Biological Forum – An International Journal **7**(1): 1205-1210(2015)

ISSN No. (Print): 0975-1130
ISSN No. (Online): 2249-3239

Forecasting Monthly and Annual Flow Rate of Jarrahi River using Stochastic Model

Ebrahim Nohani
Department of hydraulic Structures,
Dezful Branch, Islamic Azad University, Dezful, IRAN.

(Corresponding author: Ebrahim Nohani)
(Received 22 March, 2015, Accepted 26 April, 2015)
(Published by Research Trend, Website: www.researchtrend.net nohani_e@yahoo.com*)*

ABSTRACT: Limited available water resources, increased demand and water sources pollution in recent decades have increased the need for water resources management. In this regard, prediction and modeling of hydrologic systems are considered as important management tools to predict the future values of these systems. Stochastic methods can be mentioned as such models. In this study, using stochastic modeling, we predicted the annual and monthly flow rate values of Jarrahi River in Khuzestan province. The models presented in this study included simultaneous autoregressive moving average for modeling annual data and multivariate autoregressive moving average for monthly data. The SAMS 2007 software was used to implement all of the above models.

Keywords: Water resources, Modeling, Stochastic, Jarrahi River, Hydrologic systems

INTRODUCTION

Over the last few decades, predicting future values of hydrologic systems have come to the focus of the researchers' consideration for planning and management of water resources. For this purpose, various methods including stochastic models can be used as a management tool to predict future values of these systems. Stochastic methods are based on time series, and in some areas, we are faced with a lack of historical data. The methods proposed by Box and Jenkins (1970) have been used more widely, which are based on the combination of autoregressive moving average (ARMA) methods. Meanwhile, similar efforts have been made on using monthly autoregressive models and ARMA approaches for modeling of hydrological processes (Rao *et al.,* 1982, Ubertini, 1978). Kelps (1990) used the conceptual multivariable autoregressive moving average model for monthly runoff at 9 stations in southern Italy (Finzi *et al.,* 1975). This multi-station model is a continuous-time ARMA model (CARMA), which considers the spatial and periodical correlation of runoff separately. However, Salas *et al.* (1985) studied the runoff in annual scale and introduced two stochastic models for the re-production process. These researchers hypothesized that the runoff components are attributed to more than one year, and with this assumption, proved the validity of stochastic model (ARMA, 1, 1). Also, Abrishamchi *et al.* (2006) developed the annual regional stochastic model (AR, 1) for annual flow of basin of Karkheh, Karun and Dez Rivers. To this end, they used daily rivers flow data from hydrometric stations since the establishment of stations to the water year of 1997-1998, daily temperature and precipitation data from 53 stations and producing annual data. Claps (1990) officially concluded the ARMA model (1, 1) from a conceptual watershed model with non-correlated Gaussian inputs. It should be noted that the researchers considered the overall rainfall as the input of the conceptual system. Therefore, the relationships between stochastic and conceptual parameters are different as mentioned by Claps and Rossi (1991) for the univariate case; although they determine the range of the similar parameters. Given the importance of Jarrahi River and its role in the region water resources, in this research, we tried to estimate the flow rate using data from Gorgor hydrometric station located on Jarrahi River. The following models were used in this regard: Autoregressive, simultaneous autoregressive moving average for modeling annual, univariate periodic autoregressive moving average and multivariate autoregressive moving average for monthly data.

MATERIALS AND METHODS

Jarrahi River is formed of two main branches of Maroon and Ramhormuz rivers and is one of the rivers the in southwestern of Iran, which flows in the provinces of Khuzestan and Kohgiloyeh and Boyerahmad. With the source elevation of 2200 meters and 438 km in length, it is known as the eleventh long river in Iran.

The information on this project was collected from hydrometric Gorgor station. A mathematical model represents a stochastic process, which is also called stochastic model or the time series model. The model has a certain mathematical structure and a set of parameters. For example, if X has a normal distribution with the mean of and variance of , the time series model can be written as follows:

$$y_t = \mu + \sigma \varepsilon_t \quad (1)$$

Where ε_1 also has a normal distribution with a mean of zero and variance value of one.

