20th International Conference on Renewable Energies and Power Quality (ICREPQ'22)
Vigo (Spain), 27th to 29th July 2022
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.20, September 2022

Modeling the dependency relationship between wind speed and wind power generation: An application of copula theory

Tuany Esthefany Barcellos de Carvalho Silva[1], Reinaldo Castro Souza[1], Fernando Luiz Cyrino Oliveira[1] and Marco Aurélio Sanfins[2]

[1] Industrial Engineering Department
PUC-RJ, Pontifical Catholic University of Rio de Janeiro
Gávea - Rio de Janeiro (Brazil)
Phone/Fax number: +55 21 3527-1286, e-mail: tuanybarcellos@id.uff.br, reinaldo@puc-rio.br, cyrino@puc-rio.br

[2] Department of Statistics
UFF, Federal Fluminense University
São Domingos – Niterói – Rio de Janeiro (Brazil)
Phone/Fax number: +55 21 2629-2005, e-mail: marcosanfins@id.uff.br

Abstract. The concern with global warming and pollution has increased interest in the development of renewable energy sources, which are less aggressive to the environment. Wind energy can provide adequate solutions to the above-mentioned problems. The use of this energy eliminates unwanted waste that is harmful to health and the environment from other energy sources such as coal and nuclear power plants. This work aims to analyse the dependence relationship between wind speed and wind energy production, a rather complex relationship, so this study seeks to understand the stochastic nature of both phenomena. As a methodological tool the copula theory was used. A copula function is used as a general method, which consists of formulating multivariate distributions so that different dependency structures can be represented. That is, the study is based on the analysis and modelling of the dependence between wind speed data and electrical energy generation, for an hourly database of a wind farm in the state of Bahia, collected in the entire year of 2017. Thus, this paper proposes a study aiming the search for the copula function referring to the data in the period mentioned.

Key words. Copula; Simulation; Wind power generation

1. Introduction

Encouraged by a growing demand for energy, aimed at preserving the environment, the search for clean alternatives for obtaining energy has intensified over the years. Some renewable energy sources emerged in the past as a useful alternative to the scarce resources of fossil fuels, one of these renewable sources is wind energy, i.e., that derived from wind speed.

According to the International Energy Agency [1], energy consumption grows intensively and rapidly, reaching approximately 524 quadrillion British thermal units (Btu) in recent years, with an increase of 56% expected by 2040. As a result, numerous issues arise such as the depletion of traditional energy sources, for example hydraulics and fossil fuels. According to Aguilar [2] there are concerns about the environment, and the safety of energy supply systems, which causes a significant growth in the search for wind energy generation. Studies carried out by the Brazilian Wind Energy Association (ABEEólica) [3] show that Brazil has an energy matrix with a vast hydroelectric potential, and numerous renewable energy sources to be explored, hydroelectric plants represent about 58% of the country's energy production. Followed by wind power generation with 10.3

According to the Global Wind Energy Council (GWEC) [4], Brazil is a highly promising country for wind energy generation, research addressing the topic has been presented in recent years in order to contribute and accompany the growth of the generation of this renewable source. As it is, it is a generation dependent on natural factors, the main one being the wind speed, there are numerous elements that directly affect the generation of wind resources. That said, it is a variable with complex stochastic behavior, different studies were carried out in order to explore the wind potential for different regions of Brazil and the world. However, there is a limitation in relation to the necessary information about the variables of interest, often this information is confidential and with strictly limited access, making studies and analysis difficult.

The dependence relationship between wind speed and wind power generation has been the focus of research such as Using Copulas for Modeling Dependence in Wind Power by the author Karakaş [5] and Modeling Wind Energy Using Copula by the author Bahraoui [6]. These studies point to the methodology of Copulas, as something innovative and efficient for the analysis of dependencies of the variables mentioned here.

In this study, the objective is to use the copula theory to deal with the dependence structure, in addition to the correlation between the natural variable of wind speed and wind energy generation. The literature is highly expressive about the individual modeling of these variables; however the multivariate model is not so explored, which makes this research relevant. The implementation of Copula Theory makes it possible to model variables that are not normally distributed, making it possible to robustly obtain the multivariate distribution of the variables of interest, and consequently calculate any probability of this vector as the marginal, conditional, bivariate probabilities, among others. This research seeks to present the copula family corresponding to the aforementioned variables, evidencing their dependence structure. Studies using this theory can provide significant operational, social and economic benefits, since by showing the dependence structure of the variables, it is possible to simulate any scenario of interest, which allows an effective management of wind energy.

