A Fuzzy Adaptive Probabilistic Wind Power Prediction Framework Using Diffusion Kernel Density Estimators

B. Khorramdel, *Student Member IEEE*, C. Y. Chung, *Fellow, IEEE*, N. Safari, *Student Member IEEE*, and G. C. D. Price

***Abstract*—The inherent uncertainty in predicting wind power generation makes the operation and control of power systems very challenging. Probabilistic measurement of wind power uncertainty in the form of a reliable and sharp interval is of utmost importance, but construction of such high-quality prediction intervals (PIs) is difficult because wind power time series are non-stationary. In this paper, a framework based on the concept of bandwidth (BW) selection for a new and flexible kernel density estimator is proposed. Unlike previous related works, the proposed framework uses neither a cost function-based optimization problem nor point prediction results; rather, a diffusion-based kernel density estimator (DiE) is utilized to achieve high-quality PIs for non-stationary wind power time series. Moreover, to adaptively capture the uncertainties of both the prediction model and wind power time series in different seasons, the DiE is equipped with a fuzzy inference system and a tri-level adaptation function. The proposed framework is also founded based on a parallel computing procedure to promote the computational efficiency for practical applications in power systems. Simulation results demonstrate the efficiency of the proposed framework compared to well-known conventional benchmarks using real wind power datasets from Canada and Spain.**

***Index Terms*— Kernel density estimation, prediction intervals, probabilistic wind power prediction, wind power time series.**

I. NOMENCLATURE

A. Functions

$\kappa(\cdot)$	Kernel function.
$\varphi(\cdot)$	Gaussian kernel function.
$\hat{f}_t(\cdot), \hat{F}_t(\cdot)$	Estimated PDF and CDF for subinterval t.
$\delta(\cdot)$	Dirac delta function.
$\tau(\cdot)$	Pilot bandwidth function.
$\hat{\psi}(\cdot)$	Plug-in estimator function.
$\sigma(\cdot)$	Diffusion coefficient function.
$\boldsymbol{\xi}(\cdot)$	Adaptation function.
$\lambda(\cdot)$	Fuzzy function.
μ_L, μ_M, μ_H	Fuzzy membership functions.

B. Parameters

$h^*_{(\cdot)}$	Optimal bandwidth.
X_i	i^{th} wind power sample of variable x.
X_j	j^{th} wind power sample of variable y.
N_s	Number of samples inside each subinterval.
N_{test}	Number of future test samples.
N	Total number of ELM training sets.
N_{lag}	Number of time lags.
N_{sub}	Number of subintervals.
$\widetilde{N}$	Number of ELM hidden nodes.
$1-\alpha$	Nominal coverage probability of PIs.
m	Number of ELM outputs.
w_1, w_2	Adjusting parameters of fuzzy membership functions.

C. Variables

t	Variable indicating time.
$x\ , y$	Variables indicating wind power values.
h	Bandwidth of kernel functions.
γ	Bandwidth growth factor.
$\bar{w}$	Mean of samples inside each subinterval.
$Q_t^{(\alpha_l)}, Q_t^{(\alpha_u)}$	Lower and upper quantiles for subinterval t.

D. Matrices & Vectors

$\boldsymbol{\beta}$	Output weights matrix of ELM.
$\mathbf{G}$	Output vector of ELM.
$\mathbf{H}$	Input matrix of ELM.
$\boldsymbol{x}_i\ , \boldsymbol{g}_i$	Input and output vectors of ELM as i^{th} training set.

II. INTRODUCTION

WIND energy, one of the most widely used renewable energy sources around the world, brings huge uncertainty into power systems in a high penetration scenario, and thereby makes optimal decision-making problematic. This uncertainty originates from: (i) uncertainty in wind speed resulting from chaotic weather systems and (ii) nonlinear and uncertain characteristics of actual wind power curves. As such, wind power generation is uncertain and is represented by non-stationary time series [1]-[3]. Therefore, the prediction of wind power, as an essential part of modern power systems, is challenging.

Although diverse techniques have been proposed to reduce it, the unavoidable prediction error in point prediction approaches remains a problem that must be addressed [3]-[6]. The increasing penetration of wind power generation in existing power systems has resulted in the proposal of approaches to quantify wind power prediction (WPP) uncertainty, allowing power system operators to make optimal decisions to mitigate prediction error. One of the most well-known and widely-used methods of uncertainty representation is a probabilistic prediction approach that estimates a probability density function (PDF) or an interval for the uncertainty of wind power generation prediction [7]-[18]. Compared to point prediction, probabilistic prediction of future wind power generation provides much more meaningful and beneficial information for various decision-making problems in power systems such as economic dispatch, reserve allocation, optimal sizing of energy storage systems, wind farm control, stochastic unit commitment, and frequency dynamics constrained unit commitment [19]-[28]. In this context, prediction intervals (PIs) with specific confidence levels (CLs) from 90 to 99% and certain prediction horizons from minutes to days can be efficiently used for optimal operation of power systems using three different optimization strategies: robust optimization, interval optimization, and adjustable interval optimization [23]-[25].

Various PI construction approaches have been proposed to date; however, concerns regarding the quality of PIs remain. For example, lower upper bound estimation (LUBE) approach proposes a nonlinear and multi-objective function that should be minimized through heuristic optimization algorithms [7], [8]. It might be solved effi-

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Saskatchewan Power Corporation (SaskPower). Benyamin Khorramdel, C. Y. Chung, and Nima Safari 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). G. C. D Price is with Operations – System Control Centre, SaskPower, Regina, SK S4P 0S1 (e-mail: gprice@saskpower.com).

ciently, but dealing with nonlinear and multi-objective functions is a challenging issue because they might decrease the computational efficiency and be entrapped in local minima for some highly volatile time series. The quantile regression (QR) approach is also based on the minimization of a nonlinear objective function that is usually solved with particle swarm optimization (PSO) or genetic algorithm [9], [10]. However, the objective function can be linearized to be efficiently solved using a linear programming (LP)-based optimization problem [11]. The bootstrap-based extreme learning machine (BELM) approach uses point prediction results and assumes a standard Gaussian distribution for data noise and prediction model uncertainty [13]. Although BELM uses the excellent generalization ability and extremely low learning effort of ELM for single-hidden layer feedforward neural networks (SLFNs), it depends on the type of bootstrap method and the number of replicates, which make it computationally unproductive for large datasets. Autoregressive integrated moving average (ARIMA) as a classical time series-based approach does not utilize neural networks and can be used for probabilistic prediction as well [6], [16]. Conditional kernel density estimation (KDE) techniques utilizing *Nadaraya-Watson* kernel smoother or joint PDF formulation are other methods of probabilistic prediction. The main concern is the selection of a proper kernel function to avoid boundary effects and/or ignoring multi-modality, local uniformity, and long tail features of wind power PDFs [14], [15]. Furthermore, the bandwidth (BW) of these kernels is traditionally selected with plug-in techniques [29]-[32]. Importantly, the performance of this techniques strongly depends on the point prediction results. However, conditional density estimation using parametric distributions, i.e., versatile, truncated versatile, and beta distributions proposed for wind power PDF estimation, can be also used for PI construction [26]-[28]. A PI construction approach based on decomposition of original time series into trend, cycle, and noise components has also been proposed [16]. The first two components are predicted with deterministic approaches, and the lower/upper bounds of the noise component are provided by a probabilistic approach such as LUBE. Based on regular vine copulas, an advanced technology is used in [18] to model the dependence structures among wind farms for probabilistic prediction. However, it requires large datasets because its probabilistic WPP is based on historical point forecasts and real measurements.

In power systems with high penetration of wind power generation, an efficient probabilistic WPP approach is required to offer a highly reliable and sharp PI with low computational burden for practical applications. To achieve this goal, this paper focuses on a new nonparametric density estimation (NDE)-based approach. NDE is a very important tool for statistical analysis of power systems' data, and has a great potential for efficiently estimating any statistical features such as multimodality, high or low skewness, local uniformity, local modes, and other structures in the distribution of the data that are of value [29]. This paper refers to KDE as the most well-known NDE approach with BW as an important parameter [30]. Despite the huge body of literature, KDE suffers from three main problems: (i) the use of *normal reference rule* as a preliminary assumption in conventional BW selection techniques (i.e., plug-in technique) contradicts the motivation for using NDE [31], [32], (ii) conventional KDE approaches result in a tendency to ignore the peaks and valleys of the true density [33], and (iii) *boundary effects* might lead to invalid densities [34]-[35]. Although these problems have been mitigated to some extent using more advanced estimators, e.g., *balloon* estimators, *nearest neighbor* estimators [33], [36], *sample point adaptive* estimators [29], and *boundary kernel* estimators [37], these solutions are still unsatisfactory due to the high computational burden and/or invalid densities. In the context of wind power density estimation, the above problems would lead to unsatisfactory results mainly because wind power datasets are double-bounded and presents special features that change over time.

To address these problems, this paper uses a flexible density estimator called the diffusion-based kernel density estimator (DiE), which is based on the smoothing properties of linear diffusion processes [38]. A novel wind power PI construction framework is also proposed based on the DiE that can present highly reliable and sharp PIs. The proposed framework is formed based on a PDF estimation procedure over consecutive short time intervals (also referred to as subintervals) of wind power time series. Because PDFs can present complete information (e.g., mean, variance, lower/upper quantiles, etc.) of wind power samples (WPSs) inside each subinterval, they are highly beneficial for analyzing non-stationary wind power time series. After completing the PDF estimation procedure for each subinterval, lower/upper quantiles are obtained and stored in a database. To this end, a historical wind power dataset is initially subdivided into numerous subintervals to create a historical piecewise dataset. Then, the DiE along with an efficient BW selection technique is implemented for each subinterval to estimate historical wind power PDFs. Finally, lower/upper quantiles of each PDF are obtained to create a historical quantiles dataset. In this context and in contrast to kernel functions used in the literature, the DiE can flexibly deal with amorphous wind power PDFs with changing features using a well-defined fuzzy inference system. In addition, through a parallel computing process, the proposed fuzzy DiE uses an adaptation function, with a parameter called BW growth factor, to adapt the DiE's BW to capture the uncertainty of the prediction model and consider the seasonality of wind power time series on the PDF estimation procedure. Hence, this paper proposes a fuzzy and adaptive DiE-based PI construction framework (FADiE) to deliver high quality PIs. The proposed framework employs a completely different strategy than previous works [26]-[28] to deal with the historical data. The approach in [26]-[28] requires at least one year of historical data, and the forecast range [0,1] (p.u.) is divided into a few forecast bins, e.g., 25. Then, based on a forecast value in the next step, a PDF is fitted to the historical error samples inside the related bin.

