World Academy of Science, Engineering and Technology
International Journal of Agricultural and Biosystems Engineering
Vol:12, No:3, 2018

Flood Predicting in Karkheh River Basin Using Stochastic ARIMA Model

Karim Hamidi Machekposhti, Hossein Sedghi, Abdolrasoul Telvari, Hossein Babazadeh

***Abstract*—**Floods have huge environmental and economic impact. Therefore, flood prediction is given a lot of attention due to its importance. This study analysed the annual maximum streamflow (discharge) (AMS or AMD) of Karkheh River in Karkheh River Basin for flood predicting using ARIMA model. For this purpose, we use the Box-Jenkins approach, which contains four-stage method model identification, parameter estimation, diagnostic checking and forecasting (predicting). The main tool used in ARIMA modelling was the SAS and SPSS software. Model identification was done by visual inspection on the ACF and PACF. SAS software computed the model parameters using the ML, CLS and ULS methods. The diagnostic checking tests, AIC criterion, RACF graph and RPACF graphs, were used for selected model verification. In this study, the best ARIMA models for Annual Maximum Discharge (AMD) time series was (4,1,1) with their AIC value of 88.87. The RACF and RPACF showed residuals' independence. To forecast AMD for 10 future years, this model showed the ability of the model to predict floods of the river under study in the Karkheh River Basin. Model accuracy was checked by comparing the predicted and observation series by using coefficient of determination (R^2).

***Keywords*—**Time series modelling, stochastic processes, ARIMA model, Karkheh River.

I. Introduction

FLOODS are a natural process influenced largely by the weather and driven by the amount of precipitation and length of time it. After heavy precipitation, rivers and catchments may overflow. This type of flooding is most common in Iran and is known as riverine flooding. Also, floods have a huge environmental and economic impact. Hydrologists use flood analysis to estimate and predict flood occurrence in the future in order to design and operation of hydraulic structures such as dams, and reservoirs, etc. Flood analysis at hydrological phenomena is a form of extreme value analysis such as AMS values. Hydrologists predict extreme hydrological events such as flooding using different methods and models, such as probability distributions, flood frequency analysis, and stochastic models, etc.

One of the best models for the prediction of floods is stochastic models such as ARIMA (time series models). These models are commonly used in water resource management and hydrology. The beauty of time series modelling is that future values of a variable can be estimated using its historical values (past values). Ahmad et al. applied time series modelling of the annual maximum flow of river Indus at Sukkur, they found that ARIMA (2,1,1) was appropriate for this series [10]. Srikanthan et al. used time series models to analyze annual flow of Australian streamflows. ACF and PACF were used to specify the suitable form of ARIMA models (Box-Jenkins time series models) [11], [4]. Nguyen et al. applied a statistical approach for prediction of annual maximum rainfall data at 15 rain gage stations in Quebec (Canada) for the 1961-1990 period [7].

Stojković et al. suggested that the annual streamflows (discharge) simulated by the stochastic ARIMA model were suitable for hydrological simulations in large European rivers [12]. Mirzavand and Ghazavi used time series modelling for forecasting of groundwater level in an arid environment and suggested the AR (2) model is suitable [5]. Naeem studied stochastic modelling of the daily rainfall for the period 1981–2010 in Pakistan [6]. Gargano et al. used a stochastic model for daily residential water [3]. Also, Nigam et al. forecasted river runoff based on the modelling of time series [8].

The research attempts to demonstrate the occurrences of the rainfall and river streamflow and predict this two phenomenon using stochastic ARIMA models (time series models) [11], [4]. The specific emphasis has been given for the accurate flood predicting and warning for an effective management of flood tragedy if needed [9].

The applicability of data based on stochastic analysis is studied for the 3rd largest river in Iran and perennial medium size river named Karkheh. The river spans in Karkheh basin in the west of the Iran, located in the central and southern regions of the Zagros mountain range. The Karkheh basin has a catchment of about 50,000 km^2 and river 900 kilometers long. Hydrologically, the basin is divided into five sub-basins, Gamasiab, Qarasou, Kashkan, Saymareh and south Karkheh. Fig. 1 shows the location of the hydrometric stations in the studied basin.

Data pertaining to the river streamflow (AMD) have been collected from hydrometric station situated Jelogir Majin. The Jelogir Majin Station is located upstream of the Karkheh dam reservoir and between longitudes 32^0 58′ N and latitudes 47^0 48′ E. The time plot of the AMD time series has been given in Fig. 3. The data ranges have been taken as annual cumulative and beginning from 1958 till 2005. We collected this series from the Iran Water Resources Management Company (IWRMC).

Karim Hamidi Machekposhti is Ph.D. Student in Water Sciences and Engineering Department, Science and Research Branch, Islamic Azad University, Tehran, Iran (phone: 00989197704935- fax: 00982144865179-82; e-mail: karim.hamidi@srbiau.ac.ir).

