Variations Detection of Bivariate Dependence Based on Copulas Model

Yan-ling Li, Yun-peng Zhang, Ling Zhou

Abstract—**Copulas theory provides a convenient way to express joint distributions of two random variables. This paper presents an introduction to detect variations for bivariate relationship, based on the methods of sliding window and copula function, illustrated by the case of monthly precipitation and streamflow sequence in Xianyang station of Weihe River nearly 60. Euclidean distance criterion is presented for selecting an appropriate model, estimating its parameters, and checking its goodness-of-fit. We obtain the combined probability density copula function of relationship between precipitation and streamflow before and after the variations. Result shows that the relationship variant between precipitation and streamflow occurred in 1993, and at different times, precipitation and streamflow follow the different probability distribution function. The goal of this paper is to put forward copula-based in the field of variations detection, so as to provide a stepping stone exploring variations detection research of the bivariate dependence.**

Index Terms—**Variations, Dependence structure, Copulas, Joint distribution**

I. Introduction

VARIABLES are often correlative and hence require the joint modeling of several random variables in hydrology[1]. Traditionally, the dependence between variables such as precipitation, streamflow depth and volume has been described using classical families of multivariate distributions. Perhaps the most common models occurring in the hydrological variables are the univariate, or bivariate normal, gamma, lognormal distributions [2]. The limitation above mentioned methods are that the variables must then be characterized by identical distribution [3]. Global climate changes and human activities are strongly affecting the patterns of river runoff and other key hydrological variables [4] (Birsan et al., 2005). Especially in the last decades, there are significantly changes in climate, such as temperature, rainfall, and runoff. The dependence of between the variables has been increasingly emphasized [5]. To date, most dependence analyses are focused on identical distribution. In most cases transformation will lose part information of compared with the original variable analysis. In addition hydrological data that are assumed to be multivariate normal distributed are highly developed, but general approaches for joint nonlinear modeling of nonnormal data are not well developed [6]. But unfortunately a lot of variable distribution is not the same; we have to a transform variables to the same distribution.

Copula function analyses, which avoid this restriction, are just beginning to make their way into the dependence multivariable analysis [7, 8]. The theorem of the copulas was introduced in a 1959 article by Sklar [9]. Copulas are functions that connect multivariate distributions to their one-dimensional margins. It can separate the joint distribution into two contributions: the marginal distributions of each variable by itself, and the copula that combines these into a joint distribution. Copulas represent a substantial advantage of over recently proposed simulation based approaches to joint modeling. Copulas have proved useful in a variety of modeling situations[10]. Copulas analysis has been successfully applied to climate and weather related research[11], to various multivariate simulation studies in civil [12] , to being used for Warranty data analysis [13], to modeling turbulent partially premixed combustion, which is common in practical combustors[14], to the analysis of neuronal dependencies [15], and to various simulation-based performance studies[16].

This paper presents several common copulas, introduces methods for selecting which copulas may be most appropriate. in addition to this, the behavior of the copulas in tail dependence can be used to distinguish among joint distributions. Based on the methods of sliding window, we detect dependence variations of the precipitation and streamflow of Weihe River in Xianyan station from 1951 to 2010, and pointed out the bivariate distribution pattern before and after the variations. The purpose of this paper Provide an overview of two-dimensional copula-based constructions in Variations detection.

II. Copula Function

A. Copula theory

Let F(x), G(y) be random variables marginally uniformly distributed on [0, 1]. A 2-dimensional copula function C : $C:[0,1]^2 \rightarrow [0,1]$ is a joint distribution:

$$C(\mathrm{F}(x),\mathrm{G}(y)) = \mathrm{P}(\mathrm{X} \le x, Y \le y). \tag{1}$$

Manuscript received June 06, 2015; revised July 28, 2015. The work was supported by Natural Science Foundation of China (51190093), the Ministry of Education in the new century talents program (NCET-10-0933).

Yan-ling Li is with Department of Mathematics & Information Science, North China University of Water Resources and Electric Power, Henan, China. (corresponding author: 86-037169310037; fax 86-037169310038; liyanling@ncwu.edu.cn)

Yun-peng Zhang is with Department of Electric Power, North China University of Water Resources and Electric Power, Henan, China

Ling Zhou is with Department of Mathematics & Information Science, North China University of Water Resources and Electric Power, Henan, China.

