



Simulation of Seasonal precipitation using ANN and ARIMA Models: A case study of (Iran) Khozestan

Mohammad Reza Golabi¹, Feridon Radmanesh², Ali Mohammad Akhondali² and Mahmood Kashefipoor²

¹Water Resource Engineering, Faculty of Water Engineering.

²Faculty of Water Engineering, Shahid Chamran University, Ahvaz, Iran.

ARTICLE INFO

Article history:

Received: 6 December 2012;

Received in revised form:

13 February 2013;

Accepted: 14 February 2013;

Keywords

ARIMA model,
Neural Networks,
Rainfall forecast,
Time series.

ABSTRACT

Accurate rainfall prediction is of great interest for water management and flood control. In reality, physical processes influencing the occurrence of rainfall are highly complex, uncertain and nonlinear. In this paper, we present tools for modeling and predicting the behavioral pattern in rainfall phenomena based on past observations. The aim of this paper is to predict the seasonal rainfall of (Iran) khozestan using artificial neural network (ANN) and autoregressive integrated moving average (ARIMA) models. In order to evaluate the prediction efficiency, we made use of 33 years of seasonal rainfall data from year 1976 to 2008 of Khozestan Province (Iran). The models were trained with 28 years of seasonal rainfall data. The ANN and the ARIMA approaches are applied to the data to derive the weights and the regression coefficients respectively. The performance of the model was evaluated by using remaining 5 years of data. The study reveals that ANN model can be used as an appropriate forecasting tool to predict the rainfall, which out performs the ARIMA model.

© 2013 Elixir All rights reserved.

1. Introduction

Accurate quantitative rainfall forecasting is one of the most desired aspects of weather prediction to the general community. Rainfall is natural climatic phenomena whose prediction is challenging and demanding. Its forecast is of particular relevance to agriculture sector, which contributes significantly to the economy of the nation. On a worldwide scale, numerous attempts have been made to predict its behavioral pattern using various techniques (Somvanishi et al., 2006). The numerical modeler is faced with the problem of predicting a physical process that could be sensitive to any of a number of factors, such as wind, temperature, and humidity, in highly nonlinear ways, including some ways that are not completely understood at the present (Rangno and Hobbs 1994). Another important factor is introduced in regions with highly variable physiography (surface features such as topography, land-water boundaries, vegetation, and soil moisture) on small scales, which can have a profound influence on any of the factors mentioned above. The general objective of this study was to simulate rainfall using autoregressive integrated moving average (ARIMA) and artificial neural network (ANN). Artificial neural networks are mathematical models, the architecture of which has been inspired by biological neural networks (Erzin et al., 2007). ANNs are very appropriate for the modeling of nonlinear processes, such as the case of rainfall. This paper convincingly demonstrates the advantages of using ANN over that of ARIMA technique to model the rainfall behavior. The study of rainfall time series is a topic of great interest in the field of climatology and hydrology. Some significant examples in such areas include Singh (1998). Both univariate (e.g. Soltani et al., 2007) and multivariate (Grimaldi et al., 2005) approaches have been attempted to model the rainfall time series. Impact of other atmospheric variables on rainfall has been discussed in various

literatures (Chattopadhyay, 2007b). The association between rainfall and agrometeorological processes is well discussed (e.g. Jhajharia et al., 2009; Chattopadhyay et al., 2009). Several stochastic models were attempted to forecast the occurrence of rainfall, to investigate its seasonal variability and to forecast monthly/yearly rainfall over some given geographical area. Study of the rainfall is interesting because of the associated problems, such as forecasting, corrosion effects and climate variability and various literatures have discussed these issues (Tzanis and Varotsos, 2008). Chaotic features associated with the atmospheric phenomena have attracted the attention of modern scientists (Bandyopadhyay and Chattopadhyay, 2007). Mathematical tools based on the theoretical concepts underlying the methodologies for detection and modelling of dynamical and chaotic components within a hydrological time series have been studied extensively by various scientists like Islam and Sivakumar (2002) and Jayawardena and Lai (1994). Phase space reconstruction and artificial neural networks (ANN) are non-linear predictive tools that have been proposed in the modern literature as effective mathematical methodologies to be useful to hydrological time series characterized by chaotic features (Chattopadhyay and Chattopadhyay, 2008a; Elsner and Tsonis, 1992; Khan et al., 2005). Applicability of ANN to rainfall time series is well documented in the literature. Prediction of atmospheric events, especially rainfall, has benefited significantly by voluminous developments in the application field of ANN and rainfall events and quantities have been predicted statistically (e.g. DelSole and Shukla, 2002; Mohanty and Mohapatra, 2007). Guhathakurata (2008) generated an ANN based model that captured the input-output non-linear relationship and predicted the monsoon rainfall in India quite accurately.

Tele:

E-mail addresses: hamid_golabi65@yahoo.com

© 2013 Elixir All rights reserved

2. Materials and methods

2.1. Study area and rainfall data

Khuzestan Province is in south-western Iran, it covers an area of 63633/6 km² between latitudes 29° 57' -33° 4' N and longitudes 47° 40' - 50° 33' E. The climate of the province is affected by weather systems from the Mediterranean and the Persian Gulf so that the weather is typically that of a semi-arid/temperate climate. Basically, the province of Khuzestan can be divided into two regions, the plains and mountainous regions. Winters in this zone are short and moderate, while the summers are long and hot. In this research, the rainfall data of stations Ahvaz, Abadan and Dezful of Khuzestan province (Iran) has been used for studying the rainfall conditions of the province, and The data for the analysis are on a seasonal basis for the period of 33 years from 1976 to 2008. In this study, the first 112 seasonal of rainfall data were used for model training. The remaining 20 seasonal of rainfall data were used for verification of the model prediction results.

