Renewable Energy 35 (2010) 1991–2000

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Using Copulas for analysis of large datasets in renewable distributed generation: PV and wind power integration in Iran

H. Valizadeh Haghi*, M. Tavakoli Bina, M.A. Golkar, S.M. Moghaddas-Tafreshi

Faculty of Electrical and Computer Engineering, K. N. Toosi University of Technology, Seyed Khandan, P. O. Box 16315-1355, Tehran-16314, Iran

ARTICLE INFO

Article history:
Received 22 August 2009
Accepted 30 January 2010
Available online 25 February 2010

Keywords:
Copula
Correlation
Dependence
Distributed generation
Monte Carlo simulation
Photovoltaics (PV)
Wind energy

ABSTRACT

Renewable distributed generation introduced as an environmental friendly alternative energy supply while it provided the power system with ever-growing technical benefits such as loss reduction and feeder voltage improvement. The evaluation of the effects of small residential photovoltaic and wind DG systems on various system operating indices and the system net load is complicated by both the probabilistic nature of their output and the variety of their spatial allocations. The increasing penetration of renewable distributed generation in power systems necessitates the modeling of this stochastic structure in operation and planning studies. An advanced stochastic modeling of the system requires multivariate uncertainty analysis involving non-normal correlated random variables. Such an analysis is to epitomize the aggregate uncertainty corresponding to spatially spread stochastic variables. In this paper, an integration study of photovoltaics and wind turbines, distributed in a distribution network, is investigated based on the stochastic modeling using Archimedean copulas as a new efficient tool. The basic theory concerning the use of copulas for dependence modeling is presented and focus is given on an Archimedean algorithm. A comprehensive case study for Davarzan area in Iran is presented after reviewing Iran's renewable energy status. This study shows an application of the presented technique when large datasets, assuming 10-min interval between data points of PV, wind and load profiles, are involved where a deterministic study is not trivial.

1. Introduction

Distributed generation (DG) devices interact with the operation, protection and control of the distribution feeder at which they are installed. The produced electrical power via a variety of these generation units is stochastic by a non-dispatchable primary energy source. Therefore, DG systems inherently provide some benefits and produce some potentially unwanted effects. They may improve the load curve and the voltage profile across the feeder, may reduce the loading level of branches and substation transformers, and provide environmental benefits by offsetting the pollutant emissions. Utility economic benefits also include loss reduction, avoided costs of energy production, generation capacity, distribution and transmission capacity investment deferral, reducing risk from uncertain fuel prices, green pricing benefits, etc [1]. However, there are some issues remained to be solved such as the high capital costs of renewable energy technologies and poor reliability in stand-alone remote supply systems [2]. Nonetheless, it is strongly expected that the renewable DG systems will play an important role in future power systems.

Although the practical capacity of these systems is smaller than the conventional generation units, their integration may significantly alter the behavior of the system across which they are installed. Deterministic modeling of such a system with stochastic non-dispatchable DG units (e.g. wind or photovoltaics) is not trivial, due to the following reasons.

- The aggregate wind or PV power outputs are stochastic in a time-independent manner;
- the daily load profiles of a distribution network are stochastic in a time-dependent manner;
- the system configuration and device types have uncertainty at the planning stage;
- there are always a large datasets to consider that may cause an extreme computational burden; and
- a high dependence exists between the system and the renewable PV or wind DG units which is deterministically uncharacteristic resulting from an aggregate uncertainty.

Therefore, the use of stochastic methods (e.g. statistical data analysis and Monte Carlo simulation) is necessarily unavoidable in addition to the basic deterministic methods.

* Corresponding author. Tel.: +98(21)88462174 7; fax: +98(21)88462066.
E-mail address: valizadeh@ee.kntu.ac.ir (H. Valizadeh Haghi).

0960-1481/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.renene.2010.01.031

Just as consumer demands are smoothed by aggregation, so is the output from wind or PV plant, and geographic dispersion dramatically reduces the wind speed or solar radiation fluctuations. However, the DG-enhanced distribution system planning (e.g. for the calculation of the system net load distribution, which is a main concern of this paper) should completely take into account the dependence structure between the relevant determinants as follows.

(1) Wind speed in different locations;
(2) converted wind power in the wind turbine output if not available;
(3) PV output power; and
(4) system load curve data in different locations.

Not taking these dependence structures properly into account by assuming independence or by simplified modeling (e.g. using Normal distributions) would contribute to different results. This paper obtains the system net load distributions which can be used for calculating the capacity credit added to the generation system due to the integration of wind and PV distributed power. In addition, using this method coupled with the system design programs (such as probabilistic load flow algorithms) reveals the necessary system reinforcements and policy changes due to the incorporation of renewable distributed PV and wind powers. The dependence can be modeled separately from the marginal distributions, linking them using a copula function. Using copulas for modeling purposes includes two straightforward steps: first, the marginal distributions along with their correlation matrix should be modeled; and second, a proper copula should be selected and fitted to the data. It should be mentioned that it is an obscure task to find a multivariate distribution and fit it to the data. Besides, the use of copulas is practical as some good software packages have already provided its complete implementation (such as [3,4]).

This paper proposes an Archimedean copula algorithm for case studies. Such a copula approach was successfully applied to the simulation of the three-phase voltage unbalance in distribution feeders [5,6]. The stochasticity of dependent chaotic variables and time series of power systems can be efficiently modeled using copulas [6,7]. Besides, an appropriate choice should be made between the Archimedean and Elliptical copula families considering their specific applications [9]. As well, using Gaussian copulas impose some simplifications to the model that seems to be acceptable in some cases where either a threshold analysis is required [6] or a lower precision is chosen over a more detailed algorithm [8]. Besides, Gaussian copulas do not model the tail dependence which is discussed in the following section. On the other hand, Archimedean copulas can efficiently used to produce non-conventional multivariate distributions for Monte Carlo studies as presented in Sections 2 and 3.

In the following text, the main theorems and procedures for using copulas are presented focusing on an Archimedean approach. Afterwards, the basic discussions regarding the integration of distributed wind and solar systems in a distribution network are presented. Then, an updated review of Iran's renewable energy resources and policies is presented focusing on wind and PV systems. The discussions are followed by introducing the Davarzan area and the presented case study. The recorded wind speed, solar radiation, and load profiles of a radial network are used to implement the proposed copula analysis.

2. Principles of copulas and dependence

2.1. Basic definitions

According to [9], copulas are "functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions" or equivalently in terms of mathematical representation [10], a copula is a function (C) of n variables on the unit n-cube $[0,1]^n$ with the following properties.

(1) The range of C is the unit interval [0,1];
(2) $C(\mathbf{u})$ is zero for all $u = (u_1, \ldots, u_n)$ in $[0,1]^n$ for which at least one coordinate equals zero;
(3) $C(u) = u_k$ if all coordinate of $\mathbf{u}$ are 1 except the k-th one;
(4) C is n-increasing in the sense that for every $\mathbf{a} \leq \mathbf{b}$ in $[0,1]^n$ the measure ΔC_a^b assigned by C to the n-box $[a,b] = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is non-negative, i.e.

$$\Delta C_a^b := \sum_{(\epsilon_1,\ldots,\epsilon_n)\in\{0,1\}^n} (-1)^{\sum_{i=1}^{n}\epsilon_i} C(\epsilon_1 a_1 + (1-\epsilon_1)b_1, \ldots, \epsilon_n a_n + (1-\epsilon_n)b_n) \geq 0 \tag{1}$$

where, n is the number of dependent outcomes that should be modeled and all marginal distributions of the random vector $u = (u_1, \ldots, u_n)$ are uniform. It can be illustrated from the definition that copulas have many useful properties, such as uniform continuity and existence of all partial derivatives. Samples of copulas from different families are shown in Fig. 1 in bivariate form.

