MODELLING RAINFALL IN KARKHEH DAM RESERVOIR OF IRAN USING TIME SERIES ANALYSIS (STOCHASTIC ARIMA MODELS)

Karim Hamidi Machekposhti[1], Hossein Sedghi[2], Abdolrasoul Telvari[3] & Hossein Babazadeh[4]
[1]Department of Water Sciences and Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, PO Box 1477893855
[2]Department of Water Sciences and Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, PO Box 1477893855
[3]Department of Civil Engineering, Islamic Azad University, Ahvaz, Iran, PO Box 3733361349, Iran
[4]Department of Water Sciences and Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran, PO Box 1477893855
h.sedghi1320@gmail.com

ABSTRACT

Machekposhti, Karim Hamidi, Hossein Sedghi, Abdolrasoul Telvari & Hossein Babazadeh. 2017. Modelling Rainfall in Karkheh Dam Reservoir of Iran using Time Series Analysis (Stochastic ARIMA Models). Lebanese Science Journal. Vol. 18, No. 2: 204-218.

Time series analysis and prediction has become a major tool in different applications in meteorological and hydrological phenomena, such as rainfall, temperature, evaporation, flood, drought etc. Among the most effective approaches for analyzing time series data is the auto regressive integrated moving average (ARIMA) model introduced by Box and Jenkins. In this study we used Box-Jenkins methodology to build non-seasonal ARIMA model for annual rainfall data of Karkheh dam reservoir in Iran for Jelogir Majin and Pole Zal stations (upstream of Karkheh dam reservoir) for the period 1966-2015. In this paper, ARIMA 8.1.1 and 9.1.1 models were found adequate for annual rainfall at Jelogir Majin and Pole Zal stations, respectively, and these models were used to predict the annual rainfall for the coming ten years to help decision makers to establish priorities in terms of water demand management. The statistical analysis system (SAS) and statistical package for the social science (SPSS) softwares were used to determine the best model to use for these series.

Keywords: time series analysis, ARIMA model, rainfall prediction, Karkheh dam reservoir.

INTRODUCTION

Water is one of the most important natural resources and a key element in the socio- economic development of a country. Generally, water resources of the world and in Iran are under heavy stress due to increased demand and limitation of available quantity. Proper water management is the only option that ensures a squeezed gap between the demand and supply. Rainfall is the major component of the hydrologic cycle and this is the primary source of runoff. Worldwide, many attempts have been made to model and predict rainfall behavior using various empirical, statistical, numerical and deterministic techniques. They are still in research stage and needs more focused empirical approaches to estimate and predict rainfall accurately. Runoff generated by rainfall is very important in various activities of water resources development and management. The method of transformation of rainfall to runoff is highly complex, dynamic, nonlinear, and exhibits temporal and spatial variability. It is further affected by many parameters and often inter-related physical factors. Determining a robust relationship between rainfall and runoff for a watershed has been one of the most important problems for hydrologists, engineers and agriculturists. Furthermore, rainfall is essentially required to fulfil various demands including agriculture, hydropower, industries, environment and ecology. It is implicit that the rainfall is a natural phenomenon occurring due to atmospheric and oceanic circulation. It has large variability at different spatial and temporal scales. However, this input is subjected to uncertainty and stochastic errors (Mahsin *et al*., 2012).

http://dx.doi.org/10.22453/LSJ-018.2.204-218

lsj.cnrs.edu.lb/vol-18-no-2-2017/

Prediction of annual rainfall is significantly important in water resources management and crop pattern design. The annual rainfall data may be used to predict rainfall by time series analysis. The main development of time series models is done by Box and Jenkins (1976). Many attempts have been made in the recent past to model and predict rainfall using various techniques, with the use of time series techniques proving to be the most common. In time series analysis it is assumed that the data consists of a systematic pattern (usually a set of identifiable components) and random noise (error) which usually makes the pattern difficult to identify. Time series analysis techniques usually involve some method of filtering out noise in order to make the pattern more salient. The time series patterns can be described in terms of two basic classes of components: trend and seasonality. The trend represents a general systematic linear or (most often) nonlinear component that changes over time and does not repeat or at least does not repeat within the time range captured by the data. The seasonality may have a formally similar nature; however, it repeats itself in systematic intervals over time. Those two general classes of time series components may coexist in real-life data. The ARIMA model is an important predicting tool, and is the basis of many fundamental ideas in time series analysis. An auto-regressive model of order p is conventionally classified as AR (p) and a moving average model with q terms is known as MA (q). A combined model that contains p auto-regressive terms and q moving average terms is called ARMA (p.q). If the object series differed in time to achieve stationarity, the model is classified as ARIMA (p.d.q), where the letter "I" signifies "integrated". Thus, an ARIMA model is a combination of an auto-regressive (AR) process and a moving average (MA) process applied to a non-stationary data series. ARIMA modelling has been successfully applied in various water and environmental management applications (Meher and Jha, 2013).

Gurudeo and Mahbub (2010) applied time series analysis for rainfall and temperature interactions in coastal catchments of Queensland, Australian. They implied that ARIMA model is suitable for prediction of these series. Eni and Adeyeye (2015) applied seasonal ARIMA modelling for forecast of rainfall in Warri town, Nigeria. The ARIMA 1.1.1 $(0.1.1)_{12}$ model fitted to this series with AIC value of 281. Model adequacy checks showed that the model was appropriate. Coefficient of the fitted model was finalized by the residual tests (Eni and Adeyeye, 2015). Wang *et al.* (2014) used the improved ARIMA model to predict the monthly precipitation at the Lanzhou station in Lanzhou, China. The results showed that the accuracy of the improved model is significantly higher than the seasonal model (Wang *et al*., 2014). Mahsin *et al*. (2012) used Box-Jenkins methodology to build seasonal ARIMA model for monthly rainfall data taken for Dhaka station for the period 1981-2010 with a total of 354 readings (Mahsin *et al*., 2012). Mirmousavi *et al*. (2014) studied precipitation behavior in Khoi meteorological station using statistical methods (Mirmousavi *et al*., 2014). They found that ARIMA 1.1.0 model was the best fitted to annual precipitation. Based on this model, annual precipitation was predicted in 95 percent level by 2016 in this station. During recent decades, several researchers have developed methods of analyzing stochastic characteristics of rainfall time series (Ansari, 2013; Akpanta *et al*., 2015; Sayemuzzaman and Jha, 2014; Soltani *et al*., 2007 and Srikanthan and McMahon, 2001).