(Advance online publication: 23 August 2017)

Sklar's theorem [9] states that the joint distribution $H(x,y)$ can be represented as a copula function $C(\cdot)$ of its marginal distributions:

$$H(x,y)=C[F(x),G(y)] \tag{2}$$

The function $C(\cdot)$ is called a copula. For many bivariate distributions, the copula form is the easiest way to express and generate the joint probabilities.

B. Families of copulas

Some well-known copulas distributions are reviewed here [17].

The bivariate Gaussian copula $C^{Ga}(u,v;\rho)$ is given in the following form:

$$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\rho^2}}\times\exp\left[-\frac{s^2-2\rho st+t^2}{2(1-\rho^2)}\right]dsdt \tag{3}$$

Where Φ^{-1} is the inverse cumulative distribution function of a standard Gaussian and the correlation parameter ρ approaches −1 and 1.

The Gaussian copula is flexible in that it allows for equal degrees of positive and negative dependence.

The t-distribution with k degrees of freedom and correlation ρ takes the form, $C^t(u,v;\rho,k)$ is given in the following form:

$$\int_{-\infty}^{t_k^{-1}(u)}\int_{-\infty}^{t_k^{-1}(v)}\frac{1}{2\pi\sqrt{1-\rho^2}}\left[1+\frac{s^2-2\rho st+t^2}{k(1-\rho^2)}\right]^{-(k+2)/2}dsdt \tag{4}$$

where t^{-1} denotes the inverse of the cumulative distribution function of the t-distribution.

Archimedean copulas are a set of functions which are popular because they allow modeling dependence in arbitrarily high dimensions with only one parameter, governing the strength of dependence. Most common Archimedean copulas admit an explicit formula. Commonly used functions contain Gumbel Copula, Clayton Copula and Frank Copula. Distribution functions are described as follow:

$$C^{Cl}(u,v;\theta)=\max[(u^{-\theta}+v^{-\theta}-1)^{\frac{1}{\theta}}] \tag{5}$$

Where, $\theta\in(0,\infty)$

$$C^{Fr}(u,v;\theta)=-\frac{1}{\theta}\ln[1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{(e^{-\theta}-1)}] \tag{6}$$

Where, $\theta\in(-\infty,\infty)\setminus\{0\}$

$$C^{Gu}(u,v;\theta)=\exp\{-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{\frac{1}{\theta}}\} \tag{7}$$

Where, $\theta\in[1,\infty)$

The Clayton copula cannot account for negative dependence. It has been used to study correlated risks because it exhibits strong left tail dependence and relatively weak right tail dependence. As θ approaches zero, the marginal distribution become independent. The Frank copula is popular for several merit reasons. First, unlike some other copulas, The Frank copula permits negative dependence between the marginal distribution, and dependence is symmetric in both tails. Parameter values of 1 correspond to independence. Gumbel does not allow negative dependence, but it contrast to Clayton, Gumbel exhibits strong right tail dependence and relatively weak left tail dependence.

C. Tails dependence of copulas

In some cases the concordance between tail values of random variables is of interest. It requires a dependence measure for upper and lower tails of the distribution. The tail dependence measure is essentially related to the conditional probability and the upper and lower tails of the distribution can be defined as follow,

$$\lambda^{up}=\lim_{u\to1^-}P[Y>G^{-1}(u)\,|\,X>F^{-1}(u)] \tag{8}$$

$$\lambda^{low}=\lim_{u\to0^+}P[Y<G^{-1}(u)\,|\,X<F^{-1}(u)] \tag{9}$$

The measure λ^{up} and λ^{low} are widely used in drought or floods applications of extreme value theory to handle the probability that one event is extreme conditional on another extreme event. For copulas with analytical expressions, the computation of λ^{up} and λ^{low} can be straight-forward. For example, for the Gumbel-copula λ^{up} equals $2-2^{\theta}$.

