

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Improved Stochastic Modeling: An Essential Tool for Power System Scheduling in the Presence of Uncertain Renewables

Sajjad Abedi, Gholam Hossein Riahy,
Seyed Hossein Hosseinian and Mehdi Farhadkhani

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/45849>

1. Introduction

Nowadays, governments are developing ambitious goals toward the future green and sustainable sources of energy. In the U.S., the penetration level of wind energy is expected to be 20% by the year 2030 [1]. Several European countries already exhibit the adoption level in the range of 5%–20% of the entire annual demand. Also with further developments in the solar cells technology and lower manufacturing costs, the outlook is that the photovoltaic (PV) power will possess a larger share of electric power generation in the near future. Grid-connected PV is ranked as the fastest-growing power generation technology [2]. PV generates pollution-free and very cost-effective power which relies on a free and abundant source of energy.

Due to the increasing wind and solar penetrations in power systems, the impact of system variability has been receiving increasing research focus from market participants, regulators, system operators and planners with the aim to improve the controllability and predictability of the available power from the uncertain resources. The produced power from these resources is often treated as non-dispatchable and takes the highest priority of meeting demand, leaving conventional units to meet the remaining or net demand. This issue makes the optimum scheduling of power plants in power system cumbersome as embeds the stochastic parameters into the problem to be handled. The unpredictability along with potential sudden changes in the net demand, may face operators with technical challenges such as ramp up and down adaptation and reserve requirement problems [3-4].

Several investigations aiming at handling the uncertain nature of wind and solar energy resources have been reported. Basically, the methods found in the literature can be classified into three groups: methods that deal with the prediction of uncertain variables as an input data pre-processing, methods that use stochastic scenario-based approach within the optimization procedure to cover all the outcomes per the probable range of uncertain variables, and methods based on a combination of these two approaches. The studies presented in [5-7] can be mentioned as one of the most recent efforts lying in the first group. In [5-6] an Artificial Neural Network (ANN) forecast technique is employed and followed by risk analysis based on the error in the forecast data. Then, the so called pre-processed data is directly taken as the input to the optimization process. Relying on the forecast tools, such methods suffer from high inaccuracy or ex-ante underestimation of the available power which increases the scheduled generation and reserve costs. Anyway, this approach is useful as it accounts for the temporal correlation between the random variables representative of each time step of the scheduling period, in terms of time-series models. On the other hand, in [8-9] which belong to the second group, the focus is on the stochastic scenario analysis rather than the forecasting methods. The usage of this approach also has its own advantages, as it tries to model the likely range of values for the random variables. However, the efficiency of this approach largely depends on the accuracy and reliability of their probabilistic analysis; based on which the potential scenarios are built.

The most effective approach is associated with the third group, which applies the advantages of both forecast techniques and scenario-based optimization approach. Reference [10] presents a computational framework for integrating a numerical weather prediction (NWP) model in stochastic unit commitment/economic dispatch formulations that describes the wind power uncertainty. In [11], the importance of stochastic optimization tools from the viewpoint of the profit maximization of power generation companies is investigated. The exposed financial losses regarding the wind speed forecast errors are discussed. A stochastic model is also presented in [12] which uses a heuristic optimization method for the reduction of random wind power scenarios. The wind speed data is assumed to follow the normal PDF. A similar approach is introduced in [13] whereas the wind speed error distribution is considered as a constant percentage of the forecasted data. In [14], the Auto-Regressive Moving Average (ARMA) time series model was chosen to estimate the wind speed volatility. Based on the model, the temporal correlation of wind speed at a time step with respect to the prior time steps is well analyzed.

In this chapter, the authors present a framework for stochastic modeling of random processes including wind speed and solar irradiation which are involved in the power generation scheduling optimization problems. Based on a thorough statistical analysis of the accessible historical observations of the random variables, a set of scenarios representing the available level of wind and solar power for each time step of scheduling are estimated. To this aim, the Kernel Density Estimation (KDE) method is proposed to improve the accuracy in modeling the Probability Distribution Function (PDF) of wind and solar random variables. In addition, the concept of aggregation of multi-area wind/solar farms is analyzed using Copula method. Taking the advantage of this method, we can reflect the interdependency and spa-

tial correlation of the power generated by several wind farms or PV farms that are spread over different locations in the power system. A final framework is developed to perform the stochastic analysis of the random variables to be input into the stochastic optimization process, as discussed in the following sections.

2. Methodology of data processing

2.1. Probability distribution function and data sampling

In order to generate sample data for random variables, the random behavior should be simulated somehow that the model follows the historical data pattern with the most homology to real data. In order to specify the pattern of a random variable, the PDF should be obtained. There are two classes of methods to determine the PDF of a random variable including parametric and non-parametric methods [15]. In parametric methods, the data samples are fitted to one of the well-known standard PDFs (such as Normal, Beta, Weibull, etc.) so that the most possible adaptation between the PDF and the existing data is achieved. The values associated with the PDF parameters are evaluated using Goodness of Fit (GoF) methods such as Kolmogorov-Smirnov test [16]. On the other hand, the nonparametric methods do not employ specific well-known PDF models.

The use of parametric methods in some studies in which simulation of probabilistic models for wind and solar data is included have been reported, as in [6, 12]. Similarly, authors in [9] employ a fixed experimental equation to represent the PDF of wind data. However, this approach to PDF estimation can bring about some defects as follows:

1. The parametric methods may show significant deviation to the actual distribution of data, mainly because the actual distribution does not characterize the underlying symmetry in the standard PDFs. As an example, Figure 1 shows the distribution function for yearly solar irradiation sample data at 11 AM in a region. As seen in the figure, the parametric distribution fittings are not capable of modeling the right side skewness in the actual distribution, which will reveal considerable error in the outcoming samples.
2. Some random variables in general and particularly solar irradiation and wind speed are very time-dependent in behavior. In other words, their patterns change with different time periods, months and seasons. Hence, the nonparametric approach is advantageous in terms of time period adaptation, because it does not consider a specific type of distribution. However, the parametric approach tries to nominate a certain type of PDF to each random phenomenon in all circumstances. For instance, it is common to associate a Weibull pattern to wind speed data, which may not be the most appropriate option to be generalized to all time periods.

Based on the aforementioned facts, in this study, it is desired to obtain the most accurate distribution model taking the advantage of Kernel Density Estimation (KDE), categorized as a non-parametric method.

