The Modeling and Forecasting of Annual Precipitation in Iran using ARIMA Method

Ali Bayat
Ph. D Candidate of Climatology, Kharazmi University, Tehran, Iran.

ABSTRACT

The rainfall is a climatic element whose level is continuous in places, but discrete in time. Among the climatic events, the rainfall is important due to its crucial role. The rainfall is one of the key variables to assess the potential availability of water resources in Iran, but its temporal and spatial distribution is very uneven; therefore, the distribution of water resources is not uniform. The statistical methods are useful tools for understanding and evaluating the behavior of climate, especially rainfall. From the most widely used statistical models, the ARIMA family models are used widely in the modeling and prediction of climatic elements such as rainfall. In this study, the total annual rainfall’s time series of 26 synoptic stations of Iran was modeled and forecasted using ARIMA method. The polynomial models were used to model the processes and ARIMA models were used for prediction. The results showed that during the study period, the rainfall in Iran experienced two drop and constant behaviors in different areas. Also according to the selected ARIMA models for each time series, it was found that the past behaviors of Iran’s rainfall time series were a function of rainfall in the past 1 to 5 years and the random components of past 1 to 3 years. Also, the indicators of MSE and AIC were used for performance evaluation of selected ARIMA models for and diagnostic evaluation of remaining data. The rainfall level of next 5 years and 10 years was predicted using optimized ARIMA models.

INTRODUCTION

The rainfall is of particular importance due to creating surface flows, its impact on underground aquifers, and as the main source of springs’ nutrition. It is one of the climatic elements which may be effective in determining the distribution and roles of other elements. The rainfall is one of the key variables to assess the potential availability of water resources in Iran, but its temporal and spatial distribution is very uneven; therefore, the distribution of water resources is not uniform. The maintenance and management of water resources is a function of received rainfall and also depends on the variability of rainfall level. The smaller the spatial variability of rainfall, the more water sources will be homogeneous. On the other hand, when time variability of rainfall is less, the water resources will be more stable and permanent water supply will be possible. The medium and long-term hydrological forecasts seem necessary for water projects design, planning the water supply, use of water energy, irrigation system, water quality management, and so on. Several methods and models are provided to assess the medium-term and long-term hydrological and climatic forecasts. The most important of these models are random models [23], the theory of fuzzy sets [14], artificial neural network models and ARIMA models [8]. Among these models, the ARIMA model is based on random theory and is used for hydrological predictions and display random components of hydrological data [15,29]. ARIMA model has a systematic method for each model at each stage (determination, assessment, and diagnosis) [10]. However, the ARIMA model assumes a linear correlation structure between the values of time series observations [30]. The climate models that are based on statistical-probability principles have special importance and numerous applications. A three-step strategy is used in this type of modeling includes recognition (identification), fitting and testing the validity of the model, and prediction with some specified certainty [3]. The ARIMA is one of the suitable models which fit the climatic processes such as rainfall.
Along with the development and spread of computer models and tools, the statistical modeling has been developed using advanced statistical methods based on the long-term characteristics of phenomenon, elements, and major climatic factors. These types of studies are conducted worldwide. The ARIMA model was originally introduced by Box and Jenkins [8]. Some of the authors have used ARIMA model alone and some others have used it simultaneously with several other models such as artificial neural network and simple spectrum analysis. For example, Mohan and Vedula [21] applied ARIMA models to model rainfall in India. El-Fandy et al. [13] used ARIMA model for modeling rainfall in Ethiopia and flooding in the Nile. Nickerson and Madsen [22] used ARIMA model and nonlinear regression for modeling the chemical composition of rainfall in East Central Florida. Kumar et al. [18] modeled the surface flow of humid subtropical areas using ARIMA. Chattopadhyay and Chattopadhyay [11] modeled the summer seasonal rainfall’s time series using ARIMA model and artificial neural network model; they compared the results of this modeling. In Iran, few studies have been conducted to model the rainfall using ARIMA and most of the studies are related to the temperature element of stations as case studies. The studies on rainfall in Iran include: Sharifan and Gahra man [24] predicted 10-day and monthly rainfall levels at some stations in Golestan province using seasonal ARIMA (SARIMA) and time series models. Then, based on the best statistical distribution, they divided the predicted amount of monthly rainfall to 10 days rainfalls and calculated the annual rainfall by summing the 10 days’ rainfall level per year. Bayat [6] studied the time series of annual rainfall in Zanjan and modeled and predicted the next 10 years rainfall level using ARIMA model. The results showed that the annual rainfall level in Zanjan would decrease over the next decade. Due to the lack of a comprehensive study in this area and the need to conduct such a study, the annual rainfall in Iran was modeled using ARIMA model. In this study, the total annual rainfall’s time series in 26 synoptic stations in Iran will be modeled using ARIMA model and the next 5 years and 10 years rainfall level in Iran will be predicted.

Data:

In this study, the average annual rainfall of 26 synoptic stations in Iran which at least had 30 years full statistics until 2008 were studied and modeled. Figure 1 shows the distribution of the studied stations. As can be seen, the station densities are more in North and North West of Iran than the South and East. This distribution is proportional to the population arrangement of the cities. Depending on rainfall and water availability, the northern and western parts have high density and the East and South have low density. The standard normal homogeneity test [1] and the Von Neumann test [28] were used to test the homogeneity of time series’ average and homogeneity of variance, respectively.

