

Transferable Deep Kernel Emulator for Probabilistic Load Margin Assessment with Topology Changes, Uncertain Renewable Generations and Loads

Bendong Tan, *Student Member, IEEE*, Junbo Zhao, *Senior Member, IEEE*, Le Xie, *Fellow, IEEE*

Abstract—The increasing uncertainties caused by the high-penetration of stochastic renewable generation resources and flexible loads pose challenges to the power system voltage stability. To address this issue, this paper proposes a probabilistic transferable deep kernel emulator (DKE) to extract the hidden relationship between uncertain sources, i.e., wind generations and loads, and load margin for probabilistic load margin assessment (PLMA). This emulator extends the Gaussian process kernel to the deep neural network (DNN) structure and thus gains the advantages of DNN in dealing with high-dimension uncertain inputs and the uncertainty quantification capability of the Gaussian process. A transfer learning framework is also developed to reduce the invariant representation space distance between the old topology and new one. It allows the DKE to be quickly fine tuned with only a few samples under the new topology. Numerical results carried out on the modified IEEE 39-bus and 118-bus power systems demonstrate the strong robustness of the proposed transferable DKE to uncertain wind and load power as well as topology changes while maintaining high accuracy.

Index Terms—Voltage stability, uncertain resources, probabilistic load margin assessment, deep kernel emulator, transfer learning, power system stability.

NOMENCLATURE

A. Load Margin Assessment Model

$P_{G,i}, Q_{G,i}$	Active and reactive power injections at generator bus i .
$P_{L,i}, Q_{L,i}$	Active and reactive power injections at load bus i .
θ, V	Bus voltage angle and magnitude.
μ_L, σ_L	Mean and standard deviation of load distribution φ
a, b	Scale and shape parameters of Weibull distribution ϕ
$v_w, P_w, P_{\text{rated}}$	Wind speed and power, rated wind power.
v_{ci}, v_{rd}, v_{co}	Cut-in, rated and cut-out wind speeds.
$\Lambda(\cdot)$	Parameterized power flow equations.
$V_{\min}, V_{\max}$	Minimum and maximum bus voltage magnitudes.
$P_{G\min}, P_{G\max}$	Minimum and maximum active power outputs of synchronous generators.

$Q_{G\min}, Q_{G\max}$	Minimum and maximum reactive power outputs of synchronous generators.
$\lambda_{\min}$	The minimum load margin among N_c contingencies.
$P_{\max}$	Maximum power transferred in SMLB system.
V_G	Generator bus voltage in SMLB system.
V_L	Load bus voltage in SMLB system.
X_e	Equivalent reactance in SMLB system.
$\mathcal{M}(\cdot)$	Load margin assessment model.
$f(\cdot)$	Surrogate model.
A	Admittance matrix.

B. Deep Kernel Emulator

D	PLMA dataset.
$\mu_X, K_{X,X}, H$	Mean function, covariance function matrix and quadratic basis function.
I_{N_s}	N_s -dimensional identity matrix.
$\mathcal{L}$	Marginal likelihood of DKE.
$\Gamma, \delta, \omega, \sigma^2$	Hyperparameters in DKE.
η	Learning rate.

C. Transfer Learning

D_s, D_t	PLMA dataset in source and target domain.
$G_f(\cdot)$	Feature transformer.
$G_r(\cdot)$	Regressor.
α	Invariant representation.
U_s, U_t	Orthonormal base matrix in source and target domain.
Σ_s, Σ_t	Singular value matrix in source and target domain.
W	Unitary matrix.
ϱ	Principal angles.
B_s, B_t	Weight matrix in source and target domain.
$\mathcal{L}_{\text{tran}}$	Loss function of transfer learning.
N_s	The number of labeled samples in source domain.
N_t	The number of unlabeled samples in target domain.
N'_t	The number of labeled samples in target domain.

D. Uncertainty Quantification

$\vartheta(\cdot)$	Kernel density estimator.
n_e	Sample size for the future generation dispatch interval.

This work is supported by National Science Foundation under ECCS 1917308 and Department of Energy Advanced Grid Modernization program.

B. Tan and J. Zhao are with the Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT. (email: bendong.tan@uconn.edu, junbo@uconn.edu).

L. Xie is with Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, USA (email: le.xie@tamu.edu).

h	Bandwidth of kernel density estimator.
$\mathcal{B}$	Kernel smoothing function.
π_p	Predicted probabilistic distributions with the surrogate model.
π_{MCS}	Predicted probabilistic distributions with MCS.

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 not properly analyzed, inappropriate actions may be taken, leading to power system instability, e.g., static voltage stability. 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 [1]. 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, i.e., probabilistic load margin assessment (PLMA).

To quantify the uncertainties for load margin assessment, several model-based and data-driven PLMA approaches have been proposed. For the model-based methods, the Monte Carlo sampling (MCS)-based one is widely used [2]-[3]. It obtains the load margin distribution via a large number of continuation power flow (CPF) evaluations [4]. 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 [5] has been introduced to reduce the required sample size, the accuracy of this technique is sacrificed. [6] employs the Cumulant method to extract the explicit relationship between loads and load margin by linearizing the power flow equations, but this may induce large bias under stressed system operating conditions, which is the case for voltage stability. Although the two-point estimation approach is utilized in [7] to approximate the statistical moments of load margin without model simplification, its performance deteriorates for large-scale systems. Furthermore, assuming load parameters obey the Gaussian distribution, [8] derives the analytical expression of load margin calculation, but it is not scalable to large-scale systems and the Gaussian assumption is difficult to hold.

Recently, some data-driven load margin assessment methods are proposed to estimate the voltage stability/total transfer capability (TTC) with higher computational efficiency as compared to the model-based solutions. These methods include regression tree [9], local regression [10], random forest [11], support vector regressor (SVR) [12] and deep Neural Network (DNN) [13]. However, the uncertainties from renewable generations or loads are not considered. By modeling the uncertainties from wind generations, [14] proposes a probabilistic TTC assessment approach based on the online measurements but it can't quantify the uncertainties from wind generations and loads. For the fast global sensitivity analysis, [15] employs polynomial chaos expansion (PCE) to be the surrogate model for load margin calculation but it is subject to the

curse of dimensionality issue. Although the extreme learning machine (ELM) used in [16] tackles that issue, it requires much more training samples and is not robust to topology changes. Combined with uncertainty distribution inference, [17]-[18] further develop a nonparametric and reduced-order approximation method, Gaussian process emulator (GPE), for the nonlinear CPF model. In [19], by merging the discrete probabilistic variables, i.e., line outages, into sparse polynomial chaos expansion (SPCE), uncertainty propagation through CPF can be analyzed under disturbances, but the surrogate model learning will be intractable for larger systems as the number of predefined line outages is large. On the other hand, when the system topology changes, how to timely update the surrogate model for PLMA has not been investigated.

This paper proposes a computationally efficient PLMA that is robust to N-1 contingencies and topology changes. The main contributions are summarized as follows:

- A probabilistic deep kernel emulator (DKE) that extends the Gaussian process kernel to the deep neural network (DNN) structure is developed to extract the relationship between uncertain sources and load margin. The DNN structure allows dealing with high-dimension uncertain inputs and complicated nonlinear mapping relationship while the Gaussian process enables natural uncertainty quantification capability for PLMA. The proposed DKE requires less numbers of samples as compared to other DNN approaches while achieving similar accuracy. Thanks to the strong learning capability via DNN structure, all layers of DNN used in the proposed DKE serve to distill and learn increasingly complicated features with the exception of the output layer. Therefore, DKE has much better performance than the original Gaussian process method.
- An efficient transfer learning approach is developed to allow quick adaptation of the proposed DKE for different topology changes. The key idea is to minimize the invariant representation space distance between old topology and new one with a few new samples. To the best knowledge of us, this is the first time the topology changes for PLMA task have been considered as compared to existing data-driven approaches.

The remaining of this paper is organized as follows. Section II shows the problem statement. The DKE and its enhanced version with transfer learning for solving the CPF is developed in Section III. Section IV shows and analyzes simulation results on the modified 39-bus and 118-bus systems and finally Section V concludes the paper.

II. PROBLEM STATEMENT

Given a power system with N_b buses, the power flow equations are as follows:

$$\begin{cases} P_{G,i} - P_{L,i} - P_i(\mathbf{z}) = 0 \\ Q_{G,i} - Q_{L,i} - Q_i(\mathbf{z}) = 0 \end{cases} \quad (1)$$

where $i = 1, \dots, N_b$ and $\mathbf{z} = [\boldsymbol{\theta} \ \mathbf{V}]^T$; $\boldsymbol{\theta}$ and $\mathbf{V}$ are respectively the bus voltage magnitudes and angles; $P_{G,i}$ and $Q_{G,i}$ are the active and reactive power injections at generator bus i ; $P_{L,i}$ and $Q_{L,i}$ are the active and reactive power injections

at load bus i . Defining respectively the incremental power $\Delta P_{G,i}$, $\Delta P_{L,i}$ and $\Delta Q_{L,i}$ for generations and loads, the load margin can be obtained via the CPF approach:

$$\mathbf{A}(\mathbf{z}, \lambda) = \begin{bmatrix} \tilde{P}_{G,i} - \tilde{P}_{L,i} - P_i(\mathbf{z}) \\ Q_{G,i} - \tilde{Q}_{L,i} - Q_i(\mathbf{z}) \end{bmatrix} = \mathbf{0} \quad (2)$$

$$\begin{cases} P_{G,i} + \lambda \Delta P_{G,i} = \tilde{P}_{G,i} \\ P_{L,i} + \lambda \Delta P_{L,i} = \tilde{P}_{L,i} \\ Q_{L,i} + \lambda \Delta Q_{L,i} = \tilde{Q}_{L,i} \end{cases} \quad (3)$$

where the load margin is the maximum value of λ without violating physical constraints, i.e., voltage constraints. $\mathbf{A}(\cdot)$ represents the parameterized power flow equations in the CPF model. By continually increasing the loading level, CPF involves multiple predictive and corrective directions to obtain the load margin. Note that only the active power of generators is parameterized and this should be for the generators that cannot control reactive power.

A. Uncertainty Modeling

Due to the integration of uncertain renewable energy and stochastic loads, the CPF results are no longer deterministic and the PLMA is needed. To this end, the probabilistic characteristics of uncertain sources, such as wind generations and loads, are required.

1) *Load uncertainty modeling*: following existing literature, the loads are assumed to obey the Gaussian distribution [20], whose probability density is

$$\varphi(P_L) = \frac{1}{\sqrt{2\pi}\sigma_L} e^{-(P_L - \mu_L)^2 / 2\sigma_L^2} \quad (4)$$

where μ_L and σ_L are respectively the mean and standard deviations of P_L .

