Bayesian Deep Learning for Dynamic Power System State Prediction Considering Renewable Energy Uncertainty

Shiyao Zhang and James J. Q. Yu

***Abstract*—Modern power systems are incorporated with distributed energy sources to be environmental-friendly and cost-effective. However, due to the uncertainties of the system integrated with renewable energy sources, effective strategies need to be adopted to stabilize the entire power systems. Hence, the system operators need accurate prediction tools to forecast the dynamic system states effectively. In this paper, we propose a Bayesian deep learning approach to predict the dynamic system state in a general power system. First, the input system dataset with multiple system features requires the data pre-processing stage. Second, we obtain the dynamic state matrix of a general power system through the Newton-Raphson power flow model. Third, by incorporating the state matrix with the system features, we propose a Bayesian long short-term memory (BLSTM) network to predict the dynamic system state variables accurately. Simulation results show that the accurate prediction can be achieved at different scales of power systems through the proposed Bayesian deep learning approach.**

***Index Terms*—Bayesian deep learning, data analytics, Newton-Raphson power flow, renewable energy source, system state.**

I. Introduction

WITH the development of technology, the modern power system has become more diversified than the conventional power system, such as the increase in the utilization of renewable energy sources (RESs). Although RESs contribute to improving the environmental impacts to prevent global warming, the system uncertainties occur due to the uncertain and intermittent renewable power generations. In practice, the integration of RESs in the system results in sudden instability phenomena of the entire power system. Meanwhile, the power system state variables may also undergo drastic changes. To tackle such issues, the accurate state prediction for power system is crucial to capture the massive system uncertainties from RESs to support the system operations and services. For instance, a suitable prediction technique can be deployed for system data analytics and contributes to the stable operation and planning in the power system.

Existing solutions to predict the dynamic system state in the power system mainly involve in deploying the traditional prediction tools [1], [2]. For instance, in [1], a robust approach, namely extended Kalman filter, is developed to monitor system state dynamics in a faster and more reliable manner. Besides, a new robust generalized maximum-likelihood-type unscented Kalman filter in [2] is proposed to predict the state innovation vectors in the system. Although the previous studies propose the statistical techniques on system state prediction, they do not consider the penetration of RESs in the power system. By considering the RESs deployed in the system, some existing researches deploy sufficient statistical tools on predicting system state [3]-[5]. For example, a Markov model with the Viterbi algorithm is proposed in [3] for state prediction in different power systems with the penetration of RESs. Furthermore, in [4], a statistical Gaussian mixture model is developed to estimate the real-time system measurements. Nonetheless, these approaches are not practical since such models already assume that both the system and models are linear. Lastly, a novel interval state estimation algorithm is devised in [5] to consider different uncertainties of distributed generation outputs, as well as line parameters in unbalanced distribution systems. In practice, complex power systems have typical non-linear features, such as time-varying loads and multiple power generations. Hence, to handle these non-linear features, several studies are involved in deploying non-parametric methods for prediction in the system. In [6], a non-parametric kernel estimation method is applied to determine the state condition to characterize the stability issues of the system.

However, there exists a research gap in the state prediction problem of traditional dynamic system. The above researches unilaterally consider linear or non-linear features in the power system. In practice, a dynamic power system consists of various types of linear and non-linear features. For example, the electrical loads in the system are either linear or non-linear loads according to the network structure. As an emerging technique to handle the problems in complex sys-

Manuscript received: December 31, 2020; revised: March 2, 2021; accepted: May 17, 2021. Date of CrossCheck: May 17, 2021. Date of online publication: July 28, 2021.

This work was supported by the General Program of Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515011032) and the Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation (No. 2020B121201001).

S. Zhang is with the Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: syzhang@ieee.org).

J. J. Q. Yu (corresponding author) is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: yujq3@sustech.edu.cn).

DOI: 10.35833/MPCE.2020.000939

tems, machine learning approaches can be a good candidate to fully respect both linear and non-linear factors in a general power system. Some studies utilize machine learning tools for system data prediction problems [7]-[9]. For instance, a novel machine learning tool, support vector machine, can be applied for time-series prediction in the power system [7]. Besides, a low-rank tensor learning model can be utilized to predict system measurements such as solar power generation [8]. Last but not least, the use of decision trees can handle a large amount of wide-area information to keep the stability of the power system [9].

Beyond the aforementioned studies, as a subset of machine learning, deep learning has been widely utilized in the related research fields [10]. In the power system, the implementation of deep learning techniques can use multiple neural network layers to extract latent system features, which can further assist power system operations in practical scenarios. Some existing research works have solved power system problems through deep learning tools [11]-[16]. In [11], a method based on artificial neural network (ANN) is proposed to predict the long-term voltage status to ensure the margin of the voltage stability. Additionally, an intelligent system strategy is proposed in [12] to predict the dynamic voltage deviation to observe the short-term instability of the system based on the ensemble learning of neural networks. However, these studies lack multiple degrees of data interpretability in the power system. For example, the one-layer ANN cannot deeply learn the representation of large-scale power system data due to its simple structure and parameter settings. To tackle such issues, in [13], a model-specific deep neural network (DNN) based power system state estimation scheme is proposed to estimate real-time power system state. Furthermore, the long short-term memory (LSTM) model could be deployed to accurately capture the time-varying dynamic behaviors of active distribution networks [14] and to forecast the solar energy output [15]. Finally, by implementing a surrogate model, a novel LSTM model is applied to capture the time-varying consecutive states [16].

Although the aforementioned studies have demonstrated that deep learning is superior to perform prediction tasks, these studies are indeed based on the deterministic models so that they lack the ability of capturing uncertainty. In the power system, the uncertainties from various sources may expose potential safety issues to cause huge economic losses [17]. Even though several non-deep-learning approaches can deal with system uncertainties, e. g., [3] and [18], their schemes become complex and time-consuming with the increasing size of power networks. In this paper, a probabilistic deep learning model, i.e., Bayesian deep learning, is adopted to incorporate the power system uncertainties into the state estimation framework aiming at a more interpretable model with reliable prediction by means of probability theory in an efficient manner. There are several research works that justify the feasibility of the Bayesian deep learning methods in power system studies [19], [20]. For instance, a probabilistic Bayesian deep learning model is proposed for day-ahead load forecasting to capture the model and load uncertainties [19]. In addition, in [20], the state estimation could be achieved through Bayesian deep learning with consideration of the bad-data circumstance. However, there is no recent study using Bayesian deep learning model to consider the dynamic system state of the entire power system as far as we are concerned.

To bridge the research gaps, we propose a Bayesian deep learning approach to perform state prediction in the general power system considering RES and model uncertainties. The major efforts of this paper are shown as follows.

1) A Newton-Raphson power flow model is deployed to estimate the dynamic system state, and we devise a new state matrix to aggregate with all historical information.