In the standard normal homogeneity test, the K mean of first year is compared with the n-k mean of next year to obtain the statistic T (K); in other words [2]:

\[ T(K) = kZ_k^2 + (n-k)Z_{n-k}^2 \]

In this equation:

\[ Z_k = \frac{1}{K} \sum_{i=1}^{k} (Y_i - \bar{Y}) / S \]

\[ Z_{n-k} = \frac{1}{n-k} \sum_{i=k+1}^{n} (Y_i - \bar{Y}) / S \]

In the above equation:

\[ Y_i \] = Values of annual series (from 1 to n), \( \bar{Y} \) = the mean of series, S= Standard Deviation. The test statistic to reject the null hypothesis is defined as:

\[ T_k = \max(T(K)) = \max(kZ_k^2 + (n-k)Z_{n-k}^2) \]

1 ≤ k ≤ n-1

If \( T_k \) is larger than a certain critical level, then, the null hypothesis will be rejected at the considered significance level. The critical values which depend on the serious length and level of significance are provided in Table 1.

Von Neumann test was used to determine the homogeneity of variance which is defined as follows: [28]:

\[ N = \frac{\sum_{i=1}^{n} (Y_i - Y) / \sum_{i=1}^{n} (Y_i - Y)^2}{\sum_{i=1}^{n} (Y_i - Y)^2} \]

When the sample variance is homogeneous, the Von Neumann ratio (N) will be 2 or very close to 2. If the Von Neumann ratio is less than 2, the sample will have mutation and heterogeneous variance. The threshold values for Von Neumann ratio is presented in Appendix in Table 2. If Von Neumann ratio (N) are greater than these threshold values, the serious will have homogeneous variance. If Von Neumann ratio (N) are less than the threshold values, the data serious will have heterogeneous variance. Table 3 shows the results of homogeneity tests. Among the studied stations, the data of 11 stations in terms of mean and the data of 8 stations in terms of
variance are heterogeneous. In order to make stable the heterogeneous time series, however, specific tests were used. Then, the rainfall of each station was modeled.

![Map of Iran with stations marked](image)

**Fig. 1:** Distribution of studied stations in Iran.

<table>
<thead>
<tr>
<th>No</th>
<th>Station</th>
<th>Period</th>
<th>SNHT</th>
<th>Von Neumann Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ahwaz</td>
<td>1960-2008</td>
<td>Inhomogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>2</td>
<td>Arak</td>
<td>1956-2008</td>
<td>Homogenous</td>
<td>Homogen</td>
</tr>
<tr>
<td>3</td>
<td>Asehbi</td>
<td>1977-2008</td>
<td>Inhomogen</td>
<td>Inhomogen</td>
</tr>
<tr>
<td>4</td>
<td>Bandar Abbas</td>
<td>1957-2008</td>
<td>Homogen</td>
<td>Inhomogen</td>
</tr>
<tr>
<td>5</td>
<td>Birjand</td>
<td>1956-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>6</td>
<td>Bojnurd</td>
<td>1978-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>7</td>
<td>Bushehr</td>
<td>1951-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>8</td>
<td>Esfahan</td>
<td>1951-2008</td>
<td>Inhomogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>9</td>
<td>Gharazin</td>
<td>1959-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>10</td>
<td>Gorgan</td>
<td>1953-2008</td>
<td>Inhomogen</td>
<td>Inhomogen</td>
</tr>
<tr>
<td>11</td>
<td>Hamedan</td>
<td>1951-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>12</td>
<td>Kerman</td>
<td>1951-2008</td>
<td>Inhomogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>13</td>
<td>Kermanshah</td>
<td>1951-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>14</td>
<td>Khoram Abad</td>
<td>1951-2005</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>15</td>
<td>Mashhad</td>
<td>1951-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>16</td>
<td>Orumiyeh</td>
<td>1951-2008</td>
<td>Inhomogen</td>
<td>Inhomogen</td>
</tr>
<tr>
<td>17</td>
<td>Ramsar</td>
<td>1956-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>18</td>
<td>Sanandaj</td>
<td>1960-2008</td>
<td>Inhomogen</td>
<td>Inhomogen</td>
</tr>
<tr>
<td>19</td>
<td>Semnan</td>
<td>1965-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>20</td>
<td>Shahr Kord</td>
<td>1956-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>21</td>
<td>Shiraz</td>
<td>1951-2008</td>
<td>Inhomogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>22</td>
<td>Tabriz</td>
<td>1951-2008</td>
<td>Inhomogen</td>
<td>Inhomogen</td>
</tr>
<tr>
<td>23</td>
<td>Tehran</td>
<td>1951-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>24</td>
<td>Yazd</td>
<td>1953-2008</td>
<td>Homogen</td>
<td>Homogen</td>
</tr>
<tr>
<td>25</td>
<td>Zahedan</td>
<td>1951-2008</td>
<td>Inhomogen</td>
<td>Inhomogen</td>
</tr>
<tr>
<td>26</td>
<td>Zanjan</td>
<td>1956-2008</td>
<td>Inhomogen</td>
<td>Inhomogen</td>
</tr>
</tbody>
</table>
Methods:

In the ARIMA model, the function is formed by three parametric linear components: Auto-correlated (AR), integrated (I), and moving average (MA) (Box and Jenkins, 1970). A (p, d, q) model of ARIMA is expressed as follows [30]:

\[ y_t = \theta_0 - \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \]  

(4)

Where, \( y_t \) is a real value, \( \varepsilon_t \) is random error at time \( t \), \( \phi \) and \( \theta \) are the coefficients, and \( p \) and \( q \) are integers related to auto-correlated polynomials and moving average, respectively. Usually, it is assumed that \( \varepsilon \) have zero mean and constant variance \( \sigma^2 \).

If \( q = 0 \), the equation 4 will be an auto-receive model (AR) of order \( P \). When \( P = 0 \), the model is reduced to moving average model of order \( q \). The stability is the necessary condition for prediction using ARIMA model. When the time series has non-stability or trend in the mean and variance, the differentiation and power conversion (Box-Cox transformation) will be often used to remove the trend and stabilize the mean and variance before modeling.

