evaluations, restricting their practical applications. Although the Latin-hypercube-sampling [4] has been introduced to reduce the required sample size, the accuracy of this technique is

sacrificed. Besides, with an assumption that the uncertainties of load parameters obey the Gaussian distribution, [5] proposes an analytical approach to calculate load margin with improved computational efficiency. However, it is not scalable to larger-scale systems. Recently, some data-driven PLMA methods are proposed to assess the probabilistic load margin with much higher computational efficiency as compared to the model-based solutions. For the fast global sensitivity analysis, [6] employs polynomial chaos expansion to be the surrogate model for load margin calculation but it is subject to the curse of dimensionality issue. Combined with uncertainty distribution inference, [7] and [8] further propose a nonparametric and reduced-order approximation method, Gaussian process emulator (GPE), for the nonlinear CPF model. In [9], the data-driven polynomial chaos expansion is improved to handle the discrete probabilistic variable in the CPF calculation of N-1 contingency analysis. Note that the interpretability of these surrogate models have not been investigated thoroughly. Specifically, the global sensitivity analysis can only give the variable importance while the importance of uncertain inputs on the load margin assessment that describes a trend between the uncertain inputs and load margin has not been quantified.

This paper proposes an data-driven and interpretable PLMA under uncertain wind generations and loads. The key idea is to develop a probabilistic deep kernel learning to extract the relationship between uncertain sources and load margin. The uncertainty quantification capability of the probabilistic deep kernel learning allows natural PLMA. Unlike other deep learning approaches, a kernel SHAP provides the sensitivity analysis as well as interpretability of the inputs to outputs influences. This allows identifying the critical factors that affect load margin so that corrective control can be initiated for stability enhancement.

II. PROBLEM FORMULATION

In this section, the probabilistic load margin formulation will be briefly introduced. Given the uncertain input vector $\mathbf{x} = [x_1, x_2, \dots, x_n]$, which consists of wind generations as well as active power and reactive power of loads in this paper, the relationship between the uncertain inputs and the corresponding load margin y is:

$$y = f(\mathbf{x}) \quad (1)$$

where n is the dimension of the uncertain inputs $\mathbf{x}$ and each element in $\mathbf{x}$ is corresponding to an uncertain source.

This work is partially supported by U.S. Department of Energy under GMLC program.

B. Tan and J. Zhao are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT; Weijia Liu is with the National Renewable Energy Laboratory, Golden, CO; Nan Duan is with the Lawrence Livermore National Laboratory, Livermore, CA. (email: bendong.tan@uconn.edu, junbo@uconn.edu).

Also, each element in $\mathbf{x}$ is called a feature in this paper. $f(\cdot)$ represents the CPF model. Specifically, by continually increasing the load level in an adaptive step size, CPF involves multiple predictive and corrective directions, leading to CPF calculations. Similar to other work, in this paper, all loads follow the normal distribution while the statistical distribution of wind speed follows the Weibull distribution

$$\varphi(v_w, \lambda, \alpha) = \frac{\alpha}{\lambda} \left(\frac{v_w}{\lambda} \right)^{\alpha-1} e^{-(v_w/\lambda)^\alpha} \quad (2)$$

where λ and α are respectively the shape parameter and scale parameter; v_w is the wind speed. With this uncertain modeling of wind speed, the active power of the wind generator can be determined by the speed-power curve of wind turbine:

$$P_w = \begin{cases} 0, & (\nu_w < \nu_{ci}, \nu_w > \nu_{co}) \\ P_{\text{rated}} \cdot \left(\frac{\nu_w - \nu_{ci}}{\nu_{rd} - \nu_{ci}} \right), & (\nu_{ci} \leq \nu_w \leq \nu_{rd}) \\ P_{\text{rated}}, & (\nu_{rd} < \nu_w < \nu_{co}) \end{cases} \quad (3)$$

where P_w is the active power of the wind generator; P_{rated} is the rated active power; ν_{ci} , ν_{rd} and ν_{co} are the cut-in, rated and cut-out wind speeds, respectively.

