Simulation of Spatially Correlated Wind Power in Small Geographic Areas—Sampling Methods and Evaluation ☆

Guzmán Díaz[a,*], Pedro G. Casielles[a], José Coto[a]

[a] *Dep. of Electrical Engineering, University of Oviedo, Campus de Viesques, s/n, 33204 Spain*

Abstract

In clusters of wind generators spread over small geographic areas, the spatial correlation of wind power production is strong. Simulation of joint power production in such cases—such as for instance for determining the available power in a microgrid—is flawed if the correlation is not properly defined.

Several methods have been proposed in the literature for producing scenarios of correlated samples; mostly focused on wind speed. In this paper we analyze three popular choices: classical Monte Carlo (with correlation induced by Cholesky factorization), Latin Hypercube Sampling (with correlation induced by rank sorting), and the recent copula theory. We put together a variety of statistical tools to transform an uncorrelated multivariate sample into a correlated one; and supplement other works by introducing a detailed definition of the wind power distribution and by expanding the Archimedean copula analysis to dimensions beyond the bivariate case analyzed in some related works.

We analyze a year of wind production of 210 wind site from NREL data base. We cluster them to give a view of prospective microgrids, and employ several statistical techniques to measure the adequacy of the simulated samples to the original measured data.

Our results show that, for generation in small geographic areas, the higher the number of generators, the better the wind power dependence structure is described by LHS. On the contrary, copulas—Gumbel or Gaussian for two- and three-dimensional problems, and Gaussian for higher dimensions—are better suited for representing correlated wind speed. The results are different when the generators are spread over large geographic areas.

Compared with LHS endowed with rank sorting for inducing correlation, copula theory is in some sense cumbersome to apply for modeling and simulating wind power data. However, simulations can be performed in prospective microgrids in small geographical areas with larger accuracy by means of LHS if wind power is analyzed rather than wind speed. This advantage is lost for large distances or when small number of generators is considered.

Keywords: Wind power, Microgrids, multivariate dependence, Spatial correlation, Copulas, Monte Carlo, LHS, DD-plot, k-sample tests

1. Introduction

The concept of microgrid has been evolving over the last decade to encompass a new paradigm of power system—that of a small in size cluster of distributed generation and loads, which can be managed to obtain economic and technical advantages. From the early discussions about modeling issues and grid connected versus islanded control [1, 2, 3, 4] the research has focused in the last years on more realistic problems in which the microgrids might play a role in supporting the main grid [5], would harness renewable energy resources, optimize their operational costs [6, 7], or take part in electricity markets [8, 9, 10]; with demonstration projects such as in [10, 9].

Our paper is focused on wind power in microgrids, and more particularly on the simulation of wind power to obtain scenarios of interest for stochastic programming approaches. Wind power generation introduces an uncertainty in the

☆This work was supported by the Spanish Ministry of Science under Grant ENE2010-14899.

*Corresponding author

Email addresses: `guzman@uniovi.es` (Guzmán Díaz), `pcasiell@uniovi.es` (Pedro G. Casielles), `jcoto@uniovi.es` (José Coto)

model that prevents from analyzing the aforementioned problems in a deterministic way. See for instance [7], where Monte Carlo simulation of parametric (Weibull) distributions had to be employed to account for the uncertainty of wind energy in a unit commitment problem. Or the instance in [11], where Monte Carlo samples were also employed to define an interesting index, from balancing voltage security and loss reduction, which allowed determining the optimal reactive droop coefficients.

Though in general stochastic programming is still being initially and progressively introduced into microgrid analyses, see for instance [11, 12, 13, 14], in bulk power systems it has been thoroughly employed in the past; and the need for simulated wind scenario generation has been thus recognized. Moreover, spatial correlation has been recognized as a fundamental concept in determining accurate scenarios [15, 16, 17, 18, 19, 20, 21, 22]. Mostly when as for instance is reported in [23, 24, 25, 26] (and as we later justify) the correlation among close wind generators—those of a microgrid—is not negligibly but really strong. This is the subject matter of our paper.

The first of the techniques that we shall discuss is based on the notorious Monte Carlo simulation method. Some authors have proposed that uncorrelated Monte Carlo samples of wind speed can be transformed into correlated samples by means of a Cholesky factorization of the covariance matrix [15, 19, 20]. It is a relatively computationally inexpensive and simple method, which nevertheless has been already reported to have some problems with the production of unreal negative wind speeds. But more importantly, its major flaw may be in that the method is best suited for representing normal random variables, and thus its accuracy for modeling non-normal wind data may be thus compromised.

The second technique, Latin Hypercube Sampling (LHS), is considered in many related literature as an improvement over crude Monte Carlo. LHS is a technique that reduces the variance with respect to Monte Carlo and, maybe more important, is distribution free. In our case, the correlation must be again induced into the uncorrelated sample. We will follow the technique proposed by Iman and Conover in [27], which has been already employed to induce wind correlation samples in [28, 29].

Lastly, we will address copulas. As we shall show, it is a sophisticated technique that has been employed in recent wind speed related papers [30, 25, 22, 31, 32]. The proposition of copula theory is that the dependence structure among wind powers can be split into independent univariate marginals and a joint distribution function with uniformly distributed marginals, namely the copula. This dissociation is really advantageous, because marginals and copula can be managed independently. This is possible, however, at the cost of greater complexity in some analyses. For instance, Gumbel copula has been rarely employed, and only recently it has been analyzed in some papers, with favorable assessment about their accuracy in representing the dependence structure of wind power [30, 32]. Nonetheless, the works in [30, 32] have only addressed *bivariate* copulas—i.e., the dependence structure between two sites. Herein, we will show a procedure to expand the analysis to larger dimensions, so that a microgrid composed of several wind generators can be analyzed.

Specifically the following are novelties addressed in this paper, which to our knowledge have not been yet covered in the literature:

1. We introduce wind power as the random variable of interest, proposing an explicit formulation of the probability integral transform. Most of the literature deals with wind speed rather than wind power. However, directly employing wind power is of interest for some problems, such as those in which convolution of random variables is required—for instance to aggregate all produced random power in a microgrid or wind farm [33]. But the use of wind power as a random variable is complicated by the existence of "jumps" in the distribution function because of the constant or null power produced by the turbine at some speed ranges. This entails that for instance the direct application of the probability integral transform makes a copula definition non-unique. We then propose a mixed discrete-continuous formulation, derived from Rüschendorf's work in [34], which permits a simple transformation of wind power into uniform variables.

2. Our analysis is cross-sectional—rather than based on time series—and multivariate. This means that we study spatial correlation, which has been the subject matter of some papers cited above. Again, this is necessary for analyzing clusters of wind generators spread over small geographic areas—such as wind farms and microgrids. In this respect recently some papers have introduced copula theory. The novelty in our paper is that we extend the analysis to *multivariate Gumbel* copulas. So far only *multivariate Gaussian* and *bivariate Gumbel* copulas had been employed; though it is remarkable that in [30, 32] it was concluded that Gumbel copulas are best suited in many cases to represent the dependence structure of wind speed than Gaussian copulas. It thus seems

interesting to know whether this is also true in scenarios involving more than two generators. Specifically, we have not found any published work related to energy analysis in which a non-iterative procedure for modeling multivariate Gumbel copula is detailed.

3. After addressing the modeling issues, we make a comparison of three methods for random simulation: LHS, Monte Carlo, and copulas. This is of concern for stochastic programing, where wind power scenarios must be obtained. Nonetheless, we have not either found in the literature any guidance about which of the three methods performs best. Monte Carlo is the most simple and copula the most demanding. So the question that we address is: Is the complex copula theory required to properly simulate wind power/speed scenarios?

4. We have not found in energy related literature ways of assessing the goodness of cross-sectional wind simulated data. We propose the use of two methods. One for bivariate data, based on two-sample tests, and other for multivariate data, based on DD-plots first introduced in [35].

5. Finally following point 3, we provide guidance on which of the methods is best suited for representing dependence structures in a wind scenario generation framework. Our results, based on real data, indicate that the answer depends on the distance among generator sites, the dimension of the problem (number of generators), and the random variable analyzed (wind power or wind speed). We conclude that though more complex, copula theory is not always the best choice—for instance, high dimensional problems analyzing wind power are better represented by the less complex LHS.

The paper is then structured as follows. Section 2 shows the techniques employed to generate simulated samples of spatially correlated wind powers. Section 3 describes the quality assessment techniques and evaluates the goodness of each approach. Last Section concludes.

