MODELING THE STOCHASTIC BEHAVIOR OF THE FARS RIVERS

N. Samani[1], E. Raeissi[2], and A. R. Soltani[2]

[1] *Department of Geology, Shiraz University, Shiraz, Islamic Republic of Iran*
[2] *Department of Statistics, Shiraz University, Shiraz, Islamic Republic of Iran*

Abstract

Historical records for rivers in Fars Province are inadequate in comparison with the design period of hydraulic structures. In this study, time series techniques are applied to the records of three Iranian rivers in the Fars Province in order to generate forecast values of the mean monthly river flows. The autoregressive models (AR), moving average models (MA) and autoregressive moving average models (ARMA) are fitted to the stationary series and the optimum model for each river is formulated. Data generation is done and the synthetic sequences are tested individually against the corresponding historical data, and the optimum length of synthetic data is specified. Statistical tests including means, standard deviations, spectral diagrams and Hurst's coefficient are also provided.

Introduction

Design of hydraulic structures and operational analysis for many hydrological purposes are based on long sequences of data. However, the records for rivers in Fars Province do not go back far enough. For a realistic design and decision, it is necessary to utilize not only the historical data but synthetic data that could possibly occur in the future.

The conventional methods of river flow synthesis from recorded river flows are known to be inadequate. A significant improvement has been achieved through data generation by means of mathematical models both conceptual-deterministic and stochastic. Stochastic models are based on time series techniques that preserve the relevant statistical parameters of the past records. An attempt has been made to formulate the stochastic model (as a data-generating tool) for each river in the Fars Province. Utilizing AR generated data for the prediction of reservoir storage design is recommended by Kendall and Dracup [6] and demonstrated by Samani *et al.* [9]. Details of the three rivers investigated are listed in Table 1.

A map of the river basins is to be found in Figures 1a and 1b. Figures 2 and 3 illustrate monthly means and standard deviations of the three rivers. These figures reflect the climatic conditions (semi-arid) prevailing in the river basins. The bulk of the basins yields occur in the fall and winter from storm rainfall. The peak stream flows that are the result of storm rainfall alone are produced in February. The rainless period starts from early to mid spring, and stream

Keywords: Stochastic; Model; Hydrologic time series

Table 1. Rivers and selected stations for the analysis

River	Station	Catchment area (km^2)	Length of record (years)	Data of record
Shahpour	Sad-Abad	4070	25	1966-90
Mond	Tang-e-Karzin	12360	15	1976-90
	Ghantareh	30650	18	1973-90
Ghara-Aghaj	Band-e-Bahman	2410	15	1976-90

