ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

PARAMETER ESTIMATION OF THE MARGINAL CUMULATIVE DISTRIBUTION FUNCTIONS AND OF THE COPULA APPLIED IN HYDROLOGY

Daniel Ciuiu*, Romică Trandafir** and Radu Drobot***

Abstract

In this paper we will estimate the parameters of the cumulative distribution functions for marginals for bi-variates and, at the same time, those of the copula that connects them, using a sample of a uniform random variable that depends on these parameters. We will also use the $\chi^2_{1;0.99} = 9.6349$ test for the validation of the model.

In fact, the proposed method is a generalization of the PWM (Probability Weighted Moments) method used in literature. For the PWM method there are also used some moments of a uniform random variable, but only for a marginal. In our method we use the uniform random variable for both the marginal and the copula.

As an application, we will consider the maximum discharges and the volumes of the floods on the Danube River, connected by a copula in a given gauge station. In this case study, we will estimate the parameters for both the marginal distributions and the copula using the method presented in this paper.

Keywords: Archimedean copula, method of the uniform random variable moments, isolines of exceeding probabilities, model's validation, bi-variates, maximum discharge and volume of the floods
Mathematics Subject Classification: 62F10, 62G10, 62P30

1. Introduction

A copula "couples" the marginal distribution functions to form multivariate distribution functions. Sklar (see [16]) first used this word in his paper in 1959. Initially the study of copulas was involved in the development of the theory of probabilistic metric spaces (see [12]). Later, more attention was paid to the study of the dependence structure and the construction of families of multivariate distribution (see [8]).

Copulas are now widely applied to a number of fields as econometrics, economics and finance (see [2]), political science (see [4]), biostatistics (see [14]), medical research (see [9]), hydrology (see [11]) etc.

The hydraulic structures, such as dams and dykes, are designed, among other conditions, as geological or geotechnical characteristics on the hydrologic characteristics of rivers. For the design of these structures a synthetic flood, named also the design flood, is defined based on the characteristics of the registered floods.

* *Department of Mathematics and Computer Science, Technical University of Civil Engineering of Bucharest, Romania, and Romanian Institute for Economic Forecasting*
Bucharest, Romania, E-mail: dciuiu@yahoo.com
** *Department of Mathematics and Computer Science, Technical University of Civil Engineering of Bucharest, Romania, E-mail: romica@utcb.ro*
*** *Department of Hydrotechnic Engineering, Technical University of Civil Engineering of Bucharest, Romania, E-mail: drobot@utcb.ro*

The parameters of the design floods are: the maximum discharge characterized by a standard PE (probability of exceedance PE), e.g. 10%; 1% and 0.1%, the flood volume, the duration of the increasing limb of the flood hydrograph and the total duration of the flood. Usually, these parameters are independently considered, while in fact there is a very strong correlation between some of them, for example between the maximum discharge and the flood volume. This means that, in current practice, both the maximum discharge having the PE P% and the volume for the same value of PE are used for different purposes, for example to design the reservoir volume for flood protection, the spillways of the dams or the crest level of the dykes. Yet, a flood having the same PE P% both for the maximum discharge and the volume actually corresponds to a flood characterized by a lower PE. It seems necessary to derive for each PE isolines defined by the couple of values maximum discharge-volume. An infinite number of combinations of maximum discharge- flood volume results for each PE. According to the purpose of mathematical modeling (hydraulic simulation of the flood wave propagation, seepage computation or stability of the dykes) two remarkable floods can be defined: the flood characterized by the maximum discharge and the corresponding volume and the flood characterized by the maximum volume and the corresponding discharge.

The purpose of this paper is to investigate the copula as an instrument to create the isolines P%, as a basic step for defining the design floods.

In [18, 19] the methods to simulate Archimedean copulas are described, and in [3] there is a presentation of the algorithms to simulate queuing systems with one channel with arrivals and services depending on copulas.

This paper presents some copula functions as a method to derive bivariate distributions used in hydrology. The copula functions allow the construction of unknown multivariate distributions based on known marginals. We consider some cumulative distribution functions for marginals (of the discharges and the volumes of the Danube River) and some copula families to obtain bivariate cumulative distribution functions for modeling the river behavior.

The paper is organized as follows: in the next section we give some basic results on copulas. In Section 3 we design a method to estimate the parameters of the cumulative distribution functions for marginals and of the copula in the Archimedean case using the moments of the random variable uniformly distributed on $(0,1)$. For this reason, we also use a well-known result (see [5]) on Archimedean copulas (theorem 2). In Section 4 we apply the above described method to estimate the involved parameters in the case of the maximum values of the discharges and volumes of the Danube River at the gauging station in Budapest. Some appropriate conclusions are given in the last section of the paper.

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

2. Theoretical background

In the following part, we emphasize some definitions and theoretical results about copulas.

