Nonlinear Modeling of Temporal Wind Power Variations

Hamed Valizadeh Haghi, *Member, IEEE*, M. Tavakoli Bina, *Senior Member, IEEE*, and Masoud Aliakbar Golkar

***Abstract*—Modeling wind speed time series (WSTS) is an essential part of network planning studies in order to generate synthetic wind power time series (WPTS). Hence, this paper proposes a methodology to equip planners with accurate simulation of wind speed and power variations as well as complete temporal dependence structure based on the copula theory. Unlike traditional autoregressive and Markov chain methods, the suggested technique is well-prepared to deal with "nonlinear long-memory temporal dependence" and "non-Gaussian empirical probability distributions" of the WSTS. Meanwhile, the proposed statistical modeling framework is compatible with the scenario-based analysis of active networks as well. Furthermore, a case study for optimal sizing of an autonomous wind/photovoltaic/battery system is presented. The purpose of the presented study is to fully examine the accuracy and effectiveness of the copula-based model of wind generation for planning nonmemoryless power systems. Among a list of commercially available system devices, the optimal number and type of units are calculated ensuring both a minimum 20-year round total system cost and a perfect reliability. The genetic algorithm is used in four wind generation scenarios consisting of real and simulated WPTS. Then, considering the corresponding optimal solutions, the autoregressive moving average (ARMA), nonparametric Markov and proposed copula-based simulations are compared against real data.**

***Index Terms*—Copula, temporal dependence, time series, wind power.**

I. Introduction

THE intermittency of wind generation seldom creates problems when used to supply up to 20% of total electricity demand, but as the proportion increases, problems arise such as a need to use storage or a lowered ability to replace conventional generation. Given the significant growth and penetration of wind energy and other forms of stochastic generation, more active network strategies will have to be tailored to meet the needs of aggregated system balancing. Power networks will need to be designed specifically to deal with such high wind penetrations both to manage the frequency deviation and minimize curtailment of wind [1]–[5].

Before wind generation can be integrated into a power network, it is necessary to obtain good estimates of its potential contribution to that network. Hence, capturing the variability of the wind power time series (WPTS) would be required for the assumed geographic diversity [5]–[12]. Then, the wind variation in combination with load variations should be examined coupled with actual historic data and simulations [13]–[16]. These two procedures are to be used in planning simulations and modeling system characteristics and response. Because the length of historical data records are limited (obtained from the site under study), a conventional approach to long-term planning of combined wind and storage systems is to use parameterized time-series models [15], [12]. Subsequently, the models are simulated to produce alternative versions of the time series, representing what might happen over any arbitrary time periods in the future. This potentially provides many years of synthetic wind data that presumably resemble what a long-term dataset would look like. This is different from the forecasting process that describes the likely outcomes of the time series in the immediate future by performing on knowledge of the most recent outcomes.

A. Wind Modeling and Simulation

There are various techniques for modeling either distributional or temporal variations of the wind speed time series (WSTS) as well as the WPTS. Among them, Markov chain [9]–[12] and autoregressive moving average (ARMA) models and their generalizations [5]–[8] are recognized for modeling temporal dependence structure of wind speed; whereas, some multivariate probability distribution functions have been used for modeling spatial distributional dependence of wind speed variations [14], [16].

The WPTS usually obtained from modeling the recorded WSTS by using a turbine power curve. Some literature, however, recommend modeling the WPTS either directly from field records [8] or indirectly from converted WSTS measurements [12]. The first approach, i.e., transforming the modeled WSTS to WPTS by applying a suitable wind turbine power curve, is preferred by most of the literature mainly because wind speed data are usually easier to get for any given region [5].

The approach proposed here focuses on developing a statistical model of univariate WSTS, and then performing a wind-speed-to-power conversion. The proposal is intended to appropriately model the complete profiles of temporal wind power variations including seasonal, diurnal, nonstationary, and long-term fluctuations.

B. Motivation: Long-Memory Dependence and Nonlinear Transformation of Wind Power

Simulation of time series is the generation of synthetic time series with the same persistence properties (or temporal dependence) as the observed series. The term temporal dependence refers to the nonlinear dependence among WSTS values

Manuscript received September 14, 2012; revised January 17, 2013; accepted March 03, 2013. Date of publication April 01, 2013; date of current version September 16, 2013.

The authors are with the Faculty of Electrical Engineering, K. N. Toosi University of Technology, Tehran, 16314, Iran (e-mail: valizadeh@ieee.org; tavakoli_bina@ieee.org; golkar@eetd.kntu.ac.ir).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2013.2252433

at different time periods, i.e., the data cannot be assumed to be independent and identically distributed (i.i.d.). Both strong time-dependent behavior and instantaneous power balance of active power systems (generations = losses + demands) necessitates having an accurate model of temporal dependence for the WSTS; this will further make it possible to plan a higher wind penetration. Hence, a good practical wind time series model should take into account the following features.

1) A wind model should be valid statistically. Incorrect assumptions such as simple random sampling, linearity, or normality and even more complex fully parametric assumptions are among various causes for concern.
2) Wind time series exhibit strong signs of long-term or long-memory dependence. For long-memory processes, the variances of the sample mean and autocorrelations are not of asymptotic order n^{-1} (which is the usual rate for short-memory processes).
3) Wind time series exhibit seasonal variations. It should be noted that the seasonal variation is different from both the long-memory and nonstationary characteristics. Capturing nonstationary behavior should also be an objective.
4) Model of the WSTS dependence structure, which is nonlinear itself, should be robust against the nonlinear transformation to the WPTS output by applying a wind turbine power curve. In other words, the WSTS dependence structure should be transformed to a correct WPTS dependence structure. Nonetheless, the linear dependence measures neither completely characterize nor robustly preserve the variation patterns of the WSTS especially over long-term.

These mentioned features have to be taken into account in order to avoid disastrous effects on statistical inference in wind power integration planning. Available models for the WSTS (to date) ignore some of these features while retaining practical applicability. The main shortcomings are as follows:

- The ARMA model and its generalizations consider a few basic assumptions, assigning a linear temporal dependence and identically distributed random variables.
- The ARMA models do not necessarily retain the probability distribution of the original data. The majority of the existing literature assumes the i.i.d. data which is unrealistic.
- The real wind inherently is correlated over a longer period of time, needs long-memory modeling in a practical implementation [17]–[20]. Hence, for example, even well-fitted conventional Markov models, in particular under a short time-step, underestimate the amount of energy storage necessary [21].
- The seasonal variations should be removed by a filter for successful application of existing approaches; also, transforming to normality is sometimes necessary. However, common filtering techniques or transforming to Gaussian distributions usually lead to rejection of the i.i.d. hypothesis due to model misidentification. The immediate problem arising from the empirical observations is how to deal with seasonal variations of the data with respect to the observed volatile structure [5], [6]. However, the study by [22] indicates that removing seasonal patterns, by any means, has a strong impact on the analysis of distributional dependence and on the interpretational power of common dependence measures. This is the case even if the autocorrelation functions (ACFs) of the residual time series are approximated to a white noise [5].
- Transforming modeled WSTS through a wind turbine power curve (a nonlinear transformation) would destroy the original dependence structure. This appears as an example in the WPTS if ARMA or conventional Markov models are used.

Some papers propose using the fractional ARMA approach (e.g., [17] for a comprehensive WSTS modeling) to deal with the long-memory effects and using some transforms to remove seasonal variations; however, the above-mentioned issues remain because of the inherent characteristics of the ARMA or conventional Markov models.