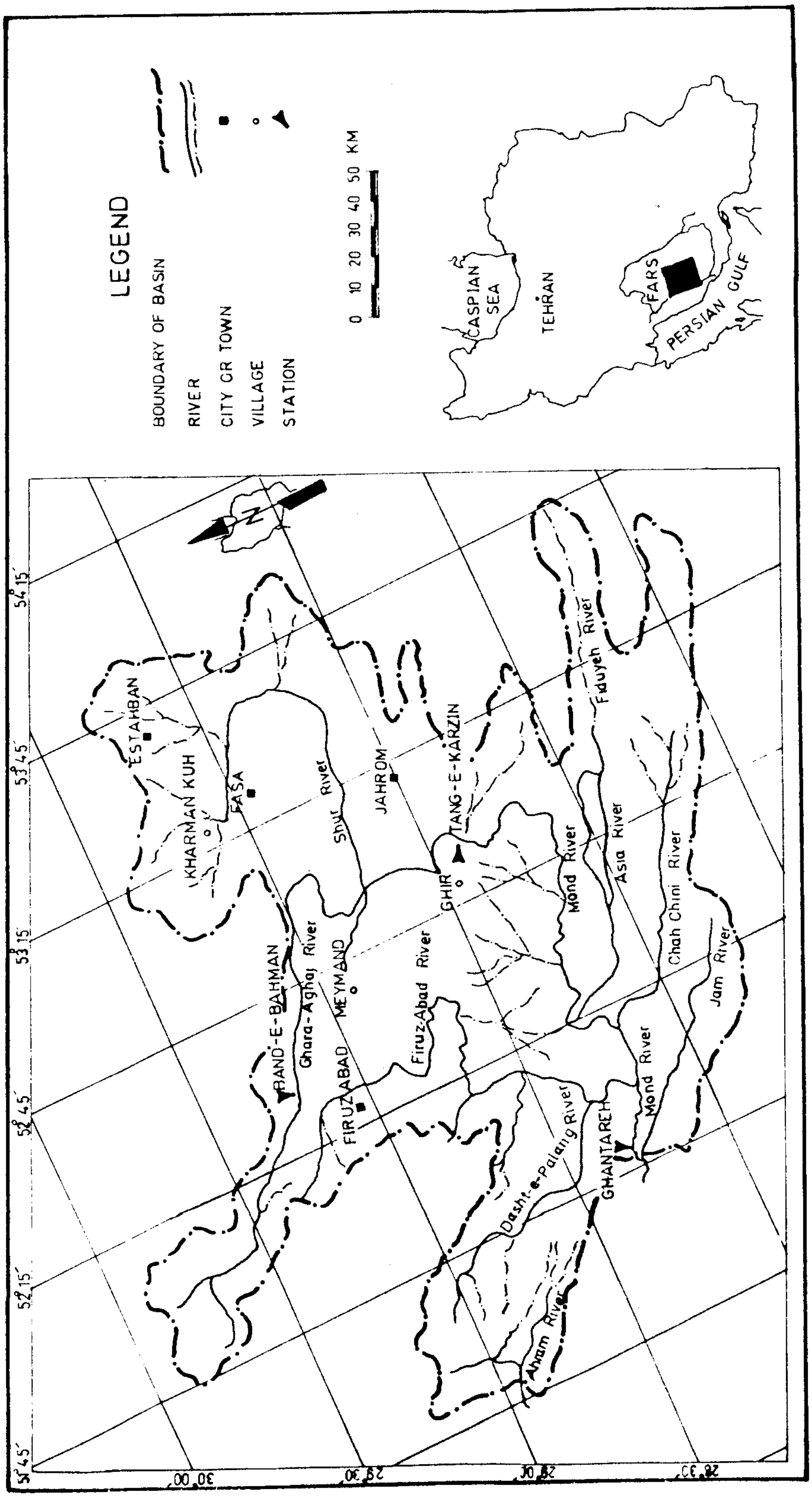

Figure 1a. The map of study area (Ghara-Aghaj and Mond rivers basins)

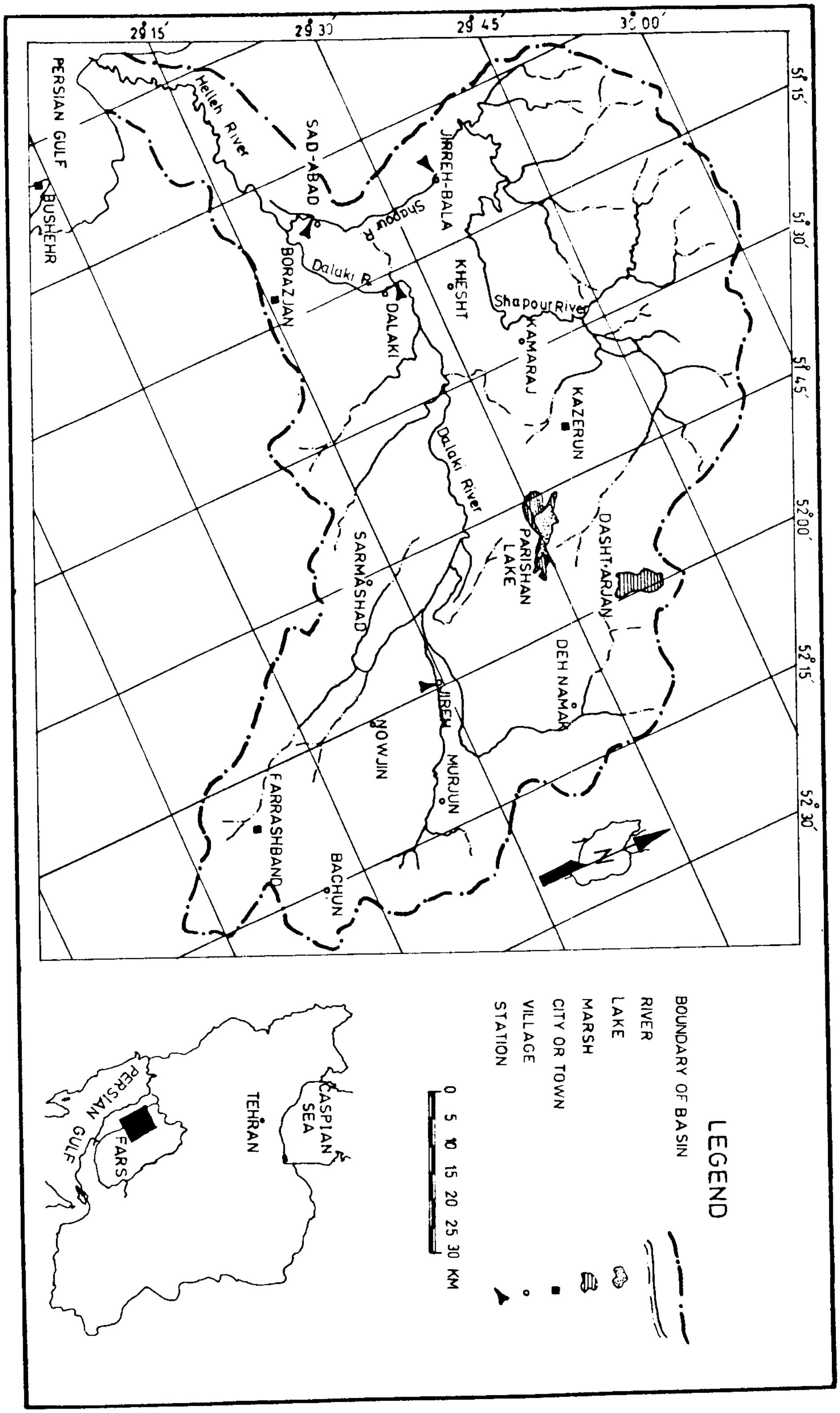

Figure 1b. The map of study area (Shahpour river basin).