2.2. ARIMA model

The general form of the ARIMA model is (Vandaele 1983)

$$\phi(B)x_t = \theta(B)a_t \quad (1)$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ is the non-seasonal autoregressive polynomial; B is the backward shift operator; ϕ_p represents the autoregressive parameters of the model; p is the order of autoregressive polynomial; x_t is the stationary series after differencing, $x_t = (1 - B)^d$; d is the number of non-seasonal differencing; X_t is the dependent variable; $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ is the non-seasonal moving average polynomial; θ_q is the moving average of the model; q is the order of moving average polynomial; and a_t is the white noise process. In an autoregressive integrated moving average model (ARIMA), the future value of a variable is assumed to be a linear function of several past observations and random errors. An ARIMA model can be explained as ARIMA(p, d, q)(P, D, Q)_s, where (p, d, q) is the non seasonal part of the model and (P, D, Q)_s is the seasonal part of the model which is mentioned below

$$\phi_p(B)\Phi_P(B^s)\nabla^d \nabla_S^d z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (2)$$

where p is the order of non-seasonal autoregression, d is the number of regular differencing, q is the order of non-seasonal MA, P is the order of seasonal autoregression, D is the number of seasonal differencing, Q is the order of seasonal MA, s is the length of season (periodicity), ϕ is the AR operator of order p , Φ is the seasonal AR parameter of order P , ∇^d is the differencing operator ∇_S^d is the seasonal differencing operator, z_t is the observed value at time point t , θ is the MA operator of order q , Θ is the seasonal MA parameter of order Q and a_t is the noise component of the stochastic model assumed to be $NID(0, \sigma^2)$. The ARIMA modeling approach involves the following three steps: model identification, parameter estimation, diagnostic checking. Identification of the general form of a model includes two stages: (1) if it is necessary, appropriate differencing of the series is performed to achieve stationary and normality; (2) the temporal correlation structure of the transformed data is identified by examining its autocorrelation (ACF) and partial autocorrelation (PACF) functions (Mishra and Desai, 2005). The ACF is a useful statistical tool that measures if earlier values in the series have some relation to later values. PACF is the amount of correlation between a variable and a lag of itself that is not explained by

correlations at all low order lags. Considering the ACF and PACF graphs of seasonal rainfall series, different ARIMA models are identified to model selection. The model that gives the minimum Akaike Information Criterion (AIC) is selected as the best fit model. The mathematical formulation for the AIC is developed as

$$AIC = n \left(\ln \left(\frac{2\pi RSS}{n} \right) + 1 \right) + 2m \quad (3)$$

where $m = (p + q + P + Q)$ is the number of terms estimated in the model and RSS denotes the sum of squared residuals (Hirotsugu, 1974). After the functions of the ARIMA model have been specified, the parameters of these functions must be estimated. Once an appropriate model is chosen and its parameters are estimated, the Box-Jenkins methodology requires examining the residuals of the model to verify that the model is an adequate one for the series. Several tests are employed for diagnostic check to determine whether the residuals of the selected ARIMA models from the ACF and PACF graphs are independent, homoscedastic and normally distributed. If the homoscedasticity and normality assumptions are not provided, the observations are transformed by a Box-Cox transformation (Wei, 1990). For a good forecasting model, the residuals, left over after fitting the model, must satisfy the requirements of a white noise process (uncorrelated and normally distributed around a zero mean). In order to determine whether seasonal rainfall time series are independent, the residual autocorrelation (RACF) function of the series is studied. There are several useful tests related to RACF for the independence of residuals. The first one is the correlograms drawn by plotting the residual ACF function against lag number. If the ARIMA model is correct, the estimated autocorrelations of the residuals are uncorrelated and distributed approximately normally about zero. The second one is Ljung-Box-Pierce statistics. In order to test the null hypothesis that a current set of autocorrelations is white noise, test statistics are calculated for different total numbers of successive lagged autocorrelations using the Ljung-Box-Pierce statistics (Q(r) test) to test the adequacy of the model. Q(r) values are compared to a critical test value (χ^2) distribution with respective degree of freedom at a 5% significant level. The third one is the cumulative periodogram, employed to diagnose the residuals for a white noise sequence. When modeling seasonal time series line the one in the present study, the periodic characteristics of seasonal rainfall time series might not be taken into account, therefore, the periodicities in the residuals should be investigated (El-Din and Smith, 2002).

2.3. Neural network model

An ANN is a massively parallel-distributed processor that has a natural propensity for storing the experimental knowledge and making it available for further use. It resembles the human brain whose speed and efficiency has been always fascinating to researchers for quite a long time. The quest to understand these processes and to solve the associated problems has led to the development of ANN technique. Neural networks essentially involve a nonlinear modeling approach that provides a fairly accurate universal approximation to any function. Its power comes from the parallel processing of the information from data. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data. Single hidden layer feedforward network is the most widely used model form for time series modeling and forecasting. The backpropagation network (BPN) is one of the neural network algorithm which is formalized by Parker, (1986), Lippmann (1987) and Rummelhart

& McClelland (1986) etc. It has been extensively used for inversion, prediction. An example of a network topology is shown in Figure 1.

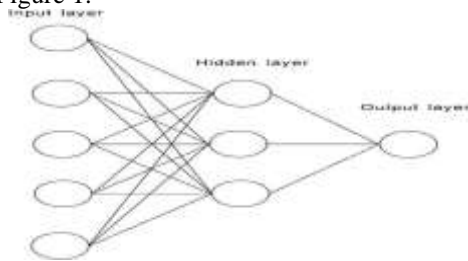


Figure 1. An example of an artificial neural network topology with one input layer, one hidden layer and one output layer

A neural network must be trained to determine the values of the weights that will produce the correct outputs. In a training step, a set of input data is used for training and presented to the network many times. The performance of the network is tested after the training step is stopped. The backpropagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient). This is the direction in which the performance function is decreasing most rapidly. It turns out that although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. Therefore, the basic gradient descent training algorithm is inefficient owing to its slow convergent speed and at times the poor accuracy in model predictions (Huang et al., 2004). From an optimization point of view, training a neural network can be considered as equivalent to minimizing a multivariable global error function of the network weights. There are several optimized training algorithms, as described by Haykin (1999), such as resilient backpropagation, Levenberg–Marquardt and conjugated gradient backpropagation. One of the optimized methods developed by Moller (1993) is the scaled conjugate gradient (SCG) algorithm. The SCG training algorithm was developed to avoid the time-consuming line search. In the conjugate gradient algorithm a search is performed along conjugate directions, which produces faster convergence than steepest descent directions. The standard backpropagation algorithm, traditionally employed in neural network learning, evaluates the gradient of the global error function with respect to the weights, $f(W^k)$, at each iteration k , and updates the weights according to