Since wind generations and loads are subject to uncertainties, according to $f(\cdot)$, the system load margin also follows a probability distribution. To describe this probability distribution $\mathbf{Y} = \{y^{(1)}, y^{(2)}, \dots, y^{(N)}\}$, MCS performs N times of calculation of $f(\cdot)$ with corresponding N random samples of $\mathbf{x}$, namely $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$. To approximate the true distribution of y accurately, a large N is required, which is time-consuming. To avoid such heavy computational burden, data-driven PLMA uses a reduced-order surrogate model $\mathcal{M}(\cdot)$, i.e., polynomial chaos expansion and GPE, to calculate load margin in a computationally-cheap manner. However, $\mathcal{M}(\cdot)$ is a black-box model, which is difficult to extract explicit relationship between inputs and outputs. Therefore, the goal of this paper is to develop an appropriate scheme to find how the uncertain inputs (wind generations and loads) affect the trend of the load margin under a black-box surrogate model.

III. INTERPRETABLE DEEP KERNEL LEARNING FOR PLMA

This paper develops the probabilistic deep kernel learning enabled surrogate model to extract the hidden relationship between uncertain sources and load margin for PLMA. The interpretability of it is also investigated via the SHAP.

A. Probabilistic Deep Kernel Learning

Given the PLMA dataset $D = \{\mathbf{x}^{(i)}, y^{(i)}\}$, where $\mathbf{x}^{(i)} \in \mathbf{X}$, $y^{(i)} \in \mathbf{Y}$, $i = 1, 2, \dots, N$, the relationship between $\mathbf{x}$ and y can be formulated as a regression problem for load margin prediction. As mentioned before, a surrogate model can be employed to extract this relationship. Since polynomial chaos expansion suffers from curse of dimensionality problem and GPE has limited capability of feature extraction, deep kernel learning (DKL) [10] is employed to obtain the surrogate model. This is because DKL can perform deep feature extraction while the strong capability of Bayesian inference,

as Gaussian process, is maintained to learn the probability distribution of load margin.

Since the Bayesian posterior distribution of an unknown system output can be inferred from a Bayesian prior distribution, we assume the Bayesian prior distribution of load margin with N observations $D = \{\mathbf{X}, \mathbf{Y}\}$ obeys joint multivariate normal distribution:

$$\mathbf{F}(\mathbf{X}) = [f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(N)})]^\top \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}_{\mathbf{X}, \mathbf{X}}) \quad (4)$$

where $\boldsymbol{\mu} = [\mu(\mathbf{x}^{(1)}), \dots, \mu(\mathbf{x}^{(N)})]^\top$ is the mean function and $\mathbf{K}_{\mathbf{X}, \mathbf{X}}$ is the covariance function, which is

$$\mathbf{K}_{\mathbf{X}, \mathbf{X}} = \begin{bmatrix} k(g(\mathbf{x}^{(1)}), g(\mathbf{x}^{(1)})) & \dots & k(g(\mathbf{x}^{(1)}), g(\mathbf{x}^{(N)})) \\ \vdots & \ddots & \vdots \\ k(g(\mathbf{x}^{(N)}), g(\mathbf{x}^{(1)})) & \dots & k(g(\mathbf{x}^{(N)}), g(\mathbf{x}^{(N)})) \end{bmatrix} \quad (5)$$

where $g(\mathbf{x}|\boldsymbol{\omega})$ denotes a neural network, where $\boldsymbol{\omega}$ is its parameter vector. $k(\cdot, \cdot)$ is the kernel function, which is $k(g(\mathbf{x}), g(\mathbf{x}')) = \exp(-\frac{1}{2} \|g(\mathbf{x}) - g(\mathbf{x}')\|^2 / \ell^2)$ in this paper and ℓ is the hyperparameter in the kernel function. Let observations $\mathbf{Y}(\mathbf{X})$ represent the system output $\mathbf{F}(\mathbf{X})$ with the additive Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N)$:

$$\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 \mathbf{I}_N) \quad (6)$$

where σ is the standard deviation and $\mathbf{I}_N$ is a N -dimensional identity matrix.

