Probabilistic Power Flow with Correlated Wind Sources

Juan M. Morales, Luis Baringo, Antonio J. Conejo and Roberto Mínguez[1]

Department of Electrical Engineering, Univ. Castilla – La Mancha, Campus Universitario s/n, 13071 Ciudad Real, Spain

Abstract

This paper proposes a probabilistic power flow model that takes into account spatially correlated power sources and loads. It is particularly appropriate to assess the impact of intermittent generators, such as wind power ones, on a power network. The proposed model is solved using an extended point estimate method that accounts for dependencies among the input random variables, i.e., loads and power sources. The proposed probabilistic power flow model is illustrated through a 24-bus case study. Finally, conclusions are duly drawn.

Keywords: Power flow analysis, wind power, intermittent sources, point-estimate methods, Monte Carlo simulation, uncertainty.

[1]Roberto Mínguez is currently at the Environmental Hydraulics Institute "IH Cantabria", Univ. Cantabria, Avenida de los Castros s/n, 39005 Santander, Spain.

1 Introduction

1.1 Motivation and problem description

The probabilistic power flow (PPF) is an efficacious tool to assess the performance of a power network over most of its working conditions, [1, 2]. Using appropriate descriptions of the input random variables, i.e., active and reactive power demand distribution functions at load buses and, active and reactive power generation distribution functions at generation buses, it is possible to characterize the distribution functions of output random variables such as voltage magnitudes, line flows, etc. The PPF is a relevant planning tool as it allows assessing network functioning over a variety of working conditions in a computationally efficient manner. It is widely used for network planning studies.

Demands and generation sources are considered generally independent from a statistical viewpoint, which results in simple algorithms to solve the PPF. However, demands at different locations throughout a power network are statistically dependent and non-dispatchable sources are dependent too.

Renewable sources, which are essentially intermittent and random, are increasingly relevant in the operation of current electric energy systems, [3,4]. Moreover, these sources are spatially correlated within a given geographical area in a very significant manner, as they are influenced often by the same physical phenomena, e.g., wind levels, [5].

Thus, there is a clear need to enhance current PPF solution algorithms to

include an appropriate treatment of the dependencies of the input variables. This need motivates the work reported in this paper.

1.2 Procedure

We focus on two types of dependent input variables for the PPF. On the one hand, loads, whose correlations are explicitly considered in the methodology proposed. On the other hand, wind power sources, whose spatial correlations are also explicitly taken into account. For the sake of simplicity, other intermittent sources, such as PV plants, are not explicitly considered. Similarly, the availability of conventional production units, such as thermal power plants, is neither modeled. However, note that the uncertainty models pertaining to these sources can be easily incorporated into the methodology proposed in this paper.

We propose a solution technique for the PPF based on an extended point-estimate method. The traditional point-estimate method is extended to consider dependencies among the input random variables. Dependencies are taken into account using suitable linear transformations. Particularly, an orthogonal transformation, which is computationally efficient, is used to transform the set of dependent input random variables into a set of independent ones, which can be processed readily via point-estimate methods. Obtained results are then transformed back to the original space. In a general mathematical framework, the use of an orthogonal transformation to deal with correlated random variables via point estimates was first proposed by Harr [6].

Table 1: Methodologies to solve the PPF

Methodologies	Analytical methods	Approximate techniques	Heuristic procedures	Monte Carlo simulation
Examples	FFT, CM	FOSMM, PEMs	Fuzzy logic	—
References	[8], [9–11]	[12]	[13]	[14,15]

FFT: Fast Fourier Transform; CM: Cumulant Method; FOSMM: First-Order Second-Moment Method; PEMs: Point-Estimate Methods

The proposed technique is validated by comparing it with a cumbersome Monte Carlo procedure, which is fed with statistically dependent samples. The generation of such samples is not straightforward and constitutes a spinoff contribution of this paper. A method to produce correlated samples from a set of normally distributed random variables modeling load uncertainties in power systems can be found in [7]. Nonetheless, in this paper such a method is extended to make it applicable to a much broader set of marginal distributions apart from the Gaussian one.

1.3 Literature review

The probabilistic power flow problem was first introduced by Borkowska in 1974, [1]. Since then, a number of different methodologies have been proposed in the technical literature to solve it in an efficient and accurate manner. Broadly speaking, these methodologies can be classified into the four groups listed in Table 1.

Point-estimate methods [6,16–19] fit into the family of approximate techniques. Su [20] was the first to tackle the probabilistic power flow problem by applying a point-estimate method. Specifically, he used the $2m$ scheme de-

veloped by Hong in [18]. Subsequently, Morales and Pérez-Ruiz [21] pointed out the inadequate performance exhibited by this method if the number of input random variables is large, and advocated the use of an alternative $2m + 1$ scheme to overcome such a limitation. However, both Su [20] and Morales and Pérez-Ruiz [21] sidestepped the problem of how to deal with correlated input variables, a gap that this paper is intended to fill.

1.4 Contributions

The contributions of this paper are fourfold:

1. To formulate and characterize a probabilistic power flow considering statistical dependencies among its input random variables, i.e., power sources and loads.

2. To exhaustively describe an algorithm enabling the application of point-estimate methods to solve the problem in point 1) above.

3. To provide a generalized procedure for the generation of the statistically dependent samples from input random variables required to solve the problem in point 1) via Monte Carlo Simulation.

4. To compare comprehensively the solution methods in points 2) and 3).

1.5 Paper organization

The rest of this paper is organized as follows. Section 2 reviews the PPF, including how to solve it via point-estimate methods. Section 3 develops a

modeling framework to address dependencies affecting the input variables of a PPF, and provides an extended point-estimate technique to solve a PPF with dependent input variables. Additionally, a Monte Carlo simulation procedure is described along with a technique to generate dependent samples to feed the corresponding Monte Carlo algorithm. Results from two case studies, based on the IEEE 24-bus and IEEE 118-bus test systems, are presented in Sections 4 and 5, respectively. Section 6 concludes the paper providing some relevant conclusions.

2 Probabilistic power flow

2.1 Overview

The power flow problem consists in determining the steady-state operating conditions of a power system. In more detail, given the load demanded at consumption buses and the power supplied by generating units, the aim is to obtain the voltages (magnitude and angle) at all system nodes and the active and reactive power flowing through every network branch.

Mathematically, the power flow problem can be represented as a vectorial function where the vector of input variables consists of the power injection at every bus except the slack and the vector of output variables $\boldsymbol{z}$ is made up of all bus voltages (except the voltage at the slack node) and all complex power flows. In fact, from a broader perspective, the power flow solution $\boldsymbol{z}$ is contingent on every factor able to modify the steady-state operating condition of the power system, namely, network topology and generation

and load values. If we denote the set of such factors by $\boldsymbol{p}$, then we can state that $\boldsymbol{z} = \boldsymbol{F}(\boldsymbol{p})$.

