Downloaded from orbit.dtu.dk on: Nov 07, 2026

DTU Library

Probabilistic forecasting of wind power generation using extreme learning machine.

Wan, Can; Xu, Zhao; Pinson, Pierre; Dong, Zhao Yang; Wong, Kit Po

Published in:
IEEE Transactions on Power Systems

Link to article, DOI:
10.1109/TPWRS.2013.2287871

Publication date:
2014

Link back to DTU Orbit

Citation (APA):
Wan, C., Xu, Z., Pinson, P., Dong, Z. Y., & Wong, K. P. (2014). Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, *29*(3). https://doi.org/10.1109/TPWRS.2013.2287871

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Probabilistic Forecasting of Wind Power Generation Using Extreme Learning Machine

Can Wan, *Student Member, IEEE*, Zhao Xu, *Senior Member, IEEE*, Pierre Pinson, *Senior Member, IEEE*, Zhao Yang Dong, *Senior Member, IEEE*, and Kit Po Wong, *Fellow, IEEE*

(Invited Paper)

***Abstract*—Accurate and reliable forecast of wind power is essential to power system operation and control. However, due to the nonstationarity of wind power series, traditional point forecasting can hardly be accurate, leading to increased uncertainties and risks for system operation. This paper proposes an extreme learning machine (ELM)-based probabilistic forecasting method for wind power generation. To account for the uncertainties in the forecasting results, several bootstrap methods have been compared for modeling the regression uncertainty, based on which the pairs bootstrap method is identified with the best performance. Consequently, a new method for prediction intervals formulation based on the ELM and the pairs bootstrap is developed. Wind power forecasting has been conducted in different seasons using the proposed approach with the historical wind power time series as the inputs alone. The results demonstrate that the proposed method is effective for probabilistic forecasting of wind power generation with a high potential for practical applications in power systems.**

***Index Terms*—Bootstrap, extreme learning machine (ELM), forecasting, prediction interval, wind power.**

NOMENCLATURE

$b, \hat{b}$	Biases of ELM, approximated biases.
B	Number of bootstrap replications.
B_M	Number of bootstrap replications for model uncertainty estimation.
B_N	Number of bootstrap replications for noise variance estimation.
c	Indicator of predication intervals coverage probability.
C_{BP}	Cost function of traditional neural networks.
D_t	Dataset used for training the models of uncertainty estimation.
D_ε	Dataset used for training the models of noise variance estimation.
$E[\,]$	Expectation.
$f(\,)$	Output function of ELM.
$\mathrm{g}(\,)$	Activation function of ELM.
$h_\varepsilon(\,)$	Output function of ELM for noise variance estimation.
$\mathbf{H}$	Hidden layer output matrix of ELM.
$\mathbf{H}^\dagger$	Moore-Penrose generalized inverse of ELM's hidden layer output matrix.
i, j, l	Common indices.
I_M^α	Model uncertainty interval.
I_t^α	Prediction interval.
K	Number of ELM's hidden nodes.
L_M^α	Lower bound of model uncertainty interval.
L_t^α	Lower bound of prediction interval.
m	Dimension of the output vector of ELM.
n	Dimension of the input vector of ELM.
N	Number of training samples.
$N_G(\,)$	Gaussian distribution.
N_{test}	Number of test samples.
$P(\,)$	Probability operator.
$\hat{r}(\,)$	Output of the ELM for noise variance estimation.
Sc_t^α	Score for predication intervals evaluation.
$\mathbf{T}, \mathbf{t}, t$	Regression model outputs/targets.
$U(\,)$	Uniform distribution.

Manuscript received October 07, 2012; revised March 03, 2013, June 05, 2013, August 20, 2013; accepted October 18, 2013. This work was supported in part by Hong Kong RGC GRF under Grants 515110, 528412, and T23-407/13-N. The work of C. Wan was supported by a Hong Kong Ph.D. Fellowship. Paper no. T-TST-SPI-06-2013-00075.

C. Wan is with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, and also with the Centre for Electric Power and Energy (CEE), Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (e-mail: can.wan@connect.polyu.hk).

Z. Xu is with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, and also with The University of Newcastle, Callaghan NSW 2308, Australia (e-mail: eezhaoxu@polyu.edu.hk).

P. Pinson is with the Centre for Electric Power and Energy (CEE), Technical University of Denmark, 2800 Kgs. Lyngby, Denmark (e-mail: ppin@elektro.dtu.dk).

Z. Y. Dong is with School of Electrical and Information Engineering, The University of Sydney, NSW 2006, Australia (e-mail: zydong@ieee.org).

K. P. Wong is with School of Electrical, Electronic and Computer Engineering, The University of Western Australia, WA 6009, Australia (e-mail: kitpo@ieee.org).

Digital Object Identifier 10.1109/TPWRS.2013.2287871

0885-8950 © 2013 IEEE

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

U_M^α	Upper bound of model uncertainty interval.
U_t^α	Upper bound of prediction interval.
$\mathbf{w}, \hat{\mathbf{w}}$	Input weights of ELM, approximated input weights.
$\mathbf{x}$	Regression model input variables.
$y(\,)$	True regression.
$\hat{y}(\,)$	Estimation of true regression.
$z_{1-\alpha/2}$	Critical value of standard Gaussian distribution.
$(1-\alpha)$	Nominal coverage probability of predication intervals.
$\beta, \hat{\beta}$	Output weights of ELM, approximated output weights.
ε	Regression noise.
σ_t^2	Variance of total prediction errors.
$\sigma_{\hat{y}}^2$	Variance of model misspecification uncertainty.
σ_ε^2	Variance of regression noise.
$\hat{\sigma}_\varepsilon^2$	Estimated noise variance.
$\sigma_{\hat{r}}^2$	Variance of model uncertainty for noise variance estimation.
δ_t^α	Width of prediction interval.

I. Introduction

WIND energy is considered to be the most efficient renewable energy source for electricity generation in modern power systems. In the past decades, wind power has experienced a rapid growth worldwide. In particular, wind power can supply up to 20% of the electricity consumption in Denmark. Wind power brings many more uncertainties than conventional generation. Accurate and reliable wind power forecasting becomes extremely important to optimize the operation cost and improve the reliability of the power system with increased wind penetration. Many industrial wind power forecasting systems have been developed and applied, including Wind Power Prediction Tool (WPPT), Previento, Sipreólico, Armines Wind Power Prediction System (AWPPS), Prediktor, and Zephyr [1].

Traditionally, most research has focused on developing accurate point forecasting methods for wind power [2], [3]. Due to the chaotic nature of the weather system, errors in wind power forecasting are simply unavoidable and quite often can be significant. Wind power forecasting error is statistically analyzed and is modeled by Beta distribution in [4]. Recently, different approaches for probabilistic wind power generation forecasts have been developed to construct prediction intervals (PIs) quantifying the forecasting uncertainty. The quantile regression can be used to estimate different wind power forecasting quantiles [5], [6]. Wind power point forecasts and probability density function of associated uncertainty are obtained from weather ensemble predictions generated by an atmospheric model incorporating calibration and kernel smoothing approaches in [7]. The novel Nadaraya–Watson estimator [8] and time-adaptive quantile-copula estimator [9] are developed for kernel density forecasts of wind power generation. Exponential smoothing method (ESM) is proposed for multistep density forecasts of wind power, with efficient and satisfactory performance [10]. With nonparametric probability forecasts of wind power, the statistical scenarios of short-term wind power can be generated [11]. Based on the point forecasting results of, e.g., AWPPS, WPPT, and Sipreólico, PIs are constructed through a combined nonparametric probability forecasts and adaptive resampling approach [12]. In [13], radial basis function has been implemented to derive quantile forecasts of wind power based on point prediction results, weather conditions, etc. With high penetration of wind power, the knowledge of uncertainties ahead can be extremely valuable to a number of power system operation and management procedures, including but not limited to, optimal operation reserve determination [14], [15], system steady-state security assessment [15], economic generation scheduling and dispatch [16], unit commitment [16], [17], electricity market trading [18], and so forth. Based on the PIs with associated confidence level, the quantified uncertainties of wind power forecasts can provide useful information to decision makers to well prepare for the worst and the best conditions ahead.

In this paper, a new probabilistic wind power forecasting approach is proposed based on the extreme learning machine (ELM), which is a novel learning algorithm proposed for training single-hidden layer feedforward neural networks (SLFNs). It randomly chooses the input weights of hidden layer neurons and analytically determines the output weights through simple matrix computations, therefore featuring an extremely faster learning speed than for most popular learning algorithms such as Back-propagation [19]. ELM has also demonstrated excellent generalization capability and outperformed traditional NNs. In practice, ELM has been used in many different applications, including both regression and classification tasks [20]–[22].

