Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017
doi:10.20944/preprints201709.0053.v1

Article

Spatial Dependence Modeling of Wind Resource under Uncertainty Using C-Vine Copulas and Its Impact on Solar-Wind Energy Co-Generation

Apurva Narayan [1,] *, Kumaraswamy Ponnambalam [1] and Sheree A. Pagsuyoin[2]

[1] University of Waterloo, Ontario, Canada N2L 3G1; {a22naray, ponnu}@uwaterloo.ca
[2] University of Massachusetts Lowell, Massachusetts, USA 01854; Sheree_Pagsuyoin@uml.edu
* Correspondence: a22naray@uwaterloo.ca, Tel.: +1-519-888-4567 x31450

Abstract: Investments in wind and solar power are driven by the aim to maximize the utilization of renewable energy (RE). This results in an increased concentration of wind farms at locations with higher average wind speeds and of solar panel installations at sites with higher average solar insolation. This is unfavourable for energy suppliers and for the overall economy when large power output fluctuations occur. Thus, when evaluating investment options for spatially distributed RE systems, it is necessary to model resource fluctuations and power output correlations between locations. In this paper, we propose a methodology for analyzing the spatial dependence, accurate modeling, and forecasting of wind power systems with special consideration to spatial dispersion of installation sites. We combine vine-copulas with the Kumaraswamy distribution to improve accuracy in forecasting wind power from spatially dispersed wind turbines and to model solar power generated at each location. We then integrate these methods to formulate an optimization model for allocating wind turbines and solar panels spatially, with an end goal of maximizing overall power generation while minimizing the variability in power output. A case study of wind and solar power systems in Central Ontario, Canada is also presented.

Keywords: renewable energy; wind and solar power; Kumaraswamy distribution; C-Vine copula

1. Introduction

Wind power is one of the world's largest and most accessible high intensity renewable energy resource, with solar power fast becoming a widely implemented renewable resource [1]. Globally, there are increasing efforts to tap more into these renewable energy sources; however, their intermittent availability presents one barrier for the renewable energy-based systems to entirely meet energy demands [2]. Wind fluctuations can be abrupt and significant, causing problems with the ability to generate steady energy outputs. Also, due to the stochastic nature of wind, it is difficult to accurately forecast wind power generation by considering only temporal wind behavior when other factors such as wind farm topology and turbine characteristics are equally important [3]. On the other hand, while the availability of solar energy is relatively constant, solar power output exhibits high sensitivity to slight changes in solar insolation [4].

Another challenge associated with renewable energy systems is their integration into the main power grid. Renewable energy installations can be geographically sparsely distributed despite being part of the same power grid, leading to sub-optimal power transactions within the grid [5]. Decisions on where to place

© 2017 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

these installations are often based on the availability of wind and solar resources in order to maximize perceived power outputs [6]. This leads to localized concentrations of installations in areas of high wind or solar availability, which can become highly unfavourable for energy suppliers due to increased power fluctuations and overall system instability [6].

Previous researchers have examined the possibility of smoothing fluctuations in wind power generation through employing geographically dispersed systems, or by interconnecting existing dispersed systems [7, 8]. For example, system reliability has been found to increase with turbine size in wind farms [9], while interconnection has been shown to greatly impact the reliability and stability of renewable energy systems [10].

In this paper, we propose a methodology for analyzing the spatial dependence, accurate modeling, and forecasting of wind power generation with special consideration to temporal variations in power output and spatial dispersion of installation sites. The rest of this paper is organized as follows: Section 2 provides a brief review of literature related to wind and solar power systems modeling. Section 3 introduces the three mathematical concepts that serve as the foundations of our proposed methodology: the Kumaraswamy distribution, the theory of copulas, and vine-copulas. Section 4 discusses our proposed model for the optimal allocation of renewable energy generation technologies while Section 5 presents a case application of the proposed methodology. Finally, Section 6 summarizes the findings and intellectual contributions of this study.

2. Literature Review

Linear correlation coefficients provide general information about the interdependence of wind power generated at spatially distributed sites [11]; however, they do not uniquely describe the structure of this dependence [11]. Further, they do not provide actionable information that is helpful to system planners and operators. For example, linear correlation coefficients are not useful in determining the duration in a year when the aggregate wind power in a system will be above or below a specified threshold value even when coupled with data on the marginal distributions of wind power at each installation, as dependence relations are nonlinear. A potential method for mathematically describing the dependency structure among wind power systems involves using joint distribution functions [12]. However, multivariate distribution models are currently not available for such systems, and common joint distributions do not accurately fit wind power data [12]. A possible workaround suggested by Kroese et al [13] involves decomposing the assumed correlation matrix using Cholesky decomposition, but this is only applicable if random variables are linearly correlated.

It has been demonstrated that wind speeds are characterized by non-normal distributions and non-linear dependence [6]. This becomes problematic in multivariate analysis; when multivariate data are not normally distributed, accurate quantiles of the sums of margins cannot be calculated from the sums of variances and covariances which makes modeling these random variables (wind speeds in our case) more challenging.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

Goethe and Schnieders [6] modeled the univariate time series of wind speed at several wind farms in Germany using a seasonal autoregressive moving average (ARMA) model proposed by Benth and Benth [14]. To model the correlation between multiple wind farm locations, they analyzed the correlation between the residuals of the various univariate time series and fit copulas to the residuals, thus developing copula-GARCH (generalized autoregressive conditional heteroskedasticity) models.

A more appropriate approach for modeling non-linear, non-normal and more complex dependency structure in data is by directly using appropriate copulas [6, 15-18]. Copulas are applied widely in finance [19, 20], and they possess unique characteristics that make them highly attractive in modeling wind power [18]. Of these characteristics, the most important is the ability of copulas to model the dependence structure of data independent of the marginal distributions of the participating variables. This feature is very critical because wind power outputs at different locations are often significant at the grid nodes they are infused and modeling them using single marginal distribution is not possible. Therefore, finding this dependence in power outputs independent of marginal distributions is of great advantage for system planners as it allows modeling wind power generation more accurately.