The notion of probabilistic power flow comes up from the consideration of the uncertainty intrinsic to the knowledge of the input vector $\boldsymbol{p}$. This uncertainty is transferred to the solution vector $\boldsymbol{z}$ through the vector function $\boldsymbol{F}(\cdot)$. To deal with the variable and uncertain nature of $\boldsymbol{p}$, both the input vector $\boldsymbol{p}$ itself and the solution vector $\boldsymbol{z}$ are treated as vectors of random variables. As a result, the aim of a probabilistic power flow analysis is to characterize the random behavior of the solution $\boldsymbol{z}$ from the statistical information on $\boldsymbol{p}$ that is at the analyst's disposal.

2.2 Solution method

In order to solve the probabilistic power flow problem, we use the $2m + 1$ point-estimate method developed in [18]. This method shows an appropriate performance in the comprehensive comparison carried out in [21].

The aim of any point-estimate method is to compute the moments of each random variable z_i that is function F_i of m input random variables, i.e.,

$$z_i = F_i(p_1, p_2, \ldots, p_m). \tag{1}$$

For this purpose, just commonly available information on the random behavior of input variables p_l, $l = 1, \ldots, m$, is required, in particular, their first few statistical moments. This feature turns point-estimate methods into useful tools for statistical analysis in applications where an accurate characterization of the input random variables is arduous.

The $2m+1$ point-estimate method used in this paper *concentrates* the statistical information provided by the first four central moments of each input random variable p_l, namely, its mean, variance, and coefficients of skewness and kurtosis, into three points often referred to as *concentrations*. More precisely, each point or concentration is a pair $(p_{l,k}, w_{l,k}), k = 1, 2, 3$, made up of a location $p_{l,k}$ at which function $F_i(\cdot)$ is to be evaluated and a weighting factor $w_{l,k}$ measuring the impact of this evaluation on the random behavior of output variable z_i.

2.2.1 Locations and weights

According to the $2m+1$ concentration scheme, each input random variable p_l gives rise to three locations $p_{l,k}$, with k = 1, 2 and 3, expressed as:

$$p_{l,k} = \mu_{p_l} + \xi_{l,k}\sigma_{p_l}, \quad k = 1, 2, 3, \tag{2}$$

where $\xi_{l,k}$ is the standard location, and μ_{p_l} and σ_{p_l} (input data) are the mean and standard deviation of p_l. Standard locations $\xi_{l,k}$ are given by, [18, 20]:

$$\xi_{l,k} = \frac{\lambda_{p_l,3}}{2} + (-1)^{3-k}\sqrt{\lambda_{p_l,4} - \frac{3}{4}\lambda_{p_l,3}^2}, \quad k = 1, 2, \tag{3}$$

$$\xi_{l,3} = 0. \tag{4}$$

Input parameters $\lambda_{p_l,3}$ and $\lambda_{p_l,4}$ are, respectively, the third and fourth standardized central moments of p_l, also known as coefficients of skewness and kurtosis.

Each location $p_{l,k}$ is coupled with a weighting factor $w_{l,k}$ computed as,

[18, 20]:

$$w_{l,k} = \frac{(-1)^{3-k}}{\xi_{l,k}\left(\xi_{l,1} - \xi_{l,2}\right)}, \quad k = 1, 2, \tag{5}$$

$$w_{l,3} = \frac{1}{m} - \frac{1}{\lambda_{p_l,4} - \lambda_{p_l,3}^2}, \tag{6}$$

with m being the number of input random variables involved.

2.2.2 Estimation of means and standard deviations

For each concentration $(p_{l,k}, w_{l,k})$, function $\boldsymbol{F}(\cdot)$ is evaluated at the point consisting of the location $p_{l,k}$ and the means of the $m-1$ remaining input variables. If we denote the solution vector of such an evaluation as $\boldsymbol{Z}(l,k)$, then its i-th component is calculated as follows:

$$\begin{aligned} Z_i(l,k) &= F_i(\mu_{p_1}, \ldots, \mu_{p_{l-1}}, p_{l,k}, \mu_{p_{l+1}}, \ldots, \mu_{p_m}) \\ &= F_i(\mu_{p_1}, \ldots, \mu_{p_{l-1}}, \mu_{p_l} + \xi_{l,k}\sigma_{p_l}, \mu_{p_{l+1}}, \ldots, \mu_{p_m}), \\ & \quad l = 1, 2, \ldots, m, \quad k = 1, 2, 3. \end{aligned} \tag{7}$$

By using weighting factors $w_{l,k}$ and $Z_i(l,k)$ values, the j-th raw moment of the output random variable z_i can be estimated as, [18, 20]:

$$E[z_i^j] \approx \sum_{l=1}^{m} \sum_{k=1}^{3} w_{l,k} (Z_i(l,k))^j, \tag{8}$$

where $E[\cdot]$ stands for the expectation operator. Therefore, approximations of the mean and the standard deviation of z_i, denoted by μ_{z_i} and σ_{z_i}, respec-

tively, can be easily obtained from (8) as follows:

$$\mu_{z_i} = E[z_i] \approx \sum_{l=1}^{m} \sum_{k=1}^{3} w_{l,k}\ Z_i(l,k), \tag{9}$$

$$E[z_i^2] \approx \sum_{l=1}^{m} \sum_{k=1}^{3} w_{l,k} \left(Z_i(l,k)\right)^2, \tag{10}$$

$$\sigma_{z_i} = \sqrt{E[z_i^2] - \left(E[z_i]\right)^2} = \sqrt{E[z_i^2] - \mu_{z_i}^2}. \tag{11}$$

Note that the number of evaluations of function $\boldsymbol{F}(\cdot)$ to be performed is equal to $3m$. However, it can be inferred from expression (4) that the third location of every input random variable p_l coincides with its mean, i.e., $p_{l,3} = \mu_{p_l}$, because $\xi_{l,3} = 0$. As a result, m of these $3m$ evaluations are to be carried out at the same point $(\mu_{p_1}, \mu_{p_2}, \ldots, \mu_{p_l}, \ldots, \mu_{p_m})$. Being so, it is possible, and computationally advantageous, to run just one evaluation of $\boldsymbol{F}(\cdot)$ at that point as long as the associated weight is updated to the value w_0 given by:

$$w_0 = \sum_{l=1}^{m} w_{l,3} = 1 - \sum_{l=1}^{m} \frac{1}{\lambda_{p_l,4} - \lambda_{p_l,3}^2}. \tag{12}$$

This is the reason why this concentration scheme is referred to as $2m+1$ point-estimate method despite the fact that it actually makes use of three points per input variable.

