Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences
Open Access
Discussions

COPULA AND ARMA BASED STUDY OF CONTROLLED OUTFLOW AT FARAKKA BARRAGE

Uttam Singh; Venkappayya R. Desai; Pramod K. Sharma; and Chandra S.P. Ojha

Research Scholar[1]; Professor[2]; Associate Professor[3]; Professor[4]

Email: uttamsingh426@gmail.com; venkapd@civil.iitkgp.ernet.in; drpksharma07@gmail.com; cspojha@gmail.com

[1,3,4]Department of Civil Engineering, Indian Institute of Technology Roorkee-247667

[2]Department of Civil Engineering, Indian Institute of Technology Kharagpur-721302

*Corresponding Author: Email: uttamsingh426@gmail.com

COPULA AND ARMA BASED STUDY OF CONTROLLED OUTFLOW AT FARAKKA BARRAGE

Uttam Singh[1]; Venkappayya R. Desai[2]; Pramod K. Sharma[3]; and Chandra S.P. Ojha[4]

[1,3,4]Department of Civil Engineering, Indian Institute of Technology Roorkee-247667

[2]Department of Civil Engineering, Indian Institute of Technology Kharagpur-721302

Abstract

In this study, 25 years mean monthly out flow discharge data of Farakka barrage was used (i.e., from 1949 to 1968). Farakka barrage is located between on Ganga River. Spatial and temporal variation in flow rate for any particular area is very common due to various meteorological and other factors existing in nature. But large variations in these factors cause extreme events (e. g., floods and droughts).Monthly outflow discharge for a particular critical month are predicted using statistical models (ARMA Model and Copula Model). Different Copulas (i.e., Normal, t, Frank, Clayton, Gumbel-Hoggard, Ali-Mikhail-Haq) are used for this purpose and the copula model is selected based on distribution functions (Normal distribution, Lognormal distribution, Extreme value type-1 distribution, Generalized Extreme value type, Gamma distribution, Weibull distribution, Exponential distribution). The distribution is selected based on the Mean square error (MSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The model parameters were computed using the Maximum Likelihood (ML) estimation method.

Key words: Farakka barrage; ARMA, Copulas, Simulation; Discharge.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences Discussions
Open Access

1. **Introduction**

An accurate flood-frequency analysis is critical for the design of many civil infrastructures such as drainage system and flood proof walls. Copula word is taken from Latin language and the meaning of copula is link and the concept of copula was introduced in mathematical and statistical manner by Sklar (1959) in a theorem that describes a copula as a function. Afterwards, many researchers such as Genest and MacKay (1986), Genest and Rivest (1993) and Nelsen (1999), Favre *et al*. (2004), Genest and Favre (2007) and Salvadori and De Michele (2007) used in hydrology applications. Crucial steps for copulas modeling are driving the bivariate distribution of peak flow and volume, volume and duration, peak flow and duration (Zhang and Singh, 2006). Archimedean copulas (Clayton, Frank, Gumbel-Hoggard, Ali-Mikhail-Haq, Indpendance and Joe) can be used for bivariate modeling peak flow and volume, volume and duration, peak flow and duration.

Dependence structure of data set is captured by copulas, thus they are used for describing the dependence of o extreme output values and also useful for depandence non parametric measurement. Statistical dependence among three random variables two copulas are used for modeling. The Archimedean copulas are prepared by association measurement of Kendall's *tau* (Osorio et al. 2009). The probability density function for the two-dimensional random variable representing volume and time is given in graphic form. The graphs both represents Clayton copula and Gumbel-Hougaard functions. The Gumbel-Hougaard copula was best suited for this study because it shows lower value in selection criterion function. Gumbel-Hougaard copula shows better matching of empirical and theoretical distribution function. The results obtained in the study, risk values at extreme analyzed values of controlled discharge and flood control capacity are not monotonic. It represents that simulations were completed for sets of only 10000 cycle elements and only 10000 cycles (Twaróg, 2016). Peak flow and hydrograph volume both can be jointly studied by bivariate approach (e.g., Goel et al. 1998; Yue et al. 1999; Favre et al. 2004; Shiau et al. 2007). The selection of the, different criterion should be consider among the candidate copula (Chowdhary et al. 2011; Requena et al. 2013). The first criterion is the goodness-of-fit test which relates the ability of copula to characterize the data (Genest et al. 2009), The second criterion is estimation of kendall's tau return period estimation by copula. It relates the adequacy of copula, for a large copula value t ϵ[0, 1], which is based on the Kendall's function $K_C(t) = P[C_\Theta(u_1, u_2) \leq t]$ (Genest and Rivest,1993). The third criterion is the estimation of Akaike Information Criterion (AIC) (e.g., Zhang and Singh, 2006). A copula-based model and a distributed hydro-meteorological model and a copula-based model can be studied by combining extension of observed flood series (Requena, et al. 2015). Significant number of researchers found in their research that Gumbel-Hougaard copula as the most suitable choice to model the dependence structure relating to the peak flow discharge and the flood volume (De Michele et al., 2005; Zhang and Singh, 2007, Karmakar and Simonovic, 2009 and Li et al., 2013). A copula-based approach was used to derive a bivariate distribution function of two constituent flood variables, with regard to a real-world case study. It was found to provide an effective and straightforward strategy for inferring probability functions from multivariate sample data. Powerful tests developed inside copula framework allowed to investigate the empirical dependence structure in an accurate manner, especially with respect to the evaluation of tail dependencies (Balistrocchi, 2017). The dependence of copula model between intensity and rain fall duration, both properties of marginal distribution and dependence between intensity and storm duration were preserved. The Joint cumulative distribution functions represents dependence between independent variables of their marginal distribution of copula (Joe, 1997 and Nelsen, 2006). Gaussian copula was used for generation of 1020 synthetic data sets. Among the data sets, 21 data sets lies beyond the range of acceptance so these data sets were omitted. Of course it is not possible to cover all input-output cases in trained models the extrapolation limit are required (Hooshyaripor et al. 2014). Best copula model can be selected by coarse grid model selection with supposedly known marginal parameters in which 15 families of copulas were divided into 4 categories and selection with uncertain marginal parameters (Parent et al. 2013). Copula is a tool for modeling multivariate distribution in which input is the marginal distribution. Multivariate distribution function couples to the corresponding marginal distribution. (Poulin *et al*., 2007; Salvadori *et al*., 2007). The monsoon rainfall of Assam, Meghalaya and Nagaland, Manipur, Mizoram, Tripura, Gumbel–Hoggard copula model was well simulates for rain fall estimation (Ghosh, 2010). Marginal distributions and correlations values are used to simulate the Gaussian model. They were taken four case studies to demonstrate its usefulness in the reference of determination of field significance analysis, analysis of regional risk , frequency analysis and design of hydrograph derivation by QdF models. (Renard et al. 2007). Copulas are very good tool to model multivariate data and they are very useful in financial economics as well and in the analysis of multivariate survival data. Dependent variables are very useful Monte Carlo simulations for copula model. It estimates the structural dependence of the data set and describe accurately for dependence of extreme out come. (Muhaisen, et al. 2006). Multivariate probability distributions with arbitrary marginal can be constructed in a flexible manner with the introduction of copulas (Wang et al. 2001). Major issue of a copula is the compatibility with dimensions though they were successfully tested and applied on several hydrological problems. (Kao and Govindaraju, 2008). Application of copula in the engineering problem need moderate and minimal computational effort and accuracy of the output is also satisfactory (Kao et al., 2012). For two copula approach the spatial dependence of rainfall dependence in sub-basins decreases up to 18 %. To predict decrease runoff error spatial rainfall dependence could be recommended for copula modeling (Razmkhah, 2016).

The aim of this paper is to generate the out flow discharge data at Farakka barrage using Copulas. In this study, Normal Copula, T- Copula, Frank Copula, Clayton Copula, Gumbel-Hoggard (GH) copula, Ali-Mikhail-Haq(AMH) copula are used and best copula is selected for generation of discharge data based on copula parameters, Mean square error(MSE), Akaike Information criterion(AIC), Bayesian Information criterion (BIC).

ARIMA model was developed to forecast monthly inflow discharge in a reservoir system (Mohan et al., 1955). Criteria for model selection are residual variance(Katz et al. 1981), Akaike information criteria (Akaike 1974) and Posterior probability criteria (Kashyap 1977).

2. Copulas used for study

Copulas are alternative methods for dealing with multivariate extremes, and these are very popular in recent times. Consider a moment pair of random variables U and V, with their distribution functions $F(u) = P[U \le u]$ and $G(v) = P[V \le v]$, respectively, and a joint distribution function $H(u, v) = P[U \le u, V \le v]$. Each pairs having of real numbers (u, v), associated three numbers: F(u), G(v), and H (u, v) and each numbers are lie in the interval [0,1]. In other words, each pair of real numbers i.e. (u, v) leads to a point {F(u), G(v)} in the unit square $[0, 1]\times[0, 1]$, and this ordered pair in turn corresponds to a number H(u, v) in [0,1]. We will show that this correspondence, those values are assign in the joint distribution function to each values of ordered pair in the individual distribution functions. Such functions are named as copulas.

A copula is used as a tool in modeling multivariate distribution in which marginal distributions are input data and neglect restrictions mentioned in pervious text. Copula means couples or joins

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences
Open Access
Discussions

multivariate distribution functions to their corresponding distribution functions of their corresponding marginal distribution functions (Poulin *et al*., 2007; Salvadori *et al*., 2007). Definition which is given below is given by Sklar (1959), if p-dimensional distribution function then F can be written as:

$$F\,(u_1, u_2, u_3\ldots\ldots\ldots\ldots, u_p) = C(F(u_1), F(u_2), F(u_3)) \qquad (1)$$

where F_1, ……., F_p= Marginal distribution functions. If F_1, ……. F_p are continuous then the copula C is unique and has the representation (Poulin *et al*., 2007):

$$C(x_1, x_2,\ldots\ldots\,.\; x_p) = F(F^{-1}(x_1),F^{-1}(x_2),\ldots\ldots\ldots\ldots\; F^{-1}(x_p)), \qquad (2)$$

$$0 \le x_1, \ldots\ldots\ldots,x_p \le 1$$

Copula is expressed for two random variables**,** U and V, with their CDFs, respectively, as $F_u(u)$ and $F_v(v)$, let $X = F_u(u)$ and $Y = F_v(v)$, Where, X and Y are random variables which is uniformly distributed with their values x and y. The list copulas and its equations with generating function is shown in Table 1.

Table 1. : List of Copulas and its equation, generating function and relation with τ.

