An entropy approach for diagnostic checking in time series analysis

Türkay Baran[1]* and Ülker G Bacanli[2]

[1]Dokuz Eylül University, Faculty of Engineering, Civil Engineering Department, Kaynaklar Campus, 35160, İzmir – Turkey

[2]Pamukkale University, Faculty of Engineering, Civil Engineering Department, Kınıklı Campus, Denizli – Turkey

Abstract

It is a fact that a hydrological time series cannot be defined as a true model in practice. One of the important problems in stochastic hydrology is to determine the most appropriate model, and therefore modellers have certain flexibilities in exercising their subjective judgment in model identification. For this purpose, autocorrelation function [ACF], minimum residual variance [Min Var(e)], and Akaike Information Criterion [AIC- AICC-modified AIC- and FPE-final prediction error-] are widely used for testing the goodness of fit (model identification or diagnostic check) in time series modelling. The objective of this paper is to investigate diagnostic checking criteria, to compare their performance for linear autoregressive (AR) models, and to define a new entropy-based criterion (transinformation).

In the presented study, observed and synthetic data sets are modelled and recognised criteria are evaluated in order to compare the diagnostic checking. All data sets are investigated for AR(1), AR(2), AR(3), ARMA(1,1) and ARMA(1,2) models which are mostly used in hydrology. The results showed that the performance of the transinformation criterion is superior to the other investigated diagnostic checking criteria.

Keywords: time series modelling, diagnostic checking, order determination, entropy, transinformation

Introduction

The main purpose of water resource engineering is to provide maximum benefit from the limited water resources and to protect us from possible damages. Evaluations of hydrological data have great importance in the development, planning and management of water resources. In fact, the available data usually do not represent the population; therefore the process needs to be modelled. The models can be used to generate data for planning and designing of hydraulic structures or forecasting.

The purpose of constructing 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, the planning and management of water resources could possibly be produced by considering not only the observed sample but also the other samples which come from the same population. So, the system behaviour can be investigated not only according to the available data but also with the aid of synthetic series (Salas et al., 1981).

It is obvious that if more reliable scenarios can be produced by the selection of the best model for the time series, then decision making will be more reliable. In time series analysis, various types of models are used according to the usage purpose and the type of the streamflow series in hydrology. The proposed models include the autoregressive (AR), fractional Gaussian noise, autoregressive and moving average (MA) (ARMA), broken line, ARMAMarkov, and shifting level models. Hence, an important problem in stochastic hydrology is to select or identify the most appropriate model to best represent the hydrological time series in question. 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 (Salas and Smith, 1981).

In general, time series modelling can be organised in four stages:

- The selection of the type of model
- The identification of the form of the model
- The estimation of the model parameters
- The diagnostic checking of the model. The diagnostic checking is only partially described by the words, 'testing the goodness of fit' (Box and Jenkins, 1976).

Various statistical tests are available for testing hypotheses in hydrological time series modelling. Although mathematical statistics comprise various parametric and non-parametric tests, practice has shown that a small number of these procedures would satisfy the needs in the analysis and modelling of hydrological time series. Salas et al. (1985) suggested using the Anderson test of the correlogram and Porte Manteau Lack of Fit test for independence in time. They classified the cumulative periodogram test, and the tests for normality, under the title of testing the goodness of fit. They also suggested testing for over-fitting, tests for the parsimony of parameters after the identification and estimation of the model parameters for testing the goodness of fit (Salas et al., 1985). Goodness of fit tests can also be used for comparison of parametric and nonparametric models; in order for compilation of models; model identification or diagnostic checking of the models (Fan and Yao, 2003).

Autocorrelation function [ACF], minimum residual variance [Min Var(e)], and Akaike Information Criterion [AIC, AICC-modified AIC, and FPE-final prediction error-] are widely used for testing the goodness of fit (model identification or diagnostic

* To whom all correspondence should be addressed.
☎ +90 (232) 4127030; fax: +90 (232) 4534279;
e-mail: turkay.baran@deu.edu.tr.
Received 24 March 2006; accepted in revised form 16 March 2007.

checking) in time series modelling. The objective of this paper is to investigate and to test diagnostic checking criteria, to compare their performance and to define a new entropy-based criterion. For this purpose, autoregressive (AR) and autoregressive – moving average (ARMA) models which are widely used in stochastic hydrology, are investigated. To compare the proposed entropy-based criterion with well-known methods, synthetically generated and observed annual flow series were used.

Time series modeling

Linear autoregressive models

The general expression of Markov or linear autoregressive models can be defined as:

$$x_t = \mu + \sum_{j=1}^{p} \phi_j (x_{t-j} - \mu) + \varepsilon_t \quad (1)$$

where:

- x_t = flow of t^{th} year
- μ = mean
- ϕ_j = autoregressive coefficients (model parameters)
- ε_t = a normally distributed random variable which constitutes an independent process with mean zero ($\mu = 0$) and variance σ_ε^2 (noise, error term)
- p = order of model

p^{th} order Markov model is often denoted as the AR(p) or simply AR model in which the flow of any year is dependent on previous p year flows (Salas et al., 1980; Salas, 1992).

The AR(p) model contains (p+2) unknown parameters (μ, ϕ_1,....., ϕ_p and σ_ε^2), which in practice have to be estimated from the data (Box and Jenkins, 1976; Salas et al., 1980; Salas, 1992). The first and second order Markov models [AR(1), AR(2)] are widely used in hydrology (Bayazıt, 1981).

Stationary autoregressive-moving average models

General expression of autoregressive (AR), moving average (MA) and combining 'autoregressive-moving average' (ARMA) models can be defined as:

$$x_t = \mu + \sum_{j=1}^{p} \phi_j (x_{t-j} - \mu) + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \quad (2)$$

where:

- x_t = flow of t^{th} year
- μ = mean
- ϕ_j = autoregressive coefficients (model parameters)
- θ_j = moving average coefficients (model parameters)
- ε_t = a normally distributed random variable which constitutes an independent process with mean zero (μ =0) and variance σ_ε^2 (noise, error term)

with

- p = number of autoregressive (ϕ_1,....., ϕ_p) parameter
- q = number of moving average (θ_1,....., θ_q) parameter

These models are defined with the expression ARMA or ARMA (p, q).The parameter number for ARMA (p, q) th order model are (p + q+ 2) number (μ, ϕ_1,....., ϕ_p, θ_1,....., θ_q and σ_ε^2) (Salas et al., 1980; Bayazıt, 1981; Salas,1992).

In hydrological 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 base flows in a river mainly result from groundwater and the flow at any particular time is a fraction of previous flows during the recession (low-flow) 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 a moving average scheme (Salas et al., 1985).

Diagnostic checking or order determination

It is a fact that hydrological time series cannot be defined by a true model in practice. Therefore modellers have certain flexibility in exercising their subjective judgment in model identification.

