Using Copulas for Modeling Stochastic Dependence in Power System Uncertainty Analysis

George Papaefthymiou, *Member, IEEE*, and Dorota Kurowicka

Abstract—The increasing penetration of renewable generation in power systems necessitates the modeling of this stochastic system infeed in operation and planning studies. The system analysis leads to multivariate uncertainty analysis problems, involving non-Normal correlated random variables. In this context, the modeling of stochastic dependence is paramount for obtaining accurate results; it corresponds to the concurrent behavior of the random variables, having a major impact to the aggregate uncertainty (in problems where the random variables correspond to spatially spread stochastic infeeds) or their evolution in time (in problems where the random variables correspond to infeeds over specific time-periods).

In order to investigate, measure and model stochastic dependence, one should transform all different random variables to a common domain, the rank/uniform domain, by applying the cumulative distribution function transformation. In this domain, special functions, copulae, can be used for modeling dependence. In this contribution the basic theory concerning the use of these functions for dependence modeling is presented and focus is given on a basic function, the Normal copula. The case study shows the application of the technique for the study of the large-scale integration of wind power in the Netherlands.

Index Terms—Copula, correlation, Monte Carlo simulation, stochastic dependence, stochastic generation, uncertainty analysis, wind power.

I. INTRODUCTION

THE integration of stochastic generation in the form of renewable energy sources in the power system necessitates the incorporation of uncertainty analysis as a integral part of the power system planning and operation [1]. In system planning, this power generation uncertainty should be quantified for the determination of the variability of the system power flows, which is central for the system dimensioning. In system operation, this uncertainty is translated into forecast uncertainty; the incorporation of forecast uncertainty in the system management is central for the optimal operation of power systems with high penetration of stochastic energy sources. In both cases, the system analysis leads to a multivariate uncertainty analysis

problem, that involves correlated random variables that follow different distributions. To clarify this decomposition into a multivariate uncertainty analysis problem, the case of wind power will be discussed.

For long-term planning studies, Weibull distributions may be used for modeling the wind activity in each generation site. The wind power infeed is further obtained as a nonlinear transformation of the Weibull wind speed distribution (wind speed/wind power characteristic of the wind turbine generators) yielding a nonstandard power distribution. The wind speeds between different locations present a significant degree of correlation due to their mutual dependence on wind activity which is translated into correlation between the system wind power infeeds. Not taking into account this correlation leads to a severe underestimation of the variability of the system power flows and consequently to an underestimation of the risks related to specific design decisions [2].

For the operation of power systems with high penetration of wind power, estimates of the wind power production for a number of hours ahead are used in the form of point forecasts [3]. However, a large part of the research efforts is concentrated on developing methods for giving the prediction uncertainty, most commonly in the form of probabilistic forecasts of wind generation [4]. They correspond to nonparametric probabilistic predictions, i.e., for which no restrictive assumption is made on the shape of predictive densities and may take the form of quantiles, interval or density forecasts. Considering different geographic areas and different prediction horizons, the prediction errors appear to be correlated, due to the movement of meteorological fronts or generally by inertia of meteorological systems [5], [6].

In the related literature, focus is mainly given on the modeling of time-dependent stochasticity in stochastic (renewable) energy sources (see, e.g., [7]–[9]). The problem is treated for each hour separately and the generation uncertainty is generally modeled as a normal distribution with a specific standard deviation around a given mean. In [7], a clustering technique for the grouping and the modeling of stochastic generation is presented, treating stochastic generation in a time-dependent framework. The case of linearly dependent normally distributed nodal power injections was treated in [8] and [9]. An alternative approach for the modeling of linearly correlated system inputs has been presented in [10] and further developed in [11], where the Gram–Schmidt orthogonalization is used in order to transform the linearly correlated system inputs into a weighted sum of independent random variables. The time-dependent approach is suitable for the modeling of the system load, which by nature presents a high time-dependence on seasonality and

daily patterns or for performing rough system analyses where a multistate unit approach is enough for the modeling of stochastic generation [7]. However, in many cases the time information will not lead to a significant reduction of the uncertainty, e.g., the wind power output for working day/noon/summer may vary from zero to nominal, depending on the weather characteristics. In this case, the problem is decomposed to the modeling of correlated non-Normal random variables. This necessitates the employment of algorithms that extend the concept of linear dependence and can be applied to the case of non-Normality.

fwrIn this contribution, a Monte Carlo simulation framework is presented for the treatment of this problem, based on the separation of the modeling of the dependence structure from the marginal distributions. Although the importance of marginal distributions is well recognized, the role of stochastic dependence for the system analysis is crucial, since it corresponds to the concurrent behavior of the respective random variables, having a major impact on their aggregate behavior. By separating the modeling of dependence, a better understanding of its contribution on the output is achieved. The next section contains an introduction to concepts of dependence.

II. ROLE OF STOCHASTIC DEPENDENCE AND IMPACT OF MARGINALS

Stochastic dependence refers to the behavior of a random variable relative to another. In this context, two random variables are independent if the occurrence of a specific value of the one has no impact on the probability of occurrence of any value of the other. On the contrary, perfect positive dependence refers to the situation where the occurrence of high (low) values for the one random variable implies that the other takes also high (low) values. To analyze the role of stochastic dependence, we will further focus on these two dependence concepts.

Let U_1 and U_2 two random variables that are uniformly distributed in [0, 1]. In Fig. 1, the scatter diagrams, the cobweb plots and the distribution of the sum of the random variables for the two dependence concepts are presented. Independence leads to a random matching of the occurrences in the domain of each random variable as can be seen by the cobweb plot of Fig. 1(c), leading to a uniform covering of the area in the scatter plot of Fig. 1(a). Perfect dependence leads to a matching between the occurrences of the random variables as can be seen in the cobweb plot of Fig. 1(d) interpreted as a linear relationship in the scatter plot of Fig. 1(b).

