

Evaluation of Suitability Criteria in Stochastic Modeling

T. Baran¹, Ü.G. Bacanlı²

¹ Assoc. Prof. Dr. Dokuz Eylül University, Faculty of Engineering, Civil Eng.Dept., Tinaztepe Campus, İzmir – Turkey
e-mail: turkay.baran@deu.edu.tr

² Ph. D. in Civil Eng. Pamukkale University, Faculty of Engineering, Civil Eng.Dept., Kınıklı Campus, Denizli – Turkey
e-mail: ugbacanlı@pamukkale.edu.tr

Abstract: Evaluation, gauging and analysis of hydrologic data have great importance in the development planning and management of water resources. Actually, the available data usually do not represent the population; therefore the process should be modeled. The models can be used for generating data for planning and design of hydraulic structures or forecasting. In stochastic hydrology, one of the important problems is the choice of appropriate model type among various models. In the presented study, observed and synthetic data are modeled and different methods used in the choice of the best fit model are compared. For this purpose, AR(1), AR(2), AR(3), ARMA(1,1) and ARMA(1,2) models which are mostly used in hydrology are investigated.

Key words: Stochastic Models, Suitability Criteria in Stochastic Modeling.

1. INTRODUCTION

Hydrometric data evaluations have vitally importance in water resources development, planning and management. In general, available data do not represent the population of process. Hence, it is also necessary to modeled hydrologic time series. The models can be used to generate data for planning and designing or forecasting.

Almost all hydrologic time series of daily, weekly and monthly values have deterministic components occurring due to astronomic cycle and therefore they are periodic-stochastic series in nature; on the other hand none of the hydrologic time series are purely deterministic or periodic. The deterministic components in time series can be mainly classified as transient components (trends and jumps) and periodic components.

Several types of stochastic models have been proposed during the last two decades for the stochastic modeling of hydrologic time series, in general, and streamflow time series in particular. The proposed models include the autoregressive, fractional Gaussian noise, autoregressive and moving average, ARIMA and FARMA. Hence, an important problem in stochastic hydrology is to select or identify the type of model for representing the hydrologic time series at hand. In common practice such model identification is usually done by judgment, experience, or personal preference. In some cases, though, the statistical properties of the various alternative models as well as the statistical characteristics of the sample time series are used for identifying the most appropriate type of model for the particular case at hand. It is, of course, desirable that in addition to the above factors, physical considerations must be used for aiding in the identification of the model type (Salas & Smith, 1981).

Actually, the exact mathematical models of a hydrologic time series are never known. The inferred population model is only an approximation. The exact model parameters are also never known in hydrology; they must be estimated from limited data. Identification of models and estimation of their parameters from available data are often referred in the literature as time series modeling or stochastic modeling of hydrologic series (Salas et al., 1981).

The purpose of constructing the models of stochastic processes is to generate synthetic processes for the considered variable with the aid of these models. With the use of generated processes, it can

be possible for the investigations of planning and management of water resources to consider for flows not only the observed sample but also the other samples which come from the same population. So, the system behavior can be investigated not only according to the available sample but also with aid of synthetic series (Bayazit, 1981).

In the presented study, the Saint Lawrence River annual runoff series, which was previously studied in literature, are used. Evaluation was made with three different data groups appropriate to the observed and statistical properties of AR(3) model which is chosen as the population of the observed data.

In the first evaluation, according to the observed data, synthetic data generation was taken for 97 years length, and the results were compared with the observed data.

In the second evaluation, 3600 years length annual mean flow series is generated. Then this data set is divided into 30 subgroups and modeled. The goodness of fit for each one subgroup is tested by using VAR(e) and AIC Akaike information criterion (AIC).

In the third evaluation, the 600 years length data was divided into subgroups of 100 ... 600 years. By this way the effect of data length on the suitability criteria is investigated.

2. TIME SERIES MODELING

2.1. Linear Autoregressive Models

General expression of Markov models can be defined as:

$$x_i = \sum_{j=1}^m \alpha_j x_{i-j} + \varepsilon_i \quad (1)$$

in which, x_i : flow of i. th year, α_j : autoregressive coefficients (model parameters), ε_i : a normally distributed variable which constituted an independent process, m : order of model. m th order Markov model means which the flow of any year is depend on previous m year flows (Bayazit, 1981).