By repeating the property of the Bayesian prior distribution, the prediction distribution $\mathbf{Y}_*$ at M unseen points indexed by $\mathbf{X}_*$ can be related with N observations:

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{Y}_* | \mathbf{X}_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}(\mathbf{X}) \\ \boldsymbol{\mu}(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\mathbf{X}, \mathbf{X}} & \mathbf{K}_{\mathbf{X}, \mathbf{X}_*} \\ \mathbf{K}_{\mathbf{X}_*, \mathbf{X}} & \mathbf{K}_{\mathbf{X}_*, \mathbf{X}_*} \end{bmatrix} \right) \quad (7)$$

As a result, $\mathbf{Y}_*$ can be formulated as:

$$\mathbf{Y}_* | \mathbf{X}_*, \mathbf{X}, \mathbf{Y}, \ell, \boldsymbol{\omega}, \sigma^2 \sim \mathcal{N}(\mathbb{E}[\mathbf{Y}_*], \text{cov}(\mathbf{Y}_*)) \quad (8)$$

$$\mathbb{E}[\mathbf{Y}_*] = \boldsymbol{\mu}_{\mathbf{X}_*} + \mathbf{K}_{\mathbf{X}_*, \mathbf{X}} [\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 \mathbf{I}_N]^{-1} (\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{X}}) \quad (9)$$

$$\text{cov}(\mathbf{f}_*) = \mathbf{K}_{\mathbf{X}_*, \mathbf{X}_*} - \mathbf{K}_{\mathbf{X}_*, \mathbf{X}} [\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 \mathbf{I}_N]^{-1} \mathbf{K}_{\mathbf{X}, \mathbf{X}_*} \quad (10)$$

To get the optimal hyperparameters $\boldsymbol{\theta} = \{\ell, \boldsymbol{\omega}\}$, we employ stochastic gradient decreasing algorithm, i.e., Adam [12], to maximize the marginal likelihood for target $\mathbf{Y}_*$:

$$\log p(\mathbf{Y}_* | \boldsymbol{\theta}, \mathbf{X}) \propto -\mathbf{Y}_*^\top (\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{Y}_* - \log |\mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 \mathbf{I}_N| \quad (11)$$

Note that, all hyperparameters are inside the kernel function, where the neural network is embedded, and this is the reason that the algorithm is called deep kernel learning. With the kernel learning, the surrogate model $\mathcal{M}(\cdot)$, namely DKL, can be obtained.

B. Interpretable Kernel SHAP

For the PLMA, it is important to indicate how each uncertain source affects the load margin so that TSOs can give appropriate actions to maintain power system stability. However, the surrogate models are in black-box, leading to the difficulties to find hidden mechanism between uncertain sources and load margin. Although some researches can give the sensitivity between them, it is not quite understandable for TSOs. Thus, Kernel SHAP [11] is employed to explicitly extract rules between uncertain sources and load margin, e.g., in which operation interval and how does uncertainties critically influence the trend of load margin?

The main idea of Kernel SHAP is to use an additive model to fit the trained surrogate model:

$$\gamma(\mathbf{x}) = \phi_0 + \sum_{j=1}^n \phi_j \quad (12)$$

where ϕ_0 is the the basic prediction, which is calculated by the expected value of predictions for all samples. ϕ_j is the Shapley value for the j -th feature of $\mathbf{x}$. Thus, the final prediction of any sample is denoted by the sum of all samples' expected prediction and Shapley values for all features of this sample.