The correlations among wind power generated at different locations are usually estimated from parameters such as separation distance and averaging period, among others [21, 22, 23]. If only basic information is available about the locations of wind turbines, an accurate model of the dependency structure of wind power generated at these locations can be produced using copulas. Consequently, the selection of an appropriate copula function is very important, as inappropriate selection can lead to unacceptable errors. Of all copulas, the Gaussian copula is the most commonly used copula due to its computational convenience; however, its suitability in wind power analysis has not been rigorously investigated. The standard Gaussian copula has been previously used to model wind power in Europe based on a qualitative assessment of Q-Q plots [17]. Louie (2012) adopted a more comprehensive approach by first testing a number of standard copulas on wind speed data, and then eventually selecting Archimedean copulas [23].

In modeling wind power, copulas have the highest utility in forecasting and in generating scenarios for optimization simulations [19]. These scenarios are necessary in stochastic programming, which is a critical decision tool in power systems analysis and planning research. For example, Gaussian copulas have been used to evaluate short-term scenarios for wind power generation [24], while empirical copulas have been used in modeling the dependency structure between the wind speed and the wind power output [25]. A quantile-copula kernel density estimator has also been used to improve probabilistic wind power forecasts [26].

With respect to solar energy, temporal modeling of solar power generation has been done using generalized distribution functions that were subsequently optimized to ensure reliable and higher power outputs [27]. Solar irradiation is most often modeled using the Hollands and Huggets distribution, which can be approximated by the Gamma distribution [28]. To our knowledge, there has been no attempt to date

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

to model solar power generation using other types of probability distributions. Similar to wind power generation, the dependence structure of solar systems is usually quantified by measures of association such as linear correlation coefficients [29]. But in contrast to wind power generation, the spatial variability in solar power generation in reasonably sized grids is not significant; thus modeling the spatial dependence of solar power generated between dispersed locations is not necessary [30] but can be done with the method we propose for wind power.

3. Methodology

Wind speed patterns and their spatial dependencies are generally non-Gaussian and non-linearly correlated [14]. Since system planners are more interested in modeling wind power generation than wind speeds, this presents a challenge because there is no standardization in modeling wind power using a specific probability distribution. Therefore, in the present study, we use the Kumaraswamy distribution for the temporal modeling of wind power generated at each site, and applied the concept of vine-copulas to model wind power dependencies.

3.1 Kumarawamy Distribution

First introduced in 1980, the double bounded Kumaraswamy distribution is a continuous probability distribution that was originally developed for hydrology applications [26]. It is equivalent to the Beta distribution but has a simpler analytical formulation, making it more efficient in computational simulations. More importantly for this study, the Kumaraswamy distribution is selected because (i) renewable power is a non-linear transformation of its resource (ex. wind power from wind speed) and (ii) its simple analytical form allows for its easy integration with copulas.

The probability density function (PDF, ($f(x)$) and cumulative density function (CDF, ($F(x)$) formulations of the Kumaraswamy distribution are given in Eq. 3.1 and Eq. 3.2, respectively, where a and b are shape parameters describing the distribution.

$$f(x) = abx^{a-1}(1-x^a)^{b-1} \tag{3.1}$$

Where,

$$a > 0,\ b > 0 \text{ and } x\,\epsilon[0,1] \tag{3.2}$$

It has many of the same properties as the Beta distribution but has some advantages in terms of tractability. The Kumaraswamy densities have similar behavior as the Beta densities such as they are unimodal, uniantimodal, increasing, decreasing or constant depending on the parameters. Therefore, based on the values of the shape parameters the densities take specific shape and exhibit certain properties such as, if $a > 1$ and $b > 1$ then the density is unimodal, if $a > 1$ and $b \leq 1$ then the density is increasing, $a < 1$and $b < 1$ then the density is uni-antimodal, and $a \leq 1$ and $b > 1$ then the density is decreasing. The densities are log-concave if and only if the shape parameters are greater than or equal to 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

In addition to hydrology, Kumaraswamy distribution is now widely used including in finance, statistical design centering of integrated systems, among others [31, 32] .

3.2 Methodology

3.2.1 Copulas and the Sklar Theorem

Copulas were first introduced in 1959 by the mathematician Abe Sklar [33] and have since become popular in describing the dependencies between random variables. Copulas are mathematical functions that allow us to combine univariate distributions to obtain a joint distribution with a particular dependence structure. The utility of a copula is most easily demonstrated in the use of distributions in probabilistic analysis. To illustrate, recall that the CDF of a distribution is used to draw a random variate. Most commonly, to draw a random value from a distribution, one starts by sampling from a uniform distribution, $U(0,1)$. This sample is treated as an observation of the variable`s CDF; a sample can be drawn from the PDF by generating a uniform random number and transforming it using the CDF to a random value.

Sklar's theorem is the theoretical foundation of copulas. It states that for a given joint multivariate distribution function and relevant marginal distributions for the corresponding random variables, there always exists a copula function that relates the marginal distributions of the variables. Mathematically, this can be derived as follows.

Let F_{xy} be a joint distribution with margins F_x and $F_{y.}$; then there exists a copula $C: [0,1]^2 \rightarrow [0,1]$ such that

$$f(x) = abx^{a-1}(1-x^a)^{b-1} \qquad 3.3$$

If the random variables, **X** and **Y** are continuous, then copula, C is unique; otherwise, C is uniquely determined on the (range of X) × (range of Y).

Conversely if C is a copula and F_x and F_y are distribution functions, then the function F_{xy} is a joint distribution with margins F_x and F_y.

C must be a function of particular type with certain properties as described by [33] and explained further in [19].

The copula is further defined as follows.