This paper investigates the temporal dependence measures with respect to various models, utilizing a real recorded data set. The main focus, however, will be on a *copula-based* method to retain the real temporal and distributional characteristics of the WSTS with regard to the above-mentioned issues. A detailed examination of a micro-grid shows that unlike the copula-based method, the available methods inherit the danger of wrong conclusions from inappropriate dependence measures. Then, various models are analytically compared in terms of the stated dependence and distributional structure. The well-known nonparametric Kruskal–Wallis test (KWt) was also carried out to examine various models and verify the proposed method. The ACF of the WSTS based on scale-invariant (copula-based) dependence measures is used to provide new insights into the interplay between distributional and temporal dependence of the WSTS as well as the WPTS. This could be made possible because each multivariate distribution function can be split up into its univariate marginal distribution functions and a copula function. In other words, copulas enable us to study the distributional and temporal dependence structure of random vectors irrespective of their marginal distributions.

II. Conceptual Modeling of Wind Characteristics

A. Wind Speed Multivariate Temporal Structure

The WSTS, like other stochastic processes, could be generally treated as a family of random variables defined on a given probability space. It is indexed by the time variable t, where t varies over an index set T. An illustration of this definition is presented in Fig. 1(a), assuming an hourly time index over a 24-h period. Random variables could be completely different at various times.

Although the random values of a stochastic process may be independent random variables at different times, in most commonly considered situations they exhibit complicated statistical dependences. Such a dependence structure is illustrated in Fig. 1(b) for the observed hourly wind speed data over one year. In planning context, it should be useful to simulate the wind speed resembling these general characteristics. Indeed, this is a standard format to express any other stochastic process as well.

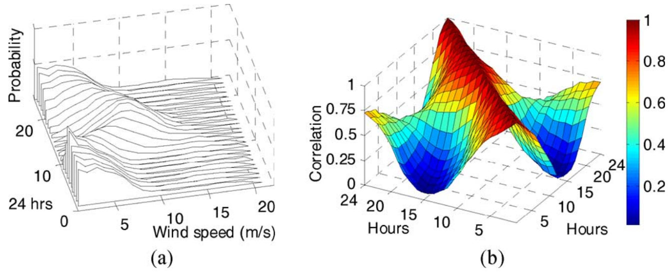


Fig. 1. (a) Hourly empirical probability distributions of wind speed values based on the recorded data of Manjil site over one year. (b) Temporal correlation structure of a typical WSTS with $T = 24$ h.

B. Modeling Methodology

Copulas provide a way to create distributions that model correlated multivariate data, in particular, when there are various general types of dependence structures, effectively more than three correlated variables, or variables with different or empirical distribution functions [23]. Copula methods are popular in many different areas where many different correlated factors must be modeled jointly. The application of copulas in power system problems is relatively new [24]–[30]. Copulas often help to perform large-scale multivariate simulations of random vectors that would be difficult to perform using other multivariate fitting and simulation methods.

The multivariate temporal structure of Fig. 1(a) demonstrates typical empirical distributions of wind speed in order to be correlated using a copula. In other words, the main goal is to represent the stochastic process of Fig. 1(a), a 24-variate probability distribution with 24 empirical marginal distributions, in a way to get a dependence structure like that of Fig. 1(b). This can be generalized by varying the time resolution, in general, an n-variate copula over an index set T. For example, considering $t = 10$ min (time resolution) over $T = 24$ h results in $n = 144$.

It is shown in [23] that the copula function C is uniquely determined by the multivariate cumulative distribution function (cdf) F and n marginal cdf F_i $(i = 1, \ldots, n)$ if all F_i are continuous (Sklar's Theorem) as follows:

$$C[F_1(x_1), F_2(x_2), \ldots, F_n(x_n)] = F(x_1, x_2, \ldots, x_n). \quad (1)$$

Let f and $f_i, i = 1, \ldots, n$ be the corresponding joint density and marginal densities, respectively. Then, copula density c is defined by

$$c(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial F_1(x_1) \ldots \partial F_n(x_n)}. \quad (2)$$

Consequently, f can be expressed by

$$f(x_1, x_2, \ldots, x_n) = c(F_1(x_1), \ldots, F_n(x_n)) \prod_{i=1}^{n} f_i(x_i). \quad (3)$$

Different families of copulas present various dependence details. Elliptical copulas (the Gaussian and t copulas) can easily be generalized to higher dimensions, allowing any marginal distributions and any positive definite dependence matrix. Archimedean copulas provide the capacity to represent more complicated dependence structures [14]. However, this letter uses the Gaussian copula since it performs suitably in modeling wind speed and can be easily put in practice. Nonetheless, any other type of copula can be applied in the proposed algorithm given that the following conditions are considered:

- The copula should be able to represent/model the negative correlation as well as the positive correlation. For example, some of Archimedean copulas such as Gumbel or Clayton represent data only in terms of positive correlation; whereas, the Frank copula which is an Archimedean copula as well, models the whole range of correlation [14].
- Since the number of variates in the proposed method is higher than 3, sample generation and manipulation algorithms depending on the type of the chosen copula may need complicated procedures [14]. In this regard, Gaussian and t copulas provide simple and practically attractive computer codes [31]. It is more convenient to implement Archimedean copulas in higher dimensions by commercially available software packages [32].

The correlation matrix is constructed by the Spearman rank correlation that is more appropriate to fulfill the desirable characteristics of a measure of dependence [3]. Furthermore, estimated empirical cdf with kernel smoothing have been used for F_i. Fig. 2 illustrates the full procedure by a pictorial flowchart.

The fitting procedure (step 3 of Fig. 2) is detailed by the following procedure as a practical way to generate sample data using Gaussian copula:

1) Fit marginal distributions to the columns of the modeling matrix $W_{n_s \times n_v}$ built in step 2 in Fig. 2.
2) Use empirical cdf functions to transform W to U_W, so that U_W has values between 0 and 1.
3) Calculate an estimation of the dependence parameter of the copula considering data in U_W. Note that the marginal distributions are all inverse cdf's fitted in steps 1 and 2 above, hence, the dependence parameter of the copula is the main estimation to be made at this stage. The maximum likelihood method is applied to obtain the copula parameters estimates. This may be referred to as the pair-wise correlation matching problem that computes the most fitting positive semidefinite matrix of the dependence parameter simply by applying the maximum likelihood method [33].
4) Generate new data $\widehat{U_W}$ from the copula by traditional sampling.
5) Use appropriate inverse cdf functions to transform $\widehat{U_W}$ to $\widehat{W}$.

In brief, the proposed method simulates wind speed using a Gaussian copula as follows:

1) Choose the time resolution that gives the total number of sample points n, modeling index set or basic modeling time T_m, and total modeling time T.
2) Divide the whole data into n_s subsets and prepare $n_v = n/n_s$ empirical cdf from the history of wind speed data.
3) Calculate $n_v \times n_v$ rank correlation matrix.
4) Fit the copula to the n_v-variate data.
5) Simulate using the fitted copula and n_v marginal cdf.

III. Verification of the Suggested Model

The suggested approach is applied to the wind speed data gathered from Manjil Wind Park in northern Iran (2007–2008).

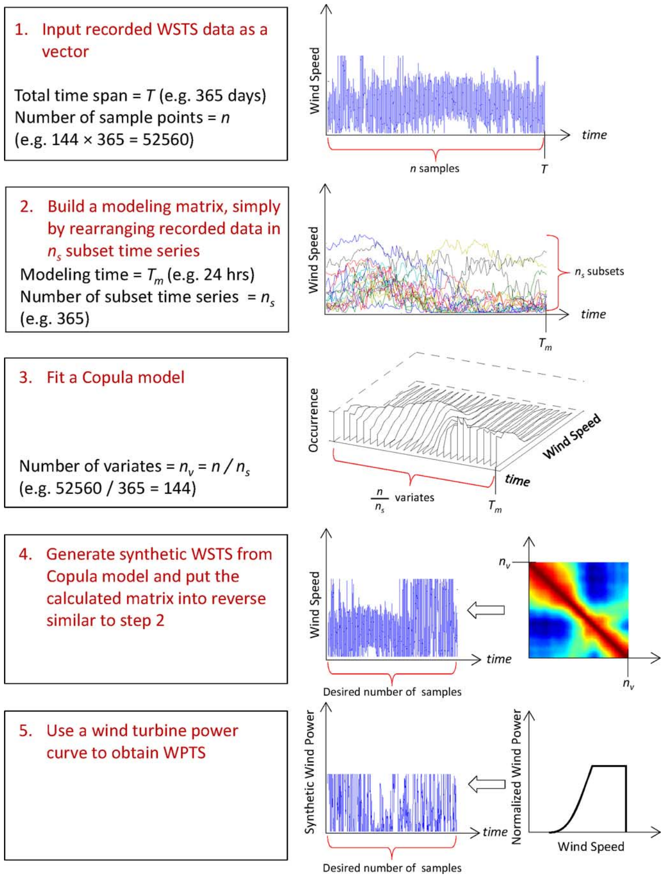


Fig. 2. Schematic representation of the proposed copula-based wind modeling.