Definition 1 [6, 7, 4] *A copula is a function* $C:[0,1]^n \to [0,1]$ *so that*

1) If there is i *so that* $x_i = 0$ *then* $C(x_1,...,x_n) = 0$.

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

3) C *is increasing in each argument.*

The following theorem is well known (see [17, 7, 13]).

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

Definition 2 [17,18,19] *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 are a (* $n-1$ *) Archimedean copula* C_1 *and a* 2 *Archimedean copula* C_2 *so that* $C(u_1,...,u_n) = C_2(C_1(u_1,...,u_{n-1}), u_n)$.

Consider a function $\varphi:[0,1] \to R$ decreasing and convex with $\varphi(1) = 0$ and its pseudo-inverse g ($g(y) = x = \sup\{z|\varphi(z) \le y|$: see [10,1]). It is known (see [5, 17,6]) that a copula C is Archimedean if and only if there is a function φ as the one above, so that for any $x, y \in [0,1]$ there is

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

Definition 3 *The above mentioned function* φ *is called the generating function of the copula* C .

In this paper we will use the Monte Carlo method to solve the non-linear system involved in the determination of the parameters of marginals and copula, namely the system (4). To obtain the non-linear system of equations whose solution consists in the unknown parameters of the marginals and of the copula we use the following theorem (see [5]).

Theorem 2 *If* X` *and* Y *are uniform random variables connected by the Archimedean copula* C *given by* φ, *like in* (1), *the random variables* $Z_1 = \dfrac{\varphi(X)}{\varphi(X) + \varphi(Y)}$ *and* $Z_2 = C(X,Y)$ *are independent.* Z_1 *is a uniform random* *variable on* $[0,1]$ *and* Z_2 *has the cdf* K *, where* $K(v)=v-\lambda(v)$ *and* $\lambda(v)=\frac{\varphi(v)}{\varphi'(v)}$ *for any* $v\in[0,1]$.

In [15] the PWM (Probability Weighted Moments) method is presented, in order to estimate the parameters of a generalized Pareto distribution. The following equations are obtained:

$$\begin{cases} a=\frac{W_0-8W_1-9W_2}{-W_0+4W_1-3W_2} \\ b=\frac{(W_0-2W_1)(W_0-3W_2)(-4W_1+6W_2)}{(-W_0+4W_1-3W_2)^2} \\ c=\frac{2W_0W_1-6W_0W_2+6W_1W_2}{-W_0+4W_1-3W_2} \end{cases}, \text{ where} \tag{2}$$

$$\begin{cases} W_r=E\left(\left(c+\frac{b}{a}\left(1-(1-F(X))^a\right)\right)\cdot(1-F(X))^r\right) \\ F(x;a,b,c)=1-\left(1-\frac{a(x-c)}{b}\right)^{\frac{1}{a}} \end{cases}. \tag{2'}$$

Remark 1 *According to the lemma of Khintchine, if* X *is a continuous random variable having the cdf* F *, then* $F(X)$ *is a continuous uniform random variable on* $[0,1]$. *In fact, in the equations* (2) *and* $(2')$ *we use the uniform random variable* $1-F(X)$ *(also used in* [20] *to generate the exponential random variable by the inverse method). The value* W_r *can be computed if we know the parameters and the moments of the orders* r .

3. The proposed method

Suppose that the random variables X and Y have the cdfs $F(x;a_1,\ldots,a_p)$ depending on the p parameters and $G(y;b_1,\ldots,b_q)$ depending on the q parameters respectively. Also consider the case of an Archimedean copula C with known $\varphi(u;\theta_1,\ldots,\theta_r)$ depending on the r parameters. Using Theorem 2 we compute for the observed data $(X_i,Y_i)_{i=\overline{1,n}}$

$$U_i(a_1,\ldots,a_p;b_1,\ldots,b_q;\theta_1,\ldots,\theta_r)=\frac{\varphi(F(X_i;a_1,\ldots,a_p);\theta_1,\ldots,\theta_r)}{\varphi(F(X_i;a_1,\ldots,a_p);\theta_1,\ldots,\theta_r)+\varphi(G(Y_i;b_1,\ldots,b_q);\theta_1,\ldots,\theta_r)}. \tag{3}$$

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

Using the moments of the uniform random variable, we obtain the nonlinear system

$$\overline{U_i^k(a_1,...,a_p;b_1,...,b_q;\theta_1,...,\theta_r)} = \frac{1}{k+1},\ 1 \le k \le p+q+r. \qquad (4)$$

We denote in the above system for $k = \overline{1, p+q+r}$

$$\Psi_k(a_1,...,a_p;b_1,...,b_q;\theta_1,...,\theta_r) = \overline{U_i^k(a_1,...,a_p;b_1,...,b_q;\theta_1,...,\theta_r)} - \frac{1}{k+1}.$$