In Fig. 1(e) and (f), one can see how different the distributions of the sum for the two dependence structures are. Both distributions have the same mean value 1 and the same domain [0, 2],[1] but different shapes. This can be easily explained as in the case of independence the occurrence of high (or low) values for both random variables is scarce, while in the case of perfect dependence these values are always combined, leading to a higher occurrence of values towards the tails of the distribution. Respectively, every other dependence structure will lead to a different relative matching of the values and therefore a different distribution for the sum of the random variables.
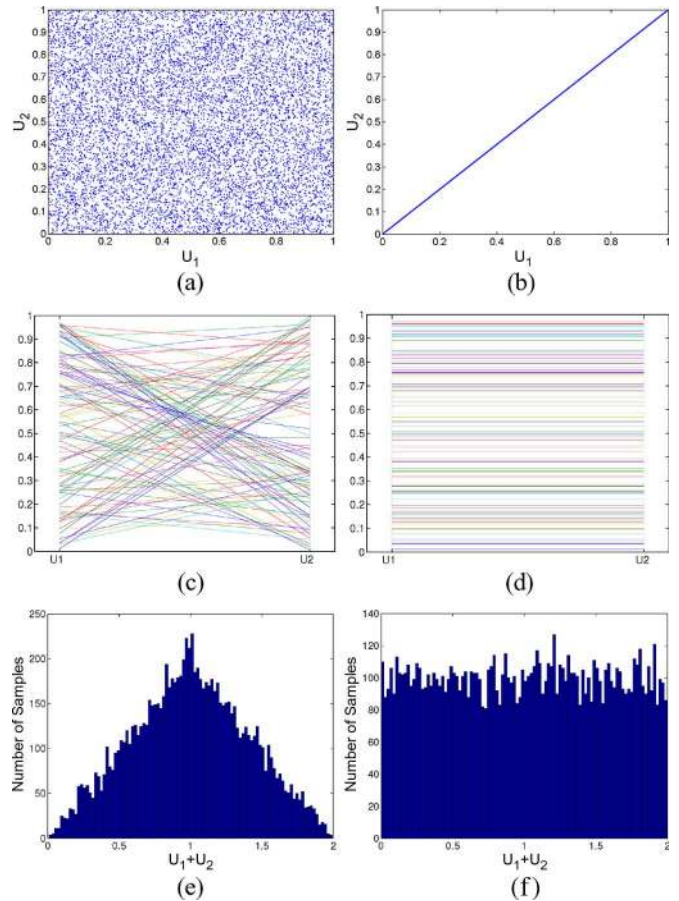

Fig. 1. Independence and perfect dependence between two uniform random variables. (a) Independence-scatter plot. (b) Perfect dependence-scatter plot. (c) Independence-cobweb plot. (d) Perfect dependence-cobweb plot. (e) Independence-sum. (f) Perfect dependence-sum.

The plots in Fig. 1 refer to uniform marginal distributions. In Fig. 2, the scatter diagrams for the same dependence concepts for the case of a Normal N_1 and a Weibull random variable W_2[2] [Fig. 2(a)] are presented. Independence leads to the scatter plot in Fig. 2(b), while the perfect dependence [Fig. 2(c)] gives a monotonic relationship between N_1 and W_2.

Independence is easily recognized for uniform variables [Fig. 1(a)]; however for different marginal distributions scatter plots representing independence would change depending on the marginals [see, e.g., Fig. 2(b)]. The same conclusion is drawn for the case of perfect dependence, where the linear relationship between uniforms [Fig. 1(b)] is becoming a monotonic one between random variables with different marginals [Fig. 2(c)]. Hence, the main efforts in the area of stochastic dependence modeling have been focused on isolating the effect of marginals from the dependence structure.

For stochastic dependence, one is interested in the behavior of the ranks of the respective random variables rather than the actual values they take. Intuitively, two random variables will be highly dependent if "when the one takes large (small) values in its domain, the other one takes large (small) values in its domain as well" and not necessarily if "when the one takes the value

---

[1]As known from basic probability theory, the mean value of the sum of random variables always equals the sum of the mean values of the respective random variables [12].

[2]The mean value and standard deviation for N_1 are $\mu_1 = 20$ and $\sigma_1 = 3$, respectively. The scale and shape parameter for W_2 are $\eta_2 = 11$ and $\beta_2 = 2$, respectively.
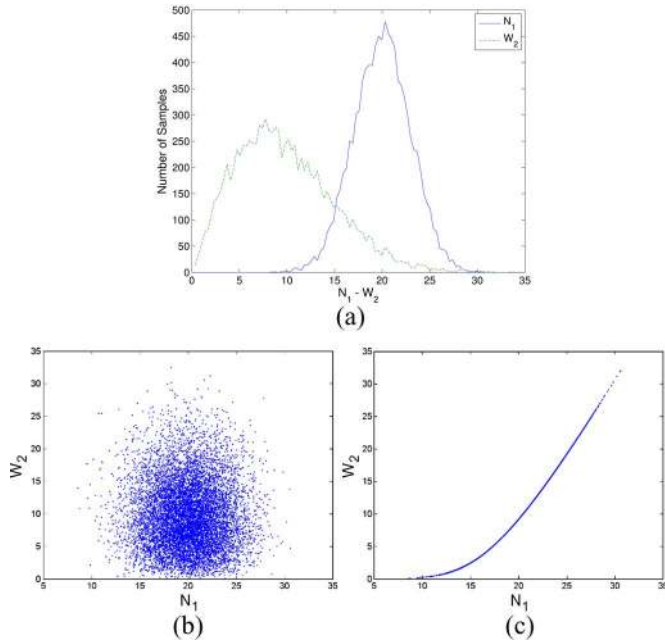
Fig. 2. Independence and perfect dependence between a Normal N_1 and a Weibull W_2 random variable. (a) pdfs. (b) Independence-scatter plot. (c) Perfect dependence-scatter plot.

100, the other takes the value 100 also." Therefore, the measuring and modeling of stochastic dependence between arbitrary random variables should be performed based on the respective ranks of the random variables. By applying specific transformations, all random variables in the problem are converted to ranks or uniforms. In this common uniform/rank domain the measurement and modeling of stochastic dependence takes place. The respective theory is presented in the next section.

III. STOCHASTIC DEPENDENCE MODELING USING COPULAS

A. From Marginals to Uniforms/Ranks: The cdf Transformation

An arbitrary continuous random variable X can be transformed to uniform by the application of the *cumulative distribution function (cdf)* transformation. This uniform variable corresponds to the respective ranks of the distribution. Accordingly, the inverse cdf transformation (if it exists) can be used for transforming the uniform random variable to its actual domain.

By definition, for a random variable X with an *invertible* cdf $F_X(x) = P(X \le x)$, the random variable $F_X(X)$ follows a uniform distribution on the interval [0, 1].[3] Thus, if F_X is invertible the following relation holds:

$$X = F_X^{-1}(U) \Leftrightarrow U = F_X(X) \tag{1}$$

where U is a random variable that is uniformly distributed on [0, 1]. In Fig. 3, the relation (1) is presented schematically. This relationship forms the basis for the sampling of a random variable in Monte Carlo simulation. In particular, the general

[3]The proof of this statement is as follows:

$$For \ r \in [0,1] : P\left(F_X(X) \le r\right) = P\left(X \le F_X^{-1}(r)\right)$$
$$= F_X\left[F_X^{-1}(r)\right] = r.$$

