Journal of
Applied Research in Water and Wastewater
Journal homepage: www.arww.razi.ac.ir

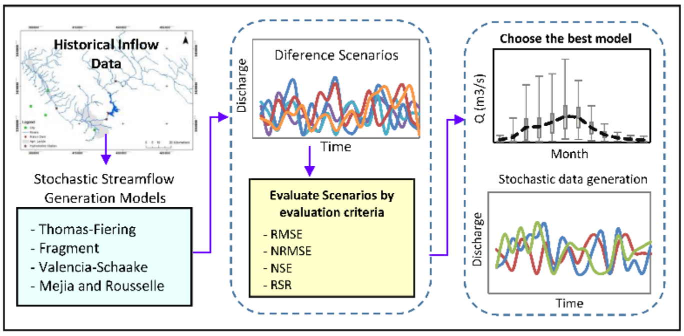
Original paper

Comparison of the performance of stochastic models in the generation of synthetic monthly flows data: A case study on Marun river

Mostafa Bayesteh, Arash Azari*

Department of Water Engineering, Faculty of Science and Agricultural Engineering, Razi University, Kermanshah, Iran.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 30 September 2019
Received in revised form 1 November 2019
Accepted 4 November 2019

Keywords:
Synthetic data generation
Parametric methods
Thomas-Fiering
Monthly flow
Marun river

ABSTRACT

One of the most important issues in planning and managing water resources is the accurate estimation of monthly input discharge of the reservoirs in the future years, which is always associated with uncertainty. To cover these uncertainties, synthetic stream flow data generation models have been used by various researchers to generate stochastic time series. The computational basis of different stochastic models for generating monthly data has been different and this can have a significant effect on their performance. Therefore, selection of the best model of stochastic data generation for accurate planning and management of a water resource system is one of the major concerns of water resources specialists. In this research, the performance of parametric models of synthetic stream flow generation including Thomas-Fiering, Fragment and ARMA (1,1) and ARMA (1,2) combined with Valencia-Schaake and Mejia and Rousselle models were compared and evaluated. For this purpose, 30 years data of monthly discharge of Marun river in Khuzestan province were used and 900 synthetic monthly flow time series were generated using each of the models mentioned above. Based on the obtained results, the ARMA (1,2) model combined with the Valencia-Schaake model was recognized as the best model, considering the very desired performance in preserving the statistical parameters of historical data and generating maximum and minimum discharges related to wet and dry periods in different probabilities. This model can be used with greater confidence to analyze river systems and reservoirs, manage drought and apply water rationing rules in future drought conditions.

1. Introduction

Considering the impossibility of dam construction on all rivers and consequently the impossibility of water regulation, operation of rivers in the low flow conditions will be of particular importance. Any mistake in estimating river discharge, especially in drought conditions, can cause irreparable damage to various sectors such as agriculture, drinking, etc., which will cause many economic and social problems. Therefore, accurate knowledge of river flow, especially in drought conditions, is essential for the management of water resources utilization. On the other hand, since rainfall and consequently river flows are non-deterministic phenomenon, most of the activities of hydrologists and water resources experts have focused on the study, prediction and generation of surface flows. Therefore, many decisions regarding water resource utilization are based on forecasting and time series analysis of hydrological processes. Therefore, it is important to analyze, identify and study the behavior of different models of synthetic time series generation. The application of different models of prediction and generation of synthetic data of hydrological processes for the proper management and planning of water resources is under uncertainty

*Corresponding author Email: a.azari@razi.ac.ir

Please cite this article as: M. Bayesteh, A. Azari, Comparison of the performance of stochastic models in the generation of synthetic monthly flows data: A case study on Marun river, Journal of Applied Research in Water and Wastewater, 6 (2), 2019, 117-125.

conditions. These models are used to generate annual, monthly and daily synthetic rainfall and river discharge data. But since most studies and planning of water resources systems, including utilization of dam reservoirs and climate change issues, etc. are often monthly, the use of monthly flow generation models is more than annual and daily models. Given the importance of this issue in managing water resources under flow uncertainty conditions, several studies have been conducted in recent years on time series generation models. One of the earliest studies of time series is Hazen's research (Hazen. 1914). Spolita and Chander (1974) with the AR model predicted the flow of the Beas River by 100 series. The results showed the good performance of the model in predicting the river flow. Savic et al. (1989) investigated the Thomas Fearing, Fragment, and disaggregation models for synthetic stream flow generation in Yugoslavia. Maheepala and Perera (1996) studied the synthetic stream flow generation of the single-station and multi-station systems in Australia using disaggregated models. The results showed that the models have retained the statistical parameters of the historical data. Prairie et al. (2008) used Markov and KNN chain models to predict the Colorado River flow. The simulation results show that the proposed models retain the statistical properties of the historical flow data. Gualdal and Tongal (2010) investigated stochastic models with synthetic neural network models to predict changes in lake water level in Turkey. Lohani et al. (2012) compared neural network, autoregressive and ANFIS models in modeling and prediction of SUTLEG river flow. The results showed that the ANFIS model performed better than the other two models. Adeli et al. (2014) used the estimated flow of the Talog River to generate synthetic time series of inlet flow of the Tallogo Dam reservoir. This study investigated various stochastic models including autoregressive (AR), moving average (MA) and autoregressive moving average (ARMA) models. The results show the superiority of the ARMA (2,3) model over the other models. Dashora et al. (2015) modeled the monthly flow of the Narmada River in India using the Thomas-Fiering, ANN and SARIMA models. The results of this study showed that the performance of the SARIMA model was better than the other two models. The Thomas Faring model performs better in high-discharge streams and the neural network model in the low flow rates. Montaseri and Heidari (2016) compared the performance of four types of data generation models of Bootstrap, Valencia, Fragment and Thomas-Fiering for synthetic data generation.

Moeini et al. (2017) predicted monthly flow of Gamishan river with ARIMA, ANN and ANFIS models and found that ANFIS model is more capable of detecting effective time delays than ANN model but ARIMA model has high capability in predicting discharge with low values. Aksoy and Dahamsheh (2018) predicted arid regional monthly rainfall in Jordan using a combination of neural network and Markov chain models. The results showed that the combination of neural network model with Markov chain has good performance in predicting monthly rainfall in arid regions.