2. Copula Theory

This section introduces the definition of copulas as well as their properties. The copula function can be defined as being a link function between the univariate marginal distributions that results in multivariate distributions. In Nelsen [7], this theory is defined by two points of view: the first approaches copulas as being functions that join or couple multivariate distributions to their marginal distributions. The other form of interpretation states that copulas are multivariate distributions whose marginal distributions are uniform in the interval [0,1].

The copula theory, as a function of dependence between random variables, was introduced by Sklar (1959), however, its applications are recent developments in several areas. In his research Sklar studied the joint three-dimensional distribution, and with that he introduced auxiliary functions defined in the unit support, linking the distribution function to its marginals. According to Nelsen [6], a copula is equivalent to a multivariate distribution function with uniform marginals at [0, 1], they contain the entire dependency structure between the observed random variables.

This question has been widely used for bivariate cases, which is the focus of the study, given that the main objective is to observe the dependence structure between two variable variables, namely wind speed and wind power generation. Thus, the copula $\boldsymbol{C(u,v)}$ is a distribution function of two variables u and v. Therefore, a two-dimensional copula for any $0 \leq u1 \leq u2 \leq 1$ and $0 \leq v1 \leq v2 \leq 1$, has the following characteristics:

1. $\boldsymbol{C(u,0) = \int_0^u \int_0^0 c(u,v)dvdu = 0 = \int_0^v \int_0^0 c(u,v)dudv = C(0,v)}$;
2. $\boldsymbol{C(u,1) = u = \int_0^u \int_0^1 c(u,v)dvdu \; e \; C(1,v) = v \int_0^v \int_0^1 c(u,v)dudv}$;
3. $\boldsymbol{C(u_2,v_2) - C(u_1,v_2) - C(u_2,v_1) + C(u_1,v_1) \geq 0}$

Then, features (1) and (2) show that the copula has marginal distributions in [0, 1]. Feature (3) shows that the copula is an increasing function. Thus, the entire concept of copulas above comes from Sklar theorem (1959). In this way, the copula was defined as a real function presenting some specific characteristics. As mentioned before, its probabilistic interpretation will be given through Sklar Theorem.

2.1. Sklar's Theorem

The copula theory is based on Sklar's theorem (1959), according to which a multivariate distribution can be treated by its dependence structure, the copula, and its marginals.

Then, for a joint n-dimensional distribution $\boldsymbol{F}$ of a vector of random variables $\boldsymbol{X = (X_1, \dots, X_n) \in \mathbb{R}^n}$ with marginals $\boldsymbol{F_1(x_1), \dots, F_n(x_n)}$, there is a copula function $\boldsymbol{C}$ such that:

$$\boldsymbol{F(x_1, \dots, x_n) = C\big(F_1(x_1), \dots, F_n(x_n)\big)} \quad \mathbf{(1)}$$

In which, $\boldsymbol{F_1, \dots, F_n}$, are continuous, the function $\boldsymbol{C}$ is unique.

From the above result, an integral probability transformation is performed on the marginal distributions, and thus the copula function is obtained, represented here by $\boldsymbol{C}$.

$$\boldsymbol{C(u_1, \dots, u_n) = F\left(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)\right), \forall \; u = (u_1, \dots, u_n) \in [0,1]^n} \quad \mathbf{(2)}$$

Being, $\boldsymbol{F_i^{-1}}$, a generalized inverse function, that is, $\boldsymbol{F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)}$, are the inverse functions of $\boldsymbol{F_1^{-1}(x_1), \dots, F_n^{-1}(x_n)}$. Throughout this paper it will be possible to observe that the inversion method is one of the main methods for creating copulas.

