A Hybrid Probabilistic Wind Power Prediction Based on An Improved Decomposition Technique and Kernel Density Estimation

B. Khorramdel, *Student Member IEEE*, M. Azizi, N. Safari, *Student Member IEEE*, C. Y. Chung, *Fellow, IEEE*, and S. M. Mazhari, *Member IEEE*

***Abstract*—The high uncertainty of non-stationary wind power time series is a challenging issue in optimal operation and planning of power systems. An efficient way to show wind power uncertainty is to use high-quality prediction intervals (PIs). This paper proposes a hybrid probabilistic wind power prediction model based on improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) technique, extreme learning machine (ELM) and kernel density estimation (KDE). First, using ICEEMDAN, the original wind power time series is decomposed to components with different frequency ranges. Then, sample entropy (SampEn) technique is employed to group components to three main time series *trend*, *cycle*, and *noise* with diverse complexity levels. The first two components are deterministically predicted while the noise component is probabilistically predicted using the combination of KDE technique and *direct plug-in* as a well-known bandwidth selection technique. The lower and upper bounds of final PI are found using the summation of lower and upper bounds of noise component with trend and cycle predicted points. The efficacy of the proposed prediction model is depicted by generating reliable and sharp PIs for real wind power datasets in Canada and comparing with other conventional PI construction approaches.**

***Index Terms*—Empirical mode decomposition, extreme learning machine, kernel density estimation, prediction interval, sample entropy.**

I. Introduction

THE increasing penetration level of wind power generation in electrical networks has caused a real challenge in today's power systems [1-3]. Due to the high intermittency and stochasticity of wind power generation by which high level of uncertainty is created in power systems, highly accurate and reliable wind power prediction approaches are strongly recommended to avoid unprecedented problems in decision making conditions like economic load dispatch, unit commitment, electricity market, etc. [4]. To ensure high reliability and optimality of wind power integrated power systems, this level of uncertainty needs to be quantified by probabilistic approaches. Unlike deterministic prediction approaches, which includes an unavoidable prediction error, probabilistic approaches aim to represent a lower and an upper bound or a prediction interval (PI) by which the future wind power observation is enclosed with a predetermined confidence level [5-10]. The mean-variance, persistence, Quantile regression (QR), lower-upper bound estimation (LUBE) [5-8], hybrid intelligent algorithm (HIA) [9], and conditional density estimation based on *Nadaraya-Watson* estimator [10] are instances of conventional probabilistic approaches which suffer from some drawbacks. For instance, the mean-variance approach uses neural networks (NNs) to estimate the conditional distribution of a target with extra Gaussian noise and a variable variance. It underestimates the variance of data, leading to unreliable PIs. LUBE approach for PI estimation is based on traditionally tuned NNs and a new cost function [5]. Considering both width and coverage probability of intervals is the main advantage of this approach. However, using traditional approaches to train NNs leads to some unavoidable limitations, such as a high computational load, slow learning procedure, and overtraining. QR is used to estimate different quantiles of a predictive distribution in different applications, but it involves a cumbersome optimization process to minimize its defined cost function. Also, HIA generates PIs by direct optimization of both sharpness and reliability with a multi objective cost function. The abovementioned approaches either assume a parametric distribution such as Gaussian, t-student, Beta, etc. for prediction error or include a heuristic optimization algorithm that might converge to local minima. There are other methods in which decomposition techniques such as wavelet, empirical mode decomposition (EMD), and ensemble EMD (EEMD) are used to promote the accuracy of the prediction [1-3], but these techniques are not efficient due to creation of time series reconstruction error.

There are still much improvements which should be fulfilled in this field to propose a comprehensive approach with satisfactory performance for real situations.

In this paper, a PI construction approach is proposed based on improved complete ensemble empirical mode decomposition of wind power time series with adaptive noise (ICEEMDAN) to avoid significant residual noise, kernel density estimation (KDE) and extreme learning machine (ELM). To achieve this, wind power time series is first decomposed into several components with different complexities. The components are then clustered into three

This work was supported in part by the Natural Sciences and Engineering Re-search Council (NSERC) of Canada and the Saskatchewan Power Corporation (SaskPower).