$$W^{k+1} = W^k - \alpha^k \nabla_f(W^k) \tag{4}$$

The step size $\alpha^k > 0$ is a user-selected learning rate parameter, which affects the performance of the learning algorithm to a great extent. In all cases, the backpropagation algorithm may follow a zigzag path to the minimum, typical for a steepest gradient descent method (Falas and Stafylopatis, 2005). A conjugate gradient algorithm avoids the zigzag approach to the minimum point by incorporating a special relationship between the direction and gradient vector at each iteration. If D^k represents the direction vector at iteration k of the algorithm, then the weight vector is updated according to the rule

$$W^{k+1} = W^k + \alpha^k D^k \tag{5}$$

Given values of W^k and D^k , a particular values of α^k that reduce the objective function as much as possible needs to be found. After a small number of iterations, the search along the line direction to find the optimum step size for the actual minimum should stop. Estimating the optimum step size with

scaled conjugate gradient(SCG) training algorithm increases the learning speed and eliminates the dependence on critical user-selected parameters. The main idea behind the algorithm is the use of a factor ρ which is raised or lowered with in each iteration during the execution of the algorithm, looking at the sign of the quantity δ^k , which reveals if the Hessian matrix is not positive definite. A brief algorithm of SCG in neural network is given as follows (Falas and Stafylopatis, 2005).

1. Initialization: At $k=0$, choose an initial weight vector W^0 , and set the initial direction vector to the negative gradient vector $D^0 = G^0 = -\nabla_f(W^0)$.

Set the scalars $0 < \delta < 10^{-4}, 0 < \rho^0 < 10^{-6}, \rho^{-0}$, set the boolean success=true.

2. If success=true, then calculate second order information:

$$\sigma^k = \frac{\sigma}{|D^k|}, S^k = \frac{(\nabla_f(W^k + \sigma^k D^k) - \nabla_f(W^k))}{\sigma^k}, \delta^k = \text{transpose}(D^k) + S^k$$

3. Scale $\delta^k: \delta^k = \delta^k + (\rho^k - \rho^{-k})|D^k|^2$, look at the sign of δ^k for each iteration adjusting ρ^k . If $\delta^k \leq 0$ then ρ^k and S^k is estimated again.

4. If $\delta^k \leq 0$ then make the Hessian positive definite

$$\rho^{-k} = 2 \left(\rho^k - \frac{\delta^k}{|D^k|^2} \right), \quad \delta^k = -\delta^k + \rho^k |D^k|^2, \quad \rho^k = \rho^{-k}$$

5. Calculate the step size: $\xi^k = \text{transpose}(D^k)G^k, \alpha^k = \frac{\xi^k}{\delta^k}$

the values of ρ^k directly scale the step size in the way, that the bigger ρ^k , the smaller the step size.

6. Calculate the comparison parameter

$$c^k: c^k = \frac{2\delta^k [f(W^k) - f(W^k + \alpha^k D^k)]}{(\xi^k)^2}$$

7. Weight and direction update: If $c^k \geq 0$, then a successful update can be made:

$$W^{k+1} = W^k + \alpha^k D^k, G^{k+1} = -\nabla f(W^{k+1}), \rho^{-k} = 0, \text{ success} = \text{true.}$$

If $k \bmod N=0$ then restart algorithm with $D^{k+1} = G^{k+1}$ else $\beta^k = \frac{(|G^{k+1}|^2 - G^{k+1T}G^k)}{\xi^k}, D^{k+1} = G^{k+1} + \beta^k D^k$. If $c^k \geq 0.75$

then reduce the scale parameter to $\rho^k = \frac{1}{4}\rho^k$ else $\rho^{-k} = \rho^k, \text{ success} = \text{false}$.

8. If $c^k \geq 0.25$ then increase the scale parameter to

$$\rho^k = \rho^k + \frac{\delta^k(1 - c^k)}{|D^k|^2}$$

9. Repetition: If the steepest descent direction $G^k \neq 0$; set $k = k+1$ and go back to step 2 else terminate and return W^{k+1} as the desired minimum.

2.5. Model verification and comparison methods

Three different forecast consistency measures are used in order to compare the performances of obtained ARIMA and artificial neural network (ANN): root mean square error (RMSE), the mean absolute percentage error (MAE) and the correlation coefficient (r).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - Q_i)^2} \tag{6}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - Q_i| \tag{7}$$

$$r = \sqrt{1 - \frac{\sum_{i=1}^n (P_i - Q_i)^2}{\sum_{i=1}^n Q_i^2 - \frac{\sum_{i=1}^n P_i^2}{n}}} \tag{8}$$

Here P_i and O_i are the predicted and observed values, respectively. And n is the total number of observations.

3. Results and discussion

3.1. ARIMA modeling

In the present study, several trials were made to choose the optimal ARIMA model parameters. The model parameters that satisfy the statistical residual diagnostic checking were chosen in the ARIMA forecasting model. The ARIMA models were used to predict seasonal rainfall time series over the period between 1976 and 2008. The seasonal rainfall data for the period between 1976 and 2003 were used for model calibration and to obtain the best model fit for each station. The data for the period between 2004 and 2008 were used for model verification and comparisons for prediction purposes. In the ARIMA modeling process, the input and output seasonal rainfall data sets were normalized to the range of $[0, 1]$. To fit ARIMA model to the available seasonal rainfall time series data, three-stage procedure of model identification, estimation of model parameters and diagnostic checking of the estimated parameters was employed. In the identification stage, to determine the possible persistence structure in the time series data, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) were used (Figs. 2, 3, 4, 5, 6, 7). Using Akaike Information Criteria (AIC), the best fitted model has been identified out of the various competing models. As demonstrated in Table1, the seasonal components (P, D, Q) of best fit ARIMA models are (2,1,0) for seasonal rainfall of station Ahvaz, (0,1,1) for station Abadan and (0,1,1) for station Dezful. The nonseasonal components (p, d, q) are (0,1,2), (0,1,2) and (0,1,1) for stations Ahvaz, Abadan and Dezful, respectively. The AICs for the best fit ARIMA models are -81.5, -35.22 and -51.9 for station Ahvaz, station Abadan and station Dezful, respectively (Table1). In the estimation of model parameters stage, the best fit ARIMA model statistical parameters were estimated. The computational method outlined by Box and Jenkins (1976) was employed to estimate model parameters. In the diagnostic checking of the estimated parameters stage, diagnostic checks were done to insure that the best fit model was selected by checking that assumptions of ARIMA modeling such as independence, homoscedastic (constant variance) and normality of the residual a_t were satisfied. In order to check the independence of residuals, the residual autocorrelation function (RACF), Ljung-Box-Pierce statistics and cumulative periodograms were used. The values of residual autocorrelation functions (RACF) were well settled within confidence limits except very few individual correlations appear large compared to the confidence limits, which were acceptable among 48 lags. The results exhibited no significant correlation between the residuals of the each seasonal rainfall. The values of Ljung-Box-Pierce $Q(r)$ statistic is shown in Table1 and has a value of 26.53, 29.45 and 22.6 for seasonal rainfall of stations Ahvaz, Abadan and Dezful, respectively. The values of $Q(r)$ were compared to a critical test value (χ^2) distribution with respective degree of freedom at a 5% significant level. It was obvious that the computed values were less than the actual (χ^2) values, which indicated that the residuals from the best models were white noise (Table1). Cumulative periodograms confirmed that no significant periodicity was available in the residual series at 95% confidence level and indicated that the points were clustering closely about the theoretical line and there was no evidence of periodic characteristics buried in the residual series. In order to check the normality of residuals, the histograms and normal