On the other hand, the purpose of model identification should be to select a model that is statistically sound and practically meaningful. A parsimonious model is always preferable when two candidate models appear to be equally acceptable. Routine steps of model identification processes are listed below from a data-analytical point of view (Fan and Yao, 2003):

i. The 1st step is **examination of time series plot.** This step aims to identify obvious trends, seasonal components, jumps, etc.
ii. The 2nd step is **examination of correlogram** (ACF). The residual trend and/or seasonal components may show up in a correlogram
iii. The 3rd step is **determining the MA-order from the ACF and the AR-order from the partial autocorrelation function (PACF)**
iv. The last step is **determining the orders using AIC or other information criteria.**

The Akaike Information Criterion (AIC) proposed by Akaike (1974) is the mathematical formulation that considers the principle of parsimony (Salas et al., 1985). In order to compare computing ARMA models, Akaike recommends:

$$AIC = N * \ln(\hat{\sigma}_\varepsilon^2) + 2 * k \quad (3)$$

where:

- N = sample size
- $\hat{\sigma}_\varepsilon^2$ = maximum likelihood estimate of the residual variance
- p = number of autoregressive (AR) coefficients
- q = number of moving average (MA) coefficients
- k = number of distribution parameters (k = p + q)

The model which gives the minimum AIC number is the one to be selected.

Hurvich and Tsai (1989) used the Kullback-Leibler information with its unbiased estimators and Modified Akaike Information Criterion (AICC). The modified criterion is expressed as:

$$AICC(p,q) = N * \ln(\hat{\sigma}_\varepsilon^2) + \left(\frac{2(p+q+1) * N}{N-p-q-2} \right) \quad (4)$$

The model which gives the minimum AICC (min AICC) number is the one to be selected.

An alternative procedure for order determination in AR modelling is the Final Prediction Error Criterion (FPE) due to Akaike (1969), the basic idea of which is very simple. The FPE criterion is defined as:

$$FPE(p) = \hat{\sigma}_\varepsilon^2 \frac{N+p}{N-p} \quad (5)$$

The model which gives the minimum FPE (min FPE) number is the one to be selected.

Various procedures described by Akaike have been proposed to modify the criterion in order to obtain a consistent estimator and to find a field of application. All criteria [FPE, AIC, AICC] proposed by Akaike are asymptotically equivalent (Brockwell and Davis, 2002; Fan and Yao, 2003). Hence, the values of criteria converge with increasing data number.

These criteria called 'information based methodology' proposed by Akaike are used (1974, 1985) for diagnostic checking. The AIC is based on two basic ideas. The first is the adoption of the expected predictive performance for the evaluation of a model. The second is the use of the log likelihood as the measure of the goodness of the model (Akaike, 2003).

One of the other methods which can be used for diagnostic checking is to calculate the variance of error terms. The variance of error terms (e) is determined from the difference of model prediction ($\hat{X}_i$) and observed data (X_i):

$$e_i = X_i - \hat{X}_i \qquad (6)$$

The decision criterion is based on the variance, consequently this method can also be defined as 'information based'.

Concept of entropy and transinformation criterion

The concept of entropy is originated from the classical thermodynamics and plays a major role in thermodynamics and statistical mechanics. The word 'entropy' was introduced by Clausius and derived from 'transformation'. Later, Boltzmann attained a new definition to entropy by analysing microscopic states of a thermodynamic system. According to Boltzmann, 'the macroscopic maximum entropy state corresponded to a thermodynamic configuration which could be realised by maximum number of different micro-states'. Later, Von Neumann generalised the classical expressions of Boltzmann to quantum mechanics by using the concept of a density matrix (Wehrl, 1978; Schrader, 2000).

In a seminal article Shannon (1948) adopted the concept of entropy into information theory. Through his significant contributions to communications theory (later known as information theory), Shannon showed that entropy describes the amount of uncertainty in any probability as a measure of information in probabilistic terms.

Similar to the role of thermodynamic entropy in physics, informational entropy has found a wide area of application in various fields including water resource engineering. In these applications, uncertainties associated with (or information conveyed by) systems of concern have been measured by the probabilistic definition of entropy. The versatile uses of the concept in water resources are essentially based on informational entropy (Harmancıoğlu et al., 1992). The entropy theory has been applied in hydrology and water resources for measuring the information content of random processes, evaluation of information transfer between hydrological processes, assessment of recharge systems for a river basin, evaluation of data acquisition systems, assessment of model performance, assessment of regional information on floods, and designing water quality monitoring network. A comprehensive review of the application of entropy theory in hydrology and water resources is given by Singh (1997; 2000; 2003).

The entropy concept is a fairly objective criterion in comparing various mathematical models and has proved to give successful results in hydrological applications performed. This characteristic of entropy makes it a very useful tool in the selection of the most appropriate model, and in the evaluation of the degree of completeness and efficiency of a selected model to represent natural phenomena, which, in fact, are some of the major problems of synthetic hydrology.

The entropy of a random variable is a measure of the information gained or reduction of uncertainty. The measures of information are defined by the marginal entropy [H(X)], joint entropy [H(X,Y)], conditional [H(X|Y)] entropy and transinformation [T(X,Y)]. The marginal, joint, conditional entropy and transinformation is defined in information theory such as (Van der Lubbe, 1997; Karmeshu and Pal, 2003):

$$H(X) = -\sum_{i=1}^{n} p(x_i)\log p(x_i) \qquad (7)$$

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{J=1}^{m} p(x_i,y_j)\log p(x_i,y_j) \qquad (8)$$

$$H(X|Y) = -\sum_{i=1}^{n}\sum_{J=1}^{m} p(x_i,y_j)\log p(x_i|y_j) \qquad (9)$$

$$T(X,Y) = -\sum_{i=1}^{n}\sum_{J=1}^{m} p(x_i,y_j)\log\left(\frac{p(x_i,y_j)}{p(x_i)p(y_j)}\right) \qquad (10)$$

where:

x and *y* = two independent variables defined in the same probability space with values x_i and y_j
i = 1, 2,…,*n*
j = 1, 2, …, *m*.
$p(x_i)$, $p(x_i,y_j)$ and $p(x_i|y_j)$ terms are defined as discrete, joint and conditional probabilities.

The specific features of the transformation concept have led to its use in evaluation of model performance and selection of the most appropriate model to represent a hydrological process or system. Accordingly, the model that produces the highest transinformation between observed and simulated data is considered to produce the best fit. Amorocho and Espildora (1973) initiated such an approach and showed the limitations and merits of using the entropy criteria in model evaluation. They have also discussed the selection of class interval size Δx, which is used in the case of continuous variables for $p(x_i)$ of entropy equations approximated by $f(x_i)$. Dx, being the relative class frequency and the length of class intervals, with respect to accuracy of entropy calculations. Further, they pointed out that 'the concepts involved in the measure of uncertainty require that the probability frequency distributions of the outcomes of a process be bounded. In practical applications the unbounded frequency distributions should be truncated; thus they will define a finite region of uncertainty' (Amorocho and Espildora, 1973).