2. Preliminaries and description of methods

2.1. Mixed formulation of the wind power distribution

A fundamental and common aspect of the methods analyzed herein is their use of uniform random variables for a number of sampling computations. This is a requirement in every method analyzed here. In our analysis of copulas, we will transform the wind power into uniform scale-less variables to fit copula families and simulate (uniformly distributed) samples. Thereafter, we shall return to the wind power true value. In LHS simulations, we shall depart from stratified uniform samples and transform the obtained quantiles into wind power. Similarly, we will obtain uniform samples to produce wind power by transformation, and thereafter introduce correlation.

If we let $F_{P_i}(P_i)$ be the cumulative distribution function (CDF) of wind power, a random variable, the probability integral transform for continuous variables indicates that $F_{P_i}(P_i) \sim U(0,1)$ [36]. That is, the random wind power P_i may be transformed into a uniform random variable U_i. (Indeed, we might transform $P_i \sim F_{P_i}$ into any other distribution, but $U(0,1)$ is particularly convenient because it is parameter free.)

So first we must obtain from the original sample of the wind power a probability model—a distribution function—of the wind power marginal CDFs, to go back and forth from wind power to uniform random variables. This estimation of the CDFs can be done parametrically or non-parametrically. In this paper for brevity, we shall follow the non-parametric approach on the NREL data, mindful that as in [29] we could have employed a given parametric CDF (with its parameters obtained either through sample fitting or from a transformation of a Weibull distribution through the turbine curve; see for instance [37]) with similar results.

Specifically, we have employed a kernel estimation of the CDF. Kernel estimation is a well-known variety of methods for softening the discrete observations of a sample to produce a CDF estimation. In our work we employed the method by Botev *et al.* in [38].

Some illustrative results are presented in Fig. 1. Obviously, they are site dependent. Thus, because the probability of calm winds at site 2 is larger than at site 1, then the power production will more probably be zero at site 2 than at site 1. This "pushes" the CDF upwards, as seen in Fig. 1. Obviously, the CDF is also generator dependent. This can be appreciated in Fig. 1 by comparing the CDF of an Enercon E40 with a GWP47, of similar output power. (Note that though of different output power rating, both have been rescaled 1 p.u. to include them in the same graphic.)

But the most remarkable features of Fig. 1 are the two discrete "jumps" at both extremes. They are a consequence of the existence of cut-in and cut-off wind speeds (below 3–5 m/s and above approximately 25 m/s the power output is null) and a rated wind speed (above 12–15 m/s the power production is locked to 1 p.u.). These discontinuities make the transformation from P_i to U_i especially delicate, because the probability integral transform can not be directly applied in its classical form. (Just note that for $P_i = 0$ the value of U_i is not unique.) Moreover, as a consequence of the mixed discrete/continuous CDF, the ensuing copula would not be unique [39].

Nonetheless, an extension of the probability integral transform can produce a unique copula representation of a multivariate distribution despite of the nature of the marginal CDFs, as demonstrated in [40] by using grade transformations, and in [34] by means of a specific distributional transform. This latter technique, proposed by Rüschendorf, states that if $X \sim F_X$ and we let $V \sim \mathrm{U}(0,1)$ be independent of X, then the distributional transform is

$$U = F_X(X-) + V[F_X(X) - F_X(X-)], \tag{1}$$

where $U \sim \mathrm{U}(0,1)$ and $X = F_X^{-1}(X)$.

Following Rüschendorf's approach, we have restated the wind power transformation problem as follows. F_{P_i} has a left limit everywhere because it is non-decreasing. More rigorously, $F_{P_i}(p_i-) := \lim_{\epsilon\to 0} F_{P_i}(p_i - \epsilon)$. It follows then that at a discontinuity in F_{P_i},

$$F_{P_i}(p_i) = F_{P_i}(p_i-) + \mathrm{pr}\{P_i = p_i\}\,. \tag{2}$$

In our problem we have two discontinuities, in $P_i = 0$ and in $P_i = 1$. So as stated above, we must compute the probabilities in these two points, which we can infer from the wind speed distribution—and not from the wind power distribution. More precisely:

$$\mathrm{pr}\{P_i = 0\} = \mathrm{pr}\{W_i \le w_{\mathrm{i}i}\} + \mathrm{pr}\{W_i \ge w_{\mathrm{f}i}\} = 1 + F_{W_i}(w_{\mathrm{i}i}) - F_{W_i}(w_{\mathrm{f}i}) \tag{3}$$

$$\mathrm{pr}\{P_i = 1\} = \mathrm{pr}\{W_i \le w_{\mathrm{f}i}\} - \mathrm{pr}\{W_i \ge w_{\mathrm{r}i}\} = F_{W_i}(w_{\mathrm{f}i}) - F_{W_i}(w_{\mathrm{r}_i}) \tag{4}$$

Now, we can account for the two discontinuities in the wind power distribution by stating an alternative, practical expression according to (1):

$$F^*_{P_i} = \begin{cases} \mathrm{U}(0, \mathrm{pr}\{P_i = 0\}), & \text{if } P_i = 0 \\ \mathrm{U}(1 - \mathrm{pr}\{P_i = 1\}, 1), & \text{if } P_i = 1 \\ F_{P_i}(P_i), & \text{otherwise} \end{cases} \tag{5}$$

The first line in (5) ensures that for null wind power the result, the corresponding uniform variable, *is not always* equal to the probability of power being zero; instead, it is any random uniform number between 0 and the probability of power being zero. The second line refers to the uniform distribution of the wind power when the wind power is maximum, i.e. 1.0 p.u. It means that the value of the uniform random variable can be any value—following a uniform distribution—from 1 down to a value obtained by subtracting the probability of maximum wind power. A qq-plot demonstrates the goodness of this approach. If erroneously $F_{P_i}(P_i)$ is used for every value of wind power, including unity and zero power, the qq-plot demonstrates large deviations in the first and fourth quartile, where the error of power estimates accumulate. (Also, given the particular representation of uniform random data, this issue is easy to corroborate in a histogram, where the erroneous transformation of $F_{P_i}(P)$ into a uniform random variable by neglecting the first two lines of (5) gives a histogram that is not uniform in the extreme quartiles.)

2.2. Monte Carlo and Cholesky factorization

First, we will explore the technique presented in some papers, based on a Monte Carlo sample extraction and its transformation by means of a Cholesky factor.

The sample extraction is quite simple. For every i-th site ($i = 1, \ldots, d$), it suffices to draw a uniform sample $U_i \sim \mathrm{U}(0,1)$. This sample is then transformed to wind power by employing the kernel CDF estimation of F_{P_i}, represented in Fig. Fig. 1; that is, $P_i = F_{P_i}^{-1}(U_i)$.

The correlation between the wind powers—which is embedded into the covariance or correlation matrix, $\mathbf{R}$—is then enforced by employing a Cholesky factorization. This technique factors the covariance matrix into a left triangular matrix times its transpose, $\mathbf{R} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$; which is possible because $\mathbf{R}$ is positive-definite. Thus the transformed random variable $\mathbf{P} = (P_1, \ldots, P_d)$ is put into correlated form by stating the linear combination $\mathbf{Y} = \mathbf{PL}^{\mathrm{T}} + \boldsymbol{\mu}$, with $\boldsymbol{\mu}$ the sample mean.

This technique is particularly interesting when $\mathbf{P}$ is a normal variable, because the linear transformation of independent normal variables still is a normal variable [41, p.81]. For other marginals, however, the transformation is not that straightforward [42, Sec. 7.4], and the ensuing convolution of their individual distributions entails a much more complex approach. For instance, a solution can be found in [43], where it is proposed to estimate a new correlation matrix, $\mathbf{R}'$, for a multivariate normal probability density function with zero mean and unit standard deviations, which would relate the nonnormal marginals to the unknown multivariate density. Thereafter, the procedure may follow as in the normal case, by employing the Cholesky decomposition. But unfortunately, the required estimation of $\boldsymbol{\Sigma}'$ is really involved, because it entails solving a dth-order integral, which must be performed by numerical methods. Obviously this limits the method to low dimensional problems.

Herein we shall employ the raw method outlined above, without referring to the particular, complex approach proposed in [43]. Therefore, it is foreseeable that it will not be as accurate as desired; but as long as it is really a simple method, we want to know up to which extent this lack of accuracy worsens the results when compared with other methods.

2.3. Latin Hypercube Sampling

LHS is a stratified sampling method. This means that from $U_i \sim \mathrm{U}(0,1)$ N observations are obtained from N equal segments (and then mapped to wind power through the CDF, as in Monte Carlo) [44, §6.2]. LHS thus provides better spread of the sample points over the sampling space than uniformly distributed sampling by conventional Monte Carlo methods. Crude Monte Carlo method, lacking of stratification, permits that by chance the sample points might be clustered together in different points of the unit hypercube after two different runs. By contrast, LHS sampling is stratified, with equal probability of every underlying uniform realization of the sample, what ensures that the sample points are well spread throughout the unit hypercube, entailing lower variation than Monte Carlo between different runs [28, 29]. Therefore, it is expected that LHS in the long run will perform better than some Monte Carlo simulation, in which the simulation might be biased in the case of Monte Carlo sampling.