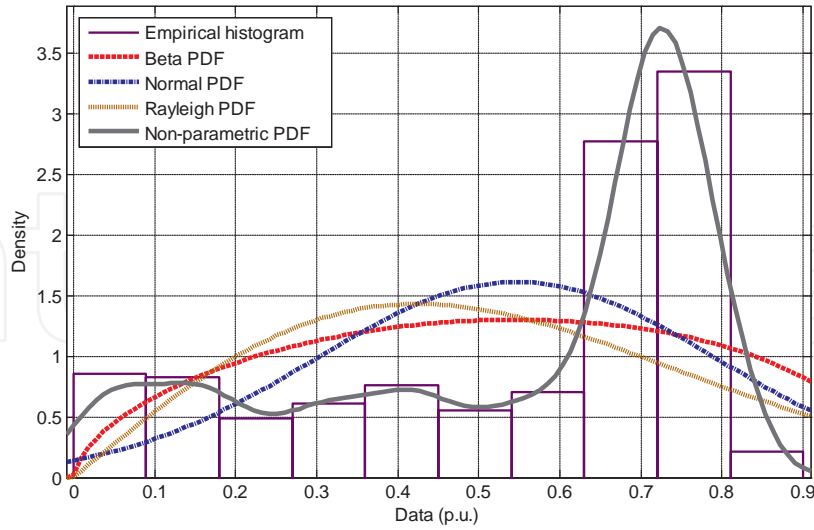


Figure 1. Solar irradiation data histogram for samples at 11 AM during a season and the parametric and nonparametric PDF fits to the data

2.2. Kernel Density Estimation (KDE)

The simplest and most frequently used nonparametric method is to use a histogram of historical samples. As a brief description of the method, the distance covering the range of samples is divided into equal sections called "bin". For each bin, a sample value is considered as the kernel of that bin. A number of rectangular blocks equal to the number of samples in each bin, each with unity area, are located on each bin. In this way, a discrete curve is obtained that somewhat describes the probability distribution of samples. However, the overall curve is largely dependent on the size of bins and their marginal points, because with the alterations in the bin size, the number of samples in each bin will be changed [17]. Besides, the obtained curve suffers from high raggedness. Hence, KDE method was introduced to solve the mentioned drawbacks. In this method, considering the samples as the kernels of each bin, the blocks are with a unity width and a height equal to the inverse of the number of samples for each sample value ($\frac{1}{n}$) totally gaining a block with unity area. The accumulation of these blocks builds the PDF curve. In order to smooth the curve and eliminate the dependency to the block width, continuous kernels such as Gaussian or Cosine along with kernel smoother methods are used [18]. Figure 2 demonstrates a KDE with Gaussian kernel. The overall PDF will be obtained by:

$$\hat{f}(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{h_i} \cdot K\left(\frac{x - X_i}{h}\right) \quad (1)$$

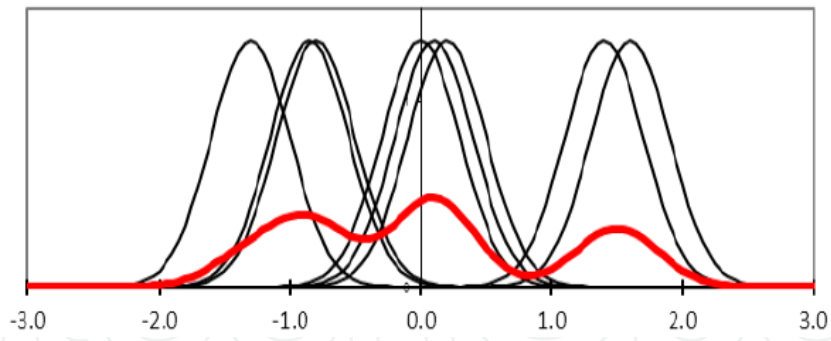


Figure 2. Demonstration of a KDE with Gaussian kernel

where h is the smoothing factor. The kernel function in Gaussian case is given by:

$$K(x) = \frac{1}{N} \sum_{i=1}^N \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-X_i)^2}{2h^2}} \quad (2)$$

In the present study, the KDE method is used to obtain the PDF of the seasonally wind speed and solar irradiation data for each hour in a day. The method is implemented using the *dfit tool* function in MATLAB.

2.3. Correlation of random variables

Sample generation from a random variable is possible simply using a Monte Carlo simulation of its corresponding PDF. However, this is more cumbersome for a group of random variables which may have underlying dependence or correlation. Neglecting the correlation will result in the inaccurate multivariate PDF and then to irrelevant and deviated samples.

There are several correlation coefficients to quantify the correlation among a number of random variables, among which the most famous one is the Pearson coefficient:

$$\rho = \frac{\text{cov}(i, j)}{\sqrt{\text{cov}(i, i) \cdot \text{cov}(j, j)}} \quad (3)$$

where *cov* is the covariance function. This analysis reflects only the *linear* correlation among the random variables. Nevertheless, in many cases, random variables reveal nonlinear correlation and more complicated relationships among themselves, especially when the PDF of the variables are not of similar patterns [19]. In a case with a large number of random variables, neglecting the nonlinear relation will result in more significant deviation of output samples from what they actually should be. For instance, the solar irradiation behavior in two different regions in power system may not establish a linear relation between each other, though they are not completely independent. The nonlinear correlation concept is presented in Figure 3.

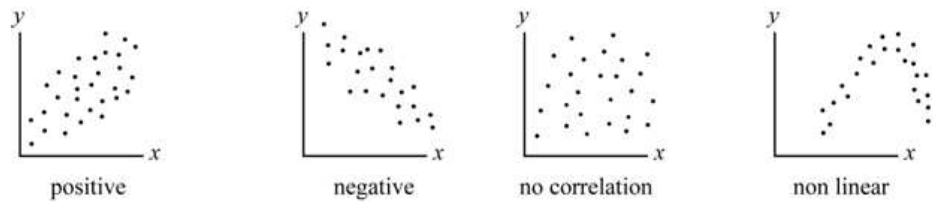


Figure 3. A representation of linear vs. nonlinear correlation between two random variables.

In the problem under study, i.e. power system scheduling in the presence of uncertain renewables, we consider the presence of multiple wind farms and solar farms throughout the power system. The solar power and wind power as well as load demand are three distinct stochastic processes. They can be discriminated into 24 random variables representing 24 hours of the day.