To the best of the authors' knowledge, the DiE with its efficient BW selection technique has not been used with respect to wind power datasets and PI construction. Because the FADiE framework does not require widely-used optimization techniques (i.e., (non)linear programming and heuristic optimization algorithms), point prediction results, and any assumptions regarding prediction error and data noise distribution, it can be simply used in practical applications. The main contributions of this paper can be summarized as follows:

- The concept of optimal BW selection for a new and flexible density estimator is introduced for the first time to construct high-quality PIs for wind power time series.
- A piecewise wind power PDF estimation procedure is introduced using piecewise and successive wind power sample sets.
- Three trapezoidal fuzzy sets are proposed to tune the flexibility of the proposed kernel density estimator for double-bounded wind power time series to avoid boundary effects.
- A tri-level adaptation function is proposed to model the uncertainty of the prediction model and variability (seasonality) of wind power time series.
- A parallel computing process is proposed to increase the computational efficiency and remove the widely-used cost function-based optimal PI construction methodologies.

This paper is organized as follows. In Section III, the main problems regarding wind power PDF estimation with conventional and modern KDE techniques are presented. Section IV presents optimal BW selection techniques for double-bounded and non-stationary wind power time series. The proposed FADiE framework is described in Section V. Comprehensive simulation results are analyzed

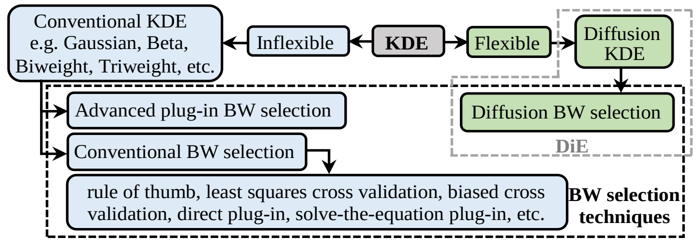


Fig. 1. Diagram of KDE technique variants and corresponding BW selection techniques.

in Section VI. Finally, Section VII concludes the paper.

III. Problem Statement for Wind Power PDF Estimation Using KDE Techniques

This section briefly introduces the main characteristics of wind power time series and the existing problems for wind power PDF estimation, then presents conventional and new KDE techniques.

A. Wind Power Time Series Characteristics and PDF Estimation Problems

Wind power time series have four main characteristics: (i) non-stationary, (ii) double-bounded ([0,1] (p.u.)), (iii) PDFs containing special features, and (iv) prediction error with high skewness and kurtosis [2], [28]. Under such conditions, the PDF of WPSs (for a subinterval) and WPP error (for a dataset) might be estimated by two methods: (a) using parametric distributions, such as Gaussian, beta, t-student, and so on and (b) using conventional KDE techniques. Although the latter has more satisfactory performance than the former, both methods fail to accurately estimate the underlying PDF in a dataset and can lead to inefficient results [14], [15], [17], as discussed in the Introduction. The classification of different kinds of KDE techniques with related BW selection techniques is illustrated in Fig. 1. The performance of KDE-based PI construction approaches depends on the type of kernel and properly adjusting the BW; therefore, inflexible KDE techniques (as opposed to flexible ones) encounter three main problems, $\mathbf{P_1}$ to $\mathbf{P_3}$, as described in [29], [30], [32], [38].

$\mathbf{P_1}$: They use either an inherently false assumption, i.e., normal reference rule, or an inefficient BW selector in the conventional selection techniques. $\mathbf{P_2}$: They tend to flatten the wind power PDF's peaks and valleys. $\mathbf{P_3}$: They suffer from boundary effects. These problems are addressed in the proposed framework in Sections IV and V in more detail.

B. Conventional Kernel Density Estimate of Wind Power

Given N_s independent stochastic WPSs $\mathcal{X}_{N_s} \equiv \{X_1, \dots, X_{N_s}\}$ over subinterval t, the unknown underlying PDF of the wind power is estimated by the kernel density estimate of f as:

$$\hat{f}_t(x) = \frac{1}{N_s}\sum_{i=1}^{N_s} \kappa(x, X_i; h) \qquad x \in \mathbb{x} = [0,1] \tag{1}$$

where $\kappa(\cdot)$ is the kernel function with parameter h as the BW. Because wind power time series present a double-bounded dataset, WPSs take values between zero and a nominal capacity (1 p.u.). The Gaussian kernel function in (2) has been widely used in the literature [17], [30], [38]. In this paper, the Gaussian kernel density estimator (GE) is denoted as a conventional KDE technique and compared with the DiE.

$$\kappa(x, X_i; h) = \varphi(x, X_i; h) = \frac{\exp\left(-\left((x - X_i)/\sqrt{2}h\right)^2\right)}{\sqrt{2\pi}\,h} \tag{2}$$

C. Diffusion Kernel Density Estimate of Wind Power

In this flexible kernel density estimator, the unknown wind power PDF f is approximated by the kernel density estimator $\hat{f}$, which is based on the smoothing properties of the general linear partial differential equation (PDE) in (3).

$$\frac{\partial \hat{f}(x;t)}{\partial t} = L\left(\hat{f}(x;t)\right) \qquad t > 0\ ,\ x \in \mathbb{x} = [0,1] \tag{3}$$

$$\hat{f}(x;0) = \Delta x = \tfrac{1}{N_s}\textstyle\sum_{i=1}^{N_s} \delta(x - X_i) \tag{4}$$

where Δx in the initial condition (4) is the empirical density of data on $\mathbb{x}$. The linear differential operator $L(\cdot)$, expressed by (5), includes two arbitrary positive functions, $a(x)$ and $p(x)$, that meaningfully affect the performance of the DiE. The boundary condition (6) should also be considered to ensure that $\hat{f}$ integrates to unity [38].

$$L(\cdot) = d\left(a(x)\, d(\cdot/p(x))/dx\right)/2dx \tag{5}$$

$$\partial\left(\hat{f}(x;t)/p(x)\right)/\partial x\big|_{\partial\mathbb{x}} = 0 \tag{6}$$

The solution to (3) is a flexible kernel density estimator if $p(x)$ is a valid PDF on the dataset $\mathbb{x}$ and $p(x) = \lim_{t\to\infty} \hat{f}(x;t)$. The function $p(x)$ is estimated by GE and named the pilot PDF. In (5), the operator $L(\cdot)$ can be adjusted when $a(x) = p(x)^\lambda$ and $\lambda \in [0,1]$. Therefore, a general solution to (3) is written in the form of (7), in which PDEs (8)-(9) are satisfied for fixed values of y and x, respectively.

$$\hat{f}(x;t) = \tfrac{1}{N_s}\textstyle\sum_{i=1}^{N_s} \kappa_{\mathrm{DiE}}(x, X_i; t) \tag{7}$$

$$\partial\left(\kappa_{\mathrm{DiE}}(x,y;t)\right)/\partial t = L\left(\kappa_{\mathrm{DiE}}(x,y;t)\right),\ x \in \mathbb{x}\ ,\ t > 0 \tag{8}$$

$$\partial\left(\kappa_{\mathrm{DiE}}(x,y;t)\right)/\partial t = L^*\left(\kappa_{\mathrm{DiE}}(x,y;t)\right),\ y \in \mathbb{x}\ ,\ t > 0 \tag{9}$$

where $L^*(\cdot) = \partial\left(a(y)\,\partial(\cdot)/\partial y\right)/\partial y/2p(y)$ is the adjoint operator of $L(\cdot)$. For wind power time series, boundary conditions (10) and (11) should be applied to guarantee that $\hat{f}(x;t)$ is a valid PDF that integrates to one. The general form of the DiE in which the general parameters x and y are defined to satisfy constraints (8)-(11) on $\mathbb{x}$ is expressed in the form of (12).

$$\partial\left(\kappa_{\mathrm{DiE}}(x,y;t)/p(x)\right)/\partial x|_{\partial\mathbb{x}} = 0 \tag{10}$$

$$\partial\left(\kappa_{\mathrm{DiE}}(x,y;t)\right)/\partial y|_{\partial\mathbb{x}} = 0 \tag{11}$$

$$\kappa_{\mathrm{DiE}}(x, y;\ h_{\mathrm{DiE}}) = \frac{p(x).\exp\left(-\int_y^x \sigma^{-1}(z)\,dz \Big/ \left(\sqrt{2}\,h_{\mathrm{DiE}}\right)^2\right)}{\sqrt{2\pi}\ h_{\mathrm{DiE}}\ \sqrt[4]{p(x)a(x)p(y)a(y)}} \tag{12}$$

where $h_{\mathrm{DiE}} = \sqrt{t}$ is the BW of diffusion kernel κ_{DiE}, and $\kappa_{\mathrm{DiE}}(x,y;0) = \delta(x-y)$. Also, $\sigma(x) = \sqrt{a(x)/p(x)}$ which includes $a(x)$ and $p(x)$ is called the diffusion coefficient and assists the diffusion kernel (12) in diffusing the initial density Δx at a different rate to provide a plausible smoothing property to extract the important features of the wind power PDF. If $a(x) = p(x) = 1$, the PDE in (3) would be the well-known Fourier heat equation with the conventional inflexible GE in (2) as its solution. The approach that estimates $p(x)$ and $a(x)$ is explained in the next section. The flexible KDE technique introduced herein, referred to as DiE, can solve the problems related to practical applications of KDE-based approaches. Because the performance of all kernel density estimators crucially depends on optimal BW selection as the cornerstone of wind power PDF estimation, different BW selection techniques are explained in the next section.

IV. Optimal BW Selection Techniques for Wind Power PDF Estimation

With the intent to implement the DiE for proposing a novel wind power PI construction framework, this section first describes two well-known BW selection techniques, i.e., conventional direct plug-in (DPI) and advanced plug-in (API). An efficient BW selection technique is then introduced for the DiE. Finally, the performance of kernel density estimators is evaluated using these techniques through wind power PDF and quantiles estimation.