After an uncorrelated d-dimensional sample has been obtained—including a *random pairing* of the samples $U_1, \ldots, U_d$ in each generation site—the statistical dependencies can be incorporated by following the rank sorting approaches described by Iman and Conover in [27] or by Stein in [45]. Stein's method outperforms Iman's and Conover's when the input variables are related in a non-monotonic way. This is not our case, and thus we can use Iman's and Conover's method that, ulnlike Stein's, does not require that producing i.i.d. samples with definite joint distribution. As in the previous Monte Carlo, the method requires transformation by means of Cholesky factorization of the covariance matrix, followed by a rank ordering. Though the method is of easy implementation, requiring only sorting and matrix manipulations, it is of rather lengthy explanation. So we refer the reader to [29, §II.b], where it is thoroughly detailed.

2.4. Copulas

A d-dimensional copula $C : [0,1]^d :\rightarrow [0,1]$ is a probability distribution with uniform marginal probability distributions on $[0,1]$ that is defined as

$$C(u_1, \ldots, u_d) = \mathrm{pr}\{U_1 \leq u_1, \ldots, U_d \leq u_d\}, \tag{6}$$

where u_i represents an observation of a standard uniform random variable U_i.

For given continuous marginal CDFs $F_{P_1}, \ldots, F_{P_d}$, Sklar's theorem states that there is a unique copula such that

$$C(u_1, \ldots, u_d) = F(F_{P_1}^{-1}(u_1), \ldots, F_{P_d}^{-1}(u_d)), \tag{7}$$

where $F(p_1, \ldots, p_d)$ is the the joint distribution function of the random wind power at each considered site. And that inversely,

$$F(p_1, \ldots, p_d) = C(F_{P_1}, \ldots, F_{P_d}). \tag{8}$$

Unlike the joint distribution F, which embodies the information of correlation and marginal behavior, it is observed in the previous statements that copulas allow *separately* modeling the marginal distributions and the dependence structure of multivariate random variables. And importantly, the functional dependence between the underlying random variables is not affected by the marginal form. In this respect, copulas allow for more sophisticated definitions of joint distributions than those conventionally provided by Gaussian and t-Student multivariate distributions. For instance, they can account for asymmetrical correlations, giving more weight when required to the correlation in the tails of some distributions. Moreover, the marginals can be of any form, and indeed it is not required that they are of the same distribution family.

However attractive their prospects appear from these basic observations, obtaining a simulated sample by means of copula requires a relatively lengthy process to separate marginals and dependence.

2.4.1. Copula family

The process starts by first selecting the appropriate copula. We shall focus on one-parameter copulas, where the degree of association between the marginals is defined by means of one only parameter, θ. More precisely, we will restrict our analysis to Gaussian and Gumbel copulas (elliptical- and Archimedean-type, respectively [46]), because in a recent analysis of 200 wind generation sites by Louie [32], it was found that these two copulas produced the best model of the dependence structure.

2.4.2. Parameter estimation

Also known as calibration, the second step consists in determining the value of the parameter θ. For continuous marginals, θ can be related to dependence measures such as Spearman or Kendall rank coefficients. However, for mixed or discrete marginals—as is the case with wind power cumulative distribution—Marshall in [47] argued that the computation of θ from rank correlations is not straightforward.

The alternative to obtaining θ from a dependence measure is the log-likelihood minimization. In our problem, the joint density of wind power in d sites can be related to each marginal density, f_{P_i}, and to the copula density, c, as follows:

$$f(p_1,\ldots,p_d) = \prod_{i=1}^{d} f_{P_i}(p_i|\theta_i)\cdot c(F_{P_1}(p_1|\theta_1),\ldots,F_{P_d}(p_d|\theta_d)|\theta_C). \tag{9}$$

The copula and marginal calibration is then obtained by maximizing the log-likelihood for an N-sized sample:

$$\max_{\theta_1,\ldots,\theta_p,\theta_C} \sum_{k=1}^{N}\sum_{i=1}^{d} \log f_{P_i}(p_i^{(k)}\theta_i) + \sum_{k=1}^{N} \log c(F_{P_1}(p_1^{(k)}|\theta_1),\ldots,F_{P_d}(p_d^{(k)}|\theta_d)|\theta_C) \tag{10}$$

This technique, though more complex, is preferred to other copula inference methods when the margins are known, according to [48] where it was verified that the best performance respect to precision was shown by the maximum-likelihood estimator. However, the parameter estimation is relatively hard to do in high dimensional spaces, when d is large, because not only a multidimensional copula must be characterized but also every marginal, introducing a large number of parameters to be estimated (namely, $\theta_1,\ldots,\theta_d$, and θ_C). In those cases it may be preferable a pseudo-maximum likelihood estimate, as presented in [49, p.188]. It consists in splitting the problem (10) into two subproblems by (i) first producing a parametric or non-parametric estimation of each CDF, namely $\widehat{F}_{P_i} = F_{P_i}(p_i;\hat{\theta}_i)$, and (ii) thereafter proceed to solve $\max_\theta \sum_{k=1}^{N} \log c\left(\widehat{F}_{P_1}(p_1^{(k)}),\ldots,\widehat{F}_{P_d}(p_d^{(k)})|\theta_C\right)$, where only one parameter needs to be estimated from an N-sized measured sample. For the first subproblem, Ruppert in [49] proposed employing empirical distributions to produce $\widehat{F}_{P_i}$; we instead employed the kernel estimation described above. For the second subproblem, some important remarks must be done about the definition of the copula density, $c\left(\widehat{F}_{P_1}(p_1^{(k)}),\ldots,\widehat{F}_{P_d}(p_d^{(k)})|\theta_C\right)$.

2.4.3. Copula density and high dimensional variates

The copula in (6) is a multivariate joint distribution with uniformly distributed marginal. And similarly to its univariate counterparts, the copula density in (9) is the derivative of the copula distribution, $c(u_1,\ldots,u_d) = \frac{\partial^d C(u_1,\ldots,u_d)}{\partial u_1\ldots\partial u_d}$.

The most known copula distributions come into two major families, namely elliptical and Archimedean. The elliptical copulas are obtained from joint distributions, from which they get their names—Gaussian and t-student—and have relatively simple close expressions of their densities. More in detail, the copula density is defined as

$$c\left(F(p_1),\ldots,F(p_n)\right) = \frac{f(p_1,\ldots,p_d)}{\prod_{i=1}^{d} f_d(p_d)}, \tag{11}$$

and as long as the Gaussian joint density is known in closed form,

$$f(\mathbf{p}) = \frac{e^{-\frac{1}{2}\mathbf{p}'\mathbf{R}^{-1}\mathbf{p}}}{(2\pi)^{\frac{d}{2}}|\mathbf{R}|^{\frac{1}{2}}}, \tag{12}$$

where $\mathbf{p} = (p_1, \ldots, p_d)$ and $\mathbf{R}$ is the correlation matrix, we are spared the trouble of obtaining a high dimensional derivative of the copula distribution (which moreover has not closed form representation) to produce the copula density.

Differently, Archimedean copula distributions have closed form representation, because they are constructed by a single-valued Archimedean generator, $\psi(t)$, as follows:

$$C(u_1, \ldots, u_p) = \psi^{-1}\left(\psi(u_1) + \cdots + \psi(u_p)\right). \tag{13}$$

This is a nice property, which simplifies the computation of any distribution by entering the appropriate ψ.

However, to obtain its density we must compute the d-th order derivative of C in (13); therefore, the d-th order derivative of the inverse of the generating function, i.e. $\frac{\partial^d}{\partial t^d}\psi^{-1}(t)$. From Archimedean copulas, we will specifically address the Gumbel family, with $\psi(t) = (-\ln t)^\theta$. In the bivariate problem, the expression is already $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\psi^{-1}(t) = t^{\frac{1}{\theta}}(\theta + t^{\frac{1}{\theta}} - 1)\exp\left(-t^{\frac{1}{\theta}}\right)(\theta t)^{-2}$. It is readily observed that when d—the number of wind sites—is larger than 2, the computation of the derivatives is arduous.