D. Empirical copula

Empirical copula function can help in choosing a copula that is appropriate for modeling given data. The bivariate empirical copula for a bivariate sample of length n is defined for random variables X and Y as: let $(x_i,y_i)(i=1,2,\cdots,n)$ be samples taken from the two dimension random variable (X,Y), and $F_n(x)$ and $G_n(y)$ are empirical distribution functions X and Y respectively.

$$\hat{C}_n(u,v)=\frac{1}{n}\sum_{i=1}^{n}I_{[F_n(x_i)\le u]}I_{[G_n(y_i)\le v]},\quad u,v\in[0,1] \tag{10}$$

In the formula (10) $I_{[\bullet]}$ is indicator function, when $F_n(x_i)\le u$, $I_{[F_n(x_i)\le u]}=1$ $I_{[F_n(x_i)\le u]}=1$ $I_{[F_n(x_i)\le u]}=1$ else $I_{[F_n(x_i)\le u]}=0$ .The Euclidean distance of E-Copula and bivariate copula is defined as equation (11),

$$d=\sqrt{\sum_{i=1}^{n}\left|C(u_i,v_i)-\hat{C}_n(u_i,v_i)\right|^2} \tag{11}$$

Euclidean distance can reflect goodness of fit, the smaller the value, the copula function is more appropriate.

E. Sliding copula function

The correlation coefficient will obvious change when changes in the relationship between two random variables. The variation analysis steps two random variables are as follows:

(1)Choose the length of the sliding window W, window length of two random variables should be consistent.

(2)Select sliding step length L, and sliding window W remains the same, starting from the first time series data with sliding step moving window W, L until the end of the time series.

(3)Calculate the parameters of the two variables' copulas within each window function, get along with the parameter sequence.

(4)Analyze parameters sequence each window copulas;

(5)Determine the best copulas types before and after variation based on Euclidean distance criteria.

III. Case Study

We analyze variation relationship of precipitation and runoff of Xianyang station in weihe river basin in this paper. The Xianyang station is located in the mainstream. The data analyzed here are annual average steamflow and runoff collected from 1951 to 2010 archived by the Weihe River Hydrological Bureau. Both sequence length of precipitation and runoff are 720. The steamflow and runoff were affected by global climate changes and human activities are strongly affecting the spatial and temporal patterns of river runoff and other key hydrological variables. Due to human activities and climate factors, the relationship of rainfall and runoff changed, and showed a trend of decrease in Xianyang.

A. Correlation coefficient

To illustrate the variation relationship between rainfall and runoff, first we calculate correlation coefficient of monthly precipitation and runoff in the different window. The sliding step length is 12, and the length of the moving window is 12, 24, 36, 48, 60 respectively, as shown in figure 1.

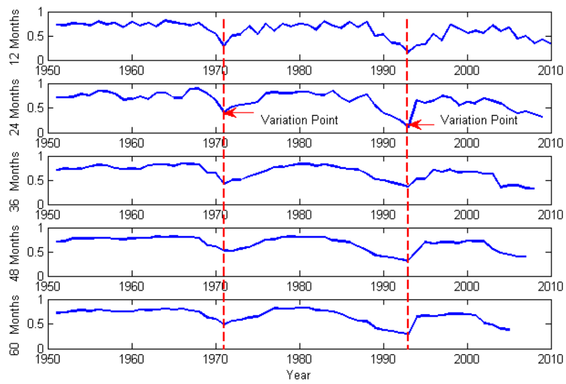

Fig. 1. The correlation coefficient of precipitation and streamflow different moving window.

TABLE I
Copula Function Parameters

	1751.1-1971.12	1972.1-1993.12	1994.1-2010.12
Gauss	$\rho=0.72$	$\rho=0.72$	$\rho=0.64$
T	$\rho=0.72$ $k=4$	$\rho=0.72$ $k=7$	$\rho=0.64$ $k=13$
Gumbel	θ=1.95	θ=1.88	θ=1.65
Clayton	θ=1.78	θ=1.75	θ=1.46
Frank	θ=5.72	θ=5.35	θ=4.33

In order to further explain the occurrence of the abrupt change points of precipitation and streamflow, the correlation coefficient is calculated, as shown in figure 1. The size of the window have play an important role in smoothing, and we choose the window size is W=12, step length L=12 using sliding copula function in this research [18,19].