The objective of this study was modelling and prediction of the annual rainfall data of Karkheh dam reservoir in Iran for Jelogir Majin and Pole Zal stations (upstream of Karkheh dam reservoir) by stochastic ARIMA model using Box-Jenkins approach in order to predict future rainfall values by the best ARIMA model and identify whether the annual rainfall had significantly changed during the period 1966 to 2015 under the impacts of climate change and human activities. In addition, ARIMA models which were found adequate, were used to predict the annual rainfall for the coming ten years to help decision makers to establish priorities in terms of water demand management.

Study Area and Data Collection

Karkheh dam is a multipurpose hydro development designed to control the Karkheh river flow for irrigation, electric power generation and for partial accumulation of extreme Karkheh river inflows into Karkheh dam. Krakheh dam was constructed on the Karkheh River in the West of Iran, 24 kilometer upstream from Andimeshk town in Khuzestan province, northwest Iran. In 2001 the project was completed. The project generates 400 Mw of electrical power from performing its flood control function. Khuzestan province of Iran gets the benefit of irrigation water from its reservoir. Figure 1 shows the study area location. The mean rainfall in Karkheh dam reservoir for the entire period of record of 50 years from 1966 to 2015 in Jelogir Majin and Pole Zal hydrometric stations (stations number 9 and 10 in figure 1) are plotted as shown in Figures 2 and 3. This data were obtained from Iran Water Resources Management Organization (IWRMO). The basin's climate is best described as Mediterranean, having mild/wet winters and hot/dry summers, with mean annual precipitation ranging from 150 mm in the southern arid plains to 750 mm in the northern mountains. The Karkheh River is directly connected to the Karkheh dam, the largest surface reservoir in the region, which has an important role in supplying water to the region.

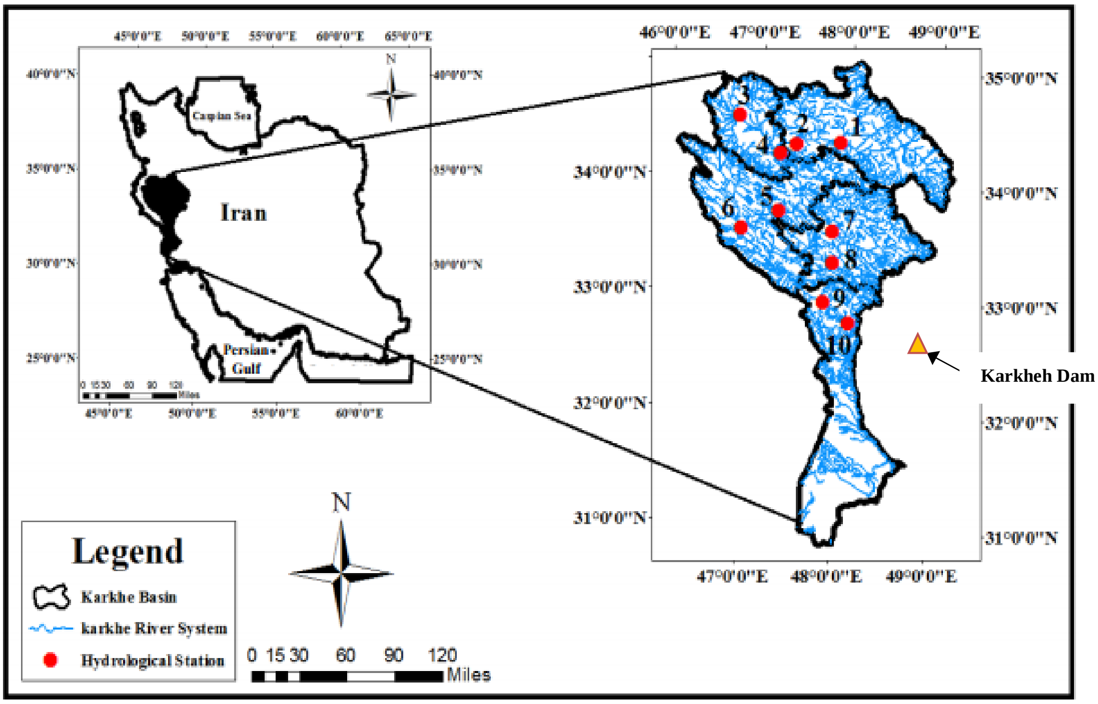

Figure 1. The location of the investigated area

Materials and Methods

Typically various statistical tools are applied to auto-correlate time series of data for modelling and prediction of future values in the annual rainfall series. In general, auto-regressive (AR), moving average (MA), auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA) models are applied to time series of rainfall data. A model which depends only on previous outputs of a system to predict an output is called an auto-regressive (AR) model. While a model which depends only on inputs to the system to predict an output is called a moving average (MA) model. The model derived from auto-regressive and moving average processes may be a mixture of these two and of higher order than one as well, which is termed as a stationary ARMA model with its random shocks independent and normally distributed with zero mean and constant variance. An ARMA model can be noted as ARMA (p, q) where p is the number of AR parameters and q is the number of MA parameters. It models only stationary data set i.e. the data set for which residuals are independent for all time period. It is understood that all hydrologic data sets are not stationary in real life. Some data sets follow some sort of persistence or auto-correlation which is inherently present in it and cannot be removed. Some sorts of trend and cycle are also associated with such data sets. Modelling of non-stationary data sets with ARMA model does not look appropriate because ARMA model is not applicable to non-stationary series. Non-stationary data sets can be modeled through auto-regressive and moving average processes through differentiation i.e. the data set is differentiated until it becomes stationary. For real life data required differentiation at maximum order of three to make it stationary. In other words, ARMA models rely upon the assumption that the time series is stationary. Techniques have been developed to convert (normalize) non-stationary time series into stationary ones so that the classic theory can be applied. As an alternative, the ARIMA has been developed as a generalization of the classical ARMA models when there is evidence of non-stationarity. ARIMA models have been incorporated into the Box-Jenkins set of time series analysis models. ARIMA model is an extension of ARMA model in the sense that by including auto-regressive and moving average it has an extra part of differentiating the time series. If a data set exhibits long term variations such as trend, seasonality and cyclic components, differentiating of dataset in ARIMA allows the model to deal with such long term variations. Two common processes of ARIMA in identifying patterns in time series data and forecasting are: auto-regressive and moving average.

Auto-regressive process: Most time series consist of elements that are serially dependent at the time of occurrence in the sense that one can estimate a coefficient or a set of coefficients that describe consecutive elements of the series from specific, time-lagged (previous) elements. Each observation of the time series is made up of a random error component (random shock, ξ) and a linear combination of prior observations.

Moving average process: Independent from the auto-regressive process, each element in the series can also be affected by the past error (or random shock) which is not considered by the auto-regressive component. Each observation of the time series is made up of a random error component (random shock, ξ) and a linear combination of prior random shocks.