The high order derivatives of $\psi^{-1}(t)$ have been conventionally computed by numerical or symbolic methods. This unfortunately made it difficult to employ multivariate Gumbel copulas in simulation studies, least in log-likelihood fitting. However, recently Hofert *et al.* have offered a solution by presenting closed-form expressions of the derivatives of any order [50] for five copulas. Particularly in the case of Gumbel copula, the sought derivative can be constructed as a polynomial,

$$\frac{\mathrm{d}^d}{\mathrm{d}t^d}\psi^{-1}(t) = \frac{\psi(t)}{t^d}\left(a_1 t^{\frac{1}{\theta_C}} + \cdots + a_d t^{\frac{d}{\theta_C}}\right), \tag{14}$$

where the polynomial coefficients are defined as

$$a_k = \sum_{j=k}^{d}(-1)^{j-k}\theta_C^{-j}\begin{bmatrix} d \\ j \end{bmatrix}\begin{Bmatrix} j \\ k \end{Bmatrix}, \tag{15}$$

and the Stirling numbers of first and second kind that define the polynomial coefficients may be obtained from a look-up table, or they can be computed without iterative procedures using the following recurrence expressions:

$$\begin{bmatrix} d+1 \\ j \end{bmatrix} = d\begin{bmatrix} d \\ j \end{bmatrix} + \begin{bmatrix} d \\ j-1 \end{bmatrix} \tag{16}$$

$$\begin{Bmatrix} j+1 \\ k \end{Bmatrix} = k\begin{Bmatrix} j \\ k \end{Bmatrix} + \begin{Bmatrix} j \\ k-1 \end{Bmatrix}, \tag{17}$$

with the initial conditions $\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1$, $\begin{bmatrix} n \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ n \end{bmatrix} = 0$, $\begin{Bmatrix} 0 \\ 0 \end{Bmatrix} = 1$, and $\begin{Bmatrix} n \\ 0 \end{Bmatrix} = \begin{Bmatrix} 0 \\ n \end{Bmatrix} = 0$.

Therefore importantly, the copula density can be computed without iterative procedures for any dimension; which makes the expressions in Hofert *et al.*'s paper ideal for inferring the parameter θ_C by means of the maximization of the log-likelihood.

2.4.4. Simulation

The last step once the copula is characterized is simulation. For that purpose, a popular choice is the conditional distribution method explained in detail in [46, §2.9]. It is a method that is based on the iterative calculation of the inverse conditional copulas up to the number of dimensions of the copula, which gives satisfactory results for bivariate. Unfortunately, the procedure turns out to be inefficient for high dimensions, because we need to compute successive derivatives of the copula distribution (an inconvenient which is aggravated in the Gumbel case by the lack of a explicit form of the required quantile curve).

Alternatively, Marshal and Olkin proposed in [51] a procedure based on first obtaining pseudo-random variables from a distribution function G such that its Laplace-Stieltjes transformation equals the inverse of the generator function; i.e., $\mathcal{LS}(G) = \psi^{-1}$. For Gumbel copulas $\mathcal{LS}(G) = e^{-t^{\frac{1}{\theta}}}$, and G is the stable distribution $\mathrm{St}(\theta^{-1}, 1, \cos^{\theta} \frac{\pi}{2\theta}, 0)$. Though somehow more involved in it its formulation, this procedure is however acceptable for large dimensions, because the value of d does not appreciably affect the computational burden. The exact procedure is as follows:

1. We independently produce a sample $V \sim G = \mathcal{LS}^{-1}(\psi)$ and an i.i.d. $X_1, \ldots, X_d \sim \mathrm{U}(0, 1)$.
2. The simulated sample is then $(U_1, \ldots, U_d)$, where $U_i = \psi^{-1}\left(-\frac{\ln X_i}{V}\right)$, $i \in 1, \ldots, d$.

Fig. 2 summarizes the process from the measured sample ()

3. Wind power simulation tests

In the previous Section we defined the procedures to simulate multivariate correlated samples of wind power from Monte Carlo, LHS, and copulas. The question hereafter is, How do we test for the best approximation to the original data?

An approach that we considered first was a goodness-of-fit test. We are given, by any of the above specified simulation methods, a data set of N simulated observations in d sites, $\{P_1, \ldots, P_d\}_1^N$. A goodness-of-fit test—such as for instance Cramer-von Mises test—tests the hypothesis that $F(P_i) = F^*(P_i^*)$, where $F(P_i)$ is the underlying joint distribution of the measured wind power at the d sites, and $F^*(P_i^*$ is the joint distribution of the simulated data. Given enough power, the test would discriminate all the distributions not holding that $F(P_i) = F^*(P_i)^*$. In other words, goodness-of-fit tests test the hypothesis that an empirical distribution is a member of some class of distribution function. (A graphical equivalent is employed in [22] for testing the goodness-of-fit of bivariate Gaussian copulas by using QQ-plots; again needing a parametric estimation of the underlaying distribution.) Hence, this prevents us from using this approach in our case, where the originally measured and the LHS and Monte Carlo simulated samples have not *a priori* known class of joint distribution.

We found a solution to the question of measuring the goodness of the simulation in the field of Information Theory, where often the similarity between two distributions is assessed using several sample-based approaches. Particularly, the two-sample Kolmogorov-Smirnov test that measures the distance between the two samples as

$$D_{\mathrm{KS}} = \sup |F_N(P_i) - F_M(P_i^*)| \tag{18}$$

does not have this flaw. It is bounded in the interval [0, 1], because it measures the distance between the two empirical distributions—the measured and the simulated. Moreover, it has not the cardinality restrictions of KL divergence. However, it also has limitations. The test is of easy application to test the distance between two univariate samples, where each empirical distribution (EDF) is uniquely defined. For higher dimensions, $d \geq 2$, the definition of the joint CDF is non-unique, however, because the direction of ordering of the observations is arbitrary, and this makes it difficult to define a proper methodology accounting for all possible variations. This issue was solved in [52] by Peacock, who proposed an extension to the univariate KS two-sample test for searching for D_{KS} between two bivariate EDFs. But the implementation of his test requires that samples of size N be evaluated in N^2 points of the type (P_i, P_j^*), with $i, j = 1, \ldots, N$. In our investigation, this would be computationally prohibitive. We therefore implemented the more cumbersome but also more efficient variant proposed in [53] by Fasano and Franscechini, that requires only N evaluations.

3.1. *Bivariate analysis*

By testing bivariate samples by means of Fasano's and Franscechini's method, we are implicitly assessing the goodness of the correlation model of a pair of wind powers. The procedure that we have followed is detailed next.

From the NREL data base (http://www.nrel.gov/wind/), we processed the wind speed of 210 sites—those inside the box limited by latitudes 48.575 and 45.275, and longitudes -103.175 and -100.775; which are represented in Fig. 3. We chose this "box" because the wind sites in the region are oddly located, with few geographical regular patterns, and the distances involved go up to the hundreds of kilometers. It also covered a large degree of correlations, as shown in Fig. 4; though in the range of distances in which we are interested, up to 35 km as shown in the inset figure, the correlation is remarkably high—a circumstance that would be expected in any microgrid setting.

For the 210 sites, we selected the data of year 2006, and averaged 3 observations per day, thus obtaining 1094 wind speed observations per year and site. We then processed the data by assuming Enercon E40 turbines at each site, so as to obtain the wind power production.

The sites were paired according to distance categories. We arbitrarily defined six categories, which are shown in Table 1 together with the number of site pairs that lay within the distance range. This amounted to testing 1586 bivariate samples, which appear in the inset of Fig. 4.

Thereafter, we performed a batch of simulations that entailed:

- For a Monte Carlo and Cholesky decomposition, (i) the calculation of the rank correlation coefficient of the NREL processed sample, (ii) obtaining bivariate Monte Carlo samples of uniform variables, (iii) the transformation into wind power, (iv) inducing the correlation by means of Cholesky-based transformation, and (v) the calculation of the KS statistic;

- For correlated LHS, (i) the calculation of the rank correlation coefficient of the NREL processed sample, (ii) the simulation of an Iman-Conover-based correlated wind power sample, and (iii) the calculation of the KS statistic; and

- For copulas, (i) the transformation of wind power into uniform distributions, (ii) the fitting of the NREL processed data to bivariate Gumbel and Gaussian copulas, (iii) the simulation of the copula, (iv) the transformation of the simulated sample into wind power, and (v) the calculation of KS statistic.

The results are summarized in Fig. 5, where it is clear that Monte Carlo simulations with correlations enforced by means of a Cholesky factorization of the covariance matrix performs poorly when compared with the other methods. Moreover, there is an evident positive skewness in the distribution of the KS statistic, indicating that there is a large probability of extreme errors above the represented mean. And this error is emphasized when the geographical separation of the generators is increased.