2) By adopting Bayesian LSTM (BLSTM) network in the system, both the model and data uncertainties can be remarkably captured.

3) Accurate predictions can be achieved by means of the proposed BLSTM network for different scales of systems.

The remainder of this paper is organized as follows. Section II presents and illustrates the proposed methodology. In Section III, we formulate a Newton-Raphson power flow model to generate the state matrix of a general power system. The deep learning approach, BLSTM network, is then proposed in Section IV. In Section V, we conduct the performance evaluation of the proposed model with a general power system. Finally, we conclude this paper in Section VI.

II. Proposed Methodology

In this section, we present the proposed methodology, which is summarized in Fig. 1.

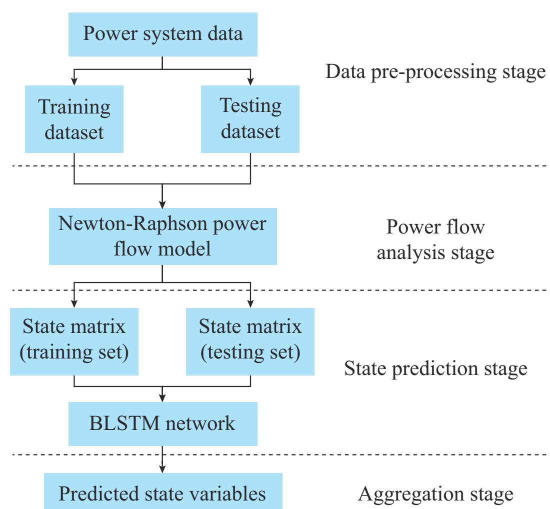

Fig. 1. Workflow of proposed methodology.

The proposed methodology consists of four main stages: data pre-processing, power flow analysis, state prediction, and aggregation. The first stage involves the dataset preparation process, where we collect various types of power system data, such as various renewable power generation, time-varying loads, and dynamic power state information. Then, the input dataset is segregated into a training dataset and a testing dataset. The size of the training dataset is determined by multiple potential features of the entire dynamic power system. After that, both the training and testing datasets are used for the power flow analysis. Once the state matrix is obtained and separated into training and testing sets, we incorporate them into the complete training and testing datasets to fit in the proposed BLSTM network. By properly training the learning system, the dynamic system states of general power systems can be obtained via online inference. Finally, the aggregation stage clusters all the individual probabilistic prediction through convolution with the previously saved weights to gain the final probabilistic dynamic state variables.

A. Dynamic State Assessment

The dynamic state assessment is gained by using the Newton-Raphson power flow model. Note that we consider the alternating-current (AC) power system in the proposed methodology. Hence, we perform AC power flow analysis so as to estimate the AC power state information. The power system network can be modeled as a graph that is defined as $\mathcal{G}=(\mathcal{N},\mathcal{E})$, where $\mathcal{N}=\{1,2,\ldots,N\}$ is the bus set, and N is the total number of buses; and $\mathcal{E}\subseteq\mathcal{N}\times\mathcal{N}$ is the transmission line set. In the power network, we can model a branch $(i,j)\in\mathcal{E}$ with two common buses $i,j\in\mathcal{N}$ as one equivalent π circuit. Hence, the line admittance of this circuit can be expressed as $y_{ij}=g_{ij}+\mathrm{j}b_{ij}$, where $g_{ij}\geq 0$ and $b_{ij}\leq 0$ are the branch conductance and branch susceptance, respectively. In addition, the shunt capacitance of branch (i,j) can be denoted as $c_{ij}=c_{ji}$, which is used for voltage and reactive power control. These aforementioned parameters are the key components to develop the AC power flow model for a general power system.

Through the power flow analysis in the system, the dynamic state variables can be estimated. The collection of these variables can be sorted as the state matrix of the system. Since we need to fit this information into the neural network model, the large amount of historical dynamic state estimations are covered and divided into training and testing sets.

B. Uncertainty Investigation

Data uncertainty in general power systems is typically related to the stochastic uncertainty of the power generation and loads injected as power sources to the system, such as load variability and intermittent power generation. It is more challenging to handle both load and intermittent power generation prediction than either of these individual tasks. Specifically, the stochastic uncertainty of load and intermittent power generation reflects the uncertainty characteristics from different injected sources resulted from the variability of weather and unpredictable human activities. Besides, most of the existing approaches only predict the upper and lower bounds of the forecasting power without the detailed information about the power distribution at every time step. Additionally, most probabilistic prediction methods are indeed deterministic approaches, which cannot fully capture the stochastic uncertainty. In this paper, as mentioned above, we consider the integration with time-varying loads and RESs in AC power systems. Hence, the data uncertainty is taken into account in our proposed model. As existing probabilistic prediction techniques are primarily derived from deterministic models, their capability in capturing the stochastic uncertainty is limited. Hence, most of the existing models cannot represent the data uncertainty due to the stochastic time-varying loads and RESs. It is important to develop a generic probabilistic deep learning model to handle a large number of such uncertainties to provide the confidence bounds for decision-making.

Besides the data uncertainty, we need to investigate the uncertainty of the model related to the output results on dealing with dynamic state prediction, which is defined as model uncertainty. Besides the stochastic uncertainty, the model uncertainty also plays an important role in the probabilistic prediction task, which is used to identify the amount of output uncertainty. Intuitively, the model uncertainty refers to the uncertainty of the model parameters and network structure. The challenge of such uncertainty is to know how much the chosen combinations affect the results of dynamic state prediction in different circumstances. Thus, the aim is to investigate the degree of uncertainty of this model corresponding to the outputs. Since we consider a large amount of model parameters and numerous variations of structures evaluated for different combinations, it is important to observe how the accurate dynamic system state prediction varies under different circumstances, e.g., days, weeks, seasons, and social factors. In addition, to tackle such system conditions, the proposed Bayesian deep learning model is developed by implementing the related parameters inside. The remainder of this paper will introduce how our proposed Bayesian deep learning model can effectively handle both the data and model uncertainties in a detailed manner.

III. Newton-Raphson Power Flow Model

This section presents the problem formulation of Newton-Raphson power flow model. As shown in Fig. 1, the AC power flow model can be utilized to obtain the corresponding state matrix for the power system after the data pre-processing stage, which is for the model training and testing in the proposed BLSTM network. The procedure of AC power flow analysis is shown below in a detailed manner [21].

