Application of Copulas in Analysis of Drought and Irrigation

Milan Cisty, Lubomir Celar and Anna Becova
Department of Land and Water Resources Management
Slovak University of Technology
Bratislava, Slovakia
Email: milan.cisty@stuba.sk, lubomir.celar@stuba.sk, anna.becova@stuba.sk

***Abstract*—This paper presents probabilistic analysis of the occurrence of irrigation needs. We have conducted a joint analysis of the severity and duration of the most demanding potential annual irrigation periods by a bivariate copula methodology. The characteristics of these periods are derived from both temperature and precipitation. The maximum annual length of the potential irrigation periods and the corresponding rainfall deficit were inferred from basic climatic variables such as inputs for a two-dimensional probability analysis by a copula methodology. The results of this work indicate the suitability of the proposed methodology for an analysis of irrigation needs with greater benefits than in the case of the usual one-dimensional analysis of individual climatic variables. A case study with the aim of testing the methodology was accomplished in southwest Slovakia, where a frequency analysis of the need for irrigation was estimated. The results indicate, e.g., that every second year, a period can be expected in which temperatures above 25°C occur and which lasts one month with a moisture deficit of about 30 mm. Even more significant periods of drought can be expected, for example, with a 5 or 10-year return period. These phenomena cause significant damage to agriculture yields in the territory investigated, so a requirement for irrigation structures in this area is indicated by the proposed methodology.**

***Keywords*–drought; irrigation; copula; precipitation; temperature**

I. Introduction

Water scarcity and droughts have a direct impact on the inhabitants and various economic sectors of a region which use and depend on water, such as agriculture, tourism, industry, energy or transport. Quantifying the expected probability characteristics of droughts assists in the planning and management of water resources, such as the design and maintenance of irrigation systems. The issue of how to characterize a drought is often dealt with through the help of various drought identification indices [1]. The numerous indices of drought that may be mentioned include, for example, the decile index (DI) [2], the percentage of normality (PN), the standardized precipitation index (SPI) [3], the Palmer PDSI index [4], and the effective drought index (EDI) [5]. Among the above–mentioned drought indices, the standardized precipitation index (SPI) is most frequently used.

Research on the probabilistic characterizations of droughts was formerly conducted using a univariate analysis [1], [6]. However, drought is a multidimensional phenomenon characterized by, for example, its severity, duration and intensity, so it is necessary to examine the properties of dry episodes using multidimensional methods [7]. For this reason, the traditional drought risk assessment based on univariate frequency analyses may lead to erroneous or incomplete conclusions about the occurrence of drought events [8]. Over the last decade, copulas have emerged as a method for addressing multivariate problems in several disciplines. Probabilistic analysis using a copula method has various positive features; the main one is that it does not assume that the variables have the same types of probability distribution functions [9]. Copulas have been adopted for hydrological studies of multivariate flood frequency analyses [10]–[14] and rainfall frequency analyses [15]–[17].

In this paper, we have chosen a different approach, which is intended for an analysis of irrigation needs and is oriented towards the point of view of the necessity for the construction of an irrigation system in a given area. Various drought indices used in previous studies are designed for the identification of drought months, not for the identification of the necessity to irrigate, which should be analyzed on a timescale of days or weeks, not months. We applied a novel approach and directed our research towards an analysis of the severity and duration of the most demanding annual potential irrigation periods.

Although previous studies have used multivariate analysis, they usually only investigated the lack of precipitation, - e.g., the duration, severity or intensity of dry periods [3], [6], [13]. In the present paper, both the temperature and precipitation are included in the analyses as will be explained in the methodology part. To evaluate the expected occurrence of periods with an increased need for irrigation, a two–dimensional analysis has been applied to the distribution of two variables, which together characterize expectations about the occurrence of episodes which require irrigation. The first variable is the length of the maximal potential irrigation period mentioned, i.e., the maximum number of consecutive summer days in a year. The second variable is the rainfall deficit during this time interval.