With this, the multivariate distribution of the transformed variables, is called a copula. The density associated with a copula is defined by the partial derivatives with respect to u of the function $\boldsymbol{C(u_1, \dots, u_n)}$.

$$\boldsymbol{C(u_1, \dots, u_n) = \frac{\partial C^n(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}} \quad \mathbf{(3)}$$

That said, according to Sklar's Theorem (1959), the density of the random vector X can be written as follows:

$$\boldsymbol{f(x_1, \dots x_n) = c(u_1, \dots, u_n) \prod_{i=1}^{n} f_i(x_i)} \quad \mathbf{(4)}$$

In which, $\boldsymbol{u_1 = F_1(x_1; \theta_1), \dots, u_n = F_n(x_n; \theta_n)}$ where $\boldsymbol{\theta_i}$ are the $\boldsymbol{p_i}$ dimensional vectors of the parameters of the marginals, $\boldsymbol{i = 1, \dots, n.}$

2.2. Major Copula Families

According to Sklar's Theorem, a multivariate distribution can be treated through its dependency structure, copula, and its marginals. With this treatment, the high flexibility of modeling via copulas functions can be explained. According to Cherubini [8], this occurs both in empirical applications and from a theoretical point of view. Thus, there are numerous copula functions proposed in the literature. This section presents the main types of bivariate copulas and which families they belong to [9].

- **Elliptical Copulas**

The class of elliptic distributions provides a variety of multivariate distributions, which share numerous properties with the multivariate normal distribution, which provides an ease of simulating data from these distributions [12].

Let $\boldsymbol{H}$ be a bivariate elliptic distribution function with marginals $\boldsymbol{F_X}$ and $\boldsymbol{F_y}$. Then, $\boldsymbol{C}$ is an elliptic copula, with dependence parameter $\boldsymbol{\theta}$, given by the following equation:

$$\boldsymbol{C(u,v) = H\left(F_X^{-1}(u), F_Y^{-1}(v); \theta\right) \quad \forall\, (u,v) \in [0,1]^2} \quad (5)$$

- **Gaussian Copula**

The bivariate Gaussian copula is determined initially taking the bivariate Normal distribution function as a base, after which the variables are replaced by their generalized inverse, as shown above. Thus, this copula is classified as an elliptical copula. Then, the Gaussian copula has the following expression:

$$\boldsymbol{C_G = (u,v;\rho) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)}\ exp\left\{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}\right\} dvdu} \quad (6)$$

Where, $\boldsymbol{(u,v)}$ is a two-dimensional random vector $\in [\boldsymbol{0},\boldsymbol{1}]^2, \boldsymbol{\rho}$ is the linear correlation coefficient, and $\boldsymbol{\phi^{-1}}$ represents the inverse of the univariate standard Normal distribution.
The Gaussian copula has correlation parameter $\boldsymbol{-1 \le \rho \le 1}$ with symmetry; the dependence increases with the value of ρ and has no dependence on the tail, indicating that random variables do not have dependence on events considered extreme.

- **T-Student copulation**

The T-Student copula, like the normal copula, is a symmetrical copula, but its advantage is that it enables efficient modeling of a greater degree of dependence at its ends. Its equation is expressed by:

$$\boldsymbol{C_T(u,v;\rho,\nu) = \int_{-\infty}^{T_\nu^{-1}(u)} \int_{-\infty}^{T_\nu^{-1}(v)} \frac{\Gamma\frac{\nu+2}{2}}{\Gamma\left(\frac{\nu}{2}\sqrt{(\pi\nu^2)(1-\rho^2)}\right)}\ exp\left\{1 + \frac{x^2-2\rho xy+y^2}{\nu(1-\rho^2)}\right\} dvdu} \quad (7)$$

In which, ν represents the degrees of freedom, $\boldsymbol{T^n}$ is the multivariate distribution $\boldsymbol{T_\nu - student}$, whereas $\boldsymbol{T_\nu^{-1}}$ is the inverse of the distribution $\boldsymbol{T_\nu - student}$.

The t-Student copula has correlation parameter $\boldsymbol{-1 \le \rho \le 1}$, and the parameter $\boldsymbol{\nu > 0}$, this being the degrees of freedom, the closer to infinity, the closer the t-copula is to the Gaussian copula. In this way, the implementation of this copula makes it possible to model a greater degree of dependence present in the tails. Dependency coefficients are given by:

$$\boldsymbol{\lambda_l \lambda_u = 2\bar{t}_{\nu+1}\left(\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\right)} \quad (8)$$

Where $\boldsymbol{\bar{t}_{\nu+1} = 1 - t_{\nu+1}}$ indicates symmetry and the dependence coefficient on the tail is increasing in $\boldsymbol{\rho}$ and decreasing in $\boldsymbol{\nu}$.