2) *Wind generation uncertainty modeling*: the wind speed follows the Weibull distribution [21]:

$$\phi(\nu_w, a, b) = \frac{b}{a} \left(\frac{\nu_w}{a} \right)^{b-1} e^{-(\nu_w/a)^b} \quad (5)$$

where a and b are respectively the scale and shape parameters; ν_w is the wind speed. After that, the output of the wind generator can be determined by the speed-power curve of wind turbine [22]:

$$P_w = \begin{cases} 0 & (\nu_w < \nu_{ci}, \nu_w > \nu_{co}) \\ P_{\text{rated}} \cdot \left(\frac{\nu_w - \nu_{ci}}{\nu_{rd} - \nu_{ci}} \right) & (\nu_{ci} \leq \nu_w \leq \nu_{rd}) \\ P_{\text{rated}} & (\nu_{rd} < \nu_w < \nu_{co}) \end{cases} \quad (6)$$

where P_w is the output of the wind generator; P_{rated} is the rated active power of the wind generator; ν_{ci} , ν_{rd} and ν_{co} are the cut-in, rated and cut-out wind speeds, respectively. The wind speed is sampled via probabilistic distribution (5) and then wind power can be obtained via (6). It should be noted that only wind generation is considered here, the approach can also deal with solar generation uncertainty when the proper probability distribution model for solar energy is used.

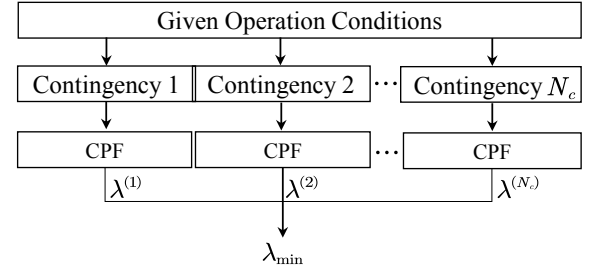


Fig. 1. Load margin under N-1 contingencies.

B. Probabilistic Load Margin under N-1 Contingencies

Since all loads and wind power are assumed to contain uncertainties, deterministic parameterized power flow equations $\mathbf{A}(\mathbf{z}, \lambda)$ needs to be reformulated to be a probabilistic ones, $\mathbf{A}(\mathbf{z}, \mathbf{P}_w, \mathbf{P}_L, \lambda)$. Assume that there are n_w wind generators connected to the system, we have $\mathbf{P}_w = [P_{G, N_b - n_w + 1}, \dots, P_{G, N_b}]^T$. Given N_c predefined contingencies, the objective function of the PLMA model under each scenario is to maximize the load margin $\lambda^{(m)}$, i.e., the maximum power that can be supplied without violating operational constraints, which can be written as [19]:

$$\max \lambda^{(m)} \quad (7)$$

subjected to the following constraints:

$$\mathbf{A}(\mathbf{z}, \mathbf{P}_w, \mathbf{P}_L, \lambda^{(m)}) = \mathbf{0} \quad (8)$$

$$\mathbf{V}_{\min} \leq \mathbf{V}(\mathbf{z}, \mathbf{P}_w, \mathbf{P}_L, \lambda^{(m)}) \leq \mathbf{V}_{\max} \quad (9)$$

$$\mathbf{P}_{G\min} \leq \mathbf{P}_G(\mathbf{z}, \mathbf{P}_w, \mathbf{P}_L, \lambda^{(m)}) \leq \mathbf{P}_{G\max} \quad (10)$$

$$\mathbf{Q}_{G\min} \leq \mathbf{Q}_G(\mathbf{z}, \mathbf{P}_w, \mathbf{P}_L, \lambda^{(m)}) \leq \mathbf{Q}_{G\max} \quad (11)$$

where $m = 1, 2, \dots, N_c$; equations (9)-(11) respectively denote the voltage, generation active power and reactive power constraints; $\mathbf{V}_{\min}$ and $\mathbf{V}_{\max}$ are respectively the minimum and maximum bus voltage magnitudes; $\mathbf{P}_{G\min}$ and $\mathbf{P}_{G\max}$ are respectively the minimum and maximum active power outputs of synchronous generators; $\mathbf{Q}_{G\min}$ and $\mathbf{Q}_{G\max}$ are respectively the minimum and maximum reactive power outputs of synchronous generators.

To determine whether the power system is operated under secure status or not, the minimum load margin among N_c contingencies is taken as the index:

$$\lambda_{\min} = \min\{1, \dots, \lambda^{(m)}, \dots, \lambda^{(N_c)}\} \quad (12)$$

To better illustrate the impacts of topology change to load margin, the single machine load bus (SMLB) system is used,

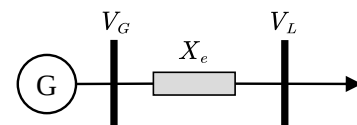


Fig. 2. The single machine load bus system.

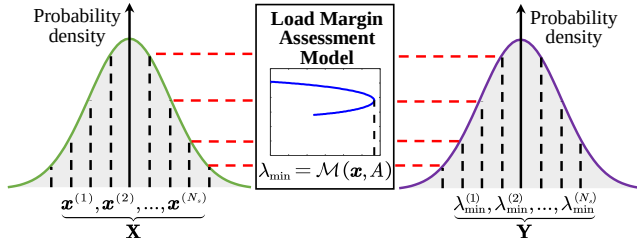


Fig. 3. Uncertainty propagation from uncertain sources to load margin.

see Fig. 2, where the maximum power transferred (load margin) $P_{\max}$ can be calculated via [23]:

$$P_{\max} = \frac{V_G V_L}{X_e} \quad (13)$$

where V_G is the generator bus voltage and V_L is the load bus voltage, and V_G and V_L are constant. X_e is the equivalent reactance between V_G and V_L . In this system, the admittance matrix $\mathbf{A}$ is a two-order matrix calculated based on X_e . Assume X_e is the equivalent reactance of multiple lines, if a line is added or tripped, X_e will decrease or increase, leading to the increase or decrease of load margin. In a large power system, $\mathbf{A}$ plays a similar role as $1/X_e$ to calculate load margin. Since topology of power system determines the admittance matrix $\mathbf{A}$ in the CPF calculation, topology changes can be reflected in $\mathbf{A}$ and therefore affect the load margin. According to (7)-(12), the load margin assessment model, namely the relationship between load margin and $\mathbf{A}$, uncertain input $\mathbf{x} = [P_w \ P_L]^T$, can be defined as:

$$\lambda_{\min} = \mathcal{M}(\mathbf{x}, \mathbf{A}) \quad (14)$$

Since wind generations and loads are subject to uncertainties, $\lambda_{\min}$ also follows a probability distribution with uncertainty propagated from $\mathcal{M}(\cdot)$, see Fig. 3. To describe this probability distribution $\mathbf{Y} = \{\lambda_{\min}^{(1)}, \lambda_{\min}^{(2)}, \dots, \lambda_{\min}^{(N_s)}\}$, MCS performs N_s times of CPF calculations with corresponding N_s random samples of $\mathbf{x}$, namely $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N_s)}\}$. Consequently, PLMA model is

$$\mathbf{Y} = \mathcal{M}(\mathbf{X}, \mathbf{A}) \quad (15)$$

(15) can be equivalently written as $\mathbf{Y} = \mathcal{M}(\mathbf{X})$, if the topology is not changed.

To approximate the true distribution of load margin accurately, a large N_s is required, which is time-consuming. To avoid such heavy computational burden, data-driven PLMA uses a reduced-order surrogate model $f(\cdot)$, i.e., PCE and GPE, to substitute $\mathcal{M}(\cdot)$ to calculate load margin in a computationally-cheap manner. However, they are not scalable to large-scale systems with high dimensional uncertain inputs.

Furthermore, the system is subject to contingencies or topology changes and the constructed surrogate model will yield large PLMA errors. For transmission systems, the topology change does not occur often and even with topology changes, only small part of elements in $\mathbf{A}$ will be changed (i.e., those corresponding to the changed topology). Consequently, the main parts of $\mathbf{A}$ before and after topology change are the same.

In other words, the shared features for different topologies have large similarities. By utilizing these shared features, this paper develops a transferable DKE to address these issues.

III. PROPOSED TRANSFERABLE DEEP KERNEL EMULATOR

In this section, the DKE will be first developed for PLMA. Next, DKE will be enhanced via the transfer learning to allow strong adaptations to contingencies and topology changes.

A. Deep Kernel Emulator for PLMA

Given the PLMA dataset $\mathbf{D} = \{\mathbf{x}^{(i)}, \lambda_{\min}^{(i)}\}$, where $\mathbf{x}^{(i)} \in \mathbf{X}$, $\lambda_{\min}^{(i)} \in \mathbf{Y}$, $i = 1, 2, \dots, N_s$, the hidden relationship between uncertain sources and the load margin, can be extracted by a surrogate model $\lambda_{\min} = f(\mathbf{x})$. Since PCE and GPE are subject to curse of dimensionality issue, the DKE is proposed that extends the GPE kernel into the DNN structure, resulting in both strong capability of handling high-dimensional uncertain inputs and uncertainty quantification.

Similar to the Bayesian posterior distribution inference in Gaussian process, DKE also assumes N_s realizations $\{f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \dots, f(\mathbf{x}^{(N_s)})\}$ of $f(\mathbf{x})$ follow the joint multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \mathbf{K}_{\mathbf{X}, \mathbf{X}})$, which is regarded as the Bayesian prior distribution:

$$\mathbf{F}(\mathbf{X}) = [f(\mathbf{x}^{(1)}), \dots, f(\mathbf{x}^{(N_s)})]^T \sim \mathcal{N}_f(\boldsymbol{\mu}_{\mathbf{X}}, \mathbf{K}_{\mathbf{X}, \mathbf{X}}) \quad (16)$$

where $\boldsymbol{\mu}_{\mathbf{X}} = [\mu(\mathbf{x}^{(1)}), \dots, \mu(\mathbf{x}^{(N_s)})]^T$ is the mean function and $\mu(\mathbf{x}) = \mathbf{H}(\mathbf{x})\boldsymbol{\Gamma}$. To provide the prior information to approximate the nonlinear CPF model, $\mathbf{H}(\mathbf{x})$ is defined as the quadratic basis function [24]:

$$\mathbf{H}(\mathbf{x}) = [1, x_1, \dots, x_d, x_1^2, \dots, x_d^2] \quad (17)$$

d is the dimension of $\mathbf{x}$; $\mathbf{K}_{\mathbf{X}, \mathbf{X}}$ is the covariance function matrix, i.e.,

$$\begin{bmatrix} k(g(\mathbf{x}^{(1)}), g(\mathbf{x}^{(1)})) & \dots & k(g(\mathbf{x}^{(1)}), g(\mathbf{x}^{(N_s)})) \\ \vdots & \ddots & \vdots \\ k(g(\mathbf{x}^{(N_s)}), g(\mathbf{x}^{(1)})) & \dots & k(g(\mathbf{x}^{(N_s)}), g(\mathbf{x}^{(N_s)})) \end{bmatrix} \quad (18)$$

where $g(\mathbf{x}|\boldsymbol{\omega})$ denotes a neural network and its parameter vector $\boldsymbol{\omega}$. This is the key component in the DKE since $g(\mathbf{x}|\boldsymbol{\omega})$ plays a critical role in feature extraction for providing powerful generalization for DKE. $k(\cdot, \cdot)$ is the kernel function, i.e., radial basis function (RBF) kernel $k(g(\mathbf{x}), g(\mathbf{x}')) = \exp(-\frac{1}{2} \|g(\mathbf{x}) - g(\mathbf{x}')\|^2 / \ell^2)$ and ℓ is the corresponding hyperparameter. Without loss of generality, the hyperparameters in the kernel function are represented as $\boldsymbol{\delta}$ in this paper.