Uslu and Tanrıöver (1979) analysed the entropy concept for the delineation of optimum sampling intervals in data collection systems, both in space and time. Harmancıoğlu (1981) investigated the transfer of information between observations of two stream gauging stations and also showed that the serial-dependence structure of hydrological series can be evaluated by entropy to determine the required order of dependence models. Chapman (1986) extended the original use of the method of Amorocho and Espildora to evaluate the reduction of uncertainty in hydrological data due to application of a model. Chapman proposed a complementary approach to overcome the limitations of the technique. Particularly for the selection of Δx, Chapman claimed that one should better use proportional rather than fixed class intervals. Chapman gave general equations for the proportional class interval and solved them for assumed log-normal and gamma distributions and extended to data series with zero values. Chapman proposed a more general criterion of model performance to be the ratio of the transinformation to the marginal entropy of the observed data. Along similar lines, Baran and Harmancıoğlu (1993) compared the goodness of fit of three models in representing the recession period flows at a stream-gauging station where a significant portion of runoff is made up of effluents originating from karst springs.

The transinformation criterion can also be used as a criterion for diagnostic checking (order determination) in time series modelling like the methods called as 'information based methodology' which defined by Akaike (1974; 1985). The transinformation can be obtained as (Mogheir et al., 2004)

$$T(X,Y) = -\frac{1}{2}\ln(1 - r_{xy}^2) \quad (11)$$

where:

x and y = defined as random variables

r_{xy} = the sample correlation coefficient

The transinformation criterion can also be obtained from Eq. (10) by x and y defined as the time series and the selected model output. This criterion can also be used for selecting the best model among several models (diagnostic checking) as an entropy approach.

Data sets

Annual mean flow series of the Saint Lawrence stream gauging station (SGS) (in USA) and Hurman Creek – Tanır/Gözlerüstü (Nr: 2015/25) SGS (in Turkey) are used for modelling and diagnostic checking (order determination) with various criteria. Annual runoff series can be accepted as stationary process if the trend and jump components are removed. Although some annual flow series have very small autoregressive components, annual runoff series can be accepted as independent process (Yevjevich, 1972; Salas et al., 1985).

The Saint Lawrence River is located in the middle of North America and forms a part of the international boundary between Canada and the USA (Fig. 1). The Saint Lawrence River rises at the outflow of Lake Ontario and the selected SGS is situated there as well. Hence, although both AR(p) and ARMA(p,q) types models are investigated, it is expected that Saint Lawrence River annual mean flow is defined by AR type models. The runs

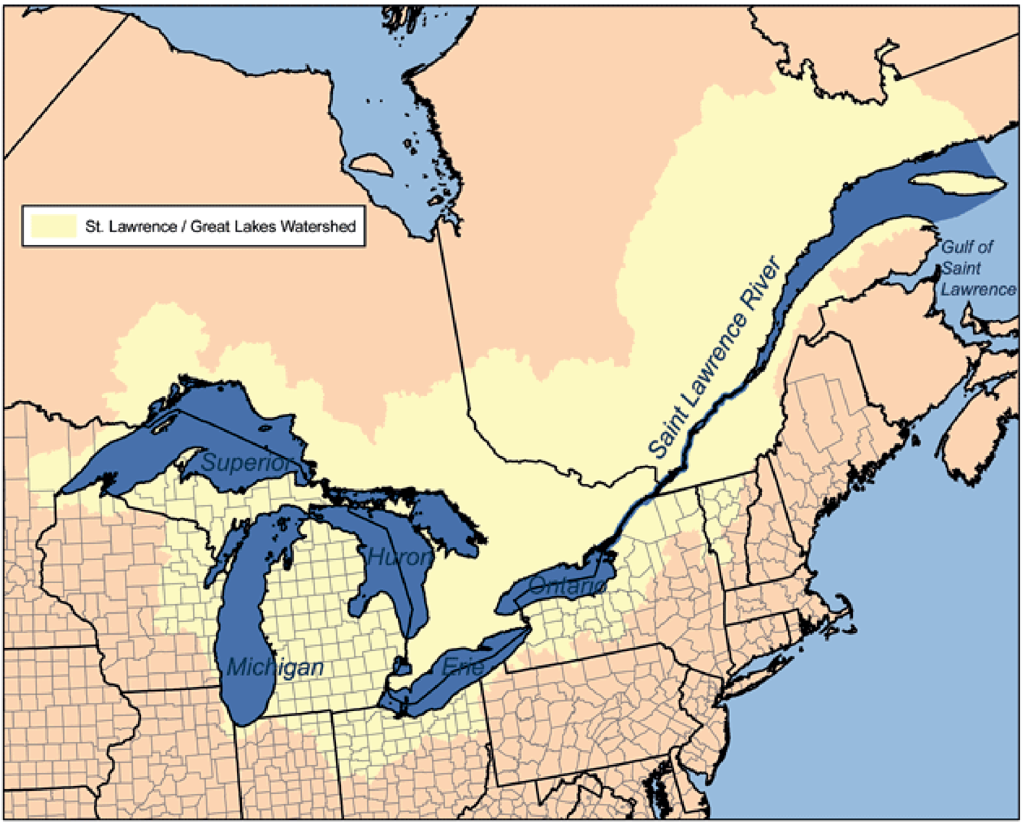

Figure 1
Location of Saint Lawrence River (Wikipedia, 2005)

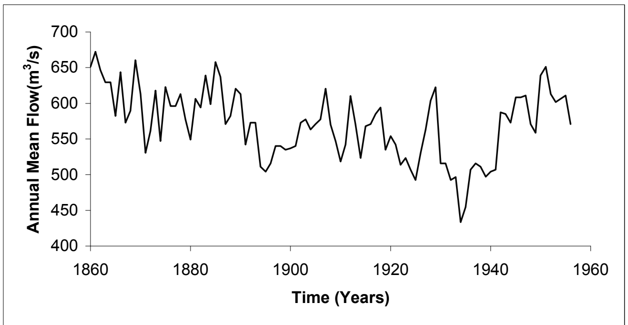

Figure 2
Observed annual mean flows of Saint Lawrence River for the period 1860 to 1956

of the annual mean flows of the Saint Lawrence SGS are shown in Fig. 2.

Hurman Creek is located in the Ceyhan River basin in Turkey (Fig. 2). Tanır/Gözlerüstü SGS (Nr.2015/25) is located upstream of the Ceyhan River of which karst springs contribute 40 to 45% of runoff (Baran et al., 1987; 1995). Although both AR(p) and ARMA(p,q) type models are investigated, it is also expected that Tanır/Gözlerüstü SGS annual mean flow is defined by AR type models. The runs of the annual mean flows of Tanır/Gözlerüstü SGSs are presented in Fig. 4.