In general, time series modeling include the following steps: Selecting the type of model, determining the structure of the model, estimating the model parameters and evaluating the usefulness of an organizable model. Basically, river flow modeling follows two methods: (1) deterministic method or physical simulation of the hydraulic system, (2) statistical method or stochastic simulation of the system. In the first method, the hydrological system is described and expressed by physical and theoretical equations, and there is always a unique correspondence between input (e.g. precipitation) and output (flow). On the other hand, stochastic methods are models with the aim to describe most statistical properties of the time series. The AR model is widely used in these methods as well.

A. Introduction of SAMS-2007 software

The SAMS software involves stochastic analysis, modeling and simulation of hydrological time series. The software has three primary uses, including data statistical analysis, fitting of stochastic model and generating fabricated series. All models in the SAMS 2007 (except for GAR 1 model) are based on the assumption that the data used are normally distributed. The GAR (1) makes models by the assumption of data gamma distribution. Therefore, normalizing of data is necessary for all other models. There are two normality tests in the software, called as normality skewness test and probability plot correlation coefficient test Filin (1975), both of which are used for the 10% confidence level. These tests can be sued for seasonal and annual data. There are plenty of univariate models in the SAMS, including annual models such as autoregressive - moving average (ARMA) model (p, q), the GAR (1) model for modeling of processes with gamma distribution, SM model for modeling processes with pattern transfer on average, and PARMA (p, q) to model periodic processes, such as quarterly or monthly data. Analysis and modeling of multivariate time series often appear to be needed in the hydrology (Rossi and Silvagni (1980). There are complete multivariate models for modeling complex structures and dependent on time and place for multiple delays. The multivariate models in this software include multiple autoregressive MAR (P) model, simultaneous autoregressive moving average as (p, q) model (CARMA), SM mixed model, CARMA (p, q)) as CSM-CARMA (p, q) and the seasonal- periodic multivariate autoregressive model MPAR (p). The fitted model in the software needs to be tested in terms of correspondence with the considered assumptions and ability to reproduce the statistical properties of available historical data. The information criteria used in the application to select the best model include modified Akaike information criterion (AICC) and Schwarz information criterion (SIC).

RESULTS

A. Annual data

In this section, using the 41- year statistics of Jarrahi River flow rate (since the water year of 1967-1968 to 2007-2008), the modeling (annual and monthly) were described and the results were analyzed. Fig. 1 shows the variation of flow per time. Analyzing the 41-year statistics of Jarrahi River shows that the average flow of the river is 63 cubic meters per second. Since lack of data pattern and normal distribution are as essential conditions in the stochastic modeling, thus, these parameters were evaluated in the first stage. Data analysis revealed the lack of data pattern. However, the normality test results on data showed that the raw data are not normally distributed. Thus, logarithmic, exponential and other transformation functions were used to normalize the data. The data normalization results using the mentioned conversion functions show that the series produced using an exponential function provides better results than the logarithmic function. In the second phase, autocorrelation function and partial auto-correlation function were drawn in order to detect determined the model order that the results show the minimum order of the model as one. However, for more reliability, models with higher orders were also investigated and compared with the Akaike information and Schwarz criteria. Finally, a model with the lowest values of the above criteria was introduced to the software to produce the new series. In the third stage, to detect the model usefulness, the new series were compared with the original series regarding data values as well as predicting the maximum drought period, drought flow rate, the maximum of high water period, the maximum flow rate of the high water period and the storage capacity.

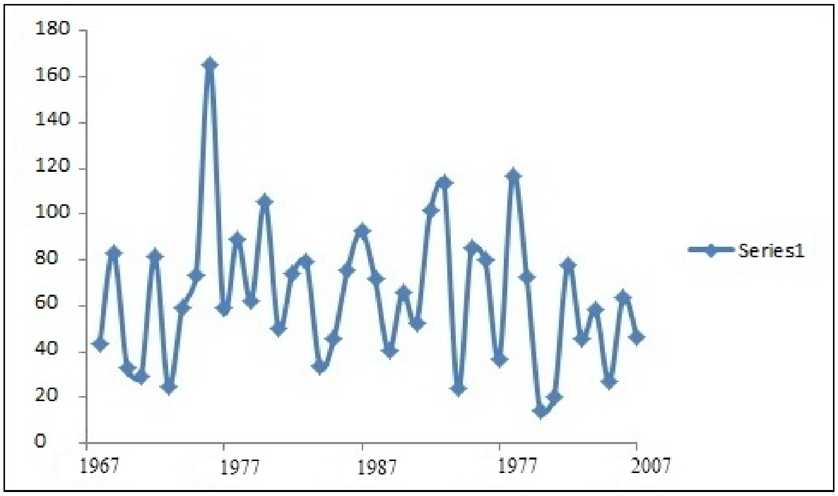

Fig. 1. Flow rate changes per the time of annual data.