To complete the construction of copula, a set of arbitrary marginal distribution functions can be assumed and therefore, the C defines a multivariate distribution function evaluated at x_1, $x_2, \ldots, x_n$ as:

$$C[F_1(x_1), F_2(x_2), \ldots, F_n(x_n)] = F(x_1, x_2, \ldots, x_n). \tag{2}$$

Sklar [11] showed that any multivariate distribution function F can be written in the form of (2) that is copula representation. He also showed that if the marginal distributions are continuous, there is a unique copula representation. The aforementioned statements are the key theorem of copulas referred to as *Sklar's theorem* and clarify the relations of dependence and the copula of a distribution. It should be mentioned that constructing *multivariate distributions without the concept of copula* has some drawbacks such as:

- Different families are necessarily needed for different marginal distributions.
- Extension of multivariate distributions to thrivariate cases and above is unclear.
- Measures of dependence often appear in the marginal distributions.

2.2. Correlation measures

To continue with the concepts of applying copula fits to the simulation of dependency, a brief reminding of *correlation* and its measures seems to be necessary. The familiar form of correlation is the Pearson's pairwise linear coefficient and defined as

$$\rho(X,Y) = \frac{\mathrm{cov}[X,Y]}{\sqrt{\sigma^2[X]\sigma^2[Y]}}. \tag{3}$$

This correlation has some shortcomings to measure the dependence except the linear case; first, it is non-ideal for a dependence measure of heavy-tailed marginal distributions. Second, assume correlation of a jointly *non-normal* distribution with a non-linear relationship is zero if relationship (3) is applied. Empirically, it is shown that there are certain data that introduce nonzero correlation in the contrary to Eq. (3); and third, when a non-linear scale transformation is applied to the

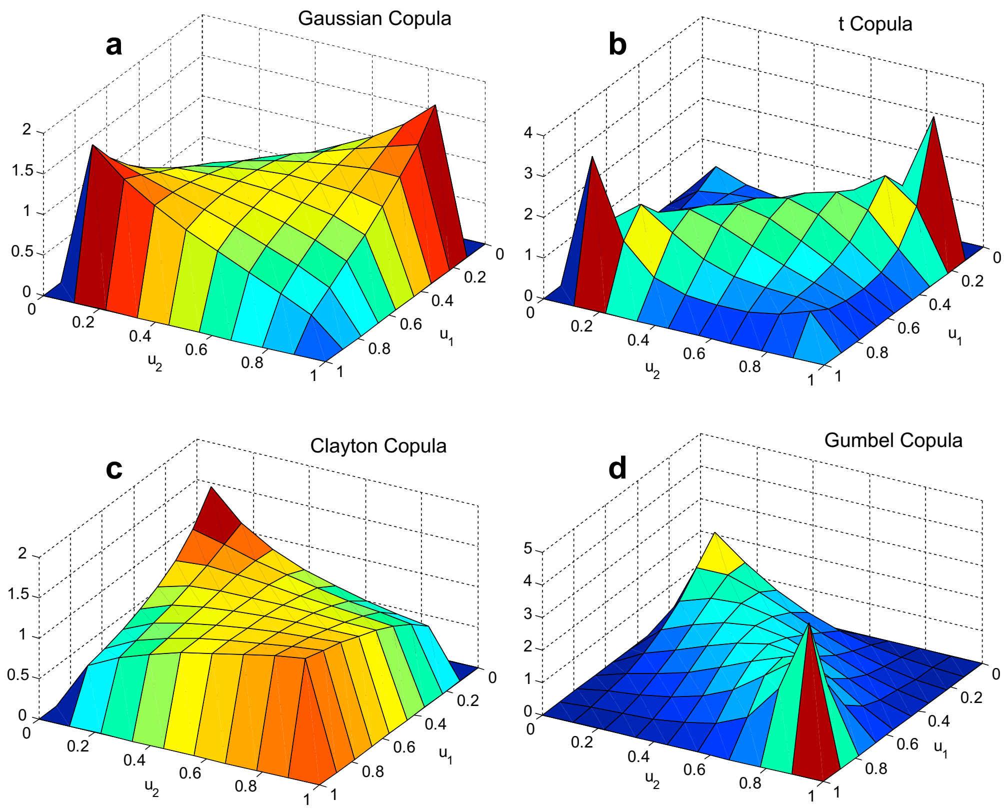

Fig. 1. Samples of different bivariate copula functions: (a) Gayssian copula; (b) t copula; (c) Clayton copula; and (d) Gumbel copula.

multivariate distributions, the calculated correlations before and after the transformation using Eq. (3) results in different values. Therefore, a rank correlation coefficient, such as Kendall's τ or Spearman's ρ, is more appropriate to fulfill the desirable characteristics of a measure of dependence. Spearman's rank correlation has been used in the succeeding analysis which is given by

$$\rho_S(X,Y) = \rho(F_X(X), F_Y(Y)) \tag{4}$$

where, $F_X(X)$ and $F_Y(Y)$ are the distribution functions of the random variables X and Y respectively. It should be mentioned that for the jointly normal distribution, Spearman's rank correlation is almost identical to the linear correlation; however, this is not true when a certain transformation is applied to the available data.

2.3. Modeling of stochastic dependence

There are various situations in the applications of power system analysis where we might wish to simulate dependent random vectors and configurations (as evidenced by Monte Carlo algorithms). Examples of such applications are in the noise modeling, reliability studies, materials and natural phenomena uncertainty analysis, risk assessment, complex modeling, etc. Randomly behaving variables of such circumstances may be assumed completely dependent, linearly correlated, superposed, or completely independent; the most appropriate choice is influenced by several factors such as the characteristics of the system and the required accuracy. In the power system problems, anyhow, many cases involve high levels of dependency. Therefore, it is very tempting to approach the problem in the following way:

(1) Estimate matrix of pairwise rank correlations,
(2) estimate marginal distributions,
(3) combine this information using a copula.

To perform a simulation, therefore, the following information should be specified from the measured or calculated data:

(1) the copula family and any required shape parameters,
(2) the rank correlations among variables, and
(3) the marginal distributions for each variable.

It should be mentioned that the copula is assumed to be chosen by the designer based on their experiences. The most commonly used copulas are the Gaussian copula for linear correlation, Gumbel copula for extreme distributions, and the Archimedean copula and the t-copula for dependence in tail [9,12].

Furthermore, a realistic correlation matrix must be positive semi-definite, real-valued, and symmetric. Proper modeling of such a matrix based on the realistic data is an important stage of the algorithm; specifically when there is some data misplacement or the values are noisy, unavailable, or unreliable. There are practical methods to deal with the problems associated with the non-proper correlation matrices [13]. However, these methods generally apply to abnormal data recordings and become redundant in Gaussian copula construction algorithms.

Another key point in a reliable dependency modeling is building the marginal distributions. One could fit a parametric model separately to each dataset, and use those estimates as the marginal distributions; however, a parametric model may not be sufficiently flexible. Instead, a nonparametric model to transform to the marginal distributions seems to be appropriate. Meanwhile, using empirical cumulative distributions results in a discrete representation which may not desirable for a continuous distribution. Therefore, it is advisable to apply a smoothing technique such as kernel smoothing or interpolate between the midpoints of the steps with a piecewise linear function.