The available data of Saint Lawrence annual mean flows have been recorded for the 1860 to 1956 period (Bayazıt, 1981) and Tanır/Gözlerüstü annual mean flows have been recorded over the 1957 to 2000 period (EIE 2000; 2003). Statistical prop-

Available on website http://www.wrc.org.za
ISSN 0378-4738 = Water SA Vol. 33 No. 4 July 2007
ISSN 1816-7950 = Water SA (on-line)

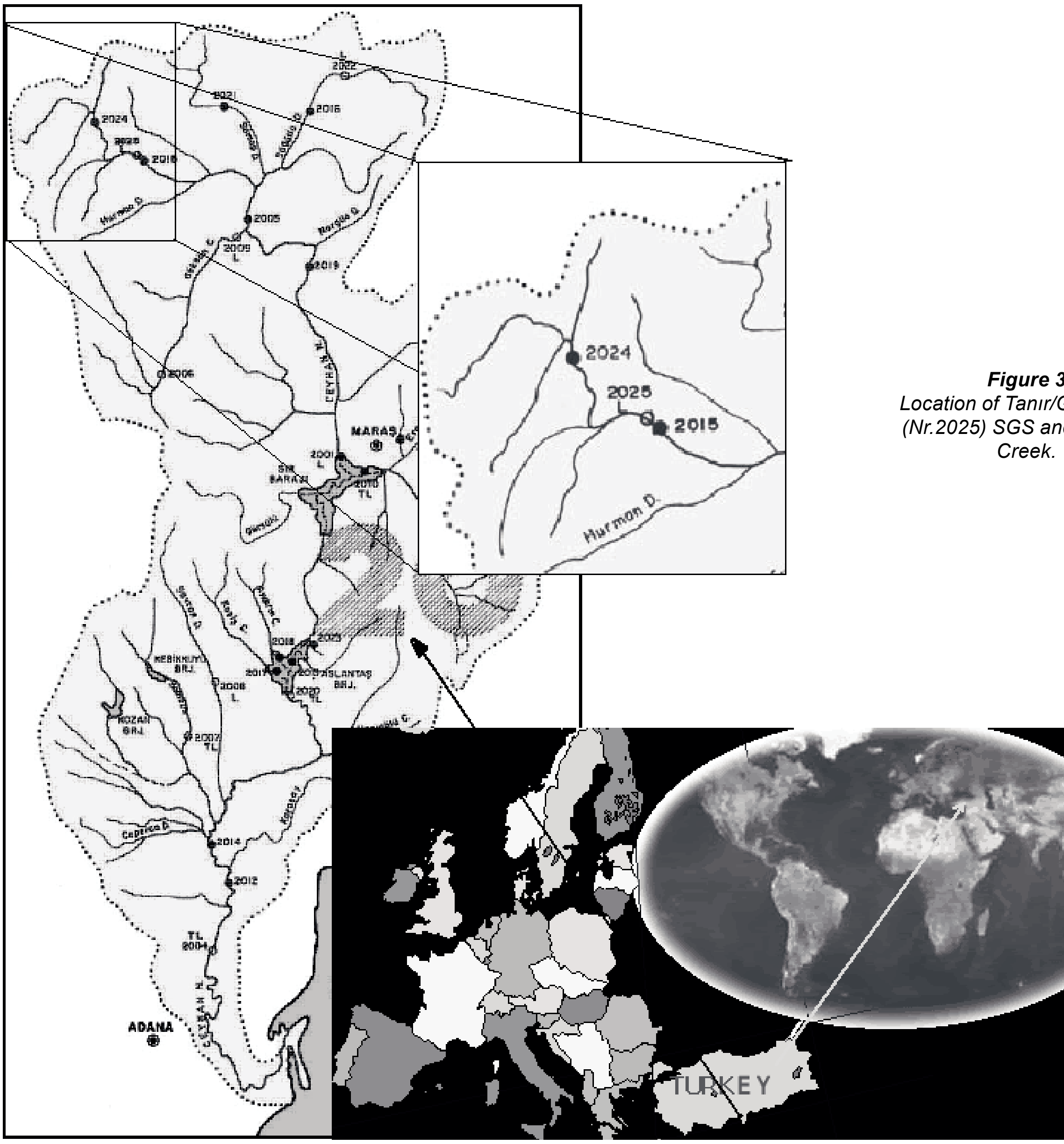

Figure 3
Location of Tanır/Gözlerüstü (Nr.2025) SGS and Hurman Creek.

erties of observed data are presented in Table 1.

The parameters of investigated AR(p) and ARMA(p, q) type models for Saint Lawrence and Tanır/Gözlerüstü SGS annual mean flows are presented in Table 2 and Table 3.

The other data sets are designed according to the observed data's statistical properties and formed as generated synthetic series. The 1st synthetic data set groups are generated as 97 years and 44 years annual mean flow records according to the observed data.

The 2nd synthetic data sets are generated as 600 years length annual mean flows, which is investigated by dividing into sub-groups of 100 … 600 years. In this step, the effect of data length on diagnostic checking criterion is investigated.

The 3rd and last synthetic data sets consisted of 30 subgroups of generated data all containing 97 and 44 members and all of them are evaluated as observed Saint Lawrence and Tanır/Gözlerüstü SGS's annual mean flow series. Thus, evaluation of diagnostic checking criterion is also investigated with three synthetic data groups appropriate to the observations of Saint Lawrence and Tanır/Gözlerüstü SGSs and the selected model parameters.

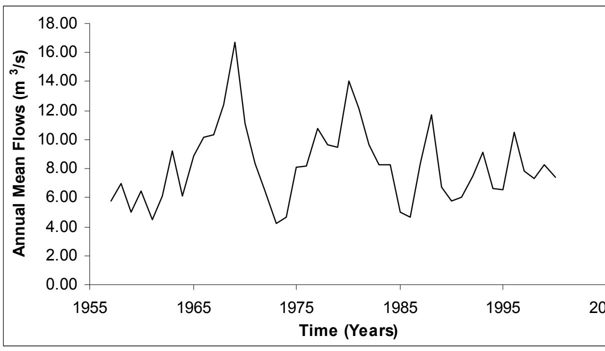

Figure 4
Observed annual mean flows of Hurman Creek Tanır/ Gözlerüstü SGS for the period 1957 to 2000

TABLE 1 **Statistical properties of annual mean flows in the period of 1860-1956 for Saint Lawrence and Tanır/Gözlerüstü SGS'**			
Statistical Properties		**St. Lawrence**	**Tanır/ Gözlerüstü**
Mean	(m^3/s)	570.62	8.21
Standard Deviation	(m^3/s)	48.71	2.67
Coefficient of Variation		0.0854	0.3255
Maximum Value	(m^3/s)	672.08	16.74
Minimum Value	(m^3/s)	433.33	4.22
Skewness Coefficient	(C_s)	0.085	0.917
Kurtosis	(k)	2.5454	1.2218

