WIND POWER PROBABILISTIC PREDICTION AND UNCERTAINTY MODELING FOR OPERATION OF LARGE-SCALE POWER SYSTEMS

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy
In the Department of Electrical and Computer Engineering
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

By

BENYAMIN KHORRAMDEL

Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis in whole or part should be addressed to:

Head of the Department of Electrical and Computer Engineering
University of Saskatchewan
57 Campus Drive
Saskatoon, Saskatchewan, S7N 5A9
Canada

OR

Dean of the College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9
Canada

Abstract

Over the last decade, large scale renewable energy generation has been integrated into power systems. Wind power generation is known as a widely-used and interesting kind of renewable energy generation around the world. However, the high uncertainty of wind power generation leads to some unavoidable error in wind power prediction process; consequently, it makes the optimal operation and control of power systems very challenging. Since wind power prediction error cannot be entirely removed, providing accurate models for wind power uncertainty can assist power system operators in mitigating its negative effects on decision making conditions. There are efficient ways to show the wind power uncertainty, (i) accurate wind power prediction error probability distribution modeling in the form of probability density functions and (ii) construction of reliable and sharp prediction intervals. Construction of accurate probability density functions and high-quality prediction intervals are difficult because wind power time series is non-stationary. In addition, incorporation of probability density functions and prediction intervals in power systems' decision-making problems are challenging. In this thesis, the goal is to propose comprehensive frameworks for wind power uncertainty modeling in the form of both probability density functions and prediction intervals and incorporation of each model in power systems' decision-making problems such as look-ahead economic dispatch.

To accurately quantify the uncertainty of wind power generation, different approaches are studied, and a comprehensive framework is then proposed to construct the probability density functions using a mixture of beta kernels. The framework outperforms benchmarks because it can validly capture the actual features of wind power probability density function such as main mass, boundaries, high skewness, and fat tails from the wind power sample moments. Also, using the proposed framework, a generic convex model is proposed for chance-constrained look-ahead economic dispatch problems. It allows power system operators to use piecewise linearization techniques to convert the problem to a mixed-integer linear programming problem. Numerical simulations using IEEE 118-bus test system show that compared with widely used sequential linear programming approaches, the proposed mixed-integer linear programming model leads to less system's total cost.

A framework based on the concept of bandwidth selection for a new and flexible kernel density estimator is proposed for construction of prediction intervals. 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 is utilized to achieve high-quality prediction intervals for non-stationary wind power time series. The proposed prediction interval construction

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 high performance of the proposed framework compared to well-known conventional benchmarks such as bootstrap extreme learning machine, lower upper bound estimation, quantile regression, auto-regressive integrated moving average, and linear programming-based quantile regression.

Finally, a new adjustable robust optimization approach is used to incorporate the constructed prediction intervals with the proposed fuzzy and adaptive diffusion estimator-based prediction interval construction framework. However, to accurately model the correlation and dependence structure of wind farms, especially in high dimensional cases, C-Vine copula models are used for prediction interval construction. The simulation results show that uncertainty modeling using C-Vine copula can lead the system operators to get more realistic sense about the level of overall uncertainty in the system, and consequently more conservative results for energy and reserve scheduling are obtained.

Acknowledgments

First and foremost, I deeply thank God Almighty for the blessings He has bestowed upon me and for giving me the strength and wisdom to achieve this dream.

This doctoral thesis is due to the support and encouragement of many people. It is a pleasure to express my sincere thanks to all those who helped me for the success of this study.

I would like to express my sincere gratitude to my supervisor Prof. Tony C. Y. Chung for the continuous support of my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D. study. I would also like to thank Mr. George C. D. Price from SaskPower company who provided us very useful comments and datasets. In addition, I am grateful to the all members of the Smart GEN Lab for providing their support and their invaluable assistance.

Also, I would like to thank my thesis advisory committee members for their insightful comments and encouragement which helped me a lot to promote the quality of this research.

Last but not least, I would like to thank my lovely parents, my brothers, and my sister who supported and encouraged me a lot throughout my entire life to achieve this goal. In addition, I would like to thank my dear wife, Mahsa, who supported me spiritually during my married life especially my Ph.D. study.

Table of Contents

List of Tables

List of Figures

List of Symbols and Abbreviations

Abbreviations

BKDR	Beta kernel density representation
CCED	Chance-constrained economic dispatch
CCF	Conventional cost function
CCSO	Chance-constrained stochastic optimization
CDF	Cumulative distribution function
CL	Confidence level
CPPs	Conventional power plants
DRCC	Distributionally robust chance constraints
GMM	Gaussian mixture model
GGMM	Generalized gaussian mixture model
LP	Linear programming
MILP	Mixed-integer linear programming
NCF	New cost function
PB	Power bin
PDF	Probability density function
PIs	Prediction intervals
RMSE	Root mean square error
RO	Robust optimization
SSA	Simplified sequentially adaptive
SLP	Sequential linear programming
SO	Stochastic optimization
TVD	Truncated versatile distribution
VMD	Versatile mixture distribution
WFs	Wind farms

Functions

$B(\cdot)$	Beta function
$C_{t,i}^{g}$	Generation cost of unit i at hour t

$C_{t,i}^{r}$	Reserve cost of unit i at hour t
C_t^{dw}	Direct cost of wind power generation at hour t
$\widehat{EC}_t^{UP}, \widehat{EC}_t^{LS},$ $\widehat{EC}_t^{DN}, \widehat{EC}_t^{WC}$	Estimated expected cost for upward (UP) reserve, load shedding (LS), downward (DN) reserve, and wind curtailment (WC) at hour t
$F(z)$	A typical non-linear function with variable z
$\hat{f}_t(\cdot), \hat{F}_t(\cdot)$	Estimated PDF and CDF at hour t
$\hat{F}_t^{-1}(\cdot)$	Inverse of the estimated CDF at hour t
f_B	Beta kernel function
F_e	Empirical CDF
$\kappa(\cdot)$	Kernel function
$F(\boldsymbol{x}), f(\boldsymbol{x})$	Joint CDF/PDF of the set of random variables $\boldsymbol{x}$
$F(x_i), f(x_i)$	Marginal CDF/PDF of random variable x_i
$F_{x_i\|\cdot}(x_i\|\cdot)$	Conditional distribution of random variable x_i
$C(\cdot), c(\cdot)$	CDF/PDF of copula
u_i	Uniform distribution of random variable x_i
$F^{-1}(u_i)$	CDF Inverse of random variable x_i
$\boldsymbol{g}(\boldsymbol{z})$	General objective function of RO problem
$\kappa(\cdot)$	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

Parameters and Coefficients

$A_q/A'_q/B_q$ $/B'_q$	Estimated UP reserve/LS/ DN reserve/WC cost coefficients of wind power generation

CL^{UP}/CL^{DN}	Confidence levels for enough UP/DN reserve
c_i^{UP}/c_i^{DN}	UP/DN reserve cost coefficients of unit i
$\underline{h}/\overline{h}$	Lower/upper values of beta kernel bandwidth
I	Number of beta kernel components
L_t^b	Base load of system at hour t
N	Number of sample moments
N_{PB}, NG, NW	Number of power bins, CPPs, and WFs
$\underline{P_i}/\overline{P}_i$	Lower /upper limit of generation for unit i
$\overline{\Delta P_i}^{UP}/\overline{\Delta P_i}^{DN}$	UP/DN maximum ramp limits
$\overline{r}_i^{UP}/\overline{r}_i^{DN}$	Maximum value of UP/DN reserve for unit i
μ_n	n^{th} sample moment of a random variable
Λ	Number of linearization segments
$\psi_\lambda, m_\lambda, n_\lambda$	Constant parameters of piecewise linearization
γ^{UP}/γ^{LS} $/\gamma^{DN}/\gamma^{WC}$	UP reserve/LS/ DN reserve/WC penalty factors for wind power uncertainty
$h^*_{(\cdot)}$	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
a_i, b_i, c_i	CPPs' cost coefficients
c_i^{UP}/c_i^{DN}	UP/DN reserve cost coefficients of unit i

$K_{l,i}/K_{l,j}$	Power transmission shift factors of transmission line l considering unit i and WF j
L_t	Base load of system at hour t
NG, NW	Number of CPPs and WFs
$p_{j,t}^{LB}, p_{j,t}^{UB}$	Lower and upper bound of prediction interval for WF j at hour t
$\underline{P_i}/\overline{P_i}$	Lower /upper limit of generation for unit i
$\overline{\Delta P_i}^{UP}/\overline{\Delta P_i}^{DN}$	UP/DN maximum ramp limits
$\overline{r}_i^{UP}/\overline{r}_i^{DN}$	Maximum value of UP/DN reserve for unit i
λ_j	Penalty factor considered for j^{th} WF for potential wind power curtailment
$\hat{\tau}_{ij}$	Empirical pairwise Kendall's value of x_i and x_j
$\underline{TL_l}, \overline{TL_l}$	Lower/upper limit of power flow on transmission line l
Γ_i	Solution robustness controller

Variables

$\widehat{D}$	Main-mass interval for PDF construction
h	Bandwidth of beta kernel for PDF construction
$P_{t,i}$	Scheduled generation of unit i at hour t
$r_{t,i}^{UP}/r_{t,i}^{DN}$	Scheduled UP/DN reserve of unit i at hour t
$R_{t,j}^{UP}/R_{t,j}^{DN}$	Scheduled UP/DN reserve for managing the uncertainty of wind farm j at hour t
υ_i, ζ_i	Parameters of i^{th} beta component
ω	Total wind power generation random variable
ω_t	Scheduled total wind power generation at hour t
$w_{t,j}$	Scheduled generation of wind farm j at hour t
$w_{r,j}$	Generation capacity of wind farm j
x_i/p_i	Location/ weight of i^{th} beta component
$\delta_\lambda/\Delta_\lambda$	Continuous/binary auxiliary variables of piecewise linearization technique
t	Variable indicating time
x, y	Variables indicating wind power values
h	Bandwidth of kernel functions
γ	Bandwidth growth factor

$\overline{w}$	Mean of samples inside each subinterval
$Q_t^{(\alpha_l)}, Q_t^{(\alpha_u)}$	Lower and upper quantiles for subinterval t
$\boldsymbol{z}_{\mathcal{M}\times 1}$	RO problem decision variables
$P_{i,t}$	Scheduled generation of unit i at hour t
$P_{j,t}^{w}$	Scheduled generation of WF j at hour t
$g_{j,t}^{w}$	Auxiliary variable of WF j at hour t
$\underline{\hat{p}}_{j,t}^{w}$ / $\hat{\overline{p}}_{j,t}^{w}$	Allowable lower/upper bounds of generation for WF j at hour t
$\Delta R_{i,t}^{UP}$ /$\Delta R_{i,t}^{DN}$	Deployed UP/DN reserve of unit i at hour t
$R_{i,t}^{UP}$ /$R_{i,t}^{DN}$	Allocated UP/DN reserve capacity for unit i at hour t
w_i, y_j, p_{ij}	Auxiliary variables of linear RO constraints

Matrices & Vectors

$\boldsymbol{C}_{N\times I}$	Beta kernel moments matrix.
H	Bandwidth vector for optimal beta kernel bandwidth selection.
$\boldsymbol{p}, \boldsymbol{v}, \boldsymbol{\zeta}$	PDF reconstruction vectors for a typical PB.
$\mathbf{P}, \boldsymbol{\Upsilon}, \mathbf{Z}$	PDF reconstruction matrices for all PBs.
$\boldsymbol{\mu}$	Moment vector of a random sample.
$\boldsymbol{\beta}$	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.
$\widetilde{\mathbf{A}}_{\mathcal{I}\times\mathcal{M}}, \mathbf{B}_{\mathcal{I}\times 1}$	Problem matrices for uncertain inequality constraints
$\mathbf{C}_{\mathcal{K}\times\mathcal{M}}, \mathbf{D}_{\mathcal{K}\times 1}$	Problem matrices for certain inequality constraints
$\mathbf{E}_{\mathcal{L}\times\mathcal{M}}, \mathbf{H}_{\mathcal{L}\times 1}$	Problem matrices for certain equality constraints

Chapter 1

Introduction

1.1 Motivation

Recent studies show that the population of the world has tripled in size over the last century [1]. The quick development of the required infrastructures for emerged societies has led to an ever-growing demand for energy. According to BP's *Statistical Review of World Energy*, world primary energy consumption reached 13,276 million tons of oil equivalent (Mtoe) in 2016. Over the period 2005-2016, world primary energy consumption increased by annual average rate of 5.8%. In 2016, the consumed energy in the world came from oil with 4,418 Mtoe (33.3% with average growth of 1.1%), coal with 3,732 Mtoe (28.1% with annual average growth of 1.6%), natural gas with 3,204 Mtoe (24.1% with annual average growth of 2.3%), nuclear electricity with 592 Mtoe (4.5% with annual average growth of 0.5%), hydroelectricity with 910 Mtoe (6.9% with annual average growth of 3%), wind and solar electricity with 292 Mtoe (2.2% with annual average growth of 25.3%), and geothermal, biomass, and other renewable sources with 127 Mtoe (0.9% with annual average growth of 7.4%) [2]. These changes in energy consumption led to increment of world dioxide carbon (CO2) emission by an annual average rate of 1.4% from 2005 to 2016.

Over the last decades, there have been a global concern about the global warming and climate change originated from greenhouse gas (GHG) emission in the atmosphere. A GHG is any gaseous compound like CO2 in the atmosphere that is capable of absorbing infrared radiation, thereby trapping heat in the atmosphere. By increment of the heat in the atmosphere, GHGs are in charge of the greenhouse effect, which eventually leads to global warming and serious environmental threats.

One possible solution to protect the environment is to reduce the number of conventional energy resources operating in power systems because they emit a huge amount of GHGs. Instead, the installed capacity of renewable energy resources, especially wind power as one of the most promising technologies, should be increased all over the world.

Due to the randomness and intermittency of wind power generation, it is known as non-dispatchable power generation by which a great uncertainty is imported into power systems. The uncertainty of wind power generation affects power system operation in various aspects, e.g. power system stability, operation, ancillary service, reliability, and power quality. In this context, wind power prediction (WPP) as an efficient technique is highly recommended to alleviate this challenge and integrate more wind power into power systems. Accurate WPP can also support wind power trading in electricity markets, thereby bringing significant economic benefits. Due to inescapable error in WPP process, probabilistic measurement of wind power prediction error (WPPE) and wind power uncertainty quantification have been recently proposed in power system applications. Therefore, probabilistic WPP (PWPP) and uncertainty modeling as important tools for power system stochastic optimization are assessed in this research.

1.2 Literature Review

Over the recent years, wind power has been becoming an efficient renewable energy source which supplies increasing demand around the world, but the inherent uncertainty of wind power generation makes the operation and control of power systems very challenging. In this context, WPP technology is known as one of the most useful ways to alleviate the negative impacts from wind power uncertainty [3]. Most of WPP approaches could only estimate a single point of future level of wind power generation, so-called deterministic prediction [4]. Deterministic WPP gives the most probable value of future wind power generation and do not provide any information about the prediction error [5]. However, there are many factors which may adversely affect deterministic WPP results. Some of these factors are numerical weather prediction (NWP) results, WPP model, the accuracy of weather and power measurement data, and the operation of wind turbines. Since WPPE is unavoidable, deterministic WPP without any information about the level of error or existing uncertainty may not meet the requirements of power systems for risk management in different decision-making problems.

It is of great importance to determine the time and the level of WPPE in the future [6]. This challenge has drawn the attention of many researchers to propose PWPP approaches. PWPP, as the extension of deterministic WPP, tries to estimate the range of WPPE or the range by which the future wind power observations would be enclosed by a certain probability. In [7], the benefits of PWPP are assessed with analyzing the optimal operation of wind and hydro power plants in power

systems. Reactive power dispatching in a wind farm is accomplished using interval prediction results in [8]. In [9], using uncertainty analysis of wind farms, a strategy is defined to verify the increasing penetration of wind power generation. Uncertainty analysis, as the cornerstone of power systems operation, is applied in many aspects such as dynamic spinning reserves, optimal operation of wind and hydro power plants, and pricing strategy in power market [10]. In [11], the influence of weather parameters and power curve on prediction uncertainty are analyzed. Authors in [4] classified on-line probabilistic prediction approaches into different types according to model inputs and outputs. Based on the type of uncertainty representation, different approaches are classified into three categories in [12], and then new developments and requirements are also presented for PWPP.

Researchers take the advantages of stochastic optimization to use the uncertainty information of wind power generation in various decision-making problems such as optimal dispatch of wind-hydro power plants [7], unit commitment [13-15], reserve scheduling [16-18], energy storage sizing [19], and wind power trading [20-23]. According to these studies, wind power penetration can significantly increase in power systems by accurate modeling of WPP uncertainty in power system analysis process.

The performance of deterministic WPP approaches can be evaluated using some criteria such as mean absolute error (MAE), mean square error (MSE), and root mean square error (RMSE) [24]. This evaluation is based on the difference between predictive and actual wind power values. However, to assess the performance of PWPP approaches, there are other criteria such as reliability, sharpness, and skill score because the existing information on predictive densities cannot be directly compared with the measured value. This is a well-known challenge which has drawn the attention of many researchers in this field.

1.2.1 Main Challenge: Wind Power Generation Uncertainty

It is widely believed that wind power uncertainty modeling and prediction should be dynamically carried out in short-term, medium-term, and long-term because wind speed time series, as the closest meteorological variable to wind power, shows high fluctuations in different frequency ranges, as shown in Table 1.1 [25]. Also, the lack of knowledge about atmospheric behavior and non-linear characteristics of wind turbines (i.e., wind speed to power conversion process) leads wind power generation to be a nonlinear, non-stable, and low predictable process. Therefore, since

the level of existing uncertainty differs for various prediction horizons from minutes to days, there should be accurate WPP approaches and uncertainty representation techniques.

Table 1.1 Frequency range of wind speed variations.

Frequency domain	Time domain	Caused by
Very low	Months	Climatic changes and human activities
Low	Days	General changes of weather pattern
Middle	Hours	Thermal exchange between ground and atmosphere
High	Minutes	Local meteorological effect
Very high	Seconds	Turbulence effects of wind speed

Generally, different sets of information are required for WPP process: (i) historical measurements of wind power, (ii) historical measurements of explanatory variables, and (iii) forecasts of explanatory variables (e.g., NWP). It suffices to use historical measurements of wind power for very short-term prediction horizon while for longer prediction horizons, explanatory variables and/or NWP may be used to improve the accuracy of WPP tools.

1.2.2 Uncertainty Representation Techniques for Wind Power

In wind power uncertainty modeling, three main uncertainty representations are usually considered, i.e. probabilistic prediction [26], risk index [27], and scenario prediction [28]. Generally, it is not important to determine which one is the best uncertainty representation. However, based on the different applications in power systems, the focus would be on the most suitable representation technique for that specific application.

1.2.2.1 Probabilistic Wind Power Prediction (PWPP)

PWPP, as the most widely-used uncertainty representation, can provide power system operators with additional quantitative information on wind power generation uncertainty. There are four kinds of PWPP representation: (i) predictive densities in the form of probability density functions (PDFs) or cumulative distribution functions (CDFs), (ii) quantiles or prediction intervals (PIs), (iii) discrete probabilities, and (iv) moments of probability distributions. To assess the overall quality

of PWPP approaches, different scoring rules and criteria have been proposed to ensure that the best PWPP performance can be obtained under the best value of the related score.

For a specific stochastic optimization technique, PDFs (or CDFs) and PIs of wind power generation are widely used for decision-making problems of power system operation. Similar to deterministic prediction, PWPP can be also classified into four categories in terms of prediction horizon, i.e. very short-term, short-term, medium-term, and long-term prediction. Table 1.2 shows the application of different PWPP horizons in power systems.

Table 1.2 Applications of forecasting horizons in power systems.

Prediction horizon	**Long-Term**	**Medium-Term**	**Short-Term**	**Very Short-Term**
Time Scale	Months-Years	Days-Months	Hours-Days	Seconds-Hour
Application	• Power System Planning • Wind Power Planning	• Unit Commitment • Maintenance Scheduling	• Economic Dispatch • Reserve Scheduling • Day-ahead electricity market	• Frequency Control (AGC control) • Wind Turbine Control • Real Time Dispatch

Predictive densities can be constructed using both parametric and non-parametric approaches. Parametric approaches are developed based on some well-known distributions such as Gaussian, beta, Weibull, hyperbolic, stable, and versatile with analytical expressions [29-34]. There are different non-parametric approaches that are proposed for PWPP. The adaptive resampling method proposed by Pinson et al. in [35] is one of the widely-used approaches. In non-parametric approaches, there is no assumption about the shape of the predictive density; instead, they take the advantage of some useful methods like quantile regression (QR) [36], kernel density estimation (KDE) [37], and artificial intelligence (AI) [38].

The main advantage of parametric approaches is that few parameters are required for the distributions which results in low computational costs. However, for some cases, parametric approaches may be inefficient because it is difficult to estimate multi-modal, fat-tail, and highly

skewed distributions using parametric approaches. Therefore, nonparametric approaches which can estimate any kind of distributions with especial features can be used. Table 1.3 summarizes different PWPP approaches with related remarks.

Table 1.3 Parametric and non-parametric approaches for PWPP.

Group	Remarks
Parametric (e.g., Gaussian, beta, Weibull, stable, etc.)	• Assumes predefined distribution shapes. • Low computational burden. • Competitive for very short-term prediction horizon.
Non-parametric (e.g., quantile regression, kernel density estimation, ensemble forecasting, artificial intelligence)	• No assumption about distribution shapes. • High computational burden. • Competitive for short and medium-term prediction horizon. • A data driven approach.

Three widely-used nonparametric approaches are briefly explained in the following.

- Quantile regression approach

QR is a non-parametric approach without any assumption of distribution shape. In QR approach, several quantiles over the range [0, 1] are estimated. Generally, there are three main QR models: (i) Local Quantile Regression (LQR) [26], [39], Spline Quantile Regression (SQR) [40], and Quantile Regression Forest (QRF) [41]. In LQR, using a linear regression, the dependence of estimated quantiles on explanatory variables is modeled. The estimated quantiles in SQR can be also modeled by a series of nonlinear smooth functions such as cubic B-Spline functions [42], [43]. In [44] a QRF model is proposed using classifications and regression trees. In QR, the independent training of each QR model for each quantile may lead to crossing quantiles. Therefore, to avoid this phenomenon, special constraints are implemented in QR model development.

- Kernel density estimation approach

Non-parametric density estimation 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 [45]. In this field, for estimating the PDF of random variables, KDE, as a well-known non-parametric density estimation approach, uses a data-driven-based methodology with an important parameter called bandwidth (BW) [46]. KDE generates a smooth curve using the contribution of all sample points to the overall PDF. A kernel which is a valid PDF and generates a smooth curve is placed at each data point. Then, the final PDF can be obtained by adding up the values of all kernels of the data points. In [37], [47-49] four types of kernel functions are proposed as shown in Table 1.4; however, there are other kernel functions which are more efficient than these kernels [50].

Table 1.4 Proper kernel functions for density estimation of different variables.

Random variable	**Example**	**Kernel function**
Variables bounded between [0,1]	Wind power	Beta
Variables bounded between [0,+∞)	Wind speed	Gamma
Unbounded variables	Temperature	Gaussian or bi-weight
Periodic variables	Wind direction	Von Mises

The smoothing parameter BW has a great influence on the accuracy of PDF estimation. A high value of BW leads to over smoothing and ignore the main features of the actual distribution. A small BW leads to a PDF with many peaks and valleys. In [51] an asymptotic mean integrated square error (AMISE) is employed as the suitable criterion for optimal BW selection. The multivariate plug-in selector was applied to solve the optimization problem of minimizing AMISE. The modified KDE approach is presented in [52] to predict a discrete-continuous PDF of wind power generation. In order to enhance the numerical stability of KDE approach, a novel KDE based on copula method is recommended in [47], [48]. Also, for on-line applications, to simultaneously update wind power PDFs, an exponential forgetting process is implemented in KDE approach in [48], [49], [53].

- Artificial intelligence approach

With the successful implementation of AI techniques such as artificial neural networks (ANNs), support vector machine (SVM) and extreme learning machine (ELM) in point prediction approaches, they can be also used for PWPP. ANNs are black boxes which can model highly nonlinear systems. The outputs of ANNs can be directly set as PIs or two quantiles so-called lower

and upper quantiles. By minimizing a predefined cost function ANN models are trained with historical data. A new cost function, so-called coverage width-based criterion (CWC) is proposed in [54-56] for the evaluation of PIs. There is an adjustable penalty coefficient in CWC which leads to generate high quality PIs. For example, CWC set greater penalty coefficients for the situations that PIs are not probabilistically calibrated so that narrower PIs have high values of CWC.

1.2.2.2 Risk Index

Generally, unlike deterministic WPP, the information received by PWPP cannot be directly used for power system operation. Risk index, as a user-oriented representation technique, is easy to understand by end-users of PWPP because it shows a single value as the expected value of WPPE [57-59]. Higher value of risk index means that the prediction error is high, so the predicted point is less reliable. The risk index can be conveyed using well-known colors such as red, green, yellow or integer values from 1 to 5. Power system operators may increase the reserve capacity when receiving high value of risk index or warm colors like red to reduce the risks of WPPE.

1.2.2.3 Scenario Prediction

In dynamic decision-making problems in power systems, it is of great importance to take the spatial-temporal correlation of uncertainty sources into account for scenario prediction because there are many multi-stage and time-dependent decision-making conditions in power systems such as dynamic economic dispatch, stochastic unit commitment, the coordination of wind farms and energy storage systems, and wind power trading in multi-market with different closing times [60-62].

When wind power uncertainty representation is independently done for uncertainty sources for every look-ahead time instant, it means that the uncertainty representation is a time-independent process without providing extra information about the variations of uncertainty level over the future time. Indeed, the temporal correlation of uncertainty sources over look-ahead time instants can be modeled by joint PDFs. Although the direct use of joint PDFs for decision-making problems is difficult, scenario prediction is a proper approach for such time dependent decision-making problems. Using a series of predicted points over future time, a series of scenarios are generated. The spatial correlation among uncertainty sources in different geographical locations can also be considered since the complete information about the spatial-temporal correlation among

uncertainty sources is very beneficial for congestion management and probabilistic load flow. However, the question is how to model the interdependence structure of wind farms as uncertainty sources for scenario prediction process. In this context, Gaussian copula method is the main approach to model such interdependence structures [63]. On one hand, to accurately represent wind power uncertainty, an abundant number of scenarios should be produced via Monte Carlo simulation. On the other hand, too many scenarios increase the computational burden when solving the power systems' dynamic stochastic optimization models. Thus, scenario reduction techniques are proposed by researchers to provide a tradeoff between the computation time and the accuracy of the optimization models [64], [65]. In this research field, the techniques that are developed based on clustering [64] and Kantorovich distance [65] are two commonly used scenario reduction techniques.

1.2.3 Application of Wind Power Uncertainty Models in Stochastic Optimization of Power Systems

In real-world power system applications, many researchers have proposed various techniques to manage the uncertainty of wind generation through optimization. In [66]-[68], a stochastic optimization (SO) is proposed for look-ahead ED to handle the uncertainty of intermittent generation. In the proposed SO, several scenarios are extracted from a predetermined probability distribution to consider the possible values of uncertain variables. However, the optimal solution and computational efficiency strongly depend on the number of scenarios and the corresponding probabilities. Alternatively, chance-constrained SO (CCSO) is used in [29], [30], [69]-[71] as a more efficient method by setting a predefined risk level. However, accurate distribution models for wind power sources are required in this method to preserve the security of the system with a certain probability level. In [72]-[77] distributionally robust chance-constrained (DRCC) optimization is used, where a moment-based ambiguity set is defined to cover a family of distributions for uncertainty sources. This approach can be used in different areas such as AC optimal power flow [73], [75]-[77], distribution system planning [72], and energy and reserve dispatch [74]. The keen readers are referred to [77] for mathematical details of DRCC.

In the early 1970s, RO was proposed by Soyster to develop a model that is immune to the data of a convex uncertainty set [78], and then it was expanded by Ben-Tal and Nemirovski [79], [80]. In RO technique, the operator can improve the security of the system by letting wind power change

over a predetermined prediction interval [81]-[83]. But RO will lead to a highly conservative solution if it is not adjusted well because it minimizes the cost of the worst-case scenario. A two-stage RO is formulated in [84] for an economic dispatch problem in which the load is specified with an uncertainty set. In [85], an adjustable RO (ARO) approach is proposed for large-scale wind power integration considering a prespecified uncertainty set. The co-optimization of the security and operational cost is fulfilled in [86] using a robust ED accounting for the automatic generation control affine recourse process. An adjustable wind power uncertainty set by defining a dynamic confidence level is used to mitigate the over-conservativeness of the RO in [87] in a look-ahead power dispatch. Also, interval optimization (IO) uses prediction intervals (PIs) and central prediction point to minimize the system operating cost. Compared to CCSO and RO, IO is respectively more conservative and less precise [87]-[90]. A robust admissibility assessment is presented in [88] to measure the amount of wind generation that can be integrated into a large-scale system considering the operational risk. Similarly, a robust risk-constrained day-ahead UC is presented in [89] using an adjustable uncertainty set to avoid the conservative solutions. Reference [90] modeled wind generation by an interval and proposed a two-stage RO for sizing energy storage systems to mitigate the wind curtailment in a second-order cone programming. In recent years, Copula theory has drawn the attention of researchers to model the uncertainty of wind generation in power system operation [91]-[93]. An efficient algorithm based on Copula theory and dependent discrete convolution is proposed in [91] to tackle the “curse of dimensionality” of high-dimensional wind generation uncertainty modeling.

1.3 Research Objectives

The primary goal of this thesis is to accurately model the uncertainty of wind power generation by proposing two separate frameworks for PI and PDF construction. To this end, this study focuses on efficient uncertainty modeling techniques such as non-parametric models, mixture models, and vine copula models. The second goal of this research is to propose efficient optimization frameworks such as stochastic optimization and robust optimization for implementing the modeled uncertainties for look-ahead operation of large-scale power systems. The proposed optimization frameworks aim at reducing the total cost of the system while managing the uncertainty of integrated wind generation.

According to the literature review, accurate wind power PDF modeling and probabilistic

measurement of wind power uncertainty in the form of reliable and sharp PIs are of utmost importance in power systems with large scale wind power generation containing high level of uncertainty because they can assist power system operators in mitigating its negative effects on decision making conditions. However, the construction of such high-quality PIs and accurate WPPE models are difficult because wind power time series are non-stationary and highly chaotic. In this research, efficient frameworks are proposed to achieve this objective. For PI construction, the proposed framework is based on the concept of bandwidth (BW) selection for a new and flexible kernel density estimator. 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. For WPPE modeling objective, the proposed framework uses wind power sample moments. To validly capture the actual features of wind power PDF such as main mass, boundaries, high skewness, and tails from the moments, an efficient moment problem is presented and solved using beta kernel density representation (BKDR) technique. Accurately estimating these features assists power system operators to optimally schedule the system using a chance-constrained economic dispatch (CCED) problem. Importantly, a generic convex cost function is proposed for the CCED problem that eliminates the need for an analytical cumulative distribution function and enables us to use a highly accurate linearization technique. Using this technique, the non-linear CCED problem can be converted to a mixed-integer linear programming (MILP) problem which can be solved to global optimality via off-the-shelf mathematical programming solvers. Numerical simulations using IEEE 118-bus test system show that compared with widely used sequential linear programming (SLP) approaches, the proposed MILP model leads to less system's total cost.

For the proposed RO framework, an efficient wind generation dependence structure (DS) modeling is used to consider the existence of numerous geographically diverse and correlated wind farms in a power system. This part of study presents a new ARO framework to account for the complex DS

of wind farms in energy and reserve dispatch problems. Efficient DS modeling and its incorporation in generation and reserve scheduling can prevent wind curtailment and load shedding events. The proposed wind generation DS modeling uses canonical vine (C-Vine) copula as a flexible and accurate statistical model to obtain both joint and conditional distributions of any number of wind farms. Then, the results of the proposed C-Vine copula-based DS modeling is used as effective inputs for the proposed ARO-based economic energy and reserve dispatch problem. A case study of 20 wind farms is used to show the effectiveness of the proposed framework through comprehensive simulations on IEEE 118-bus test system.

1.4 Contributions and Organization of the Thesis

In this study, new and practical PI and PDF construction frameworks are proposed by which the uncertainty of integrated wind power generation can be quantified well. Two highly efficient optimization frameworks (i.e., chance-constrained optimization and adjustable robust optimization) are then proposed in order to implement the quantified uncertainties. The desirable performance of the proposed frameworks proved that the research done in this thesis, contributed to lowering the total operating cost of the power systems and giving more adjustability to power system operators.

Briefly speaking, the contribution of this study depends on the following technical descriptions.

1- For the first time, an efficient wind power distribution modeling is proposed using finite sample moments (e.g., 10 to 20) and the beta kernel density representation (BKDR) technique.
2- Because the BKDR technique represents the target distribution using several beta components with restricted parameters' range, it avoids boundary effects while capturing certain features of wind power distribution without reflecting overfitting problem.
3- An efficient methodology for construction of a new cost function (NCF) is proposed for nonlinear CCED models by which power system operators can use parametric and non-parametric distribution models for wind power.
4- Because of the proposed NCF, a highly accurate linearization technique, i.e., piecewise linearization, can be used to convert the non-linear CCED into an efficient MILP-based model.
5- 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.
6- A piecewise wind power PDF estimation procedure is introduced using piecewise and successive wind power sample sets.

7- 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.
8- A tri-level adaptation function is proposed to model the uncertainty of the prediction model and variability (seasonality) of wind power time series.
9- A parallel computing process is proposed to increase the computational efficiency and remove the widely-used cost function-based optimal PI construction methodologies.
10- A new ARO is proposed in which the constraints of energy and reserve dispatch problem are linear; so, the cost function is directly minimized just in one stage and one iteration.
11- The proposed ARO problem can control the conservativeness of the solution using two degrees of freedom.
12- An efficient DS modeling is proposed for large-scale wind generation uncertainty modeling using C-Vine copula to account for the spatiotemporal correlation of all wind farms.

The remainder of the report is organized as follows:

A comprehensive overview about the parametric and non-parametric distribution models for wind power probability distributions is done in Chapter 2.

In Chapter 3, the goal is to propose a new wind power PDF modeling and implement the constructed PDFs in a real application. The proposed model is based on a mixture of beta distributions. Also, for implementing in a real application, a generic convex model for CCED problems is proposed to alleviate the existing challenges of stochastic economic dispatch problems. The advantages of the models over other approaches are shown using simulation results.

In Chapter 4, the goal is to assess the performance of signal processing-based PI construction approaches. Therefore, a hybrid wind power PI construction framework is proposed based on the application of signal decomposition techniques, extreme learning machine, and kernel density estimation techniques.

In Chapter 5, the goal is to propose an efficient PI construction framework to construct high-quality PIs. Thus, a new fuzzy-adaptive PI construction framework is suggested to remove the challenges of the conventional PI construction approaches. The high superiority of the proposed framework is shown with comprehensive simulation results.

In Chapter 6, the goal is to handle the curse of dimensionality for wind power uncertainty modeling. Thus, an efficient wind generation dependence structure modeling is proposed for a new ARO technique to account for the spatiotemporal correlation of wind farms in high dimensional cases.

Finally, the conclusion and the future works of this study are presented in chapter 7.

Chapter 2

A Nonparametric Probability Distribution Model for Short-Term Wind Power Prediction Error

2.1 Introduction

Wind energy is going to be globally captured more and more to generate electricity and reduce the general concerns of greenhouse effect and the depletion of fossil fuels sources. As a result, high level of uncertainty is imported into today's power systems due to the inherent uncertainty of wind power generation [94]. Using current technologies, wind power prediction error (WPPE) which originates from uncertainty of prediction is inevitable and can reach to as high as 30% under some events such as shut down events [25]. At each time instant within measured wind power time series, there is a predicted value and the corresponding error. So, WPPE is a random variable with a certain probability distribution. Therefore, it is an essential part of modern power systems to provide power system operators with accurate model of WPPE probability distribution to accordingly act in decision making conditions such as economic dispatch (ED), unit commitment (UC), and energy storage sizing [29], [95], [96]. The real challenge is how to well estimate the probability density function (PDF) of WPPE so that the existing features, like high skewness and kurtosis and heavy tails, inside the distribution can be well represented. In previous works, WPPE is usually considered as a parametric distribution like Normal, Beta, Gamma, Lévy, Cauchy, Stable, and Hyperbolic distributions [30], [31]. However, some studies have found that these are not valid distributions and may not properly show the abovementioned features [97], [98]. Recent studies have shown that the uncertainty of WPP changes with the level of generated wind power; so, there is a conditional relation between WPPE and the prediction value. A conditional WPPE model is proposed in [99] based on the wind speed prediction error and wind turbine power curve. In [18] various gamma-like functions are used to approximate WPPE distribution for different levels of generated wind power.

Generally, the detailed modeling of WPPE enables power system operators to allocate reserves

dynamically for different prediction levels, rather than to use reserve statically. The importance of presenting an efficient WPPE distribution model is shown in probabilistic reserve sizing and power system reliability studies by transmission system operators [100]. Due to underestimation of heavy tails by some models, the frequency of maximum value of WPPE may not be estimated, and accordingly a wise decision may not be taken in operational reserve sizing. Beside the technical aspect, WPPE has a significant effect on economical aspect when wind power is traded in electricity markets. For example, in [101] WPPE cost may take up to 10% of total income from wind power trading. Moreover, if WPPE is more accurately considered in decision making conditions, up to half of the extra cost due to WPPE can be eliminated [20]. Therefore, due to high importance of WPPE in different aspects such as ED, UC, reserve allocation, electricity markets, and energy storage sizing, in this chapter WPPE is thoroughly evaluated and the statistical features are extracted for different levels of wind power generation.

Unlike previous works that proposed parametric distributions for WPPE, in this chapter, a nonparametric approach is adopted based on kernel density estimation (KDE) with an efficient bandwidth (BW) selection technique. The proposed BW selection technique can easily deal with distributions containing different features. The merit of the proposed approach is that it just needs the wind power time series and historical predicted values for a year. The predicted values can be obtained by any deterministic WPP approach. Then, the conditional distribution of WPPE without improving the prediction skill is estimated.

2.2 Parametric Probability Distributions for WPPE

There are a number of various parametric distributions that have been used in WPPE probability distribution estimation – Normal, Weibull, Beta, Gamma, Hyperbolic, Cauchy, and Lévy α-stable to name a few. In this section, an overview of basic properties of some well-known distributions are presented. It is not the aim of this section to provide a comprehensive overview of above distributions.

2.2.1 Normal Distributions

Normal distribution $N(\mu, \sigma^2)$ is one of the most well-known and traditionally used ones for probability distribution fitting of wind power samples. It has two different parameters called mean (μ) and the standard deviation (σ) calculated from the data samples. The PDF and CDF of Normal

distribution is shown by (2.1) and (2.2), respectively. The effects of two parameters σ and μ on the PDF of Normal distribution are shown in Figure 2.1.

$$\varphi(x|\mu,\sigma) = \frac{\exp(-(x-\mu)^2/2\sigma^2)}{\sqrt{2\pi\sigma^2}} \tag{2.1}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt \tag{2.2}$$

The main drawback of Normal distribution is the lack of flexibility since it cannot be accurately fitted to wind power and its prediction error samples. Therefore, recent works have tried to present more flexible distributions like stable distributions.

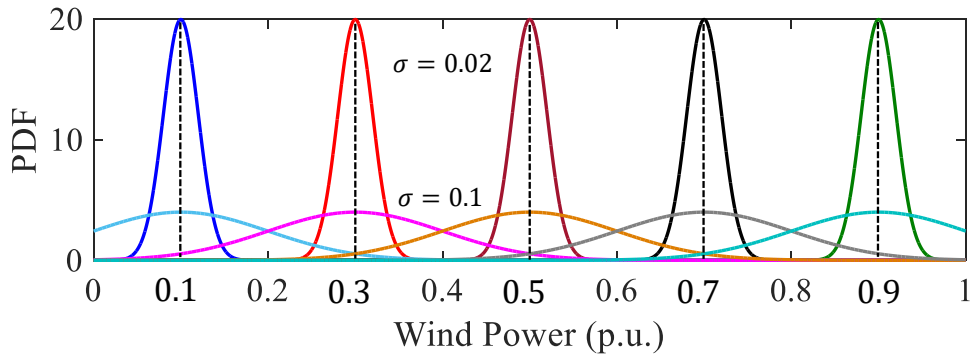


Figure 2.1 Illustration of Normal distribution.