Due to the excellent approximation and generalization capabilities, neural networks (NNs) are widely used for wind power forecasts [23]–[25], irrespective of some drawbacks like local minima, overtraining, and high computational costs. Generally speaking, NN-based forecasting methods cannot provide satisfactory predictions if the training data are chaotic or too noisy. Usually the prediction performance cannot be improved by changing the NN structure or increasing the training iteration. To effectively account for forecasting uncertainties, several approaches have been developed to obtain PIs for NN based methods, including delta, Bayesian, bootstrap, and mean-variance estimation methods [26]–[29]. Comparing with other approaches, the bootstrap approach is able to flexibly approximate the non-constant variance and heterogeneous noises thus providing reliable performance [30], [31], which is used for wind power forecasting recently [32]. In addition, it avoids the calculations of complicated derivatives and the Hessian matrix involved in delta and Bayesian methods [30]. However, due to the limitations of traditional NNs, the bootstrap approach suffers from significantly high computational

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

burden, especially for large datasets. Furthermore, the bootstrap technique for traditional NNs cannot be applicable to the case of ELM, since the associated learning process is very different from that for conventional NN learning algorithms. Therefore, a bootstrap-based ELM approach (BELM) is newly developed to construct PIs taking the heteroscedasticity of wind power time series into account. The proposed BELM method can rapidly formulate the PIs through extremely fast learning by ELM. Notably, though with high extendibility, the work in this paper focuses on a simplified approach with fast speed, using the historical wind power data alone while providing satisfactory performance for hourly ahead and intra-hour forecasting, which is significant for power system operation and control in practice. For instance, in the Nord Pool market in Scandinavia, the hourly market plays a key role in maintaining system balance.

The proposed BELM method has been tested using the measurement data of a wind farm in Australia. The reliability and overall skill of the forecasting results have been comprehensively evaluated to demonstrate the effectiveness of the proposed method. The proposed approach gives a general framework for short-term probabilistic wind power forecasting. With high reliability, efficiency and flexibility, the proposed approach can potentially provide an online tool for power system operation and planning, e.g., to assist Transmission System Operator (TSO) in determining the required reserves in advance to avoid either higher costs or excessive risks under traditional deterministic reserve dispatch and to help suppliers manage risks facing in electricity market trading through strategic bidding.

II. ELM and Prediction Interval Formulation

A. Extreme Learning Machine

ELM is simply a single hidden-layer feedforward neural network [19], [21], [22]. Instead of using traditional gradient-based learning methods that can involve many iterations, the ELM randomly chooses the input weights and biases and subsequently determines the output weights through simple matrix computations. Given N arbitrary distinct samples $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbf{R}^n$ with $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T$ and $\mathbf{t}_i \in \mathbf{R}^m$ with $\mathbf{t}_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$, ELM with K hidden nodes and activation function $\mathrm{g}(\cdot)$ can be mathematically modeled as

$$f(\mathbf{x}_j; \mathbf{w}, b, \beta) = \sum_{i=1}^{K} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i), \quad j = 1, \ldots, N \tag{1}$$

where $\mathbf{w}_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ is the weight vector connecting the ith hidden node and the input nodes and $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is the weight vector connecting the ith hidden node and the output nodes, and b_i is the threshold of the ith hidden node.

The standard ELM with K hidden nodes and activation function $\mathrm{g}(\cdot)$ can approximate the N samples with zero error, meaning that

$$\sum_{i=1}^{K} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j = 1, \ldots, N. \tag{2}$$

The above N equations can be rewritten as

$$\mathbf{H}\beta = \mathbf{T} \tag{3}$$

where $\mathbf{H}$ is the hidden-layer output matrix of the ELM

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_K \cdot \mathbf{x}_1 + b_K) \\ \vdots & \cdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_K \cdot \mathbf{x}_N + b_K) \end{bmatrix}_{N \times K}. \tag{4}$$

$\beta = [\beta_1, \beta_2, \ldots, \beta_K]^T$ denotes the matrix of output weights, and $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_m]^T$ denotes the matrix of targets.

The input weights $\mathbf{w}_i$ and the hidden-layer biases b_i are randomly generated using continuous probability distributions and are, in fact, not necessarily tuned. The hidden-layer output matrix $\mathbf{H}$ can actually remain unchanged once random values have been assigned to these parameters in the beginning of learning. Find specific parameters $\hat{\mathbf{w}}_i, \hat{b}_i$ and$\hat{\beta}$, such that

$$\begin{aligned} &\|\mathbf{H}(\hat{\mathbf{w}}_1, \ldots, \hat{\mathbf{w}}_K, \hat{b}_1, \ldots, \hat{b}_K)\hat{\beta} - \mathbf{T}\| \\ &\quad = \min \|\mathbf{H}(\mathbf{w}_1, \ldots, \mathbf{w}_K, b_1, \ldots, b_K)\beta - \mathbf{T}\| \end{aligned} \tag{5}$$

which is equivalent to minimizing the cost function of the traditional gradient-based learning algorithms used in back-propagation (BP) learning

$$C_{\mathrm{BP}} = \sum_{j=1}^{N} \left[\sum_{i=1}^{K} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) - t_j \right]^2. \tag{6}$$

Given that the input weights and the hidden-layer biases are randomly assigned and fixed, training an SLFN is simply equivalent to finding a least-squares solution of the linear system. The smallest norm least-squares solution of the above linear system is

$$\hat{\beta} = \mathbf{H}^{\dagger}\mathbf{T} \tag{7}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$. The singular value decomposition (SVD) method is generally used to obtain $\mathbf{H}^{\dagger}$.

The advantages of the ELM algorithm are significant [19], [21], [22]. Without iterative gradient-based training, it avoids many limitations of conventional gradient-based NN training algorithms, such as the local minima, the overtraining, and the high computing burdens. For any infinitely differentiable activation function, the ELM with N hidden-layer neurons can learn N distinct samples exactly with zero error. In addition, ELM training can always guarantee the best results according to the assigned input weights. The training speed is extremely fast due to the simple matrix operation in (7). ELM also distinguishes from traditional NNs in superior generalization capability without the overtraining issue.

B. Uncertainties in Forecasting

The uncertainty of NN-based prediction is mainly due to the noise of training data and the misspecification of NN model for regression.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

1) Uncertainty in NN Model: Misspecifications in model structure and parameters account for the uncertainty of neural network forecasting, which may be caused by the local minima in the training process, the randomly generated input weights, and so on. In addition, even if the global minimum can be reached, the misspecification of model structure also introduces non-negligible uncertainties in prediction results. The model uncertainty also comes from another fact that training based on finite samples can never guarantee consistent generalization performance of NN for the unseen future. Particularly, in the study of wind power forecasting herein, it is impossible to find perfect information to reduce uncertainties of predictions. These factors are collectively termed as model uncertainty. Because of the model uncertainty, the output uncertainty of neural networks should be well addressed in order to produce accurate estimation.

2) Uncertainty in Data: Except for the model uncertainty, the data noise also contributes to the prediction uncertainty. If the data exhibit stochastic characteristics, it is extremely difficult to model them in a deterministic manner. Especially, when dealing with nonstationary time series, the data noise has significant influences on the prediction results. In the study, wind power data is highly chaotic. Determining the variance of the data noise is critical in constructing prediction intervals.

Both model misspecification and data noise are the major sources of uncertainties that affect the forecasting results. Therefore, the main task of probabilistic forecasting is to quantify the prediction intervals with associated confidences taking the two uncertainties into account.

C. Prediction Interval

Theoretically, multilayered feedforward neural networks are universal approximators and, as such, have an excellent ability to approximate any nonlinear mapping to any degree of accuracy [33]. In this paper, the SLFN-based ELM is applied for the regression task to estimate the underlying mathematical relationship between input and output variables based on a finite set of training data possibly corrupted by noises. Given a set of distinct pairs $\{(\mathbf{x}_i, t_i)\}$, the measured data can be modeled by

$$t_i = f(\mathbf{x}_i; \mathbf{w}, b, \beta) + \varepsilon(\mathbf{x}_i) = y(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i) \tag{8}$$

where t_i is the ith measured target, $\mathbf{x}_i$ denotes relevant input variables that can include historical wind power and wind speed, numerical weather predictions (NWPs) including wind speed and wind direction, and so forth for wind power forecasts, $\varepsilon(\mathbf{x}_i)$ denotes the noise with zero mean, and $y(\mathbf{x}_i) = f(\mathbf{x}_i; \mathbf{w}, b, \beta)$ is the true regression mean. The error term moves the target away from its true regression mean $y(\mathbf{x}_i)$ toward the measured value t_i. We assume that the noise is more or less Gaussian distributed with variance σ_ε^2 that may depend on the input vector $\mathbf{x}_i$, i.e.,

$$\varepsilon(\mathbf{x}_i) \sim N_G\left(0, \sigma_\varepsilon^2(\mathbf{x}_i)\right). \tag{9}$$

Actually, in the study, the censored Gaussian distribution is used to model the wind power prediction uncertainty, with potential concentration of probability mass at the bounds of the unit interval [0, 1], which maintains PIs within the wind power capacity range [34]. To some extent, the censored Gaussian distribution can fit different skewnesses, i.e., different shapes of probability distributions [35]. In addition, it has been studied in [36] that, even if the actual error distribution is non-Gaussian, the time series models based on Gaussian distribution assumption can still be applied with satisfactory performance. In the following sections, the censored Gaussian assumption will also be proved to be reasonable and acceptable by generating reliable PIs based on actual wind power data.