The random variables within each random process have their own temporal relation which can be modeled by time series prediction methods [14]. However, there may also be spatial correlation among the random variables from different processes, although it might seem unlikely. Here, we are going to deal with the tangible nonlinear correlation between the hourly random variables for a wind farm and another farm located in a close region, as well as for a PV farm and another one located in a close region. The interested reader may examine other possible dependence structures between random variables / processes. Obviously, taking into account these relations results in more accuracy and enhancement of the models and solutions. Figure 4 presents how neglecting the correlations and directly using single variable PDFs to generate samples for a multivariate process may lead to model malfunction. For two random variables with similar Log-normal distribution and Pearson correlation of 0.7 (with diagonal covariance matrix), 1000 samples have been generated considering total independence (Figure 4 (a)) and linear dependence (Figure 4 (b)). It is observed in Figure 4 (b) that X_1 values tend to be closer to X_2 values especially in the upper range, in comparison with Figure 4 (a).

In order to describe the correlations between random variables including nonlinear correlations, a method named Copula can be employed which is described in the following section.

2.3.1. Copula method

The correlation between random variables or samples is measured by the Copula concept. Embrechts & McNeil introduced Copula functions for application in financial risk and portfolio assessment problems [20]. Recently, much attention is being paid to this method in statistical modeling and simulation problems.

Copulas provide a way to generate distribution functions that model the correlated multivariate processes and describe the dependence structure between the components. The cumulative distribution function of a vector of random variables can be expressed in terms of marginal distribution functions of each component and a copula function.

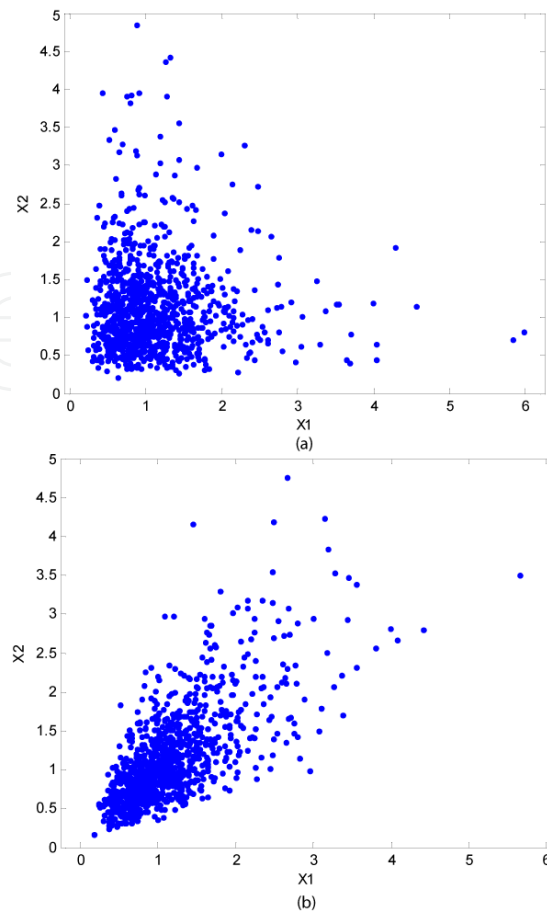


Figure 4. The effect of correlation on generation of random samples: (a) the samples are generated assuming complete independence and (b) the samples are generated considering the correlation.

The basic idea behind the copula method is described as the Sklar's theorem [21]. It shows that a multivariate cumulative distribution function (CDF) can be expressed in terms of a multivariate uniform distribution function with marginal density functions $U(0,1)$. In fact, if we have n random variables with an n -variable CDF, F , with margins F_1, F_2, \dots, F_n , there is an n -variable distribution function, C , given by:

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \quad (4)$$

This equation can be rewritten to extract the Copula of the joint distribution function of the random variables, as follows:

$$C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \quad (5)$$

where $F_i^{-1}(u_i)$ is the inverse CDF. If F is a continuous multivariate PDF with continuous single-variable PDFs, the implicit copula distribution function is obtained as follows:

$$c(u) = \frac{f(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))}{f(F_1^{-1}(u_1)) \cdot f(F_2^{-1}(u_2)) \dots f(F_n^{-1}(u_n))} \quad (6)$$

This equation can be restated as:

$$c(F_1(x_1), \dots, F_n(x_n)) = \frac{f(x_1, \dots, x_n)}{f_1(x_1) \dots f_n(x_n)} \quad (7)$$

where c is the PDF corresponding to C . Therefore, a multivariate PDF can be written in terms of the product of its single-variable marginal distributions and its underlying copula (c):

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \cdot f_1(x_1) \dots f_n(x_n) \quad (8)$$

Various copula functions are introduced by present. They are generally classified into explicit and implicit types. The implicit copulas are inspired by standard distribution functions and have complicated equations, whereas the explicit ones are simpler and do not follow the specific functions. Among the most widely used implicit copulas, Gaussian copula and t-Student copula and among the explicit ones, Clayton copula and Gumbel copula can be mentioned. The selection of the most appropriate copula is a complicated problem itself. Here, the t-Student copula is employed because of its simplicity and flexibility. The t-Student copula is formulated as [21]:

$$C_{\rho, \nu}(u, v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{ 1 + \frac{s^2 - 2\rho st + t^2}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2} ds dt \quad (9)$$

where (ρ, ν) are the copula parameters, $t_v^{-1}(\cdot)$ is the T distribution function with ν degrees of freedom, mean of zero and variance of $\frac{\nu}{\nu-2}$. The best values for these parameters can be estimated using Inference Functions for Margins (IFM) or Canonical Maximum Likelihood (CML) methods. In both methods, at first the parameters of the single-variable marginal distribution functions are computationally or experimentally determined. Then, by substitution of these functions into Copula likelihood functions, the Copula functions are calculated so that the Copula likelihood functions are maximized. Further discussion on the mathematical background and the calculation methods can be found in [21-23] which is suggested to be pursued by the interested reader.

In the current study, the authors employed the two-dimensional Copula method to present the correlation of the wind speed patterns between two wind farms and the correlation of the solar irradiation patterns between two PV farms, for every hour of the day. The available data for three years are initially normalized:

$$s = \frac{x - \mu_x}{\sigma_x} \tag{10}$$

where μ_x , σ_x are the mean and standard deviation of data x , respectively. The simulation results for wind speed data distribution in farm 1 (x axis) and farm 2 (y axis) for 7 and 11 AM in fall season are shown in Figure 5. Figure 6 shows the samples for solar irradiation data distribution on two farms. Also, the values of linear correlation coefficients, the Kendall's tau correlation and the t-Student copula parameters for wind speed are presented in Table 1. The results state that the correlation of wind speed (and similarly solar irradiation) between two farms may not be negligible since they have similarities in their climatic conditions.