probability plot of residuals were investigated and they clearly supported the assumption of normality. In order to investigate homoscedasticity of the residuals, a plot of residuals versus fitted values were examined and the plots showed a random scatter around zero. In other words, the residuals were evenly distributed around mean, which explains the models were adequate.

The trained ARIMA model was then tested using season rainfall data set for the period of 20 seasonals. As shown in Figs. 8, 9, 10, although the ARIMA models generally vary with the range of most of the seasonal rainfall data, the model predictions are not quite satisfied. The correlation coefficient values between models predicted values and observed data for stations Ahvaz, Abadan and Dezful are 0.874, 0.852 and 0.935, respectively, which are not satisfactory in common model applications (Figs. 11, 12, 13). Although the ARIMA models were able to show the cycles of the high and low seasonal rainfall values, they were not able to provide good predictions of the seasonal rainfall value magnitudes, which changed from seasonal to seasonal. This limitation was due mainly to the limitations of the linear modeling algorithm in the ARIMA model, the performance of which was generally not quite satisfactory in recognizing and reproducing the nonlinear time series of seasonal rainfall data.

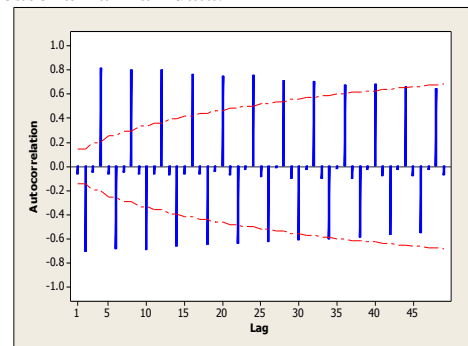


Fig.2. Plot of ACF for seasonal rainfall data for station Ahvaz.

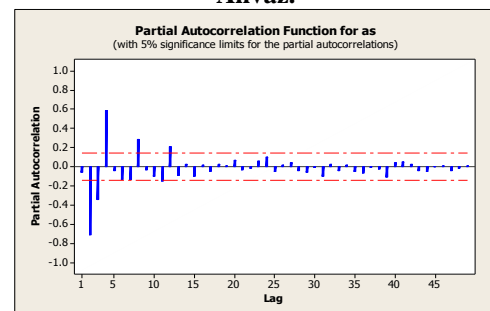


Fig.3. Plot of PACF for seasonal rainfall data for station Ahvaz.

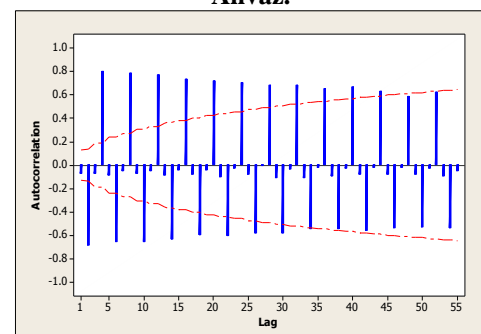


Fig.4. Plot of ACF for seasonal rainfall data for station Abadan

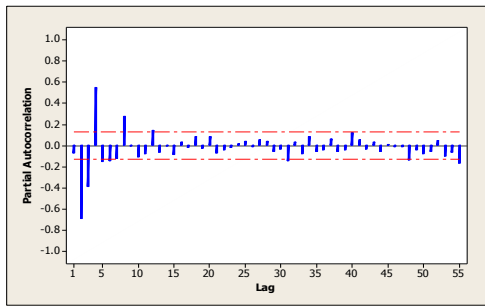


Fig.5. Plot of PACF for seasonal rainfall data for station Abadan

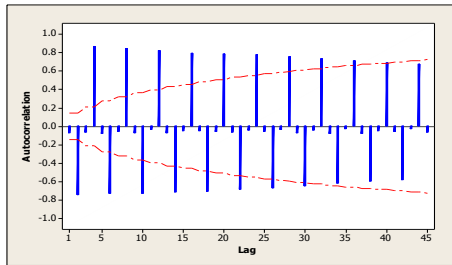


Fig.6. Plot of ACF for seasonal rainfall data for station Dezful.

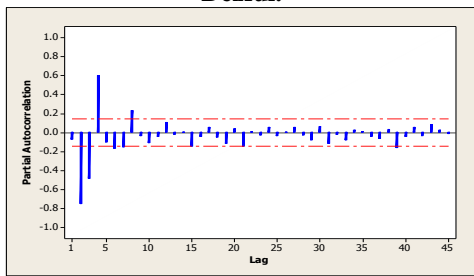


Fig.7. Plot of PACF for seasonal rainfall data for station Dezful.