S. No.	Copula	Equation	Generating function	Relation with τ
1	Normal	$C(x_1,x_2,\ldots\ldots\ldots\; x_p) = P[U_1 \le F^{-1}{}_1(x_1),U_2 \le F^{-1}{}_2(x_2),\ldots\ldots, U_p \le F^{-1}{}_p(x_p)]$		
2	T	$C(x_1, .\;\;.\;\;.\;\;, x_d) = F(F^{-1}{}_1\,(x_1), .\;\;.\;\;.\;\;, F^{-1}{}_d\,(x_d))$		
3	Frank	$C_\Theta(x,y) = \frac{1}{\theta}\ln[1+\frac{[\exp(\theta x)-1][\exp(\theta y)-1]}{\exp(\theta)-1}$	$\Phi(t) = \ln[\frac{\exp(\theta t)-1}{\exp(\theta)-1}]$	$\tau = 1-\frac{4}{\theta}[D_1(-\Theta) -1]$
4	Clayton	$C_\Theta(x,y) = [x^{-\Theta} + y^{-\Theta} -1]^{-1/\Theta}$	$\Phi(t) = t^{-\Theta} - 1$	$\tau = \frac{\theta}{\theta+1}$
5	Gumbel-Hoggard	$C_\Theta(x, y) = \exp\{-[(-\ln x)^{\Theta} +(-\ln y\,)^{\Theta}]^{1/\Theta}\}$	$\Phi(t) = (-\ln t)^{\Theta}$	$\tau = 1- \Theta^{-1}$
6	Ali-Mikhail-Haq	$C_\Theta(x,y) = \frac{xy}{1-\theta(1-x)(1-y)}$	$\Phi(t) =\ln[\frac{1-\theta(1-t)}{t}]$	$\tau = (\frac{3\theta-2}{\theta}) -2/3\,(1-1/\Theta)^2 \ln(1-\Theta)$

Where,

ϴ= Parameter which controlling the dependence between x and y.

Φ= Generator of the copulas.

Debye function is expressed as follows.

$$D_n(\beta,x) = \frac{n}{x^n}\int_0^x \frac{t^n}{(e^t-1)^\beta}\,dt \quad (3)$$

$$D_1(1,\Theta) = \frac{1}{\Theta}\int_0^\Theta \frac{t}{e^t-1}\,dt \quad (4)$$

3. Dataset used for Copulas

Mean monthly discharge at Farakka barrage data set about twenty-five years from 1949 to 1973 data has taken from Water Resources Information System of India at Farakka barrage project, Farakka, West Bengal, India.

The observed data set are divided into two parts. One part contains twenty years' data (from 1949 to 1968) has been used for parameter estimation i.e. in model calibration, next five years' data (from1969 to 1973) has been used for model validation and testing. Parameter estimation data is arranged such a way that pre-monsoon (December to May) and post monsoon (June to November) data is separated and making two series of dataset for copulas.

4. Selection of distribution for Copulas

For modeling of controlled outflow, bivariate Copula has taken in this study. As Copula accepts CDF of variables, distribution functions of two variables, should be known. The distribution functions are chosen on the basis of AIC, BIC values, k-s test and probability plots. The distributions that are tested to know the parent distribution of two variables are normal distribution, lognormal distribution, extreme value type I distribution, generalized extreme value distribution, gamma distribution, weibull and exponential distributions. We used data set for different times i.e., from Dec. -May 1949 to Dec. -May 1968 (Figure 1), from Jun. -Nov. 1949 to Jun. -Nov. 1968 (Figure 2), Dec. -May 1949 to Dec. -May 1968 (Figure 3), Jun. -Nov. 1949 to Jun. -Nov. 1968 (Figure 4), Dec. -May 1949 to Dec. -May 1968 (Figure 5), Jun.-Nov.1949 to Jun. -Nov. 1968 (Figure 6).

The violet colour represents the data set for different times and red colour represents normal distribution, green colour represents lognormal distribution, etc as shown in Figures 1-6. Figure 1 represents cumulative distribution function of data points along with all distributions in pre monsoon seasons (Dec.- May 1949 to Dec.-May 1968).

Figure 2 represents cumulative distribution function of data points along with all distributions in post monsoon seasons (Jun. - Nov. 1949 to Jun. - Nov 1968). Figure 3 represents Probability density function of data points along with all distributions in post monsoon seasons (Dec.- May 1949 to Dec.-May 1968). Figure 4 represents Probability density function of data points along with all distributions in post monsoon seasons (Jun. -Nov. 1949 to Jun. -Nov. 1968). Select the standard distribution which is best fit for original data sets. Violet colour of Figure 5 represents the data points and other colour represents the various distributions of mean monthly discharge (Dec. -May 1949 to Dec. -May 1968) Select the best fit standard probability distribution. Violet colour of this Figure 6 represents the data points and other colour represents the various distributions of mean monthly discharge (Jun.-Nov.1949 to Jun. -Nov. 1968). Selection of the distribution function can be based the best fit for original data sets.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences
Open Access
Discussions

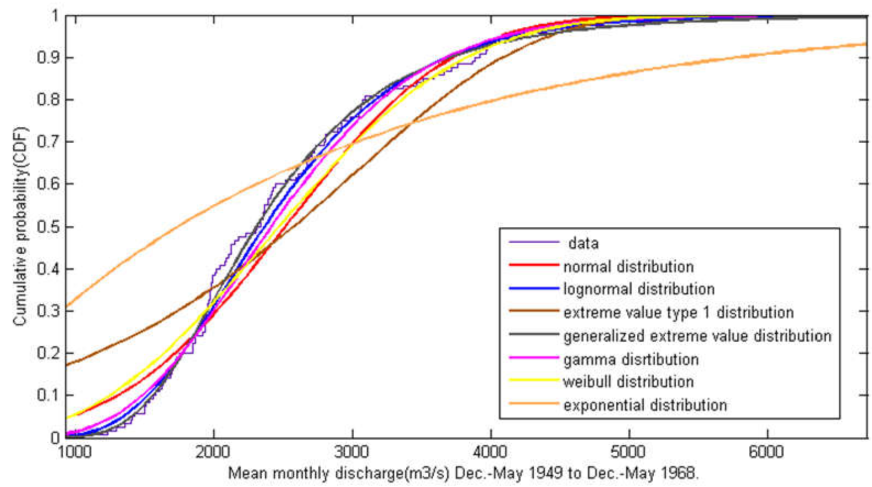


Figure 1. CDF of mean monthly discharge(m^3/s) Dec. -May 1949 to Dec. -May 1968.

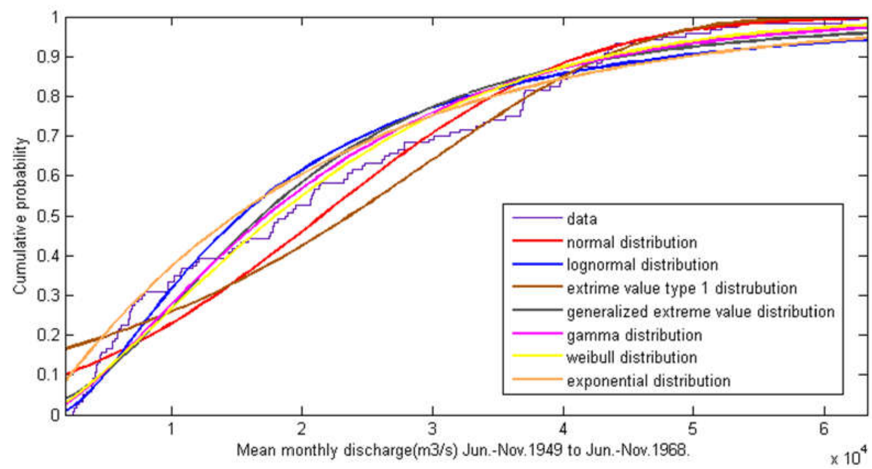


Figure 2. CDF of mean monthly discharge (m^3/s) Jun. -Nov. 1949 to Jun. -Nov. 1968.

Hydrology and Earth System Sciences
Open Access
Discussions

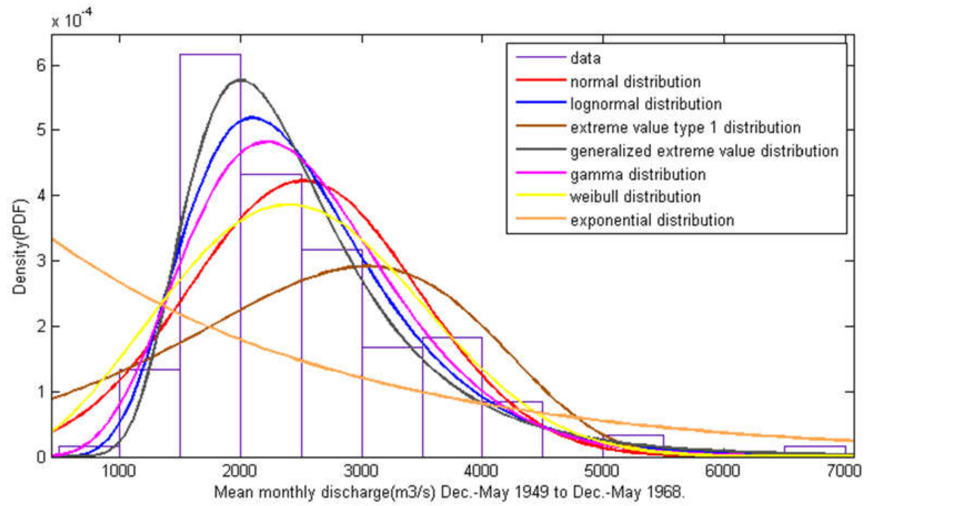


Figure 3. PDF of mean monthly discharge(m^3/s) Dec. -May 1949 to Dec. -May 1968.

192
193

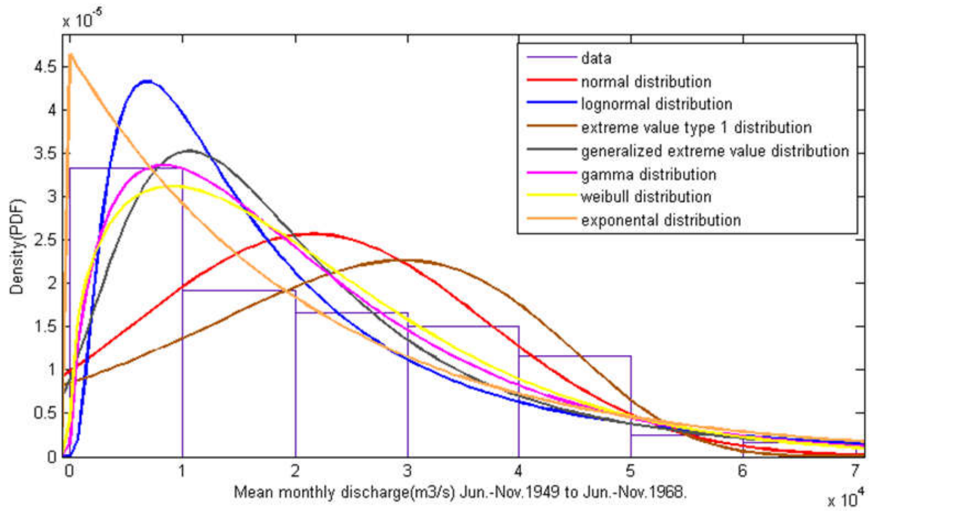


Figure 4. PDF of mean monthly discharge (m^3/s) Jun. -Nov. 1949 to Jun. -Nov. 1968.