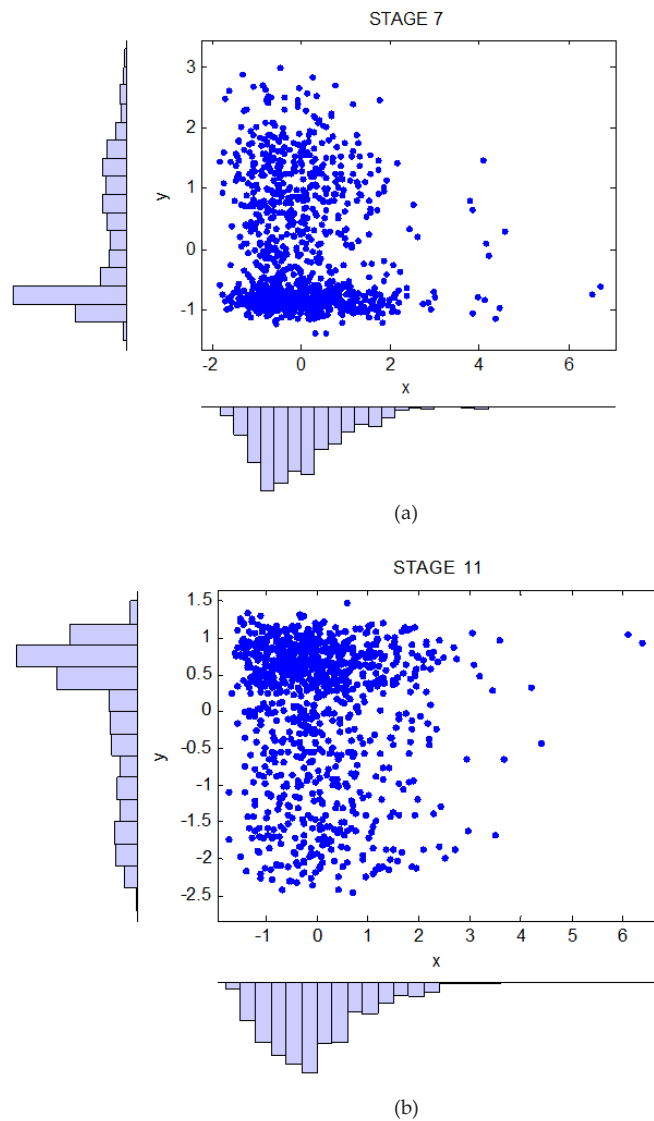


Figure 5. Correlated samples for wind speed of two farms for 7 and 11 AM in fall season.

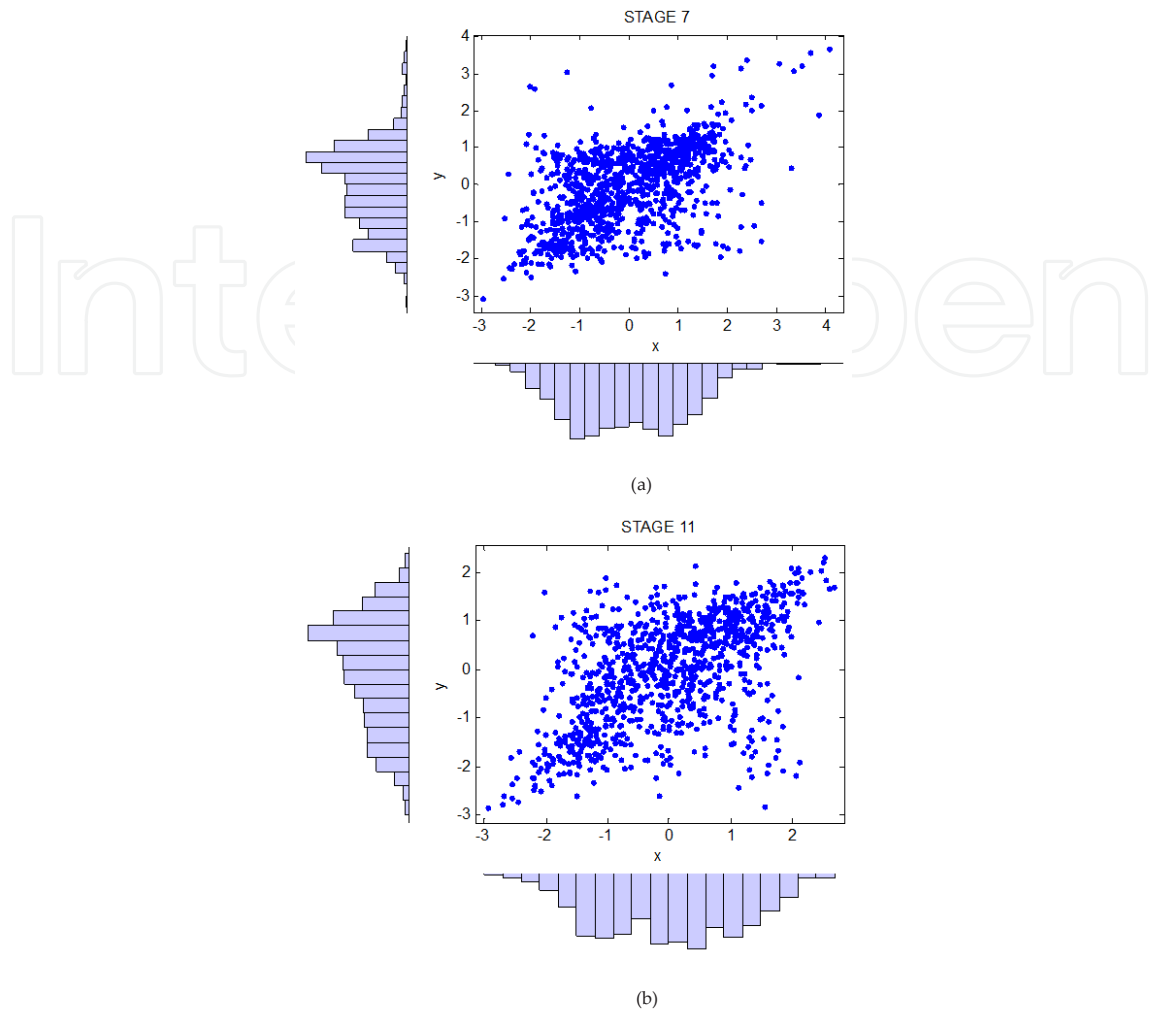


Figure 6. Correlated samples for solar irradiation of two farms for 7 and 11 AM in fall season

| Kendall's tau correlation | Freedom degree (ν) | Linear correlation (ρ) | Hour |
|---------------------------|--------------------------|-------------------------------|------|
| -0.14 | 3.7 | -0.21 | 7 |
| -0.12 | 15.58 | -0.19 | 10 |
| -0.08 | 7.39 | -0.13 | 11 |
| -0.12 | 24.31 | -0.18 | 16 |

Table 1. The parameters of the two-variable t-Student copula distribution for wind speed of two farms at different hours of the day in fall season

2.4. Time-series prediction of wind speed and solar irradiation

As mentioned earlier, besides the spatial correlation among different farms, the wind speed and solar irradiation random variables assigned to the scheduling time steps exhibit temporal correlation, i.e., they are dependent on the condition of random variables at previous

time steps (hours). In order to take into account the temporal correlation, time-series prediction models can be used. Here, For the purpose of day-ahead scheduling of power system, an initial prediction of random variables should be performed using ANN. Other forecast tools such as ARMA model are reported [6, 24], but ANN is preferred due to its capability of reflecting nonlinear relations among the time-series samples and better performance for long-term applications. Afterwards, the distribution of forecast errors is analyzed to determine the confidence interval around the forecasted values for the upcoming potential wind speed and solar irradiation data on the scheduling day.