A. AC Power Flow Model

Based on the dynamic state assessment mentioned in Section II, we further develop the power flow model of a general power system. Considering the nodal equations at each bus, the nodal admittance matrix in the system can be denoted as $\boldsymbol{Y}$. Through an one-line diagram of the whole circuit, for each bus, the current and voltage matrices are denoted as $\boldsymbol{I}$ and $\boldsymbol{V}$, respectively. By Ohm's law, the nodal admittance matrix $\boldsymbol{Y}$ of the entire system follows (1).

$$\boldsymbol{I}=\begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix}=\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & & \vdots \\ y_{N1} & y_{N2} & \ldots & y_{NN} \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix}=\boldsymbol{YV} \tag{1}$$

For the representations of power injections in the system, we denote P_i and Q_i as the active and reactive power injections at bus i, respectively, which are modeled as:

$$P_i = V_i \sum_{j \in \mathcal{N}_i} V_j \; (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}) \tag{2}$$

$$Q_i = V_i \sum_{j \in \mathcal{N}_i} V_j \; (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}) \tag{3}$$

where $\mathcal{N}_i$ is the set of nearby buses connected with bus i; and θ_{ij} is the phase difference of branch (i,j).

In addition, the active and reactive branch power flows of branch (i,j) are modeled as:

$$P_{ij} = V_i^2 (g_{si} + g_{ij}) - V_i V_j (g_{ij} \cos\theta_{ij} + b_{ij} \sin\theta_{ij}) \tag{4}$$

$$Q_{ij} = -V_i^2 (b_{si} + b_{ij}) - V_i V_j (g_{ij} \sin\theta_{ij} - b_{ij} \cos\theta_{ij}) \tag{5}$$

where g_{si} and b_{si} are the shunt conductance and shunt susceptance at bus i, respectively.

For the specific implementation at bus i of the system, it involves in both the renewable power generation and the load demand of bus i, which can be denoted as P_{G_i} and P_{L_i}, respectively. The active and reactive power injections at bus i can also be expressed as:

$$P_i = P_{G_i} - P_{L_i} \quad \forall i \in \mathcal{N} \tag{6}$$

$$Q_i = Q_{G_i} - Q_{L_i} \quad \forall i \in \mathcal{N} \tag{7}$$

In order to ensure the stable system operation, the active and reactive power generation at time t of bus $i \in \mathcal{N}$ should follow their operation limits.

$$P_{G_i,\min} \le P_{G_i} \le P_{G_i,\max} \tag{8}$$

$$Q_{G_i,\min} \le Q_{G_i} \le Q_{G_i,\max} \tag{9}$$

B. State Matrix

The actual measurements for the state estimation in the power system are obtained by means of the Newton-Raphson power flow model. As mentioned before, a new state matrix is devised to aggregate with all the historical information. Through the power flow analysis, the crucial components in the system can be obtained, such as the voltage magnitude, phase angle, active power, and reactive power at each bus. Based on the previous definitions, we denote the voltage magnitude and phase angle at bus i as $|V_i|$ and θ_i, respectively. There are totally $2N-1$ state variables related to the common bus voltage in the system, which explicitly includes N voltage magnitudes and $N-1$ phase angles, where the phase angle for each bus is defined related to the reference bus. By considering the active and reactive power injections at bus i, we define the state vector in the system as:

$$\boldsymbol{x}_i^{\mathrm{T}} = [\,|V_i| \;\; \theta_i \;\; P_i \;\; Q_i\,] \quad \forall i \in \mathcal{N} \tag{10}$$

Then, the state matrix of a general power system can be expressed as:

$$\boldsymbol{X} = [\boldsymbol{x}_1 \;\; \boldsymbol{x}_2 \;\; \dots \;\; \boldsymbol{x}_N] \tag{11}$$

The satisfaction of AC power balance conditions can be modeled as the non-linear equation, which can be expressed as:

$$\boldsymbol{g}(\boldsymbol{X}) = \begin{bmatrix} \Delta \boldsymbol{P} \\ \Delta \boldsymbol{Q} \end{bmatrix} = \boldsymbol{0} \tag{12}$$

where $\Delta\boldsymbol{P}$ and $\Delta\boldsymbol{Q}$ are the matrices of the active and reactive power imbalance, respectively.

According to [21], the utilization of the Newton-Raphson power flow model is capable of solving the above non-linear equation through iterations. After the initial guess of state matrix, the solution can be determined through an iterative manner by means of the Newton-Raphson power flow model. Specifically, given the non-singular Jacobian matrix, the Newton-Raphson power flow model can be converged quadratically from a sufficient initial guess [22]. The actual measurements of the general power system can be estimated through the power flow analysis. The complete process of Newton-Raphson state estimation method is presented in Algorithm 1.

Algorithm 1: Newton-Raphson state estimation method

```
1:  Initial setup for all buses in the system with reference to the slack bus
2:  for k = 1:M_k
3:    Obtain ΔP^k and ΔQ^k through power flow equations in [22]
4:    Develop the Jacobian matrix of the system
5:    Compute Δ|V^k| and Δθ^k and form the state matrix X^k
6:    if the stopping criterion is met then
7:      Update the final solution as X = X^k and then close the for loop
8:    else
9:      Return to Step 3 in the for loop for another iteration and set k = k + 1.
10:   end if
11: end for
```

For every iteration k, the active and reactive power imbalance can be obtained, denoted as $\Delta\boldsymbol{P}^k$ and $\Delta\boldsymbol{Q}^k$, respectively. In addition, by means of Jacobian matrix, the state matrix $\boldsymbol{X}^k$ can be computed. To satisfy the stopping criterion, the iteration number is set as $k = M_k$, where M_k is the maximum iteration number. For the tolerance of the convergence, Algorithm 1 follows the rule as:

$$|\Delta P_i^k| \le \epsilon \quad \forall i \in \mathcal{N} \tag{13}$$

$$|\Delta Q_i^k| \le \epsilon \quad \forall i \in \mathcal{N} \tag{14}$$

where $\epsilon = 10^{-5}$ is a small positive number.

IV. BLSTM Network

This section introduces the proposed BLSTM network. By handling the segregated datasets introduced in Section II, we can extract and analyze the input data features. The first step is associated with the data pre-processing. Then, through the power flow analysis stage, we aggregate all historical information to fit in the proposed BLSTM network. The steps for network training are covered below in detail.

A. Data Pre-processing

Based on the proposed methodology mentioned in Section II, we consider a predetermined period for both renewable power generation and load demands, which can be denoted as $\mathcal{T} = \{1, 2, \dots, N_T\}$ with N_T time slots. Given the time period $\mathcal{T}$, we use $\boldsymbol{R}_t$ to denote the collection of renewable power generation and load demand features in a general power system, where $t \in \mathcal{T}$. As mentioned in Section II, the complex AC power system consists of various entries such as RESs and time-varying loads. To tackle the complex system, the implementation of deep learning approach can be a suitable candidate for accurate state prediction. Specifically, we collect the input data from December 2015 to December 2016. For solar generation, wind generation, and residential load, the profiles during this period are captured with a time duration of 15 min in a cumulative manner. Furthermore, the time-varying loads fluctuate due to the practical electricity usage patterns. In this paper, we collect the input power data associated with 37 European countries. Specifically, using the topology of the city power grid, the related power generation and loads are implemented. By sorting out the input data, we can extract multiple different features of the power grid. In addition, since the training and testing datasets are constructed by capturing the periodical changes of the profiles, a cross-validation step is also required for the data pre-processing stage to assess the exact separation of training and testing sets. Besides, in order to improve the robustness of the proposed deep learning model, the sufficient amount of training dataset is considered. Furthermore, by considering the state matrix, the extracted features can rapidly increase with the number of buses in the power system. Thus, the total number of the features from the input data depends on general power network information. Besides, before we fit the input data to the BLSTM network, we first normalize the input dataset to $[0,1]$ with min-max normalization. For the missing values in the input dataset, we utilize linear interpolation for recovering the values.