LHS and copula simulations show remarkably better scores than Monte Carlo. Indeed, the bad scores ensuing from Monte Carlo has the unwanted effect of exaggerating the vertical scale of the plot in Fig. 5, which obscures the analysis. Thus, we have resorted to apply different vertical axis scales. The closer look at the LHS and copula methods shows that LHS exhibits better accuracy when the geographical distances involved are small. This evidence is also confirmed for higher dimensions. Gaussian and Gumbel copulas do not show that trend and consistently give similar scores of the KS statistic, despite the distance between generators. Our results also confirm the conclusions in Louie's paper [32], in that Gumbel and Gaussian copulas compete for accuracy in the bivariate representation of dependence structures. (Not so Clayton copula, which we have also processed for illustrative purposes.)

3.2. Multivariate assessment

Unfortunately, for $d \geq 3$ the KS two-sample test is computationally prohibitive by any of the two methods. Peacock's method requires $2^{d+1}N^d$ evaluations, and Fasano's and Franscechini's requires $2^{d+1}N$. So for $d \geq 3$ we alternatively employed a graphical procedure to analyze the goodness of each correlation model. In [35] Liu *et al.* introduced the DD-plot, short for data-depth plot, intended to graphically show the similitude between two multivariate samples; which may be considered an expansion of the bivariate PP-plot employed to analyze the goodness-of-fit in [30]. The technique is based on a measure of centrality (depth) of a given observation with respect to a multivariate distribution. In our case, the procedure consists in "pooling" the measured and the simulated sample. Then, we can obtain by a depth measure the depth of every observation in the measured and simulated samples, $\{\mathbf{P}^*\}_1^M$ and $\{\mathbf{P}\}_1^N$, against the pooled sample $\{\mathbf{P}^*\}_1^M \cup \{\mathbf{P}\}_1^N$. When the two samples come from the same underlying distribution, the depths of both samples in the pooled sample should be similar, and thus the data should be plotted near a diagonal of the DD-plot.

More rigorously, the DD-plot consists of the set of points defined as follows:

$$DD(F^*, F) = \left\{ (D_{F^*}(\mathbf{x}), D_F(\mathbf{x}))\,,\ \mathbf{x} \in \{\mathbf{P}^*\}_1^M \cup \{\mathbf{P}\}_1^N \right\}, \tag{19}$$

where D_F is a distance measure. Particularly from the several depth measures proposed in [35] we employed Mahalanobis depth which is simply computed as:

$$D_F(\mathbf{x}) := \frac{1}{1 + (\mathbf{x} - \boldsymbol{\mu}_F)\boldsymbol{\Sigma}_F^{-1}(\mathbf{x} - \boldsymbol{\mu}_F)}, \tag{20}$$

where $\boldsymbol{\mu}_F$ and $\boldsymbol{\Sigma}_F$ are the sample estimates of the mean vector and covariance matrix.

The first row of Fig. 6 represents a combination of five sites, where the distances between pairs are in the range of 5 to 25 km. This would be representative of a microgrid configuration. The graphic patterns are quite different in the four simulated methods. Liu *et al.* explained in [35] that these different patterns allude to distributional differences, such as location, scale, skewness, or kurtosis differences. More specifically, the Monte Carlo-Cholesky simulations (top left plot) shows "a form of pulling down from the point $(0.25, 0.25)$ to $(0, 0)$, leaving the upper right corner empty and spreading out the points around the midrange of the diagonal line, as if fitting a heart-shaped leaf on the diagonal pointing at $(0, 0)$." According to Liu *et al.* this is a clear case of location shift. And this location shift happened every time we tested the Monte Carlo-Cholesky technique; see for instance the left column of Fig. 6. It means that the distribution is clearly shifted in the form $F(\mathbf{P}) = F^*(\mathbf{P}^* - \alpha)$.

The behavior of the LHS-Iman & Conover simulations was remarkably more favorable than Monte Carlo in all our simulations, despite of the site combination or the problem dimension. Thus, in the first row of Fig. 6 the points are closely gathered around the diagonal line, indicating an accurate representation of the original data. In the second row (second box) this is not so, however, and the points of the LHS simulations are pulled up and away from the diagonal, thus losing accuracy. Yet, the simulation is quite accurate than its Monte Carlo counterpart.

The reason for LHS in the second row of Fig. 6 being less accurate than in the first row is that it shows processed wind speed instead of wind power. In our simulations the LHS method consistently performed worse when the simulation represented wind speed. The form of an "arch above the diagonal line in the shape of an early half moon" indicates, according to Liu *et al.*, that $F(\mathbf{P})$ and $F^*(\mathbf{P}^*)$ have the same center, but is more spread out than $F^*(\mathbf{P}^*)$. That is, it indicates a scale difference between the measured and the simulated samples. Additionally, a degree of skewness is also observed.

When the wind power is analyzed, the scale difference is evident in the copula simulations rather than in the LHS, and even more acutely in the Gumbel copula simulations; where according again to [35], a degree of augmented kurtosis can also be observed. This spread of points in the DD-plots of the copula simulations turns out to be even more evident when the dimension of the problem—the number of sites—increases. However, LHS simulations kept a reasonable accuracy even by augmenting the dimension.

It is remarkable to note the change of roles depending on the employed variable. When the variable was wind speed, Gaussian copula simulations were preferred in all our multivariate tests. Conversely, when the variable was wind power, LHS did better in most of the tests (though not in all; for instance in some scenarios representing $d \leq 4$) than Gaussian copula.

The last row represents a scenario of five distant generators. The distance between pairs was in the range from 250 km to 350 km. With these plots, we aim to close this paper by suggesting the distinctive simulation needs when the generators are spread over large distances in comparison to microgrid settings. It is observed There is not an appreciable error in selecting LHS or copula simulations when the geographical separation is large. Both are equivalent, though if we weighted the plots to magnify the region of depths lower than 0.5, we would see that LHS is not so accurate as Gaussian copulas.

4. Conclusions

In this paper we have discussed the guidelines for producing simulated samples of wind power in small power systems, such as microgrids. We have considered three methods. The first two methods, Monte Carlo with enforced correlation by means of Cholesky factorization and LHS modified by Iman's and Conover's methodology, are relatively easy to implement. The third method, copula simulation, requires more preparation. For instance, we have shown that the wind power empirical distribution is a mixed discrete-continuous distribution that requires a special interpretation. We have applied recent formulations found in statistics literature to represent multivariate Archimidean copula densities and draw samples from them. And finally we have discussed how to measure the goodness of the simulated samples, mindful that conventional goodness-of-fit tests do not apply.

On the whole, our results show that the most complex approach, copula theory, is not the best choice in every microgrid simulation. Gaussian copula simulations are better suited for representing wind power scenarios when the generator sites are substantially far apart. When they are instead clustered in some tens of kilometers, our results show that consistently the LHS simulations are more accurate. Only for low dimensions, less than four generators in a prospective microgrid, copula simulations should be preferred. In those cases Gumbel copula—which emphasizes an asymmetric correlation, favoring high wind power correlation—outperforms in a significant number of cases the accuracy of Gaussian copula; mostly when $d = 2$. Our results show also that the accuracy of copula simulations is deteriorated in small clusters when the number of considered generators is increased. Contrary to this trend, LHS simulations maintain a considerable accuracy.

Monte Carlo simulations followed by a transformation based on the Cholesky factorization of the covariance matrix is not advised. It is a simple and straightforward method that unfortunately leads to huge errors, because of the nonnormality of the wind power data.

Our results have additionally shown that the simulation of wind speed—instead of wind power—reverses the preference of LHS over Gaussian copula. This is quite remarkable, because by definition LHS (and Monte Carlo) is not a method which can be employed as a joint distribution function. LHS only serves the purpose of obtaining samples, but for instance it cannot be employed to compute conditional probabilities or, for that matter, to simply give a value of joint probability at a multivariate point. In such cases, copulas are the correct choice, and then the results that we have presented apply to the analysis of the copula limitations. That is, copula models loss accuracy when (i) wind power rather than wind speed is analyzed, (ii) the dimension of the problem is high, and (iii) the generators are closely gathered. In short, according to our results copula models should be avoided in the analysis of microgrids with more than three wind generators, if the wind power is chosen as the variable of interest. If possible, wind speed must be preferred as the random variable, because copula models (especially Gaussian) show to fit appreciably well even for large dimensions.

References

[1] N. Pogaku, M. Prodanovic, T. C. Green, Modeling, analysis and testing of autonomous operation of an inverter-based microgrid, IEEE Transactions on Power Electronics 22 (2007) 613–625.