A review of research shows that one of the most important issues in water resources planning and management is the accurate estimation of monthly input discharge of the reservoirs in the future years, which is always associated with uncertainty. To cover these uncertainties, synthetic stream flow data generation models have been used by researchers to generate stochastic time series. On the other hand, the computational basis of different stochastic models for generating monthly data is different and, this can have a significant effect on their performance. Therefore, the purpose of this study was to select the best model of stochastic monthly flow generation for accurate planning and management of the water resource system of the Marun river.

2. Material and methods

2.1. Study area

The Marun river is one of the major tributaries of the Jarahi River in southwest of Iran that flows into the Marun dam after a distance of 120 km. This study used the historical data of Idnak hydrometric station on Marun river (prepared by khuzestan water and power authority (KWPA)). Considering the role of Marun dam in the supply of downstream agricultural water, urban drinking water, supplying a large part of the environmental needs of the downstream Shadegan wetland, as well as the generation of electricity, predicting long-term inflow time series for future planning is very important. The location of the Marun Basin and the Idnak hydrometric station have been shown in Fig. 1.

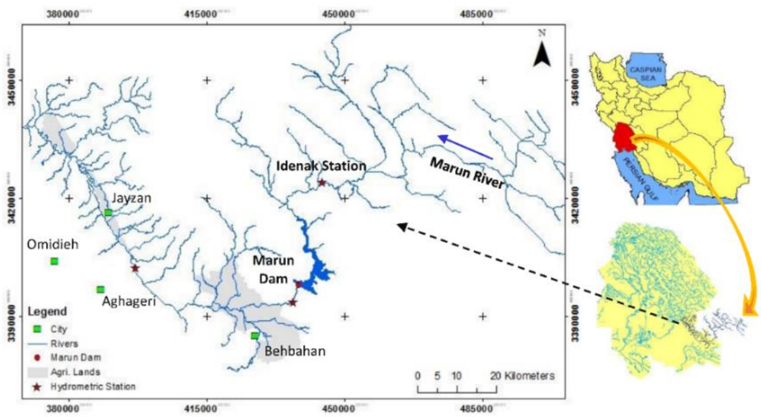


Fig. 1. Location of the study area, rivers and hydrometric stations.

2.2. Synthetic stream-flow generation of the river

In order to manage and decide on water resources systems, it is necessary to have basic information on the amount of surface and underground flows. If the system is a river, the watering flux rate is an important parameter for the management and analysis of surface water utilization. Therefore, considering the importance of estimating the flow rate especially in the future years and months, it is necessary to use appropriate methods of modeling and generation of river flow. Prediction values of these models are always associated with uncertainties and, the number of different time series with these models must inevitably be created in order to cover all uncertainties. Synthetic data generation models allow for the generation of alternative hydrological data sets that are likely to occur in the future, leading to more accurate and realistic results in water resources studies (Srikanthan and McMahon. 2001). In modeling the behavior of different hydrological systems such as catchments, rivers, etc., as more data is available, the mod el is closer to natural conditions and its use in water resources issues leads to more accurate results. However, historical data on hydrological phenomena usually have a very short statistical period of about 20 to 30 years and cannot be an accurate representative of a hydrological system with a life span of several thousand years (Montaseri and Adeloye. 1999). Synthetic data generation models are divided into two main parametric and nonparametric groups. In nonparametric models, synthetic data are directly generated by re-sampling historical recorded data. But parametric models are based on mathematical relationships, which are often defined by linear relationships. The classification of the types of synthetic data generation models has been shown in Table 1. Due to sampling by replacing historical data itself, nonparametric methods generate flow data in the minimum and maximum historical data range and are unable to generate flows out of this range. These methods are incompatible with the expected probability in nature because they cannot generate dry and wet flow series outside this range. Therefore, they were excluded in this study. Several types of parametric methods used in this research will be presented below. Using each of these methods, 900 synthetic stream flow time series (30 30-year monthly flow series) were generated.

2.2.1. Thomas-Fiering model

Please cite this article as: M. Bayesteh, A. Azari, Comparison of stochastic models performance in data generation of synthetic monthly flows : A case study on Marun river, Journal of Applied Research in Water and Wastewater, 6 (2), 2019, 117-125.

Thomas and Fiering (1962) presented this model based on the Markov chain as the relationship (1). This model is used to generate monthly data.

$$Q_{i+1} = \bar{Q}_{j+1} + b_j Q_i - \bar{Q}_j) + e_i s_{j+1}\sqrt{1 - r_j^2} \tag{1}$$

where, Q_i and Q_{i+1} are the monthly flows generated in the months i and i + 1, $\bar{Q}_j$ and $\bar{Q}_{j+1}$ are the historical average monthly flows in the months j, $j+1$ and bj represents the least squares correlation coefficient for the calculation of the month *j* and $j+1$, calculated from relationship 2.

$$r_j = b_j\left[\frac{s_{j+1}}{s_j}\right] \tag{2}$$

where, r_j is the correlation coefficient between j and $j+1$ months, S_j and S_{j+1} are the standard deviation of the months j and $j+1$ and e_i is a random variable with standard normal distribution with mean zero and standard deviation of one. The above model assumes that the process is static in mean and variance.

2.2.2. Autoregressive (AR) model

The Autoregressive model is one of the most common models in stochastic discussions and is obtained by regression on historical series. The general equation of this model is the relationship 3.

$$z_t = \emptyset_1 z_{t-1} + \emptyset_2 z_{t-2} + \cdots \emptyset_p z_{t-p} + \varepsilon_t \tag{3}$$

where, z_t is the flow value per month t and $\emptyset_p$, ..., $\emptyset_2$, $\emptyset_1$ are the parameters, and P is the order of the model. ε_t is a random variable that has a normal distribution with a mean of zero.

2.2.3. Autoregressive- moving average (ARMA (p, q)) model

One of the problems in autoregressive models for modeling hydrological time series is the variation of series under specific conditions. River flow under dry and wet water conditions follows a certain behavior that is not well modeled when using AR models alone. For this reason, another section called the moving average (MA) is added to the autoregressive model called the autoregressive moving average (ARMA) (pq) model (relationship 4).

$$z_t = \emptyset_1 z_{t-1} + \emptyset_2 z_{t-2} + \cdots \emptyset_p z_{t-p} + \varepsilon_t - \theta_1 z\varepsilon_{t-1} - \cdots \theta_q z\varepsilon_{t-q} \tag{4}$$

where, $\emptyset_p$ and θ_p are the parameters of the model, z_t represents the discharge rate at the time t, p and q are the model orders. This series of models generates annual data, and fragment, Valencia-Schaake or Mejia and Rousselle methods must be used to convert these data into monthly flows.

