Journal of Electrical Engineering & Technology
https://doi.org/10.1007/s42835-021-00863-w

ORIGINAL ARTICLE

Probabilistic Analysis To Analyze Uncertainty Incorporating Copula Theory

Bin Li[1] · Muhammad Shahzad[2] · Hafiz Mudassir Munir[3] · Asif Nawaz[4] · Nabeel Abdelhadi Mohamed Fahal[1] · Muhammad Yousaf Ali Khan[5] · Sheeraz Ahmed[6]

Received: 14 February 2021 / Revised: 13 July 2021 / Accepted: 27 July 2021

Abstract

The emerging trend of distribution generation with existing power system network leads uncertainty factor. To handle this uncertainty, it is a provocation for the power system control, planning, and operation engineers. Although there are numerous techniques to model and evaluate these uncertainties, but in this paper the integration of Copula theory with Improved Latin-hypercube Sampling (ILHS) are incorporated for Probabilistic load Flow (PLF) evaluation. In probabilistic research approaches, the dominant interest is to achieve appropriate modelling of input random variables and reduce the computational burden. To address the said problem, Copula theory is applied to execute the modelling and interaction among input random variables of the active power system network. Considering the real discrete data, the ILHS is adopted. The load flow accessibility of the power system is carefully modeled by considering the dependence and uncertainty factors. Modified IEEE 14-bus system is employed to analyze the efficiency and performance of the proposed model using active power system network. Output power of two wind energy farms situated in New Jersey are obtained for accuracy comparison. The proposed technique shows the superiority in PLF evaluation.

Keywords Copula Theory · Correlation · Modified Latin-Hypercube Sampling · Monte Carlo Simulation · Probabilistic Load Flow

1 Introduction

For the operation of future power systems renewable energy sources (RESs) provide high level of uncertainties due to its chaotic dynamics and lack of uniformity [1]. It is a vital task for power system engineers to assess the influence of uncertainties on power system control, planning, analysis and operation. Power distribution system planning with demand uncertainty is considered in [2–5], for transmission expansion planning [6].

Steady state performance is ultimate need for the power system network. To fulfill the need a highly robust method probabilistic load flow (PLF) under possible uncertainties is adopted. The main advantage of the PLF analysis is to provide the solution of line loading and bus overloading probabilities. This analysis is suitable for identifying the weak points and potential crises in power system network. In order to provide the valuable information to power systems, the requirement is to model the uncertainty with input random variable (IRV) more accurately and comprehensively. Several methodologies have put forward in literature to solve

✉ Muhammad Shahzad
shahzadpansota@hotmail.com

1 School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, People's Republic of China

2 Department of Electrical Engineering, Muhammad Nawaz Sharif University of Engineering and Technology, Multan 66000, Pakistan

3 Department of Electrical Engineering, Sukkur IBA University, Sukkur 65200, Pakistan

4 Department of Electrical Engineering, Higher Colleges of Technology, Dubai, UAE

5 Department of Electrical Engineering, Faculty of Engineering & Technology, Gomal University, D. I. Khan, Pakistan

6 Department of Computer Science, Iqra National University, Peshawar, Pakistan

problems related to correlated IRVs together with Gaussian mixture model [7, 8], convolution-based method [8–10], point estimation method [11, 12], Monte Carlo simulation (MCS) method [12–14], MCS method based on second-order design sensitivity-assisted [15] and cumulant methods [16–18]. Most of these methods only considered the linear correlation among the IRVs. To demonstrate the degree dependence Pearson's linear correlation coefficient ρ had been used. A fact is that, correlation factors in the power system are not limited to linear correlation but also depend on non-linear dependence, such as due to environmental and social impact, loads in the similar zone have a tendency to increase or decrease [19]. Also, various factors affect the IRVs for example, the control strategy of a wind energy turbine, wind speed, wind direction and power curve, these factors might be linear or non-linear relationship, i.e. strongly correlated neighbouring wind energy farms [19–21]. For obtaining the comprehensive information about the correlation between IRVs. It should be needed to get the degree of correlation as well as correlation structure of IRVs.

MCS use non-linear load flow equations without any mathematical computation and simplification of miscellaneous problems. The amount of simulation need to attain precise solution by using MCS is not dependent on system size [22]. Additional for PLF study, MCS combine with simple random sampling (SRS) is utmost common technique with purely mathematical background and has been widely employed for the analysis of power system. With a sequence of deterministic calculations this technique simulates several uncertainties to solve the required problems [23]. This is to be considered as most highly robust, flexible and accurate for PLF study when the sample size is enough high and often adopted as a touchstone for other PLF methods [24]. Though, because of a huge number of repeated simulations it suffers high computational burden [24–26]. Latin hypercube sampling (LHS) is an effective sampling method to produce the sample and this is due to the marginal cumulative distribution function (CDF) of IRV [19, 27, 28]. LHS can obtain highly accurate results as compare to SRS. However, MCS with LHS techniques always take into consideration by assuming that: a typical marginal distribution (MD) has to follow the IRV and CDF is known. Indeed, most of the RESs output and input data is not follow any marginal type of distribution. For example, wind energy farm output data is not followed any regular MD.

Copula theory (CT) is proposed by establishing the probability distribution (PD) of the correlated IRVs. This concept can obtain the whole information of IRVs with the prospect to deal with linear and non-linear correlation relationship flexible and robustly. The dependency degree among IRVs and correlation structure between different IRVs included with complete information. Due to the discrete data as input variable, ILHS proven to be very effective. The results obtained from proposed method are compared with two methods. Firstly, with MCS with Copula and SRS known as MCS-C. Secondly, with LHS with genetic algorithm and local search known as (ILHS-GA). Proposed method is tested on modified IEEE-14 bus system. The results obtained by MCS-IS is compared with MCS-C and ILHS-GA with respects to the criteria of both robustness and computational time.

The rest of the paper is organized as follows. Modelling of correlated IRVs adopted by Copula theory is presented in Sect. 2. An ILHS is in Sect. 3. MCS-IS and computational calculations is presented in Sect. 4. Section 5 present results and discussion. Conclusion is drawn in Sect. 6.

2 Formulation

2.1 Limitation of Correlation Coefficient ρ and Copula Theory

Person's linear correlation coefficient (ρ) can only measure the degree of linear correlation between IRVs that contains various advantages, but the major two are given as. (i) simple calculation (ii) suitable for elliptically and normally distribution random variables (RVs). The drawbacks are: (i) it only exists for those distribution whose standard deviation is defined (ii) it cannot solve the non-linear correlation among IRVs (iii) noninvariant under non-linear resolutely growing transformation of RVs [19].