C is a copula if $C: [0,1]^2 \rightarrow [0,1]$ and

$$C(0, u_m) = C(v_m, 0) = 0$$

$$C(1, u_m) = C(u_m, 1) = u_m$$

$$C(u_{m2}, v_{m2}) - C(u_{m1}, v_{m2}) - C(u_{m2}, v_{m1}) + C(u_{m1}, v_{m1}) \geq 0 \text{ for all } v_{m1} < v_{m2},\ u_{m1} < u_{m2}$$

If C is differentiable once in its first argument and once in its second then, C is equivalent to $\int_{v_{m1}}^{v_{m2}} \int_{u_{m1}}^{u_{m2}} \frac{\partial^2 C}{\partial u_m \partial v_m} du_m dv_m \geq 0$ for all $v_{m1} < v_{m2},\ u_{m1} < u_{m2}$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

where u_m, v_m, u_{mi}, v_{mi} are marginal distribution functions.

This definition of a copula simply states that a copula is itself a distribution function, defined on $[0,1]^2$ with a uniform marginal. Each of the marginal distributions produces a probability of one-dimensional events. The copula function takes these probabilities and maps them to a joint probability, enforcing a relationship on the probabilities. Therefore, using copulas to build multivariate distributions is a very flexible and powerful technique as it separates choice of dependence from the choice of marginal [19].

Sklar's theorem establishes one of the easiest ways of constructing copulas. In this case, if F_x and F_y are the marginal distributions, then a copula is given by the formulation in Eq. 3.4.

$$C(u_m, v_m) = F_{XY}(F_X^{-1}(u_m), F_Y^{-1}(v_m)) \qquad 3.4$$

3.2.2 Selection of the Appropriate Copula

A critical step in modeling data using copulas is the selection of the appropriate copula function from among the family of copulas that best describes the given data set. The selection process is often based on the analytical tractability of the copula function [34]. Three types of copulas are considered in this study: Gumbel, Joe-Frank, and the Student t. The Gumbel copula is most suited for extreme distributions while the Joe-Frank and Student t copulas are more suited for applications with heavy dependence on tails [19, 35].

The Gumbel copula is a bivariate Archimedean copula. It is an asymmetric copula that exhibits greater dependence on the positive tail than on the negative tail. This copula is given by Eq. 3.5, where δ is the parameter controlling the dependence between the marginal distributions u and v.

$$C_\delta(u,v) = \exp(-[(-\log u)^\delta + (-\log v)^\delta]^{1/\delta}) \qquad 3.5$$

The Joe-Frank copula, sometimes called the BB8 copula, is a two-parameter copula also from the Archimedean family of copulas. The copula CDF is given by Eq. 3.6, where the parameter δ illustrates the degree of dependence between the marginal distributions u and v, and the parameter ϑ is the degree of freedom.

$$C_{\vartheta,\delta}(u,v) = \delta^{-1}\left(1 - \{1 - \eta^{-1}[1-(1-\delta u)^\vartheta][1-(1-\delta v)^\vartheta]\}^{1/\vartheta}\right) \qquad 3.6$$

where

$$\vartheta \geq 1 \; and \; 0 \leq \delta \leq 1$$

$$\eta = 1 - (1-\delta)^\vartheta \text{ and } 0 \leq u, v \leq 1$$

The Student t copula allows for joint fat tails and an increased probability of joint extreme events compared with the other copulas. Increasing the value of ϑ decreases the tendency to exhibit extreme co-movements. The copula formulation is expressed in Eq. 3.7, where ρ and ϑ are the parameters of the copula, and t_ϑ^{-1} is the inverse of the standard univariate t-copula with ϑ degrees of freedom, expectation 0 and variance $\frac{\vartheta}{\vartheta-2}$ [35]. The variables *s* and *t* are the random vectors obtained from the two marginal distributions.

$$C_{\rho,\vartheta}(u,v) = \int_{-\infty}^{t_\vartheta^{-1}(u)} \int_{-\infty}^{t_\vartheta^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{1 + \frac{u^2 - 2\rho uv + v^2}{v(1-\rho^2)}\right\}^{-(\vartheta+2)/2} dsdt \qquad 3.7$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

3.2.3 C-Vine Copulas

Joe [35] presented the first construction of a multivariate copula using (conditional) bivariate copulas, while Bedford and Cooke [36] developed a more general construction method of multivariate densities and introduced regular vines to organize different pair-copula constructions (PCCs). Vines are a graphical representation of constraints in high dimensional probability distributions. They are used to specify so-called PCCs, as introduced by Aas et al. [37].

Conventionally, a copula model is limited to a 1-parameter or 2-parameter specification of the dependence structure, which represents a potentially severe empirical constraint. Clearly, when modeling the joint distribution of multiple variables, such limited parameter models are unlikely to adequately capture the dependence structure between variables. For example, the Gaussian copula lacks tail dependence. Similarly, while the multivariate Student t copula is able to generate different tail dependence for each pair of variables, it imposes the same upper and lower tail dependence across all pairs. These limitations are overcome by the canonical vine (C-vine) model by building bivariate copulas of conditional distributions. C-vine copulas are flexible multivariate copulas that are generated via hierarchical construction and can be decomposed into a cascade of bivariate copulas. The basic principle is to model dependence using simple local building blocks (pair-copulas).

4. Spatial and Temporal Modeling of Renewable Energy Resources

4.1 Algorithm for Temporal Modeling of Wind Power and for Scenario Generation

This section discusses the procedure for modeling wind power generation in various spatially dispersed sites using the generalized Kumaraswamy distribution and C-Vine copulas.

Firstly, given temporal data sets (daily and seasonal) on the wind power generated at different installations, we use the Kumaraswamy model to describe the probability distribution of each data set. We obtain the model parameters for each hour of the day and for three seasons in the year lets call them Season 1, Season 2, and Season 3 using the Maximum Likelihood Estimate (MLE) method for distribution fitting with historical data. This ensures that both hourly and seasonal variations are embodied in the distribution models. Therefore, we create for each hour of the day within each of the three seasons a distribution from all the measurements at that hour of day for all days within that season across a number of year. This leads us to create 3 (season) × 24 (hours) = 72 distributions for each location. These distributions are then used to choose a C-Vine model for the installation site under consideration.