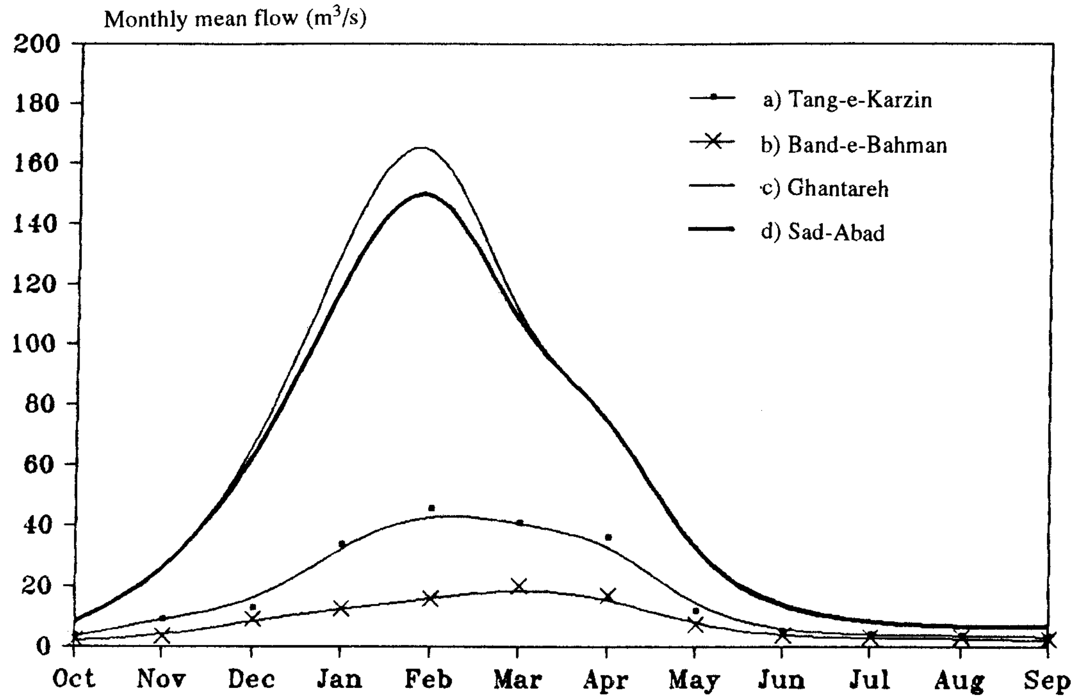

Figure 2. Monthly means

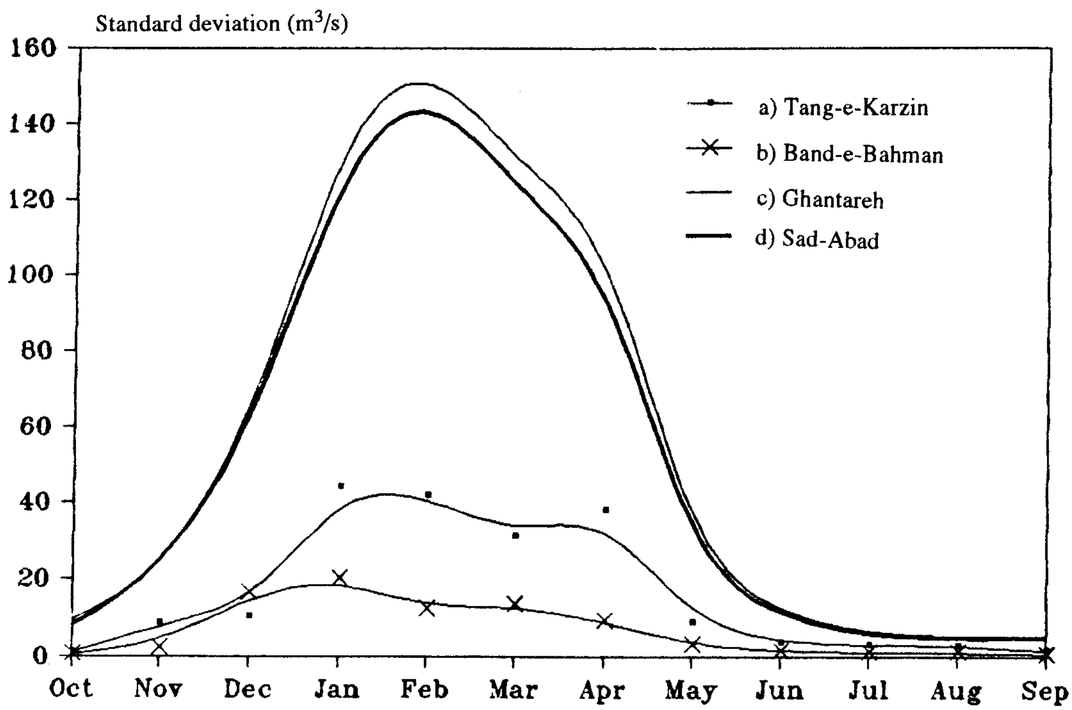

Figure 3. Standard deviations

flows experience a long dry period from May to October. The very low stream flows during this period are produced by the irrigation return flows.

Methodology

The statistical methods for analyzing time series are well established [1, 2, 7, 10, 11]. The traditional methods are mainly concerned with isolating the deterministic (trend and periodic) components and stochastic components. The latter may then be explained in terms of stochastic models such as moving average or autoregressive models. If time series techniques result in a model truly representative, the synthetic data should provide records, some of which are more critical than those observed in the past, and these are combined with the historical records for design purposes.

A brief summary of the methods used to screen and to remove the components of time series is followed in steps:

Trend

The presence of trend or long-term fluctuations can be easily distinguished by inspection of the series of annual mean against the sample mean. Other procedures, such as weighted moving average, have also been used. Application of both methods revealed that the trend components in the records of the three rivers are absent.