Let observations $\mathbf{Y}$ represent the system output $\mathbf{F}(\mathbf{X})$ with the additive Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{N_s})$, we have

$$\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}_{\mathbf{X}, \mathbf{X}} + \sigma^2 \mathbf{I}_{N_s}) \quad (19)$$

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

In case of unseen samples, i.e., during prediction stage, it is assumed that the predictions and the existing observations follows the joint multivariate Gaussian distribution. Specifically,

the prediction distribution $\mathbf{Y}_*$ at M unseen points indexed by $\mathbf{X}_*$ can be related with N_s existing observations:

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{Y}_* | \mathbf{X}_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_{X_*} \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{X,X} & \mathbf{K}_{X,X_*} \\ \mathbf{K}_{X_*,X} & \mathbf{K}_{X_*,X_*} \end{bmatrix} \right) \quad (20)$$

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

$$\mathbf{Y}_* | \{\mathbf{X}_*, \mathbf{X}, \mathbf{Y}, \boldsymbol{\delta}, \boldsymbol{\omega}, \sigma^2\} \sim \mathcal{N}(\mathbb{E}[\mathbf{Y}_*], \text{cov}(\mathbf{Y}_*)) \quad (21)$$

$$\mathbb{E}[\mathbf{Y}_*] = \boldsymbol{\mu}_{X_*} + \mathbf{K}_{X_*,X} [\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s}]^{-1} (\mathbf{Y} - \boldsymbol{\mu}_X) \quad (22)$$

$$\text{cov}(\mathbf{Y}_*) = \mathbf{K}_{X_*,X_*} - \mathbf{K}_{X_*,X} [\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s}]^{-1} \mathbf{K}_{X,X_*} \quad (23)$$

In this paper, the expectations in (22) are utilized as the predictions while (23) is employed to quantify the prediction uncertainties. Note that (20) is formulated based on the Gaussian process, but this doesn't mean DKE is not suitable for data sampled from non-Gaussian distribution. According to [25], a finite number of observations can be covered by the Gaussian process and thus DKE can be applied to data from any distribution.

Optimization of Hyperparameters: all the hyperparameters $\boldsymbol{\Psi} = \{\boldsymbol{\Gamma}, \boldsymbol{\delta}, \boldsymbol{\omega}, \sigma^2\}$ for DKE should be tuned properly. We employ stochastic gradient decreasing algorithm to maximize the marginal likelihood for target $\mathbf{Y}$:

$$\begin{aligned} \mathcal{L} &= \log p(\mathbf{Y} | \boldsymbol{\Psi}, \mathbf{X}) \\ &\propto -(\mathbf{Y} - \boldsymbol{\mu}_X)^\top (\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s})^{-1} (\mathbf{Y} - \boldsymbol{\mu}_X) \\ &\quad - \log |\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s}| \\ &= -[\mathbf{Y} - \mathbf{H}(\mathbf{X})\boldsymbol{\Gamma}]^\top (\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s})^{-1} [\mathbf{Y} - \mathbf{H}(\mathbf{X})\boldsymbol{\Gamma}] \\ &\quad - \log |\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s}| \end{aligned} \quad (24)$$

In this paper, the hyperparameters in mean function, kernel function and neural network are jointly learnt according to the their corresponding gradients, which can be derived by the following chain rules:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\delta}} = \frac{\partial \mathcal{L}}{\partial \mathbf{K}_{X,X}} \frac{\partial \mathbf{K}_{X,X}}{\partial \boldsymbol{\delta}} \quad (25)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\omega}} = \frac{\partial \mathcal{L}}{\partial \mathbf{K}_{X,X}} \frac{\partial \mathbf{K}_{X,X}}{\partial g(\mathbf{x}|\boldsymbol{\omega})} \frac{\partial g(\mathbf{x}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \quad (26)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\Gamma}} = -2[\mathbf{Y} - \mathbf{H}(\mathbf{X})\boldsymbol{\Gamma}]^\top (\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s})^{-1} \mathbf{H}(\mathbf{X}) \quad (27)$$

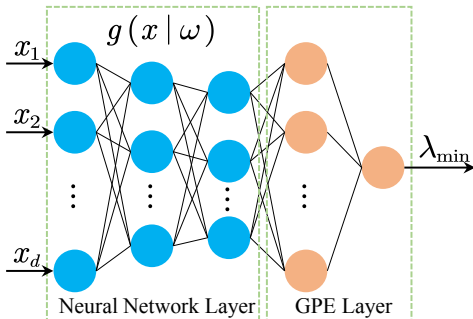


Fig. 4. Proposed deep kernel emulator for PLMA.

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = \text{trace} \left((\mathbf{K}_{X,X} + \sigma^2 \mathbf{I}_{N_s})^{-1} \right) \quad (28)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{K}_{X,X}} = \frac{1}{2} \left(\mathbf{K}_{X,X}^{-1} \mathbf{Y} \mathbf{Y}^\top \mathbf{K}_{X,X}^{-1} - \mathbf{K}_{X,X}^{-1} \right) \quad (29)$$

Thanks to the auto-gradient function in the Pytorch [26], $\frac{\partial \mathbf{K}_{X,X}}{\partial \boldsymbol{\delta}}$ and $\frac{\partial g(\mathbf{x}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}$ can be automatically calculated. Consequently, all parameters are updated according to the gradient descent algorithm:

$$\begin{cases} \boldsymbol{\delta} = \boldsymbol{\delta} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{\delta}} \\ \boldsymbol{\omega} = \boldsymbol{\omega} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{\omega}} \\ \boldsymbol{\Gamma} = \boldsymbol{\Gamma} - \eta \frac{\partial \mathcal{L}}{\partial \boldsymbol{\Gamma}} \\ \sigma^2 = \sigma^2 - \eta \frac{\partial \mathcal{L}}{\partial \sigma^2} \end{cases} \quad (30)$$

where η is the learning rate. With the kernel and mean function learning, the surrogate model $f(\cdot)$, namely DKE, can be obtained, as shown in Fig. 4. The key idea here is to leverage the neural network for extracting the high-dimensional data features and use them as prior information for GPE. After feature extractions, the input dimensions for GPE has been significantly reduced, yielding improved computational efficiency. Note that for GPE, a proper prior information is essential to achieve good performance and this is addressed in the proposed DKE method.

B. Adaptiveness Enhancement of DKE via Transfer Learning

Although the DKE-based surrogate model can achieve satisfactory performance for PLMA, it is vulnerable to topology changes, i.e., line or generator outages or switching due to maintenance or dispatching requirement. This is because the hidden relationship between uncertain sources and load margin is different from the one before topology changes, yielding different load margin distributions. Consequently, the surrogate model needs to be adaptive to new topology. This paper develops the transfer learning-based approach to deal with that.

1) *Domain Adaptation Regression:* It is worth pointing out that the load margin distributions are not completely different under two topology since the voltage stability involves extensively the local phenomena. This means that the surrogate models under two topologies share some common domain knowledge, a kind of invariant representation. If the shared domain knowledge can be utilized, the amount of samples to obtain the new surrogate model can be drastically reduced for PLMA.

Define the domain knowledge before topology changes as the source domain $\mathbf{D}_s = \{\mathbf{x}_s^{(i)}, \lambda_{\min,s}^{(i)}\}$ while that after topology changes is called the target domain $\mathbf{D}_t = \{\mathbf{x}_t^{(j)}, \lambda_{\min,t}^{(j)}\}$, where $\mathbf{x}_s^{(i)} \in \mathbf{X}_s$, $\lambda_{\min,s}^{(i)} \in \mathbf{Y}_s$, $\mathbf{x}_t^{(j)} \in \mathbf{X}_t$, $\lambda_{\min,t}^{(j)} \in \mathbf{Y}_t$, $i = 1, 2, \dots, N_s$ and $j = 1, 2, \dots, N_t$. The subscript “s” means the variables in the source domain, and subscript “t” means the variables in the target domain. To lower the generalization error of PLMA in the target source, learning transferable representations by minimizing the domain shift

between invariant representations of source and target domains is the core idea of the transfer learning in this paper. Since PLMA is essentially a regression task, it can be seen as an implementation of domain adaptation regression.

Inspired by [27], the invariant representation α for two domains are extracted by a feature transformer $G_f(\cdot)$, namely $\alpha = G_f(x)$. Mathematically, the invariant representation is a transformation of inputs, i.e., the uncertain sources. To measure the representation space distance between the source domain and the target domain, one way is to calculate the similarity of the orthonormal bases from two space globally. Therefore, the singular value decomposition (SVD) is used to obtain orthonormal bases that form the two representation spaces. Formally, we have

$$\Upsilon_s = \mathbf{U}_s \Sigma_s (\mathbf{W}_s)^\top, \quad \Upsilon_t = \mathbf{U}_t \Sigma_t (\mathbf{W}_t)^\top \quad (31)$$

where $\Upsilon = [\alpha^{(1)}, \dots, \alpha^{(N_t)}]$; $\mathbf{U} = [\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(N_t)}]$ denotes the orthonormal base matrix; Σ is the singular value matrix; $\mathbf{W}$ is the unitary matrix.

