The Investigation and Forecasting of Relative Humidity Variation of Pars Abad-e-Moghan, North-West of Iran, by ARIMA Model

[1]Ghita Shiri, [2]Bromand Salahi, [3]Rasool Samadzadeh and [4]Mohammadreza Shiri
[1]Islamic Azad University of Ardebil, Ardebil, Iran
[2]University of Mohaghegh Ardabili, Ardebil, Iran
[3]Islamic Azad University, Ardebil Branch, Iran
[4]Islamic Azad University, Parsabad Mogan Branch, Iran

Abstract: The use of statistical methods to study and predict changes in climate parameters has found wide application. One of these is Autoregressive Integrated Moving Average (ARIMA) model. In this study, the average monthly relative humidity of Pars Abad-e-Moghan station investigated based on above-mentioned model in a 25 years statistical period (1984-2010). Time series plot was used to investigate whether the series is stationary or non-stationary. Because the series was not stationary, so it was converted to a stationary series by Differencing order 1 (D = 1). Functions such as Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), Root Mean Square Error (RMSE), Mean Absolute Deviation (MAD) and Akaikes Information Criterion (AIC) were calculated to determine the order of Autoregressive (AR (p)) and Moving Average (MA (q)) parameters (Q, q, P, p). With considering above-mentioned criteria, seasonal ARIMA (p = 1, d = 0, q = 1) (P = 1, D = 1, Q = 1) modal was chosen as best modal. So, it was used to predict future values of relative humidity. Before using the model for forecasting, it checked for adequacy based on plot of standardized residuals, ACF plot of residuals, PACF plot of residuals. The relative humidity of Pars Abad-e-Moghan station was forecasted based on selected modal up to 2014. The results of which indicate the relative humidity in May, June and September will be faced with an increasing trend while in the rest of months will be not observed any trend.

Key words: ARIMA, forecasting, relative humidity, SARIMA, trend, Iran

INTRODUCTION

Considering the increasing role of human in climate changing of scientists and politicians awareness about climate change status is very important. Climatic changes have occurred from long ago in different aspects of time and location. There are many scientific signs showing existence of a different weather in the past climates from nowadays condition. Generally, emission of green house gases is counted as the main reason to climate change. In addition, climatic changes considerably affect agricultural products especially cereals. Finally, plants get susceptible to diseases as temperature and humidity change. Then, it ends in their damage and waste. In addition to importance of forecasting climatic elements changes in agriculture, one can notice the effect of this parameter in dam building and other construction projects.

To do these studies, time series methods and a process named ARIMA model with orders of (p, d, q) are often employed. A Time Series (TS) is a sequence of observations ordered in time. Mostly these observations are collected at equally spaced, discreet time intervals. A basic assumption in any time series analyzing/modeling is that some aspects of the past pattern will continue to remain in the future.

However, the background of studies related to climatic parameters changes returns to recent few decades; numerous researchers around the world including Iran because of its importance have studied it. Researchers can name some examples of these studies as follow: Jahanbakhsh and Baser (2003) used ARIMA model to study Tabriz monthly average temperature fluctuation from 1959-1998 (for 40 years).

They showed that Tabriz monthly average temperature fluctuates around a nonlinear axis with an increasing trend indicating temperature rise. They also used ARIMA (0, 0, 1) (0, 1,1) 12 model which is a combination of two parts, seasonal and non-seasonal as their calculational model to forecast Tabriz station monthly average temperatures up to 2010. Investigating temperature and rainfall changes for different locations in

Iran, Jahanbakhsh and Torabi (2004) concluded that SARIMA model is suitable to predict the future quantities of these parameters. It was also found out that the changes in minimum temperature and rainfall are not meaningful in most of stations during the statistical period under study but changes in maximum temperature are meaningful in all stations except Bandar-e-Lengeh. Ramezani (2001) investigated rainfalls in four stations in Mazandaran province with time series. As an example, recommended seasonal model $(1, 1, 0) (1, 1, 0)$ with natural logarithmic transformation to anticipate rainfall for Babolsar station in coming years.