Periodicities

In order to have a stationary series, one has to remove periodic components, among which seasonality is the most important. In the following we describe

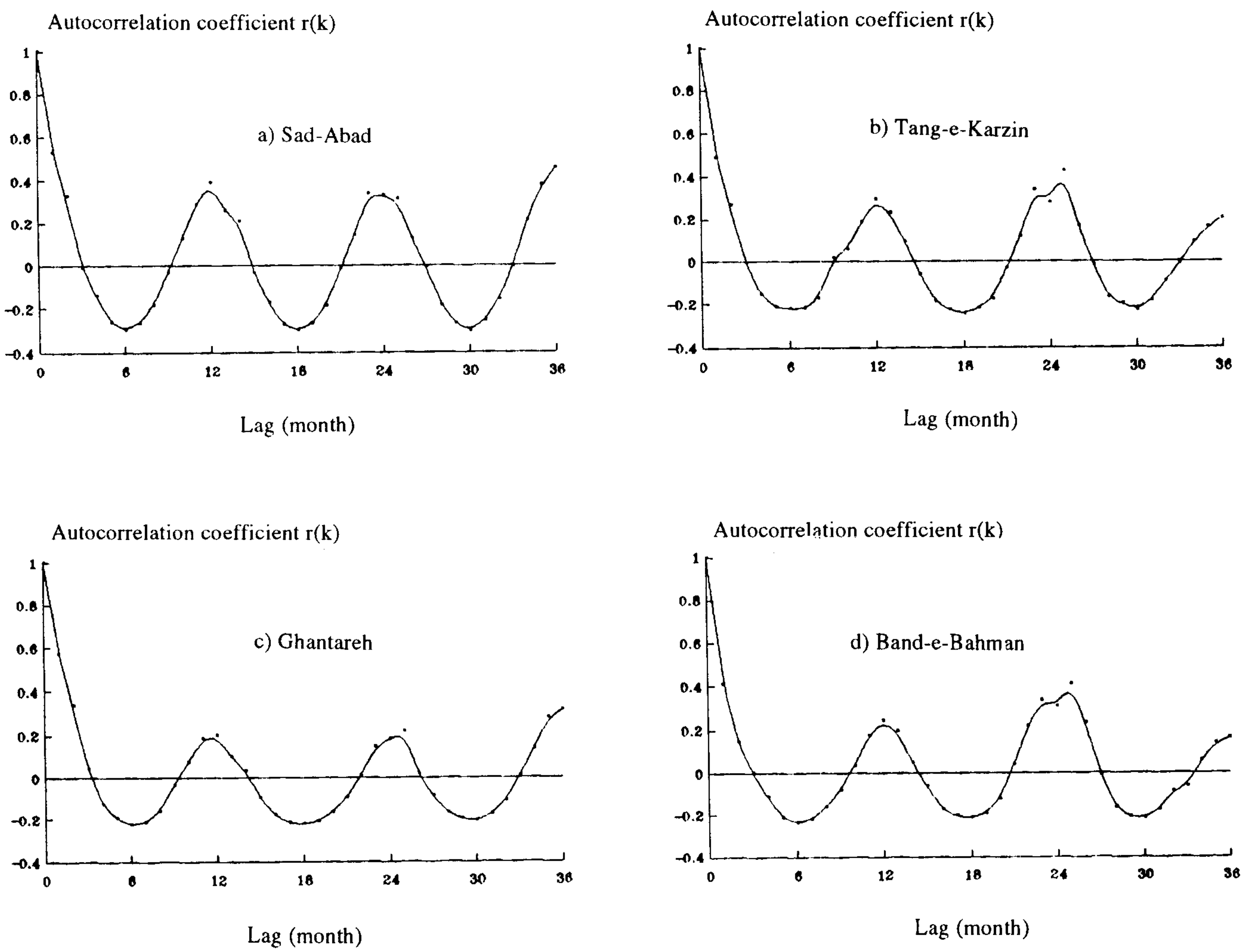

Figure 4. Correlograms of raw data

how to search for seasonal components and describe the method for removing seasonality.

Correlogram

The correlogram is a graphical representation of serial-correlation coefficients r(k) as a function of lag k, where the values of r(k) are plotted as ordinates against their respective values of k as abscissas. Plotted points are joined by straight lines. The serial-correlation coefficients of lag k are calculated by (2 & 7):

$$r(k)=\frac{C(k)}{C(0)} \qquad (1)$$

where C(k) is the autocovariance of a sample of N values of q with an estimated mean q^- at lag k and defined as:

$$C(k)=\frac{1}{N}\sum_{t=1}^{N-k}(q_t-q^-)(q_{t+k}-q^-) \qquad (2)$$

and C(0) is the estimation of the variance of the process

$$C(0)=\frac{1}{N}\sum_{t=1}^{N}(q_t-q^-) \qquad (3)$$

In a practical application, such as computer programming, the numerator of equation 1 can be written as:

$$C(k)=\frac{1}{N-k}\sum_{t=1}^{N-k}q_t q_{t+k}-\frac{1}{(N-k)^2}\left(\sum_{t=1}^{N-k}q_t\right)\left(\sum_{t=1}^{N-k}q_{t+k}\right) \qquad (4)$$

and its denominator as:

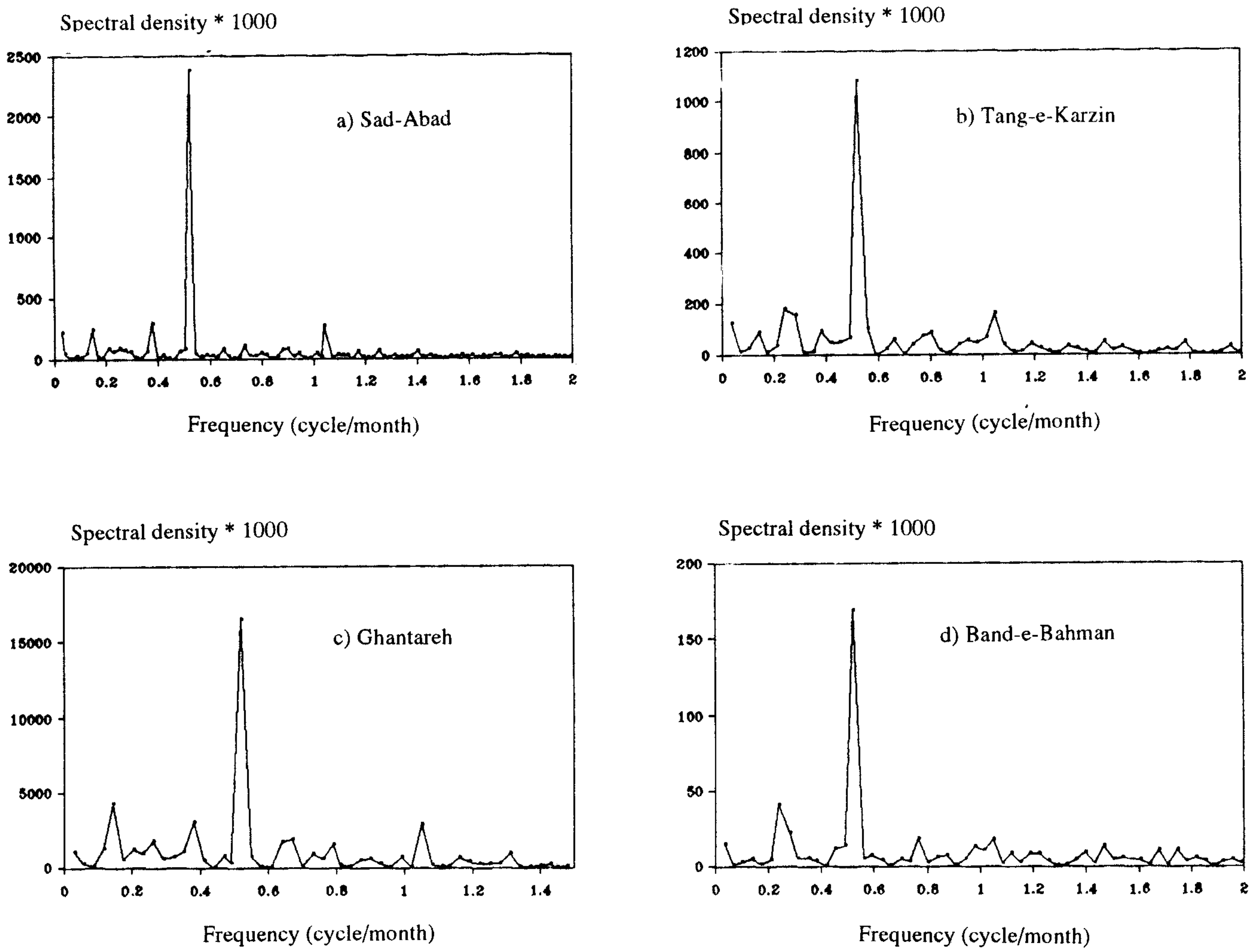

Figure 5. Spectra of raw data

$$C(0)=\left[\frac{1}{N-k}\sum_{t=1}^{N-k} q_t^2-\frac{1}{(N-k)^2}\left(\sum_{t=1}^{N-k} q_t\right)^2\right]^{\frac{1}{2}}\left[\frac{1}{N-k}\sum_{t=1}^{N-k} q_{t+k}^2-\frac{1}{(N-k)^2}\left(\sum_{t=1}^{N-k} q_{t+k}\right)^2\right]^{\frac{1}{2}} \quad (5)$$

Computer programs were written to compute the serial-correlation coefficient for records of the three rivers. The correlograms are shown in Figure 4, all of them are oscillating without any indication of damping, thus revealing the presence of harmonic (periodic) components in all time series.

Power Spectrum

The spectrum is another diagnostic tool for time series analysis in the frequency domain that can help determine the periods of the harmonic components. It is a plot of the spectral density against frequency. In the case of an infinite series, Jenkins and Watt [5] show that the spectral density is:

$$S(f)=\int_{-\infty}^{-\infty} C(i)\cos 2\pi f\, d_i \quad (6)$$

and for a finite series, the estimated spectral density defined (7) as:

$$S(k)=\frac{1}{n}\left[C(0)+2\sum_{i=1}^{n-1} C(i)\,.\,\mathrm{Cos}\frac{\pi k i}{n}+C(n)\,.\mathrm{Cos}\,\pi k\right] \quad (7)$$

Where frequency $f=k/(2n\delta_i)$, n=maximum number of lags investigated, which is taken as less than N/10 (3) and δ_i is the time interval of one month. If the time series contains periodic terms, the frequencies of these

