, Nr. , 2011.

Trandafir, Romica and Ciuiu, Daniel and Drobot, Radu

Technical University of Civil Engineering, Bucharest, Romania, Romanian Institute for Economic Forecasting; Technical University of Civil Engineering, Bucharest, Romania, Technical University of Civil Engineering, Bucharest, Romania

August 2010

THE UTILIZATION OF COPULA IN HYDROLOGY

Romica Trandafir, professor, Technical University of Civil Engineering, Bucharest, Mathematics and Computer Science Department, e-mail: romica@utcb.ro
Daniel Ciuiu, assistant professor, Technical University of Civil Engineering Bucharest, Mathematics and Computer Science Department; associate researcher, Romanian Institute for Economic Forecasting, e-mail: dciuiu@yahoo.com
Radu Drobot, professor, Technical University of Civil Engineering, Bucharest, Hydrology Department, e-mail: drobot@utcb.ro

Abstract

In this paper the parameters of the generalized Pareto cumulative distribution functions of the marginals and the parameter θ of the connecting copula for the water maximum discharges and water volumes are obtained. The isolines for $C(F(x),G(y))=1-\varepsilon$ and for $C^*\left(\overline{F}(x),\overline{G}(y)\right)=\varepsilon$ will be drawn.
Keywords: copula, isolines, water discharges and volumes

1. Introduction

In hydrology the floods are usually characterized by the maximum discharge corresponding to a certain return period (or to the corresponding probability of exceedance). The value of the maximum discharge is used to establish the crest of the dykes, the overtopping being the most frequent cause of dykes failure.
Still, in an important number of cases the dykes are destroyed due to internal erosion of the dyke itself or of the foundation**;** finally, the lost of stability of the inner or outer slope is another cause of the dykes failure. In these cases, the flood duration is the triggering factor of the dyke failure. Given the maximum discharge, the flood duration depends directly on the flood volume.
Although in many cases the maximum discharge and the flood volume are considered as being independent statistical variables, in fact they can be treated as bi-variates. A copula couples the marginal cumulative distribution functions (cdfs) to obtain multivariate distribution functions based on the theorem of Sklar [12]. Initially the copulas were used in the theory of probabilistic metric spaces, but they are now widely applied in many fields as: econometrics and finance, political science, biostatistics, medical research, hydrology etc.
The paper is organized as follows: in the following we present some definitions and results about copulas, the section 2 presents different methods to estimate the parameter θ of the connecting copula. A numerical application of the presented method is given in section 3. Some suitable conclusions are presented in last section.

Definition 1 ([10,7,11]). *A copula is a function* $C:[0,1]^n \to [0,1]$ *such that*
1) If there exists i *such that* $x_i=0$ *then* $C(x_1,\ldots,x_n)=0$.

2) If $x_j = 1$ *for all* $j \neq i$ *then* $C(x_1,\ldots,x_n) = x_i$.

3) C *is increasing in each argument.*

The following theorem (see [10,7,11]) represents the basis of the multivariate cumulative distribution functions using copulas:

Theorem 1 (Sklar). *Let* X_1, $X_2,\ldots,$ X_n *be random variables with the cumulative distribution functions* F_1, F_2 ,..., F_n, *and the common cdf* $H(x_1,\ldots,x_n) = P(X_1 \leq x_1,\ldots,X_n \leq x_n)$. *In this case there exists a copula* $C(u_1,\ldots,u_n)$ *such that* $H(x_1,\ldots,x_n) = C(F_1(x_1),\ldots,F_n(x_n))$. *The copula* C *is well defined on the Cartesian product of the images of the marginals* F_1, $F_2,\ldots,$ F_n.

An important class of copulas having many practical applications is the family of Archimedean copulas.