The ARIMA model includes three types of parameters which are: the auto-regressive parameters (p), the number of differencing passes (d), and moving average parameters (q). In the notation introduced by Box and Jenkins, models are summarized as ARIMA p.d.q. For example, a model described as ARIMA 1.1.1 means it contains 1 auto-regressive parameter and 1 moving average parameter for the time series after it is differentiated once to attain stationary. The general form of the ARIMA model describing the current value X_t of a time series is:

$$\varphi_1 (B)\ (1\text{-}B)\ X_t = \theta_1 (B)\ e_t \qquad \text{(Equation 1)}$$

Where φ_1= Auto-regressive parameter, and θ_1= Moving average parameter.

Four basic stages of ARIMA in identifying patterns in time series data and prediction are model identification, parameter estimation, diagnostic checking and forecasting.

Model identification stage

In this stage, number of auto-regressive (p) and moving average (q) parameters necessary to yield an effective model of the process are decided. The data are examined to check for the most appropriate class of ARIMA processes through selecting the order of the regular and non-seasonal differentiation required to make the series stationary, as well as through specifying the number of regular and auto-regressive and moving average parameters necessary to adequately represent the time series model. The major tools used in the identification phase are plots of the series, correlograms (plot of auto-correlation and partial auto-correlation verses lag) of auto-correlation function (ACF) and partial auto-correlation function (PACF). The ACF is the correlation between neighboring observations in a time series. When determining if an autocorrelation exists, the original time series is compared to the lagged series. This lagged series is simply the original series moved one time period forward. Suppose there are five time-based observations: 10, 20, 30, 40, and 50. When lag =1, the original series is moved forward one time period. When lag =2, the original series is moved forward two time periods. In the other words, the theoretical ACF measures amount of linear dependence between observations in a time series that are separated by k time lags. The PACF is used to compute and plot the partial auto-correlations between the original series and the lags. However, PACF eliminates all linear dependence in the time series beyond the specified lag. The PACF plot helps to determine how many auto-regressive terms are necessary to reveal trend either in the mean level or in the variance level of the series.

Parameter estimation stage

At parameter estimation stage, the parameters are estimated using by the Maximum Likelihood (ML), Conditional Least Square (CLS) and Unconditional Least Square (ULS) methods. Among these methods, maximum likelihood seems to be the best (Box and Jenkins, 1976). The parameters should be statistically significant at α=p% and satisfy two conditions, namely stationary and invertibility for auto-regressive and moving average models, respectively. In this stage, several models are tentatively chosen and then were compute the values of Akaike Information Criterion (AIC). The model structure which has the minimum AIC value, among others model structures, will be chosen as the best model. Equation 2 describes the formula to compute AIC. In this Equations Tp is the number of AR, I and MA parameters.

$$AIC = -2Ln(Max.Likelihood) + \frac{2T_p}{n - T_p - 1} \qquad \text{(Equation 2)}$$

Diagnostic checking stage

After fitting a provisional time series model, we can assess its adequacy in various ways. Most diagnostic tests deal with the residual assumptions in order to determine whether the residuals from fitted model are independent, have a constant variance, and are normally distributed. Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit of the tentative model to the time series data. The approaches that can be used to evaluate the adequacy of a model are the ACF residual and Port Manteau lack of fit test. The auto-correlations function (ACF) of the series can be used to examine whether the residual of the fitted model is white noise or not. Under the null hypothesis that residual follows a white noise process, roughly 95% of the auto-correlation coefficient should fall within the range $\pm$ 1.96/$\sqrt{T}$. In Port Manteau lack of fit test (Equation 3), the values of p-value exceed the 5%, it indicates that residuals have significant departure from white noise. If the selected model fails to pass Port Manteau lack of fit test, the modeler return to select alternative model and follow the same procedure until satisfactory model results are obtained.

$$Q = \left(N - d - DS\right)\sum_{k=1}^{M} r_k^2 a_t \qquad \text{(Equation 3)}$$

In Equation 3, r_k (a_t) is the auto-correlation coefficient of the residual (a_t) at lag k, and M is the maximum lag considered (about N/4), ARIMA model is considered adequate if p>chi square is greater than 0.05 where 0.05 is the probability level of significance.

Forecasting stage

In this stage, the estimated parameters are used to calculate new values of the time series and confidence intervals for those predicted values. The estimation process is performed on transformed (differentiated) data, hence before the forecasts are generated; the series needs to be integrated to cancel out the effect of differentiation so that the forecasts are expressed in values compatible with the input data. This automatic integration feature is represented by the letter “I” in the name of the methodology (ARIMA = Auto Regressive Integrated Moving Average). In the present study, to identify the best fitting model, the predicted values using the several different ARIMA models are compared to the observed data of the validation period (1966-2015). The data of 1966 to 2005 used for modelling and the data of 2006-2015 used for evaluation of the best model and model's ability to predict. In the present work, to evaluate the performance of the best ARIMA model at each station, coefficient of determination (R^2) was used to select the best model. R^2 gives impartial result as it takes mean values of both the observed and predicted data.

RESULTS AND DISCUSSION

Before initiating modelling, we need to conduct several tests on original data series such as normality, homogeneity and adequacy tests. In this study we tested all annual rainfall with normal distribution and we found that the all series were not normal, and therefore natural log transformation was required for the normalization. After normality test, we plotted annual rainfall data with time (Figures 2 and 3). There was a little decreasing trend of the series and the series were not stationary. In order to fit an ARIMA model, stationary data in both variance and mean were needed. Variance stationarity was attained by having log transformation and differentiation of the original data to attain stationary in the mean. For our data, we need to have first differential, d=1, of the original data in order to have stationary series. After that, we need to test the ACF and PACF for the differential series to check stationarity. As shown in Figures 4 and 5, the ACF and PACF for the differential rainfall data were almost stable which support the assumption that the series is stationary in both the mean and the variance after having 1st order non-seasonal differential. Therefore, an ARIMA (p.1.q) model could be identified for the differentiated rainfall data. After ARIMA model was identified above, the p and q parameters need to be identified for our model. According to the ACF and PACF plots of annual rainfall in Jelogir Majin station, the suggested ARIMA model is 8.1.1, because ACF plot is suggesting AR (8) term whereas PACF is suggesting MA (1) term. Also the suggested model for Pole Zal was ARIMA

9.1.1 model, because ACF plot is suggesting AR (9) term whereas PACF is suggesting MA (1) term too. For ARIMA model selection, different ARIMA models structures were estimated to find the best model structure.