Hossein Sedghi and Hossein Babazadeh are with Water Sciences and Engineering Department, Science and Research Branch, Islamic Azad University, Tehran, Iran (e-mail: h.sedghi1320@gmail.com, h_babazadeh@srbiau.ac.ir).

Abdolrasoul Telvari is with the Department of Civil Engineering, Islamic Azad University, Ahvaz, Iran, (e-mail: telvari@gmail.com).

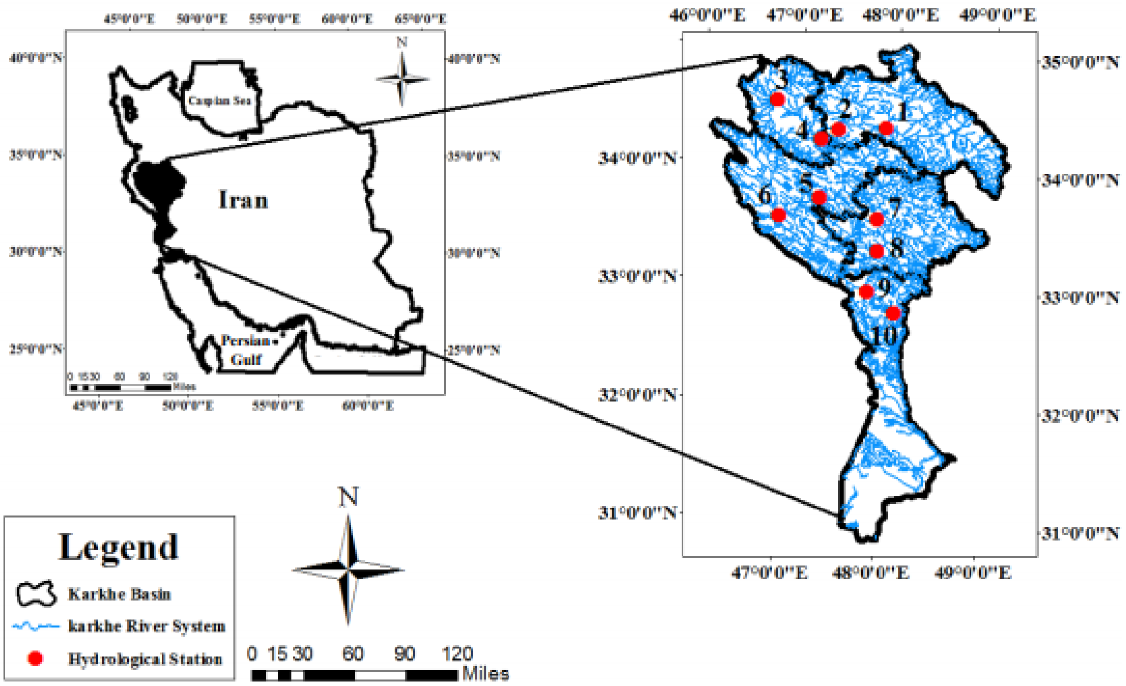

Fig. 1 The location of the hydrometric stations in Karkheh basin, Iran

II. Methods

Rainfall-runoff and the flood frequency analysis are two main approaches in a flood analysis. The rainfall-runoff method uses rainfall statistics and a catchment model to predict floods; whereas, flood frequency analysis produces a flood frequency curve and uses only peak flow data to make the prediction. Also, another method of predicting floods is by using stochastic modelling (ARMA/ARIMA model). A time series (stochastic model) has four main components which are the trend component, the periodic component, the catastrophic component and the random component. An ARMA/ARIMA model generates an artificial series for prediction. In this method, we use the values of a phenomenon at past times in order to modelling extreme hydrological events such as floods.

An ARIMA model has four basic steps: identification, parameters estimation, diagnostic checking, and forecasting. The form of non-seasonal ARIMA model is ARIMA (p,d,q) and can be written as:

$$\phi_P(B)W_t = \theta_q(B)a_t \qquad (1)$$

or

$$\left(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p\right)\nabla^d z_t = \left(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q\right)a_t \qquad (2)$$

where

$$W_t = \nabla^d Z_t = Z_t - Z_{t-d}$$

a_t denotes the residual series, B backward shift operator defined as $BZ_t=Z_{t-1}$, $B^2Z_t=Z_{t-2}$ and so on, d=degrees of differencing and the terms of ϕ and θ denotes coefficient value of an AR and MA process of order p and q, respectively [2].

The visual displays of the series, such as plotting of data against time, autocorrelation function (ACF) and partial correlation function (PACF), are the main tools used for identification of the model. Stationarity of a series is determined with ACF and PACF. If a series recognizes non-stationarity, we need differencing by trial and error. Differencing is commonly selected as one (d=1), if a series has non-stationarity.