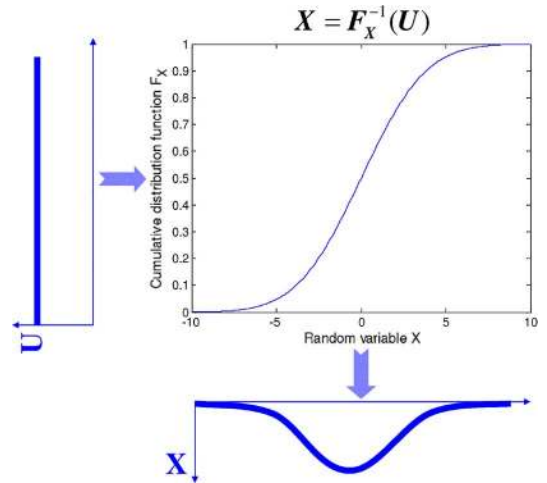
Fig. 3. Sampling of an arbitrary random variable.
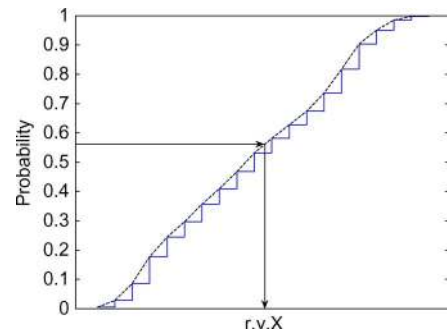
Fig. 4. Sampling of a random variable based on measured data.

Fig. 5. Increasing transformations and respective ranks.

random number generation algorithms provide samples that are uniformly distributed. By applying (1), the generated uniform samples are transformed to the desired distribution.

In cases when F_X is not available in closed form, the empirical cdf F_X^e can be used. When sampling with the empirical cdf we may first perform an interpolation to get a smooth distribution (the case of linear interpolation is presented in Fig. 4) or we can simply sample the empirical distribution by choosing one of the border values.

Transforming data to ranks is simple. One should just sort the data and replace their values by their positions in the sorted file. Fig. 5 presents the case of the application of two different increasing transformations, X^2 and $2 \cdot X$ to a sample population. One can see that although the actual numbers change, the ranks remain unchanged.

The significance of the cdf transformation lies in the property that its application does not affect the dependence structure. In particular, the ranking (and consequently the stochastic

dependence) of a sample population remains the same under increasing transformations. This is investigated in the next section.

B. Rank Correlation

To measure the strength of dependence between random variables, a measure of dependence is required. The known dependence measure, the *product moment (linear) correlation* ρ can be applied in this setting. It measures the degree of linear relationship between the random variables and is invariant to linear transformations [12]. However, since it is applied in the actual domain of the random variables, it is affected by the marginal distributions and is not invariant to increasing transformations as the ones presented in the previous section. To avoid these shortcomings, one should apply this measure to the respective ranks of the random variables.

The product moment correlation of ranks is called the *rank correlation* ρ_r. The rank correlation of the random variables X and Y with cdf F_X and F_Y is defined as

$$\rho_r(X, Y) = \rho\left(F_X(X), F_Y(Y)\right). \qquad (2)$$

Since rank correlation is the product moment correlation of uniform variables (see (1)), it always exists and it takes values in the interval $[-1, 1]$. It measures the *monotonic* relationship between the random variables. Its equality to 1 corresponds to perfect positive dependence, hence the monotonic relationship between variables [Fig. 1(b) and Fig. 2(c)]. By using the measure in the uniform/rank domain, the effect of the marginal distributions on the value of the measure is removed. One can see this by applying the two measures for the sample populations presented in Fig. 5. In particular, the product moment correlation yields $\rho(X, Y) = 0.31$, while $\rho(X^2, 2Y) = 0.43$. In both cases, the rank correlation is the same and equal to 0.30.

For uniform variables, linear and rank correlations are the same ($\rho = \rho_r$), but in general they are different. For the *joint Normal distribution*, the relationship between ρ and ρ_r is known: let (X,Y) be random vectors with joint Normal distribution, then

$$\rho(X, Y) = 2\sin\left(\frac{\pi}{6}\rho_r(X, Y)\right). \qquad (3)$$

For multivariate problems, all mutual rank correlations between the random variables are gathered in the form of a rank correlation matrix $(\mathbf{R_r})$.

C. Copulas

1) General: By applying the cdf transformation to all random variables of the problem the transition to a common domain (rank/uniform domain) is achieved. The dependence modeling takes place in this common domain and the random variables can be further transformed back to their original distributions by the inverse-cdf transformation without any loss of information. Fig. 6 provides a representation of this procedure. This transformation yields a decomposition of the stochastic modeling into two basic components:
- the one-dimensional marginal distributions;
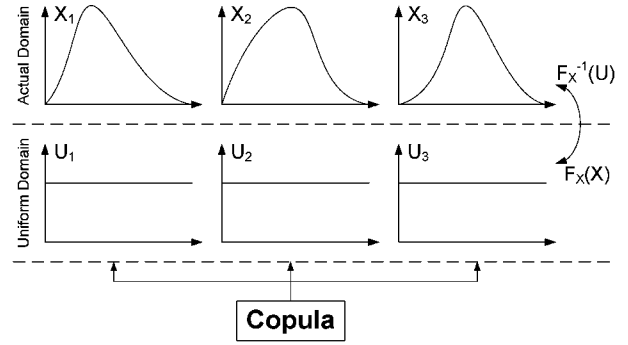- the multidimensional stochastic dependence structure between the ranks of input random variables.

Fig. 6. Copula modeling mechanism.

Copulas are functions that join or 'couple' one-dimensional marginal distributions in multivariate distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional marginals are uniform on the interval [0, 1] [13]. Sklar's theorem [14] offers the bridge between copula and joint distribution of a number of random variables: The random variables X and Y with cdf F_X and F_Y, are joined by copula C if their joint distribution can be written

$$F_{XY}(x, y) = C\left(F_X(x), F_Y(y)\right). \qquad (4)$$

If F_X and F_Y are continuous, then C is unique.

From (1) it can be written $F_X(x) = u$ and $F_Y(y) = v$, where u and v are realizations of the uniform random variables U and V, respectively. In this case, the (4) will become

$$C_{UV}(u, v) = F\left(F_X^{-1}(u), F_Y^{-1}(v)\right). \qquad (5)$$

This equation can be used for the construction of the copula function from the respective multivariate distribution function.

2) Families of Copulas: Many families of copulas are known in the literature. According to the relation (4) every joint distribution yields corresponding copula, e.g., the Normal (Gaussian) copula, the Elliptical copulas. There are families of copulas that are not related to any specific family of multivariate distributions, e.g., the Archimedian copulas, the diagonal band copula. For details on the properties of the different families of copulas, one may refer to [13], [15], and [16].

In Fig. 7, the scatter diagrams for four different copulas with the same rank correlation are presented. Although the same rank correlation is kept, the scatter diagrams are different. For the Normal, Frank's and the minimum information copulas differences are not pronounced. However the scatter plot for diagonal band copula differs significantly from the others.