Benyamin Khorramdel, Chi Yung Chung, Nima Safari and Seyed Mahdi Mazhari are with the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada (e-mail: benyamin.khorramdel@usask.ca; c.y.chung@usask.ca; n.safari@usask.ca, s.m.mazhari@usask.ca). Mrs. Mahsa Azizi is with the Department of Computer Science, University of Saskatchewan, Saskatoon, SK S7N 5C9, Canada (e-mail: maa790 @mail.usask.ca).

978-1-5386-7703-2/18/$31.00 ©2018 IEEE

main time series using sample entropy (SampEn) technique; (i) *trend*, (ii) *cycle*, and (iii) *noise* [1]. Due to the high predictability of the first two time series, they are deterministically predicted by ELM, but the noise time series is probabilistically predicted through KDE and ELM. Eventually, the final PI is constructed by the combination of deterministic and probabilistic prediction results.

This paper is organized as follows. Section II briefly reviews ICEEMDAN and SampEn technique for non-stationary time series decomposition and grouping purposes, respectively. Section III explains KDE with a well-known bandwidth selection algorithm. The ELM as an efficient learning algorithm, is briefly reviewed in section IV. Section V presents the general structure of the proposed hybrid probabilistic wind power prediction approach. Simulation results are shown in section VI. Finally, the conclusion is drawn in section VII.

II. A Brief Review on ICEEMDAN and SampEn

EMD is known as an efficient and adaptive method for fast decomposition of non-linear and non-stationary time series [11]. It consists of a fully data-driven separation procedure in which the original time series is expressed as a summation of intrinsic mode functions (IMFs) or simply modes plus a final trend. However, it suffers from mode mixing problem which is the creation of similar scales oscillations in different modes or disparate scales oscillations in one mode. To alleviate this problem, EEMD was introduced in recent years [12] in which the decomposition is performed over an ensemble of original signals mixed with a white Gaussian noise. But, EEMD also creates new problems which is the existence of a residual noise when the original time series is reconstructed. The ICEEMDAN is a highly improved version of EEMD in which the reconstruction error is negligible. In the following, first the EEMD and then the ICEEMDAN is presented.

A. EEMD

Suppose $p(t)$ is the original wind power time series and $w_i(t)|_{i=1}^{n}$ is n different realizations of white noise with Gaussian distribution. First, an ensemble of noisy versions of original signal, $p_i(t)=p(t)+\rho.w_i(t)|_{i=1}^{n}$ is generated ($\rho>0$). Then, each $p_i(t)$ is decomposed by EMD into K number of IMFs denoted by $IMF_i^{(k)}$. Finally, the k^{th} mode of $p(t)$ is calculated as follows.

$$\overline{IMF}^{(k)}=\sum_{i=1}^{n} IMF_i^{(k)}/n \tag{1}$$

B. ICEEMDAN

This technique is based on the calculation of residues for noisy versions of original time series [13]. There are five steps as follows.

Step 1: The local means of $p_i(t)=p(t)+\rho_0.E_1\big(w_i(t)\big)\big|_{i=1}^{n}$ is obtained using EMD to find the first residue by (2).

$$r_1(t)=\sum_{i=1}^{n} M(p_i(t))/n \tag{2}$$

It should be noted the operators $E_k(\cdot)$ and $M(\cdot)$ find the k^{th} mode and the local mean of the input time series, respectively.

Step 2: Calculate the first mode as $\widetilde{IMF}^{(1)}=p(t)-r_1(t)$.

In CEEMDAN, the modes are denoted as $\widetilde{IMF}^{(k)}$.

Step 3: Obtain the second residue $r_2(t)$ by calculating the average of local means of $r_1(t)+\rho_1.E_2\big(w_i(t)\big)\big|_{i=1}^{n}$ and obtain the second mode as follows.

$$\widetilde{IMF}^{(2)}=r_1(t)-\frac{\sum_{i=1}^{n} M\Big(r_1(t)+\rho_1.E_2\big(w_i(t)\big)\Big)}{n} \tag{3}$$

Step 4: Calculate the k^{th} residue for $k=3,\dots,K$ as (4). Note that K is the total number of modes.

$$r_k(t)=\frac{\sum_{i=1}^{n} M\Big(r_{k-1}(t)+\rho_{k-1}.E_k\big(w_i(t)\big)\Big)}{n} \tag{4}$$

Step 5: Calculate k^{th} mode as $\widetilde{IMF}^{(k)}=r_{k-1}(t)-r_k(t)$. Repeat steps 4 and 5 to obtain K modes.