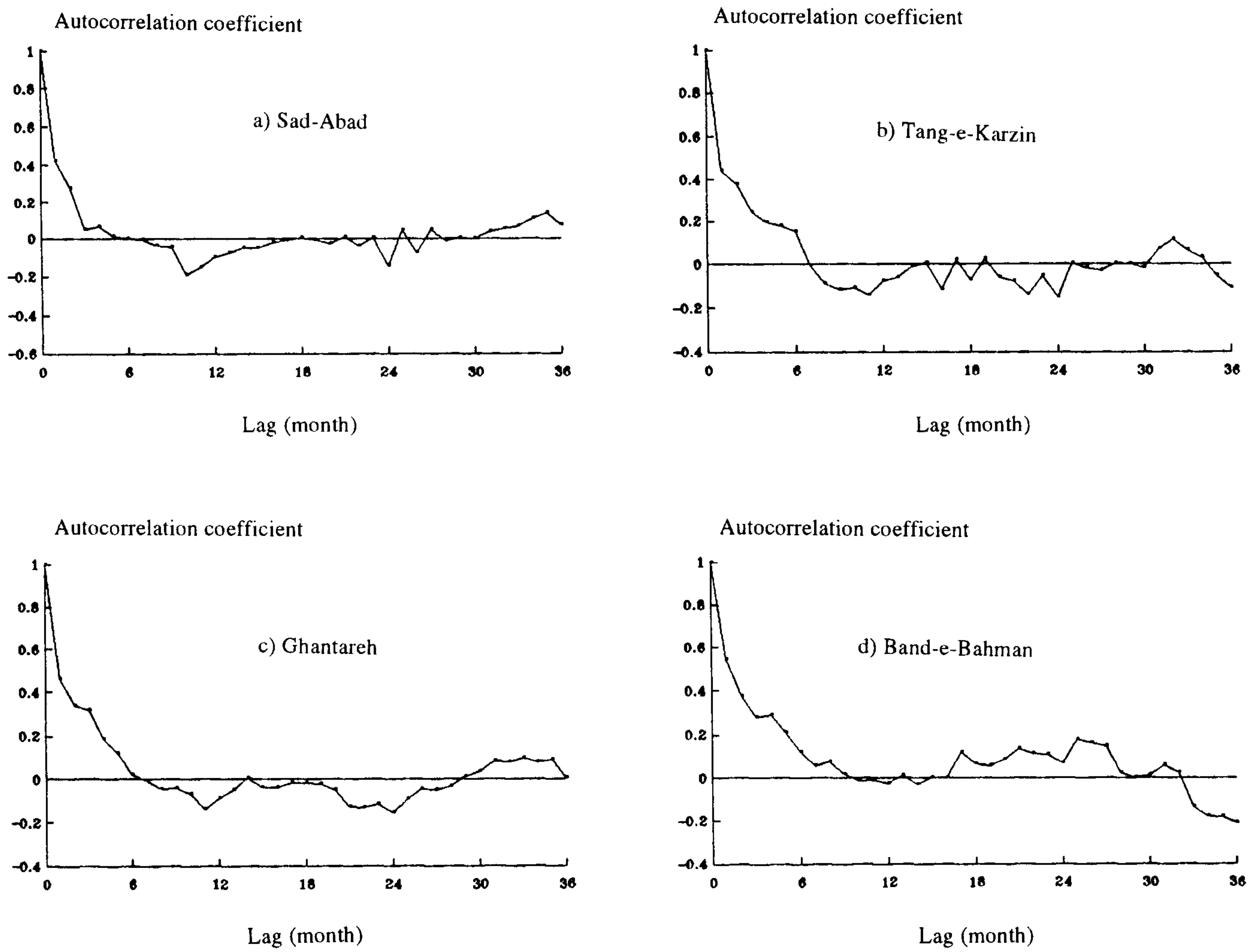

Figure 6. Correlograms of stationary series

terms will appear as high and sharp peaks in the estimated spectrum. The spectra of the three rivers are shown in Figure 5. The spectra of Figure 5 exhibit the highest peaks at 0.52 cycle per month, equivalent to 12 months. From this observation, it may be assumed that periodic components have a 12 month periodicity (annual cycle). This apparently reflects the influence of the regular annual variation implied in the means and standard deviation of stream flows depicted in Figures 2 and 3.

To remove the annual cycle, each individual observation was subtracted from the corresponding observation in the previous water year. Harmonic components revealed by correlograms (Fig. 4) are interpreted as seasonal effects, which are removed by the simple procedure of calculating the 12 monthly averages and subtracting the appropriate one from each individual record and then dividing the result by the standard deviation for the month.

Since the study of data provides no evidence of trend, and annual and seasonal effects have been removed, the remaining components are defined as the residual stochastic or stationary series. Correlograms of stationary series are plotted as shown in Figure 6.

Formulation of Models

Having estimated the serial correlation coefficient of stationary series, we should have some idea as to which stochastic process will provide a suitable model (see for instance 2). Three well-known stochastic models are autoregressive (AR), moving average (MA), and autoregressive moving average (ARMA) models. The theory of stochastic processes is quite well developed. Elements of time series analysis are