2.2 Stationary Autoregressive-Moving Average Models

Autoregressive (AR), moving average (MA) and combining ARMA models are statistically based on the assumption that a random normally distributed input (ε_t) is transformed into a dependent stationary output (Z_t) by a linear system. Denoting the random input (random shocks or white noise) with zero mean and variance σ_ε^2 by ε_t and output (linear stochastic process) with zero mean and variance $\sigma_z^2 = \sigma_x^2$ by \tilde{Z}_t it can be written as:

$$\tilde{Z}_t = X_t - \mu_t = \varepsilon_t + \Psi_1 \varepsilon_{t-1} + \Psi_2 \varepsilon_{t-2} + \dots \quad (2)$$

or, introducing the backward shift operator, $B^m \varepsilon_t = \varepsilon_{t-m}$,

$$\tilde{Z}_t = \Psi(B) \varepsilon_t \quad (3)$$

where X_t is the time series generated at time t and μ_x is the level (or mean), and

$$\Psi(B) = 1 + \Psi_1 B + \Psi_2 B^2 + \dots \quad (4)$$

is the linear operator (or transfer function of the filter) that transforms ε_t into \tilde{Z}_t (Box & Jenkins, 1976).

Under suitable conditions \tilde{Z}_t can be defined as a weighted sum of its own past values:

$$\varepsilon_t = \tilde{Z}_t - \Pi_1 \tilde{Z}_{t-1} - \Pi_2 \tilde{Z}_{t-2} - \dots \quad \text{or} \quad \varepsilon_t = \Pi(B) \tilde{Z}_t \tag{5}$$

where $\Pi(B)$ is the inverse of the transfer function $\Psi(B)$,

$$\Pi(B) = \Psi^{-1}(B) = 1 - \Pi_1 B - \Pi_2 B^2 - \dots \tag{6}$$

To ensure that stochastic process generated by the Ψ weights is stationary, the infinite series $\Psi(B)$ given by Eq. (6) must converge for $|B| \leq 1$. On the other hand, the process is invertible (Eq. 5) only if the $\Pi(B) = \Psi^{-1}(B)$ series is convergent for all $|B| \leq 1$ (Box & Jenkins, 1976).

The above given formulation of linear stationary stochastic processes is more general and the model given Eq. (5) is an infinite order moving average process while Eq. (6) is an infinite order autoregressive process. These models proposed first by Yule in 1927 are commonly applied in time series modeling with finite number of Ψ or Π weights.

A mixture of finite number of moving average Ψ and autoregressive Π weights of linear stationary stochastic processes is so called ‘‘autoregressive-moving average’’ (ARMA) processes. Denoting Π_j weights by α_j and Ψ_j weights by $-\theta_j$ an ARMA (p,q) process with finite number of autoregressive (p) and moving average (q) parameters can be written as:

$$\tilde{Z}_t = \sum_{j=1}^p \alpha_j \tilde{Z}_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j} \tag{7}$$

Various estimating methods can be used for parameter estimation. Moments, maximum likelihood and least squares methods are widely used estimation methods. In the presented study, model parameters are calculated with moments and maximum likelihood methods.

In hydrologic practice p and q are seldom greater than 2, and in most cases an ARMA (1,1) model is found satisfactory. The stationary ARMA models have a physical justification in hydrology. The low flows in a river mainly result from groundwater and the flow at a particular time is a fraction of previous flows during the recession period, which may be represented by an autoregressive dependence structure. The high flows during the wet season are formed mainly by heavy rainfall or snowmelts or both, and therefore may be represented by moving average scheme (Salas et. al., 1985).

3. OBSERVED DATA

In the presented study, annual flows of Saint Lawrence River are used. Statistical properties of observed data are presented at Table 1 and runs of the annual flows can be shown at Figure 1.

Table 1: Statistical properties of annual flows of Saint Lawrence River for 1860 – 1956 period (Bacanli & Baran, 2004).

Statistical Properties			
Mean	(μ)	(m ³ /sn)	570,62
Standard Deviation	(σ)	(m ³ /sn)	48,71
Variation Coefficient	(C _v)		0,0854
Maximum Value		(m ³ /sn)	672,08
Minimum Value		(m ³ /sn)	433,33
Skewness Coefficient	(C _s)		0,085
Kurtosis	(k)		2,5454