The similarity of orthonormal bases from two space is measured by the principal angles $\varrho = [\varrho^{(1)}, \dots, \varrho^{(N_t)}]$, i.e.,

$$\left\{ \begin{array}{l} \varrho^{(1)} = \min_{\mathbf{u}_s^{(1)}, \mathbf{u}_t^{(1)}} \arccos \left(\frac{(\mathbf{u}_s^{(1)})^\top \mathbf{u}_t^{(1)}}{\|\mathbf{u}_s^{(1)}\| \|\mathbf{u}_t^{(1)}\|} \right) \\ \varrho^{(2)} = \min_{\mathbf{u}_s^{(2)}, \mathbf{u}_t^{(2)}} \arccos \left(\frac{(\mathbf{u}_s^{(2)})^\top \mathbf{u}_t^{(2)}}{\|\mathbf{u}_s^{(2)}\| \|\mathbf{u}_t^{(2)}\|} \right) \\ \vdots \\ \varrho^{(N_t)} = \min_{\mathbf{u}_s^{(N_t)}, \mathbf{u}_t^{(N_t)}} \arccos \left(\frac{(\mathbf{u}_s^{(N_t)})^\top \mathbf{u}_t^{(N_t)}}{\|\mathbf{u}_s^{(N_t)}\| \|\mathbf{u}_t^{(N_t)}\|} \right) \end{array} \right. \quad (32)$$

(32) can be further simplified as:

$$(\mathbf{U}_s)^\top \mathbf{U}_t = \mathbf{B}_s (\text{diag}(\cos \varrho)) (\mathbf{B}_t)^\top \quad (33)$$

where $\mathbf{B}$ is the weight matrix. To minimize the representation space distance between the source domain and the target domain, an end-to-end optimization scheme based on neural network is developed in this paper, see Fig. 5. It consists of two main parts: the feature transformer $G_f(\cdot)$ and the regressor $G_r(\cdot)$. Specifically, the loss function is defined as:

$$\mathcal{L}_{tran} = \gamma \mathcal{L}_1 + \beta \mathcal{L}_2 + \mathcal{L}_3 \quad (34)$$

where $\mathcal{L}_1 = \|\sin \varrho\|^2$; $\mathcal{L}_2 = \|\mathbf{B}_s \odot \mathbf{B}_s - \mathbf{B}_t \odot \mathbf{B}_t\|^2$; $\mathcal{L}_3 = \|G_r(G_f(\mathbf{X}_s)) - \mathbf{Y}_s\|^2$; γ and β are respectively the coefficients for the corresponding loss terms; $\mathcal{L}_1$ is utilized to reflect the representation space distance between the source domain and the target domain via principal angles; $\mathcal{L}_2$ pushes the feature transformer to match orthonormal bases in source domain and target domain with similar weights; $\mathcal{L}_3$ denotes the regression loss in the source domain so that the transferable representations can be extracted for PLMA.

In this paper, the Adam optimizer [28] is advocated to minimize $\mathcal{L}_{tran}$. **Note that no labels are required for the target source and we can transfer the knowledge under old**

topology to new one quickly without the time-consuming labeling process. This also justifies why only a few samples are needed to quickly update the surrogate model.

2) *Algorithm Implementation for PLMA:* Since uncertain sources are wind power and loads, it is necessary to infer their probabilistic distributions from historical data. In this paper, the Copula statistics is employed to infer the distributions of wind speed and loads. This allows generating wind power and loads data based on the identified Copula structure. More details about Copula can refer to [29].

Assuming a surrogate model DKE has been trained in the old topology with datasets $\mathbf{D}_s = \{\mathbf{X}_s, \mathbf{Y}_s\}$ generated from the identified Copula structure, a three-stage framework is proposed to perform transfer learning for PLMA.

(1) *Stage one: uncertainty modeling* Similar to uncertainty modeling in the old topology, the uncertain inputs $\mathbf{X}_t$ can also be generated based on the identified Copula structure, resulting in the probabilistic distribution modeling for wind power and loads under new topology.

(2) *Stage two: transfer learning* By passing $\mathbf{X}_t$ and $\mathbf{X}_s$ through $G_f(\cdot)$, the invariant representations for old topology and new topology, Υ_s and Υ_t , are obtained. According to (31)-(33), they are further transformed by two SVD operations to calculate the loss of (34). By minimizing (34), the training of $G_f(\cdot)$ and $G_r(\cdot)$ can be performed with $\mathbf{D}_s$ and $\mathbf{X}_t$. Consequently, the invariant representation space distance under two topology can be reduced and thus the knowledge on the old topology is transferred to the new one.

(3) *Stage three: fine tuning* Once the training in the stage two is completed, all other layers except the final one in $G_r(\cdot)$ and $G_f(\cdot)$ can be merged into the GPE framework. As a result, DKE model for the new topology is developed. Since the transfer learning in the stage two has provided good priori information for the DKE model, much less amount of samples is required as compared to a raw DKE model, which will be demonstrated in the numerical results section. Therefore, N'_t uncertain inputs $\tilde{\mathbf{X}}_t$ (its invariant representation is $\tilde{\Upsilon}_t$), can be sampled from $\mathbf{X}_t$ and their corresponding load margin $\tilde{\lambda}_{\min,t}$ can be calculated by CPF. With these small amount of samples under new topology, the constructed DKE model is fine-tuned via (24)-(30) and thus quickly updated under new topology for PLMA.

Based on the fine-tuned DKE model $f(\cdot)$, the predicted probabilistic distribution of $\tilde{\lambda}'_{\min,t}$ corresponding to $\mathbf{X}_t$ can be established by a non-parametric inference method using a kernel density estimator [30]:

$$\vartheta(\tilde{\lambda}'_{\min,t}) = \frac{1}{n_e h} \sum_{i=1}^{n_e} \mathcal{B} \left(\frac{\tilde{\lambda}'_{\min,t} - \tilde{\lambda}'_{\min,t}^{(i)}}{h} \right) \quad (35)$$

where n_e is the sample size for the future generation dispatch interval; h is the bandwidth of the estimator and it is generally set as $1.06\tau n_e^{-0.2}$, where τ is the estimated sample standard deviation; $\tilde{\lambda}'_{\min,t}^{(i)}$ is the predicted (using the fine-tuned surrogate model) $\lambda'_{\min,t}$ for the i -th sample; $\mathcal{B}$ is the kernel smoothing function, i.e., standard Gaussian kernel in this paper. Note that, other methods, i.e., MCS and other surrogate

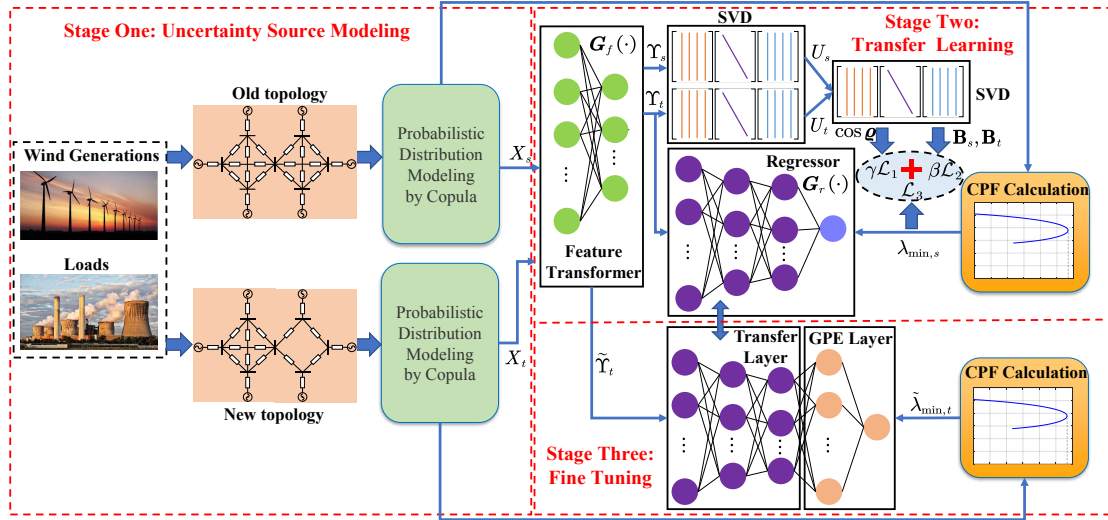


Fig. 5. The proposed domain knowledge adaptation-based transfer learning method for PLMA.

models, can also utilize this kernel density estimator to obtain their corresponding probabilistic load margin distributions.

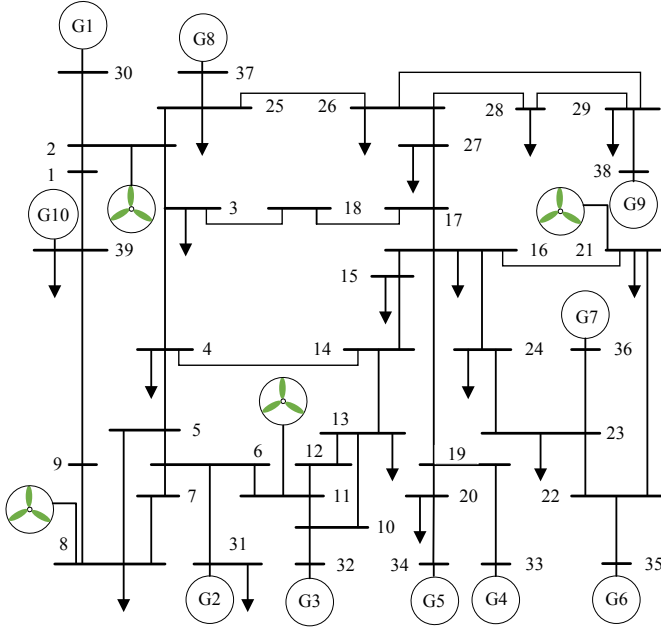


Fig. 6. The diagram of IEEE 39-bus power system.

IV. NUMERICAL RESULTS

The effectiveness of the proposed method is first demonstrated on the modified IEEE 39-bus power system, see Fig. 6, where four wind farms are connected to buses 2, 8, 11 and 21 respectively. The generation capacities of these four wind farms are 350 MW, 150 MW, 400MW and 500 MW, respectively. The random parameter settings for the uncertainty sources are as follows [31]:

- The shape parameter and scale parameter of the Weibull distribution are respectively set as $a = 20$ and $b = 2$. Besides, the cut-in, rated and cut-out wind speeds are

respectively set as $\nu_{ci} = 3$ m/s, $\nu_{co} = 25$ m/s and $\nu_{rd} = 12$ m/s.

- The mean μ_L and standard deviation σ_L of all loads are assumed as original load values in IEEE 39-bus system and $\sigma_L = 0.1\mu_L$ p.u. respectively.

As a result, 25-dimension uncertain inputs (4 wind generators and 21 loads) are constructed. All surrogate models and the transfer learning framework are built on a Python library, Pytorch [32], and all calculations are performed on a computer with a 3.2-GHz Intel Core i9-12900KF CPU, NVIDIA GeForce RTX 3090 GPU and 128 GB RAM. The load margin under contingencies are calculated via the built-in CPF module of MATPOWER 7.1 [33].

Based on the above uncertainty settings, database is generated by using MCS. The number of samples used to approximate the true load margin distribution depends on the scale of the power system. To determine this amount, the variance and mean value of λ_{min} are calculated during Monte Carlo sampling. From Fig. 7, it can be seen that MCS converges when the number of samples is about 10,000, which is consistent with the conclusion in [34]. Although MCS can approximate the true PLMA, it is very time-consuming and thus is treated as the benchmark in this paper to verify the performance of all other methods.