After selecting the best order of differencing (d), we required to identify the order of the p and q parameters. The ACF and PACF of differenced series, help to indentify the order of p and q. Also, it was recommended to suggest a few different values of p and q to get the best model.

After selecting the order of p, q and d, we need to estimation parameters. For this purpose, three methods, unconditional least square (ULS), conditional least square (CLS) and maximum likelihood (ML), were used for the estimation of parameters of selected models. The maximum likelihood (ML), conditional least square (CLS) and unconditional least square (ULS) methods are used to estimate the model parameters. The parameters of a good model have two stationary and invertibility conditions. The models which have these conditions were suitable for entrance to the next stage (diagnostic checking).

Diagnostic checking is applied to see if the model is adequate or not. There are two tests for in this stage: Port Manteau Lack-of-Fit test and Residual Autocorrelation Function Test (RACF and RPACF). In the first test, if the p>chi squared value was greater than the level of significant (0.05), the ARIMA model is considered adequate. In the second test, the ARIMA model is considered adequate if the residuals have independence. In other words, Autocorrelation and Partial Autocorrelation Function (RACF and RPACF) were not significant, which means the value of residuals at any lag will not affect the value of residual at the next lag.

The best model has parsimony. Akaike Information Criteria (AIC) was used for the parsimony of parameters. The model with the minimum AIC was selected as the best model. SAS and SPSS software can find the best model based on the AIC values calculated for a range of p and q.

At the end, the best model for predicting of hydrologic phenomenon was the model that passed the diagnostic checking and has the minimum AIC. The final step was to generate a prediction of future values. Then, we compared the predicted and observation (original) data series. The basic methodology of ARIMA development is shown in Fig. 3.

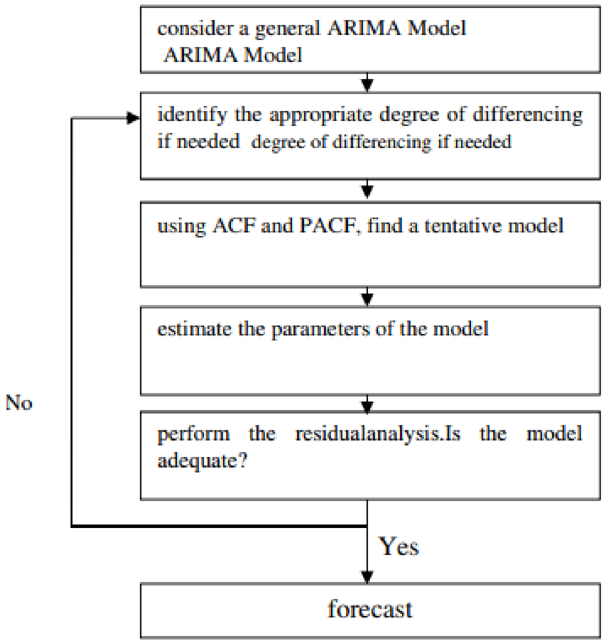

Fig. 2 ARIMA model development

III. Result and Discussion

The objective of this study was to identify a suitable ARIMA model based on the Box-Jenkins approach. Since annual river runoff (streamflow) is a non-seasonal phenomenon, we need to identify the order (p,d,q) for a non-seasonal univariate model. Then, the least square estimates of the parameters of time series models (ARIMA model) are used for predicting the river streamflow. The four-stage method contains model Identification, parameters estimation, diagnostic checking and forecasting (predicting) that are fitted to time series models (ARIMA model).

Before the start of modeling, we must plot the observations of natural data against time, as this will allow to show the important aspects of a time series such as seasonality, trend, and outliers, etc. [1].

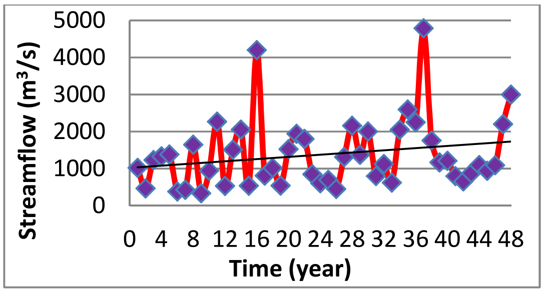

Fig. 3 Time series of natural AMD at Karkheh River

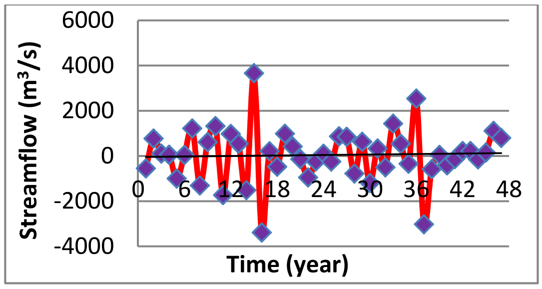

Fig. 4 Time series of AMD at Karkheh River (d=1)