2.4. Simulation algorithm

For the simulation, it is a good idea to experiment with different copulas and correlations. The two main simulation strategies are the Archimedean and compounding methods [14]. Both methods can be easily implemented for more than two dimensions (multivariate case). Nonetheless, the compounding algorithms are computationally more straightforward than the conditional distribution approach used in Archimedean methods. Meanwhile, it requires the generation of an additional variable which can be computationally expensive in applications. Because the power system problems typically require extensive calculations, addition of extra variates may be unacceptable. Therefore, as popularly used in most software [3], the Archimedean construction is used in this paper. One method is briefly as follows [14]:

(1) Generate independent uniform random numbers $U_1, U_2,\ldots,U_n$.
(2) Set $X_1 = F_1^{-1}(U_1)$ and $c_0 = 0$.
(3) For $k = 2,\ldots,n$, recursively calculate X_k by

$$U_k = F_k(X_k|x_1,\ldots,x_{k-1}) = \frac{\Phi^{-1(k-1)}\{c_{k-1}+\Phi[F_k(x_k)]\}}{\Phi^{-1(k-1)}(c_{k-1})}. \quad (5)$$

Where,

$$c_k = \Phi[F_1(x_1)]+\cdots+\Phi[F_k(x_k)]\text{.}$$

This algorithm is to generate $X_1, X_2,\ldots,X_n$ having modeled distribution function of (1), the copula is

$$C(u_1,u_2,\ldots,u_n) = \Phi^{-1}[\Phi(u_1)+\cdots+,\Phi(u_n)], \quad (6)$$

Equation (6) defines a class of copulas known as Archimedean. The Archimedean representation allows us to reduce the study of a multivariate copula to a single univariate function. The function Φ is a generator of the copula and uniquely determines it [15].

Given a dataset, choosing a copula to fit the data is an important but difficult problem [16]. Since the real data generation mechanism is unknown, it is possible that several candidate copulas fit the data reasonably well or that none of the candidate fits the data well. When maximum likelihood method is used, the general practice is to fit the data with all the candidate copulas and choose the ones with the highest likelihood [17]. A graphical tool to choose among Archimedean copulas is based on the Kendall's process [18]. Indeed, copula selection is an ongoing research area.

Considering the maximum likelihood, the Frank copula is chosen in this paper because it fits the studied data well. It should be mentioned that Frank's family permits negative as well as positive dependence. Nonetheless, other types of Archimedean copulas permit only non-negative correlations because of the limited dependence parameter space. This is the other reason to use Frank copula in the following section.

The technique for random vectors can be applied for *time series* as well [19]. A moving window with a certain number of vectors is taken as a sample vector for a stationary time series. The marginal distributions and the copula are then estimated with this sample according to the above algorithm.

3. Wind and PV renewable dispersed generation focusing on Iran

3.1. PV/wind systems in distribution networks

As mentioned in the introduction, renewable energy resources essentially have unpredictable stochastic behaviors. However, some of them like solar radiation and wind speed, have complementary profiles. Stand-alone dispersed hybrid systems usually take advantage of this particular characteristic combining PV panels and wind turbines, in conjunction with a diesel-powered backup generator or battery storage based on economical considerations [20]. Since storage cost still represents the major economic restraint, usually PV/wind systems are appropriately sized to minimize its requirements. Also, wind power is lower in cost than PV power approximately by a factor of five, so it often gets the main role in generation [21].

In radial networks, distributed generators on the feeder effectively decrease the load active power demand, change the power flow along the feeder and improve local voltage conditions. The layout of a distribution feeder in Davarzan area in north-east of Iran with one possible allocation of distributed generators, is shown in Fig. 2. This area and the recorded data are used in the following section's case study. Fig. 3 shows how the effective load is decreased in the presence of a PV generator, when the peaks of PV generation and load demand are matched. The capacity of a PV generator in this example was 20% of the nominal load [1].

3.2. Iran's wind/PV power resources

The renewable energy technologies relevant to the Iranian context are mainly those for grid-connected power generation. They include onshore wind farms, solar photovoltaics, small and medium hydropower, geothermal power, concentrating solar power, and landfill gas. We review here the wind and solar energy potentials.

Wind resources in Iran are plentiful. Based on a mesoscale wind map [22], it has been estimated that more than 10,000 MW of wind power can be installed in Iran. According to Iran Renewable Energy Organization (SUNA), wind speed in Khaf (Khorasan province) is exceptional; in Manjil (Gilan province) is excellent; and Zabol-Loutak (Sistan va Balochestan province) is good; whereas the remaining sites may be more questionable. Even so, there may possibly be economic sites on hill ridges within the more questionable area, such as Namin (Ardabil province) and Davarzan (Khorasan province), and in any case, a proper economic evaluation requires a more thorough analysis of the costs of alternative power generation. In addition, the cost of grid integration, match with the demand, and site access would have to be evaluated to get a more complete picture.

Currently, Iran has a number of wind farms using modern wind turbines, primarily near Manjil (100 MW) initially using 300 kW and 550 kW Nordtank stall controlled turbines from the mid 1990s, and more recently Vestas V47, 660 kW pitch-controlled turbines. The sites in use include Manjil, Roodbar, Herzeville and Siahpoosh. Another wind farm of 28.3 MW is operating at Binalood in Khorasan province.

Current wind farms are owned and operated by the government-owned electrical utility, but power purchasing agreements for about 600 MW are currently in progress with private developers. However, it seems that present wind energy tariffs are insufficient to drive a commercial market for wind energy [23].

On the other hand, the solar energy is outstanding. Solar radiation intensity is the highest in the arid regions of Iran, approximately below 30 degrees latitude [24]. Vast expanse of dessert

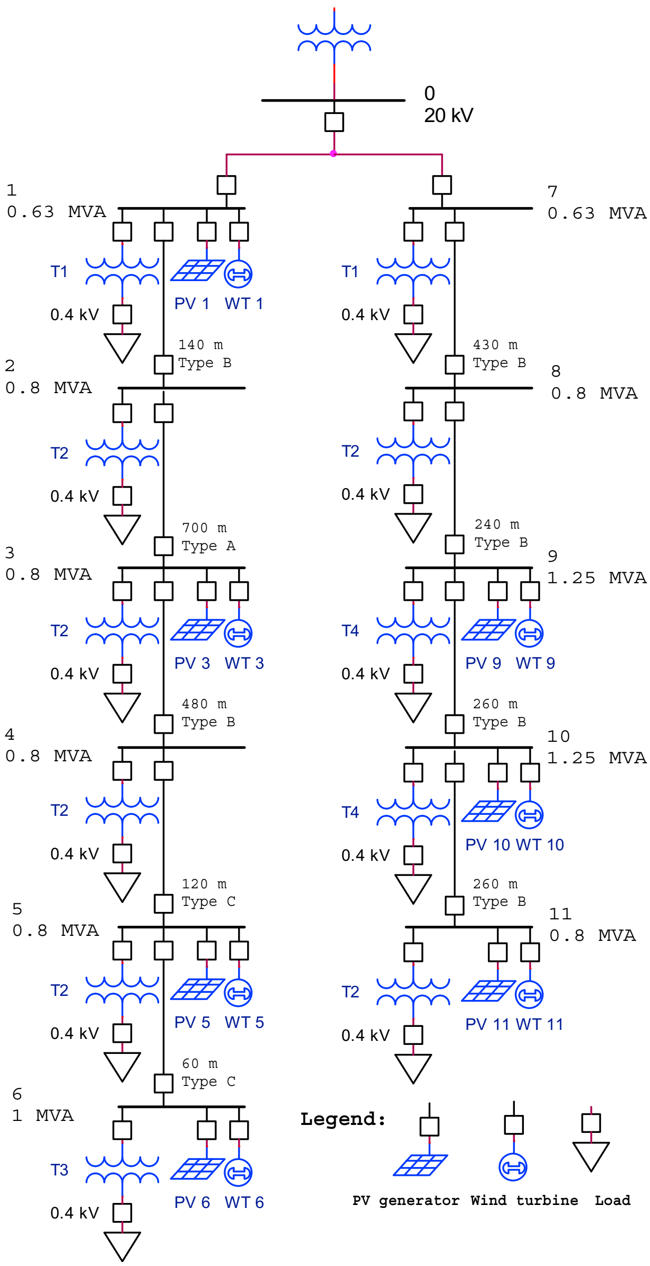

Fig. 2. The distribution network under test.