The coefficient ρ_k can choose the signal to noise ratio. Regarding the amplitude of the added noise, reference [12] suggests to use small coefficients for time series disturbed by high frequency signals, and vice versa.

C. Sample Entropy Technique

SampEn is a revised version of approximate entropy (ApEn) which are usually used for physiological time series complexity analysis [14]. SampEn is a well-known technique due to three advantages over ApEn: (i) data length independence, (ii) relatively trouble-free implementation, and (iii) excluding self-similar patterns.

Suppose m is an embedding dimension, r is a selection tolerance and N is the number of data points. SampEn is then represented by $SampEn(m,r,N)$ and defined as the negative logarithm of the probability that if two sets of data points of length m show a distance smaller than r then by adding the next point they still keep the distance smaller than r. In this context, we assume these sets of data points are drawn from k^{th} mode of wind power time series obtained by ICEEMDAN denoted by $\widetilde{IMF}^{(k)}$. Following steps need to be performed to calculate $SampEn(m,r,N)$ for a time series with N samples.

Step 1: Make the sub-sample vectors $X_m(i)$ with dimension m.

$$X_m(i)=[x_i,x_{i+1},x_{i+2},\dots,x_{i+m-1}] \quad i=1,\dots,N-m+1 \tag{5}$$

Step 2: Calculate the Chebyshev or Euclidean distance between $X_m(i)$ and $X_m(j)$ $(d_m[X_m(i),X_m(j)]\ i\neq j)$.

Step 3: Count the number of vectors by which the distance with $X_m(i)$ is less than r, and denote as B_i (for dimension $m+1$ it is denoted by A_i). Then, calculate B_i^a as follows:

$$B_i^a=(\textstyle\sum_{i=1}^{N-m} B_i)/N-m-1 \tag{6}$$

Step 4: Increase the size of sub-sample vectors to $m+1$ and repeat steps 1 to 3 to obtain A_i^a.

$$A_i^a=(\textstyle\sum_{i=1}^{N-m} A_i)/N-m-1 \tag{7}$$

Step 5: Define B^a and A^a as follows:

$$B^a=(\textstyle\sum_{i=1}^{N-m} B_i^a)/N-m \tag{8}$$

$$A^a=(\textstyle\sum_{i=1}^{N-m} A_i^a)/N-m \tag{9}$$

Step 6: Calculate the SampEn by (10).

$$SampEn(m,r,N)=-\ln[A^a/B^a] \tag{10}$$

According to [14], the value of m and r can be set as $1\le m\le 2$ and $0.1\sigma\le r\le 0.25\sigma$ where σ is the standard deviation of the time series under analysis. In this paper, $m=2$ and $r=0.2\sigma$.

Generally, in EMD technique, IMFs with high order number contain low frequency oscillations; therefore, they present low complexity level than the preceding IMFs. Based on this prior knowledge, SampEn value is employed to group IMFs into three main time series *trend*, *cycle*, and *noise*. The SampEn of original time series is calculated besides the

SampEn of IMFs. The IMFs with $SampEn_{IMF} \ll SampEn_{org}$ are grouped into trend time series. Cycle time series is obtained using IMFs with $SampEn_{org} - \delta \le SampEn_{IMF} \le SampEn_{org} + \delta$ while for noise time series $SampEn_{IMF} \gg SampEn_{org}$. The value of δ is approximately determined based on the value of $SampEn_{org}$ and $SampEn_{IMF}$ because the time series grouping procedure is not that sensitive to δ. In this paper δ is set on 0.15.