In practice, the trained neural network $\hat{y}(\mathbf{x}_i)$ could be regarded as an estimation of the true regression $y(\mathbf{x}_i)$. In principle, NNs generate the averaged values of targets conditioned on input variables vector $\mathbf{x}_i$, $E[t_i \mid \mathbf{x}_i]$ [37]

$$\hat{y}(\mathbf{x}_i) = f(\mathbf{x}_i; \hat{\mathbf{w}}, \hat{b}, \hat{\beta}) = E[t_i \mid \mathbf{x}_i]. \tag{10}$$

According to the two uncertainties discussed in the preceding section, we can divide the prediction errors into two components, including the one involved in the estimation of the true regression and the other involved in the estimation of the measured targets. Then, the prediction error can be expressed as

$$t_i - \hat{y}(\mathbf{x}_i) = [y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i)] + \varepsilon(\mathbf{x}_i) \tag{11}$$

where $t_i - \hat{y}(\mathbf{x}_i)$ denotes the total prediction error, and $y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i)$ denotes the error of the neural network estimation of the true regression. To account for model uncertainties, model uncertainty intervals (MUIs) can be used to quantify the uncertainty between the neural network estimation and the true regression $P(y(\mathbf{x}_i) \mid \hat{y}(\mathbf{x}_i))$. In contrast, prediction intervals aim to quantify the uncertainty associated with the difference between the measured values t_i and the predicted values$\hat{y}(\mathbf{x}_i)$, i.e., $P(t_i \mid \hat{y}(\mathbf{x}_i))$. Accordingly, PIs will be wider than MUIs and will enclose them.

Assuming two error components in (11) are statistically independent, the variance of the total prediction errors σ_t^2 can be mathematically obtained based on the variance of model uncertainty $\sigma_{\hat{y}}^2$ and the variance of data noise σ_ε^2, as

$$\sigma_t^2(\mathbf{x}_i) = \sigma_{\hat{y}}^2(\mathbf{x}_i) + \sigma_\varepsilon^2(\mathbf{x}_i). \tag{12}$$

Given a real process, an $100(1-\alpha)\%$ confidence level PI of the measured target t_i is a stochastic interval $I_t^\alpha(\mathbf{x}_i)$ expressed as

$$I_t^\alpha(\mathbf{x}_i) = [L_t^\alpha(\mathbf{x}_i), U_t^\alpha(\mathbf{x}_i)] \tag{13}$$

such that the coverage rate $P(t_i \in I_t^\alpha) = 100(1-\alpha)\%$, where the lower bound $L_t^\alpha(\mathbf{x}_i)$ and the upper bound $U_t^\alpha(\mathbf{x}_i)$ can be obtained by

$$L_t^\alpha(\mathbf{x}_i) = \hat{y}(\mathbf{x}_i) - z_{1-\alpha/2}\sqrt{\sigma_t^2(\mathbf{x}_i)} \tag{14}$$

$$U_t^\alpha(\mathbf{x}_i) = \hat{y}(\mathbf{x}_i) + z_{1-\alpha/2}\sqrt{\sigma_t^2(\mathbf{x}_i)} \tag{15}$$

where $z_{1-\alpha/2}$ is the critical value of the standard Gaussian distribution, which depends on the desired confidence level $100(1-\alpha)\%$. When the bounds of $I_t^\alpha(\mathbf{x}_i)$ go beyond the unit interval [0, 1], they should be adjusted to the corresponding

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

lower or upper constraint bounds to ensure the constructed wind power PIs within the capacity range, and the corresponding probability mass is added to the adjusted bounds.

III. Prediction Intervals Construction

Here, the prediction intervals for ELM forecasting are developed based on the bootstrap method. Several bootstrap methods have been compared to identify the most suitable one for PIs construction of ELM-based wind power forecasting.

A. Bootstrap Methods

Bootstrap is regarded as a general approach of statistical inference based on building a sampling distribution by uniform sampling with replacements from the original data [38], [39]. It is widely applied as a robust alternative to the statistical inference based on the parametric assumptions, which can be unreliable and even impossible due to the sophistications involved in computing the standard errors in some conditions.

Three different bootstrap algorithms can be applied for regression analysis [39], [40], including the pairs bootstrap, the standard residuals bootstrap, and the wild residuals bootstrap (wild bootstrap). The pairs bootstrap can be applied according to the algorithm shown by the following steps.

Step 1) Obtain the training samples $\{(\mathbf{x}_i, t_i)\}_{i=1}^N$.

Step 2) Generate bootstrapped pairs $\{(\mathbf{x}_i^*, t_i^*)\}_{i=1}^N$ by uniform sampling with replacement from the original training data $\{(\mathbf{x}_i, t_i)\}_{i=1}^N$.

Step 3) Estimate the ELM $\hat{y}_l(\mathbf{x}_i^*)$ from the lth bootstrapped dataset $\{(\mathbf{x}_i^*, t_i^*)\}_{i=1}^N$.

Step 4) Repeat steps 2)–3) to obtain B bootstrap replicates.

The two other methods differ from the pairs bootstrap method mainly in sampling the residuals, of which details can be found in [39]. These bootstrap methods have been implemented in our case study to identify the best one for ELM-based wind power forecast.

For the three bootstrap approaches, when training an ELM on particular bootstrap samples, the model parameters are estimated in order to minimize the errors on the training data. Based on the B bootstrap replicates, we can train and obtain B ELMs ready for wind power forecasting.

B. Model Uncertainty Variance

The MUIs can quantify the confidence in the network estimation $\hat{y}(\mathbf{x}_i)$for the true regression $y(\mathbf{x}_i)$. The bootstrap-based approach assumes that an ensemble of NN models will reach a relatively less biased approximation of true regression of the measured targets. Given the training data set

$$D_t = \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^N \tag{16}$$

the B_M training data sets are resampled from the original training data with replacement. The average output of the ensemble of B_M ELMs is taken as the estimation of the true regression, expressed as

$$\hat{y}(\mathbf{x}_i) = \frac{1}{B_M}\sum_{l=1}^{B_M} \hat{y}_l(\mathbf{x}_i) \tag{17}$$

where $\hat{y}_l(\mathbf{x}_i)$ is the prediction value of the input samples generated by the lth bootstrapped ELM.

The variance of model misspecification uncertainty $P(y(\mathbf{x}_i) \,|\, \hat{y}(\mathbf{x}_i))$ can be estimated from the variance in the outputs of the trained B_M ELMs as

$$\sigma_{\hat{y}}^2(\mathbf{x}_i) = \frac{1}{B_M - 1}\sum_{l=1}^{B_M} \left(\hat{y}_l(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i)\right)^2. \tag{18}$$

Following the bootstrap procedures, MUIs of ELM forecasts $I_M^\alpha(\mathbf{x}_i) = [L_M^\alpha(\mathbf{x}_i), U_M^\alpha(\mathbf{x}_i)]$ can be obtained through

$$L_M^\alpha(\mathbf{x}_i) = \hat{y}(\mathbf{x}_i) - z_{1-\alpha/2}\sqrt{\sigma_{\hat{y}}^2(\mathbf{x}_i)} \tag{19}$$

$$U_M^\alpha(\mathbf{x}_i) = \hat{y}(\mathbf{x}_i) + z_{1-\alpha/2}\sqrt{\sigma_{\hat{y}}^2(\mathbf{x}_i)}. \tag{20}$$

C. Data Noise Variance

In addition to the model uncertainty of ELM forecasting, the uncertainty caused by the data noise is analyzed in this section. Due to the heteroscedasticity, with only one observation of wind power at each time point, it is challenging to estimate the data uncertainty. According to the variance definition in [41], the variance of the measured target t_i conditioned on the input variables $\mathbf{x}_i$, can be calculated from

$$\sigma_\varepsilon^2(t_i \,|\, \mathbf{x}_i) = E[(t_i - E[t_i \,|\, \mathbf{x}_i])^2 \,|\, \mathbf{x}_i]. \tag{21}$$

Given the training data $\{(\mathbf{x}_i, t_i)\}_{i=1}^N$, as can be seen from (10), the outputs of ELMs produce averaged values of the targets conditioned on input variables $\mathbf{x}_i$, i.e., $\hat{y}(\mathbf{x}_i) = E[t_i \,|\, \mathbf{x}_i]$. Then, the values of $E[t_i \,|\, \mathbf{x}_i]$ in (21) can be derived based on the trained ELM.