TABLE 2
Linear autoregressive (AR) and autoregressive – moving average (ARMA) model parameters for Saint Lawrence SGS annual mean flows (Baran and Bacanlı, 2004)

	AR(1)	AR(2)	AR(3)	ARMA(1,1)	ARMA(1,2)
ϕ_1	0.689	0.651	0.634	0.731	0.956
ϕ_2	-	0.055	-0.090	-	-
ϕ_3	-	-	0.220	-	-
θ_1	-	-	-	0.081	0.362
θ_2	-	-	-	-	0.321

TABLE 3
Linear autoregressive (AR) and autoregressive – moving average (ARMA) model parameters for Tanır/Gözlerüstü SGS annual mean flows

	AR(1)	AR(2)	AR(3)	ARMA(1,1)	ARMA(1,2)
ϕ_1	0.565	0.652	0.655	0.381	0.314
ϕ_2	-	-0.153	-0.166	-	-
ϕ_3	-	-	0.021	-	-
θ_1	-	-	-	-0.278	-0.341
θ_2	-	-	-	-	-0.057

Application and results

Modelling of observed annual mean flows of Saint Lawrence and Tanır/Gözlerüstü SGSs

AR(1), AR(2), AR(3), ARMA(1,1) and ARMA(1,2) models are investigated for the determination of diagnostic checking. The ordinates of ACF: autocorrelation function of observed data up to k = 40 and PACF: partial autocorrelation functions for Saint Lawrence SGS annual mean flows are presented in Figs. 5 and 6 respectively. The ordinates of ACF and PACF for Tanır/Gözlerüstü SGS annual mean flows are presented in Figs. 7 and 8.

It is accepted that, ε_i forms an independent process in time series modelling. To check the suitability of selected model, the residual of model ε_i is obtained. The autocorrelation function (ACF-correlogram) is used for determining whether ε_i process is independent or not. The Anderson test is frequently used to test the hypothesis if the correlogram belongs to an independent process. Mean and variance of the sampling distribution of autocorrelation coefficient (r_k) is given below:

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

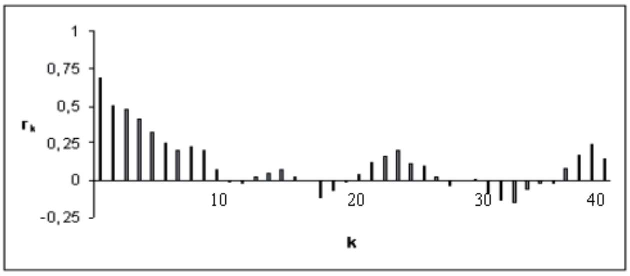

Figure 5
Lag-one autocorrelation coefficients for Saint Lawrence annual mean flows (Baran and Bacanlı, 2004)

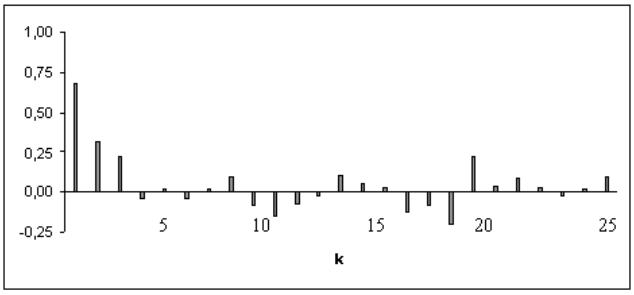

Figure 6
Partial autocorrelation coefficients for Saint Lawrence annual mean flows (Baran and Bacanlı, 2004)

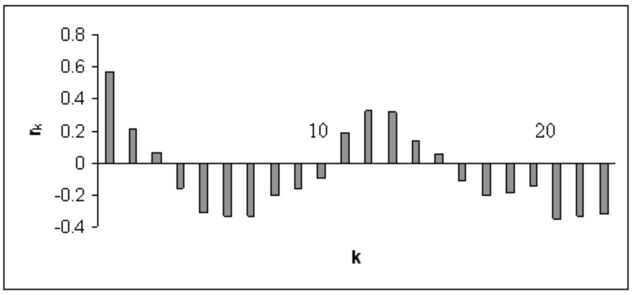

Figure 7
Lag-one autocorrelation coefficients for Tanır/Gözlerüstü annual mean flows

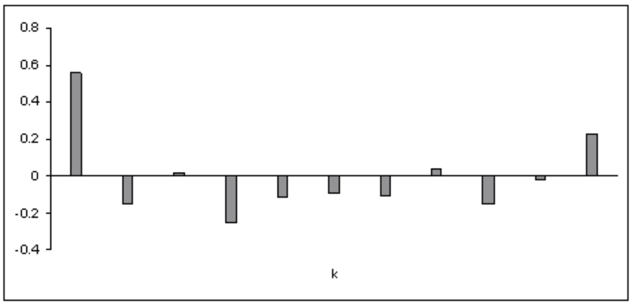

Figure 8
Partial autocorrelation coefficients for Tanır/Gözlerüstü annual mean flows

where:
N = the sample size
k = the lag number

By using these expressions, the confidence limits of the correlogram can be determined. If the percentage of calculated r_k values which are outside of this area is smaller then α, the

Available on website http://www.wrc.org.za
ISSN 0378-4738 = Water SA Vol. 33 No. 4 July 2007
ISSN 1816-7950 = Water SA (on-line)

independence hypothesis is accepted and the selected model is suitable (Salas et al., 1985).

The Box-Pierce Porte Manteau Lack of Fit test is also used to test the independence of the ε_i series. This test is used by statistical formulation of Q (Bayazıt, 1981; Salas et al., 1985):

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

where:

$r_k(\varepsilon)$ = correlogram of residual (ε_i)

L = maximum lag

The statistics Q is approximately χ^i (L - p - q)

In the presented studies of observed and synthetic series, both Anderson and the Box-Pierce Porte Manteau Lack of Fit tests are used to test the independence of ε_i series for all investigated AR(p) and ARMA(p,q) type models. The residuals ε_i are also checked to establish whether they are independent and normal or not for all evaluated models.

Figure 5 shows that the autocorrelation coefficients slowly approach zero after certain k values, and it does not suddenly become zero even for large k values for Saint Lawrence annual mean flows. It means that the autoregressive model terms are preponderate over the moving average terms. Therefore, the selected model must only be formed by autoregressive model terms. Partial autocorrelation coefficients (Fig. 6) converge to zero for k bigger than 3. So, the best order of the AR model is assumed to be 3 (p=3).

The models with the best fit are investigated by diagnostic checking criteria, after modelling observations of annual mean flow series of Saint Lawrence SGS over 97 years (1860 to 1956). The diagnostic checking results are summarised in Table 4. Although the diagnostic checking criteria are pointed out in the ARMA(1,2) model, as it is explained above, only AR(p) type models were taken into consideration for Saint Lawrence SGS. From this point of view, the results confirm that AR (3) can be assumed as the model with the best fit.