areas in these regions could provide ample room for concentrating solar power (CSP) installations. In Shiraz and Yazd, for example, where solar thermal power projects are being carried out or planned, the direct normal insolation (DNI), which is the portion of global solar radiation usable by CSP, were measured to be 6.8 and 5.6 kWh/m^2/day, respectively. The DNI of 5 kWh/m^2/day is generally considered to be the threshold for CSP applications.

Unlike CSP, PV can work well almost anywhere in Iran, as it uses all components of global radiation. To date, a total of 500 kW PV installations have been made in Iran which is mainly include stand-alone and pilot applications. In recent years, emphasis of the PV program has become increasingly focused on offgrid rural electrification. The only grid-connected application installed so far is a 30 kW system in Taleghan province which is for operational experience only. It has been estimated by SUNA that the cost of PV installations in Iran is about 10,000 $/kW which is somewhat higher than internationally quoted figures of about 6000–8000

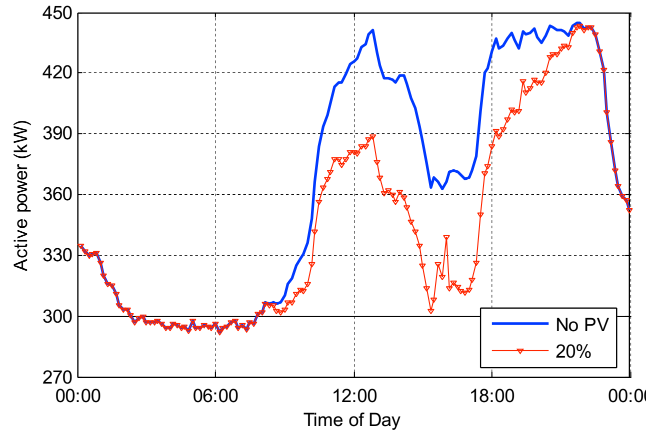

Fig. 3. The effects of dispersed PV on a load profile of the case study.

$/kW for silicon wafer based PV. Accordingly, due to the high costs of grid-connected PV compared to other renewable energy alternatives, SUNA has focused on the use of PV for offgrid distributed electrification. For dispersed users requiring relatively small amounts of power, PV is already a competitive or least cost technology. It is rugged, modular and requires little maintenance and thus highly suitable for remote rural users. For relatively larger rural productive loads, PV systems can be hybridized with wind or provided with storage batteries.

4. Case study: Davarzan area in Iran

Davarzan area is situated in Khorasan, North-east of Iran on a terrain suitable for wind activity (Fig. 4). This area reveals a good wind and solar capabilities and due to its geographical and socio-logical conditions has potentially the benefits of distributed wind/PV systems. As a numerical example, the distribution feeder of Fig. 2 is considered, with a few modifications in the placement of loads. The network data is presented in Table 1. Fig. 2 also shows the locations of loads and one possible distribution of PV and wind generators. The daily load profile is obtained from the actual utility data for a Central-Iran small city. The patterns of the loads at the feeder are different following the real recorded diversity of consumers. Fig. 5 illustrates five different typical load profiles (TLPs) corresponding to five groups of consumers. It is better to mention that, calculating these typical load profiles for a group of consumers is based on field measurements of the individual consumer's load curves. Typically, there are two types of methods for deriving TLPs: one is based on load-survey systems [25], and second, is according to predefined consumers' groups or groups identified during the process of TLPs determination [26]. The second group is obtained by identifying TLPs depending on the shape of the load curves using various pattern-recognition methods [26]. For the first TLP-determination approach, measurements need to be performed over a long time period. For the second approach, typical customer groups represented by the TLPs need to be formed. Here, we used the second method, based on [27].

The measured utility data spans one whole year in 10-min intervals, which is also the period and the resolution considered in this study for wind speed and solar radiation. The availability of higher resolution data would allow more precise detection of the stochastic dependence, as more information describing the interaction between the PV/wind output and load will be available.

The application of the Archimedean copula algorithm, as mentioned in Section 2, requires the computation of the following statistics:

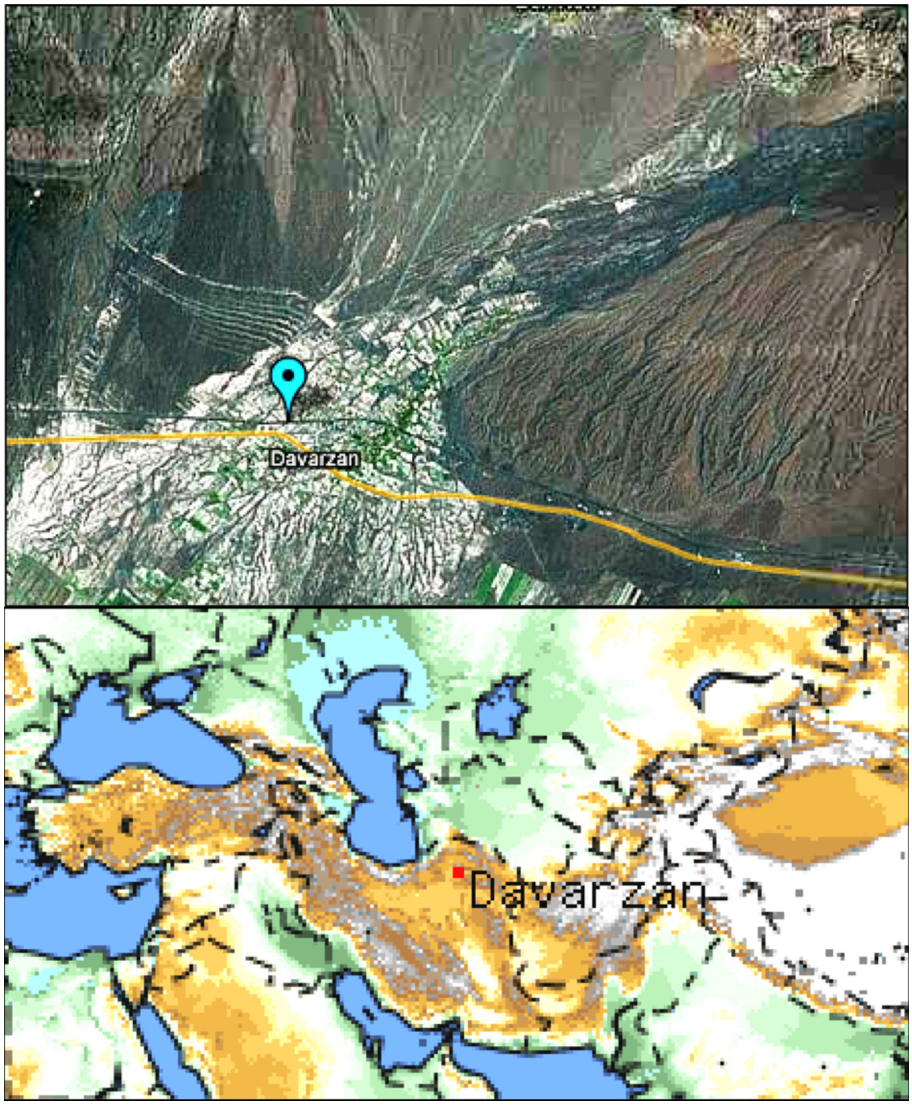

Fig. 4. Davarzan area with the wind and solar recorded data for the presented case study test system.