194
195

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences
Discussions

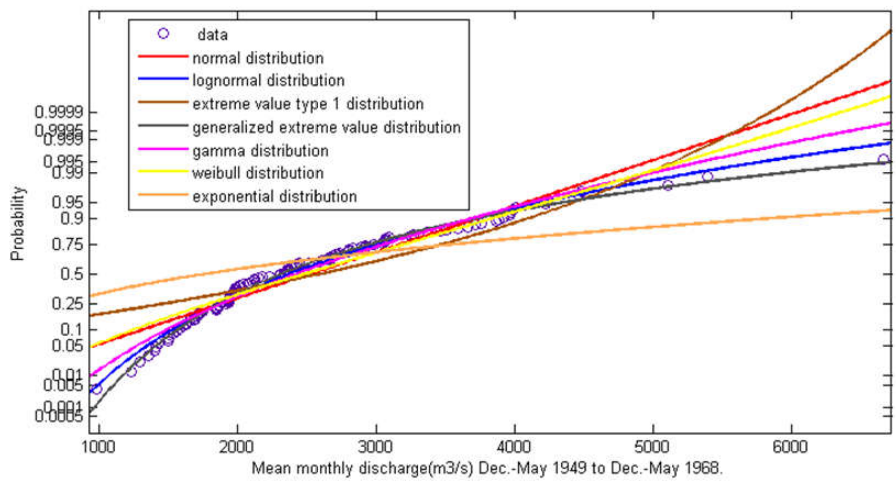


Figure 5. Probability of mean monthly discharge (m^3/s) Dec. -May 1949 to Dec. -May 1968.

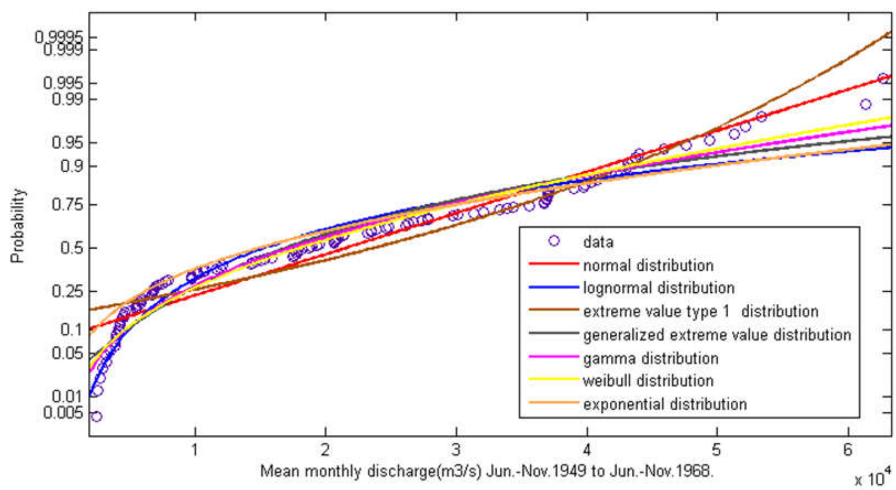


Figure 6. Probability of mean monthly discharge(m^3/s) Jun.-Nov.1949 to Jun. -Nov. 1968.

4.1 Mean square Error (MSE) or Mean squared deviation (MSD)

It is measurement of the mean of the squares of the errors or deviations i.e., the difference between the estimator and what is estimated value (Table 2). MSE represents the risk function corresponding to the expected value of the squared error loss or quadratic loss. The difference in the MSE because of randomness. Lowest value of AIC is good for model.

$$\text{MSE} = \Sigma\frac{(e_{cdf} - p_{cdf})^2}{n} \quad (5)$$

Where,

e_{cdf} = Empirical Cumulative Density Function

P_{cdf} = Predicted Cumulative Density Function

4.2 Akaike Information Criterion (AIC)

For a given data set and given set of models . AIC measures relative quality of statistical methods and it compute the each model's quality, relative to other models quality (Table 2). Hence, AIC criteria is used for model selection and lowest value of AIC is proffered for model. .

$$\text{AIC} = n*\ln(\text{MSE}) + 2K + \frac{2K*(K+1)}{n-K-1} \quad (6)$$

Where,

n = Number of data points.

K = Number of parameters.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

4.3 Bayesian Information Criterion (BIC)

It is a model selection criterion , model is selected among the finite set of model. Model with lowest value of BIC is preferred (Table 2). It is mainly based on likelihood function and it having approximate same conditions as Akaike information criterion (AIC).

$$BIC = n*\ln(MSE) + K*\ln(n) \quad (7)$$

Where,

n = Number of data points.

K = Number of parameters.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Table 2. Statistic of distributions of data.

Data	Distribution	MSE	AIC	BIC
June-Nov.	Normal	0.19185140	-194.02150	-188.54911
Dec. -May	Normal	0.18211572	-200.27095	-194.79857
June-Nov.	Lognormal	0.19240251	-193.67728	-188.20489
Dec. -May	Lognormal	0.19197824	-193.94219	-188.46980
June-Nov.	Extreme value type 1	0.17340721	-206.15091	-200.67853
Dec. -May	Extreme value type 1	0.14660442	-226.29948	-220.82709
June-Nov.	**Gen.extreme value**	**0.09875020**	**-271.61252**	**-263.45694**
Dec. -May	**Gen. extreme value**	**0.09914008**	**-271.13967**	**-262.98409**
June-Nov.	Gamma	0.19490464	-192.12678	-186.65439
Dec.-May	Gamma	0.19024585	-195.02997	-189.55758
June-Nov.	Weibull	0.19556132	-191.72315	-186.25077
Dec.-May	Weibull	0.17272273	-206.62552	-201.15314
June-Nov.	Exponential	0.16679494	-212.88491	-210.13132
Dec.-May	Exponential	0.11907597	-253.32532	-250.57173
June-Nov.	Kernel_normal	0.16931844	-211.08299	-208.32939
June-Nov.	Kernel_box	0.16880085	-211.45037	-208.69678
June-Nov.	Kernel_triangle	0.16920205	-211.16550	-208.41191
June-Nov.	Kernel_epanechnikov	0.16901130	-211.30086	-208.54727
Dec. -May	Kernel_normal	0.18343705	-201.47214	-198.71855
Dec. -May	Kernel_box	0.18311280	-201.68445	-198.93085
Dec. -May	Kernel_triangle	0.18327642	-201.57727	-19882367
Dec. -May	Kernel_epanechnikov	0.18321604	-201.61681	-198.86322

4.4 Kolmogorov – Smirnov test

The Kolmogorov–Smirnov test (K–S test or KS test) is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test) (Table 3). The two-sample K–S test is one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples. In the Figure 7 and 8, green colour shows empirical CDF and red colour shows generalized extreme value of CDF. On the basis of Figure 7, generalized extreme value distribution is representing best fit for cumulative distribution function (Jun.-Nov.1949 to Jun. -Nov. 1968). Further, on the basis of Figure 8, generalized extreme value distribution is represents best fit for cumulative distribution function (Dec. -May 1949 to Dec. -May 1968).

Table 3. k-s statistics of distributions of data.

			K - S	Test	
Data	**Distribution**	**H**	**p**	**k-s**	**cv**
June-Nov.	Normal	0	0.0509	0.1222	0.1225
Dec.-May	Normal	1	0.0336	0.129	0.1225
June-Nov.	Lognormal	0	0.0691	0.117	01225
Dec.-May	Lognormal	0	0.3695	0.0824	0.1225
June-Nov.	Extreme value type 1	1	0.0015	0.1712	0.1225
Dec.-May	Extreme value type 1	1	0.000055	0.207	0.1225
June-Nov.	**Gen. extreme value**	**0**	**0.092**	**0.1118**	**0.1225**
Dec.-May	**Gen. extreme value**	**0**	**0. 669**	**0.0649**	**0.1225**
June-Nov.	Gamma	0	0.1526	0.1021	0.1225
Dec.-May	Gamma	0	0.1784	0.0989	0.1225

June-Nov.	Weibull	0	0.1162	0.1075	0.1225
Dec.-May	Weibull	0	0.1046	0.1095	0.1225
June-Nov.	Exponential	0	0. 0748	0.1156	0.1225
Dec.-May	Exponential	1	6.16E-17	0.3939	0.1225
June-Nov.	Kernel_normal	1	0. 0286	0.1315	0.1225
June-Nov.	Kernel_box	1	0. 0141	0.1421	0.1225
June-Nov.	Kernel_triangle	1	0.0255	0.1333	0.1225
June-Nov.	Kernel_epanechnikov	1	0.0195	0.1374	0.1225
Dec.-May	Kernel_normal	0	0.8141	0.0567	0.1225
Dec.-May	Kernel_box	0	0.8095	0.057	0.1225
Dec.-May	Kernel_triangle	0	0.8221	0. 0562	0. 1225
Dec.-May	Kernel_epanechnikov	0	0.8074	0.0571	0.1225

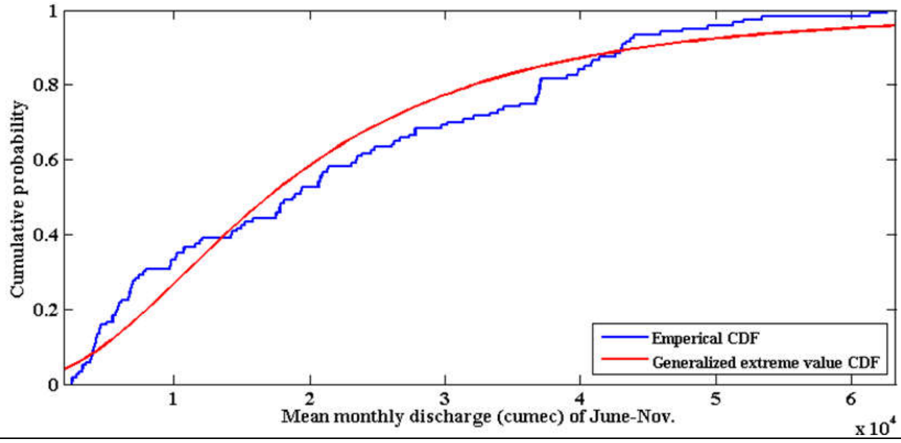


Figure 7. CDF of GEV distribution for Jun.-Nov.1949 to Jun. -Nov. 1968. data.