A. Performance Comparisons without Topology Changes

To fully demonstrate the performance of the proposed method when there is no topology change, several state-of-the-art methods are compared, including the shadow learning methods (SVR [35], GPE and PCE) as well as DNN [36]). SVR is constructed in the scikit-learn [37], which is a powerful machine learning library in Python. Its regularization parameter and base kernel in SVR are respectively chosen as 100 and the radial basis function; PCE is built on the UQ-LAB platform [38], where 2nd-order truncated PCE functions are selected; both GPE and DNN are implemented on Pytorch, while the basis function of the former is RBF and the number of neural neurons in each layer of DNN is “25-50-20-1”. In

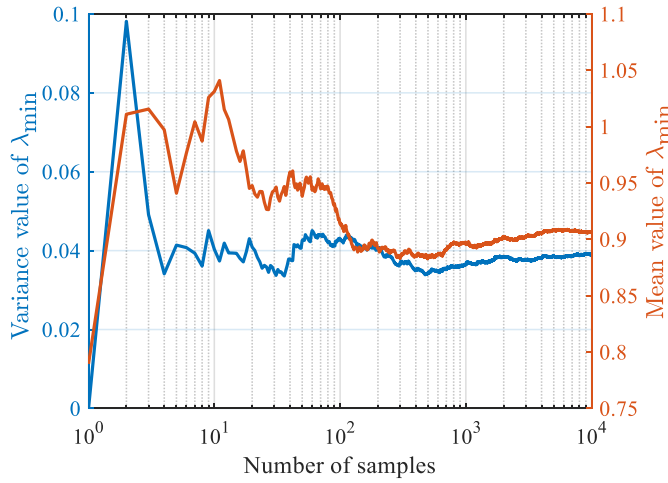
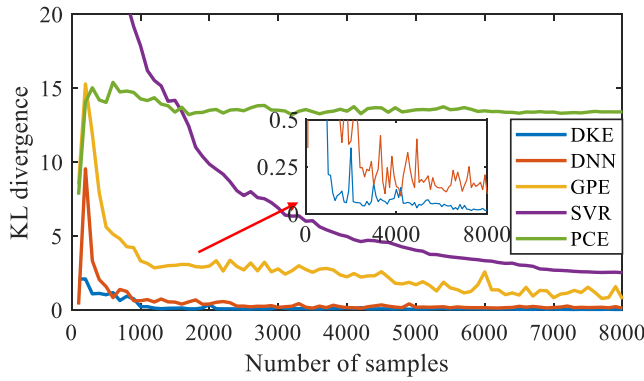
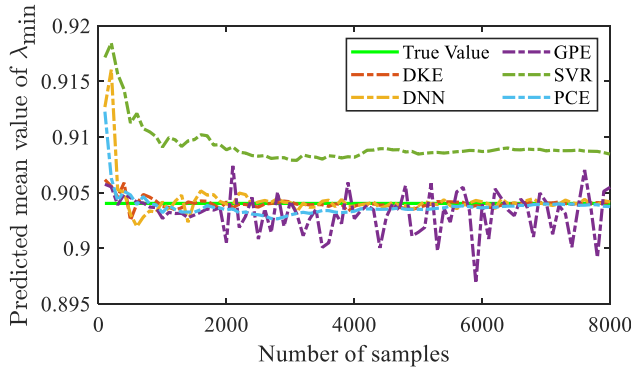


Fig. 7. Convergence curve of the MCS method.



(a)



(b)

Fig. 8. Statistical values of load margin versus the number of samples. (a) KL divergence versus the number of samples; (b) Predicted mean value of $\lambda_{\min}$ versus the number of samples.

the proposed DKE, the structure of its neural network layer is set as “25-50-20”. Hyperparameters in all methods are tuned with grid searching to show their best performance in terms of fair comparisons.

To quantify the performance of each method, we employ the well-known Kullback Leibler (KL) divergence as the index. It can measure the difference between two probabilistic distributions. In this paper, the difference between the predicted probabilistic distribution and the probabilistic distribution ob-

TABLE I
PERFORMANCE COMPARISON FOR VARIOUS METHODS WITHOUT TOPOLOGY CHANGE IN 39-BUS SYSTEM.

Performance index	MCS	DKE	SVR
KL divergence	0	0.0574	2.6752
Number of samples	10000	4300	7000
Simulation time	≈ 33 h	< 1 s	< 1 s
Performance index	GPE	DNN	PCE
KL divergence	2.8310	0.1680	13.3639
Number of samples	1200	5000	1500
Simulation time	< 1 s	< 1 s	< 1 s

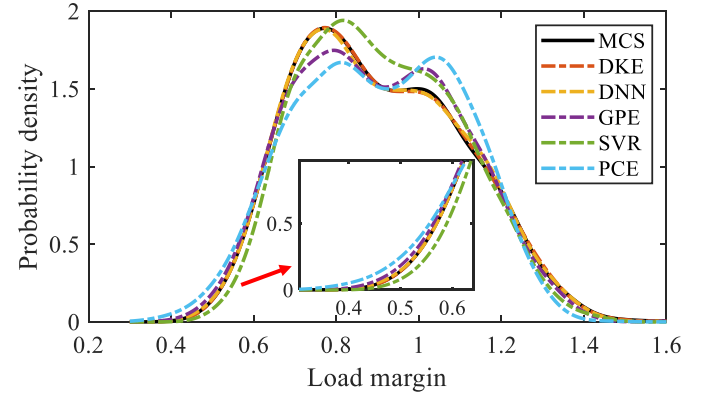


Fig. 9. Predicted probability density of different methods without topology changes in the 39-bus system.

tained by MCS is quantified. KL divergence can be formulated as [39]:

$$KL(\pi_p \parallel \pi_{MCS}) = \int \pi_p \log \frac{\pi_p}{\pi_{MCS}} \quad (36)$$

where π_p and π_{MCS} are the predicted probabilistic distributions with the surrogate models and MCS, respectively. The samples are divided into training dataset and testing dataset with the ratio of 4:1 and only the testing dataset is utilized to quantify the KL divergence rather than using all samples of MCS in [17]. This allows verifying the generalization capability of the surrogate models.

Figs. 8(a) and (b) show the statistical values versus the number of samples. It can be seen that the mean of the load margin converges much faster than the KL divergence since it is global index while the KL divergence depicts the differences between two probabilistic distribution comprehensively. The deep learning based models, i.e., DKE and DNN achieve much higher prediction accuracy than GPE, SVR and PCE, though GPE and PCE require much less samples to converge, see Table I. This demonstrates the strong capability of DKE and DNN for feature extraction and accurate PLMA. Compared with DNN, DKE requires less amount of samples without loss of accuracy. This is because the prior Bayesian inference in DKE can reduce the number of model parameters to be optimized.

Fig. 9 shows the predicted probability density of load margin under different methods when there is no topology change. Thanks to their powerful feature extraction capabilities, DKE and DNN achieve accurate assessment of probabilistic load

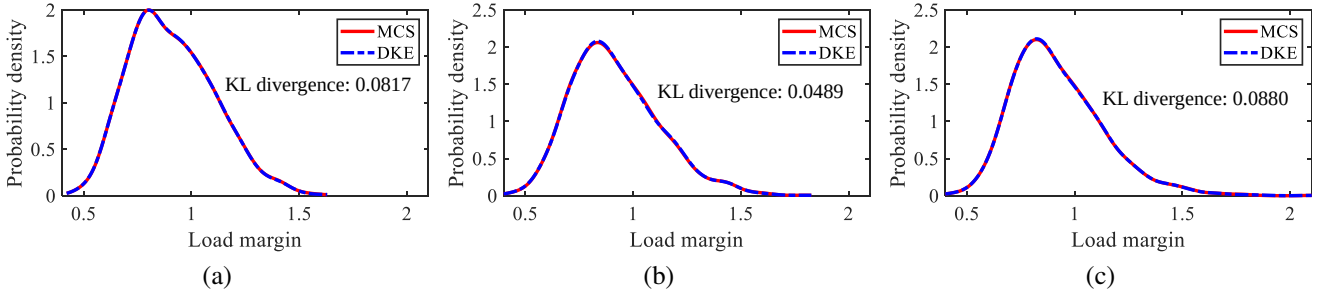


Fig. 10. Performance of DKE under various correlations among uncertain sources. (a) Correlation = 0.2; (b) Correlation = 0.4; (c) Correlation = 0.6.

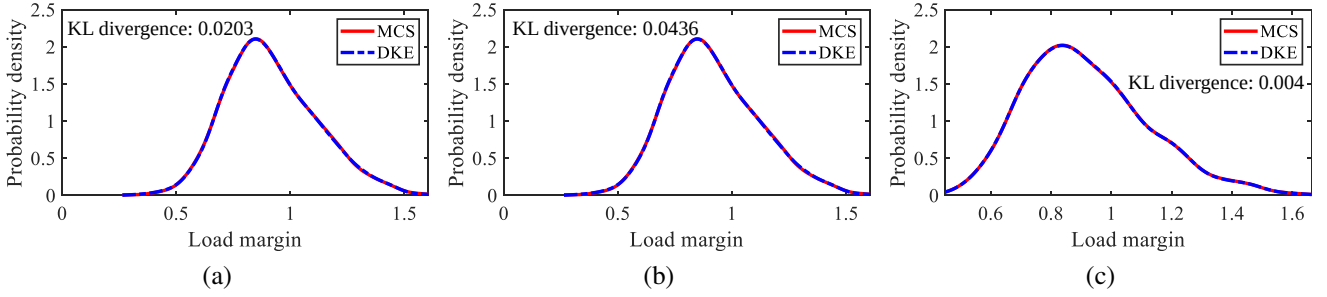


Fig. 11. Performance of DKE under various distributions of loads. (a) Gamma distribution; (b) Gumbel distribution; (c) Uniform distribution.

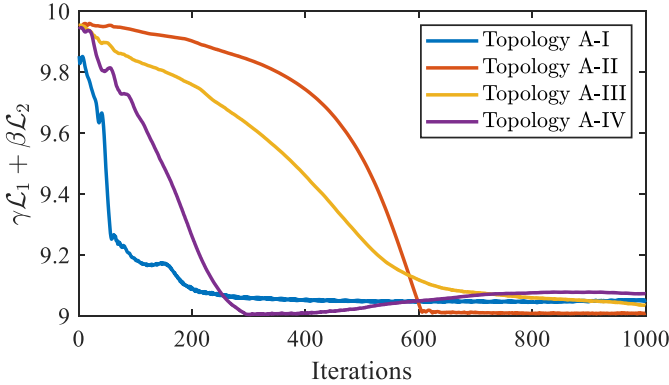


Fig. 12. The trend of invariant representation distance loss during transfer learning.

margin. However, the shadow learning based methods, i.e., GPE, SVR and PCE, have large bias on the prediction at the tails and summits of the load margin distribution. This may lead to underestimate of voltage stability, yielding possible cascading failure.