In Section 2 of the paper, a description of the area and data studied follows; then in Section 3, assessments of both one–dimensional and bivariate probability distribution functions are described. The results are presented in tabular and graphic forms in Section 4, and the paper ends with the conclusions (Section 5) from the research presented.

II. Study area and data

The analysis was carried out on an agricultural area in Slovakia with a warm and relatively dry climate–the area of the Danubian Lowland, namely, its central part around the municipality of Hurbanovo (Figure 1). The weather in this area

 ISBN: 978-1-61208-419-0

is a transition between oceanic and terrestrial influences. The annual average temperature of a substantial part of the lowland ranges between 9°C and 10°C. In terms of precipitation, it is the driest part of Slovakia with an average annual rainfall of 550 mm to 650 mm.

The analyses were accomplished using climatic data from the period 1930–2013. In the analyses, daily temperature and precipitation data were used.

Figure 1. Study area location

Basic climatic variables (temperature and precipitation) were used to determine the two derived and actually used variables which, in this paper, characterize dry and hot periods requiring irrigation. As already mentioned, such a period is defined by its duration and the rainfall deficit with respect to the normal period (1960–1990). For each year, the hot and dry periods that lasted the longest were identified. The duration was derived from the number of consecutive days with temperatures above 25°C. The hot period identified was extended by precipitation-free days before and after it. In the following, this variable is referred to as the maximum annual length of the potential irrigation period. Although plants have, to a certain extent, the ability to adapt to periods with a lack of moisture, a long duration of this period usually requires irrigation if a reduction in yields is to be prevented, especially if these periods occur in the important growth stages of plants.

III. Methodology

The procedure used to achieve the objective of this paper, e.g., a joint analysis of the two variables introduced in the previous text, was the following: 1) the preparation of the datasets of the variables investigated; 2) a verification of the dependence and relationships between the variables; 3) the identification of one-dimensional distributions of the selected variables; 4) identification of the expected class of the copulas forming a two-dimensional probabilistic dependence; 5) the determination of the copula parameters, e.g., the fitting and evaluation of the best suitable copula; and 6) the specification of the return periods with critical temperature and rainfall-deficit characteristics.

A. Specification of the one-dimensional probability distribution functions

One-dimensional distributions are required to determine the probabilistic characteristics of the individual variables that are used to describe the properties of dry and hot periods with potential irrigation requirements. In the context of this paper, they serve as the means to determine the so-called marginal functions needed in defining a two-dimensional probability.

The fitting process can be divided into three steps: 1) selecting an appropriate probability distribution function; 2) determining its parameters; 3) verifying the quality of the fitting by the appropriate statistical characteristics.

A preliminary selection of the candidate probability distribution functions was performed based on a data analysis employing descriptive statistics and graphic techniques as well as on the existing literature on the probability distribution fittings of the variables describing a drought [1], [18]–[20].

The parameters of the probability functions were determined using the maximum likelihood method (MLE) [21].

The quality of the selection of the type of distribution functions and its fitting could be evaluated by the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) [21]. The quality of the fitting is also verified in the paper using the Kolmogorov-Smirnov test and the Anderson-Darling test [21].

B. Joint probability distribution specification using copulas

When modelling with copula operators, it is first necessary to conduct certain tests of the relationship between the variables under study. The application part of this paper uses the Kendall correlation test and a multivariate test of independence based on an empirical copula process which was proposed by [22] and is often used to test independence in copula modelling (for example, [23] or [24]).

The advantages of copulas for constructing multivariate distributions lie in the fact that the multidimensional modelling of a distribution can be decomposed into a separate determination of the one-dimensional marginal functions of the variables examined and a separately conducted search of the dependencies between them using copulas [25].

The essence of modelling a two-dimensional relationship between two variables by means of copulas is based on Sklar's theorem (1959), which mathematically justifies the intuitive principle specified in the previous paragraph.

For two variables, Sklar's theorem [25] states that if $F_{X,Y}(x, y)$ is a joint distribution function of bivariate random variables (X, Y) with marginal distributions $F_X x$ and $F_Y y$ respectively, then there exists a copula function $C(.)$ such that:

$$F_{X,Y}(x, y) = C(F_x(x), F_Y(y)) \qquad (1)$$

If both $F_X(x)$ and $F_Y(y)$ are continuous distributions, then this copula is unique for the particular joint distribution.