Copulas allow realization of full range of values for the rank correlation. In Fig. 8, the scatter diagrams for the diagonal band copula with different values of ρ_r ranging from -1 to 0.8 are presented.

To conclude and summarize, we may notice that it is more convenient to investigate and model stochastic dependence not in the actual domain but in the common uniform/rank domain. The following steps should therefore be followed when modeling dependent random variables:

1) *Rank correlations*: the rank correlations between the random variables should be calculated form data.
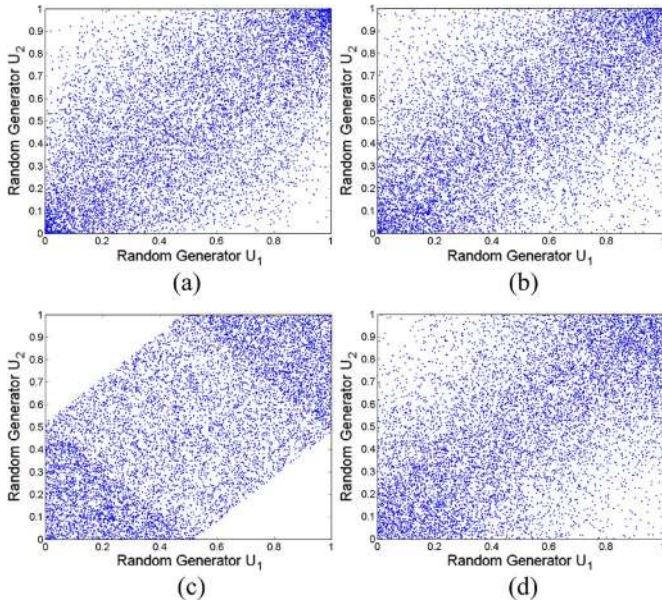
Fig. 7.   Scatter diagrams for four types of copulas under the same rank correlation. (a) Normal copula $\rho_r = 0.6$. (b) Frank's copula $\rho_r = 0.6$. (c) Diagonal band copula $\rho_r = 0.6$. (d) Minimum information copula $\rho_r = 0.6$.
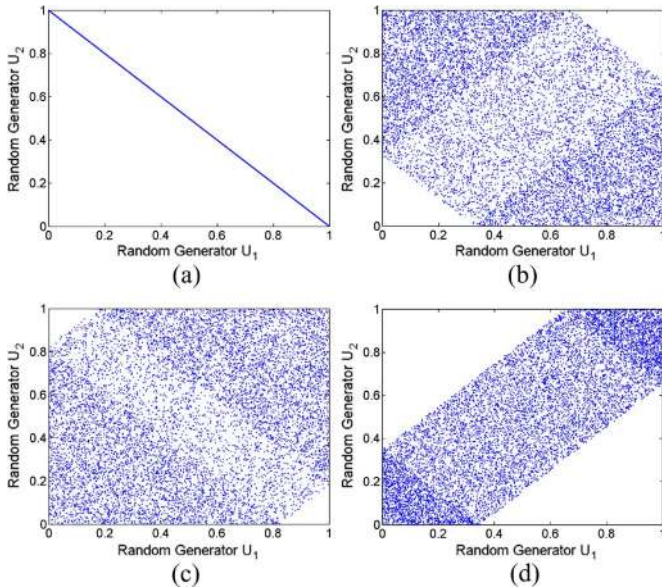

Fig. 8.   Scatter diagrams for the diagonal band copula for different rank correlations. (a) $\rho_r = -1$. (b) $\rho_r = -0.4$. (c) $\rho_r = 0.2$. (d) $\rho_r = 0.8$.

2) *Copula*: a copula with given rank correlation should be used as a dependence function between ranks of the random variables.

3) *Inverse-cdf*: the correlated ranks should be transformed into the given marginal distribution.

In the following, the modeling algorithm for the case of the most commonly used copula, the *Normal/Gaussian copula* is presented in detail.

D. Normal/Gaussian Copula: Joint Normal Transform Methodology

The Normal/Gaussian copula is the copula that corresponds to the multivariate Normal distribution. The copula can be con-
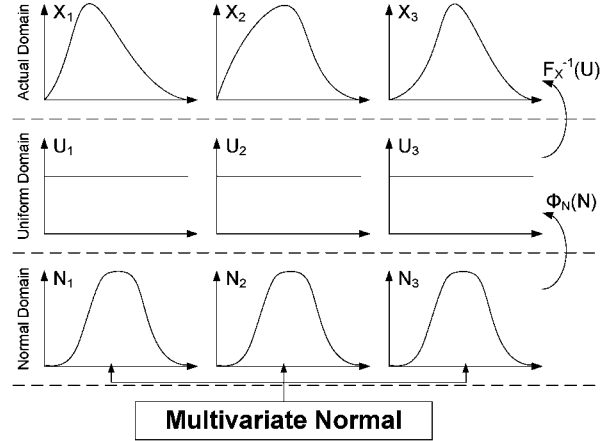

Fig. 9.   Joint Normal transform methodology.

structed from the multivariate Normal distribution using (5). The density function is further obtained from the density function of the Normal distribution by differentiating. The formula for the bivariate case is

$$c_N(u_1, u_2) = \frac{f\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)}{\Phi^{-1}(u_1) \cdot \Phi^{-1}(u_2)} \tag{6}$$

where f is the bivariate density of the Normal distribution with correlation ρ and Φ^{-1} is the inverse of the standard Normal distribution. Notice that ρ is the parameter of the Normal copula in the bivariate case. In the multivariate case the linear correlation matrix $\mathbf{R}$ of multivariate Normal forms the set of parameters of the Normal copula. To sample a distribution with given margins and the Normal copula we can use standard theory of the multivariate Normal distribution. A set of Normals correlated according to linear correlation matrix $\mathbf{R}$ can be generated if $\mathbf{R}$ is positive semi-definite. The margins are transformed into uniforms by applying the Normal-cdf transformation Φ, according to the theory presented in Section III-A. Further, based on the standard copula theory, the inverse-cdf transformation should be applied for obtaining the actual marginal distributions of the random variables. This technique is called the *joint Normal transform (JNT)* methodology [16], [17]. In Fig. 9, a schematic representation of this procedure is presented, in accordance to Fig. 6.

With the above methodology, the obtained random variables will be correlated according to the product moment correlation matrix $\mathbf{R}$. In order to obtain the desired rank correlation matrix $\mathbf{R_r}$, one should start from an appropriate matrix $\mathbf{R}$. This matrix can be calculated from the relationship between linear and rank correlation for the joint Normal distribution presented in (3). Note however that this calculation can lead to obtaining a nonpositive definite matrix [1], [16].