[2] E. Barklund, N. Pogaku, M. Prodanovic, C. Hernandez-Aramburo, T. C. Green, Energy management in autonomous microgrid using stability-constrained droop control of inverters, IEEE Transactions on Power Electronics 23 (2008) 2346–2352.

[3] G. Diaz, C. Gonzalez-Moran, C. Viescas, Operating point of islanded microgrids consisting of conventional doubly fed induction generators and distributed supporting units, IET Renewable Power Generation 6 (2012) 303–314.

[4] G. Diaz, C. Gonzalez-Moran, J. Gomez-Aleixandre, A. Diez, Complex-valued state matrices for simple representation of large autonomous microgrids supplied by pq and vf generation, IEEE Transactions on Power Systems 24 (2009) 1720–1730.

[5] C. Yuen, A. Oudalov, A. Timbus, The provision of frequency control reserves from multiple microgrids, Industrial Electronics, IEEE Transactions on 58 (2011) 173–183.

[6] T. Niknam, F. Golestaneh, A. Malekpour, Probabilistic energy and operation management of a microgrid containing wind/photovoltaic/fuel cell generation and energy storage devices based on point estimate method and self-adaptive gravitational search algorithm, Energy 43 (2012) 427–437.

[7] A. D. Hawkes, M. A. Leach, Modelling high level system design and unit commitment for a microgrid, Applied Energy 86 (2009) 1253–1265.

[8] A. L. Dimeas, N. D. Hatziargyriou, Operation of a multiagent system for microgrid control, IEEE Transactions on Power Systems 20 (2005) 1447–1455.

[9] Y.-H. Chen, S.-Y. Lu, Y.-R. Chang, T.-T. Lee, M.-C. Hu, Economic analysis and optimal energy management models for microgrid systems: A case study in taiwan, Applied Energy 103 (2013) 145–154.

[10] M. Marzband, A. Sumper, A. Ruiz-lvarez, J. L. Domnguez-Garca, B. Tomoiag, Experimental evaluation of a real time energy management system for stand-alone microgrids in day-ahead markets, Applied Energy 106 (2013) 365–376.

[11] B. Khorramdel, M. Raoofat, Optimal stochastic reactive power scheduling in a microgrid considering voltage droop scheme of dgs and uncertainty of wind farms, Energy 45 (2012) 994–1006.

[12] E. Handschin, F. Neise, H. Neumann, R. Schultz, Optimal operation of dispersed generation under uncertainty using mathematical programming, International Journal of Electrical Power & Energy Systems 28 (2006) 618–626.

[13] S. Mohammadi, A. Mohammadi, Stochastic scenario-based model and investigating size of battery energy storage and thermal energy storage for micro-grid, International Journal of Electrical Power & Energy Systems 61 (2014) 531–546.

[14] T. Malakar, S. Goswami, A. Sinha, Optimum scheduling of micro grid connected wind-pumped storage hydro plant in a frequency based pricing environment, International Journal of Electrical Power & Energy Systems 54 (2014) 341–351.

[15] A. E. Feijoo, J. Cidras, J. L. G. Dornelas, Wind speed simulation in wind farms for steady-state security assessment of electrical power systems, IEEE Transaction on Energy Conversion 14 (1999) 1582–1588.

[16] J. M. Morales, R. Mnguez, A. J. Conejo, A methodology to generate statistically dependent wind speed scenarios, Applied Energy 87 (2010) 843–855.
[17] R. Khattree, C. Rao, Handbook of Statistics: Statistics in Industry, 2003.
[18] L. Baringo, A. J. Conejo, Correlated wind-power production and electric load scenarios for investment decisions, Applied Energy 101 (2013) 475–482.
[19] A. Feijo, D. Villanueva, J. L. Pazos, R. Sobolewski, Simulation of correlated wind speeds: A review, Renewable and Sustainable Energy Reviews 15 (2011) 2826–2832.
[20] R. A. Sobolewski, A. E. Feijo, Estimation of wind farms aggregated power output distributions, International Journal of Electrical Power & Energy Systems 46 (2013) 241–249.
[21] I. Segura-Heras, G. Escriv-Escriv, M. Alczar-Ortega, Wind farm electrical power production model for load flow analysis, Renewable Energy 36 (2011) 1008–1013.
[22] S. Hagspiel, A. Papaemannouil, M. Schmid, G. Andersson, Copula-based modeling of stochastic wind power in europe and implications for the swiss power grid, Applied Energy 96 (2012) 33–44.
[23] H. Holttinen, Hourly wind power variations in the nordic countries, Wind Energy 8 (2004) 173–195.
[24] N. B. Negra, O. Holmstrom, B. Bak-Jensen, P. Sorensen, Aspects of relevance in offshore wind farm reliability assessment, Energy Conversion, IEEE Transactions on 22 (2007) 159–166.
[25] O. Grothe, J. Schnieders, Spatial dependence in wind and optimal wind power allocation: A copula-based analysis, Energy Policy 39 (2011) 4742–4754.
[26] C. Delgado, J. Domínguez-Navarro, Point estimate method for probabilistic load flow of an unbalanced power distribution system with correlated wind and solar sources, International Journal of Electrical Power & Energy Systems 61 (2014) 267–278.
[27] R. L. Iman, W. J. Conover, A distribution-free approach to inducing rank correlation among input variables, Communications in Statistics - Simulation and Computation 11 (1982) 311–334.
[28] S. Zhen, P. Jirutitijaroen, Latin hypercube sampling techniques for power systems reliability analysis with renewable energy sources, IEEE Transactions on Power Systems 26 (2011) 2066–2073.
[29] G. Diaz, A. M. Abd-el Motaleb, V. Mier, On the capacity factor of distributed wind generation in droop-regulated microgrids, IEEE Transactions on Power Systems 28 (2012) 1738 – 1746.
[30] K. Xie, Y. Li, W. Li, Modelling wind speed dependence in system reliability assessment using copulas, Renewable Power Generation, IET 6 (2012) 392–399.
[31] H. Valizadeh Haghi, M. Tavakoli Bina, M. A. Golkar, S. M. Moghaddas-Tafreshi, Using copulas for analysis of large datasets in renewable distributed generation: Pv and wind power integration in iran, Renewable Energy 35 (2010) 1991–2000.
[32] H. Louie, Evaluation of bivariate archimedean and elliptical copulas to model wind power dependency structures, Wind Energy (http://dx.doi.org/10.1002/we.1571, 2012).
[33] G. Díaz, J. Gómez-Aleixandre, J. Coto, Statistical characterization of aggregated wind power from small clusters of generators, International Journal of Electrical Power & Energy Systems 62 (2014) 273–283.
[34] L. Ruschendorf, On the distributional transform, sklar's theorem, and the empirical copula process, Journal of Statistical Planning and Inference 139 (2009) 3921–3927.
[35] R. Y. Liu, J. M. Parelius, K. Singh, Multivariate analysis by data depth: descriptive statistics, graphics and inference,(with discussion and a rejoinder by liu and singh), The Annals of Statistics 27 (1999) 783–858.
[36] J. E. Angus, The probability integral transform and related results, SIAM Review 36 (1994) 652–654.
[37] S. H. Jangamshetti, V. G. Ran, Optimum siting of wind turbine generators, IEEE Transactions on Energy Conversion 16 (2001) 8–13.
[38] Z. Botev, J. Grotowski, D. Kroese, Kernel density estimation via diffusion, The Annals of Statistics 38 (2010) 2916–2957.
[39] N. Kolev, U. dos Anjos, B. V. d. M. Mendes, Copulas: A review and recent developments, Stochastic Models 22 (2006) 617–660.
[40] M. Niewiadomska-Bugaj, T. Kowalczyk, On grade transformation and its implications for copulas, Brazilian Journal of Probability and Statistics 19 (2005) 125–137.
[41] A. Basilevsky, Statistical Factor Analysis and Related Methods: Theory and Applications, Wiley, 1994.
[42] W. Press, Numerical Recipes 3rd Edition: The Art of Scientific Computing, Cambridge University Press, 2007.
[43] A. Der Kiureghian, P.-L. Liu, Structural reliability under incomplete probability information, Journal of Engineering Mechanics 112 (1986) 85–104.
[44] C. Lemieux, Monte Carlo and Quasi-Monte Carlo Sampling, Springer-Verlag New York, 2009.
[45] M. Stein, Large sample properties of simulations using latin hypercube sampling, Technometrics 29 (1987) 143–151.
[46] R. B. Nelsen, An introduction to copulas, Springer series in statistics, Springer, New York, 2nd edition, 2006.
[47] A. W. Marshall, Copulas, marginals, and joint distributions, Lecture Notes-Monograph Series 28 (1996) 213–222.
[48] H. Marius, M. Martin, J. M. Alexander, Estimators for archimedean copulas in high dimensions, CORD Conference Proceedings (2012).
[49] D. Ruppert, Statistics and data analysis for financial engineering, Springer texts in statistics, Springer, New York, 2011.
[50] M. Hofert, M. Mchler, A. J. McNeil, Likelihood inference for archimedean copulas in high dimensions under known margins, Journal of Multivariate Analysis 110 (2012) 133–150.
[51] A. W. Marshall, I. Olkin, Families of multivariate distributions, Journal of the American Statistical Association 83 (1988) 834–841.
[52] J. Peacock, Two-dimensional goodness-of-fit testing in astronomy, Monthly Notices of the Royal Astronomical Society 202 (1983) 615–627.
[53] G. Fasano, A. Franceschini, A multidimensional version of the kolmogorov-smirnov test, Monthly Notices of the Royal Astronomical Society 225 (1987) 155–170.