The copula is described as " a function that pair multivariate joint CDF to its one dimensional (1D) marginal CDFs". The main function of CT is intended to separates the function of multivariate joint distribution (MJD) into its correlation structure and MD. For example, two variable case, if r_1 and r_2 are two RVs, $F_1(r_1)$ and $F_2(r_2)$ are the marginal CDFs of r_1 and r_2 respectively. $F_{12}(r_1, r_2)$ is the joint CDF. As proved in Sklar theory, a Copula function (CF) C exists and describe in Eq. (1)

$$F_{12}(r_{1,}r_2) = \mathrm{C}(F_1(r_1), F_2(\mathrm{r})) \tag{1}$$

If $F_1(r_1)$ and $F_2(r_2)$ are continuous marginal CDFs then CF is unique. If it is not continuous then it will be unique in the range of marginal CDFs. If a RV r has an invertible CDF $F(r)$ than RV, $U = F(r)$ also follow the uniform distribution in the range of [0, 1], according to Eq. (2). For example, two-variable case, if $U_1 = F_1(r_1)$ and $U_2 = F_2(r_2)$ are marginal CDFs of RVs r_1 and r_2 respectively than their joint CDF $F_{12}(U_1, U_2)$ can be obtained by using Eq. (3). In such a case, the range of the Copula function, that are joint CDF of RVs are [0, 1].

$$U \in [0,1] : P(U \le u) = p(F_w \le u) = P(r \le F^{-1}(u)) = F[F^{-1}(u)] = u \tag{2}$$

$$F_{12}(U_1, U_2) = P(U_1 \le u_1, U_2 \le u_2) = F_{12}(F_1^{-1}(u_1), F_2^{-1}(u_2)) = C(u_1, u_2) \tag{3}$$

where $F_1^{-1}(u_1) \& F_2^{-1}(u_2)$ are the inverse function of $F_1(r_1)$ & $F_2(r_2)$ respectively. The range of the inputs u_1 and u_2 are [0, 1]. Similarly, it is true for joint probability density function (PDF), if r_1 and r_2 are two RVs. $f_1(r_1)$ and $f_2(r_2)$ are the marginal CDFs of r_1 and r_2 respectively, and $f_{12}(r_1, r_2)$ is the joint PDF, following Eq. (4) can be derived by Eq. (1)

$$(r_1, r_2) = \frac{\partial^2 F_{12}(r_1, r_2)}{\partial r_1 \partial r_2} = \frac{\partial^2 C(F_1(r_1), F_2(r_2))}{\partial r_1 \partial r_2} = \frac{\partial^2 C(F_1(r_1), F_2(r_2))}{\partial F_{1(r_1)\partial F_2(r_2)}} \cdot \frac{\partial F_1(r_1)}{\partial r_1} \cdot \frac{\partial F_{2(r_2)}}{\partial r_2} \cdots = C(F_1(r_1), F_2(r_2)) f_1(r_1) f_2(r_2) \tag{4}$$

where C$(F_1(r_1), F_2(r_2))$ is the Copula density function. Equation (5) is for multivariate random vector (MRV) case.

$$f(r_1, \ldots, r_m) = C(F_1(r_1), \ldots \ldots ..., F_m(r_m)). \prod_{i=1}^{m} f_i(r_i) \tag{5}$$

where

$$C(F1(r1), F2(r2), \ldots \ldots \ldots .., Fm(rm)) = \frac{\partial^m C(F_1(r_1), \ldots ...., F_m(r_m))}{\partial F1(r1), F2(r2), \ldots \ldots \ldots .., Fm(rm)} \tag{6}$$

is the Copula density function. $f(r_1, r_2, \ldots\ldots\ldots, r_m)$ is the joint PDF of MRV $\boldsymbol{r}_{m\times 1} = [r_1, r_2, \ldots\ldots\ldots\ldots, r_m]$.

^{T}. T denotes the transpose of the vector and $F_1(r_1)$, $F_2(r_2), \ldots\ldots\ldots, F_m(r_m)$ are the marginal CDFs of RVs $r_1, r_2, \ldots \ldots\ldots, r_m$, m is the number of variables. More, information about this theory is available [29].

2.2 PDF Modelling of Correlated IRVs

Two steps are required for modelling the probability distribution function (PDF) of IRVs. First step is to attain the marginal CDFs and secondly is to choose the suitable CF. In the first step, if a RV r, follows any regular MD, then by using estimation theory its CDF $u = f(r)$ and its inverse function $r = F^{-1}(u)$ could be attain. If RV r does not pursue any common MD both empirical CDF and inverse function are implemented as marginal CDF and an inverse function due to the Law of large numbers. When the discrete data vector of RV r is given $\boldsymbol{r}_{1\times n} = [r_1, r_2 \ldots\ldots\ldots\ldots\ldots, r_n]$. The probability of it can be compute as 1/n. The empirical CDF and inverse function may be computed as.

(I) Classify the elements in descending order $\boldsymbol{r}_{1\times n} = [r_1, r_2, \ldots\ldots\ldots\ldots\ldots, r_n]$, and change the name as $\boldsymbol{r}'_{1\times n} = [r'_1, \ldots\ldots\ldots\ldots\ldots\ldots, r'_n]$.

(II) The empirical CDF $u = F_e(r)$ and inverse function $r = F^{-1}(u)$ of the RV r is obtained by Eqs. (7) and (8), respectively.

$$u = F_e(r) = \begin{cases} 0, & r < r'_1 \\ \frac{k}{n}, & r'_p \le r < r'_{p+1} (p = 1, \ldots n-1) \\ 1, & r \ge r'_n \end{cases} \tag{7}$$

$$\begin{cases} q = Round(u.n) \\ r = F_e^{-1}(u) = \begin{cases} r'_1, q = 0 \\ r'_q, q = 1, \ldots, n \end{cases} \end{cases} \tag{8}$$

Step 2, Suitable Copula function is selected. There is no fixed rule for selection of suitable Copula function. In fact, selection of suitable Copula function was and is an ongoing research area, but steps of general procedure for selection of Copula is give as:

For the selection of commonly used copula function, estimation theory is appropriate, i.e. Maximum likelihood estimation method.
To select optimal Copula function.

In this work, shortest distance method is adopted for selection of suitable Copula function that depends upon empirical Copula and Euclidean distance. It can be express as

$$C_e\left(\frac{i_1}{n}, \ldots, \frac{i_m}{n}\right) = \frac{1}{n} \sum_{i=0}^{n} I(r_{1j} \le r_1^{i1}, \ldots, r_{mj} \le r_m^{im}) \tag{9}$$

If the condition in the Eq. (9) (Bracketed) is fulfilled then it is equivalent to 1 and termed as indicator function. Otherwise, it is 0. C_e is the empirical function of $r_{m\times n} = [r_{1j}, \ldots \ldots\ldots, r_{mj}]^{T}$. $j = 1, 2, \ldots\ldots\ldots, n$, n is samples of m RVs. r_m^{im} is the order statistics and $1 \le i_1, \ldots\ldots\ldots, i_m \le n$ is the vector of I consist of variable m. The function consists of Euclidean distance between C *and* C_e could be computed as:

$$d_n(C, C_e) = \{\sum_{i_n=1}^{n} \ldots . \sum_{i_m=1}^{n} \left(C\left(\frac{i_1}{n}, \ldots \frac{i_m}{n}\right) C_e\left(\frac{i_1}{n}, \ldots, \frac{i_m}{n}\right)\right)^2 \}^{\frac{1}{2}} \tag{10}$$

In the given Eq. (10), C is the theoretical Copula function, C_e is the empirical Copula function and d_n is the Euclidean distance. If d_n is smallest value with proposed Copula function than this Copula function is considered as an optimal Copula function for certain set of data.