Although by definition a correlation matrix is positive semi-definite, problems with positive semi-definiteness can be sometimes encountered in real applications. It often happens that correlations are estimated by noisy procedures. It may thus arise that a "*measured*" correlation matrix is nonpositive semi-definite. This hinders the applicability of the method. In this case, the method may be applied after repairing the violations of positive semi-definiteness or other methods

should be employed that do not require the full definition of the correlation matrix [16]. Repairing the violations of positive semi-definiteness comes down to the transformation of the initial nonpositive semi-definite matrix to a positive semi-definite matrix that is as close as possible to the original one. In particular, a method has to be used that:

1) is guaranteed to produce a positive semi-definite matrix;
2) does not require a preexisting acceptable (positive semi-definite) matrix to begin with;
3) is fast to implement even for large matrices;
4) allows the determination of a feasible matrix that most closely approximates a target real symmetric (but non-positive-definite) matrix in a well-defined and quantifiable sense.

In [18], two main methods for repairing violations of positive semi-definiteness are presented; the *hypersphere decomposition*, which satisfies all above properties and a second, faster method, the *spectral decomposition*, which shares the first three properties but is not guaranteed to satisfy the fourth property. However, as mentioned in [18], empirical studies show that the results obtained using this second approach are extremely close, albeit not identical, to the ones obtained using the first technique.

Another disadvantage of the Normal copula is that it fails to represent asymptotic tail dependence, i.e., dependence between the extreme values of the random variables. In particular, as shown in [19], the Normal copula gives asymptotic independence, provided that $\rho < 1$. Regardless of how high a correlation we choose, if we go far enough into the tail, extreme events appear to occur independently of each other. In order to represent tail dependence, e.g., elliptical copulas can be used.

In the following we demonstrate the applicability of the JNT methodology in a comprehensive example: the impact of large-scale integration of wind power in the power system of the Netherlands. In this example we present how measurement data from different data sources can be combined.

IV. CASE-STUDY: WIND POWER INTEGRATION IN THE NETHERLANDS

The study case involves the evaluation of the impact of the integration of a total capacity of 1.5 GW onshore and 3.5 GW offshore wind power in the power system of the Netherlands in 2020.

For the analysis, three different data sets are used: measurements for onshore locations, measurements for offshore locations and load data. The availability and adequacy of the respective data may be categorized as follows:

1) *Well-documented data*: in the specific study case, we consider ten locations for the onshore wind parks (sites 1–9 and 13 in Fig. 10); for each location, hourly wind speed values are available for a period of 20 years from January 1, 1984 until December 31, 2003 [20]. This data set forms a well documented entity, with a small number of missing values. The planned capacities for each onshore wind park are 100 MW for the locations 1–5 and 7–8, 200 MW for location 9, and 300 MW for locations 6 and 13. The total planned onshore capacity is 1500 MW.
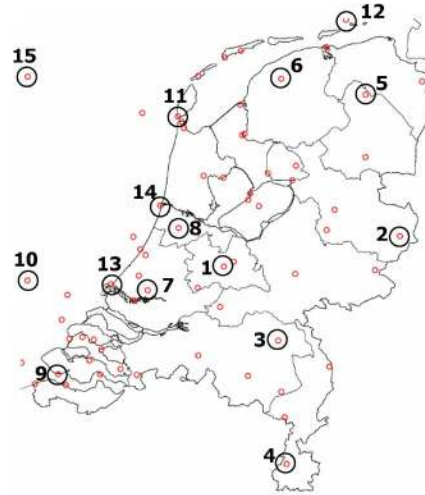

Fig. 10. Wind park locations in the Netherlands.

2) *Not well-documented data*: measurements for five offshore sites are available, namely for the sites 11, 12, 14 (near-shore wind parks) and 10 and 15 (offshore wind parks) in Fig. 10. The available data set corresponds again to hourly measurements for the same period as for the onshore wind parks. This data set however is not so well documented, presenting a relatively large number of missing values. The total planned offshore capacity is 3500 MW. In the following matrix, the planned capacities for each of these locations are presented together with the percentage of missing values for each data set.

Wind Park	10	11	12	14	15
Capacity (MW)	1000	500	500	500	1000
Missing values (%)	8.3	13.8	5.8	6.4	8.8

3) *Different data system*: the load data set corresponds to a different recording period and sampling frequency, in particular 15-min averages for a period of one year, i.e., between January 1 and December 31, 2003.

The data sets involved in the analysis are not uniform; they come from different data systems and span different periods with different sampling frequencies. Such a non-uniformity of the data sets is usual for real applications, due to the use of different databases and measuring systems. In such cases, the analysis should be based on extraction and utilization of maximum information from the data sets. The application of the JNT methodology requires the computation of the following statistics:

1) *Marginal distributions*: the wind speed/power distributions at each generation site and the system load distribution.
2) *Dependence structure*: the rank correlation matrix between the wind speed random variables and the system load.

A. Marginal Distributions

1) Wind Speed/Power Distributions: The wind speed distributions for the sites where data are available are obtained after adapting the measurement height (10 m) to the wind turbine hub height. Typical pitch-controlled wind turbine generators (WTG)
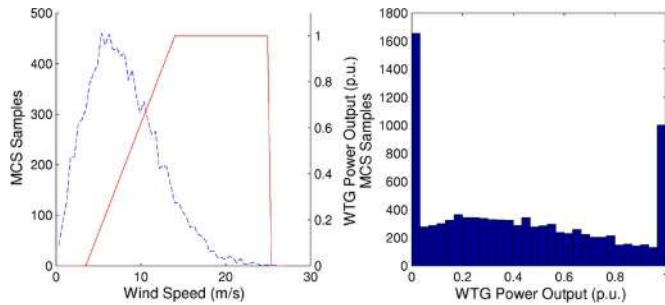
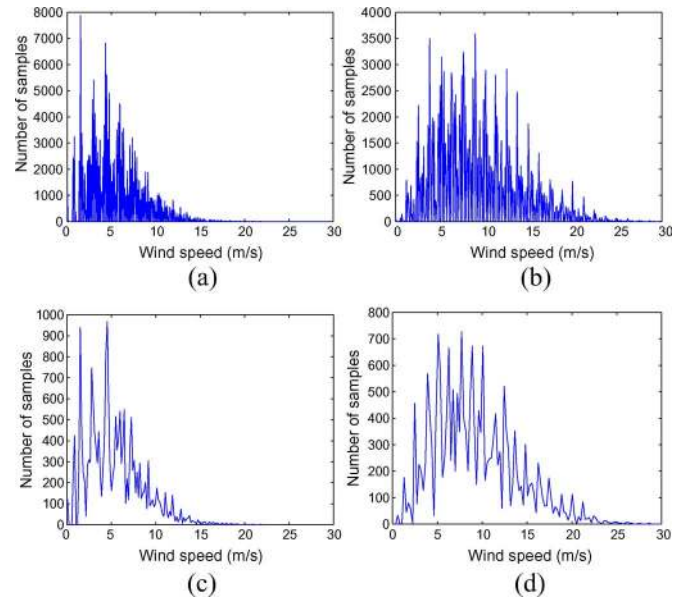Fig. 11.   WTG wind speed/power characteristic and distributions.