Definition 2 ([10,13,14]). *If* $n = 2$ *the copula* C *is Archimedean if* $C(u,u) < u$ *for any* $u \in (0,1)$ *and* $C(C(u,v),w) = C(u,C(v,w))$ *for any* $u,v,w \in [0,1]$. *If* $n > 2$ *the copula* C *is Archimedean if there exists a* $n-1$ *Archimedean copula* C_1 *and a* 2*-Archimedean copula* C_2 *such that* $C(u_1,\ldots,u_n) = C_2(C_1(u_1,\ldots,u_{n-1}),u_n)$.

Consider a function $\varphi : [0,1] \to R$ decreasing and convex with $\varphi(1) = 0$ and its pseudo-inverse g ($g(y)$ has the value x if there exists x such that $\varphi(x) = y$ and 0 in the contrary case). We know (see [5,10]) that a copula C is Archimedean if and only if there exists a function φ as above such that for any $x, y \in [0,1]$ we have

$$C(x,y) = g(\varphi(x) + \varphi(y)). \tag{1}$$

For any $n-$copula C we have (see [1])

$$W(x_1,\ldots,x_n) \leq C(x_1,\ldots,x_n) \leq \min(x_1,\ldots,x_n), \text{ where} \tag{2}$$

$$W(x_1,\ldots,x_n) = \sum_{i=1}^{n} x_i - n + 1 \tag{2'}$$

is the lower Fréchet bound, and min is the upper Fréchet bound.

In [2] we have used in order to generate the copulas W and $\min$ the fact that if X and Y are connected by the copula min there exists a function $f : [0,1] \to [0,1]$ increasing such that $f(X) = Y$. If the copula is W then f is decreasing, and if the copula is $\Pr od$ the variables are independent (see [10,7]). The Fréchet copulas (see [1,10,7]) are generated by the mixture method (see [2]).

In [4] there are found analytical formulae for the copulas that connect the number of customers in a Gordon and Newell queuing network, and their corresponding Spearman ρ and Kendall τ. This value is (see [7]):

$$\begin{gathered}\tau = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0) = \\ 4\int_0^1\int_0^1 C(u,v)\tfrac{\partial^2 C}{\partial u \partial v} dudv - 1 = 1 - 4\int_0^1\int_0^1 \tfrac{\partial C}{\partial u} \cdot \tfrac{\partial C}{\partial v} dudv.\end{gathered} \tag{3}$$

Sometimes we need the overlay probabilities, and we need in this case the notion of co-copula

$$C^*(u_1,\ldots,u_n) = C(1-u_1,\ldots,1-u_n) + \sum_{i=1}^{n} u_i - n + 1. \tag{4}$$

2. Estimation of parameters

The marginal parameters are estimated using the moments' method (see [9]).

First we estimate τ using the empirical probabilities in the above formula, and next we compute the last term: we find τ in function of θ. For instance, in the case of Farlie-Gumbel-Morgestern copula (see [7,10,8]) we find

$$\tau = \frac{2\theta}{9}, \tag{5}$$

and from here

$$\theta = \frac{9\tau}{2}. \tag{5'}$$

For the Fréchet family the copula is a mixture between the upper Fréchet bound, min and the copula product (the independence case) with the weights θ, respectively $1-\theta$. Due to the fact that in the min case we have $\tau = 1$, and in the product case we have $\tau = 0$ we obtain

$$\theta = \tau. \tag{6}$$

When the copula is Archimedean and we know the function φ in (1) we use the variables change $x = \varphi(u)$ and $y = \varphi(v)$, and finally we obtain

$$\tau = 1 - 4 \cdot \int_0^{\varphi(0)} \int_0^{\varphi(0)} \left(g'(x+y)\right)^2 dxdy. \tag{3'}$$

In the case of Clayton family we have

$$C(u,v) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-\frac{1}{\theta}}. \tag{7}$$