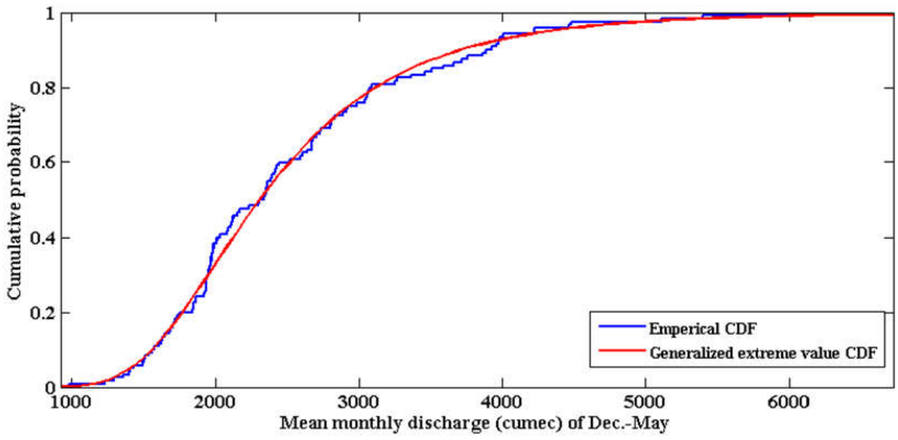


Figure 8. CDF of GEV distribution for (Dec. -May 1949 to Dec. -May 1968) data.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences Discussions

5. Copula parameter estimation

Table 4. : Copulas and its parameter.

	Parameter					
Copula model	rho	nu	MLE	MSE	AIC	BIC
Gaussian	-0.2338		3.38	0.002125	-736.448	-732.69
t	-0.2344	3.79E+06	3.37	0.002126	-734.331	-731.65
Frank	-1.1424		4.557	0.00206	-740.190	-736.44
AMH	-1		2.284	0.002072	-739.477	-735.72
Clayton	1. 45E-06		0.6685	0.002209	-731.817	-728.06
GH	1		-7.2E-07	0.002201	-732.216	-728.46

For a best copula model MLE should be high and MSE, AIC, BIC should be minimum from the above data frank is best model for predicting the data (Table 4). Figure 9 shows the probability density variation from green to red, green having lowest probability density and green colour having maximum probability density. It also represents the probability density function and cumulative distribution function for frank copula which is best for prediction of discharge data.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

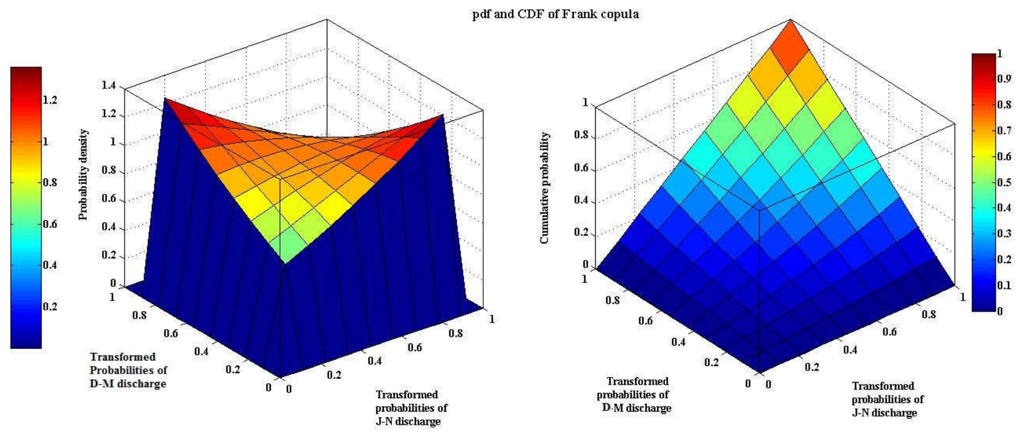


Figure 9. PDF and CDF for Frank copula.

6. Validation test of Copula

Validation test of Frank Copula is performed by comparing observed and empirical CDF in calibration and validation test. Here, observed CDF is CDF of Frank Copula and Empirical CDF is taken from some non-parametric method (Table 5). Formula of empirical CDF of copula is given below. In the Figure 10, the blue points shows data points at calibration and validation state. Blue points represents data points in calibrated and validation stage by Frank copula as shown in Figure 10.

Table 5. **: Statistics in calibration and validation test.**

Statistics	**Calibration Test**	**Validation Test**
MSE	0. 00206	0. 00147
R^2	0. 94	0.9

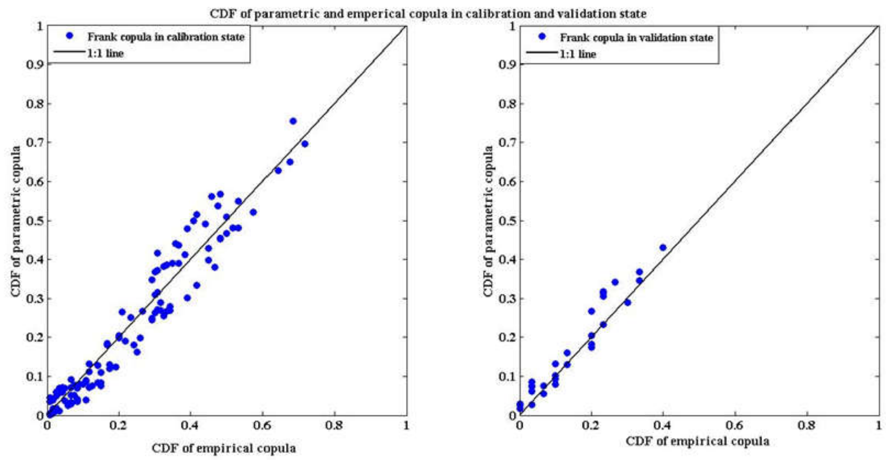


Figure 10. CDF of observed and empirical copula in calibration and validation state.

7. Statistical Approach for ARMA

In this approach linear type stationary ARMA models are fitted by observed discharge data where stationary means the ARMA models that are generated from a time series does not changing its underlying probability distribution function (pdf) from which different values of time series are pulled out. In loose sense stationarity indicates time series has constant mean and variance throughout the process where time series is the collection of random variables, plotted corresponding of its time, follow on their own distribution (figure. 11). In ARMA model AR i.e. auto regressive term indicates lag of time series value and moving average is the lag in error term. Generally, ARIMA is conventional class of model where "I" integration term indicates order of difference required to do the time series stationary but in this study it is done by normalizing all the discharge data through its long term mean and standard deviation (figure. 12). The mathematical form of normalization is given below.

$$\boldsymbol{Z_i} = \frac{X_i - \bar{X}}{\sigma_i} \;\ldots\ldots\ldots \qquad (8)$$

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Where, Xi=Value of mean monthly discharge, $\overline{X}$=Long term average,σi=Long term standard deviation, i=1 to N, N is total number of data point in monthly step. The normalization or differencing in the data is not only make it stationary but also removes periodicity from the time series where periodicity can be defined as correlation i.e. linear association of data with the previous some lag value of data. As we are interested to only capture unknown information from a process which are unknown due to noise or random term (stochastic factor in the process), so deterministic part in terms of long term mean, periodicity, seasonality, trend, sudden drop or jump is necessary to remove from the time series since these deterministic terms already reflects known information about the process, are not required to model. Generally monthly discharge time series shows periodicity and seasonality in the data set and it is necessary to remove before calibrate (finding Parameter of model) to ARMA model as this type of model is developed to capture unknown information from noise i.e. random process.

The observed data set are divided into two parts. One part contains twenty years' data (from 1949 to 1968) has been used for parameter estimation i.e. in model calibration, next five years' data (from1969 to 1973) has been used for model validation and testing. The mean monthly discharge data used for model calibration may have serial correlation i.e. any data in particular time step depends on its previous adjacent data and may follow so on. The time series plot of observe discharge depicts this serial correlation, seasonality or periodicity in terms of information contain in the series by showing some regularity or similar oscillation of the series.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

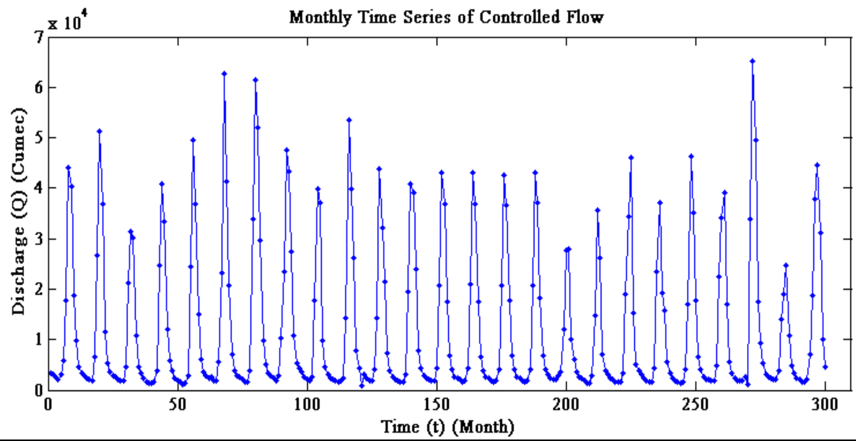


Figure 11: Time series plot of discharge (cumec) starting from (Jan. 1949 to Dec. 1973).

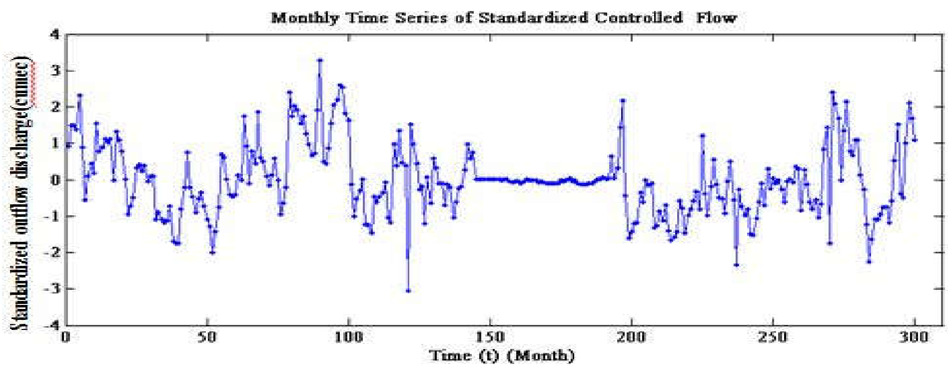


Figure 12: Time series plot of Standardized discharge (cumec) starting from (Jan. 1949 to Dec. 1973).