Keeping the input $\mathbf{x}_i$ and replacing the targets t_i with $(\hat{y}(\mathbf{x}_i) - t_i)^2$, we can obtain the transformed training dataset

$$D_\varepsilon = \{(\mathbf{x}_i, (\hat{y}(\mathbf{x}_i) - t_i))\}_{i=1}^N. \tag{22}$$

The objective variance can be estimated by training a separate ELM$h_\varepsilon(\mathbf{x}_i; \mathbf{w}, b, \beta)$, mathematically expressed as

$$h_\varepsilon(\mathbf{x}_i; \mathbf{w}, b, \beta) = (\hat{y}(\mathbf{x}_i) - t_i)^2, \quad i = 1, \ldots, N. \tag{23}$$

The output of the trained ELM $\hat{r}(\mathbf{x}_i)$ can be represented as

$$\hat{r}(\mathbf{x}_i) = E[(\hat{y}(\mathbf{x}_i) - t_i)^2 \,|\, \mathbf{x}_i]. \tag{24}$$

The model uncertainty associated with ELM $h_\varepsilon(\mathbf{x}_i; \mathbf{w}, b, \beta)$, represented by $\sigma_{\hat{r}}^2(\mathbf{x}_i)$, also should be taken into account. It can be calculated through the bootstrap-based method similar to

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

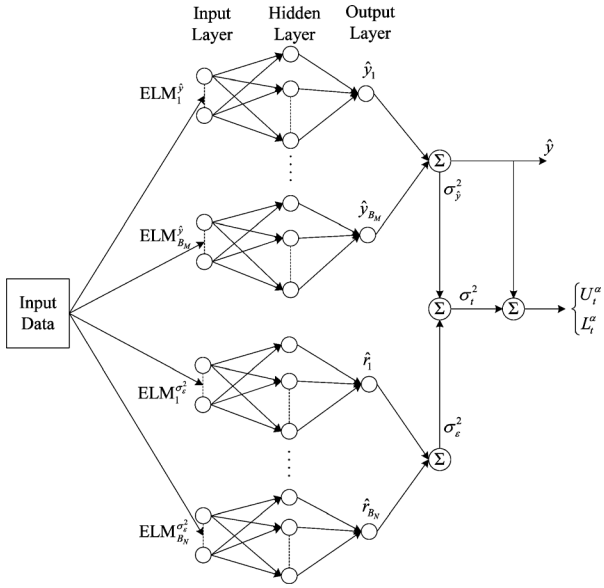

Fig. 1. Framework for PI construction of the proposed BELM approach.

the procedure of deriving the model uncertainty variance. Supposing B_N bootstrap replicates are implemented, we can obtain, respectively, the estimated noise variance $\hat{\sigma}^2_{\varepsilon}(t_i \,|\, \mathbf{x}_i)$ and the variance of regression model uncertainty $\sigma^2_{\hat{r}}(\mathbf{x}_i)$ as

$$\hat{\sigma}^2_{\varepsilon}(t_i \,|\, \mathbf{x}_i) = \hat{r}(\mathbf{x}_i) = \frac{1}{B_N} \sum_{l=1}^{B_N} \hat{r}_l(\mathbf{x}_i) \tag{25}$$

$$\sigma^2_{\hat{r}}(\mathbf{x}_i) = \frac{1}{B_N - 1} \sum_{l=1}^{B_N} \left(\hat{r}_l(\mathbf{x}_i) - \hat{r}(\mathbf{x}_i)\right)^2. \tag{26}$$

The variance of data noise can be obtained through

$$\sigma^2_{\varepsilon}(\mathbf{x}_i) = \hat{\sigma}^2_{\varepsilon}(t_i \,|\, \mathbf{x}_i) + \sigma^2_{\hat{r}}(\mathbf{x}_i). \tag{27}$$

With the model uncertainty variance and data noise variance, the total variance of the prediction intervals can be obtained based on (12).

For PIs construction of ELM forecasting using the proposed algorithm, $B_M + B_N$ ELM models are required in total. The overall framework for the proposed bootstrap-based approach for ELM probabilistic forecasting is explicitly displayed in Fig. 1. If traditional NNs are used, intensive computational efforts are required since the bootstrap-based forecasting approach involves a great number of bootstrap replicates. With the extremely fast learning speed, the proposed BELM approach can effectively and efficiently provide the probabilistic forecasting for wind power production.

IV. PI-Based Forecasts Evaluation

Here, several performance indices are introduced to assess the quality of PIs derived by the proposed BELM approach.

A. Reliability

By definition, from (14) and (15), the future targets are expected to lie within the bounds of constructed PIs with a prescribed probability termed as the PI nominal confidence (PINC) $100(1-\alpha)\%$. It is expected that the coverage probability of obtained PIs will asymptotically reach the nominal level of confidence over the full test data. PI coverage probability (PICP) is a critical measure for the reliability of the constructed PIs, which is defined by

$$\mathrm{PICP} = \frac{1}{N_{\mathrm{test}}} \sum_{i=1}^{N_{\mathrm{test}}} c_i \tag{28}$$

where N_{test} is the number of test samples, and c_i is the indicator of PICP and is defined as

$$c_i = \begin{cases} 1, & t_i \in I_t^{\alpha}(\mathbf{x}_i) \\ 0, & t_i \notin I_t^{\alpha}(\mathbf{x}_i) \end{cases}. \tag{29}$$

For reliable PIs, the examined PICP should be close to its corresponding PINC. Another assessment index, average coverage error (ACE), is defined by

$$\mathrm{ACE} = \mathrm{PICP} - \mathrm{PINC}. \tag{30}$$

Generally, to ensure PIs with high reliability, the ACE should be as close to zero as possible, i.e., smaller absolute ACE indicates more reliable PIs.

B. Sharpness

Obviously, PICP is directly related with the sharpness of PIs. High level PICP can be easily reached via widening PIs. However, such PIs are meaningless in practice since they do not express the actual variation of the measured wind power. The width of PI for the ith target, $\delta_t^{\alpha}(\mathbf{x}_i)$, is expressed as

$$\delta_t^{\alpha}(\mathbf{x}_i) = U_t^{\alpha}(\mathbf{x}_i) - L_t^{\alpha}(\mathbf{x}_i). \tag{31}$$

The interval score can be employed to comprehensively evaluate the overall skill of wind power PIs to assess the sharpness [42]. The interval score of the PI $I_t^{\alpha}(\mathbf{x}_i)$ with PINC $100(1-\alpha)\%$ is defined by

$$Sc_t^{\alpha}(\mathbf{x}_i) = \begin{cases} -2\alpha\delta_t^{\alpha}(\mathbf{x}_i) - 4\left[L_t^{\alpha}(\mathbf{x}_i) - t_i\right], & \text{if } t_i < L_t^{\alpha}(\mathbf{x}_i) \\ -2\alpha\delta_t^{\alpha}(\mathbf{x}_i), & \text{if } t_i \in I_t^{\alpha}(\mathbf{x}_i) \\ -2\alpha\delta_t^{\alpha}(\mathbf{x}_i) - 4\left[t_i - U_t^{\alpha}(\mathbf{x}_i)\right], & \text{if } t_i > U_t^{\alpha}(\mathbf{x}_i). \end{cases} \tag{32}$$

Based on the test dataset, the overall score value Sc_t^{α} can be obtained and given as

$$Sc_t^{\alpha} = \frac{1}{N_{\mathrm{test}}} \sum_{i=1}^{N_{\mathrm{test}}} Sc_t^{\alpha}(\mathbf{x}_i). \tag{33}$$

The interval score rewards the narrow PI and gives penalty if the target does not lie within the estimated PI. The score Sc_t^{α} can be used to evaluate the overall skill of constructed PIs by taking all aspects of PI quality into consideration. Generally, given a particular PINC and similar PICPs, PIs with the larger interval score have a relatively higher overall skill. However, the unique interval score does not distinguish the specific contributions of reliability or sharpness to the skill. In the evaluation process, we reasonably give a higher priority to the reliability, since it is the key feature reflecting the correctness of the constructed PIs.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Based on the prior analysis of PIs reliability, the interval score can be used to assess PIs from the perspective of sharpness.

V. Case Studies

A. Description of Experiment Data

In the study, the proposed BELM approach has been tested using the wind power data from Cathedral Rocks wind farm, South Australia. The wind farm has nominal generation capacity P_c of 66 MW combined with 33 wind turbines of 2 MW. The wind power data with one hour temporal resolution from June 2008 until June 2012 are used for the case study.