To estimated all Shapley values, the following loss function can be built for a certain instance in the samples:

$$L(\mathcal{M}, \beta, \pi_x) = \sum_{\mathbf{z} \in \mathcal{Z}} [\mathcal{M}(h(\mathbf{z})) - \beta(\mathbf{z})]^2 \pi_x(\mathbf{z}) \quad (13)$$

$$\pi_x(\mathbf{z}) = \frac{n-1}{C_n^{|\mathbf{z}|} |\mathbf{z}| (n-|\mathbf{z}|)} \quad (14)$$

$$\beta(\mathbf{z}) = \phi_0 + \sum_{j=1}^n \phi_j z_j \quad (15)$$

where $\mathbf{z}$ is an indicator vector and each element only has value of 0 or 1. For example, for the value of 0 for j -th element z_j , $h(z_j)$ means the j -th feature of $\mathbf{x}$ is randomly replaced by that of another instance, while $h(z_j)$ means the j -th feature of $\mathbf{x}$ is remained with the value of 0 of the j -th element z_j . Besides, $|\mathbf{z}|$ is equal to number of 1 in $\mathbf{z}$ and C is a combination function. (13) can also be solved by the gradient-based optimization algorithm with randomly generated $\mathcal{Z}$, which consists of m indicator vectors.

For a specific sample, the larger the Shapley value of a feature (one of uncertain sources) is, the larger impacts of this feature is on the load margin. Also, a negative Shapley value of a feature means that this feature will reduce load margin while a positive Shapley value of a feature can help increase load margin.

C. Algorithm Implementation

To implement the proposed interpretable PLMA, the offline training and online application are needed as shown in Fig. 1.

1) *Offline training*: To obtain a comprehensive dataset for training DKL, active power of wind generations and the real and reactive power of loads are sampled from their corresponding probabilistic distributions. However, the exact probabilistic

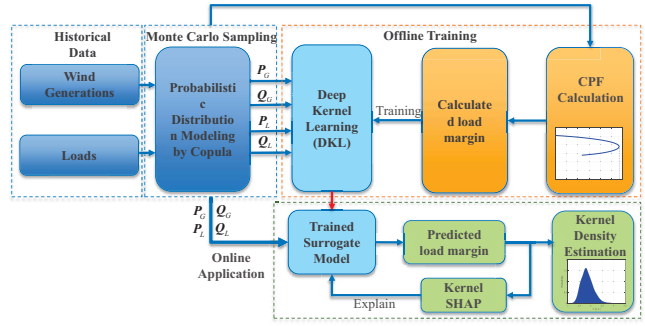


Fig. 1. Framework of the proposed scheme.

distributions of these uncertain sources may be unknown since only historical data is provided. Thus, we employ C-vine Copula [13] to construct the joint probabilistic distributions from the historical data. Based on the inferred joint probabilistic distributions, active power of wind generations and the real and reactive power of loads can be sampled by MCS. As a result, corresponding load margins are calculated by CPF. Thus, DKL can be trained with the dataset, where active power of wind generations, active power of synchronous generations as well as active power and reactive power of loads are as inputs and corresponding load margins are as outputs.

2) *Online application*: Once DKL is trained, wind power and loads can be sampled predict their corresponding load margins. Based on these predictions, the probabilistic distribution of load margin can be established by a non-parametric inference method using a kernel density estimator [14]:

$$\vartheta(y) = \frac{1}{ch} \sum_{i=1}^c \Phi \left(\frac{y - \hat{y}^{(i)}}{h} \right) \quad (16)$$

where c is the sampling size; h is the bandwidth of the estimator and it is generally set as $1.06\sigma c^{-0.2}$ and σ is the estimated sample standard deviation; y_i is the estimated (using surrogate model) or real (using MCS) y for the i -th sample; Φ is the kernel smoothing function, i.e., standard Gaussian kernel utilized in the paper. As a result, proper predictive control schemes can be made based on the probabilistic distribution. After that, comprehensive rules can be extracted by Kernel SHAP, which allows TSOs to determine proper actions to control voltage stability.