The random number matrix (RNM) $\boldsymbol{R}_{m\times N}$ of correlated multivariate random vector (MRV) vector $\boldsymbol{r}_{m\times 1} = [r_1, \ldots\ldots\ldots., r_m]^T f$ can be produced by defining the succeeding steps.

(I) For the selection of Copula function, optimum procedure has to adopt as designate above.

(II) Generate RNM, $U_{m\times N}=[u_{1s},\ldots\ldots, u_{ms}]^{\mathrm{T}}$. In this vector matrix, the first variable with N samples is u_{1s} and $u_{is}=[u_{i1},\ldots\ldots,u_{iN}]$, where $i=1,2,\ldots\ldots.,m$. The technique used to generate these samples is SRS.

(III) Inverse function of individual RV has to be determined that associate to $r_{m\times 1}$ vector.

(IV) RNM $R_{m\times N}=[r_{1s},\ldots\ldots\ldots.,r_{ms}]^{\mathrm{T}}$ of multivariate RV vector $r_{m\times 1}$ which preserve the MD, as well as correlation affiliation of $\mathrm{r}_{\mathrm{m}\times 1}$ vector, created by Eq. (11).

$$r_{is} = F_i^{-1}\left(u_{is}\right) \tag{11}$$

3 Improved Latin-Hypercube Sampling

As described in literature [13, 14, 25, 26], LHS has higher sampling efficiency as well as better robustness as compared to SRS techniques. LHS needs marginal CDF of invertible random variable. Actually, the historical data obtain by supervisory control data acquisition (SCADA) or wide area measurement system (WAMS) always in discrete data form. It limits the application of LHS technique in some circumferences. So, ILHS is proposed in this paper because ILHS depend upon the CDF future of IRVs. Let's say, for the invariant case, the vector of discrete data is r and formulated as $r_{1\times n}=[r_1,\ldots\ldots\ldots,r_n]$ is known. Each discrete data has the probability of 1/n. Each n/N interval has an equal dimension of the discrete data if marginal CDF $Y=F(r)$ range is divided into interval of N non-overlapping data. Due to this condition the following steps can be used to generate random number vector (RNV):

(I) Categories the vector of discrete data with $r_{\boldsymbol{1\times n}}=[r_1,\ldots\ldots\ldots,r_n]$ from small to large numbers and rename it $\boldsymbol{r}'_{\boldsymbol{1\times n}}=[r'_1,\ldots\ldots\ldots,r'_n]$.

(II) k is the position parameter of Pth $(P=1,\ldots\ldots\ldots.,N)$ of random number (RN) is to be computed and r''_p of RV r in the vector $\boldsymbol{r}_{\boldsymbol{1\times n}}=[r_1,\ldots\ldots\ldots,r_n]$ by following Eq. (12):

$$k = Round\left(P.\frac{n}{N}\right) \tag{12}$$

(III) To obtain the final RNV $\boldsymbol{r}''_{\boldsymbol{1\times n}}$, randomly arrange the vector $\boldsymbol{r}''_{\boldsymbol{1\times n}}$ that is obtained by Eq. (12).

In a similar mode, multivariate random matrix (MRM) $R_{m\times N}$ may be created by following steps:

(I) Create the RNV $\boldsymbol{r}''_{1\times n}=[r''_{11},\ldots\ldots\ldots,r''_{1N}]$, of RV r_1 in consonance with its vector of discrete data, $\boldsymbol{r_1}=[r_{11},\ldots\ldots\ldots,r_{1n}]$ by the ILHS.

(II) Position parameter k is computed for the P_{th} $(P=1,\ldots\ldots\ldots.,N)$ of RN r''_{1p} of the RV r_1 in vector $\boldsymbol{r_1}$ by Eq. (12).

Same for other RVs $r_h(h=2,3,\ldots\ldots\ldots,m)$, its RNV is $\boldsymbol{r}''_{\boldsymbol{h}}=[r''_{h1},\ldots\ldots\ldots,r''_{hN}]$ and $\boldsymbol{r_h}=[r_{h1},\ldots\ldots\ldots\ldots,r_{hn}]$ is discrete data vector of RV r_h. These steps to obtaining the RNM $\boldsymbol{R}''_{\boldsymbol{m\times N}}=[r''_1,\ldots\ldots\ldots.,r''_m]$, granted that it will preserve the correlation between MRV $\boldsymbol{r}_{m\times 1}$.

4 Computational Procedure of MCS-IS

Firstly, need to generate the RNs that preserve the MD and correlation between IRVs. In this work, for the calculations of deterministic load flow (DLF) equations related to exact non-linear load flow are implemented as:

$$\begin{cases} x = f(w) \\ z = g(x) = g(f(w)) \end{cases} \tag{13}$$

where input vector of the active and reactive power injection is denoted as w, the state vector of nodal voltage and angle is denoted as x, the line flows output vector is denoted as z, the flow function of the nodal power line is denoted as f & g respectively [30].

The procedure of computational calculations for MCS-IS is outlined as follows:

(I) Set the necessary data needed for DLF, power injection vector, sample size, and so forth.

(II) Choose the appropriate Copula function as described in Sect. 2.3, Generate the RNM $U_{m\times N}=[u_{1s},\ldots\ldots, u_{ms}]^{\mathrm{T}}$, here in this vector matrix, the first variable with N samples is u_{1s} $u_{is}=[u_{i1},\ldots\ldots,u_{iN}]$, where $i=1,\ldots\ldots.,m$. the technique used to generate these samples is /;'p[-SRS and in the range of [0, 1] with suitable Copula function.

(III) This is for MCS-IS, get the RNM $U''_{m\times N} = [u_{1s},\ldots\ldots, u_{ms}]^{T}$, here in this vector matrix, the first variable with N samples is u_{1s} and $u_{is} = [u_{i1},\ldots\ldots,u_{iN}]$, *and* $i = 1,2,\ldots\ldots,m$, by the ILHS.

(IV) The vector of power injection is $r_{m\times 1}$ and according to its data compute inverse function $r = F^{-1}(u)$ with its marginal CDF's. Generate RNM $R_{m\times N} = [r_{1s},\ldots\ldots, r_{ms}]^{T}(\boldsymbol{r_{is}} = [r_{i1},\ldots\ldots, r_{iN}])$ vector of power injection $r_{m\times 1}$ and with matrix $U''_{m\times N}$ produced in Step III.

(V) Method of Newton Raphson load flow is implemented and Run the DLF using Eq. (13) for N times, and then calculate the state vector $\boldsymbol{x}$ and line flow output vector $\boldsymbol{z}$.

(VI) Determine the probabilistic distribution of x and z by using estimation theory with statistical properties.

5 Results and Discussions

5.1 Accuracy Assessment of IRVs

Two wind energy farms situated in the city of New Jersey United States, that have total active output power (P_{WF}), which is to assess the accurateness of probability distribution of correlated IRVs based on Copula theory. In this case study, 52,559 discrete data (Eastern wind energy data set) P_{WF} output of two wind farm's from 1st Jan to 31st Dec 2006 with 10 min average interval are adopted [31]. Scatter plot with a histogram of P_{WF} output net-capacity of mentioned wind energy farms P_{WF1} & P_{WF2} respectively is shown in Fig. 1. From this figure, it is clearly shown that P_{WF1} & P_{WF2} are strongly correlated with $\rho = 0.978$. It is also shown in the histogram that P_{WF1} and P_{WF2} is not follow common MD.