From $\frac{\frac{\partial C}{\partial u}}{\frac{\partial C}{\partial v}} = \frac{\varphi'(u)}{\varphi'(v)}$ we obtain first $\varphi'(u) = -u^{-\theta-1}$, and from here

$$\varphi(u) = \frac{u^{-\theta} - 1}{\theta}, \text{ and} \tag{7'}$$

$$g(w) = (\theta w + 1)^{-\frac{1}{\theta}}. \tag{7''}$$

Using (3') we obtain

$$\tau = \frac{\theta}{\theta + 2}, \tag{8}$$

and from here

$$\theta = \frac{2 \cdot \tau}{1 - \tau}. \tag{8'}$$

Other family of Archimedean copulas presented in [5,6,7] and simulated in [2] is the Frank family. In this case for $\theta \in \mathrm{R}^*$ we have

$$C(u,v) = -\frac{1}{\theta} \cdot \ln\left(\frac{e^{-\theta(u+v)} - e^{-\theta u} - e^{-\theta v} + e^{-\theta}}{e^{-\theta} - 1}\right). \tag{9}$$

We obtain also the copula *Prod* for $\theta = 0$ and the copula min for $\theta \to \infty$. For $\theta \to -\infty$ we obtain the lower Fréchet bound W.

From $\frac{\frac{\partial C}{\partial u}}{\frac{\partial C}{\partial v}} = \frac{\varphi'(u)}{\varphi'(v)}$ we obtain first $\varphi'(u) = \frac{\theta e^{-\theta u}}{e^{-\theta u} - 1}$, and from here

$$\varphi(u) = \ln \frac{1 - e^{-\theta}}{1 - e^{-\theta \cdot u}}, \text{ and} \tag{9'}$$

$$g(w) = -\frac{1}{\theta}\ln\left(\gamma e^{-w}+1\right),\ \textit{where}\ \gamma = e^{-\theta}-1. \tag{9''}$$

For this family we obtain

$$\tau = 1 - 4\cdot I\text{, where} \tag{10}$$

$$I = \frac{1}{\ln^2(1+\gamma)}\cdot\int_0^{\gamma}\frac{\ln(1+x)}{x}+\frac{1}{1+x}dx. \tag{10'}$$

In the case $\theta \neq 0$ we multiply the relation (10') by $\ln^2(1+\gamma)$, and in the case $\theta = \gamma = \tau = 0$ and $I = \frac{1}{4}$ we compute $I'(0) = \frac{1}{36}$ using the Taylor series for $\ln(1+x)$ and $\frac{1}{1+x}$. We obtain the Cauchy problem

$$\begin{cases} \gamma'(\mathrm{I}) = \dfrac{\ln^2(1+\gamma(\mathrm{I}))}{\dfrac{\ln(1+\gamma(\mathrm{I}))}{\gamma(\mathrm{I})} - \dfrac{1}{1+\gamma(\mathrm{I})} + \dfrac{2\cdot\mathrm{I}\cdot\ln(1+\gamma(\mathrm{I}))}{1+\gamma(\mathrm{I})}} \text{ for } \mathrm{I} \neq \frac{1}{4} \\ \gamma'\left(\frac{1}{4}\right) = 36 \\ \gamma\left(\frac{1}{4}\right) = 0 \end{cases}.$$

Because $I = \dfrac{1-\tau}{4}$ we obtain the Cauchy problem

$$\begin{cases} \gamma'(\tau) = \dfrac{\ln^2(1+\gamma(\tau))}{4\cdot\dfrac{\ln(1+\gamma(\tau))}{\gamma(\tau)} - \dfrac{4}{1+\gamma(\tau)} + \dfrac{2\cdot(1-\tau)\cdot\ln(1+\gamma(\tau))}{1+\gamma(\tau)}} \text{ for } \tau \neq 0 \\ \gamma'(0) = -9 \\ \gamma(0) = 0 \end{cases}.$$

Finally we take into account that $\gamma(\theta) = e^{-\theta}-1$ and $\gamma'(\tau) = -e^{-\theta}\cdot\theta'(\tau)$. We obtain

$$\begin{cases} \theta'(\tau) = \dfrac{\theta^2}{2\cdot(1-\tau)+4-\dfrac{4\cdot\theta}{e^{\theta}-1}} \text{ for } \tau \neq 0 \\ \theta'(0) = 9 \\ \theta(0) = 0 \end{cases}. \tag{11}$$

The above Cauchy problem is solved using the Runge-Kutta method.