2.2.2 Stable Distribution

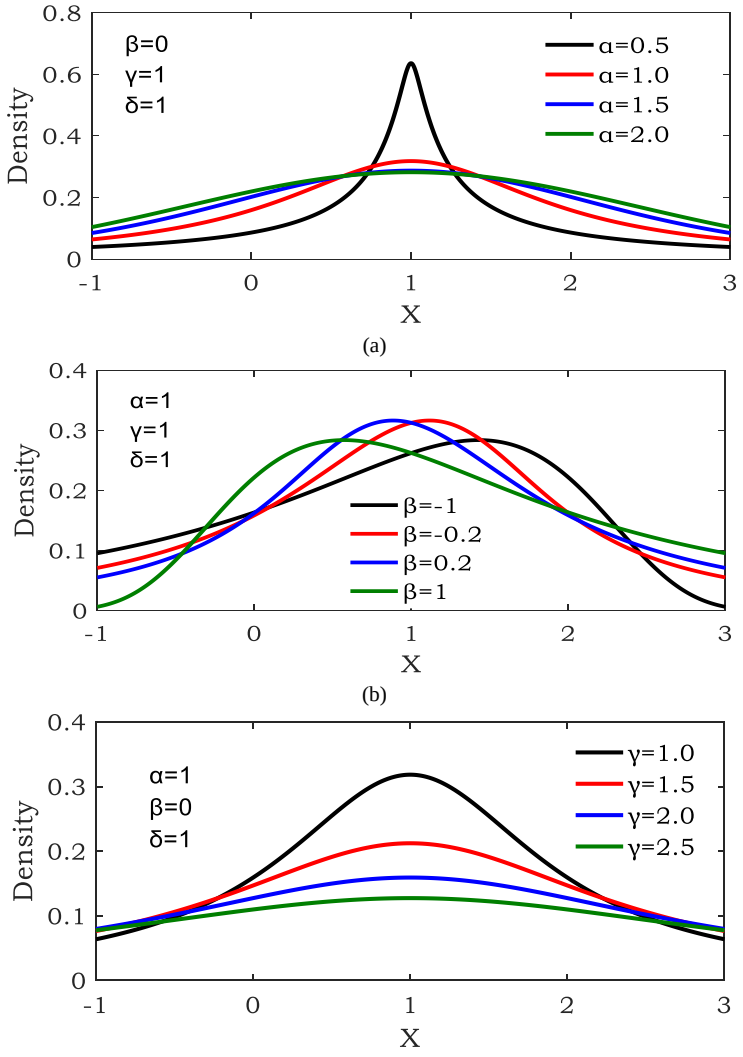
Stable distributions have a common important property. They are unchanged under linear addition of several independent and identically distributed stable distributed random variables. They are also sometimes referred to as the Lévy α-stable or Pareto-Lévy stable distributions [100]. Stable distributions are suitable for modeling heavy tails and high skewness features of WPPE. Although there are no analytical expressions for PDF of this kind of distribution, characteristic function $\Theta(x|\alpha,\beta,\gamma,\delta)$ shown in (2.3)-(2.4) can be used for random stable variables to calculate the PDF of stable distributions using (2.5). In (2.3)-(2.4), $\alpha \in (0,2]$ is the stability parameter, $\beta \in [-1,1]$ is skewness parameter, $\gamma \in (0,\infty)$ is scale parameter, and $\delta \in \mathbb{R}$ is location parameter.

$$\Theta(x|\alpha,\beta,\gamma,\delta) = \exp\left(ix\delta - |\gamma x|^{\alpha}(1 - i\beta sign(x)\Phi)\right) \tag{2.3}$$

$$\Phi = \begin{cases} \tan(\pi\alpha/2) & \alpha \neq 1 \\ -2\log|x|/\pi & \alpha = 1 \end{cases} \tag{2.4}$$

$$f(\tau) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Theta(x)\, e^{-ix\tau} dx \tag{2.5}$$

It is worth noting that the stable distribution $S(\alpha,\beta,\gamma,\delta)$ has three special forms: Normal, Cauchy, and Lévy distributions with closed-form PDFs. Normal is shown with $S\left(2,0,\sigma/\sqrt{2},\mu\right)$, Cauchy is stated with $S(1,0,\gamma,\delta)$, and $S(0.5,1,\gamma,\gamma+\delta)$ denotes Lévy distribution. Figure 2.2 shows the PDF of stable distribution when one specific parameter is altered, and others are kept fixed. It visually shows the effect of different parameters on stable distribution.



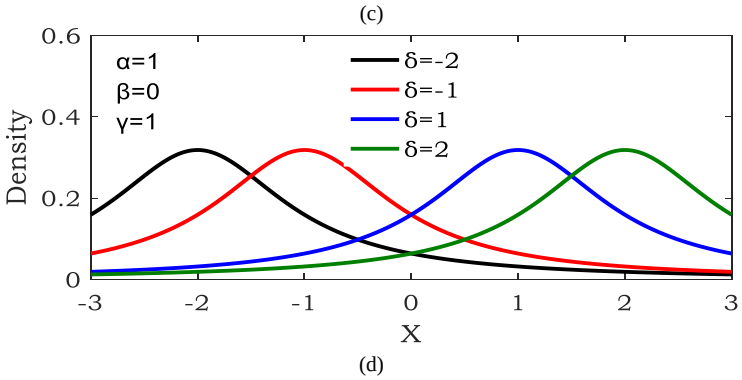


Figure 2.2 Effect of four parameters α, β, γ, and δ on stable distribution.

2.2.3 Beta Distribution

The PDF and CDF of beta distribution is shown by (2.6) and (2.7), respectively. Beta distribution as one of the most useful distribution models for wind power generation modeling has three advantages over other distribution models. First, like wind power, it is bounded between 0 (as lower boundary) and 1 (as upper boundary). Second, the shape of beta distribution symmetrically changes so that can well imitate the changes of wind power distribution from lower boundary region to upper boundary. Third, using a closed form, the set of beta parameters (υ_i, ζ_i) are related to the set of location and variance parameters (x_i, h).

To calculate the main parameters υ_i and ζ_i, a natural choice is to use mode and variance expressions of f_B in (2.8)-(2.9). Thus, for a given set of values (x_i, h), the parameters (υ_i, ζ_i) are obtained by solving the algebraic system (2.8)-(2.9). Figure 2.3 shows beta distributions with five sets of parameters (υ_i, ζ_i) over the range [0,1].

$$f_B(x|\upsilon,\zeta) = \frac{1}{B(\upsilon,\zeta)}\, x^{(\upsilon-1)}(1-x)^{(\zeta-1)} \tag{2.6}$$

$$F_B(x) = \frac{\int_0^x t^{(\upsilon-1)}(1-t)^{(\zeta-1)}\, dt}{B(\upsilon,\zeta)} \tag{2.7}$$

$$x_i = mode\big(f_B(x|\upsilon_i,\zeta_i)\big) = \frac{\upsilon_i - 1}{\upsilon_i + \zeta_i - 2} \tag{2.8}$$

$$h^2 = variance\big(f_B(x|\upsilon_i,\zeta_i)\big) = \frac{\upsilon_i\, \zeta_i}{(\upsilon_i + \zeta_i)^2(\upsilon_i + \zeta_i + 1)} \tag{2.9}$$

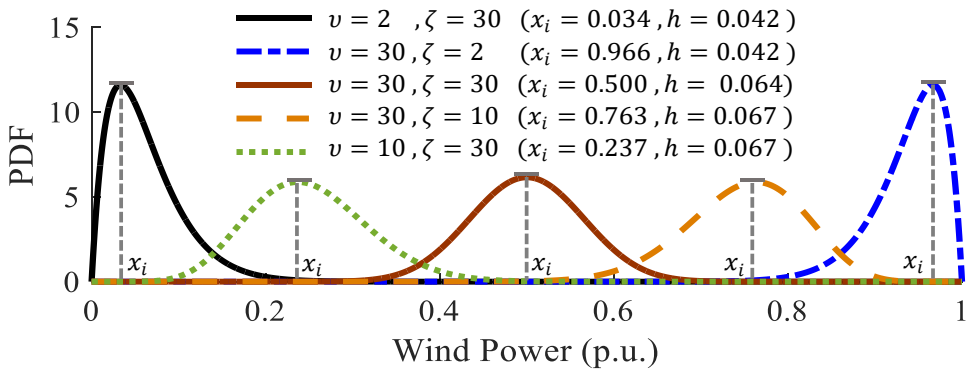


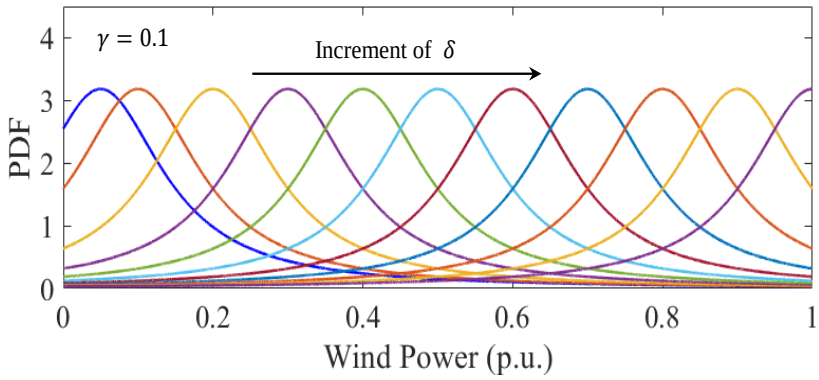
Figure 2.3 Illustration of beta distribution over the range [0,1].

2.2.4 Cauchy Distribution

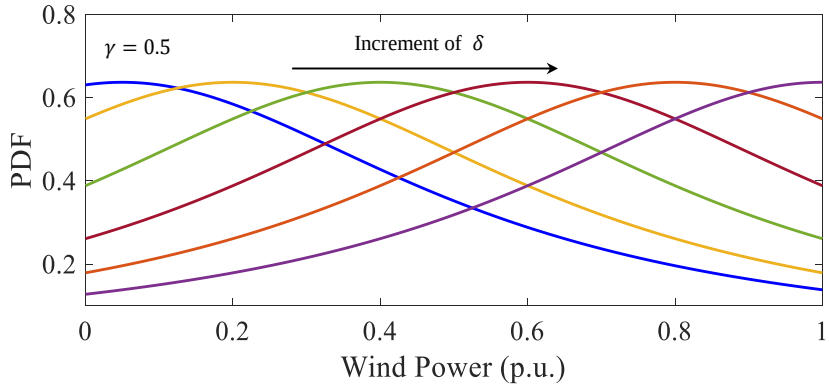
Cauchy distribution is a special case of stable distribution $S(\alpha, \beta, \gamma, \delta)$ as shown in Figure 2.4 with $S(1,0,\gamma,\delta)$. The PDF and CDF of Cauchy distribution are expressed with (2.10)-(2.11).

$$f(x|\delta,\gamma) = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x-\delta)^2+\gamma^2}\right] \tag{2.10}$$

$$F(x) = \frac{1}{\pi}\arctan\left(\frac{x-\delta}{\gamma}\right) + \frac{1}{2} \tag{2.11}$$



(a)



(b)

Figure 2.4 Illustration of Cauchy distribution.

2.2.5 Truncated Versatile Distribution (TVD)

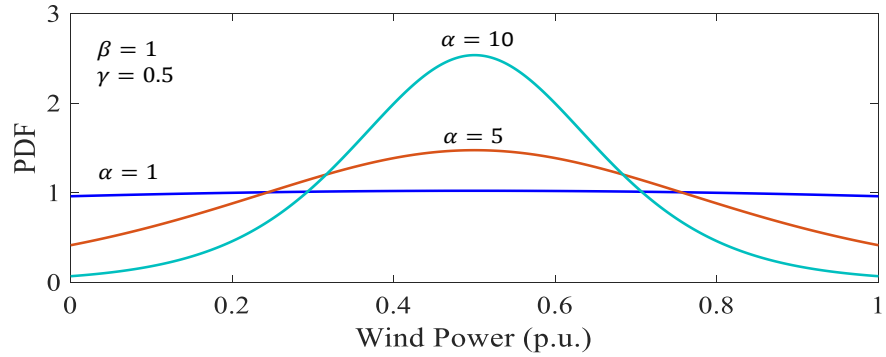
For random variable x which follows TVD distribution with shape parameters α, β, and γ, it is denoted with $X \sim V(\alpha, \beta, \gamma)$ and the PDF, CDF and CDF inverse are expressed with (2.12)-(2.14) where $\alpha > 0, \beta > 0, -\infty < \gamma < \infty$, and the normalization constant k is obtained using (2.15). The PDF of TVD distribution is shown in Figure 2.5 for different values of parameters α, β, and γ.

$$f(x|\alpha,\beta,\gamma) = \begin{cases} \frac{1}{k}\left[\frac{\alpha.\beta.\exp\left(-\alpha(x-\gamma)\right)}{(1+\exp(-\alpha(x-\gamma)))^{\beta+1}}\right] & m \le x \le n \\ 0 & others \end{cases} \tag{2.12}$$

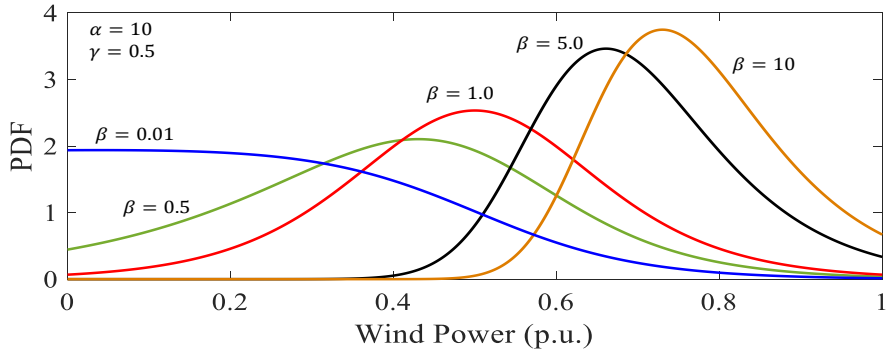
$$F(x) = \begin{cases} 0 & x < m \\ \frac{1}{k}\left[(1+\exp(-\alpha(x-\gamma)))^{-\beta} - (1+\exp(-\alpha(m-\gamma)))^{-\beta}\right] & m \le x \le n \\ 1 & x > n \end{cases} \tag{2.13}$$

$$F^{-1}(c) = \gamma - \frac{1}{\alpha}\ln\left(\left(k.c + (1+\exp\left(-\alpha(m-\gamma)\right))^{-\beta}\right)^{-1/\beta} - 1\right) \tag{2.14}$$

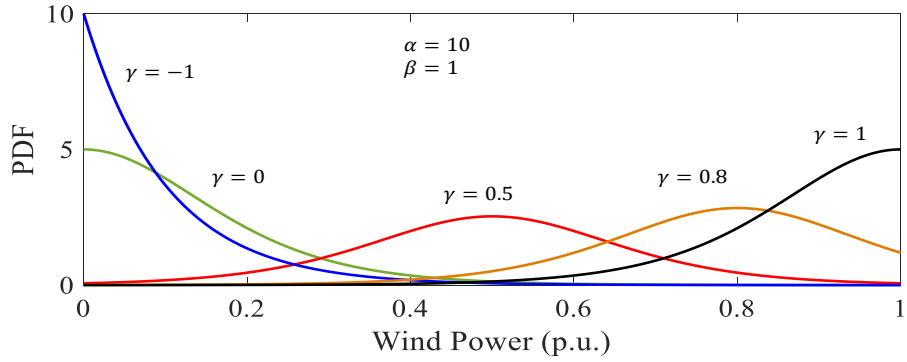
$$k = (1+\exp(-\alpha(n-\gamma)))^{-\beta} - (1+\exp(-\alpha(m-\gamma)))^{-\beta} \tag{2.15}$$



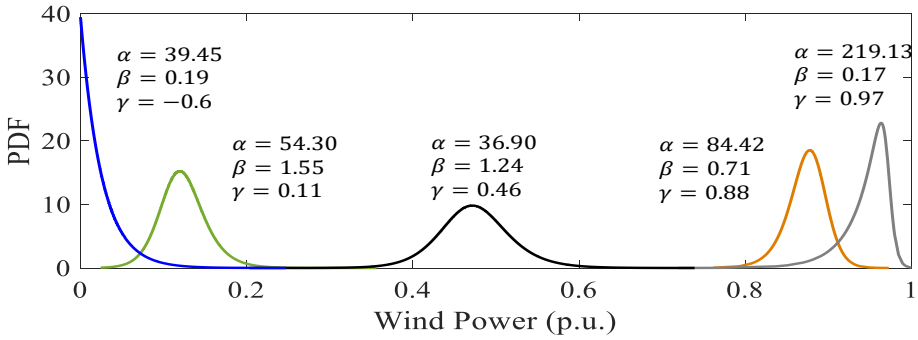
(a)



(b)



©



(d)

Figure 2.5 Illustration of TVD distribution.

2.3 An Overview of Nonparametric Density Estimation

This section briefly explains nonparametric techniques as data-driven approaches for estimating the PDF of random variables.

Suppose Y_t be the random variable for wind power generation at time t with the PDF f_t and CDF F_t. As it is shown in Figure 2.6, the quantile $q_t^{(\alpha_i)}$ with level α_i is expressed as follows [102], [103]:

$$q_t^{(\alpha_i)} = F_t^{-1}(\alpha_i) \tag{2.16}$$

$$\mathbb{Pr}\left(y_t \leq q_t^{(\alpha_i)}\right) = \alpha_i \tag{2.17}$$

The predicted quantile $\hat{q}_{t+k|t}^{(\alpha_i)}$ is an estimate of $q_{t+k}^{(\alpha_i)}$ for lead time $t+k$ based on information available at time t. A single quantile cannot give all information on prediction uncertainties; therefore, m number of quantiles, with nominal levels $\alpha_i, i = 1, \dots, m$ are considered to yield a nonparametric PDF forecast as follows:

$$\hat{f}_{t+k|t} = \left\{\hat{q}_{t+k|t}^{(\alpha_i)}\,; 0 \leq \alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m \leq 1\right\} \tag{2.18}$$

Since $\hat{f}_{t+k|t}$ is made of limited quantiles, the full continuous PDF can be obtained by linear or nonlinear interpolation of predetermined predicted quantiles. After prediction of quantiles and then construction of predictive densities, prediction intervals (PIs) which represent lower quantile $\hat{q}_{t+k|t}^{(\underline{\alpha})}$ with nominal proportion $\underline{\alpha}$ and upper quantile $\hat{q}_{t+k|t}^{(\overline{\alpha})}$ with nominal proportion $\overline{\alpha}$ for future observations with a certain probability can be obtained as (2.19)-(2.20) where β is the level of risk by which the reliability of PIs are determined.

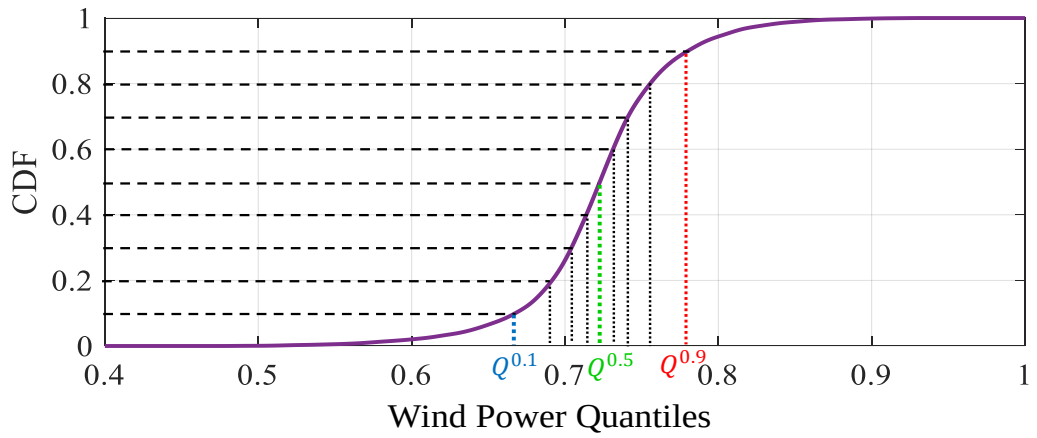


Figure 2.6 Illustration of quantiles of a typical CDF.

$$\hat{I}_{t+k|t}^{(1-\beta)} = \left[\hat{q}_{t+k|t}^{(\overline{\alpha})}, \hat{q}_{t+k|t}^{(\underline{\alpha})}\right] \tag{2.19}$$

$$1 - \beta = \overline{\alpha} - \underline{\alpha} \tag{2.20}$$

There are different ways to obtain the lower and upper quantiles for future observations of wind power such as heuristic optimization-based quantile regression (QR), linear programming-based QR, kernel density estimation (KDE), combination of artificial intelligence (AI) and heuristic optimization techniques. The focus of this chapter is on QR and KDE as basics of non-parametric estimation of wind power PDFs.

2.3.1 Quantile Regression Approach

Quantile regression is an efficient but conventional approach to approximate the conditional probability distribution of any kind of random variable using quantiles. The quantiles are modeled as functions of explanatory variables. To be more specific, the quantile with proportion α is estimated through minimizing the cost function expressed in (2.21) where T is the size of dataset for construction of the model [102], [103]. The cost function includes a non-linear function shown in Figure 2.7.

$$\min \sum_{t=1}^{T} \rho_{\alpha}(y_t - q_t^{(\alpha)}) \tag{2.21}$$

$$\rho_\alpha(x) = \begin{cases} \alpha x & \text{if } x \geq 0 \\ (\alpha - 1)x & \text{if } x < 0 \end{cases} \tag{2.22}$$

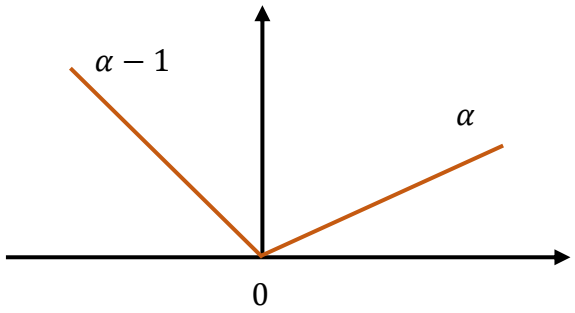


Figure 2.7 Cost function of quantile regression approach.

Note that quantiles obtained by (2.21) are unconditional while for estimation of conditional quantiles, the optimization process (2.21) should be replaced with (2.23) where γ is the main parameter of the model for mapping input x_t to output $q_t^{(\alpha)}$.

$$\min \sum_{t=1}^{T} \rho_\alpha\left(y_t - q_t^{(\alpha)}(x_t, \gamma)\right) \tag{2.23}$$

In (2.23), $q_t^{(\alpha)}(x_t, \gamma)$ reflects a nonlinear model which can be trained based on the historical data. In this research work, extreme learning machine (ELM) model shown with $g(x_t, w^\alpha)$ is replaced with $q_t^{(\alpha)}(x_t, \gamma)$. w^α is the output weights matrix of trained ELM when quantiles with nominal proportion α are used as the outputs of ELM.

2.3.2 The Basics of Kernel Density Estimators

Let $\{x_i\}_{i=1}^N$ be a set of independent samples from a random variable with unknown density $f(x)$. The kernel density estimate of f at point x is shown by $\hat{f}$ and expressed via (2.24).

$$\hat{f}(x; h) = \frac{1}{Nh} \sum_{i=1}^{N} \kappa((x_i - x)/h) \qquad x \in \mathbb{R} \tag{2.24}$$

where $\kappa(u)$ is a kernel function with BW parameter h. Generally, kernel function $\kappa(u)$ is a unimodal probability density that is symmetric about zero and satisfies constraints (2.25)-(2.27).

$$\int_{\mathbb{R}} \kappa(u)du = 1 \quad , \ \kappa(u) > 0 \tag{2.25}$$

$$\int_{\mathbb{R}} u\,\kappa(u)du = 0 \tag{2.26}$$

$$0 < \mu_2(\kappa) = \int_{\mathbb{R}} u^2\kappa(u)du < \infty \tag{2.27}$$

where $\mu_{2(\kappa)}$ is the second moment of the kernel $\kappa(u)$.

There are both symmetric and asymmetric kernel functions by which an unknown probability density can be estimated. Most nonparametric estimations are based on symmetric kernel functions. So far, different kernel functions have been used, including Normal, Epanechnicov, Biweight, Triweight, and Uniform. In this chapter, Normal kernel function in (2.1) is used for introducing an efficient nonparametric PDF estimation because it has highest smoothing properties among others [104]. In (2.1), standard deviation σ denotes the smoothing parameter (also referred to as BW) of Normal kernel and μ is the location parameter which is equivalent to the value of each random sample. The standard Normal kernel is expressed by (2.28).

$$\varphi(u) = \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \tag{2.28}$$

The efficiency of kernel density estimators highly depends on the optimal selection of BW parameter h in (2.24). Figure 2.8 shows the effect of increasing BW on the estimated PDF of a typical WPPE with six samples. In the next section, an advanced BW selection technique is utilized to gain the optimal value of h and accurately estimate the PDF of WPPE.

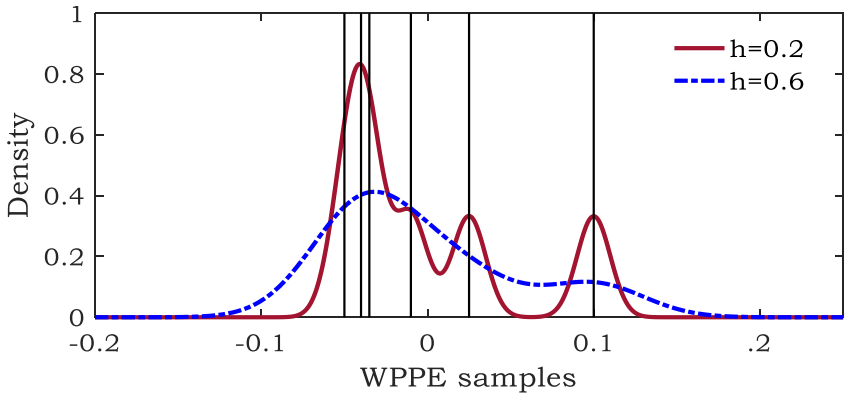


Figure 2.8 Illustration of BW effect on the estimated PDF of WPPE with six samples.

2.3.2.1 An Advanced Plug-In (API) Bandwidth Selection Technique for PDF Estimation

In the previous section, the PDF of WPPE with only six samples obtained using trial and error as shown in Figure 2.8. It is found that the value of BW highly affects the estimated PDF, and for the WPPE with huge number of samples we need a criterion to optimally estimate WPPE PDF. The well-defined criterion $MISE$ shown in (2.29) is generally used for optimal selection of h in (2.24). It includes two components, the integrated squared bias and integrated variance that can measure the expected error between estimated ($\hat{f}$) and true (f) densities. Thus, the aim is to find the value of h by which $MISE$ is minimized. For a detailed discussion, the reader is referred to [104].

$$MISE\left(\hat{f}(x;h)\right) = \mathbb{E}_f\left\{\int\left(\hat{f}(x;h) - f(x)\right)^2 dx\right\} = \int Bias\left(\hat{f}(x;h)\right)^2 dx + \int Var_f\left(\hat{f}(x;h)\right) dx \tag{2.29}$$

API BW selection technique finds the minimum value of the asymptotic approximation (2.29) using h_{opt} through following easy-to-implement l-stage ($l > 2$) algorithm [50].

1: Choose l, e.g. $l = 4$.

2: Set $\rho_l = 0.001$, and find the plug-in estimator $\|\widehat{g^{(l)}}\|^2$ using (2.30).

3: Find the pilot BW $\hat{\rho}_J$ via (2.31).

4: Find again the plug-in estimator $\|\widehat{g^{(l-1)}}\|^2$ using (2.30) and $\hat{\rho}_J$ obtained from the previous stage, and continue this procedure until $\|\widehat{g^{(2)}}\|^2$ and consequently $\hat{\rho}_1$ are acquired.

5: If $|\hat{\rho}_1 - \rho_l| < \varepsilon$, the optimal BW for Normal estimator (2.1) equals $h_{opt} = \sqrt{0.9(\hat{\rho}_1)}$; else go to step 2 with $\rho_l = 0.9(\hat{\rho}_1)$.

$$\|\widehat{g^{(J)}}\|^2 = \frac{(-1)^J}{N^2}\sum_{k=1}^{N}\sum_{m=1}^{N}\varphi^{(2J)}\left(X_k, X_m; 2\rho_J\right),\ J \geq 1 \tag{2.30}$$

$$\hat{\rho}_J = \left[\frac{\left(1+1/2^{J+0.5}\right)\left((2J-1)\times\ldots\times3\times1\right)}{\left(3N\sqrt{\pi/2}\|\widehat{g^{(J+1)}}\|^2\right)}\right]^{\frac{2}{3+2J}} \tag{2.31}$$

2.4 Proposed Methodology for Conditional WPPE Distribution Modeling

In this section, the proposed methodology utilizes two steps. First, historical real and predicted values of wind power for the whole year is obtained, and wind power dataset is divided into different generation levels (GLs). Then, in the second step, since the PDF of WPPE depends on wind power GL (i.e., prediction value), conditional distribution of WPPE is calculated based on proposed BW selection technique and KDE technique with Normal kernel function.

2.4.1 Step 1: Extraction of WPPE samples in Historical Wind Power Dataset for a Given Prediction Value

In this chapter, the prediction range [0,1] (p.u.) is divided into 20 GLs with width of 0.05 for each level ($N_{GL} = 20$). According to Figure 2.9, at each time instant in the historical dataset, there are two values; real and predicted values. In each GL, there are a series of real samples with corresponding prediction values. Therefore, based on the prediction value in one step ahead, the associated GL is identified and the PDF of real wind power and WPPE in that GL is estimated using KDE technique.

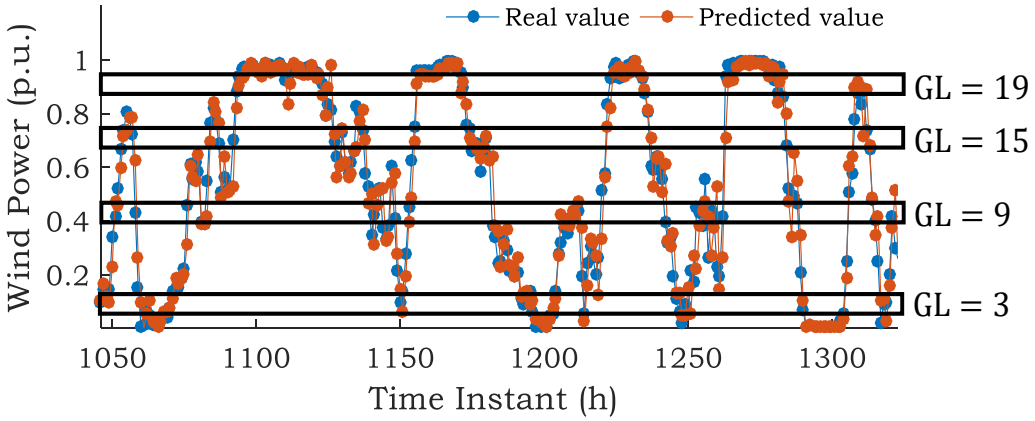


Figure 2.9 Historical real and predicted wind power dataset of Centennial wind farm for a short time period in 2015.

2.4.2 Step 2: Conditional PDF Estimation for WPPE

Because of the stochastic nature of wind power, there is not any fixed relationship between WPP time series and real wind power time series; however, they are statistically dependent on each other with high correlation. Therefore, by changing the prediction points, the distribution of real wind

power and WPPE varies. There are two different ways to quantify the conditional PDF of WPPE: (i) the first one is to use real wind power samples for the GL that encloses point prediction $\hat{p}$. Then, PDF of WPPE is obtained by shifting the estimated wind power PDF to left by $\hat{p}$, and (ii) the second one is to directly use the WPPE samples for that specific GL.

Given the point prediction $\hat{p}$, first the corresponding GL is identified then the conditional PDF of WPPE is obtained by (2.32).

$$\hat{f}_j(e|\hat{p}) = \frac{1}{N_j h_{opt_j}} \sum_{i=1}^{N_j} \varphi\left(\left(e_{i,j} - e\right)/h_{opt_j}\right) \tag{2.32}$$

where $j = 1, \dots, \mathrm{N_{GL}}$ and $e \in [-\hat{p}, 1 - \hat{p}]$.

2.5 Simulation Results

2.5.1 Dataset

In this chapter, the proposed approach is implemented using wind power dataset from Canada. The historical data of the whole year of 2015 from Centennial wind farm in south Saskatchewan with 150 MW capacity and 1-min data resolution is used [5], [105]. Also, the real wind power dataset and the corresponding predicted values are selected for 1-hour prediction horizon. Of course, without loss of generality, other prediction horizons could be considered.

2.5.2 WPPE and Real Wind Power Distribution Models for a Given Prediction Value

In this section, based on the proposed methodology, the PDF of WPPE and real wind power for different prediction values, i.e., 0.1, 0.4, 0.7, and 0.9 (p.u.), are constructed to examine how well various distribution models can fit the actual PDF. The actual PDF is represented by a probability density histogram (PDH). PDHs, introduced by Karl Pearson [106], are usually used as a typical way to represent the distribution of numerical data by estimating the probability density of a continuous random variable. The first step to build a PDH is to divide the whole range of numerical data into K bins with width Δ. The bins are usually consecutive, distinct, and equally sized. Then, the number of samples that fall into each bin should be accounted. K may be chosen directly by

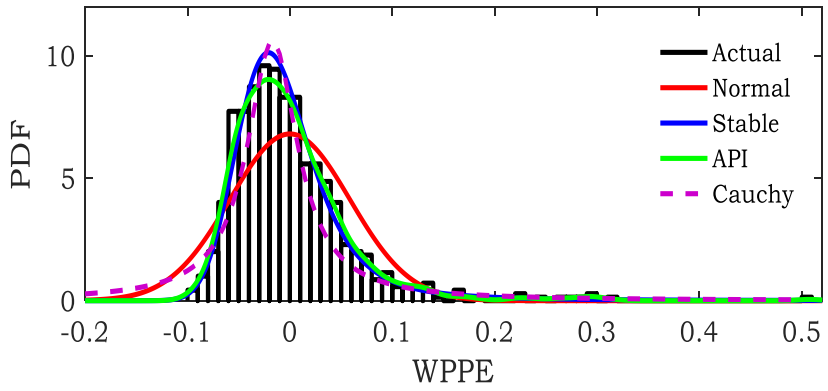
end-user or may be calculated using a formula which depends to Δ as shown in (2.33).

$$\mathrm{K} = \frac{\max(x) - \min(x)}{\Delta} \tag{2.33}$$

There are different approaches to build a PDH, e.g., Scott's normal reference rule, Freedman–Diaconis rule, Doane's formula, and Rice rule. In this chapter, Freedman–Diaconis rule, shown in (2.34), is used because it is less sensitive than other approaches to outliers in the dataset [107].

$$\Delta = \frac{2\,\mathrm{IQR}(x)}{M^{1/3}} \;,\; \mathrm{IQR}(x) = Q(x)^{75\%} - Q(x)^{25\%} \tag{2.34}$$

where M is the number of data points, and IQR is a measure of statistical dispersion that equals to the difference between 75^{th} and 25^{th} percentiles. Figure 2.10 to Figure 2.13 illustrate the PDH (as actual distribution) and fitted distributions, i.e., Normal, Stable, Cauchy, and API, of WPPE for prediction values 0.1, 0.4, 0.7, and 0.9, respectively. The prediction values are independently associated to GLs 3, 9, 15, and 19. As can be seen, because of more flexibility, Stable parametric model can fit the actual distribution of WPPE better than Normal and Cauchy. However, API model as a nonparametric model outperforms Stable distribution model. Unlike other fitting models, API can detect different peaks over the main mass and tail areas. The actual and API model of real wind power distribution are shown in Figure 2.10 to Figure 2.13 for comparison purpose. It is found that the distribution of WPPE is not exactly the same as shifted distribution of real wind power if we use the second way; however, they can be equivalent to each other.



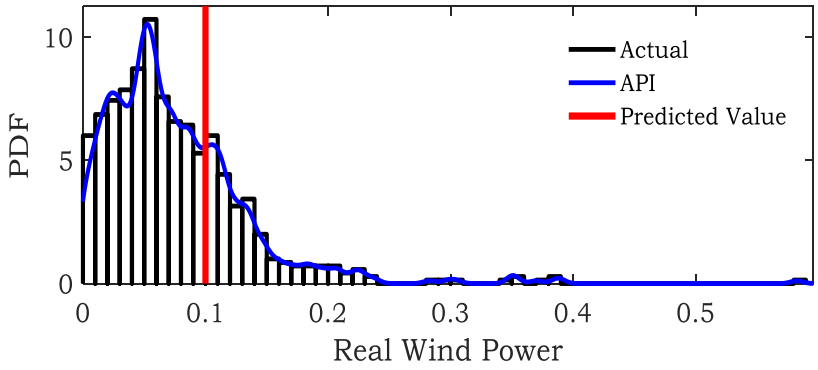


Figure 2.10 Probability distributions for prediction value 0.1 (p.u.). Upper: WPPE, and Lower: real wind power.

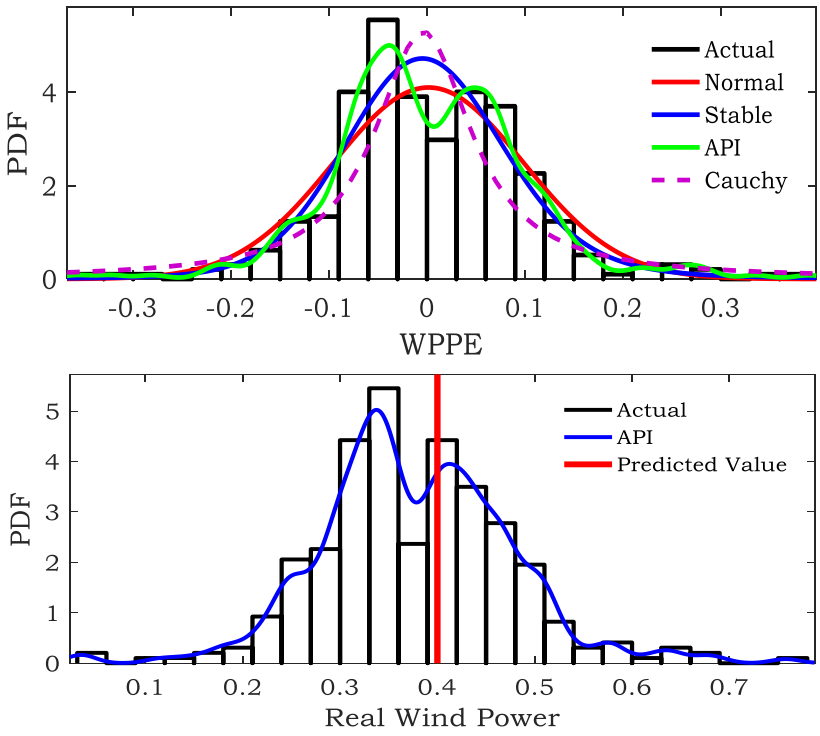


Figure 2.11 The probability distributions for prediction value 0.4 (p.u.). Upper: WPPE, and Lower: real wind power.

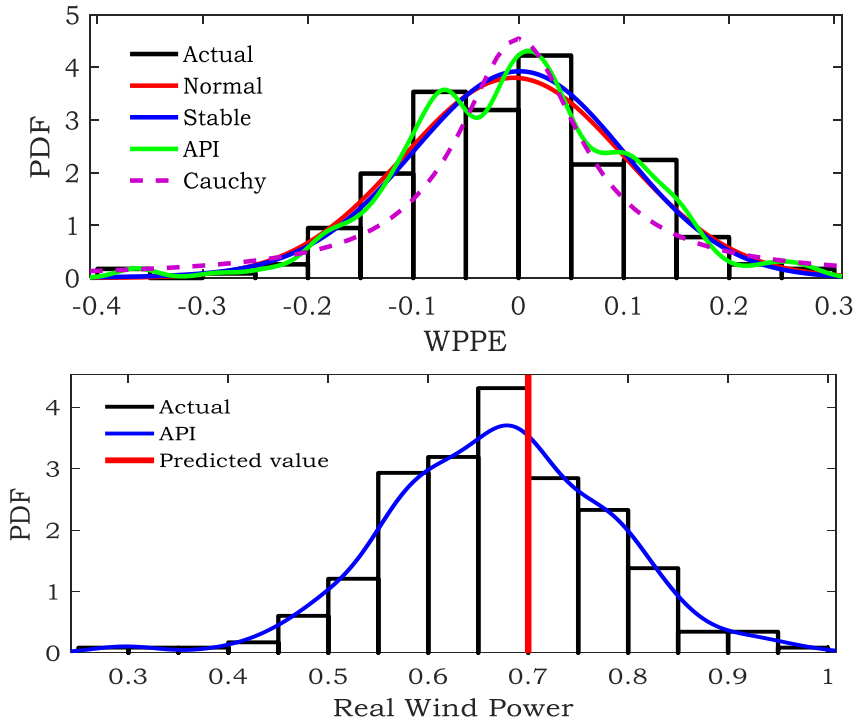


Figure 2.12 The probability distributions for prediction value 0.7 (p.u.). Upper: WPPE, and Lower: real wind power.

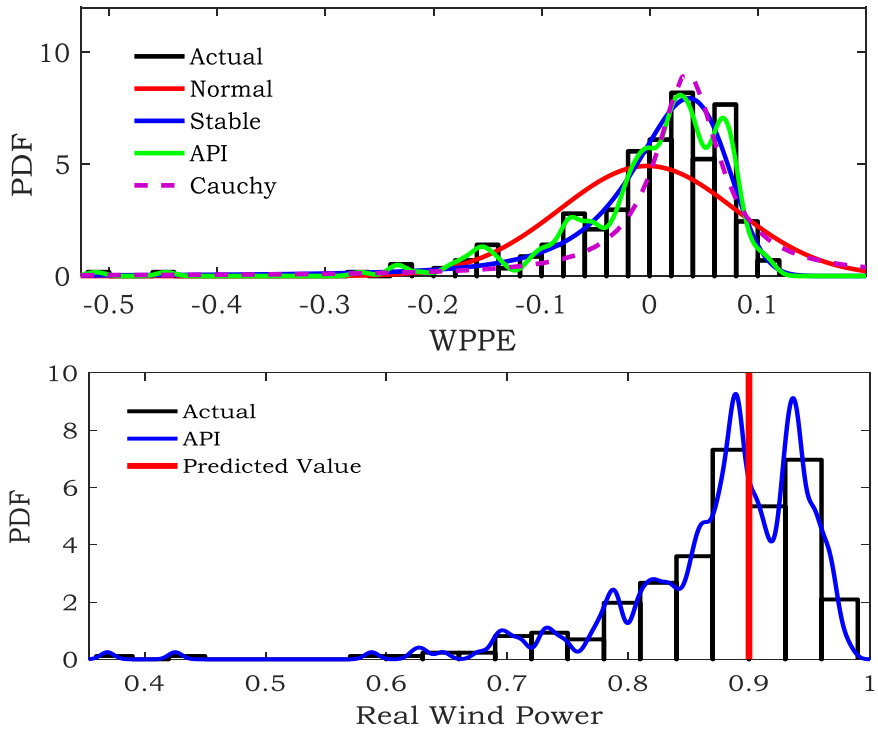


Figure 2.13 Probability distributions for prediction value 0.9 (p.u.). Upper: WPPE, and Lower: real wind power.