To perform probabilistic analyses of a drought, it is necessary to select an appropriate copula function on the basis of certain principles. There are many varieties of copulas which, based on common features, belong to several classes. Among the most widely used are included, for example, elliptical copulas, the Ali-Mikhail-Haq (AMH) copula, the Clayton, Frank, Galambos, Gaussian, Gumbel-Hougaard, Joe and Plackett copulas [25].

The choice of the proper copula is based on various factors, such as the scope of the dependence to be described by the copula. In this paper, we selected one parametric copula, specified by parameter Θ.

For each choice it is necessary to optimally determine the Θ copula parameter. In this paper, we used the values of the em-

 ISBN: 978-1-61208-419-0

pirical marginal one-dimensional distribution functions so that the choice of the family of the parametric marginal distribution does not affect the search for the Θ copula parameter. This includes finding the ranks of the individual values of the data and scaling them to the interval (0, 1). The actual calculations were performed using the copula and acopula packages for the R language [26], [27]. There are several methods of copula fitting such as the maximum likelihood method, the inversion of Kendall's τ, and the inversion of Spearman's ρ. This work uses the maximum likelihood method.

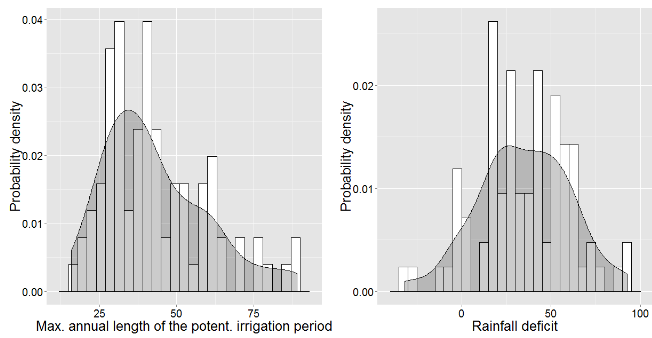

Figure 2. Histogram and kernel-estimated nonparametric probability distribution functions

The best fitting of a copula model to describe the relationship between both characteristics of a drought was verified using multiple criteria, namely by application of the AIC and the BIC. There are several alternatives for testing the goodness of fit, and active research is being done in this field. We used the test referenced in [28], which also provides an overview of the other tests for this purpose. In this procedure a nonparametric empirical copula was computed and compared with the values of the parametric copulas. The parametric copula that was closest to the empirical copula was defined as the most appropriate choice [29].

IV. Results

The analyzed data, which are the two time series derived from the temperatures and precipitation at the Hurbanovo climatological measuring station, have been described in the "Study area and data" section. Figure 2 contains a histogram with kernel–estimated nonparametric probability distributions of both variables. On that basis, it could be expected that, due to the different shapes of the kernel distribution function, these two variables will be described by different parametric probability distribution functions. This means that the joint probability of these variables should be determined by a copula methodology (as opposed to standard multivariate probability distributions, which assume the same distribution for all jointly evaluated variables).

TABLE I. The Coefficients of the Correlation Between the Variables Studied

Correlation coefficient	
Pearson	0.392
Kendall	0.287
Spearman	0.399

The two data series examined were successfully tested for independence using a nonparametric test based on an empirical copula according to [22], in which the output statistics of Harald Cramér are used [30]. The Mann-Kendall