B. LSTM

The pre-processing input data with different items are fed in the Bayesian deep learning model for the purpose of model training. In this paper, we focus on the BLSTM network, which is inherently a probabilistic model for handling time-series data. The network parameters in the BLSTM network are expressed by conditional probabilities, which are different from the typical LSTM network with fixed parameters. Since the input data cover the features of renewable power generation, it is crucial to capture the characteristics of RESs. Hence, our proposed BLSTM network can capture both the model uncertainty and stochastic uncertainty [10].

To illustrate the basic architecture of the proposed BLSTM network, we first introduce the structure of one LSTM cell shown in Fig. 2.

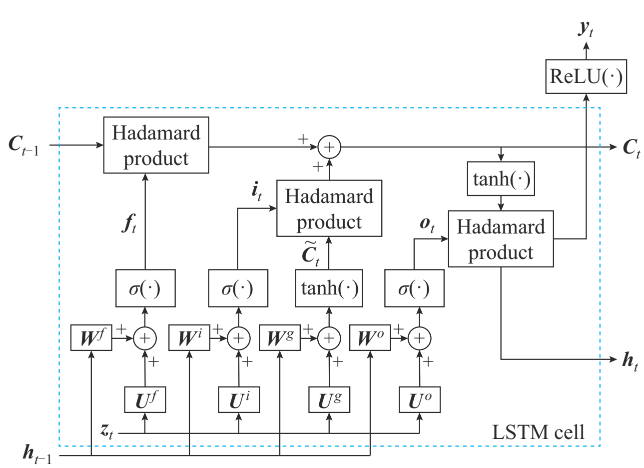

Fig. 2. Structure of LSTM cell.

Let z_t be the input of the neural network, which is formed by:

$$z_t = Concatenate(\boldsymbol{R}_t, \boldsymbol{X}_t) \tag{15}$$

where $Concatenate(\cdot)$ is the function to concatenate a list of inputs.

For the time sequence number in the LSTM model, we consider the model input as the observations over the past L time slots. The vector of observations can be denoted as $\boldsymbol{z}_{t-L,t} = [\boldsymbol{z}_{t-L}, \boldsymbol{z}_{t-L+1}, \ldots, \boldsymbol{z}_t]$. The LSTM network topology contains the fully self-connected hidden layers with the memory cells and related gate units installed. LSTM units implement the hidden layers, and all units have direct connections with each other. In Fig. 2, each LSTM cell is formed as the chain structure and covers four interacting layers with individual communication links, which are different from the traditional recurrent neural network (RNN) with a single neural network structure. The key function for each LSTM cell consists of an input gate, forget gate, and output gate. First of all, the input gate in the LSTM network at time t $\boldsymbol{i}_t$ can be expressed as:

$$\boldsymbol{i}_t = \sigma(z_t \boldsymbol{U}^i + \boldsymbol{h}_{t-1} \boldsymbol{W}^i) \tag{16}$$

where $\boldsymbol{U}^i$ and $\boldsymbol{W}^i$ are the coefficient and weighted matrices of the input gate, respectively; and $\boldsymbol{h}_{t-1}$ is the hidden state for the previous time slot. By using the sigmoid function $\sigma(\cdot)$, the non-linearity of the hidden layers is shown. The backup cell state of LSTM network at time t $\tilde{\boldsymbol{C}}_t$ can be denoted as:

$$\tilde{\boldsymbol{C}}_t = \tanh(z_t \boldsymbol{U}^g + \boldsymbol{h}_{t-1} \boldsymbol{W}^g) \tag{17}$$

where $\tanh(\cdot)$ is the hyperbolic tangent activation function; $\boldsymbol{U}^g$ and $\boldsymbol{W}^g$ are the coefficient and weighted matrices of the backup cell state, respectively. Then, the forget gate of LSTM network at time t $\boldsymbol{f}_t$ can be denoted as:

$$\boldsymbol{f}_t = \sigma(z_t \boldsymbol{U}^f + \boldsymbol{h}_{t-1} \boldsymbol{W}^f) \tag{18}$$

where $\boldsymbol{U}^f$ and $\boldsymbol{W}^f$ are the coefficient and weighted matrices of the forget gate, respectively. The function of the memory cell is to activate the forget gate to decide whether to delete the useless information from the previous time slot. The output gate of LSTM network at time t $\boldsymbol{o}_t$ can be expressed as:

$$\boldsymbol{o}_t = \sigma(z_t \boldsymbol{U}^o + \boldsymbol{h}_{t-1} \boldsymbol{W}^o) \tag{19}$$

where $\boldsymbol{U}^o$ and $\boldsymbol{W}^o$ are the coefficient and weighted matrices of the output gate, respectively. The function of the output gate is to prevent from storing long time lag memories. Besides, the hidden state of LSTM network at time t $\boldsymbol{h}_t$ can be expressed as:

$$\boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh(\boldsymbol{C}_t) \tag{20}$$

where $\odot$ denotes the Hadamard product operation. The cell state of LSTM network at time t $\boldsymbol{C}_t$ can be denoted as:

$$\boldsymbol{C}_t = \boldsymbol{f}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{C}}_t \tag{21}$$

where $\boldsymbol{C}_{t-1}$ denotes the cell state from the previous time slot, which can also be defined as the memory cell state.

C. BLSTM

The utilization of LSTM network has demonstrated that the point prediction can capture the long-term dependencies of the input dataset [23]. However, the standard LSTM network cannot capture the uncertainty. Hence, we propose the BLSTM network for probabilistic prediction tasks. Specifically, the BLSTM network involves in structure and parameter uncertainties. On one hand, the structure uncertainty refers to the way of choosing model structure to interpolate or extrapolate the selected data. On the other hand, the parameter uncertainty is associated with the set of weight parameters in the network to indicate the uncertain observations. For Bayesian neural networks, the prior distributions should be relevant to the distribution of the network parameters, e.g., weights and bias. The reason is that they are hard to be identified due to the unclear meanings of these parameters. According to the demonstrations in [24], deploying standard parametric distributions is effective when the prior is hard to be identified.