The differentiation is one of the most important tools to convert non-stationary time series of average and achieve a stationary series. The difference operator is shown as \( (1-B)^d \). Where, \( d \) shows the order of differentiation. In this study, this technique has been used in some stations [5]:

\[ (1-B)x_t = x_t - x_{t-1} \]  

(5)

\[ (1-B^2)x_t = x_t - x_{t-1} - (x_{t-1} - x_{t-2}) \]  

(6)

The power conversion introduced by Cox and Box can be used to convert the variance as follows:

\[ T(X) = \left\{ \begin{array}{ll} X^\lambda - 1, & \lambda \neq 0 \\ \ln(X), & \lambda = 0 \end{array} \right. \]  

(7)

The \( \lambda \) is called conversion parameter.

The partial autocorrelation function is used to calculate the order of \( p \) and the autocorrelation function is calculated to determine the order of \( q \). Autocorrelation function which is shown by \( \rho_k \) is defined as follows [27]:

\[ \rho_k = \frac{cov(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)} \sqrt{\text{Var}(Z_{t+k})}} \]  

(8)

In this equation, \( \gamma_k \) is called auto-covariance function and \( \rho_k \) is called auto-correlation function (ACF), because they show the covariance and correlation between \( Z_t \) and \( Z_{t+k} \) of a process which are separated only by a time delay \( k \). The correlation between \( Z_t \) and \( Z_{t+i} \) after removing the common linear dependence of \( Z_{t+1} \), \( Z_{t+2} \), \ldots, and \( Z_{t+i-1} \) variables is called and partial auto-correlation function. Thus, the partial auto-correlation between \( Z_t \) and \( Z_{t+j+k} \) will be equal to routine auto-correlation between \( (Z_t - \hat{Z}_t) \) and \( (Z_{t+j+k} - \hat{Z}_{t+j+k}) \). Therefore, if \( P_k \) will be the partial auto-correlation between \( Z_t \) and \( Z_{t+j+k} \), we have

(Wei, 2007):

\[ P_k = \frac{\text{cov}(Z_t - \hat{Z}_t, Z_{t+j+k} - \hat{Z}_{t+j+k})}{\sqrt{\text{Var}(Z_t - \hat{Z}_t)} \sqrt{\text{Var}(Z_{t+j+k} - \hat{Z}_{t+j+k})}} \]  

(9)

The least squares method is used to predict \( \phi_i \)s \( (i = 1, 2, \ldots, p) \) and \( \theta_j \)s \( (j = 1, 2, \ldots, q) \). In the parameters prediction, the significance tests were used for determining the significance of parameters. The insignificance parameters would be removed from the model. The most important tests are \( p \) statistic and t-student.
The Ljung-Box statistics is used to check the adequacy of model. For example, this statistic is used to determine whether the residuals are white noise (i.e., residuals have zero mean, constant variance, mutual uncorrelation, and normal distribution). When several models have different suitable parameters, the Akaike’s Information Criterion (AIC) is used to select the best and most appropriate model [25]:

$$AIC = n \ln n(S_a^2) + 2(m)$$

(10)

Where, $S_a^2$ is the maximum likelihood estimation, $\alpha_i$ and n are serious length, and m is the number of model’s parameters.

One application and criteria for acceptance of model in ARIMA modeling is prediction of future values of the data. In other words, the last criterion for selecting a model is that it will provide suitable behavior for prediction. This characteristic was used to predict the rainfall level of next 5 years and 10 years. There are several criteria for evaluating the effectiveness of model and the most important ones are the correlation coefficient and the coefficient of determination, the mean square of errors, and optimum efficiency. In this study, the mean square of errors is used to evaluate the effectiveness of model. This index measures the difference between the values predicted by the model and the actual values of observations and is defined as follows [30]:

$$MSE = \frac{1}{n} \sum_{t=1}^{n} (A_t - F_t)^2$$

(11)

Where, $A_t$ is the values observed at time t and $F_t$ is the predicted values. The lower values of mean square of errors index indicate a better prediction.

Finally, in some sections of this paper, the Kriging method was used to interpolate annual rainfall and interpolate the predicted rainfall levels in next 5 and 10 years in order to estimate the total rainfall in Iran based on rainfall levels of 26 points.

The Kriging method is an optimal method for interpolation. The main basis for interpolation by this method is variogram calculation by adjacent points which is used to estimate the unknown values. In this way, a specific statistical weight is considered for each of the stations inside and outside the area in terms of its distance and position such that the variance of estimation will be minimal. Variogram function is defined as follows [3]:

$$\gamma(h) = \frac{1}{m(h)} \sum_{i=1}^{m(h)}[z(x_i) - z(x_i + h)]^2$$

(12)

In the above equation:

$\gamma(h)$: The variogram value for the pairs of points which are at distance h from each other.

$m(h)$: the number of pairs of points which are at distance h from each other.

$z(x_i)$: The observed variable at point x

$z(x_i + h)$: The observed variable of the variable which are at distance h from x.