As can be seen from the figure 1, the correlation coefficient sequence of precipitation and streamflow is divided into three sections by two points of 1971 and 1993, so can be preliminarily ascertained two variation points occurred in 1971 and 1993 at Xianyang from 1951 to 2010. In addition, the variation tendency of the correlation coefficient is consistent in different sliding window.

As shown in figure 1, the correlation coefficient of precipitation and streamflow significantly declined in the Xianyang, and the correlation coefficient of precipitation and streamflow is divided into three parts by two points, and significant changes in the correlation coefficient occurred at 1971and 1993.

B. Fitting copula function

Three stages of copulas connect function of various parameters such as shown in table I, only considering the correlation coefficient.

TABLE I
Copula Function Parameters

	1751.1-1971.12	1972.1-1993.12	1994.1-2010.12
Gauss	$\rho=0.72$	$\rho=0.72$	$\rho=0.64$
T	$\rho=0.72$ $k=4$	$\rho=0.72$ $k=7$	$\rho=0.64$ $k=13$
Gumbel	θ=1.95	θ=1.88	θ=1.65
Clayton	θ=1.78	θ=1.75	θ=1.46
Frank	θ=5.72	θ=5.35	θ=4.33

Calculate tail-dependence coefficient of three stages, as shown in figure 2:

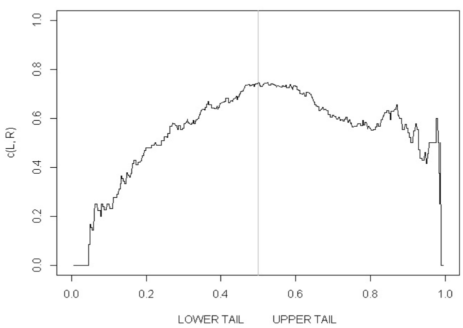

Fig. 2-1. Tail-dependence coefficient from 1951 to 1971.

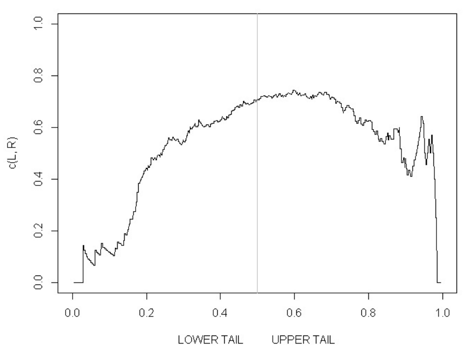

Fig. 2-2. Tail-dependence coefficient from 1972 to 1993.

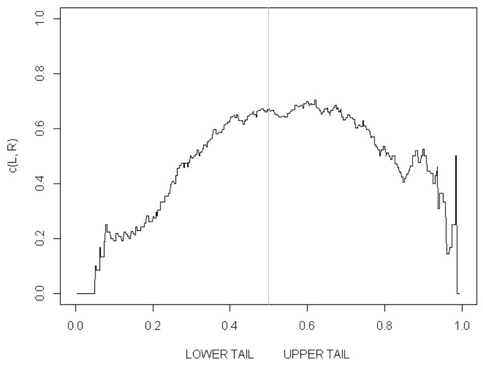

Fig. 2-3. Tail-dependence coefficient from 1994 to 2010.
Fig. 2. Tail dependence coefficient from 1951 to 2010 of Xianyang station.

Fit copula-based bivariate distributions to bivariate data, and estimate both the marginal and the copula parameters (the correlation coefficient, the number of degrees of freedom, the upper tail dependence coefficient, the coefficient of lower tail dependence, the empirical distributions etc.) of precipitation and streamflow copulas within each window function. According to the section 1, slide copulas connect method step, we choose the window size is W=12, step length L=12 in this paper.

Table I and figure 2 shows, the first stage (1951-1971) and the second stage (1972-1993) of the various parameters of precipitation and streamflow are close, whether the relationship between precipitation and streamflow vary in 1971, has yet to be further verified. Euclidean distance of E-Copula and b cty b cc8vgvbivariate copula can reflect goodness of fit.