2.5.3 Quantitative Comparison of Distribution Models

To quantitatively distinguish the best fitting model for a PDH, the root mean-square error (RMSE) between the actual CDF and the benchmarks' CDF is used via (2.35).

$$\mathrm{RMSE_{GL}} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(CDF_{act}(x_m) - CDF_{ben}(x_m))^2} \tag{2.35}$$

Table 2.1 summarizes the RMSEs for Normal, stable and API distribution models for four GLs that can cover the whole range of real wind power ([0, 1] p.u.). It is seen that the value of RMSE in the proposed nonparametric distribution model, API, is much smaller than benchmarks. Therefore, API can satisfactorily fit the actual WPPE distribution for different prediction values.

Table 2.1 RMSE of various distribution models for WPPE for 1-hour ahead prediction.

GL	Prediction value range (p.u.)	RMSE (%)		
		Normal	Stable	API
3	[0.10, 0.15]	2.70	0.45	**0.24**
9	[0.40, 0.45]	1.10	0.80	**0.25**
15	[0.70, 0.75]	0.88	1.00	**0.80**
19	[0.90, 0.95]	3.20	0.80	**0.18**

2.6 Summary

The uncertainty of generated wind power in power systems should be appropriately modeled for probabilistic assessment of power systems like ED, UC, stability, and reliability problems using efficient probability distributions. The parametric distributions such as Normal, Stable, etc. cannot well represent the main features of actual WPPE and real wind power distributions. In this chapter a nonparametric distribution modeling using an advanced bandwidth selection technique, API, was proposed to properly fit the actual distributions. The results show that compared with other parametric distribution models, the proposed model can well fit the actual distributions of WPPE

and real wind power for different prediction horizons and generation levels; thereby, benefit the power system operators in stochastic ED and UC problems as well as probabilistic security assessment analysis.

Chapter 3

A Convex Model for a Look-Ahead Economic Dispatch Problem With an Efficient Wind Distribution Modeling

3.1 Introduction

The increasing penetration of uncertain wind power in overall generation around the world can adversely affect the operation, flexibility, and security of large-scale power systems in terms of reserve depletion, transmission line overloading, and incremental cost of total generation [108]. Accurate wind power generation uncertainty modeling should be accomplished as the first step to handle the uncertainty in power systems [29-34], [71], [91], [109-112]. In the second step, optimal reserve and generation scheduling in a look-ahead economic dispatch (ED) problem should be carried out using an efficient stochastic methodology. Chance-constrained ED (CCED) is one such efficient stochastic methodology for uncertainty management [29], [30], [71], [109]. Because the conventional cost function (CCF) in the CCED problem includes the expected values of wind power overestimation and underestimation costs expressed by a non-linear and non-convex structure, it is not efficient to solve the CCED problem in a conventional optimization framework. In addition, parametric distribution models containing closed forms for cumulative distribution functions (CDFs) are usually used for wind power [29], [30]-[34], [71], [109]. This structure of CCF prevents power system operators from using more accurate distribution models (e.g., non-parametric models) and efficient optimization algorithms such as mixed-integer linear programming (MILP). Therefore, sequential linear programming (SLP) based on Taylor series expansion is often used as an optimization approach for the CCED problem [29], [30], [71], [109]. The main motivation of this chapter is to propose a generic convex model that is convex with respect to the continuous variables of the problem. Also, it is adaptable to both parametric and non-parametric wind power distribution models and can provide the opportunity to formulate the CCED model like an MILP model while a finite convergence to a global optimal solution for the linearized model is guaranteed.

Wind power distribution modeling is the cornerstone of overall uncertainty modeling in power systems [110]. Recent studies have proposed parametric probability density functions (PDFs) (e.g., Cauchy [31], Weibull, hyperbolic [32], Gaussian [33], and beta [34]) for wind power generation. In [31], a more flexible and complex distribution, i.e., the Levy α-stable distribution, is also proposed for probabilistic reserve sizing in power systems. Efficient versatile and truncated versatile distributions (TVD) for wind power forecast error are proposed and implemented in a look-ahead stochastic ED in [29], [30]. Conditional wind power forecast error modeling is presented in [110] using a Gaussian copula function for generation scheduling. Alternatively, a Gaussian mixture model (GMM) and a versatile mixture distribution (VMD) are adopted in [71], [109], to represent more accurate wind power distribution models. However, the constructed distributions might show boundary effects because Gaussian and versatile components are not bounded between 0 and 1 p.u. In recent researches, a generalized gaussian mixture model (GGMM) is proposed by [111], [112] for statistical representation and probabilistic forecasting of wind power ramps with especial features like duration, rate, and magnitude. In [91], the Copula Theory is used as an efficient approach to model the dependency structure among wind resources. It uses dependent discrete convolution to obtain the distribution of aggregate wind power in high-dimensional cases. Although dependency structure modeling is important, it is a complex process and is not in the main scope of this chapter for overall uncertainty modeling. To deal with the uncertainty of wind power in real-world power system applications, many researchers have proposed different methodologies. Stochastic optimization (SO) is proposed for look-ahead ED in [66-68] to manage the uncertainty of renewable energy resources. In SO, several scenarios are extracted from a predetermined probability distribution to consider the possible values of uncertain variables. However, the optimal solution and computational efficiency strongly depend on the number of scenarios and the corresponding probabilities. In [29], [30], [70], [71], [109], [113], chance-constrained SO (CCSO) is used as a more efficient method by adjusting a predefined level of risk. In this method, accurate distribution models for wind power sources are required to preserve the security of the system with a certain probability level. In [72-77] distributionally robust chance-constrained (DRCC) optimization is used, where instead of accurately estimating the wind power probability distribution, a moment-based ambiguity set is defined to cover a family of distributions for uncertainty sources. It can be used in diverse areas such as distribution system planning [72], AC optimal power flow [73], [75], [76], and energy and reserve dispatch [74]. The

keen readers are referred to [77] for mathematical details of DRCC. Robust optimization (RO) as an alternative method that improves the security of the system allows wind power to vary in a given uncertainty range [81-83]. However, RO leads to an overly conservative solution because it minimizes the cost of the worst-case scenario. Interval optimization (IO) uses prediction intervals (PIs) and central prediction point to minimize the system operating cost. Compared to CCSO and RO, IO is respectively more conservative and less precise [70], [113], [81-83],[114]. In [90], a two-stage optimization process is proposed for battery energy storage capacity for alleviating wind curtailment in a second-order cone programming framework while wind power is modeled by lower and upper intervals. Considering large intra-interval variations of wind power, an intra-interval security dispatch is proposed in [115] to provide a trade-off between economics and security. A strategic reserve purchasing is proposed in [116] to mitigate wind power uncertainty in a real time market and avoid a predetermined penalty for wind power producers (WPPs). In [117], a convex model is proposed for a risk-based unit-commitment that considers wind power uncertainty to manage wind curtailment, load shedding, and line overflow. This chapter focuses on CCSO technique for wind power uncertainty management. Based on the literature, the approaches utilized in CCED have used approximate calculations for partial derivatives of the CCF, which might lead to a non-optimal solution. The above-mentioned methodologies for wind power uncertainty modeling and handling in ED problems have their own merits and demerits. The literature, however, lacks an efficient CCED model for simultaneous incorporation of non-parametric distribution models and highly accurate linearization techniques for the cost function. Figure 3.1 shows an overview of the proposed MILP-based CCED model in this study where the building blocks (a) and (b), as the main novelty of this study, are proposed for an efficient rolling dispatch system. All details of each block are explained in Section 3.2 to Section 3.4 through representation of Algorithm 1, Algorithm 2, Algorithm 3, and the general structure of the proposed framework.

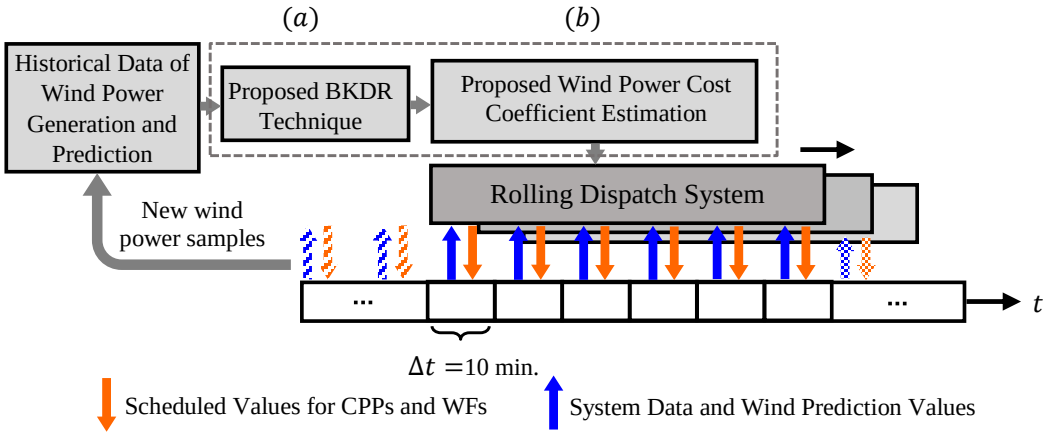


Figure 3.1 Overview of the economic dispatch model with the proposed building blocks: (a) BKDR block and (b) wind power cost coefficient estimation block.

This chapter proposes an efficient CCED model and makes the following contributions:

- An efficient methodology for construction of a new cost function (NCF) is proposed for nonlinear CCED models by which power system operators can use parametric and non-parametric distribution models for wind power.
- Because of the proposed NCF, a highly accurate linearization technique, i.e., piecewise linearization, can be used to convert the non-linear CCED into an efficient MILP-based model.
- For the first time, an efficient wind power distribution modeling is proposed using finite sample moments (e.g., 10 to 20) and the beta kernel density representation (BKDR) technique.
- Because the BKDR technique represents the target distribution using several beta components with restricted parameters' range, it avoids boundary effects while capturing certain features of wind power distribution without reflecting overfitting problem.

3.2 Wind Power Distribution Modeling Using Beta Kernel Density Representation Technique

The reconstruction of a probability distribution $f(x)$ using a limited number of moment data μ_n is called the truncated moment problem first proposed by *Stieltjes* and stated as follows [118].

$$\mu_n = \int x^n f(x)\, dx \quad , \quad n = 0,1,2, \dots, N \tag{3.1}$$

The question is how to use the information available in μ_n to recover the corresponding target function $f(x)$ with high accuracy. If the target PDF $f(x)$ is bounded in interval $x \in [a, b]$, the truncated *Hausdorff* moment problem is realized [119]. To obtain a satisfactory solution to this problem, two steps should be followed. First, an appropriate representation methodology, i.e., kernel density representation (KDR), for the function to be recovered should be chosen. Second, some important features or *a priori* knowledge about the target PDF $f(x)$, such as boundary conditions, tail behavior, modality information, and the main-mass interval, need to be determined to improve the solution [119]. For a truncated *Hausdorff* moment problem, it suffices to determine the main-mass interval referred to as $\widehat{D}$.

3.2.1 KDR-Based Wind Power PDF Representation

KDR is a parametric representation of PDF $f(x)$ by means of a weighted sum of known non-negative kernel density functions (KDFs), as shown in (3.2)-(3.3) [118], [119].

$$\hat{f}(x; \boldsymbol{p}) = \sum_{i=1}^{I} p_i\, \kappa(x; x_i, h) \qquad x \in [0,1] \tag{3.2}$$

$$\sum_{i=1}^{I} p_i = 1 \quad , \quad p_i \geq 0 \quad , \quad \boldsymbol{p} = [p_1, \dots, p_I]^{\mathrm{T}} \tag{3.3}$$

Bandwidth parameter h, which controls the smoothness of the overall fit, can be determined either from a predetermined optimal range or by minimizing the estimated error of the KDR model as it is proposed in this work. Equation (3.2) means that an unknown density function $f(x)$ can be represented by I kernels placed at uniformly distributed locations x_i of the sample space [0,1], where p_i measures the contribution of the i^{th} kernel to the overall density evaluation and should meet the constraints in (3.3). According to (3.2), the KDF $\kappa(\cdot)$, the kernel weights p_i, the kernel

locations x_i, and bandwidth h should be optimally determined to efficiently reconstruct $f(x)$. The optimal values of parameters p_i, x_i, and $,h$ are obtained through minimizing a performance criterion, e.g., $\|\hat{f} - f\|^2$. Also, instead of using the whole interval [0,1] in (3.2), a main-mass interval, e.g., $\widehat{D}$, can be estimated by which the performance of the proposed technique is improved. The details of each part are provided in sections 3.2.2, 3.2.3, and 3.2.4, respectively.

3.2.2 A Proper Choice for Beta Distribution

As mentioned in Section 2.2.3, beta distribution is an efficient parametric distribution which can properly mimic the changes in wind power distribution. The features and equations of beta distribution are mentioned again in this section for the sake of convenience.

For $x \in [0,1]$, the beta PDF with two main parameters (υ, ζ) expressed by (3.4) is an efficient choice for the KDR process because of four main advantages expressed as follows: (i) it is bounded between 0 and 1 like normalized wind power distribution; thus, by summation of several beta distributions over the range [0,1] a bona fide PDF is obtained without occurring boundary effects; (ii) the set of beta parameters (υ_i, ζ_i) is easily related to the set of location and bandwidth parameters (x_i, h) using a closed form (i.e., algebraic system (3.6)-(3.7)); (iii) there is a simple function by which the moments of beta distribution can be calculated (i.e., equation (3.9)); (iv) the flexible shape of the beta distribution symmetrically changes so that it coincides with the skewness of wind power PDF from the lower boundary region to the upper boundary region [34]. Thus, according to (3.2), the target PDF, $f(x)$, can be approximated by (3.5). To calculate the parameters υ_i and ζ_i, a natural choice is to use mode and variance expressions of f_B in (3.6)-(3.7). Thus, for a given set of values (x_i, h), the parameters (υ_i, ζ_i) are obtained by solving the algebraic system (3.6)-(3.7). Figure 3.2 shows beta distributions with five sets of parameters (υ_i, ζ_i) over the range [0,1].

$$\kappa(x; x_i, h) = f_B(x; \upsilon, \zeta) = \frac{1}{B(\upsilon,\zeta)}\, x^{(\upsilon-1)}(1-x)^{(\zeta-1)} \tag{3.4}$$

$$\hat{f}(x; \boldsymbol{p}, \boldsymbol{\upsilon}, \boldsymbol{\zeta}) = \sum_{i=1}^{I} \frac{p_i}{B(\upsilon_i,\zeta_i)}\, x^{(\upsilon_i-1)}(1-x)^{(\zeta_i-1)} \tag{3.5}$$

$$x_i = mode\big(f_B(x; \upsilon_i, \zeta_i)\big) = \frac{\upsilon_i - 1}{\upsilon_i + \zeta_i - 2} \tag{3.6}$$

$$h^2 = variance\big(f_B(x; \upsilon_i, \zeta_i)\big) = \frac{\upsilon_i\, \zeta_i}{(\upsilon_i+\zeta_i)^2(\upsilon_i+\zeta_i+1)} \tag{3.7}$$

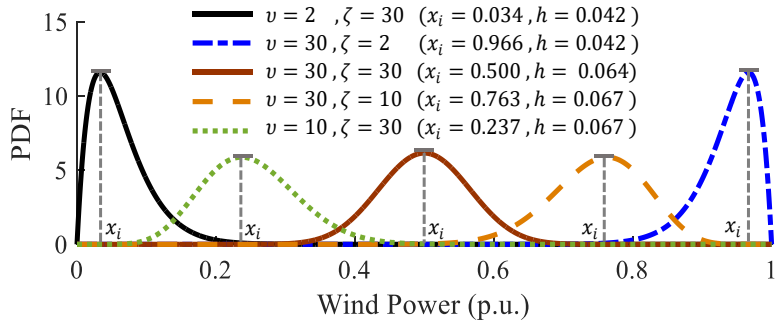


Figure 3.2 Illustration of several beta distribution over the range [0,1].

3.2.3 Optimal Calculation of KDF Weights

Given a finite number of sample moments $\boldsymbol{\mu} = [\mu_1, \dots, \mu_N]^{\mathrm{T}}$ of an unknown target PDF $f(x)$, I locations on the sample space [0,1] ($I \leq N$), and bandwidth value h, the weights p_i are calculated by solving the simple minimization problem (3.8). For the numerical solution of (3.8), the function *lsqnonneg* of the MATLAB optimization package is employed. The matrix $\boldsymbol{C}_{N\times I}$, which contains N moments of the beta kernel on each location, is obtained by (3.9) [119]. Therefore, the set of (υ_i, ζ_i) or the locations x_i and bandwidth h should be optimally selected to achieve an optimal solution for (3.8).

$$\widehat{\boldsymbol{p}} = \arg\min_{p_i} \frac{1}{2}\|\boldsymbol{\mu} - \boldsymbol{C}\boldsymbol{p}\|^2 \qquad \text{s.t. (3.3)} \tag{3.8}$$

$$\boldsymbol{C}_{n,i} = \int_0^1 x^n f_B(x; \upsilon_i, \zeta_i)\, dx = \prod_{s=0}^{n-1} \frac{\upsilon_i + s}{\upsilon_i + \zeta_i + s} \qquad n = 0,1,\dots,N \;\;,\;\; i = 1,2,\dots,I \tag{3.9}$$

$$RMSE = \left[\frac{1}{N_s}\sum_{i=1}^{N_s}\left(\hat{F}(x_i) - F_e(x_i)\right)^2\right]^{1/2} \tag{3.10}$$

3.2.4 Optimal Parameters of Beta KDF

Simulations show that kernel locations x_i and bandwidth h have a sizable impact on the quality of the overall estimation. It is much easier to find the optimal set (x_i, h) than (υ_i, ζ_i) because the ranges of x_i and h are definite. To find the above-mentioned parameters, a strategy is adopted in this section and shown by Algorithm 1 in Figure 3.3. It uses an iterative simplified sequentially adaptive (SSA) algorithm to find optimal locations x_i and a parallel computing-based procedure

to find the optimal value of bandwidth h [118]. The basic idea behind the SSA algorithm is to propagate an initial set of kernel locations based on the values of their weights and the distance between successive locations. SSA algorithm is based on two criteria: (i) the death of insignificantly weighted kernels and (ii) the birth of new kernel locations. The first criterion is fulfilled by defining a small threshold value denoted by ε_d. Using this criterion, the kernels with weights smaller than ε_d are removed. The second criterion is met by identifying the kernels with the highest weight and adding two location points nearby. The progressive impoverishment of the resulting KDR model is avoided by considering a small number of iterations k (e.g., $\mathbb{K}$ =5). Algorithm 1 finds the optimal bandwidth h_{opt} through a parallel-computing process based on the minimization of the root mean square error ($RMSE$) in (3.10).

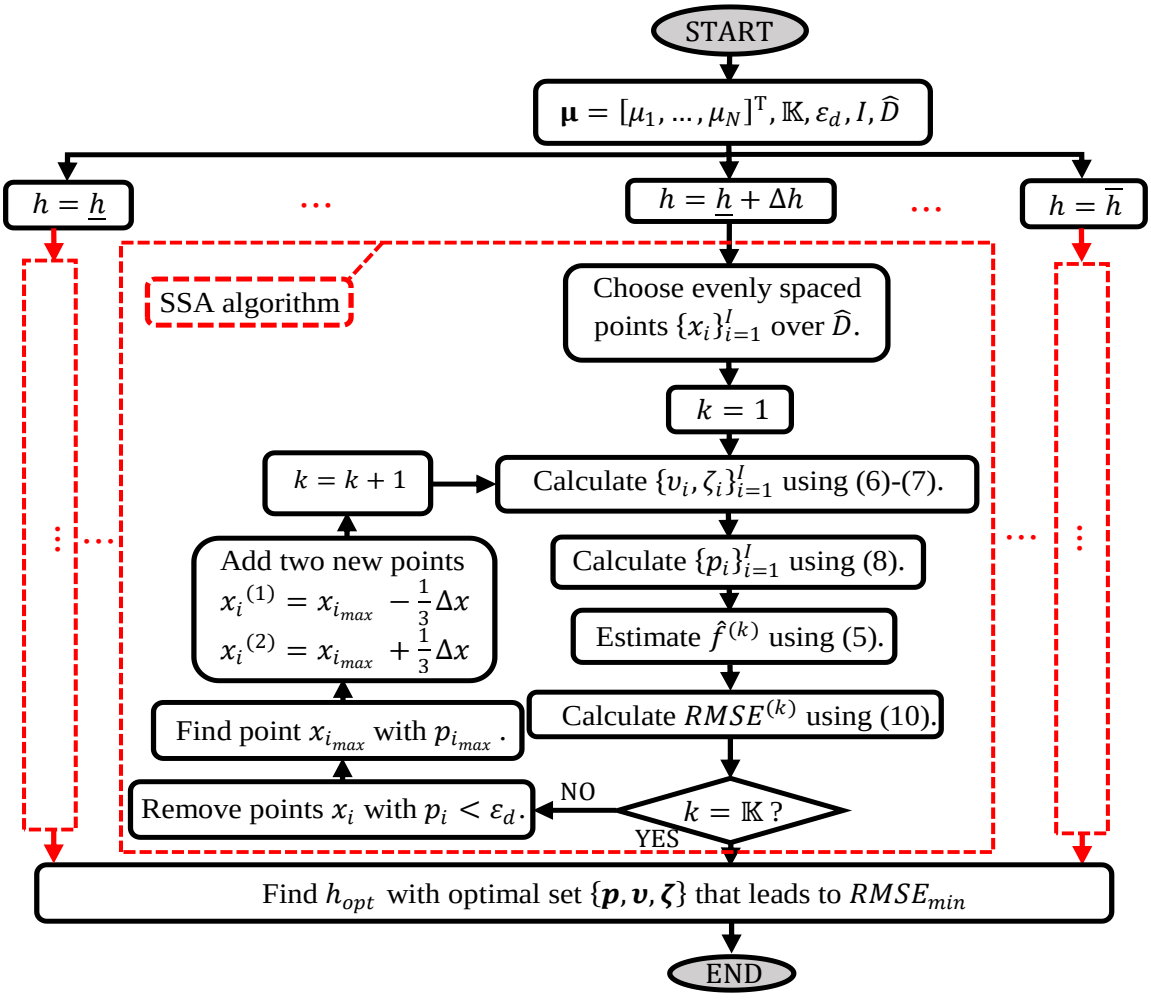


Figure 3.3 Flowchart of Algorithm 1 which determines optimal values of beta kernels parameters p_i, x_i, and h where $\Delta x = min\{x_{i_{max}} - x_{i_{max}-1}, x_{i_{max}+1} - x_{i_{max}}\}$.

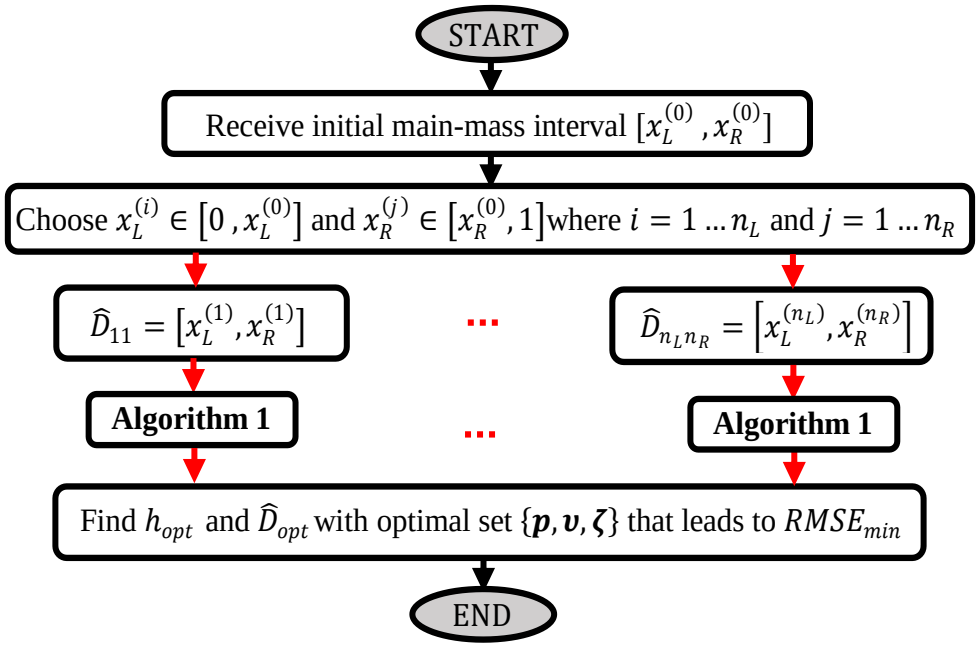


Figure 3.4. Flowchart of Algorithm 2 which determines main-mass interval $\widehat{D}$.

The $RMSE_{min}$ is based on CDF estimation and ensures that the estimated CDF contains existing features of the empirical CDF. The $RMSE$ minimization procedure is very fast and simple. Suppose h lies in the range $H = [\underline{h}, \overline{h}]$, and let $h_j, j = 1, \dots, m$ be m evenly spaced points in H. The SSA algorithm is performed for each h_j and iteration k to estimate $\widehat{F}$ and compute $RMSE$ to find optimal set $\{\boldsymbol{p}, \boldsymbol{v}, \boldsymbol{\zeta}\}$. In order to avoid PDFs with many peaks and valleys as an over-fitted modeling, three input parameters I, h, and $\mathbb{K}$ in the proposed SSA algorithm are set on reasonable ranges $I \in [5,20]$, $h \in [0.01,0.10]$, and $\mathbb{K} \in [2,5]$. Therefore, high computational efficiency and accuracy of distribution modeling can be assured without overfitting. As indicated, the overall improvement of the PDF reconstruction procedure is achieved by determining the main-mass interval $\widehat{D}$ in which the beta functions are placed. In this work, instead of setting the main-mass interval to be equal to the support (i.e., $\widehat{D} = [0,1]$) and varying the number of location points I based on some criteria, an initial main-mass interval $[x_L^{(0)}, x_R^{(0)}]$ is chosen, then the endpoints of the main-mass interval change. In this way, a set of intervals are obtained, and Algorithm 1 can be applied to each interval. A reasonable and simple choice for endpoints $[x_L^{(0)}, x_R^{(0)}]$ that covers the main part of the probability mass is based on the empirical PDF. The main steps of this procedure are given by Algorithm 2 in which Algorithm 1 is the main part as shown in Figure 3.4.

3.2.5 Proposed Conditional Modeling of Wind Power CDF

As the level of wind power uncertainty greatly depends on the forecast value, the wind power PDF has a conditional relationship with the wind power forecast value [32], [110]. In this section, the conditional model of wind power PDF is presented based on the proposed wind power distribution modeling. However, the performance of the proposed probability distribution modeling does not depend on the deterministic prediction accuracy since the proposed model tunes the range of input parameters I, h, $\widehat{D}$, and $\mathbb{K}$ by pre-processing of received actual and predicted time series. For given forecast value $\widehat{\omega}_t$, the wind power PDF is specified using (3.11) with predetermined optimal sets of $\boldsymbol{p}$, $\boldsymbol{v}$, and $\boldsymbol{\zeta}$. To assign a certain set of $\boldsymbol{p}$, $\boldsymbol{v}$, and $\boldsymbol{\zeta}$ to forecast value $\widehat{\omega}_t$, the forecast range $[0,1]$ p.u. is first divided into several equally sized power bins (PBs). N_{PB} depends on the length of the wind power time series under study, which usually equals to 20 if there are sufficient samples (i.e., X_i) inside each PB. Then, using the wind power samples inside each PB, the sample moments (i.e., μ_n) of wind power are calculated via (3.12). Finally, using Algorithm 2, the optimal sets $\mathbf{P}$, $\boldsymbol{\Upsilon}$, and $\mathbf{Z}$, expressed by (3.13), are obtained for the related time series. Therefore, different PBs have different sets of $\boldsymbol{p}$, $\boldsymbol{v}$, and $\boldsymbol{\zeta}$ that lead to PDFs with diverse features. The matrices $\{\mathbf{P}, \boldsymbol{\Upsilon}, \mathbf{Z}\}$ can be updated weekly. The proposed wind power CDF modeling might be run every day, every week, etc. whenever enough new wind power samples are available for updating the inputs.

In this chapter, for the sake of simplicity, the time series of real and predicted wind power are supposed to have 10-min resolution with 10-min prediction horizon. However, for conditional PDF modeling, the evolution of estimated PDFs depends on the prediction horizon and resolution. The prediction horizon and resolution of the time series under study should be consistent with the power system generation and reserve dispatch because, for example, the PDF of wind power looking 10-min ahead is different from 1-hour ahead. If the dispatch system is run every hour with 10-min resolution, the prediction horizon of historical data can be one hour with the same resolution. It is worth noting that, without loss of generality, the proposed conditional PDF modeling can be done for various prediction horizons and resolutions.

$$f(x|\widehat{\omega}_t) = f(x|\boldsymbol{p}, \boldsymbol{v}, \boldsymbol{\zeta}) = \hat{f}(x; \boldsymbol{p}, \boldsymbol{v}, \boldsymbol{\zeta}) \tag{3.11}$$

$$\mu_n = \frac{1}{N_s}\sum_{i=1}^{N_s} X_i^n \quad , \qquad n = 0,1,2,\dots,N \tag{3.12}$$

$$\mathbf{P} = \{\boldsymbol{p_k}\}_{k=1}^{N_{PB}} \ , \boldsymbol{\Upsilon} = \{\boldsymbol{v_k}\}_{k=1}^{N_{PB}} \ , \mathbf{Z} = \{\boldsymbol{\zeta_k}\}_{k=1}^{N_{PB}} \tag{3.13}$$

3.2.6 Comparison with Other Distribution Models

To show the superiority of the proposed wind power distribution modeling, a quantitative analysis is provided to compare with TVD, GMM, VMD, and GGMM [30-32], [71], [109], [111], [112] using three wind power datasets. **Dataset 1**- Canada's Alberta Electric System Operator (AESO) dataset with 1463 MW capacity (WF1); **Dataset 2**- Canada's Centennial wind farm in Saskatchewan province with 150 MW capacity (WF2); and **Dataset 3**- Spain's Sotavento wind farm with 17 MW capacity (WF3) [120], [121].

This study uses the $RMSE$ index to calculate the closeness of approximate CDF models to actual distributions. Table 3.1 shows that the BKDR technique outperforms others by presenting the minimum value of $RMSE$ for the related PB of three different wind farms. However, BKDR and GMM show the same performance for $\{\mathrm{PB}_{19}\}$ of WF1, $\{\mathrm{PB}_3, \mathrm{PB}_{18}\}$ of WF2, and $\{\mathrm{PB}_9, \mathrm{PB}_{10}, \mathrm{PB}_{18}, \mathrm{PB}_{20}\}$ of WF3. Importantly, GMM cannot give a reasonable solution for PB_1 of WF2 and WF3. Two datasets WF2+WF3 and WF1+WF2+WF3 are used to demonstrate the performance of BKDR, GGMM, and VMD for aggregate wind power generation in Table 3.2. Because these approaches use approximately the same strategy for distribution modeling, they reflect the same performance for generation levels far from boundaries, and for boundary areas such as PB1, PB19, and PB20, BKDR makes a little difference. A qualitative comparison is shown in Figures 3.5 and 3.6 to have a visual sense of BKDR performance. Figure 3.5 shows the CDF of WF1, WF2, and WF3 obtained by TVD and BKDR along with the actual distribution for different levels of generation. Because the CDFs acquired by GMM are very close to those obtained by BKDR, they are not shown in Figure 3.5 for the sake of clarity. Figure 3.5 shows that the BKDR can follow the variations of actual distributions quite well. Figure 3.6 indicates the actual histogram and fitted distribution models for datasets WF2+WF3, WF1+WF2+WF3, and WF1. It is obvious that single-mode Gaussian, beta, and TVD models cannot compete with GMM, VMD, GGMM, and BKDR while the performance of the last three models is quite close to each other. Table 3.3 shows the computation time for different distribution models. Compared with other models, the proposed BKDR model has a reasonable computation time for real applications in wind power generation modeling.

Table 3.1 RMSE index for different levels of wind power generation.

PB_i	RMSE(%)								
	WF1			WF2			WF3		
	TVD	**GMM**	**BKDR**	**TVD**	**GMM**	**BKDR**	**TVD**	**GMM**	**BKDR**
PB_1	3.50	2.50	**1.67**	2.80	---	**1.06**	1.85	---	**0.83**
PB_2	2.27	0.61	**0.54**	0.81	0.71	**0.63**	0.33	0.28	**0.27**
PB_3	1.58	0.65	**0.63**	0.62	**0.57**	**0.57**	0.28	0.27	**0.25**
PB_9	1.22	0.78	**0.70**	0.84	0.66	**0.32**	0.67	**0.26**	**0.26**
PB_{10}	1.15	0.52	**0.50**	1.05	0.74	**0.70**	0.83	**0.57**	**0.57**
PB_{14}	0.60	0.40	**0.34**	1.21	1.15	**1.14**	0.52	0.50	**0.49**
PB_{15}	0.98	0.68	**0.53**	1.10	0.88	**0.78**	0.60	0.45	**0.42**
PB_{18}	1.15	0.57	**0.46**	1.84	**0.90**	**0.90**	1.17	**0.65**	**0.65**
PB_{19}	1.91	**0.44**	**0.44**	1.56	0.83	**0.69**	1.70	0.79	**0.78**
PB_{20}	1.31	1.00	**0.98**	1.85	0.81	**0.72**	3.87	**1.00**	**1.00**

Table 3.2 RMSE index for different levels of aggregate wind power generation.

PB_i	RMSE(%)					
	WF2+WF3			WF1+WF2+WF3		
	VMD	**GGMM**	**BKDR**	**VMD**	**GGMM**	**BKDR**
PB_1	0.72	0.70	**0.67**	0.60	**0.58**	**0.58**
PB_3	**0.25**	**0.25**	**0.25**	**0.23**	**0.23**	**0.23**
PB_{15}	**0.33**	**0.33**	0.34	**0.33**	**0.33**	**0.33**
PB_{19}	0.42	**0.40**	**0.40**	0.40	0.39	**0.38**
PB_{20}	0.92	0.92	**0.90**	0.91	0.90	**0.87**

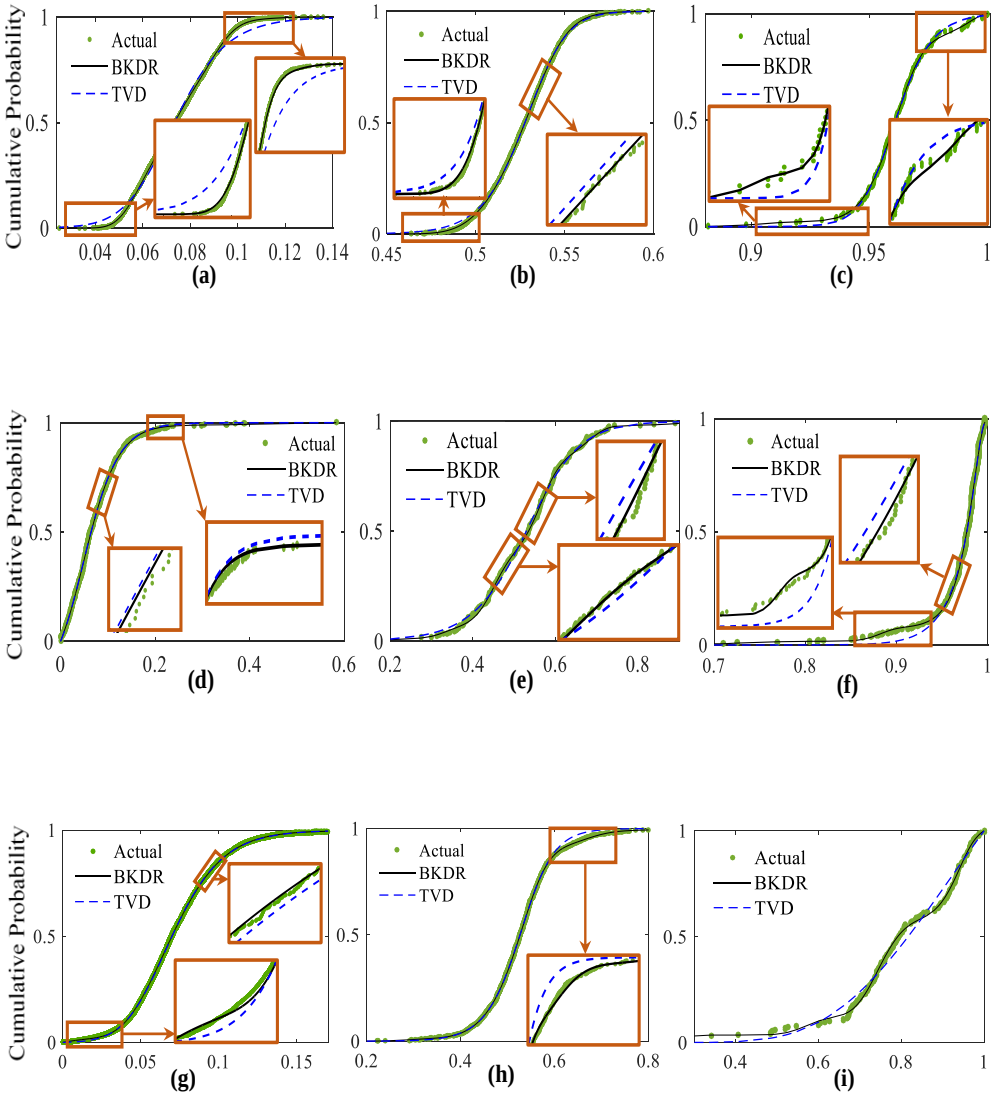


Figure 3.5 Actual, BKDR, and TVD fits of the CDF for WF1 (a)-(c), WF2 (d)-(f), and WF3 (g)-(i) for wind power forecast values of 0.05 (a-d-g), 0.5 (b-e-h), and 0.98 (c-f-i) p.u., respectively.

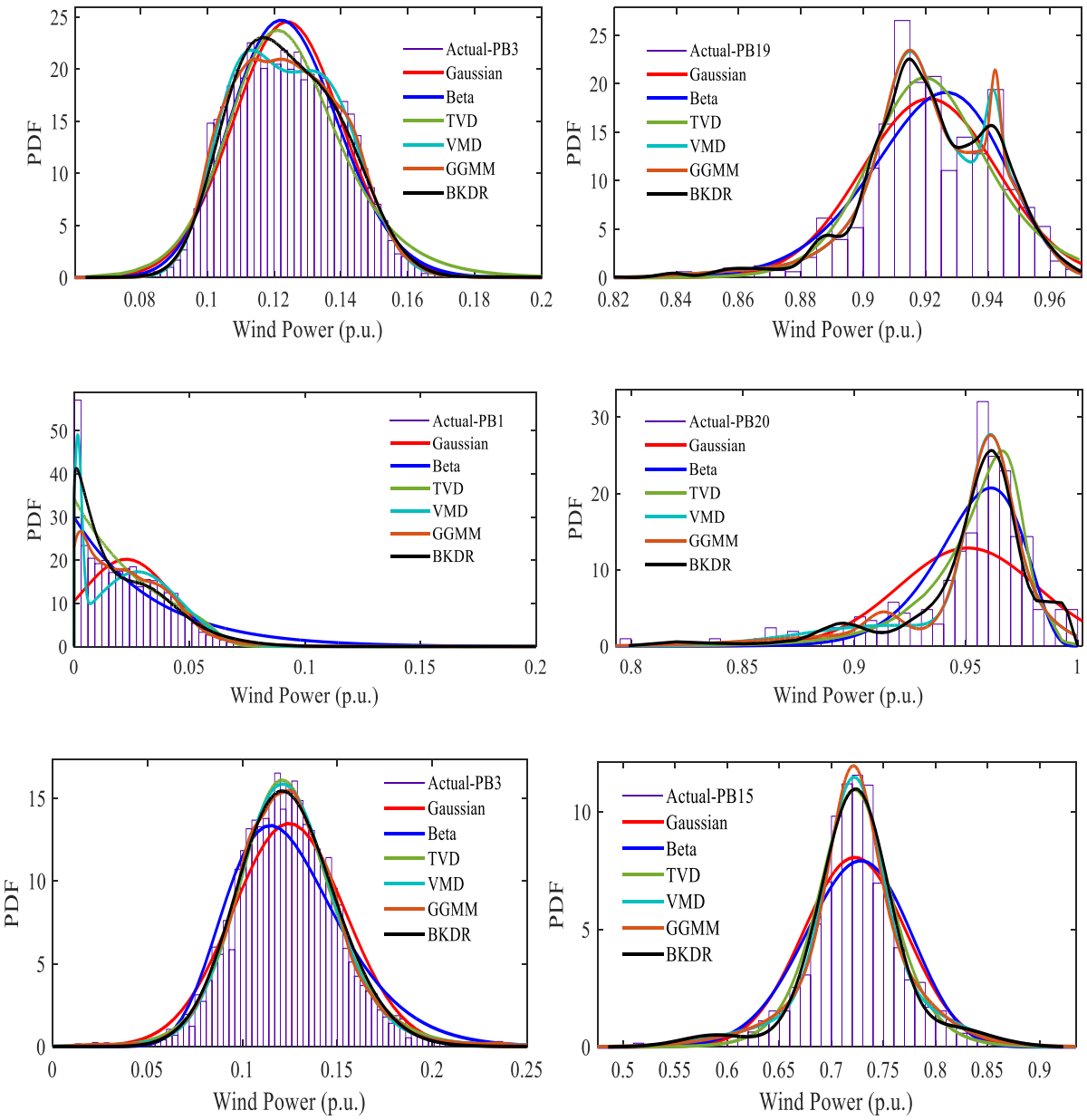


Figure 3.6 Comparison of PDFs for WF2+WF3 dataset (upper figures), W1+WF2+WF3 dataset (middle figures), and WF1 dataset (lower figures).

Table 3.3 Computation time for construction of different distribution models.