The literature covers the selection of different suitable Copula functions, so according to this four Copula function is proposed that is suitable for this work. By using Eq. (10). the proposed four Copula functions, Euclidean distance dn is calculated individually. The Gumbel Copula has smallest value of $dn = 0.713$ and the worst one is Clayton Copula with the value of $dn = 5.041$. For Frank and Gaussian Copula, the values of dn are between of them. It is concluded that the suitable function to fit for given dataset is Gumbel Copula. Figure 2 shows the scatter plot of u_1 and u_2, where u_1 and u_2 are the CDFs of P_{WF1} & P_{WF2} respectively. In Fig. 3 the results of simulation with different Copula functions are presented. It is shown that almost all Copula functions are looking to be fit at given dataset. Among all of these, the most suitable one is Gumbel Copula function for given dataset. Gumbel copula function is selected for further study.

There are two types of errors, average root mean square (ARMS) ζ error index of CDF and relative error index ε_s consist of statistical properties for P_{WF} of wind energy farms are expressed in Eqs. (14) and (15) respectively:

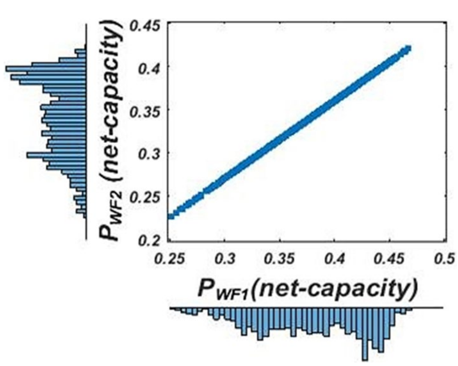


Fig. 1 Scatter and histogram plot for both wind farms net-capacity

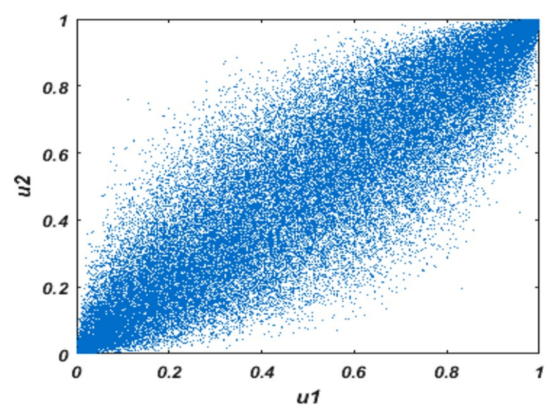


Fig. 2 Both Wind farm's scatter plot (output active power CDF)

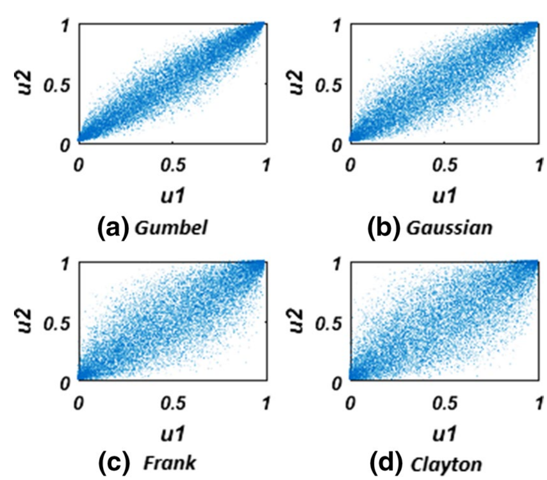


Fig. 3 Both wind farm's scatter plot (Generated CDF with Different Copula Function)

$$\zeta = \frac{\sqrt{\sum_{i=1}^{N} \left(P_{WFgs} - P_{WFri}\right)^2}}{N} \times 100\% \tag{14}$$

$$\varepsilon_s = \left| \frac{P_{WFgs} - P_{WFrs}}{P_{WFrs}} \right| \times 100\% \tag{15}$$

In term ε_s, the subscript s means statistical properties of wind energy farm's P_{WF} such as kurtosis, skewness, mean and standard deviation. P_{WFgs} & P_{WFri} are produced and actual values of P_{WF}, correspondingly. P_{WFgi} & P_{WFri} are the *ith* values on the CDFs are produced with actual values of P_{WF}, simulation points numbers are given as N [19]. The results of both indices are shown in Tables 1, 2 respectively. The most suitable one is Gumbel Copula function as shown in the results for the output datasheet of wind energy farm.

CDFs of P_{WF} are shown in Fig. 4 of correlated IRVs. The simulation results of Fig. 4 shows that the correlation influence on the synchronization of the P_{WF1} & P_{WF2} and further affect the PDF of wind energy farm's P_{WF}. When the degree of correlation is large. It means strong synchronization of wind energy farms and shows P_{WF} of the wind energy farm is increased. Also, shows that misleading results will occur without dependence factor. More, Fig. 4 also demonstrate that the results obtained by Copula theory of correlated IRVs are accurate.

5.2 MCS-C, ILHS-GA and MCS-IS Performance Evaluation

The performance of proposed method with compare with MCS-C and ILHS-GA is assessed by modified IEEE 14-bus. Test systems deterministic data are conferred in [30]. Loads active and reactive power are modelled as Gaussian distribution. Correlation between load and demand is demonstrated as Gaussian Copula function. Gaussian distribution parameters such that mean and standard deviation are equal to its deterministic value and arbitrary value respectively. More, to control the wind energy farm constant power factor approach is used. The results obtained by MCS-C method is adopted as accurate and reference for MCS-IS method. To validate the distribution accuracy of output RVs relative error index is implemented. as in [19, 23, 24].

$$\varepsilon_{\mu}^{*} = \left| \frac{\mu_a - \mu_s}{\mu_a} \right| \times 100\% \tag{16}$$

$$\varepsilon_{\sigma}^{*} = \left| \frac{\sigma_a - \sigma_s}{\sigma_a} \right| \times 100\% \tag{17}$$

where* represent the type of output RV and divided into four parts in this case study, such as: line active power P and line reactive power Q, nodal voltage magnitude V and nodal voltage phase angle θ. The mean and standard deviation are given as μ and σ, refer to statistical properties of output RVs [19]. Here, the subscript a mean is actual value of output RV obtained by MCS-C essentially considered in this work as a reference. RV output value obtained by MCS-IS and ILHS-GA is written as subscript s.