2.2.4. Fragment model

This model presented by Harm and Campbell (1967) is one of the disaggregated parametric models in the generation of monthly data. Porter and Pink (1991) developed this method and introduced it as the Method of Synthetic Fragment. To use this model, we first calculate the monthly to annual historical data ratio for the same year. The values obtained from this ratio are called Fragment. Then the historical annual data is sorted and the fragments of each year are kept in new sorting. Next, using a suitable model, new annual data are generated and disaggregated based on the fragments available in the months of the year.

2.2.5. Valencia-Schaake model

In this method, like the Fragment model, the annual data is first generated with a suitable parametric model, and in the next step, the annual data is disaggregated in the months of the year. The general relationship of this model is as follows.

$$Q_{i+1} = \bar{Q} + \rho \times (Q_i - \bar{Q}) + v_i \times s \times \sqrt{1 - \rho^2} \tag{5}$$

where, Q_i and Q_{i+1} are annual flows generated at time step i and i + 1, $\bar{Q}$ is average annual observational flows, v_i is random variable from standard normal distribution with mean zero and unit variance, ρ is self-correlation coefficient with delay of one, s is standard deviation of annual observational data. After generation, annual data are disaggregated in the months of the year using the relationship 6.

$$Y_v = AX_v + B\varepsilon_v \tag{6}$$

Y_v is a (12 × 1) vector of monthly data for year v, ε_v is a (12 × 1) vector of random variables from the standard normal distribution which is independent of X_v, X_v is the annual data generated, A and B are the coefficients matrix whose dimensions are (12 × 1) and (12 × 12), respectively. The values of A and B are obtained using annual and monthly historical data.

2.2.6. Mejia and Rousselle model

This model was introduced by (Mejia and Rousselle. 1976) for disaggregation of the generated annual data in the months of the year. The general equation for this method is as follows.

$$Y_v = AX_v + B\varepsilon_v + CY_{v-1} \tag{7}$$

C denotes the coefficient matrix with the dimensions of (12 × 12) and Y_{v-1} is the monthly data generated in the year $v-1$. The other coefficients and parameters in this method are similar to the Valencia-Schaake method.

2.3. Data sufficiency test

Hurst coefficient calculation is one of the methods used to assess the adequacy of time series for modeling and generation of synthetic stream flow. This coefficient is used to measure long-term memory of a time series. Long-term memory is defined on the basis of observations of extreme events over a given period of time. If the Hurst coefficient for a time series is greater than 0.5, the length of the time series is sufficient to generation the data (Karamooz and Araghinejad. 2006; Turcotte. 1997). Otherwise, the length of the time series should be increased by different methods. The general relation of this coefficient is as follows.

$$h = \frac{\log(\frac{R}{\sigma})}{\log(\frac{N}{2})} \tag{8}$$

where, h is the Hurst coefficient, N is the number of data, σ is the standard deviation of the data, and R is the difference between the maximum positive value and the lowest negative value of the deviation from the mean time series calculated cumulatively.

Table 1. Classification of different types of synthetic data generation models.

Data generation models	**1. Nonparametric**	1.1 .Bootstrap 1.2. Moving window		
	2. Parametric	2.1 .monthly	2.1.1. Directly	A. Thomas-fiering
			2.1.2. Indirectly	A .Proration method B. Fragment C. Valencia-Schaake D. Mejia and Rousselle
		2.2. annual	2.2.1. AR (P) 2.2.2. ARMA (p,q) 2.2.3. FGN 2.2.4. BL	

2.3.2. Data normalization method

Most probability theories and statistical methods used in hydrology in general and time series analysis have been specifically developed with the assumption that their variables have a normal distribution. Most hydrologic variables do not follow the normal distribution and must be normalized using different methods of data conversion. In this study, the normality of flow discharge data at the Adenak hydrometric station was investigated as a first step of the modeling steps. Given that the historical data did not follow the normal distribution, the Johnson

Please cite this article as: M. Bayesteh, A. Azari, Comparison of stochastic models performance in data generation of synthetic monthly flows : A case study on Marun river, Journal of Applied Research in Water and Wastewater, 6 (2), 2019, 117-125.

transform method was used to normalize the data. Fig. 2 shows the normality test of the historical series of the hydrometric station before and after normalization. According to Fig. 2, after applying Johnson transform to historical data, they are within the upper and lower limits of the normal distribution graph and are normalized to the 95 % significance level.

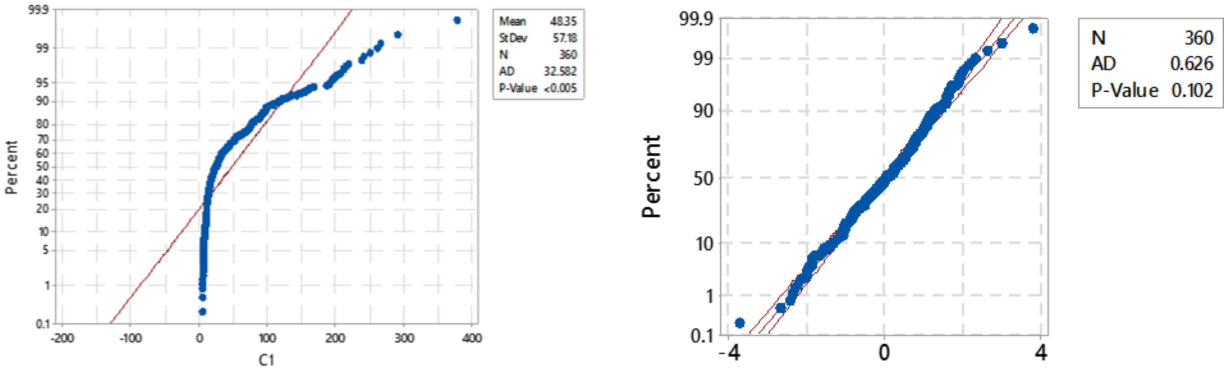


Fig. 2. Distribution of historical data (a) before and (b) after normalization.

