İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION İSTATİSTİK
Vol. 7, No. 3, December 2014, pp. 63–70
ISSN 1300-4077 | 14 | 3 | 63 | 70

BIVARIATE ANALYSIS OF PRECIPITATION AND RUNOFF IN THE HIRFANLI DAM BASIN, TURKEY, USING COPULAS

Emel Kızılok Kara*
Kirikkale University, Faculty of Arts and Sciences, Actuarial Science Department, 71451 Kirikkale/Turkey

Osman Yıldız
Kirikkale University, Engineering Faculty, Civil Engineering Department, 71451 Kirikkale/Turkey

Abstract: A bivariate analysis of precipitation and runoff in the Hirfanli dam basin, Turkey, using copulas is performed in this study. Two elliptical (Gaussian and Student-t) and three Archimedean family copulas (Clayton, Frank and Gumbel) are tested in modeling of the dependence structure between these hydrological variables. The regionally averaged annual precipitation depths and runoff volumes measured at the entrance of the Hirfanli dam reservoir between 1954 and 2003 are utilized for the parameter estimation and modeling. Different graphical tools and numerical techniques are employed for the appropriate model selection, parameter estimation and goodness-of-fit tests.

Key words: Bivariate analysis; Copulas; Goodness-of-fit; Precipitation; Runoff; Kizilirmak, the Hirfanli reservoir.
History: Submitted: 4 September 2014; Revised: 26 May 2015; Accepted: 10 June 2015

1. Introduction

Copula functions are widely used for modeling the dependence structure in the statistical literature such as in survival analysis, actuarial science and finance. In recent years, numerous successful applications of copulas have been reported in hydrological area [1]. Some authors used Archimedean copulas in modeling of the dependence structure between peak and volume variables [2, 3, 4].

The main advantage of this approach for the hydrological applications is that the selection of an appropriate model for the dependence between multiple variables can proceed independently from the choice of the marginal distributions by the copula. The other purpose of using copula for the hydrological data is related to dependence measure. The linear correlation assumes that marginals of variables are normal distributions. In general, distributions of the hydrological data such as the rainfall depths and runoff volumes have fatter tails than normal distributions. Hence, the dependencies between these variables are described by the dependency parameters included in copula functions instead of linear correlation.

This study investigates appropriate copula selection for the annual precipitation depths and runoff volumes in the Hirfanli dam basin using the data between 1954 and 2003. The Maximum Likelihood Estimation (MLE) and Inference Function for Margins (IFM) methods are respectively used for estimating the parameters of marginal distributions and copula functions. Then, the best fit copula function to describe the dependence structure of the bivariate hydrological data is determined with goodness-of-fit tests according to Akaike Information Criterion (AIC), Schwartz Information Criterion (SIC) and Hanna-Quinn Information Criterion (HQIC).

* Corresponding author. E-mail address: emel.kizilok@gmail.com (E. K. Kara)

2. Materials and Methods

2.1. Study area and hydrological data

With a drainage area of nearly 26700 km^2 the study area is located at the upstream of the Kizilirmak river, one of the major water resources of the Turkey (Figure 1). The Hirfanli dam basin is generally characterized by high topography ranging from 780 to 2400 m and has a typical dry climate with a mean temperature of 10°C and a mean annual precipitation of about 400 mm which is about 63% of the country average. Convective and frontal rainfall systems dominate across the basin. Built in 1959 with 263 km^2 surface area and 5980 hm^3 volume at normal water surface level, the Hirfanli dam reservoir is operated for hydropower, water supply and flood control purposes [5].

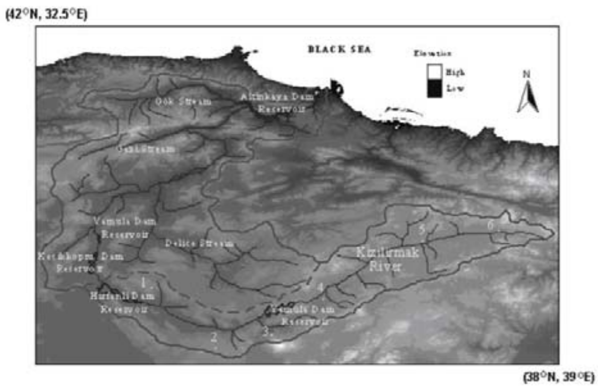

FIGURE 1. The Hirfanli dam basin (separated by the dashed line) and the meteorology stations within the basin

The precipitation time series obtained from the Turkish State Meteorology Directorate (http://www.dmi.gov.tr) were utilized in the study. The regionally averaged annual precipitations were calculated by the Thiessen polygons from 6 meteorology stations within the basin (see Table 1).