B. Performance Comparisons under Different Correlations and Distributions

Due to the locations or environmental factors of loads and

wind farms, there may be correlations among them. In this section, the correlations of 0.2, 0.4 and 0.6 are tested. Figs. 10(a)-10(c) show that the proposed DKE can perform accurate PLMA under various correlations. An interesting phenomenon is that the long tail of large load margin happens when correlations among uncertain sources increase.

Since loads may not follow Gaussian distribution in practice, the performance of DKE under different distributions is also investigated. Specifically, three distributions, namely Gamma, Gumbel and Uniform distributions, are employed to depict the probabilistic characteristics of loads respectively [40]-[41], while the wind speed of each wind farm still obeys the Weibull distribution. The mean and standard deviation of them are the same as Gaussian distribution. Without losing generalization, the correlation between each uncertainty source is set as 0.4. The proposed DKE also performs well under different distributions of uncertain sources, as shown in Fig. 11. This is because the finite realizations can be covered by Gaussian process, which is basic rationale of the proposed DKE. Consequently, the performance of DKE is not affected by the specific distribution of inputs.

C. Performance Comparison under Topology Changes

Once the power system topology has changed, DKE has to be retrained since the CPF model is changed. However, this is

TABLE II
PERFORMANCE COMPARISON UNDER VARIOUS TOPOLOGY CHANGE SCENARIOS FOR 39-BUS SYSTEM.

Performance index	Topology A-I	Topology A-II	Topology A-III	Topology A-IV
KL divergence (transferred DKE)	0.0933	0.0236	0.0477	0.0840
Number of samples (transferred DKE)	400	300	400	700
KL divergence (untransferred DKE)	458.431	166.536	15.475	15.272
KL divergence (Retrained DKE)	6.342	15.286	9.263	2.616
Training time of fine tuning	34.33s	35.44s	35.57s	34.77s

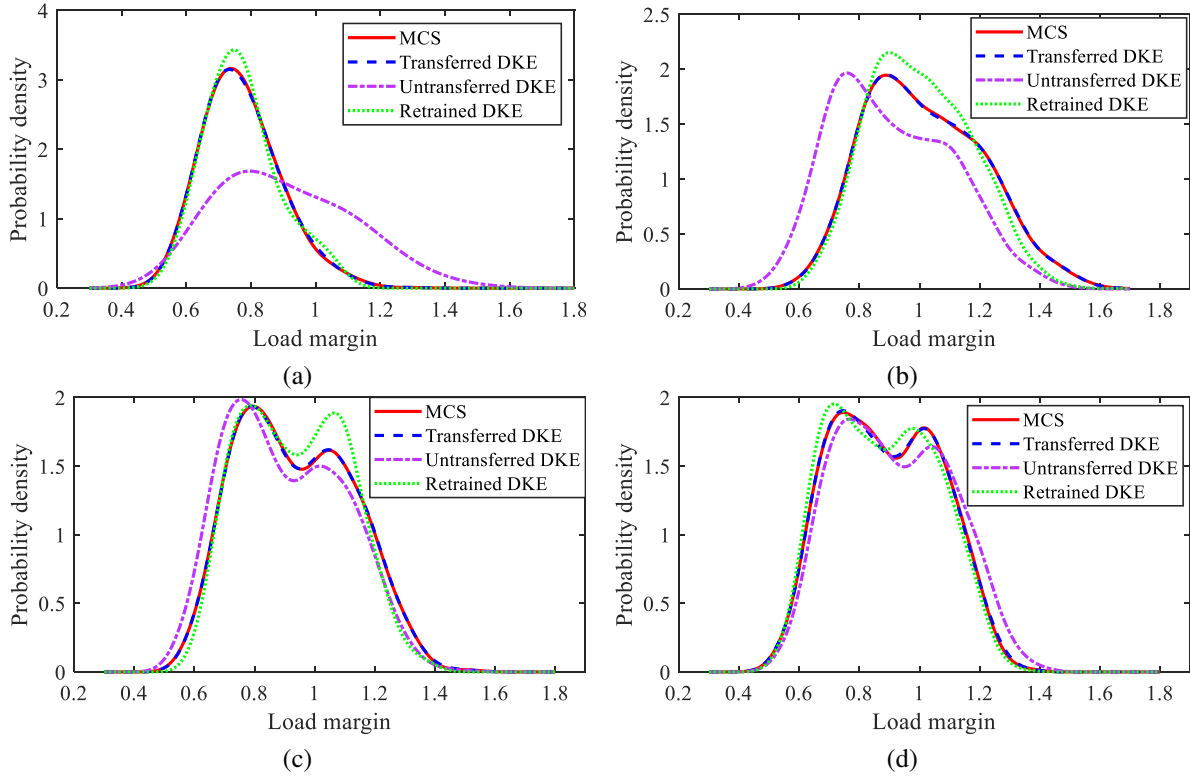


Fig. 13. Load margin probability density prediction after transfer learning under various topology scenarios. (a) Topology A-I; (b) Topology A-II; (c) Topology A-III; (d) Topology A-IV.

time-consuming as thousands of load margin re-calculations are needed. Consequently, the transfer learning for DKE is needed to timely update the surrogate model for quick PLMA. Since DKE has been demonstrated to be the best approach in the previous section, we only show the comparison results among the original DKE (untransferred DKE), the DKE with transfer learning (transferred DKE) and without transfer learning (retrained DKE). This allows us to verify the effectiveness of the proposed transfer learning approach. Four topology change scenarios have been considered in the modified 39-bus system:

- **Topology A-I:** Line 5-6 is removed;
- **Topology A-II:** Line 24-28 is added;
- **Topology A-III:** A 100 MW generator is connected to Bus 4;
- **Topology A-IV:** Generator 37 is removed.

The dataset generation is same as that in Section IV-A. γ and β are set as 10^{-3} and 10^{-4} , respectively. Besides, the training iterations are set as 1000. The structure of the feature transformer $G_f(\cdot)$ and the regressor $G_r(\cdot)$ are “25-10” and “10-50-20-10-1”, respectively. Note that the output dimension of $G_f(\cdot)$ is smaller than its input dimension as we intend to extract its invariant subspace.

Fig. 12 shows the trend of invariant representation distance loss between two topology during transfer learning. It can be seen that the invariant representation distance is significantly reduced via transfer learning. It also indicates that the larger the $\gamma\mathcal{L}_1 + \beta\mathcal{L}_2$ is, more samples are required to fine tune the DKE. Therefore, $\gamma\mathcal{L}_1 + \beta\mathcal{L}_2$ can be used as an transferable

index. The performance comparisons for each method and the probability density function estimation results are shown in Table II and Fig. 13. In Table II, “Retrained DKE” means the reduced data under the new topology to train a new DKE model, where the amount of the reduced data is same as that used to perform transfer learning. It can be seen that, in general, when topology change occurs, the load margin distribution changes. It is interesting to find that the line change has higher impacts on the load margin distribution changes than the generator changes. This may be due to the fact that voltage stability is a local phenomena and line changes may affect more the reactive power change to support local voltage in the modified IEEE 39-bus power system. If the original DKE is directly used for PLMA without considering the topology changes, significant errors are obtained. By contrast, with the proposed transfer learning framework, only a few hundred of samples are required to update DKE model to yield accurate PLMA under new topology. This is only 10% data of the original sample size and thus allows the quick adaptation of the proposed method in PLMA. Besides, the training time for fine tuning is about around 35 seconds, resulting in feasibility of online applications, i.e., in a 15 or 30-minute’s dispatch interval.

Although there is no closed form to formulate the difference between two domains since CPF model is solved in an iterated way, we can quantify the difference between two domains under different topologies via their distribution KL divergence. This is because their domain space can be depicted by their corresponding datasets. Specifically, since the proposed surrogate model can accurately approximate the CPF model under

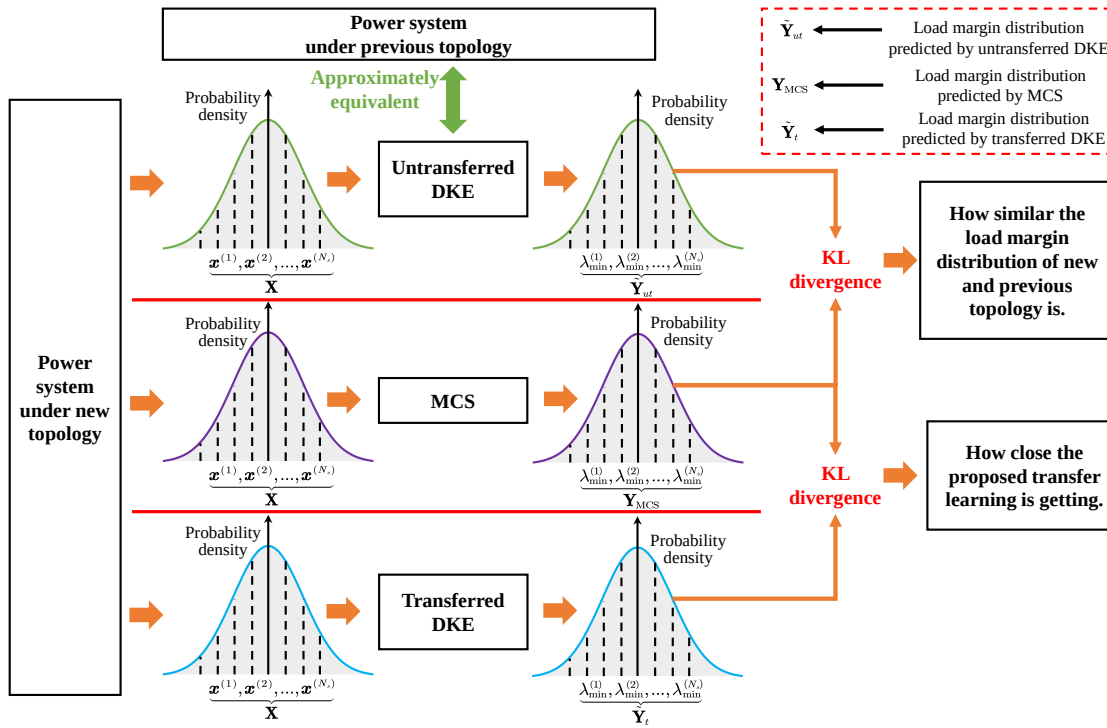


Fig. 14. Explanation for the difference of load margin distribution under different topologies.