The forecast process is performed using two Multi-Layer Perceptron (MLP) neural networks [25] for wind speed and solar irradiation. Each network is configured with three layers including one hidden layer. The input is a 24 hour structure in which a vector of 90 data samples forms its arrays (representing each hour of the day for three month of a season). The available data is divided into three groups proportional to 70%, 15% and 15% for training, validation and test steps, respectively. The hourly data of wind speed and solar irradiation for the first farm are presented in Figure 7 and Figure 8, respectively. The plots of forecast results along with the actual data for one week are followed in Figure 9 and Figure 10.

2.5. Estimation of the confidence interval and risk analysis for wind speed and solar irradiation scenarios

From the power system planning viewpoint, the important aim is to reduce as far as possible the uncertainty and risk associated with generation and power supply. The risk is more crucial when the ex-ante planned generation is less than the ex-post actual generation. The error in the forecast data can be estimated with a level of confidence (LC), in order to determine a reliable level of generation to be considered in the planning stage. Here, the confidence interval method [26], known in risk assessment problems is proposed as a constraint to specify a lower and upper band for the wind speed and solar irradiation scenarios. For example, $LC=90\%$ means that the probability of forecast error ($Power_{risk}$) being less than a definite value obtained from the distribution of forecast error

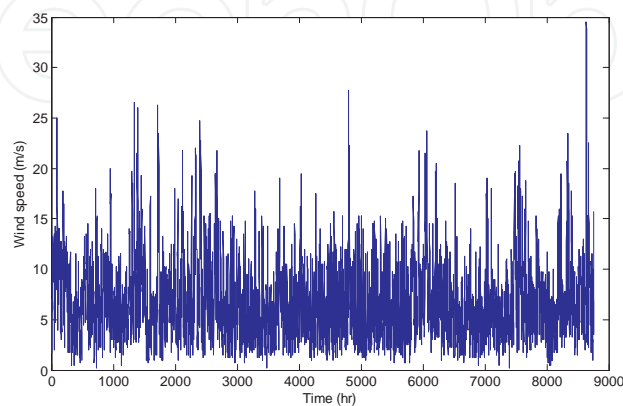


Figure 7. Hourly wind speed data in farm 1

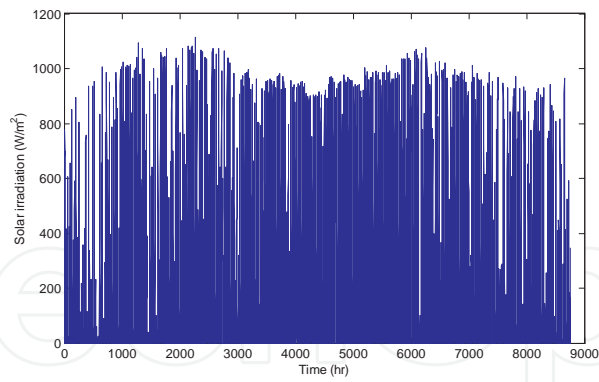


Figure 8. Hourly solar irradiation data in farm 1

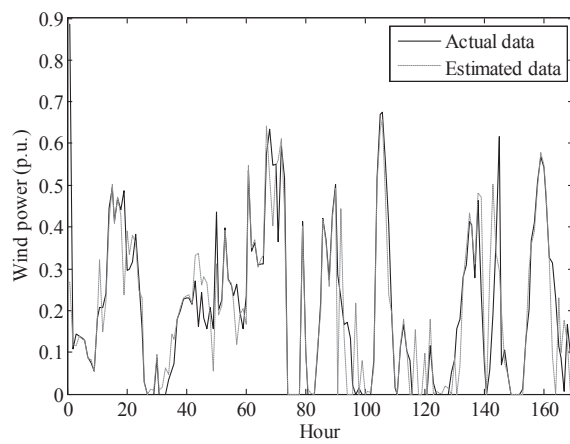


Figure 9. forecast results of wind power along with the actual data for one week

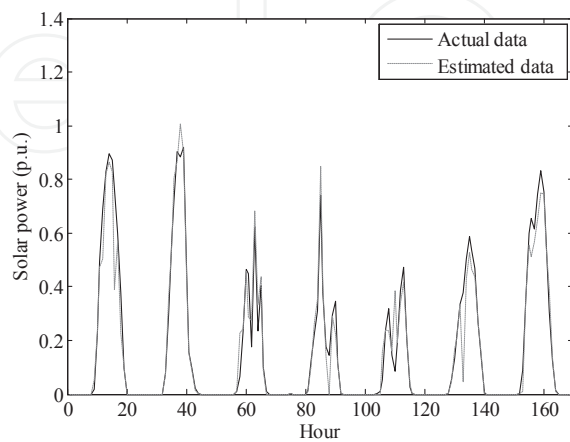


Figure 10. forecast results of wind power along with the actual data for one week

($Power_{risk,max}$) is more than 90%. The value of $Power_{risk,max}$ is referred to as the "Value at Risk" (VAR).

$$P [Power_{risk} \leq Power_{risk,max}] > LC \tag{11}$$

In order to calculate the VAR of forecast based on the defined LC, the PDF of forecast error should be analyzed. In fact, the maximum value of $Power_{risk,max}$ that satisfies equation (11) is chosen as the VAR. Given the forecast error PDF to be a normal distribution with mean μ_e and standard deviation σ_e , the VAR is calculated by (Figure 11):

$$\alpha = 100 - LC \tag{12}$$

$$P[e \geq \mu_e + z_\alpha \sigma_e] < \frac{\alpha}{100} \tag{13}$$

$$Power_{risk,max} = \tilde{e} = \mu_e + z_\alpha \sigma_e \tag{14}$$

where \tilde{e} is the maximum forecast error at which the condition is satisfied. The error is expressed in terms of μ_e and σ_e where z_α is the coefficient of σ_e . Then, this value is subtracted from and added to the forecasted data to give the lower band and upper band of the confidence interval, respectively. This band limits the scenarios generated for wind speed and solar irradiation and provides a reliable range for generation of scenarios that will be considered in the stochastic optimization process. Figure 12 and Figure 13 present an example of the distribution of forecast error of wind speed and solar irradiation at 11 AM, respectively. A representation of the confidence interval method applied to the wind speed data is depicted in Figure 14.