We refer the reader to references [18] and [21] for a complete review on Hong's point-estimate methods. Likewise, references [20] and [21] provide a detailed description of the algorithm to be run for the application of these methods.

2.3 Monte Carlo method

The Monte Carlo method is a well-known computational technique to characterize the random behavior of a physical system by simulating the uncertain nature of its inputs. This technique has been used in the technical literature to solve the probabilistic power flow problem (see, for instance, [15]). Since the underlying algorithm is based on a repeated random sampling, the associated computational burden is usually high. In this sense, point-estimate methods constitute a much more efficient solution approach, [20, 21].

In this paper, the Monte-Carlo method fed with statistically dependent samples is employed as a benchmark to test and validate the results provided by the $2m + 1$ point-estimate scheme.

2.4 Uncertainty characterization

In this paper, two sources of uncertainty are considered, namely, the wind power production and the power consumed at load buses. For the sake of simplicity, we focus the power flow analysis on those variables susceptible of being strongly dependent, and consequently, we treat conventional generation plants (e.g., thermal units) as deterministic. The incorporation of generation forced outages into the PPF problem is straightforward under a point-estimate or Monte Carlo approach (see, e.g., [21]).

The amount and quality of the statistical information required to characterize these sources of uncertainty is different depending on the method, point estimate or Monte Carlo, used to tackle a probabilistic power flow analysis.

In this respect, point-estimate techniques solely require the first few statistical moments of the input random variables. Specifically, the $2m + 1$ scheme just needs the means, standard deviations, and skewness and kurtosis coefficients of such variables. These parameters are generally easy to estimate accurately from historical records. In contrast, the application of the Monte Carlo method necessitates the knowledge of the whole probability distributions of input variables, information that, at least, is hard to get in a precise manner.

In order to use the Monte Carlo technique as a benchmark, we generate samples from Weibull distributions modeling wind speed uncertainty at wind farms, which are subsequently converted into power production by using the power curve of the considered turbine model. This transformation is valid on the assumption that the characteristics of the wind are the same all over the wind plant at each instant. In addition, we also assume that all the turbines comprising wind farms are available. Note that we need to resort to these assumptions due to the lack of public data on the power production of particular wind farms. On the other hand, Normal distributions are used to simulate the active and reactive power consumed at load buses as in [20, 21].

3 Managing spatial correlations

An appropriate modeling of the wind power generation requires recognizing the likely spatial correlation existing among wind sites. This correlation can have a significant impact on the reliability and security of power systems [22, 23], an impact that a probabilistic power flow analysis should

reflect. Likewise, correlation among loads in power systems is apparent and needs to be taken into account. For this purpose, an enhanced $2m+1$ point-estimate method capable of managing the spatial correlations among loads and among intermittent sources is introduced in this paper. This improvement is achieved through an orthogonal (rotational) transformation that allows us to convert the set of correlated input variables into an uncorrelated one. Equations (2)–(6) are then applied to this uncorrelated set of input variables, thus obtaining the *transformed* concentrations. Finally, these concentrations are *untransformed*, and from this point on, the rest of the algorithm runs conventionally as described in [21]. The above process is described in detail below.

3.1 Orthogonal transformation

Let us consider m random variables:

$$\boldsymbol{p} = \begin{pmatrix} p_1 & \cdots & p_m \end{pmatrix}^T,$$

with a mean vector $\boldsymbol{\mu_p}$:

$$\boldsymbol{\mu_p} = \begin{pmatrix} \mu_1 & \cdots & \mu_m \end{pmatrix}^T,$$

and a variance-covariance matrix $\boldsymbol{C_p}$:

$$\boldsymbol{C_p} = \begin{pmatrix} \sigma_{p_1}^2 & \sigma_{p_1 p_2} & \cdots & \sigma_{p_1 p_m} \\ \sigma_{p_2 p_1} & \sigma_{p_2}^2 & \cdots & \sigma_{p_2 p_m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p_m p_1} & \sigma_{p_m p_2} & \cdots & \sigma_{p_m}^2 \end{pmatrix}.$$

Superscript T denotes the transpose of a matrix.

In addition, input variables $\boldsymbol{p}$ are further characterized by a matrix of skewness coefficients $\boldsymbol{\lambda_3}$:

$$\boldsymbol{\lambda_3} = diag\left(\lambda_{p_1,3} \quad \cdots \quad \lambda_{p_l,3} \quad \cdots \quad \lambda_{p_m,3}\right)$$

and a matrix of kurtosis coefficients $\boldsymbol{\lambda_4}$:

$$\boldsymbol{\lambda_4} = diag\left(\lambda_{p_1,4} \quad \cdots \quad \lambda_{p_l,4} \quad \cdots \quad \lambda_{p_m,4}\right),$$

where *diag* stands for the operator that yields a diagonal matrix of the elements specified in brackets. Therefore, the non-diagonal elements of matrices $\boldsymbol{\lambda_3}$ and $\boldsymbol{\lambda_4}$ are assumed to be nil, or equivalently, the crossed moments of an order higher than two are disregarded.

Matrix $\boldsymbol{C_p}$ is symmetric by definition. As a result, there always exists a matrix $\boldsymbol{B}$ of an orthogonal transformation through which the set $\boldsymbol{p}$ of correlated variables can be transformed into a new set $\boldsymbol{q}$ of uncorrelated ones as follows:

$$\boldsymbol{q} = \boldsymbol{B}\boldsymbol{p}. \tag{13}$$

According to this transformation, the variance-covariance matrix $\boldsymbol{C_q}$ of the new set $\boldsymbol{q}$ of input variables is equal to the m-dimensional identity matrix $\boldsymbol{I}$.

In most engineering applications, matrix $\boldsymbol{C_p}$, besides being symmetric, is also positive definite, and as such, can be decomposed through the computationally advantageous Cholesky decomposition, which avoids the calculation of eigenvalues and eigenvectors. The Cholesky decomposition can be stated as

$$\boldsymbol{C_p} = \boldsymbol{L}\boldsymbol{L}^T, \tag{14}$$

where $\boldsymbol{L}$ is an inferior triangular matrix whose inverse turns out to be the orthogonal matrix $\boldsymbol{B}$ required for transformation (13), i.e., $\boldsymbol{B} = \boldsymbol{L}^{-1}$, as shown below.

The variance-covariance matrix $\boldsymbol{C_q}$ can be computed as

$$\begin{aligned} \boldsymbol{C_q} &= cov\left(\boldsymbol{q}, \boldsymbol{q}^T\right) = cov(\boldsymbol{B}\boldsymbol{p}, \boldsymbol{p}^T\boldsymbol{B}^T) \\ &= \boldsymbol{B}\, cov(\boldsymbol{p}, \boldsymbol{p}^T)\boldsymbol{B}^T = \boldsymbol{B}\,\boldsymbol{C_p}\boldsymbol{B}^T = \boldsymbol{I}, \end{aligned} \tag{15}$$

where $cov(\cdot, \cdot)$ stands for the covariance operator.