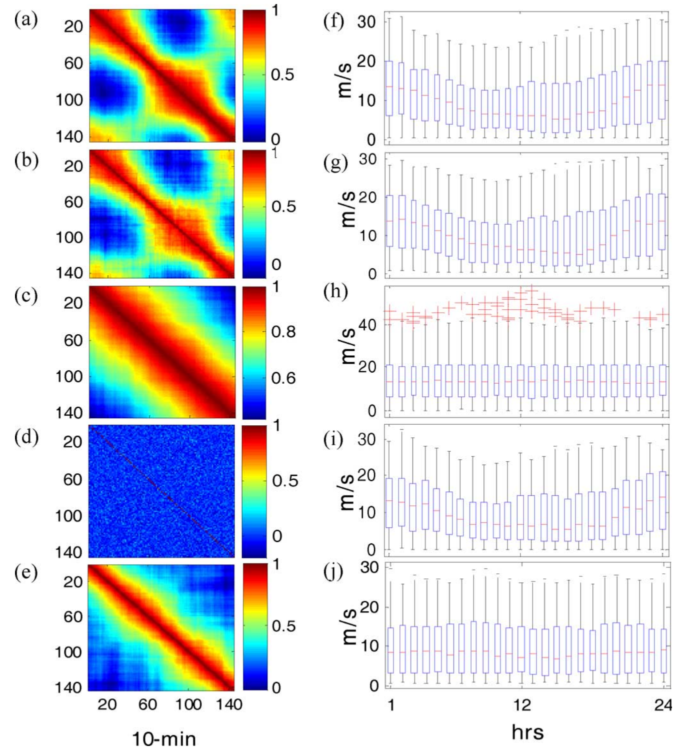


Fig. 3. Temporal correlation structure (left column) and "box and whisker" plots (right column) of wind speed based on (a) and (f) exact data, (b) and (g) the proposed method, (c) and (h) ARMA(4,3), (d) and (i) marginal distributions apart from correlations, and (e) and (j) nonparametric Markov chain with 21 states.

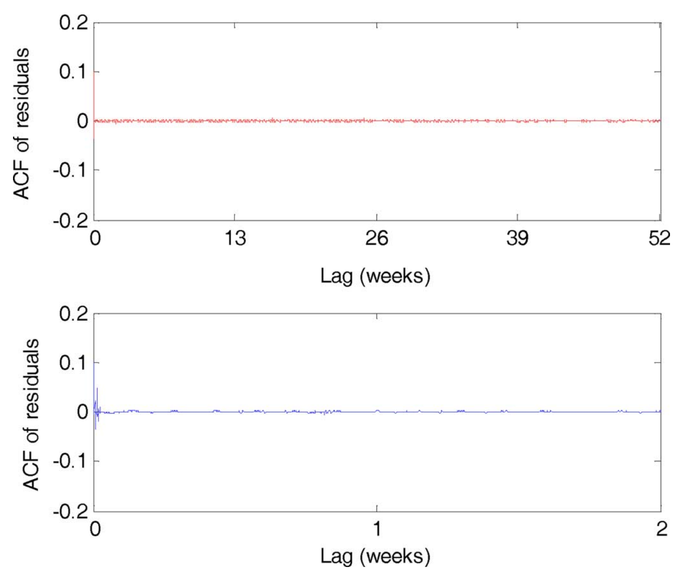


Fig. 4. ACF of residuals of the employed ARMA for Manjil area, showing adequate ARMA model estimation.

The simulation was carried out under a 10-min time resolution, i.e., $n = 52\,560$, $n_v = 144$, $n_s = 350$, $T_m = 24$ h, and $T = 1$ year. Thus, Fig. 3(a)–(e) presents the dependence structure of the simulated wind speed. Interestingly, while the ARMA model assumes an overestimated linear correlation, the proposed method captures complete temporal structure of wind speed. On the other hand, the Markov model does not replicate the correct dependence in corners, i.e., an important time evolution characteristic. It should be mentioned that the application of the employed ARMA model includes determining the model order, seasonal correction, and verification tests. For instance, Fig. 4 shows that the residuals of the simulation from the employed ARMA model approximate to white noise; therefore, the model can be accepted as satisfactory.

Fig. 3(f)–(j) (the box and whisker plots) compares actual data with simulations performed by the proposed method, ARMA, Markov chain, and marginal-only simulations (steps 1 and 2 of the proposed method only). A box and whisker plot is a way of graphically comparing distributions between several sets of data through their five-number summaries. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually by a plus sign. Fig. 3(f)–(j) shows that the distributions of the WSTS at any time are closest to the simulated WSTS by the proposed method. Hence, these graphs demonstrate that the proposed method, unlike the ARMA and Markov chain, retains the probability distribution of the original data.

Fig. 5 shows the dependence structure of the WPTS corresponding to the data in Fig. 3. Wind speed measurements and simulations are converted to power generation using a curve developed for a generic turbine by the National Renewable Energy Laboratory [34]. It is obvious that converting simulated wind speed to wind power using the nonlinear power curve of the turbine would destroy the original dependence structure when linear correlation measures are implicitly or explicitly used in the model. This is the case when using the ARMA and conventional Markov models. However, the proposed model employs

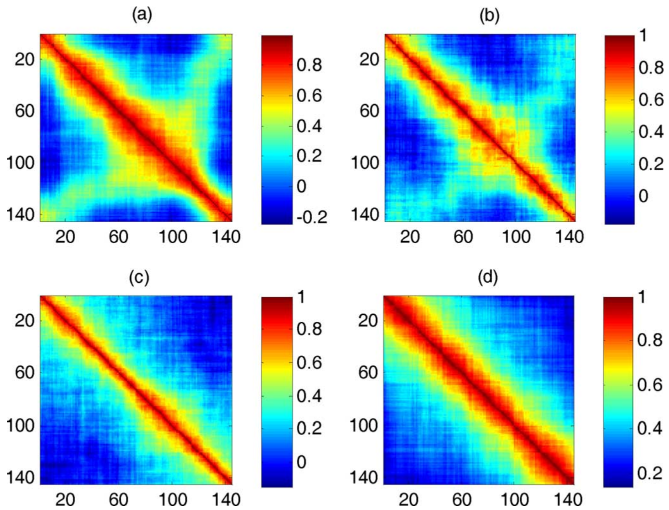


Fig. 5. Temporal correlation structure of converted wind generation based on (a) exact data, (b) the proposed method, (c) nonparametric Markov chain with 21 states, and (d) ARMA(4,3). (a) Real. (b) Copula. (c) NP Markov. (d) ARMA.