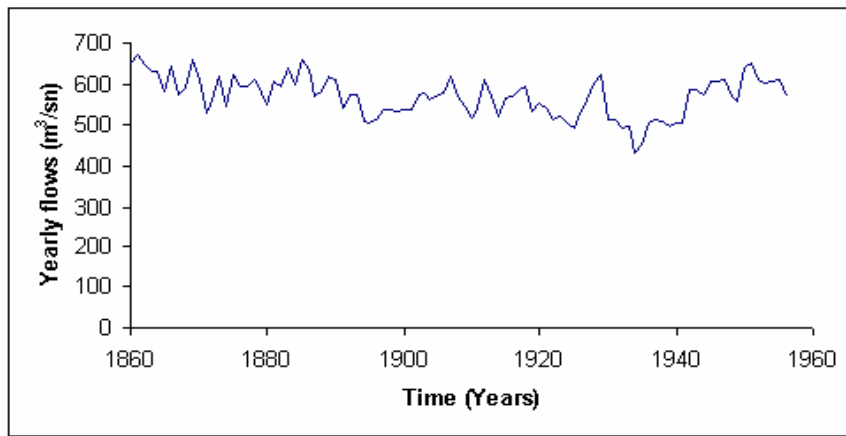


Figure 1: Observed annual flows of Saint Lawrence River for 1860 – 1956 years period (Bayazit 1981).

For the determination of the best fitted model of observed flows, ordinates of ACF: autocorrelation function (Figure 2) of observed data up to $k = 40$ and PACF: partial autocorrelation function (Figure 3) up to $k = 25$ are calculated. Figure 2 and Figure 3 shows that the autocorrelation coefficients slowly approach to zero after certain k value not being suddenly zero even for large k value. It is demonstrated that the terms of autoregressive models terms are in the model. Partial autocorrelation coefficients come close to zero for k bigger than 3.

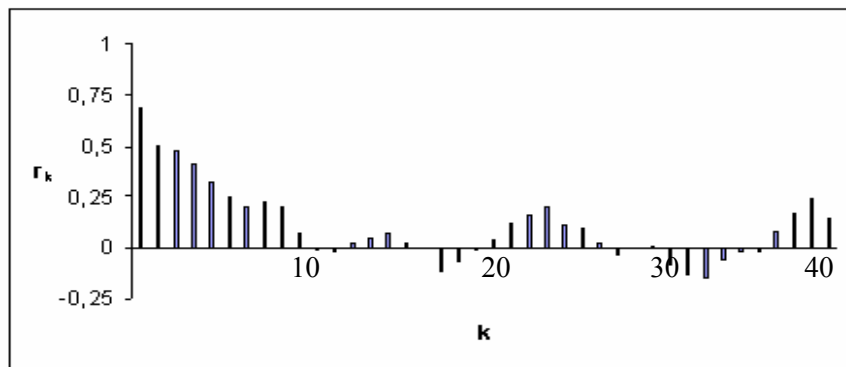


Figure 2: Lag-one autocorrelation coefficients of annual flows.

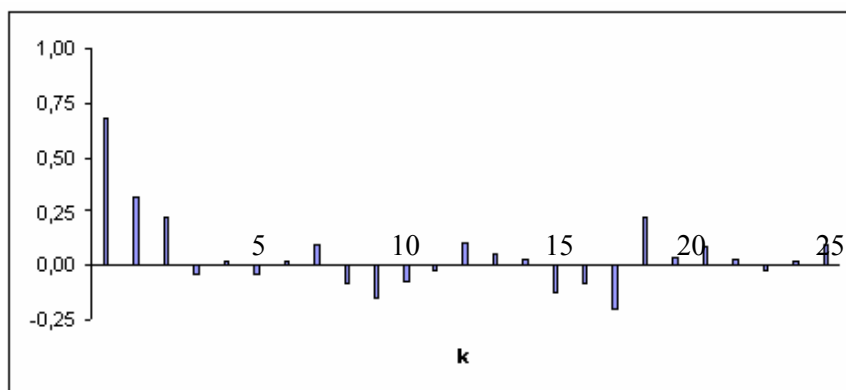


Figure 3: Partial autocorrelation coefficients of annual flows.

4. METHODOLOGY

Annual runoff series can be accepted as stationary after run and jump variable components were cleaned. Although, some annual flow series have very small autoregressive components, annual runoff series can be accepted to be independent process (Yevjevich, 1972; Salas et al., 1985).

In the presented study, the first evaluation, according to the observed data, synthetic data generation was taken for 97 years length, and the results were compared with the observed data.

In the second evaluation, 3600 years length annual mean flow series are generated. Then this data set is divided into 30 subgroups and modeled. The goodness of fit for each subgroup is tested by Var(e) and AIC .