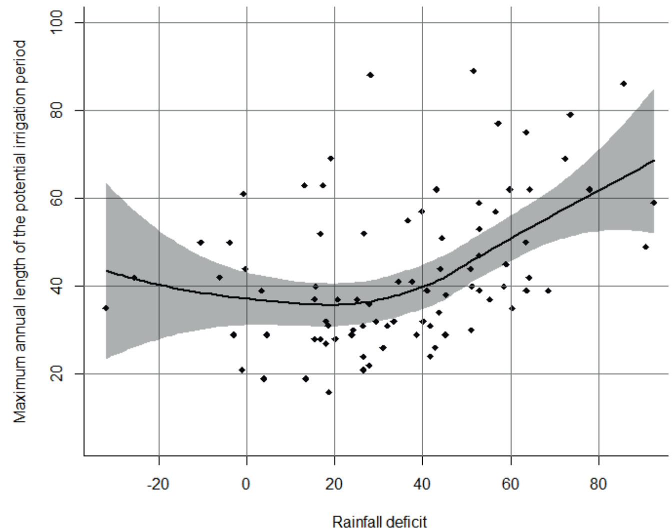

Figure 3. Data compared by a nonparametric Loess regression line

test for correlation was also carried out with the alternative hypothesis that the true τ is greater than zero, which was accepted (for the null hypothesis of the equality of the correlation to zero) on the basis of the p-value = 6.37810^{-05}.

TABLE II. Evaluation of the Distribution Function for the Rainfall Deficit in the Maximum Potential Annual Irrigation Period

Evaluation of the rainfall deficit fitting						
	Statistical indicator				p-value	
Distribution:	KS Test	AD Test	AIC	BIC	AD Test	KS Test
GEV	0.10	1.34	799	804	0.22	0.30
Normal	**0.06**	**0.23**	**784**	**789**	**0.98**	**0.92**
Log-Logistic	0.07	0.32	786	791	0.92	0.79
Cauchy	0.09	1.51	817	822	0.17	0.52
Gumbel	0.10	1.34	799	804	0.22	0.30

The values of the correlation are shown in Table 1; in Figure 3, the two variables are plotted with a graphic representation of the non-linear dependence using a Loess curve with a confidence level of 0.95. Both the table and the picture show that the relationship between the variables is not very strong but can be detected.

The results of this testing could be summarized as follows: the simultaneous probability of these two variables will not be equal to their product, but should be modelled using joint bivariate probability distributions (copulas).

The next step in the analysis is to determine the one-dimensional probability distribution functions of both variables studied. Based on the above reasons, the probability distribution functions according to Table 2 were selected as the candidate probability distributions for the rainfall deficit. Table 2 also includes an evaluation using the methods described in the methodological section of this article. Table 3 serves the same purpose, except it is for the "maximum annual length of the potential irrigation period" variable. In both tables, the bold typeface indicates the selected distribution based on the values of the statistical indicators and p-values; a normal distribution was chosen for the rainfall deficit, and the logarithmic Pearson type III distribution was chosen for the length of the maximum potential annual irrigation period.

To construct an associated probability distribution function, several one-parameter copulas were evaluated (Gumbel, Clayton, Frank, Ali-Mikhail-Haq, Joe, Normal, t-copula, Plackett and Hussler-Reiss). The results and evaluation of the copula function fitting calculations are shown in Table 4.

The main criterion for the selection of a suitable copula operator from Table 4 was the p-value obtained from the goodness-of-fit test according to [28], which uses a parametric bootstrap and is mentioned in the methodological part about fitting the copula function. The copula parameter was determined in the test using the inverse Kendall's τ method. The values of the AIC and BIC served as the auxiliary criteria. The table shows that the Gumbel, Joe, and Husler-Reiss copulas are the admissible copula functions. On the basis of the highest p-value, the Joe copula was selected to describe the dependence.

TABLE III. EVALUATION OF THE DISTRIBUTION FUNCTION FOR THE LENGTH OF THE MAXIMUM POTENTIAL ANNUAL IRRIGATION PERIOD

Evaluation of the rainfall deficit fitting						
	Statistical indicator				p-value	
Distribution:	KS Test	AD Test	AIC	BIC	AD Test	KS Test
Gumbel	0.08	0.39	700	705	0.86	0.72
Pearson III	0.08	0.55	701	706	0.70	0.59
Gamma	0.08	0.55	701	706	0.70	0.59
Logn w.3p	0.06	0.28	700	705	0.95	0.95
Log-P.III	**0.06**	**0.28**	**705**	**705**	**0.95**	**0.88**