Remark 2 *The above random variables* U_i *, depending on the parameters* a_1 *,...,* a_p *,* b_1 *,...,* b_q *,* θ_1 *,...,* θ_r *are independent uniform random variables. This is because* $F(X_i;a_1,...,a_p)$ *and* $G(Y_i;b_1,...,b_q)$ *are uniform due to the lemma of Khintchine, hence the random variables* U_i *are uniform due to Theorem 2.*
Therefore, the method to estimate the parameters of the marginal and of the copula at the same time using the non-linear system (4) *is analogue to the above mentioned PWM method. Both methods use the moments of the uniform random variable. The differences between the two methods are that in the former we use the moments of the uniform random variable, not combinations of two moments, whereas in the latter we use equations for all the involved parameters (for the two marginals and for the copula).*

The system (4) is solved by the Monte Carlo method. We first generate 1000 sets of parameters in the following way. For the parameters for which we know a bounded interval (such as the third parameter of a Pareto distribution $c \in \left(0, \min_{i=1,n} X_i\right)$) we generate a uniform on this interval. If the parameter is greater than 0 (as λ for the exponential distribution) we generate the U uniform on $(0,1)$, and the generated parameter is $\frac{1}{U} - 1$. For a parameter that can take any real value (as μ for the log-normal distribution) we first generate a positive parameter and next a random sign.

For a copula with unbounded parameter we first generate the Kendall τ (see [5, 1, 7]) and we compute the parameter to be generated as a function of τ. In the case of the Frank copula, when we do not know an analytical formula $\theta = \theta(\tau)$ we first generate the parameter as for the Clayton family $(\theta > 0)$, and then a random sign.

After we generate a set of parameters we compute the left sides of the equations $\Psi_k(a_1,...,a_p;b_1,...,b_q;\theta_1,...,\theta_r) = 0$, with $k = \overline{1, p+q+r}$ in (4), $F\left(\min_{i=1,n} X_i\right)$, $1 - F\left(\max_{i=1,n} X_i\right)$, $G\left(\min_{i=1,n} Y_i\right)$ and $1 - G\left(\max_{i=1,n} Y_i\right)$. The estimated parameters are in such a way that the sum of squares of the values above is minimum.

In the hydrological literature the most common cdfs for the statistical processing of the maximum discharges and flood volumes are:

a) For the partial series of maximum annual discharges: Generalized Extreme Values (GEV), LogPearson3, Gamma2, Generalized Gamma, Lognormal etc

b) For the partial series of maximum discharges over a threshold: Generalized Pareto Distribution, Weibull, LogPearson3, Gamma2, Generalized Gamma.

At the same time, the next copulas are used in hydrology for modeling the connection between the maximum discharges and the floods volume [4]: Frank, Gumbel-Hougaard, Gumbel-Barnett, Ali-Mikhail-Haq and Nelsen Ten.

Based on the partial series of maximum discharges and volumes, a number of $p+q+r$ coefficients is determined, where p, q are the numbers of the parameters for the first and the second marginal respectively, while r is the number of the parameters of the copula. In the considered cases, $r=1$.

Next we will present some types of marginals and of Archimedean copula that we consider in the following. As marginal cdfs we consider the exponential distribution, the translated exponential distribution, the generalized Pareto distribution and the log-normal distribution.

The one-parameter exponential cdf is for $\lambda > 0$ and $x \ge 0$

$$F(x) = 1 - e^{-\lambda \cdot x}. \tag{5}$$

The two-parameter exponential cdf is for $b > 0$ and $x \ge c$

$$F(x) = 1 - e^{-\frac{x-c}{b}}. \tag{6}$$

The generalized Pareto cdf (with the above particular case for $a = 0$) is for $a \in \mathbb{R}$, $b > 0$ and $x \ge c$ (see [15])

$$F(x) = 1 - \left(1 - \frac{a(x-c)}{b}\right)^{\frac{1}{a}}. \tag{7}$$

The log-normal cdf is for $\mu, \gamma \in \mathbb{R}$, $\sigma > 0$ and $x > \gamma$

$$F(x) = \Phi\left(\frac{\ln(x-\gamma) - \mu}{\sigma}\right), \tag{8}$$

where Φ is the cdf of the standard normal random variable.

In the following part, we will determine the generating copula functions for some copula families. Then we use these generating functions to obtain the corresponding system (4). Those solutions represent the copula's parameters and the parameters of the marginal cdfs.

In the case of the Clayton family, for $\theta > 0$ there is (see [5,4])

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

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

For $\theta \to 0$ we obtain the copula *Prod* (independence case), and for $\theta \to \infty$ we obtain the upper Fréchet bound min.