Distribution Model	Gaussian	Beta	TVD	VMD	GGMM	BKDR
Average Time (sec.)	24.50	26.20	17.30	22.50	21.70	21.30

3.3 Proposed Chance-Constrained ED Problem

A look-ahead ED problem, which affects the online operation of power systems, is a submodule of multiple time-scale coordinated active power control systems. It is performed once per hour to determine the active power output of all generation units over the forthcoming four hours with a time resolution of 10 or 15 minutes [29], [30], [71], [82], [109]. With high penetration of wind generation, a cost function is minimized while satisfying several constraints in a look-ahead CCED problem as an efficient stochastic optimization methodology that allows the uncertainties to be handled by mitigating wind power underestimation and overestimation impacts. In [29], [30], [71], [109] efficient formulations for generation and reserve scheduling are proposed in look-ahead and real-time CCED models. In the next sections, these formulations are expanded and explained to reach the proposed NCF and MILP-based CCED model.

3.3.1 Cost Function and Constraints of the CCED Problem

To make decisions under uncertain situations, a classical two-stage stochastic problem is recommended where some of the decisions must be made before the uncertainty is realized (first stage), and then the recourse decisions can be made after the realizations (second stage) [122]. The classical two-stage linear stochastic problems can be formulated as $\min_{x \in X} A^T x + \mathbb{E}[Z(x,\xi)]$ where $x \in \mathbb{R}^n$ is the first-stage decision vector, ξ is the data of the second stage, and $Z(x,\xi)$ is the optimal solution of the second stage defined as $\min_{y} z^T y$ subject to $Bx + Cy \leq D$ where $y \in \mathbb{R}^m$ is the second-stage decision vector and $\xi = (z, B, C, D)$. In this kind of formulation, a "here-and-now" decision should be made for x at the first stage before knowing the realization of uncertain data ξ. At the second stage, after a realization of ξ, another optimization problem should be solved to optimize the decision-making procedure. Therefore, the solution of the second-stage problem is viewed as a recourse action where the term Cy mitigates the inconsistency of $Bx \leq D$, and $z^T y$ would be the cost of this recourse action. If the random variable ξ has a finite support, a linear programming equivalent to the two-stage model can be used as $\min_{x} A^T x + \int Z(x,\xi) f(\xi) d\xi$ which after solving this problem, an optimal solution which can cover all possible scenarios of ξ is found. In this chapter, similarly, the cost function of the stochastic look-ahead ED problem is shown in (3.14)-(3.21). The costs related to the uncertainty

of wind power generation originate from the overestimation and underestimation of wind power. Figure 3.7 shows four main areas of wind power generation uncertainty as well as the actual and fitted distributions for the forecast value $\hat{\omega}_t$ =0.1 p.u. The main goal of Figure 3.7 is to show the main areas of wind power distribution model and their relations with the scheduled value. To mitigate the impact of wind power overestimation, one strategy is to take upward reserve. If the upward reserve does not suffice, a load shedding strategy will be added. Similarly, wind power underestimation is mainly alleviated using downward reserve, and a wind curtailment strategy is added when downward reserve is not adequate. Therefore, wind power overestimation cost equals the expected upward reserve cost EC_t^{UP} with penalty factor γ^{UP}plus the expected load shedding cost EC_t^{LS} with penalty factor γ^{LS}shown in (3.18) and (3.19), respectively. Also, wind power underestimation cost equals the expected downward reserve cost EC_t^{DN} with penalty factor γ^{DN} plus the expected wind curtailment cost EC_t^{WC} with penalty factor γ^{WC} expressed by (3.20) and (3.21), respectively. Although not all electricity markets apply penalty factors γ^{UP} and γ^{DN} for deviation of wind power generation from the scheduled values, in this chapter these penalty factors are considered for generalization of the proposed model [29], [30], [71], [109].

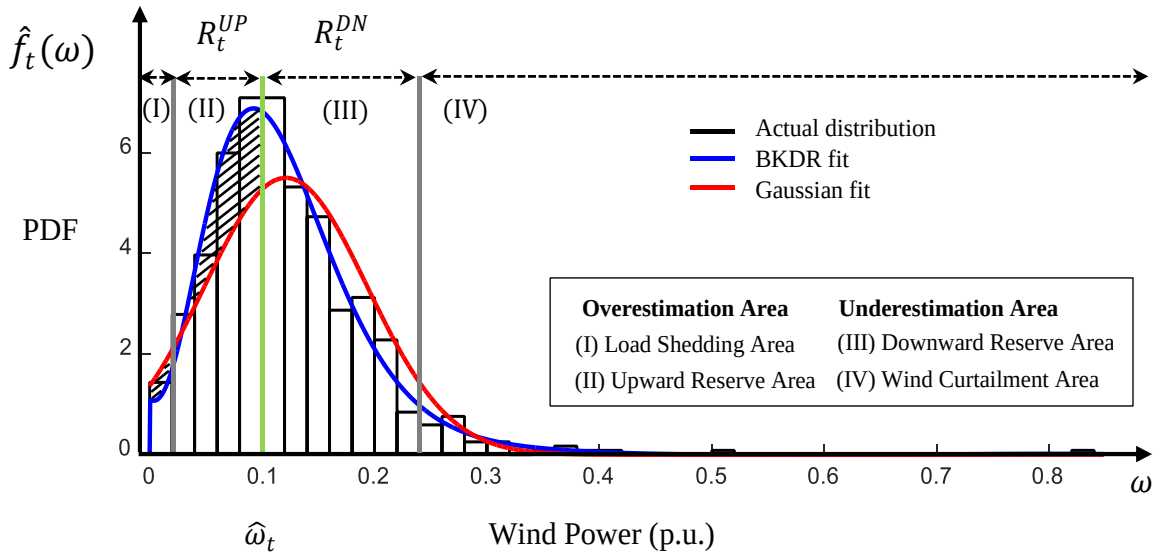


Figure 3.7 Illustration of wind power uncertainty and overestimation/underestimation areas for forecast value $\hat{\omega}_t = 0.1\ (\text{p.u.})$ for PB_3 of WF2.

$$Minimize \ \sum_{t=1}^{T} \Delta t.\left(\sum_{i=1}^{NG} C_{t,i}^{g} + \sum_{i=1}^{NG} C_{t,i}^{r} + C_t^{dw} + EC_t^{UP} + EC_t^{LS} + EC_t^{DN} + EC_t^{WC}\right) \quad (3.14)$$

$$C_{t,i}^{g} = a_i\left(P_{t,i}{}^2\right) + b_i\left(P_{t,i}\right) + c_i \ \ \forall t, i \quad (3.15)$$

$$C_{t,i}^{r} = c_i^{UP}\left(r_{t,i}^{UP}\right) + c_i^{DN}\left(r_{t,i}^{DN}\right), \forall t, i \quad (3.16)$$

$$C_t^{dw} = d_t(\omega_t) \ \ , \ \forall t \quad (3.17)$$

$$EC_t^{UP} = \gamma^{UP}\left(\int_{\omega_t - R_t^{UP}}^{\omega_t} (\omega_t - \omega)\hat{f}_t(\omega)d\omega + \int_{0}^{\omega_t - R_t^{UP}} R_t^{UP}\hat{f}_t(\omega)d\omega\right) \ , \ \ \forall t \quad (3.18)$$

$$EC_t^{LS} = \gamma^{LS}\left(\int_{0}^{\omega_t - R_t^{UP}} \left((\omega_t - R_t^{UP}) - \omega\right)\hat{f}_t(\omega)d\omega\right) \ , \forall t \quad (3.19)$$

$$EC_t^{DN} = \gamma^{DN}\left(\int_{\omega_t}^{\omega_t + R_t^{DN}} (\omega - \omega_t)\hat{f}_t(\omega)d\omega + \int_{\omega_t + R_t^{DN}}^{\omega_r} R_t^{DN}\hat{f}_t(\omega)d\omega\right), \ \forall t \quad (3.20)$$

$$EC_t^{WC} = \gamma^{WC}\left(\int_{\omega_t + R_t^{DN}}^{\omega_r} \left(\omega - (\omega_t + R_t^{DN})\right)\hat{f}_t(\omega)d\omega\right) \ , \forall t \quad (3.21)$$

s.t.

$$\sum_{i=1}^{NG} P_{t,i} + \omega_t = L_t^b \qquad \forall t \quad (3.22)$$

$$\sum_{j=1}^{NW} w_{t,j} = \omega_t \qquad \forall t \quad (3.23)$$

$$0 \le w_{t,j} \le w_{r,j} \qquad \forall t, j \quad (3.24)$$

$$\begin{cases} \sum_{i=1}^{NG} r_{t,i}^{UP} = R_t^{UP} \\ \sum_{i=1}^{NG} r_{t,i}^{DN} = R_t^{DN} \end{cases} \qquad \forall t \quad (3.25)$$

$$\begin{cases} 0 \le R_t^{UP} \le \omega_t \\ 0 \le R_t^{DN} \le \ \omega_r - \omega_t \end{cases} \qquad \forall t \quad (3.26)$$

$$\begin{cases} \mathbb{P}\mathbb{r}\{R_t^{UP} \ge \omega_t - \omega\} \ge CL^{UP} \\ \mathbb{P}\mathbb{r}\{R_t^{DN} \ge \omega - \omega_t\} \ge CL^{DN} \end{cases} \qquad \forall t \quad (3.27)$$

$$\underline{P_i} + r_{t,i}^{DN} \le P_{t,i} \le \overline{P}_i - r_{t,i}^{UP} \qquad \forall t, i \quad (3.28)$$

$$\begin{cases} 0 \le r_{t,i}^{UP} \le \overline{r}_i^{UP} \\ 0 \le r_{t,i}^{DN} \le \overline{r}_i^{DN} \end{cases} \qquad \forall t, i \quad (3.29)$$

$$\begin{cases} P_{t,i} - P_{t-1,i} \le \overline{\Delta P}_i^{UP} \\ P_{t-1,i} - P_{t,i} \le \overline{\Delta P}_i^{DN} \end{cases} \qquad \forall t, i \quad (3.30)$$

where (3.22) reflects system's power balance constraints, (3.23) shows that the summation of WFs' scheduled wind power equals the scheduled aggregate wind power, and (3.24) defines the range of scheduled WFs' generation. The constraints in (3.25) show the equality of total CPPs reserves with required reserves for aggregate wind power, and (3.26)-(3.27) denote the range of

required reserve at each hour. The uncertain random variable of the system shown by ω reflects the possible values of actual wind power generation. The operator $\mathbb{P}\mathrm{r}\{\cdot\}$ in (3.27) is a probability measure and indicates the chance constraints applied to the upward and downward reserves considered for aggregate wind power. The generation and reserve constraints as well as ramp-rate constraints are represented by (3.28)-(3.30). Note that the chance-constrained limits in (3.27) can be rewritten as (3.31).

$$\begin{cases} R_t^{UP} \geq \omega_t - \hat{F}_t^{-1}(1 - CL^{UP}) \\ R_t^{DN} \geq \hat{F}_t^{-1}(CL^{DN}) - \omega_t \end{cases} \qquad \forall t \tag{3.31}$$

3.4 Challenges and Solutions for CCED Problem

3.4.1 Existing Challenges

Regarding the implementation of the CCED problem, there are two main challenges as follows. **1)** The distribution model of wind power, shown by $\hat{f}_t(\omega)$, should reflect the main features of the actual distribution such as mode, long tail, high skewness, etc. while avoiding boundary effects. Figure 3.7 compares the appropriate BKDR fit (proposed model) and unsuited Gaussian fit and highlights the existing differences in overestimation area with two hatched areas. Considering equations (3.18), (3.19), the difference between the actual and fitted distributions leads to the miscalculation of the upward reserve and the load shedding costs, as will be shown in Section 3.5. Likewise, considering (3.20), (3.21), the existing mismatch in the underestimation area of the probability distribution causes a misjudgment about the downward reserve and wind curtailment costs. As a result, the solutions of the CCED problem might not be optimal if fed by the distribution models such as Gaussian, beta, stable, versatile, etc., that suffer from either boundary effects or low flexibility in terms of showing the above main features [29], [30]-[34], [71], [109]. However, the boundary effects challenge is alleviated using TVD in [30].

2) The most important challenge is that the CCF in (3.14) contains the non-linear functions (3.18)-(3.21) defined by integrals, which makes tackling such an optimization problem difficult. This mainly originates from the dependency of the integral operators' boundaries in (3.18), (3.19) on the pair decision variables $\{R_t^{UP}, \omega_t\}$ and dependency of integral operators' boundaries in (3.20), (3.21) on $\{R_t^{DN}, \omega_t\}$, such that there is not a closed form for the CCED cost function versus the

decision variables. Consequently, the partial derivatives of the CCF with respect to these variables (as required for interior-point and SLP optimization techniques) might not be correctly derived. Nevertheless, based on partial derivative formulations in [29], [30], [71], [109], the boundaries of integrals are ignored, and approximate partial derivatives are proposed for integral functions that might lead to misleading results.

3.4.2 Potential Solutions

The first challenge is remediated by the proposed BKDR technique in Section 3.2. The proposed BKDR technique can well estimate the PDF of wind power generation and avoid boundary effect. To address the second challenge, an NCF is explicitly proposed for the CCED problem; specifically, it can be formulated like an MILP model and efficiently solved using standard MILP solvers. To this end, EC_t^{UP} and EC_t^{LS} in (3.18), (3.19) are explicitly reformulated with respect to the decision variables R_t^{UP} and ω_t, and EC_t^{DN} and EC_t^{WC} in (3.20), (3.21) are rewritten versus R_t^{DN} and ω_t. These reformulated functions are expressed using polynomial functions with degree q $(4 \leq q \leq 8)$ to accurately model wind power overestimation and underestimation costs. The polynomials are single-variable functions with the variables R_t^{UP} or R_t^{DN}, and just the coefficients depend on variable ω_t. The polynomial fitting procedure is done for every value of wind power over the range [0,1] with a certain resolution. For this procedure, first, suppose ω_t takes on a constant value over PB_k where $k = 1 \dots N_{\mathrm{PB}}$, and the integral operators in (3.18)-(3.21) are expanded independently while the main variables R_t^{UP} and R_t^{DN} take on the valid values over the specific ranges defined by (3.26) and (3.31). By repeating this procedure for each value of ω_t, the integral-based functions can be represented by few lookup tables. Then, they can be precisely expressed in the form of an octic function with decision variables R_t^{UP} or R_t^{DN}, as shown in (3.32)-(3.35) where n=8. Also, the cost coefficients A_q, A'_q, B_q, and B'_q of the octic functions only depend on ω_t, and the hat sign shows that the cost functions are approximate functions. Note, for each value of wind power inside each PB, a separate octic function is fitted. As an example, the fitted cost functions (3.32)-(3.35) are shown in Figure 3.8 for all wind power values in PB_{10} of WF2 dataset. It shows the nonlinearity and convexity of each function for a certain level of generation; however, it is not guaranteed that the summation of these functions for different generation levels to be a convex function. The fitting error of approximate underestimation and overestimation cost functions are shown in Figure 3.9 for WF1 to WF3. The fitting error for wind farms with high

generation capacity is more than those with low generation capacity because the non-linearity of the cost functions in (3.18)-(3.21) increases proportional to wind farm generation capacity.

$$\widehat{EC}_t^{UP} = \sum_{q=0}^{n} A_q \,.\, (R_t^{UP})^q \tag{3.32}$$

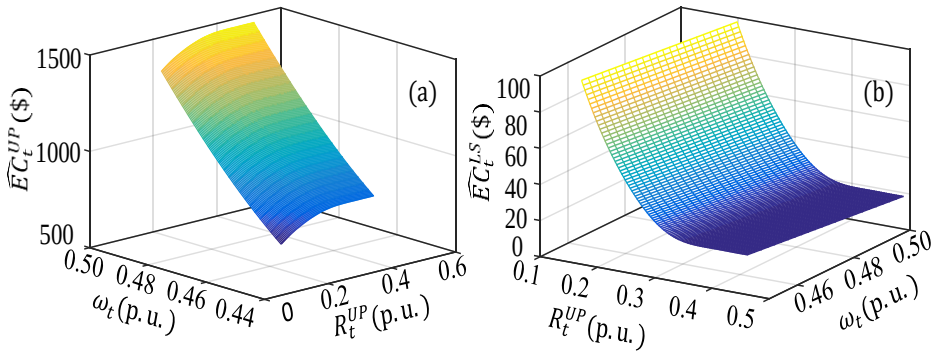
$$\widehat{EC}_t^{LS} = \sum_{q=0}^{n} A'_q \,.\, (R_t^{UP})^q \tag{3.33}$$

$$\widehat{EC}_t^{DN} = \sum_{q=0}^{n} B_q \,.\, (R_t^{DN})^q \tag{3.34}$$

$$\widehat{EC}_t^{WC} = \sum_{q=0}^{n} B'_q \,.\, (R_t^{DN})^q \tag{3.35}$$

$$NCF = \sum_{t=1}^{T} \Delta t. \left(\sum_{i=1}^{NG} C_{t,i}^{g} + \sum_{i=1}^{NG} C_{t,i}^{r} + C_t^{dw} + \widehat{EC}_t^{UP} + \widehat{EC}_t^{LS} + \widehat{EC}_t^{DN} + \widehat{EC}_t^{WC}\right) \tag{3.36}$$

The main process of obtaining the proposed NCF in (3.36) is based on the estimation of cost coefficients A_q, A'_q, B_q, and B'_q as shown by Algorithm 3 in Figure 3.10. One of the main building blocks of Algorithm 3 is Algorithm 2 which estimates the wind power PDF of PBs. The estimated cost coefficients for total wind power generation ω_t with a certain resolution over the range [0,1] are stored in a database. In the proposed MILP-based CCED model, the corresponding cost coefficients for each wind power forecast value are chosen from the database, and estimated functions (3.32)-(3.35) can be explicitly formed versus decision variables R_t^{UP} and R_t^{DN}. The structure of the proposed MILP-based CCED model is depicted in Figure 3.11 where the utilized linearization technique is presented in Section 3.3.3 and the detailed solution methodology is presented in Section 3.3.4.



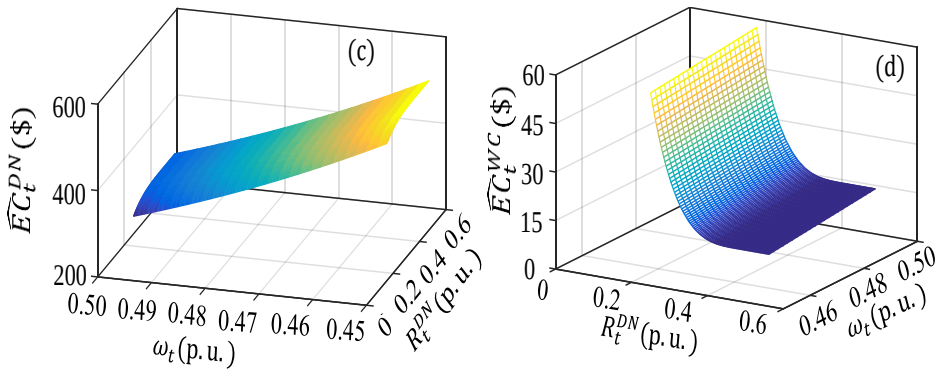


Figure 3.8 Illustration of wind power overestimation cost variations ((a) upward reserve cost, (b) load shedding cost) and underestimation cost variations ((c) downward reserve cost, (d) wind curtailment cost) for PB_{10} of WF2.

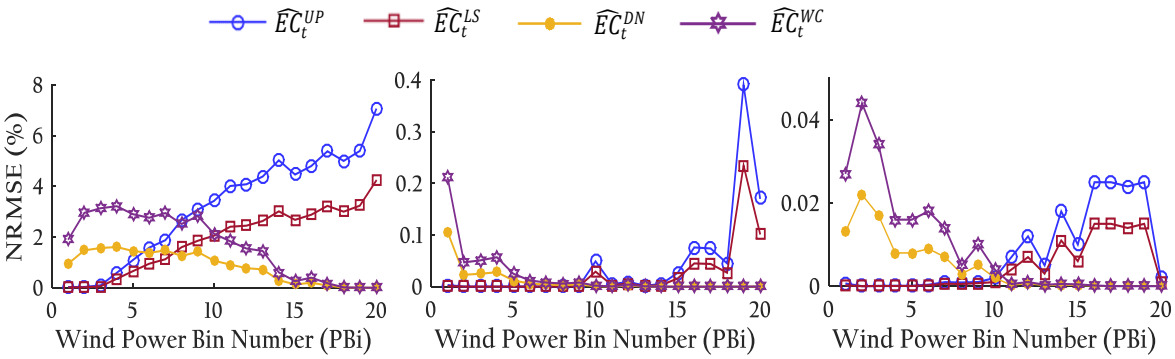


Figure 3.9 Fitting error of four parts of the proposed NCF for three different wind farms WF1, WF2, and WF3 from left to right, respectively.

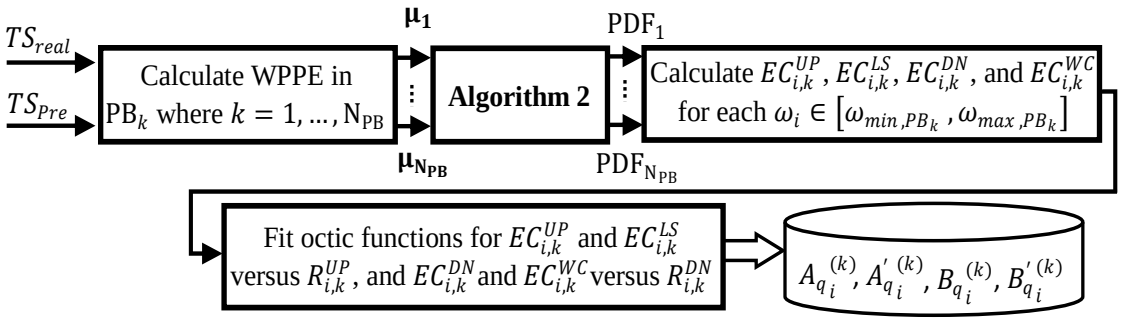


Figure 3.10 Structure of Algorithm 3 for wind power cost coefficients estimation.

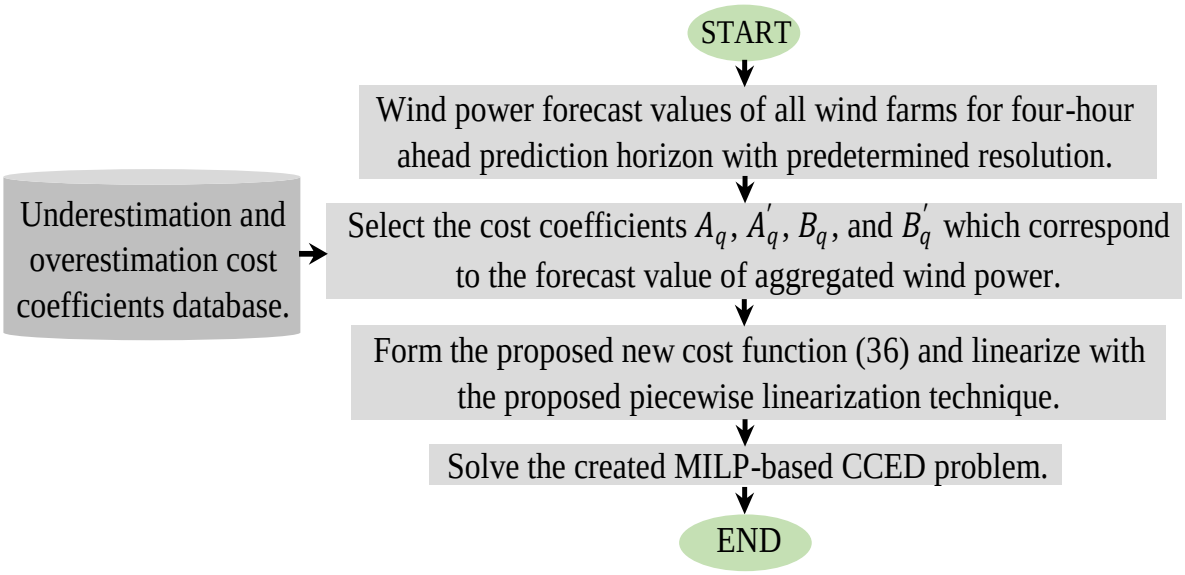


Figure 3.11 General structure of the solution algorithm for the proposed MILP-based CCED model.

3.4.3 Piecewise Linearization of the Proposed NCF

The CCF is a non-linear function without a closed form that makes achieving the global optimal solution difficult [30], [71], [109]. Using the proposed reformulation process, an NCF is reproduced to lead to a solution closer to the global solution. To overcome the non-linearity of the proposed NCF, a highly accurate piecewise linearization technique is used [72]. By doing so, the CCED problem is converted to a tractable MILP-based model that can be solved using off-the-shelf mathematical programming solvers such as CPLEX and Gurobi. The non-linearity of the cost function arises from the quadratic function in (3.15) as well as the proposed octic functions (3.32)-(3.35). Let us express each of these non-linear functions as $F(z)$. Using a highly accurate piecewise linearization technique, $F(z)$ can be linearized as (3.37)-(3.41), where $\overline{z}$ and $\underline{z}$ represent the upper and lower limits of the variable z, respectively.

$$F(z) = \sum_{\lambda=1}^{\Lambda}(m_\lambda \delta_\lambda + n_\lambda \Delta_\lambda) \tag{3.37}$$

$$z = \sum_{\lambda=1}^{\Lambda} \delta_\lambda \tag{3.38}$$

$$\psi_{\lambda-1}\Delta_\lambda \leq \delta_\lambda \leq \psi_\lambda \Delta_\lambda \ , \quad \lambda = 1, \dots, \Lambda \tag{3.39}$$

$$\sum_{\lambda=1}^{\Lambda} \Delta_\lambda \leq 1 \tag{3.40}$$

$$\delta_\lambda \geq 0 \ , \ \Delta_\lambda \in \{0,1\} , \quad \lambda = 1, \dots, \Lambda \tag{3.41}$$

where δ_λ and Δ_λ are required to obtain the piecewise linear representation of $F(z)$, and constant parameters ψ_λ, m_λ, and n_λ can be obtained from (3.42)-(3.44).

$$\psi_\lambda = \underline{z} + \lambda(1/\Lambda)\left(\overline{z} - \underline{z}\right) \quad , \lambda = 1, \dots, \Lambda \tag{3.42}$$

$$m_\lambda = [F(\psi_\lambda) - F(\psi_{\lambda-1})]/[\psi_\lambda - \psi_{\lambda-1}] \quad , \lambda = 1, \dots, \Lambda \tag{3.43}$$

$$n_\lambda = F(\psi_\lambda) - m_\lambda \psi_\lambda \quad , \lambda = 1, \dots, \Lambda \tag{3.44}$$

To shed light on the proposed linearization technique, the piecewise linear approximation of a typical non-linear function is illustrated in Figure 3.12. The feasible range of the variable z is partitioned into Λ segments. Then, a line with a slope of m_λ and intercept of n_λ is considered corresponding to each segment λ. Finally, using the binary variables denoted by Δ_λ, only one of the lines is chosen to represent the non-linear function $F(z)$. Note that the parameter Λ determines the number of additional variables and constraints required to linearize $F(z)$. Therefore, the approximation error will obviously decrease as this parameter increases. Using the above-described linearization technique, the proposed CCED problem is now converted to an MILP-based model, which guarantees the solution optimality and computational tractability.

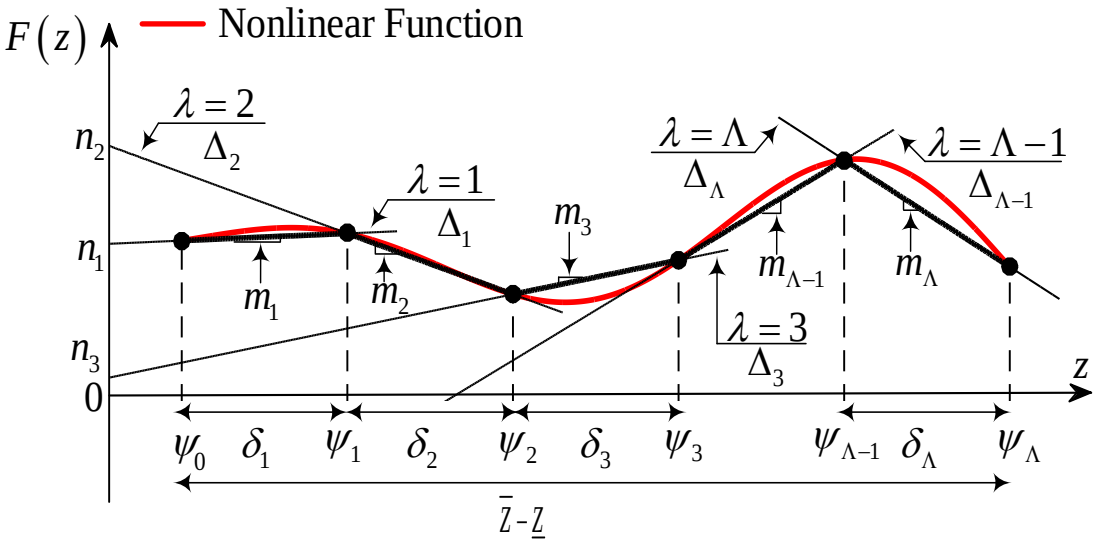


Figure 3.12 Piecewise linear approximation of a typical non-linear function.

3.4.4 Solution Methodology

The detailed steps of the proposed solution methodology are as follows.

Step (1) **Initialization:** Set decision vector $\boldsymbol{x}$ on zero where $\Lambda = 10$, $CL^{UP} = CL^{DN} = 0.95$, UB $= +\infty$, LB $= -\infty$, MILP gap tolerance $\varepsilon = 1\text{e} - 3$, and $\boldsymbol{x} = [P_{t,i}, \omega_{t,j}, r_{t,i}^{UP}, r_{t,i}^{DN}, R_t^{UP}, R_t^{DN}, \delta_\lambda, \Delta_\lambda]$.

Step (2) **Receiving system data:** NG, NW, d_t, L_t^b, c_i^{UP}, c_i^{DN}, $\underline{P}_i$, $\overline{P}_i$, $\overline{\Delta P}_i^{UP}$, $\overline{\Delta P}_i^{DN}$, $\overline{r}_i^{UP}$, $\overline{r}_i^{DN}$, γ^{UP}, γ^{LS}, γ^{DN}, and γ^{WC}.

Step (3) **Modeling NCF:** Obtain A_q, A'_q, B_q, and B'_q of aggregate wind generation based on the wind power forecast values and CL^{UP} and CL^{DN} (Algorithm 3), and obtain $\widehat{EC}_t^{UP}$, $\widehat{EC}_t^{LS}$, $\widehat{EC}_t^{DN}$, and $\widehat{EC}_t^{WC}$ to create NCF as shown by (3.36).

Step (4) **Linearizing NCF:** linearize $C_{t,i}^g$, $\widehat{EC}_t^{UP}$, $\widehat{EC}_t^{LS}$, $\widehat{EC}_t^{DN}$, and $\widehat{EC}_t^{WC}$ simultaneously using piecewise linearization in (3.37)-(3.44).

Step (5) **Solving linearized NCF:** Minimize the linearized NCF with CPLEX (i.e., *cplexmilp*) subject to constraints (3.22)-(3.31) where $\mathbf{NCF} = \sum_{t=1}^{T} \Delta t. \left(\sum_{i=1}^{NG} C_{t,i}^g + \sum_{i=1}^{NG} C_{t,i}^r + C_t^{dw} + \widehat{EC}_t^{UP} + \widehat{EC}_t^{LS} + \widehat{EC}_t^{DN} + \widehat{EC}_t^{WC}\right)$.

Step (6) **Updating cost coefficients of aggregate wind generation:** After a specific time period (e.g., one week), update the coefficients A_q, A'_q, B_q, and B'_q by receiving new wind power samples that lead to new wind power distribution $\hat{f}_t(\omega)$, then go to *Step (3)*.

3.5 Case Studies

3.5.1 Test System

To show the efficiency of the proposed approach, the widely used IEEE 118-bus test system, shown in Figure 3.13, with 54 CPPs is simulated in this study. The total base loads in the system

considered over a 4-hour scheduling period are 3.6, 3.9, 4.1, and 4.2 GW. The developed MILP-based CCED model is solved by CPLEX 12.6.1 using MATLAB R2015a on a Corei7-6700 CPU@3.40 GHz personal computer with 16GB RAM.

3.5.2 Experimental Datasets and Wind Power Penetration Scenarios

Three different wind power datasets, introduced in Section 3.2.5, are used to examine the proposed MILP-based CCED problem. Three penetration scenarios with different combinations of wind power datasets are considered in the system under study. **Case 1**- 150 MW wind power generation using WF2; **Case 2**- 167 MW wind generation using two different wind farms (WF2 and WF3); and **Case 3**- High penetration of wind power is assessed in this case with WF1, WF2, and WF3 with total capacity of 1630 MW. The correlation level among wind farms in Case 1 to Case 3 are not considered. To demonstrate the effect of correlation on total cost of the system, Case 4 and Case 5 are defined. **Case 4** is like Case 2 with three values $\rho = 0$, 0.50, and 0.95 as correlation coefficients. **Case 5** is like Case 3 considering two correlation matrices $\mathbb{R}_1$ and $\mathbb{R}_2$ shown below. Note that to construct wind power datasets with abovementioned correlation coefficients in Case 4 and Case 5, wind power time series WF1 to WF3 are checked and appropriate time lags are found in each time series to construct desired correlated time series.

$$\mathbb{R}_{\mathbf{1}} = \begin{bmatrix} 1.00 & 0.00 & 0.07 \\ 0.00 & 1.00 & 0.06 \\ 0.07 & 0.06 & 1.00 \end{bmatrix} \qquad \mathbb{R}_2 = \begin{bmatrix} 1.00 & 0.95 & 0.90 \\ 0.95 & 1.00 & 0.95 \\ 0.90 & 0.95 & 1.00 \end{bmatrix}$$

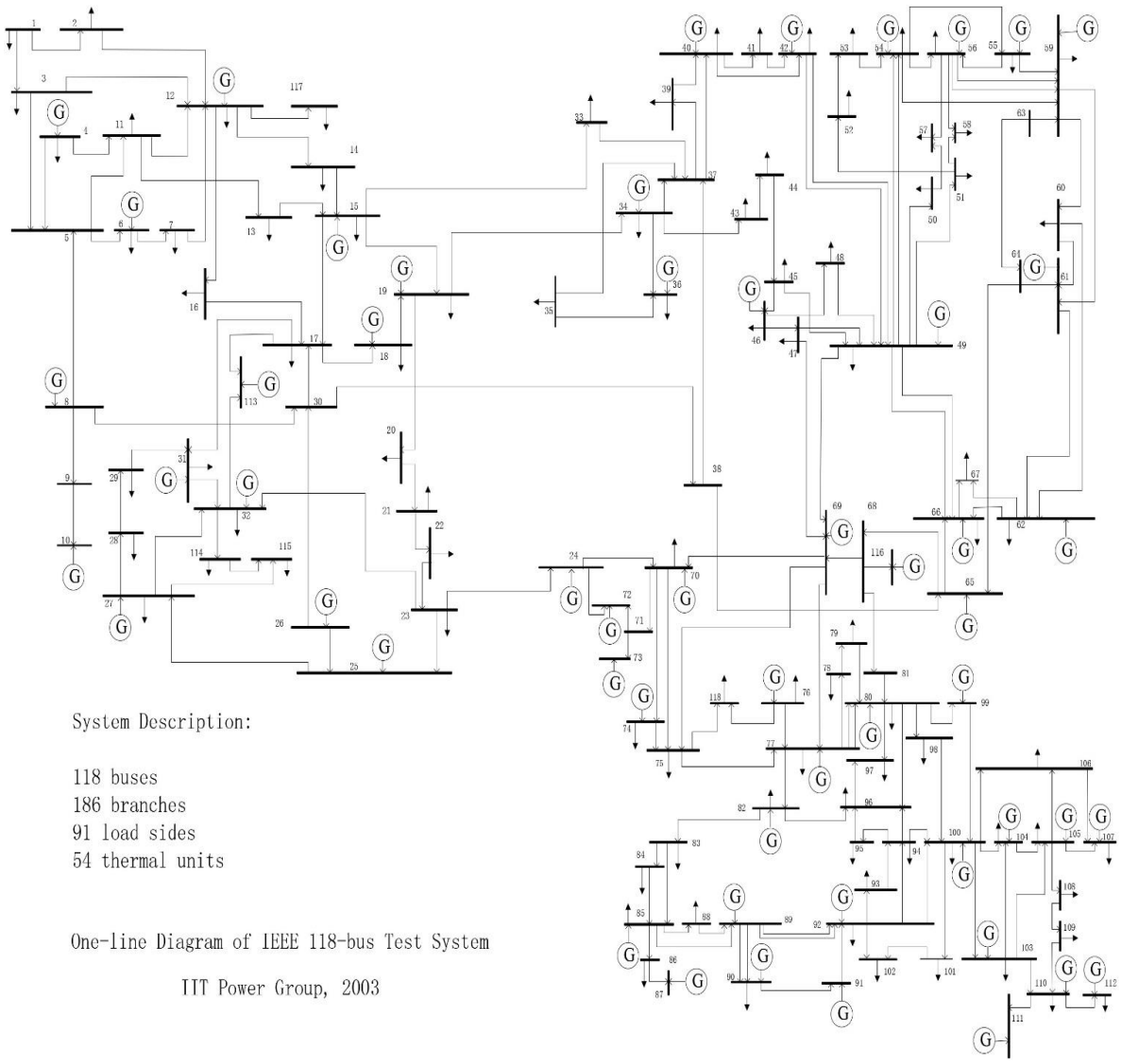


Figure 3.13 Single-line diagram of IEEE 118-bus test system.

3.5.3 Simulation Results of the Proposed MILP-based CCED Problem

This section examines the proposed model using the IEEE 118-bus test system with Cases 1 to 5 with the proposed model. The results are compared with conventional approaches in which either an SLP optimization technique or TVD model (as a parametric distribution model) is used. In this study, c_i^{UP} and c_i^{DN} are considered to be \$15/MWh, and γ^{UP}, γ^{LS}, γ^{DN}, and γ^{WC} are set at \$120, \$200, \$60, and \$120/MWh, respectively. The generation of CPPs and wind farms as well as the system's required reserve are determined for the next 4-hour scheduling period with 10-min resolution. It is assumed that wind farms' forecast values change from 0.06 to 0.94 p.u. Tables 3.4, 3.5, and 3.6 show the value of different parts of the CCED cost function in (3.14) for Cases 1, 2,

and 3, respectively. The results of the proposed MILP-based model (i.e., MILP-NCF-BKDR) are shown in the first numerical column. The second numerical column reflects the results of the approach, which includes NCF and BKDR while it is solved using SLP technique (i.e., SLP-NCF-BKDR). The third numerical column indicates the results of the CCED model comprised of the CCF in (3.14) and TVD model and solved by SLP technique (i.e., SLP-CCF-TVD). The comparisons show that if the proposed NCF is linearized with the piecewise linearization technique, compared to SLP-NCF-BKDR, the developed MILP-based model can decrease the total cost of the system by $985, $1,147, and $2,474 over just four hours for Cases 1 to 3, respectively. Compared with SLP-CCF-TVD, the reductions in total cost are, respectively, $3,150, $2,303, and $6,154 for Cases 1 to 3 over four hours. A significant reduction in the system's total cost will occur over a year if power system operators use the proposed model for two reasons. First, the proposed model, based on wind power forecast values, incorporates accurate wind power distribution models and efficiently identifies four main areas of wind power distribution (Figure 3.7) so that there is no miscalculation for upward reserve, load shedding, downward reserve, and wind curtailment costs. Second, the NCF is used that allows the system operators to convert the CCED problem to an MILP-based problem and efficiently solve it using powerful off-the-shelf solvers to obtain the more optimal solution. For example, unlike the upward reserve and wind curtailment costs, the costs of load shedding and downward reserve in Case 1 obtained by MILP-NCF-BKDR are more than those obtained by SLP-CCF-TVD. However, the total cost of wind power generation and its uncertainty for the proposed approach is $5,781 while for the other two approaches are $5,874 and $6,010, respectively. The same interpretation can be made for Cases 2 and 3. In Case 2, the total costs of wind power generation for MILP-NCF-BKDR, SLP-NCF-BKDR, and SLP-CCF-TVD are $3,313, $3,303, and $3,423, respectively.

These costs for Case 3 are higher at $22,453, $21,494, and $23,150, respectively. Although wind power related cost in the proposed model is more than other models in Cases 2 and 3, the generation and reserve costs of CPPs are less. Also, increasing wind power penetration, if appropriately managed, leads to less CPPs' generation cost. It is worth mentioning that if the chance constraints (3.27) are applied for each individual wind farm without considering the joint distribution of wind farms, the solution would be over conservative. In such a case, Case 2 and Case 3 result in $227,172 and $212,443 as total cost, respectively. The high accuracy of the piecewise linearization technique guarantees that the solution of the MILP-based model is the more

optimal solution of the initial non-linear CCED problem. Tables 3.7 and 3.8 examine the effect of correlation among wind farms on each part of total cost. The results of MILP-based model show that high levels of correlation necessitate more reserve cost, i.e., C^r and more expected cost of wind power misestimation, i.e., ΣEC, and the other costs, i.e., C^g and C^{dw} are almost unchanged. However, with the SLP-based model, such statement cannot be concluded because SLP-based model might not reach the optimal solution for different correlation levels. The computation time for three utilized CCED model is indicated in Table 3.9. Note that for the linearization purpose in the proposed MILP-based model, the number of segments and the MILP gap equal 15 and 1%, respectively.

Table 3.10 reflects the effect of the number of linearization segments and MILP gap on the convergence of the proposed MILP-based algorithm. Also, Figure 3.14 shows that, by increasing the number of segments Λ for linearization purposes, the accuracy of the MILP-based model increases, but it reaches a plateau after a certain value. Figure 3.15 shows the total cost of the system versus total base load changes and different values for confidence level over the range [0.90, 0.99]. The total base load varies with predefined base load coefficients from 0.75 to 1 for Case 1 and from 0.9 to 1.2 for Case 2 and Case 3. The red arrows in Figure 3.15 show the results presented in Tables 3.4 to 3.6, which are for CL=95%, and the base load coefficient equals one. Also, Figure 3.16 shows the performance of MILP-based and SLP-based CCED models with different values of CL and correlation level. The red arrows in Figure 3.16 reflect the results in Table 3.7 and Table 3.8. The Figures 3.15 and 3.16 indicate the superiority of the proposed MILP-based model under all conditions. Figure 3.17 shows the scheduled and forecast values in Case 2 for the proposed MILP-based and SLP-based models.