In the case of the Gumbel-Hougaard family (see [5,7,11,8]) we have for $\theta \geq 1$ and $\beta = \frac{1}{\theta}$

$$C(u,v) = e^{-\left((-\ln u)^{\theta} + (-\ln v)^{\theta}\right)^{\beta}}. \tag{12}$$

For $\theta = 1$ we obtain the copula *Prod* and for $\theta \to \infty$ we obtain the copula min.

From $\frac{\frac{\partial C}{\partial u}}{\frac{\partial C}{\partial v}} = \frac{\varphi'(u)}{\varphi'(v)}$ we obtain first $\varphi'(u) = -\frac{\theta(-\ln u)^{\theta-1}}{u}$, and from here

$$\varphi(u) = (-\ln u)^{\theta}, \text{ and} \tag{12'}$$

$$g(x) = e^{-x^{\beta}}. \tag{12''}$$

For this family we obtain

$$\tau = 1 - \frac{1}{\theta}, \text{ and from here} \tag{13}$$

$$\theta = \frac{1}{1-\tau}. \tag{13'}$$

The Gumbel-Barnett copula is

$$C(u,v) = u \cdot v \cdot e^{-\left(\theta(\ln u)(\ln v)\right)}, \text{ with } 0 < \theta \leq 1. \tag{14}$$

We notice that we have also the copula product (independence) for $\theta \to 0$.

From $\frac{\frac{\partial C}{\partial u}}{\frac{\partial C}{\partial v}} = \frac{\varphi'(u)}{\varphi'(v)}$ we obtain first $\varphi'(u) = -\frac{1}{u(1-\theta \ln u)}$, and from here

$$\varphi(u) = \frac{\ln(1-\theta \ln u)}{\theta}, \text{ and} \tag{14'}$$

$$g(x) = e^{\frac{1-e^{\theta x}}{\theta}}. \tag{14''}$$

Using (3’) we obtain

$$\tau = -e^{\beta} \cdot \int_{\beta}^{\infty} \frac{e^{-x}}{x} dx < 0. \tag{15}$$

where $\beta = \frac{2}{\theta}$.

The Ali-Mikhail-Haq copula is

$$C(u,v) = \frac{u \cdot v}{1-\theta(1-u)(1-v)}, \text{ with } -1 \leq \theta \leq 1. \tag{16}$$

We notice that we have the copula *Prod* (independence) for $\theta = 0$.

From $\frac{\frac{\partial C}{\partial u}}{\frac{\partial C}{\partial v}} = \frac{\varphi'(u)}{\varphi'(v)}$ we obtain first $\varphi'(u) = -\frac{1}{u(1-\theta(1-u))}$, and from here

$$\varphi(u) = \frac{1}{1-\theta} \cdot \ln\left(\theta + \frac{1-\theta}{u}\right), \text{ and} \tag{16'}$$

$$g(x) = \frac{1-\theta}{e^{(1-\theta)x} - \theta}. \tag{16''}$$

Using (3’) we obtain

$$\tau = 1 - \frac{2(1-\theta)^2 \ln(1-\theta)}{3\theta^2} - \frac{2}{3\theta}. \quad (17)$$

In the above formula τ is increasing on θ, and we have $\tau(-1) = \frac{5 - 8\ln 2}{3}$ and $\tau(1) = \frac{1}{3}$. If we know τ we obtain θ using the bisection method.

3. Case study

Let consider the annual maximum discharges and the floods volumes over a given threshold at Budapest gauge station, for a period of 85 years.