III. Kernel Density Estimation

A. Gaussian Kernel Density Estimator

Assume N independent samples $\{x_i\}_{i=1}^{N}$ drawn from a random noise samples with unknown probability density function (PDF) f. The kernel density estimate $\hat{f}$ is expressed by (11).

$$\hat{f}(\boldsymbol{x};h) = \frac{1}{N}\sum_{i=1}^{N} \kappa(\boldsymbol{x}, x_i; h) \qquad \boldsymbol{x} \in \mathbb{R} \tag{11}$$

where $\kappa(\cdot)$ is a kernel with bandwidth h. Gaussian kernel $\Phi(\cdot)$ has been widely used in the literature [15].

$$\Phi(\boldsymbol{x}, x_i; h) = \frac{\exp(-((\boldsymbol{x} - x_i)/h)^2/2)}{\sqrt{2\pi}\, h} \tag{12}$$

B. Direct Plug-In Bandwidth Selection Technique

Since the practical implementation of all kernel density estimators depends on bandwidth parameter h, an efficient bandwidth selection technique, *direct plug-in*, which tends to give good answers for wide range of underlying functions is used here [15]. A well-defined criterion for optimal selection of h, shown in (13), is mean integrated squared error (MISE) with two components, the integrated squared bias and integrated variance. Hence, h is determined such that MISE is minimized. The mathematical proof is found in [15].

$$h = \left(\frac{\int_{\mathbb{R}} \kappa(x)^2 dx}{\left(N.\left(\int_{\mathbb{R}} x^2\kappa(x)dx\right)^2.\hat{\psi}_4(q)\right)}\right)^{\frac{2}{5}} \tag{13}$$

$$\hat{\psi}_z(q) = \tfrac{1}{N^2}\textstyle\sum_{i=1}^{N}\sum_{j=1}^{N} \kappa_q^{(z)}\left(x_i - x_j\right) \tag{14}$$

$$q = \left(-2\kappa_q^{(z)}(0)/N.\left(\int_{\mathbb{R}} x^2\kappa(x)dx\right).\psi_{z+2}\right)^{1/z+3} \tag{15}$$

This method depends on the pilot bandwidth q and z^{th} derivative of standard Gaussian kernel $\kappa_q^{(z)}(z \in 2n, z \ge 6)$; therefore, an l-stage bandwidth selector is implemented as bellow.

Step 1: Choose l, the number of stages for selection of h.

Step 2: Set $z = 2l + 4$.

Step 3: Find the initial value of the kernel estimator ψ_z through $\psi_z^{ini} = (-1)^{z/2}\, z!/\left(\sqrt{\pi}(2\sigma)^{z+1}(z/2)!\right)$, where σ is the standard deviation of samples under analysis.

Step 4: Find pilot bandwidth q by (15).

Step 5: Calculate $\hat{\psi}_4(q)$ by (14) to calculate h. If $z \ne 4$ go to Step 4 and continue until $\hat{\psi}_4(q)$ is obtained.

C. Quantiles of a PDF

After estimating PDF of noise samples, using the associated cumulative distribution functions $F(x)$ obtained by (16), the lower/upper quantiles $L^{(\alpha)}$ and $U^{(\alpha)}$ corresponding to confidence level $100(1-\alpha)\%$ are obtained via (17). These quantiles are used as outputs of ELM in training procedure.

$$F(x) = \int_0^x f(\tau)d\tau \tag{16}$$

$$U^{(\alpha)} = F^{-1}\left(\tfrac{\alpha}{2}\right)\ ,\ L^{(\alpha)} = F^{-1}\left(1 - \tfrac{\alpha}{2}\right) \tag{17}$$

IV. A Brief Review on ELM

A fast and efficient learning approach is employed in this paper to train a single hidden layer neural network. ELM is known to have extremely low learning burden, high generalization ability, and high tendency to avoid local minima and overfitting [16].

Assume N different samples $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ drawn from a dataset, ELMs with $\widetilde{N}$ hidden nodes and infinitely differentiable activation function $\Phi(x) = 1/1 + \exp(-x)$ are expressed as bellow.