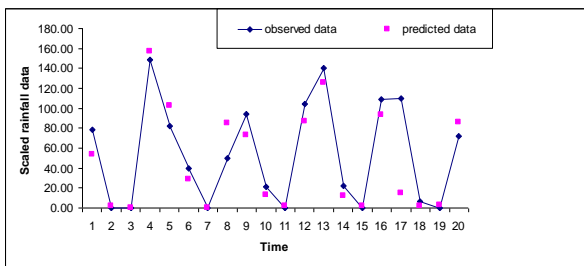


Fig. 8. ARIMA model verification for station Ahvaz

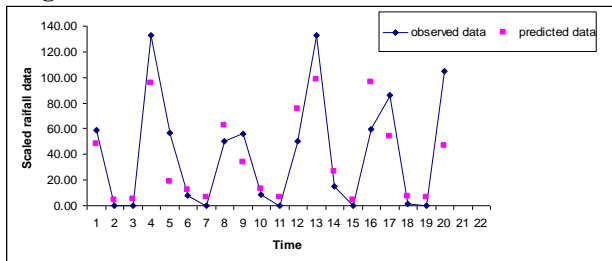


Fig. 9. ARIMA model verification for station Abadan

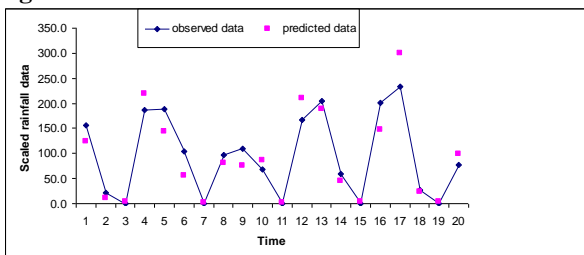


Fig. 10. ARIMA model verification for station Dezful

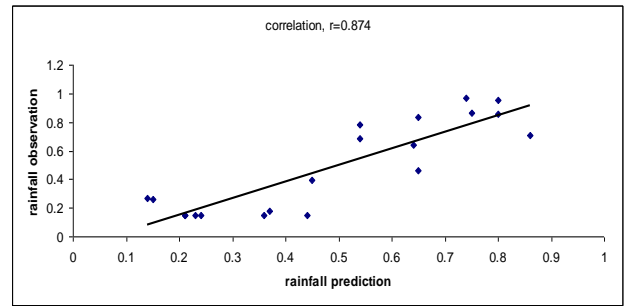


Fig. 11. Observed versus ARIMA predicted data for station Ahvaz

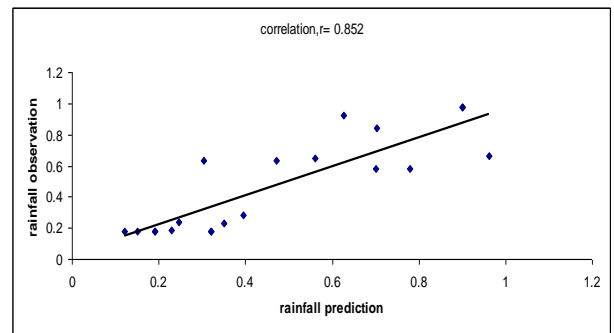


Fig. 12. Observed versus ARIMA predicted data for station Abadan

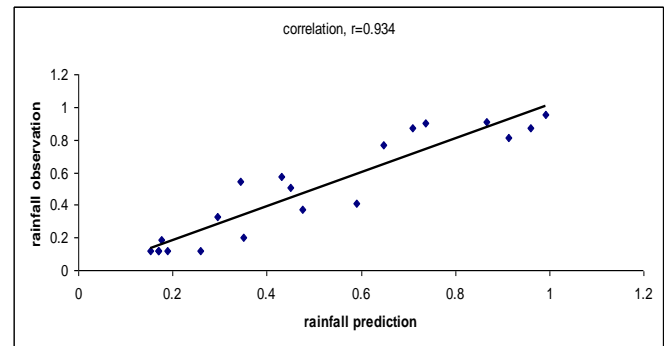


Fig. 13. Observed versus ARIMA predicted data for station Dezful

3.2. Neural network modeling

A three-layer feedforward neural network model was developed for the prediction of seasonal rainfall using an optimized back-propagation training algorithm. In the present study, the scaled conjugated gradient algorithm was selected as the optimized training method. In the following part, artificial neural network model performances were validated for rainfall prediction under seasonally time-step condition. The data for the period between 1976 and 2008 were available for the modeling purposes. Seasonal rainfall time series data were divided into two independent data sets. The first data set was used for model training, and the second data set was used for model verification purposes. In the ANN modeling process, the input and output seasonal rainfall data sets for each station were normalized to the range of [0,1]. Figs. 14, 15, 16 compares the model predictions for seasonal rainfall with the observations. The verifications stage indicate that the model prediction results reasonably match the observed seasonal rainfall. The correlation coefficient between the ANN model predicted values and observed data for stations Ahvaz, Abadan and Dezful are 0.949, 0.922 and 0.945, respectively, which are satisfactory in common model applications (Figs. 17, 18, 19). These results indicate that the neural network model is able to recognize the pattern of the seasonal rainfall to provide good predictions of the seasonally

variations of seasonal rainfall data of the Khozestan Province (Iran).