We obtain the Pareto marginals with $a_1 = 0.24115$, $b_1 = 1243.99617$ and $c_1 = 5094.76968$ for discharges, respectively $a_2 = -0.21028$, $b_2 = 1083.46865$ and $c_2 = 65.52616$ for volumes. The Kendal τ is 0.09524. The parameter θ depending on the copula type is as in the following table.

Type	Constraints on $\tau \in [-1,1]$	θ
Clayton: $\theta > 0$	$\tau > 0$	0.21053
Frank: $\theta \neq 0$	$\tau \neq 0$	0.00146
Gumbel-Hougaard: $\theta \geq 1$	$\tau \geq 0$	1.10526
Gumbel-Barnett: $0 < \theta \leq 1$	$\tau < 0$	not our case
Ali-Mikhail-Haq: $-1 \leq \theta \leq 1$	$\frac{5-8\ln 2}{3} \leq \tau \leq \frac{1}{3}$	0.38451
FGM: $-1 \leq \theta \leq 1$	$\lvert\tau\rvert \leq \frac{2}{9}$	0.42857
Fréchet: $\theta \geq 0$	$\tau > 0$	0.09524

In the following graph (fig. 1) the isolines $C(F(x), G(y)) = 1 - \varepsilon = 0.99$ corresponding to a probability of non-exceedance of 99% for the Clayton copula (blue), Frank copula (green), Gumbel-Hougaard copula (red), Ali-Mikhail-Haq copula (light blue), Farlie-Gumbel-Morgestern copula (magenta) and Fréchet (light green) are drawn.

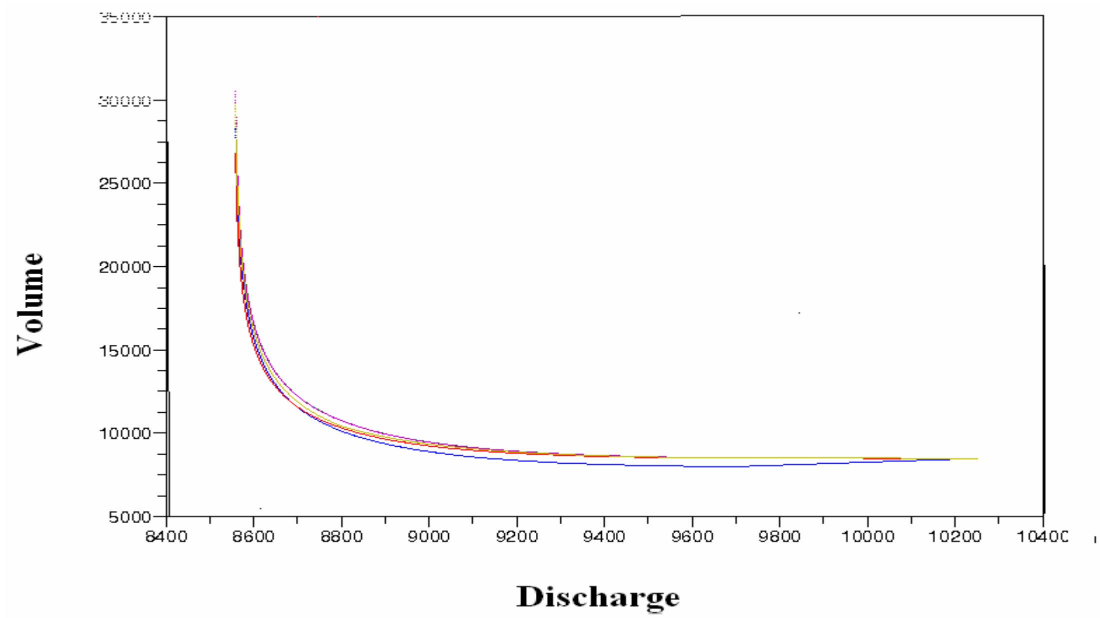

Fig. 1. Isolines of the probability of non-exceedance of 99%

Because in the engineering practice the values corresponding to the probability of exceedance are necessary, the isolines $C^*\left(\overline{F}(x),\overline{G}(y)\right)=\varepsilon=0.01$ for the above families of copula are presented in the same order in the following graph (fig. 2).