8. Spectral Analysis

The observe time series is analyzed in frequency domain to indicate exactly in which months periodicity present in the data that is only indicates by correlogram. In this frequency domain

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

analysis an assumption is taken as time series is a random sample of a process over time and is made up of oscillations of all possible frequencies. The time series is approximated by signal process contains deterministic term in wave form and noise or random term by which the information is extracted from time series and shows prominent spike in variance spectrum plot. The contributing equations for spectral analysis are given below

$$X_t = \alpha_0 + \sum_{k=1}^{n-1/2,n/2}[\alpha_k \cos(2\pi f_k t) + \beta_k \sin(2\pi f_k t)] + \varepsilon_t \qquad (9a)$$

$$f_k = \frac{k}{N}\ ;\ P = \frac{1}{f_k};\ \alpha_0 = \bar{x} \qquad (9b)$$

$$\alpha_k = \frac{2}{N}\sum_{i=1}^{n} x_t \cos(2\pi f_k) \qquad k = 1,2,3,\ldots\ldots M \qquad (9c)$$

$$\beta_k = \frac{2}{N}\sum_{i=1}^{n} x_t \sin(2\pi f_k) \qquad k = 1, 2, 3, \ldots\ldots M \qquad (9d)$$

Where;

N = Observation numbers, X_t = Observe rainfall data, t= Time step in month

P = Periodicity in the data, $\overline{X}$ = Mean of the series (average monthly rainfall)

α_k = Cosine wave form, β_k = Sine wave form of time series.

M = Maximum lag typically consider 0.25N.

Values of α_k and β_k in equation number 7 are valid up to k = N/2.

8.1. Line spectrum

The spike in the line spectrum confirms the presence of particular month periodicity in the data (Figure 13 and Table 6)and lime spectrum is plot between spectral density versus angular

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

frequency. It is also known as variance spectrum. Line spectrum plot is drawn by using discharge data and standardized discharge data.

$$I_k = \frac{N}{2}[\,\alpha_k^2 + \beta_k^2\,]; \qquad k = 1, 2,3, \ldots\ldots\ldots, M \tag{10a}$$

$$\alpha_k = \frac{2}{N}\sum_{i=1}^{n} x_t \cos(2\pi f_k) \qquad k = 1,2,3,\ldots\ldots M \tag{10b}$$

$$\beta_k = \frac{2}{N}\sum_{i=1}^{n} x_t \sin(2\pi f_k) \qquad k = 1, 2, 3, \ldots\ldots M \tag{10c}$$

$$\omega_k = \frac{2\pi k}{N}; \quad k = 1, 2, \ldots\ldots\ldots, M \tag{11}$$

Where, ω_k = Angular frequency and I_k = Spectral density.

N = Observation numbers,

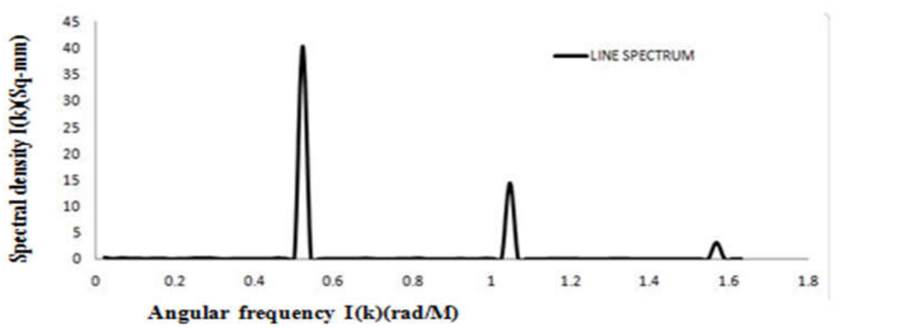


Figure 13: Plot of spectral density versus angular frequency.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

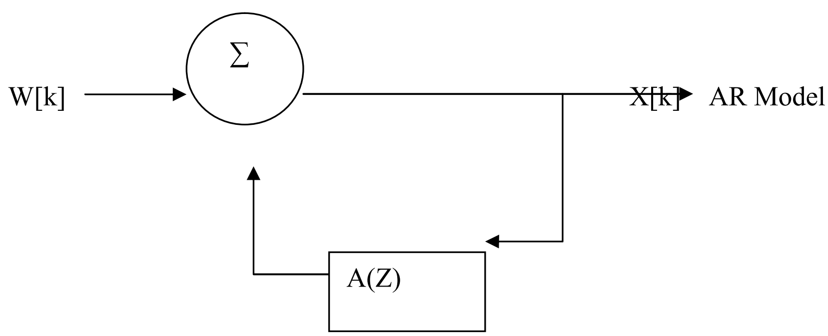
Hydrology and Earth System Sciences
Open Access
Discussions

Table 6: Showing the spectral density and frequency data corresponding to spikes.

Spike	Spectral density (I_k) (sq-mm)	Angular frequency (ω_k)(rad/M)	Periodicity (Months)
1	$4.05 * 10^{10}$	0.52	12
2	$4.48 * 10^{9}$	1.05	6
3	$4.60 * 10^{9}$	1.6	4

9. Model Description

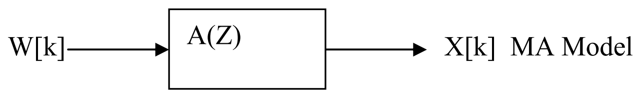
Auto regressive moving average models are developed using white noise series. In the present study the information form observed time series has captured not only developing ARMA (p, q) model but also by pure AR (p) and MA (q) model. The block diagram for AR (p), MA (q) and ARMA (p, q) process are shown below.



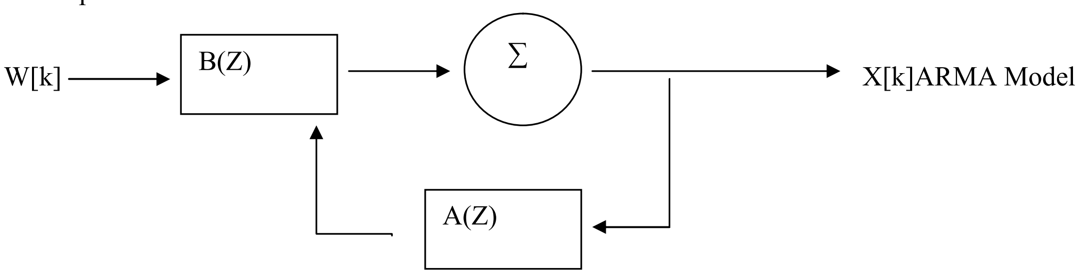
Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences
Open Access
Discussions

X[k] = ∑an*X[k-n] + W[k], where X[k] = Discrete value or kth sequence of random variable (discharge), an = AR parameter for nth order sum over n=1to N, N = number of data point, X[k-n] = nth lag of random variable (discharge), A(Z) = AR polynomial equation, W[k] = Error term associated in the model prediction.



X[k]=∑bn*W[k-n], where X[k]=Discrete value or kth sequence random variable (rainfall), bn = MA parameter for nth order sum over n = 0 to M-1, M = number of error point, W[k-n] = nth lag of white noise or error term, B(Z)=MA polynomial equation, W[k] = Error term associated in the model prediction.



X[k] = ∑an*X[k-n] + ∑bn*W[k-n], where X[k] = Discrete value or kth sequence of random variable (discharge), an = AR parameter for nth order, X[k-n] = nth lag of random variable (discharge) sum over for n=1to N, N=number of data point, a (Z)=AR polynomial equation, W[k] = Error term associated in the model prediction (white noise), bn = MA parameter for nth order sum over n=0 to M-1, M=number of error point, W[k-n] = nth lag of white noise or error term, B(Z)=MA polynomial equation.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

An ARMA (p, q) model which having autoregressive order i.e. p and moving average order i.e. q can be expressed as following from of equation.

$$X_t - \Phi_1 X_{t-1} - \Phi_2 X_{t-2} - \ldots - \Phi_p X_{t-p} = \varepsilon_t + \Theta_1\varepsilon_{t-1} + \Theta_2\varepsilon_{t-2} + \ldots + \Theta_p\varepsilon_{t-p4} \quad (12)$$

$$\Phi(L)X_t = \Theta(L)\,\varepsilon_t\,,\ \Phi(L) = 1 - \sum_{j=1}^{p} \Phi_j L^j \text{ and } \Theta(L) = \sum_{j=0}^{q} \Theta_j L^j \quad (13)$$

Where, Back shift operator $L^j X_t = X_{t-j}$, it shift the value for j th lag.

10. Model Calibration

Two types of model (prediction model) has developed using white noise series but model identification and parameter estimation are not done by conventional Box-Jenkins and Yule-Walker method. In this present study model identification has done by picking up some candidate ARMA model of order up to ten and five for AR and MA process as for most hydrologic cases AR parameter (table 7) and MA parameter (table 8). The model selection is based maximum likelihood estimate (MLE) criteria for prediction model. The underlying equations for MLE criteria for model selection which are used for present study has given in following form.

$$\text{MLE criteria: MLE} = -\frac{N}{2}\ln(\sigma_i) - n_i \quad (14)$$

Where, N is the total data sets those are used for model calibration, σ_i is the variance of residual series where residual is the difference between observe data and corresponding to model output and n_i is the total number of parameter of a model.

The parameter, MLE values for candidate models are shown in table below.