By using the Cholesky decomposition (14), equation (15) yields:

$$\boldsymbol{C_q} = \big(\boldsymbol{B}\boldsymbol{L}\big)\big(\boldsymbol{L}^T\boldsymbol{B}^T\big) = \big(\boldsymbol{B}\boldsymbol{L}\big)\big(\boldsymbol{B}\boldsymbol{L}\big)^T = \boldsymbol{I} \Rightarrow \boldsymbol{B} = \boldsymbol{L}^{-1}.$$

On the other hand, the transformed vector $\boldsymbol{q}$ satisfies the following properties:

1. It has a mean vector $\boldsymbol{\mu_q}$ computed as:

$$\boldsymbol{\mu_q} = \boldsymbol{L}^{-1}\boldsymbol{\mu_p}. \tag{16}$$

2. It has a variance-covariance matrix $\boldsymbol{C_q}$ equal to the identity matrix:

$$\boldsymbol{C_q} = \boldsymbol{L}^{-1}\boldsymbol{C_p}(\boldsymbol{L}^{-1})^T = \boldsymbol{I}. \tag{17}$$

3. It can be easily shown that the coefficients of skewness and kurtosis of its component variables q_l are given by equations (18) and (19), respectively, under the assumption that the crossed moments of an order higher than two are zero.

$$\lambda_{q_l,3} = \sum_{r=1}^{m} \left(L_{lr}^{-1}\right)^3 \lambda_{p_r,3}\, \sigma_{p_r}^3, \tag{18}$$

$$\lambda_{q_l,4} = \sum_{r=1}^{m} \left(L_{lr}^{-1}\right)^4 \lambda_{p_r,4}\, \sigma_{p_r}^4. \tag{19}$$

Scalar L_{lr} represents the element located at the l-th row and r-th column of the matrix $\boldsymbol{L}$ resulting from the Cholesky decomposition (14).

3.2 Point-estimate for correlated input variables

The algorithm to solve the probabilistic power flow problem with correlated input variables by means of the $2m+1$ point-estimate method is as follows:

1. Given the variance-covariance matrix $\boldsymbol{C_p}$ of the input random variables, obtain the matrix $\boldsymbol{B}$ of the orthogonal transformation (13) by using the Cholesky decomposition (14).

2. Transform the first four central moments of the input variables as stated in equations (16)–(19).

3. Compute the concentrations $(q_{l,k}, w_{l,k})$ defining the $2m+1$ scheme in the transformed space according to equations (2)–(6).

4. Construct the $2m+1$ *transformed* points in the form $(\mu_{q_1}, \ldots, q_{l,k}, \ldots, \mu_{q_m})$.

5. Take the points defined in step 4 above to the original space by applying the inverse transformation of (13), i.e., $\boldsymbol{p} = \boldsymbol{B}^{-1}\, \boldsymbol{q}$.

6. Solve a deterministic power flow problem for each one of the points resulting from step 5). This steps yields the solution vectors $\boldsymbol{Z}(l,k)$.

7. Estimate the means and standard deviations of output random variables z_i as expressed in (9) and (11).

3.3 The Monte Carlo method for correlated input variables

Dealing with correlated input variables in the Monte Carlo method entails the generation of vectors of random numbers preserving both the correlation and the marginal distributions of such variables. To this end, we propose a procedure based on a *normal* transformation through which these vectors of random numbers are obtained from samples of correlated standard normal distributions. The crux of the matter lies then in the determination of the correlation matrix characterizing these standard normal distributions. We provide below the theoretical guidelines required to build the suggested procedure.

We can standardize the vector of input random variables $\boldsymbol{p}$ as follows:

$$\boldsymbol{p'} = \boldsymbol{D}_{\boldsymbol{p}}^{-\frac{1}{2}}(\boldsymbol{p} - \boldsymbol{\mu_p}),$$

where $\boldsymbol{D_p} = diag(\sigma_{p_1}^2, \ldots, \sigma_{p_m}^2)$. In consequence, the new set $\boldsymbol{p'}$ of input variables has a zero mean vector and a variance-covariance matrix $\boldsymbol{C_{p'}}$:

$$\boldsymbol{C_{p'}} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & 1 & \cdots & \rho_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & 1 \end{pmatrix},$$

with $-1 \leq \rho_{lr} \leq 1, l = 1, \ldots, m$, and $r = 1, \ldots, m$. Matrix $\boldsymbol{C_{p'}}$ is also known as the correlation coefficient matrix of the original set of input variables $\boldsymbol{p}$, and as such, is renamed as $\boldsymbol{R_p}$ hereinafter, i.e., $\boldsymbol{R_p} = \boldsymbol{C_{p'}}$. In the context of this paper, the non-diagonal elements ρ_{lr} $(l \neq r)$ of matrix $\boldsymbol{R_p}$ constitute a non-dimensional measure of the spatial interrelations existing among wind sites or among loads.

Let us consider next a series of standard normal variables $\boldsymbol{y}$ =(y_1, ..., y_l, ..., y_m) obtained by marginal transformations of the original set of input variables $\boldsymbol{p} = (p_1, \ldots, p_l, \ldots, p_m)$:

$$y_l = \Phi^{-1}\left[H_{p_l}(p_l)\right], \tag{20}$$

where H_{p_l} is the cumulative distribution function (CDF) of the input variable p_l and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable (with zero mean and unit standard deviation). The aim is to create samples of $\boldsymbol{p}$ from samples of $\boldsymbol{y}$ by reverting the *normal* transformation (20), i.e.:

$$p_l = H_{p_l}^{-1}\left[\Phi(y_l)\right]. \tag{21}$$

The m-dimensional normal vector $\boldsymbol{y}$ is characterized by a correlation coefficient matrix $\boldsymbol{R_y}$ with the following structure:

$$\boldsymbol{R_y} = \begin{pmatrix} 1 & \rho'_{12} & \cdots & \rho'_{1m} \\ \rho'_{21} & 1 & \cdots & \rho'_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \rho'_{m1} & \rho'_{m2} & \cdots & 1 \end{pmatrix}.$$

It is possible to establish a multiplicative factor G relating each correlation coefficient ρ'_{lr} in $\boldsymbol{R_y}$ to its corresponding counterpart ρ_{lr} in $\boldsymbol{R_p}$, i.e.:

$$\rho'_{lr} = G(\rho_{lr})\ \rho_{lr}. \tag{22}$$

In fact, factor $G(\cdot)$ is a function of the correlation coefficient ρ_{lr} itself, the type of probability distributions associated with the correlated variables p_l and p_r, and the parameters determining such distributions. Reference [24] provides mathematical expressions to compute the value of G for a number of probability distributions.