The parameters estimation values for several models by ML, CLS and ULS method are shown in Tables 1 and 2, and suggested that all selected models except ARIMA 1.1.1 and 4.1.1 models in ML and ULS method were suitable for entrance to next stage because it satisfied two conditions: stationary and invertibility. In addition, the Port Manteau values lack fit test for annual rainfall for Jelogir Majin and Pole Zal stations, as shown in Tables 3 and 4. These tables showed that all three selected models were adequate for prediction of studied series data. The plots of the ACF and PACF for the residuals of selected models are shown in Figures 6 and 7. These figures showed that the suggested model can be considered as appropriate model. Because the results of both tests suggest that the residuals were white noise, therefore the models are suitable for prediction. Several models in three methods were recognized for forecast, therefore it is essential to use the AIC index to select the best model. The values of AIC for the different ARIMA models are shown in Table 5 (AIC=321.128 in CLS estimation method for annual rainfall in Jelogir Majin station and AIC=316.938 in ML estimation method for annual rainfall in Pole Zal station). The model that gives the minimum AIC is selected as best fit model. Obviously, the ARIMA 8.1.1 and 9.1.1 models have the smallest values of AIC for prediction of annual rainfall in Jelogir Majin and Pole Zal stations, respectively, thus these models are suitable for prediction in the future. In order to evaluate the performance of the models, ten years forecasts were generated for the testing period from 2006 to 2015 and presented in Table 6. The hydrograph between real and predicted (forecasted) rainfall data using ARIMA models are shown in Figure 8. It was evident that the values of actual and forecasted rainfall data were close to each other, thus it can be concluded that the chosen models are suitable for each station. In addition, the coefficient of determination (R^2) between actual and forecasted values was high (R^2=0.70 and 0.80 for annual rainfall in Jelogir Majin and Pole Zal stations, respectively). The results obtained confirmed that the ARIMA 8.1.1 and 9.1.1 models were adequate for the studied series. Furthermore, these models are better than other models because ARIMA model gives low error value and is in good fit with the observed data. The Jelogir Majin station is larger than Pole Zal station and it is affected by more area than the Karkheh river basin. Therefore, the selected model at Jelogir Majin station could better predict rainfall in the region.

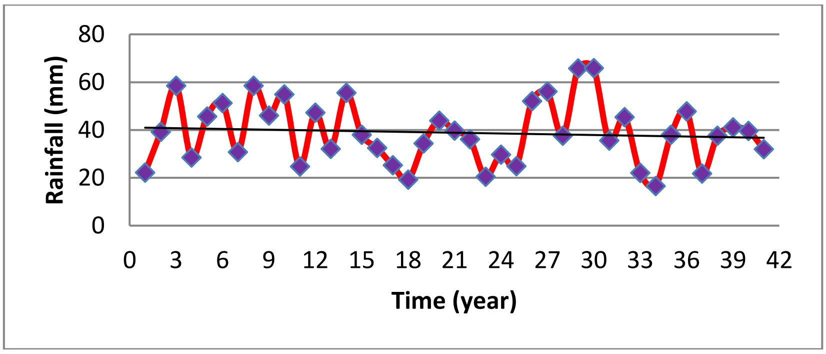

Figure 2. Annual rainfall data for Jelogir Majin station (1966-2005).

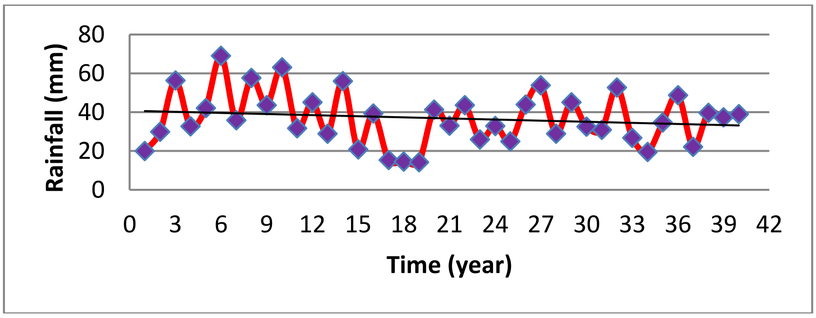

Figure 3. Annual rainfall data for Pole Zal station (1966-2005).

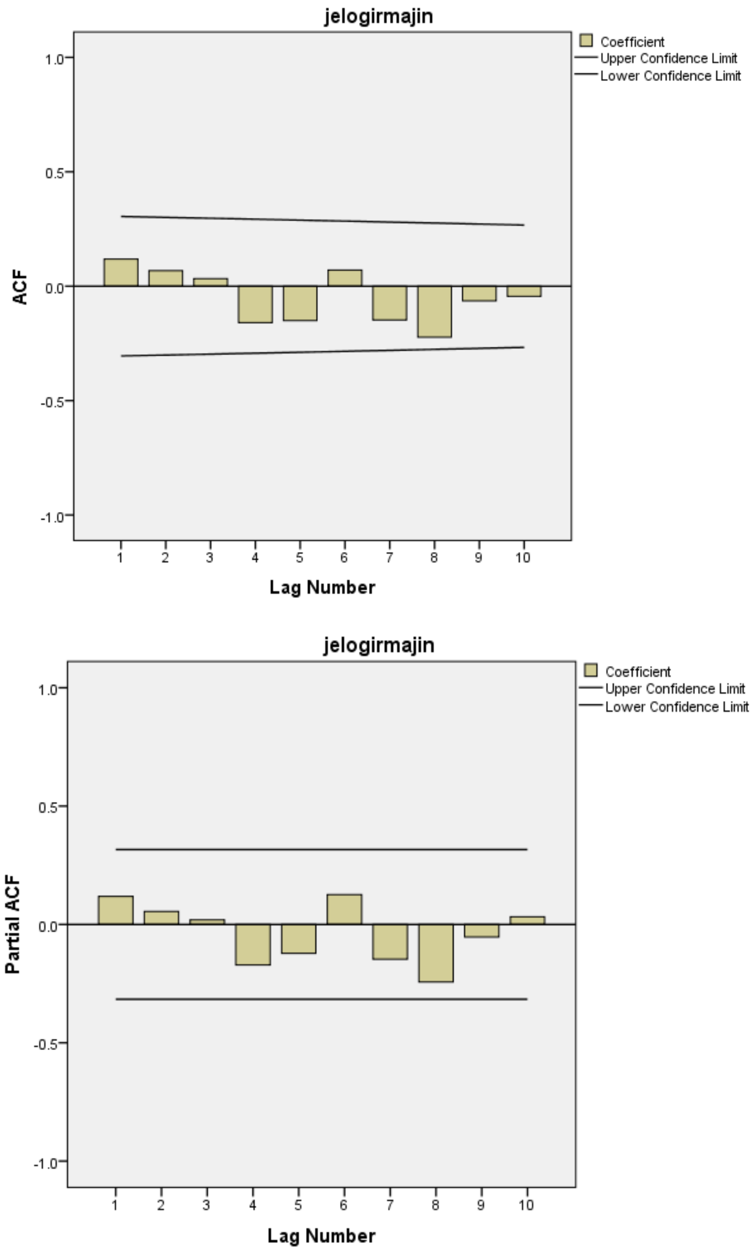

Figure 4. Plot of ACF and PACF for first order differential (d=1) of original rainfall data in Jelogir Majin station.