403 **Table 7: Showing only AR parameter for ARMA model.**

			AR PARAMETERS							
MODELS	ϕ1	ϕ2	ϕ3	ϕ4	ϕ5	ϕ6	ϕ7	ϕ8	ϕ9	ϕ10
ARMA(1,0)	0.66451									
ARMA(2,0)	0.60492	0.09057								
ARMA(3,0)	0.60295	0.07795	0.02135							
ARMA(4,0)	0.60080	0.06671	-0.06864	0.15228						
ARMA(5,0)	0.58985	0.07236	-0.07197	0.10985	0.07035					
ARMA(6,0)	0.58063	0.05453	-0.05759	0.10015	-0.01679	0.14551				
ARMA(7,0)	0.58787	0.05451	-0.05024	0.09341	-0.01130	0.18096	-0.06163			
ARMA(8,0)	0.58702	0.05674	-0.05001	0.09497	-0.01299	0.18255	-0.05265	-0.01534		
ARMA(9,0)	0.58727	0.05783	-0.05312	0.09458	-0.01517	0.18504	-0.05505	-0.02823	0.02195	
ARMA(10,0)	0.58604	0.05918	-0.04949	0.08478	-0.01712	0.17845	-0.04629	-0.03598	-0.01817	0.06831
ARMA(1,1)	0.75312									
ARMA(2,1)	1.39970	-0.43496								
ARMA(3,1)	-0.05225	0.52430	0.00826							
ARMA(4,1)	1.24323	-0.31924	-0.11346	0.12318						
ARMA(5,1)	1.33450	-0.37355	-0.12367	0.15158	-0.03725					
ARMA(6,1)	0.28522	0.22950	-0.03610	0.07931	0.01672	0.16687				
ARMA(7,1)	0.74461	-0.03680	-0.05859	0.10082	-0.02648	0.18399	-0.08480			
ARMA(8,1)	-0.20826	0.52546	-0.00582	0.05756	0.05380	0.17932	0.10911	-0.08011		
ARMA(9,1)	-0.19884	0.51957	-0.00903	0.05670	0.05344	0.17970	0.09762	-0.07628	0.01668	
ARMA(10,1)	0.56699	0.07040	-0.04841	0.08390	-0.01537	0.17811	-0.04268	-0.03706	-0.01876	0.06879
ARMA(10,2)	-0.26579	-0.29836	0.50422	0.09413	0.02310	0.25209	0.10769	0.09484	-0.09784	0.05943
ARMA(10,3)	0.51149	-0.07838	0.78081	-0.33277	-0.05451	0.24153	-0.11466	-0.01912	-0.17039	0.13406
ARMA(10,4)	0.38327	0.00193	0.77221	-0.22377	-0.10340	0.23160	-0.08697	-0.02519	-0.17577	0.11709
ARMA(10,5)	-1.13146	-0.06556	0.95650	0.77413	0.07620	-0.10191	0.33892	0.13060	-0.20583	-0.18946
ARMA(1,2)	0.88700									
ARMA(1,3)	0.93403									
ARMA(1,4)	0.93193									
ARMA(1,5)	0.93484									
ARMA(2,2)	-0.02219	0.53793								
ARMA(2,3)	0.21627	0.65017								
ARMA(2,4)	0.19993	0.67908								

ARMA(2,5)	0.23180	0.66432								
ARMA(3,2)	0.64304	0.68470	-0.38491							
ARMA(3,3)	0.42446	0.65453	-0.16819							
ARMA(3,4)	-0.21319	0.15327	0.78757							
ARMA(3,5)	-0.72403	0.74153	0.76954							
ARMA(4,2)	0.69864	0.48430	-0.37004	0.11205						
ARMA(4,3)	-0.37744	1.04098	0.50283	-0.33690						
ARMA(4,4)	-0.42158	1.00669	0.53456	-0.29754						
ARMA(4,5)	0.49824	-0.28955	0.00477	0.62246						
ARMA(5,2)	-0.50475	-0.14324	0.64171	0.13615	0.05173					
ARMA(5,3)	-0.39712	-0.33977	0.55777	0.19782	-0.01648					
ARMA(5,4)	-0.66573	0.97770	0.86047	-0.19813	-0.14756					
ARMA(5,5)	-0.17019	0.61541	-0.04977	-0.06419	0.40870					
ARMA(6,2)	0.35887	-0.35880	0.29042	0.13706	-0.04217	0.24540				
ARMA(6,3)	0.42395	-0.24326	0.70088	-0.18135	-0.08643	0.20389				
ARMA(6,4)	0.39509	-0.20829	0.69730	-0.14510	-0.11156	0.19805				
ARMA(6,5)	0.08679	0.26106	-0.29711	0.48029	0.60613	-0.29824				
ARMA(7,2)	-0.17285	-0.32911	0.46577	0.11014	0.03296	0.19965	0.14650			
ARMA(7,3)	-0.00080	-0.29558	0.54497	0.01053	0.00388	0.20120	0.12022			
ARMA(7,4)	1.36622	-0.64456	0.93643	-0.80636	0.06862	0.28351	-0.21098			
ARMA(7,5)	0.07913	0.23442	-0.31250	0.48226	0.60847	-0.29135	0.02455			
ARMA(8,2)	-0.06562	0.56493	-0.08504	0.05863	0.04780	0.16703	0.08868	-0.09679		
ARMA(8,3)	-0.66504	-0.38677	0.35814	0.33139	0.05602	0.22303	0.23437	0.17652		
ARMA(8,4)	0.96323	-0.34874	0.82988	-0.53430	0.01236	0.20510	-0.08258	-0.08917		
ARMA(8,5)	-1.07677	-0.28472	0.64295	0.49656	0.08941	-0.03977	0.43053	0.33915		
ARMA(9,2)	-0.28074	-0.29950	0.52283	0.11113	0.02688	0.25629	0.13284	0.07980	-0.11560	
ARMA(9,3)	-0.40831	-0.32266	0.46804	0.18178	0.04158	0.25586	0.16121	0.10094	-0.10421	
ARMA(9,4)	-0.62826	0.98023	0.80696	-0.29996	-0.08852	0.21088	0.05200	-0.13123	-0.07542	
ARMA(9,5)	-1.1315	0.65017	0.78081	0.48226	-0.11156	0.28351	0.08868	-0.08917	-0.1758	

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

406 **Table 8: Showing only MA parameter and constant for ARMA model.**

			MA Parameters			Constant
MODELS	θ1	θ2	θ3	θ4	θ5	
ARMA(1,0)						-0.00051
ARMA(2,0)						-0.00129
ARMA(3,0)						-0.00139
ARMA(4,0)						-0.00351
ARMA(5,0)						-0.00456
ARMA(6,0)						-0.00525
ARMA(7,0)						-0.00497
ARMA(8,0)						-0.00477
ARMA(9,0)						-0.00501
ARMA(10,0)						-0.00559
ARMA(1,1)	-0.16221					-0.00215
ARMA(2,1)	-0.82815					-0.00338
ARMA(3,1)	0.66709					0.00311
ARMA(4,1)	-0.66822					-0.00463
ARMA(5,1)	-0.75677					-0.00408
ARMA(6,1)	0.30166					-0.00591
ARMA(7,1)	-0.15794					-0.00470
ARMA(8,1)	0.79691					-0.00455
ARMA(9,1)	0.78870					-0.00475
ARMA(10,1)	0.01917					-0.00562

ARMA(10,2)	0.87471	0.90483				0.00162
ARMA(10,3)	0.19815	-0.77084	0.45834			-0.01064
ARMA(10,4)	0.21564	0.19305	-0.75702	-0.11842		-0.01140
ARMA(10,5)	1.78713	1.23513	-0.25609	-1.00000	-0.67359	-0.03063
ARMA(1,2)	-0.33688	-0.16404				-0.00522
ARMA(1,3)	-0.35536	-0.15329	-0.15629			-0.00608
ARMA(1,4)	-0.35436	-0.15683	-0.15869	0.01476		-0.00611
ARMA(1,5)	-0.35372	-0.15800	-0.15768	0.01995	-0.02017	-0.00608
ARMA(2,2)	0.62654	-0.03828				0.00221
ARMA(2,3)	0.35863	-0.42746	-0.26765			-0.01054
ARMA(2,4)	0.37873	-0.42551	-0.28864	-0.05125		-0.01094
ARMA(2,5)	0.36628	-0.40484	-0.26811	-0.07566	-0.09411	-0.01076
ARMA(3,2)	-0.09104	-0.63486				-0.00590
ARMA(3,3)	0.15828	-0.52132	-0.18806			-0.00865
ARMA(3,4)	0.77956	0.30867	-0.57369	-0.18270		-0.01575
ARMA(3,5)	1.35415	0.09708	-0.76904	-0.36178	-0.14700	-0.01603
ARMA(4,2)	-0.11722	-0.50226				-0.00684
ARMA(4,3)	1.01092	-0.42902	-0.79243			-0.01703
ARMA(4,4)	1.06432	-0.37089	-0.81084	-0.04031		-0.01738
ARMA(4,5)	0.08596	0.39643	0.17633	-0.41673	-0.30529	-0.00931
ARMA(5,2)	1.12739	0.92096				0.01845
ARMA(5,3)	1.02365	1.06681	0.12434			0.01720
ARMA(5,4)	1.31475	-0.15571	-1.00000	-0.28767		-0.01530

ARMA(5,5)	0.78848	-0.10724	-0.07385	0.08787	-0.31487	-0.01845
ARMA(6,2)	0.23672	0.56680				0.00669
ARMA(6,3)	0.16843	0.41609	-0.56455			-0.00732
ARMA(6,4)	0.19811	0.39653	-0.57059	-0.04465		-0.00843
ARMA(6,5)	0.54966	0.11588	0.33089	-0.21262	-0.80621	-0.01506
ARMA(7,2)	0.77331	0.87867				0.00266
ARMA(7,3)	0.59455	0.73423	-0.18396			-0.00062
ARMA(7,4)	-0.78780	0.25034	-0.96854	0.50600		-0.00029
ARMA(7,5)	0.56501	0.13511	0.35062	-0.18961	-0.79343	-0.01516
ARMA(8,2)	0.65268	-0.12361				-0.00555
ARMA(8,3)	1.26941	1.23172	0.38231			0.02268
ARMA(8,4)	-0.37589	0.20412	-0.86433	0.21865		-0.00413
ARMA(8,5)	1.78003	1.50000	0.18756	-0.54292	-0.54283	0.04868
ARMA(9,2)	0.88633	0.91781				0.00390
ARMA(9,3)	1.01837	1.02326	0.13147			0.01121
ARMA(9,4)	1.28045	-0.18837	-1.00000	-0.27298		-0.01673
ARMA(9,5)	1.35415	-0.1533	-0.757	-0.3618	-0.67359	-0.011398

10.1 Maximum likelihood rule

A likelihood value for every of the candidate models (table 9)is calculated which model represents highest likelihood value is chosen for data generation. Gaussian process, general expression of log-likelihood function for the ith model is given below.

$$L_i = \ln(\, p[z, \hat{\varphi}_i]) - n_i \tag{15}$$

It may be approximated as-

$$L_i = -\frac{N}{2}\ln(\sigma_i) - n_1 \quad (16)$$

Where, Li is the likelihood value, z represents a vector of historical series i.e. parameter vector, MA and parameters (θ_1, θ_2,……..; φ_1,φ_2,……;σ_i), σ_i represents the residual variance and n_i is the number of parameters. As the number of parameters increase, the likelihood value decreases.

Table 9: Showing MLE values constant for ARMA model.