3. Analysis and Results

The data used for analysis refer to wind speed and wind power generation in the state of Bahia in the year 2017. These data present observations every 10 minutes, totaling more than 50,000 measurements. For the first analysis covering the full year 2017, the hourly average was considered, reducing the base to a total of 8,723 observations. Atypical values resulting from measurement or storage errors were classified as unrealistic wind speeds and negative wind energy. Following the approach taken in Duca [1], seeking not to lose information, it was decided to adopt a strategy for wind speed variable above 30 m/s, limiting them to this value, but for negative power values, the decision was to exclude them, for being incompatible results with the variable, and for missing values it was used the Seasonal adjustment method together with Linear Interpolation.
To understand the stochastic behavior of each variable, a brief descriptive analysis of the data was performed. In Figure 1, one can observe the behavior of the hourly series of power in KW of the wind farm in the state of Bahia.

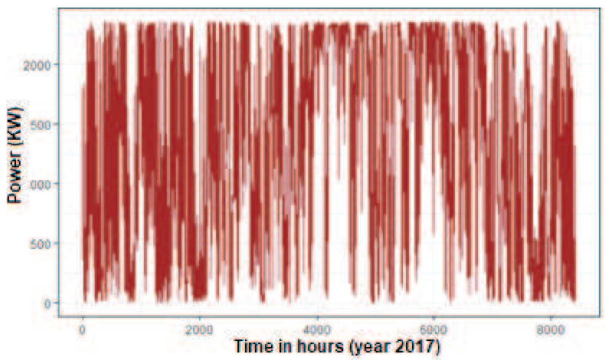

Fig.1. Wind power generation in KW
Source: Own elaboration based on analyzed data.

The figure 2 shows the wind speed series in m/s. It can be seen that the highest point of wind speed occurs in the middle of the year between the months of June and July.

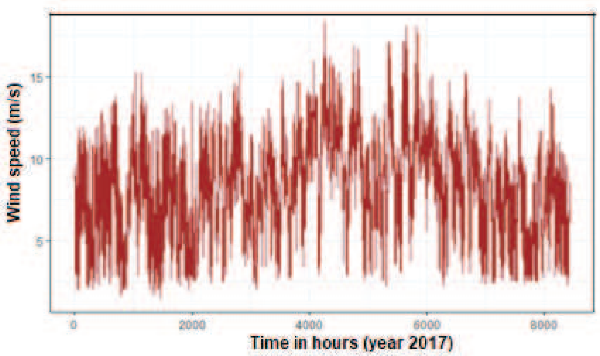

Fig.2. Wind speed in m/s
Source: Own elaboration based on analyzed data.

Table 1 shows the descriptive statistics of the data. It is important to note that in this analysis the hourly average is being addressed, so it is noted that on average the maximum wind speed is approximately 18.32 m/s in this region and the maximum power is 2.361.3 KW.

Table I: Descriptive statistics

Statistical measures	Power (KW)	Speed m/s
Minimum	0	1.53
Median	1,417.3	8.487
Mean	1,346.6	8.356
Maximum	2,361.3	18.326

Source: Own elaboration based on analyzed data.

According to the literature, the Weibull distribution is significantly efficient in modeling wind speed and power generation data. In Figure 3, the density of the data can be observed, it is noted that graphically there is evidence that points to the Weibull distribution as a good fit.

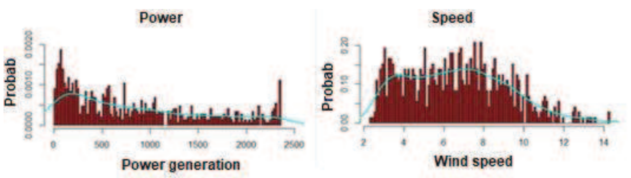

Fig.3. Density of variables
Source: Own elaboration based on analyzed data.