$$\boldsymbol{y}_i = \textstyle\sum_{j=1}^{\widetilde{N}} \beta_j\, \Phi\left(\boldsymbol{\omega}_j.\boldsymbol{x}_i + b_j\right), \qquad i = 1, \dots, N \tag{18}$$

where input vector $\boldsymbol{x}_i = [x_{i1}, \dots, x_{in}]^T$ and the output vector $\boldsymbol{y}_i = [y_{i1}, \dots, y_{im}]^T$ should be determined for training procedure. Also, $\boldsymbol{\omega}_j = \left[\omega_{j1}, \dots, \omega_{jn}\right]^T$ is the input weight vector, $\boldsymbol{\beta}_j = \left[\beta_{j1}, \dots, \beta_{jm}\right]^T$ is the output weight vector, and $\boldsymbol{b} = [b_1, \dots, b_{\widetilde{N}}]^T$ is the hidden layer biases vector. By randomly choosing values of $\boldsymbol{\omega}$ and $\boldsymbol{b}$, ELM is converted to a linear system, and the set of equations (18) can be mathematically written by a simple matrix multiplication as $\mathbf{H\beta} = \mathbf{Y}$. The input matrix $\mathbf{H}$ is expressed as (19) where the output weight matrix $\boldsymbol{\beta}$ and the target matrix $\mathbf{Y}$ are illustrated by (20).

$$\mathbf{H} = \begin{bmatrix} \Phi(\boldsymbol{\omega}_1.\boldsymbol{x}_1 + b_1) & \cdots & \Phi(\boldsymbol{\omega}_{\widetilde{N}}.\boldsymbol{x}_1 + b_{\widetilde{N}}) \\ \vdots & \ddots & \vdots \\ \Phi(\boldsymbol{\omega}_1.\boldsymbol{x}_N + b_1) & \cdots & \Phi(\boldsymbol{\omega}_{\widetilde{N}}.\boldsymbol{x}_N + b_{\widetilde{N}}) \end{bmatrix}_{N \times \widetilde{N}} \tag{19}$$

$$\boldsymbol{\beta} = \left[\boldsymbol{\beta}_1^T \quad \dots \quad \boldsymbol{\beta}_{\widetilde{N}}^T\right]^T,\ \mathbf{Y} = \left[\boldsymbol{y}_1^T \quad \dots \quad \boldsymbol{y}_N^T\right]^T \tag{20}$$

After calculation of $\mathbf{H}$, matrix $\boldsymbol{\beta}$ is obtained using $\boldsymbol{\beta} = \mathbf{H}^{\dagger}\mathbf{Y}$ where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$ [17].

V. Proposed PI Construction Model

The general structure of the proposed wind power PI construction approach is shown in Fig. 1. As it can be seen there are four main building blocks in this approach; ICEEMDAN, SampEn, ELM, and KDE. First, wind power time series is received, preprocessed and normalized. Then, ICEEMDAN decomposes it into several IMFs. Using SampEn technique, three main time series trend, cycle, and noise with different complexity levels are generated and independently predicted with ELM prediction model.

The point prediction approach is used for trend and cycle time series because these time series contain low complexity contents and are more predictable than noise time series. They are deterministically predicted by ELM to produce predicted points PP_t and PP_c, respectively. To assess the performance of ELM for trend and cycle time series prediction, two evaluation criteria (21)-(22) are generally used, mean absolute error (MAE) and root-mean-square error (RMSE), where y_i and $\hat{y}_i$ are real and predicted test sample points, respectively.

$$MAE = \frac{1}{N_{ts}}\sum_{i=1}^{N_{ts}}|y_i - \hat{y}_i| \quad (21)$$

$$RMSE = \left(\frac{1}{N_{ts}}\sum_{i=1}^{N_{ts}}(y_i - \hat{y}_i)^2\right)^{1/2} \quad (22)$$

For noise time series, a probabilistic prediction approach is developed here to constitute the lower and upper bounds $PI_n = [L_n, U_n]$ based on KDE and ELM training for a prespecified confidence level. Eventually, the final PI, i.e., $PI_f = [L_f, U_f]$ is constructed by the combination of PI_n, PP_t and PP_c such that it can meet the related evaluation criteria. To evaluate the performance of the proposed probabilistic approach, PI_f should be tested using two important indices, reliability and sharpness, explained as follows.

1. Reliability: The PI coverage probability ($PICP$) in (23) should be too close to PI nominal coverage ($PINC$) to meet the reliability criterion as the most important feature.

$$PICP = \frac{1}{N_{ts}}\sum_{i=1}^{N_{ts}}\delta_i \quad (23)$$

$$\delta_i = \begin{cases} 1 \quad , \quad L_i^\alpha \le y_i \le U_i^\alpha \\ 0 \quad , \quad y_i < L_i^\alpha \; or \; y_i > U_i^\alpha \end{cases} \quad (24)$$

where the number of test samples is denoted by N_{ts}. L_i^α and U_i^α are respectively the lower and upper bounds of the PI related to the prediction target y_i.