Figure 7 shows that the autocorrelation coefficients approach to zero in second k values, for Tanır/Gözlerüstü annual mean flows. It shows that the model formed by autoregressive models terms like Saint Lawrence annual mean flows. Partial autocorrelation coefficients (Fig. 8) converge to zero for k bigger than 2. So, the best order of the AR model can be assumed as 2 (p=2).

The models with the best fit are investigated by diagnostic checking criteria, after modelling observations of annual mean flow series of the Tanır/Gözlerüstü SGS over 44 years (1957 to 2000). The results of diagnostic checking are summarised in Table 5. The results of diagnostic checking criteria confirm that AR(2) can be selected as the model with the best fit.

Synthetic series

To determine the accuracy of the entropy approach as a diagnostic checking criterion in time series modelling, generated synthetic series were also evaluated. Three different types of synthetic series are generated for testing the accuracy of transinformation in diagnostic checking. All generated series are constituted as AR(2) and AR(3), the model parameters are defined as the same calculated AR(3) model parameters of Saint Lawrence SGS (Table 2), and Tanır/Gözlerüstü SGS annual mean flows (Table 3).

The following generated series are investigated: AR(1), AR(2), AR(3), ARMA(1,1) and ARMA(1,2) models reflected the observed flows. All generated data sets are remodelled such as observed data. After the determination of the investigated linear autoregressive (AR) and linear autoregressive – moving average (ARMA) type model parameters, Anderson and Box-Pierce Porte Manteau Lack of Fit Tests are applied for each investigated model.

The 1st group of generated data sets were defined as Saint Lawrence SGS annual mean flows [AR(3)] and as Tanır/Gözlerüstü SGS annual mean flows [AR(2)] over time periods of the same length. The results of diagnostic checking for all criteria are presented in Table 6 and 7 (next page).

The 2nd generated data sets were defined as long-term Saint Lawrence and Tanır/Gözlerüstü SGS annual mean flows. Both data sets had 600 members with the same parameters AR (3) of Saint Lawrence and AR (2) of Tanır/Gözlerüstü SGS. The generated data sets were divided into subgroups of 100, 200… 600 years length and re-modelled as observed data sets for determining the variation of performances diagnostic checking criteria vs. data number. The results of diagnostic checking for all subgroups are presented in Tables 8 and 9 (next page).

The last and the 3rd synthetic data sets consisted of data generated by 30 subgroups. In this step, the synthetic data sets were defined as a population of Saint Lawrence –AR(3)- and Tanır/Gözlerüstü –AR(2)- SGS annual mean flows. Both data sets were generated with 3 600 members. The subseries were defined as 97 members for Saint Lawrence and 44 members for Tanır/Gözlerüstü and evaluated as observed annual mean flow series.

TABLE 4
The results diagnostic checking for annual mean flows of St. Lawrence SGS

Diagnostic checking criterion	ACF	VAR(e)	AIC	AICC	FPE	T(X,Y)
Best model (determined from all investigated models)	(*)	ARMA(1,2)	ARMA(1,2)	ARMA(1,2)	-	ARMA(1,2)
Best model (determined only from investigated AR type models)	(*)	AR(3)	AR(1) -AR(3)	AR(3)	AR(2)	AR(3)

(*) All models are assumed suitable for 95% confidence limit

TABLE 5
The results diagnostic checking for annual mean flows of Tanır/Gözlerüstü SGS

Diagnostic checking criterion	ACF	VAR(e)	AIC	AICC	FPE	T(X,Y)
Best model (determined from all investigated models)	(*)	AR(2)-AR(3)	AR(2)	AR(2)-AR(3)	-	AR(2)
Best model (determined only from investigated AR type models)	(*)	AR(2)-AR(3)	AR(2)	AR(2)-AR(3)	AR(2)	AR(2)

(*) All models are assumed suitable for 95% confidence limit

TABLE 6
The results of diagnostic checking for generated AR(3) synthetic data with 97 members

Diagnostic checking criterion	ACF	VAR(e)	AIC	AICC	FPE	T(X,Y)
Best model (determined from all investigated models)	(*)	ARMA(1,2)	ARMA(1,2)	ARMA(1,2)	-	AR(3)
Best model (determined only from investigated AR type models)	(*)	AR(3)	AR(3)	AR(1)	AR(3)	AR(3)

(*) All models are suitable for %95 confidence limit

TABLE 7
The results of diagnostic checking for generated AR(2) synthetic data with 44 members

Diagnostic checking criterion	ACF	VAR(e)	AIC	AICC	FPE	T(X,Y)
Best model (determined from all investigated models)	(*)	ARMA(1,2)	ARMA(1,2)	ARMA(1,2)	-	AR(2)
Best model (determined only from investigated AR type models)	(*)	AR(2)	AR(2)	AR(2)	AR(1)	AR(2)

(*) All models are suitable for %95 confidence limit

TABLE 8
The results of diagnostic checking for generated AR(3) synthetic data set and subgroups

N	Best fitted models determined from all investigated AR(p) and ARMA(p,q) models					Best fitted models determined from investigated AR(p) models					
	VAR(e)	AIC	ACF	AICC	T(X,Y)	VAR(e)	AIC	ACF	AICC	FPE	T(X,Y)
100	ARMA(1,2)	ARMA(1,2)	(*)	ARMA(1,2)	AR(3)	AR(1)	AR(1)	(*)	AR(1)	AR(2)	AR(3)
200	ARMA(1,2)	AR(1)	(*)	ARMA(1,2)	AR(2)	AR(2)	AR(1)	(*)	AR(1) or AR(2)	AR(3)	AR(2)
300	ARMA(1,2)	ARMA(1,2)	(*)	ARMA(1,2)	AR(2)	AR(2)	AR(1)	(*)	AR(2)	AR(3)	AR(2)
400	ARMA(1,2)	AR(3)	(*)	ARMA(1,2)	AR(2)	AR(2)	AR(3)	(*)	AR(2)	AR(3)	AR(2)
500	ARMA(1,2)	AR(3)	(*)	ARMA(1,2)	AR(2)	AR(2)	AR(3)	(*)	AR(2)	AR(3)	AR(2)
600	ARMA(1,2)	AR(3)	(*)	ARMA(1,2)	AR(3)	AR(2)	AR(3)	(*)	AR(2)	AR(3)	AR(3)

(*) All models are suitable for 95% confidence limit

TABLE 9
The results of diagnostic checking for generated AR(2) synthetic data set and subgroups