Fig. 12.   Measured and simulated wind speed distributions at site 1 (onshore) and 14 (offshore). (a) Measurements site 1. (b) Measurements site 14. (c) Simulation site 1. (d) Simulation site 14.
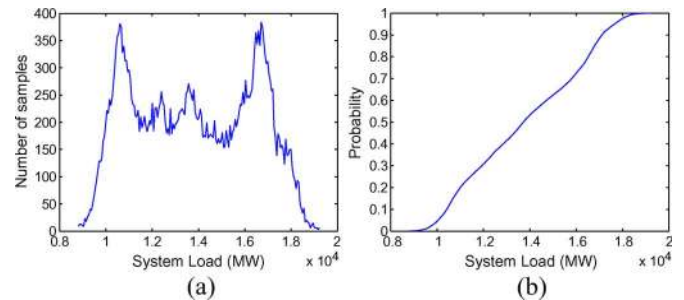

Fig. 13.   System load distributions for the year 2020. (a) pdf. (b) cdf.

are considered for this project, with a hub height of 80 m, a nominal power of 2 MW and cut-in, nominal and cut-out wind speed values of 3.5, 13.5 and 25 [m/s]. In Fig. 11, the wind speed and wind power distributions for a such a wind turbine generator are presented as obtained by a 10 000-sample Monte Carlo simulation. On the left graph, the wind speed distribution for a typical site is presented (discontinuous line), together with the wind speed/power wind turbine generator characteristic (continuous line). In the right graph, the obtained wind power distribution is presented. An accumulation of probability masses at the zero and nominal wind power is observed. The zero values correspond to wind speeds lower than the cut-in and higher than the cut-out wind speed values. The nominal power output values correspond to wind speeds between the nominal and the cut-out values. The aggregate power curve of the wind park is consider similar to the one of a single wind turbine. This simplistic approximation is employed here in order to give emphasis to the impact of stochastic dependence. Alternatively, one may use a multiturbine power curve approach from the related literature (see, e.g., [21]) without rendering the applicability of the methodology.

The sampling of the wind speed distributions has been performed based on the empirical distribution obtained by data, according to the methodology presented in Fig. 4. In Fig. 12, the measured wind speed distributions for two characteristic sites, the onshore site 1 [Fig. 12(a)] and the offshore site 14 [Fig. 12(b)] are compared to the simulated ones obtained from a 20 000-sample Monte Carlo simulation. The simulated distributions yield very accurate approximations of the measured ones.

2) System Load Distribution: In Fig. 13, the pdf and cdf for the extrapolated system load for 2020 are presented. The extrapolation is performed based on a yearly load growth rate of 2%. This results in an expected system load variation for 2020 ranging approximately from 9 GW to 19 GW. As may be seen, the planned wind power capacity corresponds to a significant part of the system demand.

B. Dependence Structure: Correlation Matrices

For the application of the method, the transformed correlation matrix $\mathbf{R}$ should be calculated from the rank correlation matrix $\mathbf{R_r}$, based on (3). Since different datasets are involved in the analysis, the matrix $\mathbf{R_r}$ is calculated in parts, based on the maximum time-intersection of the respective datasets. In particular, the data for the period of 20 years are used for the derivation of the mutual wind speed rank correlations between onshore sites,

a revised version of this dataset without the missing values is used for the estimation of the rank correlations between offshore sites and between onshore and offshore, whereas the rank correlations between the wind speeds and system load are calculated for the common year of measurements, according to the lower measurement frequency, i.e., hourly mean values.

The 16×16 rank correlation matrix is presented in Table I. The first row and column correspond to the system load, while the numbering of the other rows is in accordance to the locations in Fig. 10. A low correlation between the wind activity and the system load is observed, ranging from zero to 0.24. The onshore wind presents a low rank correlation to the system load, while the offshore wind appears to be practically independent. The rank correlations between the wind speeds range between 0.47 and 0.91. Hence, the wind resources are correlated even for distant locations throughout the country.

C. The 5-GW Integration Scenario

According to the simulation procedure presented in Section III-D, first the transformed correlation matrix $\mathbf{R}$ is calculated from the rank correlation matrix $\mathbf{R_r}$ according to (2). Both matrices are positive semi-definite, so the method can be applied directly.

TABLE I
RANK CORRELATION MATRIX $\mathbf{R_r}$ FOR THE CASE STUDY

| 1.00	0.24	0.18	0.19	0.13	0.19	0.18	0.19	0.19	0.09	0.01	0.03	0.00	0.05	0.03	0.00
0.24	1.00	0.82	0.85	0.74	0.78	0.77	0.88	0.89	0.76	0.58	0.57	0.54	0.76	0.68	0.52
0.18	0.82	1.00	0.83	0.74	0.82	0.79	0.79	0.80	0.69	0.54	0.56	0.59	0.69	0.64	0.51
0.19	0.85	0.83	1.00	0.81	0.74	0.73	0.82	0.82	0.75	0.56	0.53	0.52	0.72	0.63	0.48
0.13	0.74	0.74	0.81	1.00	0.65	0.63	0.76	0.73	0.75	0.57	0.48	0.47	0.69	0.60	0.47
0.19	0.78	0.82	0.74	0.65	1.00	0.89	0.77	0.79	0.63	0.52	0.58	0.70	0.66	0.65	0.53
0.18	0.77	0.79	0.73	0.63	0.89	1.00	0.77	0.81	0.64	0.54	0.63	0.71	0.69	0.66	0.57
0.19	0.88	0.79	0.82	0.76	0.77	0.77	1.00	0.91	0.83	0.67	0.59	0.55	0.82	0.73	0.58
0.19	0.89	0.80	0.82	0.73	0.79	0.81	0.91	1.00	0.79	0.64	0.62	0.59	0.81	0.75	0.58
0.09	0.76	0.69	0.75	0.75	0.63	0.64	0.83	0.79	1.00	0.69	0.56	0.51	0.81	0.66	0.54
0.01	0.58	0.54	0.56	0.57	0.52	0.54	0.67	0.64	0.69	1.00	0.50	0.48	0.68	0.62	0.70
0.03	0.57	0.56	0.53	0.48	0.58	0.63	0.59	0.62	0.56	0.50	1.00	0.56	0.64	0.58	0.48
0.00	0.54	0.59	0.52	0.47	0.70	0.71	0.55	0.59	0.51	0.48	0.56	1.00	0.60	0.61	0.50
0.05	0.76	0.69	0.72	0.69	0.66	0.69	0.82	0.81	0.81	0.68	0.64	0.60	1.00	0.72	0.54
0.03	0.68	0.64	0.63	0.60	0.65	0.66	0.73	0.75	0.66	0.62	0.58	0.61	0.72	1.00	0.62
0.00	0.52	0.51	0.48	0.47	0.53	0.57	0.58	0.58	0.54	0.70	0.48	0.50	0.54	0.62	1.00
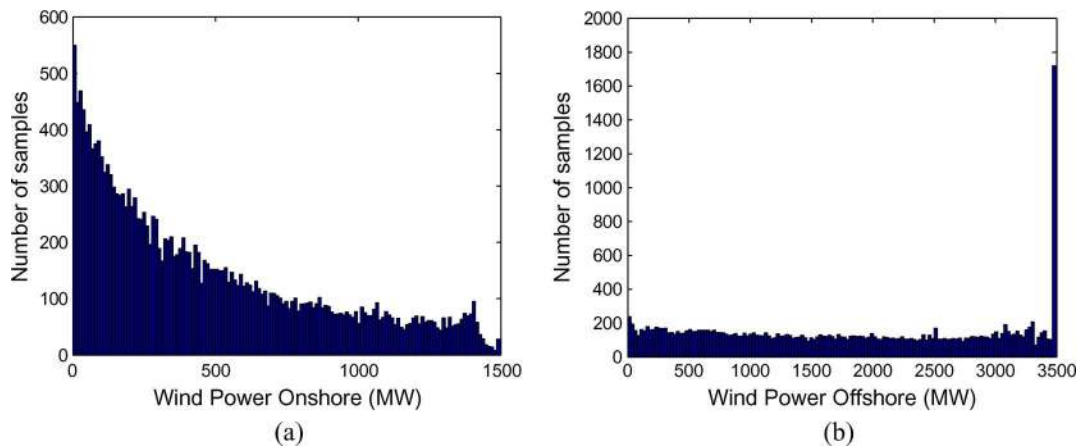
Fig. 14. Onshore and offshore wind power distributions for the 5-GW integration scenario. (a) Onshore. (b) Offshore.