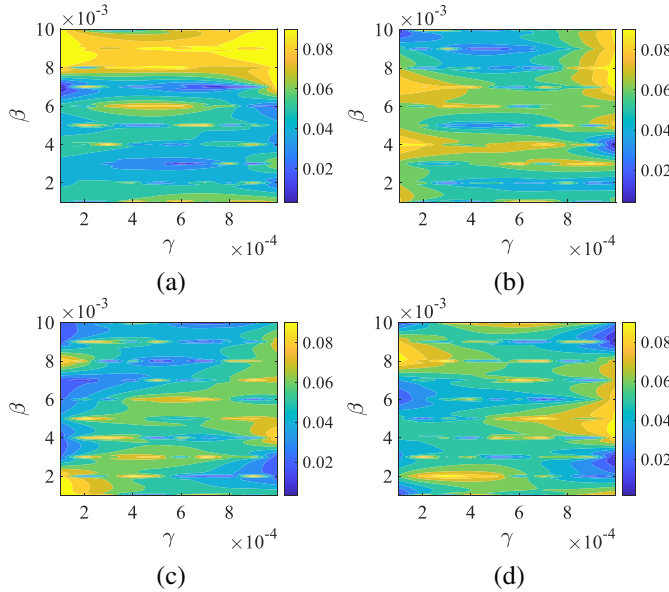


Fig. 15. Hyperparameter sensitivity analysis under various topology change scenarios. (a) Topology A-I; (b) Topology A-II; (c) Topology A-III; (d) Topology A-IV.

certain topology, the untransferred DKE can represent the CPF model under previous topology. Consequently, under the same distribution of input, the KL divergence between load margin distributions obtained by untransferred DKE and MCS can be utilized to quantify the difference between two domains, see Fig. 14. Besides, the distribution KL divergence between the transferred DKE and MCS under new topology can describe how close we are getting. As shown in Table II, the transferred DKE always achieves a KL divergence near to zero toward

MCS, demonstrating the effectiveness of the proposed transfer learning.

D. Hyperparameter Sensitivity Analysis

The hyperparameters can have impacts on the performance of the transfer learning approach and thus this section shows the sensitivity analysis. For the four topology change scenarios in the modified IEEE 39-bus power system, Figs. 15 (a)-(d) present the trend of KL divergences versus various values of β and γ . Note that the values of color bars denote values of KL divergences. It can be observed that the KL divergence of the proposed transfer learning scheme remains below 0.1 when it transfers to different topology. This means that the proposed method is not sensitive to hyperparameters, demonstrating the robustness of the proposed method.

E. Larger-Scale System Testing Results

To demonstrate the scalability of the proposed method, new tests are performed on the modified IEEE 118-bus power system. 11 wind farms with rated power 45 MW, 60 MW, 50 MW, 50 MW, 90 MW, 75 MW, 30 MW, 30 MW, 30 MW, 30 MW, and 90 MW are connected to buses 3, 7, 13, 16, 37, 38, 45, 50, 93, 94, and 114, respectively. Consequently, 110-dimension uncertain inputs (11 wind generators and 99 loads) are constructed. The hyperparameter setting is similar to Section IV-A. The structure of regressor $G_r(\cdot)$ is “110-10”. The following topology change scenarios are investigated:

- **Topology B-I:** Line 30-38 is removed from the modified IEEE 118-bus power system;
- **Topology B-II:** Line 40-117 is added to the modified IEEE 118-bus power system;

TABLE III
TEST RESULTS FOR DIFFERENT TOPOLOGY CHANGE SCENARIOS IN THE 118-BUS SYSTEM.

Performance index	Topology B-I	Topology B-II	Topology B-III	Topology B-IV
KL divergence (transferred DKE)	0.0674	0.0868	0.0716	0.0359
Number of samples (transferred DKE)	500	600	400	500
KL divergence (untransferred DKE)	208.390	59.641	11.893	57.436
KL divergence (Retrained DKE)	6.381	8.148	17.518	3.293
Training time of fine tuning	38.24s	37.32s	37.87s	37.75s

TABLE IV
PERFORMANCE COMPARISON BETWEEN MCS AND PROPOSED METHOD FOR THE 118-BUS SYSTEM WITHOUT DIFFERENT TOPOLOGY CHANGE SCENARIOS.

Performance index	MCS	DKE
KL divergence	0	0.0783
Number of samples	10000	4100
Simulation time	≈ 80 h	< 1 s

- **Topology B-III:** A 100 MW generator is connected to Bus 17;
- **Topology B-IV:** Generator 25 is removed from the modified IEEE 118-bus power system.

The test results for different topology change scenarios in the modified 118-bus system are shown in Tables III and IV. It can be observed that in the high-dimension uncertain input scenarios, the proposed DKE still needs much less samples as compared to MCS and the KL divergence remains at a very low level. The load margin probability density prediction

after transfer learning under various topology change scenarios is displayed in Fig. 16. It can be found that the proposed transfer learning scheme achieves high accuracy and only requires around 10% data under the new topology to update the original surrogate model for PLMA. It is also noticed that the line changes have larger impacts than the generators, which is consistent with the conclusion in the 39-bus system.

F. Performance under $N-k$ Topology Changes

The above results are for $N-1$ topology changes. To verify the performance of the proposed method under $N-k$ topology changes, up to four lines are removed/added simultaneously in the IEEE 118-bus system. Nine $N-k$ topology change scenarios are considered, as shown in Table V, where “Topology C-II-2” means two lines are changed under Topology C-II, for example. Other simulation settings are same as those in Section IV-E.

Table VI and Fig. 17 show the performance of the proposed method under $N-k$ topology changes. From Figs. 17(a)-17(c), it can be observed that, with more lines tripped, the load margin becomes smaller, resulting in a larger probability

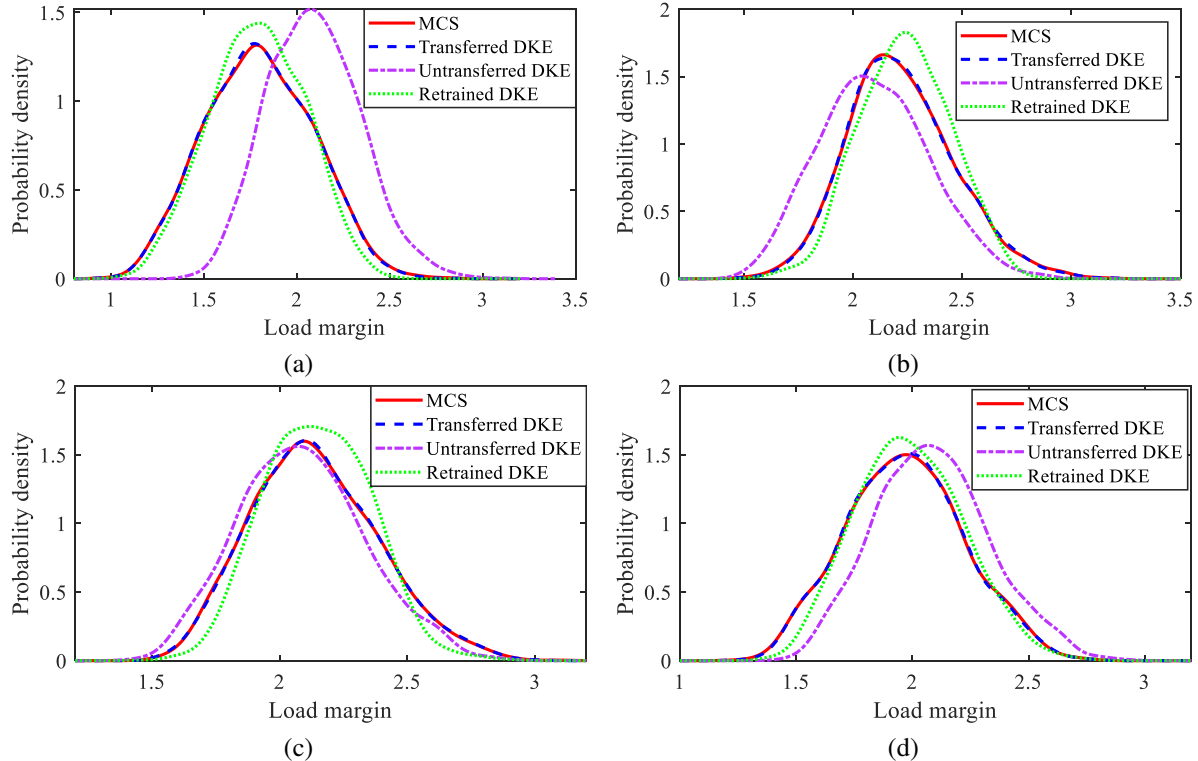


Fig. 16. Load margin probability density prediction after transfer learning under various topology change scenarios. (a) Topology B-I; (b) Topology B-II; (c) Topology B-III; (d) Topology B-IV.

TABLE V
N-K TOPOLOGY CHANGE SCENARIOS IN THE 118-BUS SYSTEM.

Number of changes	Topology C-I	Topology C-II	Topology C-III
2	Line 30-38 removed Line 69-75 removed	Line 30-38 removed Line 40-117 added	Line 40-117 added Line 18-72 added
3	Line 30-38 removed Line 69-75 removed Line 19-34 removed	Line 30-38 removed Line 69-75 removed Line 40-117 added	Line 40-117 added Line 18-72 added Line 20-24 added
4	Line 30-38 removed Line 69-75 removed Line 19-34 removed Line 33-37 removed	Line 30-38 removed Line 69-75 removed Line 40-117 added Line 18-72 added	Line 40-117 added Line 18-72 added Line 20-24 added Line 21-24 added

TABLE VI
TEST RESULTS FOR N-K TOPOLOGY CHANGE SCENARIOS IN THE 118-BUS SYSTEM.

Performance index	Topology C-I-2	Topology C-I-3	Topology C-I-4	Topology C-II-2	Topology C-II-3
KL divergence (transferred DKE)	0.0269	0.0259	0.0230	0.086	0.085
Number of samples (transferred DKE)	500	500	500	600	500
KL divergence (untransferred DKE)	136.913	326.859	2936.4	47.612	66.156
KL divergence (retrained DKE)	4.044	5.889	7.380	1.392	4.120
Training time of fine tuning	40.49s	39.91s	38.17s	41.56s	40.46s
Performance index	Topology C-II-4	Topology C-III-2	Topology C-III-3	Topology C-III-4	
KL divergence (transferred DKE)	0.0695	0.0528	0.0516	0.0897	
Number of samples (transferred DKE)	500	500	500	500	
KL divergence (untransferred DKE)	18.084	84.669	86.536	86.000	
KL divergence (retrained DKE)	3.435	5.257	4.405	4.824	
Training time of fine tuning	39.90s	38.23s	38.30s	38.51s	

of voltage instability. Besides, comparing Figs. 17(a)-17(c) with Figs. 17(g)-17(i), it can be seen that the removal of lines may lead to more violent distribution changes of load margin than adding of lines. It is not easy to accurately quantify how much distribution changes the proposed method can tolerate, since there is no closed mathematical form between topology changes and load margin distribution. However, the proposed method can perform accurate PLMA with only 500 samples to adapt to new topology even if there are not much common parts between the load margin distributions after topology changes, as illustrated in Fig. 17(c). This indicates that the proposed method is adequate for the practical application.