To determine the generating function φ for this type of copula we compute $\frac{\frac{\partial C}{\partial u}}{\frac{\partial C}{\partial v}}=\frac{\varphi'(u)}{\varphi'(v)}$. It results that $\varphi'(u)=-u^{-\theta-1}$, and from here

$$\varphi(u)=\frac{u^{-\theta}-1}{\theta}. \quad (9')$$

In the same way we have obtained the generating functions for some copulas, the results being presented in the Table 1.

4. Application

The above mentioned method is applied in the following, to determine the suitable copulas to model the dependence between the maximum discharges and the volumes of the Danube River at the gauging station in Budapest (85 data of the time series). The results are in table 2 in the appendix B. The first column is for the connecting copula, the second column for the discharges' marginal, the third column for the volumes' marginal, and the next seven columns are for the values of the estimated parameters (seven is the maximum number of parameters: for instance, if both marginals are Pareto, we have $3+3+1=7$ parameters). The last column contains the minimum sum of squares for the left sides of the equation $\Psi_i\left(a_1,...,a_p;b_1,...,b_q;\theta_1,...,\theta_r\right)=0$ with $i=\overline{1,p+q+r}$ in (4), for $F\left(\min_{i=\overline{1,n}} X_i\right)$ and $G\left(\min_{i=\overline{1,n}} Y_i\right)$, and finally for $1-F\left(\max_{i=\overline{1,n}} X_i\right)$ and $1-G\left(\max_{i=\overline{1,n}} Y_i\right)$ as measure of errors. The selected cases are so that the above minimum sums of squares are less than 0.25. The order of parameters is a,b,c in the case of the generalized Pareto distribution (hence b,c in the particular case of the translated exponential, i.e. the generalized Pareto distribution with $a=0$), whereas in the case of the log-normal distribution the order is μ,σ,γ.

The best combination between marginal distributions and copula is for the generalized Pareto distribution of the discharges, the generalized Pareto distribution of the volumes and the Frank copula having the minimum sum of squares 0.00097. The above mentioned combination of marginal distributions (the generalized Pareto distribution of the discharges, the generalized Pareto distribution of the volumes) is also the best in the case of the Clayton copula having the error 0.08492.

We will represent the isoline of exceedance probabilities for 1% in the best case (generalized Pareto discharges, generalized Pareto volumes and the Frank copula) in Figure 1, Appendix *A*. In the mentioned graphics the first point (the bottom-right corner) is $(Q_1 = 8756, c_2 = 201)$, and the last point is $(c_1 = 5130, Q_2 = 4526)$.

To apply the χ^2 test for the model's validation we first divide the interval $(0,1)$, which contains the obtained uniform random variables, in 20 sub-intervals having the same length. For having 5 values in each interval, we link some neighboring intervals and we finally obtain only 13 intervals. Since we have estimated 7 parameters, the number of freedom degrees is $13-1-7=5$. We obtain the statistics value $\chi^2_{calc} = 12.60784$, an inferior value to $\chi^2_{7;0.99} = 18.4753$. This means that the model above can be accepted with the first degree error $\varepsilon = 0.01$.

If we consider the exponential discharges, the sum of squares is between 0.45216 (exponential volumes and Frank copula) and 0.66189 (log-normal volumes and Gumbel-Hougaard copula). We conclude that the discharges are not exponential. If we are not convinced that one of these models is wrong, we can check it using the χ^2 test. As an exemplification, we test the model with the minimum of the above mentioned sum of squares (0.45216 for exponential discharges, exponential volumes and Frank copula). If the number of the initial interval with the same length in which we divide the interval $(0,1)$ is between 20 and 35 inclusively, we obtain at most 4 intervals by joining neighboring intervals. In this case the number of degrees of freedom is at most $4-1-3=0$. For 36 initial intervals, in the end we obtain 5 intervals. There is $\chi^2_{calc} = 956.97861$, a much greater value than $\chi^2_{1;0.99} = 9.6349$, hence we reject the model. The results of the χ^2 test for the considered cases in Table 2 (appendix *B*) are contained in Table 3 , appendix *C*.

According to the obtained results for the analyzed case study, we can conclude that the best combination between marginal distributions and copula is for the generalized Pareto distribution of the discharges, the generalized Pareto distribution of the volumes and the Frank copula.

The above mentioned combination of marginal distributions (the generalized Pareto distribution of the discharges, the generalized Pareto distribution of the volumes) is also the best in the case of the Clayton copula.

The resulted bi-variates cdf allow us to obtain two major characteristics of the synthetic floods (maximum discharge and volume). The other two parameters (increasing time of the flood hydrograph and the total duration) are obtained processing the registered floods.

Having these parameters, the synthetic floods can be drawn. As mentioned before, the synthetic floods are used for the design of the hydraulic structures: dykes, reservoirs, spillways etc.