Table 1: Pairing of sites of Fig. 3.

Separation (km)	0–5	5–10	10–15	15–20	20–25	25–30
Number of sites	173	189	222	310	343	349

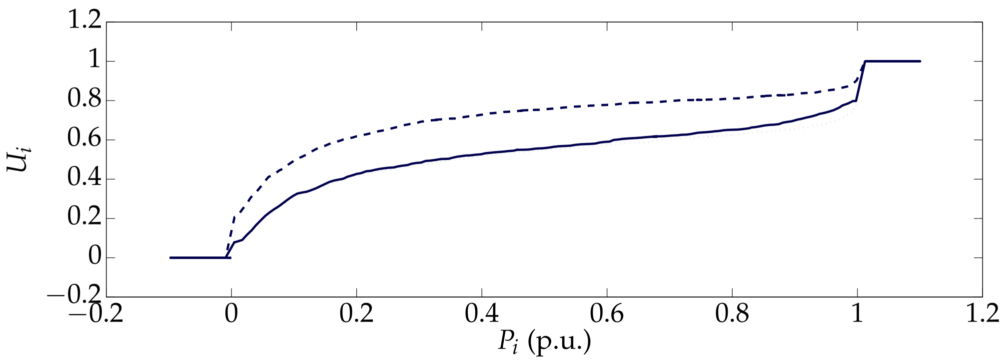

Figure 1: Kernel CDF estimation of wind power. Line styles: Solid, Enercon E40; dashed, Enercon E40 at a site 15 km apart; and dotted, GWP47 at the first site.

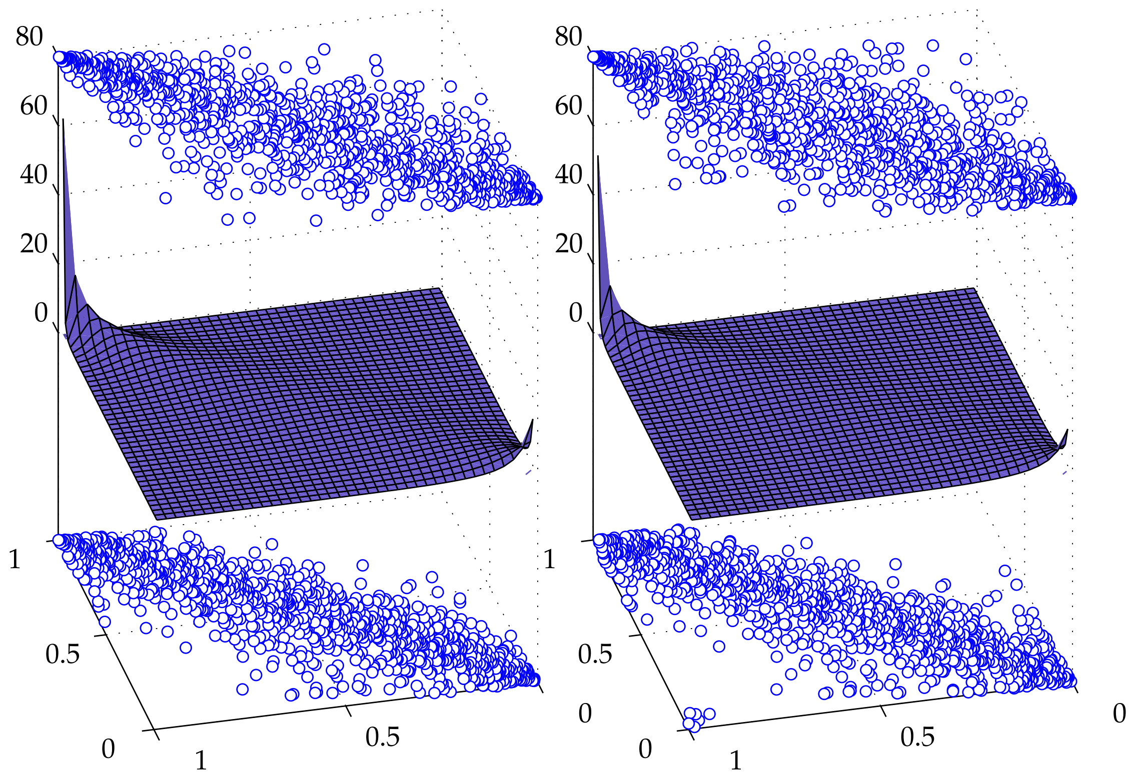

Figure 2: Left, wind speed; right, wind power obtained from those wind speeds by Enercon E40. The bottom scatter points represent the measured sample. From this sample, the copula density in the middle of the plot is obtained through maximum log-likelihood parameter estimation, after selecting a copula family. Here Gumbel copula is selected, leading to $\theta = 1.7016$ for wind speed and $\theta = 1.4674$ for wind power. Though copulas isolate the dependence structure from the marginals, the two are different because the transformation function (wind versus power of Enercon E40) is not *strictly* monotone. Finally, from the copula density the simulated samples (top scatter plots) are obtained by means of Marshal's and Olkin's method [51]. A sample of the same size has been produced in order to provide a fair comparison between measured and simulated samples.

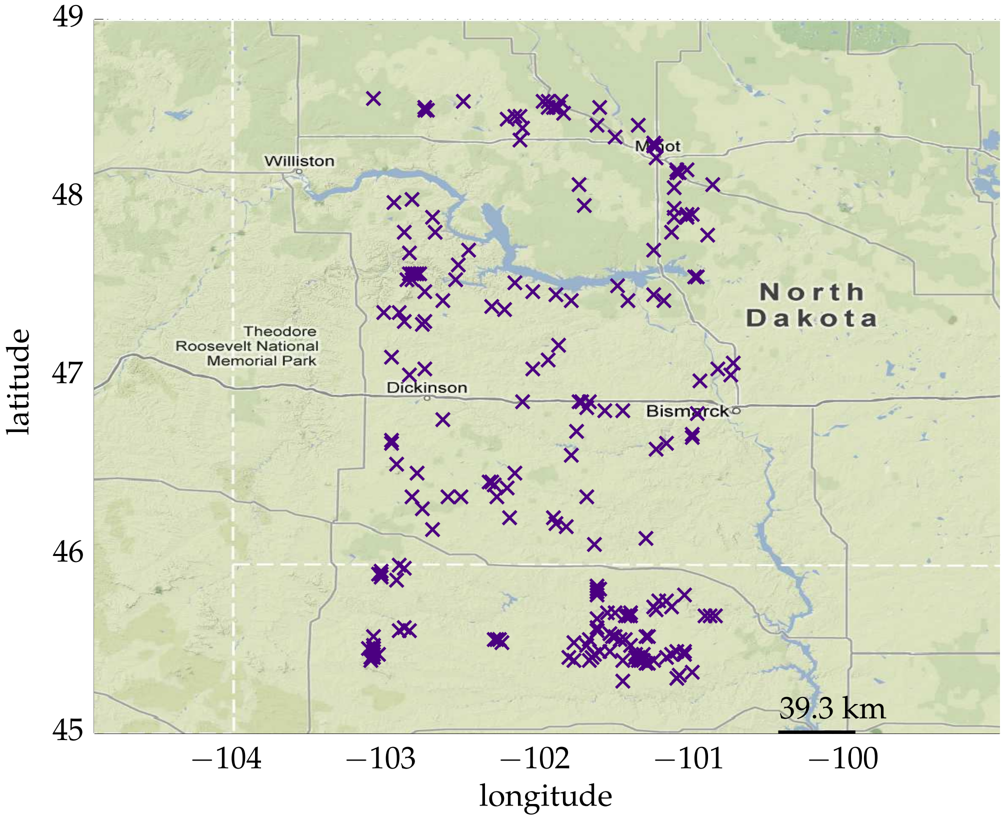

Figure 3: Location of the 210 analyzed sites.