As mentioned above, in order to capture the model uncertainty, a prior distribution that each parameter is set as a standard normal distribution with zero mean can be placed over weight parameters. Note that the posterior distribution is used to produce the samples of forecasts after training the BLSTM network. We define $p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train})$ as the posterior distribution of the network weights with training data, where $\boldsymbol{W}$ is the weighted matrix for the network; and $\boldsymbol{Z}_{train}$ and $\boldsymbol{Y}_{train}$ are the training sets. According to Bayes's theorem, the posterior distribution can be expressed as:

$$p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train}) = \frac{p(\boldsymbol{Y}_{train}|\boldsymbol{Z}_{train}, \boldsymbol{W})p(\boldsymbol{W})}{p(\boldsymbol{Y}_{train}|\boldsymbol{Z}_{train})} \tag{22}$$

Then, we can have:

$$p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{Z}_{train}, \boldsymbol{Y}_{train}) = \int p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{W}) p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train}) \mathrm{d}\boldsymbol{W} \tag{23}$$

where $\boldsymbol{z}$ and $\boldsymbol{y}$ are the given input and output points, respectively.

For Bayesian deep learning, the network parameters are assumed to follow the posterior probability distributions. Hence, after the training process of the Bayesian neural network, the queries for the unseen data can be predicted by:

$$\begin{aligned} p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}) = E_{p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train})}(p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}, \boldsymbol{Z}_{train}, \boldsymbol{Y}_{train})) = \\ \int p(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}, \boldsymbol{Z}_{train}, \boldsymbol{Y}_{train}) p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train}) \mathrm{d}\boldsymbol{W} \end{aligned} \tag{24}$$

where $E(\cdot)$ is the expectation over the posterior probability distribution $p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train})$; and $\hat{}$ represents the prediction valnes. In this case, the Bayesian neural network is equivalent to taking the average predictions from an ensemble of neural networks weighted by the posterior probabilities of the parameters $\boldsymbol{W}$.

Note that the true posterior is usually intractable for the LSTM network. By considering the complexity of the posterior distribution of the network parameters, (24) is hard to be performed due to its intractable computation. Therefore, we introduce $q_\delta(\boldsymbol{W})$ with parameter δ as an approximating variational distribution to ensure the optimal distribution by minimizing the Kullback-Leibler (KL) divergence according to [25]. To solve this issue, the variational inference can be utilized to gain the latent parameters δ on $q_\delta(\boldsymbol{W})$ as:

$$\begin{aligned} \delta^* = \arg\min_{\delta} KL(q_\delta(\boldsymbol{W}) \| p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train})) = \\ \arg\min_{\delta} \int q_\delta(\boldsymbol{W}) \lg \frac{q_\delta(\boldsymbol{W})}{p(\boldsymbol{W})p(\boldsymbol{W}|\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train})} \mathrm{d}\boldsymbol{W} = \\ \arg\min_{\delta} [KL(q_\delta(\boldsymbol{W}) \| p(\boldsymbol{W}) - E_{q_\delta(\boldsymbol{W})}(\lg p(\boldsymbol{Z}_{train}, \boldsymbol{Y}_{train}|\boldsymbol{W}))] \end{aligned} \tag{25}$$

Hence, given the optimal distribution, the predictive distribution is calculated approximately by:

$$p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{Z}_{train}, \boldsymbol{Y}_{train}) = \int p(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{W}) q_\delta^*(\boldsymbol{W}) \mathrm{d}\boldsymbol{W} = q_\delta^*(\boldsymbol{y}|\boldsymbol{z}) \tag{26}$$

where $q_\delta^*(\boldsymbol{W})$ is the optimal distribution. Meanwhile, considering the dimension of the distribution of the parameters, we follow the assumption in [26].

By repeating the calculation with the sample weight T_s, the predictive mean of the samples can be calculated to present the model parameter uncertainty, which is shown as:

$$\exp(\boldsymbol{y}) = \frac{1}{T_s} \boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}) \tag{27}$$

where $\boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z})$ is the stochastic forward pass with its weight in the network. Besides, given that $\boldsymbol{W}_t \sim q_\delta^*(\boldsymbol{W})$ and $p(\boldsymbol{y}|\boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z})) = N(\boldsymbol{y}; \boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}), \sigma), \sigma > 0$, we have the following estimator:

$$\exp(\boldsymbol{y}^{\mathrm{T}} \boldsymbol{y}) = \frac{1}{T_s} \sum_{t=1}^{T_s} (\boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}))^{\mathrm{T}} \boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}) + \sigma \tag{28}$$

where σ is the data uncertainty and refers to the inherent noise in the data. The measurement of this value can be determined by the residual sum of squares on the independent validation sets of data. Then, the predictive variance can be obtained as:

$$Var(\boldsymbol{y}) = \frac{1}{T_s} \sum_{t=1}^{T_s} (\boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}))^{\mathrm{T}} \boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}) - \frac{1}{T_s^2} \sum_{t=1}^{T_s} (\boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}))^{\mathrm{T}} \sum_{t=1}^{T_s} \boldsymbol{f}^{\boldsymbol{W}_t}(\boldsymbol{z}) + \sigma^2 \tag{29}$$

Besides, the loss function in the BLSTM network can be defined as:

$$\mathcal{L}(\delta) = \frac{1}{T_{train}^2} \sum_{t=1}^{T_{train}} \frac{1}{2\sigma^2} \|y_i - f(z_i)\|^2 + \frac{1}{2} \lg \sigma^2 \tag{30}$$

To combine the data and model uncertainties, we define a new expression for the output of this model by considering the predictive mean and model precision as $[\hat{\boldsymbol{z}}, \hat{\sigma}^2] = \boldsymbol{f}_{BLSTM}^{\boldsymbol{W}_t}(\boldsymbol{z})$, where $\boldsymbol{f}_{BLSTM}^{\boldsymbol{W}_t}(\boldsymbol{z})$ is the proposed BLSTM network that is parameterized by $\boldsymbol{W}_t \sim q_\delta^*(\boldsymbol{W})$. Thus, the final loss function can be expressed as:

$$\mathcal{L}_{BLSTM}(\delta) = \frac{1}{T_{train}^2} \sum_{t=1}^{T_{train}} \frac{1}{2\sigma^2} \|y_i - \hat{y}_i\|^2 + \frac{1}{2} \lg \hat{\sigma}_i^2 \tag{31}$$

Last but not least, the confidence interval (CI) can be calculated as:

$$CI = [\hat{\mu} - t_{\beta/2}\hat{\sigma}, \ \hat{\mu} + t_{\beta/2}\hat{\sigma}] \tag{32}$$

where μ is the expection value; and $t_{\beta/2}$ is the t-score in the table of t-distribution.