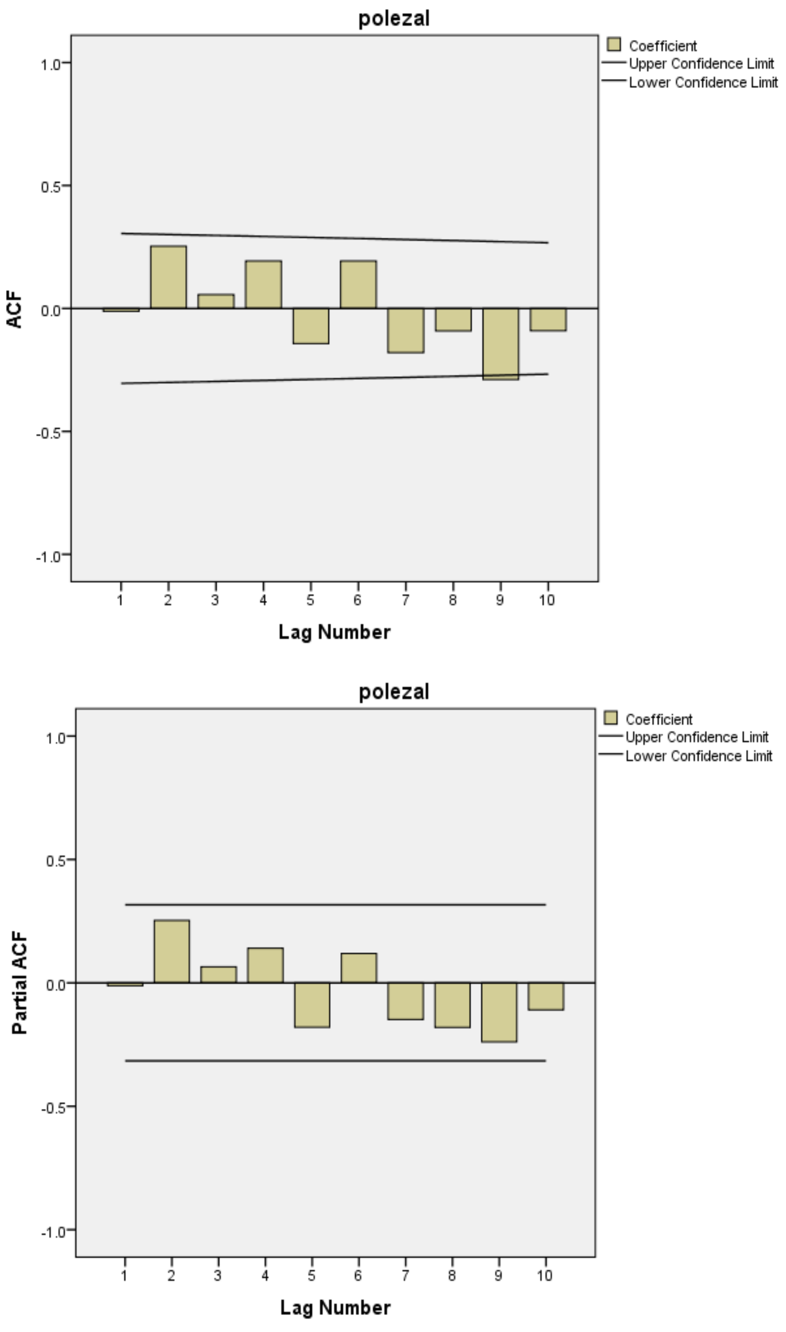

Figure 5. Plot of ACF and PACF for first order differential (d=1) of original rainfall data at Pole Zal station.

Table 1. Parameters estimation of the selected ARIMA model for annual rainfall at Jelogir Majin station.

Estimation method	Type (order) and values of ARIMA(p,1,q) parameters	Std. error coefficient	Absolute value of t	Probability of t	Stationary condition	Invertion condition
ML	p(1) = -0.45831	0.14281	-3.21	0.0013	Satisfactory	
	q(0)					
CLS	p(1) = -0.45857	0.14418	-3.18	0.0029	Satisfactory	
	q(0)					
ULS	p(1) = -0.47025	0.14318	-3.28	0.0022	Satisfactory	
	q(0)					
ML	p(1) = 0.21071	0.18855	1.12	0.2638	Satisfactory	
	q(1) = 0.99971	28.2487	0.04	0. 9718		Not satisfactory
CLS	p(1) = 0.19615	0.21063	0.93	0.3578	Satisfactory	
	q(1) = 0.83593	0.11764	7.11	0.0001<		Satisfactory
ULS	p(1) = 0.22697	0.16853	1.35	0.1863	Satisfactory	
	q(1) = 0.99998	0.33052	3.03	0.0045		Not satisfactory
ML	p(8) = -0.25552	0.16684	-1.53	0.1257	Satisfactory	
	q(1) = 0.99976	28.6089	0.03	0.9721		Not satisfactory
CLS	p(8) = -0.21750	0.18156	-1.2	0.2385	Satisfactory	
	q(1) = 0.70035	0.11755	5.96	0.0001<		Satisfactory
ULS	p(8) = -0.33925	0.17363	-1.95	0.0583	Satisfactory	
	q(1) = 0.99998	0.32887	3.04	0.0043		Not satisfactory

ML= Maximum likelihood CLS= Conditional least square ULS= Unconditional least square

Table 2. Parameters estimation of the selected ARIMA model for annual rainfall at Pole Zal station.