Usually, with SLP-NCF-BKDR model, the scheduled wind power generation is lower than the corresponding forecast values for lower confidence levels (i.e., 90%) while for higher confidence levels (i.e., 99%) the scheduled wind power generation is greater than the forecast values. On the contrary, for all confidence levels in the proposed MILP-based model, the scheduled wind power generation is very close to the forecast values. The average values of RMSE for the proposed MILP-based model are 1.26, 0.87, and 0%, for 90, 95, and 99% confidence levels, respectively. In contrast, the average values of RMSE for the SLP-based model reach higher values of 2.97, 3.14, and 3.01%, respectively.

Table 3.4 Total cost of IEEE 118-bus test system in Case 1.

Cost ($) (CL=95%)	MILP-NCF-BKDR	SLP-NCF-BKDR	SLP-CCF-TVD
EC^{UP}	3,668	3,725	3,731
EC^{LS}	204	171	193
EC^{DN}	1,093	1,041	1,082
EC^{WC}	233	337	396
C^{dw}	583	600	608
C^{r}	2,673	2,614	3,510
C^{g}	218,571	219,522	220,655
Total	**227,025**	**228,010**	**230,175**

Table 3.5 Total cost of IEEE 118-bus test system in Case 2.

Cost ($) (CL=95%)	MILP-NCF-BKDR	SLP-NCF-BKDR	SLP-CCF-TVD
EC^{UP}	1,844	1,742	1,805
EC^{LS}	113	174	179
EC^{DN}	579	557	568
EC^{WC}	131	176	192
C^{dw}	646	654	679
C^{r}	1,219	1,353	1,397
C^{g}	218,079	219,102	220,094
Total	**222,611**	**223,758**	**224,914**

Table 3.6 Total cost of IEEE 118-bus test system in Case 3.

Cost ($) (CL=95%)	MILP-NCF-BKDR	SLP-NCF-BKDR	SLP-CCF-TVD
EC^{UP}	11,705	8,390	9,560
EC^{LS}	366	2,355	2,354
EC^{DN}	3,806	3,604	3,764
EC^{WC}	333	736	892
C^{dw}	6,243	6,409	6,580
C^{r}	5,970	7,243	7,435
C^{g}	178,000	180,160	181,992
Total	**206,423**	**208,897**	**212,577**

Table 3.7 Total cost of IEEE 118-bus test system in Case 4.

CL=95% Base Load Coefficient=1		$\rho = 0$	$\rho = 0.5$	$\rho = 0.95$
MILP-NCF-BKDR	C^g	218,099	218,091	218,098
	C^r	**1,218**	**1,284**	**1,319**
	C^{dw}	643	645	644
	ΣEC	**2,708**	**2,772**	**2,840**
	Total	**222,668**	**222,792**	**222,901**
SLP-NCF-BKDR	C^g	219,090	219,093	219,068
	C^r	1,285	1,415	1293
	C^{dw}	657	657	660
	ΣEC	2,718	2,775	2,888
	Total	**223,750**	**223,940**	**223,909**

Table 3.8 Total cost of IEEE 118-bus test system in Case 5.

CL=95% Base Load Coefficient=1		$\mathbb{R}_1$	$\mathbb{R}_2$
MILP-NCF-BKDR	C^g	177,966	177,966
	C^r	**11,635**	**12,631**
	C^{dw}	6,249	6249
	ΣEC	**26,390**	**27,771**
	Total	**222,240**	**224,617**
SLP-NCF-BKDR	C^g	180,164	180,703
	C^r	11,635	12,780
	C^{dw}	6,408	6,324
	ΣEC	27,034	26,974
	Total	**225,242**	**226,783**

Table 3.9 Computation time of CCED models for IEEE 118-bus test system.

CCED Model	MILP-NCF BKDR	SLP-NCF-BKDR	SLP-CCF-TVD
Average Time (sec.)	8.15	4.28	14.20

Table 3.10 Average convergence time in the proposed MILP-based approach.

MILP Gap / Λ	$\varepsilon = 1e\text{-}3$	$\varepsilon = 5e\text{-}3$	$\varepsilon = 1e\text{-}2$	$\varepsilon = 5e\text{-}2$
5	1.31s	1.27s	1.25s	1.26s
10	5.6s	5.4s	5.5s	5.6s

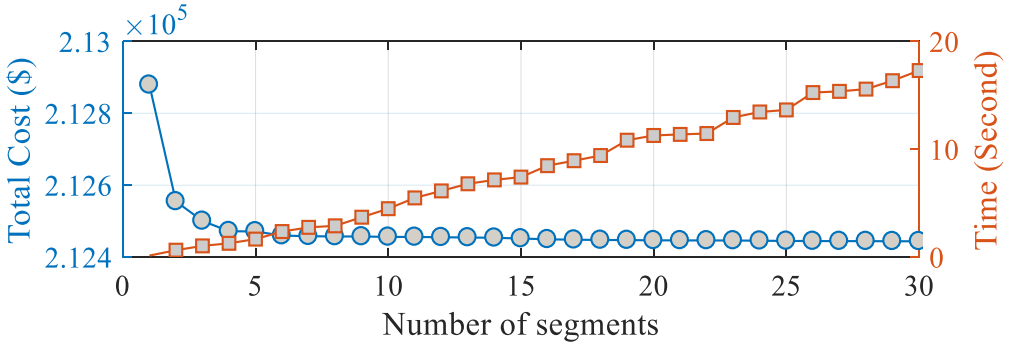


Figure 3.14 Total cost and solution time of the proposed MILP-based CCED problem versus the number of linearization segments.

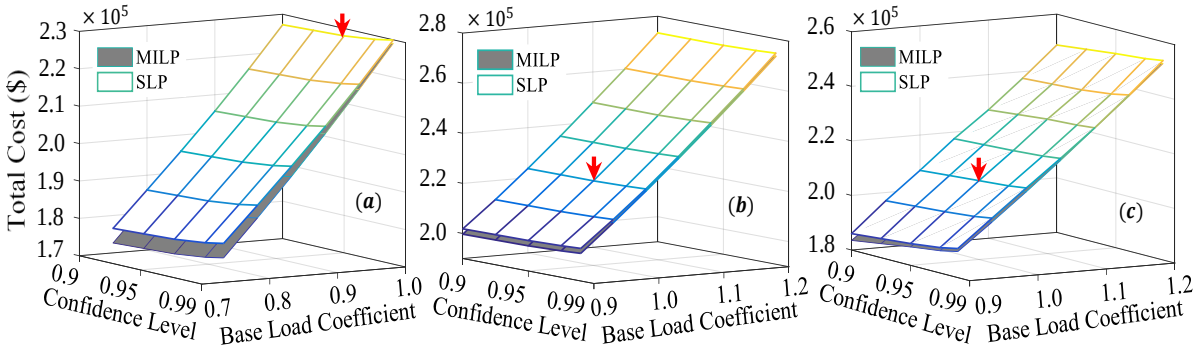


Figure 3.15 Illustration of total cost hyperplane vs. confidence level and system load level for, (a) Case 1, (b) Case 2, and (c) Case 3.

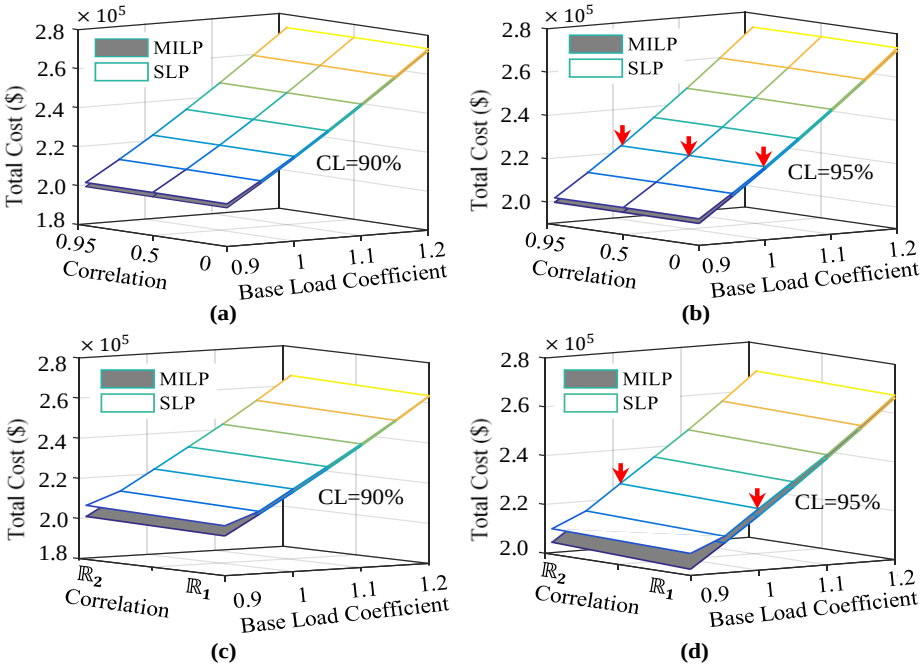


Figure 3.16 Illustration of total cost hyperplane vs. correlation level and system load level, (a)-(b) Case 4, and (c)-(d) Case 5.

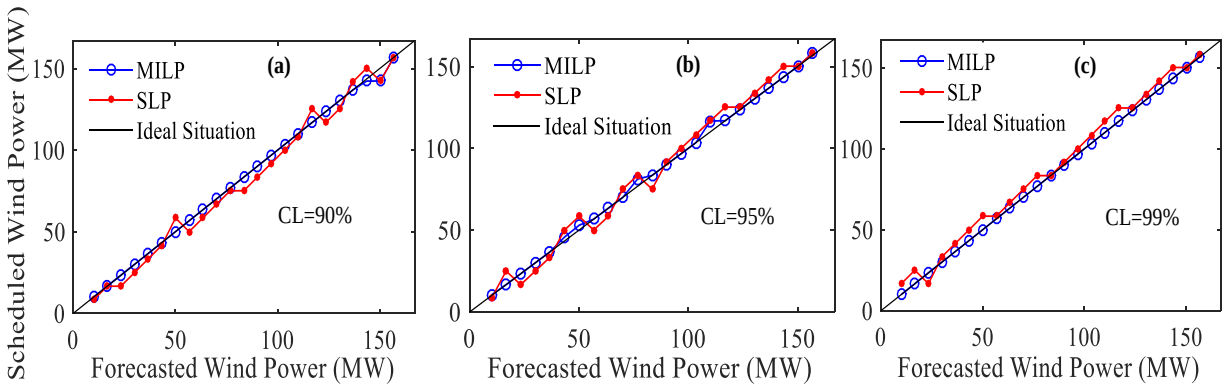


Figure 3.17 Scheduled vs. forecasted wind power generation in Case 2.

3.6 Summary

This chapter proposes an efficient wind power probability distribution modeling using sample moments of wind power data and BKDR technique. The proposed model shows high accuracy and low computational burden for practical applications. Also, an efficient methodology for obtaining an NCF is proposed to be able to convert the conventional non-linear CCED model to an MILP-based model using an accurate piecewise linearization technique. The proposed MILP-based CCED model is convex with respect to the continuous variables of the system and can effectively minimize the total cost. Moreover, the proposed CCED model enables power system operators to use both parametric and non-parametric models of wind power CDF. The NCF includes a set of cost coefficients which are stored in a database and calculated through CDF calculation before running CCED model. As a result, the solution process is computationally efficient and much closer to the global solution of the initial non-linear CCED problem. In addition, the proposed MILP-based model can consistently reflect the effect of WFs' correlation on the total reserve and wind power misestimation costs. Furthermore, because the nonlinearity of the CCED cost function increases proportional to wind power generation capacity, the proposed model is very efficient in high wind power penetration scenarios. Compared to the existing approaches, the proposed CCED model leads to a more optimal generation and reserve scheduling, thereby achieving a reasonable reduction in total system costs. As a future work, in the proposed MILP-based CCED model, the uncertainty of solar generation and electricity load can also be efficiently considered. Without loss of generality, the proposed NCF and MILP-based CCED model can be formed after probability

distribution modeling of correlated uncertain resources such as wind power, solar power, and electricity load. Moreover, as wind farms might have different levels of uncertainty, certain penalty factors and chance constraints might be defined for individual wind farms owning by different producers. Therefore, the required reserve for managing the underestimation and overestimation of each wind farm can be efficiently determined proportional to its level of uncertainty without calculating the aggregate wind power generation. In addition, the proposed MILP-based model might be adopted by look-ahead CCED problems using DRCC optimization approaches. In DRCC, instead of accurate distribution modeling, by defining an ambiguity set for wind power generation, a family of simple distributions are covered to compensate inaccurate distribution modeling.

Chapter 4

A Hybrid Probabilistic Wind Power Prediction Based on Kernel Density Estimation

4.1 Introduction

The increasing penetration level of wind power generation in electrical networks has caused a real challenge in today's power systems [5], [105], [123]. 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, and electricity market [124]. 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 [26], [37], [38], [125-127]. The mean-variance, persistence, Quantile regression (QR), lower-upper bound estimation (LUBE) [26], [38], [125-126], hybrid intelligent algorithm (HIA) [127], and conditional density estimation based on *Nadaraya-Watson* estimator [37] 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 [38]. 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 [5], [105] [123], 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 chapter, 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 main time series using sample entropy (SampEn) technique; (i) *trend*, (ii) *cycle*, and (iii) *noise* [123]. 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.

4.2 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 [128]. 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 [129] 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.

4.2.1 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.

$$\overline{IMF}^{(k)} = \sum_{i=1}^{n} IMF_i^{(k)}/n \tag{4.1}$$

4.2.2 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 (4.2).

$$r_1(t) = \sum_{i=1}^{n} M(p_i(t))/n \tag{4.2}$$

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.

$$\widetilde{IMF}^{(2)} = r_1(t) - \frac{\sum_{i=1}^{n} M\left(r_1(t) + \rho_1 . E_2\big(w_i(t)\big)\right)}{n} \tag{4.3}$$

Step 4: Calculate the k^{th} residue for $k = 3, \dots, K$ as (4.4). Note that *K* is the total number of modes.

$$r_k(t) = \frac{\sum_{i=1}^{n} M\left(r_{k-1}(t) + \rho_{k-1} . E_k\big(w_i(t)\big)\right)}{n} \tag{4.4}$$

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 [129] suggests to use small coefficients for time series disturbed by high frequency signals, and vice versa.

4.2.3 Sample Entropy Technique

SampEn is a revised version of approximate entropy (ApEn) which are usually used for physiological time series complexity analysis [130]. 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.

$$X_m(i) = [x_i, x_{i+1}, x_{i+2}, \dots, x_{i+m-1}] \quad i = 1, \dots, N - m + 1 \tag{4.5}$$

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:

$$B_i^a = (\sum_{i=1}^{N-m} B_i)/N - m - 1 \tag{4.6}$$

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

$$A_i^a = (\sum_{i=1}^{N-m} A_i)/N - m - 1 \tag{4.7}$$

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

$$B^a = (\sum_{i=1}^{N-m} B_i^a)/N - m \quad (4.8)$$

$$A^a = (\sum_{i=1}^{N-m} A_i^a)/N - m \quad (4.9)$$

Step 6: Calculate the SampEn by (10).

$$SampEn(m, r, N) = -\ln[A^a/B^a] \quad (4.10)$$

According to [14], the value of m and r can be set as $1 \leq m \leq 2$ and $0.1\sigma \leq r \leq 0.25\sigma$ where σ is the standard deviation of the time series under analysis. In this chapter, $m = 2$ 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 $SampEn_{org} - \delta \leq SampEn_{IMF} \leq SampEn_{org} + \delta$ 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 chapter δ is set on 0.15.

4.3 Direct Plug-In Bandwidth Selection for Kernel Density Estimation

Since the practical implementation of all kernel density estimators, like the kernel density estimator (2.24) with Gaussian kernel (2.1), 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 [46]. A well-defined criterion for optimal selection of h, shown in (4.13), 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 [46].

$$h = \left(\frac{\int_{\mathbb{R}} \kappa(x)^2 dx}{\left(N. \left(\int_{\mathbb{R}} x^2 \kappa(x) dx \right)^2 . \hat{\psi}_4(q) \right)} \right)^{\frac{2}{5}} \tag{4.13}$$

$$\hat{\psi}_z(q) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \kappa_q^{(z)}(x_i - x_j) \tag{4.14}$$

$$q = \left(-2\kappa_q^{(z)}(0)/N. \left(\int_{\mathbb{R}} x^2 \kappa(x) dx \right). \psi_{z+2} \right)^{1/z+3} \tag{4.15}$$

This method depends on the pilot bandwidth q and z^{th} derivative of standard Gaussian kernel $\kappa_q^{(z)}(z \in 2n\,, z \geq 6)$; 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 (4.15).

Step 5: Calculate $\hat{\psi}_4(q)$ by (4.14) to calculate h. If $z \neq 4$ go to Step 4 and continue until $\widehat{\psi}_4(q)$ is obtained.

4.3.1 Quantile Calculation

After estimating PDF of noise samples, using the associated cumulative distribution functions $F(x)$ obtained by (2.16), the lower/upper quantiles $L^{(\alpha)}$ and $U^{(\alpha)}$ corresponding to confidence level $100(1-\alpha)\%$ are obtained via (2.17). These quantiles are used as outputs of ELM in training procedure.

$$F(x) = \int_0^x f(\tau) d\tau \tag{4.16}$$

$$U^{(\alpha)} = F^{-1}\left(\frac{\alpha}{2}\right)\ ,\ L^{(\alpha)} = F^{-1}\left(1 - \frac{\alpha}{2}\right) \tag{4.17}$$

4.4 A Brief Review on Extreme Learning Machine (ELM)

A fast and efficient learning approach is employed in this chapter 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 [131]. 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.

$$\boldsymbol{y}_i = \sum_{j=1}^{\widetilde{N}} \beta_j \, \Phi\left(\boldsymbol{\omega}_j . \boldsymbol{x}_i + b_j\right), \qquad i = 1, \dots, N \tag{4.18}$$

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 (4.18) can be mathematically written by a simple matrix multiplication as $\mathbf{H\beta} = \mathbf{Y}$. The input matrix $\mathbf{H}$ is expressed as (4.19) where the output weight matrix $\boldsymbol{\beta}$ and the target matrix $\mathbf{Y}$ are illustrated by (4.20).

$$\mathbf{H} = \begin{bmatrix} \Phi(\boldsymbol{\omega}_1 . \boldsymbol{x}_1 + b_1) & \cdots & \Phi(\boldsymbol{\omega}_{\widetilde{N}} . \boldsymbol{x}_1 + b_{\widetilde{N}}) \\ \vdots & \ddots & \vdots \\ \Phi(\boldsymbol{\omega}_1 . \boldsymbol{x}_N + b_1) & \cdots & \Phi(\boldsymbol{\omega}_{\widetilde{N}} . \boldsymbol{x}_N + b_{\widetilde{N}}) \end{bmatrix}_{\boldsymbol{N \times \widetilde{N}}} \tag{4.19}$$

$$\boldsymbol{\beta} = \left[\boldsymbol{\beta}_{\mathbf{1}}^{\boldsymbol{T}} \quad \dots \quad \boldsymbol{\beta}_{\widetilde{\boldsymbol{N}}}^{\boldsymbol{T}}\right]^{\boldsymbol{T}}, \ \mathbf{Y} = [\boldsymbol{y}_{\mathbf{1}}^{\boldsymbol{T}} \quad \dots \quad \boldsymbol{y}_{\boldsymbol{N}}^{\boldsymbol{T}}]^{\boldsymbol{T}} \tag{4.20}$$

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 $\mathbf{H}$ [132].

4.5 Proposed Hybrid PI Construction Framework

The general structure of the proposed wind power PI construction approach is shown in Figure 4.1. 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 (4.21)-(4.22) 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.

$$MAE = \frac{1}{N_{ts}} \sum_{i=1}^{N_{ts}} |y_i - \hat{y}_i| \tag{4.21}$$

$$RMSE = \left(\frac{1}{N_{ts}} \sum_{i=1}^{N_{ts}} (y_i - \hat{y}_i)^2 \right)^{1/2} \tag{4.22}$$

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.

Reliability: The PI coverage probability ($PICP$) in (4.23) should be too close to PI nominal coverage ($PINC$) to meet the reliability criterion as the most important feature.

$$PICP = \frac{1}{N_{ts}} \sum_{i=1}^{N_{ts}} \delta_i \tag{4.23}$$

$$\delta_i = \begin{cases} 1 & , \quad L_i^\alpha \le y_i \le U_i^\alpha \\ 0 & , \quad y_i < L_i^\alpha \ or \ y_i > U_i^\alpha \end{cases} \tag{4.24}$$

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.

Sharpness: To present meaningful information by a PI, the PI normalized average width ($PINAW$) in (4.25) should take small values as much as achievable.

$$PINAW = \frac{1}{R.N_{ts}} \sum_{i=1}^{N_{ts}} (U_i^\alpha - L_i^\alpha) \tag{4.25}$$

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

In this chapter, to deterministically predict trend and cycle time series, ELM_{trend} and ELM_{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.

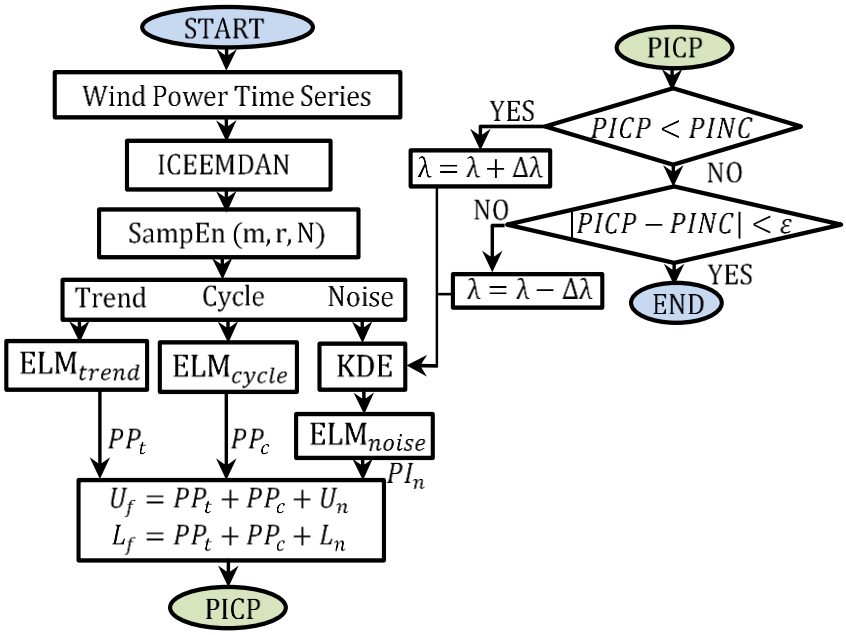


Figure 4.1 Structure of the proposed wind power prediction model.

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, ELM_{noise} is trained considering lower and upper quantiles of estimated PDFs as outputs and noise samples as inputs. After completing training procedure, ELM_{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 $h^* = \lambda.h$ 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 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.

4.6 Case Studies

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 [120]. 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 [133].

4.6.1 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.

4.6.2 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 Figure 4.2. 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 Figure 4.3, 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. Figure 4.4 and Figure 4.5 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 4.1 for Centennial dataset as a highly chaotic case study [5]. 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.

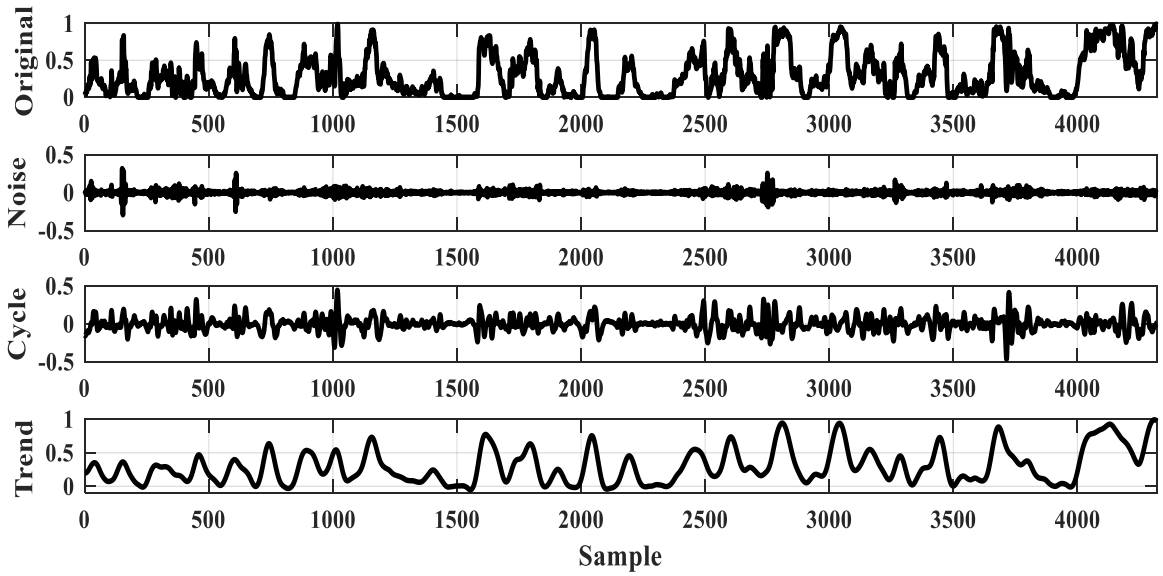


Figure 4.2 Original and main components of Centennial time series.

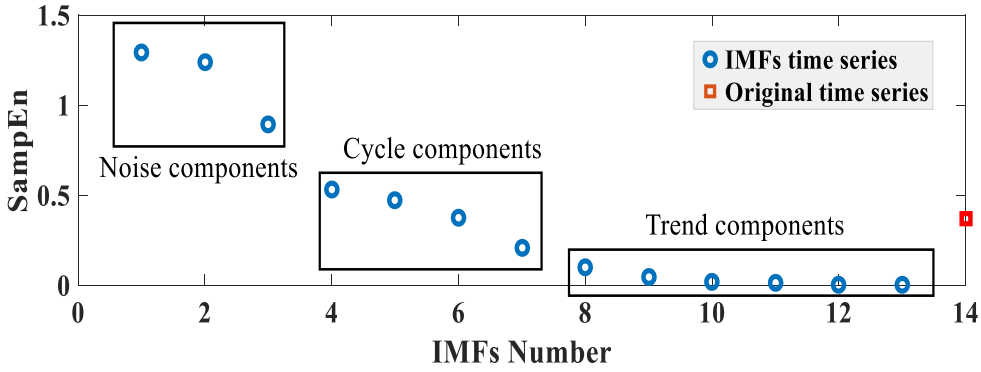


Figure 4.3 Value of SampEn for IMFs obtained by ICEEMDAN.

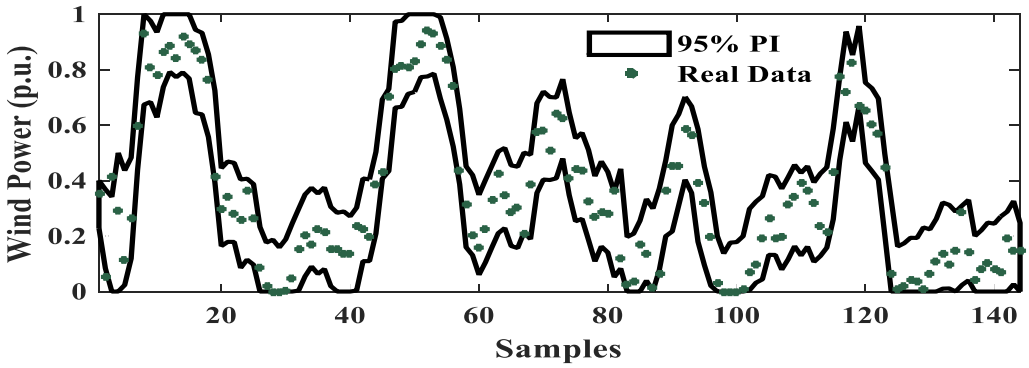


Figure 4.4 Constructed PI at 95% confidence level for Centennial.

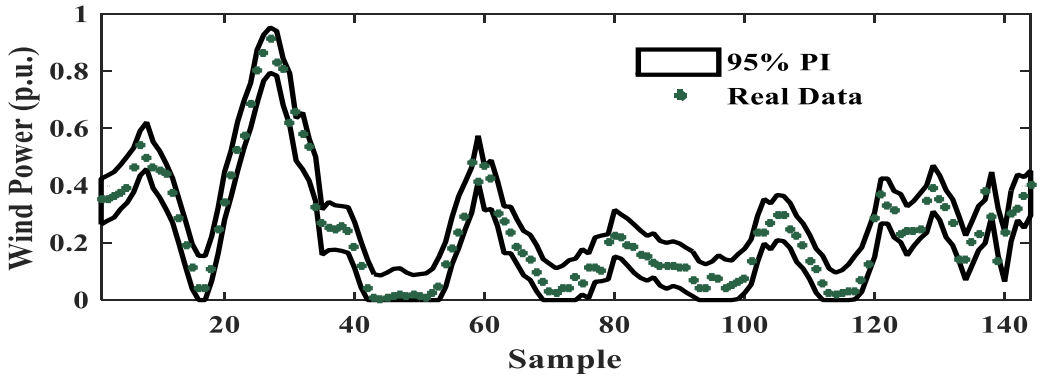


Figure 4.5 Constructed PI at 95% confidence level for AESO.

Table 4.1 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**

4.7 Summary

Due to the high importance of wind power prediction models in modern power systems, in this chapter 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.

Chapter 5

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

5.1 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 [5], [134], [135]. 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 [5], [136-138]. 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 [36-38], [54], [123], [125], [139-144]. 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 [82], [83], [96], [145-147]. 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 [82], [83], [108].

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 non-linear and multi-objective function that should be minimized through heuristic optimization algorithms [38], [54]. It might be solved efficiently 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 [36], [125]. However, the objective function can be linearized to be efficiently solved using a linear programming (LP)-based optimization problem [139]. 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 [141]. 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 [123], [138]. 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 [37], [142]. Furthermore, the bandwidth (BW) of these kernels is traditionally selected with plug-in techniques [45], [46], [148], [149]. 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 [29], [30], [34]. A PI construction approach based on decomposition of original time series into trend, cycle, and noise components has also been proposed [123]. 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 [144] 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 chapter 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 [45]. This paper refers to KDE as the most well-known NDE approach with BW as an important parameter [46]. 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 [148], [149], (ii) conventional KDE approaches result in a tendency to ignore the peaks and valleys of the true density [150], and (iii) *boundary effects* might lead to invalid densities [151], [152]. Although these problems have been mitigated to some extent using more advanced estimators, e.g., *balloon* estimators, *nearest neighbor* estimators [150], [153], *sample point adaptive* estimators [45], and *boundary kernel* estimators [154], 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 chapter 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, and lower/upper quantiles) 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 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 chapter 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 to deal with the historical data. The approach in [29], [30], [34] 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 chapter 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.

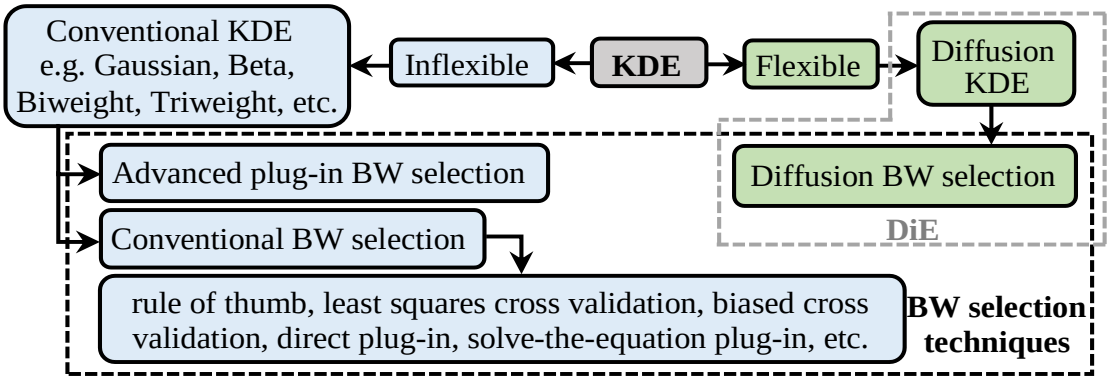


Figure 5.1 Diagram of KDE technique variants and corresponding BW selection techniques.

5.2 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.

5.2.1 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 [12], [34]. 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 [37], [142], as discussed in the Introduction. The classification of different kinds of KDE techniques with related BW selection techniques is illustrated in Figure 5.1. 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 [45], [46], [50], [139].

$\mathbf{P_1}$: 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. **P_3**: They suffer from boundary effects. These problems are addressed in the proposed framework in Sections 5.3 and 5.4 in more detail.

5.2.2 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:

$$\hat{f}_t(x) = \frac{1}{N_s}\sum_{i=1}^{N_s} \kappa(x, X_i; h) \qquad x \in \mathbb{x} = [0,1] \tag{5.1}$$

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 (5.2) has been widely used in the literature [143], [46], [50]. In this chapter, the Gaussian kernel density estimator (GE) is denoted as a conventional KDE technique and compared with the DiE.

$$\kappa(x, X_i; h) = \varphi(x, X_i; h) = \frac{\exp\left(-\left((x - X_i)/\sqrt{2}h\right)^2\right)}{\sqrt{2\pi}\, h} \tag{5.2}$$

5.2.3 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 (5.3).

$$\frac{\partial \hat{f}(x; t)}{\partial t} = L\left(\hat{f}(x; t)\right) \qquad t > 0 \ , \ \ x \in \mathbb{x} = [0,1] \tag{5.3}$$

$$\hat{f}(x; 0) = \Delta x = \frac{1}{N_s}\sum_{i=1}^{N_s} \delta(x - X_i) \tag{5.4}$$

where Δx in the initial condition (5.4) is the empirical density of data on $\mathbb{x}$. The linear differential operator $L(\cdot)$, expressed by (5.5), includes two arbitrary positive functions, $a(x)$ and $p(x)$, that meaningfully affect the performance of the DiE. The boundary condition (5.6) should also be

considered to ensure that $\hat{f}$ integrates to unity [50].

$$L(\cdot) = d\,(a(x)\,d(\cdot/p(x))/dx)/2dx \tag{5.5}$$

$$\partial\left(\hat{f}(x;t)/p(x)\right)/\partial x\big|_{\partial\mathbb{x}} = 0 \tag{5.6}$$

The solution to (5.3) 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 (5.5), the operator $L(\cdot)$ can be adjusted when $a(x) = p(x)^{\lambda}$ and $\lambda \in [0,1]$. Therefore, a general solution to (5.3) is written in the form of (5.7), in which PDEs (5.8)-(5.9) are satisfied for fixed values of y and x, respectively.

$$\hat{f}(x;t) = \frac{1}{N_s}\sum_{i=1}^{N_s} \kappa_{\mathrm{DiE}}(x, X_i; t) \tag{5.7}$$

$$\partial\big(\kappa_{\mathrm{DiE}}(x,y;t)\big)/\partial t = L\big(\kappa_{\mathrm{DiE}}(x,y;t)\big),\ x \in \mathbb{x}\,,\ t > 0 \tag{5.8}$$

$$\partial\big(\kappa_{\mathrm{DiE}}(x,y;t)\big)/\partial t = L^{*}\big(\kappa_{\mathrm{DiE}}(x,y;t)\big),\ y \in \mathbb{x}\,,\ t > 0 \tag{5.9}$$

where $L^{*}(\cdot) = \partial\,(a(y)\,\partial(\cdot)/\partial y)/\partial y/2p(y)$ is the adjoint operator of $L(\cdot)$. For wind power time series, boundary conditions (5.10) and (5.11) 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 (5.8)-(5.11) on $\mathbb{x}$ is expressed in the form of (5.12).

$$\partial(\kappa_{\mathrm{DiE}}(x,y;t)/p(x))/\partial x|_{\partial\mathbb{x}} = 0 \tag{5.10}$$

$$\partial(\kappa_{\mathrm{DiE}}(x,y;t))/\partial y|_{\partial\mathbb{x}} = 0 \tag{5.11}$$

$$\kappa_{\mathrm{DiE}}(x,y;\ h_{\mathrm{DiE}}) = \frac{p(x).\exp\left(-\int_y^x \sigma^{-1}(z)\,dz\ /\left(\sqrt{2}\,h_{\mathrm{DiE}}\right)^2\right)}{\sqrt{2\pi}\ \ h_{\mathrm{DiE}}\ \sqrt[4]{p(x)a(x)p(y)a(y)}} \tag{5.12}$$

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 (5.12) 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 (5.3) would be the well-known Fourier heat equation with the conventional inflexible

GE in (5.2) 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.

5.3 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.

5.3.1 Optimal BW Selection: Criterion and Techniques

A well-defined criterion for optimal BW selection is the mean integrated squared error ($MISE$) expressed by (5.13), which can be divided into two components: the integrated squared bias and integrated variance. It is proven with technical details in [45], [46], [50] that an optimal BW minimizes the first-order asymptotic approximation of $MISE$:

$$MISE = \int \left(\mathbb{E}_f\left(\hat{f}_t(x)\right) - f_t(x)\right)^2 dx + \int Var_f\left[\hat{f}_t(x)\right]dx \tag{5.13}$$

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 [45], as will be shown in Section 5.3.2.

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.

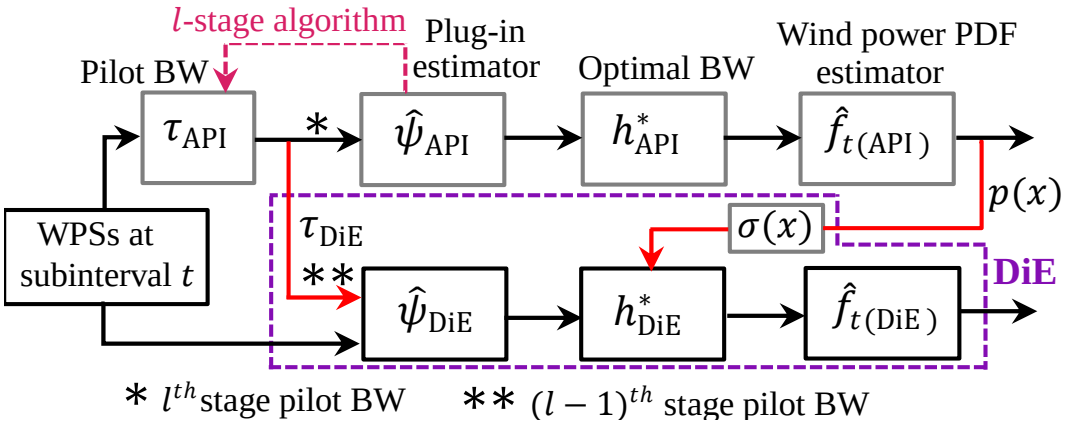


Figure 5.2 General diagram of wind power PDF estimation via optimal BW selection techniques.

5.3.1.1 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 (5.14), (5.15). Because the DPI BW selection technique depends on the pilot BW τ_{DPI} in (5.14) and the r^{th} derivative of standard Gaussian kernel $\kappa_{\text{G}}^{(r)}$ with second moment $\mu_{2(\kappa_{\text{G}})} = \int_{\mathbb{R}} x^2 \kappa_{\text{G}}(x) dx$, an l-stage DPI BW selector is developed in Algorithm 1 [46].

$$\tau_{\text{DPI}} = \left(-2\, \kappa_{\text{G}}^{(r)}(0) / N_s\, \mu_{2(\kappa_{\text{G}})}\, \hat{\psi}_{\text{DPI}}^{(r+2)}\right)^{1/r+3} \tag{5.14}$$

$$\hat{\psi}_{\text{DPI}}^{(r)} = \frac{1}{{N_s}^2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} \kappa_{\text{G}}^{(r)}\left(X_i, X_j; \tau_{\text{DPI}}\right) \tag{5.15}$$

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 chapter, the Gaussian estimator that uses the DPI technique is denoted by GE_{DPI}, and the estimated PDF is indicated by $\hat{f}_{t(\text{DPI})}(x)$.

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 $\hat{\psi}_{\mathrm{DPI}}^{(ini)} = (-1)^{r/2}\, r!/\left(\sqrt{\pi}(2\sigma_s)^{r+1}(r/2)!\right)$.

3: Find pilot BW τ_{DPI} using (5.14) and then $\hat{\psi}_{\mathrm{DPI}}^{(r)}$ using (5.15).

4: Continue the process to obtain $\hat{\psi}_{\mathrm{DPI}}^{(4)}$, then use (5.14) to obtain the optimal value of BW as $h_{\mathrm{DPI}}^{*} = \tau_{\mathrm{DPI}}|_{r=2}$.

5.3.1.2 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 (5.16), where an l-stage algorithm is stated in detail to calculate the pilot BW $\tau_{\mathrm{API}}^{(1)}$ using (5.17), (5.18). 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 (5.18), the j^{th} derivative of the Gaussian kernel φ is shown by $\varphi^{(j)}$ [50].

$$h_{\mathrm{API}}^{*} \cong \left(0.9\,\tau_{\mathrm{API}}^{(1)}\right)^{\frac{1}{2}} \tag{5.16}$$

$$\tau_{\mathrm{API}}^{(J)} = \left[\frac{\left(1+1/\left(2^{J+0.5}\right)\right)\left((2J-1)\times\ldots\times3\times1\right)}{\left(3N_s\sqrt{\pi/2}\,\hat{\psi}_{\mathrm{API}}^{(J+1)}\right)}\right]^{\frac{2}{3+2J}} \tag{5.17}$$

$$\hat{\psi}_{\mathrm{API}}^{(J)} = \frac{(-1)^J}{{N_s}^2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s}\varphi^{(2J)}\left(X_i, X_j; 2\tau_{\mathrm{API}}^{(J)}\right) \tag{5.18}$$

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 (5.18).