In Fig. 14, the aggregate onshore and offshore wind power are presented. As may be seen, the different resources and dependence structures for onshore and offshore wind power lead to completely different distributions for the aggregate wind power. The offshore wind power distribution presents a uniform behavior for power values from zero to nominal and the accumulation of probability mass in the nominal value (which is the sum of the nominal outputs from all wind parks, i.e., 3500 MW), corresponding to the cases when all offshore wind parks produce simultaneously in their nominal value. This "steep" accumulation of probability mass is due to the fact that the wind parks are modeled as single wind turbines, as discussed in Section IV-A1. In this case, for wind speed values between nominal and cut-out, the wind park produces nominal power output. Therefore, when the wind speeds in all wind parks are in this area (which is often the case due to the high correlation values), the aggregate power output will correspond to the sum of the nominal power outputs of the wind parks. The choice of a multiturbine approach for modeling the wind park power curve would result to the dispersion of samples close to the nominal wind power area for the wind park instead of accumulating the probability mass in one point. As discussed in Section IV-A1, the choice of this simple wind park model does not render the applicability of the methodology, but instead allows the comprehension of the

TABLE II
CAPACITY FACTOR FOR THE WIND PARKS IN THE NETHERLANDS

Windpark	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
η_W	17.8	18.1	18.3	24.7	25.9	29.1	29.1	31
Windpark	W_9	W_{10}	W_{11}	W_{12}	W_{13}	W_{14}	W_{15}	
η_W	34.4	51.8	50.2	51.8	47.1	47.7	54.2	

aggregation effects of high dependence, as is the case here. In Table II, the capacity factor η_W for the wind parks is presented. As may be seen, the onshore wind parks present a lower capacity factor than the offshore ones. The maximum capacity factor of 0.54 is obtained in wind park location 15.

In Fig. 15, the aggregate wind power for the system is presented, together with the net load distribution.[4] The onshore and offshore wind power distributions add up to a distribution that has a uniform shape. The system net load distribution [Fig. 15(b)], obtained as the difference between the distribution in Fig. 13(a) and the aggregate wind power distribution [Fig. 15(a)] is changed compared to the situation before the incorporation of 5 GW of wind power [Fig. 13(a)]. The way these distributions are subtracted is dictated by the correlation between wind speed and load and also by the transformation of

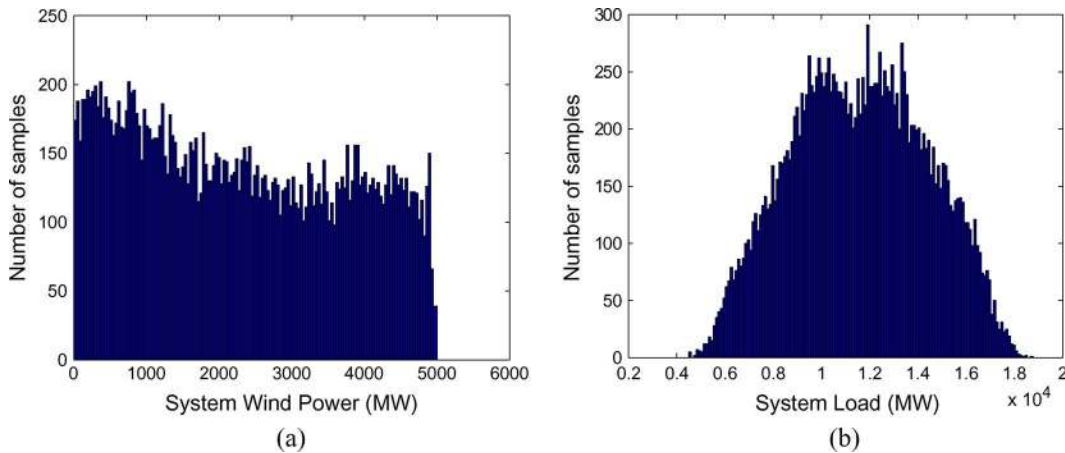[4]System load minus generation is considered here as net load.

Fig. 15.   Total system wind power and system net load distributions for the 5-GW integration scenario. (a) System wind power. (b) System net load.

wind speed to power. This complex operation may sometimes lead to counterintuitive results.

The analysis presented here allows for the calculation of the system net load distribution, taking into account the dependence structure between the wind activity in different locations in the system, as well as with the system load. As discussed in Section II, not taking this dependence structure into account by assuming independence would lead to different results. Focusing in the specific case study, the obtained distributions are of major importance for calculating the capacity credit added to the generation system due to the integration of wind power. The same sampling approach can also be used for the repetitive calculation of the system steady state model, allowing for the estimation of the system power flow distributions. This study enables the system designer to calculate the necessary system reinforcements due to the incorporation of wind power. This subject will be investigated in future work.