To perform the research, the annual data were first divided into two education (80%) and validation (20%) categories. Then, using the education data, different models such as autoregressive model with p(AR (p) order, the autoregressive moving average model of order P and Q (ARMA (p, q)) and the simultaneous autoregressive moving average model (CARMA (p, q)) were fitted on the data. In better words, the considered models include AR (1), AR (2), AR (3), ARMA (1,1), ARMA (1,1), ARMA (1,1) as well as CARMA (1,1), CARMA (1,2) and CARMA (1,3). In Table 1, the informational criteria calculated for different models are given.

According to the table above, it can be concluded that among the autoregressive models, the AR (1) model is the proper one, while among the autoregressive moving average models, the ARMA (1,1) is the model of choice; and the appropriate order for the simultaneous autoregressive moving average model is CARMA (1,1).

Table 1: Informational criteria for AR, ARMA and CARMA models for training data.

Model	Akaike Informational criteria	Schwarz criteria Informational
AR(1)	35.523	34.574
AR(2)	37.822	37.896
AR(3)	40.091	41.007
ARMA(1,1)	37.905	37.979
ARMA(1, 2)	40.611	41.527
ARMA(1, 3)	43.427	44.983
CARMA(1,1)	-36.541	-36.467
CARMA(1, 2)	-33.838	-32.922
CARMA(1, 3)	-31.004	-29.449

Table 2: Comparison of statistical parameters for AR (1), ARMA (1, 1), CARMA (1, 1). Models.

Model	Training			Validation		
	R2	RMSE	MAE	R2	RMSE	MAE
AR(1)	0.04	38.869	27.059	0.0	27.514	20.495
ARMA(1,1)	-0.09	43.908	36.298	0.15	17.603	13.466
CARMA(1,1)	-0.09	43.908	36.298	0.15	14.603	13.466

The results of tables (1) and (2) show that according to Akaike and Schwarz criteria and due to statistical parameters of autoregressive models, autoregressive moving average and the simultaneous autoregressive moving average model, the models of AR (1), ARMA (1,1) and CARMA (1,1) are respectively the best models. In total, the CARMA (1,1) was selected as the best model and used to generate a time series (Table 3 and Fig. 2).

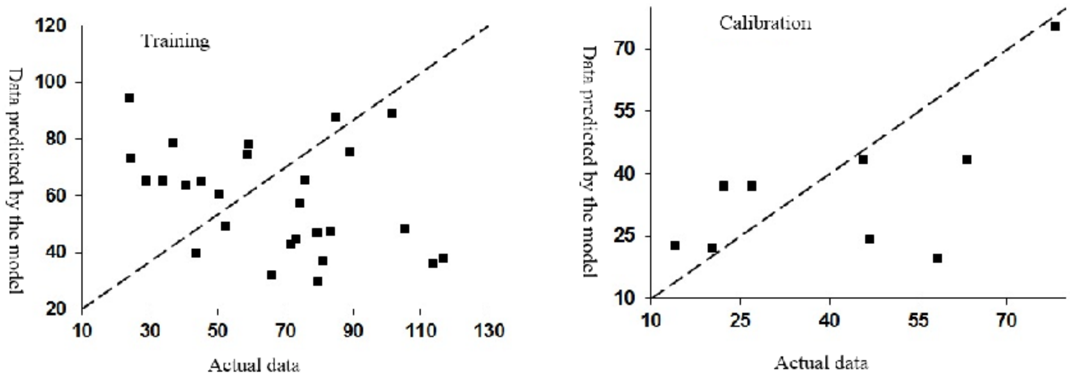

Fig. 2. Distribution of calculation and observation data in the training and validation stages.