Seeking to verify whether this distribution really fits the data efficiently, the parameters were estimated using the maximum likelihood method. That said, the kolmogorov Smirnov test was applied, where the hypotheses tested were:

$$\begin{cases} \boldsymbol{H_0}: \textbf{\textit{Possibly the data are extracted from the tested distribution}} \\ \boldsymbol{H_1}: \textbf{\textit{Possibly the data are not extracted from the tested distribution}} \end{cases}$$

Considering a $\boldsymbol{\alpha = 0.01}$, test for the Weibull distribution showed a p-value equal to 0.03 for the generation variable and a p-value of 0.09 for the velocity variable. Thus at a significance level of 1%, the null hypothesis, that is, possibly the data are extracted from the tested distribution, in this case the Weibull, is not rejected. It is noted that the result is in line with what is reported in the literature.

After selecting the marginal distributions corresponding to each random variable, the next step is to choose the copula function. The estimation of the copula parameters was performed using the maximum likelihood method, and the one that presented the best fit to the data was selected, to be evaluated by graphic analysis and by other methods of evaluating the fit of copulas to the data of a sample, such as BIC and AIC.

The BIC and AIC results showed significantly low values for the bivariate Tawn type 2 copula when compared to the others, this being the copula that best fitted the hourly database for the full year of 2017. of a flexible class copula of dependency models, consisting of bivariate building blocks. The Tawn copula is an extension of the Gumbel copula with three parameters. For simplicity, two versions of the Tawn copula with two parameters were implemented. Each type has one of the asymmetry parameters set to 1, so that the corresponding copula density is skewed to the left or right (with respect to the main diagonal). In the case of the analyzed data, the copula is Tawn type 2, that is, with density sloping to the right. In figure 4, it is possible to observe the density of the copula found, it presents the first parameter of 12.41, and the second of 0.99, with a tau of 0.91.

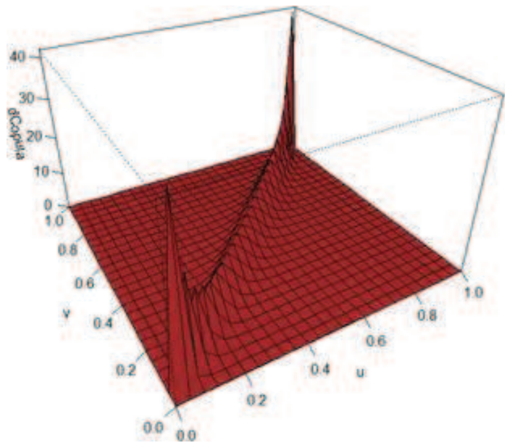

Fig.4. Density of Copula
Source: Own elaboration based on analyzed data.

Therefore, using the copula that best fitted the data, 8,723 observations were simulated. Figure 5 presents the final scatter plot of the data, under the assumption of marginal weibull distributions and the Tawn type 2 copula for the dependence structure, as it can be seen, this copula presents the simulated results close to the real observations. thus evidencing that the simulated values adjust well to the observed values.

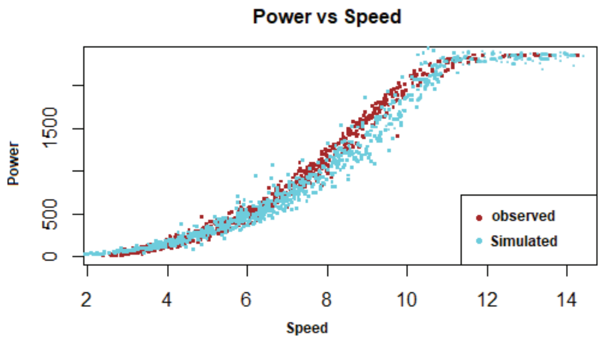

Fig.5. Simulation via copulas (Power vs Speed)
Source: Own elaboration based on analyzed data.

Then, a graphical analysis was performed (Figure 6), where the quality of fit can be observed. The contours of the Tawn Type 2 Copula are compared with the contours of the empirical Copula that was calculated with data of this study. In Figure 6 it is observed that the contours of the copula that was found fit the curves empirical copula, thus showing a good fit.