Estimation method	Type (order) and values of ARIMA(p,1,q) parameters	Std. error coefficient	Absolute value of t	Probability of t	Stationary condition	Invertion condition
ML	p(1) = -0.62039	0.12452	-4.98	0.0001<	Satisfactory	
	q(0)					
CLS	p(1) = -0.63144	0.12581	-5.02	0.0001<	Satisfactory	
	q(0)					
ULS	p(1) = -0.63594	0.12521	-5.08	0.0001<	Satisfactory	
	q(0)					
ML	p(1) = -0.31071	0.20247	-1.53	0.1249	Satisfactory	
	q(1) = 0.59798	0.17659	3.39	0.0007		Satisfactory
CLS	p(1) = -0.37590	0.21428	-1.75	0.0877	Satisfactory	
	q(1) = 0.45739	0.20593	2.22	0.0325<		Satisfactory
ULS	p(1) = -0.30706	0.20147	-1.52	0.1360	Satisfactory	
	q(1) = 0.63238	0.16534	3.82	0.0005		Satisfactory
ML	p(9) = -0.36711	0.15570	-2.36	0.0184	Satisfactory	
	q(1) = 0.89646	0.10410	8.61	0.0001<		Satisfactory
CLS	p(9) = -0.30225	0.17256	-1.75	0.0881	Satisfactory	
	q(1) = 0.71765	0.11474	6.25	0.0001<		Satisfactory
ULS	p(9) = -0.45312	0.15926	-2.85	0.0072	Satisfactory	
	q(1) = 0.99998	0.32942	3.04	0.0044		Not satisfactory

ML= Maximum likelihood CLS= Conditional least square ULS= Unconditional least square

Table 3. Result of autocorrelation check of residual annual rainfall at Jelogir Majin station.

ARIMA model	Estimation method	To lag	Df	Chi-square	Pr>chi square	Adequacy for modelling
ARIMA(1,1,0)	ML	6	5	7.57	0.1819	Satisfactory
		12	11	10.49	0.4871	
		18	17	12.02	0.7991	
		24	23	17.23	0.798	
	CLS	6	5	7.57	0.1814	Satisfactory
		12	11	10.44	0.4912	
		18	17	12.01	0.7998	
		24	23	17.27	0.7956	
	ULS	6	5	7.64	0.1771	Satisfactory
		12	11	10.58	0.4792	
		18	17	12.12	0.7926	
		24	23	17.26	0.7963	
ARIMA(1,1,1)	CLS	6	4	2.43	0.6568	Satisfactory
		12	10	4.41	0.9269	
		18	16	6.57	0.9807	
		24	22	12.71	0.9411	
ARIMA(8,1,1)	CLS	6	4	4.25	0.3729	Satisfactory
		12	10	4.9	0.898	
		18	16	7.03	0.9727	
		24	22	11.73	0.9626	

ML= Maximum likelihood CLS= Conditional least square ULS= Unconditional least square

Table 4. Result of autocorrelation check of residual annual rainfall at Pole Zal station.

ARIMA model	Estimation method	To lag	Df	Chi-square	P>chi square	Adequacy for modelling
ARIMA(1.1.0)	ML	6	5	6.76	0.2389	Satisfactory
		12	11	8.78	0.6424	
		18	17	10.40	0.8861	
		24	23	11.29	0.9801	
	CLS	6	5	6.75	0.2401	Satisfactory
		12	11	8.83	0.6375	
		18	17	10.54	0.8796	
		24	23	11.44	0.9783	
	ULS	6	5	6.82	0.2342	Satisfactory
		12	11	8.90	0.6312	
		18	17	10.54	0.8793	
		24	23	11.42	0.9785	
ARIMA(1.1.1)	ML	6	4	2.80	0.5924	Satisfactory
		12	10	10.42	0.4043	
		18	16	14.10	0.5916	
		24	22	15.42	0.8437	

	CLS	6	4	2.71	0.6080	
		12	10	7.87	0.6420	
		18	16	11.50	0.7778	
		24	22	12.64	0.9438	
	ULS	6	4	2.89	0.5759	
		12	10	11.63	0.3105	
		18	16	15.54	0.4855	
		24	22	16.99	0.7642	
ARIMA(9.1.1)	ML	6	4	6.78	0.1482	Satisfactory
		12	10	9.77	0.4613	
		18	16	12.47	0.7114	
		24	22	14.75	0.8726	
	CLS	6	4	7.07	0.1325	
		12	10	8.83	0.5487	
		18	16	10.55	0.8366	
		24	22	12.79	0.9388	

ML= Maximum likelihood CLS= Conditional least square ULS= Unconditional least square

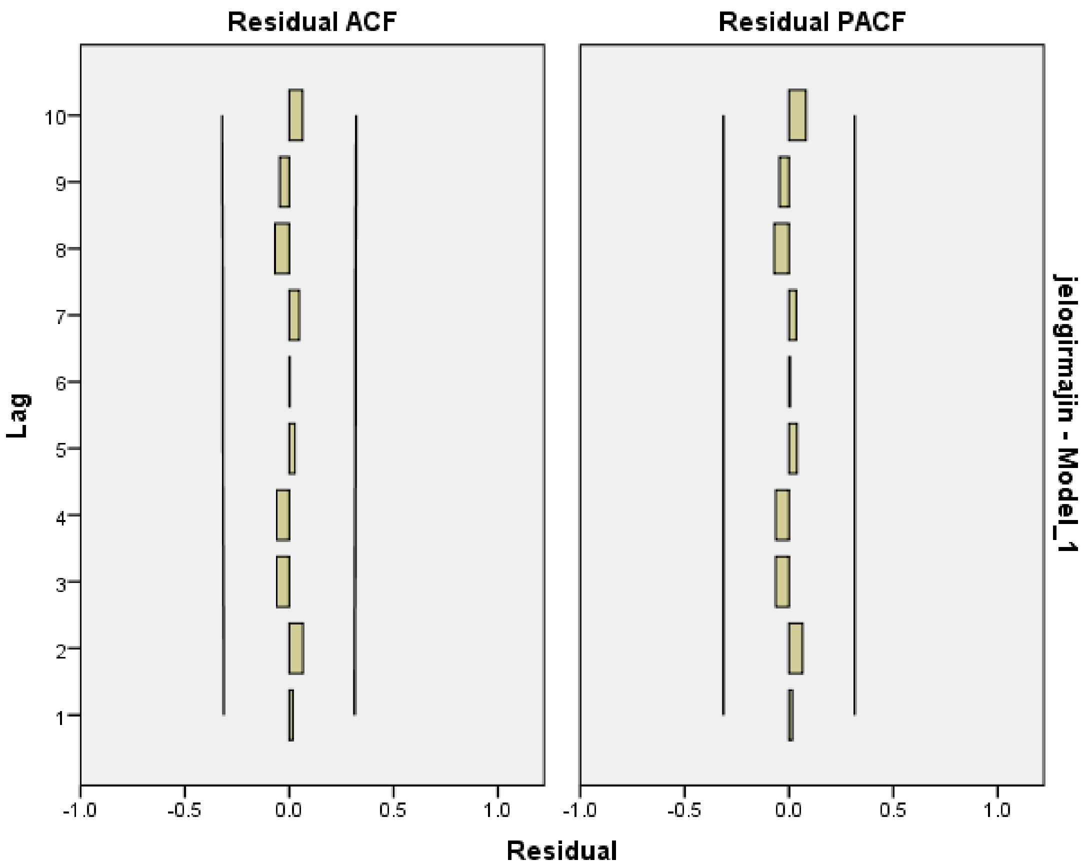

Figure 6. Autocorrelogram of residual series parameter for annual rainfall at Jelogir Majin station.