Note that $\boldsymbol{R_y}$ and $\boldsymbol{R_p}$ are the two sides of the same coin, i.e., $\boldsymbol{R_y}$ is the transform of $\boldsymbol{R_p}$ through (20), and conversely, $\boldsymbol{R_p}$ is the transform of $\boldsymbol{R_y}$ by (21). This means that every sample of standard normal distributions correlated according to $\boldsymbol{R_y}$ is transformed into a sample of $\boldsymbol{p}$ by using (21). Therefore, once the correlation matrix $\boldsymbol{R_y}$ has been obtained by applying the transform (22) to the elements ρ_{lr} of the input matrix $\boldsymbol{R_p}$, the problem boils down to generating samples from a set of m standard normal variables that are precisely correlated in accordance with $\boldsymbol{R_y}$. However, this is a simple task to accomplish by means of the orthogonal transformation studied in Section 3.1. Specifically, we can use this transformation to produce such samples from a vector $\boldsymbol{w}$ of independent standard normal variables.

Therefore, the algorithm to solve the probabilistic power flow problem with correlated input variables via the Monte Carlo method is as follows:

1. Given the marginal distributions of input variables p_l and their associated correlation matrix $\boldsymbol{R_p}$, build matrix $\boldsymbol{R_y}$ by using (22). The value of G for each conversion $\rho_{lr} \rightarrow \rho'_{lr}$ can be obtained from the expressions provided in [24]. In particular, if p_l and p_r follow both Weibull distributions, factor G is given by:

$$\begin{aligned} G =& 1.063 - 0.004\rho_{lr} - 0.200(\gamma_l + \gamma_r) - 0.001\rho_{lr}^2 + 0.337(\gamma_l^2 + \gamma_r^2) \\ &+ 0.007\rho_{lr}(\gamma_l + \gamma_r) - 0.007\gamma_l\gamma_r, \end{aligned} \tag{23}$$

   where γ_l and γ_r are, respectively, the coefficients of variation of p_l and p_r, i.e., $\gamma_l = \frac{\sigma_{p_l}}{\mu_{p_l}}$ and $\gamma_r = \frac{\sigma_{p_r}}{\mu_{p_r}}$. According to [24], the maximum error incurred by using expression (23) above is below 2.6% with respect to

the exact conversion provided that $0.1 \leq \gamma_l, \gamma_r \leq 0.5$.

If p_l and p_r are normally distributed, the value of G is equal to 1.

2. Apply the Cholesky decomposition to $\boldsymbol{R_y}$ in order to obtain an orthogonal matrix $\boldsymbol{B}$ such that $\boldsymbol{w} = \boldsymbol{B}\boldsymbol{y}$, with $\boldsymbol{w}$ being a vector of independent standard normal variables.

3. Generate a sample $\boldsymbol{w}_s = (w_{1,s}, \ldots, w_{l,s}, \ldots, w_{m,s})$ of independent standard normal variables. This is a common function included in most commercially available software for statistical analysis.

4. Correlate, according to $\boldsymbol{R_y}$, the sample $\boldsymbol{w}_s$ produced in step 3) above by reversing the orthogonal transformation, that is: $\boldsymbol{y}_s = \boldsymbol{B}^{-1}\boldsymbol{w}_s$.

5. Apply transformation (21) to $\boldsymbol{y}_s$ so as to finally obtain the sample $\boldsymbol{p}_s$ of the original vector $\boldsymbol{p}$ of correlated input variables. This step is graphically represented by the bold path in Fig. 1 and is mathematically expressed as:

$$p_{l,s} = H_{p_l}^{-1}\left[\Phi(y_{l,s})\right].$$

As indicated by the direction of the arrows in this figure, the cumulative probability of the standard normal sample $y_{l,s}$ is first computed. The inverse cumulative distribution function of input variable p_l is then evaluated for this probability, thus obtaining the target sample $p_{l,s}$.

6. Solve a deterministic power flow problem for the sample vector $\boldsymbol{p}_s$ obtained in 5).

7. Repeat steps 3)-6) until a sufficient number N_S of simulations ($s = 1, 2, \ldots, N_S$) is performed.

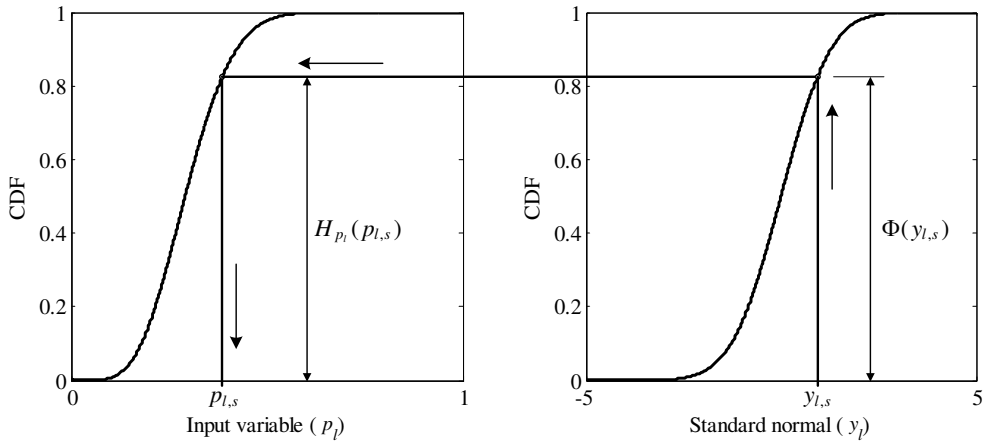

Figure 1: Graphical representation of the marginal transformation (21).

8. Compute means, standard deviations and any other statistical information of interest.

4 24-bus case study

Results from a test case based on the IEEE 24-bus Reliability Test System [25] are presented in this section. This system consists of 24 buses, 38 lines, 32 generating units, and 17 loads. The active power consumed by each load is assumed to be normally distributed with means equal to the values provided in Table 5 of [25], and standard deviations of 5 % with respect to such mean values (i.e., with coefficients of variation, CV, equal to 5%). The reactive power consumption of each load is such that its power factor is kept constant. Load power factors can be also computed from the bus load data provided in the aforementioned table.

We distinguish two regions in the 24-bus system, namely, the northern

region, which is made of nodes 1–10, and the southern one, comprising buses 11–24. Loads situated within the same area are correlated with a correlation coefficient of 0.9. This correlation drops up to 0.5 for loads in different regions.

Two 299-MW wind farms having 130 2.3-MW commercial wind generators each (model Nordex N90/230 with a hub height of 80 m) are located at bus 7. The same Weibull distribution, with scale and shape parameters, α and β, equal to 9 and 2.025, respectively, is used to model wind speed at both sites. Both wind farms are correlated with a correlation coefficient of 0.9. The power curve of the considered turbine model can be found in the manufacturer data base and is provided in [26].