N	Best fitted models determined from all investigated AR(p) and ARMA(p,q) models					Best fitted models determined from investigated AR(p) models					
	VAR(e)	AIC	ACF	AICC	T(X,Y)	VAR(e)	AIC	ACF	AICC	FPE	T(X,Y)
44	ARMA(1,2)	ARMA(1,2)	(*)	ARMA(1,2)	AR(2)	AR(1)	AR(1)	(*)	AR(1)	AR(1)	AR(2)
100	AR(2) – AR(3)	AR(2)	(*)	AR(2)	AR(2)	AR(2)-AR(3)	AR(2)	(*)	AR(2)	AR(2)	AR(2)
200	AR(2)	AR(2)	(*)	AR(2)	AR(2)	AR(2)	AR(2)	(*)	AR(2)	AR(3)	AR(2)
300	AR(2)	AR(2)	(*)	AR(2)	AR(2)	AR(2)	AR(2)	(*)	AR(2)	AR(3)	AR(2)
400	AR(2)	AR(2)	(*)	AR(2)	AR(2)	AR(2)	AR(2)	(*)	AR(2)	AR(3)	AR(2)
500	AR(2)	AR(2)	(*)	AR(2)	AR(2)	AR(2)	AR(2)	(*)	AR(2)	AR(3)	AR(2)
600	AR(2)	AR(2)	(*)	AR(2)	AR(2)	AR(2)	AR(2)	(*)	AR(2)	AR(3)	AR(2)

(*) All models are suitable for 95% confidence limit

At this stage, 30 different data sets were taken into consideration which were known to have come from the same population for AR(3) and AR(2) generated data sets. The performance of the diagnostic checking criteria can be evaluated by this approach. The results are presented in Tables 10 and 11.

Conclusions

The exact model parameters are never known during the determination stage of the mathematical models of a hydrological time series. Therefore, they must be estimated from limited observed data. The inferred population model is only an approximation. Consequently, the most important issue in stochastic hydrology is to define the model with the best fit amongst various models.

The criteria used for the diagnostic checking of the modelling of the time series in which long-term observation sets are seldom available, such as annual mean flows, should be considered with all their properties including their weaknesses.

The criteria used to determine the most suitable model (diagnostic checking) may give very different results depending on many properties such as the structure, the statistical properties of the investigated hydrological process and/or the

TABLE 10 The results of sampling distribution of diagnostic checking criterion for 97 years length AR(3) synthetic subseries (m: absolute, f: relative frequency)				
	Investigated AR(p) and ARMA(p,q) models		Investigated AR(p) models	
VAR(e)	**m**	**f**	**m**	**f**
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	**m**	**f**	**m**	**f**
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	---	---
ARMA(1,2)	11	0.37	---	---
AICC	**m**	**f**	**m**	**f**
AR(1)	1	0.03	14	0.47
AR(2)	0	0.00	2	0.06
AR(3)	10	0.33	14	0.47
ARMA(1,1)	5	0.17	---	---
ARMA(1,2)	14	0.47	---	---
FPE	**m**	**f**	**m**	**f**
AR(1)	---	---	8	0.26
AR(2)	---	---	11	0.37
AR(3)	---	---	11	0.37
ARMA(1,1)	---	---	---	---
ARMA(1,2)	---	---	---	---
T(X,Y)	**m**	**f**	**m**	**f**
AR(1)	0	0	0	0
AR(2)	0	0	6	0.20
AR(3)	24	0.80	24	0.80
ARMA(1,1)	6	0.20	---	---
ARMA(1,2)	0	0	---	---

TABLE 11 The results of sampling distribution of diagnostic checking criterion for 44 years length AR(2) synthetic subseries (m: absolute, f: relative frequency)				
	Investigated AR(p) and ARMA(p,q) models		Investigated AR(p) models	
VAR(e)	**m**	**f**	**m**	**f**
AR(1)	0	0	2	0.07
AR(2)	10	0.33	12	0.40
AR(3)	12	0.40	16	0.53
ARMA(1,1)	1	0.03	---	---
ARMA(1,2)	7	0.24	---	---
AIC	**m**	**f**	**m**	**f**
AR(1)	1	0.03	6	0.20
AR(2)	12	0.40	15	0.50
AR(3)	9	0.30	9	0.30
ARMA(1,1)	1	0.03	---	---
ARMA(1,2)	7	0.24	---	---
AICC	**m**	**f**	**m**	**f**
AR(1)	0	0.00	2	0.07
AR(2)	10	0.33	11	0.37
AR(3)	12	0.40	17	0.56
ARMA(1,1)	1	0.03	---	---
ARMA(1,2)	7	0.24	---	---
FPE	**m**	**f**	**m**	**f**
AR(1)	---	---	7	0.235
AR(2)	---	---	16	0.53
AR(3)	---	---	7	0.235
ARMA(1,1)	---	---	---	---
ARMA(1,2)	---	---	---	---
T(X,Y)	**m**	**f**	**m**	**f**
AR(1)	7	0.235	8	0.27
AR(2)	17	0.57	20	0.67
AR(3)	2	0.065	2	0.06
ARMA(1,1)	2	0.065	---	---
ARMA(1,2)	2	0.065	---	---

data length. Different criteria such as autocorrelation function [ACF], minimum residual variance [min Var(e)], Akaike Information Criterion [AIC], Modified Akaike Information Criterion [AICC], final prediction error criterion [FPE] and transinformation criterion [T(X,Y)] were evaluated and compared in order to determine their advantages and disadvantages according to one another. The results obtained are given below:

- When the correlograms (ACF an PACF) for the annual mean flows of the Saint Lawrence and Tanır/Gözlerüstü SGSs were evaluated, ACF showed that autoregressive model terms form the model for both SGSs. PACF showed that the best order model can be accepted as 3 for Saint Lawrence and 2 for Tanır/Gözlerüstü. The diagnostic checking criteria also confirmed that the model with the best fit is AR (3) for Saint Lawrence and AR (2) for Tanır/Gözlerüstü.
- Although the correlogram (ACF-PACF) is not used for diagnostic checking, they are essential in determining the suitable model type for hydrological phenomena. If necessary, they could be used as a criterion for diagnostic checking by reducing the confidence limits.
- In the short-term evaluations only the transinformation criterion seems to be reliable. The results of the first group of generated series for AR (3) and AR (2) (Tables 6 and 7) show that the transinformation criterion is the best. The transinformation criterion is also the one which gives the generated population.
- The investigations on the long-term synthetic series show that increase in the number of data affects the performances of diagnostic check criteria except the correlogram (ACF).
- The results of long-term synthetic series (Table 8 and 9) show that the transinformation criterion performance is better than others, in general. The results of AR (2) generated data sets show that the best performance is obtained from transinformation criterion for all data lengths (Table 9). On the other hand, the results of AR (3) generated data sets show (Table 8) that the performance of the AIC criterion is better than the transinformation for 400 and over number of data.
- Although it is applied to limited data belonging to a 97-year and 44-year period, to avoid speculation, two sets of data for 3 600 years were generated and 30 subgroups of 97- [AR (3)] and 44- [AR (2)] year periods were made up and investigated. The results proved that the transinformation criterion was still reliable with 57% to 80% success for

all investigated AR(p)-ARMA(p,q) models and 67% to 80% for autoregressive AR(p) models (Tables 10 and 11).