Discussion and Analysis:

General characteristics of rainfall in Iran

Iran is a rugged country and its average height is about 1,250 meters above sea level [20]. There are two mountain ranges in Iran which are its main climatic actors. Alborz Mountains form a continuous strip in north of Iran stretching from East to West and separate Iran’s north strip from central parts. Also, Zagros Mountains stretch from North-West to South-East of Iran. The topography of Iran and its total annual rainfall is presented in Figure 2. The general trend of decrease in rainfall is from north to south and from West to East. Also, the density of isohyet lines can be seen parallel to the mountains in the West, especially in much areas of the north. For example, the rainfall level in the north at Ramsar station in altitude of 20 meters is about 1224 mm, in the south at Bushehr station in altitude of 20 meters is 272 mm, in the west at Khoram Abad station in altitude of 1148 meters is about 508 mm, and in the east at Birjand station in altitude of 1491 meters is about 167 mm. In terms of level and timing of rainfall, Iran can be divided into four parts: the very low rainfall, low rainfall, high rainfall level, and very high rainfall. The parts with very low rainfall level and low rainfall level account for 83% of the country’s total area and their average rainfall level is about 188 mm. Mainly the South, East, Center, and parts of the North West are included in this territory. The parts with high rainfall level and very high rainfall level account for 17% of the country’s total area and their average rainfall level is about 570 mm. In this division, the average rainfall level in Iran is calculated about 251 mm. Compared with the global average, this region has very low rainfall level [19]. The spatial variation in Iran’s rainfall is very high. On the one hand, these differences are due to the nature of rainfall spatial behavior which is basically a rogue variable and shows strong spatial variations. On the other hand, the variation of rainfall source in different parts of Iran has caused the level of rainfalls and their time to vary. As can be seen in Figure 2, the rainfall is reduced from North to
South and from West to East. The roughness of Zagros Mountains stretching in the western border of Iran from North West to South East dynamically influence western and southwestern systems which enter Iran and force them to climb and strengthen them. That is why the West benefits from high level of rainfall than East and center. The lack of rain in the East and center of Iran is on one hand due to the high-pressure rule of subtropical area in the warm period of year and on the other hand due to being located in the rain shadow of Zagros mountain range which prevents impact of Western systems on this part of Iran. In the northern strip of Iran and the Caspian Sea border, the very cold and dry winds move over the warm waters of the Caspian and get volatile thanks to rising humidity and heat and lead to heavy rains in this region [19]. The figure 3 shows the relationship between rainfall and altitude of stations. As indicated in the figure, the rainfall level is increased at some stations with increasing altitude. However, this principle is not true everywhere. In some stations, sometimes, a reverse situation can be observed: decline in rainfall level with increasing altitude. For example, in Gorgan and Ramsar stations which are located in northern Iran and have the highest amount of rainfall level in Iran, the altitude is zero and even lower than the sea level. Moving from the north of Iran to Alborz mountain range and with increasing altitude, the level of rainfall decreases, while in the West and North West of Iran, the amount of annual rainfall level has increased with increasing altitude. So, it can be said that there is no unique relationship between received rainfall level in a location and its height from the sea level. The rainfall level increases in the windward slopes of Zagros (Western Range) with increasing altitude, however, the rainfall decreases with increasing altitude in the coast of Caspian Sea and large parts of northern slopes of Alborz. This contradiction shows that the model of rainfall lines in Iran is more substantially a function of roughness arrangement than height [20].

Fig. 2: Interpolation map of annual rainfall and topography in Iran.

Modeling of rainfall trend:

In order to provide a picture of rainfall’ time series behavior in the studied stations in Iran and estimate the change, all of the individual stations were exposed to modeling in polynomials family to determine the type of trend and slope degree of trend line. In modeling in polynomials family, it is determined which one of the linear, parabolic, or other model is the most elegant model to display the trend. A polynomial model with degree k is defined as follows [16]:

\[ y = a_k x^k + a_{k-1} x^{k-1} + \cdots + a_1 x + a_0 \]
Fig. 3: Relationship between rainfall and altitude of studied stations in Iran.

\[ Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \ldots + \beta_k t^k + e_t, \]

Where, \( Y_t \) is the response variable, \( \beta \) is the unknown parameter, and \( e_t \) is the remaining \((t = 1, 2, \ldots, n)\).