A. Optimal BW Selection: Criterion and Techniques

A well-defined criterion for optimal BW selection is the mean integrated squared error ($MISE$) expressed by (13), which can be

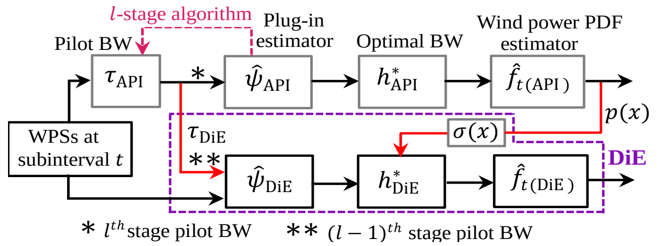


Fig. 2. General diagram of wind power PDF estimation via optimal BW selection techniques.

divided into two components: the integrated squared bias and integrated variance. It is proven with technical details in [29], [30], [38] that an optimal BW minimizes the first-order asymptotic approximation of $MISE$:

$$MISE = \int \left(\mathbb{E}_f\left(\hat{f}_t(x)\right) - f_t(x)\right)^2 dx + \int Var_f[\hat{f}_t(x)]dx \tag{13}$$

where $\mathbb{E}_f(\cdot)$ and $Var_f(\cdot)$ are respectively the expectation and variance operators. A large BW leads to over-smoothing and can result in loss of some features in the wind power PDF, while an overly small BW generates a PDF with many peaks that are not meaningful. To extract diverse features such as multi-modality, local uniformity, and long tails within the probability distribution of wind power time series, efficient BW selection techniques should be used in conjunction with flexible kernels. Otherwise, the approach fails to accurately capture the features [29], as will be shown in Section IV-B.

A common property among DPI, API, and DiE optimal BW selection techniques is the existence of an intermediate estimator called the plug-in estimator ($\hat{\psi}$) and an intermediate BW called the pilot BW (τ). The distinction between these techniques arises from diverse l-stage algorithms for calculation of τ and $\hat{\psi}$ that lead to completely different performances.

1) Direct plug-in BW selection technique for GE

In this technique, DPI optimal BW, denoted by h^*_{DPI}, which minimizes the asymptotic approximation of $MISE$, is found using (14)-(15). Because the DPI BW selection technique depends on the pilot BW τ_{DPI} in (14) and the r^{th} derivative of standard Gaussian kernel $\kappa_{\mathrm{G}}^{(r)}$ with second moment $\mu_{2(\kappa_{\mathrm{G}})} = \int_{\mathbb{R}} x^2 \kappa_{\mathrm{G}}(x)dx$, an l-stage DPI BW selector is developed in algorithm 1 [30].

$$\tau_{\mathrm{DPI}} = \left(-2\,\kappa_{\mathrm{G}}^{(r)}(0)/N_s\,\mu_{2(\kappa_{\mathrm{G}})}\,\hat{\psi}_{\mathrm{DPI}}^{(r+2)}\right)^{1/r+3} \tag{14}$$

$$\hat{\psi}_{\mathrm{DPI}}^{(r)} = \frac{1}{{N_s}^2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s} \kappa_{\mathrm{G}}^{(r)}(X_i, X_j; \tau_{\mathrm{DPI}}) \tag{15}$$

Algorithm 1 DPI BW selection ($r \in 2n, r \geq 6$)

1: Set l, e.g., l=5, and then $r = 2l + 4$.

2: Calculate σ_s as the standard deviation of N_s random samples then set the initial value of $\hat{\psi}_{\mathrm{DPI}}^{(r)}$ with $\hat{\psi}_{\mathrm{DPI}}^{(ini)} = (-1)^{r/2}\, r!/(\sqrt{\pi}(2\sigma_s)^{r+1}(r/2)!)$.

3: Find pilot BW τ_{DPI} using (14) and then $\hat{\psi}_{\mathrm{DPI}}^{(r)}$ using (15).

4: Continue the process to obtain $\hat{\psi}_{\mathrm{DPI}}^{(4)}$, then use (14) to obtain the optimal value of BW as $h^*_{\mathrm{DPI}} = \tau_{\mathrm{DPI}}|_{r=2}$.

As a conventional BW selection technique, the DPI technique is inefficient for wind power PDF estimation because it is too smooth and ignores the main features of the wind power PDF over its main interval [0,1]. In this paper, the Gaussian estimator that uses the DPI technique is denoted by $\mathrm{GE_{DPI}}$, and the estimated PDF is indicated by $\hat{f}_{t(\mathrm{DPI})}(x)$.

2) Advanced plug-in BW selection technique for GE

This technique finds the minimum value of the asymptotic approximation of $MISE$ using the optimal BW h^*_{API} shown in (16), where an l-stage algorithm is stated in detail to calculate the pilot BW $\tau_{\mathrm{API}}^{(1)}$ using (17)-(18). The wind power PDF estimated by this technique (i.e., $\hat{f}_{t(\mathrm{API})}(x)$) is then used to implement the DiE BW selection technique. In this study, the Gaussian estimator equipped with the API technique is referred to as $\mathrm{GE_{API}}$. In (18), the j^{th} derivative of the Gaussian kernel φ is shown by $\varphi^{(j)}$ [38].

$$h^*_{\mathrm{API}} \cong \left(0.9\,\tau_{\mathrm{API}}^{(1)}\right)^{\frac{1}{2}} \tag{16}$$

$$\tau_{\mathrm{API}}^{(J)} = \left[\frac{\left(1+1/(2^{J+0.5})\right)\left((2J-1)\times\ldots\times3\times1\right)}{\left(3N_s\sqrt{\pi/2}\,\hat{\psi}_{\mathrm{API}}^{(J+1)}\right)}\right]^{\frac{2}{3+2J}} \tag{17}$$

$$\hat{\psi}_{\mathrm{API}}^{(J)} = \frac{(-1)^J}{{N_s}^2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s}\varphi^{(2J)}\left(X_i, X_j; 2\tau_{\mathrm{API}}^{(J)}\right) \tag{18}$$

Algorithm 2 API BW Selection

1: Choose l, e.g., $l = 5$.

2: Set the pilot BW $\tau_{\mathrm{API}}^{(l)}$ to a small value, e.g., 0.001, and find the plug-in estimator $\hat{\psi}_{\mathrm{API}}^{(l)}$ using (18).

3: Find the pilot BW $\tau_{\mathrm{API}}^{(l-1)}$ via (17).

4: Find the plug-in estimator $\hat{\psi}_{\mathrm{API}}^{(l-1)}$ using (18) and $\tau_{\mathrm{API}}^{(l-1)}$ obtained from the previous stage, and continue this procedure until $\hat{\psi}_{\mathrm{API}}^{(2)}$ and consequently $\tau_{\mathrm{API}}^{(1)}$ are acquired.

5: If $\left|\tau_{\mathrm{API}}^{(1)} - \tau_{\mathrm{API}}^{(l)}\right| < \varepsilon$, equation (16) gives the optimal BW for the Gaussian estimator (1); else go to step 2 with $\tau_{\mathrm{API}}^{(l)} = 0.9\,\tau_{\mathrm{API}}^{(1)}$.

3) BW selection technique for DiE

DiE BW selection leads to the minimum value of asymptotic $MISE$ using the DiE optimal BW in (19) where the plug-in estimator $\hat{\psi}_{\mathrm{DiE}}$ is estimated by (20) using the pilot BW τ_{DiE} and the flexible kernel κ_{DiE}. The value of τ_{DiE} is twice that of τ_{API} in the $(l-1)^{th}$ stage of the API technique. Because $\mathbb{E}_f[\sigma^{-1}(x)]$ in (21), $L(\cdot)$, and $L^*(\cdot)$ depend on $a(x)$ and $p(x)$, algorithm 2 is first executed to estimate $p(x) = \hat{f}_{t(\mathrm{API})}(x)$. Then, by adjusting $\lambda \in [0,1]$, $a(x)$ is acquired using $a(x) = p(x)^{\lambda}$.

$$h^*_{\mathrm{DiE}} \cong \left(0.5\,\mathbb{E}_f[\sigma^{-1}(x)]/\left(N_s\sqrt{\pi}\,\hat{\psi}_{\mathrm{DiE}}\right)\right)^{\frac{1}{5}} \tag{19}$$

$$\hat{\psi}_{\mathrm{DiE}} = \frac{1}{{N_s}^2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s} L^*\left(L\left(\kappa_{\mathrm{DiE}}(X_i, X_j; \tau_{\mathrm{DiE}})\right)\right) \tag{20}$$

$$\mathbb{E}_f[\sigma^{-1}(x)] = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(a(X_i)/p(X_i)\right)^{-\frac{1}{2}} \tag{21}$$

Algorithm 3 Diffusion BW Selection

1: Find $p(x)$ using the implementation of GE with h^*_{API}.

2: Set $a(x) = p(x)^{\lambda}$ with $\lambda \in [0,1]$.

3: Calculate the value of $\mathbb{E}_f[\sigma^{-1}(x)]$ using (21).

4: Find the value of $\hat{\psi}_{\mathrm{DiE}}$ using (20), where $\tau_{\mathrm{DiE}} = 2\,\tau_{\mathrm{API}}^{(2)}$ and $\tau_{\mathrm{API}}^{(2)}$ is determined from step 4 in **Algorithm 2**.

5: Calculate h^*_{DiE} using (19) as the DiE optimal BW.

A general diagram of BW selection techniques is provided in Fig. 2. A normal reference rule is not used; instead, diffusion coefficient $\sigma(x)$ and an l-stage algorithm are utilized to efficiently tune the flexibility of DiE (problem $\mathbf{P_1}$ is removed). Due to its similarity to API process, the DPI process is not shown in Fig. 2. The DiE utilized herein, in which a diffusion-based BW selection technique is used, offers far greater flexibility in modeling a given dataset with high accuracy and consistency [38]. The PDF estimated by DiE is denoted by $\hat{f}_{t(\mathrm{DiE})}(x)$.