Table 3: Comparison of data of storage, drought and high water years of the generated series with the primary series for ARMA (1,1) model

Parameter	Initial series	Generated series
Maximum drought period (years)	6	8
Drought flow rate (m^3/s)	116.9	159.6
Maximum high water period (years)	3	2
Maximum high water flow rate (m^3/s)	112.2	66.09
Storage capacity (MCM)	194.3	325.4

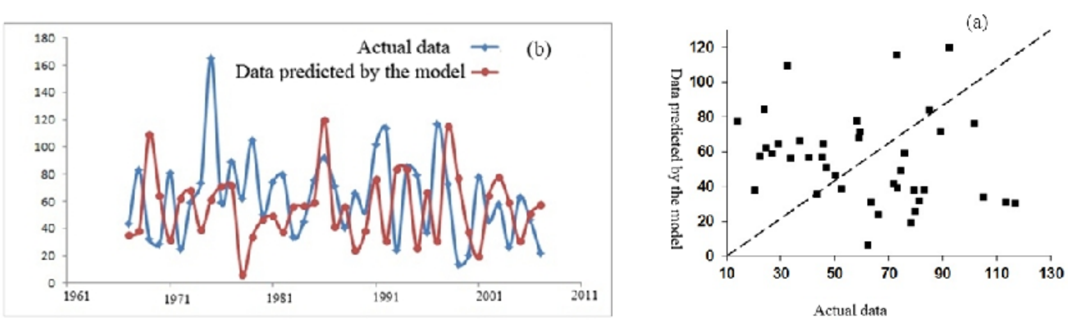

Fig. 3. (a) Distribution of computational data; (b) Comparison of the generated series and the original series for CARMA (1,1) model.

B. Monthly data

To fit the monthly models, the 41-year monthly statistics of the Jarrahi River flow rate from 1968 to 2008 were used (Fig. 4). Evaluation of pattern and distribution of monthly data normality shows that the data have no pattern; however, the normality test results indicate that the raw data are not normally distributed. Therefore, the logarithmic, exponential and other transformation functions were used to normalize the data; in this regard, the generated series using the logarithmic transformation function provided better results.

Also, autocorrelation and partial autocorrelation functions were plotted to distinguish the model order, and accordingly, the minimum model order was selected as 1; however, the models with higher-order were also implemented for ensuring and compared with Akaike and Schwarz information criteria. For annual data, different models such as PARMA and MPAR were used. In other words, the models considered included PARMA (1, 1), PARMA (2, 1), PARMA (3, 1) as well as MPAR (1), MPAR (2) and MPAR (3). The obtained results are given in Tables (4), and finally, the model with minimum information criteria values was introduced to the software for generating the new series.

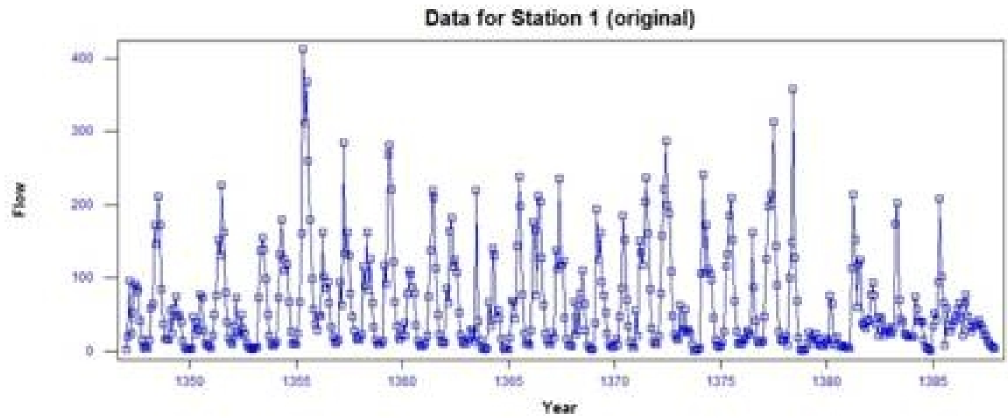

Fig. 4. Initial time series of monthly data.

Table 4: Information criteria of MPAR and PARMA models for training data.