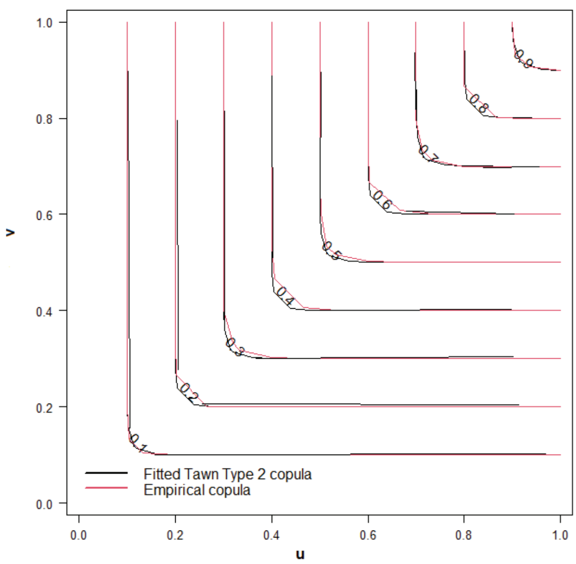

Fig.6. Adjust Copula
Source: Own elaboration based on analyzed data.

Once this is done, the Copula Town Type 2 is the one that best adjusts to the daily data of the variables of wind power generation and wind speed.

4. Conclusion

The present work presented a proposal to implement the copula theory, in order to study the dependence structure between wind speed and wind energy generation. Data from a wind farm in the state of Bahia were analyzed, the period used was from January to December 2017. After performing the kolmogorov Smirnov test, the result showed evidence that the analysed data come from a Weibull distribution, corroborating with literature. It was noted that the variables are highly positively correlated, which is an important factor for the implementation of the copula theory. The copula indicated for the analysed data was the Tawn type 2 copula (R-Vine Copula), these copulas are from a flexible class of dependency models. The Tawn copula is an extension of the Gumbel copula with two parameters, its asymmetry is fixed at 1, with density sloping to the right, it has the first parameter of 12.41, and the second of 0.99, with a tau of 0.91. With this, it is noted that the bivariate copula Tawn type 2 presents a good simulation, adjusting the simulated data in a very similar way to the observed data, it was also possible to graphically observe the quality of the fit through the level curves of the Tawn Type 2 Copula comparing with empirical copulation, indicating a good fit. For future work, it is intended to continue the study of the implementation of the Tawn type 2 copula, seeking to analyses in a more robust way the dependence structure between wind speed and wind energy generation month by month, generating scenario simulations for the variables of interest.

References

[1] World Energy Outlook (EIA) 2018, accessed 2 March 2022, <https://www.iea.org/reports/world-energy-outlook-2018>

[2] Aguilar, S. A; Esteves, G. R. T.; Maçaira, P. M.; Bastos, B. Q.; Cyrino Oliveira, F. L.; Souza, R. C. Wind Power Generation: A Review and Research Agenda (2019). Journal of Cleaner Production, v. 218, p. 850-870. DOI: https://doi.org/10.1016/j.jclepro.2019.02.015

[3] Brazilian Wind Energy Association (ABEEólica) 2020, accessed 2 March 2022, < http://abeeolica.org.br/dados-abeeolica/>

[5] Duca, V. E. A.; Fonseca, T. Cyrino Oliveira, F. C. Joint modelling wind speed and power via Bayesian Dynamical models (2022). Energy. v. 247, 123431.

[4] Global Wind Energy Concil (GWEC) 2021, accessed 2 March 2022, < https://gwec.net/wp-content/uploads/2021/03/GWEC-Global-Wind-Report-2021.pdf>

[5] Karakaaÿ, Ayse Metin. Using copulas for modeling dependence in wind power. Asian Journal of Engineering and Technology, v. 7, n. 1, 2019.

[6] Bahraoui, Zuhair et al. Modeling wind energy using Copula (2018). Open Access Library Journal, v. 5, n. 11, p. 1.

[7] Nelsen, Roger B. Dependence and order in families of Archimedean copulas (2006). Journal of Multivariate Analysis, v. 60, n. 1, p. 111-122.

[8] Cherubini, U., Luciano, E., & Vecchiato, W., Copula methods in finance (2004). John Wiley & Sons.

[9] de Almeida Pereira, G. A., & Veiga, Á. Periodic copula autoregressive model designed to multivariate streamflow time series modelling (2019). Water Resources Management.

[10] Embrechts, P., Lindskog, F., & McNeil, A., Modelling dependence with copulas. Rapport technique.