V. Conclusions

The modeling of stochastic renewable generation in power system studies brings forward the problem of modeling correlated non-Normal random variables, be it the stochastic infeeds in planning studies or forecast uncertainty in system operation. Stochastic dependence refers to the relative behavior of random variables. For this, its role is central in the calculation of the properties of the sum of a multitude of random variables; even in the case of the same marginal distributions, different dependence structures lead to totally different distribution of their sum, around the same mean value. Thus, not taking into account the stochastic dependence between stochastic infeeds in power system analysis, such as wind parks situated in different sites, will lead to errors in the calculation of the total system generation.

In order to investigate and model stochastic dependence in a multivariate context, the copula theory has been presented. The methodology proposes the application of the cdf transformation to each random variable, leading to the transition to a common uniform/rank domain. The modeling of stochastic dependence takes place in this common domain, using specific distributions with uniform marginals, the copula functions. The modeling algorithm for the Gaussian copula has been presented together

with a related application on the modeling of the impact of the large-scale integration of wind power to the system load in the Netherlands.

This contribution sets the foundations for the modeling of stochastic dependence in power system studies. Future work will focus on a variety of applications and extend the modeling algorithms for the treatment of problems where the rank correlation matrix is nonpositive definite and the investigation of the applicability of different copula families to the power system uncertainty analysis.

References

[1] G. Papaefthymiou, "Integration of stochastic generation in power systems" Ph.D. dissertation, Electrical Power Systems Group (EPS), Delft Univ. Technol., Delft, The Netherlands, 2007. [Online]. Available: http://www.library.tudelft.nl/ws/search/publications/index.htm.

[2] G. Papaefthymiou, P. H. Schavemaker, L. van der Sluis, W. L. Kling, D. Kurowicka, and R. M. Cooke, "Integration of stochastic generation in power systems," *Elect. Power Energy Syst.*, vol. 28, pp. 655–667, 2006.

[3] G. Giebel, G. Kariniotakis, and R. Brownsword, State of the Art on Short-Term Wind Power Prediction, ANEMOS, Deliverable Report D1.1, 2003. [Online]. Available: http://anemos.cma.fr.

[4] P. Pinson, "Estimation of the uncertainty in wind power forecasting" Ph.D. dissertation, Ecole des Mines de Paris, Paris, France, 2006. [Online]. Available: http://pastel.paristech.org/.

[5] P. Pinson, G. Papaefthymiou, B. Klöckl, and H. A. Nielsen, "Generation of statistical scenarios of short-term wind power production," in *Proc. PowerTech 2007 Conf.*, Lausanne, Switzerland, Jul. 1–5, 2007.

[6] G. Papaefthymiou and P. Pinson, "Modeling of spatial dependence in wind power forecast uncertainty," in *Proc. 10th Int. Conf. Probabilistic Methods Applied to Power Systems (PMAPS)*, Rincon, Puerto Rico, May 25–29, 2008.

[7] C. Singh and Y. Kim, "An efficient technique for reliability analysis of power systems including time dependent sources," *IEEE Trans. Power Syst.*, vol. 3, no. 3, pp. 1090–1096, Aug. 1988.

[8] A. M. L. da Silva, V. L. Arienti, and R. N. Allan, "Probabilistic load flow considering dependence between input nodal powers," *IEEE Trans. Power App. Syst.*, vol. PAS-103, no. 6, pp. 1524–1530, Jun. 1984.

[9] R. N. Allan, A. M. L. da Silva, and R. C. Burchett, "Evaluation methods and accuracy in probabilistic load flow solutions," *IEEE Trans. Power App. Syst.*, vol. PAS-100, no. 5, pp. 2539–2546, May 1981.

[10] M. C. Caramanis, R. D. Tabors, K. S. Nochur, and F. C. Schweppe, "The introduction of non-dispatchable technologies as decision variables in long-term generation expansion models," *IEEE Trans. Power App. Syst.*, vol. PAS-101, no. 8, pp. 2658–2666, Aug. 1982.

[11] J. A. Bloom, "Probabilistic production costing with dependent generating sources," *IEEE Trans. Power App. Syst.*, vol. PAS-104, no. 8, pp. 2064–2070, Aug. 1985.

[12] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, ser. Electrical and Computer Engineering, 4th ed. New York: McGraw-Hill, 2002.

[13] R. B. Nelsen, *An Introduction to Copulas*. New York: Springer, 1999.

[14] A. Sklar, "Fonctions de répartition á n dimensions et leurs marges," *Publications de l'Institut de Statistique de L'Université de Paris*, vol. 8, pp. 229–231, 1959.

[15] H. Joe, *Multivariate Models and Dependence Concepts*. London, U.K.: Chapman & Hall, 1997.

[16] D. Kurowicka and R. M. Cooke, *Uncertainty Analysis With High Dimensional Dependence Modelling*, ser. Probability and Statistics. New York: Wiley, 2006.

[17] S. Ghosh and S. Henderson, "Properties of the notra method in higher dimensions," in *Proc. 2002 Winter Simulation Conf.*, 2002, pp. 263–269.

[18] R. Rebonato and P. Jäckel, "The most general methodology to create a valid correlation matrix for risk management and option pricing purposes," *J. Risk*, vol. 2, no. 2, pp. 17–28, 1999.

[19] P. Embrechts, A. McNeil, and D. Straumann, "Correlation and dependence in risk management: Properties and pitfalls," *Risk Manage.: Value at Risk and Beyond*, pp. 176–223, 2002.

[20] Royal Netherlands Meteorological Institute (KNMI), , HYDRA Project: Wind Climate Assessment of The Netherlands, 2006. [Online]. Available: http://www.knmi.nl/samenw/hydra.

[21] P. Nørgaard and H. Holttinen, "A multi-turbine power curve approach," in *Proc. Nordic Wind Power Conf. (NWPC2004)*, Gothenburg, Sweden, Mar. 1–2, 2004.

George Papaefthymiou (M'03) received the Dipl-Eng. degree in electrical and computer engineering from the University of Patras, Patras, Greece in 1999 and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2007.

He is currently with the Electrical Power Systems Laboratory of the Delft University of Technology, as a Research Associate, and with Ecofys Germany GmbH, as a Consultant. His current research interests include modeling of uncertainty in power systems and design of energy systems with large-scale penetration of distributed and stochastic generation.

Dorota Kurowicka studied mathematics at Gdansk University, Gdansk, Poland, and specialized in numerical methods. She performed her Ph.D. research at the Gdynia Maritime Academy and the Delft University of Technology, Delft, The Netherlands. She received the Ph.D degree from Delft University in 2000.

Since then, she has been employed at Delft Institute of Applied Mathematics, first as an Assistant Professor and from 2006 as an Associate Professor. Her current research is carried out in the area of uncertainty and risk analysis.