- Marginal distributions: the wind speed/power distributions at each generation site, the system load distribution and the PV power distributions at each generation site.
- Dependence structure: the rank correlation matrix between the wind speed random variables, the system load and the solar radiation.

4.1. Marginal distributions

4.1.1. Wind speed/power distributions

The wind speed distributions for the sites where data are available are obtained considering the measurement height of 40 m. Typical pitch-controlled wind turbine generators (WTG) are considered for this project, with a hub height of 40 m, a nominal power of 600 kW and cut-in, nominal and cut-out wind speed values of 3, 13 and 25 [m/s]. In Fig. 6, the wind speed and wind power distributions for such a wind turbine generator are presented as obtained by a 10,000-sample Monte Carlo simulation similar to the method described in [28]. On the top graph, the wind speed distribution for a typical site is presented (discontinuous line), together with the wind speed/power wind turbine generator characteristic (continuous line). An accumulation of probability masses at the zero and nominal wind power is observed. The zero values correspond to wind speeds lower than the cut-in and higher than the cut-out wind speed values. The nominal power output values correspond to wind speeds between the nominal and the cut-out values.

The sampling of the wind speed distributions has been performed based on the empirical distribution obtained by data, according to the methodology presented in Section 2. In Fig. 7, the measured wind speed distributions are compared to the simulated ones obtained from a 10,000-sample Monte Carlo simulation. The simulated distributions yield very accurate approximations of the measured ones.

4.1.2. Solar radiation and system load distribution

From the viewpoint of marginal distributions, the pdf and cdf for the system load are presented in Fig. 8 compared to the simulated ones. Also, the match between possible PV modules output and the actual system load is illustrated by Fig 9.

4.2. Dependence structure

For the application of the method (see Section 2), the 11×11 rank correlation matrix is calculated and presented in Table 2.

Table 1
Network data of Fig. 2.

Branch No.	Sending node	Receiving node	Branch parameters		Receiving node avreage load	
			r (Ω)	x (Ω)	P (kW)	Q (kVAr)
1	0	1	1.303	0.408	410	350
2	1	2	0.358	0.124	560	420
3	2	3	1.904	0.808	580	360
4	3	4	0.987	0.456	320	280
5	4	5	0.300	0.110	380	210
6	5	6	0.150	0.055	610	430
7	0	7	1.113	0.501	330	260
8	7	8	0.902	0.414	380	200
9	8	9	0.493	0.228	390	270
10	9	10	0.512	0.232	460	230
11	10	11	0.512	0.232	320	240

Substation voltage: 20 kV.
Total load: $P = 4740$ kW, $Q = 3250$pukVAr.

Seven rows and columns correspond to the active powers in accordance with the locations in Fig. 2; three rows and columns correspond to the wind speeds at three different hub heights; and the last row and column corresponds to the solar radiation. A moderate correlation between the wind and solar activity and the system load is observed, ranging from 0.02 to 0.58. Also, it should be noted that the wind resources are correlated throughout the area and the solar activity can potentially makes a good correlation with the system load at some locations with a specific typical load profile. A visual demonstration of the data presented in Table 2 is shown in Fig. 10 for three TLPs. In this figure, the scatter plots display the mutual dependence structure while the diagonal bar plots show the marginal distributions.

In Fig. 11, the system net load distribution obtained as the difference between the distribution in Fig. 8, and the aggregate wind and solar activities' distribution. These distributions are subtracted considering the correlation between wind speed and load and also by the transformation of wind speed to power.

The presented analysis provides for an appropriate calculation of the system net load distribution, taking into account the dependence structure between the wind and solar activities in different locations throughout the system, as well as with the system load. As discussed in Section 3, not taking this dependence structure into account, either by assuming independence or by using Gaussian copulas, would contribute to different results. Focusing on the specific case study, the obtained distributions are of a major importance for considering the capacity credit added to the

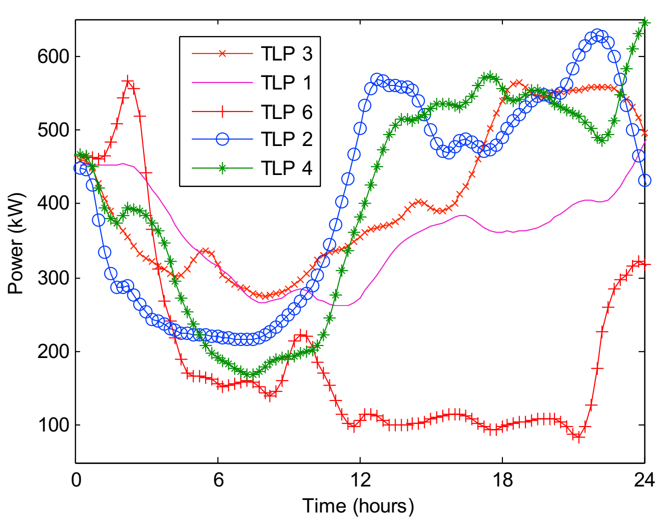

Fig. 5. Scaled patterns of group typical load profiles.

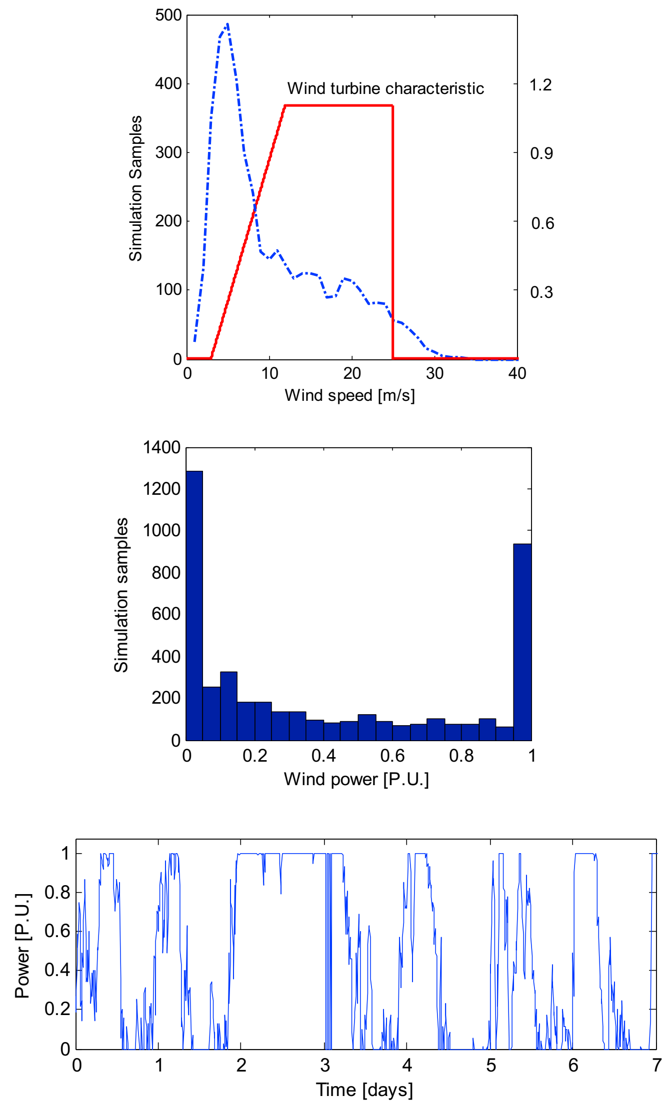

Fig. 6. Wind turbine simulated wind speed/power.

generation system due to the integration of wind and solar power. Besides, the estimation of the system power flow distributions can be performed using the same method if it is repetitively calculated in the system steady state model. These studies enable the system designer to calculate the necessary system reinforcements due to the incorporation of wind and solar powers.