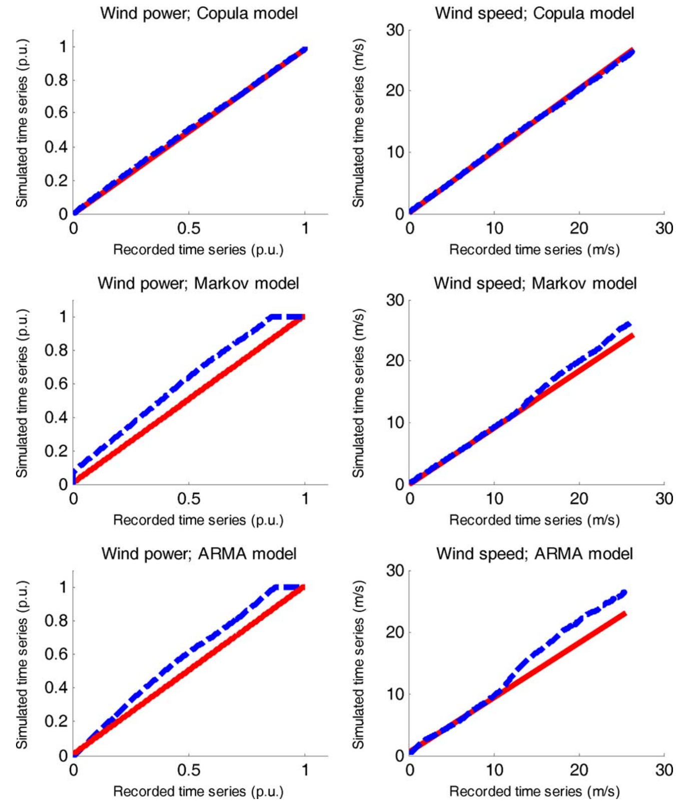
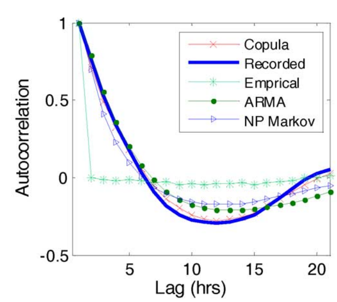
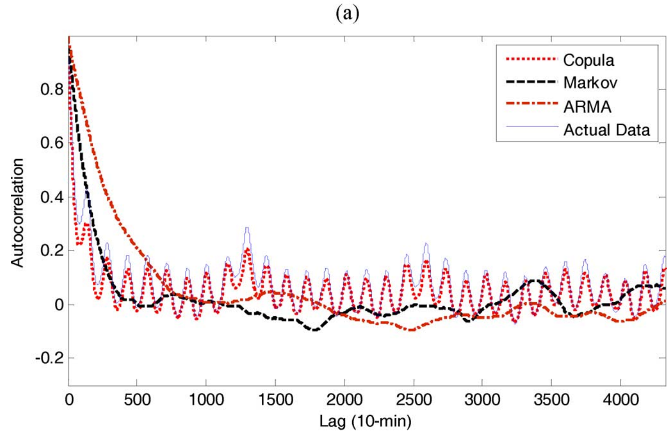


Fig. 6. Simulated WPTS and WSTS against real recordings for the three models. The dashed lines are the estimate.



(a)



(b)

Fig. 7. (a) ACF with $n_v = 24$, (b) over one year T with $n_v = 144$.

the Spearman rank correlation as stated in Section II; therefore, it measures the correlation only in terms of ranks. As a consequence, the rank correlation is preserved under any monotonic linear or nonlinear transformation. In particular, the transformation by the power curve preserves the rank correlation. Therefore, modeling the rank correlation structure of the WSTS exactly determines the rank correlation of the final transformed WPTS. While the linear correlation coefficient is needed to parameterize the underlying process, the Spearman correlation is more useful in describing the dependence between random variables, because it is invariant to the choice of marginal distributions.

An effect of transforming the real temporal dependence through the turbines power curve is shown in Fig. 6, where the simulated WPTS and WSTS are plotted against real recordings for the three models. The perfect modeling would fit the straight line.

Moreover, Fig. 7 performs a similar comparison in terms of the autocorrelation function (ACF). Fig. 7(a) shows the ACF over the basic modeling time span whereas Fig. 7(b) compares the ACF over a one-year extension of the simulated data. If a more appropriate selection of n_v, T, and T_m was performed for the available data in advance, then the accuracy of the proposed method would be even better. Thus, Fig. 7(b) illustrates significant potential of the proposed method in modeling long-memory dependence as well as periodic variations of wind speed dependence structure. The latter represents a cyclic variation known as seasonality, periodic variation, or periodic fluctuations. This variation can be either regular or semi-regular with a period of less than one year that should be measured apart from the deterministic trend (e.g., diurnal variations would be categorized as seasonal but exponential growth would not). If the exact period is known, seasonal subseries plots are a tool for detecting seasonality in a time series, whereas, if the period is not known, an autocorrelation plot or spectral plot can be used to determine it [35]. The WSTS seasonal patterns are not exactly identifiable *a priori* [36]; hence, we have used autocorrelation plots to demonstrate capability of the proposed model for representing seasonal variations. The autocorrelation plot should show peaks at lags equal to the period of the seasonal trends. For example, since there is a diurnal seasonality effect on the WSTS, it is expected to see significant peaks at lags 144, 288, 432, and so on according to

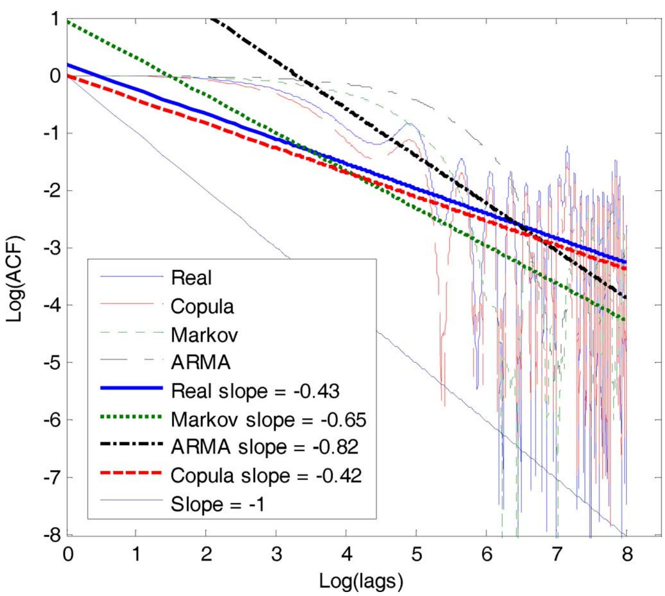


Fig. 8. ACF of simulated and recorded WSTS in the logarithmic scale showing the long-term dependence of wind.

a 10-min data resolution. Fig. 7(b) confirms this, although the amplitude of peaks decreases over time. Fig. 7(b) also shows patterns of monthly seasonality as the peak at lag 4320 (i.e., 144×30) is slightly higher than the preceding quarterly peak at lag 3744.

The presence of long memory in the WSTS can be seen in Fig. 8. Temporal dependence and autocorrelations decay proportional to $n^{-\alpha}$ ($\alpha \in (0, 1)$ and n is the lag). In other words, the temporal dependence structure considering sum of all correlations with lags $-(n-1), -(n-2), \ldots, -1, 0, 1, 2, \ldots, n-2, n-1$ must be proportional to $n^{1-\alpha}$, i.e.,

$$\sum_{k=-(n-1)}^{n-1} \rho(k) \approx C_\rho \cdot n^{1-\alpha} \quad (4)$$

where $\alpha < 1$ implies

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty. \quad (5)$$

Thus, slow decay rate of the correlations tends the summation in (5) to infinity. More specifically, (4) is correct if

$$\rho(k) \approx C_\rho |k|^{-\alpha} \quad (6)$$

where $|k|$ tends to infinity and C_ρ is a finite positive constant. The intuitive interpretation of (6) is that the process has long memory. In other words, the dependence between events that are far apart in time diminishes very slowly with increasing distance. This is in contrast to the processes with summable correlations which are also called processes with short memory or weak temporal dependence. For example, the asymptotic decay of the correlations for ARMA and Markov processes is exponential in the sense that there is an upper bound

$$|\rho(k)| \leq ba^k \quad (7)$$

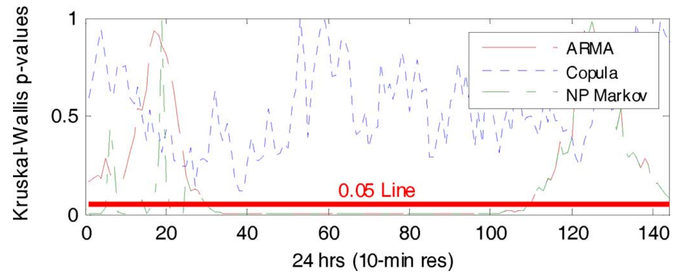


Fig. 9. KWt results for all discussed models.

