INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **27**: 819–829 (2007)
Published online 10 October 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.1427

The use of time series modeling for the determination of rainfall climates of Iran

S. Soltani,[a] R. Modarres[a,*] and S. S. Eslamian[b]
[a] *Faculty of Natural resources, Isfahan University of Technology, Iran*
[b] *Department of Water Engineering, College of Agriculture, Isfahan university of Technology, Iran*

Abstract:

In this study, regional climates of Iran were identified based on the properties of the monthly rainfall time series models of 28 main cities of Iran. The autocorrelation (ACF) and partial autocorrelation (PACF) of selected series revealed the seasonal behavior of the monthly rainfall. After the parameters of the models were estimated and the residuals of the models analysed to be time independent and the normality was checked using Portmanteau lack of fit and nonparametric tests, the multiplicative ARIMA model was fitted to monthly rainfall time series of the stations. To determine regional climates, a hierarchical cluster analysis was applied on autocorrelation coefficients at different lags and three main climatic groups were found based on the time series models, namely, simple, moderate and complex climates. The results of the time series modeling showed a high variation of the temporal pattern of the monthly rainfall over Iran except for the margins of the Caspian Sea and the Persian Gulf. The study also shows that the correlation between the seasonal autocorrelation coefficient of the rainfall time series and the rainfall coefficient of variation and elevation of the stations is significant while lag-one autocorrelation coefficient does not correlate to rainfall coefficient of variation and the elevation of the stations. Different models also imply the high variation in the spatial rainfall producing mechanism and different stationarity and periodicity characteristics of the rainfall temporal pattern over Iran. A nomenclature of the abbreviation is given at the end of the paper. Copyright © 2006 Royal Meteorological Society

KEY WORDS rainfall time series; ARIMA model; seasonal autocorrelation; cluster analysis; Iran

Received 7 January 2006; Revised 22 July 2006; Accepted 17 August 2006

INTRODUCTION

Hydrologists have always tried to classify atmospheric and hydrologic events in order to simplify the hydrologic convolutions and the observations or to save the time and the budget. Most of these methods are used for the regionalization of hydrologic phenomena like rainfall, streamflow and other components of water cycle. Multivariate techniques have been underlined as suitable and powerful tools for classifying the meteorological data such as rainfall. Principal components, factor analysis and different cluster techniques have been used to classify daily rainfall patterns and their relationship to the atmospheric circulation (Romero *et al.*, 1999); to classify flood and drought years (Singh, 1999); to classify streamflow drought (Stahl and Demuth, 1999); to classify lake annual fluctuations (Sen *et al.*, 1999); to classify streamflow regimes (Sanz and del Jalon, 2005; Harris *et al.*, 2000); water quality of the lakes (Kitpati *et al.*, 2005) and to classify storm events (Palecki *et al.*, 2005). Acerman (1985) and Acerman and Sinclair (1986) concluded that the cluster analysis has some intrinsic worth to explain the observed variation in data. Gottschalk (1985) applied cluster and principal component analysis (PCA) to the territory of Sweden and concluded that cluster analysis is an appropriate method to use on a national scale with heterogeneous hydrological regimes.

On the other hand, time series modeling is a major tool in planning, operating and decision making of water resources and investigating climatic fluctuations and has been commonly used for data generation, forecasting, estimating missing data and extending hydrologic data records (Delleur *et al.*, 1976; Salas, 1993; Hipel and McLeod, 1994). To accomplish these objectives, hydrologists and meteorologists have to construct stochastic models and the modeler has to decide on choosing the type of the model whether it is univariate or multivariate. These model types are generally based on the annual time series with homogenous mean and variance or on the seasonal series generally with periodic parameters (Salas and Obeysekera, 1982a, among others).

Autoregressive integrated moving average (ARIMA) model is the most widely used time series model in hydrologic and climatic time series modeling. Salas *et al.* (1980) reviewed all these models and described their characteristics. Hipel and McLeod (1994) also presented the ARIMA family among the other models

* Correspondence to: R. Modarres, Faculty of Natural resources, Isfahan University of Technology, Iran. E-mail: r_m5005@yahoo.com

such as broken line, fractional ARMA model (FARMA), fractional Gaussian noise (FGN) and others.

After choosing the model, the modeler should estimate its parameters and then apply diagnostic checking of the selected model. These three modeling steps were completely described by Box and Jenkins (1976) and have been applied, developed and improved by many hydrologists (e.g. Delleur *et al.*, 1976; Hipel *et al.*, 1977; McLeod *et al.*, 1977; Salas and Smith, 1982b; Srinivas *et al.*, 1982; Salas and Fernandez, 1993; Banalya *et al.*, 1998; Elek and Markus, 2004; Kallache *et al.*, 2005; Carslaw, 2005).

In recent years, classification and time series cluster analysis has become an important area of research in several fields, such as economics, marketing, business, and many other fields (e.g. Piccolo, 1990; Maharaj, 1999, 2000; Xiong and Yeung, 2004). However, time series models have not been used for classification meteorological variables yet.

The primary objectives of this study are twofold: first, to find the best time series models for monthly rainfall of the selected stations and second, to classify these stations based on the characteristics of the time series models using cluster analysis. The results of this study can show the temporal behavior of rainfall over Iran.

This paper has the following sections. In the section 'Methodology', the climate data and the methods which are used for time series modeling and cluster analysis are discussed. In the Section 'Results and Discussion', the time series models fitted to the monthly rainfall of the selected stations are presented. Then we describe the results of cluster analysis for the classification of the temporal characteristics of these models. In the last part of this section, the relationships between these climate groups and time series characteristics are discussed. A brief conclusion is then given based on the new look of this study at the use of time series modeling for the rainfall classification.

METHODOLOGY

Climate data

For this study, monthly rainfall time series for 28 major cities of Iran are selected. These stations have the only available long-term rainfall data. The selected series contain 30 year (360 month) rainfall data from 1970 to 2000, except for Ardabil station (rainfall data for 1980–2000), Hamedan (rainfall data for 1977–2000), Yasuj, Ghom and Ilam (rainfall data for 1987–2000). The spatial location of the selected stations is presented in Figure 1. The mean monthly rainfall distribution of the selected stations is also presented in Figure 2. Table I shows the annual rainfall and geographical characteristics of these stations.