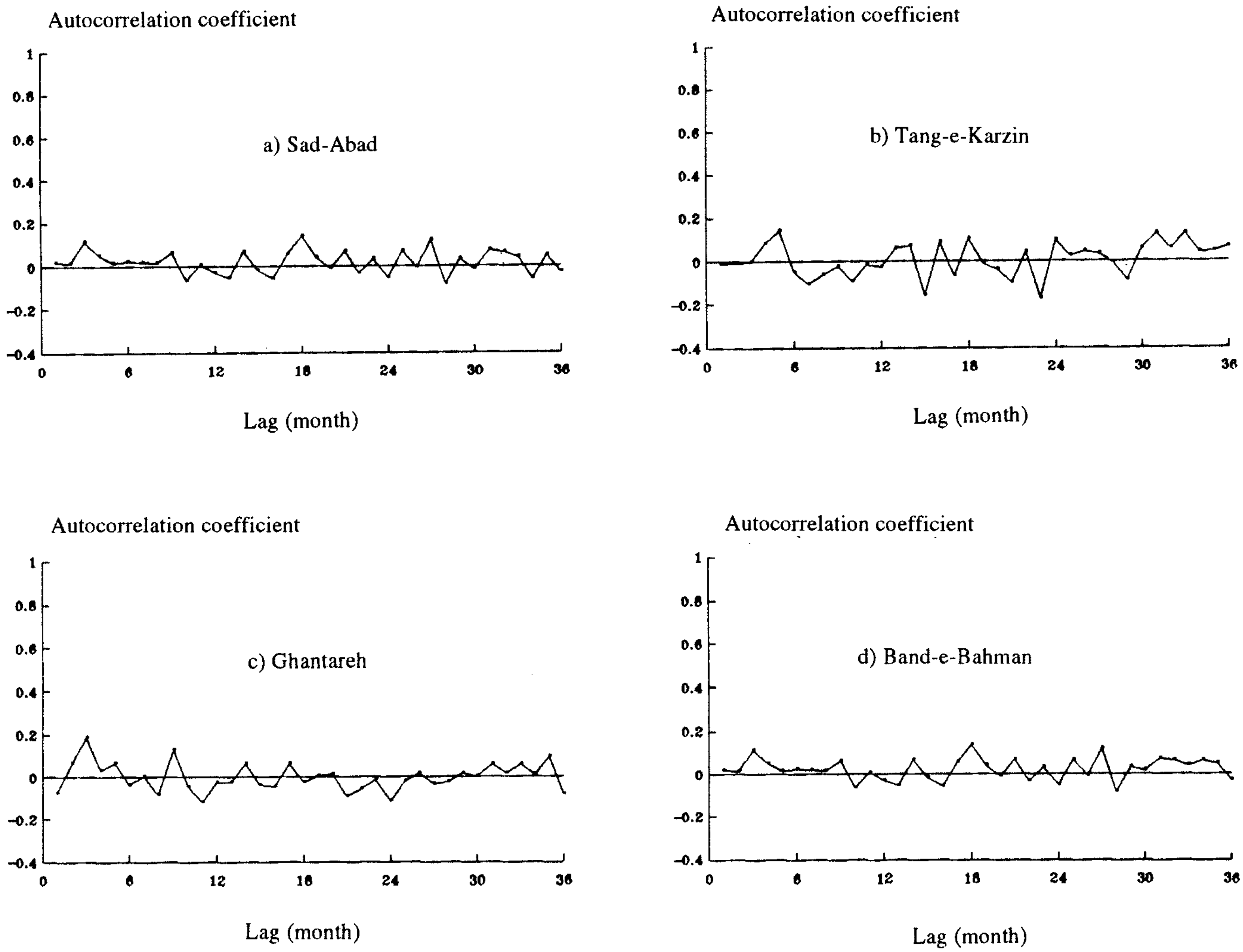

Figure 7. Correlograms of AR model residuals (Z_t)

given by Chatfield [2], and Box and Jenkins [1] discuss the subject extensively. The stationary series of the three rivers are first fitted to AR models of order one and two; these are:

$$q_t = \alpha_1 q_{t-1} + \alpha_2 q_{t-2} + Z_t \qquad (8)$$

where q_t, q_{t-1} are mean monthly flows in months t, t-1 and t-2, respectively. Z_t is the random component, $\alpha_1 = r_1$ and $\alpha_2 = 0$ for an AR1 model and:

$$\alpha_1 = \frac{r_1(1-r_2)}{1-r_1^2}, \quad \alpha_2 = \frac{r_2 - r_1^2}{1-r_1^2}$$

for an AR2 model (r_1 and r_2 are serial correlation coefficients at lag 1 and 2, respectively). The results of fitted model are presented in Table 2, and correlograms of Z_t are shown in Figure 7.

Fitting MA and ARMA (1, 1) to the stationary series has been found to be unsatisfactory. Yevjevich [10] and Kottegoda [7] agree that a moving average model cannot represent a hydrologic time series.

Table 2. Formulated model for each river

River	Station	Fitted model	Model α_1	Parameters α_2
Shahpour	Sad-Abad	AR2	0.37	0.11
Mond	Tang-e-Karzin	AR1	0.44	
	Ghantareh	AR1	0.46	
Ghara-Aghaj	Band-e-Bahman	AR1	0.41	
		AR2	0.49	0.13

Hurst's Coefficient

To justify the suitability of autoregressive models for the rivers, another criterion known to hydrologists as the Hurst's coefficient was calculated. According to Hurst *et al.* [4] and Mandelbort and Wallis [8] the ratio of range R to the standard deviation of flows in a sequence of N values is given by:

$$\text{Log}(R/s) = \text{Log}\,K + H\,\text{Log}N \qquad (9)$$

in which K and H are constants. The constant H should not be significantly different from 0.5 if an AR model based on a short memory is to be adapted. The mean value of Hurst's coefficient was found to be equal to 0.54, 0.51, 0.51 and 0.52 for the three rivers considered. These values are not considered to be significantly different from 0.5.

Generation of Data

Mean monthly flow generation based on AR models is done for each river. The length of generated series remained equal to the length of its respective historical series. The synthetic sequences are tested