In order to develop the C-Vine tree, one location must first be selected as the root node of the tree and the others its children (nodes). This is accomplished by generating the Kendall rank correlation matrix, and summing the correlations across each location with respect to the other locations. The location with the maximum value of the Kendall rank correlation is chosen as the root node.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

Once the C-vine tree is constructed, various families of bivariate copulas are then fitted to model the dependence between the root node and each one of its children. We again use the MLE method to fit the copulas, and use AIC/BIC (Akaike Information Criterion/ Bayesian Information Criteria) to evaluate the goodness of fit. In this, pair-copula construction approach, a bivariate copula is fitted to the root node and the child. Finally, we utilize the PCC-based tree to produce scenarios by drawing data from the PCC followed by Kumaraswamy distribution for each hour of the day or season.

.

The algorithm for modeling of wind power is summarized in Table 1.

Step	**Specific Action**
1	Fit Kumaraswamy distribution to each location's hourly data for the three seasons (Season 1, Season 2 and Season 3) and obtain parameters for the distribution.
2	Compute the Kendall rank correlation matrix with correlation values where Correlation Values, location with respect to location
3	Formulate the Vine tree, where root is the location with $\max(\sum_{i,:} \sigma_{ij})$.
4	Compute the Pair Copula Construction using the various copula options available
5	Generate scenarios from the PCC followed by inverse of Kumaraswamy CDF.

For scenario simulations using C-Vine, we generate the Vine matrix that defines the connections and parameter matrices containing the parameters of each of the copulas defined by each link.

4.2 Algorithm for Spatial Modeling of Solar Power Generation

Similar to the procedure outlined in Section 4.1, we also use PCC for the spatial modeling of solar power generation. Our goal here is to develop a standardized approach for spatial modeling of renewable power sources. The procedure consists of two steps, as indicated in Table 2. Firstly, given the hourly (power generation) data at each location for the three seasons (Season 1, Season 2, and Season 3), we use the Kumaraswamy distribution to model hourly solar power outputs. We then generate scenarios by drawing random variables from the Kumaraswamy CDF for each day for the three seasons. Although, solar power is given by a strong deterministic component, and limiting the upper limit of the power output for each hour in the year, moderated by a stochastic process can be justified given the random behavior in solar insolation due to cloud cover, wind direction, smog, and other environmental factors.

Table 1. Steps in Modeling Solar Power Generation

Steps	**Specific Action**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

Step 1	Fit Kumaraswamy distribution to each location's hourly data for the three seasons (Season 1, Season 2 and Season 3) and obtain parameters for the distribution.
Step 2	Generate scenarios from the Kumaraswamy CDF for each hour of day for the three seasons.

4.3 Model Optimization for the Optimal Allocation of Wind Turbines and Solar Panels

Once we obtain: (i) the model parameters for simulating data for a given site based on the marginal distributions (i.e., Kumaraswamy distribution), and (ii) the dependence structure model parameters using Vine-copulas, we can find an optimal allocation of wind power and solar power at each location in a given space. For system planners, this information is important in deciding on the number of wind turbines and solar panels that needs to be installed at each location to optimize power output.

In this study, our goal is to investigate the lower quantiles of the distribution of the overall renewable energy (wind and solar) produced within a power system. These quantiles should be maximized to design an optimal placement of the renewable energy installations. Because our approach is based on probability distribution and the persistence in the hourly wind power is found to be not so strong in our data and we can consider it to be independent for each hour. Similarly, for solar power since solar insolation follows a daily pattern and the insolation at each hour can be considered independent. This allows us to model the data using Kumaraswamy distribution and considered hourly variations as independent.

Suppose we have n locations in a given space (site). We need to make allocations of wind power and solar panel installations at each location such that the allocation maximizes the overall power generation and smoothens the total system power output. To reiterate, fluctuations in the total system output are due to the erratic nature of the renewable sources as discussed in Sections 1 and 2. Thus, the overall objective of the optimization problem is to minimize the negative effects of the erratic nature of the renewable energy (wind and solar).

The optimization model is depicted in Eq. 3.8, where W_S and W_W are weightages of solar power and wind power allocation at each location, respectively. $X_{S,l}$ and $X_{w,l}$ are solar and wind generation scenarios, respectively, for each location l in the total n locations.

$$\max Quantile_{\alpha}\left(W_{W,S}{}^{T}X\right) \qquad 3.8$$

where

$$W_{W,S}{}^{T}X = \left(W_{W_l} \times X_{w,l}\right) + \left(W_{S_l} \times X_{S,l}\right)$$

$$l \in 1 \ldots n, for\ all\ locations, \text{ s.t.}$$

$$\sum_{i=1}^{n} W_{W_i} = 1$$

$$\sum_{i=1}^{n} W_{S_i} = 1$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

$$0 \leq W_{S_i}, W_{W_i} \leq 1$$

Eq. 3.8 represents the joint quantile optimization for solar and wind power allocation at a given site. Such an approach tries to smoothen the power output in the entire power system by choosing an optimal scheme for allocating solar and wind power resources. It also results in a more accurate modeling of the renewable resources, as it considers not only the temporal but spatial features of wind power. Optimizing the formulation in Eq. 3.8 will ensure that $(1-\alpha) * 100\%$ of all cases, the total power produced will be above the α-quantile.

5. Ontario Case Study

5.1 Location

The modeling methodology proposed in Section 4 has been applied in the case study of wind and solar power systems in four sites in Central Ontario, Canada: Pearson, Toronto, BillyBishop, and Buttonville (see Figure 1). These sites are important and unique due to their proximity to a densely-populated city (City of Toronto) and a large water body (Lake Ontario), and their association with the main power grid in the Greater Toronto Region. The power demand in this region is very high (27,000 MW /day on peak demand), therefore it is critically important to achieve a stable power supply in the region. Increasing the penetration of renewable energy-based systems, specifically wind power, may lead to instability in the available power in the grid.