ARIMA models

In this section, we will explain the general form of ARIMA models. The Box–Jenkins model has two general forms:, ARIMA (p,d,q) and the multiplicative ARIMA (p,d,q) × (P,D,Q) in which p and q are non-seasonal autoregressive and moving average parameters, P and Q are the seasonal autoregressive and moving average parameters, respectively. The two other parameters, d and D, are required nonseasonal and seasonal differencing respectively, used to make the series stationary. The form of ARIMA (p,d,q) is written as,

$$\phi(B)(1-B)^d Z_t = \theta(B)\varepsilon_t \qquad (1)$$

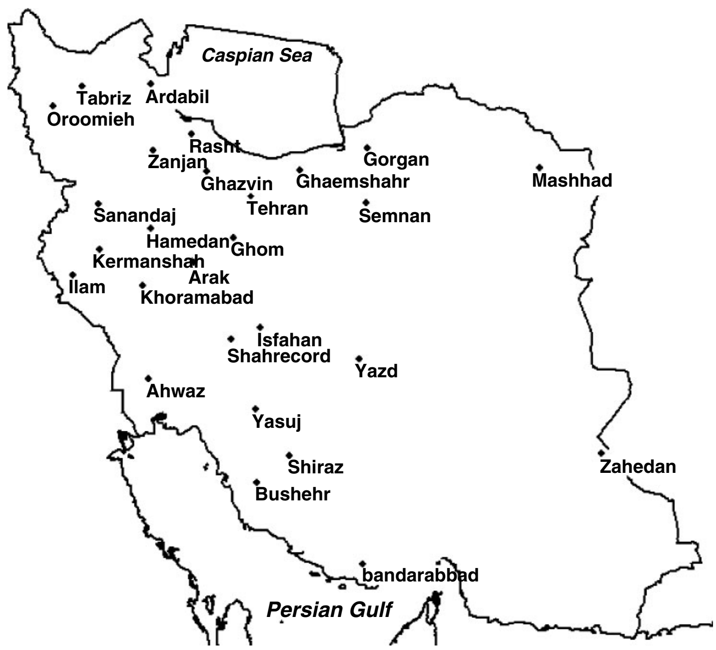

Figure 1. Spatial location of selected rainfall stations.

DOI: 10.1002/joc

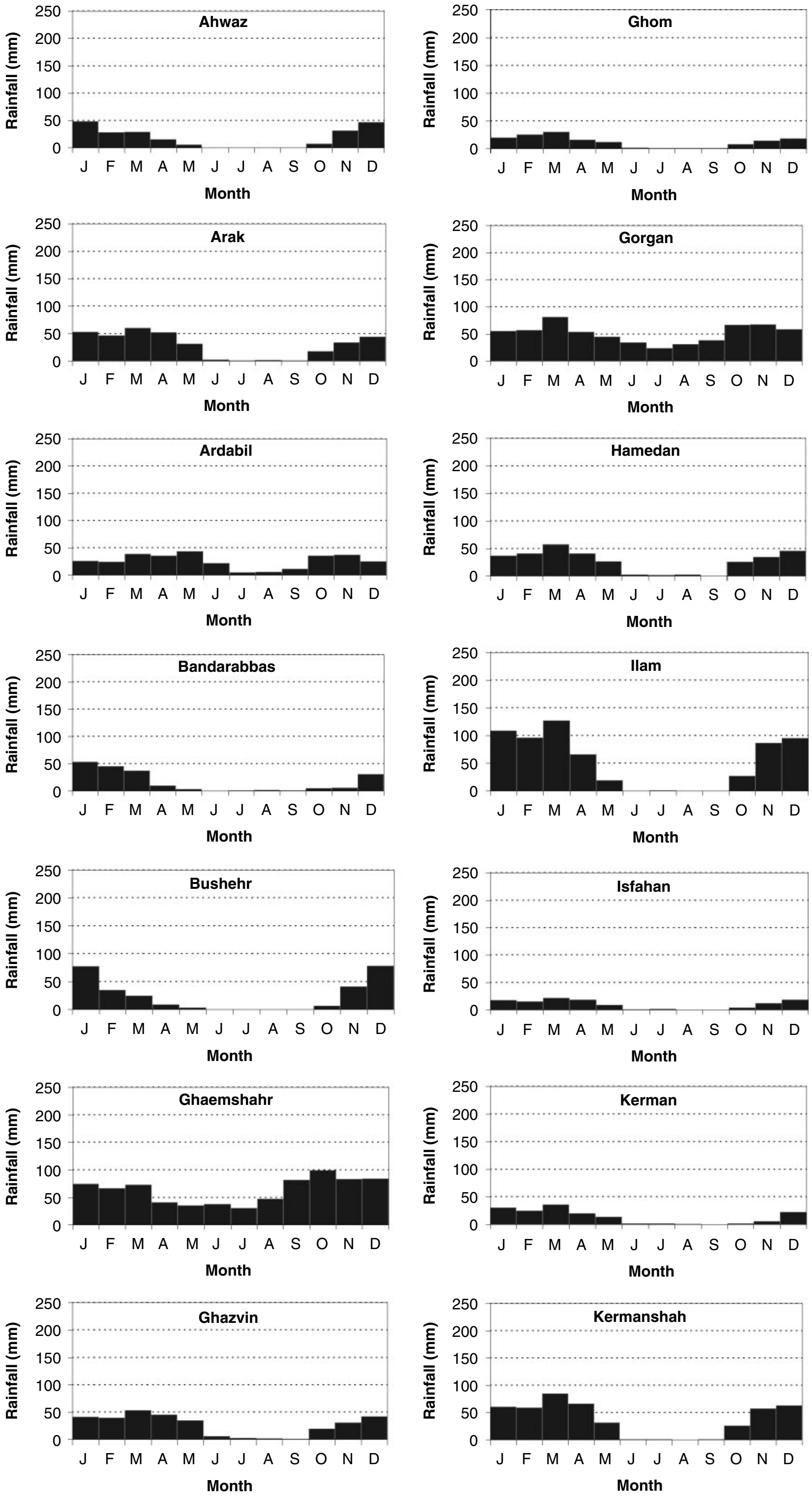

Figure 2. Mean monthly rainfall distribution of the selected stations.

and multiplicative ARIMA(p,d,q) × (P,D,Q) has the following form,

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D Z_t = \theta_q(B)\Theta_Q(B^S)\varepsilon_t \qquad (2)$$

See the nomenclature at the end of the paper for more details. For rainfall time series modeling in this study, we have to find the best model for monthly rainfall time series and to estimate the significant time series parameters for the models.