For the sake of simplicity and without loss of generality, we assume that wind plants and loads are uncorrelated.

The characteristics listed above for the 24-bus case study apply integrally hereinafter, unless explicitly stated otherwise.

4.1 Performance appraisal

The accuracy and efficiency of the proposed point-estimate methodology to handle correlated input variables is assessed by comparing the obtained results with those provided by the Monte Carlo method with 15,000 samples. This number of samples is sufficiently high to yield estimates for means and standard deviations with a degree of precision (coefficient of variation of estimates expressed in percentage) below 0.9 and 1.3 %, respectively.

Table 2: Average relative errors (%). 24-bus case study. CV = 5%

	V_i	δ_i	P_i	Q_i	P_{i-j}	Q_{i-j}
$\bar{\varepsilon}_\mu$	0.0024	0.3218	0.0085	0.1806	0.2650	0.2267
$\bar{\varepsilon}_\sigma$	2.0023	0.5652	0.0274	3.9465	0.4056	1.6057

In order to offer a general and concise overview of the conclusions drawn from the comparison, we present the average relative errors [21] for means ($\bar{\varepsilon}_\mu$) and standard deviations ($\bar{\varepsilon}_\sigma$) computed in terms of percentages of the solution values obtained from the Monte Carlo simulation. Table 2 provides the values of these errors associated with the estimates of voltages (V_i), angles (δ_i), active power injections (P_i), reactive power injections (Q_i), active power flows (P_{i-j}) and reactive power flows (Q_{i-j}). It should be noted that all these errors are smaller than 4 %, which highlights the good performance of the proposed point-estimate methodology. Note also that the average relative errors associated with the estimation of standard deviations are greater than those corresponding to the means, which is in accordance with the common behavior of point-estimate methods, whose accuracy worsens as the order of the estimated statistical moment becomes higher.

On the other hand, the CPU times required by the point-estimate method and the Monte Carlo simulation to compute the means and standard deviations of output variables are 0.46 and 164.61 seconds, respectively. That is, the Monte Carlo method takes about 350 times longer than the proposed point-estimate algorithm. Therefore, we confirm the common conclusion that the point-estimate methodology is computationally much more efficient than the Monte Carlo technique. Both methods have been implemented in MAT-

Table 3: Average relative errors (%). 24-bus case study. CV = 10%

	V_i	δ_i	P_i	Q_i	P_{i-j}	Q_{i-j}
$\bar{\varepsilon}_\mu$	0.0061	1.1555	0.8375	0.5915	0.3958	0.6917
$\bar{\varepsilon}_\sigma$	2.1428	1.3095	1.1815	4.1724	0.4553	6.5194

LAB on a Intel Pentium 1.60-GHz PC with 1 GB RAM. MATPOWER [27] has been employed to solve the deterministic power flow problems.

In order to appraise the impact of the uncertainty level on the performance of the proposed PEM, Table 3 provides the estimation errors in average and relative terms for a coefficient of variation of loads equal to 10%. From the comparison between Tables 2 and 3, it can be inferred that the uncertainty level involved in the probabilistic power flow problem has a detrimental effect on the accuracy of the estimates yielded by the proposed solution algorithm, which is, on the other hand, in line with the general results reported in the technical literature in this respect. Note that if the coefficients of variation of loads are doubled from 5% to 10%, the maximum average relative error increases from 3.95 to 6.52%, approximately.

Fig. 2 illustrates the standard deviation of the active power flow through line 7-8 estimated with both the proposed point-estimate method and the Monte Carlo simulation as a function of the correlation coefficient between wind farms. The 95 % confidence limits for the standard deviation estimator provided by the Monte Carlo simulation are also depicted. The active power flow in this line is the output variable on which correlation between wind farms impacts the most. In this sense, note that if such a correlation is increased from 0.02 to 0.93 the standard deviation of this flow grows from 158.4

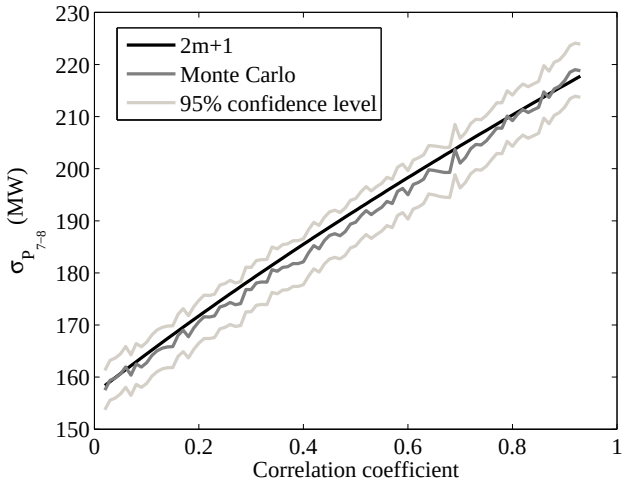

Figure 2: Correlation effect on the standard deviation of active power flow through line 7-8

to 217.8, which means a 37.5% of increase. On the other hand, observe that the Monte Carlo method exhibits an oscillatory behavior, which is inherent to this simulation process. As the number of samples used in the simulation increases, the amplitude of oscillations diminishes. Aside from this undesirable, but expected fluctuating behavior of the Monte Carlo method, what is actually important in Fig. 2 is that the point-estimate solution falls within the 95 % confidence interval associated with the standard deviation estimator provided by the Monte Carlo simulation. Therefore, statistically speaking, we can state that the solutions provided by both techniques are the same with a 95 % confidence level, albeit, in practice, the behavior of the standard deviation yielded by the point-estimate method as the correlation coefficient increases is smooth and non-oscillatory, and as such, more convenient.

For the reasons above, just the results obtained from the proposed point-estimate technique are shown henceforth.

4.2 Impact of wind site correlations

Next, we present some results to illustrate how the correlation level between wind farms has a remarkable influence on system conditions. In order to offer a general overview of such influence, we define two indices, namely, the *Voltage Profile Variability* (VPV) and the *Transmission System Available Margin* (TSAM):

$$\text{VPV}(\%) = \sum_{i=1}^{N_B} (1 - V_i)^2 \cdot \frac{1}{N_B} \cdot 100 \tag{24}$$

$$\text{TSAM}(\%) = \sum_{k=1}^{N_L} \frac{S_k^{\max} - S_k}{S_k^{\max}} \cdot \frac{1}{N_L} \cdot 100 \tag{25}$$

where N_B, N_L, S_k, and $S_k^{\max}$ are, respectively, the number of system buses, the number of transmission lines, the apparent power flow magnitude in line k, and the transmission capacity limit of line k.