Fig. 3 shows that there is little increase trend for AMD. The natural AMD series are not stationary, and for this reason, we first differentiated natural data to achieve a stationary series. This plot is shown in Fig. 4; it is stationary and there is no trend. The choice of the order of p, d and q for the identification model, in practice, the degree of differencing d is assumed one (d=1), while the autocorrelation function and partial autocorrelation function are plotted to guess the order of p and q. ACF and PACF of the AMD data for natural data (d=0) is shown in Figs. 5 and 6, and for d=1 in Figs. 7 and 8. The runoff of the Karkheh River shows a non-seasonal pattern, the same can be seen in the ACF and PACF plots and hence the flow pattern requires a non-seasonal model. An analysis of significant ACF and PACF plots implies the one and four order non-seasonal ARMA parameterization of AMD series.

The results of values for the parameters estimation methods (ML (maximum likelihood), CLS (conditional least square) and ULS methods (unconditional least square)) of suggested models (1,1,0), (1,1,1) and (4,1,1) for AMD are shown in Table I. This table shows that all three models are suitable for modelling because the parameters of all models have two stationary and invertibility conditions. We enter the diagnostic check stage in order to compare suggested models together and select the best model. The diagnostic checking test is applied to see if the model is adequate or not. The results of this test are indicated in Table II. This table shows that all three selected models are adequate for the predicting of AMD data.