5.3 Modified Test System

Modified IEEE 14-bus test system is used for this work. At bus 13 and bus 14 both wind farms are located, as shown in Fig. 5. Total input RVs consist by this test system are 26 and total no of loads are 11 that covered into two parts:

Table 1 All Copula's with Relative error index ε_s of P_{WF}

Index εs (%)	Kurtosis	Skewness	Mean	Standard deviation
Gumbel	0.194	0.3804	0.024	0.281
Gaussian	0.116	0.899	0.542	0.866
Clayton	14.65	15.981	0.082	4.381
Frank	7.197	6.862	0.381	1.225

Table 2 ARMS error index of All Copula's

Various Copula function with ARMS index (%)			
Gumbel	Gaussian	Clayton	Frank
0.154	0.276	0.593	0.488

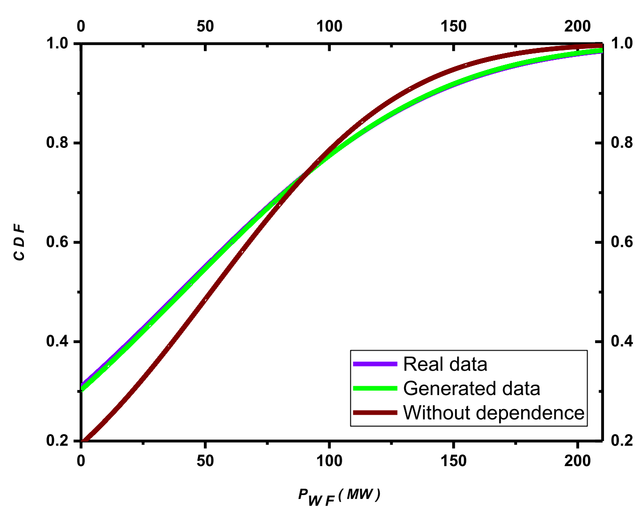
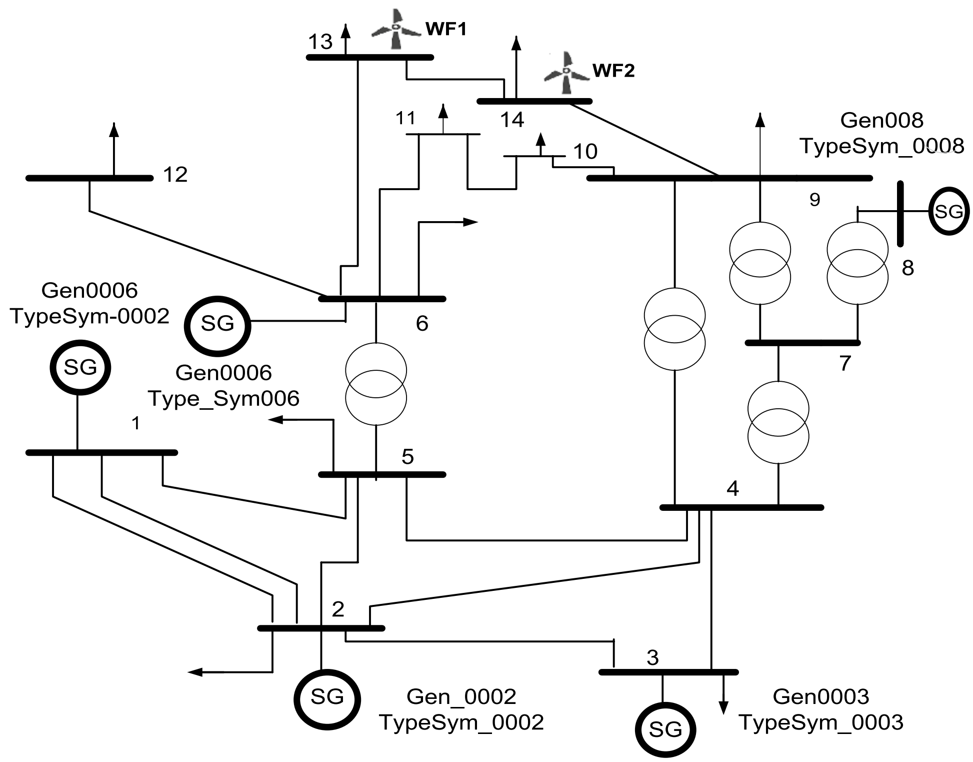


Fig. 4 CDF plot of both wind farm's P_{WF}

Fig. 5 Modified IEEE 14-bus test system



nodes 1–5 covered in part *I*, and nodes 6–14 covered in part [19]. Information about load modelling as Gaussian distribution describe in previous section. Accordingly, the parameters designed for mean values of loads and standard deviation values are deterministic and arbitrary respectively. For both parameters 15% value set in part *I* and 12% in part *II*. The correlation coefficient among loads of the same zone is set at 0.95 and in a different zone are set at 0.7.

In this proposed scheme, line active power *P* has selected as descriptive for the sake of the accuracy of output RVs. By using Eqs. (16) and (17) the error curves of ε_μ^p and ε_σ^p are calculated respectively, for MCS-C, MCS-IS and ILHS-GA methods as displayed in Figs. 6 and 7. The mean errors

Fig. 6 Mean error curves comparison of $P_{14\text{-}13}$ for both methods

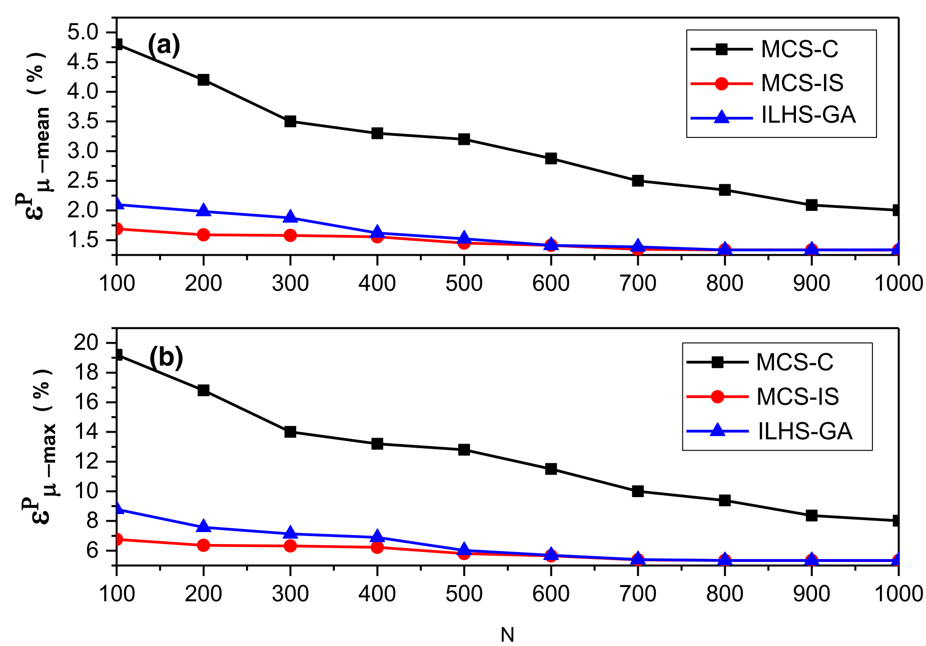
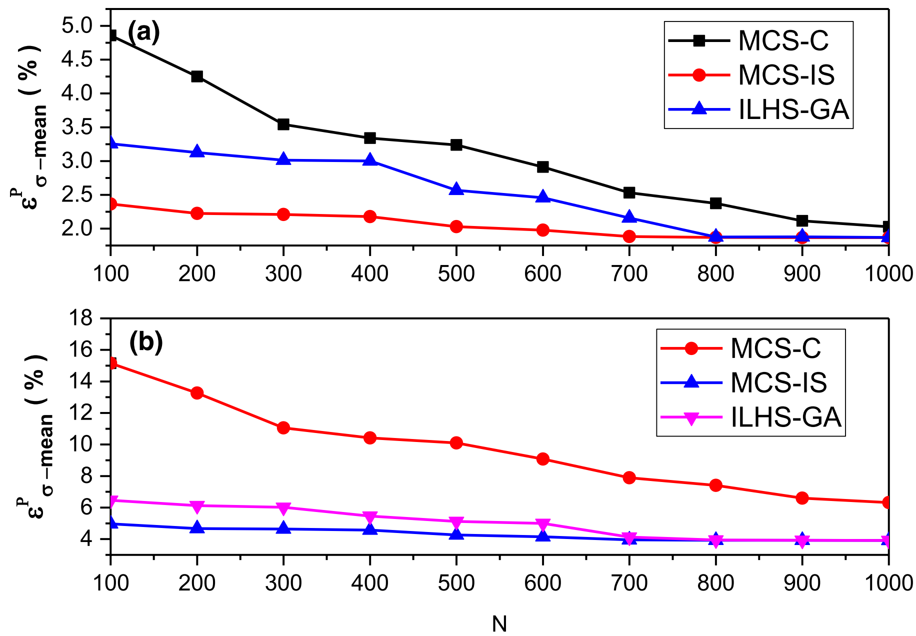