V. CONCLUSIONS

This paper proposes a transferable PLMA framework considering topology changes and uncertain wind generations and loads. A probabilistic DKE that extends the Gaussian process kernel to the DNN structure is developed to extract the relationship between uncertain sources and load margin. This allows us to gain the advantages of DNN in dealing with high-dimension uncertain inputs and the uncertainty quantification capability of the Gaussian process for PLMA. A new transfer learning approach that minimizes the invariant representation space distance between old topology and new one is proposed to quickly update the DKE model with only a few samples. Numerical results in the modified IEEE 39-bus and 118-bus power systems demonstrate that the proposed method can 1) efficiently capture the probabilistic distribution of load margin and obtain accurate PLMA, 2) can quickly update the DKE model accurately under new topology with only a

few samples and 3) achieve a better performance as compared to other approaches under various scenarios. Although the proposed method can be accurately and efficiently transferred to new topology, it lacks the analytical interpretations that how uncertain sources affect the load margin, which is still an open problem in PLMA field. Future work would explore this approach in other operational planning practice and their physical interpretation.

REFERENCES

- [1] P. Kundur, *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, *et al.*, "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. Kim, *et al.*, "Probabilistic security analysis using SOM and Monte Carlo simulation," *Conference Proceedings, New York*, 2002.
- [4] H. Chiang, *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.
- [5] J. Zhao, *et al.*, "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.
- [6] A. Schellenberg, *et al.*, "Cumulant-based stochastic nonlinear programming for variance constrained voltage stability analysis of power systems," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 579-585, 2006.
- [7] K. Liu, *et al.*, "Probabilistic evaluation of static voltage stability taking account of the variation of load and stochastic distributed generations," in *2013 International Conference on Electrical Machines and Systems (ICEMS)*, 2013, pp. 418-421.
- [8] J. Zhang, *et al.*, "Voltage stability analysis considering the uncertainties of dynamic load parameters," *IET Gener. Transmiss. Distrib.*, vol. 3, no. 10, pp. 941-948, Oct. 2009.

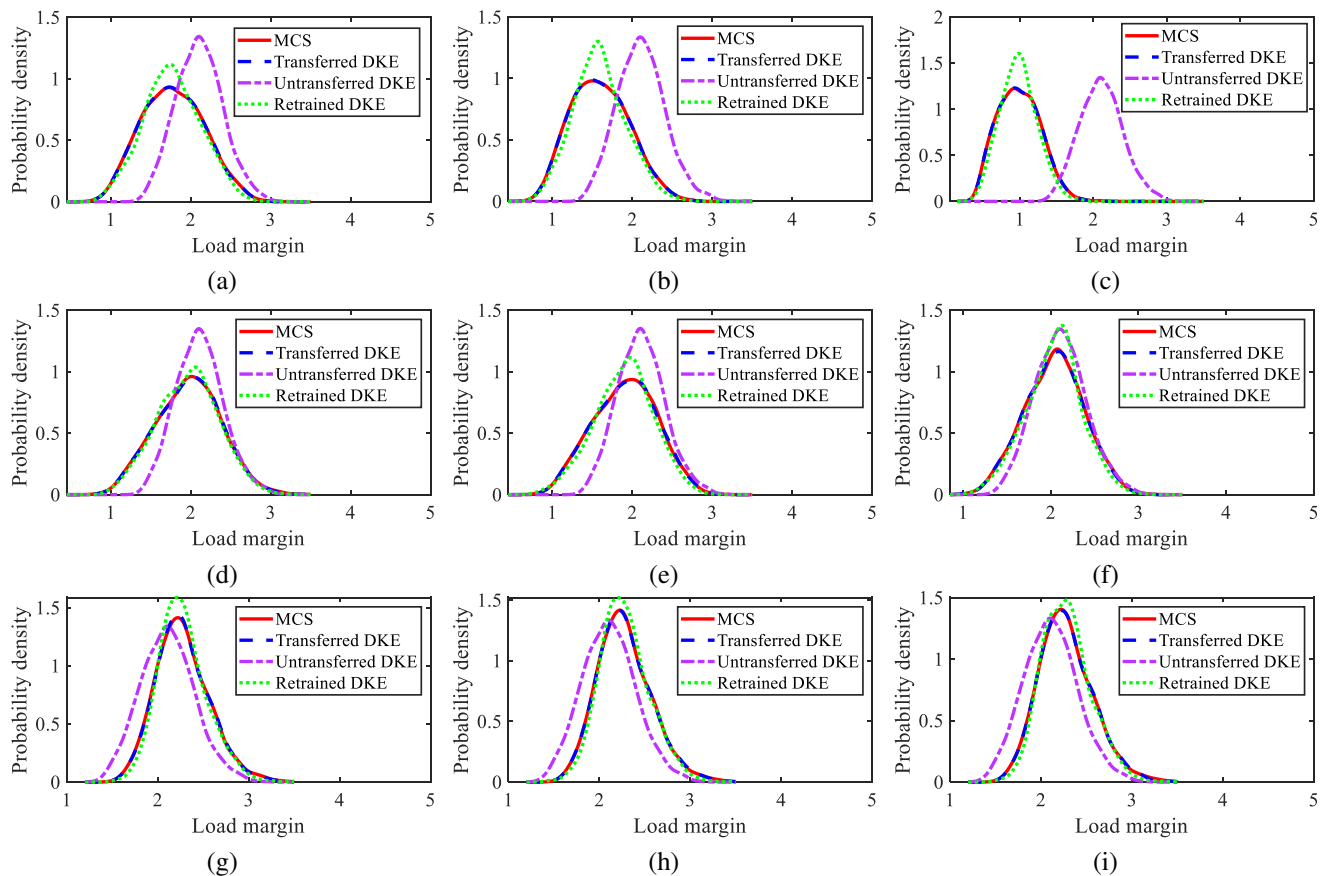


Fig. 17. Performance of DKE under various N-k topologies. (a) Topology C-I-2; (b) Topology C-I-3; (c) Topology C-I-4; (d) Topology C-II-2; (e) Topology C-II-3; (f) Topology C-II-4; (g) Topology C-III-2; (h) Topology C-III-3; (i) Topology C-III-4.

- [9] C. Zheng, *et al.*, "Regression tree for stability margin prediction using synchrophasor measurements," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1978-1987, May 2013.
- [10] S. Li, *et al.*, "Adaptive online monitoring of voltage stability margin via local regression," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 701-713, Jan. 2018.
- [11] Liu, S., *et al.*, "An integrated scheme for static voltage stability assessment based on correlation detection and random bits forest," *International Journal of Electrical Power & Energy Systems*, vol.130, pp. 106898, 2021.
- [12] A. Prasad, *et al.*, "Determination of TTC and application of SVM for demarcation of stability limits," in *2016 International Conference on Cogeneration, Small Power Plants and District Energy (ICUE)*, 2016, pp. 1-5.
- [13] J. Yan, *et al.*, "Deep learning based total transfer capability calculation model," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 952-957.
- [14] Y. Liu, *et al.*, "Online TTC estimation using nonparametric analytics considering wind power integration," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 494-505, Jan. 2019.
- [15] X. Xu, *et al.*, "Power system voltage stability evaluation considering renewable energy with correlated variabilities," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3236-3245, May 2018.
- [16] S. Nejadfar-jahromi, *et al.*, "Data-driven look-ahead voltage stability assessment of power system with correlated variables," *IET Gener. Transmiss. Distrib.*, vol.16, no.9, pp. 1795-1807, 2022.
- [17] Y. Xu, *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.
- [18] Y. Xu, *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.
- [19] X. Wang, *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.
- [20] R. Billinton, *et al.*, "Effects of load forecast uncertainty on bulk electric system reliability evaluation," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 418-425, May 2008.
- [21] H. Huang, *et al.*, "Coordinated damping control design for DFIG-based wind generation considering power output variation," *IEEE Trans. Power Syst.* vol. 27, no. 4, pp. 1916-1925, Nov. 2012.
- [22] M. Aien, *et al.*, "Probabilistic optimal power flow in correlated hybrid wind-PV power systems: A review and a new approach," *Renewable Sust. Energy Rev.*, vol. 41, pp. 1437-1446, 2015.
- [23] V. Yarlagadda, *et al.*, "Integrated voltage and generator stability analysis of a three bus system combining SMIB and SMLB Systems," in *2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 2021, pp. 1-7.
- [24] A. G. Wilson, *et al.*, "Deep kernel learning", In *Artificial intelligence and statistics*, May. 2016, pp. 370-378.
- [25] C. E. Rasmussen, *et al.*, *Gaussian processes for machine learning*. Cambridge, MA, USA: MIT Press, 2006.
- [26] A. Paszke, *et al.*, "Automatic differentiation in pytorch," Tech. Rep., 2017. [Online]. Available: <https://github.com/pytorch/pytorch/blob/master/CITATION>.
- [27] X. Chen, *et al.*, "Representation subspace distance for domain adaptation regression." In *International Conference on Machine Learning*, 2021, pp. 1749-1759.
- [28] D. Kingma, *et al.*, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2014.
- [29] R. B. Nelsen, *An Introduction to Copulas*. Berlin, Germany: Springer, 2007.
- [30] M. P. Wand and M. C. Jones, *Kernel smoothing*. London: Chapman & Hall, 1995.
- [31] S. Xia, *et al.*, "Probabilistic transient stability constrained optimal power flow for power systems with multiple correlated uncertain wind generations," *IEEE Trans. Sustainable Energy*, vol. 7, no. 3, pp. 1133-1144, Jul. 2016.

- [32] A. Paszke, *et al.*, “PyTorch: An imperative style, high-performance deep learning library,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.
- [33] R. D. Zimmerman, *et al.*, “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.
- [34] P. N. Papadopoulos, *et al.*, “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.
- [35] M. V. Suganyadevi, *et al.*, “Support vector regression model for the prediction of loadability margin of a power system,” *Applied Soft Computing*, vol. 24, pp. 304–315, Nov. 2014.
- [36] S. Kamalasadan, *et al.*, “Novel algorithm for online voltage stability assessment based on feed forward neural network,” in *2006 IEEE Power Engineering Society General Meeting*, 2006, pp. 7.
- [37] F. Pedregosa, *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [38] S. Marelli, *et al.*, “UQLab user manual–Polynomial chaos expansions,” Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, Tech. Rep. V1.1-104, 2018.
- [39] Y. Marzouk, *et al.*, “A stochastic collocation approach to Bayesian inference in inverse problems,” *Commun. Comput. Phys.*, vol. 6, no. 4, pp. 826–847, Oct. 2009.
- [40] U. H. Ramadhani, *et al.*, “Review of probabilistic load flow approaches for power distribution systems with photovoltaic generation and electric vehicle charging,” *International Journal of Electrical Power & Energy Systems*, vol.120, pp. 106003, 2020.
- [41] V. Singh, *et al.*, “Uncertainty handling techniques in power systems: a critical review,” *Electric Power Systems Research*, vol. 203, pp. 107633, 2020.