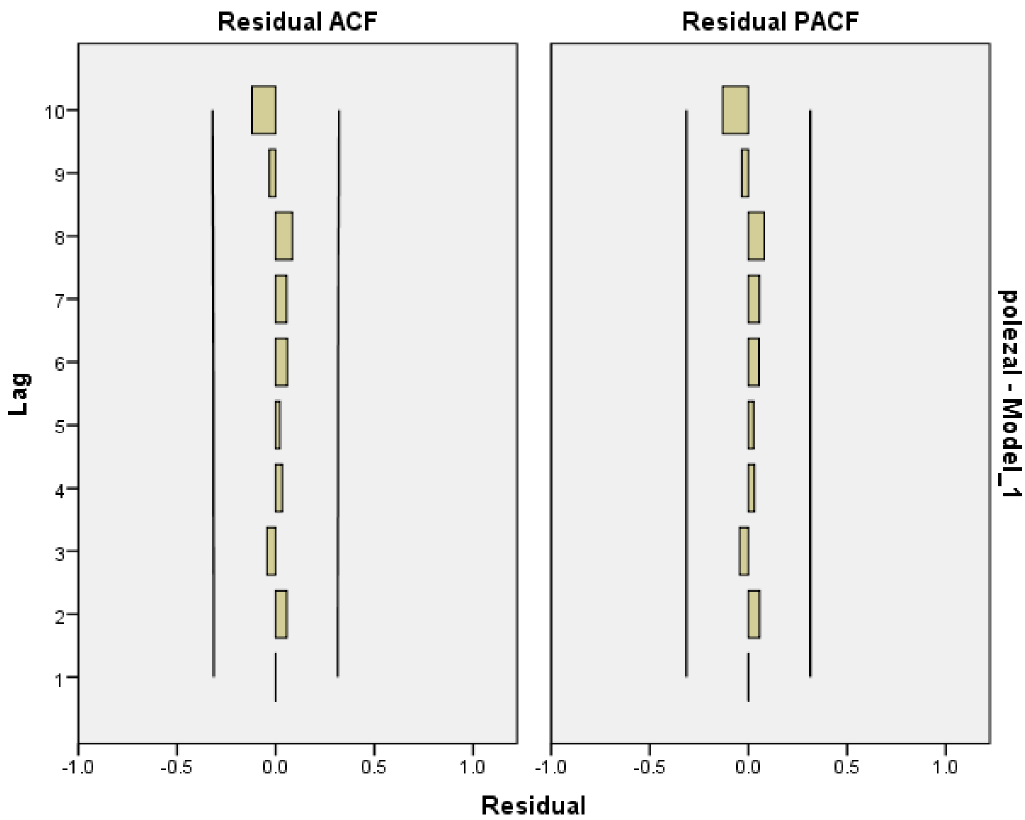

Figure 7. Autocorrelogram of residual series parameter for annual rainfall at Pole Zal station.

Table 5. Goodness of statistic fit for rainfall.

ARIMA model	Estimation method	Akaikc's statistic of rainfall	
		Jelogir Majin station	**Pole Zal station**
(1.1.0)	ML	325.576	322.965
	CLS	325.599	322.648
	ULS	325.583	322.981
(1.1.1)	ML	-	320.117
	CLS	322.981	321.385
	ULS	-	320.159
(8.1.1)	ML	-	-
	CLS	**321.128**	-
	ULS	-	-
(9.1.1)	ML	-	**316.938**
	CLS	-	320.969
	ULS	-	-

Table 6. Forecasted and actual rainfall for the period from 2006-7 to 2015-16.

Period	Rainfall at Jelogir Majin station		Rainfall at Pole Zal station	
	Forecasted (mm)	Actual (mm)	Forecasted (mm)	Actual (mm)
2006-7	43	40	41	40
2007-8	44	42	39	37
2008-9	40	38	42	39
2009-10	37	39	39	37
2010-11	43	41	38	39
2011-12	40	38	37	36
2012-13	39	37	35	34
2013-14	39	38	36	35
2014-15	38	37	34	35
2015-16	38	36	33	34

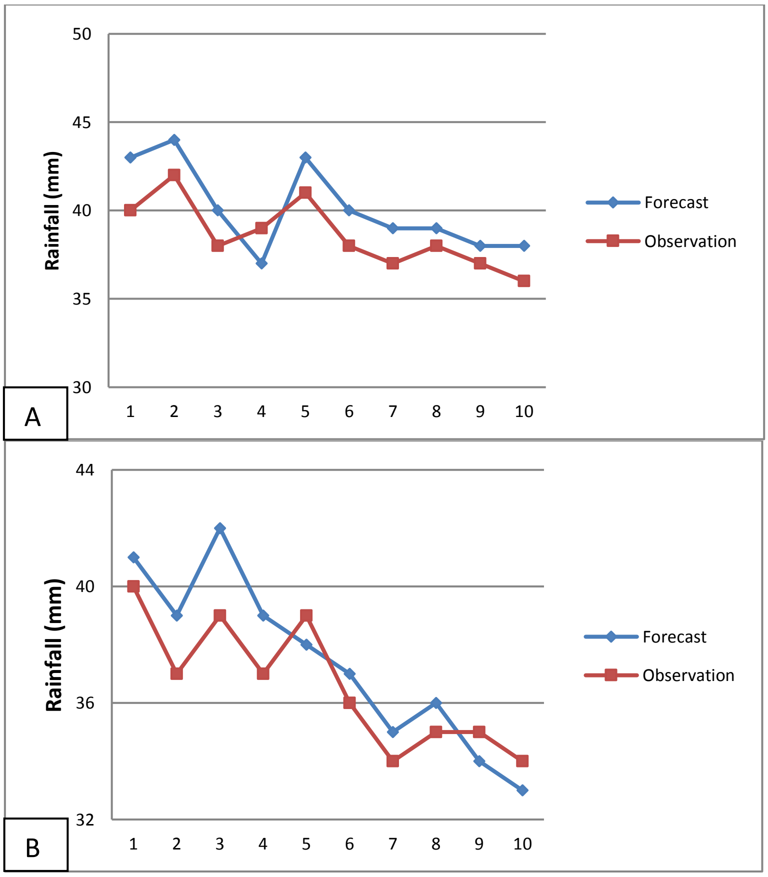

Figure 8. Comparison of forecasted and actual rainfall data for the period 2006-2015 in Jelogir Majin (A) and Pole Zal (B) stations.