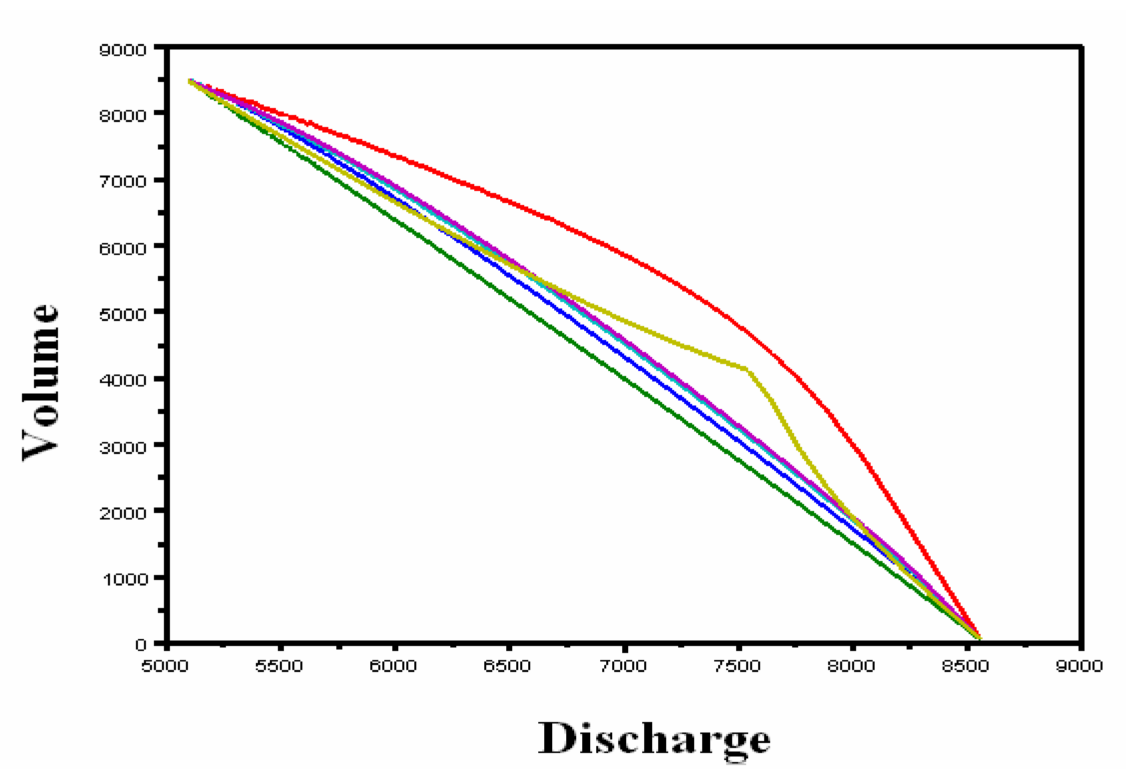

Fig. 2 Isolines of the probability of exceedance of 1%

We notice that the isolines for the Ali-Mikhail-Haq copula and for the Farlie-Gumbel-Morgestern copula are very closed. All the isolines are located between the Frank copula (green) and the Gumbel-Hougaard copula (red). The edge for the Fréchet copula (light green) can be explained by the fact that the upper Fréchet bound $\min$ is not analytical.