2.3.3. Assessment of model performance

The evaluation criteria of the models used in this study were the mean squared error, squared normalized error and absolute error, as well as preserving the statistical properties of historical data (mean, standard deviation, coefficient of variation, maximum and minimum data) by synthetic time series generation models. The mathematical formulas of these criteria have been shown in Eqs. 9 to 12.

$$RMSE = \sqrt{(\sum_{i=1}^{n}(X_i^{obs} - X_i^{sim})^2)/n} \quad (9)$$

$$NRMSE = \frac{RMSE}{X_i^{obs,max} - X_i^{obs,min}} \quad (10)$$

$$NSE = 1 - [\sum_{i=1}^{n}(X_i^{obs} - X_i^{sim})^2 / \sum_{i=1}^{n}(X_i^{obs} - X^{mean})^2] \quad (11)$$

$$\text{RSR} = \frac{RMSE}{STDEV_{obs}} = \left[\sqrt{\sum_{i=1}^{n}(X_i^{obs} - X_i^{sim})^2} \Big/ \sqrt{\sum_{i=1}^{n}(X_i^{obs} - X^{mean})^2}\right] \quad (12)$$

where, X_i^{obs} is the ith observation for the constituent being evaluated, X_i^{sim} is the ith simulated value for the constituent being evaluated, X^{mean} is the mean of observed data for the constituent being evaluated, and n is the total number of observations.

3. Results and discussion

Non-deterministic models of synthetic stream flow generation have been developed largely with the assumption that the data are static (Salas. 1993). Therefore, the models used in data generation should be able to maintain statistically significant parameters of observational time series. Therefore, one of the criteria for selecting a superior model in generating stochastic time series is that the model retains the statistical properties of historical data and also capable of generating peak discharges during the wet months and minimum discharges in dry conditions. According to the above, the value of the Hurst coefficient must be greater than 0.5 for the adequacy of the data series length. The value of Hurst coefficient in this study was 0.74, indicating that the data are sufficient for generating new series. A number of evaluation criteria were used to compare the performance of synthetic data generation models, which will be described below.

3.1. Statistical parameters of monthly flows

As stated, the study used historical monthly flow data of the Idank Station during the 30 years (1989–2018). Important statistical properties of historical data and data generated by stochastic models have been shown in Fig. 3. To generate the monthly data using the Fragment model, an appropriate annual data model has been first generated and these data has then broken into months. Valencia-Schaake and Mejia and Rousselle models also provide monthly values based on annual data. Therefore, the appropriate model must first be selected for generating annual data. Based on the evaluation criteria shown in Table 2, the annual ARMA (1,1) model has the lowest error values. So, it was obtained as the best model. However, this model generated zero discharge during the data generation phase for the river, which is unacceptable given the historical data recorded. Therefore, this model did not perform well for generating the flow in this river. Then, according to Tables 2 and 3, the ARMA (1,2) model has better performance than the other models. Also according to Fig. 3, the ARMA (1,2) model has been more successful in preserving the statistical properties of the historical flow. Valencia-Schaake and Mejia and Rousselle models performed very closely together in generating monthly data using annual data. Therefore, the Valencia-Schaake model was chosen because it was easier to compare with other models and the Mejia and Rousselle model were excluded from the comparisons. Therefore, outputs of ARMA (1,1) and ARMA (1,2) models were disaggregated monthly using Valencia-Schaake method. The following is the comparison of the results of the combination of these two models with the Valencia-Schaake method with other monthly synthetic data generation methods. According to Fig. 3, the Thomas-Fiering model in March and April have had higher average than the historical data and data generated by other models. Thus, it performs poorly in modeling during the wet months. But it has acceptable performance than the rest months of the year. Other models were able to generate data with average close to the river's historical flow, with ARMA (1,1) and ARMA (1,2) models performing better. But the ARMA model (1,1) was not suitable for some months due to the generation of zero data. According to Fig. 3, the ARMA model (1,1) generates maximum discharges near the maximum discharge of historical data and is not capable of generating peak values beyond the historical values which are not probably acceptable. The Thomas-Fiering, ARMA (1,2) and Fragment models have been able to generate maximum discharges from historical river data during the wet months of the year. But the disaggregation of these discharges in the Thomas-Fiering model is inconsistent with historical data. The Thomas-Fiering model has generated more than the minimum river discharge in all months of the year in terms of the generated minimum discharges for dry years, which is not an acceptable result. But other models have been able to model discharges less than historical data and have had satisfactory performance.

Table 2. Evaluation criteria for selecting the best model.

	ARMA (1,1)	ARMA (1,2)	AR (1)	T.F
RMSE	3.54	5.2	4.2	16.68
NRMSE	0.03	0.054	0.047	0.175
NSE	0.99	0.98	0.98	0.77
RSR	0.098	0.145	0.117	0.463

3.2. Empirical disaggregation of historical and generated data

Box plot was used to show the empirical disaggregation of historical and generated data for Idank station. These diagrams have been shown in Fig. 4 to describe the variation range of the historical series generated by the models. Accordingly, the ARMA (1,1) and ARMA (1,2) models combined with the Valencia-Schaake method have the best performance in maintaining the empirical disaggregation of historical data. The Fragment model generates minimum flow data less than the historical minimum flow values in all months. Therefore, because of the lack of coordination of the disaggregation of minimum discharge data in this model with the disaggregation of historical minimum discharge data, its performance is not appropriate in this regard. Also, the

Please cite this article as: M. Bayesteh, A. Azari, Comparison of stochastic models performance in data generation of synthetic monthly flows : A case study on Marun river, Journal of Applied Research in Water and Wastewater, 6 (2), 2019, 117-125.

disaggregation of monthly data across all generated statistical years in Fragment model is obtained on the basis of available historical records. In this method, the monthly disaggregation coefficients are constant values based on the categories of generated flows. Therefore, this model gives unreal probabilistic values due to the non-deterministic monthly disaggregation and therefore it is not suitable for generating stochastic data. A review of the Box Plot charts shows that the Thomas-Fiering model has not been successful in simulating maximum and minimum discharges in all months. In this model the values of peak flows generated in all months are higher than the historical peak flows, which is probabilistically impossible. Also, the minimum generated discharges in all months are higher than the historical recorded minimum discharges, and this model is not able to simulate drier flows compared to historical flow data in these months.