2. Sharpness: To present meaningful information by a PI, the PI normalized average width ($PINAW$) in (25) should take small values as much as achievable.

$$PINAW = \frac{1}{R.N_{ts}}\sum_{i=1}^{N_{ts}}(U_i^\alpha - L_i^\alpha) \quad (25)$$

where R is the range of targets and used to normalize the PI average width.

In this paper, to deterministically predict trend and cycle time series, $\text{ELM}_{\text{trend}}$ and $\text{ELM}_{\text{cycle}}$ are trained with one-hour ahead prediction horizon based on historical wind power data points. To this end, a 10-fold cross validation is firstly run to optimally determine the number of ELM hidden nodes and the optimal length of trend and cycle time series as training datasets. Based on the experiments, five time lags are considered as the optimal number of lags for input samples.

For noise time series, first, KDE is used to estimate the PDF of noise samples over numerous one-hour time steps with 10-min resolution. In other words, each one-hour time step in noise time series contains six random samples with an unknown PDF. Second, $\text{ELM}_{\text{noise}}$ is trained considering lower and upper quantiles of estimated PDFs as outputs and noise samples as inputs.

After completing training procedure, $\text{ELM}_{\text{noise}}$ uses validation dataset to construct PI_n with predetermined confidence level. Then, a PI_f is constructed by adding the predicted points of trend and cycle time series to PI_n. In this stage, $PICP$ and $PINAW$ are calculated and assessed in three possible situations; (i) the reliability criterion is satisfied and $|PICP - PINC|$ is small, (ii) the reliability criterion is satisfied but $|PICP - PINC|$ is large, and (iii) the reliability criterion is not satisfied. The first situation is the ideal one, but for the two other situations regulation factor λ with regulation step $\Delta\lambda$ is innovatively defined for KDE bandwidth to regulate the bandwidth h as (26) such that the constructed PI_n for validation dataset can simultaneously lead to a high quality PI. The main reason to choose h^* for reliability improvement

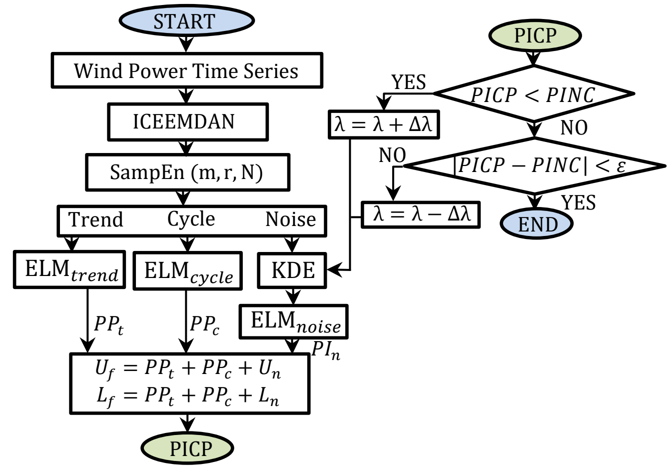

Fig. 1 The structure of the proposed wind power prediction model.

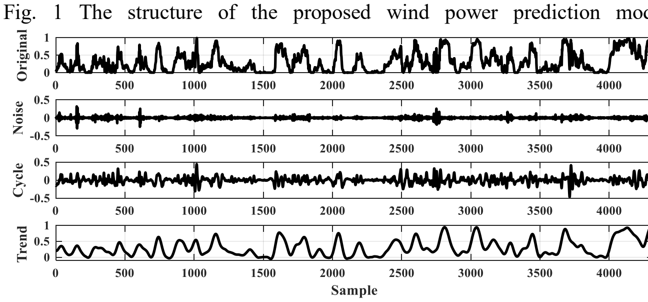

Fig. 2 The original and main components of Centennial time series.

is that a KDE with larger bandwidth can produce smoother PDFs and consequently larger intervals by which more uncertainty level can be captured. After finding optimal value of λ, it is used for test dataset to construct desired PIs.