3: Find the pilot BW $\tau_{\mathrm{API}}^{(l-1)}$via (5.17).

4: Find the plug-in estimator $\hat{\psi}_{\mathrm{API}}^{(l-1)}$ using (5.18) and $\tau_{\mathrm{API}}^{(l-1)}$ obtained from the previous stage, and

continue this procedure until $\hat{\psi}_{\text{API}}^{(2)}$ and consequently $\tau_{\text{API}}^{(1)}$ are acquired.

5: If $\left|\tau_{\text{API}}^{(1)} - \tau_{\text{API}}^{(l)}\right| < \varepsilon$, equation (5.16) gives the optimal BW for the Gaussian estimator (5.1); else go to step 2 with $\tau_{\text{API}}^{(l)} = 0.9\,\tau_{\text{API}}^{(1)}$.

5.3.1.3 BW selection technique for DiE

DiE BW selection leads to the minimum value of asymptotic $MISE$ using the DiE optimal BW in (5.19) where the plug-in estimator $\hat{\psi}_{\text{DiE}}$ is estimated by (5.20) 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 (5.21), $L(\cdot)$, and $L^*(\cdot)$ depend on $a(x)$ and $p(x)$, Algorithm 2 is first executed to estimate $p(x) = \hat{f}_{t(\text{API})}(x)$. Then, by adjusting $\lambda \in [0,1]$, $a(x)$ is acquired using $a(x) = p(x)^{\lambda}$.

$$h_{\text{DiE}}^* \cong \left(0.5\,\mathbb{E}_f[\sigma^{-1}(x)] / \left(N_s \sqrt{\pi}\,\hat{\psi}_{\text{DiE}}\right)\right)^{\frac{1}{5}} \tag{5.19}$$

$$\hat{\psi}_{\text{DiE}} = \frac{1}{{N_s}^2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} L^* \left(L\left(\kappa_{\text{DiE}}\left(X_i, X_j; \tau_{\text{DiE}}\right)\right)\right) \tag{5.20}$$

$$\mathbb{E}_f[\sigma^{-1}(x)] = \frac{1}{N_s} \sum_{i=1}^{N_s} \left(a(X_i)/p(X_i)\right)^{-\frac{1}{2}} \tag{5.21}$$

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 (5.21).

4: Find the value of $\hat{\psi}_{\text{DiE}}$ using (5.20), where $\tau_{\text{DiE}} = 2\,\tau_{\text{API}}^{(2)}$ and $\tau_{\text{API}}^{(2)}$ is determined from step 4 in **Algorithm 2**.

5: Calculate h_{DiE}^* using (5.19) as the DiE optimal BW.

A general diagram of BW selection techniques is provided in Figure 5.2. 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 Figure 5.2. 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 [50]. The PDF estimated by DiE is denoted by $\hat{f}_{t(\mathrm{DiE})}(x)$.

5.3.2 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 BWs h^*_{DPI}, h^*_{API} (with Gaussian kernel) and h^*_{DiE} (with diffusion kernel) the associated PDFs are respectively estimated as shown in Figure 5.3 (a)-(i). Thereafter, using corresponding cumulative distribution functions $F_{t(\cdot)}(x)$, obtained from (5.22), one can calculate the lower/upper quantiles $Q_t^{(\alpha_l)}$ and $Q_t^{(\alpha_u)}$ and the related interval width using (5.23) and (5.24), respectively.

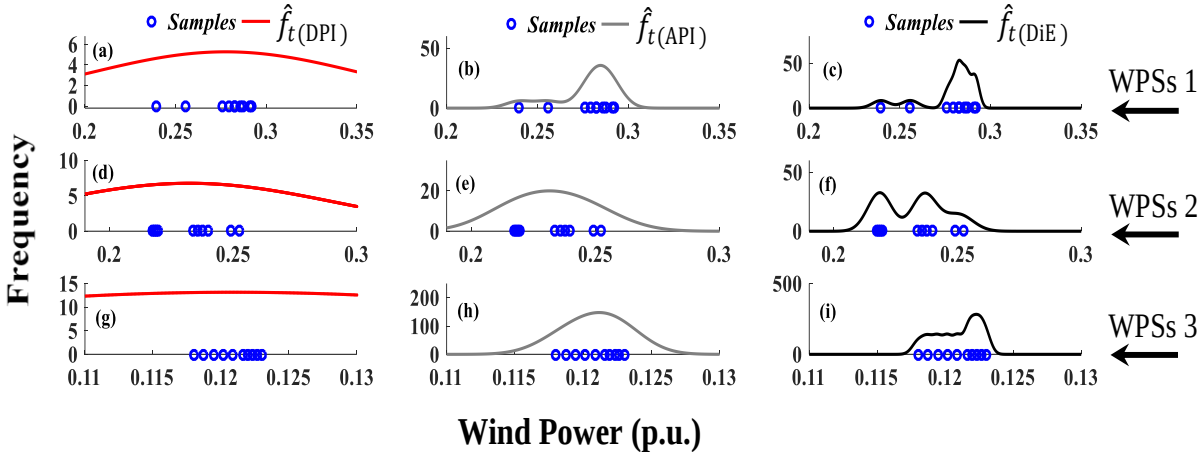


Figure 5.3 Comparison of KDE techniques for three different sets of WPSs drawn from the Centennial wind farm dataset (located in South Saskatchewan, Canada).

Table 5.1 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**

$$F_{t(\cdot)}(x) = \int_0^x \hat{f}_{t(\cdot)}(\omega)d\omega \tag{5.22}$$

$$Q_t^{(\alpha_l)} = F_{t(\cdot)}^{-1}(\alpha/2) \ , \ Q_t^{(\alpha_u)} = F_{t(\cdot)}^{-1}(1-\alpha/2) \tag{5.23}$$

$$Int. = Q_t^{(\alpha_u)} - Q_t^{(\alpha_l)} \tag{5.24}$$

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

Figure 5.3 and Table 5.1 show that, unlike GE_{DPI} and 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 5.1 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.

5.4 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) [139], [141], [123], [131], (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.

5.4.1 Wind Power PIs Evaluation Criteria

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

Reliability: The average coverage error (ACE) in (5.25), which is the deviation of the PI coverage probability ($PICP$) in (5.26) 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.

$$0 \leq ACE = PICP - PINC < \varepsilon \tag{5.25}$$

$$PICP = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \mathbb{I}_{[L_i^\alpha, U_i^\alpha]}(y_i) \tag{5.26}$$

$$\mathbb{I}_{[L_i^\alpha, U_i^\alpha]}(y_i) = \begin{cases} 1 & y_i \in [L_i^\alpha, U_i^\alpha] \\ 0 & y_i \notin [L_i^\alpha, U_i^\alpha] \end{cases} \tag{5.27}$$

where $\mathbb{I}_{[L_i^\alpha, U_i^\alpha]}(\cdot)$ 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 $[L_i^\alpha, U_i^\alpha]$.

Sharpness: To perceive meaningful information from a PI, its normalized average width ($PINAW$) in (5.28) should take small values to induce a sharp PI.

$$PINAW = \frac{1}{R.N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} (U_i^\alpha - L_i^\alpha) \tag{5.28}$$

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

Overall Score: To assess the overall skill of a PI construction approach, the overall score in (5.29) is considered in evaluation process because it simultaneously takes both reliability and sharpness aspects into account. A sharp PI presents a small value for $|Sc|$. 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.

$$Sc = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} \left[-2\alpha (U_i^\alpha - L_i^\alpha) - 4\mathbb{I}_{[0, L_i^\alpha)}(y_i)(L_i^\alpha - y_i) - 4\mathbb{I}_{(U_i^\alpha, 1]}(y_i)(y_i - U_i^\alpha) \right] \tag{5.29}$$

5.4.2 Prediction Model for Lower and 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 [141]. 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 [131]. 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 $\widetilde{N}$ hidden nodes is expressed by $\mathbf{H}_{N\times\widetilde{N}}\boldsymbol{\beta}_{\widetilde{N}\times m} = \mathbf{G}_{N\times m}$. Input and output vectors $\boldsymbol{x}_i$ and $\boldsymbol{g}_i$ are shown in (5.30) and Figure 5.4, and the output weight matrix, $\boldsymbol{\beta}$, and the target matrix, $\mathbf{G}$, are denoted by (5.31).

$$\boldsymbol{x}_i = [x_{i1}, \dots, x_{in}]^T \in \mathbb{R}^n, \boldsymbol{g}_i = [g_{i1}, \dots, g_{im}]^T \in \mathbb{R}^m \tag{5.30}$$

$$\boldsymbol{\beta} = [\,\boldsymbol{\beta}_1 \quad \dots \quad \boldsymbol{\beta}_{\widetilde{N}}]^T\,,\ \mathbf{G} = [\boldsymbol{g}_1 \quad \cdots \quad \boldsymbol{g}_N]^T \tag{5.31}$$

where $\boldsymbol{\beta}_j = \left[\beta_{j1}, \dots, \beta_{jm}\right]^T$ is the output weight vectors. After calculation of $\mathbf{H}$ according to [131] or [141], 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 $\mathbf{H}$ [132]. 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 Figure 5.4, 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 chapter, the 10-fold cross validation technique in [155] is used to identify the optimal number of time lags and hidden nodes and optimal length of the training dataset, which includes N input vectors $[\boldsymbol{x}_i]_{(N_{\mathrm{lag}}\times N_{\mathrm{s}})\times 1}$ and N output vectors $[\boldsymbol{g}_i]_{2\times 1}$ where $N = N_{\mathrm{sub}} - N_{\mathrm{lag}}$ and $N_{\mathrm{sub}} = N_{\mathrm{d}} \times 24 \times 6/N_s$.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 Figure 5.5.

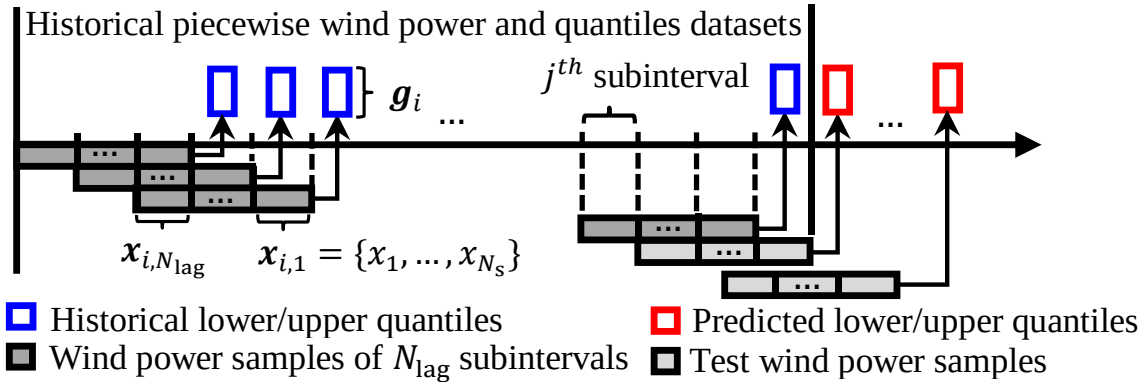


Figure 5.4 Temporal diagram of input and output data structure for ELM training.

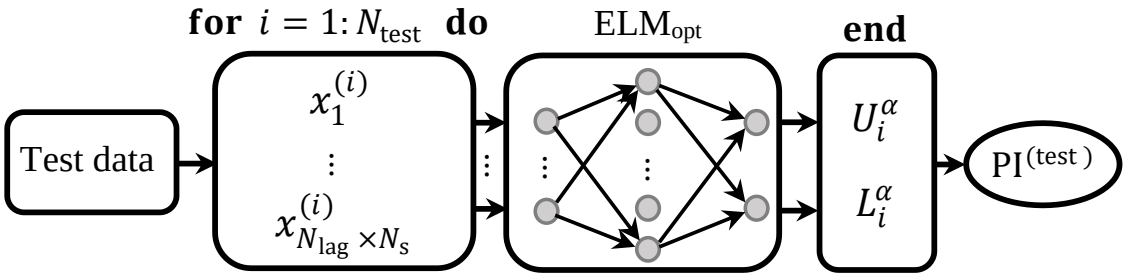


Figure 5.5 Flowchart of the main prediction process.

5.4.3 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 Figure 5.6. Using the average value of WPSs ($\bar{w}$) inside each subinterval, the fuzzy sets tune the flexibility of the DiE through (5.32) to avoid boundary effects. Figure 5.7 shows the boundary effects with an ellipse with a probability allocation for values outside of [0,1].

$$\lambda(\bar{w}, \mu) = (1 - \mu_L)\mathbb{I}_{[0, w_1]}(\bar{w}) + \mu_M \mathbb{I}_{[w_1, w_2]}(\bar{w}) + (1 - \mu_H)\mathbb{I}_{[w_2, 1]}(\bar{w}) \tag{5.32}$$

The philosophy behind the proposed trapezoidal fuzzy sets is to alleviate problem **P$_3$**. If the average of the WPSs is near boundaries, i.e., low-power ($\bar{w} \in [0, w_1^-]$) and high-power ($\bar{w} \in [w_2^+, 1]$) regions where boundary effects might happen, the DiE sets λ=0 to generate sharp and bona fide PDFs as shown in Figure 5.7. For the medium power region ($\bar{w} \in [w_1^+, w_2^-]$), where

boundary effects do not matter, the DiE chooses λ=μ_M ($0 < \mu_M \leq 1$) to estimate smoother PDFs while preserving the main features.

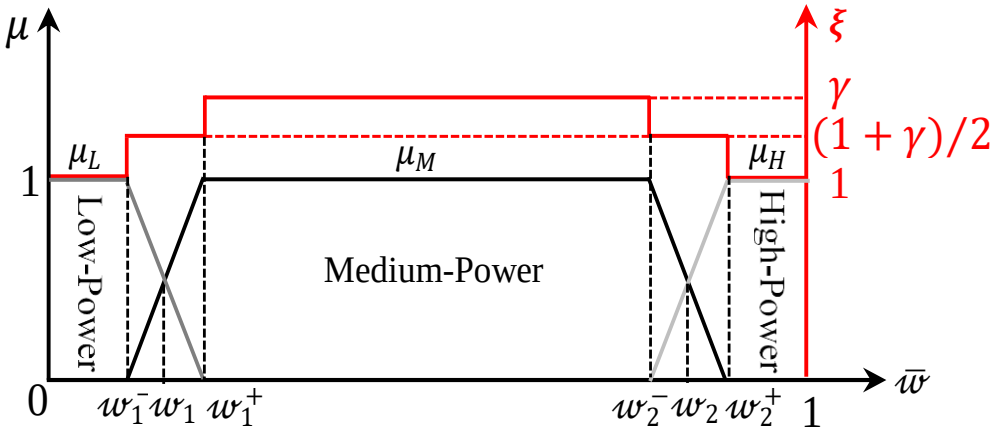


Figure 5.6 Proposed trapezoidal fuzzy sets (left axis) and tri-level adaptation function $\xi(\bar{w}, \gamma)$ (right axis) considered for the proposed DiE.

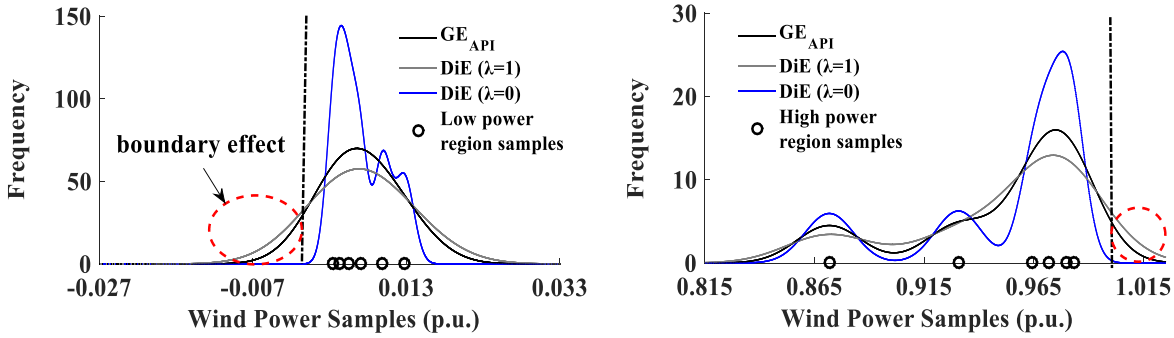


Figure 5.7 Illustration of boundary effects for WPSs near boundaries.

5.4.4 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 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 proposed adaptation function $\xi(\bar{w}, \gamma)$, shown on the right axis of Figure 5.6, aims to adaptively raise the reliability of constructed PIs under such

conditions. Using (5.33) and (5.34), the function $\boldsymbol{\xi}$ leads to the controlled growth of the DiE BW for the subintervals in which the value of $\overline{\imath w}$ is far from the boundaries and the boundary effect does not matter for the PDFs. In (5.34), γ is the BW growth factor, and $\imath w_i^-$ and $\imath w_i^+$ are the average values of wind power smaller and larger than $\imath w_i$. 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 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.

$$h_{\mathrm{DiE}} = \boldsymbol{\xi}.\, h^*_{\mathrm{DiE}} \tag{5.33}$$

$$\boldsymbol{\xi} = \mathbb{I}_{[0,\imath w_1^-]\cup[\imath w_2^+,1]}(\overline{\imath w}) + \left(\frac{1+\gamma}{2}\right)\mathbb{I}_{(\imath w_1^-,\imath w_1^+]\cup[\imath w_2^-,\imath w_2^+)}(\overline{\imath w}) + (\gamma)\mathbb{I}_{(\imath w_1^+,\imath w_2^-)}(\overline{\imath w}) \tag{5.34}$$

5.4.5 Proposed FADiE Framework Stages for Optimal Construction of 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., $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_{\mathrm{M}}]$, 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 Figure 5.8). 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 Figure 5.8 should be repeated over time after each lower/upper quantile prediction in Figure 5.5.

First Stage: In this stage, ELM_1 to $\mathrm{ELM}_{\mathrm{M}}$ 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 $\mathrm{ELM}_{\mathrm{M}}$ 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 Figure 5.9.

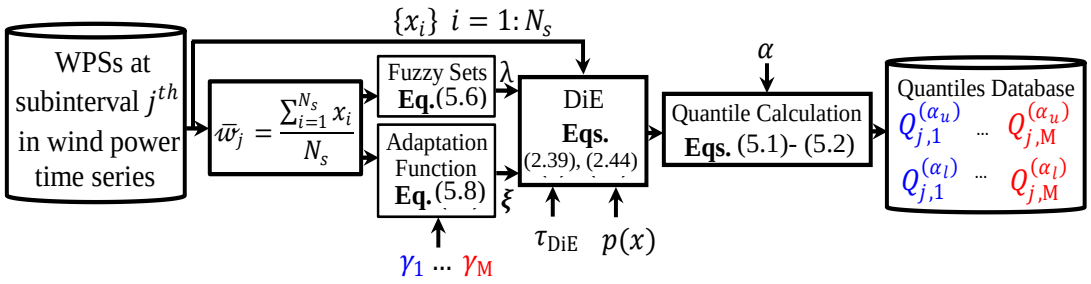


Figure 5.8 General diagram of wind power quantiles database construction.

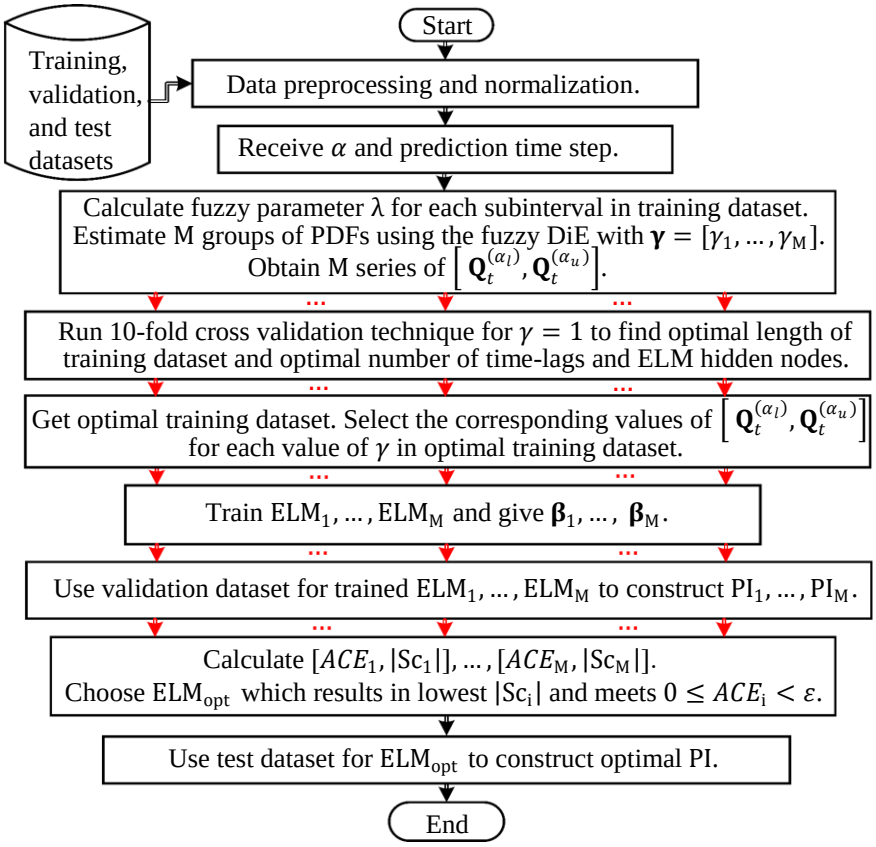


Figure 5.9 Structure of the parallel computing-based FADiE framework.

5.5 Case Studies

5.5.1 Experimental Dataset

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 [120].

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 from October to December of 2015 [121].

Case 4: AESO dataset from January to December of 2015 [120].

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 [5] 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 [82]. 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 [83], [108]. 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

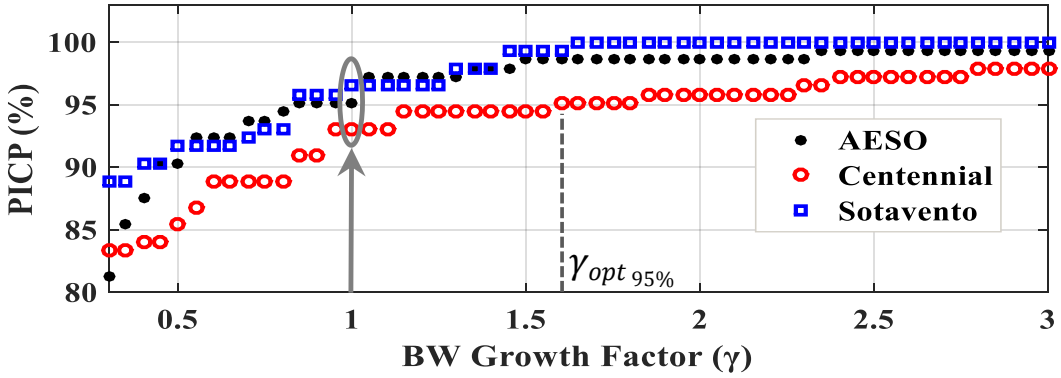
direction, temperature, and numerical weather prediction (NWP) might be needed. The optimal selection of explanatory variables is only necessary for better training of the ELM in our proposed framework [12], [36], [141]. In this study, the parameters of fuzzy sets and adaptation function are set as follows: 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, $\Delta\gamma = 0.1$, γ_1 =0.5, and γ_M =1.2, and for 1-hour horizon $\Delta\gamma = 0.1$, γ_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.

5.5.2 Analysis of Simulation Results

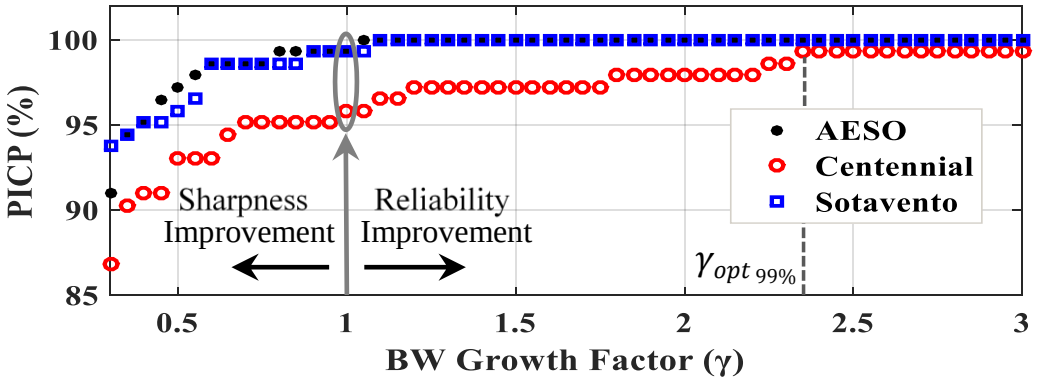
5.5.2.1 Sensitivity Analysis of the Proposed FADiE Framework

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. Figure 5.10 shows that, for a certain CL and 1-hour prediction horizon, the desired reliability of a PI might be ideally gained by $\gamma_{opt} = 1$, 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., $\gamma_{opt_{95\%}}$=1.6 and $\gamma_{opt_{99\%}}$=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.



(a)



(b)

Figure 5.10 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%.

5.5.2.2 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 5.2. 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 5.3 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.

Table 5.2 Comparison of KDE-based approaches and effect of fuzzy sets.

Approach CL=95%	**AESO 2012** (1-hour)			**Sotavento** (1-hour)		
	$PICP$ (%)	$PINAW$ (%)	$\lvert S_c \rvert$ (%)	$PICP$ (%)	$PINAW$ (%)	$\lvert S_c \rvert$ (%)
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 5.3 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**

5.5.2.3 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 5.4. 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.

Table 5.4 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		

5.5.2.4 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 chapter, $30 \leq \widetilde{N} \leq 40$, $3 \leq N_{\text{lag}} \leq 8$, 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.

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 5.5 to 5.7. 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 5.8 and 5.9 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 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 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 5.5 to 5.7, 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 5.8 and 5.9, 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).

Table 5.5 Results of PI construction for Case 1.

Horizon	PINC	Method	$PICP$ (%)	ACE (%)	$PINAW$ (%)	$\lvert S_c \rvert$ (%)
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 5.6 Results of PI construction for Case 2.

Horizon	PINC	Method	$PICP$ (%)	ACE (%)	$PINAW$ (%)	$\lvert S_c \rvert$ (%)
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

Table 5.7 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

Table 5.8 Results of PI construction for Case 4, 30-min ahead.

Prediction Horizon	30-min (PINC=95%) $\boldsymbol{\gamma_{opt-Sp} = 0.7, \gamma_{opt-Su} = 0.6, \gamma_{opt-Au} = 0.6, \gamma_{opt-Wi} = 0.7}$					
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 5.9 Results of PI construction for Case 4, 1-hour ahead.

Prediction Horizon	1-hour (PINC=95%) $\boldsymbol{\gamma_{opt-Sp}=1.0, \gamma_{opt-Su}=2.0, \gamma_{opt-Au}=1.0, \gamma_{opt-Wi}=1.6}$					
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	

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 Figures 5.11 to 5.13 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.

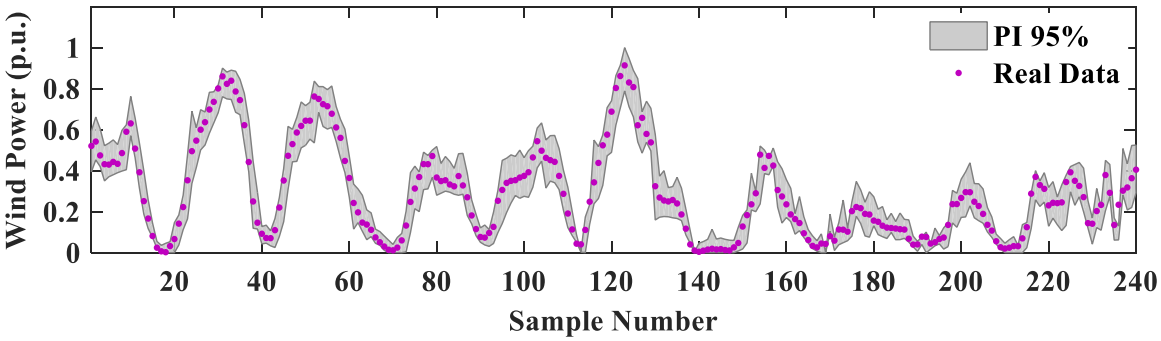


Figure 5.11 Constructed PI for 1-hour prediction horizon for Case 1 (AESO).

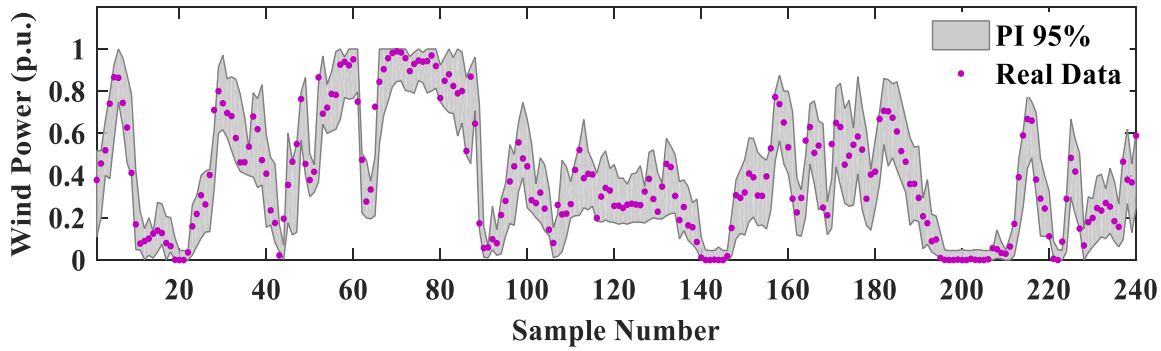


Figure 5.12 Constructed PI for 1-hour prediction horizon for Case 2 (Centennial).

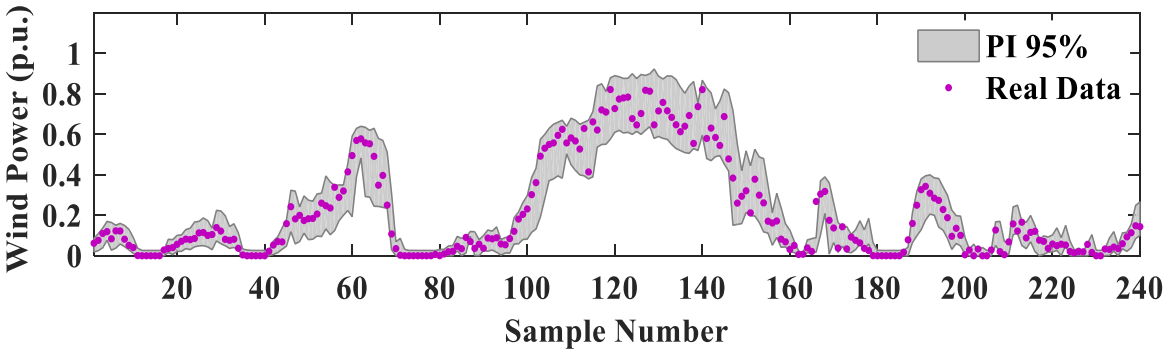


Figure 5.13 Constructed PI for 1-hour prediction horizon for Case3 (Sotavento).

5.6 Summary

This chapter 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 prediction model for longer prediction horizons.

Chapter 6

An Efficient Wind Generation Dependence Structure Modeling for A Robust Energy and Reserve Dispatch

6.1 Introduction

The economic operation, flexibility, and security of modern power systems in terms of different criteria will be negatively influenced by the uncertainty imposed by geographically diverse and correlated wind farms. This challenge would be much worse in case of numerous wind farms because of more interaction. The criteria might be static security criteria, such as bus overvoltage, transmission line overloading, and reserve depletion, or economic criterion like incremental cost of total generation, or transient stability as a dynamic security criterion [13], [29], [30], [33], [67-69], [72-74], [81-90], [100], [108]. To cope with this challenge, as the first step, precise dependence structure (DS) modeling of wind power generation should be used as an efficient way to quantify the related uncertainty [91-93]. Then, to complement uncertainty handling, in the second step, look-ahead energy and reserve scheduling might be carried out over the next few hours using efficient optimization methodology including stochastic optimization (SO), robust optimization (RO), and hybrid optimization. Because of the intractability of SO in large-scale systems with a large number of uncertainty sources, RO has been widely used for uncertainty management in power systems [81-87]-[89]. However, the type of cost function and the defined uncertainty sets in most of the robust optimization problems lead the developed formulations to be bilevel and often difficult to be solved directly in one stage. Also, RO has some biconvex quadratic programming models that require alternating convex search (ACS) technique for a theoretically optimal solution [82], [156]. Therefore, a two-stage optimization methodology is often used for RO problems. Moreover, RO might lead to highly conservative solutions if both the solution methodology and the uncertainty set are not well defined. Importantly, the dependence structure of uncertainty sources is not well-studied in RO. In the next section, some drawbacks of RO

problems are elaborated in detail.

The main motivation of this chapter is to propose an efficient RO problem for robust energy and reserve optimization with an adjustable conservativeness feature in which the DS of all uncertainty sources is precisely modeled using canonical vine (C-Vine) copula. The proposed C-Vine copula-based DS modeling can efficiently quantify the uncertainty level of any number of uncertainty sources through joint and conditional probability distributions modeling. Also, unlike existing research works [81-89], [156], the proposed optimization problem has linear constraints and can be efficiently solved in one stage using off-the-shelf mathematical programming solvers such as CPLEX and Gurobi. In addition, for those problems with nonlinear cost functions, it can provide the opportunity to use linearization techniques [72].

There are several methodologies for handling uncertainties in power systems. To manage the uncertainty of renewable energy resources, authors in [66-68] proposed SO in look-ahead ED. However, the performance of SO approach highly depends on the values and probabilities of the scenarios selected from a predetermined probability distribution. A more flexible SO, i.e., chance-constrained SO (CCSO) in which a certain level of risk is predefined is used in [29], [30], [69]-[71]. Although the conservativeness of the solution can be controlled by adjusting the level of risk, accurate probability distributions are required for the uncertainty sources to keep the security of the system with a certain probability level.

Alternatively, distributionally robust chance-constrained (DRCC) optimization is used in AC-OPF and energy and reserve dispatch [72-77], [157] where an ambiguity set is specified based on the first few moments to account for a family of distributions for each uncertainty source. In [158], [159], the second-order cone programming is used to solve the proposed DRCC model. In a practical point of view, both the probability distributions family and the moments are difficult to find for a DRCC problem. Although accurate probability distributions or precise moments are not necessary for DRCC model, the proposed semidefinite algorithm is not computationally efficient for online situations.

In the early 1970s, RO was proposed by Soyster to develop a model that is immune to the data of a convex uncertainty set [78], and then it was expanded by Ben-Tal and Nemirovski [79], [80]. In RO technique, the operator can improve the security of the system by letting wind power change over a predetermined prediction interval [81-89]. Although it might lead to an over-conservative solution, it can be controlled using efficient techniques which are discussed in the next sections.

A two-stage RO is formulated in [84] for an economic dispatch problem in which the load is specified with an uncertainty set. In [85], an adjustable RO approach is proposed for large-scale wind power integration considering a prespecified uncertainty set. The co-optimization of the security and operational cost is fulfilled in [86] using a robust ED accounting for the automatic generation control affine recourse process. An adjustable wind power uncertainty set by defining a dynamic confidence level is used to mitigate the over-conservativeness of the RO in [87] in a look-ahead power dispatch.

A robust admissibility assessment is presented in [88] to measure the amount of wind generation that can be integrated into a large-scale system considering the operational risk. Similarly, a robust risk-constrained day-ahead UC is presented in [89] using an adjustable uncertainty set to avoid the conservative solutions. Reference [90] modeled wind generation by an interval and proposed a two-stage RO for sizing energy storage systems to mitigate the wind curtailment in a second-order cone programming. In recent years, Copula theory has drawn the attention of researchers to model the uncertainty of wind generation in power system operation [91]. An efficient algorithm based on Copula theory and dependent discrete convolution is proposed in [91] to tackle the "curse of dimensionality" of high-dimensional wind generation uncertainty modeling.

This chapter focuses on an efficient ARO technique with linear constraints for uncertainty management of large-scale wind generation. Unlike the literature, the proposed framework avoids two-stage optimization methodologies and uses a precise DS modeling to account for the spatiotemporal correlation of wind farms. This chapter proposes an efficient DS modeling for an adjustable robust optimization and makes the following contributions:

- The constraints of the proposed RO-based energy and reserve dispatch problem are linear; so, the cost function is directly minimized just in one stage and one iteration.
- The proposed RO problem can control the conservativeness of the solution using two degrees of freedom: (i) violation probability, (ii) confidence level of prediction intervals.
- An efficient DS modeling is proposed for large-scale wind generation uncertainty modeling using C-Vine copula to account for the spatiotemporal correlation of all wind farms.

6.2 Dependence Structure Modeling Using C-Vine Copula

6.2.1 Background

Dependence structure reflects the behavior of an intermittent variable considering the effects of all other stochastic variables. A pair of independent stochastic variables do not affect the probability distributions of each other. In contrast, dependent variables might be partially or completely correlated so that any change in one variable can affect the value of the other one; therefore, the probability distributions depend on each other. In such a case, conditional probability distributions are not equal to marginal distributions. Also, it is extremely challenging to obtain the joint distribution of the stochastic variables specially in high-dimensional cases when there are basically different shapes of distributions. In this chapter, the output of each wind farm and the corresponding point forecasts are regarded as dependent stochastic variables. Copula theory (6.1), proposed by Sklar in 1959, remediate the abovementioned challenge by establishing a link between the multivariate joint distribution $F(\boldsymbol{x})$ and the univariate marginal distributions $F_i(x_i)$. In (6.1), the copula distribution function is shown by $C(\cdot)$ with the domain $[0,1]^d$ [160].

$$F(\boldsymbol{x}) = C\big(F_1(x_1), \dots, F_d(x_d)\big) \quad \boldsymbol{x} = [x_1, \dots, x_d]^{\mathrm{T}} \tag{6.1}$$

The Sklar's theorem can model the joint multivariate density as the multiplication of all marginal densities and multivariate DS density function $c(\cdot)$ called copula density function. Based on this theorem, it is easier and more efficient to model DS in uniform domain by variables u_i rather than actual variables x_i. Therefore, it is preferred to model the multivariate copula function rather than joint distribution function.

$$f(\boldsymbol{x}) = \frac{\partial^d F(\boldsymbol{x})}{\partial x_1 \dots \partial x_d} = c\big(F_1(x_1), \dots, F_d(x_d)\big).\prod_{i=1}^{d} f_i(x_i) \tag{6.2}$$

$$u_i = F_i(x_i) \quad , \quad u_i \in [0,1]$$

Generally, there are two popular types of copula families, *elliptical*, e.g., Gaussian and *t*-student, and *Archimedean* copula, e.g., Joe, Frank, Gumbel, and Clayton [161]. In bivariate cases, these copula functions are widely used for joint distribution modeling. However, due to the inflexibility

of the structure and parameter restrictions, copulas become intricate and challenging in high-dimensional cases. Vine copulas, first proposed by Joe in 1996, have been developed to mitigate this challenge and model complex DS patterns by using a cascade of *pair-copulas* [161-163]. Since wind generation probability distribution has some special features like heavy tails and high skewness, it is not efficient to use just well-known symmetrical Gaussian copula function. Therefore, pair-copulas can independently take any of the above distribution functions and give great flexibility for DS modeling.

6.2.2 C-Vine Copula

Vines are graphical illustration of pair-copulas specification. There are generally three kinds of vine copulas, regular vine (R-Vine), drawable vine (D-Vine), and canonical vine (C-Vine). The focus of this chapter is on C-Vine copulas. However, the proposed DS modeling can be extended using the first two types. To explain C-Vine copula, a graphical illustration of Bayesian network is shown in Figure 6.1 which represents the joint distribution of random variables A_1 to A_d by expression (3).

$$P(A_1, \dots, A_d) = \prod_{i=1}^{d} P(A_i | A_1, \dots, A_{i-1}) \tag{6.3}$$

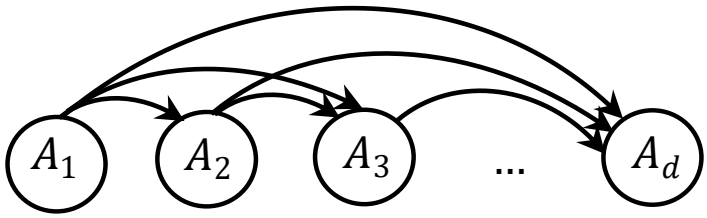


Figure 6.1 Graphical illustration of a fully connected Bayesian network.