TABLE I
MSE Based on the Sample Mean

	Proposed copula	Markov	ARMA
MSE of sample mean for one day	0.3067	3.3647	3.8499
MSE of sample mean for one month	7.6372	15.1937	12.6897
MSE of sample mean for one season	21.2035	36.2979	32.0317
MSE of sample mean for one year	68.9902	114.8487	103.0329

where $0 < b < \infty, 0 < a < 1$ are constants. Because the absolute value of α is less than 1, (5) does not hold and

$$\sum_{k=-\infty}^{\infty} \rho(k) = \text{Bounded Constant}. \quad (8)$$

Fig. 8 indicates that the slope of the fitted least squares line is equal to -0.43, far from -1 as a theoretical summable correlation. This plot of the autocorrelations in log-log coordinates suggests a slow decay of long-lasting temporal dependences. This is typically the case for processes with nonsummable correlations [37]. The complicated dependence structure as well as the slope of the fitted line is completely modeled by using the copula model; however, other models could not capture the exact dependence structure.

On the other hand, in order to evaluate the hypothesis that the observed and simulated samples are drawn from the same distribution, the Kruskal–Wallis test (KWt) provides an effective verification. Fig. 9 shows the p-values resulting from KWt, indicating that the ARMA and nonparametric Markov simulations of temporal pdf is rejected at 95% confidence level in view of null hypothesis (all samples are simulated with the same distribution). Meanwhile, the proposed modeling is strongly verified at all time instants by the KWt in Fig. 9. It should be mentioned that if the p-values, i.e., the y-axis of Fig. 9, is near zero, this casts doubt on the null hypothesis, suggesting that at least one sample median is significantly different from the others. The choice of a critical p-value to determine whether the result is judged statistically significant or strongly verified is often 0.05 or 0.01. When the p-value is less than this significance level, the test rejects the null hypothesis. The proposed method would also be verified by choosing a higher significance level of 0.15 that could be deemed an extraordinary verification within the most extreme 15% of all possible results under the null hypothesis.

Finally, the mean squared error (MSE) has been calculated based on the sample mean for the three models in comparison with the real recorded WSTS. The obtained results are included in Table I. It should be emphasized that modeling time series then simulating the model to produce alternative versions of the time series, is different from the forecasting process. The latter describes the likely outcomes of the time series in the immediate future by performing on knowledge of the most recent outcomes point-by-point. Hence, calculating the MSE would be helpful as long as we use the "sample mean" of the simulated WSTS. The sample mean or empirical mean is a statistic computed from a collection of data on one or more random variables. The sample mean is a vector defined as follows [38]. Let w_{ij} be the ith independently drawn observation $(i = 1, \ldots, n_s)$ on the jth random variable $(j = 1, \ldots, T_m)$. These observations can be arranged into n_s column vectors, each with T_m entries, with the ith observation denoted w_i. This assumption is in accordance with the proposed algorithm as $n_s = 365$ and $T_m = 24$. The sample mean vector $\bar{w}$ is a column vector whose jth element $\bar{w}_i$ is the average value of the n_s observations of the jth variable

$$\bar{w}_i = \frac{1}{n_s} \sum_{i=1}^{n_s} w_{ji}, j = 1, \ldots, T_m. \tag{9}$$

IV. Assessing Validity of the Model: Studying Optimal Sizing of a Microgrid

A considerable share of renewable distributed generation increases the uncertainty in sizing, planning, and operation of both microgrids and standalone power systems. These uncertainties affect both long-term and medium-term system planning as well as the day-ahead operation. This explains the growing importance of probabilistic tools in such applications. Considering the uncertainty in planning studies makes the whole system less expensive and more secure.

Generally speaking, many factors can introduce uncertainties for sizing microgrids. Examples are the nominal production, cost of energy, cost of gas or fuel, environmental constraints, stochastic generation, and load growth. In this section, it has been assumed that the power production of wind turbines (WTs) is the main cause of uncertainty. The other factors (cost of energy, cost of gas or fuel, and so on) are assumed to be deterministic. This assumption allows us to effectively compare the proposed modeling of WSTS with those of the ARMA and Markov models. The remote microgrid under study, as shown in Fig. 10, has several battery units to meet the required level of storage. The optimal energy storage sizing is a main determinant of a successful operation and the optimal cost as well. The storage sizing problem has been frequently addressed in the literature for remote areas [4], [39]–[41], using a variety of optimization techniques. In this paper, the optimal sizing of a hybrid standalone system is investigated with a genetic algorithm. The methodology is generally based on [41], in which the microgrid consists of some wind turbines (WTs) and photovoltaic (PV) units along with some battery storages.

It is assumed that both WT and PV units are connected to a common dc distribution network. The battery banks are to be charged from the respective PV and WT input power sources, which are usually configured in multiple power generation blocks. An inverter is used to interface the dc battery voltage to the consumer load ac requirements. The outputs of all battery chargers, the battery banks, and the inverter input terminals are

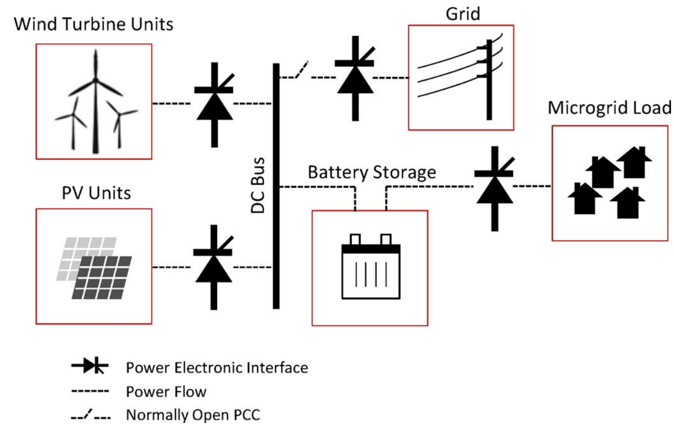


Fig. 10. Concept of the case study.

connected in parallel according to Fig. 10. The energy produced from each PV or WT source is transferred to the consumer load through the battery charger and the dc/ac inverter, while the energy surplus is used to charge the battery banks. This forms a kind of dc distribution network for integrating various generators, storage systems, and dc loads in a local dc distribution network that is interfaced with public grid by means of one or more inverters. This solution makes the islanding operation easy for the generators.

The sizing of the components can be seen as a multiobjective cost optimization problem which takes place over three main stages with the following information:

1) Input data: specifications of battery, PV modules, WTs, chargers, and inverters as well as wind speed, solar radiation, electrical load, and temperature time series.
2) System modeling and operation: wind turbine power curve that gives WPTS, modeling PV modules by considering maximum power point tracking (MPPT) and temperature that gives the PV output, and a model of battery bank operation considering charging characteristics and state-of-charge (SOC) at every time instant throughout the whole year. The modeling equations is used to verify whether a solution derived by the GA-based optimization fulfills the load power requirements during the whole year while considering SOC and other constraints.
3) Cost optimization: a multiobjective cost function of the individual system devices capital and 20-year round maintenance costs optimized using the conventional GA routine with crossover and mutation.