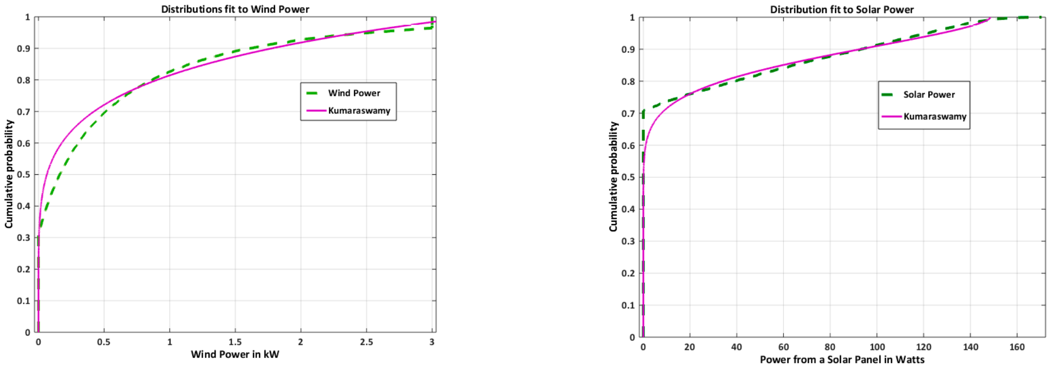
Figure 1: Central Ontario with all four locations under consideration

5.2 Fitting Probability Distribution Models to Wind and Solar Power Data

For the case study, we used data on solar power using the hourly solar insolation data for 3 years, available through RETScreen [38]) and similarly data on wind power was generated using a wind turbine [39] model for hourly wind speeds for 3 years for Toronto. Weibull is used as a standard for modeling wind speed, log normal has been used at times as well and Gamma distribution has been sparingly used for modelling solar insolation but there has been no standard distribution for fitting wind or solar power generated. It is important as both wind speed and solar insolation undergo a non-linear transformation and hence cannot be fit using either Weibull distribution for wind power or Gamma distribution for solar power. Therefore, there is a need for a generalized distribution such as the Beta distribution that be used for fitting both solar and wind power. We tested the Kumaraswamy distribution for fittingboth solar power and wind power data sets given its ability as a general distribution. It gives the flexibility and avoids the numerical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017
doi:10.20944/preprints201709.0053.v1

intractability that inhibit the use of the Beta distribution. The Kumaraswamy model best describes the probability distribution of the data (see Figure 2) and hence will be used for the further study with copulas.



(a) (b)

Figure 2: Probability distribution models fitted to data from Toronto, Ontario for Season 3 for the three year period (a) Wind Power generated from Whisper 500 3kW wind turbine (b) to solar power from a 0.18kWp solar panel

5.3 Analysis of Wind Power for Dependence Modeling

We also performed a pair-wise comparison of wind power data (by transforming the wind speed information obtained from RETScreen to power using the Whisper 500 wind turbine model at each of the four sites. It is a Type 1 wind turbine.) for the four sites to determine the correlation between power data at the sites (see Figure 3). Since the correlations appeared non-linear and data distribution was non-Gaussian, we chose Kendall rank correlation as our correlation parameter. To ensure that the wind power data were non-linearly correlated, we further grouped the data set for each paired site into 3 equal subsets. For each site pairing, data were randomly assigned to a sub-group. An analysis of the data correlations in the subsets independently revealed that the correlations varied markedly and were not constant for the sub-groups of each paired site (see Figure 3). With the exception of one sub-group for the Toronto-Pearson paired location, the Kendall rank values for all other sub-groups indicate non-linear behavior. For example, the sub-plot between Buttonville and BillyBishop has an overall correlation of 0.26 whereas the corresponding three subsets (low, medium, and high values) have varying correlation values of 0.06, 0.23 and 0.11.

5.4 C-Vine Copula Generation

Based on the observations of wind power behavior in Section 5.3, the wind power output at each site are modeled using the Kumaraswamy distribution for each of the four sites in the case study to generate pair-copula construction. We employed vine copulas to model the spatial dependence of wind power production sites; this choice of copula was influenced by the following factors, as seen in Figure 3:

- the wind power in each location was non-normally distributed;
- the Kendall rank correlations of wind power between sites varied (the correlation coefficients of the subplots of Figure 3 differ); and
- the wind power outputs at the sites were non-linearly dependent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

The Kumaraswamy distribution was fitted to historical wind power data from each site. Maximum likelihood estimates were used to obtain the distribution parameters. To model the dependence of wind power generated from the four sites with C-vine copula, we first converted the wind power data from the real domain to copula data, which lie inside the $[0,1]$ hypercube. This was accomplished by taking the Kumaraswamy CDF of the individual data series.

As described earlier in Section 4.1 (Step 3), we need to identify the root node of the C-Vine tree. To do this, we generated the Kendall correlation matrix for the sites and added together the correlation coefficients of each site (rightmost column of Table 3). This sum is an indicator of the strength of the correlation of a site's wind power output to other locations' wind power outputs. Consequently, the site with the highest the sum is selected as the root node of the C-vine tree. In our case study, the root node of the C-vine tree is the Pearson site.

Table 2: Kendall Correlation Matrix for the Four Sites

Index		**Toronto**	**Pearson**	**BillyBishop**	**ButtonVille**	**Sum**
1	**Toronto**	1.000	0.811	0.561	0.213	2.585
2	**Pearson**	0.811	1.000	0.566	0.233	**2.610**
3	**BillyBishop**	0.561	0.566	1.000	0.255	2.382
4	**Buttonville**	0.213	0.233	0.255	1.000	1.701

The C-vine tree representation of our power generation sites is shown in Figure 5. The three other sites (Toronto, BillyBishop and Buttonville) are connected to the root node (Pearson) by a link representing the pair-copula construction between the root node and the site connected to it.

We used bivariate copulas to formulate the PCC for the three links. Each link represents a copula describing the dependence in the marginal distributions of wind power at each site. We used the marginal distribution (Kumaraswamy distribution) for each of the three site pairs and generated scenarios of power production for each site. These generated data were then used to estimate the copula parameters. The choice of the copula function was based on analytical tractability and simplicity and best fit to data. Copula fitting was performed in R statistical software package (R version 3.2.4) using the CDVine package. We fitted the data generated from the scenarios to a set of 24 copulas using maximum likelihood estimates and ranked them based on the Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC).