$$h^* = \lambda.h \quad (26)$$

VI. Case Studies

A. Dataset

In this study, the proposed approach is implemented using the wind power datasets from Canada; Winter 2016 of the Centennial wind farm in Saskatchewan with 150 MW, and Spring 2012 of Alberta Electric System Operator (AESO) with 967 MW installed capacity [2]. The data resolution 10-min is used for one-hour ahead probabilistic prediction. Training, validation and test datasets include 70%, 20%, and 10% of the original dataset, respectively. The reason behind choosing just wind power time series is that statistical methods show better performance than numerical weather prediction methods for wind power prediction with prediction horizon less than few hours [18].

B. Benchmark Approaches

To assess the performance of the proposed hybrid approach, i.e., ICEEMDAN-ELM-KDE, it is compared with persistence and other well-trained approaches such as QR, EEMD-ELM-QR and ICEEMDAN-ELM-QR. In persistence approach, the predictive distribution of wind power is simply estimated using most recent observations. In QR, a cost function for the dataset under analysis is minimized to estimate lower and upper bounds. EEMD is compared with ICEEMDAN using EEMD-ELM-QR and ICEEMDAN-ELM-QR benchmarks to show the efficacy of ICEEMDAN. Then, the superiority of the proposed KDE approach is shown by comparison with ICEEMDAN-ELM-KDE.

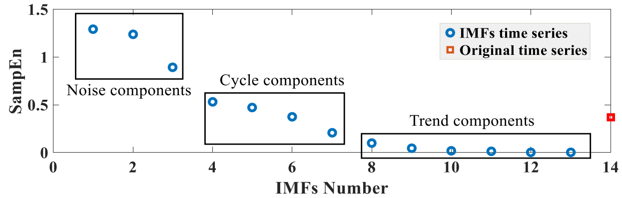

Fig. 3 The value of SampEn for IMFs obtained by ICEEMDAN.

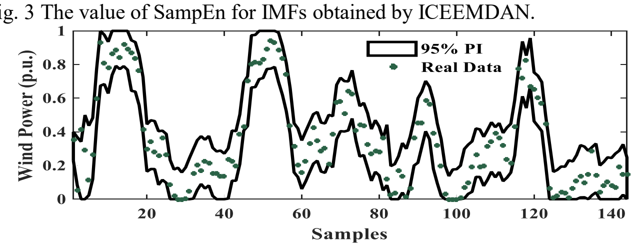

Fig. 4 The constructed PI at 95% confidence level for Centennial.

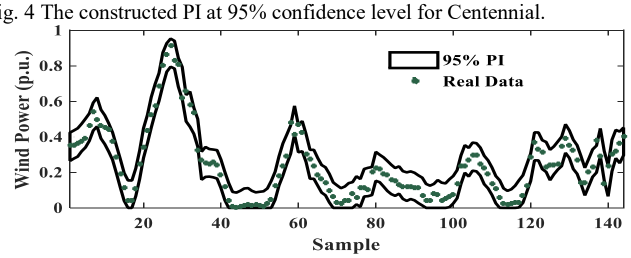

Fig. 5 The constructed PI at 95% confidence level for AESO.

TABLE I
THE RESULTS OF PI CONSTRUCTION FOR CENTENNIAL DATASET

Conf. level	Method	PICP (%)	PINAW (%)
90%	Persistence	82.92	26.60
	QR	88.33	27.45
	EEMD-ELM-QR	87.91	25.32
	ICEEMDAN-ELM-QR	89.16	26.83
	ICEEMDAN-ELM-KDE	**91.66**	**22.20**
95%	Persistence	88.33	31.10
	QR	92.50	28.82
	EEMD-ELM-QR	93.33	29.63
	ICEEMDAN-ELM-QR	94.58	28.40
	ICEEMDAN-ELM-KDE	**95.83**	**25.12**