- The performance of the transinformation criterion [T(X,Y)] is significantly better than other diagnostic checking criteria. The results of relative frequencies have shown that transinformation is quite a reliable criterion in testing of goodness of fit especially for AR type models. The results show that transinformation can effectively and precisely be used as a criterion for diagnostic checking in time series analysis.

References

AKAIKE H (1969) Fitting autoregressive model for prediction. *Ann. Inst. Stat. Math.* **21** 243-247.

AKAIKE H (1974) A new look at the statistical model identification. IEEE Trans. Autom. Control **19** 716-723.

AKAIKE H (1985) Prediction and entropy. In: Atkinson AC and Fienberg SE (eds.) *A Celebration of Statistics*. Springer Verlag, Heidelberg, Germany.

AKAIKE H (2003) On the art of modeling: Illustrated with the analysis of the golf swing motion. Genome Inf. **14** 263-265.

AMOROCHO J and ESPILDORA B (1973) Entropy in the assessment of uncertainty of hydrologic systems and models. *Water Resour. Res.* **9** (6) 1511-1522.

BARAN T and BACANLI ÜG (2004) Evaluation of suitability criteria in stochastic modeling. *Proc. EWRA Symposium on Water Resources Management Risks and Challenges for the 21st Century.* **VII** Theme XI Modeling as a Tool in Decision Making, 2-4 September, Izmir, Turkey. 937-948.

BARAN T and HARMANCIOĞLU NB (1993) Assessment of mathematical models with exponential functions describing karstic spring discharges. UKAM, IAHS & IAH, International Symposium and Field Seminar on Hydrogeologic Processes in Karst Terrains, Session X on Modeling. September 1990, Antalya, Turkey. In: PE LaMoreaux, AI Assaad and A McCorley (eds.) *Hydrogeological Processes in Karst Terrains.* IAHS **207** 231-241.

BARAN T, HARMANCIOĞLU NB and ÖZİŞ Ü (1987) Türkiye'nin akarsu havzalarında karst pınar katkıları (Karst spring effluents in Turkey). Ankara, *İnşaat Mühendisleri Odası, Türkiye İnşaat Mühendisliği IX. Teknik Kongresi,* **II** 299-311.

BARAN T, HARMANCIOĞLU NB and ÖZİŞ Ü (1995) Average base flow rates of karst spring effluents in Turkey. Proc. UKAM, IAHS & IAH, International Symposium and Field Seminar on Karst Waters and Environmental Impacts. September 10 - 20, Abstracts. Antalya, Turkey, 105.

BAYAZIT M (1981) *Statistical Methods in Hydrology.* İTÜ, Istanbul. 223 pp.

BOX GEP and JENKINS GM (1976) *Time Series Analysis, Forecasting and Control*. Holden Day, San Francisco. 575 pp.

BROCKWELL PJ and DAVIS RA (2002) *Introduction to Time Series and Forecasting.* Springer, New York. 435 pp + CD.

CHAPMAN TG (1986) Entropy as a measure of hydrologic data uncertainty. *J. Hydrol.* **85** 111-126.

EIE (2000) Su akım aylık ortalama (Monthly mean flows) 1935-1995. General Directorate of Electrical Power Resources Survey and Development Administration, Ankara. 742 pp.

EIE (2003) Su akımları yıllığı (Yearly Book of daily mean flows)1996-2000. General Directorate of Electrical Power Resources Survey and Development Administration, Ankara.

FAN J and YAO Q (2003) *Nonlinear Time Series Nonparametric and Parametric Methods.* Springer-Verlag, New York. 551 pp.

HURVICH CM and TSAI CL (1989) Regression and time series model selection in small samples. *Biometrika* **76** 297-307.

HARMANCIOĞLU NB (1981) Measuring the information content of hydrological processes by the entropy concept. *Centennial of Ataturk's Birth, J. Civil Engineering Special edn*., Faculty of Ege University, Izmir, Turkey. 13-88.

HARMANCIOĞLU N, SINGH VP and ALPASLAN N (1992) Versatile uses of the entropy concept in water resources. In: Singh VP and Fiorentino M (eds.) *Entropy and Energy Dissipation in Water Resources*. Kluwer, Netherlands. 91-117.

KARMESHU J and PAL NR (2003) Uncertainty, entropy and maximum entropy principal – An overwiew. In: Karmeshu J (ed.) *Entropy Measures, Maximum Entropy Principle and Emerging Applications.* Springer-Verlag, Berlin, Germany. 1-53.

MOGHEIR Y, DE LIMA JLMP and SINGH VP (2004) Characterizing the spatial variability of groundwater quality using the entropy theory: I. Synthetic data. *Hydrol. Proc.* **18** 2165-2179.

SALAS JD and SMITH RA (1981) Physical basis of stochastic models of annual flows. *Water Resour. Res.* **17** (2) 428-430.

SALAS JD, OBEYSEKERA JTB and SMITH RA (1981) Identification of streamflow stochastic models. *ASCE J. Hydraul. Div.* **107** HY7: 853-868.

SALAS JD, DELLEUR JW, YEVJEVICH V and LANE WL (1985) *Applied Modeling of Hydrologic Time Series*. Water Resources Publications (First Printing 1980), Littleton. 484 pp.

SCHRADER R (2000) On a quantum version of Shannon's conditional entropy. *Fortschr. Physics* **48** 747-762.

SHANNON CE (1948) A mathematical theory of information. *Bell System Tech. J.* **27** 379-423 and 623-656 and reprinted in: Shannon CE and Weawer W (1964) *The Mathematical Theory of Information*. The University of Illinois Press, Urbana, Illinois.

SINGH VP (1997) The use of entropy in hydrology and water resources. *Hydrol. Proc.* **11** 587-626.

SINGH VP (2000) The entropy theory as a tool for modelling and decision-making in environmental and water resources. *Water SA* **26** (1) 1-12.

SINGH VP (2003) The entropy theory as a decision making tool in environmental and water resources. In: Karmeshu J (ed.) *Entropy Measures, Maximum Entropy Principle and Emerging Applications.* Springer-Verlag, Berlin, Germany. 261-297.

USLU O and TANRIÖVER A (1979) Hidrolojik süreçlerin bilgi içeriğinin ölçülmesi. Measuring the information content of hydrological processes. *Proc. 1st Natl. Congress on Hydrology*, November 12-14, Istanbul, Turkey. 437-443.

VAN DER LUBBE JCA (1997) *Information Theory.* Cambridge University Press, Cambridge. 362 pp.

WEHRL A (1978) General properties of entropy. *Rev. Mod. Phys.* **50** 221-260.

WIKIPEDIA (2005) http://en.wikipedia.org/wiki/Image:Grlakes_lawrence_map.png (Accessed in Feb. 2006).

YEVJEVICH V (1972) *Probability and Statistics in Hydrology.* Water Resources Publications, Fort Collins, Colorado. 302 pp.