TABLE 1. The meteorology stations in the Hirfanli dam basin

Index No	Station Name	Altitude (m)	Latitude (°N)	Longitude (°E)
1	Kirsehir	1007	39.09	34.10
2	Nevsehir	1260	38.35	34.40
3	Kayseri	1093	38.44	35.29
4	Gemerek	1173	39.11	36.04
5	Sivas	1285	39.45	37.01
6	Zara	1348	39.54	37.45

The runoff volumes measured at the entrance of the Hirfanli dam reservoir were obtained from the Turkish State Water Works Directorate (http://www.dsi.gov.tr). Both hydrological time series had no missing data and were found homogeneous. The annual precipitation depths and runoff volumes between 1954 and 2003 were utilized for the bivariate analysis using copulas.

Figure 2 clearly displays the dependence structure between precipitation depths (X_i) and runoff volumes (Y_i) in the Hirfanli dam basin. The scatter plots of 50 pairs (X_i, Y_i) and the corresponding pairs (U_i, Vi) are shown in the figure.

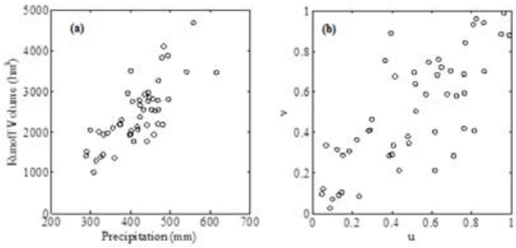

FIGURE 2. The scatter plots of pairs (X_i, Y_i) and (U_i, V_i)

The descriptive statistics, correlation and normality test results are given in Table 2. Here, one can see that the annual precipitations are normal whereas the annual runoff volumes seem to be nearly non-normal with a relatively higher skewness coefficient. The Pearson as well as ranked based Spearman's ρ and Kendal's τ coefficients are utilized for the assessment of dependence. They are respectively calculated as 0.7609,0.7361 and 0.5559, which reveal the presence of a relatively high dependence between the two hydrological variables under consideration.

TABLE 2. Descriptive statistics, correlation and normality test results for the hydrological variables

Statistics	**Precipitation (mm)**		**Runoff volume (hm^3)**	
Sample Number	50		50	
Mean	416.7738		2439.8	
Std. deviation	70.5433		775.5209	
Skewness	0.1849		0.6495	
Excess of Kurtosis	3.1358		3.3289	
Tests	**p-value**	**χ^2 stat.**	**p-value**	**χ^2 stat.**
Jarque-Bera	0.50*	0.3233	0.0771*	3.7404
Correlation tests	**Kendall's τ**	**Spearman's ρ**	**Pearson**	**p-value**
	0.5559	0.7361	0.7609	0.0000**

The p-value is the upper value of significance level of 5%.
(*) the null hypothesis of normal distribution is admitted.
(**) the correlation is significantly different from zero.

Table 3 shows the goodness of fit test results for the marginals of precipitation and runoff volume data. It is shown that the precipitation can be represented by the Gaussian (Normal) distribution and the runoff volumes can be modelled by the Gamma distribution according to the lowest criterion values generated by different criteria and tests. Figure 3 depicts QQ-plots for the annual precipitations and runoff volumes.

TABLE 3. Goodness-of-fit test results for the marginal distributions of precipitation and runoff volume

		Precipitation (mm)					
Marginals	**Parameters**	**MLE**	**AIC**	**BIC**	χ^2	**K-S**	**A-D**
Normal	μ	416.774	**570.7719**	**574.3407**	7.9200	**0.0758**	0.4188
	σ	70.543					
Logistic	μ	417.566	571.2796	574.8483	**6.3200**	0.0759	**0.4082**
	σ	39.622					
Weibull	α	2.5829	571.7760	576.9903	8.8800	0.0993	0.5817
	β	188.55					
Laplace	μ	422.98	574.0453	577.6140	8.5600	0.0843	0.5945
	σ	77.5813					
ExtValue	a	382.242	574.8886	578.4573	11.1200	0.1369	1.1002
	b	65.095					
		Runoff volume (hm^3)					
Marginals	**Parameters**	**MLE**	**AIC**	**BIC**	χ^2	**K-S**	**A-D**
Gamma	α	7.966	**808.7518**	**813.9662**	**6.9600**	0.0957	0.3113
	β	272.86					
Inv Gauss	μ	3163.3	808.8030	814.0173	6.9600	0.0937	0.3069
	λ	53313.3					
Weibull	α	2.1556	808.8252	814.0395	8.5600	0.1046	0.3618
	β	1780.7					
Log Normal	μ	3138.9	808.8348	814.0491	6.9600	**0.0925**	**0.3047**
	σ	771.43					
Normal	μ	2439.78	810.5027	814.0714	21.3600	0.1205	0.6022
	σ	775.52					