2.6. Scenario generation and reduction

The final step of data processing before performing the stochastic optimization procedure is the scenario generation and reduction. This makes the main distinction between the deterministic programming and the stochastic programming approach. The deterministic programming deals with determined inputs and pre-defined parameters of the system model, whereas the stochastic programming combines the process of assignment of optimum values to the control variables with the stochastic models of the existing random variables. The variables with stochastic behavior are indeed represented by a bunch of scenarios reflecting the most probable situations that is likely to occur for them. Here, in the scenario generation step, using the multivariate distribution function obtained, a large set of random vector samples will be generated using the Monte Carlo simulation. Then, since all of the generated scenarios may not be useful and some of them may exhibit similarities and correlation with

other scenarios, the scenario reduction techniques are applied to eliminate low-probability scenarios and merge the similar ones to extract a limited number of scenarios keeping the whole probable region of the variables covered. Furthermore, the scenario reduction technique increases the computational efficiency of the optimization process. The most well-known methods for scenario reduction are fast backward, fast forward/backward and the fast forward method [27-28]. The first step in scenario reduction is clustering. By clustering, the scenarios which are close to each other are put in one cluster. In the following, we present a description on the fast backward method.

The backward method is initialized by selection of scenarios as the candidates to be eliminated. The selection criterion is based on the minimum value for the product of each sample's probability by the probabilistic distance of that sample to others. The probabilistic distance of each sample to others is considered as the minimum distance of that sample to each of the other samples in the same set. In the next step, the same analysis is performed on the remaining scenarios, however, in this step the product of the eliminated scenarios probabilities in the previous step by the distance of the current sample to other samples are also included. This process continues until the probabilistic distance obtained in each step (iteration) would be less than a predefined value as the convergence criterion. Mathematically speaking, the algorithm can be summarized as follows:

STEP [0]:

$$J^{[0]} := 0$$

$$Z^i = p_i \min_{i \neq j} \|\xi^i - \xi^j\|^r; \quad i = 1, \dots, N$$

$$J^{[1]} = \text{Min } Z^i \quad ; \quad i = 1, \dots, N$$

STEP [i]:

$$j_i \in \arg \text{Min}_{J^{[i-1]}} \sum_{j \in J^{[i-1]} \cup i} p_k \text{Min}_{j \in J^{[i-1]} \setminus i} \|\xi^k - \xi^i\|^r \quad (15)$$

$$\text{if } \text{Min}_{J^{[i-1]}} \sum_{j \in J^{[i-1]} \cup i} p_k \text{Min}_{j \in J^{[i-1]} \setminus i} \|\xi^k - \xi^i\|^r \leq \varepsilon$$

$$J^{[i]} = J^{[i-1]} \cup j_i$$

$$\text{Else } J = J^{[i-1]}$$

END

where N is the total number of scenarios, n is the number of remaining scenarios, and $|J| = N - n$. ξ^i is scenario i with probability p_i . Here, 1000 initial scenarios are generated based on the obtained Copula multi-variable distribution function within the calculated confidence interval. Then, the scenarios are reduced to 20 uncorrelated scenarios using SCENRED function in GAMS software. Figure 15 and Figure 16 demonstrate the final set of scenarios for a day of scheduling.

The power generation scheduling problem including units with uncertain and volatile characteristic is commonly treated as a stochastic optimization problem. In this approach, the objective function calculation is repeated per all final scenarios in each iteration of the optimization process, where the summation of these objective values commonly defines the

overall objective function value to be optimized. As a comparison, the mean square error (MSE) of the calculated scenarios with respect to the real data is calculated in two cases. In the first case, the scenarios are obtained from the Copula distribution and within the calculated confidence interval, based on the proposed framework. In the second case, the scenarios are generated from the single-variable distributions without modeling the underlying temporal and spatial correlations. Table 2 exhibits that the error in the first case is less than that in the second case in three days under study.