Taheri (1998) employed multiplicative autoregressive moving average model for modeling and forecasting temperature and rainfall for eleven weather stations in Iran. Al-Awadhi and Jolliffe (1998) predicted surface pressure of England South East by use of ARIMA family model. They presented different models for every season and chose the best one for each season.

Leite and Peixoto (1996) investigated the application of autoregressive models in studying temperature changes by use of the longest time series. The study showed existence of considerable variability values in annual and decadal scales. Box and Jenkins (1976) stated the reason behind the use of autocorrelated cumulative moving average models as existence of an autocorrelation in climatic data having seasonal trends. This research is done to investigate changes in monthly relative humidity and to forecast this parameter for Pars Abad-e-Moghan synoptic station benefiting ARIMA model.

MATERIALS AND METHODS

The selected station was Pars Abad-e-Moghan that it is limited by latitudes 39°20'N and 47°30'E with 40-50 m above from sea level and located at North-West of Iran. The climate of this region (Pars Abad-e-Moghan) is semi-arid with average annual precipitation of 271.2 mm. In addition, the averages of annual maximum temperature, minimum temperature and relative humidity are 20.5°C, 9.6°C and 72%, respectively.

The relative humidity data of 1984-2010 obtained from Adredebil meteorological organization. Autoregressive Integrated Moving Average (ARIMA) model used to determine trend and forecast future values in relative humidity of Pars Abad-e-Moghan.

In statistics, Autoregressive Moving Average (ARMA) models, sometimes called Box-Jenkins models after the iterative Box-Jenkins methodology usually used to estimate them are typically applied to auto-correlated time series data. Given a time series of data Xt, the ARMA model is a tool for understanding and predicting future values in this series. The model consists of two parts, an Autoregressive (AR) part and a Moving Average (MA) part.

The model is usually then referred to as the ARMA (p, q) model where p is the order of the autoregressive part and q is the order of the moving average part. In time series analysis, an Autoregressive Integrated Moving Average (ARIMA) model is a generalization of an Autoregressive Moving Average (ARMA) model. They are applied in some cases where data show evidence of non-stationarity where an initial differencing step (corresponding to the integrated part of the model) can be applied to remove the non-stationarity. The model is generally referred to as an ARIMA (p, d, q) model where p, d and q are integers ≥ 0 and refer to the order of the autoregressive, integrated and moving average parts of the model, respectively.

ARIMA models form an important part of the Box-Jenkins approach to time-series modeling. When one of the terms is zero, it's usual to drop AR, I or MA. A number of variations on the ARIMA model are commonly used. Sometimes a seasonal effect is suspected in the model. In this case it is often considered better to use a SARIMA (seasonal ARIMA) model than to increase the order of the AR or MA parts of the model. Finding appropriate values of p and q in the ARMA (p, q) model can be facilitated by plotting the partial autocorrelation functions for an estimate of p and likewise using the autocorrelation functions for an estimate of q. The autocorrelation of a random process describes the correlation between values of the process at different points in time as a function of the two times or of the time difference.

Partial autocorrelations measures the degree of association between X_t and X_{t-k} when the effects of other time lags 1, 2, 3..., k-1 are removed. Auto Correlation Formula (ACF) in lag k is as follows:

$$Pk = \frac{\sum_{t=1}^{n-k}(X_t - \overline{X})(X_{t-k} - \overline{X})}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

Where:

Pk = Autocorrelation function value time series at lag K, X_1

X_1-K = Values of variables or time series data in t and t-k time, respectively

$\overline{X}$ = Mean value of variables or time series data

For determining the order of parameters (Q, q, P, p), Autocorrelation (ACF) and Partial Autocorrelation (PACF) functions were used. Type and orders of modal was determined based on significant differences of lag correlation coefficients with interval. To avoid excessive of error estimation were used RMSE, AIC and MAD criteria. Assumption of independence and random of series was confirmed with residuals distribution (Moshkani, 1994; Azar and Momeny, 2001). The calculation of RMSE, AIC and MAD was done with following formula:

$$MAD = \frac{1}{n} \sum_{t=1}^{n} \left| X_t - \hat{X}_t \right|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(X_t - \hat{X}_t \right)^2}$$

$$AIC(i) = N.Ln(\sigma_{ci}^2) + 2n_i$$

Where:

X_t =   Values of time series data in t time
$\hat{X}_t$ =   Fitted values of time series data in t time
i   =   Indicates the number of selected models
n   =   The sum of parameters (the sum of orders in autoregressive and moving average models)
N   =   Number of data observations
σ_{ci} =   Standard deviation

Akaikes criteria is based on the belief that every order that is less, it is will do better fit (Modares and Eslamiyan, 2004). The data were statistically analyzed by SPSS (13) and MINTAB (15) computer programs.

RESULTS AND DISCUSSION

In first step, it must be determined whether the variance is stable or not by considering the graph of TS (time series) plot. If the variance is not stable, it can often be converted to a stable by transformation. Plotted time series for the main data showed that this series had seasonal stationary (Fig. 1). Because the series pattern variation is repeated after twelve data.

So, it is necessary for making stability in variance, a logarithmic transformation is performed on the data. Considering Fig. 2, it seems that relationship between mean and variance is partly destroyed after logarithmic transformation.

If a graph of ACF (autocorrelation function) of the time series values either cuts off fairly quickly or dies down fairly quickly then the time series values should be considered stationary. If a graph of ACF dies down
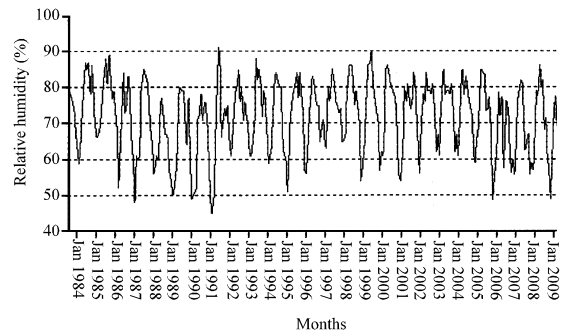

Fig. 1: Time series plot of relartive humidity data in Pars Abad-e-Moghan station
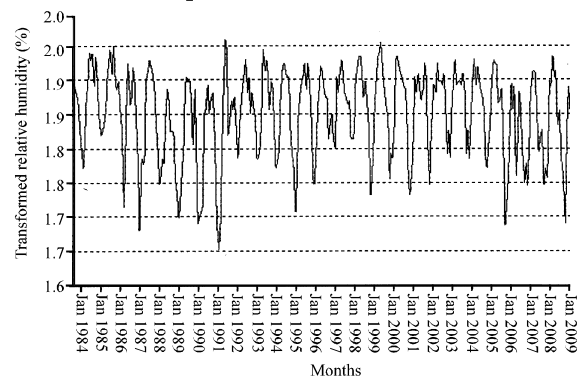

Fig. 2: Time series plot of relative humidity of Pars Abad-e-Moghan station based on tranformed data (logarithmic tranformation)

extremely slowly then the time series values should be considered non-stationary. If the series is not stationary, it can often be converted to a stationary series by differencing.

That is the original series is replaced by a series of differences. An ARMA model is then specified for the differenced series.

Differencing is done until a plot of the data indicates the series varies about a fixed level and the graph of ACF either cuts off fairly quickly or dies down fairly quickly. Plotted ACF and PACF (partial autocorrelation function) graphs showed that series had regular seasonal variation. In addition, ACF didn't fail to die out rapidly. Thus the series had seasonal non-stationary (S = 12) (Fig. 3 and 4). So, it should be converted to a stationary series by seasonal differencing.