MLE: Maximum Likelihood Estimation, AIC: Akaike Information Criterion,
BIC: Bayesian Information Criterion, χ^2: Chi-square test, K-S: Kolmogorov-Smirnov test,
A-D: Anderson–Darling test.

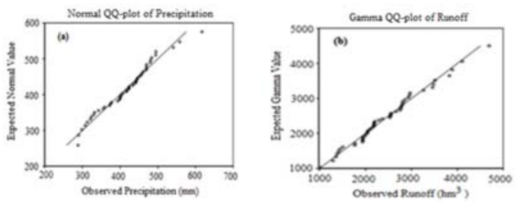

FIGURE 3. QQ-plots showing the fit of marginal models for precipitation (a) and runoff volume (b) data in the Hirfanli dam basin

2.2. Copula functions

Let X and Y be dependent continuous random variables. A two-dimensional copula is a distribution function on the unit square with uniform margins. [6] states that for any continuous random vector (X,Y), a copula, $C:[0,1]^2 \to [0,1]$, uniquely determines $F(x,y) = C(F_1(x), F_2(y))$, where F is the joint distribution function of X and Y random variables and F_i, $i = 1,2$, are the marginal distributions of X and Y, respectively [5]. Further details about copulas can be found in ([7]-[9]). There are various copulas in the literature. In our work, we concentrate on the elliptical copula family including the Gaussian and the Student-t copulas, and the Archimedean copula family such as the Clayton, the Frank and the Gumbel copulas which are widely used in the statistical and hydrological literature. Table 4 shows the copula functions and their parameter space used in this study.

Table 4. The copula functions and their parameter space used in the study

Copula	$C_\theta(u,v)$	Parameter space
Gaussian	$\Phi_\rho\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)=\int\limits_{-\infty}^{\Phi^{-1}(u)}\int\limits_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\rho^2}}exp\left(-\frac{2\rho st-s^2-t^2}{2\left(1-\rho^2\right)}\right)ds$	$\rho\in(-1,1)$
Student-t	$t_{d,\rho}(t_d^{-1}(u),t_d^{-1}(v))=\int\limits_{-\infty}^{t_d^{-1}(u)}\int\limits_{-\infty}^{t_d^{-1}(v)}\frac{1}{2\pi\sqrt{1-\rho^2}}\left(1+\frac{s^2+t^2-2\rho st}{d\left(1-\rho^2\right)}\right)^{-\frac{d+2}{2}}dsdt$	$\rho\in(-1,1),d\in(0,\infty)$
Clayton	$\left(u^{-\theta}+v^{-\theta}-1\right)^{-1/\theta}$	$\theta\in(0,\infty)$
Gumbel	$\exp\left[-\left(u^{-\theta}+v^{-\theta}\right)^{-1/\theta}\right]$	$\theta\in[1,\infty)$
Frank	$-\frac{1}{\theta}\ln\left[1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{\left(e^{-\theta}-1\right)}\right]$	$\theta\in(-\infty,\infty)\setminus\{0\}$

In Table 4, Φ_ρ is the bivariate standard normal distribution function with parameter ρ, and Φ^{-1} is the functional inverse of the univariate standard normal c.d.f., and Φ. $t_{d,\rho}$ is the bivariate Student-t distribution, and t_d^{-1} is the functional inverse of Student-t c.d.f. with d degrees of freedom t_d.