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

A synthetic flood obtained on the derived hydrological characteristics is presented in the following Figure 2:

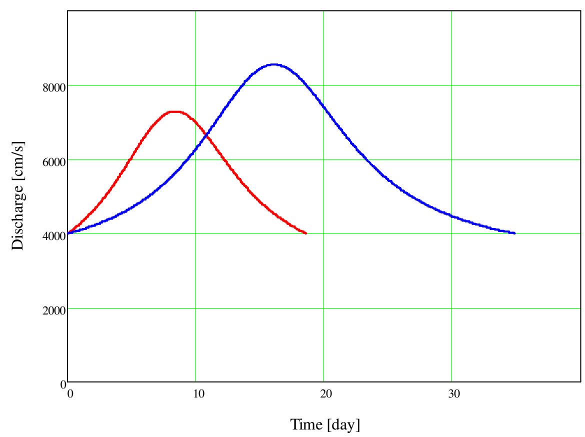

Figure 2:
red curve – 10% excedence flood (T= 10 years return period)
blue curve – 1% exceedance flood (T=100 years return period)

5. Conclusions

The method of the uniform random variable moments that we have proposed in this paper to estimate the parameters of the cdfs for marginals and the copula is equivalent to minimize the sum of squares except $F\left(\min_{i=1,n} X_i\right)$, $G\left(\min_{i=1,n} Y_i\right)$, $1-F\left(\max_{i=1,n} X_i\right)$ and $1-G\left(\max_{i=1,n} Y_i\right)$ (in the case of fulfilling the $p+q+r$ equations the sum of squares is zero). But to the sum of squares we add the four terms from the expressions above, in order to obtain good marginal distributions: even if the minimum sum of squares (without the four added terms) is for instance 0.01, if we have $F\left(\min_{i=1,n} X_i\right)=0.4$ and $F\left(\max_{i=1,n} X_i\right)=0.5$ something is wrong.

The method is general enough, because the marginals are not necessary of the same form (for instance the Pareto distribution for discharges and the exponential distribution for volumes). In our paper we have considered only copulas with only one parameter, but we can generalize to a copula with two or more parameters (for each parameter we have to add a new equation).

An open problem is to build a uniform random variable if we do not know the analytical form of φ, to apply for it a method similar to ours (for example the Farlie-

Gumbel-Morgestern copula).

The synthetic flood obtained using the proposed approach in the case of the Danube River is mainly used for hydraulic simulations, having as purpose the delineation of the flooded areas for different probabilities of exceedance. These hazard maps are necessary for the zonation of the flooded areas and for the future land use and planning. The local authorities will use these maps for territorial organization, not allowing future developments in areas frequently subjected to floods.

Acknowledgement: The authors are grateful to South East Europe (SEE) program which ensured the funding for the Danube Floodrisk project.

References

1. Dall' Aglio G., *Fréchet classes: the beginning*, in: Advances in *Probability Distributions with Given Marginals. Beyond the Copulas*, G. Dall' Aglio, S. Kotz, G. Salinetti (Eds.), Kluwer Academic Publishers, 1-12.
2. Bares P.A., Rajna G., Gyger S., *Style Consistency and Survival Probability in the Hedge Funds Industry*, EFA 2001, Barcelona Meetings and EFMA 2001, Lugano Meetings.
3. Ciuiu D., *Simulation Experiments on the Service Systems Using Arrivals and Services Depending Through Copulas*, Analele Universitatii Bucuresti, **1** (2005) 17-30.
4. Flores A.Q., *Testing Copula Functions as a Method to Derive Bivariate Weibull Distributions*, Report New York University, Nov. 4, 2009.
5. Genest C., *Statistical Inference Procedures for Bivariate Archimedean Copulas*, Journal of American Statistical Association, **88(423)** (1993) 1034-1043.
6. Kotz S., Seeger J.P., *A new approach to dependence in multivariate distributions* in*: Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, G. Dall' Aglio, S. Kotz, G. Salinetti (Eds.), Kluwer Academic Publishers, 113-127.
7. Nelsen R.B., *Copulas and association in: Advances in Probability Distributions with Given Marginals. Beyond the Copulas*, G. Dall' Aglio, S. Kotz, G. Salinetti (Eds.), Kluwer Academic Publishers, 51-74.
8. Nelsen R.B., *An Introduction to Copulas*, Springer-Verlag, New York, 1999.
9. Regnate R. et al., *Clinical Characteristic and Four-Year Outcomes of Patients in the Rhode Island Takotsubo Cardiomypaty Registry*, American Journal of Cardiology, **103(7)** (2009) 1015-1019.
10. Robert, C., Casella, G., *Monte Carlo Statistical Methods*, Springer-Verlag, 2010.
11. Salvadori G., De Michele C., *On the Use of Copulas in Hydrology: Theory and Practice,* J. Hydrologic Engrg., **12(4)** (2007) 369-380.
12. Schweizer B., Sklar A., *Probabilistic Metric Spaces*, North-Holland, New York,

1983.