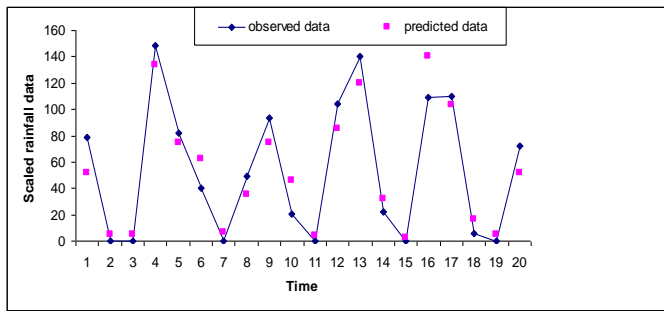


Fig. 14. ANN model verification for station Ahvaz

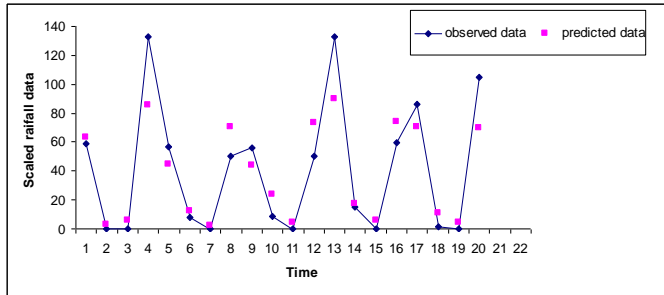


Fig. 15. ANN model verification for station Abadan

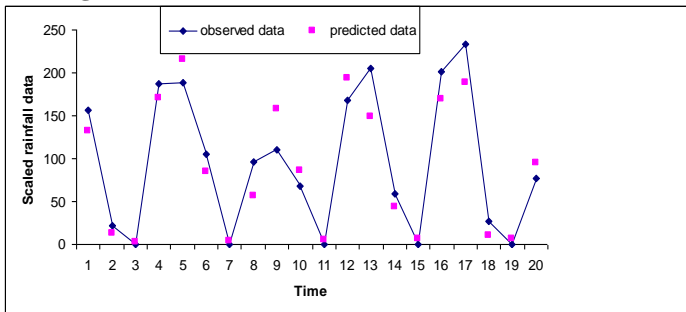


Fig. 16. ANN model verification for station Dezful

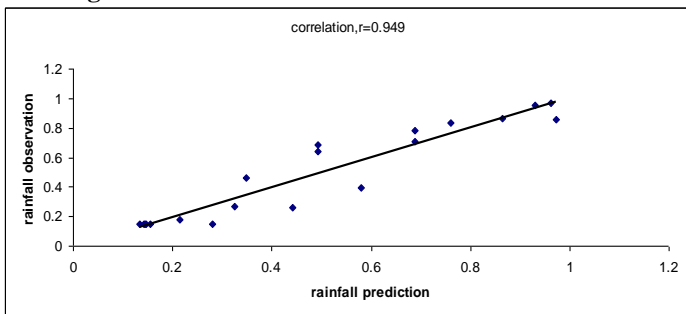


Fig. 17. Observed versus ANN predicted data for station Ahvaz.

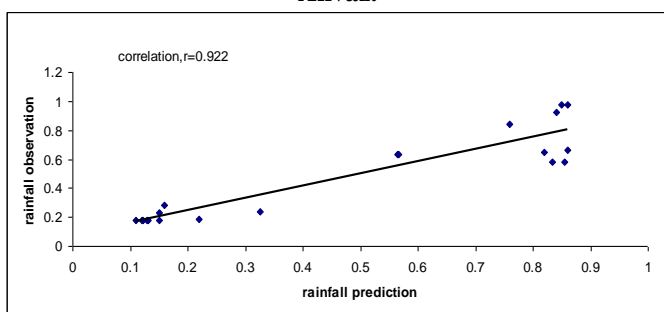


Fig. 18. Observed versus ANN predicted data for station Abadan.

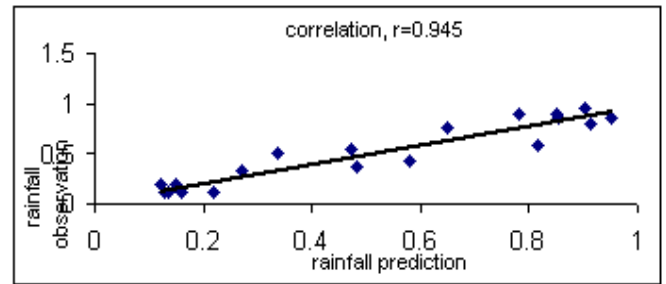


Fig. 19. Observed versus ANN predicted data for station Dezful

3.4. Comparison of model performances

Figures. 20, 21, 22 shows the plot of predicted models vs. observed values of the seasonal rainfall data from year 1976 to 2003 by both the ANN and ARIMA. The ANN model fits extremely well with the actual data values as compared to the ARIMA model. Both the models were tested using the test data set for the period 2004 to 2008, which is shown in Figs.23, 24, 25. From this figure it can be observed that the seasonal rainfall values predicted by the ANN model are quite closer to observed seasonal rainfall as compared to the ARIMA model. Employing accuracy measures (RMSE, MAE and R^2), the observed data and predicted data from the ANN and ARIMA models were compared to determine the best performed model. The predicted seasonal rainfall using the ARIMA models were not found to be in reasonable agreement with the observed data. However, the ANN approach provided reasonable precision for all stations. Tables 2, 3, 4 gives the error estimates of the three different approaches used in the study for predicting seasonal rainfall. The RMSEs between observed and predicted data were calculated in ARIMA models as 25.66, 23.97 and 30.44 for stations Ahvaz, Abadan and Dezful, respectively. In the case of ANN modeling approach, the RMSEs between observed and predicted data were computed as 16.13, 19.31 and 26.53 for stations Ahvaz, Abadan and Dezful, respectively. Furthermore, the MAEs between observed and predicted data for stations Ahvaz, Abadan and Dezful were appeared to be slightly lower for the ANN modeling approach. Prediction error statistics for the ANN approach produced MAEs of 37.63, 33.15 and 35.11 for stations Ahvaz, Abadan and Dezful, respectively. These results indicated that the ANN model performed well for adequate predicting of seasonal rainfall. Therefore, it can be concluded that the ANN modeling approach can give more reliable predictions of seasonal rainfall time series of Khozestan Province (Iran) than the ARIMA modeling approach.

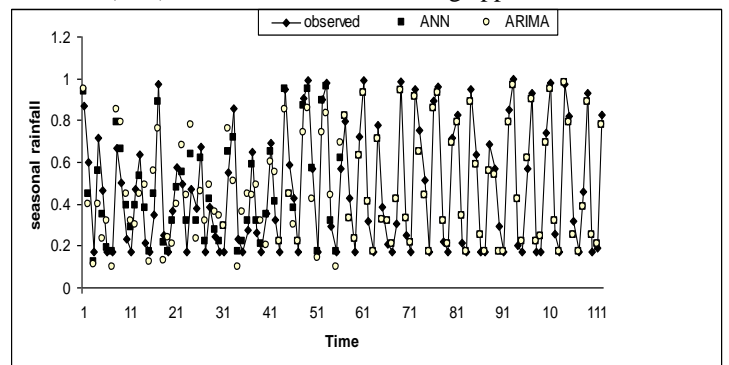


Fig.20. A linear scale plot of the predicted and observed seasonal rainfall for model data set using ANN and ARIMA model for station Ahvaz.