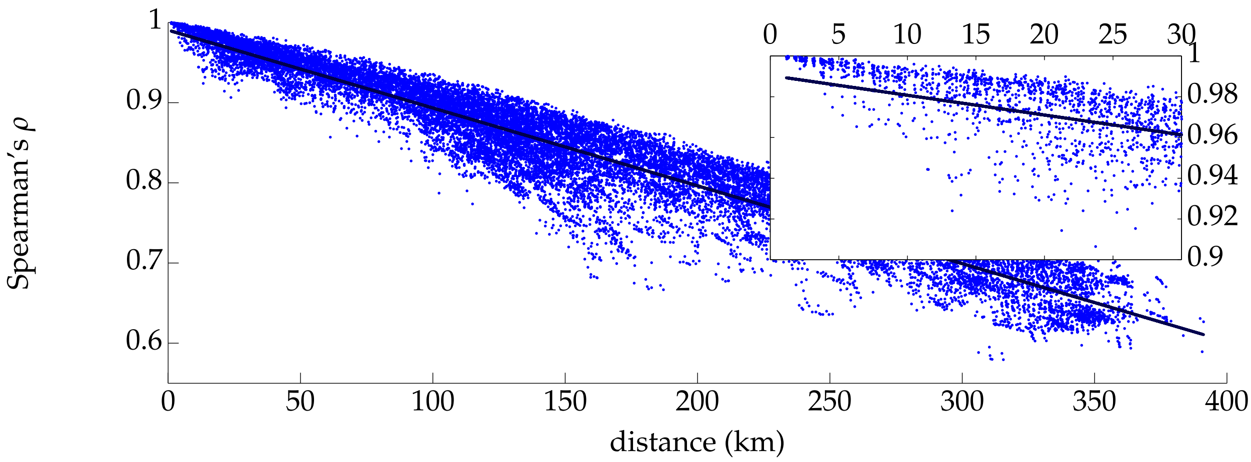

Figure 4: Spearman rank correlation as a function of the separation between pairs of sites in Fig. 3. 21945 pairs are possible.

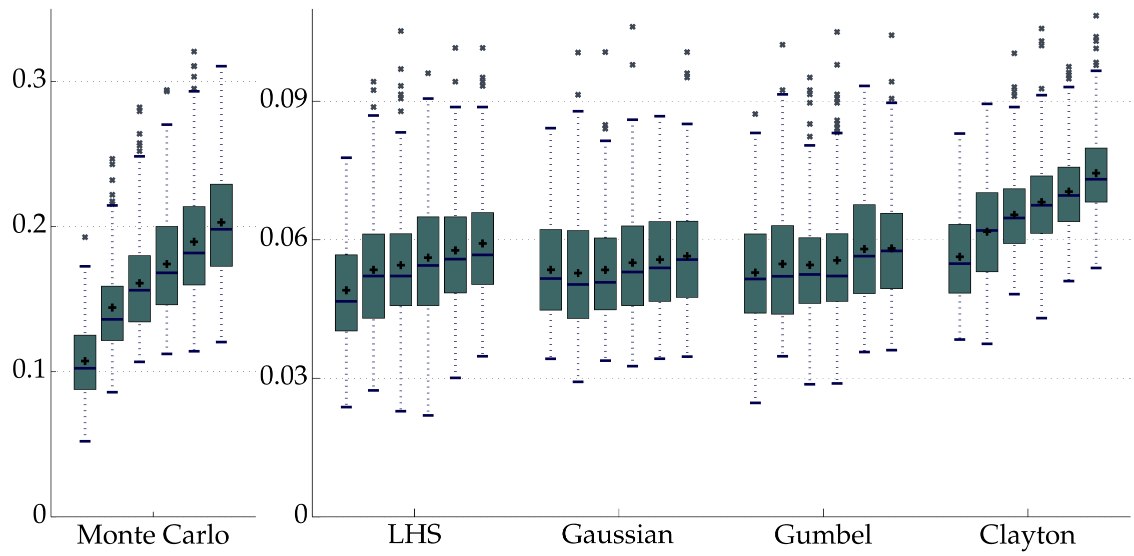

Figure 5: Box plot of the bivariate two-sample Kolmogorov-Smirnov statistic. For each of the five simulation methods analyzed, there are six instances representing, left to right, the pairings in Table 1. Monte Carlo-Cholesky method has a different scale.

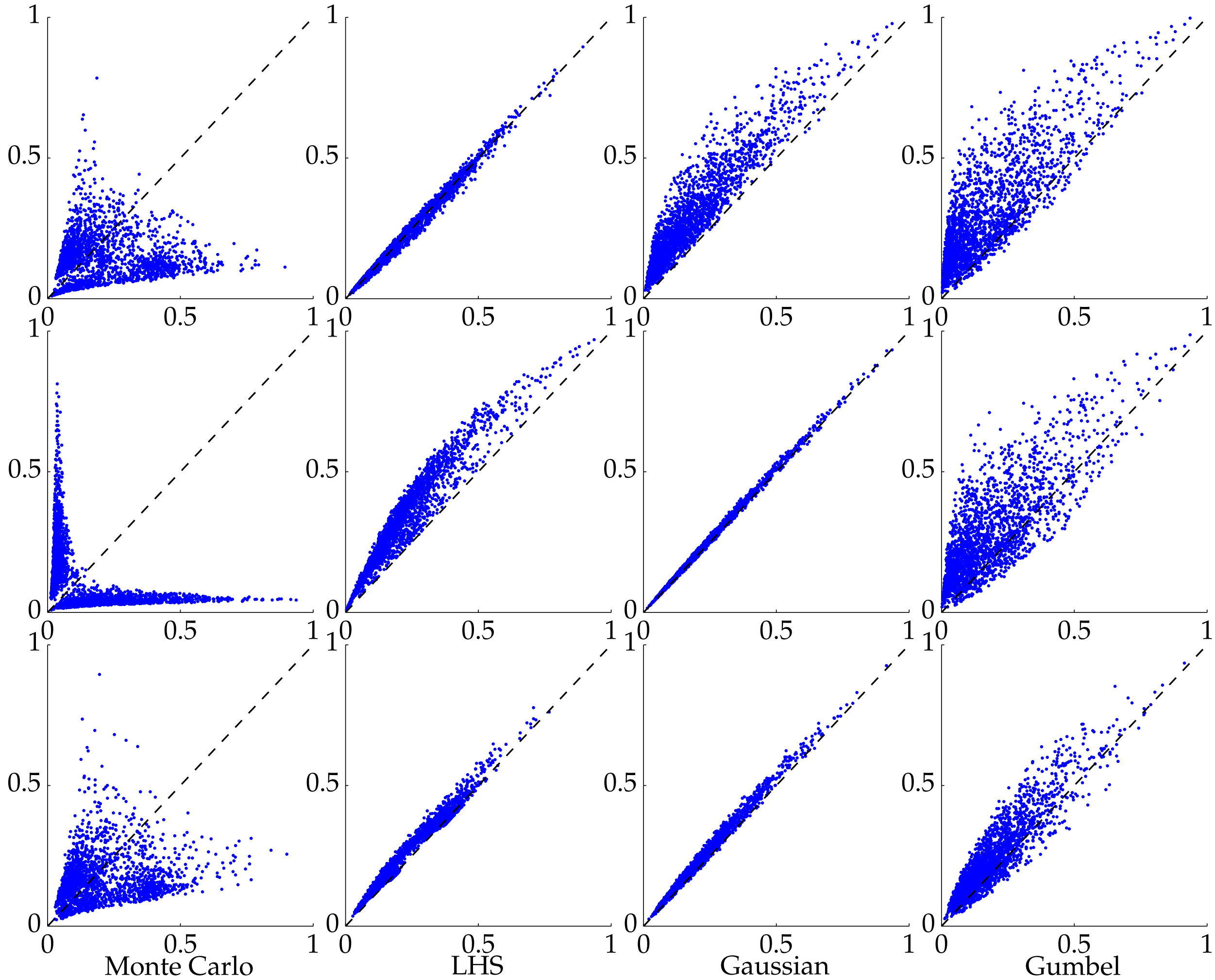

Figure 6: DD-plots [35] showing the goodness of every simulation method in a five-generator microgrid. The first row represents wind power simulation of close generators, with distances between sites ranging from 5 to 25 km. The second rows analyzes the same generator configuration, but represents wind speed instead of wind power. The third row shows the DD-plots of five generators, but differently from the first row, they are from 250 to 350 km apart.