C. Comparison of Results

In this section, the simulation results for construction of reliable and sharp PIs are presented. Three main time series of Centennial dataset, obtained by ICEEMDAN and SampEn, are illustrated in Fig. 2. It can be seen that ICEEMDAN with the help of SampEn can decompose the original time series into just three time series with completely different complexity levels. In Fig. 3, it is shown how IMFs are grouped into three main time series based on their SampEn values. Instead of predicting several IMFs, using just three time series makes a trade-off between computational burden and accuracy of prediction. Figures 4 and 5 visually illustrate how well the constructed final PIs at 95% confidence level for Centennial and AESO datasets can encompass the real wind power data. To clearly show the upper and lower bounds of the constructed PIs, they are only shown for six days. To confirm the efficacy of the proposed PI construction approach, the values of evaluation indices are shown in Table I for Centennial dataset as a highly chaotic case study [2]. The proposed approach outperforms the well-trained benchmarks in terms of generating reliable, i.e., PICP is too close to PINC, and sharp PIs, i.e., PINAW is of the least value among benchmarks.

VII. CONCLUSION

Due to the high importance of wind power prediction models in modern power systems, in this paper a hybrid probabilistic approach is proposed which takes advantages of ICEEMDAN as an efficient decomposition technique, KDE as a well-known and easy-to-implement estimation method for capturing the uncertainty of random variables, and ELM as a fast and efficient learning algorithm. Besides, SampEn technique is used to group different components of wind power time series into three main time series to make the prediction more accurate. The simulation results show that the proposed approach outperforms the utilized benchmarks for construction of reliable and sharp PIs.

REFERENCES

[1] G. Zhang et al., "An advanced approach for construction of optimal wind power prediction intervals," *IEEE Trans. Power Syst*., vol. 30, no. 5, pp. 2706–2715, Sep. 2015.

[2] N. Safari, C. Y. Chung, and G. C. D. Price, "A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis," *IEEE Trans. Power Syst.*, *Early Access.*

[3] N. Safari, Y. Chen, B. Khorramdel, L. P. Mao, and C. Y. Chung "A Spatiotemporal Wind Power Prediction based on Wavelet Decomposition, Feature Selection, and Localized Prediction" *IEEE EPEC conference*, Saskatoon, Saskatchewan, Canada, 2017.

[4] H. Khorramdel, B. Khorramdel, M. T. Khorrami, and H. Rastegar, "A multiobjective economic load dispatch considering accessibility of wind power with here-and-now approach,"*J. Oper. Autom. Power Eng*., vol. 2, no. 1, pp. 49–59, 2014.

[5] A. Kavousi-Fard, A. Khosravi, and S. Nahavadi, "A new fuzzy based combined prediction interval for wind power forecasting," *IEEE Trans. Power Syst.,* vol. 99, no. 1, pp. 1–9, Jan. 2015.

[6] A. Haque, M. Nehrir, and P. Mandal, "A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting," *IEEE Trans. Power Syst.,* vol. 29, no. 4, pp. 1663–1672, Jul. 2014.

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

[8] H. Sangrody, et al. "On the Performance of Forecasting Models in the Presence of Input Uncertainty," *North American Power Symposium (NAPS)*, Morgantown, West Virginia, USA, 2017.

[9] C. Wan, Z. Xu, P. Pinson, Z. Dong, and K. Wong, "Optimal prediction intervals of wind power generation," *IEEE Trans. Power Syst*., vol. 29, no. 3, May 2014.

[10] R. J. Bessa, et al. "Time adaptive conditional kernel density estimation for wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 660–669, Oct. 2012.

[11] N.E. Huang et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. Lond. A*, vol. 454, pp. 903–995, 1998.

[12] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise assisted data analysis method," *Advances in Adaptive Data Analysis*, vol. 1, no. 1, pp. 1–41, 2009.

[13] Marcelo A. Colominas, Gastón Schlotthauer, and María E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing" *Biomedical Signal Processing and Control*, no. 14, 2014.

[14] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol. Heart Circulat. Physiol.*, vol. 278, no. 6, pp. 2039–2049, Jun. 2000.

[15] M. P. Wand and M. C. Jones, "Kernel smoothing," London: Chapman and Hall, 1995.

[16] 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.

[17] C.R. Rao and S.K. Mitra, "Generalized inverse of matrices and its applications," New York: Wiley, 1971.

[18] G. Giebel, G. Kariniotakis, and R. Brownsword, "State of the art on short-term wind power prediction," Anemos Project Deliverable Report D1.1, Tech. Rep., 2003. [Online]. Available: http://anemos.cma.fr.