Models	**MLE Values**	**Models**	**MLE Values**	**Models**	**MLE Values**
ARMA(1,0)	6.473	ARMA(10,2)	-14.515	ARMA(5,2)	-16.810
ARMA(2,0)	53.207	ARMA(10,3)	-15.273	ARMA(5,3)	-29.974
ARMA(3,0)	4.812	ARMA(10,4)	-17.592	ARMA(5,4)	-37.916
ARMA(4,0)	4.546	ARMA(10,5)	-51.252	ARMA(5,5)	-13.428
ARMA(5,0)	3.268	ARMA(1,2)	5.395	ARMA(6,2)	-39.489
ARMA(6,0)	0.671	ARMA(1,3)	2.531	ARMA(6,3)	-9.516
ARMA(7,0)	-0.729	ARMA(1,4)	1.738	ARMA(6,4)	-9.926
ARMA(8,0)	-1.772	ARMA(1,5)	0.438	ARMA(6,5)	-18.510
ARMA(9,0)	-2.787	ARMA(2,2)	2.875	ARMA(7,2)	-22.491
ARMA(10,0)	-2.796	ARMA(2,3)	-7.629	ARMA(7,3)	-17.672
ARMA(1,1)	6.031	ARMA(2,4)	-7.240	ARMA(7,4)	-20.179
ARMA(2,1)	1.747	ARMA(2,5)	-12.101	ARMA(7,5)	-18.904
ARMA(3,1)	-30.943	ARMA(3,2)	-7.923	ARMA(8,2)	-17.325
ARMA(4,1)	1.620	ARMA(3,3)	-8.838	ARMA(8,3)	-26.489
ARMA(5,1)	0.698	ARMA(3,4)	-9.603	ARMA(8,4)	-13.709

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

ARMA(6,1)	-0.723	ARMA(3,5)	-43.814	ARMA(8,5)	-39.369
ARMA(7,1)	-1.542	ARMA(4,2)	-5.464	ARMA(9,2)	-26.378
ARMA(8,1)	-4.142	ARMA(4,3)	-29.525	ARMA(9,3)	-26.855
ARMA(9,1)	-5.747	ARMA(4,4)	-30.252	ARMA(9,4)	-44.180
ARMA(10,1)	-4.396	ARMA(4,5)	-6.041	ARMA(9,5)	-37.972

11. Model Validation

In the present study ARMA(2,0) (table 9)models has selected as one time step ahead and prediction model by Maximum MLE criteria respectively. The selected model is validate to examine whether the assumptions used for selection of the model are valid.

11.1 Significance of residual mean

This test examines the validity of the assumption that the error series e(t) has zero mean. A statistic η(e) is defined as:

$$\eta(e) = \frac{N^{1/2}\vec{e}}{\rho^{1/2}} \tag{17}$$

Where, $\vec{e}$ = Estimated residual mean.

ρ = Estimated residual variance.

The statistic η(e), approximated distribution as t(α, N–1), α represents the significance level at test is being carried out. If the value of η(e) <t (α, N–1),(table 10) then the mean of the residual series is not significantly different from zero (-)ve series passes the test.

Table 10: Showing the statistic η(e).

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

ρ	**α**	**N**	**t (α, N–1)**	**η(e)**
0.161472	.05	30	1.699	0.1

The value of η(e) <t (α, N–1) (i.e.1.22<1.699). It shows that mean of the residual series is not significantly different from zero (-) series passes the test.

12. Results and Discussion

Outflow data for future are generated by using Frank Copula. The sample size for data generation is taken 100 and 1000. These generated values are compared with observed validation data set. The comparison and the individual dependence of generated samples are also shown in Figure 15, which represents observed data in red color and generated data in blue, data generation by copula with sample size 100 and 1000 of kendall's tau 0.25 and 0.32, respectively. Figure 16 describes dependence structure for 100 generated samples i.e. Copula is a statistical theory on dependence and measurement of association.

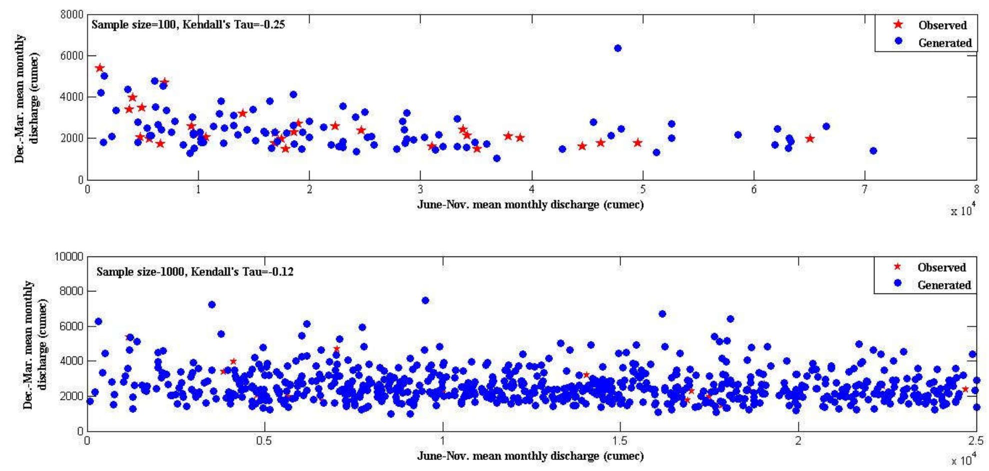


Figure 15. Data generation by Frank Copula.

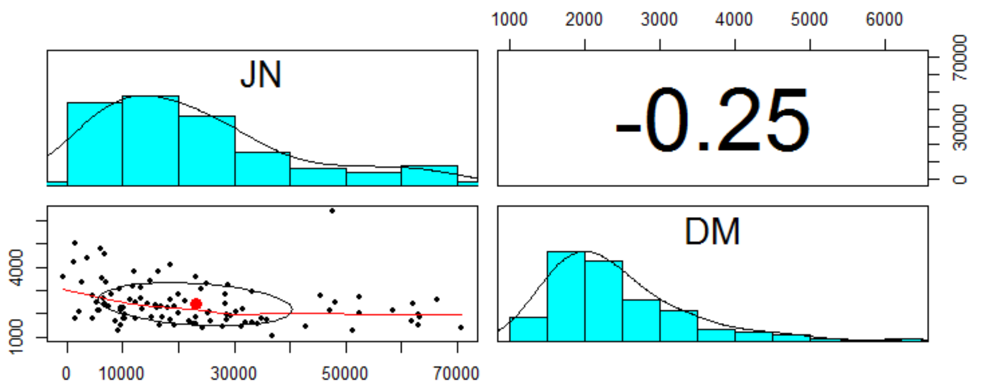


Figure 16. Dependence structure for 100 generated samples.

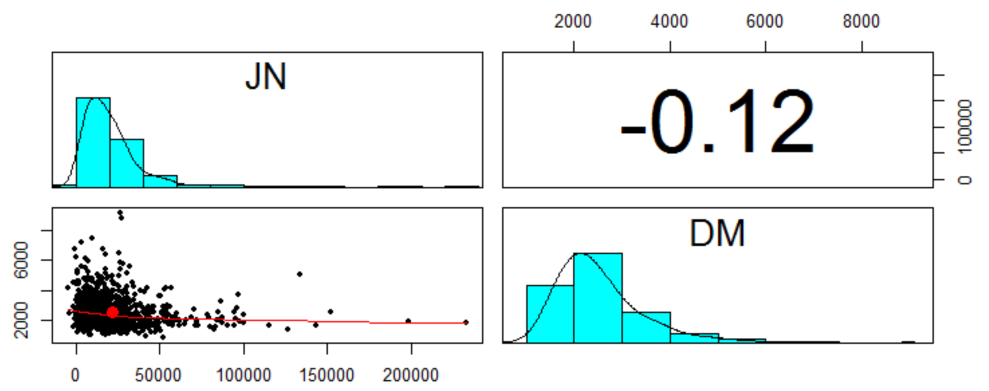


Figure 17. Dependence structure for 1000 generated samples.

Figure 17 describes dependence structure for 1000 generated samples i.e. Copula is a statistical theory on dependence and measurement of association and teal colour shows generated data set month vise. Dependence structure of a multivariate distribution is described by copula, it might be appropriate to use measures of dependence which are copula-based. Linear correlation coefficient can be opposed by the concordance measure spearman's rho and kendall's tau as well as tail dependence and it is expressed by under laying copula.

Figure 20 is a time series of discharge data, blue colour represents observed data from jan. 1968 to dec. 1973 and red colour represents generated data from jan. 1974 to dec. 2004. The green colour shows observed data set and red colour shows generated data set. When generated data set is small, it shows good results because errors incorporated are less in comparison to large data set generation.

In the study area, has analyzed that best model in ARMA (2,0) model, and Frank Copula model for generating discharge data at Farakka barrage on the basis of Mean Square Error (MSE),

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Coefficient of determination (R^2) (Figure 18 and Figure 19). Based upon all above test Frank Copula is the best model for generating outflow discharge data at Farakka barrage.

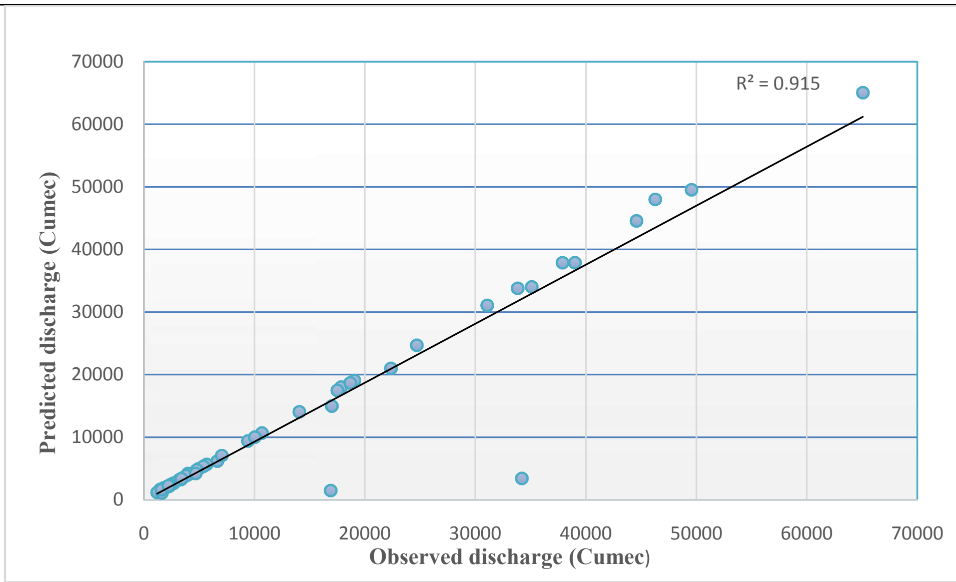


Figure18: Comparison of Observed and predicted data for Frank Copula model.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