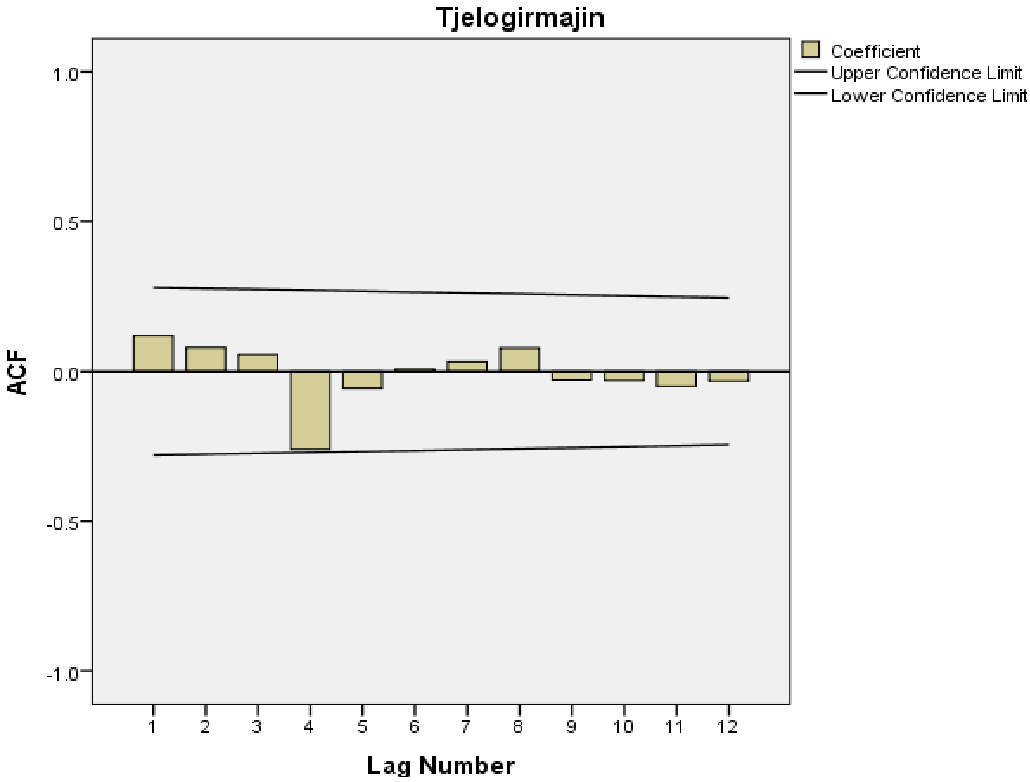

Fig. 5 Autocorrelation Function for natural AMD Series

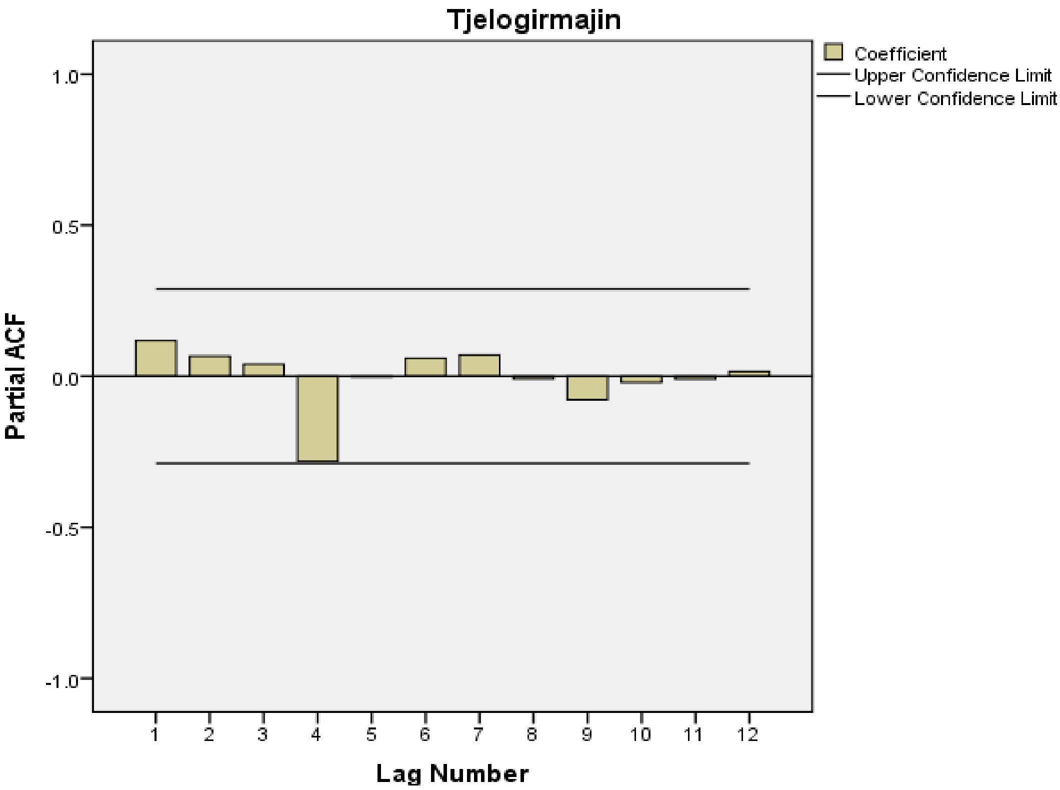

Fig. 6 Partial Autocorrelation Function for natural AMD Series

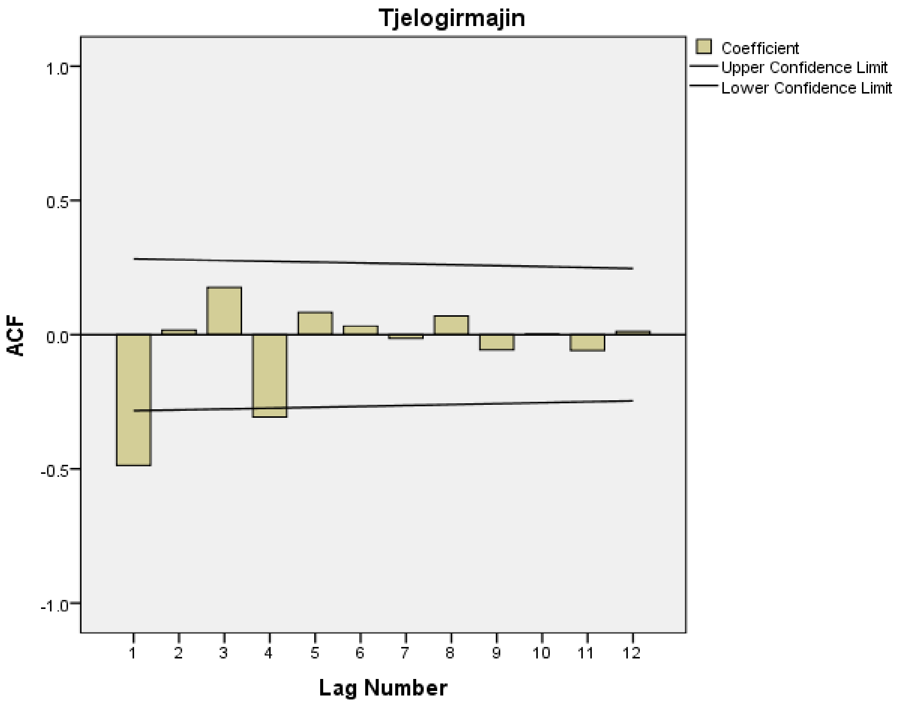

Fig. 7 Autocorrelation Function for AMD Series (d=1)