Operational planning and scheduling in modern power systems with wind power requires the forecasts of the future wind power generation according to the planning horizons. Generally, wind power forecasting can be divided into four categories of different timeframes: very short-term, short-term, medium-term and long-term forecasts [43]. Short-term and very short-term forecasts are important because of their significances to both generation and reserve dispatches and etc. As introduced in [1], statistical methods would outperform NWP-based methods for forecasting wind power with look-ahead times less than a few hours. Though having external NWP information, statistical models using historical measurements only should be preferred for such short look-ahead times [1]. Therefore, the proposed BELM approach takes only historical wind power data as inputs for hourly ahead forecasting, which is essential for, e.g., dispatching ancillary service market in practice. Other data such as the weather information can be easily included in our future work.

B. Determination of ELM Hidden Nodes Number

The ELM is based on SLFNs, of which the number of hidden-layer neurons for ELM models need to be determined properly. Since different forecasts may have different needs and properties, optimization of the ELM structure is necessary and critical to minimize the uncertainties due to model misspecifications and ensure the efficiency simultaneously. The hidden nodes number of ELMs is determined based on the cross-validation approach [19]. The ELMs' generalization performance of different structures over the validation dataset is assessed by both root-mean-square error (RMSE) and mean absolute error (MAE) as

$$\mathrm{MAE} = \frac{1}{N_{\mathrm{test}}} \sum_{i=1}^{N_{\mathrm{test}}} |t_i - \hat{y}(\mathbf{x}_i)| \tag{34}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{test}}} \sum_{i=1}^{N_{\mathrm{test}}} [t_i - \hat{y}(\mathbf{x}_i)]^2}. \tag{35}$$

Fig. 2 shows the results of validation test for hourly ahead forecasting based on the actual wind farm data. It can be seen that ELMs will have stable generalization performance once the hidden nodes exceed a certain threshold. The ELM with 63 hidden neurons can sufficiently ensure the optimal MAE and RMSE simultaneously.

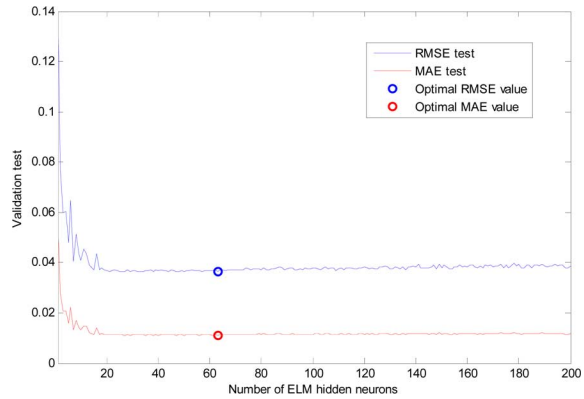

Fig. 2. Validation test for ELMs with different numbers of hidden neurons.

TABLE I
Evaluation Results of Different Bootstrap Methods

Method	PINC	PICP	ACE	Score
Pairs bootstrap	90%	93.59%	3.59%	-6.02%
	95%	96.17%	1.17%	-3.67%
	99%	98.25%	-0.75%	-1.17%
Standard residuals bootstrap	90%	92.85%	2.85%	-6.11%
	95%	95.92%	0.92%	-3.72%
	99%	98.09%	-0.91%	-1.21%
Wild bootstrap	90%	93.43%	3.43%	-6.16%
	95%	95.92%	0.92%	-3.78%
	99%	97.67%	-1.33%	-1.25%

C. Comparison of Bootstrap Methods

The MUIs of ELM regression are approximated based on the bootstrap methods. The commonly applied bootstrap methods, including pairs bootstrap, standard residuals bootstrap, and wild bootstrap, can provide different performances for different applications. The performances of the different bootstrap approaches used to construct the PIs of wind power forecasting with ELMs are compared. PICP with corresponding PINC, ACE, and interval score of different bootstrap approaches are given in Table I.

From Table I, it can be found that the pairs bootstrap provides the most reliable PIs of the measured wind power. The residuals based bootstrap relies on the errors that are representative of the true model errors. However, the nonlinear relationship for prediction is always unknown, and the model misspecification is unavoidable. If the model is either misspecified or overfitted, the pairs based bootstrap approach can be more robust [40]. As expected, the pairs bootstrap method outperforms the other two methods in the tests using the chaotic wind power data. Based on the comparisons, the pair bootstrap is applied for the proposed BELM approach.

D. Analysis of Forecasting Results

High complexity of chaotic climate systems contributes to high level of uncertainties in wind power generation. The patterns of weather conditions and wind speeds vary very much in different seasons. To examine the effectiveness and applicability of the proposed approach, the four seasons in Australia, summer (December to February), autumn (March to May), winter (June

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II
RESULTS OF PIS RELIABILITY IN DIFFERENT SEASONS

Season	PINC	BELM		BELM-Beta		Persistence		Climatology		ESM	
		PICP	ACE	PICP	ACE	PICP	ACE	PICP	ACE	PICP	ACE
Summer	90%	92.46%	2.46%	89.40%	-0.60%	86.93%	-3.07%	98.01%	8.01%	88.55%	-1.45%
	95%	95.09%	0.09%	92.46%	-2.54%	90.70%	-4.30%	98.65%	3.65%	92.09%	-2.91%
	99%	98.08%	-0.92%	95.30%	-3.70%	95.35%	-3.65%	99.72%	0.72%	96.64%	-2.36%
Autumn	90%	93.01%	3.01%	72.88%	-17.12%	83.99%	-6.01%	95.84%	5.84%	88.98%	-1.02%
	95%	95.59%	0.59%	76.46%	-18.54%	88.36%	-6.64%	98.59%	3.59%	92.90%	-2.10%
	99%	97.67%	-1.33%	79.45%	-19.55%	92.36%	-6.64%	100%	1.00%	97.04%	-1.96%
Winter	90%	91.30%	1.30%	80.22%	-9.78%	86.83%	-3.17%	96.34%	6.34%	88.97%	-1.03%
	95%	94.41%	-0.59%	83.61%	-11.39%	90.88%	-4.12%	99.45%	4.45%	92.94%	-2.06%
	99%	97.89%	-1.11%	87.45%	-11.55%	95.30%	-3.70%	100%	1.00%	96.99%	-2.01%
Spring	90%	93.19%	3.19%	84.03%	-5.97%	86.52%	-3.48%	95.46%	5.46%	88.36%	-1.64%
	95%	96.12%	1.12%	86.39%	-8.61%	90.46%	-4.54%	97.54%	2.54%	91.89%	-3.11%
	99%	98.68%	-0.32%	89.79%	-9.21%	95.00%	-4.00%	99.62%	0.62%	96.24%	-2.76%

to August), and spring (September to November) are considered respectively. Models are separately constructed for different seasons.

Considering the seasonal difference and diversity, the proposed BELM method is tested using wind power data in summer 2012, autumn 2011, winter 2011, and spring 2010. The wind power data before these test dates are used as the training data. These datasets are normalized with respect to the capacity of Cathedral Rocks wind farm before applying to the proposed models. To evaluate the proposed approach, the climatology and the persistence approaches are used for benchmarking the forecasting performance [12], [13], [44]. Climatology predictive distribution is formed based on all available wind power observations, and is a unique and unconditional probabilistic prediction. The climatology is relatively easy to outperform for the short-term probabilistic wind power forecasting. In deterministic point forecasts of wind power generation, the persistence method is considered to be the most common benchmark and difficult to outperform for short look-ahead time forecasting. The persistence based probabilistic forecast model is used as a benchmark for comparisons in the study as well. Its mean is given by the last available power measurement, and the variance is computed using the latest observations. Both the climatology and persistence methods are relatively simple. To benchmark the proposed BELM approach, an advanced model, the ESM method, proposed in [10] is applied in the study. In addition, to evaluate the impacts of forecasting error distribution model, the proposed method has also been tested using the Beta distribution for forecasting error modeling, termed as BELM-Beta [4].

The major objective of the proposed BELM method is to derive reliable PIs. Furthermore, power system operation requires useful information with high confidence levels. Therefore, it should be more practically meaningful to obtain high-confidence-level PIs to fulfill the needs of power system operation. Different levels of PINC $100(1-\alpha)\%$ ranging from 90%–99% are considered in the study. For the PIs reliability test, corresponding PICPs and ACEs are given in Table II. The interval score results of different approaches are given in Table III.

As seen in Table II, in all four seasons, the proposed method outperforms other approaches with the resultant PICPs consistently closer to the corresponding nominal confidence levels. All ACEs of the proposed method are close to zero, especially for the higher confidence levels of 95% and 99%, which indicates the high reliability of the constructed PIs, e.g., in autumn, the proposed method has absolute ACEs at confidences 95% and 99% around 1%, smaller than the other four benchmarks. Particularly, in summer, PIs obtained by the Beta distribution model have similar reliability with the censored Gaussian distribution model of the proposed BELM approach and demonstrate much better quality than in other seasons, which means that the Beta distribution modeling is much more proper to summer than to other three seasons.