13. Schweizer B, *Thirty years of copulas*, in: *Advances in Probability Distributions with Given Marginals. Beyond the Copulas,* G. Dall' Aglio, S. Kotz, G. Salinetti (Eds.), Kluwer Academic Publishers, 13-50.
14. Shih J.H., Thomas L.A., *Inference of the Association Parameter in Copula Models for Bivariate Survival Data*, Biometrics, **51(4)** (1995) 1384-1399.
15. Singh V.P., Guo H., *Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME)*, Hydrological Sciences-Journal des Sciences Hydrologiques, **40(2)** (1995) 165-181.
16. Sklar A., *Fonctions de r epartition a n dimensions et leur marges*, *Publications de l'Institut de Statistique de l'Universit e de Paris*, **8** (1959) 229-231.
17. Sungur E., Tuncer Y., *The Use of Copulas to Generate New Multivariate Distributions*, in: *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, 197-222.
18. Văduva I., *Fast Algorithms for Computer Generation of Random Vectors used in Reliability and Applications*, Preprint nr. 1603, ian. 1994, TH-Darmstadt.
19. Văduva I., *Simulation of some Multivariate Distributions*, Analele Universitatii Bucuresti, **1** (2003) 127-140.
20. Văduva I., *Modele de simulare*, Bucharest University Publishing House, 2004.

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

Appendix A

The isoline of exceedance probabilities for 1% in the case of generalized Pareto discharges, generalized Pareto volumes and the Frank copula

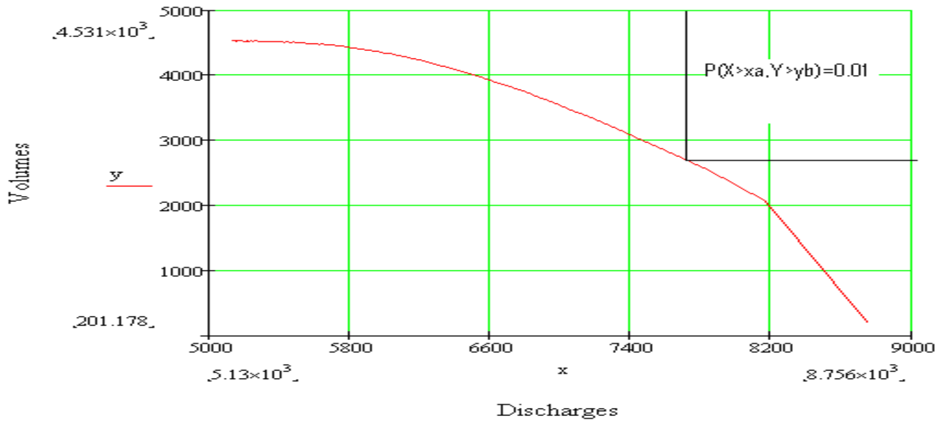

Figure 1: Graphical representation of the excendance probabilities

Appendix B
Generating functions and the estimated parameters for some families of copulas

Table 1: Some families of Archimedean copula and their generating functions

Copula family and limits	Formula for $C(u,v)$	Generating function $\varphi(u)$
Frank $(\theta \in \mathbb{R}^*)$ Prod for $\theta \to 0$; min for $\theta \to \infty$; W for $\theta \to \infty$.	$-\frac{1}{\theta}\cdot \ln\left(\frac{e^{-\theta(u+v)} - e^{-\theta u} - e^{-\theta v} + e^{-\theta}}{e^{-\theta}-1}\right)$	$\ln\frac{1-e^{-\theta}}{1-e^{-\theta \cdot u}} -$
Gumbel-Hougaard $(\theta \ge 1)$ Prod for $\theta = 1$; min for $\theta \to \infty$	$e^{-\left((-\ln u)^{\theta} + (-\ln v)^{\theta}\right)^{\frac{1}{\theta}}}$	$(-\ln u)^{\theta}$
Gumbel-Barnett $(0 < \theta \le 1)$ Prod for $\theta \to 0$	$u \cdot v \cdot e^{-(\theta(\ln u)(\ln v))}$	$\frac{\ln(1-\theta \ln u)}{\theta}$
Ali-Mikhail-Haq $(-1 \le \theta < 1)$ Prod for $\theta = 0$	$\frac{u \cdot v}{1-\theta(1-u)(1-v)}$	$\frac{1}{1-\theta}\cdot \ln\left(\theta + \frac{1-\theta}{u}\right)$
Ali-Mikhail-Haq $(\theta = 1)$	$\frac{u \cdot v}{u + v - u \cdot v}$	$\frac{1}{u} - 1$
Nelsen Ten	$\frac{u \cdot v}{\left(1+(1-u^{\theta})(1-v^{\theta})\right)^{\frac{1}{\theta}}}$	$\frac{1}{2\theta}\cdot \ln(2u^{-\theta} - 1)$