5. A discussion on the applications of the presented method

As previously mentioned, the wide use of renewable DG is supported by the continuous technological development, the effective reduction of greenhouse gas emissions, the spread of automation in control and management of electrical networks and by the liberalization of energy markets. Nevertheless, the gradual increase of DG penetration, especially of renewable type, will determine a deep revision of methodologies for planning and management of electrical distribution networks. The impact of high-penetration renewable distributed generation on electric power system planning methodologies is briefly discussed in this section, and it is outlined how the presented method could makes a contribution to enable effective integration of variable-output renewable generation sources. As a general view, all three areas

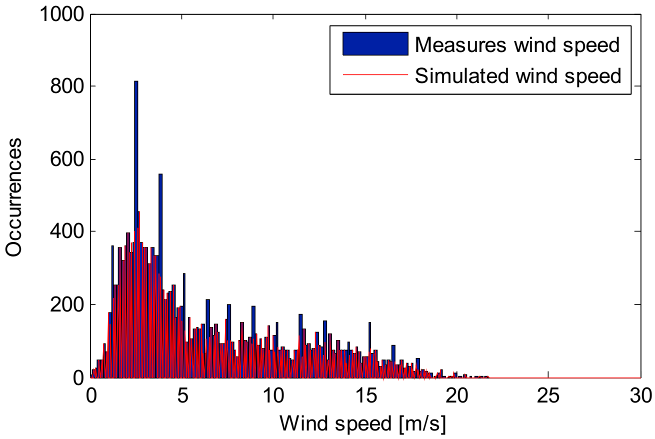

Fig. 7. Recorded vs. simulated wind speed for Davarzan area.

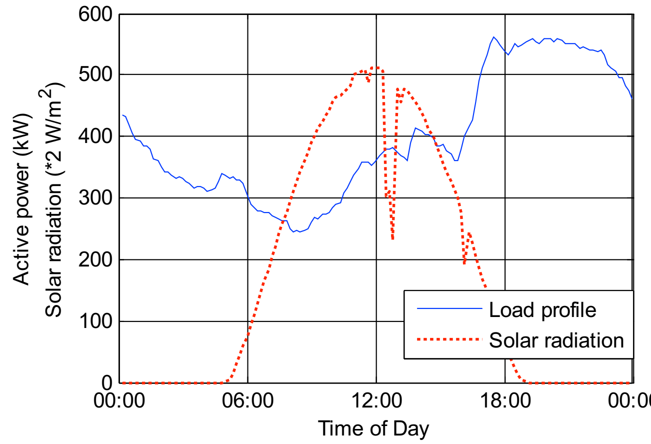

Fig. 9. System load distribution and solar radiation.

of system planning, including generation, transmission, and distribution, should be involved. Some of the potential applications are briefly outlined as followed.

5.1. Generation system planning

Generation planning is shifting from planning for peak load towards planning for system energy. This shift is centered on using net load as a basis for capacity planning and this creates a set of requirements for reliable and large sets of renewable resource data analysis [29]. Besides, the marginal dimension of this shift is to incorporate the variability of net load at the time scale of load following [30]. This needs an increased flexibility by providing effective load control, energy storage, and proper portfolio management. The results of the case study, in the presented form, contribute directly to the first case. Quantifying the variability to determine required flexibility also requires correlated historic load and resource data at the time scales that currently are not being collected. One may notice that the main aspect of using copulas, as presented in Section 2, is to modeling the underlying correlated and non-normal stochastic behavior of the renewable resources. Integration of renewable distributed resource data into generation planning is an important area of future work.

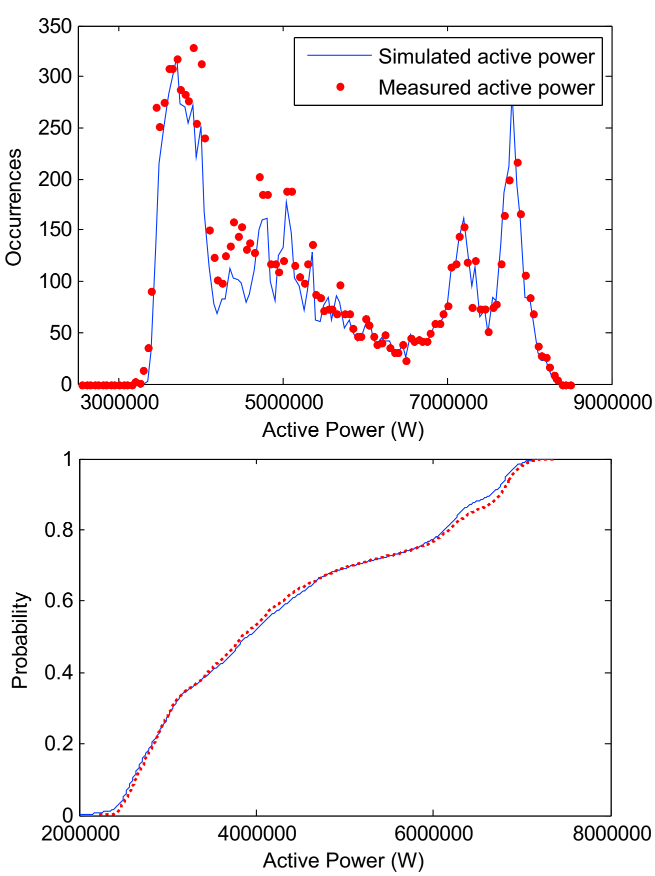

Fig. 8. System load distribution.

To further examine the net load applicability in the power system planning, consider an emerging practice to include renewable energy supply early in the planning process and consider it during energy growth forecast. In this manner, the variability of renewable generation is considered as part of the load variation. It is readily appeared that the system load is reduced to account for contribution from renewable generation. Generation and transmission are then planned relative to this net load with sufficient flexibility to meet the net load requirements. This evaluation of flexibility is a fundamentally important step, as it has a direct impact on the system's operating costs [31]. An emerging viewpoint is proposed by [31] as it is repeated by the diagram shown in Fig. 12. According to this scheme, the contribution of this paper is centered on characterizing the variability of wind/PV power output more precisely by considering the realistic interdependence structure between them and between the consumers' load profiles.

5.2. Distribution system planning and operation

Distribution planning guides already incorporate processes that allow connection of distributed generation [31]. These

Table 2
Copula rank correlation matrix for Davarzan area case study.