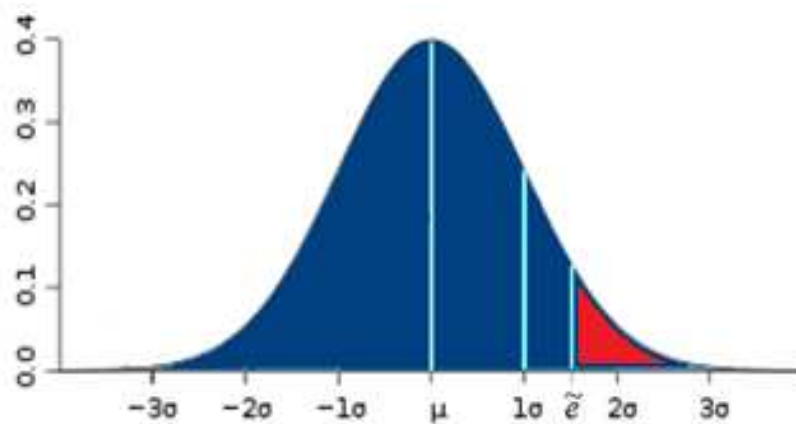


Figure 11. Value at risk calculation based on the distribution of forecast error

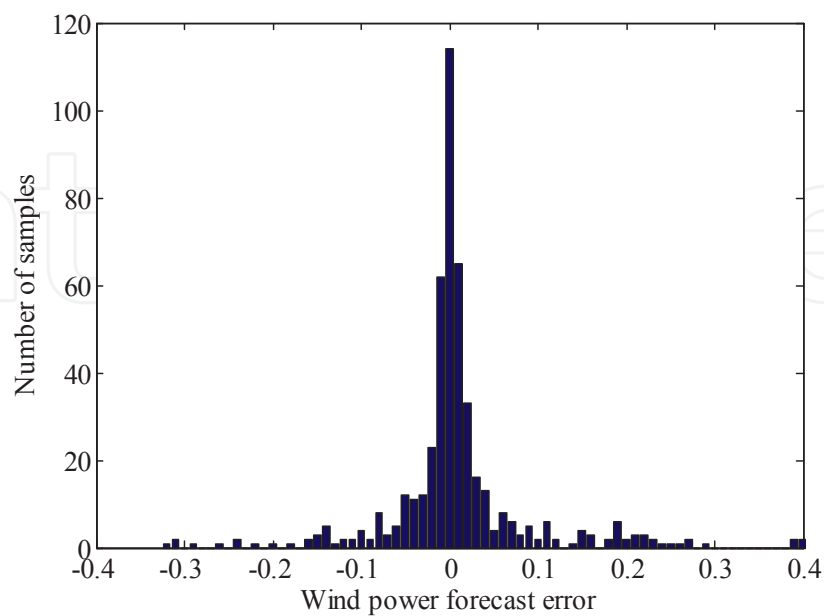


Figure 12. Distribution of error for wind speed forecast at 11 AM

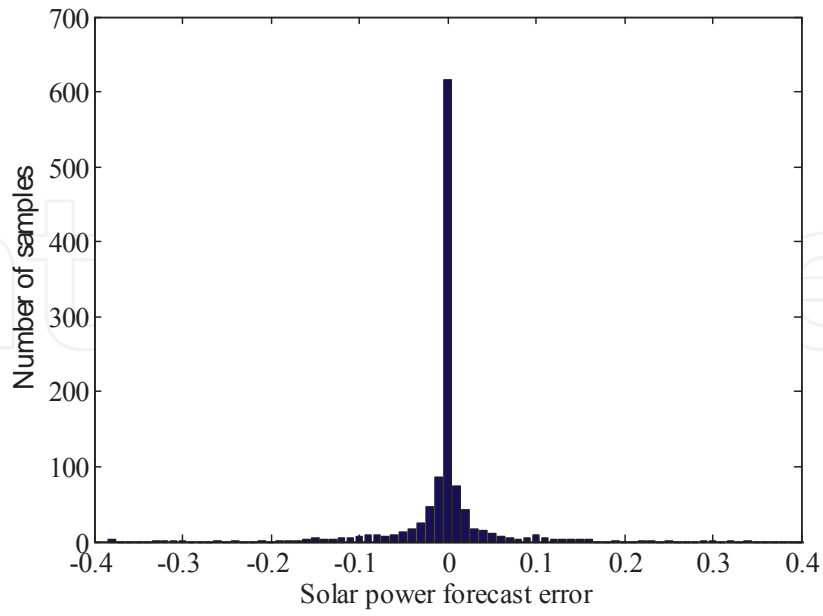


Figure 13. Distribution of error for solar irradiation forecast at 11 AM

| MSE in Case 2 | MSE in Case 1 | |
|---------------|---------------|-------|
| 1.97e-3 | 1.4e-3 | Day 1 |
| 5.99e-3 | 4.05e-3 | Day 2 |
| 4.11e-3 | 3.31e-3 | Day 3 |

Table 2. MSE in Case 1 and 2 for the total power of two wind farms

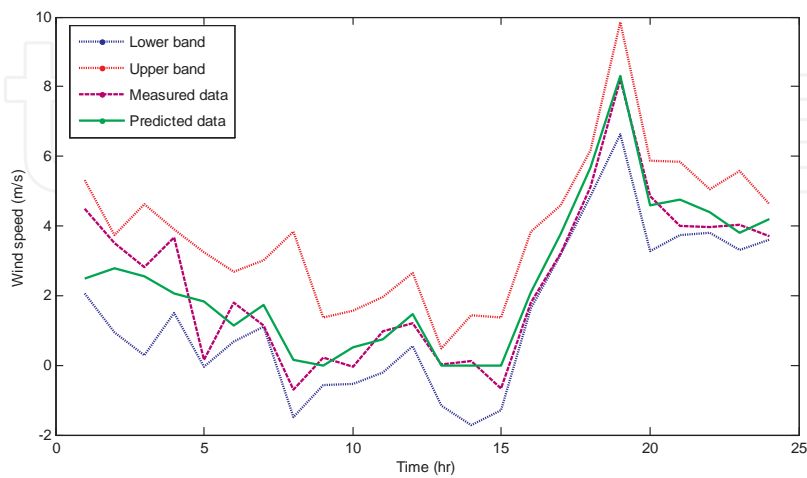


Figure 14. Upper and lower band of confidence interval for wind speed data on one day

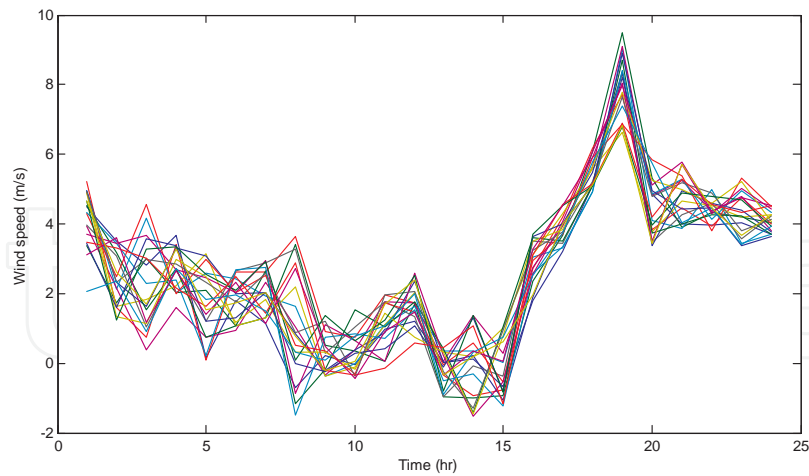


Figure 15. Final scenarios of wind speed within its confidence interval

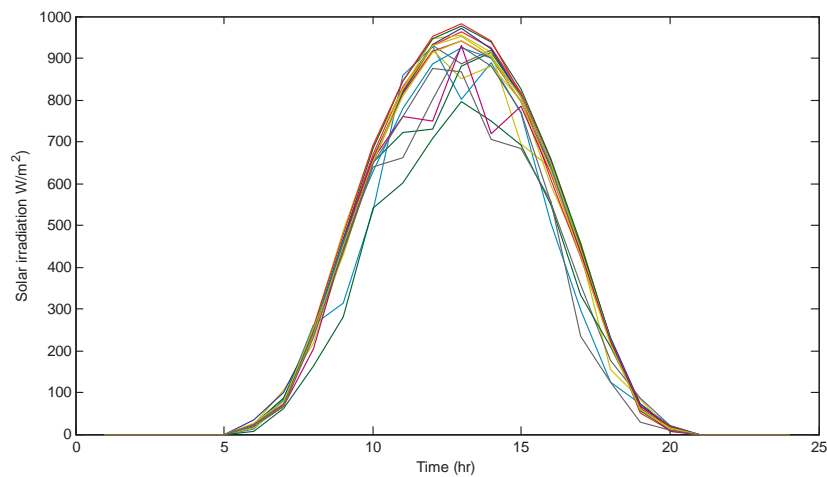


Figure 16. Final scenarios of solar irradiation within its confidence interval

3. Conclusion

Natural characteristics of wind and solar energy impose uncertainty in their design and operation. Hence, considering various possible scenarios in the model of these resources can lead to more realistic decisions. The uncertain parameters are expressed by probability distributions, showing the range of values that a random variable could take, and also accounting for the probability of the occurrence of each value in the considered range. Therefore, the way the random processes are modeled in terms of their PDF is a significant problem. The possible spatial correlations have been addressed and shown effective using the Copula