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

Table 2: The estimated parameters for the marginals and copula

Copula	Discharges' marginal	Volumes' marginal	Values of the parameters							Minimum sum of squares
Clayton	Translated exponential (b_1,c_1)	Translated exponential (b_2,c_2)	b_1 =2237	C_1=5018	b_2=5681	c_2=187	θ=0.37			0.2129
		Pareto (a_2, b_2,c_2)	b_1 =937.54	C_1=5130	a_2=-0.16294	b_2=2226.76	c_2=42	θ=0.55663		0.12591
	Pareto (a_1, b_1,c_1)	Pareto (a_2, b_2,c_2)	a_1=-0.08	b_1=2009.65643	c_1=5130	a_2=-0.71707	b_2=2308.55758	c_2=184	θ=0.91614	0.08492
12 nk	Translated exponential (b_1,c_1)	exp(λ_2)	b_1 =2533.04	C_1=5114	λ_2=0.00014	θ=-19.003				0.14583
		Translated exponential (b_2,c_2)	b_1 =1885	C_1=5041	b_2=5990	c_2=42	θ=-0.81			0.17579
		Pareto (a_2, b_2,c_2)	b_1=1935.76	C_1=5130	a_2=-0.03965	b_2=2865.02	c_2=75	θ=17.1394		0.03652
	Pareto (a_1, b_1,c_1)	exp(λ_2)	a_1=-0.0233	b_1=1792.45	c_1=5130	λ_2=0.00024	θ=17.225			0.03406
		Translated exponential (b_2,c_2)	a_1=-0.05997	b_1=2602.08	c_1=5130	b_2=7863.28	c_2=63	θ=-18.4917		0.1884
		Pareto (a_2, b_2,c_2)	**a_1=-0.03288**	**b_1=729.34**	**c_1=5130**	**a_2=-0.05245**	**b_2=830.31**	**c_2=201**	**θ=6.04409**	**0.00097**
Gumbel-Hougaard	Translated exponential (b_1,c_1)	Pareto (a_2, b_2,c_2)	b_1=1236.25	C_1=5029	a_2=-0.56648	b_2=895.24	c_2=215	θ=3.28626		0.01075
	Pareto (a_1, b_1,c_1)	Pareto (a_2, b_2,c_2)	a_1=-0.03739	b1=1615.99545	c_1=4942	a_2=-0.07185	B_2=9761.63763	c_2=196	θ=1	0.0449
Gumbel-Barnett	Translated exponential (b_1,c_1)	exp(λ_2)	b_1=2557.16	C_1=5130	λ_2=0.00013	θ=1				0.15998
		Translated exponential (b_2,c_2)	b_1=2006.83767	C_1=4990	b_2=6131.0087	c_2=71	θ=0.50993			0.17904

		Pareto (a_2, b_2, c_2)	b_1=1933.57	C_1=5129	a_2=-0.0062	b_2=2820.89	c_2=50	θ=0.15506		0.08514
	Pareto (a_1, b_1, c_1)	Translated exponential (b_2, c_2)	a_1=-0.04183	B_1=2769	c_1=5130	b_2=2151.87	c_2=24	θ=1		0.11985
		Pareto (a_2, b_2, c_2)	a_1=-0.16634	B_1=2219	c_1=5130	a_2=-0.82754	b_2=2847.24	c_2=148	θ=0.71675	0.17465
Ali-Mikhail-Haq	Translated exponential (b_1, c_1)	exp(λ_2)	b_1=1776.2	c_1=5130	λ_2=0.00026	θ=-1				0.07805
		Translated exponential (b_2, c_2)	b_1=2073	c_1=4882	b_2=6449	c_2=44	θ=-0.92596			0.21086
		Pareto (a_2, b_2, c_2)	b_1=2096	c_1=5130	a_2=-0.85041	b_2=2529	c_2=193	θ=1		0.08601
	Pareto (a_1, b_1, c_1)	Pareto (a_2, b_2, c_2)	a_1=-0.00482	b_1=2313.77	c_1=5130	a_2=-0.78039	b_2=2202.63	c_2=198	θ=-0.76231	0.13355
Nelsen Ten	Translated exponential (b_1, c_1)	Translated exponential (b_2, c_2)	b_1=1809.13	c_1=5130	b_2=5334.7	c_2=95	θ=0.60894			0.13356
		Pareto (a_2, b_2, c_2)	b_1=2236	c_1=5130	a_2=-0.65872	b_2=2552.32	c_2=150	θ=0.77198		0.11788
	Pareto (a_1, b_1, c_1)	Translated exponential (b_2, c_2)	a_1=-0.0247	b_1=1548.67	c_1=5130	b_2=1455.06	c_2=56	θ=1		0.04136
		Pareto (a_2, b_2, c_2)	a_1=-0.19496	b_1=2025.53	c_1=5130	a_2=-0.91357	b_2=2151.2	c_2=187	θ=0.34721	0.14512
	Log-normal $(\mu_1, \sigma_1, \gamma_1)$	Pareto (a_2, b_2, c_2)	μ_1=7.05	σ_1=0.3	γ_1=5039	a_2=-0.68605	b_2=2476.21	c_2=153	θ=0.31176	0.01046