In the third evaluation has been made in the 600 years length generic annual runoff series. The generic series and observed data set have the same statistical properties and parameters. The generic 600 years length data was divided into subgroups of 100 ... 600 years for to take into consideration the effect of data length on the suitability criteria.

5. MODELING OF FLOWS

During the modeling stage population known $X_{p,z}$ values are investigated as observed data. The data is divided into 100, 200, 300, 400, 500 and 600 years length subgroups and the best fitted model for each subgroup is investigated. For this purpose, the model parameters for AR(1), AR(2), AR(3), ARMA(1,1) and ARMA(1,2) are determined and X_i (model) and ε_i (residual) values are calculated according to these parameters.

The processes are standardized by,

$$Z_{i,\tau} = \frac{x_{i,\tau} - m_\tau}{\sigma_\tau} \quad (\tau=1,2,3, \dots, 12) \quad (8)$$

equation (8) and autocorrelation coefficients are calculated. The parameters of AR(1), AR(2), AR(3), ARMA(1,1) and ARMA(1,2) models are tested for stationary and invertibility. ACFs of synthetic data are shown in Figure 4 for 97 years and in Figure 5 for 600 years.

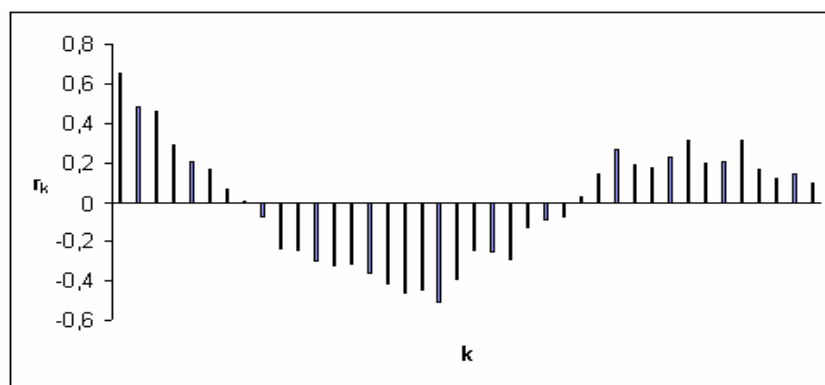


Figure 4: Autocorrelation coefficients of 97 yearly flows data.

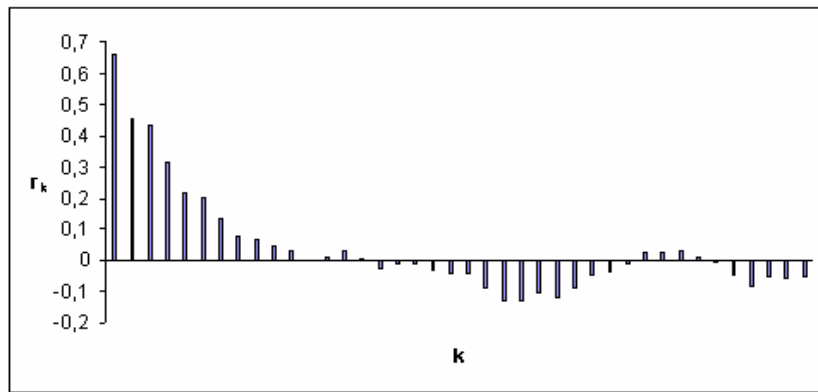


Figure 5: Autocorrelation coefficients of 600 yearly flows data

6. TEST OF GOODNESS OF FIT

6.1. Minimum Residual Variance [Var(e)]

One of the methods for testing the model suitability is the calculation of Var(e) value where:

$$e = X_{Si} - X_{Mi} \quad (9)$$

X_{Si} is observed data; X_{Mi} is estimated data for each model. Minimum Var(e) value helps for deciding the suitable model (Salas et al., 1985).

6.2. Akaike Information Criterion [AIC]

Another method for testing the model suitability is Akaike Information Criterion (AIC). Akaike Information Criterion (AIC) proposed by Akaike in 1974 is the mathematical formulation, which considers the principle of parsimony (Salas et al., 1985). In order to compare among computing ARMA models, Akaike recommends

$$AIC(p, q) = N \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q) \quad (10)$$

where N is the sample size and $\hat{\sigma}_\varepsilon^2$ is the maximum likelihood estimate of the residual variance; k is the number of distribution parameters ($k = p+q$); p is α coefficient number; q is θ coefficient number. The model, which gives the minimum AIC number, is the one to be selected.