Euclidean distance of E-Copula and bivariate copula of three stages as shown in table Ⅱ,

TABLE Ⅱ
EUCLIDEAN DISTANCE OF E-COPULA AND BIVARIATE COPULA IN XIANYANG STATION

	1951.1-1971.12	1972.1-1993.12	1994.1-2010.12
GAUSS	$d = 0.491$	$d = 0.398$	$d = 0.327$
T	$d = 0.387$	$d = 0.379$	$d = 0.492$
GUMBEL	$d = 0.357$	$d = 0.339$	$d = 0.415$
CLAYTON	$d = 0.493$	$d = 0.456$	$d = 0.527$
FRANK	$d = 0.478$	$d = 0.455$	$d = 0.369$
OPTIMAL COPULA	GUMBEL -COPULA	GUMBEL -COPULA	FRANK -COPULA

Kernel density of precipitation and streamflow as shown figure 3,

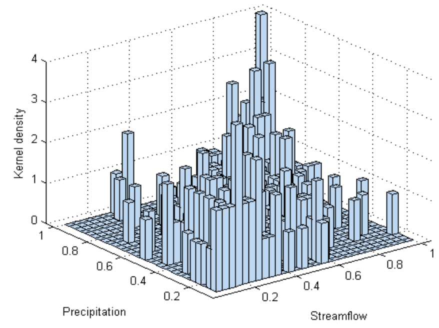

Fig. 3-1. Kernel density from 1951 to 1971.

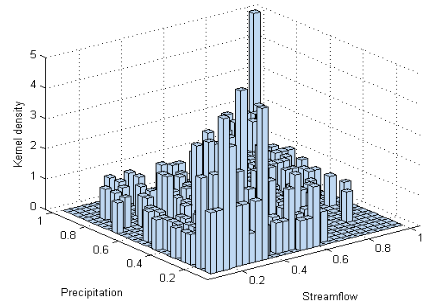

Fig. 3-2. Kernel density from 1972 to 1993.

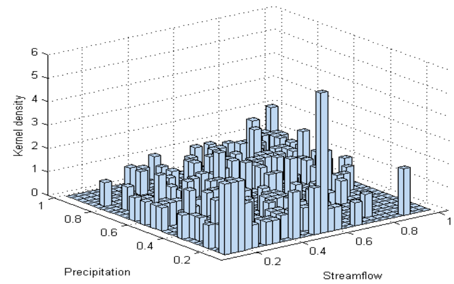

Fig. 3-3. Kernel density from 1994 to 2010.
Fig. 3. Kernel density from 1951 to 2010 of Xianyang station.

We can be seen from table 2 and figure 3, there are changes of copulas parameter of precipitation and steamflow in the first stage (1951-1971) and second stage (1972-1993), but the type of density function has not changed. The density function obeys Gumbel-copula distribution in 1951-1993 and follows Frank-Copula function distribution in 1994-2010. The optimum density functions at each stage as shown in figure 4:

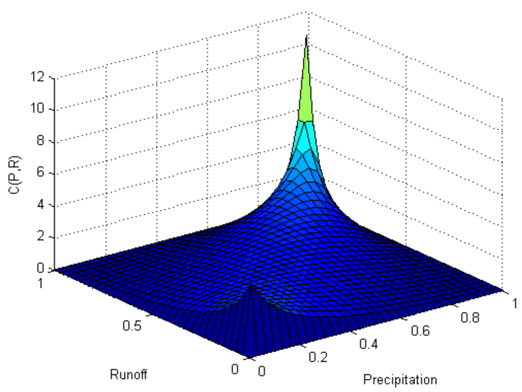

Fig. 4-1. Bivariate Gumbel-copula density function from 1951 to 1971

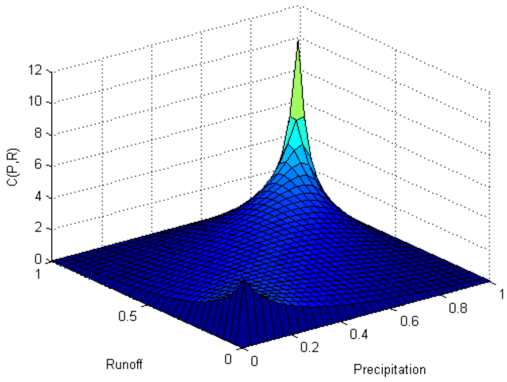

Fig. 4-2. Bivariate Gumbel-copula density function from 1972 to 1993

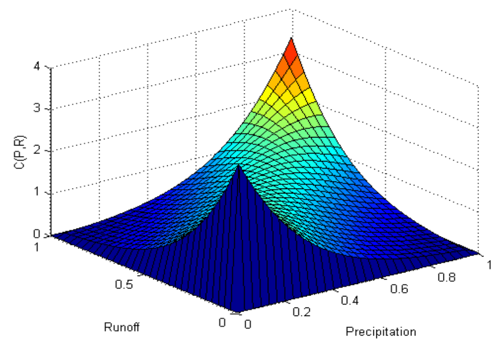

Fig. 4-3. Bivariate Frank -copula density function in from 1994 to 2010
Fig. 4. The joint probability density function of best copula functions each stage

IV. Discussions

Hydrological cycles and water resources are strongly influenced by climatic factors, such as El Niño events and global climate changes, and human activities, such as large-scale water conservation constructions and ecological restoration measures [20, 21]. In 1993 a third ENSO event occurred, while the West Pacific Subtropical high pressure system was strong, resulting in a further significant drop (14.9%) in precipitation in the Weihe River[22], which substantially contributed to a 42.4% reduction in average annual runoff. However, human activities have also markedly influenced runoff patterns. Notably, the construction of terraces, reservoirs and irrigation canals, and other water conservancy measures, reduced the efficiency and runoff of watershed source area. Continuous increases in industrial and agricultural water consumption, combined with climate change, contributed to both the abrupt change of runoff in 1993 and its tendency to decline [23, 24].

V. Conclusions

Copulas have become a popular tool in multivariate modeling successfully applied in many fields. Copula function was used to estimate the probability distribution of precipitation and streamflow in the paper. And on this basis, we discuss the variation of dependence structure between precipitation and streamflow. The copula function contains all information on the dependence structure between the components of precipitation and streamflow, whereas the marginal cumulative distribution function contains all information on the marginal distribution. The copula function, calculated as described above, show significant correlation features with significant decreasing trends in month runoff throughout the study period were detected at Xianyang hydrological station. During the period of records, the Weihe River basin has been becoming drier. Meantime, the local human activities have become more and more extensive. Because of climate changes and a series of water conservancy measures, there is one variation of precipitation and streamflow in Xianyang station. Result shows that the variant of the dependence structure between precipitation and streamflow occurred in 1993, obeyed Gumbel-copula function distribution in 1951-1993 and followed Frank-copula function distribution in 1994-2010. Many advantages of the copula function help us further understand the tail dependence of precipitation and streamflow relationship in the simulation relationship bivariate.

Acknowledgments

The work was supported by Natural Science Foundation of China (51190093, 51179149), the Ministry of Education in the new century talents program (NCET-10-0933). The author thanks Professor Zhang for helpful discussions on time in modeling, Doctor Huang for helpful discussions on copulas, and referees for helpful comments that improved the paper.

References

[1] M. Sulkava, S. Luyssaert, P. Rautio, I. A. Janssens, and J. Hollm'en; "Modeling the effects of varying data quality on trend detection in environmental monitoring," *Ecological Informatics*, vol. 2, no. 2, pp. 167-176, 2007.

[2] R. V. Andreoli , M. T. Kayano "Multi-scale variability of the sea surface temperature in the tropical Atlantic." *Journal of Geophysical Research-atmospheres*, vol. 109, no. 5, pp. 2719-2732, 2004.

[3] C. Genest , A. C. Favre, "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask," *Journal of Hydrologic Engineering*, vol. 12, no. 4, pp. 347-368, 2007.

[4] M. V. Birsan, P. Molnar, P. Burlando, and M. Pfaundler, "Streamflow trends in Switzerland." *Journal of Hydrology,* vol. 314, no. 1-4, pp. 312-329, 2005.

[5] B. Chen, "Climate change and pesticide loss in watershed systems: a simulation modeling study", Journal of Environmental Informatics, vol. 10, no. 2, pp. 55-67, 2007.

[6] P. K. Trivedi , D. M. Zimmer, "Copula Modeling: An Introduction for Practitioners", *Foundations and Trends in Econometrics*, vol . 1, no. 1 pp. 1-11, 2007.