ROMANIAN JOURNAL OF MATHEMATICS AND COMPUTER SCIENCE, 2014, VOLUME 4, ISSUE 1, p.27-43

Appendix C
Testing the models

Table 3: The results of the χ^2 test

Copula	Discharges' marginal	Volumes' marginal	χ^2_{calc}	$\chi^2_{\deg;0.99}$	Accepting the model
Clayton	Translated exponential (b_1,c_1)	Translated exponential (b_2,c_2)	150.435829	$\chi^2_{2;0.99} = 9.21034$	No
		Pareto (a_2, b_2,c_2)	53.19608	$\chi^2_{5;0.99} = 15.0863$	No
	Pareto (a_1, b_1,c_1)	Pareto (a_2, b_2,c_2)	55.3431	$\chi^2_{3;0.99} = 11.3449$	No
Frank	Translated exponential (b_1,c_1)	$\exp(\lambda_2)$	322.91597	$\chi^2_{1;0.99} = 6.6349$	No
		Translated exponential (b_2,c_2)	149.84706	$\chi^2_{1;0.99} = 6.6349$	No
		Pareto (a_2, b_2,c_2)	40.4902	$\chi^2_{5;0.99} = 15.0863$	No
	Pareto (a_1, b_1,c_1)	$\exp(\lambda_2)$	153.70588	$\chi^2_{5;0.99} = 15.0863$	No
		Translated exponential (b_2,c_2)	388.26139	$\chi^2_{1;0.99} = 6.6349$	No*
		Pareto (a_2, b_2,c_2)	**13.6667**	$\chi^2_{5;0.99} = 15.086$	**Yes**
Gumbel-Hougaard	Translated exponential (b_1,c_1)	Pareto (a_2, b_2,c_2)	15.29412	$\chi^2_{5;0.99} = 15.0863$	No
	Pareto (a_1, b_1,c_1)	Pareto (a_2, b_2,c_2)	71.23519	$\chi^2_{1;0.99} = 6.6349$	No
Gumbel-Barnett	Translated exponential (b_1,c_1)	$\exp(\lambda_2)$	214.29785	$\chi^2_{1;0.99} = 6.6349$	No
		Translated exponential (b_2,c_2)	209.91765	$\chi^2_{1;0.99} = 6.6349$	No
		Pareto (a_2, b_2,c_2)	125.38655	$\chi^2_{1;0.99} = 6.6349$	No
	Pareto (a_1, b_1,c_1)	Translated exponential (b_2,c_2)	263.83922	$\chi^2_{2;0.99} = 9.21034$	No*
		Pareto (a_2, b_2,c_2)	209.08021	$\chi^2_{1;0.99} = 6.6349$	No

Ali-Mikhail-Haq	Translated exponential (b_1,c_1)	$\exp(\lambda_2)$	146.11485	$\chi^2_{3;0.99} = 11.3449$	No
		Translated exponential (b_2,c_2)	246.67647	$\chi^2_{2;0.99} = 9.21034$	No
		Pareto (a_2, b_2,c_2)	51.90196	$\chi^2_{1;0.99} = 6.6349$	No
	Pareto (a_1, b_1,c_1)	Pareto (a_2, b_2,c_2)	178.94706	$\chi^2_{1;0.99} = 6.6349$	No**
Nelsen Ten	Translated exponential (b_1,c_1)	Translated exponential (b_2,c_2)	138.85882	$\chi^2_{1;0.99} = 6.6349$	No
		Pareto (a_2, b_2,c_2)	168.34454	$\chi^2_{1;0.99} = 6.6349$	No
	Pareto (a_1, b_1,c_1)	Translated exponential (b_2,c_2)	121.16471	$\chi^2_{1;0.99} = 6.6349$	No
		Pareto (a_2, b_2,c_2)	171.92513	$\chi^2_{2;0.99} = 9.21034$	No**
	Log-normal $(\mu_1,\sigma_1,\gamma_1)$	Pareto (a_2, b_2,c_2)	58.23529	$\chi^2_{1;0.99} = 6.6349$	No

* we need 40 initial intervals with the same length

** we need 30 initial intervals with the same length