6.3. Autocorrelation Function – Correlogram [ACF]

General expression for the mathematical models of stochastic processes can be defined as:

$$x_i = f(x_{i-1}, x_{i-2}, \dots) + \varepsilon_i \quad (11)$$

It is accepted that, ε_i is formed an independent process. To check the suitability of selected model, the residual of model ε_i is obtained.

The correlogram is used for determining whether of ε_i process is an independent. Anderson test is frequently used test for the hypothesis of if the correlogram belongs to an independent process. Mean and variance of the sampling distribution of autocorrelation coefficient (r_k) is given below:

$$E(r_k) = -\frac{1}{N-k}, \quad \text{Var}(r_k) = \frac{N-k-1}{(N-k)^2} \quad (12)$$

where N is the sample length and k is the lag number. By using these expressions, the confidence region of the correlogram in a certain confidence limits is determined. If the percentage of calculated r_k values which are outside this area is smaller than α , the independence hypothesis is accepted and the selected model is suitable (; Salas et al., 1985).

In this study furthermore, the residuals ε_i are tested, to check whether they are independent and normal. The Box-Pierce Porte Manteau lack of fit test is used for testing the independence of ε_i series. This test is used by statistical formulation of Q (Bayazit, 1981; Salas et al., 1985):

$$Q = N \sum_{k=1}^L r_k^2(\varepsilon) \quad (13)$$

where $r_k(\varepsilon)$ is correlogram of residual (ε_i) and L is maximum lag. The statistics Q is approximately $\chi^2(L-p)$.

7. APPLICATION

7.1. St. Lawrence River Annual Runoff

In the presented study, 97 years length (1860-1956) annual runoff series of St Lawrence River in USA are modeled. The results of test of goodness of fit are summarized in Table 2. The results confirmed that AR(3) can be selected as the best fit model.

Table 2: The results of goodness of fit for annual runoff series of St Lawrence River

OBSERVED DATA	VAR(e)	AIC	ACF
Evaluations for all models	ARMA(1,2)	ARMA(1,2)	All Models
Evaluations for AR type models	AR(3)	AR(3)-AR(1)	All Models

7.2. Synthetic Annual Mean Flows

7.2.1. Comparison of Synthetic and Observed Annual Mean Flows

According to the observed data, synthetic data generation was taken for 97 years length in the first evaluation. The population of the synthetic series assumed AR(3) and the results of test of goodness of fit are summarized in Table 3.

Table 3: The results of goodness of fit for 97 years length synthetic annual runoff series generated according to the statistical properties of Saint Lawrence River.

SYNTHETIC DATA	VAR(e)	AIC	ACF
Evaluations for all models	ARMA(1,2)	ARMA(1,2)	All Models
Evaluations for AR type models	AR(3)	AR(3)	All Models

7.2.2. Sampling Distributions of Synthetic Series

In the second evaluation, 3600 years length annual mean flow series are generated. Then this data set is divided into 30 subgroups (97 years data length) and modeled. The results of test of goodness of fit for each subgroup, tested by Var(e) and AIC, are summarized in Table 4.

7.2.3. Effect of Data Length

As it mentioned above, first evaluation has been made in the 600 years length generic annual runoff series. The synthetic series generated under assumption of population can be acceptable as AR(3). The results of test of goodness of fit are summarized in Table 5.

Table 4: The results of goodness of fit for 97 years length synthetic annual runoff subseries, generated according to the statistical properties of Saint Lawrence River (*m*: absolute, *f*: relative frequency)

VAR(e)	All Models		AR Models	
	<i>m</i>	<i>f</i>	<i>m</i>	<i>f</i>
AR(1)	0	0	13	0,43
AR(2)	0	0	0	0
AR(3)	12	0,40	17	0,57
ARMA(1,1)	9	0,30	---	---
ARMA(1,2)	9	0,30	---	---
AIC	<i>m</i>	<i>f</i>	<i>m</i>	<i>f</i>
AR(1)	4	0,13	19	0,63
AR(2)	0	0	0	0
AR(3)	6	0,20	11	0,37
ARMA(1,1)	9	0,30	---	0
ARMA(1,2)	11	0,37	---	0

Table 5: The results of goodness of fit for 600 years length synthetic annual runoff series generated according to the statistical properties of Saint Lawrence River.