For each link, the copulas with the largest AIC and BIC values were chosen for that particular link. Figure 6 shows the results of the copula fitting. The Gumbel copula describes the dependence of the marginal distributions of wind power between Pearson and Toronto, the Student-t copula describes the Pearson-BillyBishop pair, and the Joe-Frank copula describes the Pearson-Buttonville pair.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

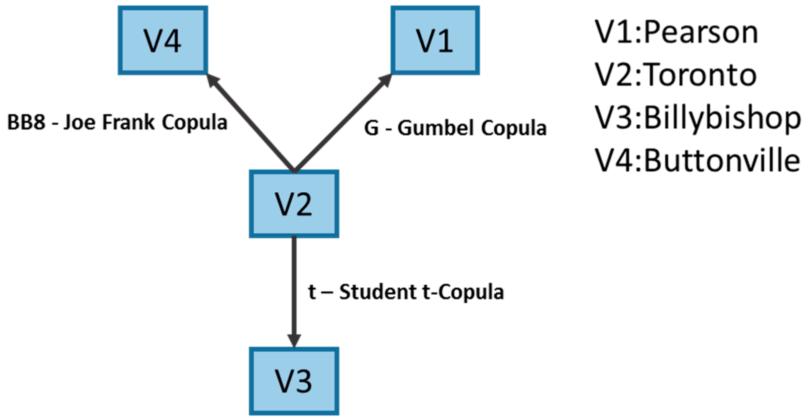


Figure 3: Tree Estimated using Maximum Likelihood Estimates

Node labels represent the four sites in this case study: V1 (Pearson), V2 (Toronto), V3(Billybishop) and V4 (Buttonville) represent the four locations in our study. Link labels represent the copula chosen for modeling the spatial dependence of wind power generation between the sites based on the maximum likelihood estimates: BB8 (Joe-Frank), G (Gumbel) and t (Student-t).

5.5 Optimization of Power Allocation

Once the model of the probability distributions of renewable power is determined (wind, using Kumaraswamy distribution and copulas; and solar, using Kumaraswamy distribution) the next step is to find an optimal allocation (distribution of the total capacity of solar panels and wind turbines among sites) of renewable energy technologies among the four sites (see Section 4.3). Table 4 shows the results of this optimization. For solar power, the distribution of weights across the four sites tends to remain constant even in cases where high reliability is desired (α values range from 0.05 to 0.10). This implies low variations in solar power generation across sites (i.e., stable power source). In contrast, for high reliability cases (α values range from 0.05 to 0.10), wind power distributions across the four sites exhibit more pronounced variations. This implies that different power allocation strategies must be implemented to achieve higher and stable power output from the 4 sites.

Table 3: Wind and Solar Allocation Weightages

α	W_{W_1}	W_{W_2}	W_{W_3}	W_{W_4}	W_{S_1}	W_{S_2}	W_{S_3}	W_{S_4}

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

0.05	0.2556	0.2541	0.2619	0.2284	0.2498	0.2498	0.2505	0.2498
0.10	0.2670	0.2537	0.2381	0.2412	0.2500	0.2500	0.2500	0.2500
0.11	0.2486	0.2507	0.2492	0.2515	0.2500	0.2500	0.2500	0.2500
0.12	0.2499	0.2457	0.2587	0.2458	0.2500	0.2500	0.2500	0.2500
0.13	0.2529	0.2498	0.2538	0.2435	0.2485	0.2546	0.2485	0.2485
0.14	0.2609	0.2402	0.2588	0.2402	0.2486	0.2519	0.2508	0.2486
0.15	0.2551	0.2498	0.2484	0.2467	0.2496	0.2511	0.2496	0.2496

α:Reliability factor, $W_{W_{1,2,3,4}}$: Weightage of allocation for wind power technology at a site and $W_{S_{1,2,3,4}}$: Weightage of allocation for solar power technology at a site, where 1: Toronto, 2: Pearson, 3:Billybishop and 4: Buttonville

To illustrate the advantage of using this technique for allocating renewable energy technologies, we compared the overall power generation (from wind and solar) when equal weightage across the four sites is used, and when optimal allocation is used for $\alpha = 0.05$. A snapshot of power production for a 24-hour period for the Season 3 months is shown in Figure 7. The optimized allocation technique results in significantly higher and accurate overall power output during periods of peak power production.