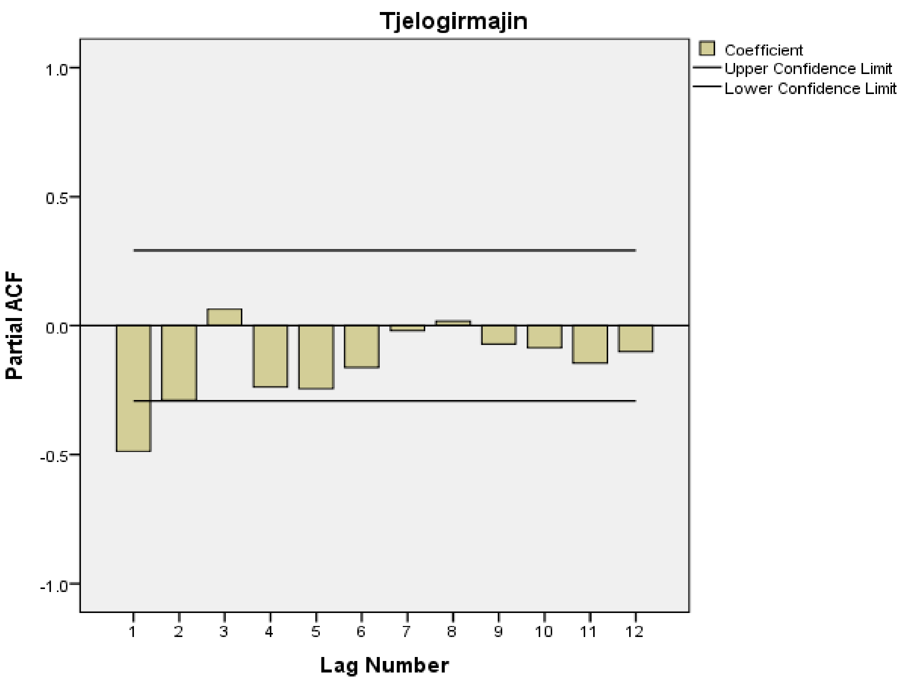

Fig. 8 Partial Autocorrelation Function for AMD Series (d=1)

Also, the independency of the resulting studied series, the correlogram of this series are computed for lag (M=N/5) is shown in Fig. 9. This figure shows that most of the computed lags lie inside the tolerance interval (±2/√N, at 95% confidence limits) and the residuals have independence. The goodness-of-fit statistic for the parsimony of series (Akaike Information Criteria (AIC)) is shown in Table III. Therefore, the ARIMA (4,1,1) model in the CLS estimation parameter method is the best model for AMD at Karkheh River. Table IV shows the predicted AMD for 10 years ahead of original data for the period from 2006 to 2015 by applying the best models. Fig. 10 shows the forecasted series for these data. The corresponding observed values are also shown in Fig. 11, and since agreement between the observed and predicted values is very good (R^2=0.84), it is confirmed that the ARIMA (4,1,1) model is adequate for the predicting of AMD.

TABLE I
VALUES OF NON-SEASONAL ARIMA MODEL PARAMETERS FOR AMD

Estimation Method	Type (Order) and Values of parameters ARIMA(p,1,q)	Absolute Value of t	Probability of t	Stationary Condition	Invertibility Condition
ML	P(1) = -0.48656 Q(0)	-3.81	0.0001	Satisfy	
CLS	P(1) = -0.48713 Q(0)	-3.78	0.0005	Satisfy	
ULS	P(1) = -0.49708 Q(0)	-3.88	0.0003	Satisfy	
ML	P(1) = 0.10744 Q(1) = 0.93539	0.5104 0.0001<	Satisfy	Satisfy	0.5104 0.0001<
CLS	P(1) = 0.11274 Q(1) = 0.96723	0.4926 0.0001<	Satisfy	Satisfy	0.4926 0.0001<
ULS	P(1) = 0.12820 Q(1) = 0.99998	0.4072 0.0001<	Satisfy	Not Satisfy	0.4072 0.0001<
ML	P(4) = -0.3317 Q(1) = 0.86679	0.0243 0.0001<	Satisfy	Satisfy	0.0243 0.0001<
CLS	P(4) = -0.33489 Q(1) = 0.86679	0.0339 0.0001<	Satisfy	Satisfy	0.0339 0.0001<
ULS	P(4) = -0.36065 Q(1) = 0.89524	0.0208 0.0001<	Satisfy	Satisfy	0.0208 0.0001<

ML: Maximum Likelihood CLS: Conditional Least Square ULS: Unconditional Least Square