IV. NUMERICAL RESULTS

The proposed method is verified on the IEEE 118-bus power system, where synchronous generators at buses 1, 4, 6, 8, 10, 12, 15 and 18 are replaced by eight wind farms with same rated power respectively. The parameter settings for uncertainty sources are as follows [15]:

- The shape and scale parameters of the Weibull distribution are respectively set as $\lambda = 20$ and $\alpha = 12$. The cut-in, rated and cut-out wind speeds are respectively set as $v_{ci} = 3$, $v_{co} = 25$ and $v_{rd} = 12$.

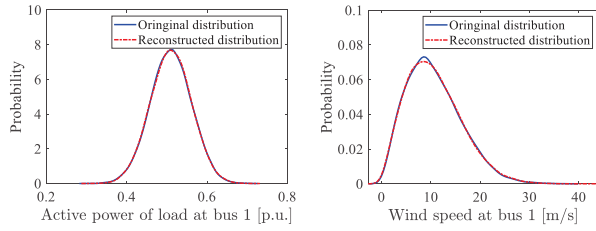


Fig. 2. Probabilistic distribution inference by Copula statistics.

- The mean and standard deviation of loads are assumed as the basic load μ_L and $\sigma_L = 0.1\mu_L$, respectively.

Based on the above settings of uncertainties, database is generated by using MCS according to Section III-C. All simulations are performed on the computer with a CPU of “Intel(R) Core(TM) i5-11500 @ 2.70GHz”.

A. Uncertainty Probabilistic Distribution Modeling

The probability distributions of wind power and loads are unknown and sampling-based approaches, i.e., MCS, are not applicable. Thus, Copula statistics are utilized to infer the joint probabilistic distribution from historical data. To verify its effectiveness, the probability distributions of wind generations and loads are the same as Section II. With 10000 samples, it can be seen that the reconstructed distribution is very close to the original distribution for both wind power and loads, as shown in Fig. 2. This allows generating more representative data for DKL training with obtained Copula statistics rather than explicitly using the specific probabilistic distributions of uncertain sources.

B. Performance Comparisons on PLMA

In this paper, MCS is used as the benchmark to assess the performance of each approach. The shadow learning methods (Support Vector Regression (SVR) and GPE) as well as deep learning methods (Deep Neural Network (DNN)) are selected to compare their PLMA performance with DKL. SVR and GPE are respectively constructed in the scikit-learn, which is a powerful machine learning library in Python. Specifically, regularization parameters and base kernel in SVR are respectively chosen as 10000 and the radial basis function. DNN utilizes a three-layer architecture of 251-100-1, where the activation function is Relu. For DKL, the same neural network architecture as DNN is used in the Gaussian process framework. Both DKL and DNN are built in Tensorflow library of Python.

Fig. 3 shows the variance and mean value curves of load margin with increased samples for MCS method. It can be seen that MCS converges with 10000 samples. This demonstrates that MCS can provide a reliable benchmark though it is a very time-consuming. Note that this takes more than three hours to finish, as shown in Table I. For the data-driven PLMA method, i.e. DKL, the online computing time is less than one second.

From Figs. 4, it can be observed that although the non-parametric method GPE requires the least number of samples,

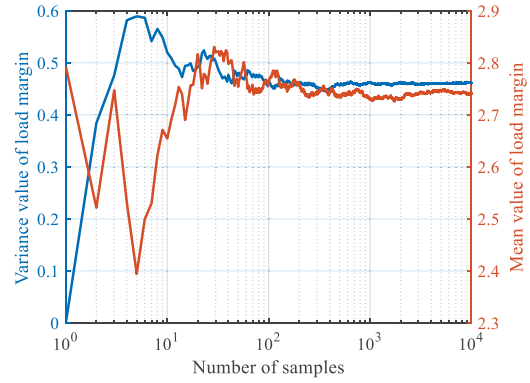


Fig. 3. Convergence curve of the MCS.