The conventional regression assumes that \( e_t \) the sequence of independent normal random variables with mathematic expectation equal to zero and their variance \( \sigma^2 \) is constant. The linear model and parabolic model correspond to \( K = 1 \) and \( k = 2 \), respectively. The figures 4, 5, and 6 show that the models fit the time series of annual rainfall. In this modeling, the significance of models’ parameters is tested by t-student test. If the absolute value of this statistic is greater than or equal to 1.96, it indicates the significance of the parameter. However, seven stations of the 26 synoptic stations in Iran have significant linear trends. Meanwhile, 5 stations including Arak, Ardabil, Urmia, Sanandaj, and Tabriz which are located in the North West of Iran have a linear decreasing trend. Also, Kerman and Zahedan which are located in the South East of Iran have a decreasing linear significant trend model. Therefore, the North-West and South-East of Iran has decreased rainfall trend during the studied 50 years. The reduction has been 2 mm to over 3 mm per year. So, Zahedan with about 3.3 mm per year has the greatest reduction and Tabriz with 3 mm is in the next rank. According to the results of the modeling, it can be said that the systems which lead to rainfall in the North West and South East of Iran have changed in recent decades. This requires further investigation and is beyond the framework of this paper. The Qazvin station has a significant parabolic trend; it has increased trend in the early period of the study and decreased trend in the final period. Other stations have not any significant models and show a constant trend. The figures of stations rainfall trend modeling (Figures 4 and 6) show that about 73% of the studied stations in Iran (19 stations) show a decreasing trend. Of those, 7 stations have significant reduction trend at 5% error; the spatial arrangement of these stations was mentioned above. However, 6 stations have increase trend which have not statistical significance at 5% error.
Fig. 4: The time series plots and their fitted models.
Fig. 5: The time series plots and their fitted models (continued).
Fig. 6: The time series plots and their fitted models (continued).
The (ACF) and (PACF) charts were used to determine the values of q and p and the most appropriate and meaningful fitted model was chosen. As it was mentioned, the autocorrelation function $\rho_k$ measures the linear relationship between the time series observations which are separated by k time interval. The $\rho_k$ is always between +1 and -1. The correlation at different delays is shown by a graph named correlogram. If the values of $\rho_k$ are close to +1 (positive and statistically significant), the observations which are separated by k intervals have a strong tendency to move together in a linear direction with a positive slope. However, if the values of $\rho_k$ are close to -1 (negative and statistically significant), the observations which are separated by k intervals have a strong tendency to move together in a linear direction with a negative slope and indicate fluctuating values. This idea is one of the specific characteristics of climatic time series such as rainfall. This means that in the climatic time series, the observations are interdependent. The autocorrelation at delay one is more affected by non-viability than other delays. So, it provides the trend in the best way. However, if the autocorrelation at delay one is minimal or insignificant, the series viability may be inferred. Also, if the linear relationship of rainfall level in a year is calculated with the same values and 4 years delay and the correlation is positive and statistically significant, the 4 years repetition can be inferred from it. In other words, it is expected that the similar annual rainfall will be repeated after every 4 years. Conversely, if for example the autocorrelation at delay 4 years is negative and significant, it is expected that a dry year will occur after 4 humid years and a wet year will occur after 4 drought years. Thus, the negative autocorrelation displays oscillatory model at time interval k. If a time series with successive observations in both sides of mean have a tendency to frequency, then the correlogram will also have tendency to frequency. The figures 7, 8 and 9 show auto-correlogram graphs and figures 10, 11, and 12 show the partial auto-correlogram graphs of annual rainfall at each station. As it can be seen, the Gorgan, Uromiyeh, Semnan, and Zahedi stations have the first significant antenae. The Bojnoord and Zanjan stations at delay 4 and negative slope have significant antenae. Thus, it can be said that rainfall at these two stations has 4 years fluctuating wet and drought situation. In other words, it is expected that a dry year will occur after 4 humid years and a wet year will follow after 4 drought years. The Sanandaj station at first and second delays with positive slope have significant antenae. In addition to trend, therefore, the identical 2-year rainfall can be inferred in Sanandaj. The Kermanshah station in delay 6 with positive slope and in delay of 10 with negative slope is significant. So, it can be said that in Kermanshah, the identical rainfalls have 6 years repetition and fluctuating rainy- drought years have 10 years repetition. The Ardabil station in delay one has positive significant antenae which indicates the existence of a trend with negative slope. Also in delay 8, it has negative significant antenae which indicates the existence of fluctuating 8 years rainy- drought years in this station. Finally, the Tabriz station has significant antena in first and fifth delays with positive slope. In addition to the trend with negative slope, the identical and similar 5- year rainfalls can be inferred from rainfall in this city. The series with trend or non- viability in the mean or variance were exposed to appropriate differentiation operation and Box - Cox conversion. Finally, different ARIMA models were fitted to the time series. At most stations, more than a meaningful and appropriate model was chosen. Several criteria were used to choose the best ARIMA model fitted to one of the stations. For this purpose, the Akaike values and MSE index were calculated for the models. From two or more candidate ARIMA model for each station, the model with minimum AIC and MSE value is the most appropriate model. These criteria were considered in selecting the most suitable model for each station. For example, the process of modeling the Sanandaj, Zahedan, and Bojnoord stations is provided in the following. Considering significant linear trend and based on equation 6 and Figure 6, one conversed time series named $\{W_t\}$ was obtained in Sanandaj station after first order differentiation. Then, auto-correlogram graph and partial auto-correlogram graph was provided for conversed time series based on relations (2) and (3) and was presented in Figures 7 and 8. Based on auto-correlogram graph, the first and second antenae are significant and based on partial auto-correlogram graph, only one antenna is significant. Thus, $\text{ARIMA}(0,1,2)$ model may be inferred from auto-correlogram and $\text{ARIMA}(1,1,0)$ may be inferred from partial auto-correlogram. According to auto-correlogram graph, one second-order moving average model $\text{ARIMA}(0,1,2)$ seems logical. In other words:

$$Z_t = Z_{t-4} - a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Using the software Minitab / win, the fitted model is obtained as follows:

$$Z_t = Z_{t-4} - a_t - 0.7016 a_{t-1} - 0.1410 a_{t-2}$$

(4.73) (0.94)

The numbers in parentheses show t-statistics values for rejection of the null hypothesis (insignificance of parameters). If $|t| > 1.96$, it can be said with 95% confidence (5% error) that the parameters in the model are significant. Otherwise, the parameter will be considered equal to zero and will be removed from the model.
According to the fitted model, since the absolute value of the t-statistic for parameter $\theta_2$ is less than 1.96, therefore $\theta_2 = 0$; this will be removed from the model. Next, the model $ARIMA(0,1,1)$ is tested. The following fitted model is obtained:

$$Z_t = Z_{t-1} - a_t - 0.7892 a_{t-1}$$

(1.71)

Considering the t-statistic value is less than 1.96, the above model is not significant and does not seem logical. By fitting different models and trial and error, we finally obtain $ARIMA(0,1,3)_{con}$ model. This seems logical model and has significant parameters. In other words:

$Z_t = \theta_0 + Z_{t-1} - a_t - \theta_0 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$

The fitted model is:

$$Z_t = -3.307 + Z_{t-1} - a_t - 0.8193 a_{t-1} + 0.4873 a_{t-2} - 0.6563 a_{t-3}$$

\((-2.15)\) \((6.13)\) \((-3.00)\) \((4.93)\)

As it can be seen, based on the auto-correlogram graph, the $ARIMA(0,1,3)_{con}$ model with a constant value $\theta_0$ is a suitable candidate and it is called $M_1$.