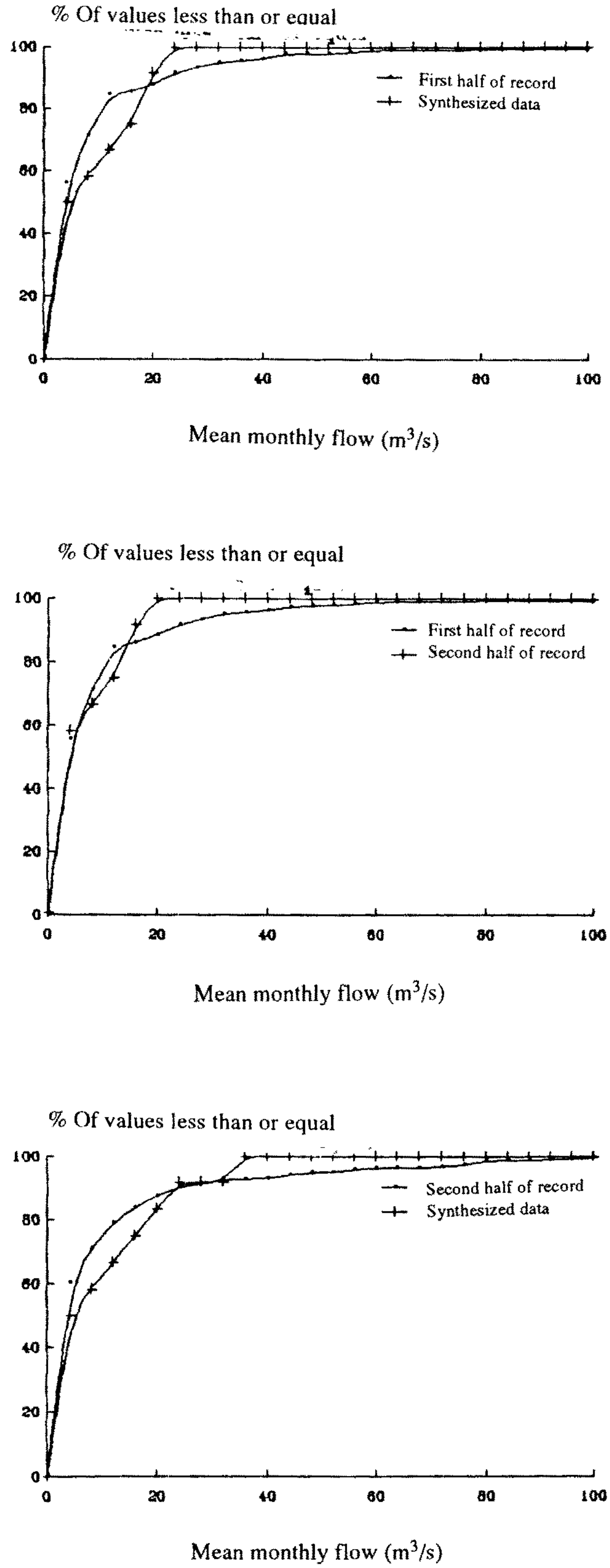

Figure 8. Cumulative distribution functions

against the historical records by plotting the cumulative distribution functions, as an example those of Band-e-Bahman are presented in Figure 8. These figures reveal the close match between the historical and synthesized data. Further statistical tests including the Kolmogorov-Smirnov two sample tests confirm that the residual of the synthesized and historical data come from the same population.

To examine the reliability of synthetic series for design purposes and to find the optimum length of reliable synthetic series, a combined series of historical and synthetic series are used for the generation of new series of different lengths. Based on the above statistical tests, the optimum length of reliable series are found and listed in Table 3 for the three rivers. All attempts to generate longer series failed. Statistical parameters of these series showed large discrepancies when compared with those of historical data.

Table 3. Optimum length of generated data

River	Station	Length of data		
		Historical	Combined	Optimum
Shahpour	Sad-Abad	25	50	70
Mond	Tang-e-Karzin	15	30	40
	Ghantareh	18	36	50
Ghara-Aghaj	Band-e-Bahman	15	30	40

Conclusion

Since long sequences, i.e. longer than 25 years of river flow data, are not available for many rivers in Iran, this type of simulation study is recommended as an integral part of water resources design in the country. In this regard, special care should be given to the optimum length of synthetic data generated by stochastic models. Considering a design life of 75 or more years, a minimum of 25 years of historical data is necessary to formulate a stochastic model with sufficient accuracy. The fact that AR models have yielded satisfactory results does not preclude further investigations into nonlinear models. The models given may require further refinement when applied to simulation studies in areas with different catchment and climatic characteristics.

Acknowledgements

The work presented in this paper is part of a research project sponsored by the Research Council of Shiraz University under the contract no. 67-SC-482-248.

References

1. Box, G. E. P. and Jenkins, G. M. *Time series analysis, forecasting and control.* San Francisco, Holden-Day, (1970).
2. Chatfield, C. *The analysis of time series: theory and practice.* Chapman and Hall, (1982).
3. Chow, V. T. Stochastic modelling of watershed system. In *Advances in hydroscience,* Vol. 11. (1978).
4. Hurst, H. E. *et al. Long term storage: an experimental study.* Constable, London, (1965).
5. Jenkins, G. M. and Watt, D. G. *Spectral analysis and its application.* Holden-Day Inc., San Francisco, (1968).
6. Kendall, D. R. and Dracup, J. A. A comparison of index sequential and AR1 generated hydrologic sequences. *J. Hydrol.,* **122**, 335-352, (1991).
7. Kottegoda, N. T. Statistical methods of river flow synthesis for water resources assessment. *Supplement of Proceedings ICE,* **18**, (1970).
8. Mandelbrot, B. B. and Wallis, J. R. Computer experiment with fractional Gaussian noises. *Wat. Resour. Res.,* **5** (Feb), (1), 228-267, (1969).
9. Samani, N., Raeissi, E. and Soltani, A. R. Stochastic synthesis of droughts for reservoir storage design, *Journal of Engineering, IRI,* (in press), (1992).
10. Yevjevich, V. *Probability and statistics in hydrology.* Water Resources Publ., Fort Collins, Colorado, (1972).
11. Yevjevich, V. *Stochastic processes in hydrology. Ibid.,* (1972).