[7] C. D. Michele, G. Salvadori, M. Canossi, A. Petaccia, and R. Rosso, "Bivariate statistical approach to check adequacy of dam spillway." *Journal of Hydrologic Engineering*, vol. 10, no. 1, pp. 50–57, 2005.

[8] Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., and Bobée, B., "Multivariate hydrological requency analysis using copulas." *Water Resour. Res*., vol. 10, no. 1, pp. 1–12, 2005.

[9] A. Sklar, "Random variables, joint distributions, and copulas," *Revue De Théologie Et De Philosophie,* vol. 9, no. 6, pp. 449–460, 1973.

[10] D. M. Zimmer, P. K. Trivedi , "Using trivariate copulas to model sample selection and treatment effects: Application to family health care demand". *Journal of Business and Economic Statistics* vol. 24, pp. 63–76, 2006.

[11] C. S. Lzel, P.Friederichs, "Multivariate non-normally distributed random variables in climate research – introduction to the copula approach". *Nonlinear Processes in Geophysics*, vol. 15, no. 5, pp. 761-772, 2008.

[12] D. B. Thompson, R. T. Kilgore, "Estimating Joint Flow Probabilities at Stream Confluences using Copulas", *Transportation Research Record* , vol. 2262, no. 20, pp. 200–206, 2012.

[13] S. Wu, "Construction of asymmetric copulas and its application in two-dimensional reliability modelling," *European Journal of Operational Research,* vol 238, no. 2, pp. 476–485, 2014.

[14] S. Ruan, "Modelling of turbulent lifted jet flames using flamelets: a priori assessment and a posteriori validation," *Combustion Theory and Modelling,* vol. 18, no. 2, pp 295–329, 2014.

[15] E. Eban, G. Rothschild, A. Mizrahi, G. Elidan, C. Carvalho, and P. Ravikumar, "Dynamic Copula Networks for Modeling Real-valued Time Series," *Journal of Machine Learning Research*, vol. 31, no. 2 pp. 397–405, Part II, 2013.

[16] H. M. Bandara, A. P. Jayasumana, "On Characteristics and Modeling of P2P Resources with Correlated Static and Dynamic Attributes". *IEEE Global Telecommunications Conference,* pp. 1–6, 2011.

[17] A. Philipp. "Bayesian Copula Distributions, with Application to Operational Risk Management-Some Comments", *Methodology and Computing in Applied Probability*, vol. 15, no. 1, pp. 105–108, 2013.

[18] Y. L. Li, J.X. Chang, Q. Huang, and Z. L. Wang, "Diagnosis of abrupt changes in precipitation and runoff relation based on sliding Copula function," *Journal of Hydroelectric Engineering*, vol. 33, no. 6 , pp. 20-24, 2014.

[19] Y. Fang, L. Madsen, L. Liu, "Comparison of Two Methods to Check Copula Fitting," *IAENG International Journal of Applied Mathematics*, vol. 44, no.1, pp. 53-61, 2014.

[20] Z. X. Bao, J. Y. Zhang , G. Q. Wang, "Attribution for decreasing streamflow of the Haihe River basin, northern China: Climate variability or human activities." *Journal of Hydrology*, vol. 407, pp. 117-129, 2012.

[21] K. Kezer , H .Matsuyama, "Decrease of river runoff in the Lake Balkhash basin in central Asia." *Hydrol. Process*, vol. 20, no.2, pp. 1407-1423, 2006.

[22] H. L. Zhang, Y. Chen , G .X. Ren, "The characteristics of precipitation variation of Weihe River Basin in Shaanxi Province during recent 50 years." *Agricultural Research in the Arid Areas*, vol. 26, pp. 236-240, 2008.

[23] S.L. Zhang, Y.H. Wang , P. T.Yu, "Impact of human activities on the spatial and temporal variation of runoff of Jinghe Basin,Northwest China." *Journal of Arid Land Resources and Environment*, vol. 25, no. 2, pp. 66-72, 2011.

[24] A. Bouchoucha, R. Bessaïh, "Natural convection and entropy generation of nanofluids in a square cavity", *International Journal of Heat and Technology*, vol. 33, no. 4, pp. 1-10, 2015.