Based on partial auto-correlogram graph, the Sanandaj differentiated time-series is significant in first antennae. So, modeling is started by $ARIMA(1,1,0)$. In other words:

$$W_t = \phi_1 W_{t-1} + a_t$$
$$W_t = -0.4766 W_{t-1} + a_t$$

Where, $W_t = Z_t - Z_{t-1}$. The T value corresponding $\phi_1$ in this model is 3.65. Therefore, the parameter is significant. The addition of another parameter ($\theta_0$ and $\phi_1$ or other parameters) to this model does not make it significantly better. Thus, the model $ARIMA(1,1,0)$ which is called $M_2$ is another good candidate for the time series of annual rainfall in Sanandaj. As previously mentioned, each accepted model should be tested in terms of normality, independence, and homogeneity of residuals variance $\{a_t\}$, because the residuals of the model should have no model. Also, the AIC and MSE indicators were used to choose a suitable model from among two or more candidate models for each series. For time series of Sanandaj which had two $M_1$ and $M_2$ candidates, the $ARIMA(0,1,3)_{con}$ model was selected as a suitable final model by testing the residuals of model and AIC and MSE indices values due to randomness and normality of residuals and having a lower MSE. Table 4 shows the fitted model of each series along with the values of AIC and MSE indices. The values of these indices in this model are obtained 595.9 and 12155, respectively.

For Bojnoord series, based on the Figure 7 in auto-correlogram graph (q), the fourth antennae is significant. So, one fourth-order moving average model $ARIMA(0,1,4)$ is fitted. In other words,

$$Z_t = Z_{t-1} - a_t - \theta_3 a_{t-1} - \theta_4 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4}$$

The fitted model is obtained as follows:

$$Z_t = Z_{t-1} - a_t - 0.8631 a_{t-1} - 0.3292 a_{t-2} + 0.3222 a_{t-3} - 0.1114 a_{t-4}$$

(4) \(1.25\) \((-1.23)\) \((0.53)\)

Given that the t-statistic values in the above fitted model for the $\theta_3$, $\theta_4$, $\theta_3$ and $\theta_4$ parameters are respectively equal to 1.25, -1.23, and 0.53 and the absolute value of these values is less than 1.96, therefore it is concluded that the parameters are not significant in this model and are excluded from the model. Now, the $ARIMA(0,1,3)$ model is fit. The T-statistic values for $\theta_2$ and $\theta_3$ parameters are obtained 0.41 and -0.50, respectively. By fitting different models and by trial and error, the $ARIMA(2,1,3)$ model is selected as optimum model for annual rainfall in Bojnoord. In other words:

$$Z_t = Z_{t-1} + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + a_t - \theta_3 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$$
$$Z_t = Z_{t-1} + 0.7289 Z_{t-1} - 0.8312 Z_{t-2} + a_t - 1.5101 a_{t-1} + 1.4112 a_{t-2} - 0.8130 a_{t-3}$$

(2.14) \((-2.30)\) \((4.34)\) \((-2.55)\) \((2.50)\)

Considering the t-statistic values for the parameters, the third-order moving average model and second-order autoregressive is selected as suitable model for Bojnoord. It should be mentioned that considering equation 6, the annual rainfall time series of Bojnoord have no significant linear or parabolic trend. However, by fitting different models and using the criteria used in this article, the $ARIMA(2,1,3)$ model was chosen as the
most appropriate model. The MSE and AIC values in this model was obtained 5739 and 352.5, respectively (Table 4).

For time series of Zahedan, the equation 6 and Figure 6 shows that the time series has significant linear trend. By first order differentiation, the auto-correlogram and partial auto-correlogram of differentiated series were calculated based on equations (2) and (3). Both graphs have first significant antennae. Therefore, \(ARIMA(0,1,1)\) and \(ARIMA(1,1,0)\) models can be inferred. Thus, the following moving average model is first fitted:

\[ Z_t = Z_{t-1} - \theta_1 a_{t-1} \]
\[ Z_t = Z_{t-1} - a_t - 0.8811 a_{t-1} \]

The t-statistic value for \(\theta_1\) in this model is obtained 14.79. Thus, this factor is significant and opposite to zero. Now, it should be considered whether the constant parameter \(\theta_0\) makes the model significantly better. The fitted model is as follows:

\[ Z_t = \theta_0 + Z_{t-1} - a_t - \theta_1 a_{t-1} \]
\[ Z_t = 0.2665 + Z_{t-1} - a_t - 0.9684 a_{t-1} \]

(1.11)  (3.88)

Where, the t-value for the test of assuming coefficients are equal to zero and for the constant factor in the 95% confidence level is less than 1.96. Therefore, the coefficient \(\theta_0\) is considered equal to zero and is excluded from the model. Now, it is considered whether \(\theta_1\) is significant in the model? Then, the following model is examined:

\[ Z_t = Z_{t-1} + \phi_1 Z_{t-1} + a_t - \theta_1 a_{t-1} \]

The fitted model is obtained as follows:

\[ Z_t = Z_{t-1} + 0.3433 Z_{t-1} + a_t - 1.0033 a_{t-1} \]

(2.76)  (180.3)

Since the t value for above coefficients is greater than 1.96, so \(\phi_1\) makes the model significantly better. The addition of another parameter does not make the above model significantly better. Therefore, the \(ARIMA(0,1,1)\) model is selected as suitable model for annual rainfall series in Zahedan. The MSE and AIC values in this model are obtained 1677 and 593.1, respectively (Table 4).