TABLE I
PERFORMANCE COMPARISON UNDER VARIOUS METHODS

Model	MCS	DKL	SVR
Computing time	>3 h	<1 s	<1 s
Number of samples	10000	2500	4000
Model	GPE	DNN	
Computing time	<1 s	<1 s	
Number of samples	2000	8000	

its accuracy of is not satisfactory compared with DKL and DNN. DKL and DNN outperform both GPE and SVM regarding to the prediction accuracy. Specifically, GPE and SVM do not work well at the summit and tail of the load margin distribution, where the latter is especially critical for voltage stability analysis. On the other hand, DKL requires much less samples than DNN due to its strong Bayesian inference capability.

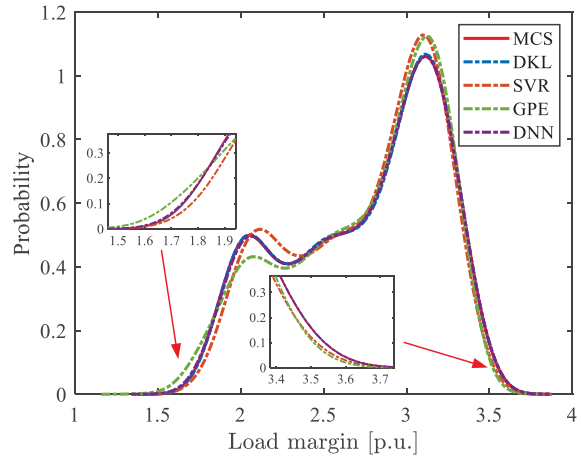


Fig. 4. Load margin distribution predicted by various methods.

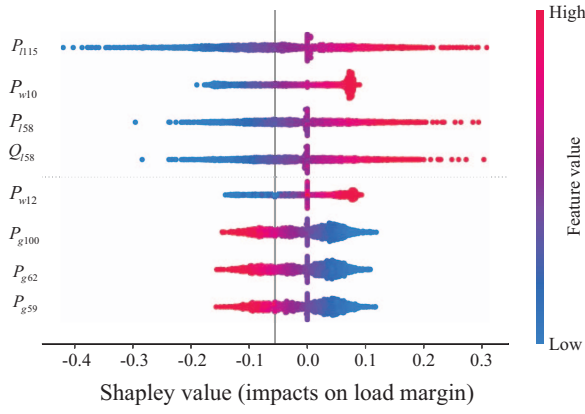


Fig. 5. Shapley values distribution of seven features.

C. Interpretability Analysis

Once DKL is trained, its results can be interpreted by Kernel SHAP as shown in Fig. 5, where seven features with the largest Shapley values are selected to analyze. In Fig. 5, higher degree of the red the feature is, the larger value of the feature is. We can find that increasing the generations at buses 59, 62 and 100 will lead to decrease of load margin while the reactive power of load at bus 58 has the positive impacts since it can provide voltage support. In Fig. 6, it shows that the wind generation at bus 10 has a positive Shapley value. Therefore, to get a large load margin, we can keep the active power at bus 10 larger than 3 p.u. This information is valuable since it can guide TSOs to perform appropriate controls.

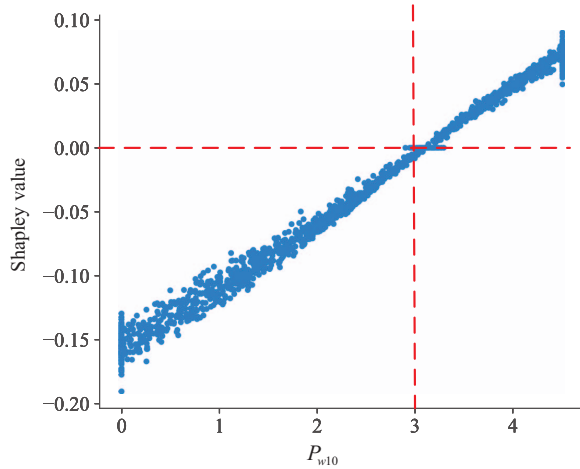


Fig. 6. Shapley values distribution for wind power at bus 10.