TABLE II
RESULT OF AUTOCORRELATION CHECK OF RESIDUALS AMD

ARIMA Model	Estimation Method	To Lag	Df	Chi-Square	Pr>Chi Square	Adequacy for Modelling
		6	5	9.99	0.0754	
	ML	12	11	11.25	0.4228	Satisfy
		18	17	14.32	0.6447	
		24	23	17.26	0.7962	
		6	5	9.68	0.0850	
ARIMA(1,1,0)	CLS	12	11	10.94	0.4484	Satisfy
		18	17	14.02	0.6659	
		24	23	16.93	0.8128	
		6	5	10.02	0.0748	
	ULS	12	11	11.29	0.4190	Satisfy
		18	17	14.38	0.6400	
		24	23	17.33	0.7928	
		6	4	5.32	0.2562	
ARIMA(1,1,1)	ML	12	10	5.93	0.8212	Satisfy
		18	16	9.21	0.9046	
		24	22	12.45	0.9473	
		6	4	4.63	0.3269	
	CLS	12	10	5.20	0.8775	Satisfy
		18	16	8.26	0.9409	
		24	22	10.89	0.9763	
		6	4	1.30	0.8608	
	ML	12	10	2.77	0.9863	Satisfy
		18	16	7.46	0.9635	
		24	22	10.70	0.9787	
ARIMA(4,1,1)		6	4	1.36	0.8504	
	CLS	12	10	2.98	0.9818	Satisfy
		18	16	7.56	0.9610	
		24	22	10.41	0.9822	
		6	4	1.25	0.8702	
	ULS	12	10	2.53	0.9904	Satisfy
		18	16	7.61	0.9596	
		24	22	11.47	0.9674	

ML: Maximum Likelihood CLS: Conditional Least Square ULS: Unconditional Least Square

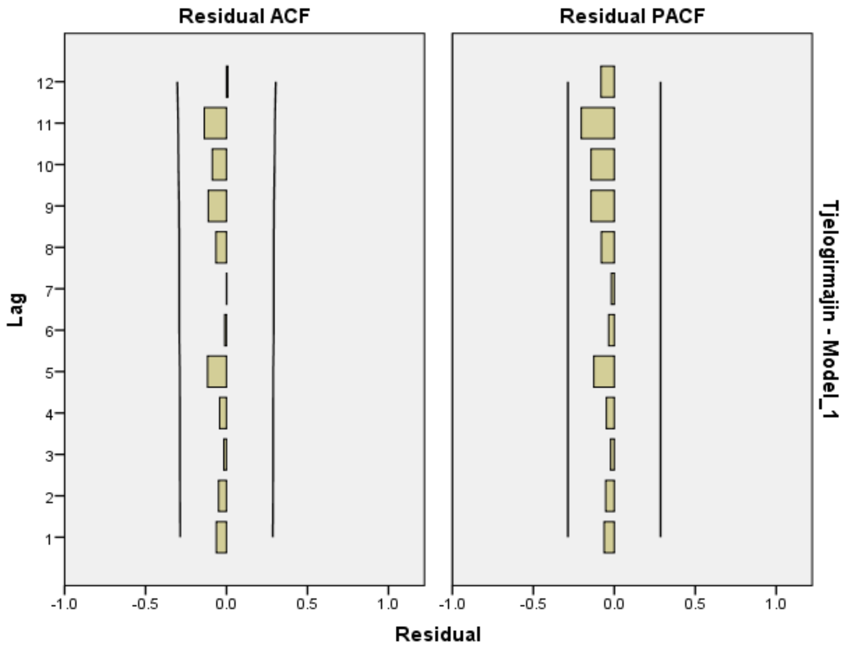

Fig. 9 Auto correlogram of Residual Series Parameter for AMD

TABLE III
GOODNESS OF FIT STATISTIC FOR AMD

ARIMA Model	Estimation Method	Akaikc's Statistic
	ML	103.4247
(1,1,0)	CLS	103.4469
	ULS	103.4316
(1,1,1)	ML	95.2824
	CLS	93.1350
	ML	90.8381
(4,1,1)	***CLS***	***88.8680***
	ULS	91.0387

TABLE IV
FORECASTS AMD FROM PERIOD 2006-7 TO 2015-16

Period	AMD	
	Forecast	Observation
2006-7	1451	1300
2007-8	1385	1323
2008-9	1092	1175
2009-10	984	1004
2010-11	1257	1253
2011-12	1276	1187
2012-13	1382	1346
2013-14	1431	1401
2014-15	1319	1330
2015-16	1312	1290

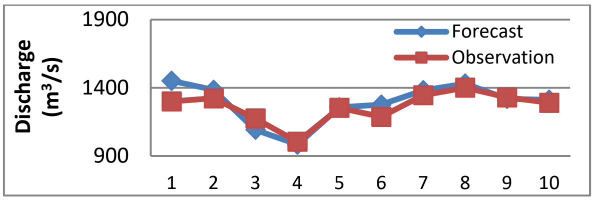

Fig. 10 Comparison of Forecasted and Observed data for AMD (2006-2015)