Time series modeling

Time series modeling includes three steps of model identification, model estimation and diagnostic checking

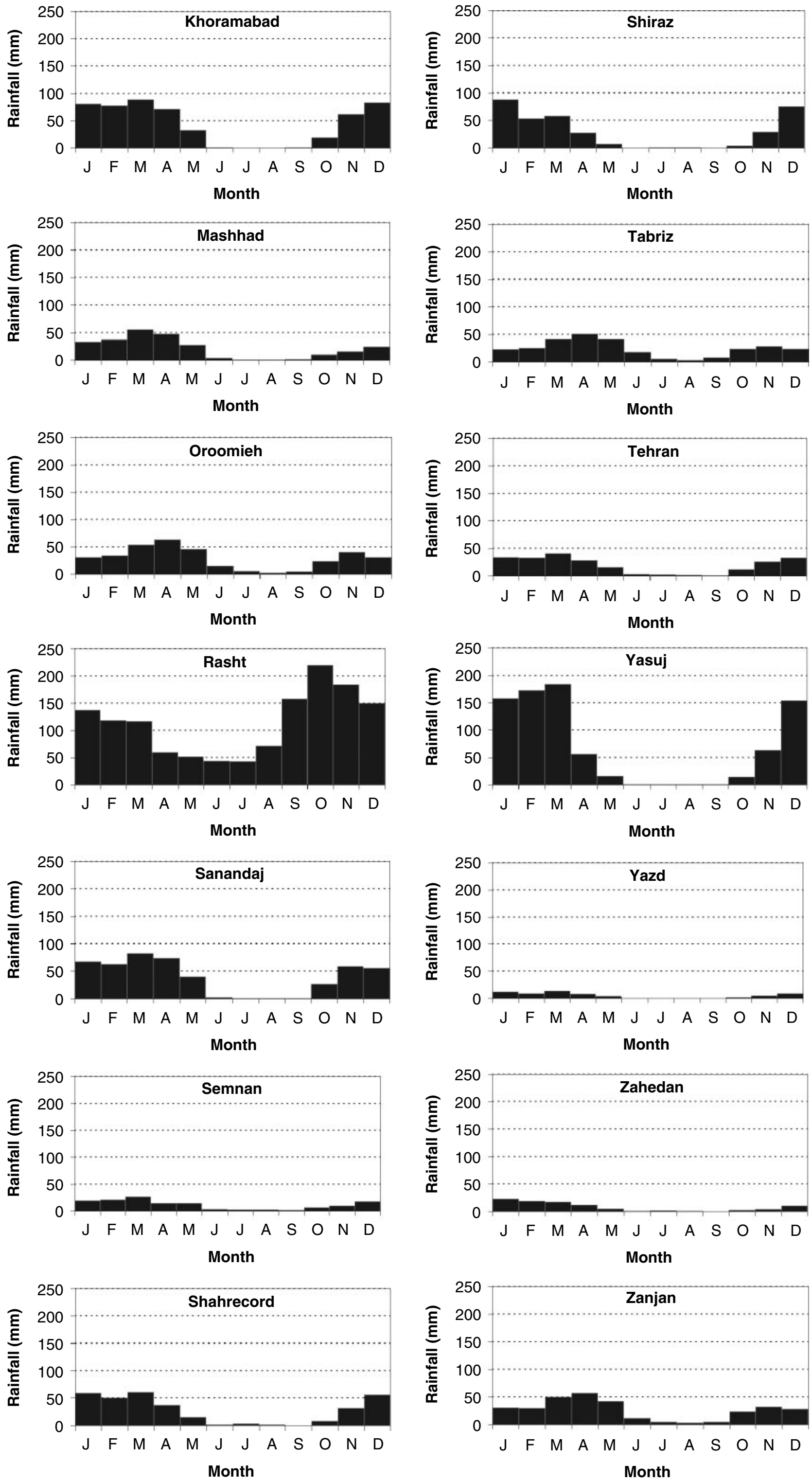

Figure 2. (*Continued*).

(goodness of fit test). In the first step, the initial models which seem to represent the behavior of the time series and are worthy for the further investigation and parameter estimation are identified following the guidelines presented by Box and Jenkins (1976), applied by Hipel *et al*. (1977); McLeod *et al*. (1977); Rao *et al*. (1982) and described by Bowerman and O'Connel (1993). These guidelines are based on the behavior of the autocorrelation (ACF) and partial autocorrelation functions (PACF).

After model identification, the modeler needs to obtain an efficient estimation of the parameters. The model parameters should satisfy two conditions, namely, stationarity and invertibility for the autoregressive and the moving average parameters, respectively. The parameters should also be tested as to whether they are statistically

Table I. Selected stations with geographical location and annual rainfall characteristics.

Stations	Mean annual rainfall (mm)	STDEV (mm)	C_V	Latitude	Longitude	Elevation (m)
Ahwaz	213.3	86.3	0.40	31°20′	48°40′	22
Arak	345	92.8	0.27	34°06′	49°46′	1708
Ardabil	309	88	0.28	38°15′	48°17′	1332
Bandar Abbas	192	121.8	0.63	27°13′	56°22′	10
Bushehr	275.6	118.8	0.43	28°59′	50°50′	19
Ghaemshahr	752.3	116.7	0.16	36°27′	52°46′	14
Gorgan	612.1	102.8	0.17	36°51′	54°16′	13
Ghazvin	315.9	89.5	0.28	36°15′	50°03′	1279
Hamedan	316.2	76.6	0.24	35°12′	48°43′	1679
Isfahan	121.4	40.1	0.33	32°37′	51°40′	1550
Ilam	627.9	170.7	0.27	33°38′	46°26′	1337
Oroomieh	349.3	98.4	0.28	37°32′	45°05′	1315
Ghom	149	47.1	0.32	34°42′	50°51′	877
Zahedn	94.8	40.1	0.42	29°28	48°40′	1370
Zanjan	317.6	72.6	0.23	36°41′	48°29′	1663
Yazd	62.1	27.9	0.45	31°54	54°17′	1237
Yasuj	822.9	183.02	0.22	30°50′	51°41′	1831
Tehran	229.2	63.92	0.28	35°41′	51°19′	1190
Tabriz	293.3	68	0.23	38°05′	46°17′	1361
Shiraz	344.7	99.7	0.29	29°32′	52°36′	1481
Shahrecord	319	86.6	0.27	32°17′	50°51′	2048
Semnan	139.9	54.2	0.39	35°35′	53°33′	1130
Sanandaj	471	118.8	0.25	35°20′	47°04′	1373
Rasht	1353	279.3	0.21	37°12′	49°39′	36
Mashhad	257.5	77.4	0.30	36°16′	59°38′	999
Khoramabad	515.1	125.6	0.24	33°26′	48°17′	1147
Kermanshah	450.8	120.4	0.27	34°21′	47°09′	1318
Kerman	158.9	50.2	0.32	30°15′	56°58′	1753

significant or not. Associated with parameter values are standard error of estimation and related t-value which are used to investigate the statistical significance of the parameters.

Goodness of fit tests verify the validity of the model by some tools. In this step, the residuals of the model are considered to be time-independent and normally distributed over time. The popular Portmanteau lack of fit test based on Ljung – Box statistic (Salas *et al.*, 1980) is written as:

$$Q^* = n'(n'+2)\sum_{k=1}^{L}(n'-l)^{-1}r_e^2(\hat{\varepsilon}) \tag{3}$$

Q^* is approximately distributed as $\chi^2(L-p-q)$ and has $k-n_p$ degrees of freedom. We have considered $L = 48$. If the probability of Q^* is less than $\alpha = 0.01$, there is a strong evidence that the model is inadequate and if the probability is greater than $\alpha = 0.05$, it is reasonable to conclude that the model is adequate. In this study, the $\alpha = 0.05$ significant level is used as the significant level for the model building.