The net present cost (NPC) of the nth component can be calculated via

$$\mathrm{NPC}_n = N_n \times \{\mathrm{CC}_n + \mathrm{RC}_n \times K_n + \mathrm{OM}_n \times \mathrm{AP}(\mathrm{IR}, R)\} \tag{10}$$

where N_n is the number/rating, CC_n is the capital cost, RC_n is the replacement cost, and OM_n is the operation and maintenance cost, all for nth equipment ($n = 1, 2, 3$ represent wind turbines, PV units, and battery storage, respectively). In addition, R is the project's lifespan, and *IR* is the so-called real interest rate that is assumed to be 0.08. Other parameters, *AP* and

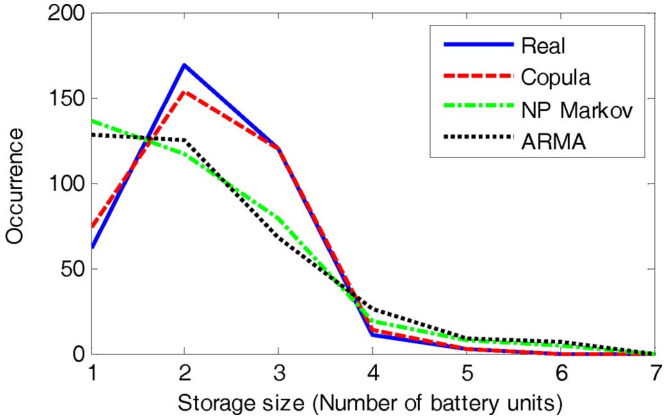


Fig. 11. Occurrences of the optimal storage estimations against real storage requirements.

K are annual payment present worth and single payment present worth, respectively, defined as

$$\mathrm{AP}(\mathrm{IR}, R) = \frac{(1+\mathrm{IR})^R - 1}{\mathrm{IR} \times (1+\mathrm{IR})^R}, K_n = \sum_{j=1}^{y_n} \frac{1}{(1+\mathrm{IR})^{j \times L_n}} \tag{11}$$

where y is the number of times each component is replaced and L_n is the lifespan of each component. Therefore, based on the definition of NPC along with the constraints, a multiobjective optimization problem could be written as

$$\begin{aligned} &\underset{x}{\text{Minimize}} \sum_n \mathrm{NPC}_n \\ &\text{with respect to } N_n \\ &\text{subject to } N_n \geq 0 \\ &0 \leq E_{\mathrm{batt}(i)} \leq E_{\max} \\ &P_{\mathrm{WT\_conv}(i)} + P_{\mathrm{PV\_conv}(i)} = P_{\mathrm{load}(i)}/\eta_{\mathrm{conv}} \\ &\mathrm{sim}(N_{\mathrm{WT}}, N_{\mathrm{PV}}, N_{\mathrm{BAT}}) = 1. \end{aligned}$$

The $\mathrm{sim}(\cdot)$ function performs the system simulation in order to verify that the system configuration fulfils the uninterrupted power supply requirement of the load during the simulation time span. If verified, the output would be 1, otherwise 0.

The optimization procedure was carried out four times. First, it was fed by the real recorded WSTS together with the other inputs. The result is assumed to be the real storage. Then, the simulated WSTS by the proposed copula, ARMA, and Markov models were fed into the program one at a time while other data and parameters were unchanged. Thus, the outcome provides four sets of optimal sizes over a 365-day period which could be used to evaluate the performance of the three models against real data.

It should be noted that all three simulated models produce one-year-long data, the very same as the length of the real recorded data. The main goal is provision of a credible ground for comparison. Once a real investment is intended, the optimal storage size would be calculated using a parameterized time-series model (e.g., one of the three studied models) as data inputs. This potentially provides many years of synthetic wind, PV, and load data. This presumably resemble what a long-term

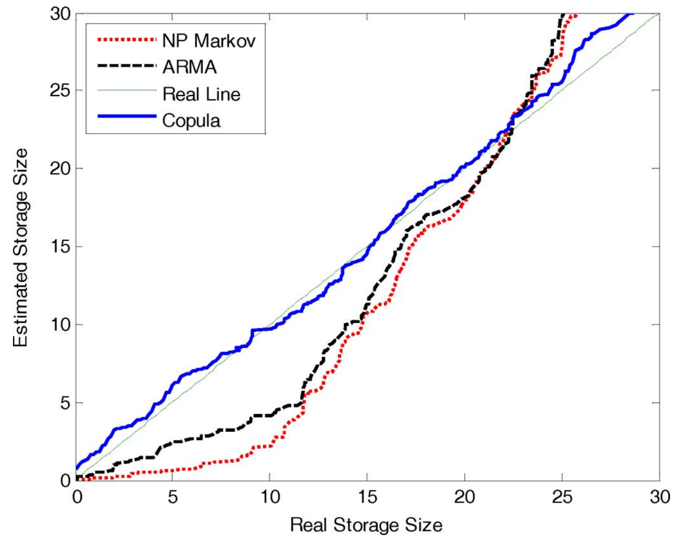


Fig. 12. Real battery sizes compared to the estimated battery sizes using the three models of the WSTS.

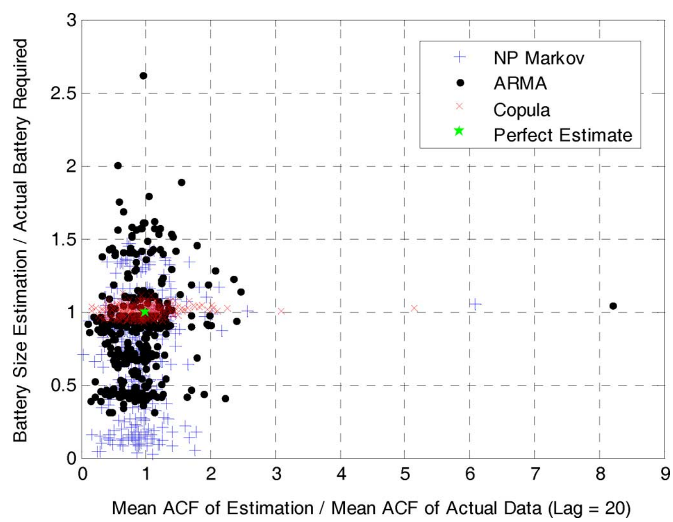
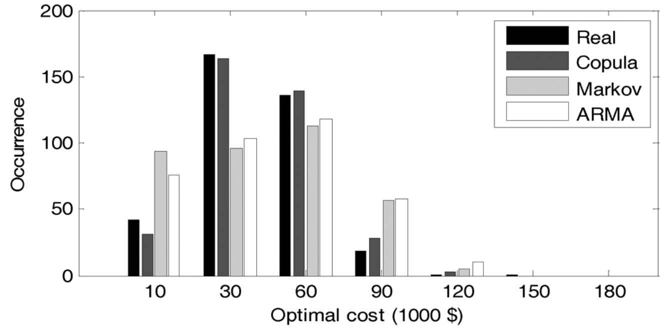


Fig. 13. Estimated versus actual storage plotted against estimated versus actual ACF.

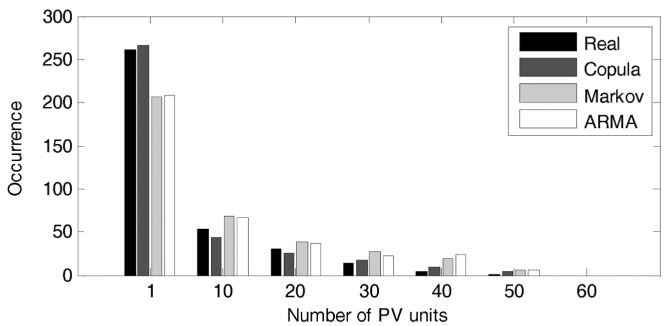
dataset would look like. In fact, using the historical data records with limited length (obtained from the site under study over a year) would be inadequate in the presented long-term planning of combined wind and storage systems in practice and, for example, the proposed copula-based time-series models should be used to produce several years long dataset.