B. Estimation of Wind Power PDFs, Quantiles and Intervals

The above-mentioned inflexible ($\mathrm{GE_{DPI}}$-$\mathrm{GE_{API}}$) and flexible (DiE) kernel density estimators are implemented and compared in this section. Using a wind power dataset, PDFs over three different subintervals are estimated to see how well the important features of WPSs can be extracted via DiE. First, using optimal

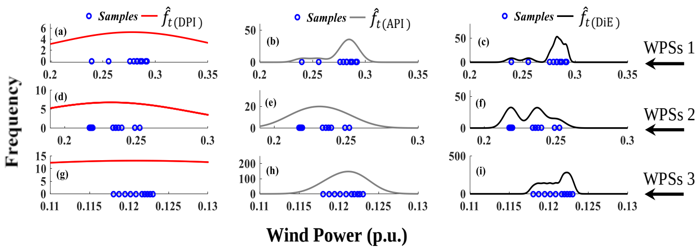


Fig. 3. Comparison of KDE techniques for three different sets of WPSs drawn from the Centennial wind farm dataset (located in South Saskatchewan, Canada).

TABLE I
THE RESULTS OF INTERVAL CALCULATION FOR CL=95%.

KDE	WPSs 1		WPSs 2		WPSs 3	
	$[Q_t^{(\alpha_l)}, Q_t^{(\alpha_u)}]$	$Int.$	$[Q_t^{(\alpha_l)}, Q_t^{(\alpha_u)}]$	$Int.$	$[Q_t^{(\alpha_l)}, Q_t^{(\alpha_u)}]$	$Int.$
$\mathrm{GE_{DPI}}$	[0.128,0.425]	0.297	[0.117,0.348]	0.231	[0.062,0.180]	0.118
$\mathrm{GE_{API}}$	[0.232,0.301]	0.069	[0.196,0.270]	0.074	[0.115,0.126]	0.011
DiE	**[0.236,0.294]**	**0.058**	**[0.211,0.257]**	**0.046**	**[0.118,0.120]**	**0.002**

BWs h^*_{DPI}, h^*_{API} (with Gaussian kernel) and h^*_{DiE} (with diffusion kernel) the associated PDFs are respectively estimated as shown in Fig. 3 (a)-(i). Thereafter, using corresponding cumulative distribution functions $F_{t(\cdot)}(x)$, obtained from (22), one can calculate the lower/upper quantiles $Q_t^{(\alpha_l)}$ and $Q_t^{(\alpha_u)}$ and the related interval width using (23) and (24), respectively.

$$F_{t(\cdot)}(x) = \int_0^x \hat{f}_{t(\cdot)}(\omega)d\omega \tag{22}$$

$$Q_t^{(\alpha_l)} = F_{t(\cdot)}^{-1}(\alpha/2)\ ,\ Q_t^{(\alpha_u)} = F_{t(\cdot)}^{-1}(1-\alpha/2) \tag{23}$$

$$Int. = Q_t^{(\alpha_u)} - Q_t^{(\alpha_l)} \tag{24}$$

where $\alpha_l = \alpha/2$ and $\alpha_u = 1-\alpha/2$ lead to CL=100×$(1-\alpha)$%.

Fig. 3 and Table I show that, unlike $\mathrm{GE_{DPI}}$ and $\mathrm{GE_{API}}$, the DiE can efficiently identify existing features of wind power PDFs (e.g., multi-modality, local uniformity, long tail, and high skewness) and consequently eliminates problem $\mathbf{P_2}$. To provide a better sense of wind power interval estimation, the widths of intervals in KDE techniques are shown in Table I for CL=95%. Observe that narrower intervals are obtained by the DiE. Based on this superior performance, sharp PIs are constructed in the next section in the context of probabilistic WPP.

V. THE PROPOSED FADIE FRAMEWORK FOR OPTIMAL WIND POWER PI CONSTRUCTION

The proposed FADiE framework aims to mitigate the drawbacks of conventional PI construction approaches, such as dependency on historical prediction results, assuming parametric distributions, and definition of certain objective functions based on reliability and sharpness of PIs in an optimization framework. It utilizes four building blocks to construct optimal PIs: (i) DiE with its efficient BW selection technique, (ii) a fast and efficient prediction model (i.e., ELM) [11], [13], [16], [39], (iii) three trapezoidal fuzzy sets, and (iv) a tri-level adaptation function. The fuzzy sets are defined according to WPSs average values and used for DiE flexibility tuning, by adjusting the parameter λ, to avoid boundary effects. The adaptation function provides an adaptive procedure for the fuzzy DiE to create diverse lower/upper quantiles datasets with different average interval widths to model time series seasonality and prediction model uncertainty. The following sections introduce wind power PI evaluation criteria, then explain the building blocks (ii)–(iv) in more detail followed by FADiE framework stages.

A. Wind Power PIs Evaluation Criteria

Three important indices are used to assess the quality of constructed wind power PIs by the proposed FADiE framework.

1) Reliability

The average coverage error (ACE) in (25), which is the deviation of the PI coverage probability ($PICP$) in (26) from its nominal coverage ($PINC$) should be positive and too close to zero to guarantee the high reliability of the PIs as the main feature.

$$0 \le ACE = PICP - PINC < \varepsilon \tag{25}$$

$$PICP = \frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}} \mathbb{1}_{[L_i^\alpha, U_i^\alpha]}(y_i) \tag{26}$$

$$\mathbb{1}_{[L_i^\alpha, U_i^\alpha]}(y_i) = \begin{cases} 1 & y_i \in [L_i^\alpha, U_i^\alpha] \\ 0 & y_i \notin [L_i^\alpha, U_i^\alpha] \end{cases} \tag{27}$$

where $\mathbb{1}_{[L_i^\alpha, U_i^\alpha]}(\cdot)$ is an indicator function, L_i^α and U_i^α are, respectively, the lower and upper bounds of wind power PI associated with the prediction target, y_i. Note that $PICP$ is generally used to illustrate the probability that future wind power, y_i, as a target, will be enclosed by the interval $[L_i^\alpha, U_i^\alpha]$.

2) Sharpness

To perceive meaningful information from a PI, its normalized average width ($PINAW$) in (28) should take small values to induce a sharp PI.

$$PINAW = \frac{1}{R.N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}(U_i^\alpha - L_i^\alpha) \tag{28}$$

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

3) Overall score

To assess the overall skill of a PI construction approach, the overall score in (29) is considered in evaluation process because it simultaneously takes both reliability and sharpness aspects into account. A sharp PI presents a small value for $|Sc|$.

$$Sc = \frac{1}{N_{\text{test}}}\sum_{i=1}^{N_{\text{test}}}\Big[-2\alpha(U_i^\alpha - L_i^\alpha) - 4\mathbb{1}_{[0,L_i^\alpha)}(y_i)(L_i^\alpha - y_i) - 4\mathbb{1}_{(U_i^\alpha,1]}(y_i)(y_i - U_i^\alpha)\Big] \tag{29}$$

Since very sharp PIs may violate the reliability criterion, the overall score index, Sc, in which the reliability is also included can reasonably reflect the real sharpness of PIs.

B. The Prediction Model for Lower/Upper Quantiles Prediction

In gradient-based traditional approaches for neural network training, some unavoidable limitations include high computational effort to tune the parameters, slow learning procedure, and overtraining [13]. Therefore, the proposed framework uses ELM as an easy-to-implement learning algorithm for training SLFNs with excellent generalization ability, extremely low learning effort, and high ability to avoid local minima and overtraining [39]. In the ELM approach, if the activation functions in the hidden layer are infinitely differentiable, by randomly selecting the input weights and biases, SLFNs can be viewed as a simple linear system with the output weights analytically determined using a generalized inverse operation.

Considering N different training sets $(\boldsymbol{x}_i, \boldsymbol{g}_i)|_{i=1}^N$ drawn from N_{d} days of a dataset, ELM with $\tilde{N}$ hidden nodes is expressed by $\mathbf{H}_{N\times\tilde{N}}\boldsymbol{\beta}_{\tilde{N}\times m} = \mathbf{G}_{N\times m}$. Input and output vectors $\boldsymbol{x}_i$ and $\boldsymbol{g}_i$ are shown in (30) and Fig. 4, and the output weight matrix, $\boldsymbol{\beta}$, and the target matrix, $\mathbf{G}$, are denoted by (31).

$$\boldsymbol{x}_i = [x_{i1}, \dots, x_{in}]^T \in \mathbb{R}^n, \boldsymbol{g}_i = [g_{i1}, \dots, g_{im}]^T \in \mathbb{R}^m \tag{30}$$

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_1 \quad \dots \quad \boldsymbol{\beta}_{\tilde{N}}]^T\ ,\ \mathbf{G} = [\boldsymbol{g}_1 \quad \dots \quad \boldsymbol{g}_N]^T \tag{31}$$

where $\boldsymbol{\beta}_j = \left[\beta_{j1}, \dots, \beta_{jm}\right]^T$ is the output weight vectors. After calculation of $\mathbf{H}$ according to [13] or [39], matrix $\boldsymbol{\beta}$ is obtained using $\boldsymbol{\beta} = \mathbf{H}^\dagger\mathbf{G}$ where $\mathbf{H}^\dagger$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$ [40]. To train ELM, the input matrix $\mathbf{H}$ and the output matrix $\mathbf{G}$ should be constructed based on input vector $\boldsymbol{x}_i$ (i.e., a set of WPSs) and output vector $\boldsymbol{g}_i$ (i.e., lower and upper quantiles). A temporal diagram of ELM input and output data is provided in Fig. 4, where historical datasets are divided into N_{sub} subintervals, $\boldsymbol{x}_i$ contains N_{lag} time lags with the same sample size N_{s}(e.g., one-hour samples for each lag), and $\boldsymbol{g}_i$ includes the lower/upper quantiles for one step ahead. In this paper, the 10-fold cross validation technique in [41] is used to identify the optimal number of time lags and hidden nodes

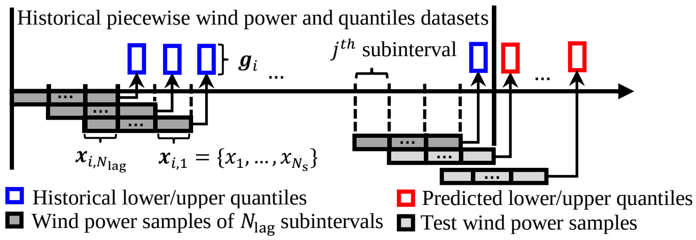


Fig. 4. Temporal diagram of input and output data structure for ELM training.