The VPV measures how far the bus voltages are from the flat profile, while the TSAM quantifies the *room* available in the network to accommodate additional power injections.

In Figs. 3 and 4, the evolution of the standard deviations of indices VPV and TSAM are represented, respectively, as the correlation coefficient between wind sites increases. We should point out that, for a correlation coefficient between 0.10 and 0.17, the standard deviation of the VPV is so small that Equations (9)-(11) yield a non-real number for the estimate of this parameter due to the estimation errors associated with the proposed PEM. In such cases, the estimated value of the standard deviation of the VPV is considered to be zero.

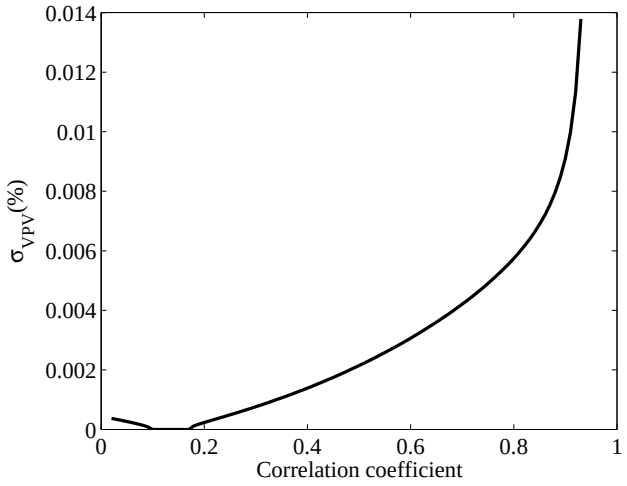

Figure 3: Correlation effect on the standard deviation of the voltage profile variability (VPV)

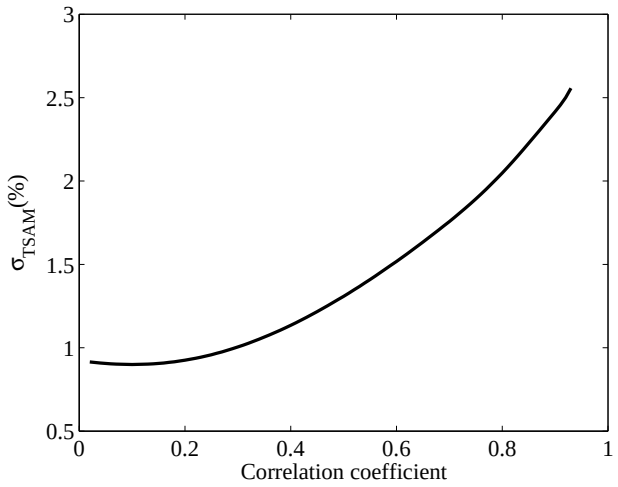

Figure 4: Correlation effect on the standard deviation of the transmission system available margin (TSAM)

The standard deviations can be seen as a measure of the uncertainty level affecting the power system, and therefore, the smaller they are, the better. Means are barely impacted by wind site correlation, with a maximum variation below 0.05 %. On the contrary, the standard deviations of VPV and TSAM are significantly affected by wind site correlation. Observe the marked rise undergone by the standard deviations of VPV and TSAM as the correlation coefficient approaches 1. In numbers, if the correlation coefficient is augmented from 0.02 to 0.93, the standard deviations of the VPV and the TSAM grow from 3.7×10^{-4} to 0.014 and from 0.91 to 2.56, respectively, which translates into relative increases of 3684% and 181%, in that order.

In general, lower/higher correlations among wind sites translate into a lower/higher variability of the total wind power injected into the network.

5 118-bus case study

In this section, the probabilistic power flow problem is solved for the IEEE 118-bus test system [28] with the only purpose of assessing how an increase in the number of uncertain variables affects the performance of the proposed PEM. This system comprises 186 transmission lines, 54 generating units, and 99 loads. We distinguish four areas covering nodes 1-31, 32-58, 59-92, and 94-118. The active power consumed by each load is modeled as a normally distributed random variable with a mean equal to the value provided in [28] and a coefficient of variation of 5%. As in the 24-bus case study above, the reactive power consumption of each load is such that its power factor is kept constant. Moreover, loads are correlated with a correlation coefficient equal

to 0.9 if they are placed within the same area, and equal to 0.5 otherwise.

For comparative purposes, wind speed uncertainty is described by the same Weibull distribution used in the previous 24-bus case study, irrespective of the wind farm location. Likewise, the same wind turbine model is considered.

Two variants of the 118-bus case study are examined, namely:

Variant 1: Two 483-MW wind farms are located at node 59, with a correlation coefficient of 0.9.

Variant 2: Six 161-MW wind farms are connected in pairs to nodes 59, 90, and 116. The correlation coefficient between two wind farms is equal to 0.9 if they are located at the same node, and 0.5 otherwise.

It should be pointed out that, in these two variants, the wind power penetration level is kept around the 20% of the average total system demand, which represents a percentage similar to that considered in the 24-bus case study.

Table 5 shows the average relative errors of the mean and standard deviation estimates provided by the proposed PEM. In this case, 50,000 samples for the Monte Carlo simulation are used in order to guarantee stable results. Note that these errors are of the same order of magnitude as those figuring in Table 2 for the 24-bus case study. Therefore, the number of uncertain input variables does not have, by itself, a clear impact on the performance of the extended $2m+1$ PEM, which is in accordance with the results provided

Table 4: Average relative errors (%). 118-bus case study

		V_i	δ_i	P_i	Q_i	P_{i-j}	Q_{i-j}
Variant 1	$\bar{\varepsilon}_\mu$	0.0001	0.1130	0.0008	0.3599	0.3278	0.1845
	$\bar{\varepsilon}_\sigma$	1.5084	0.4160	3.6607	1.5329	0.4056	0.7563
Variant 2	$\bar{\varepsilon}_\mu$	0.0001	0.1120	0.0033	0.3832	0.3568	0.1823
	$\bar{\varepsilon}_\sigma$	1.9543	0.6113	3.6739	1.4668	0.6636	1.9598

in [21]. Notwithstanding this, observe that the estimations errors pertaining to the variant 2 of the 118-bus case study, characterized by a higher number of wind farms, are, in general, greater than those of variant 1. This is mainly due to the discrete nature of the power generated by a wind plant. Consequently, the growth of the number of wind farms does have a negative, although small, effect on the accuracy of the estimates yielded by the considered PEM.

6 Conclusions

This paper considers a probabilistic power flow and models loads and wind sources as correlated random variables. If a point-estimate method is used, the consideration of these dependencies requires a significantly modified solution algorithm. If, on the other hand, a Monte Carlo simulation procedure is used, an algorithm needs to be devised to generate properly correlated samples. This paper provides both a point-estimate solution algorithm that allows treating correlated random variables and an algorithm to generate correlated samples.