1.00	0.23	0.30	0.23	0.50	0.55	0.44	0.16	0.19	0.17	0.19
0.23	1.00	0.67	0.35	0.34	0.02	0.46	0.25	0.19	0.26	0.58
0.30	0.67	1.00	0.24	0.48	0.10	0.67	0.22	0.21	0.23	0.45
0.22	0.35	0.24	1.00	0.11	0.33	0.02	0.07	0.02	0.08	0.34
0.50	0.34	0.48	0.11	1.00	0.35	0.67	0.24	0.25	0.24	0.17
0.55	0.02	0.10	0.33	0.35	1.00	0.26	0.06	0.10	0.06	0.12
0.44	0.46	0.67	0.03	0.66	0.26	1.00	0.22	0.24	0.23	0.29
0.16	0.25	0.22	0.07	0.24	0.05	0.21	1.00	0.61	0.86	0.22
0.19	0.19	0.20	0.02	0.25	0.10	0.24	0.60	1.00	0.66	0.17
0.17	0.26	0.23	0.08	0.24	0.06	0.23	0.87	0.66	1.00	0.24
0.19	0.58	0.45	0.34	0.16	0.12	0.31	0.22	0.17	0.24	1.00

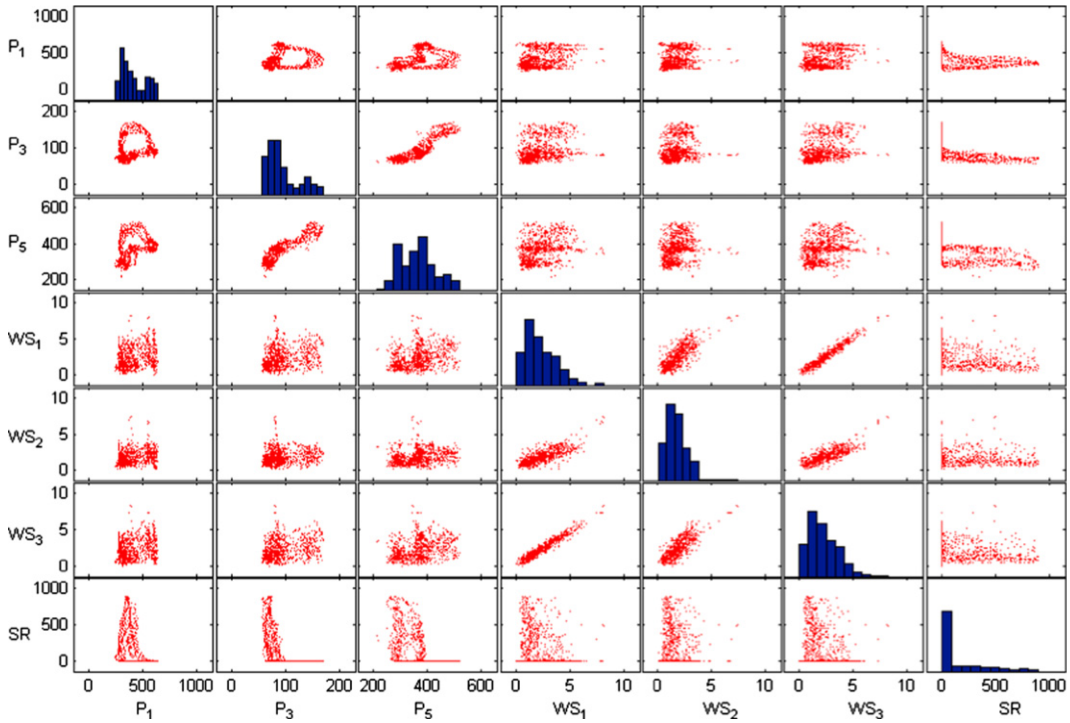

Fig. 10. Some of the simulated buses' powers, wind speeds, and solar radiation by the proposed copula approach as a visual demonstration of the data presented in Table 2.

processes were developed for integrating co-generation; however, they are not optimized for integration of small, renewable distributed resources such as PV/wind. Currently, remaining technical problems are possible to mitigate through careful analysis (stochastic/comprehensive simulations), but the analysis software should be harmonized with respect to the representation of renewable DGs such as PV/wind modules, the impact interfaced inverters have on feeder operating parameters, active and reactive losses, substation power factor, etc. Deterministic modeling of such a system with stochastic non-dispatchable DG units is not trivial, as mentioned in introduction. Therefore, statistical representations of these effects are beneficial for distribution planning and operation purposes.

In order to accurately take into account the effects of stochastic DG output, a whole system needs to be simulated over an extended period of time (typically one year) resulting in extremely large amount of datasets. This can be done via Monte Carlo simulation which may cause an extreme computational burden, as shown by an example in [1]. From this viewpoint, any effort for reducing the input data, while preserving the underlying behavior and dependence structure of data, could be invaluable. Using copulas, in such an application, could provide a predefined arbitrarily reduced set of data for the simulation of realistic renewable power output behavior. It should be mentioned that the copula approach properly models variates (both the expected and extreme values) of the renewable prime movers (such as wind speed or solar insolation) along with their dependence with each other and with the load. Fitting a suitable copula provides a model which can then be used to generate data needed for the subsequent analysis.

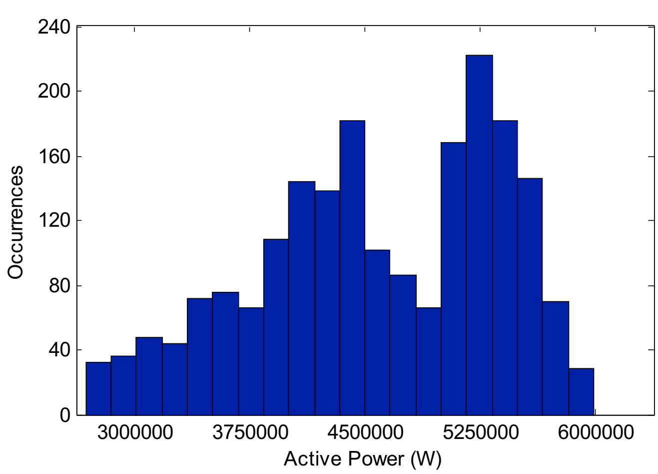

Fig. 11. System net load for the presented integration scenario.

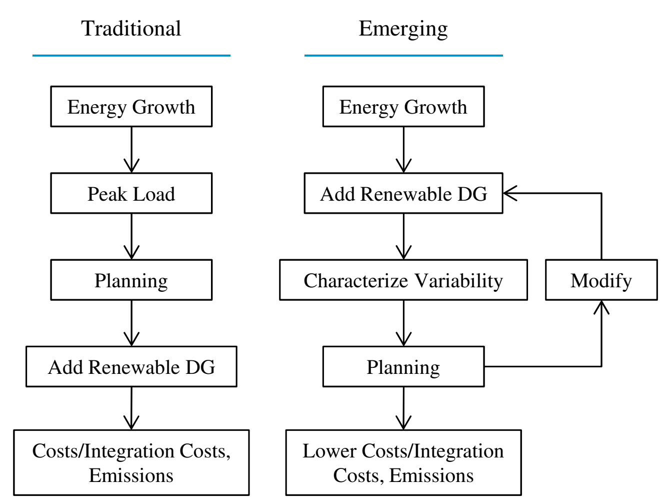

Fig. 12. Traditional and emerging practice in capacity planning [31].

5.3. Prediction uncertainties and their dependence

The non-dispatchable nature of renewable distributed generation implies that system operation depends on power prediction programs to forecast their output. Unfortunately, the prediction results have high levels of uncertainty even in the short-term time horizon. This is mainly because the specific statistical characteristics of stochastic renewable generators (such as wind/PV) impair the performance of forecasting algorithms. Probabilistic load flow (PLF) becomes especially difficult when wind/PV generation is considered. This is due to the high uncertainty of the produced power, the non-Gaussian probability density function and the clear dependence between the wind/PV dispersed generators and the consumers' load profiles. The predictions provided by short-term power prediction programs are uncertain, and this uncertainty must be modeled for an adequate assessment of these predictions through PLF or other optimization/simulation procedures.