D. Training Algorithm

In the proposed BLSTM network, we train the network by minimizing the loss function in (30) to adjust the predicted results. Also, the BLSTM network is capable of capturing the uncertainty of the input data. Considering the process of training BLSTM network, we incorporate the Adam optimizer [27] into the training algorithm for the LSTM network. In addition, the LSTM cell in the network is designed so that the updating complexity per time interval and weight does not depend on the size of the neural network, and the storage does not rely on the sequence length of the input data [28]. Besides, the historical data, known as the input data, can be regarded as the prior information for input training. When we apply the BLSTM network for the practical scenario, we should pre-train the network under different power system operations. Thus, the state of the general power system can be predicted through our proposed BLSTM network. Consequently, the training of BLSTM network can be summarized as follows.

1) Collect the power system data, and separate the data into training and testing datasets.

2) Normalize the input data and prepare the sample data ζ with normal distribution.

3) Initialize network parameters: $\delta = \mu + \sigma\zeta$.

4) Perform forward and back propagations on the batch.

5) Update μ and σ according to the gradients in the network respect to δ.

6) Fine tune the whole network with multiple trials and output the predicted result.

V. Performance Evaluation

This section shows the performance evaluation of the proposed methodology. First of all, we introduce the simulation setup. Then, we present the scenarios for comparison and performance metrics for the simulation. Finally, we present and discuss the simulation results in a detailed manner.

A. Simulation Setup

In the simulation, we evaluate the performance of the proposed model with the historical power system data. We employ the IEEE 57-bus system as the testing power system [29]. Real power system data from [30] are adopted in the subsequent case studies. In particular, 13-month data from December 2015 to December 2016 are obtained, which are divided into a training dataset (the first twelve months) and a testing dataset (the remaining month) for cross-validation. The sampling period of all profiles is aggregated into 15 min. The historical data cover multiple entries, including solar generation, wind generation, and time-varying household load. We scale the time-varying household load to fit the particular IEEE 57-bus test system. In addition, the solar and wind generators are installed at buses 13 and 37, respectively. According to the settings of the test system [29], the installed capacities for these two renewable power generators are set to be 150 MW and 70 MW, respectively.

For the purpose of simulation, the measurements of the system data are used to obtain the actual values of voltages, phase angles, and power by the Newton-Raphson power flow model. Then, we design the BLSTM network as follows. We empirically set the sequence length L (in time slots) to be 4. The epoch number is 150 and the batch size is set to be 64. The numbers of the three hidden units for the LSTM network are set to be 64, 128, and 256, respectively. The dropout rate is 0.5. The value of β is set to be 5% to investigate the 95% confidence interval for the proposed BLSTM network. The number of samples T_s is set to be 100. All the tested algorithms are implemented with Python and PyTorch [31].

B. Scenarios for Comparison

The evaluation of our proposed BLSTM network is based on the comparison with other typical prediction techniques. In this paper, we introduce six widely-adopted techniques for baseline comparisons, including multiple linear regression (MLR) [32], ANN [11], LSTM [16], unscented Kalman filter (UKF) [2], quantile regression (QR) [33], and quantile random forest (QRF) [34]. Indeed, these tools have already been widely used for the application in the power system. The former three techniques belong to point prediction techniques while the latter three refer to probabilistic models. Our proposed BLSTM and LSTM network can capture both the data and model uncertainties simultaneously. The baseline techniques are fine-tuned for optimal parameter configurations to produce the best prediction results.

C. Performance Metrics

To assess the prediction accuracy of the proposed BLSTM network, four performance metrics are employed, which are shown as:

$$MSE = \frac{1}{N_T}\sum_{t\in T}(y(t) - \hat{y}(t))^2 \tag{33}$$

$$RMSE = \sqrt{\frac{1}{N_T}\sum_{t\in T}(y(t) - \hat{y}(t))^2} \tag{34}$$

$$MAE = \frac{1}{N_T}\sum_{t\in T}\left| y(t) - \hat{y}(t) \right| \tag{35}$$

$$MAPE = \frac{1}{N_T}\sum_{t\in T}\left| \frac{y(t) - \hat{y}(t)}{y(t)} \right| \times 100\% \tag{36}$$

where $y(t)$ and $\hat{y}(t)$ are the actual and predicted values at time t, respectively; and N_T is the number of the sampling period. These metrics are used to compare the predicted value with the actual value for point prediction, namely, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). For these metrics, a smaller value indicates a better performance in the prediction.

D. Simulation Result

1) Comparison of Different Prediction Methods

As previously introduced in Section V-B, we evaluate the proposed BLSTM network compared with the other six baseline methods. Table I presents the four performance metrics in evaluating these prediction methods. It is apparent that our proposed BLSTM model outperforms the rest of the prediction methods with the smallest values in MSE, RMSE, MAE, and MAPE. Even if both UKF and QRF can have relatively low MSE and RMSE, MAE and MAPE are both much higher than BLSTM, which indicate larger errors. In addition, we can observe that since BLSTM network considers both the model and data uncertainties, the predicted result can be better even if for the point estimation. Owing to the characteristics of capturing long-term dependencies of time-series input data on LSTM network, the proposed BLSTM network can help further enhance the prediction accuracy of the output results. The time complexity of the proposed BLSTM thus only depends on the number of sampling period N_T and the number of features N_F. Hence, the time complexity of BLSTM is defined as $\mathcal{O}(N_T N_F)$. It is apparent that a lower time complexity reflects a more effective approach.

TABLE I
COMPARISON OF DIFFERENT PREDICTION TECHNIQUES IN IEEE 57-BUS SYSTEM

Method	MSE	RMSE	MAE	MAPE (%)
MLR	2.100×10^{-3}	0.0460	0.0320	467.1
ANN	2.820×10^{-4}	0.0170	0.0110	163.2
LSTM	1.870×10^{-5}	0.0043	0.0076	87.4
UKF	1.490×10^{-5}	0.0039	0.0034	35.2
QR	6.060×10^{-5}	0.0078	0.0125	28.0
QRF	1.194×10^{-5}	0.0035	0.0029	39.2
BLSTM	5.660×10^{-6}	0.0023	0.0017	24.9

2) Probabilistic Prediction of Net Active Power Imbalance

In this part, we investigate the net active power imbalance prediction for the entire month using the BLSTM network. By considering both the model and data uncertainties, the normalized prediction results can be obtained as shown in Fig. 3. By comparing with the actual values, it is apparent that the predicted results can better fit the trend of the actual profile. According to Table I, we can demonstrate that our proposed BLSTM network can outperform baseline techniques, which also yields the lowest errors through the four evaluating metrics.