TABLE III
RESULTS OF INTERVAL SCORE IN DIFFERENT SEASONS

Season	PINC	BELM	BELM-Beta	Persistence	Climatology	ESM
Summer	90%	-7.61%	-8.39%	-8.78%	-16.08%	-8.54%
	95%	-4.59%	-5.37%	-5.49%	-8.50%	-5.25%
	99%	-1.41%	-2.21%	-1.96%	-1.81%	-1.76%
Autumn	90%	-5.92%	-5.97%	-6.50%	-15.10%	-6.36%
	95%	-3.64%	-3.61%	-4.03%	-7.85%	-3.89%
	99%	-1.19%	-1.09%	-1.46%	-1.71%	-1.30%
Winter	90%	-6.91%	-6.85%	-7.60%	-15.66%	-7.36%
	95%	-4.12%	-4.06%	-4.76%	-8.19%	-4.47%
	99%	-1.28%	-1.18%	-1.65%	-1.78%	-1.40%
Spring	90%	-7.10%	-7.19%	-7.95%	-16.31%	-7.71%
	95%	-4.15%	-4.16%	-4.96%	-8.55%	-4.72%
	99%	-1.12%	-1.09%	-1.75%	-1.81%	-1.55%

According to Table III, interval scores of the proposed approach are larger than the climatology, the persistence, and the ESM methods, indicating that the proposed BELM approach outperforms these three benchmarks from the perspective of sharpness and overall skill. In addition, the proposed approach also can have similar or higher skill than the approach with Beta distribution based error modeling in some cases. It can be proved that the average interval score of the proposed method still outperforms the Beta distribution considering all the four seasons. Considering the reliability and overall skill, the proposed approach shows much better results in terms of comprehensive performance than the four benchmarks.

This is not unreasonable as Beta distribution can well reflect the long-term statistics of wind power forecasting errors [4], but it is unable to reflect seasonal variations in detail. In addition, the Beta distribution can model the forecasting errors well given different levels of wind power outputs, which is not the case

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

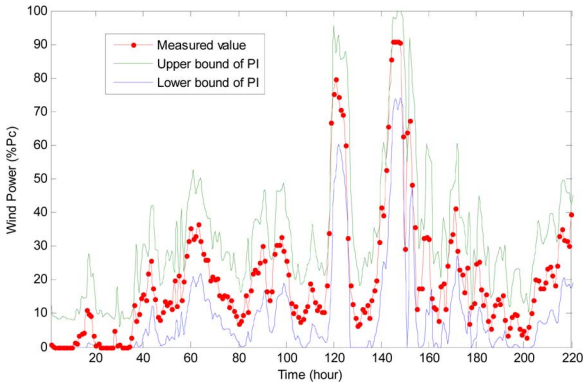

Fig. 3. PIs with nominal confidence 90% in summer 2012 obtained by the proposed BELM approach.

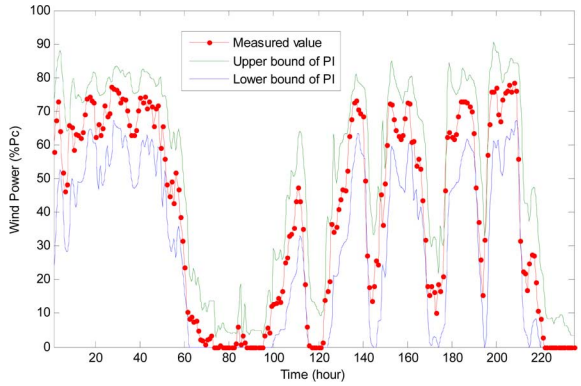

Fig. 5. PIs with nominal confidence 90% in winter 2011 obtained by the proposed BELM approach.

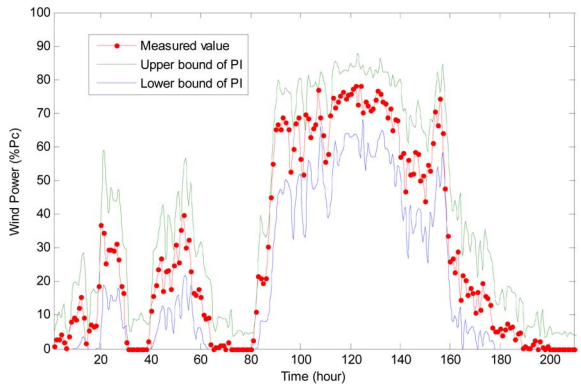

Fig. 4. PIs with nominal confidence 90% in autumn 2011 obtained by the proposed BELM approach.

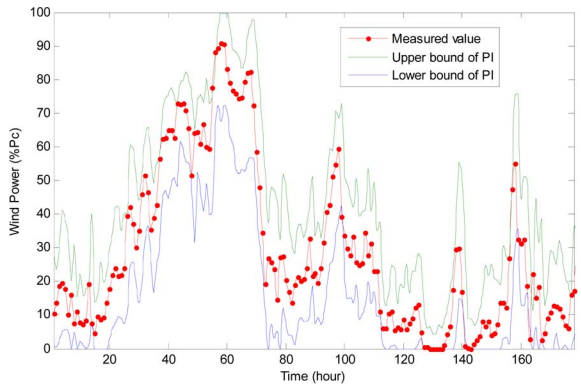

Fig. 6. PIs with nominal confidence 90% in spring 2010 obtained by the proposed BELM approach.

in our approach, where the forecasting errors are modeled statistically dependent of the input variables. Therefore the Beta distribution based approach provides less satisfactory results at high confidence levels of 90% and above. Furthermore, in the autumn season, the Beta distribution based approach gives very poor results.

The climatology is a simple unconditional prediction approach and does not consider the heteroscedasticity of wind power data. Therefore large widths of PIs are resulted at high confidence levels which are barely useful in practice. The PICP for persistence approach varies significantly in different seasons, indicating the significant seasonal variations of wind power. Due to the simple mapping, the persistence cannot obtain sufficiently satisfactory PIs. In comparisons, the ESM approach has fair results with respect to both reliability and sharpness. From Table II, it can be seen that in summer PIs reliability of the proposed method are slightly lower than the rest seasons. This is understandable as weather conditions in summer are relatively more chaotic.

The 90% confidence PIs obtained by the proposed BELM method and the actually measured wind power in the four seasons are visually displayed in Figs. 3–6, respectively. For all four seasons, the measured wind power data are perfectly enclosed by the PIs generated by the proposed method, indicating an excellent performance that can fulfill the needs of power system operation. These graphs also clearly demonstrate the non-stationary characteristics of wind power series. Notably, some PIs can have abnormal values out of capacity range of the wind farm. Therefore resultant predictive densities have been censored to concentrate probability mass outside the interval on the bounds [12].

To investigate the influence of the bootstrap replicates number on the resulted PIs, the proposed method is further tested on the wind power data in autumn 2011. Each test with given bootstrap replicates is conducted for 100 repetitive times using a PC with Intel Core Duo 3.16-GHz CPU and 4-GB RAM. The mean PICP (MPICP), the standard deviation of PICP (SDPICP), and the needed training time with different bootstrap replicates are given in Table IV. With bootstrap replicates varying between 20 and 1000, the BELM method can produce reliable PIs. Considering both the accuracy and efficiency, the 200 bootstrap replicates for generating PIs are considered the best option for the case study. Although the size of the training data set is not small, the total time needed for ELMs training using the proposed BELM approach only accounts for about 30 s, indicating a significantly high efficiency and potential for online application. On the contrary, training traditional BP NNs for hourly ahead wind power forecasting using the similar size of data can take thousands of times longer. The extremely fast model construction should benefit practical applications from several aspects under the precondition of ensuring satisfactory performance. Foremost, it saves the efforts in offline model construction which could otherwise be much time-consuming and computationally intensive. It should not be

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
RELIABILITY AND EFFICIENCY OF DIFFERENT BOOTSTRAP REPLICATES

Bootstrap Replicates	PINC	MPICP	SDPICP	Training Time (s)
20	90%	92.09%	0.38%	3.26
	95%	95.18%	0.28%	
50	90%	92.09%	0.28%	7.89
	95%	95.17%	0.21%	
100	90%	92.04%	0.24%	15.78
	95%	95.17%	0.18%	
200	90%	92.03%	0.18%	31.43
	95%	95.14%	0.16%	
300	90%	92.03%	0.18%	47.16
	95%	95.17%	0.14%	
500	90%	92.02%	0.17%	78.15
	95%	95.17%	0.15%	
1000	90%	92.01%	0.17%	166.78
	95%	95.16%	0.14%	

TABLE V
EFFECT OF BOOTSTRAPPING ON RESULTANT PIS

PINC	Method	PICP	ACE	Score
90%	BELM	92.03%	2.03%	-6.20%
	MVR-ELM	88.56%	-1.44%	-6.13%
95%	BELM	95.14%	0.14%	-3.97%
	MVR-ELM	91.00%	-4.00%	-3.94%
99%	BELM	97.75%	-1.25%	-1.33%
	MVR-ELM	93.56%	-5.44%	-1.38%

unreasonable that the characteristics of wind power series could be changed continuously or suddenly, similarly to the chaotic weather systems. Therefore, continuous online model updating can be significantly meaningful to maintain and improve the forecasting performance as far as possible, especially for the very short-term forecasting.