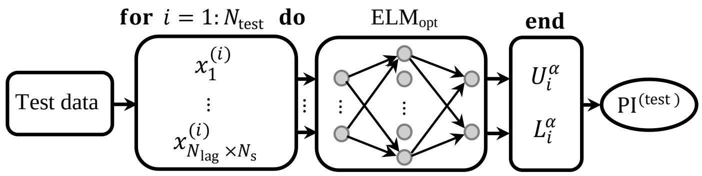


Fig. 5. Flowchart of the main prediction process.

and optimal length of the training dataset, which includes N input vectors $[\boldsymbol{x}_i]_{(N_{\text{lag}} \times N_s) \times 1}$ and N output vectors $[\boldsymbol{g}_i]_{2\times 1}$ where $N = N_{\text{sub}} - N_{\text{lag}}$ and $N_{\text{sub}} = N_{\text{d}} \times 24 \times 6 / N_s$.These optimal variables are obtained by minimizing the mean absolute error or root mean square error to avoid underfitting and overfitting. After optimal training of the ELM, the validation or test datasets can be used to construct corresponding PIs by the generation of lower and upper bounds L_i^α and U_i^α for each test sample as shown in the flowchart in Fig. 5.

C. *The Proposed Trapezoidal Fuzzy Sets for Flexibility Tuning*

To capture wind power uncertainty by estimating bona fide PDFs over successive subintervals, the DiE takes the advantage of trapezoidal fuzzy sets (μ) shown on the left axis of Fig. 6. Using the average value of WPSs (ϖ) inside each subinterval, the fuzzy sets tune the flexibility of the DiE through (32) to avoid boundary effects. Fig. 7 shows the boundary effects with an ellipse with a probability allocation for values outside of [0,1].

$$\lambda(\varpi, \mu) = (1 - \mu_L)\mathbb{1}_{[0,\varpi_1]}(\varpi) + \mu_M \mathbb{1}_{[\varpi_1,\varpi_2]}(\varpi) + (1 - \mu_H)\mathbb{1}_{[\varpi_2,1]}(\varpi) \tag{32}$$

The philosophy behind the proposed trapezoidal fuzzy sets is to alleviate problem $\mathbf{P_3}$. If the average of the WPSs is near boundaries, i.e., low-power ($\varpi \in [0, \varpi_1^-]$) and high-power ($\varpi \in [\varpi_2^+, 1]$) regions where boundary effects might happen, the DiE sets λ=0 to generate sharp and bona fide PDFs as shown in Fig. 7. For the medium power region ($\varpi \in [\varpi_1^+, \varpi_2^-]$), where boundary effects do not matter, the DiE chooses $\lambda = \mu_M$ ($0 < \mu_M \le 1$) to estimate smoother PDFs while preserving the main features.

D. *The Proposed Tri-Level Adaptation Function for Wind Power PI Reliability Improvement*

The main factors that reduce the reliability of a PI are associated with the uncertainties originating from chaotic wind power datasets over different seasons and misspecification of ELM parameters, e.g., training based on non-informative samples and randomly generated input weights and biases. The proposed adaptation function $\boldsymbol{\xi}(\varpi, \gamma)$, shown on the right axis of Fig. 6, aims to adaptively raise the reliability of constructed PIs under such conditions. Using (33)-(34), the function $\boldsymbol{\xi}$ leads to the controlled growth of the DiE BW for the subintervals in which the value of ϖ is far from the boundaries and the boundary effect does not matter for the PDFs. In (34), γ is the BW growth factor, and ϖ_i^- and ϖ_i^+ are the average values of wind power smaller and larger than ϖ_i. The value of $\boldsymbol{\xi}$ for low-power and high-power regions is set to unity to prevent boundary effects. For the medium-power region, the growth factor γ increases the DiE BW for generation of smoother PDFs or large intervals to capture the aforementioned uncertainties. Also, between medium and (low) high-power regions, the adaptation function applies an average value

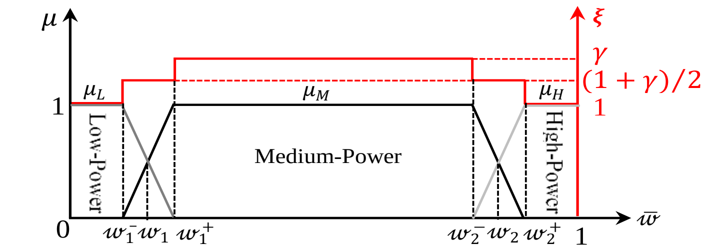


Fig. 6. Proposed trapezoidal fuzzy sets (left axis) and tri-level adaptation function $\boldsymbol{\xi}(\varpi, \gamma)$ (right axis) considered for the proposed DiE.

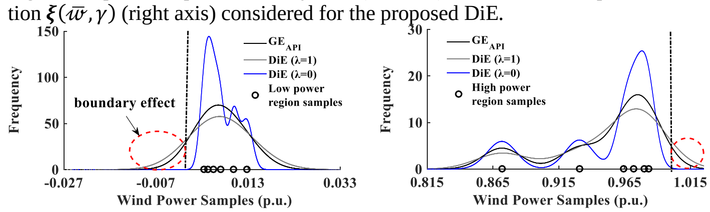


Fig. 7. Illustration of boundary effects for WPSs near boundaries.

for BW growth factor to create a trade-off between both regions. Note that the optimal value of γ that ultimately results in a reliable and sharp PI might be different for diverse datasets or even different seasons of a dataset.

$$h_{\text{DiE}} = \boldsymbol{\xi}.h_{\text{DiE}}^* \tag{33}$$

$$\boldsymbol{\xi} = \mathbb{1}_{[0,\varpi_1^-] \cup [\varpi_2^+,1]}(\varpi) + \left(\frac{1+\gamma}{2}\right)\mathbb{1}_{(\varpi_1^-,\varpi_1^+] \cup [\varpi_2^-,\varpi_2^+)}(\varpi) + (\gamma)\mathbb{1}_{(\varpi_1^+,\varpi_2^-)}(\varpi) \tag{34}$$

E. *The Proposed FADiE Framework Stages for Optimal Construction of Wind Power PI*

Three stages should be followed to implement the proposed FADiE framework for real wind power datasets. Before the first stage, the wind power dataset, including training, validation and test datasets, is preprocessed and normalized. Then, in a parallel computing process, considering M values of BW growth factor for adaptation function $\boldsymbol{\xi}$, i.e., $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_{\text{M}}]$, M groups of wind power PDFs in each subinterval in the original training dataset are estimated via the fuzzy DiE. Thereafter, M series of lower/upper quantiles with nominal coverage probability α are calculated and stored in a database (see Fig. 8). A 10-fold cross validation technique is then run for case $\gamma = 1$. Note that to update the prediction tool with the most recent quantiles data, the process in Fig. 8 should be repeated over time after each lower/upper quantile prediction in Fig. 5.

First Stage: In this stage, ELM_1 to ELM_{M} are trained with a parallel procedure using the optimal training dataset. The sensitivity analysis provided in the case studies section, shows that a limited number of ELMs should be considered even for highly chaotic time series to find γ_{opt} to simultaneously satisfy reliability and sharpness criteria. *Second Stage*: A validation dataset is first used as input data for the trained ELM_1 to ELM_{M} to generate PI_1 to PI_{M}, respectively. Then, ACE and $|S_c|$ are calculated for PI_1 to PI_{M} to identify the best ELM that results in a PI with high reliability and sharpness. *Third Stage*: The superior ELM obtained from the second stage, i.e., ELM_{opt}, is used to construct a reliable and sharp PI for the test dataset with the prespecified CL. An in-depth structure of the parallel computing-based FADiE framework is shown in Fig. 9.

VI. Case Studies

A. *Experimental Datasets*

To assess the efficiency of the proposed FADiE framework, four wind power datasets are considered.

Case 1: Canada's Alberta Electric System Operator (AESO) dataset, with P_{inst}=967 MW, from April to June of 2012 [42].

Case 2: Canada's Centennial wind farm (Saskatchewan) dataset, with P_{inst}=150 MW, from June to August of 2016.

Case 3: Spain's Sotavento wind farm dataset, with P_{inst}=17.5 MW

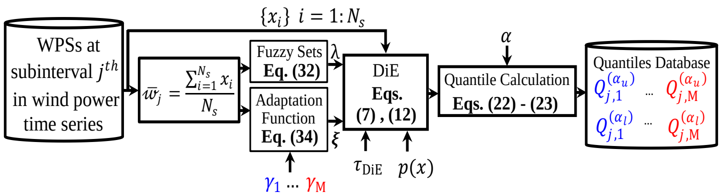


Fig. 8. General diagram of wind power quantiles database construction.

from October to December of 2015 [43].

Case 4: AESO dataset from January to December of 2015 [42].

The main reason behind the selection of these datasets is to thoroughly examine the applicability of the proposed FADiE framework with diverse wind power generation profiles. The empirical results in [3] show that the chaos in Case 2 is much higher than other cases for WPP; thus, it is a good case for testing the proposed FADiE framework. Case 4 is considered to assess the seasonality effect on the performance of the framework. The datasets are split into training (60%), validation (30%), and test (10%) datasets.

The 30-min and 1-hour very short-term prediction horizons, with respective 30- and 60-min subintervals sizes, are considered to construct optimal PIs. Different wind power prediction horizons, e.g., from minutes to days, with certain resolutions are required for diverse applications in power systems. For example, 30-min and 1-hour prediction horizons can be used for wind farm control, frequency control, and real-time economic dispatch [23]. However, short-term prediction horizons longer than one hour, can also be considered based on the corresponding applications in power systems such as look-ahead economic dispatch, reserve scheduling, unit commitment and day-ahead electricity market [24], [25]. But, in every probabilistic prediction approach, the longer the prediction horizon, the more the uncertainty in the prediction error. Without loss of generality, different prediction horizons from one minute to days can be implemented in the proposed framework if necessary. For very short-term prediction horizons, e.g., one to 60 minutes, only wind power datasets are needed, and this is referred to as a statistical approach. For longer prediction horizons, from two hours to days, other explanatory variables such as wind speed, wind direction, temperature, numerical weather prediction (NWP), etc. might be needed. The optimal selection of explanatory variables is only necessary for better training of the ELM in our proposed framework [2], [9], [13]. In this paper, the parameters of fuzzy sets and adaptation function are set as follows: w_1=0.2, w_1^-=0.15, w_1^+=0.25, w_2=0.8, w_2^-=0.75, w_2^+=0.85. For 30-min horizon, $\Delta\gamma = 0.1, \gamma_1$ =0.5, and γ_M =1.2, and for 1-hour horizon $\Delta\gamma = 0.1, \gamma_1$ =1, and γ_M =3. The simulations are performed on a Windows PC with an Intel Core i7-6700 CPU with 3.4 GHz and 16 GB RAM.