N	Acceptable – Best Fitted models					
	AR and ARMA models			AR Models		
	VAR(e)	AIC	ACF	VAR(e)	AIC	ACF
100	ARMA(1,2)	ARMA(1,2)	All Models	AR(1)	AR(1)	All Models
200	ARMA(1,2)	AR(1)	All Models	AR(2)	AR(1)	All Models
300	ARMA(1,2)	ARMA(1,2)	All Models	AR(2)	AR(1)	All Models
400	ARMA(1,2)	AR(3)	All Models	AR(2)	AR(3)	All Models
500	ARMA(1,2)	AR(3)	All Models	AR(2)	AR(3)	All Models
600	ARMA(1,2)	AR(3)	All Models	AR(2)	AR(3)	All Models

8. CONCLUSIONS

According to Table 2 and 3, the results of test of goodness of fit are showed nearly same results for observed and synthetic data. According to ACF, all investigated models can be acceptable. ACF can only be used for acceptable models. According to Var(e) and AIC, ARMA(1,2) is the best for all investigated models. If only autoregressive models take into consideration, AR(3) is the best for these criterion.

Annual runoff series can be accepted as stationary after run and jump variable components were cleaned. The evaluated data sets have no trends or jumps (Figure 1), and as shown in Table 1, the statistical properties of observed data have demonstrated that the probability distribution can be acceptable as Normal Distribution. The coefficient of variation is very small. Because of these, it is

expected that AR type model can define better than ARMA types. Hence, AR(3) is selected as the best model.

In the second evaluation, 30 data subgroups (97 years length) are modeled. The results of test of goodness of fit for each subgroup, tested by Var(e) and AIC, and the absolute and cumulative frequencies are presented in Table 4.

According to Table 4, if Var(e) criterion is used for the test of goodness of fit, the population can be predicted accurately;

- as 40% probability for all investigated models,
- as 57% probability for autoregressive models.

If AIC criterion is used for the test of goodness of fit, the population can be predicted accurately;

- as 20% probability for all investigated models,
- as 37% probability for autoregressive models.

The evaluations of long period synthetic data are approved that ARMA(1,2) is the best model, if Var(e) criterion takes into consideration. If only autoregressive models are evaluated, AR(2) appears as the best model.

According to AIC criterion, ARMA type models is better than AR types for short data length, but AR(3) is predicted as the best model when data length is extended. If only autoregressive models take into consideration, the similar consequences are obtained.

According to ACF, all investigated models can be acceptable both for observed and 97 year length synthetic series.

Although Akaike Information Criterion (AIC) was suggested to be used for deciding fit of ARMA models in relevant literature; to predict according to this criteria is more suitable than minimum residual variance for long-term data.

The sampling distributions of synthetic series have showed that, AIC criterion can be inappropriate for testing the goodness of fit for the short-term data. As it mentioned above, the second evaluation approved that minimum residual variance criterion is seem more accurate both for ARMA, AR and only autoregressive models.

ACF criterion can not predict the best model. Var(e) and AIC criterion have different advantages and disadvantages depending on the data length. Hence, it can be said that the new criterion must be investigated for the test of goodness of fit for stochastic modeling.

REFERENCES

- Akaike, H.: 1974, A New Look at the Statistical Identification Model Identification. IEEE Transactions on Automatic Control, 19: 716-723.
- Bacanlı, Ü. G., Baran, T.: 2004, Evaluation of Suitability Criteria of Stochastic Modeling in Annual Time Series (in Turkish). Proceedings of the IV. National Congress of Hydrology, İTÜ, 21-25 June 2004, İstanbul, Turkey.
- Bayazit, M.: 1981, *Statistical Methods in Hydrology* (in Turkish). İstanbul, İTÜ.
- Box, G. E. P., Jenkins, G. M. : 1976, *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco.
- Salas, J. D., Smith, R. A. (1981). Physical Basis of Stochastic Models of Annual Flows. Water Resources Research, 17, 2:428-430.
- Salas, J. D., Obeysekera, J. T. B., Smith, R. A.: 1981, Identification of Streamflow Stochastic Models. ASCE Journal of Hydraulics Division, 107, HY7: 853-868.
- Salas, J.D.; Delleur, J.W.; Yevjevich, V.; Lane, W.L. : 1985, *Applied Modeling Of Hydrologic Time Series*. Littleton: Water Resources Publications (First Printing 1980), 484 p.
- Yevjevich, V. :1972, *Probability and Statistics in Hydrology*. Water Resources Publications, Fort Collins, Colorado, 302 p.