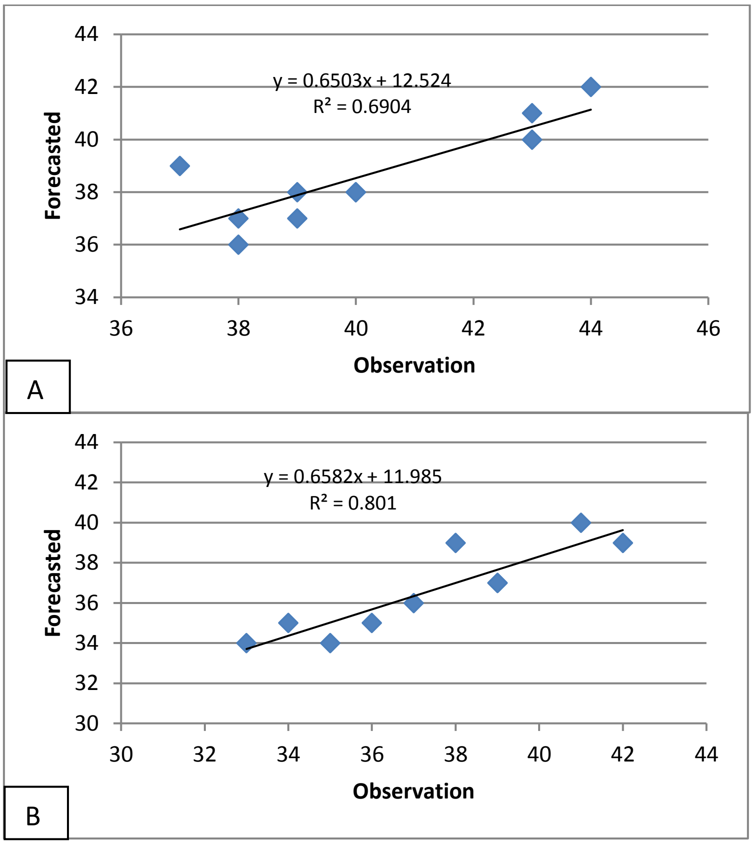

Figure 9. Correlation between actual and predicted rainfall values at Jelogir Majin (A) and Pole Zal (B) stations.

CONCLUSION

Rainfall prediction is crucial for making important decisions and performing strategic planning. The ability to predict rainfall quantitatively guides the management of water related problems such as extreme rainfall conditions such as floods and droughts, among other issues. Therefore, prediction of hydrological variables such as rainfall, floodstream and runoff flow as probabilistic events is a key issue in water resources planning. These hydrological variables are usually measured longitudinally across time, which makes time series analysis of their occurrences in discrete time appropriate for monitoring and simulating their hydrological behavior. Rainfall is among the sophisticated and challenging components of the hydrological cycle to modelling and prediction because of various dynamic and environmental factors and random variations both spatially and temporally.

In this research, we attempted to predict annual rainfall in Jelogir Majin and Pole Zal stations of Karkheh dam reservoir in Iran using ARIMA model. The ARIMA model was an appropriate tool to predict annual rainfall. The ARIMA model has a better performance than other ARMA stochastic model because it makes time series stationary, in both training and forecasting. The model 8.1.1 with AIC=321.128 in CLS estimation method for annual rainfall in Jelogir Majin station and the model 9.1.1 with AIC=316.938 in ML estimation method for annual rainfall in Pole Zal station were less than the other models. For model validation, the values of R^2 were 0.7 and 0.8 for annual rainfall in Jelogir Majin and Pole Zal stations, respectively, an indication of high ability of ARIMA model in annual rainfall prediction. According to the results obtained, R^2 between actual and forecasted (predicted) values in the studied station had downward trend and critical condition. Therefore, agriculture water management and cropping pattern must be done very carefully. Our selection of ARIMA 8.1.1 and 9.1.1 models for annual rainfall in Jelogir Majin and Pole Zal stations, respectively, gave us ten years prediction along with their 95% confidence interval that can help decision makers to establish strategies, priorities and proper use of water resources in Karkheh River in Iran. The ARIMA models are suitable for short term prediction because the ARMA family models can model short term persistence very well. These models are a finite memory model, and does not do well in long term prediction.

REFERENCES

Ansari, H. 2013. Forecasting Seasonal and Annual Rainfall Based on Nonlinear Modeling with Gamma Test in North of Iran. *International Journal of Engineering Practical Research,* 2(1): 16-29.

Akpanta, A.C. Okorie, I.E. and Okoye, N.N. 2015. SARIMA Modelling of the Frequency of monthly Rainfall in Umuahia, Abia State of Nigeria. *American Journal of Mathematics and Statistics,* 5(2): 82-87.

Box, G.E.P. and Jenkins, G.M. 1976. Time Series Analysis: Forecasting and Control. Revised Edn, Hoden Day, San Francisco.

Eni, D. and Adeyeye F.J. 2015. Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria. *Journal of Geoscience and Environment Protection,* 3: 91-98.

Gurudeo, A.T. and Mahbub, I. 2010. Time Series analysis of Rainfall and Temperature Interactions in Coastal Catchments. *Journal of Mathematics and Statistics,* 6(3): 372-380.

Mahsin, M. Akhter, Y. and Begum, M. 2012. Modeling Rainfall in Dhaka Division of Bangladesh using Time Series Analysis. *Journal of Mathematical Modelling and Application*, 1(5): 67-73.

Meher, J. Jha R. 2013. Time-series analysis of monthly rainfall data for theMahanadi River Basin, India. *Sciences in Cold and Arid Regions,* 5(1): 73-84.

Mirmousavi, S.H. Jalali, M. Garrousi, H. and Khaefi, N. 2014. Time Series Analysis of Rainfall in Khoi Meteorology Station. *Geographic Space,* 14(47): 1-17.

Sayemuzzaman, M. and Jha, K.M. 2014. Seasonal and annual precipitation time series trend analysis in North Carolina, United States. *Atmospheric Research,* 137: 183-194.

Soltani, S. Modarres, R. and Eslamianb, S.S. 2007. The use of time series modeling for the determination of rainfall climates of Iran. *International Journal of Climatology,* 27: 819–829.

Srikanthan, R. and McMahon T.A. 2001. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences,* 5(4): 653–670.

Wang, H.R. Wang, C. Lin, X. and Kang, J. 2014. An Improved ARIMA model for Precipitation Simulations. *Nonlinear Processes in Geophysics,* 21: 1159–1168.