Regionalization: multivariate methods

Multivariate techniques are common methods for classifying meteorological data such as rainfall. Principal components and cluster techniques are used in this study to classify autocorrelation coefficient of rainfall series in different groups. Let the matrix **X** ($m \times k$) consist of autocorrelation coefficients at lags $k = 1, \ldots, 12$ of m stations. $k = 12$ is chosen as the autocorrelation; coefficients of higher lags are not significant or have similar seasonal fluctuations as the first $k = 12$. A commonly used dissimilarity measure is the Euclidean distance (d_{rs}^2) which is written as follows (Jobson, 1992):

$$d_{rs}^2 = \sum_{j=1}^{k}(x_{rj} - x_{sj})^2 \tag{4}$$

where the r^th^ and s^th^ rows of the data matrix X is denoted by ($x_{r1}, x_{r2}, \ldots, x_{rk}$) and ($x_{s1}, x_{s2}, \ldots, x_{sk}$) respectively. In this study, our matrix consists of 12 lags of ACF of 28 rainfall time series. These 12 autocorrelation coefficients are selected because the coefficients in higher lags are similar to them or are not significant at very higher lags. This means a matrix of 28 columns of stations and 12 rows of autocorrelation coefficients of rainfall series.

The Euclidean distance of dissimilarity is then used in the cluster techniques. As the variables must not be correlated with each other, PCA is usually prefixed (Backhaus *et al.*, 1994). PCA was first applied to reduce a large data matrix into some important factors (principal

DOI: 10.1002/joc

components). The first principal component is the linear combination of the original variables that captures as much of the variation in the original data as possible. The second component captures the maximum variation that is uncorrelated with the first component, and so on. After having decided which autocorrelation coefficient at which lags are to be used for classification, cluster analysis based on Ward's method is applied. To choose the proper number of the clusters, one can refer to the total spatial variance for each number of the clusters. The R-Squared, squared multiple correlation or the decrease in the proportion of variance accounted for due to joining two clusters to form the current cluster, is the key to select the number of the proper clusters. In order to illustrate these clusters, we apply canonical discriminant analysis. Canonical discriminant analysis is a dimension-reduction technique related to PCA and the canonical correlation. In a canonical discriminant analysis, we find the linear combinations of the quantitative variables that provide maximum separation between the classes or the groups. Two output data sets can be produced: one containing the canonical coefficients and another containing the scored canonical variables. The scored canonical variables output data set can be used to plot pairs of canonical variables to aid visual interpretation of the group differences.

RESULTS AND DISCUSSION

Time series modeling

The process of time series modeling begins with the selection of the preliminary models interpreted from the characteristics of ACF and PACF functions using SAS ARIMA procedure (SAS/ETS, 1999). At first look, the monthly fluctuations show the seasonal behavior of the temporal pattern of the monthly rainfall due to the significant correlation coefficients at lag k = 12. For example, the ACF and PACF of the Ahwas, Isfahan and Ghaemshahr monthly rainfall series are presented in Figure 3. The parameter estimation of the preliminary selected models is then applied using the method of maximum likelihood. For example, the results of the parameter estimation and the satisfaction of the stationarity and invertibility conditions for Isfahan stations are presented in Table II. This model has been derived based on trying several models with different orders of the parameters. As all invertibility and stationarity conditions are accepted for the model, the model residuals were checked for stationarity and normality using Portmanteau lack of fit test and normal tests.

The portmanteau lack of fit test and the two normal tests (Kolmogrov–Smirnov and Anderson–Darling tests) proved the residuals to be time-independent (stationarity) and normally distributed. As a result from the above monthly rainfall time series modeling steps for Isfahan station, the best model for this station is ARIMA$(1,0,0)(0,1,1)_{12}$. Plotting the observed and the model predicted rainfall time series shows that the model performs the observed rainfall series very well (Figure 4). Following the above procedures for all the selected stations, the best model for each station was estimated and presented in column 2 of Table III. In columns 3 and 4, the lag 1 and lag 12 autocorrelation coefficient values are also shown.

It is clear from Figure 4 that the model predicted rainfall series have the same seasonal fluctuations with the observed rainfall series. For better verification of the selected models and for checking their efficiency, two criteria are used, the correlation coefficient, R^2, between observed and model predicted rainfall series and the $R^2{}_{N-S}$ criterion of Nash and Sutcliffe (1970). It is related to the sum of the squares of the differences, F, between the estimated and observed rainfall. This criterion is defined by

$$R^2{}_{N-S} = \frac{F_\circ - F}{F_\circ} \qquad (5)$$

where $F_\circ$ is the sum of the squares of differences between the observed rainfall and the mean rainfall. A value of $R^2{}_{N-S}$ greater than 90% would normally indicate a very satisfactory model performance while a value in the range 80–90% is regarded as an indication of a fairly good model. Values of $R^2{}_{N-S}$ in the range 60–80% generally indicate an unsatisfactory model fit (Shamseldin and O'Connor, 2001). The correlation coefficients and R^2 criterion of Nash and Sutcliffe are presented in columns 5 and 6 of Table III, which postulate that the rainfall predicted by the models fits correctly the observed values with $R^2 > 0.78$ and $R^2{}_{N-S} > 85\%$ to the observed rainfall and the fitted ARIMA models are satisfactory in all stations.

Regional climates: cluster analysis

In the first step, PCA was applied on the autocorrelation coefficients of the lag k = 1 to the lag k = 12 for 28 stations using SAS software (SAS/STAT, 1999). To improve the interpretation of the unrotated PCA results, the principal components are rotated using the orthogonal VARIMAX rotation. This rotation method is common where a cluster analysis is to follow PCA because the PCs are uncorrelated and the assumptions of the cluster analysis are satisfied (McGregor, 1993).