Results from simulations allow deriving the relevant conclusions below:

1. The proposed point-estimate method handles properly correlations and presents high computational efficiency.

2. The uncertainty level affecting the probabilistic power flow problem, measured by the coefficients of variation of the input random variables, has a significant impact on the performance of the extended $2m + 1$ PEM in such a way that larger variations usually lead to larger estimation errors. In contrast, the effect of the number of input random variables on the PEM performance is comparatively small and mostly caused by the number of wind farms in the power system under consideration.

3. The Monte Carlo algorithm fed with properly correlated samples gives the same results as the proposed point-estimate method, but its computational burden is high.

4. The expected values of output variables remain basically unchanged as the correlation among the input variables increases; however, the standard deviations of these output variables increase significantly as the correlation grows. The proposed methodology allows quantifying precisely these increases.

Lastly, the extended point-estimate method described in this paper enables the use of the probabilistic power flow as an efficacious power system analysis tool to address long-term studies such as transmission expansion

planning and reliability analysis related to electric energy systems with a high penetration of wind power.

7 Acknowledgments

J. M. Morales, A. J. Conejo, and R. Mínguez are partly supported by Junta de Comunidades de Castilla – La Mancha through project PCI-08-0102 and by the Ministry of Science and Technology of Spain through CICYT Project DPI2009-09573. Additionally, R. Mínguez is partly supported by the Spanish Ministry of Science and Innovation (MCI) through the "Ramon y Cajal" program.

References

[1] Borkowska, B.: 'Probabilistic load flow', IEEE Trans. Power App. Syst., 1974, PAS-93, (3), pp. 752–759

[2] Dopazo, J. F., Klitin, O. A., and Sasson, A. M.: 'Stochastic load flows', IEEE Trans. Power App. Syst., 1975, PAS-94, (2), pp. 299–309

[3] Hatziargyriou, N. and Zervos, A.: 'Wind Power Development in Europe', Proc. IEEE, 2001, 89, (12), pp. 1765–1782

[4] Global Wind Energy Council: 'Global Wind 2008 Report', available in http://www.gwec.net, accessed January 2010

[5] Morales, J. M., Conejo, A. J., and Mínguez, R.: 'A methodology to generate statistically dependent wind speed scenarios', Appl. Energy, 2010, 87, (3), pp. 843–855

[6] Harr, M. E.: 'Probabilistic estimates for multivariate analysis', Appl. Math. Model., 1989, 13, (5), pp. 313–318

[7] Li, W. and Billinton, R.: 'Effects of Bus Load Uncertainty and Correlation in Composite System Adequacy Evaluation,' IEEE Trans. Power Syst., 1991, 6, (4), pp. 1522–1529

[8] Allan, R. N., Leite da Silva, A. M., and Burchett, R. C.: 'Evaluations methods and accuracy in probabilistic load flow solutions', IEEE Trans. Power App. Syst., 1981, PAS-100, (5), pp. 2539–2546

[9] Allan, R. N., and Al-Shakarchi, M. R. G.: 'Probabilistic techniques in a.c. load flow analysis', 1977, Proc. IEEE, 124, (2), pp. 154–160

[10] Zhang, P., and Lee, S. T.: 'Probabilistic load flow computation using the method of combined cumulants and Gram-Charlier expansion', IEEE Trans. Power Syst., 2004, 19, (1), pp. 676–682

[11] Sanabria, L. A., and Dillon, T. S.: 'Stochastic power flow using cumulants and von Mises functions', Int. J. Electr. Power Energy Syst., 1986, 8, (1), pp. 47–60

[12] Madrigal, M., Ponnambalam, K., and Quintana, V. H.: 'Probabilistic optimal power flow'. Proceedings of the 1998 IEEE Canadian Conference on Electrical and Computer Engineering, Waterloo, Ontario, May 1998, pp. 385–388

[13] Miranda, V., Matos, M. A., and Saraiva, J. T.: 'Fuzzy load flow new algorithms incorporating uncertain generation and load representation'. Proc. 10th Power System Computation Conf., Graz, Austria, 1990, pp. 621–627

[14] Rubinstein, R. Y.: 'Simulation and the Monte Carlo Method' (New York: John Wiley and Sons, 1981)

[15] Leite da Silva, A. M., Arienti, V. L., and Allan, R. N.: 'Probabilistic load flow considering dependence between input nodal powers', IEEE Trans. Power App. Syst., 1984, PAS-103, (6), pp. 1524–1530

[16] Rosenblueth, E.: 'Point estimates for probability moments', Proc. Nat. Acad. Sci., 1975, 72, pp. 3812–3814

[17] Rosenblueth, E.: 'Two-point estimates in probability', Appl. Math. Model., 1981, 5, pp. 329–335

[18] Hong, H. P.: 'An efficient point estimate method for probabilistic analysis', Reliab. Eng. Syst. Saf., 1998, 59, pp. 261–267

[19] Tsai, C. W., and Franceschini, S.: 'Evaluation of probabilistic point estimate methods in uncertainty analysis for environmental engineering applications', J. Environ. Eng.-ASCE, 2005, 131, (3), pp. 387–395

[20] Su, C.-L.: 'Probabilistic load-flow computation using point estimate method', IEEE Trans. Power Syst., 2005, 20, (4), pp. 1843–1851

[21] Morales, J. M., and Pérez-Ruiz, J.: 'Point estimate schemes to solve the probabilistic power flow', IEEE Trans. Power Syst., 2007, 22, (4), pp. 1594–1601

[22] Wangdee, W., and Billinton, R.: 'Considering load-carrying capability and wind speed correlation of WECS in generation adequacy assessment', IEEE Trans. Energy Convers., 2006, 21, (3), pp. 734–741

[23] Xie, K., and Billinton, R.: 'Considering wind speed correlation of WECS in reliability evaluation using the time-shifting technique', Electr. Power Syst. Res., 2009, 79, (4), pp. 687-693

[24] Liu, P.-L., Der Kiureghian, A.: 'Multivariate distribution models with prescribed marginals and covariances', Probab. Eng. Mech., 1986, 1, (2), pp. 105–112

[25] Reliability Test System Task Force: 'The IEEE reliability test system—1996', IEEE Trans. Power Syst., 1999, 14, (3), pp. 1010–1020

[26] Danish Wind Industry Association: 'Wind Turbine Power Calculator', http://www.windpower.org, accessed October 2009

[27] Zimmerman, R., Murillo-Sánchez, C., and Gan, D.: 'MATPOWER, a MATLAB power system simulation package. Version 3.0.0'. School of Electrical Engineering, Cornell University: Power System Engineering Research Center (PSERC), 2005

[28] 'Power System Test Case Archive', http://www.ee.washington.edu/research/pstca, accessed January 2010