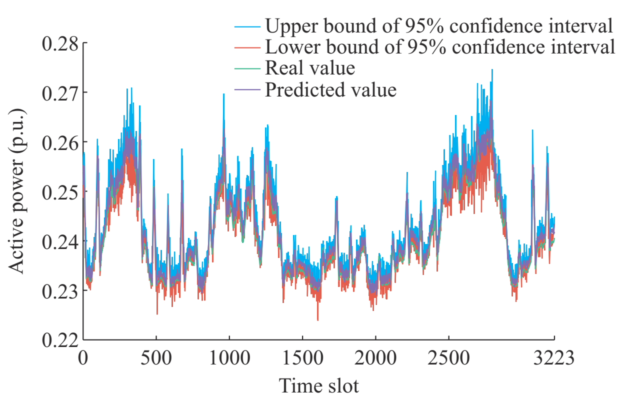

Fig. 3. Probabilistic prediction of net active power imbalance.

3) Probabilistic Prediction of Bus Voltage

In this part, we further focus on the monthly dynamic state variables of each common bus. The relatively large-scale solar power generation is integrated at bus 13. Meanwhile, we capture both the phase angle and the voltage magnitude of this bus. The normalized results are presented in Figs. 4 and 5. In Fig. 4, it is obvious that the lower and upper bounds of 95% confidence interval can better fit the trend of the real values of the phase angle at bus 13. Besides, in Fig. 5, most of the predicted values are close to the real ones of the voltage magnitude at bus 13. In addition, the similar results can also be demonstrated at bus 37, which is integrated with wind power generation. Furthermore, by observing Figs. 4 and 5, although the predicted values cannot fit well with the sudden fluctuations of the real values, the occurrence of the confidence bounds of the BLSTM network show the effectiveness of capturing such features.

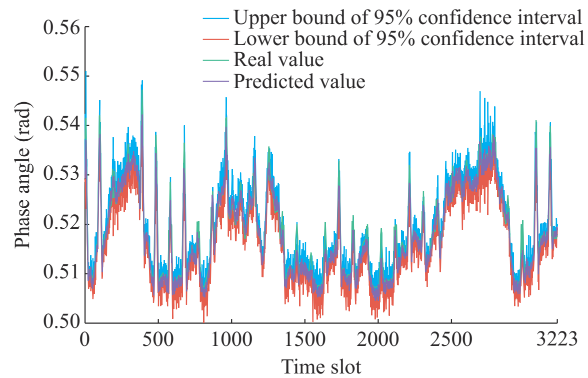

Fig. 4. Probabilistic prediction of phase angle at bus 13.

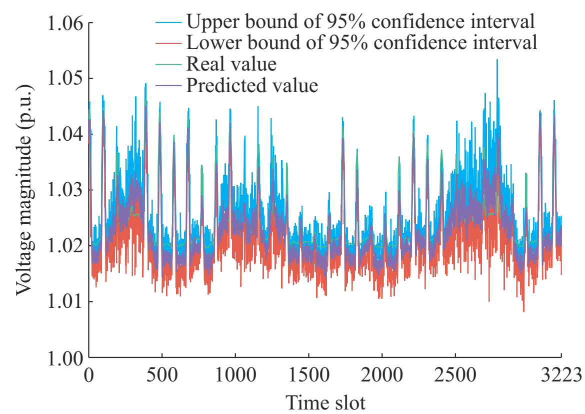

Fig. 5. Probabilistic prediction of voltage magnitude at bus 13.

4) Tractability for Large Power System

In this part, we investigate the numerical performance of the proposed BLSTM network in IEEE 118-bus system. Table II summarizes the prediction results by adopting different techniques in IEEE 118-bus system. Considering the change of test case, all tools are fine-tuned to obtain the best parameters to give the best prediction results. By comparing with these techniques, it is apparent that our proposed BLSTM network still obtain the lowest MSE, RMSE, MAE, and MAPE values, indicating the highest prediction accuracy. Besides, with the increase of the system complexity, the model and data uncertainties may rise for the six baseline techniques. However, according to the results in Table II, the accurate prediction results can still be obtained by means of BLSTM network, which shows the best performance.

TABLE II
COMPARISON OF DIFFERENT PREDICTION TECHNIQUES IN IEEE 118-BUS SYSTEM

Method	MSE	RMSE	MAE	MAPE (%)
MLR	9.20×10^{-3}	0.0960	0.05800	484.3
ANN	2.65×10^{-5}	0.0051	0.00440	36.1
LSTM	2.82×10^{-5}	0.0054	0.00210	22.8
UKF	2.90×10^{-5}	0.0054	0.00190	18.7
QR	7.32×10^{-5}	0.0086	0.00320	48.1
QRF	1.60×10^{-5}	0.0040	0.00150	11.6
BLSTM	7.45×10^{-6}	0.0027	0.00085	7.3

5) *Robustness of Proposed BLSTM Network*

We further evaluate the robustness of the proposed BLSTM network. In this part, we utilize two different systems, IEEE 57-bus and 118-bus systems, to investigate the robustness of the proposed BLSTM network. Note that, the input datasets are different between the two systems because the historical dynamic system states are generated based on the structure information of the test systems. The results are presented in Fig. 6. It is obvious that the loss curve of BLSTM in IEEE 57-bus system converges around 18 epochs. Meanwhile, the loss curve of IEEE 118-bus system converges around 27 epochs. The results indicate the effectiveness of proposed BLSTM network for different systems.

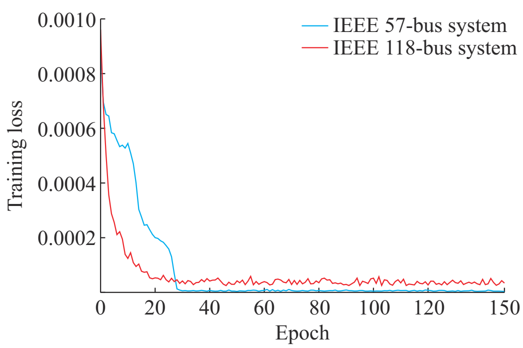

Fig. 6. Training loss of proposed BLSTM network for different systems.

VI. CONCLUSION

In this paper, we propose a Bayesian deep learning approach to predict the dynamic system states in general power systems. First of all, the model respects the implementations of renewable generation and household loads in general systems through data pre-processing. Then, by the Newton-Raphson power flow model, the system state matrix is developed. After combining multiple system features as a final complete input dataset, we train and validate the BLSTM network to generate accurate prediction results. In addition, we capture both the data and model uncertainties in the proposed model. The simulation results indicate that the accurate prediction can be obtained through our proposed BLSTM network for different scales of power systems. Future work will consider exploring a more complex deep learning model for a practical power system to further capture the potential features of such system. In addition, the proposed Bayesian deep learning model can be extended to the application of other related research fields in the power system.