Traditionally, the model uncertainty is always ignored in point forecasting [23]. Further, there have existed theoretical results for uncertainty estimation for the Gaussian linear regression [45]. However, the linear regression system of ELM is based on the randomly assigned input weights and biases. As introduced in Section II-B, the randomly generated input weights and biases are also one source of the model uncertainty of NNs. To comprehensively involve the model misspecifications and improve the forecasts accuracy, the bootstrap is applied for the proposed ELM based PI construction approach. To investigate the influence of the bootstrapping, the proposed approach with (200 replicates) or without bootstrapping are conducted for 100 times to obtain the mean forecasting reliability measured by PICP and ACE, and mean sharpness measured by interval score. The test results are shown in Table V. The proposed approach without bootstrapping just uses ELM for mean and variance regression (MVR), termed as MVR-ELM here.

We can find that the model uncertainty does have observable impacts on the resultant PIs from Table V. Though similar sharpness can be obtained, PICP can be reduced by more than 4%, if the model uncertainty is not considered by the application of bootstrap in PI formulation. Similarly, degradations of ACE due to no bootstrapping involvements are also observed in Table V. It echoes the descriptions in Section II-B that the model uncertainty is one indispensible aspect of uncertainty sources for the proposed BELM-based forecasting, and should be considered to form reliable and accurate PIs.

TABLE VI
RESULTS OF MULTISTEP INTRA-HOUR FORECASTING

Horizon	Method	PINC 90%			PINC 95%		
		PICP	ACE	Score	PICP	ACE	Score
10 min	BELM	91.51%	1.51%	-3.10%	95.72%	0.72%	-1.99%
	BELM-Beta	71.89%	-18.11%	-3.11%	75.14%	-19.86%	-2.01%
	Persistence	83.83%	-6.17%	-3.38%	87.25%	-7.75%	-2.17%
	Climatology	96.94%	6.94%	-14.83%	98.50%	3.50%	-7.75%
	ESM	90.48%	0.48%	-3.38%	93.72%	-1.28%	-2.12%
30 min	BELM	91.69%	1.69%	-5.33%	94.72%	-0.28%	-3.27%
	BELM-Beta	70.89%	-19.11%	-5.38%	73.70%	-21.30%	-3.32%
	Persistence	83.06%	-6.94%	-5.94%	86.90%	-8.10%	-3.73%
	Climatology	96.94%	6.94%	-14.83%	98.50%	3.50%	-7.75%
	ESM	90.09%	0.09%	-5.86%	93.47%	-1.53%	-3.61%
1 h	BELM	90.54%	0.54%	-6.81%	93.82%	-1.18%	-4.09%
	BELM-Beta	70.21%	-19.79%	-6.86%	73.45%	-21.55%	-4.11%
	Persistence	82.44%	-7.56%	-7.52%	86.45%	-8.55%	-4.66%
	Climatology	96.94%	6.94%	-14.83%	98.50%	3.50%	-7.75%
	ESM	89.47%	-0.53%	-7.38%	93.31%	-1.69%	-4.45%

The comprehensive numerical studies have indicated the effectiveness of the proposed BELM approach. Actually, the multi-step forecasts with look-ahead times two and three hours have been implemented using the proposed approach, and satisfactory PIs can be obtained. In addition to wind power forecasts with hourly resolution, the intra-hour prediction results are also highly concerned by TSO and wind farm controller. The higher resolution wind power, e.g., 10-min measures are very crucial to wind farm control, continuous generation and reserve dispatch, and so forth, and would have higher volatility than hourly data. Practically, in Denmark, the 10-min lead time is regarded as the most important very short-term horizon by the TSO since power fluctuations at this time horizon have the most serious impacts on the balance of power systems [46]. We study 10-min resolution forecasting with different look-ahead horizons including 10 min, 30 min, and 1 h in autumn 2009 of Cathedral Rocks wind farm. The resultant PIs with PINCs 90% and 95% are obtained and evaluated, respectively, and given in Table VI.

From Table VI, it can be found that the proposed BELM has superior performance than the ESM method and other benchmarks for intra-hour wind power forecasting. The persistence approach demonstrates relatively lower reliability than that in hourly forecasting, indicating the higher violability of wind power with 10-min resolution. Comparing with the ESM approach, which is a well-established time series model for short-term probabilistic forecasting of wind power, the proposed BELM approach has high flexibility due to the non-linear mapping capability of ELM. With successful application to short term probabilistic wind power forecasting in this paper, the proposed method can perform longer term forecasting by including NWP information as additional inputs to ensure the performance. In practice, the system-level aggregated wind power is also highly concerned by the TSO. Due to the flexibility, the proposed BELM approach provides a generalized framework for probabilistic wind power forecasting. Therefore, local NWP and historical wind power of individual wind farms can be taken as the inputs to the proposed model to forecast

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

the aggregated wind generation involving the farm-level information. With the fast speed and high flexibility, the proposed model can provide an online tool to facilitate various decision making activities by TSO and generation companies to determine the needed reserve and design proper bidding strategies against risks.

VI. Conclusion

Wind power forecasting is critical to power system operation. However, wind power forecasting errors are unavoidable to some extent due to the nonlinear and stochastic nature of the weather system. Traditional neural network based forecasting models cannot provide satisfactory performances with respect to both accuracy and computing time needed. In this paper, extreme learning machine is successfully applied for probabilistic interval forecasting of wind power. A novel statistical approach BELM is developed to construct the PIs of ELM based regression. Accurate PIs can be obtained by combining the variances of regression model uncertainty and residual noise. Different bootstrap methods have been compared and analyzed to select the best one for the developed forecasting model. The influence of bootstrap replicates on the efficiency and the quality of the constructed PIs has also been carefully investigated. Further, the effect of the model uncertainty (bootstrapping process) on resultant PIs is examined in the study, verifying its indispensability. Because of the extremely fast learning, the training of the proposed BELM forecasting method can be extremely faster than traditional NNs based approaches, demonstrating a high potential for online application. Comprehensive experiments using practical wind farm data of different seasons have demonstrated the highly satisfactory results, which indicate that efficient and accurate short term wind power forecasts can be achieved using the proposed BELM method. With fast speed, high reliability and high flexibility, the proposed BELM approach is a generalized framework for probabilistic forecasting of wind power and can provide an efficient and meaningful online tool for power system applications including probabilistic reserve determination, generation dispatch, wind farm control, electricity market trading, and so forth.

References

[1] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl, "The state-of-the-art in short-term prediction of wind Power: A literature review," Tech. Rep., ANEMOS.plus Project Deliverable Rep. D1.2, 2011.

[2] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renew. Sustain. Energy Rev.*, vol. 12, no. 6, pp. 1725–1744, Aug. 2008.

[3] L. Ma, S. Y. Luan, C. W. Jiang, H. L. Liu, and Y. Zhang, "A review on the forecasting of wind speed and generated power," *Renew. Sustain. Energy Rev.*, vol. 13, no. 4, pp. 915–920, May 2009.

[4] H. Bludszuweit, J. A. Dominguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 983–991, Aug. 2008.

[5] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004.

[6] H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts," *Wind Energy*, vol. 9, no. 1–2, pp. 95–108, 2006.

[7] J. W. Taylor, P. E. McSharry, and R. Buizza, "Wind power density forecasting using ensemble predictions and time series models," *IEEE Trans. Energy Convers.*, vol. 24, no. 3, pp. 775–782, Sep. 2009.

[8] R. J. Bessa, V. Miranda, A. Botterud, J. Wang, and E. M. Constantinescu, "Time adaptive conditional kernel density estimation for wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 660–669, Oct. 2012.

[9] R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang, "Time-adaptive quantile-copula for wind power probabilistic forecasting," *Renew. Energy*, vol. 40, no. 1, pp. 29–39, Apr. 2012.

[10] A. Lau and P. McSharry, "Approaches for multi-step density forecasts with application to aggregated wind power," *Ann. Appl. Stat.*, vol. 4, no. 3, pp. 1311–1341, 2010.

[11] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol. 12, no. 1, pp. 51–62, Jan. 2009.

[12] P. Pinson and G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1845–1856, Nov. 2010.