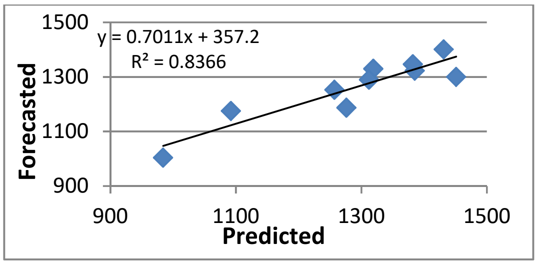

Fig. 11 Correlation between actual values and predicted values of AMD in Karkheh River

IV. CONCLUSION

Recognizing and predicting AMS (discharge) of the Karkheh River at Jelogir Majin Station during a statistical period is necessary for flood control and the planning of agricultural activities. Results from this reviewing indicated that:

- The best ARIMA models for AMD series with the least Akaike information criterion (AIC) (AIC=88.87) and which passed diagnostic checks was (4,1,1) in the CLS estimation parameter method. Its residual was independent, homoscedastic and approximately normally distributed. By comparing the models' synthetic series with the original series, their accuracies were checked. The predicted AMD data showed very good agreement with the actual recorded data (R^2=0.84). This gave increasing confidence of the selected ARIMA models.
- The study reveals that the Box-Jenkins (ARIMA) model methodology could be used as an appropriate tool to predict the flood in this river for the up-coming 10 years (2006-2015). Also, this methodology can help farmers in the area, in order to best plan agricultural activities to enlarge the areas of land to be cultivated using

Open Science Index, Agricultural and Biosystems Engineering Vol:12, No:3, 2018 publications.waset.org/10008713.pdf

supplemental irrigation.

- The significant ACF and PACF functions with high orders can be caused by factors such as area, good vegetation and snowmelt. The good vegetation of the region and the forest causes water retention in the soil surface layer and delay in the rise in surface runoff.
- The ARIMA model is suitable for short-term predicting of a series, because the ARMA family of models can model short-term durability very well. The AR model is a finite memory model, thus it does not fare well in long-term predicting.
- Model identification is the critical step in ARIMA modelling. The values of p, q and d had to be determined visually and they depended on the modeler's experience and judgment.

References

[1] Abd Saleh, Z. "Forecasting by Box-Jenkins (ARIMA) Models to Inflow of Haditha Dam," *J. of Babylon Uni., Eng. Sci.*, 2013, 5(21), 1675-1685.

[2] Box, G.E.P.; Jenkins, G.M. "*Time Series Analysis: Forecasting and Control,*" Holden Day, San Francisco, California, 1976.

[3] Gargano, R.; Tricarico, C.; Giudice, G. D.; Granata, F. "a stochastic model for daily residential water demand," *Water science and technology: water supply,* Available online ws 2016102, 2016.

[4] Hamidi machekposhti K., Sedghi H., Abdolrasoul Telvari A., Babazadeh, H. "Forecasting by Stochastic Models to Inflow of Karkheh Dam at Iran," *Civil Engineering Journal,* 2017, Vol. 3, No. 5, 340-350.

[5] Mirzavand, M.; Ghazavi, R. "A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods," *Water Resources Management,* 2015, 29(4), 1315-1328.

[6] Naeem, S. Stochastic Modelling of the Daily ainfall Frequency and Amount. *Arabian J. for Science and Eng.,* 2014, 39(7), 5691–5702.

[7] Nguyen, V. T. V.; Nguyen, T. D.; Cung, A. "A statistical approach to downscaling of sub-daily extreme rainfall processes for climate-related impact studies in urban areas," *Water science and technology: water supply,* 2007, 7(2), 183-192.

[8] Nigam, R.; Bux, S.; Nigam, S.; Pardasani, K. R.; Mittal, S. K.; Haque, R. "Time series modeling and forecast of river flow," *Current World Environment,* 2009, 4(1), 79-87.

[9] Nigam, R.; Nigam, S.; Mittal, S. K. "The river runoff forecast based on the modeling of time series," *Russian Meteorology and Hydrology,* 2014, 39(11), 750-761.

[10] Shakeel, A. M.; Idrees, A. M.; Naeem, H. M.; Sarwar, B. M. "Time Series Modelling of Annual Maximum Flow of River Indus at Sukkur," *Pakistan. Journal of Agricultural Sciences,* 1993, 30(1), 36-38.

[11] Srikanthan, R.; McMohan, T. A., Irish, J. L. "Time series analysis of annual flows of Australian streams," *J. Hydrology,* 1983.

[12] Stojković, M.; Prohaska, S.; Plavšić, J. "Stochastic structure of annual discharges of large European rivers," *J. Hydrology and Hydromechanics,* 2015, 63(1), 63–70.