B. Analysis of Simulation Results

1) Sensitivity analysis of the proposed FADiE framework

Because the FADiE framework is developed based on BW as a fundamental parameter, the sensitivity of the constructed PIs needs to be assessed versus the BW growth factor γ. Fig. 10 shows that, for a certain CL and 1-hour prediction horizon, the desired reliability of a PI might be ideally gained by $\gamma_{\text{opt}} = 1$, which results in original BW h^*_{DiE}, e.g., AESO 2012 and Sotavento datasets. For the Centennial dataset, as a highly chaotic time series, the reliability of the PIs is not satisfactory with $\gamma = 1$; therefore, γ must increase to meet the reliability criterion $0 \leq ACE < \varepsilon$, i.e., $\gamma_{\text{opt}_{95\%}}$=1.6 and $\gamma_{\text{opt}_{99\%}}$=2.4.

For the datasets containing low chaos or for short prediction horizons, the PIs might have high reliability and low sharpness with $\gamma = 1$; thus, γ should decrease to raise the sharpness while satisfying the reliability criterion. Therefore, according to this analysis, the suitable range of γ can be easily determined to train ELM_1 to ELM_M to satisfy the reliability and sharpness criteria.

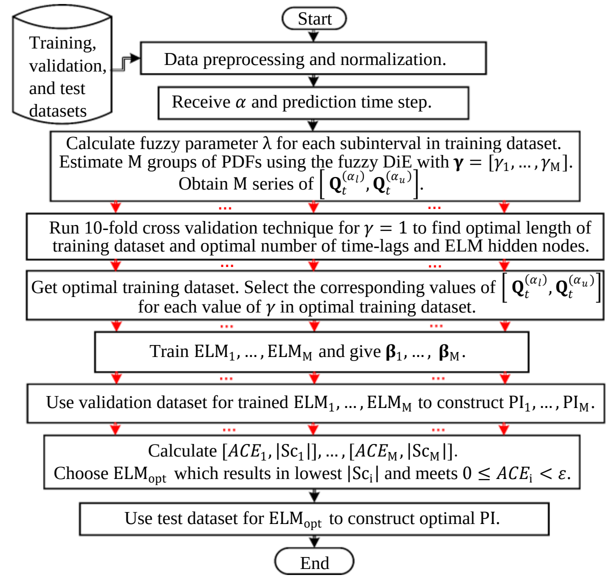


Fig. 9. Structure of the parallel computing-based FADiE framework.

2) Effect of fuzzy sets and adaptation function on wind power PIs

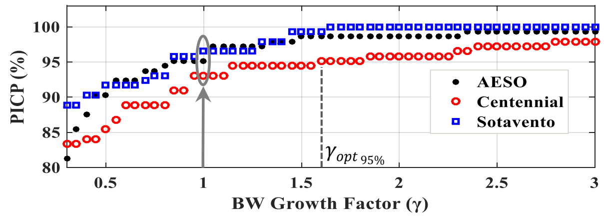
To establish the superiority of fuzzy DiE over other approaches and assess the effects of fuzzy sets on the reliability and sharpness of PIs, five PI construction approaches are considered for the AESO 2012 and Sotavento datasets in Table II. GE_DPI evidently cannot provide sharp PIs, and GE_API does not guarantee high reliability. The DiE with λ=0 produces very sharp PIs while with λ=1 generates reliable PIs. However, by setting λ=1 for all time periods, boundary effects happen for some, and the DiE cannot generate bona fide PDFs to present a real PI. While satisfying the reliability criterion, the sharpness can be further improved by applying the proposed fuzzy sets. The effect of the proposed adaptation function on PI construction results is shown for the chaotic Centennial time series in Table III where the reliability and sharpness are simultaneously satisfied. Even though $PINAW$ must increase for chaotic datasets to satisfy the reliability criterion, sharpness can still be preserved in a reasonable range by the FADiE framework.

3) Computational efficiency analysis

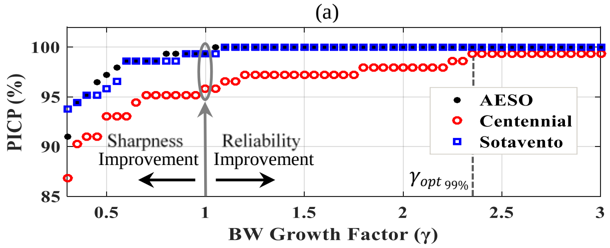
In the FADiE framework, the computation time is mainly devoted to training and validation stages for online applications. The computation time in the FADiE framework is very low compared to benchmarks due to the use of a predetermined database and a parallel processing. Moreover, the optimal BW selection procedure takes some time for a large dataset. To show the superiority of the proposed framework in online practical applications, BW selection and the total training and validation computation time for the 1-hour prediction horizons are summarized in Table IV. Based on the simulations, the FADiE is shown to be at least three times faster than LP-QR, 10 times faster than BELM, 300 times faster than PSO-QR, and 500 times faster than LUBE.

4) Comparison with benchmarks

To validate the satisfactory performance of the proposed FADiE framework, five well-known benchmarks (PSO-QR, LP-QR, LUBE, BELM, and ARIMA) are used to construct PIs using the same datasets and optimal training processes. They are also evaluated with the same criteria. However, none of the benchmark methods except ARIMA take advantage of the proposed parallel computing process because they have different training strategies. Generally, in this paper, $30 \leq \tilde{N} \leq 40$, $3 \leq N_\text{lag} \leq 8$, and $30 \leq N_\text{d} \leq 60$. Because power system operators always need reliable and sharp PIs with high confidence levels to ensure optimal generation and control of power systems, in this study PIs with CL=95% and 99% are constructed to evaluate the performance of the FADiE framework.



(a)



(b)

Fig. 10. The sensitivity of PI reliability to BW growth factor for 1-hour prediction horizon in Cases 1 to 3: (a) CL= 95%, (b) CL=99%.

TABLE II
COMPARISON OF KDE-BASED APPROACHES AND EFFECT OF FUZZY SETS.

Approach CL=95%	**AESO 2012** (1-hour)			**Sotavento** (1-hour)		
	$PICP$ (%)	$PINAW$ (%)	$\|S_c\|$ (%)	$PICP$ (%)	$PINAW$ (%)	$\|S_c\|$ (%)
GE_{DPI}	99.17	56.50	5.71	97.91	59.73	6.97
GE_{API}	91.67	11.89	1.88	94.17	19.12	3.04
DiE (λ=0)	75.83	08.70	2.66	84.58	12.15	2.96
DiE (λ=1)	96.66	15.19	1.78	97.08	19.94	2.40
Fuzzy DiE	**95.83**	**14.22**	**1.48**	**96.25**	**18.86**	**2.30**

TABLE III
EFFECT OF ADAPTATION FUNCTION ON THE FUZZY DIE.

Centennial (1-hour)	CL=95%			CL=99%		
	$PICP$ (%)	$PINAW$ (%)	$\|S_c\|$ (%)	$PICP$ (%)	$PINAW$ (%)	$\|S_c\|$ (%)
Fuzzy DiE	92.08	17.57	3.18	95.41	22.70	1.32
FADiE	**95.41**	**25.10**	**3.40**	**99.58**	**37.45**	**0.89**

The detailed simulation results for AESO 2012 (Case 1), Centennial (Case 2), and Sotavento (Case 3) datasets, including $PICP$, ACE, $PINAW$, and $|S_c|$, are respectively given in Tables V-VII. To assess the effect of seasonality on the performance of the FADiE framework and observe the variations in the optimal BW growth factor, Case 4 is used and compared with the ARIMA and LP-QR benchmarks, and the results are shown in Tables VIII and IX for 30-min and 1-hour prediction horizons, respectively. Note, since ARIMA uses parallel computing as well to predict lower/upper quantiles, its computation time is close to that of FADiE. The comparison of evaluation criteria demonstrates that the FADiE framework outperforms the benchmarks and provides a trade-off between high reliability and high sharpness of constructed PIs for both prediction horizons. As an indication of the sharpness of PIs with prior considerations of reliability, the simulation results are mainly discussed in terms of the S_c criterion. At CL=95%, the maximum value of $|S_c|$ for the worst case for 30-min and 1-hour prediction horizons across Cases 1 to 3 are 2.76 and 3.40%, respectively. Compared to the average of the benchmarks, sharpness is improved by 26.15 and 22.72%, respectively. For CL=99%, $|S_c|_{\max}$ takes smaller values of 0.78 and 0.89%, with 31.90 and 42.30% sharpness improvement for 30-min and 1-hour prediction horizons, respectively. A longer prediction horizon is found to create more uncertainty, which consequently results in lower sharpness. These results illustrate that, for the time series containing high chaos, much more meaningful PIs can still be obtained by the proposed FADiE framework than by existing approaches. From the reliability perspective, the constructed PIs in these cases satisfy the reliability criterion $0 \leq ACE < \varepsilon$. Among the benchmarks, BELM and LUBE approaches show approximately the same performance and outperform the PSO-QR

TABLE IV
CPU TIME FOR BW SELECTION AND TOTAL TRAINING AND VALIDATION

Off-line BW selection technique	**DPI**	**API**	**DiE**
1-hour time step (for 30 days)	28.00 (s)	24.50 (s)	28.80 (s)
Approach	**CPU time (s)**		
FADiE	2.93		
LP-QR	9.20		
BELM	37.82		
PSO-QR	890.20		
LUBE	1521.65		