2.3. Parameter estimation

There are different methods for estimating copula parameters. For example, Inference Function for Margins (IFM) methods depend on the marginal distributions for this purpose. However, the method based on the inversion of Kendall's τ dependence measure and the Maximum Pseudo-Likelihood (MPL) method estimate copula parameters without considering them. There is no consensus, but a large number of researchers use the rank based estimation methods. It is argued that IFM methods are non-robust against misspecification of marginal distributions, as the parameter estimation depends on the choice of the univariate marginal distributions and can be affected if such models do not fit adequately [10]. Consequently, the inversion of Kendall's τ and IFM methods are both preferred in the present work. Kendall's τ is a powerful and well-known measure of concordance. The relationship between τ and a copula C is given in Equation (2.1) as follows:

$$\tau=4\int\limits_0^1\int\limits_0^1 C(u,v)\,dC(u,v)-1=1-4\int\limits_0^1\int\limits_0^1\frac{\partial C}{\partial u}(u,v)\frac{\partial C}{\partial v}(u,v)\,dudv \tag{2.1}$$

The estimations based on the inversion of Kendall's τ method for different copulas are given in Table 5, which is taken from [9].

Table 5. The parameter estimations based on the inversion of Kendall's τ method

Copula	Gaussian	Student-t	Clayton	Gumbel	Frank
Kendall's τ	$\frac{2}{\pi}\arcsin(\theta)$	$\frac{2}{\pi}\arcsin(\theta)$	$\frac{\theta}{\theta+2}$	$1-\theta^{-1}$	$1-\frac{4}{\theta}\left[D_1(-\theta)-1\right]$

In Table 5, D_1 denotes the Debye function such that $D_k(-\theta)=D_k(\theta)+\frac{k\theta}{k+1}$ and $D_k(\theta)=\frac{k}{\theta}\int\limits_0^\theta\frac{t^k}{e^t-1}dt$.

The IFM estimates the marginal distribution parameters separately from the copula parameters. The estimation procedure of this method consists of two steps described by [9]. First, the parameters of the marginals are estimated as $\hat{\theta}_1=\arg\max_{\theta_1}\sum\limits_{t=1}^{T}\sum\limits_{i=1}^{2}\log f_i\left(x_{i,t};\theta_1\right)$ and then, given $\hat{\theta}_1$, the parameters of the copula model are estimated as $\hat{\theta}_2 = \arg\max_{\theta_2} \sum_{t=1}^{T} \log c\left(F_1(x_{1,t}), F_2(x_{2,t}); \theta_2, \hat{\theta}_1\right)$. The IFM estimator is defined as $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_2\right)'$.

2.4. Goodness-of-fit testing

The aim of a goodness-of-fit testing is to select the appropriate copula that best represents the dependence structure of variables via the observed data. AIC and SIC are widely used in the literature for this purpose [12], [13]. SIC performs better for relatively large samples, while AIC tends to be superior for relatively small samples [14]. HQIC is also used in the literature for model selection.

Actually, the copula that provides the best fit is the one that corresponds to the lowest value of the criteria used. In this study, we employed AIC, SIC and HQIC defined by Equations 2.2-2.4 to select the best fit copula to the observed data.

$$AIC = -2 \cdot LL + 2 \cdot k \tag{2.2}$$

$$SIC = -2 \cdot LL + \ln(n) \cdot k \tag{2.3}$$

$$HQIC = n \cdot \log\left(\frac{RSS}{n}\right) + 2 \cdot k \cdot \log(\log n) \tag{2.4}$$

where LL is loglikelihood, k is the number of parameters of the copula model, n is the number of observations and RSS is the residual sum of squares that results from linear regression or other statistical model.

3. Results and Discussion

A graphical test of the goodness-of-fit of the dependence structure taken in isolation is given in Figure 4, which displays the simulated random sample of size 1000 from the five selected copulas with parameters estimated by the method of IFM using the precipitation-runoff volume data. The viability of the appropriate model for the bivariate analysis can be visually assessed in Figure 5, where all five copulas seem to fit to the hydrological data.

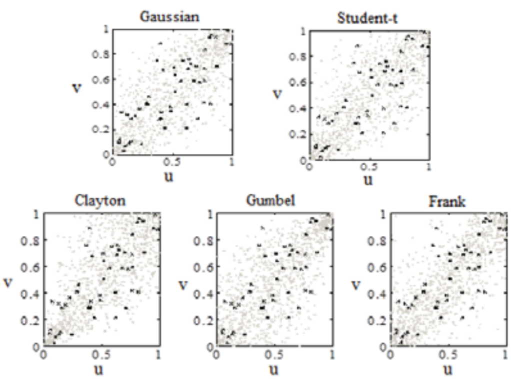

FIGURE 4. Simulated random sample of size 1000 from the five selected copulas with parameters estimated by the method of IFM using the precipitation-runoff volume data, whose pairs of ranks are indicated by an "X".