Table IV lists the VARIMAX-rotated Principal components. The first component, which has the largest loadings for the lags k = 4–8, explains 59% of the total variance between stations. The second component which has large loadings for the lags k = 1 and k = 12, explains 21% of the total variance, while the third component explains 12% of the total variance. The first two factors which describe 80% of the total between-stations variance can be called 'stationarity' and '*periodicity*'. This is because the significant autocorrelations at higher lags (higher than k-3) are the sign of stationarity in the rainfall statistical characteristics like the mean or variance and the significant autocorrelation at lag 12 shows the seasonality or the periodicity in the statistical characteristics of rainfall

DOI: 10.1002/joc

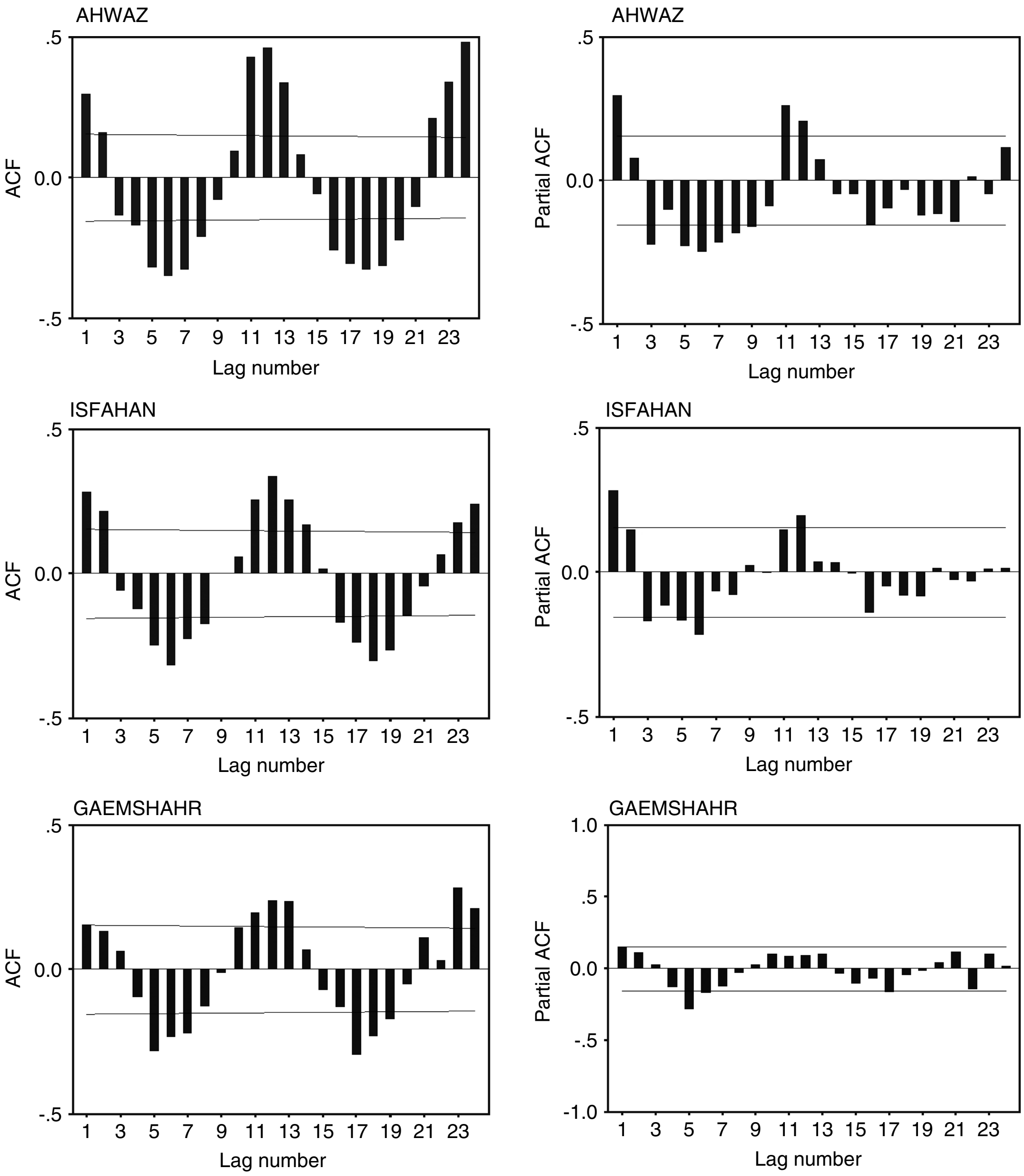

Figure 3. Autocorrelation (a) and partial autocorrelation (b) functions of some selected stations.

Table II. The parameter estimations of model ARIMA(1,0,0)(0,1,1)$_{12}$ for Isfahan station.

Estimation method	Type and order of parameters	Parameter value	Standard error	t-value	Probability of t	Invertibility and stationarity conditions
ML	Seasonal MA(1)	0.92	0.03	25.14	$P < 0.0001$	Accepted
	AR(1)	0.14	0.05	2.74	$P < 0.0061$	

(Salas, 1993). In other words, stationarity and periodicity are the two main elements by which the temporal characteristics between stations could be explained. The third factor can be called '*T-GCM*' effect. '*T-GCM*', refers to the 'topographic-general circulation model' effects on the temporal behavior of the rainfall. This means that 12% of the rainfall variations are the effect of the elevation and the atmospheric circulation patterns. However, further investigations are necessary to demonstrate these effects which are out of the scope of the present study. As the above three components explain 92% of the total variance between lags, the 28 × 3 matrix of factor scores was subjected to the hierarchical clustering based on the method of Ward's minimum variance. To choose the

DOI: 10.1002/joc

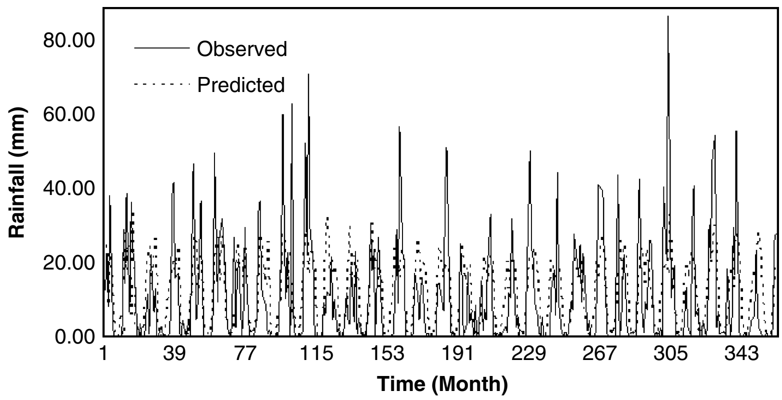

Figure 4. Time series of observed and model predicted rainfall for Isfahan station.

Table III. The best time series model, efficiency values, autocorrelation coefficients and the groups of the selected stations.