method. Similarly, the possible temporal correlations have been taken into account using a time-series analysis. In summary, the overall framework can be listed as follows:

1. Take historical data of random variables;
2. Data normalization;
3. Calculate the PDF for each random variable using KDE;
4. Calculate multivariate Copula PDF;
5. Forecast the data time-series over the scheduling horizon;
6. Calculate the confidence interval of potential scenarios for the vector of random variables based on error analysis of forecast data;
7. Generate initial set of scenarios within the confidence interval obtained from the previous step, based on the PDF of step 4;
8. Perform scenario reduction;
9. Perform stochastic optimization based on the final set of scenarios for each random process.

Author details

Sajjad Abedi¹, Gholam Hossein Riahy¹, Seyed Hossein Hosseinian¹ and Mehdi Farhadkhani²

1 Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran

2 Electricity Markets Research Group, Niroo Research Institute, Tehran, Iran

References

- [1] 20% "Wind Energy By 2030: Increasing Wind Energy's Contribution to U.S. Electricity Supply," Office of Energy Efficiency and Renewable Energy DOE/GO-102008-2567, 2009.
- [2] REN21 Steering Committee, Renewables Global Status Report 2009 Update [Online]. Available: http://sitedev.cci63.net/europe/actu/2009/0609/doc/energie_renov2009_Update.pdf
- [3] S. Abedi, *et al.*, "Risk-Constrained Unit Commitment of Power System Incorporating PV and Wind Farms," *ISRN Renewable Energy*, vol. 2011, 2011.
- [4] S. Abedi, *et al.*, "A comprehensive method for optimal power management and design of hybrid RES-based autonomous energy systems," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 1577-1587, 2012.

- [5] S. Abedi, *et al.*, "Risk-Constrained Unit Commitment of Power System Incorporating PV and Wind Farms," *ISRN Renewable Energy*, vol. 2011, p. 8, 2011.
- [6] K. Methaprayoon, *et al.*, "An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty," *Industry Applications, IEEE-Transactions on*, vol. 43, pp. 1441-1448, 2007.
- [7] B. C. Ummels, *et al.*, "Impacts of wind power on thermal generation unit commitment and dispatch," *Energy Conversion, IEEE Transactions on*, vol. 22, pp. 44-51, 2007.
- [8] C. L. Chen, "Optimal wind-thermal-generating unit commitment," *Energy Conversion, IEEE Transactions on*, vol. 23, pp. 273-280, 2008.
- [9] M. A. Ortega-Vazquez and D. S. Kirschen, "Estimating the spinning reserve requirements in systems with significant wind power generation penetration," *Power Systems, IEEE Transactions on*, vol. 24, pp. 114-124, 2009.
- [10] E. M. Constantinescu, *et al.*, "A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation," *Power Systems, IEEE Transactions on*, pp. 1-1, 2011.
- [11] [L. V. L. Abreu, *et al.*, "Risk-Constrained Coordination of Cascaded Hydro Units With Variable Wind Power Generation," *Sustainable Energy, IEEE Transactions on*, vol. 3, pp. 359-368, 2012.
- [12] V. S. Pappala, *et al.*, "A stochastic model for the optimal operation of a wind-thermal power system," *Power Systems, IEEE Transactions on*, vol. 24, pp. 940-950, 2009.
- [13] H. Siahkali and M. Vakilian, "Stochastic unit commitment of wind farms integrated in power system," *Electric Power Systems Research*, vol. 80, pp. 1006-1017, 2010.
- [14] R. Billinton, *et al.*, "Unit commitment risk analysis of wind integrated power systems," *Power Systems, IEEE Transactions on*, vol. 24, pp. 930-939, 2009.
- [15] J. W. GOETHE, "Non-Parametric Statistics," 2009.
- [16] G. Marsaglia, *et al.*, "Evaluating Kolmogorov's distribution," *Journal of Statistical Software*, vol. 8, pp. 1-4, 2003.
- [17] A. R. Mugdadi and E. Munthali, "Relative efficiency in kernel estimation of the distribution function," *J. Statist. Res*, vol. 15, pp. 579-605, 2003.
- [18] R. D. Deveaux, "Applied Smoothing Techniques for Data Analysis," *Technometrics*, vol. 41, pp. 263-263, 1999.
- [19] K. Sutiene and H. Pranevicius, "Copula Effect on Scenario Tree," *International journal of applied mathematics*, vol. 37, 2007.
- [20] P. Embrechts, *et al.*, "Correlation and dependence in risk management: properties and pitfalls," *Risk management: value at risk and beyond*, pp. 176-223, 2002.
- [21] U. Cherubini, *et al.*, *Copula methods in finance* vol. 269: Wiley, 2004.

- [22] D. Berg, "Copula goodness-of-fit testing: an overview and power comparison," *The European Journal of Finance*, vol. 15, pp. 675-701, 2009.
- [23] M. Dorey and P. Joubert, "Modelling copulas: an overview," *The Staple Inn Actuarial Society*, 2005.
- [24] G. Riahya and M. Abedi, "Short term wind speed forecasting for wind turbine applications using linear prediction method," *Renewable Energy*, vol. 33, pp. 35-41, 2008.
- [25] M. T. Hagan, *et al.*, *Neural network design*: PWS Boston, MA, 1996.
- [26] D. Y. Hsu, *et al.*, *Spatial Error Analysis: A Unified, Application-oriented Treatment*: IEEE Press, 1999.
- [27] *GAMS/SCENRED Documentation [Online]*. Available: <http://www.gams.com/docs/document.htm>
- [28] N. G.-K. J. Dupacová, and W. Römisch, "Scenario Reduction in Stochastic Programming: An Approach Using Probability Metrics," *Math. Program*, pp. 493-511, 2000.