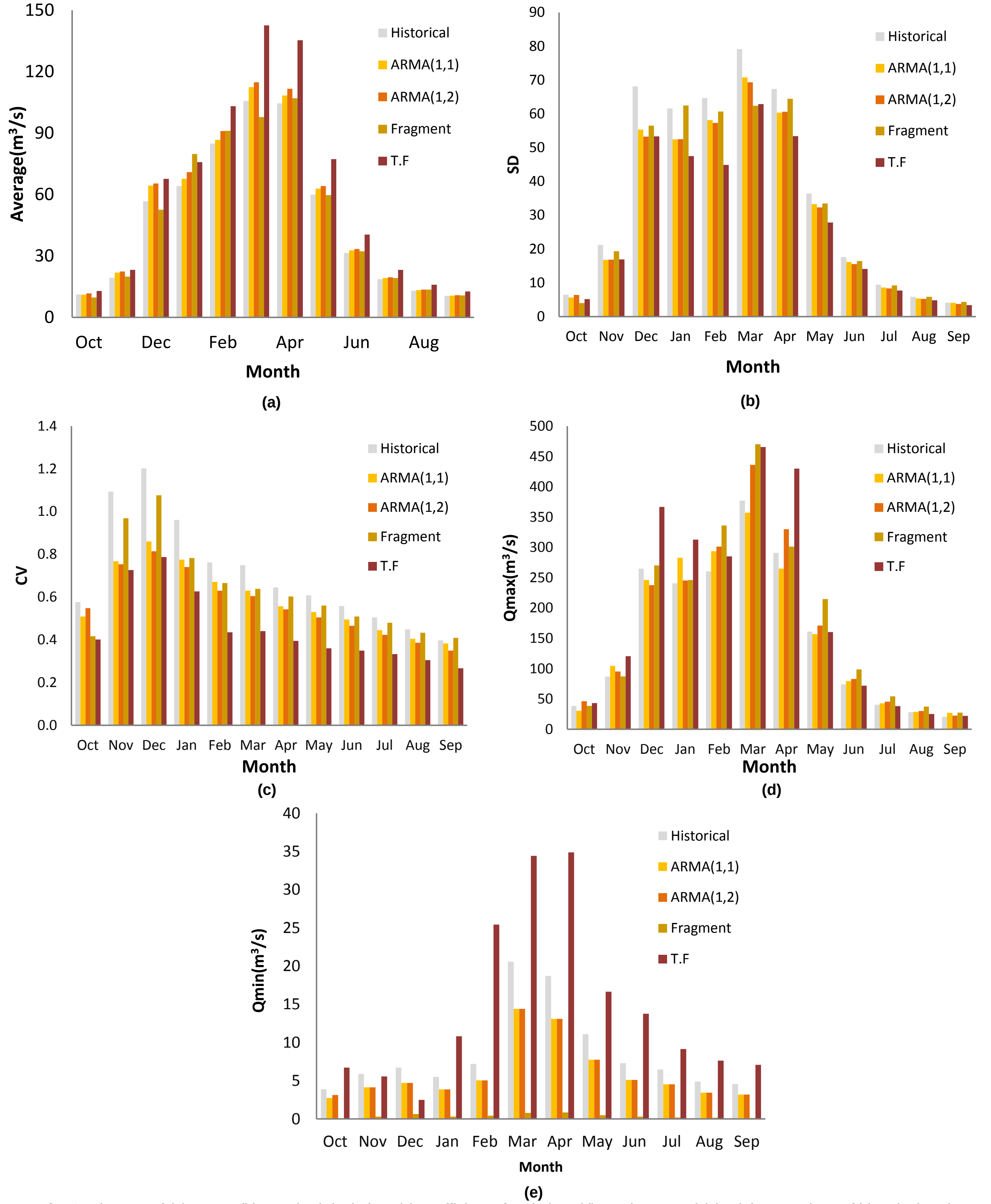
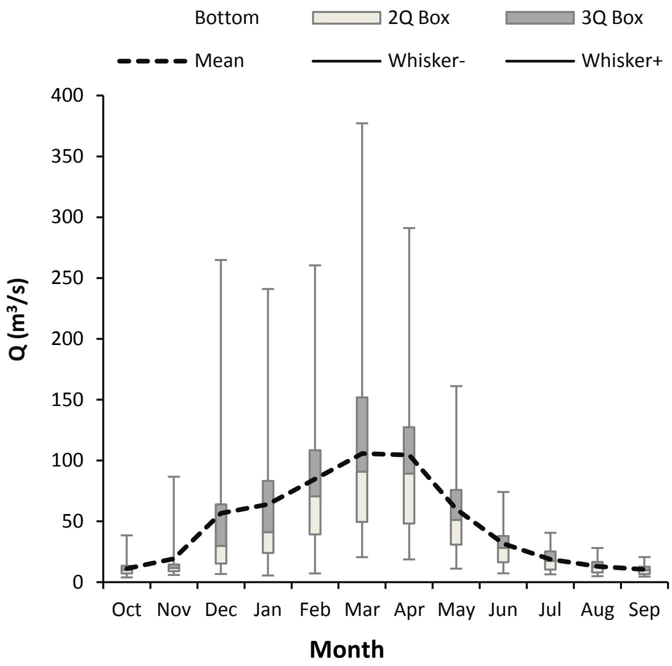


Fig. 3. Diagram of (a) mean, (b) standard deviation, (c) coefficient of variation, (d) maximum and (e) minimum values of historical and generated data.

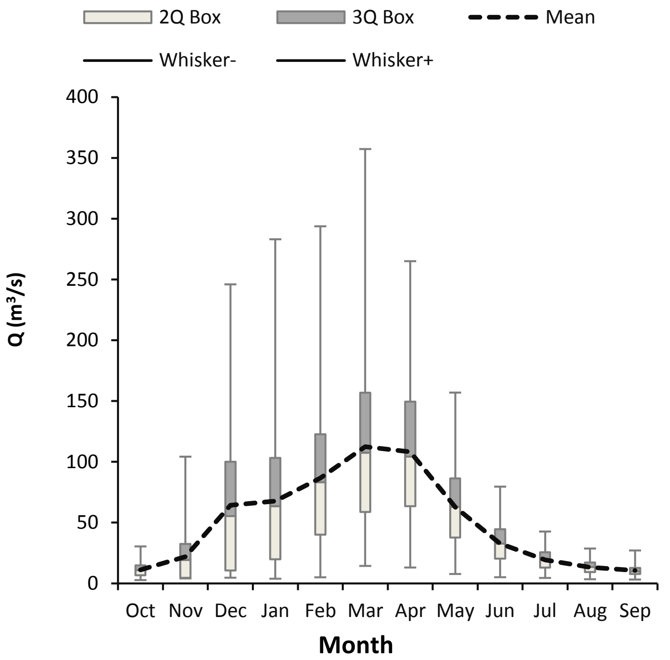
Please cite this article as: M. Bayesteh, A. Azari, Comparison of stochastic models performance in data generation of synthetic monthly flows : A case study on Marun river, Journal of Applied Research in Water and Wastewater, 6 (2), 2019, 117-125.