The annual rainfall time series of other stations were modeled in the same way. The suitable model was selected using criteria such as models’ residual testing and indicators such as MSE and AIC. Using the selected models, the rainfall was forecast for the next 5 or 10 years. It should be mentioned that in order to avoid prolongation and repetition, the process of modeling all time series is not provided. The results of modeling and the selected ARIMA models for each of the series along with MSE and AIC criteria are presented in Table 4. The first column of the table shows the studied stations. The second column shows the selected ARIMA models for each station. The third and fourth columns are respectively the MSE and AIC values for each of the stations ARIMA models. The criteria of selecting best model includes minimum MSE and AIC values for each model as well as the reviewing model residuals in terms of being independence, randomness, and normality. According to ARIMA models in Table 4, about 54 percent of rainfall in stations of Iran is a function of rainfall in past 1 to 5 years and random components of past 1 to 3 years.
Fig. 7: The auto-correlogram graphs of stations’ annual rainfall time series.
Fig. 8: The auto-correlogram graphs of stations’ annual rainfall time series.
Fig. 9: The auto-correlogram graphs of stations’ annual rainfall time series.
Fig. 10: The partial auto-correlogram graphs of stations’ annual rainfall time series.
Fig. 11: The partial auto-correlogram graphs of stations’ annual rainfall time series.
Fig. 12: The partial auto-correlogram graphs of stations’ annual rainfall time series.

The rainfall in about 31 percent of Iran’s stations is only a function of random components in the past one to three years. In addition to this rainfall, about 15 percent of Iran’s stations are a function of rainfall in the past one to five years. The figure 13 shows the spatial distribution of rainfall fitted models to each of the studied stations. As is clear in the figure, the models which are fitted to Iran’s North and North West stations rainfall have greater complexity. The model with more significant parameters is called complex model. In the fitted models of rainfall stations in this part of Iran, most of the models have 5, 6, and 7 significant parameters which indicate the complexity of models. While, in the Center, South, and South-East of Iran, the models have low complexity and have two or three significant parameters. Given that most models are simple in this part of
country and have first-order moving average, first-order autoregressive, or a combination of them, it can be concluded that rainfall in this part of Iran is a function of random components in last year or a function of rainfall in last year. While, the models have greater complexity in the northern and western parts of Iran and the rainfall is more a function of the random components in the last one to three years or a function of rainfall in the past 5 years. However, the complex integrated models are considered more in this part of country.

Table 2: Selected ARIMA models for stations.

<table>
<thead>
<tr>
<th>Station</th>
<th>Model</th>
<th>MSE</th>
<th>AIC value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahwaz</td>
<td>ARIMA(1,0,1)</td>
<td>7553</td>
<td>580.9</td>
</tr>
<tr>
<td>Arak</td>
<td>ARIMA(1,0,1)</td>
<td>10624</td>
<td>655.6</td>
</tr>
<tr>
<td>Ardebel</td>
<td>ARIMA(1,1,1)</td>
<td>6156</td>
<td>361.8</td>
</tr>
<tr>
<td>Bandar Abbas</td>
<td>ARIMA(1,0,1)</td>
<td>12858</td>
<td>646.8</td>
</tr>
<tr>
<td>Birjand</td>
<td>ARIMA(1,1,1)</td>
<td>3244</td>
<td>574.8</td>
</tr>
<tr>
<td>Bojnurd</td>
<td>ARIMA(2,1,3)</td>
<td>5739</td>
<td>352.5</td>
</tr>
<tr>
<td>Bushehr</td>
<td>ARIMA(1,0,1)</td>
<td>13428</td>
<td>764.9</td>
</tr>
<tr>
<td>Esfahan</td>
<td>ARIMA(1,0,1)</td>
<td>3175</td>
<td>641.4</td>
</tr>
<tr>
<td>Ghazvin</td>
<td>ARIMA(0,1,2)</td>
<td>7726</td>
<td>582.5</td>
</tr>
<tr>
<td>Gorgan</td>
<td>ARIMA(1,1,1)</td>
<td>30057</td>
<td>730.2</td>
</tr>
<tr>
<td>Hamedan</td>
<td>ARIMA(0,1,2)</td>
<td>6956</td>
<td>676.3</td>
</tr>
<tr>
<td>Kerman</td>
<td>ARIMA(0,1,1)</td>
<td>2932</td>
<td>625.5</td>
</tr>
<tr>
<td>Kermanshah</td>
<td>ARIMA(3,1,0)</td>
<td>16463</td>
<td>722.1</td>
</tr>
<tr>
<td>Khoram abad</td>
<td>ARIMA(0,1,1)</td>
<td>15739</td>
<td>718</td>
</tr>
<tr>
<td>Mashhad</td>
<td>ARIMA(4,1,0)</td>
<td>5443</td>
<td>659.3</td>
</tr>
<tr>
<td>Orumiyeh</td>
<td>ARIMA(0,1,1)</td>
<td>9091</td>
<td>686.3</td>
</tr>
<tr>
<td>Ramsar</td>
<td>ARIMA(2,2,3)</td>
<td>35978</td>
<td>757.7</td>
</tr>
<tr>
<td>Sanandaj</td>
<td>ARIMA(0,1,3)</td>
<td>12155</td>
<td>595.9</td>
</tr>
<tr>
<td>Semnan</td>
<td>ARIMA(1,1,1)</td>
<td>2931</td>
<td>471.8</td>
</tr>
<tr>
<td>Shahrekurd</td>
<td>ARIMA(1,0,1)</td>
<td>11503</td>
<td>639.6</td>
</tr>
<tr>
<td>Shiraz</td>
<td>ARIMA(1,0,1)</td>
<td>18229</td>
<td>740.7</td>
</tr>
<tr>
<td>Tabilriz</td>
<td>ARIMA(0,1,1)</td>
<td>5170</td>
<td>658.5</td>
</tr>
<tr>
<td>Tehran</td>
<td>ARIMA(2,2,3)</td>
<td>5913</td>
<td>665.9</td>
</tr>
<tr>
<td>Yazd</td>
<td>ARIMA(0,1,1)</td>
<td>766</td>
<td>524.9</td>
</tr>
<tr>
<td>Zahedan</td>
<td>ARIMA(1,1,1)</td>
<td>1677</td>
<td>593.1</td>
</tr>
<tr>
<td>Zanjan</td>
<td>ARIMA(3,1,0)</td>
<td>7360</td>
<td>626.1</td>
</tr>
</tbody>
</table>