TABLE IV. EVALUATION OF FITTING THE COPULA OPERATOR

Copula class	AIC	BIC	θ parameter	p-value
Gumbel	-12.88	-10.45	1.400	0.120
Clayton	-4.10	-1.67	0.800	0.001
Frank	-12.08	-9.65	2.757	0.049
AMH	-8.13	-5.70	0.915	0.013
Joe	**-12.59**	**-10.16**	**1.719**	**0.229**
Normal	-11.88	-9.46	0.434	0.027
t-copula	7.36	-4.94	0.434	0.041
Plackett	-11.58	-9.15	3.723	0.041
Husler-Reiss	-13.11	-10.69	1.074	0.137

In water resources management, the results of a probabilistic analysis are usually expressed using the concept of return periods. These correspond to the long-term average time between two successive occurrences of a certain event. For the problem considered in this paper, the return values of the two variables which are useful in assessing the need for irrigation and for dimensioning the components of an irrigation project were evaluated.

A multivariate analysis of a phenomenon in comparison with a one-dimensional analysis has a distinctive feature in that a combination of variables with different values can lead to the same joint probability and, consequently, to the same return period. In Figure 4, this feature is expressed by means of a contour plot. The chart in this figure shows the return periods of the two variables studied simultaneously in this paper, using the contour lines of the return periods. Some selected results of the joint analysis are also shown in Table 5. The probability of variable values occurring simultaneously can be expressed by the following relationship [15]:

$$T_{x,y} = \frac{1}{1 - F_{(x)} - F_{(y)} + C\left(F_{(x)}.F_{(y)}\right)} \quad (2)$$

where $F_{(x)}$ and $F_{(y)}$ are one-dimensional probability distribution functions, and $C(F_{(x)}, F_{(y)})$ represents a joint distribution function based on the Joe copula.

TABLE V. JOINT RETURN PERIODS OF SOME SELECTED VALUES OF THE VARIABLES INVESTIGATED

Probability	One-dimensional return period	Length of the potential irrigation period	Rainfall deficit	Joint return period
	years	*days*	*mm*	*years*
0.5	2	39	35	4
0.8	5	55	56	9
0.9	10	66	67	20
0.95	20	78	76	39