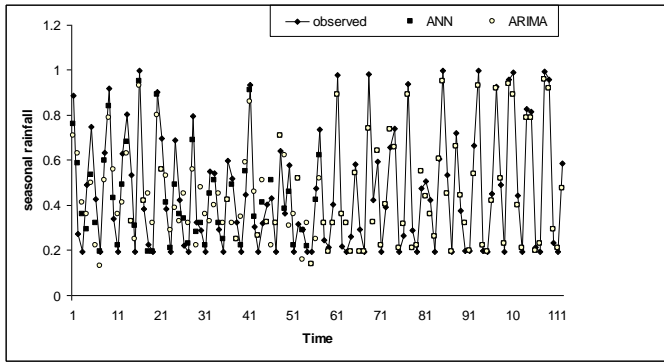


Fig.21. A linear scale plot of the predicted and observed seasonal rainfall for model data set using ANN and ARIMA model for station Abadan.

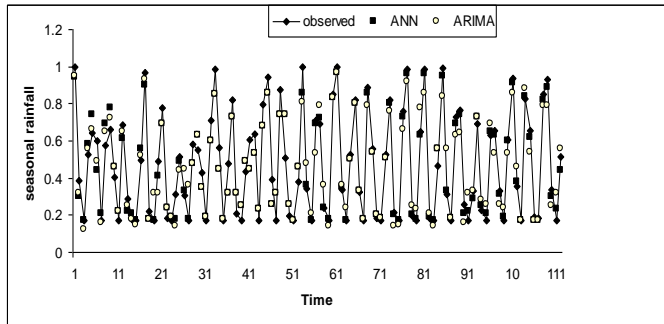


Fig. 22. A linear scale plot of the predicted and observed seasonal rainfall for model data set using ANN and ARIMA model for station Dezful

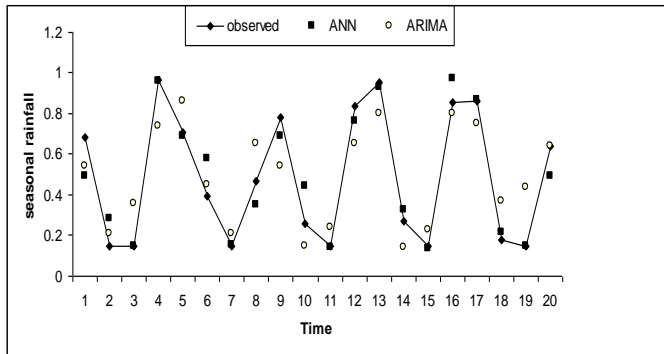


Fig. 23. A linear scale plot of the predicted and observed seasonal rainfall for test data (1976 -2003) set using ANN and ARIMA model for station Ahvaz.

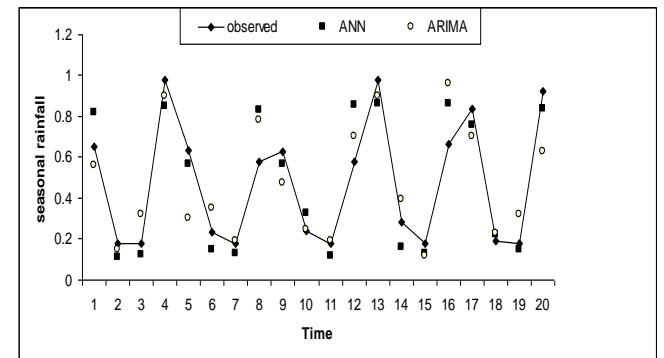


Fig. 24. A linear scale plot of the predicted and observed seasonal rainfall for test data (1976 -2003) set using ANN and ARIMA model for station Abadan.

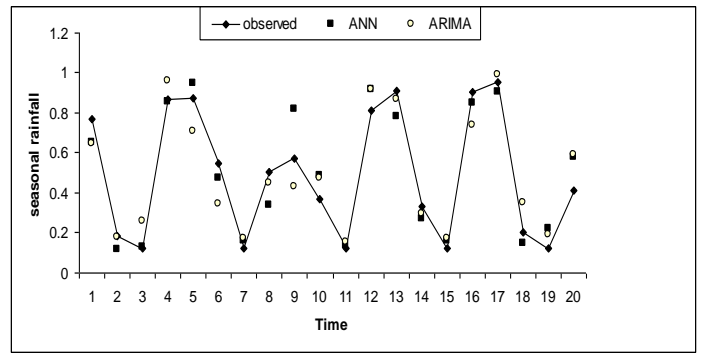


Fig.25. A linear scale plot of the predicted and observed seasonal rainfall for test data (1976 -2003) set using ANN and ARIMA model for station Dezful.

Table 2- Statistical comparison of observed and predicted data from the ANN and ARIMA modeling approaches for station Ahvaz.

Techniques	Error Measures for seasonal rainfall model data set			Error Measures for seasonal rainfall test data set		
	RMSE	MAE	R ²	RMSE	MAE	R ²
ANN	0.62	0.94	0.935	16.13	37.63	0.902
ARIMA	3.66	21.16	0.791	25.66	45.32	0.7648

Table 3- Statistical comparison of observed and predicted data from the ANN and ARIMA modeling approaches for station Abadan

Techniques	Error Measures for seasonal rainfall model data set			Error Measures for seasonal rainfall test data set		
	RMSE	MAE	R ²	RMSE	MAE	R ²
ANN	0.74	0.83	0.892	19.31	37.15	0.8502
ARIMA	3.42	21.3	0.751	23.97	40.12	0.7261

Table 4- Statistical comparison of observed and predicted data from the ANN and ARIMA modeling approaches for station Dezful.

Techniques	Error Measures for seasonal rainfall model data set			Error Measures for seasonal rainfall test data set		
	RMSE	MAE	R ²	RMSE	MAE	R ²
ANN	0.98	0.87	0.943	26.53	35.11	0.8942
ARIMA	3.8	19.2	0.891	30.44	42.14	0.8739

4. Conclusions

An empirical comparative evaluation of the performance of ANN model to the ARIMA modeling approach was presented for seasonal rainfall predictions. Investigations were conducted to examine the ANN model performance for predicting rainfall in seasonally time steps. The results from the ARIMA models poorly represented the pattern of seasonal rainfall data for stations Ahvaz and Abadan, but the model produced acceptable results for station Dezful. The results from the ANN model indicated that the modeling approach gave more reliable predictions of seasonal rainfall time series data. The predictions from ANN model were compared with that obtained from the ARIMA traditional time series approaches. Owing to its ability in recognizing time series patterns and nonlinear characteristics, the accuracy measures RMSE and MAE and R² demonstrated that the ANN model provided much better accuracy over the ARIMA methods for seasonal rainfall predictions. Therefore, the proposed ANN algorithm can be used for the Khozestan Province.