REFERENCES

[1] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3205-3216, Jul. 2017.
[2] J. Zhao, M. Netco, and L. Mili, "A robust generalized-maximum likelihood unscented Kalman filter for power system dynamic state estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 4, pp. 578-592, Aug. 2018.
[3] H. Livani, S. Jafarzadeh, C. Y. Evrenosoğlu *et al*., "A unified approach for power system predictive operations using Viterbi algorithm," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 3, pp. 757-766, Jul. 2014.
[4] X. Cao, B. Stephen, I. F. Abdulhadi *et al*., "Switching Markov Gaussian models for dynamic power system inertia estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3394-3403, Sept. 2016.
[5] Y. Zhang, J. Wang, and Z. Li, "Interval state estimation with uncertainty of distributed generation and line parameters in unbalanced distribution systems," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 762-772, Jan. 2020.
[6] J. Lv, M. Pawlak, and U. D. Annakkage, "Prediction of the transient stability boundary based on nonparametric additive modeling," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4362-4369, Nov. 2017.
[7] N. I. Sapankevych and R. Sankar, "Time series prediction using support vector machines: a survey," *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24-38, May 2009.
[8] M. H. Kapourchali, M. Sepehry, and V. Aravinthan, "Multivariate spatio-temporal solar generation forecasting: a unified approach to deal with communication failure and invisible sites," *IEEE Systems Journal*, vol. 13, no. 2, pp. 1804-1812, Jun. 2019.
[9] T. Wang, T. Bi, H. Wang *et al*., "Decision tree based online stability assessment scheme for power systems with renewable generations," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 2, pp. 53-61, Jun. 2015.
[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, May 2015.
[11] D. Q. Zhou, U. D. Annakkage, and A. D. Rajapakse, "Online monitoring of voltage stability margin using an artificial neural network," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1566-1574, Aug. 2010.
[12] Y. Xu, R. Zhang, J. Zhao *et al*., "Assessing short-term voltage stability of electric power systems by a hierarchical intelligent system," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1686-1696, Aug. 2016.
[13] L. Zhang, G. Wang, and G. B. Giannakis, "Real-time power system state estimation and forecasting via deep unrolled neural networks," *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4069-4077, Aug. 2019.
[14] C. Zheng, S. Wang, Y. Liu *et al*., "A novel equivalent model of active distribution networks based on LSTM," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2611-2624, Sept. 2019.
[15] A. Gensler, J. Henze, B. Sick *et al*., "Deep learning for solar power forecasting – an approach using autoencoder and LSTM neural networks," in *Proceedings of 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Budapest, Hungary, Oct. 2016, pp. 2858-2865.
[16] Z. Cao, Y. Wang, C. Chu *et al*., "Scalable distribution systems state estimation using long short-term memory networks as surrogates," *IEEE Access*, vol. 8, pp. 23359-23368, Jan. 2020.
[17] C. Peng, S. Lei, Y. Hou *et al*., "Uncertainty management in power system operation," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 1, pp. 23359-23368, Mar. 2015.
[18] M. Ghosal and V. Rao, "Fusion of multirate measurements for nonlinear dynamic state estimation of the power systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 216-226, Jan. 2019.
[19] M. Sun, T. Zhang, Y. Wang *et al*., "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 188-201, Jan. 2020.
[20] K. R. Mestav, J. Luengo-Rozas, and L. Tong, "Bayesian state estimation for unobservable distribution systems via deep learning," *IEEE Transactions on Power Systems*, vol. 34, no. 6, pp. 4910-4920, Nov.

2019.
[21] H. L. Nguyen, "Newton-Raphson method in complex form power system load flow analysis," *IEEE Transactions on Power Systems*, vol. 12, no. 3, pp. 1355-1359, Aug. 1997.
[22] A. Wood and B. Wollenberg, *Power Generation, Operation, and Control*. New Jersey: Wiley-Interscience, 2014.
[23] W. Kong, Z. Y. Dong, Y. Jia *et al*., "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, Jan. 2019.
[24] Y. Gal, "Uncertainty in deep learning," Ph.D. dissertation, University of Cambridge, Cambridge, 2016.
[25] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79-86, Mar. 1951.
[26] M. Gabrié, "Mean-field inference methods for neural networks," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 22, pp. 1-74, May 2020.
[27] D. Kingma and J. Ba. (2014, Dec.). Adam: a method for stochastic optimization. [Online]. Available: https://arxiv.org/abs/1412.6980v8
[28] J. Schmidhuber, "A local learning algorithm for dynamic feedforward and recurrent networks," *Connection Science*, vol. 1, pp. 403-412, Sept. 1995.
[29] UWEE. (2020, Dec.). Power systems test case archive. [Online]. Available: http://www.ee.washington.edu/research/pstca/
[30] Open Power System Data. (2020, Dec.). Data platform: time series, open power system data. [Online]. Available: http://data.open-power-system-data.org/
[31] A. Paszke, S. Gross, S. Chintala *et al*., "Automatic differentiation in pytorch," in *Proceedings of the NIPS 2017 Workshop, Co-located with the 31st Annual Conference on Neural Information Processing Systems (NIPS 2017)*, Long Beach, United States, Dec. 2017, pp. 1-4.
[32] A. C. S. Hettiarachchige-Don and V. Aravinthan, "Estimation of missing transmission line reactance data using multiple linear regression," in *Proceedings of 2017 North American Power Symposium (NAPS)*, Morgantown, USA, Sept. 2017, pp. 1-6.
[33] S. M. Mazhari, N. Safari, C. Y. Chung *et al*., "A quantile regression-based approach for online probabilistic prediction of unstable groups of coherent generators in power systems," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2240-2250, May 2019.
[34] L. Alfieri and P. De Falco, "Wavelet-based decompositions in probabilistic load forecasting," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1367-1376, Mar. 2020.

Shiyao Zhang received the B.S. degree in electrical and computer engineering from the Purdue University, West Lafayette, USA, in 2014, the M.S. degree in electrical engineering from University of Southern California, Los Angeles, USA, in 2016, and the Ph.D. degree from the University of Hong Kong, Hong Kong, China. He is currently a Post-doctoral Research Fellow with the Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, China. His research interests include smart energy systems, intelligent transportation systems, optimization theory and algorithms, and deep learning applications.

James J. Q. Yu received the B.Eng. and Ph.D. degrees in electrical and electronic engineering from the University of Hong Kong, Hong Kong, China, in 2011 and 2015, respectively. He was a Post-doctoral Fellow at the University of Hong Kong from 2015 to 2018. He is an Assistant Professor at the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China, and an Honorary Assistant Professor at the Department of Electrical and Electronic Engineering, the University of Hong Kong. He currently also serves as the Chief Research Consultant of GWGrid Inc., Zhuhai, China, and Fano Labs, Hong Kong, China. His research interests include smart city and urban computing, deep learning, intelligent transportation systems, smart energy systems, forecasting and decision-making of future transportation systems, and basic artificial intelligence techniques for industrial applications.