Fig. 11 shows the probability distribution of the calculated optimal battery size. It is seen that the ARMA and Markov models *underestimate* the required battery storage capacity in general. The difference becomes more obvious if the calculated optimal battery size values based on the simulated WSTS are compared to the real optimal storage needs in Fig. 12.

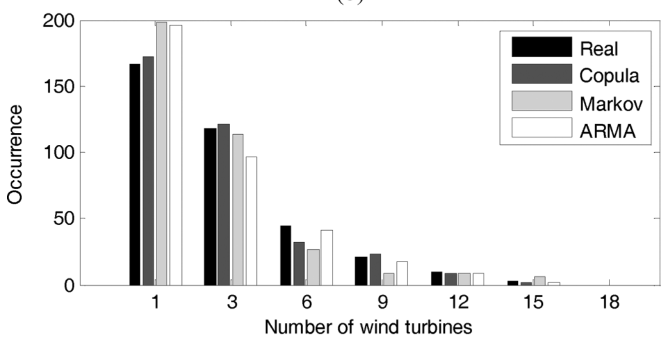
On the other hand, in order to provide a more sensitive index, the estimated storage values are divided by the real storage value, yielding a storage fraction. Such a fraction is calculated for ACFs as well. Fig. 13 is a scatter plot of these fractions for all cases. It is clear that closer and more concentrated cluster points to (1,1) introduce better fitness for the corresponding



(a)



(b)



(c)

Fig. 14. Probability distributions of the occurred optimal sizes of each component along with the optimal costs.

model. Therefore, according to Fig. 13, the copula-Markov is much better than the other two models in terms of both nearer and more accurate estimations. Finally, the optimal sizes of the other components as well as battery units along with corresponding minimal cost occurrences are shown in Fig. 14 by cumulative distribution plots.

V. Conclusion

There exist more than one simple autoregressive temporal dependence structure for a given wind speed data set. This paper proposes a copula-based approach in achieving temporal dependence as well as actual distributions. The proposed modeling approach gives both an accurate and practically simple representation of wind generation uncertainty which is crucial in optimal planning and operation of active networks with high penetrations of wind power. For example, sequential reliability studies, storage design, as well as stochastic programming decision tools rely on accurate time series models generated using available wind data. The uncertainty characterization is such that the generated scenario sets retain the main statistical properties of both the WSTS and WPTS. Several statistical tests and a comprehensively analyzed case study indicate that the proposed model performs better than those of previous literature such as the ARMA and conventional Markov. In particular, the proposed model achieves the following advantages:

- The proposed model replicates nonlinear long-memory temporal dependence of the WSTS as well as the converted WPTS. Autocorrelations are completely preserved. This means that any analysis for power systems having substantial long-term storage which adds memory to the system, could rely on the modeled WSTS more effectively.
- Empirical probability distributions of the WSTS at any time instant could be put directly into the model. This eliminates any preventable approximation errors in comparison with applying well-known distributions. Also, empirical distributions extend the use of nonparametric methods in an effort to identify the importance of accurate characterizations of wind-speed distributions.
- The nonlinear turbine power curve still preserves the exact correlation structure of the WSTS when transformed to the WPTS.
- Energy storage buffers against short- and long-term fluctuations in output from wind generation. Meanwhile, this buffering effect in storage operation should be integrated into the optimal planning of the energy content. Hence, modeling the exact temporal wind power variations would contribute to more accurate estimations of the stored energy content for both short- and long-term delivery periods.

An interesting area of future research could be the choice of an appropriate parametric copula, extension to copula-based semiparametric Markov processes of any finite order, and employing time-varying copulas. These would enable reasoning and computation with the model when the data is to be controlled or streamed in smart grid applications. For example, modeling and controlling electric vehicles and managing data from the advanced metering infrastructure (AMI) would be of interest. Furthermore, the proposed simulation approach could be used within the short-term forecasting (not simulation) framework in parallel with wind power prediction tools to further correct for errors.

References

[1] B. C. Ummels, E. Pelgrum, and W. L. Kling, “Integration of large-scale wind power and use of energy storage in the Netherlands electricity supply,” *IET Renew. Power Gen.*, vol. 2, no. 1, pp. 34–46, 2008.

[2] H. Holttinen, P. Meibom, A. Orths, F. V. Hulle, B. Lange, M. O’Malley, J. Pierik, B. Ummels, J. O. Tande, A. Estanqueiro, M. Matos, E. Gomez, L. Söder, G. Strbac, A. Shakoor, J. Ricardo, J. C. Smith, M. Milligan, and E. Ela, Design and Operation of Power Systems With Large Amounts of Wind Power, Final report, IEA WIND Task 25, Phase one 2006–2008, Jun. 2009, pp. 1–199.

[3] J. Mohammadi, A. Rahimi-Kian, and M.-S. Ghazizadeh, “Aggregated wind power and flexible load offering strategy,” *IET Renew. Power Gen.*, vol. 5, no. 6, pp. 439–447, 2011.

[4] Y. M. Atwa and E. F. El-Saadany, “Optimal allocation of ESS in distribution systems with a high penetration of wind energy,” *IEEE Trans. Power Syst.*, vol. 25, no. 4, pp. 1815–1822, Nov. 2010.

[5] D. C. Hill, D. McMillan, K. R. W. Bell, and D. Infield, “Application of auto-regressive models to U.K. wind speed data for power system impact studies,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 134–141, Jan. 2012.

[6] J. M. Morales, R. Minguez, and A. J. Conejo, “A methodology to generate statistically dependent wind speed scenarios,” *Appl. Energy*, vol. 87, pp. 843–855, 2010.

[7] R. Billinton, H. Chen, and R. Ghajar, “Time-series models for reliability evaluation of power systems including wind energy,” *Microelectron. Reliab.*, vol. 36, no. 9, pp. 1253–1261, 1996.

[8] P. Chen, T. Pedersen, B. Bak-Jensen, and Z. Chen, “ARIMA-based time series model of stochastic wind power generation,” *IEEE Trans. Power Syst.*, vol. 25, no. 2, pp. 667–685, May 2010.

[9] A. Shamshad, M. A. Bawadi, W. M. A. Wan Hussin, T. A. Majid, and S. A. M. Sanusi, “First and second order Markov chain models for synthetic generation of wind speed time series,” *Energy*, vol. 30, pp. 693–7081, 2005.

[10] F. C. Sayas and R. N. Allan, “Generation availability assessment of wind farms,” *Proc. Inst. Elect. Eng., Gen., Transm., Distrib.*, vol. 143, no. 5, pp. 507–518, 1996.

[11] N. B. Negra, O. Holmstrøm, B. Bak-Jensen, and P. Sørensen, “Model of a synthetic wind speed time series generator,” *Wind Energy*, vol. 11, pp. 193–209, 2008.

[12] G. Papaefthymiou and B. Klöckl, “MCMC for wind power simulation,” *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 234–240, Mar. 2008.

[13] R. Billinton and D. Huang, “Incorporating wind power in generating capacity reliability evaluation using different models,” *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 2509–2517, Nov. 2011.

[14] H. V. Haghi, M. T. Bina, M. A. Golkar, and S. M. M. Tafreshi, “Using copulas for analysis of large datasets in renewable distributed generation: PV and wind power integration in Iran,” *Renew. Energy*, vol. 35, pp. 1991–2000, 2009.

[15] D. S. Callaway, “Sequential reliability forecasting for wind energy: Temperature dependence and probability distributions,” *IEEE Trans. Energy Convers.*, vol. 25, no. 2, pp. 577–585, Jun. 2010.

[16] G. Papaefthymiou and D. Kurowicka, “Using copulas for modeling stochastic dependence in power system uncertainty analysis,” *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 40–49, Feb. 2009.