Model	PARMA(1,1)	PARMA(2,1)	PARMA(3,1)	MPAR(1)	MPAR(2)	MPAR(3)
Akaike Informational criteria	-5.01	-3.99	-2.63	-4.19	-4.68	-3.71
Schwarz criteria Informational	-4.94	-3.08	-1.08	-5.15	-4.61	-2.55

In the third stage, to identify the model usefulness, the new series were compared with the original series regarding data values, predicting the maximum period of drought, drought flow rate, the maximum of high water period, the maximum of high water flow rate and the storage capacity. The results in Table 4 show that among the univariate and multivariate periodic autoregressive moving average models, the PARMA (1,1) and MPAR (2) models with having the minimum values of Akaike and Schwarz information criteria were select as the appropriate models. In total, the MPAR (2) model was selected as the best model for monthly time series. Therefore, this model was used to produce the new series. Also, the following factors were compared between the initial and generated time series by MPAR (2) model in Table (5) as a sample for the second half of the year: Discharge values, prediction of the maximum period, the drought flow rate, the maximum of high water period, the maximum of high water flow rate and the storage capacity.

CONCLUSION

Annual and monthly data on the flow rate of Jarrahi River, located in the province of Khuzestan, from the 41-year statistics of Gorgor hydrometric station were used for stochastic modeling. The provided models included autoregressive models, autoregressive moving average models and the simultaneous autoregressive moving average for modeling annual data.

Table 5: Comparison of data on storage, drought and high water years in the produced series with the primary series for the MPAR (2) model.

Parameters / Month	Period	The maximum period of drought	Discharge drought (m^3/s)	Maximum period of water	Maximum Discharge the water-filled (m^3/s)	Storage capacity
October	Primary	10	70.88	6	116.6	116.6
	Production	10	71.06	2	95.61	95.56
November	Primary	4	87.56	3	102.1	112.2
	Production	9	167.9	3	531.18	390.2
December	Primary	5	173.5	3	174.8	226.1
	Production	4	170.2	3	108.2	323.6
January	Primary	7	343.5	4	346.7	343.5
	Production	6	259.0	3	148.6	348.0
February	Primary	4	213.8	3	375.9	447.3
	Production	8	514.9	6	425.2	514.0
March	Primary	9	719.7	3	320.9	719.1
	Production	7	405.2	3	426.8	440.6

The periodic univariate autoregressive moving average and multivariate autoregressive moving average were used for monthly data. Using the SAMS 2007 software, all of the above models were fitted to the data. The provided had a relatively good performance in predicting the discharge values in the years with statistics. As mentioned on the model testing, the CARMA (1,1) and MPAR (2) models were identified and determined suitable for fitting annual data and monthly data, respectively, which could well predict the values of flow rate at 20% of the experimental data. In some graphs, there were some deviations due to the large range of data variation in drought and high water periods.

REEFRENCES

Box G.E. and Jenkins G. (1970). "Time Series Analysis, Forecasting and Control. Revised Edition", Holden day, San Francisco.

Rao. A.R., Kashyap R.L. And Mao L. (1982). Optimal choice on type and order of river flow time series model. *Water Resources Publication*, Vol. **18**, No. 4, pp. 1097-1109.

Ubertini L. (1978). Metodologie statistich per l'analisi delle serie idrrologiche. CNR Publ. No. **14**, Perugia.

Finzi G., Todini E. and Wallis J.R. (1975) "Comment upon multivariate synthesis hydrology." *Water Resources Publication*, Vol. **11**, No. 6, pp 844-850.

Abrishamchi, A., Tajrishi M. And portraitist, b. (2001). "Stochastic models of watershed area West of the Year", *Water Resources Research*, Vol. **1**, No. 1. (in Persian).

Claps P., 1990. "Modelli stocastici dei deflussi Dei corsi d'acqua, Ph.D." Universita degli Studidi Napoli, Italy, 279P.

Claps P. and Rossi F. (1991). "Metodi la generazione sintetica dei deflussi." In: Medelli Idrologic Superficiali nella Pianificazione di Bacino, Politectico di Milano (in Italian).

Salas. J.D., Delleur J.W., Yevjevich V. and Lane W.L. (1985). "Applied Modeling of Hydrologic Time series". Second Edition, Water resources Publication, Littleton, Colorado.

Rossi F. and Silvagni G. (1980). "Analysis of annual runoff series." Third IAHR Symposium on Stochastic Hydraulics, Tokyo, A-18, PP.1-12.