Station (1)	Best model (2)	Values of autocorrelation coefficient		R^2 (5)	R^2_{N-S} (%) (6)	Groups (7)
		Lag-one (3)	Seasonal (lag-12) (4)			
Ahwaz	$ARIMA(3,0,0) \times (0,1,1)_{12}$	0.27	0.36	0.92	90.22	3
Arak	$ARIMA(1,0,0) \times (0,1,1)_{12}$	0.42	0.44	0.94	90.31	2
Ardabil	$ARIMA(1,0,0) \times (7,1,1)_{12}$	0.17	0.2	0.89	86.04	3
Bandarabbas	$ARIMA(1,1,1)_{12}$	0.38	0.33	0.91	90.03	1
Bushehr	$ARIMA(1,1,1)_{12}$	0.35	0.3	0.91	90.25	1
Ghaemshahr	$ARIMA(0,1,1)_{12}$	0.18	0.28	0.90	89.6	1
Gorgan	$ARIMA(0,1,1)_{12}$	0.28	0.32	0.90	89.26	1
Ghazvin	$ARIMA(0,0,1) \times (0,0,1)_{12}$	0.44	0.46	0.93	90.8	2
Hamedan	$ARIMA(8,0,11) \times (19,1,1)_{12}$	0.47	0.55	0.87	85.09	3
Isfahan	$ARIMA(1,0,0) \times (0,1,1)_{12}$	0.48	0.4	0.90	90.01	2
Ilam	$ARIMA(1,1,1)_{12}$	0.25	0.41	0.92	91.01	1
Oroumieh	$ARIMA(6,0,0) \times (0,6,1)_{12}$	0.34	0.31	0.89	90.0	3
Ghom	$ARIMA(1,0,4) \times (4,1,1)_{12}$	0.33	0.42	0.88	85.78	3
Zahedan	$ARIMA(1,0,1)_{12}$	0.27	0.17	0.94	91.1	1
Zanjan	$ARIMA(6,0,0) \times (0,1,1)_{12}$	0.41	0.46	0.91	90.3	3
Yazd	$ARIMA(1,0,1)_{12}$	0.38	0.42	0.89	89.5	1
Yasuj	$ARIMA(2,0,1) \times (0,1,1)_{12}$	0.32	0.33	0.90	90.06	2
Tehran	$ARIMA(1,0,1)_{12}$	0.38	0.4	0.91	90.7	1
Tabriz	$ARIMA(0,0,1) \times (1,0,1)$	0.29	0.18	0.91	90.54	2
Shiraz	$ARIMA(1,0,1)_{12}$	0.28	0.24	0.92	90.87	1
Shahrecord	$ARIMA(1,0,1)_{12}$	0.23	0.36	0.90	88.6	1
Semnan	$ARIMA(1,1,1)_{12}$	0.2	0.35	0.93	91.5	1
Sanandaj	$ARIMA(0,1,1)_{12}$	0.21	0.24	0.94	92.61	1
Rasht	$ARIMA(0,1,1)_{12}$	0.41	0.46	0.94	93.1	1
Mashad	$ARIMA(1,0,0) \times (1,1,1)_{12}$	0.25	0.24	0.89	86.4	2
Khoramabad	$ARIMA(1,0,1)_{12}$	0.53	0.51	0.92	89.74	1
Kermanshah	$ARIMA(1,0,1)_{12}$	0.21	0.29	0.91	90.2	1
Kerman	$ARIMA(1,1,1)_{12}$	0.32	0.34	0.91	90.25	1

proper number of clusters, the R-squared showed that 2, 3, 4 and 5 clusters explain 71, 99.2, 99.6 and 99.8% of between-group variance, respectively. It is evident that 3 clusters with 99.2% of variance is the best and the most proper number of the groups of the temporal behaviors of rainfall over Iran. The canonical variables of each group are plotted in Figure 5 using Canonical Discriminant Analysis (SAS/STAT, 1999). One can see on Figure 5 a good separation between the three groups, which clearly indicate the validity of the classification. The membership of each station to each group has been presented in column 7 of Table III.

Climate regions: time series model and rainfall temporal characteristics relationships

The different temporal characteristics of the stations and the groups can be interpreted as follows. All ACF and PACF indicate the seasonality of rainfall which means

DOI: 10.1002/joc

Table IV. VARIMAX rotated principal component (PCs) loadings of different lags (K = 1–12).

Lags	PC1	PC2	PC3
K = 1	−0.411	0.755	0.292
K = 2	−0.577	0.359	0.415
K = 3	0.004	0.073	0.174
K = 4	0.868	−0.323	0.096
K = 5	0.903	−0.197	−0.270
K = 6	0.839	−0.223	−0.358
K = 7	0.890	−0.271	−0.278
K = 8	0.869	−0.213	0.149
K = 9	−0.055	0.068	0.573
K = 10	−0.467	0.406	*0.976*
K = 11	−0.014	−0.001	−0.085
K = 12	−0.296	0.912	0.022

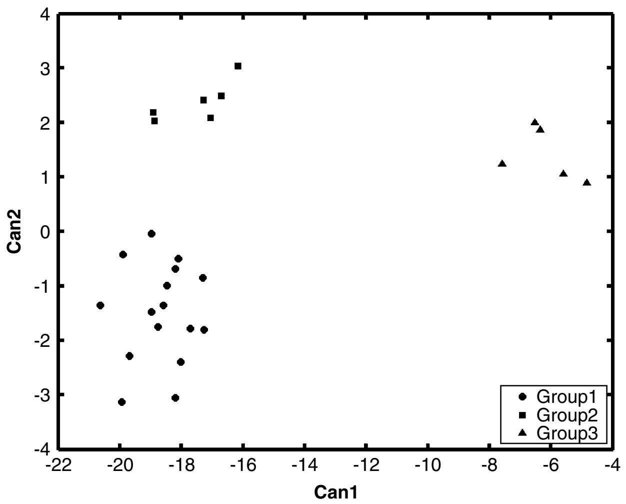

Figure 5. Illustration of the spatial separation of the canonical scores.

that the same months are correlated with each other during consecutive years. The existence of seasonal moving average parameters (see Table III, values of $MA(Q) \neq 0$) in the model confirms these seasonal fluctuations in the rainfall temporal fluctuation. It also indicates the existence of the seasonal pattern in the atmospheric conditions over Iran. On the other hand, some stations show nonseasonal moving average (see Table III, values of $MA(q) \neq 0$) which proves nonseasonal fluctuations in some stations such as Yasuj, Ghom, Hamedan, Ghazvin and Tabriz. The existence of autocorrelation parameters in some models can be interpreted as the persistence of the climate condition. For example, the persistent condition and the prevalence of the snow precipitation against rainfall in most of the humid western and north western stations, Oroumieh, Zanjan and Hamedan, is the reason for having a higher AR(p) parameter. It also suggests the parallel act of different rainfall generating mechanisms like the large-scale cyclonic conditions with a 'life span' lasting about several days or months and the effect of the elevation variation in microscale (Burlando and Rosso, 1993).

In Table V, one can see that the correlation between lag-one autocorrelation coefficient and the station elevation or the rainfall coefficient of variation (C_v) is not strong, while these correlations are significant for lag-12 (seasonal) autocorrelation coefficient with values of 0.37 and −0.48 for the station elevation and C_v, respectively. The correlations between lags 4 and 8 and the elevation are also significant. These correlations imply the effect of the station elevation on the variation of the temporal behavior of the rainfall in some stations like Zanjan, Oroumieh and Hamedan. The low order MA (q) parameter (see Table III) shows a simple and uniform precipitation generating mechanism with the uniform and periodic temporal rainfall scaling (Burlando and Rosso, 1993). This pattern can be seen in the margin of the Caspian Sea and the Persian Gulf.