The dependence between uncertainties of short-term power predictions could be modeled using copulas, by the presented algorithm in Section 2. This will provide a more accurate modeling compared to other methods such as using a linear correlation matrix, or by assuming independence.

Other applications of the presented method are also possible where a stochastic multivariate uncertainty exists. Considering the rapid movement of passive power systems towards highly active topologies forming smarter grids, it seems that the need for modeling of such uncertainties will increase. In fact, both the technical problems such as interactions between converters and the grid network in the new interconnected active system, and the financial risks, requires an adequate representation and modeling of uncertainty in a multivariate context. This could be efficiently engaged by using copulas, as it is shown for a case study in this paper.

6. Conclusions

This paper suggests a new type of analysis related to the deterministic-stochastic dependencies in power systems. The performance assessment of a distribution feeder equipped with renewable distributed generators is a challenging problem at the planning stage. The complexity arises from the uncertainties involved in predicting the DG output due to stochastic nature of its input (wind, insolation), its location on the feeder, and interaction with feeder load, which is another stochastic process. Simulating stochastic processes that drive DG output, as well as DG placement, combining it with extensive field measurements of load profiles at feeders that are candidates for DG installations, and determining their interaction, result in large datasets. In order to obtain the usual load profiles, a method is necessary which takes into account the stochastic dependence structure in a multivariate context. In order to investigate and model stochastic dependence, the copula theory has been presented. The Archimedean modeling algorithm for the Frank copula has been presented together with the modeling of interactions when the wind and the solar powers are integrated with the system load in Davarzan area in Iran.

The procedure is demonstrated on an 11-bus MV distribution feeder, with randomly placed PV generators and wind turbines. However, the procedure is not affected by the changes in feeder topology, as it operates on a set of load and DG profiles that may be applied to an arbitrary feeder. The proposed algorithms serve as a bridge between uncertainties imposed by having intermittent, stochastic DG and conventional tools for analysis of distribution systems. They provide more answers to the designer of DG-enhanced networks.

References

[1] Pregelj A, Begovic M, Rohatgi A. Quantitative techniques for analysis of large data sets in renewable distributed generation. IEEE Trans Power Syst August 2004;19(3):1277–85.

[2] McHenry MP. Why are remote western Australians installing renewable energy technologies in stand-alone power supply systems? Renew Energy 2009;34(1):1252–6.

[3] R Development Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/; 2006 (accessed 10 June 2009).

[4] Simulink, MATLAB 7.4.

[5] Valizadeh Haghi H, Tavakoli Bina M. A Study on Probabilistic Evaluation of Harmonic Levels Produced by Static Compensators. In: Proceedings of the Australasian Universities Power Engineering Conference, AUPEC, Sydney, Australia; 2008. p. 1–6.

[6] Valizadeh Haghi H, Tavakoli Bina M. Complete harmonic-domain modeling and performance evaluation of an optimal-PWM-modulated STATCOM in a realistic distribution network. Przegląd Elektrotechniczny 2009;85(1): 156–61.

[7] Papaefthymiou G, Schavemaker PH, VanDerSluis L, Kling WL, Kurowicka D, Cooke RM. Integration of stochastic generation in power systems. Int J Elect Power Energy Syst 2006;28:655–67.

[8] Papaefthymiou G. Integration of stochastic generation in power systems. PhD diss, Delft University of Technology, Delft, Netherlands; 2007. Available from: http://www.library.tudelft.nl/.

[9] Nelsen RB. An introduction to copulas. 2nd ed. New York: Springer; 2006.

[10] Pfeifer D, Neslehova J. Modeling dependence in finance and insurance: the copula approach. Blatter der DGVFM(2):177–91. Available from: http://www.staff.uni-oldenburg.de/dietmar.pfeifer/Copulas.pdf, 2003;26 (accessed 20 Oct 2008).

[11] Sklar A. Functions de repartitiona n dimensions et leurs marges. Pub Inst Statist Univ Paris 1959;8:229–30.

[12] Embrechts P, McNeil A, Straumann D. Correlation and dependence in risk management: properties. In: pitfallsDempster M, editor. Risk management: value at risk and beyond. Cambridge University Press; 2002. p. 176–223.

[13] Rebonato R, Jackel P. The most general methodology to create a valid correlation matrix for risk management and option pricing purposes. J Risk 2000;2:17–27.

[14] Frees EW, Valdez EA. Understanding relationships using copulas. In: Proceedings of the 32nd actuarial research conference, Alberta, Canada; August 1997.

[15] Genest C, MacKay J. The joy of copulas: bivariate distributions with uniform marginals. Am Statist 1986;40:280–3.

[16] Yan J. Enjoy the joy of copulas. J Stat Software 2007;21(4):1–2.

[17] Frees EW, Wang P. Credibility using copulas. North Am Actuar J 2005;9(2):31–48.

[18] Genest C, Rivest LP. Statistical inference procedures for bivariate Archimedean copulas. J Am Stat Assoc 1993;88:1034–43.

[19] Strelen JC, Nassaj F. Analysis and generation of random vectors with copulas. In: Proceedings of the Winter Simulation Conference; 2007: p. 488–496.

[20] Borowy B, Salameh Z. Methodology for optimally sizing the combination of a battery bank and PV array in a wind/PV hybrid system. IEEE Trans Energy Conv June 1996;11:367–75.

[21] Jayadev J. Harnesing the wind. IEEE Spectrum 1995;32:78–83.

[22] Ameri M, Ghadiri M, Hosseini M. Recent advances in the implementation of wind energy in Iran. In: Proceedings of the joint international conference on sustainable energy and environment (SEE-2), Bangkok, Thailand; 2006. p. 21–23.

[23] Moosavi SM, Pari MN. Data and information needed from SUNA for each major technology area. Tehran, Iran: Iranian Renewable Energy Organization (SUNA); June 2008.

[24] World Design Insolation-Solarex;2004.

[25] Jardini JA, Tahan CMV, Gouvea MR, Ahn SU, Figueiredo FM. Daily load profiles for residential, commercial and industrial low voltage consumers. IEEE Trans Power Del Jan 2000;15:375–80.

[26] Gerbec D, Gasperic S, Smon I, Gubina F. Allocation of the load profiles to consumers using probabilistic neural networks. IEEE Trans Power Syst May 2005;20:548–55.

[27] Valizadeh Haghi H, Nekoui MA, Tavakoli Bina M. Allocation of daily load profiles for a group of consumers without data loggers in distribution system of Tehran. In: Proceedings of the 24th international power system conference, PSC, Tehran, Iran; Nov 2009. p. 1–11.

[28] Papaefthymiou G, Kloeckl B. MCMC for wind power simulation. IEEE Trans Energy Conv March 2008;23(1):234–40.

[29] Willis HL. Power distribution planning reference book. 2nd ed. New York: Marcel Dekker; 2004.

[30] Parsons B, Milligan M, Smith JC, DeMeo E, Oakleaf B, Wolf K, et al. Grid impacts of wind power variability: recent assessments from a variety of utilities in the United States. European wind energy conference, Athens, Greece. Available from:, http://www.nrel.gov/docs/fy06osti/39955.pdf; 2006.

[31] Bebic J. Power system planning: emerging practices suitable for evaluating the impact of high-penetration photovoltaics. Subcontract report, NREL/SR-581-42297. Available from:, http://www.osti.gov/bridge; 2008.