MSE is one of the most important indicators of model’s performance evaluation. In fact, this indicator measures the difference between the values predicted by the selected model and the actual observed values. The larger the value of this index means that this model may better predict the future values of time series and with fewer errors. The MSE index value for each of the selected models of stations is given in Table 4. Figure 14 shows the Interpolation and zoning map of MSE index in Iran. Based on this figure, the minimum MSE index value is related to East, South East, and parts of Central Iran and the maximum value of this index is related to North and then the southwestern margin of Iran. Given that less the MSE index value means that the model is suitable and reliable for predicting future values of time series, it can be said that the predictions in eastern, central, northeastern, and northwestern parts of Iran are reliable than the values predicted for northern and southwestern parts via selected ARIMA models. The main use of modeling by ARIMA techniques is predicting future values. The necessary planning may be done by predicting to have better efficiency in different fields. The prediction should be such that it will be closer to observed actual values. In this paper, each of the selected models was used to predict rainfall of next 5 years and 10 years. Then, the predicted rainfalls of stations in the next 5 to 10 years were interpolated using kriging method. A number of 4646 pixels were obtained from interpolated rainfalls in Iran. The figures 15 and 16 respectively show the interpolated maps of rainfall in next 5 years and 10 years which are predicted through selected ARIMA models. There are many similarities in both maps. According to data of 4646 pixels in Iran, 210 mm is the average of rainfall in about next5 years and about 207 mm is for next 10 years. This shows a decrease in mean of rainfall in future years. However, this reduction is not in everywhere and there is a slight increase in some places in predicted rainfall. Given that some of the studied stations had significant reduction trend in the studied statistical period, thereby, the reduction in average rainfall seems logical.
Fig. 13: Spatial distribution of fitted models to each of the studied stations.

Fig. 14: Zoning and Interpolation Map of MSE index values in Iran.
Table 5 shows statistic features of the coming 5 or 10 year rainfall of Iran and it also shows downfalls of past 50 years. This information is result of 8389 pixels that was interpolated using Kriging method. As can be seen in table 5 the average rainfall of Iran for the past 50 years is 232 mm and average of downfall for the last decade is about 216 mm, while the average downfall of Iran for the coming 5 years is predicted to be about 210 mm and for future 10 years it might be about 208 mm. So as presented here, rainfall of Iran has enjoyed a decreasing trend and consequently the variance has decreased. This indicates that the rainfall of Iran has had a steady behavior. Figure 14 shows histogram graph of the rainfall of Iran. As table 5 and figure 14 suggest, skewness of rainfall of Iran is positive. Positive skewness in downfall expresses that number of observations is less than the mean frequency. And this is a quality of downfall in arid and semi-arid areas.
Table 5:

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Mean total</th>
<th>Mean of 10 recent years</th>
<th>5 next year</th>
<th>10 next year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>232.30</td>
<td>216.70</td>
<td>210.1</td>
<td>208.6</td>
</tr>
<tr>
<td>StDev</td>
<td>121.80</td>
<td>119.60</td>
<td>118.80</td>
<td>119.20</td>
</tr>
<tr>
<td>Variance</td>
<td>14828.80</td>
<td>14301.30</td>
<td>14124.30</td>
<td>14219.90</td>
</tr>
<tr>
<td>Q3</td>
<td>135.10</td>
<td>104.70</td>
<td>110.70</td>
<td>108.30</td>
</tr>
<tr>
<td>Median</td>
<td>208.00</td>
<td>182.20</td>
<td>176.50</td>
<td>181.90</td>
</tr>
<tr>
<td>Q3</td>
<td>304.10</td>
<td>274.80</td>
<td>276.30</td>
<td>274.20</td>
</tr>
<tr>
<td>Skewness</td>
<td>1.34</td>
<td>1.42</td>
<td>1.55</td>
<td>1.50</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.68</td>
<td>4.63</td>
<td>4.66</td>
<td>4.35</td>
</tr>
</tbody>
</table>

Fig. 17:

**Conclusion:**

In this paper, the annual rainfall of Iran was modeled with polynomial and ARIMA models. The results showed a decrease in rainfall trend for the incoming years in most parts of Iran. Given that over the past half century, the average temperature in Iran has increased one degree, this may be the reason of decreasing rainfall in some stations and lack of any increase rainfall trend in Iran. The definite judgment requires a comprehensive study of multivariate methods and digitized data. Given that over 70% of the stations that have experienced a significant decline trend are located in the North West (Arak, Ardabil, Orumiyeh, Sanandaj, and Tabriz) of Iran, it seems that the mechanisms leading to the rainfall in this area are changed and weakened. On the other hand, considering that the rainfall in Iran has increasing trend in winter and autumn and decreasing trend in summer, it can be said that in the past half-century, the rainfall in Iran has been concentrated. The use of water in such a climate requires powerful management of water.
Appendix:

Table 1: 5% critical values for the statistic $T_n$ of the single shift SNHT (Alexanderson and Moberg, 1997).

<table>
<thead>
<tr>
<th>n</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{95}$</td>
<td>5.70</td>
<td>6.95</td>
<td>7.65</td>
<td>8.10</td>
<td>8.45</td>
<td>8.65</td>
<td>8.95</td>
<td>9.15</td>
</tr>
</tbody>
</table>

Table 2: 5% critical values for the statistic $N$ of Von Neumann ratio (Buishand, 1982).

<table>
<thead>
<tr>
<th>n</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>70</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{95}$</td>
<td>1.30</td>
<td>1.42</td>
<td>1.49</td>
<td>1.54</td>
<td>1.61</td>
<td>1.67</td>
</tr>
</tbody>
</table>

REFERENCES