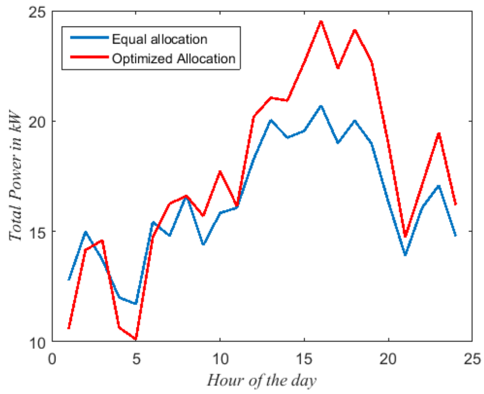


Figure 7: Power production for various allocation schemes during the Season 3

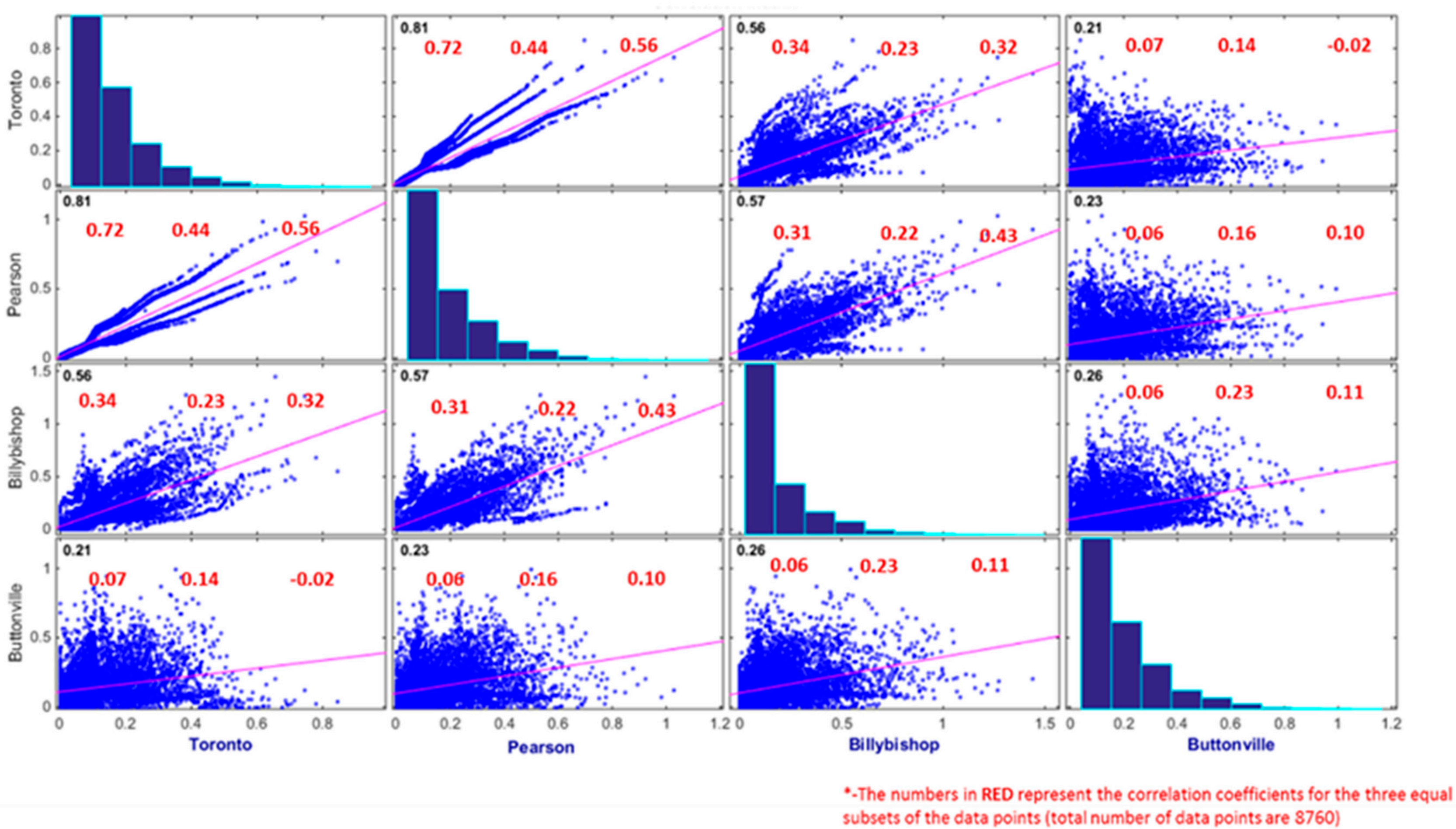


Figure 4: Pair-wise comparison of correlations in wind power data from 4 locations in Central Ontario. In the plots above the figures in black in each sub-plot represents the correlation coefficient for the entire data. Besides Pearson vs Toronto all other datasets are highly non-linear. Pink lines are least-square regression line whose slope is the correlation coefficient for the entire dataset. The panels in the main diagonal represent the histograms of the variables.

444

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017

doi:10.20944/preprints201709.0053.v1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

6. Conclusions

Copulas are one of the most sophisticated tools for modeling the dependence structure of between variables when their correlation is non-linear. In this paper, we present a methodology for modeling the non-linear spatial dependence in wind power generation using *copulas.* We modeled the temporal distributions of both wind and solar power for each individual location using the Kumaraswamy distribution. The data for solar and wind power generated from these probabilistic models is used in an optimization model for obtaining an appropriate allocation of solar and wind power technologies in a spatially dispersed landscape to maximize the overall power output and minimize the effect of random nature of the renewable sources of energy. We find that this approach is useful in increasing the overall reliability of energy production as well as accurate modeling of renewable resources.

Acknowledgement

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). NSERC had no inputs in the study design, model development and implementation, and in the preparation of this manuscript.

References

1. D.E. Newton, World Energy Crisis: A Reference Handbook, ABC-CLIO, 2013 .
2. K. GILLINGHAM and J. SWEENEY, "BARRIERS TO IMPLEMENTING LOW-CARBON TECHNOLOGIES," Climate Change Economics, vol. 03, pp. 1250019, 2012.
3. F. Ribeiro, P. Salgado and J. Barreira, "Engineering Applications of Neural Networks: 13th International Conference, EANN 2012, London, UK, September 20-23, 2012. Proceedings," pp. 254-263, 2012.
4. D.L. King, W.E. Boyson and J.A. Kratochvil, "Analysis of factors influencing the annual energy production of photovoltaic systems," in Photovoltaic Specialists Conference, 2002. Conference Record of the Twenty-Ninth IEEE, pp. 1356-1361, 2002.
5. T. Hammons, "Integrating renewable energy sources into European grids," International Journal of Electrical Power & Energy Systems, vol. 30, pp. 462-475, 2008.
6. O. Grothe and J. Schnieders, "Spatial dependence in wind and optimal wind power allocation: A copula-based analysis," Energy Policy, vol. 39, pp. 4742-4754, sep. 2011.
7. B. Hasche, "General statistics of geographically dispersed wind power," Wind Energy, vol. 13, pp. 773-784, 2010.
8. A. Gerber, M. Qadrdan, M. Chaudry, J. Ekanayake and N. Jenkins, "A 2020Â GB transmission network study using dispersed wind farm power output," Renewable Energy, vol. 37, pp. 124, 2012.
9. E. Kahn, "The reliability of distributed wind generators," Electr.Power Syst.Res., vol. 2, pp. 1, 1979.
10. A. Abdollahi and M.P. Moghaddam, "Investigation of Economic and Environmental-Driven Demand Response Measures Incorporating UC," IEEE Transactions on Smart Grids, vol. 3, pp. 12-25, 2012.
11. D. D. Le, G. Gross and A. Berizzi, "Probabilistic Modeling of Multisite Wind Farm Production for Scenario-Based Applications," IEEE Transactions on Sustainable Energy, vol. 6, pp. 748-758, July. 2015.
12. K. Veeramachaneni, X. Ye and U. O'Reilly, "Statistical Approaches for Wind Resource Assessment," Computational Intelligent Data Analysis for Sustainable Development, pp. 303, 2013.
13. D. P. Kroese, T. Taimre and Z.I. Botev, Handbook of Monte Carlo Methods, Wiley, 2013, .
14. J. S. Benth and F.E. Benth, "Analysis and modelling of wind speed in New York," Journal of Applied Statistics, vol. 37, pp. 893-909, 2010.
15. R. T. Clemen and T. Reilly, "Correlations and Copulas for Decision and Risk Analysis," Management Science, vol. 45, pp. 208-224, 1999.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 September 2017 doi:10.20944/preprints201709.0053.v1