In case of continuous random variables, the expression (6.3) is rewritten based on the corresponding density functions as shown by (6.4). As an example, a three-dimensional case is represented in this section. Suppose $\boldsymbol{x} = [x_1, x_2, x_3]^{\mathrm{T}}$ with marginal distributions $F_1(x_1)$, $F_2(x_2)$, and $F_3(x_3)$. Therefore, each conditional distribution in (6.5) can be expressed by Sklar's theorem as (6.6) and (6.7). As shown in (6.8), the expression (6.7) is more simplified by substituting $f(x_1, x_2, x_3)/f_1(x_1)$ by $f_{2,3|1}(x_2, x_3|x_1)$. Similar to (6.6), $f_{3|1}(x_3|x_1)$ in (6.8) can be written as (6.9); as a result, the expression (6.8) is written as (6.10). Based on the expression (6.5), the final

expression for joint density function $f(x_1, x_2, x_3)$ is written as (6.11). It should be noted that the decomposition in (6.5) is not unique, and there might be many structures for construction of pair-copulas $c_{i,j}$.

$$f(x_1, \dots, x_d) = \prod_{i=1}^{d} f_i(x_i|x_1, \dots, x_{i-1}) \tag{6.4}$$

$$f(x_1, x_2, x_3) = f_1(x_1) f_{2|1}(x_2|x_1) f_{3|1,2}(x_3|x_1, x_2) \tag{6.5}$$

$$f_{2|1}(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = c_{1,2}\big(F_1(x_1), F_2(x_2)\big) f_2(x_2) \tag{6.6}$$

$$f_{3|1,2}(x_3|x_1, x_2) = \frac{f(x_1, x_2, x_3)}{f(x_1, x_2)} = \frac{f(x_1, x_2, x_3)}{f_1(x_1) f_{2|1}(x_2|x_1)} \tag{6.7}$$

$$\begin{aligned} f_{3|1,2}(x_3|x_1, x_2) &= \frac{c_{2,3|1}\left(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)\right) f_{2|1}(x_2|x_1) f_{3|1}(x_3|x_1)}{f_{2|1}(x_2|x_1)} \\ &= c_{2,3|1}\left(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)\right) f_{3|1}(x_3|x_1) \end{aligned} \tag{6.8}$$

$$f_{3|1}(x_3|x_1) = c_{1,3}\big(F_1(x_1), F_3(x_3)\big) f_3(x_3) \tag{6.9}$$

$$f_{3|1,2}(x_3|x_1, x_2) = c_{2,3|1}\left(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)\right) c_{1,3}\big(F_1(x_1), F_3(x_3)\big) f_3(x_3) \tag{6.10}$$

$$\begin{aligned} &f(x_1, x_2, x_3) \\ &= c_{1,2}\big(F_1(x_1), F_2(x_2)\big) c_{1,3}\big(F_1(x_1), F_3(x_3)\big) c_{2,3|1}\left(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)\right) \\ &\qquad f_1(x_1) f_2(x_2) f_3(x_3) \end{aligned} \tag{6.11}$$

Therefore, to classify these structures, proposed a graphical illustration of d random variables with a *vine* with $d-1$ trees and $d(d-1)/2$ pair-copulas [161-163]. Figure 6.2 shows 5-dimensional C-vine trees with corresponding nodes and edges.

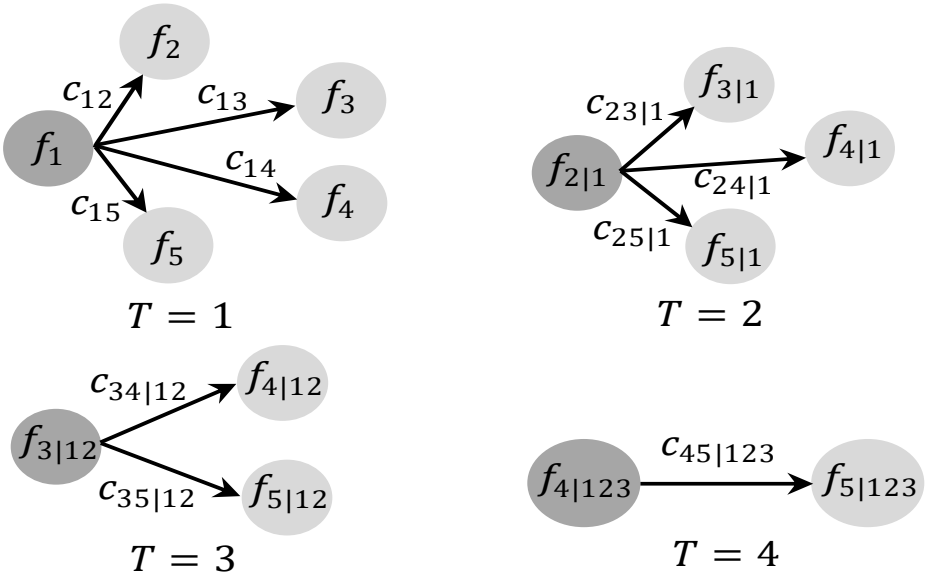


Figure 6.2 Representation of 5-dimensional C-vine trees.

In each of the above C-vine trees, the dependencies with respect to a root node are modeled using bivariate copulas. The root node in each tree is shown by a darker color in Figure 6.2. Then, conditioned on the previous root node, the pairwise dependencies with respect to the next root node are modeled again with bivariate copulas. This process is sequentially done until all dependencies are modeled. Generally, the star structure of C-vine trees leads to the following decomposition formula, i.e., (6.12), that is the generalized version of (6.11).

$$f(\boldsymbol{x}) = \prod_{i=1}^{d} f_i(x_i) \times \prod_{j=1}^{d-1} \prod_{k=1}^{d-j} c_{j,j+k|1:(j-1)}\left(F_j, F_{j+k}|\boldsymbol{\theta}_{j,j+k|1:(j-1)}\right) \tag{6.12}$$

$$F_j = F\left(x_j|x_1, \dots, x_{j-1}\right), \quad F_{j+k} = F\left(x_{j+k}|x_1, \dots, x_{j-1}\right) \tag{6.13}$$

where $\boldsymbol{\theta}_{j,j+k|1:(j-1)}$ is the parameters of bivariate copulas density $c_{j,j+k|1:(j-1)}$. In (6.12), the conditional distributions F_j and F_{j+k} are calculated through equation (6.14) where v_j reflects a random component of $\boldsymbol{v}$, and $\boldsymbol{v}_{-j}$ is the same as $\boldsymbol{v}$ excluding the component v_j.

$$F(x|\boldsymbol{v}) = \frac{\partial C_{xv_j|\boldsymbol{v}_{-j}}\left(F\left(x|\boldsymbol{v}_{-j}\right), F\left(v_j|\boldsymbol{v}_{-j}\right)|\boldsymbol{\theta}\right)}{\partial F(v_j|\boldsymbol{v}_{-j})} \tag{6.14}$$

Each pair-copula, shown as edges labels in Figure 6.2, is allowed to take any copula families such as gaussian, *t*-student, Joe, Frank, etc., so that the final multivariate copula can efficiently model complex dependence structures with special features. In addition, to have an appropriate DS modeling using C-vine copula, the correct order of the root nodes should be determined. In Figure 6.2, it is assumed that the order of root nodes in trees are 1,2,3, and 4, respectively. This ordering is an important factor mainly because there are $d!/2$ different structures that can be considered for C-Vine copula. In the next section, the process of C-Vine copula model selection, including bivariate copula family selection and C-Vine copula structure selection, is described in detail.

6.2.3 Selection of C-Vine Copula Models

According to Figure 6.2, it is extremely recommended to select the bivariate models $c_{j,k|\boldsymbol{I}}$ so that the dependence level of the bivariate conditional distribution between j and k is high while the length of conditioning variable $\boldsymbol{I}$ is small. To this end, a data driven sequential approach is proposed as follows [161-163].

Algorithm 1: Sequential C-Vine model selection

Input: Receive the d-dimensional dataset $\boldsymbol{X}$ with N observation $\boldsymbol{X} = [x_1, \dots, x_d]$, $x_i = [x_{i1}, \dots, x_{iN}]$, and set $k = 1$.

Step 1: Calculate $\hat{S}_i = \left(\sum_{j=k}^{d} |\hat{\tau}_{ij}|\right)\Big|_{i=k}^{d}$ where $\hat{\tau}_{ii} = 1$ and $\hat{\tau}_{ij}$ is the empirical pairwise Kendall's value of x_i and x_j.

Step 2: Find index $i_*^{(k)}$ by which $\hat{S}_i$ has the maximum value.

Step 3: Set $i_*^{(k)} = k$, and reorder the all variables.

Step 4: Select bivariate copula $c_{i_*^{(k)}, j+k|(i_*^{(1)},\dots,i_*^{(k-1)})}\Big|_{j=1}^{d-k}$ and corresponding parameters using **Algorithm 2**.

Step 5: Calculate conditional distributions $F_{j+k|(i_*^{(1)},\dots,i_*^{(k)})}\Big|_{j=1}^{d-k}$.

Step 6: Set $k = k + 1$ and go to **Step 1**.

In Step 4, it is required to choose a bivariate copula family for each edge of the tree. This step is an efficient procedure since it is just a 2-D sequential copula selection for modeling a high-dimensional case. Before performing the copula selection process, a hypothesis test for the independence of two random variables x_i and x_j is done based on the value of $\hat{\tau}_{ij}$. If there is not enough evidence to reject the null hypothesis, these two variables will be considered as independent variables, and the independence copula without any parameter is used. If null hypothesis is rejected, the most proper copula model should be selected among a set of given copula models. There are different tests for copula selection such as copula goodness-of-fit, Cramer-von-Mises test, likelihood ratio-based tests, and Akaike information criterion (AIC) test [161-163]. In this study, AIC test is used for model selection in Algorithm 2. AIC criterion is shown in (6.16) where k is the number of copula parameters and $\widehat{\boldsymbol{\theta}}$ denotes the estimated parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_k]^{\mathrm{T}}$. The copula parameters are estimated by maximum likelihood estimation (MLE) shown by (6.15).

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \ell(\boldsymbol{\theta}; u_i, u_j) = \sum_{r=1}^{n} \ln\left(c(u_{ri}, u_{rj}; \boldsymbol{\theta})\right)/n \right\} \tag{6.15}$$

$$AIC = -2 \sum_{r=1}^{n} \ln\left(c(u_{ri}, u_{rj}; \widehat{\boldsymbol{\theta}})\right) + 2k \tag{6.16}$$

Following the abovementioned points, Algorithm 2 is proposed here to optimally find the parameters and the copula models for each edge of C-Vine trees.

Algorithm 2: Bivariate copula model and parameter selection

Input: Receive a set of copula models and n observations of u_i and u_j which are the counterparts of x_i and x_j in uniform domain.

Step 1: Choose the first bivariate copula model $c(\cdot)$ from the list, e.g., Gaussian, *t*-student, Joe, etc. and obtain corresponding parameters $\widehat{\boldsymbol{\theta}}$ using (6.15).

Step 2: Calculate AIC criterion using (6.16).

Step 3: Repeat steps 1 and 2 for all copula models in the list.

Step 4: Find the fittest copula model for DS modeling of x_i and x_j by which AIC is minimized.

6.3 Robust Optimization Incorporating C-Vine Copula in High Dimensional Cases

In the developed conventional optimization models, it is assumed that the input data are exactly known and take on some deterministic or nominal values. However, the uncertainty of data cannot be considered in the deterministic optimization models; consequently, the feasibility of the model for uncertain input data is controversial. This is because some constraints might be violated if the input data are different from the predetermined nominal values. Therefore, it is strongly recommended to use the models that are *robust* to data uncertainty and give feasible solutions for a certain range of input data. The primary efforts for the development of linear robust models by [164] and non-linear models by [165] led to either too conservative solution or computationally inefficient methodology, respectively. These drawbacks are alleviated in [164] by proposing a linear robust optimization methodology in which the conservativeness of the solution can be easily regulated by controlling the conservatism degree on each constraint. The robust optimization model in [164] is extremely efficient in cases with high-dimensional uncertainty sets where the computational efficiency matters. In the next sections, based on the developed model in [164], an adjustable robust energy and reserve dispatch is proposed while a high-dimensional wind generation uncertainty set is adopted by C-Vine copula-based DS modeling.

6.3.1 Adjustable Robust Optimization

Assume that the cost function $\boldsymbol{g}(\boldsymbol{z})$ in (6.17) is maximized subject to a set of inequality and equality constraints (6.18)-(6.21) where $\boldsymbol{z}_{\mathcal{M}\times 1}$ is the problem decision variables, the coefficient vector $\tilde{A}_i$, containing uncertain input data, beside B_i constitute the i^{th} row of the matrices $\widetilde{\mathbf{A}}_{\mathcal{I}\times\mathcal{M}}$ and $\mathbf{B}_{\mathcal{I}\times 1}$, respectively, and the vectors C_k and D_k are k^{th} row of two matrices $\mathbf{C}_{\mathcal{K}\times\mathcal{M}}$ and $\mathbf{D}_{\mathcal{K}\times 1}$ that create the problem inequality constraints without uncertainty. Also, E_ℓ and H_ℓ reflect the ℓ^{th} row of $\mathbf{E}_{\mathcal{L}\times\mathcal{M}}$ and $\mathbf{H}_{\mathcal{L}\times 1}$ respectively to form the equality constraints. For each vector $\tilde{A}_i$, the set of uncertain coefficients $\tilde{a}_{ij}$ is shown with J_i and $\tilde{a}_{ij}$, $j \in J_i$ can have any probability distribution with expected value $\bar{a}_{ij}$ as nominal value over the interval $[a_{ij}^{LB}, a_{ij}^{UB}]$.

$$\textbf{Max. } \boldsymbol{g(z)} \tag{6.17}$$

$$\tilde{A}_i \boldsymbol{z} \leq B_i \qquad \forall\, i = 1, \dots, \mathcal{I} \tag{6.18}$$

$$C_k \boldsymbol{z} \leq D_k \qquad \forall k = 1, \dots, \mathcal{K} \tag{6.19}$$

$$E_\ell \boldsymbol{z} = H_\ell \qquad \forall\, \ell = 1, \dots, \mathcal{L} \tag{6.20}$$

$$\underline{z}_m \leq z_m \leq \overline{z}_m \quad \forall m = 1, \dots, \mathcal{M} \tag{6.21}$$

In order to adjust the robustness of the solution, the parameter Γ_i ($0 \leq \Gamma_i \leq |J_i|$), which is not necessarily an integer, is defined for every uncertain constraint i. Since it might be improbable that all uncertain parameters $\tilde{a}_{ij}$, $j \in J_i$ change simultaneously, just the uncertainty of up to $\lfloor \Gamma_i \rfloor$ coefficients are considered, and one coefficient a_{it} is assumed to change by $(\Gamma_i - \lfloor \Gamma_i \rfloor)\hat{a}_{it}$. The remaining uncertain coefficients are set on the corresponding expected values. So, based on this methodology, if there are less then $\lfloor \Gamma_i \rfloor$ uncertain coefficients that change in real situations, the robust solution is certainly feasible. Moreover, in cases with more than $\lfloor \Gamma_i \rfloor$ changing coefficients, the problem has a robust solution with high probability of feasibility. In order to obtain the robust counterpart of the main problem, the constraint (6.18) is converted to the constraints (6.22)-(6.26), and a LP problem is made as shown in [164]. In constraint (6.24), $\hat{a}_{ij}$ reflects the maximum deviation of the expected value of $\tilde{a}_{ij}$ from the lower/upper bounds of $\tilde{a}_{ij}$.

$$\sum_{m=1}^{\mathcal{M}} \bar{a}_{ij}\, z_m + \Gamma_i\, w_i + \sum_{j \in J_i} p_{ij} \leq B_i \quad \forall\, i = 1, \dots, \mathcal{I} \tag{6.22}$$

$$-y_m \leq z_m \leq y_m \;, \quad y_m \geq 0 \qquad \forall\, m = 1, \dots, \mathcal{M} \tag{6.23}$$

$$w_i + p_{ij} \geq \hat{a}_{ij}\, y_j \;, \quad p_{ij} \geq 0 \qquad \forall\, i, j \in J_i \tag{6.24}$$

$$w_i \geq 0 \qquad \forall\, i = 1, \dots, \mathcal{I} \tag{6.25}$$

$$\hat{a}_{ij} = \max\,(a_{ij}^{UB} - \bar{a}_{ij}, a_{ij}^{LB} - \bar{a}_{ij}) \tag{6.26}$$

6.3.2 Solution Robustness and Constraint Violation Probability

Since in the utilized constraint conversion, some of uncertain coefficients are supposed to take the expected values, the constraint (6.18) might be violated with the optimal solution of the problem, i.e., $\boldsymbol{z}^*$, but authors in [164] proved that the probability of violation, that depends on the value of Γ_i and $|J_i|$, is small and can be obtained by (6.27)-(6.30) where $\theta = (\Gamma_i + |J_i|)/2$. In

(6.27), the term $\mathcal{P}(\Gamma_i, |J_i|)$ reflects the violation probability bound of i^{th} constraint that is subject to uncertainty.

$$\mathbb{P}\mathrm{r}\{\tilde{A}_i \boldsymbol{z}^* > B_i\} \leq \mathcal{P}(\Gamma_i, |J_i|) \tag{6.27}$$

$$\mathcal{P}(\Gamma_i, |J_i|) \leq (1-\theta+\lfloor\theta\rfloor)\mathcal{G}(|J_i|, \lfloor\theta\rfloor) + \sum_{r=\lfloor\theta\rfloor+1}^{|J_i|} \mathcal{G}(|J_i|, r) \tag{6.28}$$

$$\mathcal{G}(|J_i|, r) = \begin{cases} 1/2^{|J_i|} & if\ r = |J_i|\ or\ r = 0 \\ \mathcal{Q} & otherwise \end{cases} \tag{6.29}$$

$$\mathcal{Q} = \sqrt{|J_i|/2\pi r(|J_i|-r)}\,.\,exp(|J_i|\ln(|J_i|/2(|J_i|-r)) + r\ln((|J_i|-r)/r)) \tag{6.30}$$

The abovementioned robust formulation enables the power system operator to have a sense about the level of robustness, and this great capability can cause a trade-off between the degree of conservatism and the probability of constraints violation. In high-dimensional cases where there are many uncertain coefficients $\tilde{a}_{ij}$ (e.g., $|J_i| > 10$) with a special dependence structure, there is a well-known challenge called *curse of dimensionality*. Considering the high-dimensional dependencies among $\tilde{a}_{ij}$, the performance of RO approaches for optimal operation of power systems will be adversely affected if the used RO approaches cannot efficiently handle the curse of dimensionality. The C-Vine copula-based DS modeling proposed in Section 6.2.3 along with the abovementioned RO methodology can handle such high-dimensional dependencies. This is done through efficient conditional probability distribution modeling of $\tilde{a}_{ij}$ shown by $F(\tilde{a}_{ij}|\tilde{a}_{is})$, $\boldsymbol{s} = \{1, \dots, |J_i|\} - \{j\}$ and controlling the robustness of the solution using handling a subset of uncertain factors. In expression (6.26), a_{ij}^{UB}, a_{ij}^{LB}, and $\bar{a}_{ij}$ are acquired by (6.31) and (6.32).

$$a_{ij}^{UB} = F^{-1}_{\tilde{a}_{ij}|\tilde{a}_{is}}(CL)\ ,\ a_{ij}^{LB} = F^{-1}_{\tilde{a}_{ij}|\tilde{a}_{is}}(1-CL) \tag{6.31}$$

$$\bar{a}_{ij} = \int x\, f(x|\tilde{a}_{is})dx \tag{6.32}$$

6.3.3 Adjustable Robust Energy and Reserve Dispatch with Large Scale Wind Power Generation

In this chapter, the cost function of the RO problem is shown by (6.33) where the first and second terms are related to energy and reserve costs of conventional power plants (CPPs), respectively. The third term which shows the cost related to the uncertainty of wind power generation originates

from the deviation of allowable upper and lower bounds from the corresponding prediction intervals. The power balance constraint and CPPs output limits are shown by (6.34)-(6.35). Also, the constraints (6.36)-(6.38) reflect the CPPs' ramping rate considering the upward and downward reserve levels. In addition, the constraint (6.39) expresses the relation between the allocated and the deployed reserve capacities for each CPP unit. The constraint (6.40) implies that the scheduled generation of each wind farm at each hour ($P_{j,t}^{w}$), should be inside the corresponding allowable lower ($\underline{\hat{p}}_{j,t}^{w}$) and upper ($\hat{\overline{p}}_{j,t}^{w}$) bounds. Furthermore, the relation between prediction intervals and allowable intervals are expressed by (6.41) to make a feasible solution. The three most important sets of constraints in this chapter that are subject to uncertainty are shown by (6.42)-(6.44). These constraints enable the operator to prevent the insecurity on transmission lines and load shedding and wind curtailment events, respectively. The uncertain input parameter $\tilde{P}_{j,t}^{w}$ as the representative of all possible scenarios for the output of wind farm j at hour t is determined by the C-Vine copula-based DS modeling in the previous section. Importantly, the auxiliary decision variable $\mathcal{g}_{j,t}^{w}$ in (6.42)-(6.44) is defined for each WF to be able to reformulate the uncertain inequality constraints of the problem in the form of (6.18); therefore, based on this definition, the variable $\mathcal{g}_{j,t}^{w}$ should equal one to avoid transmission lines overloading, load shedding, and wind curtailment.

$$\textbf{Min}.\ \sum_{t=1}^{T}\Delta t.\left(\sum_{i=1}^{NG}\left(a_i {P_{i,t}}^2+b_i P_{i,t}+c_i+c_i^{UP}R_{i,t}^{UP}+c_i^{DN}R_{i,t}^{DN}\right)+\sum_{j=1}^{NW}\left(\lambda_j\left(\left(\hat{\overline{p}}_{j,t}^{w}-p_{j,t}^{UB}\right)^2+\left(\underline{\hat{p}}_{j,t}^{w}-p_{j,t}^{LB}\right)^2\right)+c_j^{w}P_{j,t}^{w}\right)\right) \tag{6.33}$$

s.t.

$$\sum_{i=1}^{NG}P_{i,t}+\sum_{j=1}^{NW}P_{j,t}^{w}=L_t \qquad \forall t \tag{6.34}$$

$$\begin{cases}\underline{P_i}\le P_{i,t}\le \overline{P}_i\\ \underline{P_i}+R_{i,t}^{DN}\le P_{i,t}\le \overline{P}_i-R_{i,t}^{UP}\end{cases} \qquad \forall t,i \tag{6.35}$$

$$\begin{cases}P_{i,t}-P_{i,t-1}\le \overline{\Delta P_i}^{UP}.\Delta t\\ -P_{i,t}+P_{i,t-1}\le \overline{\Delta P_i}^{DN}.\Delta t\end{cases} \qquad \forall t,i \tag{6.36}$$

$$\begin{cases} P_{i,t} - P_{i,t-1} + R_{i,t}^{UP} \le \overline{\Delta P}_i^{UP} . \Delta t \\ -P_{i,t} + P_{i,t-1} + R_{i,t}^{DN} \le \overline{\Delta P}_i^{DN} . \Delta t \end{cases} \quad \forall t, i \tag{6.37}$$

$$\begin{cases} P_{i,t} - P_{i,t-1} + \Delta R_{i,t}^{UP} + \Delta R_{i,t-1}^{DN} \le \overline{\Delta P}_i^{UP} . \Delta t \\ -P_{i,t} + P_{i,t-1} + \Delta R_{i,t-1}^{UP} + \Delta R_{i,t}^{DN} \le \overline{\Delta P}_i^{DN} . \Delta t \end{cases} \quad \forall t, i \tag{6.38}$$

$$\begin{cases} 0 \le \Delta R_{i,t}^{UP} \le R_{i,t}^{UP} \\ 0 \le \Delta R_{i,t}^{DN} \le R_{i,t}^{DN} \end{cases} \quad \forall t, i \tag{6.39}$$

$$\underline{\hat{p}}_{j,t}^{w} \le P_{j,t}^{w} \le \hat{\overline{p}}_{j,t}^{w} \quad \forall t, j \tag{6.40}$$

$$\begin{cases} \hat{\overline{p}}_{j,t}^{w} \le p_{j,t}^{UB} \\ \underline{\hat{p}}_{j,t}^{w} \le p_{j,t}^{LB} \end{cases} \quad \forall t, j \tag{6.41}$$

$$\underline{TL}_l \le \sum_{i=1}^{NG} K_{l,i} P_{i,t} + \sum_{j=1}^{NG} K_{l,j} \tilde{P}_{j,t}^{w} \, \mathscr{g}_{j,t}^{w} \le \overline{TL}_l \quad \forall t, l \tag{6.42}$$

$$\sum_{i=1}^{NG} P_{i,t} + \sum_{i=1}^{NG} \Delta R_{i,t}^{UP} + \sum_{j=1}^{NW} \tilde{P}_{j,t}^{w} \, \mathscr{g}_{j,t}^{w} \ge L_t \quad \forall t \tag{6.43}$$

$$\sum_{i=1}^{NG} P_{i,t} - \sum_{i=1}^{NG} \Delta R_{i,t}^{DN} + \sum_{j=1}^{NW} \tilde{P}_{j,t}^{w} \, \mathscr{g}_{j,t}^{w} \le L_t \quad \forall t \tag{6.44}$$

$$\mathscr{g}_{j,t}^{w} = 1 \quad \forall t, j \tag{6.45}$$

6.4 Case Studies

6.4.1 Test System

In this chapter, the broadly used IEEE 118-bus test system is simulated in order to indicate the advantages of the proposed RO framework for energy and reserve dispatch. There are 54 CPPs, and the system base load over a 4-hour scheduling period are 3.6, 3.7, 4, 4.2 GW. In this chapter, 20 different wind power time series as the representative of 20 wind farms are considered with the capacities shown in Table 6.1 to assess the effect of high-dimensional cases on robust dispatch. The developed C-Vine copula-based RO model is solved by CPLEX 12.6.1 using MATLAB R2016a on a Corei7-6700 CPU@3.40 GHz personal computer with 16GB RAM.

Table 6.1 Wind farms nominal capacities located in IEEE 118-bus test system.

Wind Farms	WF1	WF2	WF3	WF4	WF5	WF6	WF7	WF8	WF9	WF10
Capacities (MW)	80	50	50	100	90	50	50	120	75	50
Bus No.	9	15	20	25	30	38	42	45	55	6
Wind Farms	WF11	WF12	WF13	WF14	WF15	WF16	WF17	WF18	WF19	WF20
Capacities (MW)	50	95	70	60	50	100	80	50	50	100
Bus No.	65	70	75	80	88	95	100	105	114	117

6.4.2 Dependence Structure Modeling of Wind Farms

According to the proposed algorithms in Section 6.2.3, using a detailed exploratory data analysis, one can choose a C-Vine copula model including the structure, the bivariate copulas, and the parameters. In this study, the utilized copula models that can cover a broad range of dependence behavior are Independence copula, Gaussian, *t*-student, Joe, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, and BB8. To make the DS modeling more flexible, the rotated versions of Clayton, Gumbel, Joe, BB1, BB6, BB7, and BB8 copulas with 90 and 180 degrees are also considered. The index number of all considered copula families are shown in Table 6.2. Also, the root nodes (i.e., root WFs) which make the strongest dependencies in terms of Kendall values, are obtained and shown in Table 6.3. Using the proposed algorithms in Section 6.2.3 for C-Vine copula modeling, the bivariate copulas of sample trees T_1, T_2, T_3, T_{16}, T_{17}, T_{18}, and T_{19} are shown in Table 6.4.

Table 6.2 Index of bivariate copula families.

Copula	**Ind.**	**Gaussian**	**t-student**	**Clayton**	**Gumbel**	**Frank**	**Joe**	**BB1**
Index	0	1	2	3	4	5	6	7
Copula	**BB6**	**BB7**	**BB8**	**180° Rotated Clayton**	**180° Rotated Gumbel**	**180° Rotated Joe**	**180° Rotated BB1**	**180° Rotated BB6**
Index	8	9	10	13	14	16	17	18
Copula	**180° Rotated BB7**	**180° Rotated BB8**	**90° Rotated Clayton**	**90° Rotated Gumbel**	**90° Rotated Joe**	**90° Rotated BB1**	**90° Rotated BB6**	**90° Rotated BB7**
Index	19	20	23	24	26	27	28	29

Table 6.3 Root nodes in each tree of C-Vine copula.

Tree	$\mathbf{T_1}$	$\mathbf{T_2}$	$\mathbf{T_3}$	$\mathbf{T_4}$	$\mathbf{T_5}$	$\mathbf{T_6}$	$\mathbf{T_7}$	$\mathbf{T_8}$	$\mathbf{T_9}$	$\mathbf{T_{10}}$
Root Node	WF3	WF8	WF9	WF14	WF12	WF1	WF2	WF5	WF10	WF11
Tree	$\mathbf{T_{11}}$	$\mathbf{T_{12}}$	$\mathbf{T_{13}}$	$\mathbf{T_{14}}$	$\mathbf{T_{15}}$	$\mathbf{T_{16}}$	$\mathbf{T_{17}}$	$\mathbf{T_{18}}$	$\mathbf{T_{19}}$	
Root Node	WF15	WF16	WF17	WF18	WF19	WF20	WF4	WF6	WF7	

Table 6.4 Bivariate copulas indices of some trees in the modeled C-Vine copula.

$\mathbf{T_1}$	$\mathbf{T_2}$	$\mathbf{T_3}$	$\mathbf{T_{16}}$	$\mathbf{T_{17}}$	$\mathbf{T_{18}}$	$\mathbf{T_{19}}$
$c_{1,2}=2$	$c_{2,3\|1}=2$	$c_{3,4\|1,2}=2$	$c_{16,17\|1,..,15}=1$	$c_{17,18\|1,..,16}=1$	$c_{18,19\|1,..,17}=1$	$c_{19,20\|1,..,18}=1$
$c_{1,3}=2$	$c_{2,4\|1}=2$	$c_{3,5\|1,2}=13$	$c_{16,18\|1,..,15}=9$	$c_{17,19\|1,..,16}=2$	$c_{18,20\|1,..,17}=2$	
$c_{1,4}=1$	$c_{2,5\|1}=19$	$c_{3,6\|1,2}=1$	$c_{16,19\|1,..,15}=2$	$c_{17,20\|1,..,16}=9$		
$c_{1,5}=19$	$c_{2,6\|1}=0$	$c_{3,7\|1,2}=5$	$c_{16,20\|1,..,15}=2$			
$c_{1,6}=0$	$c_{2,7\|1}=4$	$c_{3,8\|1,2}=19$				
$c_{1,7}=4$	$c_{2,8\|1}=19$	$c_{3,9\|1,2}=2$				
$c_{1,8}=5$	$c_{2,9\|1}=1$	$c_{3,10\|1,2}=24$				
$c_{1,9}=12$	$c_{2,10\|1}=24$	$c_{3,11\|1,2}=16$				
$c_{1,10}=19$	$c_{2,11\|1}=15$	$c_{3,12\|1,2}=13$				
$c_{1,11}=2$	$c_{2,12\|1}=15$	$c_{3,13\|1,2}=1$				
$c_{1,12}=13$	$c_{2,13\|1}=7$	$c_{3,14\|1,2}=24$				
$c_{1,13}=24$	$c_{2,14\|1}=24$	$c_{3,15\|1,2}=1$				
$c_{1,14}=2$	$c_{2,15\|1}=7$	$c_{3,16\|1,2}=17$				
$c_{1,15}=7$	$c_{2,16\|1}=1$	$c_{3,17\|1,2}=18$				
$c_{1,16}=5$	$c_{2,17\|1}=18$	$c_{3,18\|1,2}=27$				
$c_{1,17}=2$	$c_{2,18\|1}=10$	$c_{3,19\|1,2}=1$				
$c_{1,18}=2$	$c_{2,19\|1}=10$	$c_{3,20\|1,2}=1$				
$c_{1,19}=10$	$c_{2,20\|1}=2$					
$c_{1,20}=2$						

6.4.3 Simulation Results of Robust Energy and Reserve Dispatch

In this section, the proposed robust energy and reserve dispatch is simulated on IEEE 118-bus test system with two different cases.

Case1: The proposed robust energy and reserve dispatch is done without considering the correlation and dependence structure among wind farms. In this case the prediction intervals, described in Chapter 5, are used to quantify the uncertainty of each wind farm.

Case 2: The proposed robust energy and reserve dispatch is simulated considering the dependence structure modeling with C-Vine copula. In this case, the lower and upper bounds of wind farms are obtained by conditional distribution modeling through C-Vine copula modeling.

Table 6.5 reflects the total cost of the IEEE 118-bus test system and the robustness of the solution by changing the value of violation probability bound $\mathcal{P}$ from 1% to 10% to run the system with 99% to 90% theoretical robustness, respectively. Table 6.5 shows that, without dependence structure modeling in Case 1, by increasing the violation probability bound $\mathcal{P}$, the system operator requests for a less robust operation; consequently, the total cost of the system decreases while decreasing the robustness of the solution. However, with dependence structure modeling with C-Vine copula in Case 2, the operator can reach more robust energy and reserve dispatch because in this case the properly modeled correlation and dependence structure of uncertainty sources lead to more realistic results. Note that, the confidence level (CL) 95% is considered in Table 6.5. Other confidence levels like 99%, however, might be considered for upper and lower bound quantification.

Table 6.5 Total cost of IEEE 118-bus test system for different values of robustness.

$\mathcal{P}$ $(CL = 95\%)$	Case 1	Solution Robustness (%)	Case 2	Solution Robustness (%)
1%	177,806	99.05	179,090	99.90
2%	177,141	98.10	178,515	99.05
3%	176,800	97.55	178,016	98.90
4%	176,501	96.00	177,798	97.80
5%	176,186	95.70	177,435	97.15
6%	175,901	94.80	177,086	96.00
7%	175,734	93.10	176,674	95.50
8%	175,568	92.35	176,230	94.80
9%	175,401	91.15	175,994	93.75
10%	175,224	90.20	175,684	93.10

The values of upward and downward reserves for each value of violation probability bound are shown in Table 6.6 where the required reserves continue to decline by decreasing the robustness or increasing the violation probability of constraints in the system. Considering correlation and dependence structure among wind farms in Case 2, the values of the required reserves are higher than Case 1.

Table 6.6 Reserve cost of IEEE 118-bus test system for different values of robustness.

$\mathcal{P}$ $(CL = 95\%)$	Case 1		Case 2	
	UP Reserve	DN Reserve	UP Reserve	DN Reserve
1%	769	8,795	863	9,977
2%	691	8,209	725	9,518
3%	649	7,908	692	9,204
4%	612	7,645	650	8,927
5%	572	7,367	617	8,605
6%	535	7,114	583	8,300
7%	513	6,966	565	7,510
8%	491	6,817	539	6,915
9%	470	6,669	507	6,778
10%	447	6,510	488	6,650

The values of conventional generation cost, reserve cost, and wind generation cost for Case 1 and Case 2 are represented in Table 6.7 for 1% violation probability and 95% confidence level. It is seen that in Case 2 the value of reserve increases by 13.4% while there are not significant changes in generation cost of conventional power plants and wind farms. Similarly, Table 6.8 reflects the above costs for 99% confidence level and 1% violation probability. In Table 6.8, compared with Case 1, there is 11.5% increase in reserve cost of Case 2. By increasing the confidence level from 95% to 99%, the reserve cost in Case 1 increases by 15.8% while in Case 2 it increases by 13.8%.

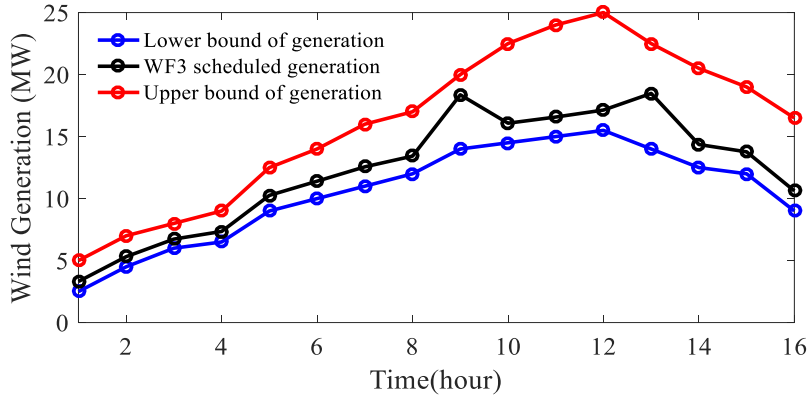
Table 6.7 Costs of conventional generation, reserve, and wind generation for IEEE 118-bus test system for $CL = 95\%$.

Cost ($) $(\mathcal{P} = 1\%)$	Case 1	Case 2
C_g	145,730	145,728
C_r	9,564	10,850
C^w	22,512	22,512
Total	**177,806**	**179,090**

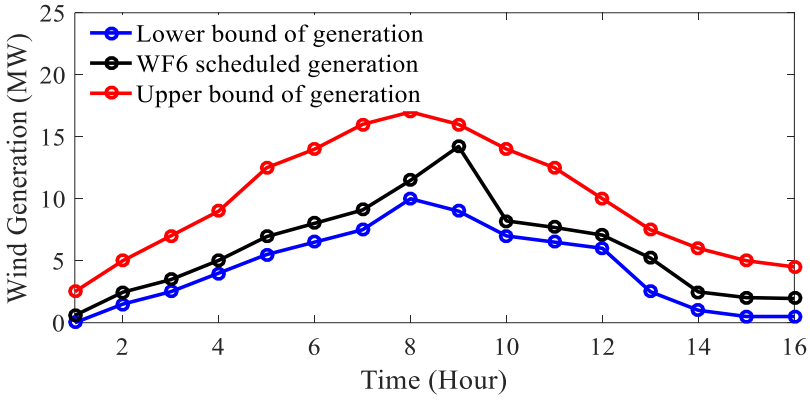
Table 6.8 Costs of conventional generation, reserve, and wind generation for IEEE 118-bus test system for $CL = 99\%$.

Cost ($) ($\mathcal{P} = 1\%$)	Case 1	Case 2
C_g	145,750	145,755
C_r	11,078	12,354
C^w	22,513	22,520
Total	**179,341**	**180,629**

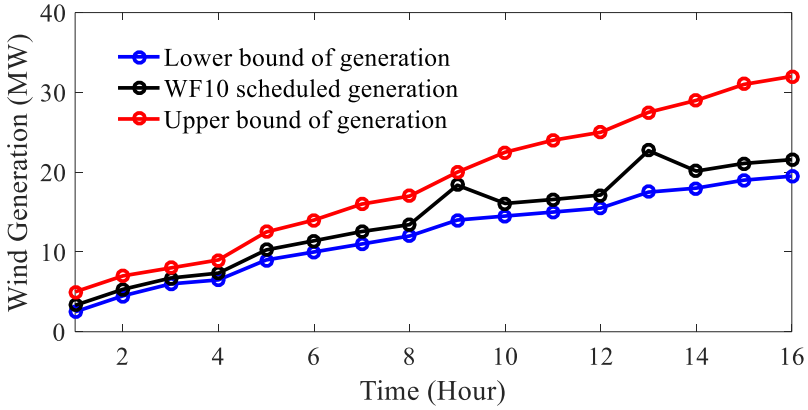
Figure 6.3 illustrates the scheduled generation, i.e., $P_{j,t}^w$, and the allowable upper (i.e., $\hat{\bar{p}}_{j,t}^w$) and lower (i.e., $\hat{\underline{p}}_{j,t}^w$) bounds for seven wind farms WF3, WF6, WF10, WF13, WF16, WF18, and WF20 over a 4-houre scheduling horizon with 15 minutes resolution. According to Figure 6.3, the proposed robust energy and reserve dispatch enforces all wind farms to generate inside the allowable interval $[\hat{\underline{p}}_{j,t}^w, \hat{\bar{p}}_{j,t}^w]$ in order to make the upper mismatch $\left(\hat{\bar{p}}_{j,t}^w - p_{j,t}^{UB}\right)$ and lower mismatch $\left(\hat{\underline{p}}_{j,t}^w - p_{j,t}^{LB}\right)$ zero.



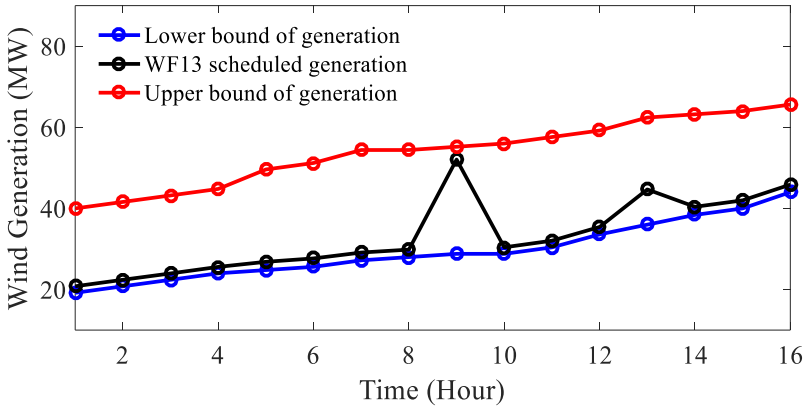
(a)



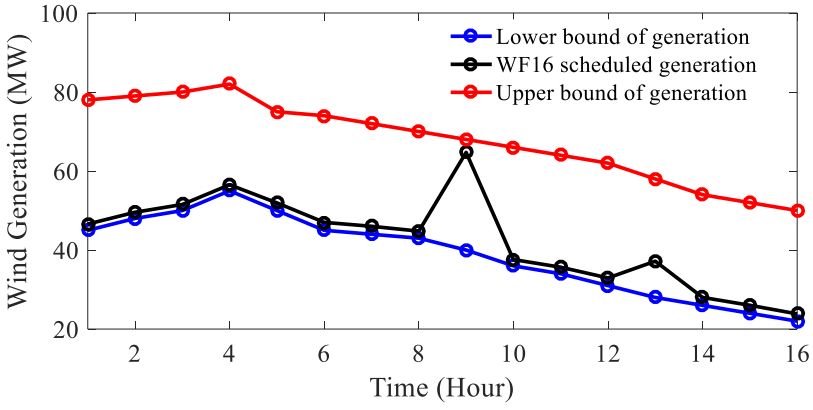
(b)


Lower bound of generation
WF10 scheduled generation
Upper bound of generation
Wind Generation (MW)
Time (Hour)
0
10
20
30
40
2
4
6
8
10
12
14
16

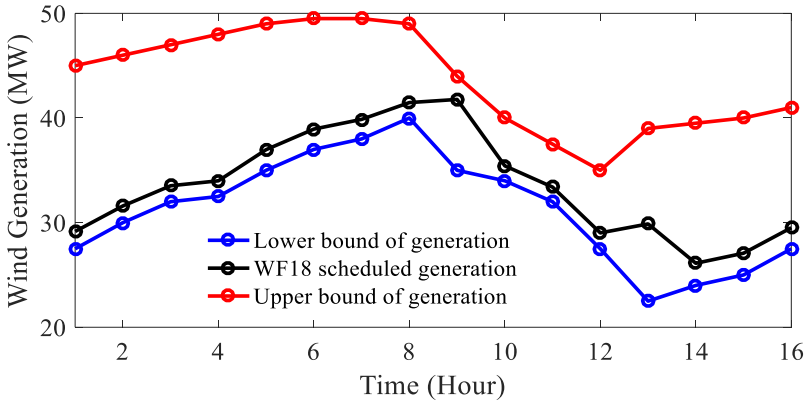
(c)


Lower bound of generation
WF13 scheduled generation
Upper bound of generation
Wind Generation (MW)
Time (Hour)
20
40
60
80
2
4
6
8
10
12
14
16

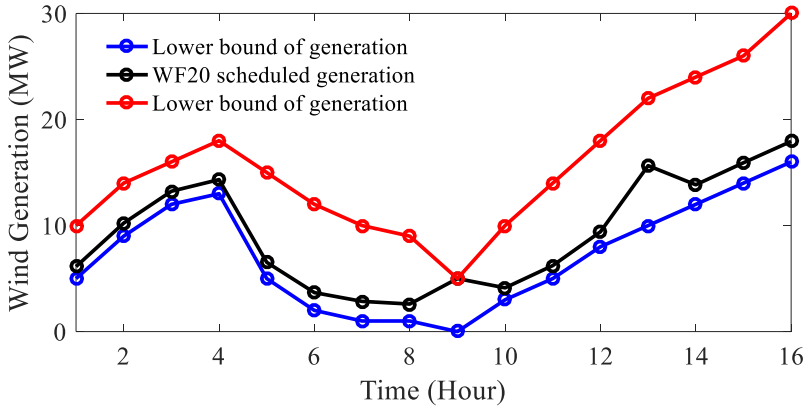
(d)


Lower bound of generation
WF16 scheduled generation
Upper bound of generation
Wind Generation (MW)
Time (Hour)
20
40
60
80
100
2
4
6
8
10
12
14
16

(e)



(f)



(g)

Figure 6.3 Scheduled wind generation and related allowable upper and lower bounds for (a) WF3, (b) WF6, (c) WF10, (d) WF13, (e) WF16, (f) WF18, (g) WF20.