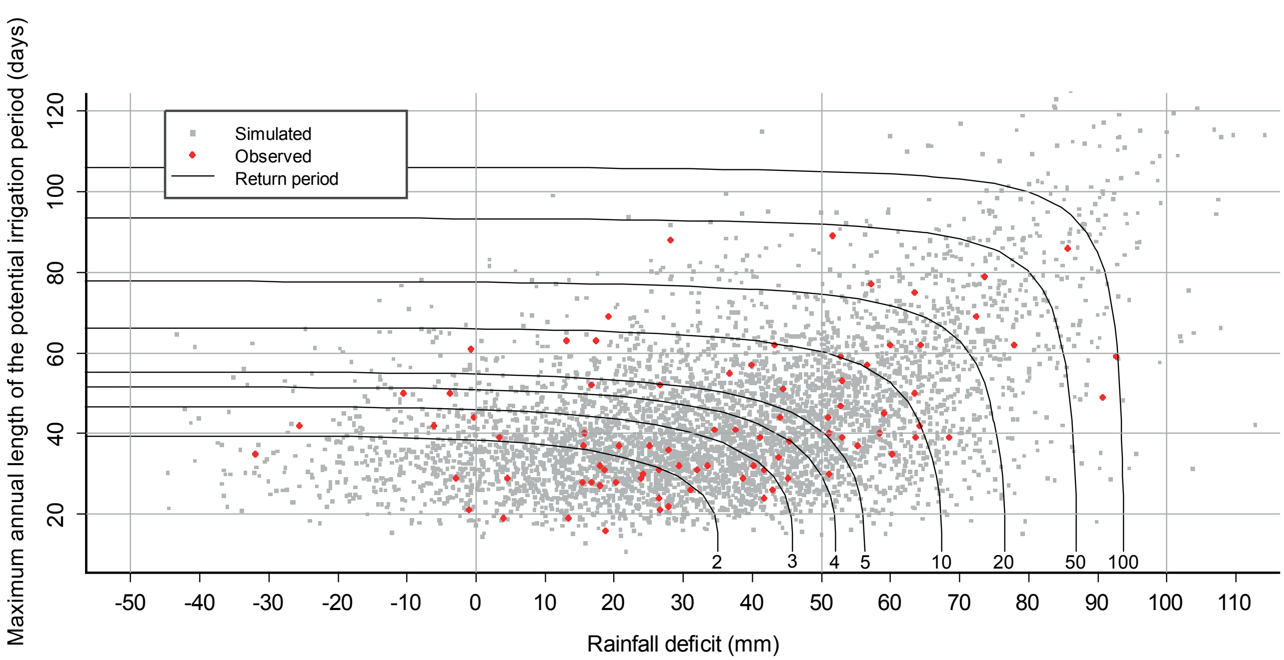

Figure 4. Joint return period of the variables studied for the simultaneous attainment of the corresponding values expressed by the return period contours (in years)

 ISBN: 978-1-61208-419-0

V. Conclusion

The results of this work (Table 5, Figure 4) indicate that in the context of the case study accomplished in south-west Slovakia, the need for irrigation occurs very often. Every second year, for example, a period can be expected in which temperatures above 25°C occur, and a dry period usually lasts one month with a moisture deficit of about 30 mm. Months of the growing season with rainfall totals smaller than 50 mm are considered to be those with irrigation needs. A precipitation of 80 mm in such a period (which would be needed to maintain this limit) occurs with a probability in the upper quartile, i.e., it is very rare. Even more significant periods of drought can be expected, for example, with a 5 or 10-year return period. These phenomena result in significant damage to agriculture yields, which, as is often declared in the domestic water management community, are greater than the investment needed for the reliable maintenance or reconstruction of irrigation systems.

Acknowledgment

This work was supported by the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences, Grant No. 1/0665/15.

References

[1] P. Angelidis, F. Maris, N. Kotsovinos, and V. Hrissanthou, "Computation of drought index SPI with alternative distribution functions," Water Resour Manag, vol. 26, 2012, pp. 2453–2473.

[2] A. Shaban, "Indicators and aspects of hydrological draught in Lebanon," Water Resour Manag, vol. 23, 2009, pp. 1875–1891.

[3] F. Yusof, F. Hui-Mean, J. Suhaila, and Z. Yusof, "Characterisation of drought properties with bivariate copula analysis," Water Resour Manag, vol. 27, 2013, pp. 4183–4207.

[4] W. Palmer, "Meteorological drought." US Department of Commerce, Weather Bureau, Tech. Rep. 45, 1965, [retrieved: June, 2015]. [Online]. Available: http://www.ncdc.noaa.gov/temp-and-precip/drought/docs/palmer.pdf

[5] S. Morid, V. Smakhtin, and M. Moghaddasi, "Comparison of seven meteorological indices for drought monitoring in Iran," Int J Climatol, vol. 26, 2006, pp. 971–985.

[6] J. Santos, M. Portela, and I. Pulido-Calvo, "Regional frequency analysis of droughts in Portugal," Water Resour Manag, vol. 25, 2011, pp. 3537–3558.

[7] J. Zhai, B. Liu, H. Hartmann, B. Su, T. Jiang, and K. Fraedrich, "Dryness/wetness variations in China during the first 50 years of the 21st century," Hydrol Earth Syst Sci Discuss, vol. 6, 2009, pp. 1385–1409.

[8] L. Vergni, F. Todisco, and F. Mannocchi, "Analysis of agricultural drought characteristics through a two–dimensional copula," Water Resour Manag, vol. 29, 2015, pp. 2819–2835.

[9] G. Salvadori and C. D. Michele, "Frequency analysis via copulas: Theoretical aspects and applications to hydrological events," Water Resour Res, vol. 40, 2004, W12511.