Fig. 7 Standard deviation error curves comparisons of $P_{14\text{-}13}$ for both methods



comparisons of active power flow through the line 14 to 13 ($P_{14\text{-}13}$) are displayed in Fig. 6a, b, correspondingly. Similarly, for standard deviation errors comparisons are shown in Fig. 7a, b. The result shows that curve obtained from proposed method is closer to touchstone method. More error is found with ILHS-GA method. The error comparisons of remaining output RVs are shown in Table. 3. The PDF and CDF of $P_{14\text{-}13}$ are shown in Fig. 7. With MCS-C method sample size is 52,559 and with MCS-IS and ILHS-GA method sample size is 600 are adopted in these plots. The index of ARMS error calculate by Eq. (16) is 0.154% for $P_{14\text{-}13}$ listed in Table. 2, with Gumbel Copula function. With the use of modified IEEE 14-bus test system, the computational time of these three methods are illustrated in Table 4. The required computational time considered for test system is 915 s for MCS-C method with sample size 52,559.

The DIgSILENT PowerFactory framework is utilized to perform simulation and examine the performance of both methods. The PC consist of following hardware configuration with processor of AMD Model No A12-9700P, RADEON R7, with 10 COMPUTOR CORES UC + 6G (4 CPU's), 2.5 GHz processing speed, and dual channel 8 GB DDR3 RAM.

In the simulations, MCS-C, MCS-IS and ILHS-GA are simulated 100 times with step size 100 up to 1000 samples. Both error indices are calculated for four output random variables. The results of these error indices are shown in Figs. 6, 7. These Figures demonstrate that convergent robustness of MCS-IS is better than MCS-C and ILHS-GA methods. The PDF and CDF plot of $P_{14\text{-}13}$ are shown in Fig. 8. It is clearly shown that both plots are close to each other rather than ILHS-GA method. From the results of Table 4, it

Table 3 Error comparison of both methods at sampling size 600

Methods	MCS-C	ILHS-GA	MCS-IS	
$\varepsilon^{v}_{\mu}(\%)$	M	0.244	0.201	0.018
	SD	0.026	0.021	0.016
	MAX	0.084	0.078	0.106
$\varepsilon^{v}_{\sigma}(\%)$	M	2.788	2.158	1.362
	SD	1.565	1.178	0.655
	MAX	8.201	5.785	3.762
$\varepsilon^{\theta}_{\mu}(\%)$	M	1.968	1.354	0.476
	SD	1.581	1.123	0.301
	MAX	7.405	3.215	1.489
$\varepsilon^{\theta}_{\sigma}(\%)$	M	2.597	1.892	1.268
	SD	1.896	1.451	0.807
	MAX	8.18	5.356	3.397
$\varepsilon^{Pij}_{\mu}(\%)$	M	3.152	2.458	0.878
	SD	2.125	1.568	0.468
	MAX	10.991	7.325	2.456
$\varepsilon^{Pij}_{\sigma}(\%)$	M	2.648	1.785	1.301
	SD	1.655	1.201	0.861
	MAX	7.781	4.325	2.704
$\varepsilon^{Qij}_{\mu}(\%)$	M	2.485	1.754	0.812
	SD	1.596	1.325	0.387
	MAX	7.612	6.214	2.141
$\varepsilon^{Qij}_{\sigma}(\%)$	M	3.4762	2.471	1.687
	SD	1.798	1.125	0.501
	MAX	9.023	5.021	2.874

[*] M for Mean, SD for standard deviation & MAX for maximum

Table 4 Computational time with 100 step size

Sample size	Methods		
	MCS-C seconds (S)	ILHS-GA seconds (S)	MCS-IS seconds (S)
100	0.69	0.58	0.72
200	1.16	0.60	0.107
300	1.17	0.95	1.26
400	1.27	1.31	1.05
500	1.33	1.46	1.45
600	1.42	1.70	1.78
700	2.34	1.86	2.15
800	2.51	2.33	2.45
900	2.85	2.71	2.78
1000	4.21	3.04	3.19

could be observed that the computational burden of MCS-IS method is lesser on average. For MCS-IS, it is a bit larger than MCS-C and ILHS-GA at 400 sample sizes. It is due to additional steps needed for ILHS and genetic algorithm. But this computational time difference is negligible even better when sample size is large. Additionally, the probability distribution of $P_{14\text{-}13}$ for MCS-C with 52,559 samples. For MCS-IS and ILHS-GA methods the sample size is 600. So, the simulation results validate that MCS-IS can attain a high accurate solution as compare to MCS-C and ILHS-GA with much less sample size. From the discussion of above results, Overall, it is summarized that MCS-IS is an accurate, flexible and robust method for probabilistic problems. Also, it has a tremendous potential for PLF analysis of wind energy sources.

6 Conclusion

In this work, a probabilistic load flow method for active power system network has been proposed. Uncertainty is increasing day by day due to penetration of distributed and stochastic generation such that solar and wind energy in power system network. PLF study is essential factor to model the uncertainty. Copula theory has been proposed to model the probability distribution of correlated IRVs. From the simulation results, this theory has the capability to deal with a correlation degree as well as correlation structure between IRVs flexibly that is a fundamental need for modelling of IRVs probability distribution. Modified Latin-hypercube sampling has implemented to conquer the limitation of LHS. From the results, ILHS can deal with discrete real data obtain by SCADA or other real data measuring techniques. More, ILHS is unconstraint by the marginal distribution type variables. MCS-IS can converge with much fewer samples as compare to MCS-C and ILHS-GA. It means, it is more efficient and robust method. Also, its computational time is almost similar to MCS-C and ILHS-GA at lesser sample size but it decreases when sample size is large. By comparing the overall performance and accuracy of these methods MCS-C, MCS-IS and ILHS-GA, it is clearly shown that MCS-IS is an auspicious method for PLF study for active power system network, especially with wind energy farms.