4. Conclusions

This paper presents a method to describe the simultaneous behavior of the maximum discharges and of the floods volume using bivariate cumulative distribution function obtained by copulas.
The estimation of the parameter θ of the copula does not depend on the marginal distributions: the cdfs of the marginal are increasing and by applying them to the data we have the same increases of the variables. This is the reason to use Kendall's τ instead of Spearman's ρ.
The points on the isolines from fig.1 are such that the probability of non-exceedance for the coupled variables (the discharges and the volumes) is fixed to a given value α. A similar interpretation is valid for the the points on the isolines from fig. 2, but for the probability of exceedance. Each point on these isolines identified by the couple $(Q_{\max}, V)_{1\%}$ represents a possible realization of the flood corresponding to a probability of exceedance of 1%. This means that there is not a unique flood corresponding to a probability of exceedance of 1% (or to a return period of 100 years), but an infinite number of such floods. The greater the maximum discharge, the smaller the volume is and vice-versa. From this infinity of floods, of outstanding interest in engineering practice is the flood corresponding to the maximum discharge and the flood corresponding to the maximum volume due to the different mechanisms of dykes' failure.

Aknowledgement: The authors are grateful to the European Commission which funded the South East Europe (SEE) program in the frame of which the Danube Floodrisk project is included.

References

[1]G. Dall' Aglio: "Fréchet classes: the beginning", in *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, Eds. G. Dall' Aglio, S. Kotz and G. Salinetti, Kluwer Academic Publishers, 1991, pg. 1-12.
[2]D. Ciuiu: "Simulation Experiments on the Service Systems Using Arrivals and Services Depending Through Copulas", *Analele Universitatii Bucuresti*, No. 1, 2005, pg. 17-30.
[3]D. Ciuiu: *Sisteme si retele de servire*, Matrix Rom, Bucuresti, 2009.
[4]D. Ciuiu: "Gordon and Newell Queueing Networks and Copulas", *Yugoslav Journal of Operations Research*, Vol. 19, No. 1, 2009, pg. 101-112.
[5]C. Genest: "Statistical Inference Procedures for Bivariate Archimedean Copulas", *Journal of American Statistical Association*, Vol. 88, No. 423, 1993, pg. 1034-1043.
[6]S. Kotz, J.P. Seeger: "A new approach to dependence in multivariate distributions", in *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, Eds. G. Dall' Aglio, S. Kotz and G. Salinetti, Kluwer Academic Publishers, 1991, pg. 113-127.
[7]R. Nelsen: "Copulas and association", in *Advances in Probability Distributions with Given*

Marginals. Beyond the Copulas, Eds. G. Dall' Aglio, S. Kotz and G. Salinetti, Kluwer Academic Publishers, 1991, pg. 51-74.

[8]A. Quiroz Flores: "Testing Copula Functions as a Method to Derive Bivariate Weibull Distribution"

[9]V.P. Singh and H. Guo: "Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME)", *Hydrological Sciences--Journal des Sciences Hydrologiques*, Vol. 40, No. 2, 1995, pg. 165-181.

[10] E. Sungur and Y. Tuncer: "The Use of Copulas to Generate New Multivariate Distributions", *The Frontiers of Statistical Computation, Simulation & Modeling. Volume I of the Proceedings of the ICOSCO-I Conference (The First International Conference on Statistical Computing*, Çeşme, Izmir, Turcia, 1987, pg. 197-222.

[11] B. Schweizer: "Thirty years of copulas", in *Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, Eds. G. Dall' Aglio, S. Kotz and G. Salinetti, Kluwer Academic Publishers, 1991, pg. 13-50.

[12] Sklar A., Fonctions de répartition à n dimensions et leur marges, Publications de l'Institut de Statistique de l'Université de Paris, 8 (1959) 229-231.

[13] I. Văduva: *Fast Algorithms for Computer Generation of Random Vectors used in Reliability and Applications*, Preprint nr. 1603, ian. 1994, TH-Darmstadt.

[14] I. Văduva: "Simulation of some Multivariate Distributions", *Analele Universitatii Bucuresti*, Nr. 1, 2003, pg. 127-140.

[15] I. Văduva: *Modele de simulare*, Bucharest University Publishing House, 2004.