3.3. Flow-frequency curves

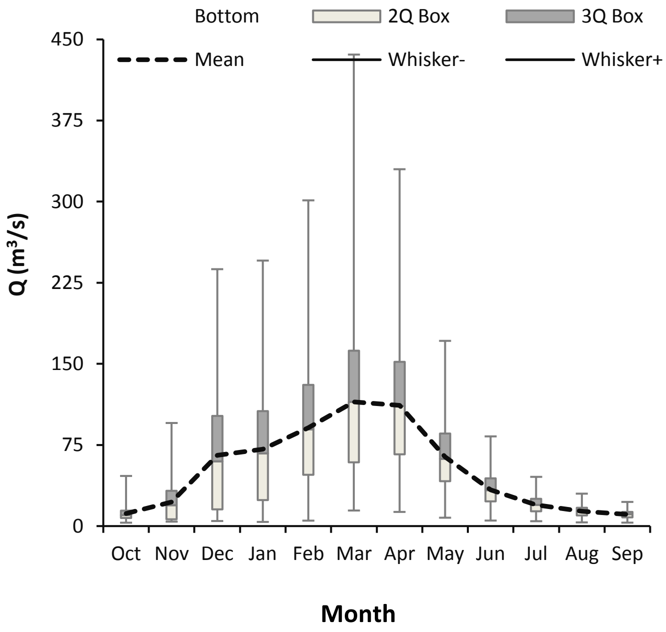
The flow-frequency curve shows the relationship between frequency and discharge rate. Discharge data are used as flow continuity curves in the design, management and planning of water resources such as design of dams and control structures, power plants and climate change topics (Castellarin et al., 2004). In this study, the discharge values with probabilities of 0.1, 0.2, 0.5, 0.8, 0.9 and 0.95 were calculated using data generated in different stochastic models. As shown in Fig. 5, most of the models used are able to well model the river flow at low probability values (0.1 to 0.5). Calculated discharges with these probabilities can be used more for water resources planning for synthetic feeding of aquifers and supplying agricultural and public consumption. But for higher probabilities (0.9 and 0.95), the Thomas-Fiering model generates discharge rate up to several times higher than historical data and also than other models. Because these discharges are related to the low water conditions of the river and are mostly used for planning of drinking water supply and industry in low water conditions, therefore, estimating of several times of the discharge rates more than actual values by Thomas-Fiering model can lead to wrong planning and incorrect decision making. In the Fragment model, due to the close proximity of the discharge values with probabilities of 0.9 and 0.95 to the historical values, this model is not able to predict the probability of occurrence of dry conditions than the present situation and is therefore weak. But in lower probability (0.1 to 0.5) its performance is desirable. Given all the probabilities, the ARMA model (1,2) has the best performance in estimating the flow-probability values.



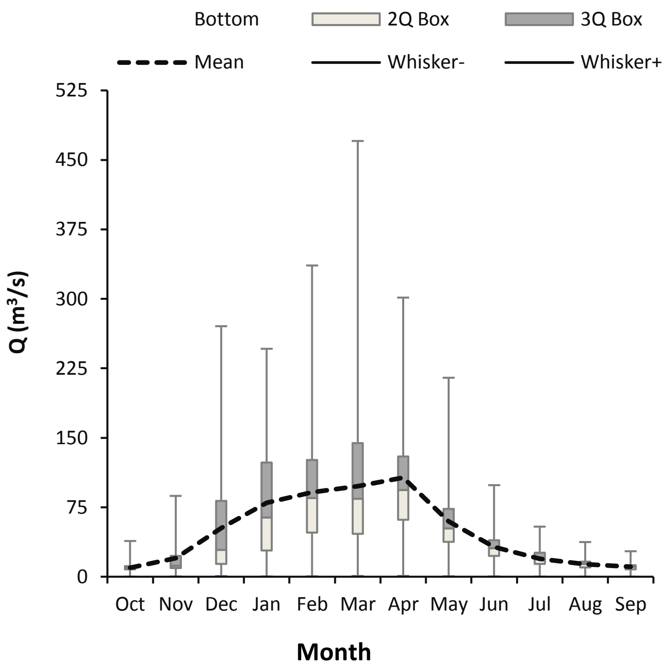
(a)



(b)

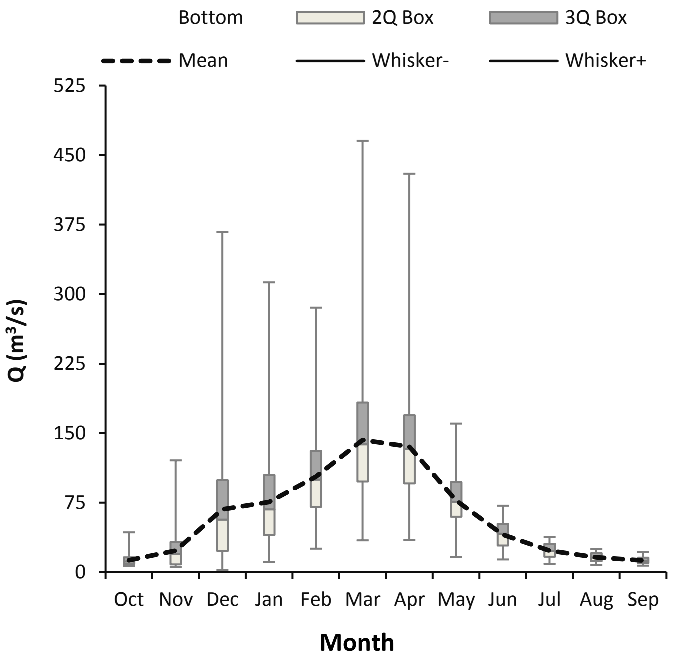


(c)