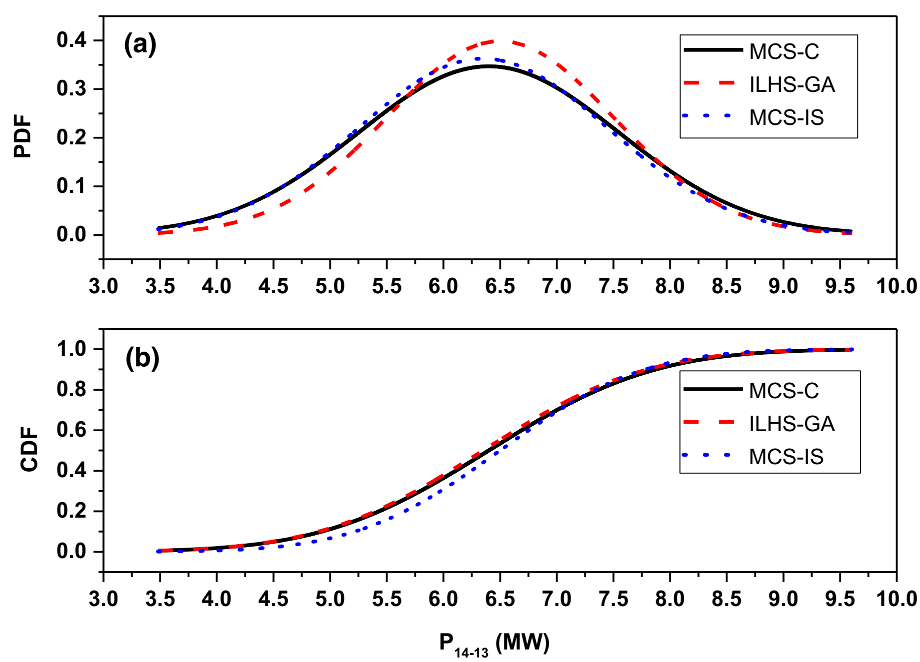


Fig. 8 PDF and CDF curves of $P_{14\text{-}13}$ for both methods

Acknowledgements I would like to thank Prof. Bin Li for their time and guidance and all co-authors. The authors express their gratitude to "North China Electric Power University, Beijing, China, and State Grid Corporation of China for their financial support on this research work".

Declaration

Conflict of Interest There is no conflict of interest.

Future Work Input random variables will be model with other parametric distributions. The result will be compared with proposed techniques and the techniques available in literature.

References

1. Wang Y, Zhang N, Kang C, Miao M, Shi R, Xia Q (2018) An efficient approach to power system uncertainty analysis with high-dimensional dependencies. IEEE Trans Power Syst 33(3):2984–2994
2. De Jong M, Papaefthymiou G, Palensky P (2018) A framework for incorporation of infeed uncertainty in power system risk-based security assessment. IEEE Trans Power Syst 33(1):613–621
3. Bin L, Shahzad M, Bing Q, Ahsan M, Shoukat MU, Khan HM, Fahal NA (2018) The probabilistic load flow analysis by considering uncertainty with correlated loads and photovoltaic generation using Copula theory. AIMS Energy 6(3):414–435
4. Bin L, Shahzad M, Bing Q, Fahal NAM, Islam MR, Shoukat MU, Ahsan M (2018) Probabilistic load flow analysis of power system network considering uncertainty with generation and correlated loads. IJSSST 19(3):1–6
5. Prusty BR, Jena D (2018) An over-limit risk assessment of PV integrated power system using probabilistic load flow based on multi-time instant uncertainty modeling. Renew Energy 116:367–383
6. Bin L, Shahzad M, Bing Q, Shoukat MU, Fahal NA, Simiyu P, Islam R (2018) Probabilistic analysis of payback period for AC–DC transmission and distribution asset expansion projects. J Eng 16:680–685
7. Nijhuis M, Gibescu M, Cobben S (2017) Gaussian mixture based probabilistic load flow for LV-network planning. IEEE Trans Power Syst 32(4):2878–2886
8. Valverde G, Saric A, Terzija V (2012) Probabilistic load flow with non-Gaussian correlated random variables using Gaussian mixture models. IET Gener Transm Distrib 6(7):701–709
9. Wang Y, Zhang N, Chen Q, Yang J, Kang C, Huang J (2017) Dependent discrete convolution based probabilistic load flow for the active distribution system. IEEE Trans Sustain Energy 8(3):1000–1009
10. Cai L (2011) A probabilistic characterization of g-harmonic functions. arxiv preprint 1108:2558
11. Kloubert, M.-L. , C. Rehtanz (2017) Of Conference *Enhancement to the combination of point estimate method and Gram-Charlier Expansion method for probabilistic load flow computations*. In: *PowerTech, 2017 IEEE Manchester*. 2017. IEEE
12. Diop, F. M. Hennebel (2017) of Conference *Probabilistic load flow methods to estimate impacts of distributed generators on a LV unbalanced distribution grid*. In: *PowerTech, 2017 IEEE Manchester*. 2017. IEEE
13. Yu, H. B. Rosehart (2011) of Conference *Probabilistic power flow considering wind speed correlation of wind farms*. In: *17th Power Systems Computation Conf., Stockholm, Sweden*
14. Cao, J., W. Du, H. Wang , L. Xiao (2011) of Conference *Probabilistic load flow using latin hypercube sampling with dependence for distribution networks*. In: *Innovative Smart Grid Technologies (ISGT Europe), 2011 2nd IEEE PES International Conference and Exhibition on*. IEEE.
15. Ren Z, Koh C-S (2013) A second-order design sensitivity-assisted Monte Carlo simulation method for reliability evaluation of the electromagnetic devices. J Electric EngTechnol (JEET) 8(4):780–786
16. Kenari MT, Sepasian MS, Nazar MS, Mohammadpour H (2017) The combined cumulants and laplace transform method for probabilistic load flow analysis. IET Gener Trans Distrib 11(4):3548–3556
17. Corbetta A, Muntean A, Vafayi K (2015) Parameter estimation of social forces in pedestrian dynamics models via a probabilistic method. Math Biosci Eng 12(2):337–356
18. Su H, Dong X, Yu X (2020) Probabilistic load flow analysis based on sparse polynomial chaotic expansion. J Electric Eng Technol 15(2):527–538
19. Cai D, Shi D, Chen J (2014) Probabilistic load flow computation using Copula and Latin hypercube sampling. IET Gener Transm Distrib 8(9):1539–1549
20. Beltran Valle O, Peña Gallardo R, Segundo Ramirez J, Wenzhong D, Muljadi E (2020) A graphical probabilistic representation for the impact assessment of wind power plants in power systems. J Electric Eng Technol 15(5):2033–2043
21. Khan M, He C, Liu T, Ullah F (2021) A new hybrid approach of clustering based probabilistic decision tree to forecast wind power on large scales. J Electric Eng Technol 16(2):697–710
22. Ruiz-Rodriguez F, Hernandez J, Jurado F (2012) Probabilistic load flow for radial distribution networks with photovoltaic generators. IET Renew Power Gener 6(2):110–121
23. Yu H, Chung CY, Wong KP, Lee HW, Zhang JH (2009) Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition. IEEE Trans Power Syst 24(2):661–667
24. Chen Y, Wen J, Cheng S (2013) Probabilistic load flow method based on Nataf transformation and Latin hypercube sampling. IEEE Trans Sustain Energy 4(2):294–301
25. Xu, X. , Z. Yan (2015) of Conference *Probabilistic load flow evaluation with hybrid Latin Hypercube Sampling and multiple linear regression*. In: *Power & Energy Society General Meeting, IEEE*. 2015
26. Hajian M, Rosehart WD, Zareipour H (2013) Probabilistic power flow by Monte Carlo simulation with Latin supercube sampling. IEEE Trans Power Syst 28(2):1550–1559
27. Bin L, Shahzad M, Bing Q, Shoukat MU, Shakeel M, Mohammedsaeed EK (2018) Probabilistic computational model for correlated wind farms using copula theory. IEEE Access 6:14179–14187
28. Merola SS, Marchitto L, Tornatore C, Valentino G (2014) Spray-combustion process characterization in a common rail diesel engine fuelled with butanol-diesel blends by conventional methods and optical diagnostics. AIMS Energy 2(2):116
29. Nelsen, R.B. (1999) Introduction, in *An Introduction to Copulas*. Springer. 1–4.
30. Cai D, Shi D, Chen J (2013) Probabilistic load flow computation with polynomial normal transformation and Latin hypercube sampling. IET Gener Transm Distrib 7(5):474–482
31. Obtaining the Eastern Wind Dataset. http://www.nrel.gov/electricity/transmission/eastern_wind_dataset.html.