16. H. Valizadeh Haghi, M. Tavakoli Bina, M.A. Golkar and S.M. Moghaddas-Tafreshi, "Using Copulas for analysis of large datasets in renewable distributed generation: PV and wind power integration in Iran," Renewable Energy, vol. 35, pp. 1991-2000, 9. 2010.
17. S. Hagspiel, A. Papaemannouil, M. Schmid and G. Andersson, "Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid," Appl.Energy, vol. 96, pp. 33, 2012.
18. G. Papaefthymiou and D. Kurowicka, "Using Copulas for Modeling Stochastic Dependence in Power System Uncertainty Analysis," Power Systems, IEEE Transactions On, vol. 24, pp. 40-49, Feb. 2009.
19. R.B. Nelsen, An introduction to copulas, New York : Springer, c2006, 2006, .
20. P. Joubert, "Modelling Copulas : An Overview," pp. 1-27, .
21. Yih-Huei Wan, "Wind power plant behaviors: Analyses of long-term wind power data," NREL., Tech. Rep. NREL/TP-500-36551, 2004.
22. Bernhard Ernst, Yih-Huei Wan, Brendan Kirby, "Short-term power fluctuation of wind turbines: Analyzing data from the German 250-MW measurement program from the ancillary services viewpoint," NREL., Tech. Rep. NREL/CP-500-26722, 1999.
23. H. Louie, "Evaluating Archimedean Copula models of wind speed for wind power modeling," in Power Engineering Society Conference and Exposition in Africa (PowerAfrica), 2012 IEEE, pp. 1-5, 2012.
24. P. Pinson and R. Girard, "Evaluating the quality of scenarios of short-term wind power generation," Appl.Energy, vol. 96, pp. 12, 2012.
25. S. Gill, B. Stephen and S. Galloway, "Wind Turbine Condition Assessment Through Power Curve Copula Modeling," Sustainable Energy, IEEE Transactions On, vol. 3, pp. 94-101, Jan. 2012.
26. R. J. Bessa, J. Mendes, V. Miranda, A. Botterud, J. Wang and Z. Zhou, "Quantile-copula density forecast for wind power uncertainty modeling," in PowerTech, 2011 IEEE Trondheim, pp. 1-8, 2011.
27. D. Heinemann, E. Lorenz and M. Girodo, "Forecasting of solar radiation," Solar Energy Resource Management for Electricity Generation from Local Level to Global Scale.Nova Science Publishers, New York, 2006.
28. G. Tina, S. Gagliano and V.A. Doria, "Probability Analysis of Weather Data for Energy Assessment of Hybrid Solar / Wind Power System University of Catania," in 4th IASME/WSEAS Interantional Conference on ENERGY, ECOSYSTEMS and SUSTAINABLE DEVELOPEMNT, pp. 217-223, 2008.
29. A. Sfetsos, "A comparison of various forecasting techniques applied to mean hourly wind speed time series," Renewable Energy, vol. 21, pp. 23-35, 2000.
30. B. Tarroja, F. Mueller and S. Samuelsen, "Solar power variability and spatial diversification: implications from an electric grid load balancing perspective," Int.J.Energy Res., vol. 37, pp. 1002-1016, 2013.
31. A. Seifi, K. Ponnambalam and J. Vlach, "A unified approach to statistical design centering of integrated circuits with correlated parameters," Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions On, vol. 46, pp. 190-196, Jan. 1999.
32. A. Seifi, K. Ponnambalam and J. Vlach, "Optimization of filter designs with dependent and asymmetrically distributed parameters," Journal of the Franklin Institute, vol. 350, pp. 378, 2013.
33. A. Sklar, "Distribution functions of n dimensions and margins," Publications of the Institute of Statistics of the University of Paris, vol. 8, pp. 229-231, 1959.
34. W. Hurlimann, "Fitting bivariate cumulative returns with copulas," Computational Statistics \& Data Analysis, vol. 45, pp. 355-372, 2004.
35. H. Joe, Multivariate Models and Multivariate Dependence Concepts, Taylor \& Francis, 1997, .
36. T. Bedford and R. Cooke, Probabilistic Risk Analysis: Foundations and Methods, Cambridge University Press, 2001, .
37. K. Aas, C. Czado, A. Frigessi and H. Bakken, "Pair-copula constructions of multiple dependence," Insurance: Mathematics and Economics, vol. 44, pp. 182-198, 2009.
38. G. J. Leng, A. Monarque, R. Alward, N. Meloche and A. Richard, "Canada's Renewable Energy Capacity Building Program & Retscreen International," in Proceedings of the World Renewable Energy Congress VII, Cologne, Germany, 2002.
39. Southwest Energy, "Specification Sheet Whisper 500,"