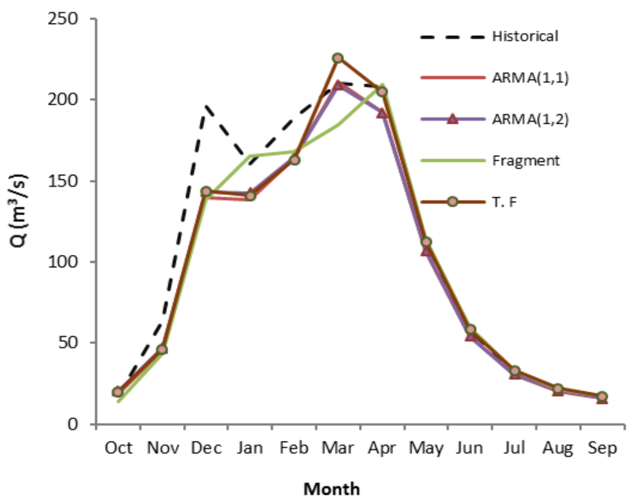
(d)

Please cite this article as: M. Bayesteh, A. Azari, Comparison of stochastic models performance in data generation of synthetic monthly flows : A case study on Marun river, Journal of Applied Research in Water and Wastewater, 6 (2), 2019, 117-125.

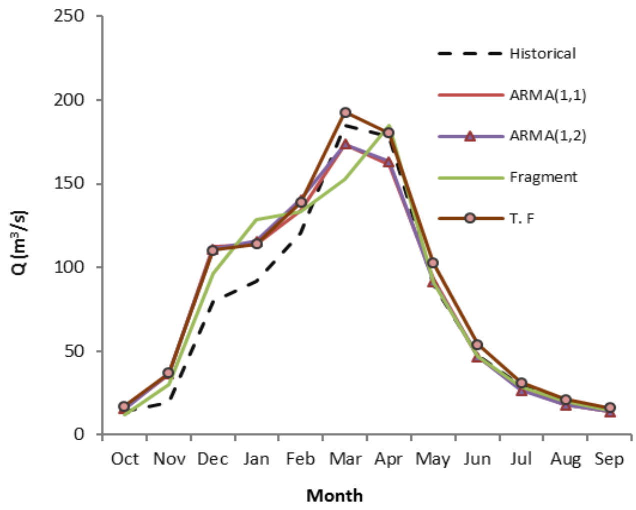


(e)

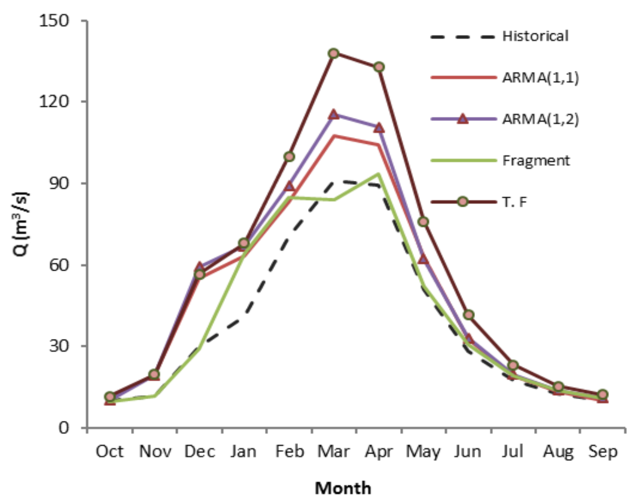
Fig. 4. Box plot of monthly flows of (a) historical data and (b, c, d and e) generation data by ARMA (1,1), ARMA (1,2), Fragment and Thomas-Fiering models.



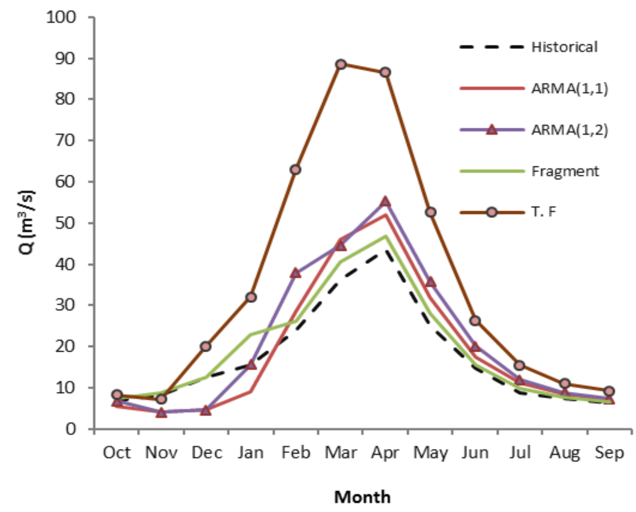
(a)



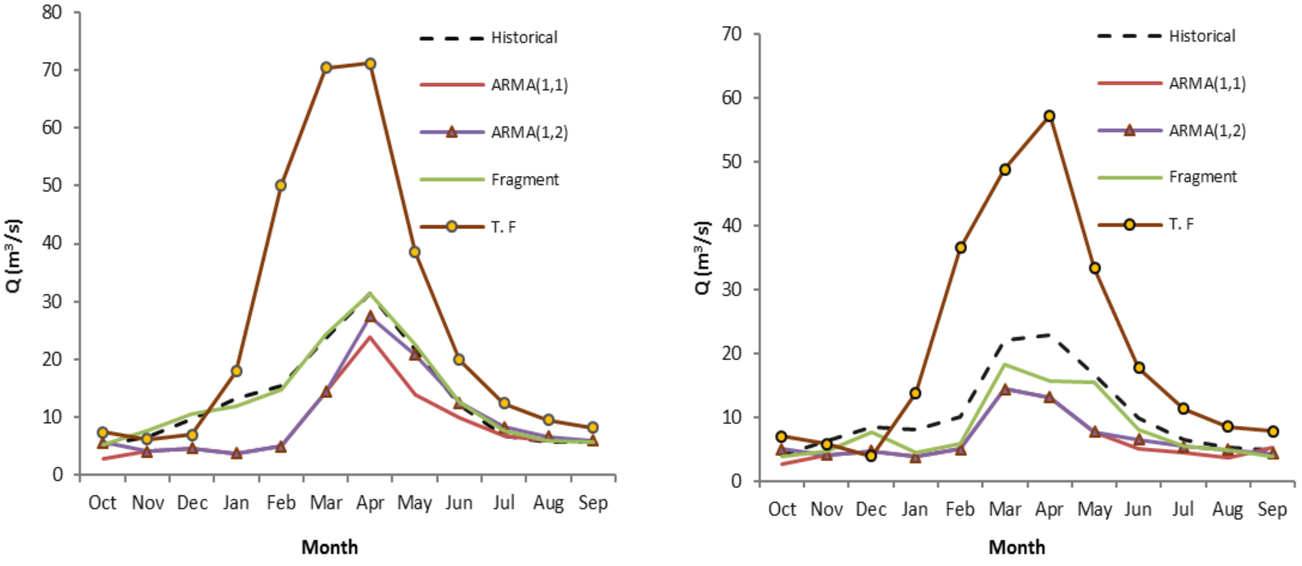
(b)