TABLE V
THE RESULTS OF PI CONSTRUCTION FOR CASE 1

Horizon	PINC	Method	$PICP$ (%)	ACE (%)	$PINAW$ (%)	$\|S_c\|$ (%)
30-min	95%	**FADiE**	**95.41**	**+0.41**	**12.16**	**1.52**
		PSO-QR	93.33	-1.67	16.96	2.50
		LUBE	94.79	-0.21	15.60	2.21
		BELM	95.83	+0.83	15.93	2.13
	99%	**FADiE**	**99.37**	**+0.37**	**18.88**	**0.50**
		PSO-QR	96.87	-2.13	28.45	1.18
		LUBE	98.33	-0.67	26.02	1.05
		BELM	99.79	+0.79	27.64	1.02
1-hour	95%	**FADiE**	**95.83**	**+0.83**	**14.22**	**1.48**
		PSO-QR	94.17	-0.83	22.64	2.84
		LUBE	95.83	+0.83	21.75	2.71
		BELM	96.25	+1.25	22.50	2.58
	99%	**FADiE**	**99.17**	**+0.17**	**29.09**	**0.69**
		PSO-QR	96.67	-2.33	34.45	1.63
		LUBE	98.33	-0.67	32.25	1.20
		BELM	98.75	-0.25	33.44	1.03

TABLE VI
THE RESULTS OF PI CONSTRUCTION FOR CASE 2

Horizon	PINC	Method	$PICP$ (%)	ACE (%)	$PINAW$ (%)	$\|S_c\|$ (%)
30-min	95%	**FADiE**	**95.83**	**+0.83**	**21.00**	**2.76**
		PSO-QR	92.71	-2.29	30.08	4.05
		LUBE	93.75	-1.25	23.65	3.55
		BELM	94.79	-0.21	25.80	3.65
	99%	**FADiE**	**99.17**	**+0.17**	**29.25**	**0.78**
		PSO-QR	97.50	-1.50	34.80	1.09
		LUBE	97.70	+1.30	31.20	1.15
		BELM	98.12	-0.88	32.54	1.03
1-hour	95%	**FADiE**	**95.41**	**+0.41**	**25.10**	**3.40**
		PSO-QR	94.58	-0.42	37.53	4.83
		LUBE	94.17	-0.83	28.75	4.20
		BELM	94.58	-0.42	30.50	4.38
	99%	**FADiE**	**99.58**	**+0.58**	**37.45**	**0.89**
		PSO-QR	97.08	-1.92	45.12	1.96
		LUBE	97.91	-1.09	41.25	1.25
		BELM	99.17	+0.17	43.20	1.43

approach. Moreover, using point prediction approach and the assumption of Gaussian distribution for data noise and prediction model uncertainty might affect the quality of PIs in BELM approach. The definition of a certain cost function in a heuristic optimization problem, with the possibility of entrapping in local minima, is one reason low-quality PIs are generated in the PSO-QR and LUBE approaches. Even if these optimization problems can be efficiently solved to give a global solution, a better solution might exist because the defined cost functions might not reflect a suitable criterion to lead to the best solution. However, to improve the results and computational efficiency of the QR approach, the cost function can be linearly formulated with the linear model of ELM and efficiently solved with a linear programming approach. However, no linear formulation has yet been suggested for LUBE.

Statistical analysis of the results for Cases 1 to 3 shows that, for CL=95% and the 30-min prediction horizon, PSO-QR, LUBE, and BELM have average reliability values of 93.05, 94.10, and 95.07%, respectively, while this value is 95.55% for the FADiE framework. In addition, $PINAW_{avg}$ for the FADiE framework is 16.14%, but it equals to 21.88, 19.41, and 20.26% for the respective benchmark methods. The same analysis for CL=95% and the 1-hour prediction

TABLE VII

THE RESULTS OF PI CONSTRUCTION FOR CASE 3

Horizon	PINC	Method	$PICP$ (%)	ACE (%)	$PINAW$ (%)	$\lvert S_c \rvert$ (%)
30-min	95%	**FADiE**	**95.41**	**+0.41**	**15.27**	**1.83**
		PSO-QR	93.12	-1.88	18.60	2.94
		LUBE	93.75	-1.25	18.98	2.76
		BELM	94.58	-0.42	19.05	2.72
	99%	**FADiE**	**99.79**	**+0.79**	**24.63**	**0.49**
		PSO-QR	96.45	-2.55	29.90	1.45
		LUBE	97.70	-1.30	27.05	1.02
		BELM	98.12	-0.88	28.15	1.05
1-hour	95%	**FADiE**	**96.25**	**+1.25**	**18.86**	**2.30**
		PSO-QR	93.75	-1.25	28.65	4.29
		LUBE	94.17	-0.83	23.03	3.01
		BELM	95.41	+0.41	25.50	3.07
	99%	**FADiE**	**99.17**	**+0.17**	**23.66**	**0.60**
		PSO-QR	96.25	-2.75	35.47	1.77
		LUBE	97.91	-1.09	28.20	1.04
		BELM	98.75	-0.25	32.44	1.00

horizon indicates average values of 94.16, 94.72, and 95.41% for $PICP$ across Cases 1 to 3 for the respective benchmarks, while an average reliability of 95.82% is obtained using FADiE. $PINAW_{avg}$ values are 29.60, 24.51, and 26.16% for the three respective benchmarks, while for FADiE is 19.39%. Although the benchmarks can achieve the desired reliability on average, they cannot generate sharp PIs compared with the FADiE framework. Based on Tables V to VII, the same analysis can be done to show the superior performance of FADiE for CL=99%.

To evaluate the ARIMA approach, the lower/upper quantiles obtained by the DiE are predicted using ARIMA(1,1,2). The results presented in Tables VIII and IX, for 30-min and 1-hour prediction horizons respectively, illustrate that the optimal value of BW growth factor, γ_{opt}, changes from season to season according to the level of wind power volatility. For 30-min horizon in Case 4, $PINAW_{avg}$ has values of 16.48, 11.46, and 8.55%, while for 1-hour horizon, it takes greater values 28.24, 19.78, and 13.90% for ARIMA(1,1,2), LP-QR, and FADiE, respectively. In both tables for all seasons, FADiE constructs very sharp PIs with desired reliability, i.e., 95%. Although ARIMA(1,1,2) can achieve the desired reliability, the PIs are not sharp. LP-QR cannot meet the ACE criterion for summer case, while it generates sharper PIs compared with ARIMA(1,1,2).

The constructed PIs with CL=95% obtained by the proposed FADiE framework and the corresponding real wind power values over a period of ten days are illustrated in Figs. 11-13 for Cases 1 to 3, respectively. These figures illustrate how well the PIs constructed by the proposed framework can preserve the sharpness and enclose the measured wind power for these three different wind power datasets with different nominal capacities. The results demonstrate the flexibility and robustness of the framework to provide high-quality PIs. The promising results show that decision-making conditions with prediction horizons ranging from minutes to hours, such as wind farm control, electricity market, optimal reserve dispatching, and so on, can benefit from the proposed WPP uncertainty quantification.

VII. CONCLUSION

This paper proposed a fast and efficient general framework for probabilistic prediction of wind power generation based on the concept of optimal bandwidth selection for a diffusion-based kernel density estimator. Because the framework avoids historical deterministic prediction results, any assumptions about prediction error and data noise, and widely-used cost-function based optimization problems in the literature, it has the potential to outperform other approaches in terms of evaluation criteria, computational efficiency, and practicality. It can also be efficiently used for probabilistic prediction of solar generation and electricity load containing special patterns in the time series. The key point of the framework is that its performance can be optimally oriented via a fuzzy inference system and a tri-level adaptation function to capture the inherent uncertainty of non-stationary wind power time series in different seasons as well as the uncertainty of the prediction model. The high efficiency of the framework is verified using simulations with datasets from different wind farms and different seasons. Compared to previous approaches, the framework can provide both reliable and very sharp PIs for power system operators. Although the framework uses simultaneous processes for construction of the output datasets for prediction model training, which might make the implementation challenging, this does not decrease the computational efficiency because parallel processing is applied. Future work could further improve the performance of the proposed framework by incorporating techniques that provide *a priori* knowledge about the chaos level of the time series under study. The combination of these techniques with parallel computing processes provides an opportunity for better training of the

TABLE VIII

THE RESULTS OF PI CONSTRUCTION FOR CASE 4, 30-MIN AHEAD.

Prediction Horizon	30-min (PINC=95%) $\gamma_{\text{opt-Sp}}=0.7, \gamma_{\text{opt-Su}}=0.6, \gamma_{\text{opt-Au}}=0.6, \gamma_{\text{opt-Wi}}=0.7$					
Method	**FADiE**		LP-QR		ARIMA(1,1,2)	
Indices	$PICP$ (%)	$\lvert S_c \rvert$ (%)	$PICP$ (%)	$\lvert S_c \rvert$ (%)	$PICP$ (%)	$\lvert S_c \rvert$ (%)
Spring	**97.29**	**0.82**	95.21	1.01	95.63	1.56
Summer	**94.58**	**1.04**	92.50	1.49	94.17	2.00
Autumn	**97.50**	**0.89**	95.63	1.36	96.04	1.59
Winter	**96.04**	**1.25**	95.42	1.78	95.00	2.07
$PINAW_{avg}$	**8.55**		11.46		16.48	

TABLE IX

THE RESULTS OF PI CONSTRUCTION FOR CASE 4, 1-HOUR AHEAD.

Prediction Horizon	1-hour (PINC=95%) $\gamma_{\text{opt-Sp}}=1.0, \gamma_{\text{opt-Su}}=2.0, \gamma_{\text{opt-Au}}=1.0, \gamma_{\text{opt-Wi}}=1.6$					
Method	**FADiE**		LP-QR		ARIMA(1,1,2)	
Indices	$PICP$ (%)	$\lvert S_c \rvert$ (%)	$PICP$ (%)	$\lvert S_c \rvert$ (%)	$PICP$ (%)	$\lvert S_c \rvert$ (%)
Spring	**96.25**	**1.17**	95.42	1.88	95.42	3.06
Summer	**95.83**	**1.77**	93.75	2.30	95.42	3.13
Autumn	**95.00**	**1.46**	95.83	2.38	94.17	3.26
Winter	**96.25**	**2.03**	94.58	3.08	95.00	3.25
$PINAW_{avg}$	**13.90**		19.78		28.24	

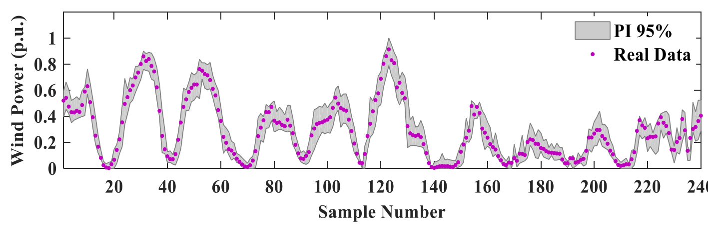


Fig. 11. The constructed PI for 1-hour prediction horizon for Case 1 (AESO).