[13] G. Sideratos and N. D. Hatziargyriou, "Probabilistic wind power forecasting using radial basis function neural networks," *IEEE Trans. Power Syst.*, vol. 27, no. 4, pp. 1788–1796, Nov. 2012.

[14] M. A. Matos and R. J. Bessa, "Setting the operating reserve using probabilistic wind power forecasts,," *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 594–603, May 2011.

[15] R. J. Bessa, M. A. Matos, I. C. Costa, L. Bremermann, I. G. Franchin, R. Pestana, N. Machado, H. P. Waldl, and C. Wichmann, "Reserve setting and steady-state security assessment using wind power uncertainty forecast: A case study," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 827–836, Oct. 2012.

[16] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Y. Huang, and K. Subbarao, "Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures," *IEEE Trans. Sustain. Energy*, vol. 2, no. 4, pp. 433–442, Oct. 2011.

[17] E. M. Constantinescu, V. M. Zavala, M. Rocklin, S. Lee, and M. Anitescu, "A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 431–441, Feb. 2011.

[18] P. Pinson, C. Chevallier, and G. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1148–1156, Aug. 2007.

[19] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.

[20] A. H. Nizar, Z. Y. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 946–955, Aug. 2008.

[21] G. B. Huang, X. J. Ding, and H. M. Zhou, "Optimization method based extreme learning machine for classification," *Neurocomputing*, vol. 74, no. 1–3, pp. 155–163, Dec. 2010.

[22] G. B. Huang, H. M. Zhou, X. J. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[23] G. Karniotakis, G. S. Stavrakakis, and E. F. Nogaret, "Wind power forecasting using advanced neural network models," *IEEE Trans. Energy Convers.*, vol. 11, no. 4, pp. 762–767, Dec. 1996.

[24] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, "Long-term wind speed and power forecasting using local recurrent neural network models," *IEEE Trans. Energy Convers.*, vol. 21, no. 1, pp. 273–284, Mar. 2006.

[25] K. Bhaskar and S. N. Singh, "AWNN-assisted wind power forecasting using feed-forward neural network," *IEEE Trans. Sustain. Energy*, vol. 3, no. 2, pp. 306–315, Apr. 2012.

[26] J. T. G. Hwang and A. A. Ding, "Prediction intervals for artificial neural networks," *J. Amer. Stat. Assoc.*, vol. 92, no. 438, pp. 748–757, Jun. 1997.

[27] D. J. C. MacKay, "The evidence framework applied to classification networks," *Neural Comput.*, vol. 4, no. 5, pp. 720–736, Sep. 1992.

[28] T. Heskes, , T. P. M. Mozer and M. Jordan, Eds., "Practical confidence and prediction intervals," in *Neural Information Processing Systems*. Cambridge, MA, USA: MIT, 1997, vol. 9, pp. 176–182.

[29] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," in *Proc. IEEE Int. Conf. Neural Netw.*, Orlando, 1994, vol. 1, pp. 55–60.

[30] R. Dybowski and S. J. Roberts, , R. Dybowski and V. Gant, Eds., "Confidence intervals and prediction intervals for feed-forward neural networks," in *Clinical Applications of Artificial Neural Networks*. Cambridge, U.K.: Cambridge Univ., 2001, pp. 298–326.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[31] A. Khosravi, S. Nahavandi, D. Creighton, and D. Srinivasan, "Optimizing the quality of bootstrap-based prediction intervals," in *Proc. Int. Joint Conf. Neural Netw.*, San Jose, CA, USA, 2011, pp. 3072–3078.
[32] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction intervals for short-term wind farm power generation forecasts," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 602–610, Jul. 2013.
[33] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.
[34] P. Pinson and H. Madsen, "Adaptive modelling and forecasting of offshore wind power fluctuations with Markov-switching autoregressive models," *J. Forecasting*, vol. 31, no. 4, pp. 281–313, Jul. 2012.
[35] B. Muthén, "Moments of the censored and truncated bivariate normal distribution," *Br. J. Math. Stat. Psychol.*, vol. 43, no. 1, pp. 131–143, 1990.
[36] J. D. Hamilton, *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press, 1994.
[37] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1996.
[38] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, no. 1, pp. 1–26, Jan. 1979.
[39] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York, NY, USA: Chapman & Hall/CRC, 1993.
[40] R. Tibshirani, "A comparison of some error estimates for neural network models," *Neural Comput.*, vol. 8, no. 1, pp. 152–163, Jan. 1996.
[41] H. Cramér, *Mathematical Methods of Statistics*. Princeton, NJ, USA: Princeton Univ., 1971.
[42] R. L. Winkler, "A decision-theoretic approach to interval estimation," *J. Amer. Stat. Ass.*, vol. 67, no. 337, pp. 187–191, Mar. 1972.
[43] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *Proc. North Amer. Power Symp.*, Arlington, TX, USA, 2010, pp. 1–8.
[44] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, "Non-parametric probabilistic forecasts of wind power: Required properties and evaluation," *Wind Energy*, vol. 10, no. 6, pp. 497–516, Nov. 2007.
[45] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*. New York, NY, USA: Wiley, 2003.
[46] V. Akhmatov, "Influence of wind direction on intense power fluctuations in large offshore wind farms in the North Sea," *Wind Eng.*, vol. 3, no. 1, pp. 59–64, Jan. 2007.

Can Wan (S'10) received the B.Eng. degree from Zhejiang University, Hangzhou, China, in 2008. He is currently working toward the Ph.D. degree from The Hong Kong Polytechnic University.

He is also a Visiting Scholar with the Center for Electric Power and Energy, Technical University of Denmark, Lyngby, Denmark. His research interests include power systems analysis and planning, grid integration of renewable energies, machine learning, and system engineering.

Zhao Xu (M'06–SM'12) received the B.Eng. degree from Zhejiang University, Hangzhou, China, in 1996, the M.Eng. degree from National University of Singapore, Singapore, in 2002, and the Ph.D. degree from The University of Queensland, Australia, in 2006.

He is now with The Hong Kong Polytechnic University. He was previously an Associate Professor with Centre for Electric Power and Energy, Technical University of Denmark. His research interest includes demand side, grid integration of renewable energies and EVs, electricity market planning and management, and AI applications in power engineering. He is an editor of *Electric Power Components and Systems*.

Pierre Pinson (M'11–SM'13) received the M.Sc. degree in applied mathematics from the National Institute for Applied Sciences, INSA, Toulouse, France, and the Ph.D. degree in energy from Ecole des Mines de Paris, Paris, France.

He is currently a Professor of Modeling of Electricity Markets with the Centre for Electric Power and Energy, Department of Electrical Engineering, Technical University of Denmark, Lyngby, Denmark. His research interests include among others forecasting, uncertainty estimation, optimization under uncertainty, decision sciences, and renewable energies. He is an editor for *Wind Energy*.

Dr. Pinson serves as an editor for the IEEE TRANSACTIONS ON POWER SYSTEMS.

Zhao Yang Dong (M'99–SM'06) received the Ph.D. degree from the University of Sydney, Sydney, Australia, in 1999.

He is now Head and Chair Professor with School of Electrical and Information Engineering, The University of Sydney, Syndey, Australia. He previously held academic and industrial positions with The University of Newcastle, The Hong Kong Polytechnic University, the University of Queensland, Australia, and Transend Networks, Australia. His research interests include smart grids, power system planning, power system security, load modeling, renewable energy systems, electricity market, and computational intelligence.

Prof. Dong serves as an an editor of the IEEE TRANSACTIONS ON SMART GRID and the IEEE POWER ENGINEERING LETTERS.

Kit Po Wong (M'87–SM'90–F'02) received the M.Sc., Ph.D., and D.Eng. degrees from University of Manchester, Institute of Science and Technology, Manchester, U.K., in 1972, 1974, and 2001, respectively.

Prof. Wong was with the University of Western Australia, Perth, Australia, from 1974 until 2004 and is now an Adjunct Professor there. Since 2002, he has been Chair Professor, and previously Head, of the Department of Electrical Engineering, The Hong Kong Polytechnic University. He served as Editor-in-Chief of and was Editor-in-Chief of *IEE Proc. Generation, Transmission and Distribution*. His current research interests include computation intelligence applications to power system analysis, planning and operations, as well as power market.

Prof. Wong is a Fellow of IET, HKIE, and IEAust. He was the recipient of three Sir John Madsen Medals (1981, 1982, and 1988) from the Institution of Engineers Australia, the 1999 Outstanding Engineer Award from IEEE Power Chapter Western Australia, and the 2000 IEEE Third Millennium Award. He was a Co-Technical Chairman of IEEE Machine Learning and Cybernetics (ICMLC) 2004 Conference and General Chairman of IEEE/CSEE PowerCon2000. He is now Editor-in-Chief of the IEEE POWER ENGINEERING LETTERS.