(c)



(d)



(e) (f)

Fig. 5. Flow-Frequency curves of historical and generated data in different probabilities (a) p=0.1, (b) p=0.2, (c) p=0.5, (d) p=0.8, (e) p=0.9 and (f) p=0.95.

4. Conclusions

Monthly stochastic data generation models have great importance because of the need to be aware of the status of water resource systems in the coming years, including storage tanks, drought monitoring, etc., which is always associated with uncertainty. The use of data generated by these models allows for a more accurate analysis of the actual behavior and performance of such systems, especially in drought conditions. The computational basis of different stochastic models for generating monthly data has been different and this can have a significant effect on their performance. This research compared and evaluated the performance of parametric models of synthetic stream flow generation including Thomas-Fiering, Fragment and ARMA (1,1) and ARMA (1,2) combined with Valencia-Schaake model. Based on the obtained results, the ARMA (1,2) model combined with the Valencia-Schaake model was recognized as the best model according to the optimum performance in preserving the statistical parameters of historical data and also the generation of maximum and minimum discharges related to wet and dry periods in different probabilities. Therefore, this model can be used with greater confidence to analyze river systems and reservoirs, manage drought and apply water rationing rules in future drought conditions. The design, management and operation of water resource systems including storage reservoirs and water rationing in drought conditions are particularly affected by minimum discharges. Therefore, comparing the range of minimum amounts of historical and generated data has a great importance in the stochastic analysis of these systems. The ARMA (1,1) model combined with the Valencia-Schaake model, despite its ability to simulate stochastic time series and retain statistical properties of historical data, generates near-zero discharge in some months, which does not correspond to river reality and available historical data. The results of this study are based on Marun River flow data in southwestern Iran with semi-arid climate. Therefore, the use of stochastic data generated by the superior model can provide important information for adjusting and applying the rationing rules of the Marun Dam reservoir properly under water deficit conditions.

References

Adeli A., Fathi Moghadam M., Musavi Jahromi H., Using stochastic models for Synthetic time series generation and predicting the input of Talogh dam reservoir, International Journal of Analytical Research on Water Resources and Development 5 (2014) 1-13.

Aksoy H., and Dahamsheh A., Markov chain-incorporated and synthetic data-supported conditional Synthetic neural network models for forecasting monthly precipitation in arid regions, Journal of Hydrology 562 (2018) 758-779.

Castellarin A., Galeati G., Brandimarte L., Montanari A., Brath A., Regional flow-duration curves: Reliability for ungauged basins, Advances in Water Resources 27 (2004) 953-965.

Dashora L., Singal S.K., Srivastav D.K., Software application for data driven prediction models for intermittent streamflow for Narmada River Basin, International Journal of Computer Applications 113 (2015) 9-17.

Güldal V., and Tongal, H., Comparison of recurrent neural network, adaptive Neuro-Fuzzy inference system and stochastic models in Egirdir lake level forecasting, Water Resources Management 24 (2010) 105–128.

Harms A.A., and Campbell T.H., An extension to the Thomas-Fiering Model for the sequential generation of stream flow, Water Resources Research 3 (1967) 653-661.

Hazen A., Storage to be provided in impounding reservoirs for municipal water supply, Transactions of the American Society of Civil Engineers 77 (1914) 1539-1669.

Karamooz W., and Araghinejad S., Advanced hydrology, 1st ed., Amir Kabir University of Technology Publication, Tehran, (2006) 264-267.

Lin G.F., and Lee F.C., An aggregation-disaggregation approach for hydrologic time series modeling, Journal of Hydrology 138 (1992) 543-557.

Lohani A.K., Kumar R., Singh R.D., Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques, Journal of Hydrology 442 (2012) 23-35.

Maheepala S., and Perera B.J.C., Monthly hydrologic data generation by disaggregation, Journal of Hydrology 178 (2012) 277-291.

Mejia J.M., and Rousselle J., Disaggregation models in hydrology revisited, Water Resources Research 12 (1976) 185–186.

Moeini H., Bonakdari H., Abdollahi S., Performance evaluation of some statistical models and soft computing in river flow forecasting, Iran Water Resources Research 3 (2017) 201-206.

Montaseri M., and Adeloye A.J., Critical period of reservoir systems for planning purpose, Journal of Hydrology 224 (1999) 115- 136.

Montaseri M., and Heydari J., Comparison of performance of stochastic models in generation of monthly flow and rainfall data, Iran Water Resources Research 3 (2016) 69-84.

Porter J.W., and Pink B.J., A method of synthetic fragments for disaggregation in stochastic data generation, Proceedings in

Hydrology and Water Resources Symposium, Australia (1991) 187-191.

Prairie J., Nowak K., Rajagopalan B., Lall U., Fulp T., A stochastic nonparametric approach for streamflow generation combining observational and paleo reconstructed data, Water Resources Research 44 (2008) 1-11.

Salas J.D., Analysis and modeling of hydrologic time series, In Handbook of Hydrology, Edited by Maidment, McGrow – Hill book CO., New York, (1993).

Savic D., Burn D.H., Zrinji Z., A comparison of stream flow generation models for reservoir capacity-yield analysis, Water Resources Bulletin 25 (1989) 977-983.

Turcotte D.L., Fractals and chaos in geology and geophysics, Cambridge university press, New York, (1997) 403.