V. CONCLUSION

In this paper, an interpretable data-driven PLMA approach is proposed. Specifically, Copula statistics are employed to infer the joint probabilistic distribution of wind generation and

loads from historical data to generate representative data. With these data, DKL is developed to extract the relationship between uncertain sources and load margin. The Kernel SHAP is utilized to interpret the characteristic of the load margin distribution. Numerical results on the IEEE 118-bus power system demonstrate that the proposed method is able to perform accurate PLMA with fewer number of historical data requirements as compared to other deep learning approaches. Thanks to the Shapley values, the proposed approach can also provide the influence of input to the model outputs quantitatively. This provides useful information for system operator for corrective controls for system voltage stability improvement.

REFERENCES

- [1] Kundur, P. and Paserba, J. and Ajarapu *et al.*, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1387-1401, Aug. 2004.
- [2] A. B. Rodrigues, R. B. Prada, and M. D. G. da Silva, "Voltage stability probabilistic assessment in composite systems: Modeling unsolvability and controllability loss," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1575-1588, Aug. 2010.
- [3] H. Chiang, A. J. Flueck, and K. S. Shah *et al.*, "CPFLOW: A practical tool for tracing power system steady-state stationary behavior due to load and generation variations," *IEEE Trans. Power Syst.*, vol. 10, pp. 623-634, May 1995.
- [4] J. Zhao, Y. Bao, and G. Chen, "Probabilistic voltage stability assessment considering stochastic load growth direction and renewable energy generation," in *Proc. IEEE Power Energy Soc. Gen. Meet.*, 2018, pp. 1-5.
- [5] J. Zhang, C. Tse, K. Wang, and C. Chung, "Voltage stability analysis considering the uncertainties of dynamic load parameters," *IET Gener. Transmiss. Distrib.*, vol. 3, no. 10, pp. 941-948, Oct. 2009.
- [6] X. Xu, Z. Yan, M. Shahidehpour, H. Wang, and S. Chen, "Power system voltage stability evaluation considering renewable energy with correlated variabilities," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3236-3245, May 2018.
- [7] Y. Xu, L. Mili and M. Korkali *et al.*, "A Data-Driven Nonparametric Approach for Probabilistic Load-Margin Assessment Considering Wind Power Penetration," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4756-4768, Nov. 2020.
- [8] Y. Xu, K. Karra and L. Mili *et al.*, "Probabilistic Load-Margin Assessment using Vine Copula and Gaussian Process Emulation," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 2020, pp. 1-5.
- [9] X. Wang, X. Wang and H. Sheng *et al.*, "A Data-Driven Sparse Polynomial Chaos Expansion Method to Assess Probabilistic Total Transfer Capability for Power Systems With Renewables," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2573-2583, May 2021.
- [10] A. G. Wilson, Z. Hu, Salakhutdinov and R. Xing *et al.*, "Deep kernel learning," in *Artificial intelligence and statistics*, May. 2016, pp. 370-378.
- [11] S. M. Lundberg, and S. I. Lee, "A unified approach to interpreting model predictions", in *Advances in neural information processing systems*, 2017, pp. 4765-4774.
- [12] D. Kinga and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learning Representations*, 2015, pp. 1-15.
- [13] M. Jan-Frederik and M. Scherer, *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*, 2nd ed. Singapore: World Scientific, 2017.
- [14] R. J. Bessa, V. Miranda and A. Botterud *et al.*, "Time adaptive conditional kernel density estimation for wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 660-669, Oct. 2012.
- [15] S. Xia, X. Luo and K. W. Chan *et al.*, "Probabilistic Transient Stability Constrained Optimal Power Flow for Power Systems With Multiple Correlated Uncertain Wind Generations," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 1133-1144, Jul. 2016.