[17] J. Haslett and A. E. Raftery, “Space-time modelling with long-memory dependence: Assessing Ireland’s wind power resource,” *J. Royal Statist. Society. Series C, Appl. Stat.*, vol. 38, no. 1, pp. 1–50, 1989.

[18] J. Beran, “Statistical methods for data with long-range dependence (with discussion),” *Statist. Sci.*, vol. 7, no. 4, pp. 404–416, 425–427, 1992.

[19] J. Beran, , E. J. Wegman, Y. H. Said, and D. W. Scott, Eds., “Long-range dependence,” in *Wiley Interdisciplinary Reviews: Computational Statistics*. New York, NY, USA: Wiley, 2010, vol. 2, pp. 26–35.

[20] H. Künsch, J. Beran, and F. Hampel, “Contrasts under long-range correlations,” *The Annals of Statistics*, vol. 21, pp. 943–964, 1993.

[21] K. Brokish and J. Kirtley, “Pitfalls of modeling wind power using Markov chains,” in *Proc. Power Syst. Conf. Expo.*, Seattle, WA, USA, 2009, pp. 1–6.

[22] N. H. Bingham and R. Schmidt, , Y. Kabanov, R. Liptser, and J. Stoyanov, Eds., “Interplay between distributional and temporal dependence,” in *From Stochastic Calculus to Mathematical Finance*. New York, NY, USA: Springer, 2006, pp. 69–90.

[23] R. B. Nelsen, *An Introduction to Copulas*. New York, NY, USA: Springer, 2006.

[24] H. V. Haghi and M. T. Bina, “A study on probabilistic evaluation of harmonic levels produced by static compensators,” in *Proc. Australasian Universities Power Eng. Conf.*, Sydney, Australia, Dec. 2008, pp. 1–6.

[25] H. V. Haghi and M. T. Bina, “Complete harmonic-domain modeling and performance evaluation of an optimal-PWM-modulated STATCOM in a realistic distribution network,” *Przeglad Elektrotech.*, vol. 85, no. 1, pp. 156–161, 2009.

[26] B. Stephen, S. J. Galloway, D. McMillan, D. C. Hill, and D. G. Infield, “A copula model of wind turbine performance,” *IEEE Trans. Power Syst.*, vol. 26, no. 2, pp. 965–966, May 2011.

[27] S. Gill, B. Stephen, and S. Galloway, “Wind turbine condition assessment through power curve copula modeling,” *IEEE Trans. Sustain. Energy*, vol. 3, no. 1, pp. 94–101, Jan. 2012.

[28] M. A. Golkar and H. V. Haghi, “Using a multivariate DOE method for congestion study in distribution systems under impacts of plug-in electric vehicles,” in *Proc. 21st Int. Conf. Electricity Distribution (CIRED)*, Frankfurt, Germany, Jun. 2011, pp. 1–4.

[29] S. Hagspiela, A. Papaemannouilb, M. Schmidb, and G. Andersson, “Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid,” *Appl. Energy*, pp. 1–12, 2011, to be published.

[30] G. Westner and R. Madlener, “Investment in new power generation under uncertainty: Benefits of CHP vs. condensing plants in a copula-based analysis,” *Energy Econ.*, vol. 34, pp. 31–44, 2012.

[31] D. Kurowicka and R. M. Cooke, *Uncertainty Analysis With High Dimensional Dependence Modeling*, ser. Probability and Statistics Series. Hoboken, NJ, USA: Wiley, 2006.

[32] [Online]. Available: http://www.vosesoftware.comaccessed Mar. 2012

[33] N. Channouf and P. L’Ecuyer, “Fitting a normal copula for a multivariate distribution with both discrete and continuous marginals,” in *Proc. 2009 Winter Simulation Conf.*, pp. 352–358.

[34] L. Fingersh, M. Hand, and A. Laxson, Wind Turbine Design Cost and Scaling Model. NREL Washington, DC, NREL Tech. Rep. 500-405662006, 2006 [Online]. Available: www.nrel.gov/wind/docs/weibull_betz5_lswt_baseline.xls

[35] W. Cleveland, *Visualizing Data*. AT&T Bell Laboratories, Murray Hill, NJ, USA: Hobart Press, 1993.

[36] N. H. Bingham and R. Schmidt, , Y. Kabanov, R. Liptser, and J. Stoyanov, Eds., “Interplay between distributional and temporal dependence,” in *From Stochastic Calculus to Mathematical Finance*. New York, NY, USA: Springer, 2006, pp. 69–90.

[37] J. Beran, *Statistics for Long Memory Processes*. London, U.K.: Taylor & Francis, 1994.

[38] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ, USA: Pearson Prentice-Hall, 2007.

[39] C. Abbey and G. Joós, “A stochastic optimization approach to rating of energy storage systems in wind-diesel isolated grids,” *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 418–426, Feb. 2009.

[40] R. S. Garcia and D. Weisser, “A wind-diesel system with hydrogen storage: Joint optimisation of design and dispatch,” *Renew. Energy*, vol. 31, no. 14, pp. 2296–2320, 2006.

[41] E. Koutroulis, D. Kolokotsa, A. Potrirakis, and K. Kalaitzakis, “Methodology for optimal sizing of stand-alone photovoltaic/wind generator systems using genetic algorithms,” *Sol. Energy*, vol. 80, no. 9, pp. 1072–1088, 2006.

Hamed Valizadeh Haghi (M’08) was born in Ardebil, Iran, in 1984. He received the B.Sc. degree (with distinction) from the University of Tabriz, Tabriz, Iran, in 2006, and the M.Sc. degree (with distinction) from K. N. Toosi University of Technology (KNTU), Tehran, Iran, in 2008, all in electrical engineering.

Currently, he is a researcher at KNTU with the Faculty of Electrical and Computer Engineering. Since November 2008, he has been a consultant and researcher, focusing on smart grid. His primary research interests include applications of statistical methods in power system analysis, active distribution networks, uncertainty modeling, stochastic optimization, model-driven software development, and distributed generation.

Mr. Haghi is an honorary member of the National Elites Foundation, and a member of the IEEE PES.

M. Tavakoli Bina (SM’07) received the B.Sc. and M.Sc. degrees from the University of Tehran and Ferdowsi, in 1988 and 1991, respectively, and the Ph.D. degree from the University of Surrey in the U.K., Guildford, in 2001, all in power electronics and power system utility applications.

From March 1992 to November 1997, he was with the K. N. Toosi University of Technology in Tehran, as a lecturer working on power systems. In July 2000, he received his Ph.D. from the Department of Electronics and Computer Engineering, University of Surrey, U.K. Since September 2001, he has been with the Faculty of Electrical and Computer Engineering, K. N. Toosi University of Technology (KNTU), Tehran, Iran. Currently, he is a full professor at the KNTU. His main research interests include power converters, modulation techniques, control and modeling of FACTS controllers, distribution systems, and power system control.

Masoud Aliakbar Golkar was born in Tehran, Iran, in 1954. He received his B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 1977, the M.Sc. degree from the Oklahoma State University, USA, in 1979, and the Ph.D. degree from the Imperial College of Science, Technology, and Medicine, The University of London, U.K., in 1986, all in electrical engineering (power systems).

Since 1979 he has been teaching and doing research at K. N. Toosi University of Technology, Tehran, Iran. He is the advisor of many electricity boards and has successfully conducted many projects for different electricity utilities in Iran. He conducted some research groups on Electrical Distribution Systems and Reactive Power studies at the Electric Power Research Center (EPRC) for more than 10 years. From Jan. 2002 to July 2005, he has served as a Senior Lecturer at Curtin University of Technology in Malaysia. His main Research areas are smart grid, distributed generation, and renewable generations studies, electric distribution systems and reactive power studies, voltage collapse studies, and load and energy management. He is the author of some books and has more than 160 papers in national and international journals and conferences. Currently, he is a professor at K. N. Toosi University of Technology in Tehran, Iran.