6.5 Summary

Since the increasing penetration of intermittent and correlated wind energy sources makes the operation of today's power systems challenging, efficient wind generation dependence structure (DS) modeling, as a big challenge, is an essential task for uncertainty handling in optimal operation of power systems containing numerous geographically diverse and correlated wind farms. This chapter proposed a robust energy and reserve dispatch using efficient DS modeling of wind farms through C-Vine copula. Efficient DS modeling and its incorporation in generation and reserve scheduling can prevent wind curtailment and load shedding events. The proposed wind generation DS modeling uses canonical vine (C-Vine) copula as a flexible and accurate statistical model to obtain both joint and conditional distributions of any number of wind farms. Then, the results of the proposed C-Vine copula-based DS modeling is used as effective inputs for the proposed

adjustable robust economic energy and reserve dispatch problem. A case study of 20 wind farms was used to show the effectiveness of the proposed framework through comprehensive simulations on IEEE 118-bus test system. The results show that, based on the preference of the power system operators, one can use efficient DS modeling to properly dispatch the system and reduced the probability of wind curtailment and load shedding events. Using the proposed robust framework, the robustness of the solution is adjustable by two degrees of freedom, (i) adjusting a violation probability bound and (ii) adjusting the confidence level considered for wind generation upper and lower bound estimation. As future works, in the proposed robust framework, the uncertainty of solar generation and electricity load as well as the dependence structure among these sources of uncertainty can be efficiently considered. Without loss of generality, the proposed robust model can be solved with SLP or QP, or it can be linearized and solved with MILP solvers.

Chapter 7

Conclusion and Suggestions for Future Works

7.1 Conclusion

In this project, comprehensive wind power generation uncertainty modeling and stochastic optimization frameworks are developed for large scale wind power integrated power systems. In the first stage, conditional probability distribution modeling of wind power is examined considering different parametric and nonparametric approaches. Also, the application of kernel density estimation (KDE) and optimal bandwidth (BW) selection are examined in this context. It is found that non-parametric approaches, in which efficient BW selection techniques such as *advanced plug-in* (API) and diffusion estimation (DiE) are used, outperformed parametric ones because they could efficiently extract the main features of wind power probability distribution such as long tail, high skewness, and multimodality. Also, a new non-parametric model has been introduced in this project based on a mixture of beta kernels with optimal values of beta parameters, i.e., beta kernel density representation (BKDR) technique. The simulation results showed that the performance of the proposed model is superior than other benchmarks.

In the second stage, the application of the proposed conditional probability distribution modeling is shown in a stochastic optimization-based look ahead economic dispatch (ED). Chance-constrained stochastic optimization (CCSO) as one of the most well-known stochastic optimization approaches is used for optimal energy and reserve dispatch of IEEE 118-bus test system. It should be mentioned that one of the main advantages of the proposed BKDR technique is that using a new CCED cost function (NCF) it enables the operator to convert the sequential linear programming (SLP) of conventional CCED problems to MILP models and solve through off-the-shelf mathematical programming solvers more efficiently.

In the third stage of the project, a hybrid probabilistic wind power prediction model is proposed to construct high-quality prediction intervals (PIs) based on improved complete ensemble

empirical mode decomposition with adaptive noise (ICEEMDAN) technique, extreme learning machine (ELM) and 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* (DPI) as a well-known BW 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. Although the proposed hybrid approach is superior to other benchmarks, because of the high computational burden, it is not efficient in real applications. Therefore, a new fuzzy adaptive framework based on DiE is proposed for the construction of high-quality PIs for wind power generation. The proposed frameworks, compared to the existing benchmarks, could create highly sharp and reliable prediction intervals for one hour ahead with an arbitrary resolution with high computational efficiency. The construction of high-quality prediction intervals is of high importance because they might be used for robust optimization in power systems. However, the proposed prediction interval construction is a kind of direct prediction of historical quantiles without considering the correlation among other wind farms. Note that, by ignoring the correlation and dependence structure among the existing uncertainty sources, the actual uncertainty might be underestimated and lead to higher operation cost, wind curtailment, and load shedding events.

Therefore, in the final stage of the project, the correlation and dependence structure among wind farms have been considered using C-Vine copula modeling, and the constructed prediction intervals are applied for robust energy and reserve dispatch. The results showed that compared with the benchmarks in which the correlation is ignored, the solution of energy and reserve dispatch is more robust.

7.2 Suggestion for Future Works

For future extension of this study, the following research works are recommended:

- For comprehensive assessment of uncertainties in power systems, the uncertainty of solar generation and electricity load can also be efficiently considered in the proposed MILP-based CCED model. The proposed NCF and MILP-based CCED model can be formed

after probability distribution modeling of correlated uncertain resources such as wind power, solar power, and electricity load.

- The application of other Vine copula models such as D-Vine copula and R-Vine copula models can be evaluated for high-dimensional dependence structure modeling in power systems economic dispatch, unit commitment, security assessment, and reliability.
- Since wind farms might have different levels of uncertainty, certain penalty factors and chance constraints might be defined for individual wind farms owning by different producers. Therefore, the required reserve for managing the underestimation and overestimation of each wind farm can be efficiently determined proportional to its level of uncertainty without calculating the aggregate wind power generation.
- Distributionally-robust chance constrained (DRCC) optimization approaches can be developed by the proposed MILP-based model for look-ahead CCED problems. Instead of accurate wind power distribution modeling, by defining an ambiguity set for uncertainty of wind power generation, a family of simple distributions are covered to compensate inaccurate distribution modeling in DRCC.
- For further improvement of the performance of the proposed FADiE framework for PI construction, some techniques can be incorporated to 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 prediction model for longer prediction horizons.
- The proposed FADiE framework can be extended to longer prediction horizons, e.g., 6-hour ahead, 12-hour ahead, etc. using the application of deep learning techniques such as deep Boltzmann machine (DBM), long short-term memory (LSTM), convolutional neural networks (CNN), and deep reinforcement learning (DRL).

List of Publications

Peer-Reviewed Journal Papers in This Thesis

[1] **Benyamin Khorramdel**, C. Y. Chung, N. Safari, and G. C. D. Price, "A fuzzy adaptive probabilistic wind power prediction framework using diffusion kernel density estimators" ***IEEE Trans. Power Systems, vol. 33, no. 6, pp. 7109-7121, Nov. 2018***.

[2] **Benyamin Khorramdel,** A. Zare, C. Y. Chung, and P. Gavriliadis, "A generic convex model for a chance-constrained look-ahead economic dispatch problem incorporating an efficient wind power distribution modeling" ***IEEE Trans. Power Systems, Early Access, 2019.***

[3] **Benyamin Khorramdel,** Mahdi Mazhari, and C. Y. Chung, "Towards efficient wind generation dependence structure modeling for an adjustable robust energy and reserve dispatch" Under preparation to be submitted to IEEE Trans. Power Systems.

Peer-Reviewed Conference Papers in This Thesis

[4] **Benyamin Khorramdel**, Mahsa Azizi, Nima Safari, Chi Yung Chung, and Seyed Mahdi Mazhari, "A Hybrid Probabilistic Wind Power Prediction Based on An Improved Decomposition Technique and Kernel Density Estimation," *2018 IEEE PES General Meeting*, Portland, USA., 1-5.

[5] **Benyamin Khorramdel**, Hossein Khorramdel, Alireza Zare, Nima Safari, Hossein Sangrody, and Chi Yung Chung, "A Nonparametric Probability Distribution Model for Short-Term Wind Power Prediction Error," *2018 IEEE CCECE*, Quebec, Canada., 1-5 (**Best student paper award**).

Miscellaneous Peer-Reviewed Journal and Conference Papers

[6] Nima Safari, Seyed Mahdi Mazhari, **Benyamin Khorramdel**, and Chi Yung Chung, "Tidal Current and Level Uncertainty Prediction via Adaptive Linear Programming," *IEEE Transactions on Sustainable Energy, **vol. 10, no. 2, pp. 748-758, April. 2019**.*

[7] Nima Safari, **Benyamin Khorramdel**, Alireza Zare, and Chi Yung Chung, "An advanced multistage multi-step tidal current speed and direction prediction model,"*2017 IEEE Electrical Power and Energy Conference (EPEC)*, Saskatoon, SK, 2017, pp. 1-6.

[8] Nima Safari, Y. Chen, **Benyamin Khorramdel**, M. L. Ping, and Chi Yung Chung, "A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction,"*2017 IEEE Electrical Power and Energy Conference (EPEC)*, Saskatoon, SK, 2017, pp. 1-6.

[9] Alireza Zare, Chi Yung Chung, **Benyamin Khorramdel**, Nima Safari, Sherif O. Faried, "A Novel Unscented Transformation-Based Framework for Distribution Network Expansion Planning Considering Smart EV Parking Lots" *2018 IEEE CCECE,* Quebec, Canada., 1-5

Appendix A

Data Related to Generators of IEEE 118-Bus Test System

Table A.1 Conventional generators data of IEEE 118-bus test system.

Unit No.	Bus No.	Unit Cost Coefficients			P_{max} (MW)	P_{min} (MW)	Ramp Limits $\Delta P_{u\,max}$, $\Delta P_{d\,max}$ (MW/h)	Reserve Limits $r_{u\,max}$, $r_{d\,max}$ MW	UP/DN Reserve Cost Coefficients c_u,c_d ($/MWh)
		c ($)	b ($/MWh)	a ($/MW^{2}h)					
1	4	31.67	26.2438	0.069663	30	5	15	2.5	15
2	6	31.67	26.2438	0.069663	30	5	15	2.5	15
3	8	31.67	26.2438	0.069663	30	5	15	2.5	15
4	10	6.78	12.8875	0.010875	300	150	150	75	15
5	12	6.78	12.8875	0.010875	300	100	150	50	15
6	15	31.67	26.2438	0.069663	30	10	15	5	15
7	18	10.15	17.8200	0.012800	100	25	50	12.5	15
8	19	31.67	26.2438	0.069663	30	5	15	2.5	15
9	24	31.67	26.2438	0.069663	30	5	15	2.5	15
10	25	6.78	12.8875	0.010875	300	100	150	50	15
11	26	32.96	10.7600	0.003000	350	100	175	50	15
12	27	31.67	26.2438	0.069663	30	8	15	4	15
13	31	31.67	26.2438	0.069663	30	8	15	4	15
14	32	10.15	17.8200	0.012800	100	25	50	12.5	15
15	34	31.67	26.2438	0.069663	30	8	15	4	15
16	36	10.15	17.8200	0.012800	100	25	50	12.5	15
17	40	31.67	26.2438	0.069663	30	8	15	4	15

18	42	31.67	26.2438	0.069663	30	8	15	4	15
19	46	10.15	17.8200	0.012800	100	25	50	12.5	15
20	49	28	12.3299	0.002401	250	50	125	25	15
21	54	28	12.3299	0.002401	250	50	125	25	15
22	55	10.15	17.8200	0.012800	100	25	50	12.5	15
23	56	10.15	17.8200	0.012800	100	25	50	12.5	15
24	59	39	13.2900	0.004400	200	50	100	25	15
25	61	39	13.2900	0.004400	200	50	100	25	15
26	62	10.15	17.8200	0.012800	100	25	50	12.5	15
27	65	64.16	8.3391	0.010590	420	100	210	50	15
28	66	64.16	8.3391	0.010590	420	100	210	50	15
29	69	6.78	12.8875	0.010875	300	80	150	40	15
30	70	74.33	15.4708	0.045923	80	30	40	15	15
31	72	31.67	26.2438	0.069663	30	10	15	5	15
32	73	31.67	26.2438	0.069663	30	5	15	2.5	15
33	74	17.95	37.6968	0.028302	20	5	10	2.5	15
34	76	10.15	17.8200	0.012800	100	25	50	12.5	15
35	77	10.15	17.8200	0.012800	100	25	50	12.5	15
36	80	6.78	12.8875	0.010875	300	150	150	75	15
37	82	10.15	17.8200	0.012800	100	25	50	12.5	15
38	85	31.67	26.2438	0.069663	30	10	15	5	15
39	87	32.96	10.7600	0.003000	300	100	150	50	15
40	89	6.78	12.8875	0.010875	200	50	100	25	15
41	90	17.95	37.6968	0.028302	20	8	10	4	15
42	91	58.81	22.9423	0.009774	50	20	25	10	15

43	92	6.78	12.8875	0.010875	300	100	150	50	15
44	99	6.78	12.8875	0.010875	300	100	150	50	15
45	100	6.78	12.8875	0.010875	300	100	150	50	15
46	103	17.95	37.6968	0.028302	20	8	10	4	15
47	104	10.15	17.8200	0.012800	100	25	50	12.5	15
48	105	10.15	17.8200	0.012800	100	25	50	12.5	15
49	107	17.95	37.6968	0.028302	20	8	10	4	15
50	110	58.81	22.9423	0.009774	50	25	25	12.5	15
51	111	10.15	17.8200	0.012800	100	25	50	12.5	15
52	112	10.15	17.8200	0.012800	100	25	50	12.5	15
53	113	10.15	17.8200	0.012800	100	25	50	12.5	15
54	116	58.81	22.9423	0.009774	50	25	25	12.5	15

Appendix B

Copyright Permission Letters from Co-Authors

To Whom It May Concern:

I, Chi Yung Chung, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled "Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems".

1. Benyamin Khorramdel, A. Zare, **C. Y. Chung**, and P. Gavriliadis, "A generic convex model for a chance-constrained look-ahead economic dispatch problem incorporating an efficient wind power distribution modeling" ***IEEE Trans. Power Systems, Under 4th Review.***
2. Benyamin Khorramdel, **C. Y. Chung**, N. Safari, and G. C. D. Price, "A fuzzy adaptive probabilistic wind power prediction framework using diffusion kernel density estimators" ***IEEE Trans. Power Systems, vol. 33, no. 6, pp. 7109-7121, Nov. 2018***.
3. Benyamin Khorramdel, Mahdi Mazhari, and **C. Y. Chung,** "Towards efficient wind generation dependence structure modeling for an adjustable robust energy and reserve dispatch" Under preparation to be submitted to IEEE Trans. Power Systems.
4. Benyamin Khorramdel, Mahsa Azizi, Nima Safari, **C. Y. Chung**, and Seyed Mahdi Mazhari, "A Hybrid Probabilistic Wind Power Prediction Based on An Improved Decomposition Technique and Kernel Density Estimation," *2018 IEEE PES General Meeting,* Portland, USA., 1-5.
5. Benyamin Khorramdel, Hossein Khorramdel, Alireza Zare, Nima Safari, Hossein Sangrody, and **C. Y. Chung**, "A Nonparametric Probability Distribution Model for Short-Term Wind Power Prediction Error," *2018 IEEE CCECE,* Quebec, Canada., 1-5 (**Best student paper award**).

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, Alireza Zare, hereby grant permission to Mr. Benyamin Khorramdel to include the following paper in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, **A. Zare**, C. Y. Chung, and P. Gavriliadis, “A generic convex model for a chance-constrained look-ahead economic dispatch problem incorporating an efficient wind power distribution modeling” ***IEEE Trans. Power Systems, Under 4th Review.***
2. Benyamin Khorramdel, Hossein Khorramdel, **Alireza Zare**, Nima Safari, Hossein Sangrody, and C. Y. Chung, “A Nonparametric Probability Distribution Model for Short-Term Wind Power Prediction Error,” *2018 IEEE CCECE,* Quebec, Canada., 1-5 (**Best student paper award**).

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, Nima Safari, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, C. Y. Chung, **N. Safari**, and G. C. D. Price, “A fuzzy adaptive probabilistic wind power prediction framework using diffusion kernel density estimators” ***IEEE Trans. Power Systems, vol. 33, no. 6, pp. 7109-7121, Nov. 2018***.
2. Benyamin Khorramdel, Mahsa Azizi, **N. Safari**, C. Y. Chung, and Seyed Mahdi Mazhari, “A Hybrid Probabilistic Wind Power Prediction Based on An Improved Decomposition Technique and Kernel Density Estimation,” *2018 IEEE PES General Meeting,* Portland, USA., 1-5.
3. Benyamin Khorramdel, Hossein Khorramdel, Alireza Zare, **N. Safari**, Hossein Sangrody, and C. Y. Chung, “A Nonparametric Probability Distribution Model for Short-Term Wind Power Prediction Error,” *2018 IEEE CCECE,* Quebec, Canada., 1-5 (**Best student paper award**).

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, G. C. D. Price, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, C. Y. Chung, N. Safari, and **G. C. D. Price**, “A fuzzy adaptive probabilistic wind power prediction framework using diffusion kernel density estimators” ***IEEE Trans. Power Systems, vol. 33, no. 6, pp. 7109-7121, Nov. 2018***.

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, Panagiotis Gavriliadis, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, A. Zare, C. Y. Chung, and **P. Gavriliadis**, “A generic convex model for a chance-constrained look-ahead economic dispatch problem incorporating an efficient wind power distribution modeling” ***IEEE Trans. Power Systems, Under 4th Review.***

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, Mahsa Azizi, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, **Mahsa Azizi,** N. Safari, C. Y. Chung, and Seyed Mahdi Mazhari, “A Hybrid Probabilistic Wind Power Prediction Based on An Improved Decomposition Technique and Kernel Density Estimation,” *2018 IEEE PES General Meeting,* Portland, USA., 1-5.

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, Hossein Khorramdel, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, **Hossein Khorramdel,** Alireza Zare, N. Safari, Hossein Sangrody, and C. Y. Chung, “A Nonparametric Probability Distribution Model for Short-Term Wind Power Prediction Error,” *2018 IEEE CCECE,* Quebec, Canada., 1-5 (**Best student paper award**).

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, Seyed Mahdi Mazhari, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, **Seyed Mahdi Mazhari**, and C. Y. Chung, “Towards efficient wind generation dependence structure modeling for an adjustable robust energy and reserve dispatch” Under preparation to be submitted to IEEE Trans. Power Systems.
2. Benyamin Khorramdel, Mahsa Azizi, N. Safari, C. Y. Chung, and **Seyed Mahdi Mazhari,** “A Hybrid Probabilistic Wind Power Prediction Based on An Improved Decomposition Technique and Kernel Density Estimation,” *2018 IEEE PES General Meeting,* Portland, USA., 1-5.

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

To Whom It May Concern:

I, Hossein Sangrody, hereby grant permission to Mr. Benyamin Khorramdel to include the following papers in his thesis titled “Wind Power Probabilistic Prediction and Uncertainty Modeling for Operation of Large-Scale Power Systems”.

1. Benyamin Khorramdel, Hossein Khorramdel, Alireza Zare, N. Safari, **Hossein Sangrody**, and C. Y. Chung, “A Nonparametric Probability Distribution Model for Short-Term Wind Power Prediction Error,” *2018 IEEE CCECE,* Quebec, Canada., 1-5 (**Best student paper award**).

I am aware that all University of Saskatchewan theses are also posted in the digital USask eCommons thesis repository, making the thesis openly available on the internet.

Date:

Signature:

References

[1] UN. World population prospects: the 2008 revision, highlights. New York: United Nations, Department of Economic and Social Affairs, Population Divi- sion; 2009.

[2] M. Li, "World Energy 2017-2050: Annual Report", Department of Economics, University of Utah, June 2017.

[3] G. Giebel, G. Kariniotakis, R. Brownsword, "The state of the art in short-term prediction of wind power – from a Danish perspective", In Proceedings of the 4th international workshop on large-scale integration of wind power and transmission networks for offshore wind farms. Billund, Denmark, 20–21 October 2003.

[4] J. Juban, L. Fugon, G. Kariniotakis, "Uncertainty estimation of wind power forecasts: comparison of probabilistic modelling approaches", European wind energy conference & exhibition EWEC 2008. Brussels: Belgium; 2008.

[5] 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.

[6] M. Lange, "Analysis of the uncertainty of wind power predictions", [Ph.D. dissertation]. Wuppertal: Carl von Ossietzky Oldenburg University; 2003.

[7] E.D. Castronuovo and J.A.P. Lopes, "On the optimization of the daily operation of a wind-hydro power plant", *IEEE Trans. Power Syst.*, vol 19, no. 3, 1599–1606, Aug. 2004.

[8] J. Yan, Y. Q. Liu, S. Han, Y. Yang, "An integration of enhanced wind power interval forecasting into reactive power dispatching", IET 2nd renewable power generation conference. Beijing China, pp. 9–11, Sep. 2013.

[9] P. Pinson, G. Papaefthymiou, B. Klöckl, HA. Nielsen, "Generation of statistical scenarios of short-term wind power production", Power Tech, 2007 IEEE Lausanne, 1–5, July 2007.

[10] P. Pinson, J. K. Møller, H. A. Nielsen, H. Madsen, "Evaluation of nonparametric probabilistic forecasts of wind power", [Informatics and mathematical modeling technical report IMM-2007-02]. Technical University of Denmark; 16 January 2007.

[11] P. Pinson, H. A. Nielsen, H. Madsen, "Methods for the Estimation of the Uncertainty of Wind

Power Forecasts", March 23, 2007.

[12] Y. Zhang, J. Wang, X. Wang, "Review on probabilistic forecasting of wind power generation", *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255-270, April 2014.

[13] J. Wang, M. Shahidehpour, Z. Li, "Security-constrained unit commitment with volatile wind power generation", *IEEE Trans Power Syst.* vol. 23, no. 3, pp. 1319–1327, Aug. 2008.

[14] J. Wang, et al., "Wind power forecasting uncertainty and unit commitment", *Applied Energy*, vol. 88, no. 11, pp. 4014–4023, Nov. 2011.

[15] A. Botterud, et al., "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: a case study of illinois", *IEEE Trans. Sustain Energy*; vol. 4, no. 1, pp. 250–261, Jan. 2013

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

[17] R. J. Bessa, et al., "Reserve setting and steady-state security assessment using wind power uncertainty forecast:a case study", *IEEE Trans. Sustain Energy*, vol. 3, no. 4, pp. 827–836, Oct. 2012.

[18] N. Menemenlis, M. Huneault, A. Robitaille, "Computation of dynamic operating balancing reserve for wind power integration for the time-horizon 1–48 h", *IEEE Trans. Sustain Energy*, vol. 3, no. 4, pp. 692–702, Oct. 2012

[19] H. Bludszuweit, J. A. Dominguez-Navarro, "A probabilistic method for energy storage sizing based on wind power forecast uncertainty", *IEEE Trans. Power Syst*., vol. 26, no. 3, pp. 1651–8. Aug. 2011.

[20] P. Pinson, C. Chevallier, G. N. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power", *IEEE Trans. Power Syst*., vol. 22, no. 3, pp. 1148–56 Aug. 2007.

[21] A. Botterud, J. Wang, V. Miranda, R. J. Bessa, "Wind power forecasting in U.S. electricity markets", *The Electricity Journal*, vol. 23, no. 3, pp. 71–82, April 2010.

[22] A. Botterud, et al., "Wind power trading under uncertainty in LMP markets," *IEEE Trans Power Syst*, vol. 27, pp. 894–903, 2012.

[23] M. Zugno, T. Jónsson, P. Pinson, "Trading wind energy on the basis of probabilistic forecasts both of wind generation and of market quantities," *Wind Energy*, vol.16, pp. 909–26, 2013.

[24] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, T. S. Nielsen, "Standardizing the performance evaluation of short-term wind power prediction models," *Wind Energy*, vol. 29, pp. 475–89, 2005.

[25] T. Ackermann, "Wind power in power systems," Wiley; Stockholm, Sweden; 2005.

[26] J. B. Bremnes, "Probabilistic wind power forecasts using local quantile regression," *Wind Energy*, vol. 7, pp. 47–54, 2004.

[27] P. Pinson, G. Kariniotakis, "On-line assessment of prediction risk for wind power production forecasts," *Wind Energy*, vol.7, pp.119–32, 2004.

[28] P. Pinson, H. Madsen, H. A. Nielsen, G. Papaefthymiou, B. Klockl, "From probabilistic forecasts to statistical scenarios of short-term wind power production," *Wind Energy*, vol.12, pp. 51–62, 2009.

[29] 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.

[30] C. 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, pp. 1755–1767, Mar. 2018.

[31] Kenneth Bruninx and Erik Delarue, "A statistical description of the error on wind power forecasts for probabilistic reserve sizing," *IEEE Trans. Sustain. Energy*, vol. 5, no. 3, pp. 995–1002, Jul. 2014.

[32] B. M. S. Hodge, E. G. Ela, and M. Milligan, "Characterizing and modeling wind power forecast errors from operational systems for use in wind integration planning studies," *Wind Energy*, vol. 36, no. 5, pp. 509–524, Oct. 2011.

[33] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 306–316, May 2008.

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

[35] P. Pinson, G. Kariniotakis, "Conditional prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 25, pp. 1845–56, 2010.

[36] 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.

[37] 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.

[38] 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.

[39] J. B. Bremnes, "A comparison of a few statistical models for making quantile wind power forecasts," *Wind Energy*, vol. 9, pp.3–11, 2006.

[40] R. A. Marrie, N. V. Dawson, A. Garland, "Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables," *J. C. Epidemiol*, vol. 62, pp. 511-517, 2009.

[41] N. Meinshausen, "Quantile regression forests," *Journal of Machine Learning Research*, vol. 7, pp. 983–999, 2006.

[42] H. A. Nielsen, H. Madsen, T.S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts, "*Wind Energy, vol.*9, pp. 95–108, 2006.

[43] J. K. Møller, H. A. Nielsen, H. Madsen, "Time-adaptive quantile regression," Comput. Stat. Data Anal., vol 52, pp. 1292–303, 2008.

[44] J. Juban, L. Fugon, G. Kariniotakis, "Uncertainty estimation of wind power forecasts: comparison of probabilistic modelling approaches," in: Proceedings of the European wind energy conference, Brussels, Belgium; 2008.

[45] B. W. Silverman, “Density estimation for statistics and data analysis,” Chapman and Hall, London. 1986.

[46] M. P. Wand and M. C. Jones, “Kernel smoothing,” London: Chapman and Hall, 1995.

[47] R. J. Bessa, J. Mendes, V. Miranda, A. Botterud, J. Wang, Z. Zhou, "Quantile-copula density forecast for wind power uncertainty modeling," in: Proceedings of the IEEE power tech conference, Trondheim, Norway; 2011.

[48] R. J. Bessa, V. Miranda, A. Botterud, Z. Zhou, J. Wang, "Time-adaptive quantile copula for wind power probabilistic forecasting," *Renewable Energy*, vol. 40, pp. 29–39, 2012.

[49] R. J. Bessa, J. Sumaili, V. Miranda, A. Botterud, J. Wang, E. Constantinescu, "Time- adaptive kernel density forecast: a new method for wind power uncertainty modeling. in: Proceedings of the 17th power systems computation conference, Stockholm, Sweden; 2010.

[50] 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.

[51] J. Zhang, S. Chowdhury, A. Messac, L. Castillo, "Multivariate and multimodal wind distribution model based on kernel density estimation. in: Proceedings of the ASME international conference on energy sustainability, Washington DC, USA; 2011.

[52] J. Juban, N. Siebert, G. N. Kariniotakis, "Probabilistic short-term wind power forecasting for the optimal management of wind generation. in: Proceedings of the IEEE powertech conference, Lausanne, Switzerland; 2007.

[53] J. Jeon, J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *J. Am. Stat. Assoc*., vol. 107, pp. 66–79, 2012.

[54] 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.

[55] A. Khosravi, S. Nahavandi, D. Creighton, A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE Trans Neural Netw*., vol. 22, pp. 337–46, 2011.

[56] A. Khosravi, S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *IEEE Trans Sustain Energy*, vol. 4, pp. 849–56, 2013.

[57] P. Pinson, G. Kariniotakis, "On-line assessment of prediction risk for wind power production forecasts," *Wind Energy*, vol. 7, pp. 119–32, 2004.

[58] P. Pinson, H. A. Nielsen, H. Madsen, G. Kariniotakis, "Skill forecasting from ensemble predictions of wind power," *Applied Energy*, vol. 86, pp. 1326–34, 2009.

[59] E. Holmgren, N. Siebert, G. Kariniotakis, "Wind power prediction risk indices based on numerical weather prediction ensembles," in: Proceedings of the European wind energy conference, Warsaw, Poland; 2010.

[60] X. Guan, J. Wu, P. Li, "Analyzing aggregated characteristics of distributed wind farms," in: Proceedings of the IEEE PES general meeting, San Diego, USA; 2012.

[61] S. Zhen, P. Jirutitijaroen, "Latin hypercube sampling techniques for power systems reliability analysis with renewable energy sources," *IEEE Trans. Power Syst*., vol. 26, pp. 2066–73, 2011.

[62] Z. L. Qin, W. Y. Li, X. F. Xiong, "Generation system reliability evaluation incorporating correlations of wind speeds with different distributions," *IEEE Trans. Power Syst.*, vol. 28, pp. 551–8, 2013.

[63] G. Papaefthymiou, P. Pinson, "Modeling of spatial dependence in wind power forecast uncertainty. in: Proceedings of the international conference on probabilistic methods applied to power systems," Rincon, Puerto Rico; 2008.

[64] J. Sumaili, H. Keko, V. Miranda, Z. Zhi, A. Botterud, J. Wang, "Finding representative wind power scenarios and their probabilities for stochastic models. in: Proceedings of 16th international conference on intelligent system application to power systems (ISAP); Hersonissos, Greek; 2011.

[65] K. C. Sharma, P. Jain, R. Bhakar, "Wind power scenario generation and reduction in stochastic programming framework," *Electric Power Compon. Syst.*, vol. 41, pp. 271–85, 2013.

[66] M. Kaut, S. W. Wallace, "Evaluation of scenario-generation methods for stochastic programming", *Stochastic Program.*, vol. 3, no. 2, pp. 257-271, 2007.

[67] A. Papavasiliou, S. S. Oren, R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework", *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2197-2206, Nov. 2011.

[68] L. Yang, M. He, V. Vittal, J. Zhang, "Stochastic optimization-based economic dispatch and interruptible load management with increased wind penetration", *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 730-739, Mar. 2016.

[69] M. Lubin, Y. Dvorkin, S. Backhaus, "A robust approach to chance constrained optimal power flow with renewable generation", *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3840-3849, Sep. 2016.

[70] Q. Wang, Y. Guan, J. Wang, "A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output", *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 206-215, Feb. 2012.

[71] C. Tang, *et al*., "A versatile mixture distribution and its application in economic dispatch with multiple wind farms," *IEEE Trans. Sustain. Energy*, vol. PP, no. 99, pp. 1–1, 2017.

[72] A. Zare, C. Y. Chung, J. Zhan, and S. O. Faried, "A distributionally robust chance-constrained MILP model for multistage distribution system planning with uncertain renewables and loads", *IEEE Trans. Power Systs.*, vol. 33, no. 5, pp. 5248-5262, 2018.

[73] C. Duan, W. Fang, L. Jiang, L. Yao, and J. Liu, "Distributionally robust chance-constrained approximate AC-OPF with Wasserstein metric", *IEEE Trans. Power Systs.*, vol. 33, no. 5, pp. 4924-4936, 2018.

[74] W. Wei, F. Liu, S. Mei, "Distributionally robust co-optimization of energy and reserve dispatch", *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 289-300, Jan. 2016.

[75] L. Roald, S. Misra, T. Krause, G. Andersson, "Corrective control to handle forecast uncertainty: A chance constrained optimal power flow", *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1626-1637, Mar. 2017.

[76] L. Roald, G. Andersson, "Chance-constrained AC optimal power flow: Reformulations and efficient algorithms", *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2906-2918, May 2018.

[77] R. Jiang, Y. Guan, "Data-driven chance constrained stochastic program", *Math. Program.*, vol. 158, pp. 291-327, 2016.

[78] A. L. Soyster, "Convex programming with set-inclusive constraints and applications to inexact linear programming," *Oper. Res.*, vol. 21, no. 5, pp. 1154–1157, Oct. 1973.

[79] A. Ben-Tal and A. Nemirovski, "Robust convex optimization," *Math. Oper. Res.*, vol. 23, pp. 769–805, 1998.

[80] A. Ben-Tal and A. Nemirovski, "Robust solutions of linear programming problems contaminated with uncertain data," *Math. Program.*, vol. 30, no. 3, pp. 1337–1350, May 2015.

[81] A. Lorca, X. A. Sun, "Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind", *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1702-1713, Jul. 2015.

[82] Z. Li, W. Wu, B. Zhang, 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.

[83] W. Wu, J. Chen, B. Zhang, 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.

[84] T. Zheng, J. Zhao, E. Litvinov, and F. Zhao, "Robust optimization and its application to power system operation," presented at the 2012 CIGRE, Paris, France, 2012.

[85] R. A. Jabr, “Adjustable robust OPF with renewable energy sources,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4742–4751, Nov. 2013.

[86] M. Yang, M. Q. Wang, F. L. Cheng, and W. J. Lee, “Robust economic dispatch considering automatic generation control with affine recourse process,” *Electr. Power Energy Syst.*, vol. 81, pp. 289–298, Oct. 2016.

[87] A. Lorca and X. A. Sun, “Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind,” *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1702–1713, Oct. 2014.

[88] C. Wang, F. Liu, J. Wang, and S. Mei, “Risk-based admissibility assessment of wind generation integrated into a bulk power system,” *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 188–196, Oct. 2016.

[89] C. Wang *et al.*, “Robust risk-constrained unit commitment with largescale wind generation: An adjustable uncertainty set approach,” *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 723–733, Jan. 2017.

[90] X. Dui, G. Zhu, and L. Yao, "Two-stage optimization of battery energy storage capacity to decrease wind power curtailment in grid-connected wind farms," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3296–3305, May 2018.

[91] Y. Wang, N. Zhang, C. Kang, M. Miao, R. Shi, Q. Xia, "An efficient approach to power system uncertainty analysis with high-dimensional dependencies", *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 2984–2994, May. 2018.

[92] G. Papaefthymiou, D. Kurowicka, "Using copulas for modeling stochastic dependence in power system uncertainty analysis", *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40-49, Feb. 2009.

[93] M. Sun, I. Konstantelos, G. Strbac, "C-vine copula mixture model for clustering of residential electrical load pattern data", *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2382-2393, May 2017.

[94] 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.

[95] C. Lowery, M. O'Malley, “Impact of wind forecast error statistics upon unit commitment”, *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 760-768, Oct. 2012.

[96] 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.

[97] B.-M. Hodge and M. Milligan, “Wind power forecasting error distributions over multiple timescales,” in *Proc. IEEE Power and Energy Society General Meeting*, Detroit, MI, USA, 2011.

[98] R. A. Jabr and B. C. Pal, “Intermittent wind generation in optimal power flow dispatching,” *IET Gener., Transm., Distrib.*, vol. 3, no. 1, pp. 66–74, 2009.

[99] M. Lange, “On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors,” *J. Solar Energy Eng.*, vol. 127, no. 2, pp. 177–184, Apr. 2005.

[100] H. Holttinen and M. Milligan, “Methodologies to determine operating reserves due to increased wind power,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 713–723, Oct. 2012.

[101] A. Fabbri, T. G. S. Román, J. R. Abbad, and V. H. M. Quezada, “Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market,” *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1440–1446, Aug. 2005.

[102] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33-50, Jan. 1978.

[103] R. Koenker, *Quantile Regression*: Cambridge University Press, 2005.

[104] A. C. Guidoum, “Kernel estimator and bandwidth selection for density and its derivatives the kedd package version 1.0.3”, Oct. 2015, [online] http://cran.us.r-project.org/web/packages/kedd/vignettes/kedd.pdf.

[105] 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.

[106] Pearson, K. (1895). “Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material”. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 186: 343–414.

[107] Freedman, David; Diaconis, P. (1981). “On the histogram as a density estimator: L_2 theory”. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete. 57 (4): 453–476.

[108] 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.

[109] Z. W. Wang, C. Shen, F. Liu, X. Y. Wu, C. C. Liu, F. Gao, "Chance-constrained economic dispatch with non-Gaussian correlated wind power uncertainty", *IEEE Trans. Power Syst.*, vol. 32, no. 6, Nov. 2017.

[110] N. Zhang, C. Kang, Q. Xia, and J. Liang, "Modeling conditional forecast error for wind power in generation scheduling," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1316–1324, May 2014.

[111] M. Cui, et al., "A data-driven methodology for probabilistic wind power ramp forecasting" *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1326-1338, March 2019.

[112] M. Cui, et al. "Statistical representation of wind power ramps using a generalized gaussian mixture model" *IEEE Trans. Sust. Energy*, vol. 9, no. 1, pp. 261-272.

[113] M. Lubin, Y. Dvorkin, S. Backhaus, "A robust approach to chance constrained optimal power flow with renewable generation", *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3840-3849, Sep. 2016.

[114] Y. Wang, Q. Xia, C. Kang, "Unit commitment with volatile node injections by using interval optimization", *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1705-1713, Aug. 2011.

[115] L. Che, et al., "Intra-interval security based dispatch for power systems with high wind penetration," *IEEE Trans. Power Syst*., vol. 34, no. 2, pp. 1243-55, March 2019.

[116] E. Du et al., "Managing wind power uncertainty through strategic reserve purchasing", *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2547-2559, Jul. 2017.

[117] N. Zhang, et al., "A convex model of risk-based unit commitment for day-ahead market clearing considering wind power uncertainty," *IEEE Trans. Power Syst*., vol. 30, no. 3, May 2015.

[118] P. N. Gavriliadis and G. A. Athanassoulis, "The truncated Stieltjes moment problem solved by using kernel density functions," *Journal of Computational and Applied Mathematics*, 236 (2012), 4193-4213.

[119] G. A. Athanassoulis and P. N. Gavriliadis, "The truncated Hausdorff moment problem solved by using kernel density functions," *Probabilistic Engineering Mechanics*, vol. 17, pp. 273-291, 2002.

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

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

[122] P. Kall, and J. Meyer, "Stochastic Linear Programming, Models, Theory, and Computation," Springer, 2005.

[123] 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.

[124] 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.

[125] 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.

[126] 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.

[127] 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.

[128] 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.

[129] 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.

[130] 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.

[131] 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.

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

[133] 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.

[134] 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.

[135] 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.

[136] 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.

[137] G. Li and H. D. Chiang, "Toward cost-oriented forecasting of wind power generation," *IEEE Trans. Smart Grid,* vol. 9, no. 4, pp. 2508-17, July 2018.

[138] 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.

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

[140] F. Golestaneh, P. Pinson, R. Azizipanah-Abarghooee, H.B. Gooi, "Ellipsoidal Prediction Regions for Multivariate Uncertainty Characterization" *IEEE Trans. Power Syst,* vol. 33, no. 4, pp. 4519-30, July 2018.

[141] 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.

[142] 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.

[143] 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.

[144] 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.

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

[146] 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.

[147] 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.

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

[149] 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.

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

[151] 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.

[152] 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.

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

[154] 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.

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

[156] Z. Li, W. Wu, B. Zhang, B. Wang, "Robust look-ahead power dispatch with adjustable conservativeness accommodating significant wind power integration", *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 781-90, July 2015.

[157] Q. Bian, H. Xin, Z. Wang, D. Gan, and K. P. Wong, “Distributionally robust solution to the reserve scheduling problem with partial information of wind power,” *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2822–2823, Sep. 2015.

[158] W. Xie and S. Ahmed, “Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation,” *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1860–1867, Mar. 2018.

[159] Y. Zhang, S. Shen, and J. L. Mathieu, “Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads,” *IEEE Trans. Power Syst.*, vol. 32, no. 2, pp. 1378–1388, Mar. 2017.

[160] M. Sklar, "Fonctions de Répartition À N Dimensions Et Leurs Marges," *Publ. inst. statist. univ. paris*, vol. 8, pp. 229-231, 1960.

[161] E.C. Brechmann, U. Schepsmeier, "Modeling dependence with C- and D-vine copulas: the R package CDVine," *J. Stat. Softw*., vol. 52, no. 3, pp. 1-27, 2013

[162] Aas K, Czado C, Frigessi A, Bakken H, "Pair-Copula Constructions of Multiple Dependence," *Insurance: Mathematics and Economics*, vol. 44, no. 2, 182-198, 2009.

[163] Czado C, Schepsmeier U, Min A (2012). “Maximum Likelihood Estimation of Mixed C-Vineswith Application to Exchange Rates.” *Statistical Modelling*,12(3), 229–255.

[164] D. Bertsimas, M. Sim, "The price of robustness", *Oper. Res.*, vol. 52, no. 1, pp. 35-53, Feb. 2004.

[165] A. Ben-Tal, A. Nemirovski, "Robust solutions of linear programming problems contaminated with uncertain data," *Math. Prog.* vol. 88, pp. 411–424, 2000.