Based on the hierarchical cluster analysis of the seasonal autoregressive parameter, we can find three main types of the rainfall time series models. We call them simple (group 1), moderate (group 2) and complex (group 3) groups. The simple group includes the pure seasonal models in the form of ARIMA(p,d,q)_{12}. The moderate group has a general form of ARIMA(p,d,q) $\times$ (P,D,Q)_{12} with a low order of AR(p) and MA(q) parameters. The third complex group consists of the same models as the second group but with a higher order of AR(p) and MA(q) parameters. In other words, these groups can be defined as characterized by simple climate conditions with regular fluctuations (group 1), simple and more persistent climate conditions with more irregularity due to the different rainfall generating mechanism (group2) and high temporal irregularity due to the interaction between large and microscale climate conditions and elevation (group 3). The spatial pattern of these three groups is presented in Figure 6.

Table V. Correlation matrix between autocorrelation coefficients (R) at lags k = 1–12 and rainfall properties (mean and STDEV).

	C_V	Elevation	Mean	STDEV
R_1	−0.29	0.28	0.29	0.35
R_2	0.11	0.18	−0.01	0.11
R_3	0.02	0.02	0.24	0.27
R_4	0.09	−0.43*	0.04	−0.04
R_5	0.02	−0.17	−0.12	−0.19
R_6	−0.01	−0.11	−0.15	−0.23
R_7	0.03	−0.17	−0.13	−0.21
R_8	0.13	−0.39*	−0.03	−0.07
R_9	0.15	−0.21	0.23	0.29
R_{10}	−0.07	−0.44*	0.39*	0.48**
R_{11}	−0.15	0.18	−0.03	−0.08
R_{12}	−0.48**	0.37**	0.33	0.36

* Significant at 95%, ** Significant at 99%.

DOI: 10.1002/joc

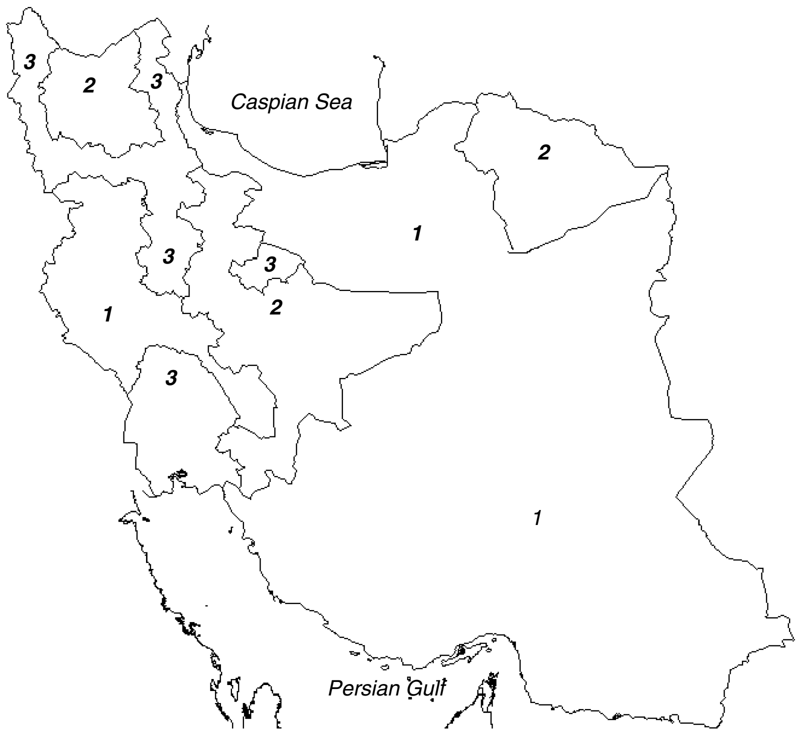

Figure 6. Spatial pattern of climate regions of Iran based on time series modeling.

CONCLUSIONS

Time series modeling of the major cities of Iran was analyzed in this study. The Box–Jenkins popular ARIMA model was applied and seemed to fit the monthly rainfall time series very well. Different ARIMA models such as pure seasonal model (ARIMA$(P,D,Q)_{12}$); multiplicative model, ARIMA(p,d,q) $\times$ $(P,D,Q)_{12}$ with low order parameters and multiplicative model, ARIMA(p,d,q) $\times$ $(P,D,Q)_{12}$ with high order parameters were fitted to the rainfall series. These models indicate temporal characteristics of rainfall generating mechanism over Iran very well. The spatial pattern of these rainfall time series models was determined using the PCA in order to delineate the temporal rainfall groups and related macroscale rainfall generating mechanisms over Iran. These macro mechanisms were divided into three groups with different temporal characteristics such as regular year to year rainfall fluctuations, rainfall with different generating mechanisms such as elevation and sea neighborhood and rainfall which is affected by the general atmospheric circulations with both seasonal and nonseasonal fluctuations.

In other words, these groups are indications of the effect of the sea neighborhood and the elevation on the variation in the temporal behavior of the rainfall from pure fluctuation to stationary conditions, respectively. The PCA analysis indicates that the autocorrelation coefficients at different lags are good indicators of the temporal relationship of the rainfall and can result in a suitable classification of the rainfall regions of Iran. The rainfall time series modes fitted to monthly rainfall of the selected stations can also be used for estimating missing rainfall values, rainfall forecasting and investigating future rainfall change due to global climate change. It would also be appropriate in future studies to investigate the relationship between temporal behavior of rainfall time series and El Nino Southern Oscillation effects such as droughts and floods using multivariate time series modeling.

ACKNOWLEDGEMENT

This work is financially supported by the research council of Isfahan University of Technology. The comments of the two anonymous reviewers contributed to significant improvements in the paper and they are appreciated.

Nomenclature

ACF	AutoCorrelation Function
ARIMA	AutoRegressive Integrated Moving Average
AR(p)	AutoRegressive parameter of order (p)
B	backward shift operator
C_v	coefficient of Variation, C_v = Standard Deviation/ Average
d	order of nonseasonal differencing operator
D	order of seasonal differencing operator
d_{rs}^2	euclidian distance
k	lag time of autocorrelation function
L	maximum lags of residual autocorrelation function
MA(q)	moving average of parameter (q)
ML	maximum likelihood

DOI: 10.1002/joc

n	the number of observation
n_p	number of parameters
n'	$n - n_p$
m	number of stations in matrix **X**
PACF	Partial AutoCorrelation Function
PC	Principal Component
PCA	Principal Component Analysis
Q*	Ljung-Box statistic
$r_e^2(\hat{\varepsilon})$	correlation function of residuals
s	seasonal difference
STDEV	standard deviation
VARIMAX	variance maximum
Z_t	observation at time t

Greek Symbols

ϕ	nonseasonal autoregressive parameter
θ	nonseasonal moving average parameter
ε_t	residual of the model at time t
Φ	seasonal autoregressive parameter
Θ	seasonal moving average parameter
∇	differencing operator
χ	chi-square distribution

REFERENCES

Acerman MC. 1985. Predicting the mean annual flood from basin characteristics in Scotland. *Hydrological Sciences Journal* **30**: 37–49.