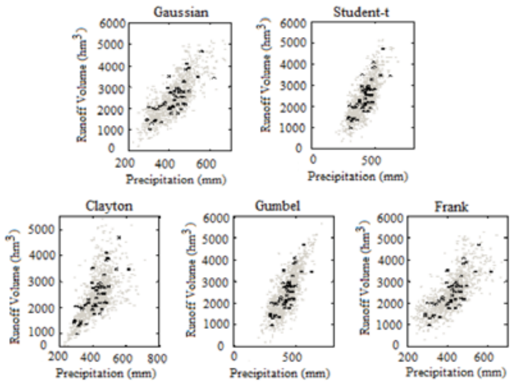

FIGURE 5. Same data as in Fig.4, upon transformation of the marginal distributions as per the selected models for the precipitation and the runoff volume data, whose pairs of observations are indicated by an "X"

According to the goodness-of-fit test results tabulated in Table 6, for a given copula, the criterion values produced by AIC, SIC and HQIC have no significant differences. As a member of elliptical family, the Student-t copula has the lower criterion values as compared to the Gaussian copula. The Clayton copula, on the other hand, emerges as the best Archimedean family copula with the lowest criterion values.

TABLE 6. Goodness-of-fit test results for the selected copula models.

Copula	Parameter	Kendall's τ	IFM	AIC	SIC	HQIC
Gaussian	ρ	0.7758	0.7747	39.72	37.89	39.07
Student-t	ρ	0.7758	0.7747	**37.44**	**33.87**	**36.24**
	d	40.00	40.00			
Clayton	θ	2.5036	2.1629	**20.10**	**16.54**	**18.90**
Gumbel	θ	2.2518	2.2391	32.09	28.52	30.89
Frank	θ	6.858	7.6756	34.26	30.70	33.16

4. Conclusions

In this study, two elliptical family copulas (Gaussian and Student-t) and three Archimedean family copulas (Clayton, Gumbel and Frank) are considered to select an appropriate copula model for the bivariate analysis of the annual precipitation and runoff volumes in the Hirfanli dam basin, Turkey. Regardless of their marginal distribution type, a bivariate analysis of the hydrological data was successfully performed with the use of different family copulas. Three separate criteria were utilized for goodness-of-fit testing to select the best fit copula function. The study results indicated that the three criterion values were found very close for each individual copula under consideration. Overall, the Student-t and the Clayton copulas, as two different family copulas, seemed to be more appropriate to modeling of the dependence structure between the precipitation and runoff volume data in the study area.

References

[1] C. Genest and A.C. Favre. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering,* July/August, 347-368.

[2] A.C. Favre, S. El Adlouni, L. Perreault, N.Thimonge and B. Bobe. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, 40(1): W01101, 1-12.

[3] J.T.Shiau. (2006). Fitting drought duration and severity with two-dimensional copulas. *Water Resources Management*, 20(5), 795-815.

[4] A. I. Requena, L. Mediero, L. And L. Garrote. (2013). Bivariate return period based on copulas for hydrologic dam design: comparison of theoretical and empirical approach. *Hydrology and Earth System Sciences Discussions*, 10(1), 557-596.

[5] O. Yildiz. (2009). Assessing temporal and spatial characteristics of droughts in the Hirfanli dam basin, Turkey. *Scientific Research and Essays*, 4(4), 249-255.

[6] A. Sklar. (1959). Fonctions de rpartitions n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris,* 8, 229-231.

[7] H. Joe. (1997). Multivariate Models and Dependence Concepts, Chapman & Hall, London.

[8] R.B. Nelsen. (2006). An Introduction to Copulas, (second ed) Springer, New York.

[9] U. Cherubini, E. Luciano, W. Vecchiato. (2004). Copula Methods in Finance, Wiley, New York.

[10] G. Kim, M.J. Silvapulle, P. Silvapulle. (2007). Comparison of semiparametric and parametric methods for estimating copulas. *Communications in Statistics. Simulation and Computation*, 51, 28362850.

[11] F. Michiels and A. De Schepper. (2008). A copula test space model how to avoid the wrong copula choice. *Kybernetika*, 44(6), 864-878.

[12] G. Schwarz. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461464.

[13] H. Akaike. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19, 716723.

[14] R.H. Shumway, D.S. Stoffer. (2011). Time series analysis and its applications: with R examples. Springer.