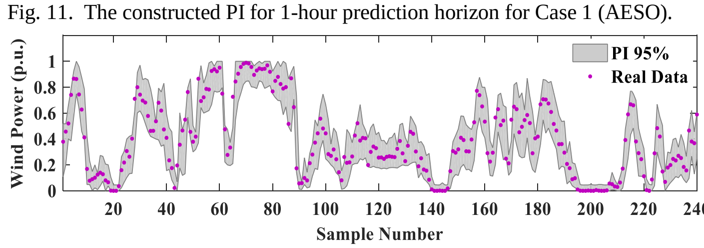


Fig. 12. The constructed PI for 1-hour prediction horizon for Case 2 (Centennial).

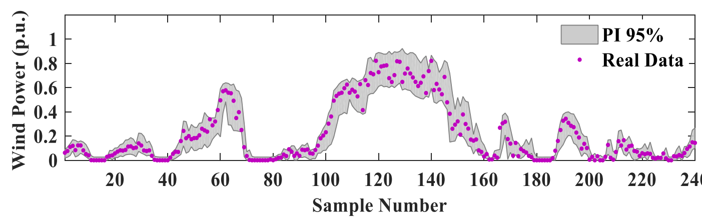


Fig. 13. The constructed PI for 1-hour prediction horizon for Case3 (Sotavento).

prediction model for longer prediction horizons.

REFERENCES

[1] N. Amjady, F. Keynia, H. Zareipour, "Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization", *IEEE Trans. Sustainable Energy*, vol. 2, no. 3, pp. 265-276, Jul. 2011.

[2] Y. Zhang, J. Wang, and X. Wang "Review on probabilistic forecasting of wind power generation," *Renew. Sust. Ener. Rev.*, vol. 32, pp. 255–270, 2014.

[3] 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*, vol. 33, no. 1, pp. 590-601, Jan. 2018.

[4] M. Khodayar, O. Kaynak, and M.E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770-2779, Dec. 2017.

[5] G. Li and H. D. Chiang, "Toward cost-oriented forecasting of wind power generation," *IEEE Trans. Smart Grid, Early Access*, 2017.

[6] David C. Hill, et al., "Application of auto-regressive models to U.K. wind speed data for power system impact studies," *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, Jan 2012.

[7] A. Kavousi-Fard, A. Khosravi, and S. Nahavadi, "A new fuzzy based combined prediction interval for wind power forecasting," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 18–26, Jan. 2016.

[8] 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.

[9] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation—with application to solar energy," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3850-3863 Sep. 2016.

[10] A. Haque, et al., "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.

[11] C. Wan, et al., "Probabilistic forecasting of photovoltaic generation: an efficient statistical approach," *IEEE Trans. Power Sys.*, vol. 32, no. 3, May 2017.

[12] F. Golestaneh, P. Pinson, R. Azizipanah-Abarghooee, H.B. Gooi, "Ellipsoidal Prediction Regions for Multivariate Uncertainty Characterization" *IEEE Trans. Power Syst, Early Access*, 2018.

[13] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1033–1044, May 2014.

[14] 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.

[15] Y. Zhang, J. Wang, and X. Luo, "Probabilistic wind power forecasting based on logarithmic transformation and boundary kernel," *Energy Convers. Manage.*, vol. 96, no. 1, pp. 440–451, May 2015.

[16] 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.

[17] Y. Zhang and J. Wang, "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting," *Int. J. Forecast.*, vol. 32, no. 3, pp. 1074–1080, Jul.–Sep. 2016.

[18] Z. Wang, W. Wang, C. Liu, Z. Wang, and Y. Hou, "Probabilistic forecast for multiple wind farms based on regular vine copulas," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 578–589, Jan. 2018.

[19] Chaoyue Zhao, and Yongpei Guan, "Data-driven stochastic unit commitment for integrating wind generation" *IEEE Trans. Power Syst.*, vol. 31, no. 4, July 2016.

[20] H. Khorramdel, J. Aghaei, B. Khorramdel, and P. Siano, "Optimal battery sizing in microgrids using probabilistic unit commitment," *IEEE Trans. Ind. Informat.*, vol. 6, no. 2, pp. 834–843, Apr. 2016.

[21] C. Uckun, A. Botterud, and J. R. Birge, "An improved stochastic unit commitment formulation to accommodate wind uncertainty," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2507–2517, Jul. 2016.

[22] Y. Wen, W. Li, G. Huang, and X. Liu, "Frequency dynamics constrained unit commitment with battery energy storage," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 5115-5125, Nov. 2016.

[23] Z. Li, W. Wu, B. Zhang, and B. Wang, "Adjustable robust real-time power dispatch with large-scale wind power integration," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 357–368, Apr. 2015.

[24] W. Wu, J. Chen, B. Zhang, and H. Sun, "A robust wind power optimization method for look-ahead power dispatch," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 507–515, Apr. 2014.

[25] M. Doostizadeh, et al., "Energy and reserve scheduling under wind power uncertainty: An adjustable interval approach," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2943–2952, Nov.2016.

[26] Z. S. Zhang, et al., "A versatile probability distribution model for wind power forecast errors and its application in economic dispatch," *IEEE Trans. Power Syst*. vol. 28, no. 3, pp. 3114–3125, Aug. 2013.

[27] Chenghui Tang, et al., "Look-ahead economic dispatch with adjustable confidence interval based on a truncated versatile distribution model for wind power" *IEEE Trans. Power Syst.*, vol. 33, no. 2, March 2018.

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

[29] B. W. Silverman, "Density estimation for statistics and data analysis," Chapman and Hall, London. 1986.

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

[31] L. Devroye, "Universal smoothing factor selection in density estimation: Theory and practice," *Test 6*, pp. 223–320, 1997.

[32] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *J. Amer. Statist. Assoc.*, vol. 91, pp. 401–407, 1996.

[33] G. R. Terrell and D. W. Scott, "Variable kernel density estimation. *Ann. Statist*, vol. 20, pp. 1236–1265, 1992.

[34] J. S. Marron and D. Ruppert, "Transformations to reduce boundary bias in kernel density-estimation," *J. Roy. Statist. Soc. Ser. B*, vol. 56, pp. 653–671, 1996.

[35] B. U. Park, S. O. Jeong, and M. C. Jones, "Adaptive variable location kernel density estimators with good performance at boundaries," *J. Nonparametr. Stat.*, vol. 15, pp. 61–75, 2003.

[36] D. O. Loftsgaarden and C. P. Quesenberry, "A nonparametric estimate of a multivariate density function," *Ann. Math. Statist.*, vol. 36, pp. 1049–1051, 1965.

[37] M. C. Jones, J. S. Marron, and S. J. Sheather, "Simple boundary correction for kernel density estimation," *Statist. Comput.*, vol. 3, pp. 135–146, 1993.

[38] Z. I. Botev, J. F. Grotowski, and D. P. Kroese "Kernel density estimation via diffusion," *The Ann. of Statist.*, Institute of Mathematical Statistics, vol. 38, no. 5, pp. 2916–2957, 2010.

[39] 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.

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

[41] G. James, D. Witten, T. Hastie, R. Tibshirani, "An introduction to statistical learning," Springer, New York Heidelberg Dordrecht London.

[42] https://www.aeso.ca/market/market-and-system-reporting/data-requests/.

[43] http://www.sotaventogalicia.com/en/real-time-data/historical.

Benyamin Khorramdel (S'14) received his B.Sc. and M.Sc. degrees in electrical engineering from Shahid Chaman University of Ahvaz, Ahvaz, Iran and Shiraz University, Shiraz, Iran, in 2009 and 2011, respectively. He is currently a Ph.D. student at the University of Saskatchewan, Saskatoon, SK, Canada. His current research interests include statistical learning, data mining, wind power forecasting, uncertainty modeling, and operation of renewable energy integrated power systems.

C. Y. Chung (M'01-SM'07-F'16) received the B.Eng. (with First Class Honors) and Ph.D. degrees in electrical engineering from The Hong Kong Polytechnic University, Hong Kong, China, in 1995 and 1999, respectively. He has worked for Powertech Labs, Inc., Surrey, BC, Canada; the University of Alberta, Edmonton, AB, Canada; and The Hong Kong Polytechnic University, China. He is currently a Professor, the NSERC/SaskPower (Senior) Industrial Research Chair in Smart Grid Technologies, and the SaskPower Chair in Power Systems Engineering in the Department of Electrical and Computer Engineering at the University of Saskatchewan, Saskatoon, SK, Canada. His research interests include smart grid technologies, renewable energy, power system stability/control, planning and operation, computational intelligence applications, power markets and electric vehicle charging. Dr. Chung is an Editor of *IEEE Transactions on Power Systems* and *IEEE Transactions on Sustainable Energy*. He is also an IEEE PES

Distinguished Lecturer and a Member-at-Large (Global Outreach) of the IEEE PES Governing Board.

N. Safari (S'16) received his B.Sc. (Hon.) and M.Sc. degrees in electrical engineering from the K. N. Toosi University of Technology and Tehran Polytechnic University, Tehran, Iran, in 2012 and 2014, respectively. Currently, he is pursuing the Ph.D. degree in electrical engineering at the University of Saskatchewan, Saskatoon, SK, Canada. His research interests include applications of signal processing, optimization, and machine learning algorithms in the integration of large-scale renewable energy generation to power systems. He is holding Dean's Scholarship, which is amongst the most prestigious scholarships at University of Saskatchewan.

G. C. D. Price received the B.Sc. in Electrical Engineering and B.Sc. in Physics (Honors) from the University of Alberta, Edmonton, Canada in 1996 and 2000, respectively and the M.Sc. in Mathematics from the University of Regina, Regina, Canada in 2009. He is currently a senior engineer with SaskPower, Regina, SK, Canada. His research interests include risk-based optimal decision-making in power systems operations and statistical decision theory.