Table 1- Summary of the statistical parameters of the best fitted multiplicative ARIMA models fitted to seasonal rainfall

AIC	λ^2	Q	Θ_2	Θ_1	Φ_2	Φ_1	θ_2	θ_1	ϕ_2	ϕ_1	Model	Station
-81.5	32.67	26.53	-	-	-0.8	-0.83	-0.1394	0.9347	-	-	(0,1,2)(2,1,0)12	Ahvaz
-35.22	32.67	29.45	-	0.5576	-	-	-0.2535	1.065	-	-	(0,1,2)(0,1,1)12	Abadan
-51.9	32.67	22.6	-	0.7138	-	-	0.4643	0.1883	-	-0.73	(1,1,2)(0,1,1)12	Dezful

References

- 1- Akaike Hirotugu., "A new look at the statistical model identification", IEEE Transactions on Automatic Control 19(6): 716-723, 1974.
- 2-Bandyopadhyay, G., Chattopadhyay, S., 2007. Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone. International Journal of Environmental Science and Technology 4, 141-149.
- 3-Chattopadhyay, S., 2007b. Feed forward Artificial Neural Network model to predict the average summer-monsoon rainfall in India. Acta Geophysica 55, 369-382.
- 4-Chattopadhyay, G., Chattopadhyay, S., 2008a. A probe into the chaotic nature of total ozone time series by Correlation dimension method. Soft Computing 12, 1007-1012.
- 5- Chattopadhyay, S, Rajni Jain and Goutami Chattopadhyay. 2009. Estimating potential evapotranspiration from limited weather data over Gangetic West Bengal, India: a neurocomputing approach, Meteorological Applications, 16, 403-411.
- 6-DelSole, T., Shukla, J., 2002. Linear prediction of Indian Monsoon rainfall. Journal of Climate 15, 3645-3658.
- 7-El-Din, A.G., Smith, D.W., 2002. A combined transfer-function noise model to predict dynamic behavior of a full-scale primary sedimentation tank. Water Research 36, 3747-3764.
- 8-Elsner, J., Tsonis, A., 1992. Nonlinear Prediction, Chaos, and Noise. Bulletin of American Meteorological Society 73, 49-60.
- 9-Falás, T., Stafylopatis, A., 2005. Implementing temporal difference learning with the scaled conjugated gradient algorithm. Neural Processing Letters 22, 361-375.
- 10-Grimaldi, S., Serinaldi, F., Tallorini, C., 2005. Multivariate linear parametric models applied to daily rainfall time series. Advances in Geosciences 2, 87-92.
- 11-Guhathakurata, P., 2008. Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network. Meteorology and Atmospheric Physics 101, 93-108.
- 12-Haykin, S., 1999. Neural Networks, a Comprehensive Foundation. Prentice Hall, Upper Saddle River, NJ, USA.
- 13-Huang, W., Bing, X., Chan-Hilton, A., 2004. Forecasting flows in Apalachicola River using neural networks. Hydrological Process 18, 2545-2564.
- 14-Islam, M.N., Sivakumar, B., 2002. Characterization and prediction of runoff dynamics: a nonlinear dynamical view. Advances in Water Resources 25, 179-190.
- 15-Jayawardena, A.W., Lai, F., 1994. Analysis and prediction of chaos in rainfall and stream flow time series. Journal of Hydrology 153, 23-52.
- 16-Jhajharia, D., Shrivastava, S.K., Sarkar, D., Sarkar, S., 2009. Temporal characteristics of pan evaporation trends under the humid conditions of northeast India. Agricultural and Forest Meteorology 149, 763-770.
- 17-Lippmann R.P., 1997. An Introduction to computing with neural nets. IEEE ASSP Magazine April 1987, pp 4-20.
- 18-Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. Stochastic Environmental Research Risk Assessment 19, 326-339.
- 19-Mohanty, U.C., Mohapatra, M., 2007. Prediction of occurrence and quantity of daily summer monsoon precipitation over Orissa (India). Meteorological Applications 14, 95-103.
- 20-Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6, 525-533.
- 21-Parker D.B., 1986. A comparison of algorithm for neuronlike cell in J. S. Denker (e.d.) AIP Conference Proceedings. 151 Neural networks for Computing, Snowbird Utah, AIP
- 22-Ramirez, M.C.V, de Campos Velho, H.F., Ferreira, N.J., 2005. Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. Journal of Hydrology 301, 146-162.
- 23-Rangno, A. L., and P. V. Hobbs, 1994: Ice particle concentrations and precipitation development in small continental cumuliform clouds. Quart. J. Roy. Meteor. Soc., 120, 573-601.
- 24-Somvanishi, V.K., Pandey, O.P., Agrawal, P.K., Kalanker, N.V., Ravi Prakash, M., and R. Chand, 2006. Modeling and prediction of rainfall using artificial neural network and ARIMA techniques, J. Ind. Geophys. Union. 10(2), 141-151.
- 25-Singh, C.V., 1998. Long term estimation of monsoon rainfall using stochastic models. Int. J. Climatol. 18 (14), 1611-1624.
- 26-Soltani, S., Modarres, R., Eslamian, S.S., 2007. The use of time series modelling for the determination of rainfall climates of Iran. Int. J. Climatol. 27 (6), 819-829.
- 27-Toth, E., Brath, A., Montanari, V., 2000. Comparison of short-term rainfall prediction models for real-time flood forecasting. J. Hydrol. 239, 132-147.
- 28-Tzani, C., Varotsos, C.A., 2008. Tropospheric aerosol forcing of climate: a case study for the greater area of Greece. International Journal of Remote Sensing 29, 2507-2517.
- 29-Vandaele, W. 1983. Applied Time Series and Box-Jenkins Models. New York: Academic Press.
- 30-Wei, W.W.S., 1990. Time series analysis. Addison-Wesley Publishing Company, Inc., New York.