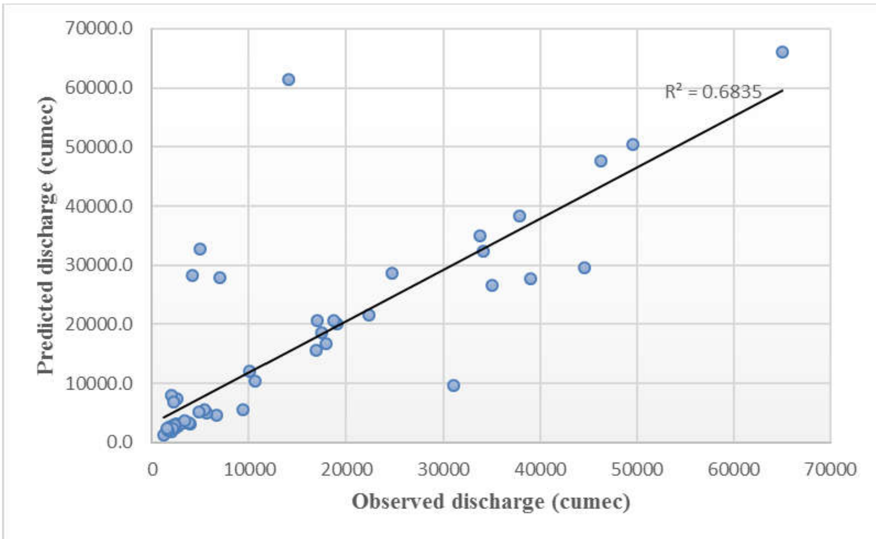


Figure **19** Comparison of Observed and predicted data for ARIMA(2,0) model.

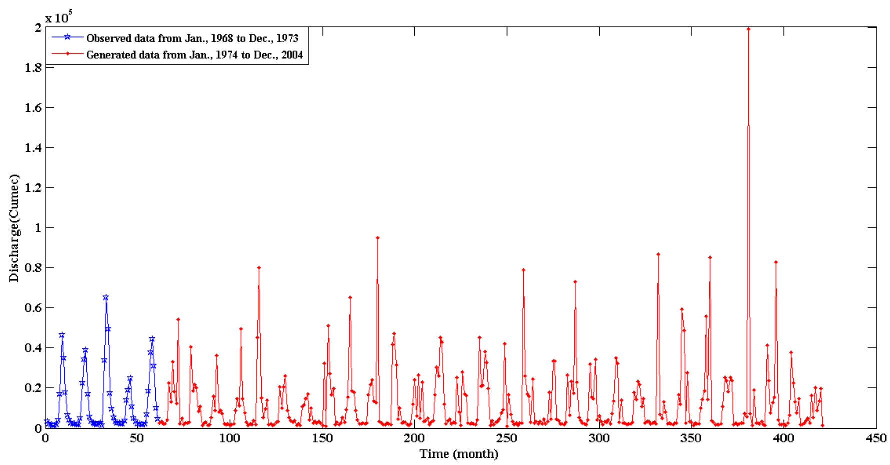


Figure 20. Time series of observed data and generated data.

13. Conclusion

Copula based study and ARMA models are used in this study Frank copula is selected among the copulas based on parameter for discharge data generation and ARMA(2,0) is selected among the ARMA models. Errors incorporated in the copula model is less in comparison to the ARMA(2,0) model and the value of Coefficient of determination (R^2) for Frank is close to one i.e 0.915. Frank copula estimated better result over ARMA(2,0).

A copula based study which can be used to derive bivariate distribution function of flow rate variables and it shows the real world case study. Best suited model for this study is frank copula among all above copula in term of non parametric tests i.e. AIC, MSE, BIC and Kolmogrov-Smirnov test. When generated data sample data set, copula shows convergence of sample data set to estimated population. Copula models are an alternative approach and in this study Frank Copula model is used for data generation at Farakka barrage. Bivariate series are prepared based on pre monsoon and post monsoon outflow data. Moreover they are very useful in this study of dependent variable. Copula is very useful for describing the dependence of extreme outcome because it captures the structural dependence of data. The autocorrelation is not captured in the bivariate model .

References

Akaike, H.: A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723. 1974.

Balistrocchi, M., Orlandini, S., Ranzi, R., & Bacchi, B.: Copula-Based Modeling of Flood Control Reservoirs, *Water Resources Research*, *53*(11), 9883-9900, 2017.

Chowdhary, H., Escobar, L. A., Singh, V. P.: Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. Hydrology Research, 42(2–3), 193–216, 2011.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., & Rosso, R.: Bivariate statistical approach to check adequacy of dam spillway. Journal of Hydrologic Engineering, 10(1), 50-57, 2005.

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B.: Multivariate hydrological frequency analysis using copulas, Water resources research, 40(1), 2004.

Genest, C., Favre, A. C., Béliveau, J., & Jacques, C.: Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. Water Resources Research, 43(9). W09401. DOI:10.1029/2006WR005275, 2007.

Genest, C., Rémillard, B., & Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and economics, 44(2), 199-213. 2009.

Genest, C., & Rivest, L. P.: Statistical inference procedures for bivariate Archimedean copulas, J of the American statistical Association, 88(423), 1034-1043, 1993.

Ghosh, S.: Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula, Hydrological Processes, 24(24), 3558-3567, 2010.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Hydrology and Earth System Sciences
Open Access
Discussions

Goel, N. K., Seth, S. M., & Chandra, S.: Multivariate modeling of flood flows. Journal of Hydraulic Engineering, 124(2), 146-155, 1998.

Hooshyaripor, F., Tahershamsi, A., & Golian, S.: Application of copula method and neural networks for predicting peak outflow from breached embankments. Journal of Hydro-Environment Research, 8(3), 292-303, 2014.

Joe, H.: Multivariate models and multivariate dependence concepts. CRC Press, Chapman and Hall, London, 1997.

Kao, S.-C., and Govindaraju, R. S.: Trivariate statistical analysis of extreme rainfall events via Plackett family of copulas, Water Resour. Res., 44(2), W02415, 2008.

Kao, S. C., & Chang, N. B.: Copula-based flood frequency analysis at ungauged basin confluences: Nashville, Tennessee, Journal of Hydrologic Engineering, 17(7), 790-799, 2011.

Karmakar, S., and Simonovic, S. P.: "Bivariate flood frequency analysis. Part 2: A copula-based approach with mixed marginal distributions", J. Flood Risk Manage., 2(1), 32–44,2009.

Kashyap, R.: A Bayesian comparison of different classes of dynamic models using empirical data, IEEE Transactions on Automatic Control, 22(5), 715-727,1977.

Katz, R. W., & Skaggs, R. H.: On the use of autoregressive-moving average processes to model meteorological time series, Monthly Weather Review, 109(3), 479-484,1981.

Li, T., Guo, S., Chen, L., & Guo, J.: Bivariate flood frequency analysis with historical information based on copula, Journal of Hydrologic Engineering, 18(8), 1018–1030,2013.

Muhaisen, O. S., Osorio, F., & García, P. A.: Two-copula based simulation for detention basin design,Civil Engineering and Environmental Systems, 26(4), 355-366,2009.

Mohan, S., & Vedula, S.: Multiplicative seasonal ARIMA model for longterm forecasting of inflows,Water resources management, 9(2), 115-126,1995.

Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-380
Manuscript under review for journal Hydrol. Earth Syst. Sci.
Discussion started: 12 December 2018

Osorio, F., Muhaisen, O., & García, P. A.: Copula-based simulation for the estimation of optimal volume for a detention basin, *Journal of Hydrologic Engineering*, *14*(12), 1378-1382,2009.

Poulin, A., Huard, D., Favre, A. C., & Pugin, S.: Importance of tail dependence in bivariate frequency analysis,Journal of Hydrologic Engineering, 12(4), 394-403,2007.

Razmkhah, H., AkhoundAli, A. M., Radmanesh, F., & Saghafian, B.: Evaluation of rainfall spatial correlation effect on rainfall-runoff modeling uncertainty, considering 2-copula,*Arabian Journal of Geosciences*, *9*(4), 323,2016.

Renard, B., & Lang, M.: Use of a Gaussian copula for multivariate extreme value analysis: some case studies in hydrology, *Advances in Water Resources*, *30*(4), 897-912,2007.

Requena, A. I., Chebana, F., & Mediero, L.: A complete procedure for multivariate index-flood model application, *Journal of Hydrology*, *535*, 559-580,2016.

Requena, A. I., Flores, I., Mediero, L., & Garrote, L.:Extension of observed flood series by combining a distributed hydro-meteorological model and a copula-based model, Stochastic environmental research and risk assessment, 30(5), 1363-1378,2016.

Requena AI, Mediero L, Garrote L.: A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation, Hydrol Earth Syst Sc 17:3023–3038,2013.

Requena, A. I., Prosdocimi, I., Kjeldsen, T. R., & Mediero, L.: A bivariate trend analysis to investigate the effect of increasing urbanisation on flood characteristics, Hydrology Research, 48(3), 802-821,2017.

Salvadori, G., & De Michele, C.: On the use of copulas in hydrology: theory and practice, Journal of Hydrologic Engineering, 12(4), 369-380,2007.

Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R.: Extremes in nature: an approach using copulas (Vol. 56), Springer Science & Business Media,2007.

Shiau, J. T., Feng, S., & Nadarajah, S.: Assessment of hydrological droughts for the Yellow River, China, using copulas, Hydrological Processes, 21(16), 2157-2163,2007.

Sklar A.: Fonction de re'partition a' n dimensions et leurs marges, vol. 8. Publications de L'Institute de Statistique, Universite' de Paris: Paris; 229–231,1959.

Twaróg, B.: An assessment of risks posed by control rule parameters implemented in a flood control reservoir, carried out with the application of elements of ruin theory and of bivariate distribution of a random variable based on the copula function, American Journal of Environmental Engineering, 6(2), 62-71,2016.

Vrugt, J. A., Diks, C. G., Gupta, H. V., Bouten, W., & Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water resources research, 41(1),2005.

Wang, Q. J.: A Bayesian joint probability approach for flood record augmentation, Water Resour. Res., 37(6), 1707–1712,2001.

Zhang L, Singh V.P.: Trivariate flood frequency analysis using the Gumbel–Hougaard copula, J. Hydrol Eng 12:431–439,2007.

Zhang, L. S. V. P., & Singh, V. P.: Bivariate flood frequency analysis using the copula method, Journal of hydrologic engineering, 11(2), 150-164,2006.