Acerman MC, Sinclair CD. 1986. Classification of drainage basins according to their physical characteristics: an application for flood frequency analysis in Scotland. *Journal of Hydrology* **84**: 365–380.

Backhaus K, Erichson B, Plinke W, Weiber R. 1994. *Multivariate Analyse methoden*. Springer: Berlin; 594.

Banalya A, De Troch F, Sebei H. 1998. Characterization and stochastic modeling of monthly rainfall time series. In *Hydrology in a Changing Environment*, Vol. 1, Wheater H, Kirby C (eds). British Hydrological Society: West Sussex, UK; 54–61.

Box GEP, Jenkins GM. 1976. *Time Series Analysis, Forecasting and Control*, Revised edn. Holden – Day: San Francisco, CA.

Bowerman BL, O'Connel RT. 1993. *Forecasting and Time Series, an Applied Approach*. Duxbury: Pasific Grove, CA.

Burlando P, Rosso R. 1993. Stochastic models of temporal rainfall: reproductibility, estimation and prediction of extreme events. In *Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization*, Macro JB, Harboe R, Salas JD (eds). NATO Advanced Study Institute, Kluwer Academic Publishers: Boston; 137–173.

Carslaw DC. 2005. On the changing seasonal cycles and trends of ozone at Mace Head, Ireland. *Atmospheric Chemical Physics* **5**: 3441–3450.

Delleur JW, Tao PC, Kavass ML. 1976. An evaluation of practicality and complexity of some rainfall and runoff time series models. *Water Resources Research* **12**: 953–970.

Elek P, Markus L. 2004. A long range dependent model with non linear innovations for simulating daily river flows. *Natural Hazards and Earth System Sciences* **4**: 277–283.

Gottschalk L. 1985. Hydrological regionalization of Sweden. *Hydrological Sciences Journal* **30**: 65–83.

Harris NM, Gurnell AM, Hannah DM, Petts GE. 2000. Classification of river regimes: a context for hydroecology. *Hydrological Processes* **14**: 2831–2848.

Hipel KW, McLeod AE. 1994. *Time Series Modeling of Water Resources and Environmental Systems*. Elsevier: Amsterdam.

Hipel KW, McLeod AE, Lennox WC. 1977. Advances in Box- Jenkins modeling, 1-model construction. *Water Resources Research* **13**: 567–575.

Jobson JD. 1992. *Applied Multivariate Data Analysis, Categorical and Multivariate Methods*, Vol. II. Springer-Verlag: New York; 731.

Kallache M, Rust HW, Kropp J. 2005. Trend assessment: applications for hydrology and climate research. *Nonlinear Processes in Geophysics* **12**: 201–210.

Kitpati S, Fuji S, Shimizu Y. 2005. Development of an integrated analysis procedure for three-ways layout field surveys data in lakes. *Water Science and Technology* **52**(9): 283–294.

Maharaj EA. 1999. Comparison and classifying of stationary multivariate time series. *Pattern Recognition* **32**: 1129–1138.

Maharaj EA. 2000. Clusters of time series. *Journal of Classification* **17**: 297–314.

McGregor GR. 1993. A multivariate approach to the evaluation of climatic regions and climatic resources of China. *Geoforum* **24**: 357–380.

McLeod AE, Hipel KW, Lennox WC. 1977. Advances in Box-Jenkins modeling, 2-Applications. *Water Resources Research* **13**: 577–586.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models. Part 1. A discussion of Principals. *Journal of Hydrology* **10**: 282–290.

Palecki MA, Angel JR, Hollinger SE. 2005. Storm precipitation in the United States. Part I: meteorological characteristics. *Journal of Applied Meteorology* **44**(6): 933–946.

Piccolo D. 1990. A distance measure for classifying ARIMA models. *Journal of Time Series Analysis* **11**: 152–164.

Rao AR, Kashyap RL, Mao LT. 1982. Optimal choice of type and order of models for river flow time series. *Water Resources Research* **18**: 1097–1109.

Romero R, Summer G, Ramis C, Genoves A. 1999. A classification of the atmospheric circulation patterns producing significant daily rainfall in the Spanish Mediterranean area. *International Journal of Climatology* **19**: 765–785.

Salas JD. 1993. Analysis and modeling of hydrologic time series. In *Handbook of Hydrology*, Maidment DR (ed.). McGraw Hill: New York; 19.1–19.72.

Salas JD, Obeysekera JTB. 1982a. ARMA model identification of hydrologic time series. *Water Resources Research* **18**: 1011–1021.

Salas JD, Smith RA. 1982b. Physical basis of stochastic models of annual flows. *Water Resources Research* **17**: 428–430.

Salas JD, Fernandez B. 1993. Models for data generation in hydrology: univariates techniques. In *Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization*, Macro JB, Harboe R, Salas JD (eds). NATO Advanced Study Institute, Kluwer Academic Publishers: Boston; 47–73.

Salas JD, Delleur JW, Yevjevich VM, Lane WL. 1980. *Applied Modeling of Hydrologic Time Series*. Water Resources Publications: Littleton, CO.

Sanz DB, del Jalon DG. 2005. Characterisation of streamflow regimes in central Spain, based on relevant hydrobiological parameters. *Journal of Hydrology* **310**(1–4): 266–279.

SAS/ETS. 1999. *User's Guide, Version 8*. SAS Inc: Cary, NC.

SAS/STAT. 1999. *User's Guide, Version 8*. SAS Institute Inc.: Cary, NC.

Sen Z, Kadioglu M, Batur E. 1999. Cluster regression model and level fluctuation features of Van Lake, Turkey. *Annales Geophysicae* **17**: 273–279.

Shamseldin AY, O'Connor KM. 2001. A non-linear neural network technique for updating of river flow forecasts. *Hydrology and Earth System Sciences* **5**(4): 577–597.

Singh CV. 1999. Principal components of monsoon rainfall in normal, flood and drought years over India. *International Journal of Climatology* **19**: 639–952.

Srinivas G, Rao R, Quilan EW. 1982. Stochastic ARIMA models for monthly streamflows. In *Time Series Methods in hydrosciences, Proceedings of an International Conference for Inland Waters*, Vol. 2, El Shaarawi AH, Esterby SR (eds). Elsevier Scientific Publication: New York; 474–486.

Stahl K, Demuth S. 1999. Methods for regional classification of streamflow drought series: cluster analysis. Technical report to the ARIDE project, No. 1. University of Freiburg, Germany.

Xiong Y, Yeung D. 2004. Time series clustering with ARMA mixtures. *Pattern Recognition* **37**: 1675–1689.