Bin Li was born in Beijing, P.R.China, in 1983. He is currently an associate professor of North China Electric Power University. His research interests include communication technology for smart grid, and electric power communication, automated demand response. Dr. Li has received the certifications from IBM Rational University Program and always working as OPNET Training Engineer. He has published more than 150 journal and conference papers, obtained more than 20 software copyright, and submitted more than 20 patents. He is a member of China Communications Standards Association (CCSA) and Chinese Society for Electrical Engineering (CSEE).

Dr. Muhammad Shahzad He received his B.Sc. and M.Sc. degrees from The Islamia University of Bahawalpur, Pakistan, in 2011 and 2013 respectively. He received his Ph.D. degree from North China Electric Power University, Beijing, P.R. China. Currently, He is Assistant Professor at Muhammad Nawaz Sharif University of Engineering and Technology, Multan, Pakistan. He has published more than 30 articles and conference papers. His research interest include Probabilistic and statistical analysis of electric Power System, Renewable Energy, and Power System operation and Control.

Hafiz Mudassir Munir received his B.S. degree in Electrical Engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2009; his M.S. degree in Electrical Power Engineering from the University of Greenwich, London, ENG, UK, in 2012; his Ph.D. degree in Electrical Engineering from the University of Electronic Sciences and Technology of China (UESTC), Chengdu, China, with research focused on power electronics including the hierarchical and cooperative control of distributed generation and active power filters. Currently, he is working as an Assistant Professor at Sukkur IBA University, Pakistan. His research interests include the ac/dc microgrids, active distribution networks, power quality, grid-connected converters for renewable energy systems, active power filters, multilevel converters, and static synchronous compensators (STATCOMs). He has published numerous articles and papers in various high-ranked journals including the Journal of Power Electronics, Sustainability, IEEE Access, IEEE Journal of Emerging and Selected Topics in Power Electronics. Dr. Munir was awarded a scholarship for graduate studies by the Government of China. He is a Member of Pakistan Engineering Council and Institute of Electrical and Electronics Engineers.

Dr. Asif Nawaz received his B.Sc in Electrical Engineering from University of Engineering and Technology Lahore in 1998, MBA-ITM from International Islamic University Islamabad in 2004, MS Computer Engineering from Lahore University of Management Sciences (LUMS) Lahore from 2006, PhD in Electromagnetic Field and Microwave Technology from Beijing University of Posts and Telecommunication (BUPT), Beijing, China in 2010, EMBA from Lahore University of Management Sciences (LUMS) in 2012. Presently he is working in Higher Colleges Technology as Asst Professor of Electrical Engineering in Faculty of Engineering Technology and Sciences. His research interests are Machine Learning and Artificial Intelligence, IOT, ICT Transformation, Cloud Architecture, NFV/SDN, Next Generation Networks; Wireless Mesh Networks (WMNs); Satellite and Mobile Adhoc Networks (MANETs); Optical Wireless Hybrid Networks; Optical Networks (OBS, OPS); High Speed Transport Protocols; MPLS/GMPLS; Mobile IP and IPV6 Deployment and Test beds, Network Management and Performance Evaluation and Telecom Project Management.

Nabeel Abdelhadi Mohamed Fahal received his B.Sc. and M.Sc. degrees from North China Electric Power University, Beijing, P.R. China, in 2012 and 2016 respectively. Currently, He is doing his Ph.D. degree at North China Electric Power University, Beijing, P.R. China. He has published Four articles and conference papers. His research interest includes Probabilistic and statistical analysis of electric Power System, Renewable Energy, and Power System operation and Control.

Dr. Muhammad Yousaf Ali Khan did his BSc Electrical Engineering from University of Engineering &Technology, Peshawar, Pakistan in 2006. He did his MS and PhD in Electrical and Electronic System Engineering from Hanyang University, South Korea in 2010 and 2016 respectively. Currently, he has been working as Assistant Professor at the department of Electrical Engineering, Gomal University, D.I.Khan, Pakistan. He has vast experience of working in Academia, Industry and Research organizations. His research contributions are published in many international and national journals. He remained session Chair for different IEEE Conferences and being a technical reviewer for different conferences and journals.

Professor Dr. Sheeraz Ahmed did his graduation in Electrical Engineering (Communication Group) from University of Engineering and Technology, Peshawar, Pakistan with Honors in 1997. He then did his Masters in Mathematics from University of Peshawar in 1999. In 2006, he enrolled in the MSCS program in City University of Science and Technology, Peshawar and completed his MS program with a major in Computer Networks. He then got enrolled in PhD program in Electrical Engineering in COMSATS Islamabad in Spring 2012 and completed his Doctoral degree in Spring 2016. During his stay in Ph.D program, he published a total of 17 publications both in IF journals of international repute and IEEE conferences. His thesis title was "Towards Cooperative Routing in Underwater and Body Area Wireless Sensor Networks". Dr. Sheeraz has been engaged in serving as Dean Faculty of Engineering and Technology in Gomal University, D.I.Khan, Pakistan in 2017. Currently he is serving as Full Professor in the department of Computer Science, Iqra National University, Peshawar. Till date, he has more than 190 publications to his credit with more than 30 IF journals, 42 non-IF journals and rest IEEE conferences. He has the honor to be the first researcher to establish a Research Academy in Pakistan by the name of Career Dynamics Research Centre. His research interest areas include Underwater Sensor Networks, Body Area Networks, Vehicular Adhoc Networks, Smart Homes, Smart Grid etc.