[10] A. Favre, S. E. Adlouni, L. Perreault, N. Thiemonge, and B. Bobée, "Multivariate hydrological frequency analysis using copulas," Water Resour Res, vol. 40, 2004, pp. 1–20.

[11] M. J. Reddy and P. Ganguli, "Bivariate flood frequency analysis of upper Godavari River flows using Archimedean copulas," Water Resour Manag, vol. 26, no. 14, 2012, pp. 3995–4018.

[12] J. Shiau and R. Modarres, "Copula-based drought severity-duration-frequency analysis in Iran," Meteorol Appl, vol. 16, 2009, pp. 481–489.

[13] N. Bezak, M. Miko, and M. Šraj, "Trivariate frequency analyses of peak discharge, hydrograph volume and suspended sediment concentration data using copulas," Water Resour Manag, vol. 28, 2014, pp. 2195–2212.

[14] S. Kao and R. Govindaraju, "A bivariate frequency analysis of extreme rainfall with implications for design," J Geophys Res-Atmos, vol. 112, 2007, D13119.

[15] C. D. Michele and G. Salvadori, "A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas," J Geophys Res-Atmos, vol. 108, 2003, pp. 1–9.

[16] L. Zhang and V. Singh, "Bivariate rainfall frequency distributions using Archimedean copulas," J Hydrol, vol. 332, 2007, pp. 93–109.

[17] U. Rauf and P. Zeephongsekul, "Analysis of rainfall severity and duration in Victoria, Australia using non-parametric copulas and marginal distributions," Water Resour Manag, vol. 28, 2014, pp. 4835–4856.

[18] X. Lana, M. Martnez, A. Burgueño, C. Serra, J. Martín-Vide, and L. Gómez, "Distributions of long dry spells in the Iberian peninsula, years 1951-1990," Int J Climatol, vol. 26, 2006, pp. 1999–2021.

[19] Q. Zhang, V. Singh, Y. Chen, and M. Xiao, "Regional frequency analysis of droughts in China: A multivariate perspective," Water Resour Manag, vol. 29, 2015, pp. 1767–1787.

[20] S. Nadarajah, "A bivariate distribution with gamma and beta marginals with application to drought data," J Appl Stat, vol. 36, 2009, pp. 277–301.

[21] A. Hayter, Probability and Statistics for Engineers and Scientists. Cengage Learning, 2006.

[22] C. Genest and B. Rémillard, "Test of independence and randomness based on the empirical copula process," Test, vol. 13, 2004, pp. 335–369.

[23] X. Wang, M. Gebremichael, and J. Yan, "Weighted likelihood copula modeling of extreme rainfall events in Connecticut," J Hydrol, vol. 390, 2010, pp. 108–115.

[24] T. Bacigal, V. Jagr, and R. Mesiar, "Non-exchangeable random variables, Archimax copulas and their fitting to real data," Kybernetika, vol. 47, 2011, pp. 519–531.

[25] R. Nelsen, An introduction to copulas. Springer, 2006.

[26] T. Bacigal, "R Package to handle Archimax or any user-defined continuous copula construction: acopula. In: Bustince H, Fernandez J, Mesiar R, Calvo T (eds) Aggregation functions in theory and in practise," Advances in intelligent systems and computing, vol. 26, 2013, pp. 75–84.

[27] I. Kojadinovic and J. Yan, "Modeling multivariate distributions with continuous margins using the copula R package," J Stat Softw, vol. 34, 2010, pp. 1–20.

[28] C. Genest, B. Rémillard, and D. Beaudoin, "Goodness-of-fit tests for copulas: A review and a power study," Mathematics and Economics, vol. 44, 2009, pp. 199–213.

[29] C. Genest and A. Favre, "Everything you always wanted to know about copula modeling but were afraid to ask," J Hydrol Eng, vol. 12, 2007, pp. 347–368.

[30] H. Cramér, "On the composition of elementary errors," Scandinavian Actuarial Journal, vol. 11, 1928, pp. 141–180.

 ISBN: 978-1-61208-419-0