Plotted time series graphs on monthly relative humidity showed that series had funded a fixed level after seasonal differencing order (D = 1) ((Fig. 5). Also ACF and PACF graphs died out sectional and exponential toward zero (Fig. 6 and 7). So, the time series seems to be stationary. About 60 autocorrelation and partial
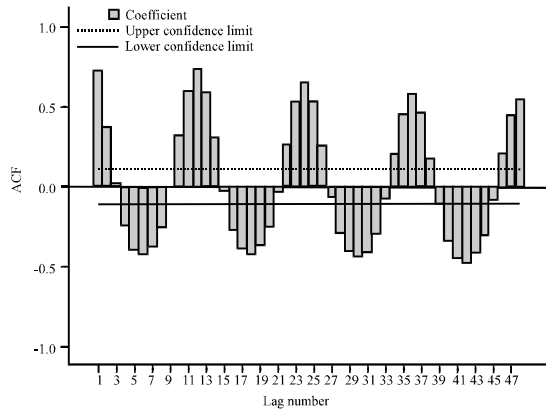
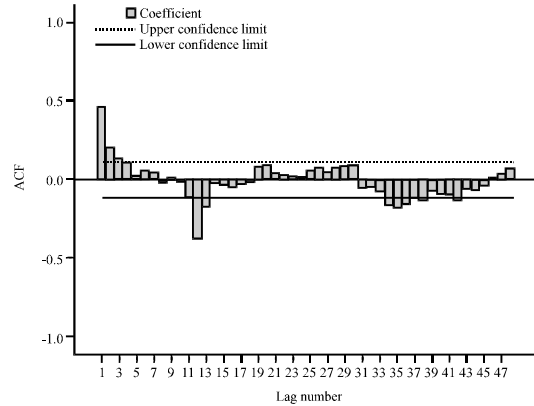Fig. 3: Autocorrelation function plot of relative humidity of Pars Abad-e-Moghan station

Fig. 6: Autocorrelation function plot of relative humidity of Pars Abad-e-Moghan station with once seasonal differncing
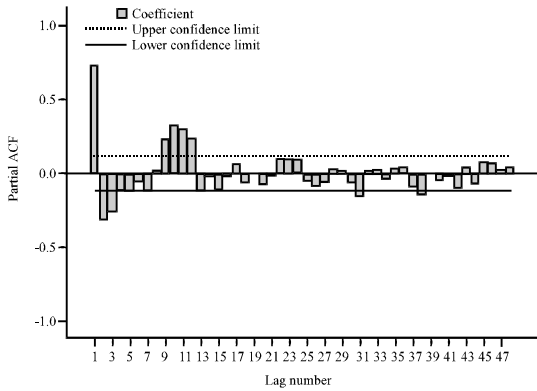

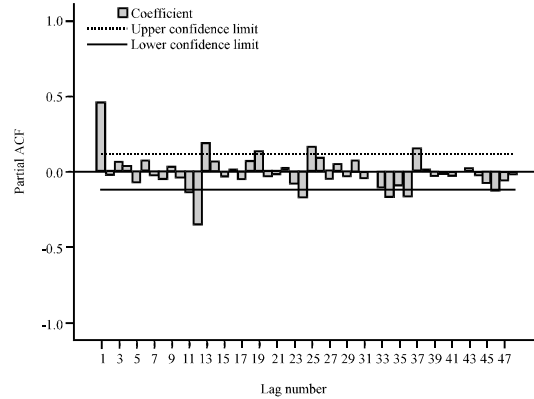Fig. 4: Partial autocorrelation function plot of relative humidity of Pars Abad-e-Moghan station

Fig. 7: Partial autocorrelation function plot of relative humidity of Pars Abad-e-Moghan station with once seasonal differncing
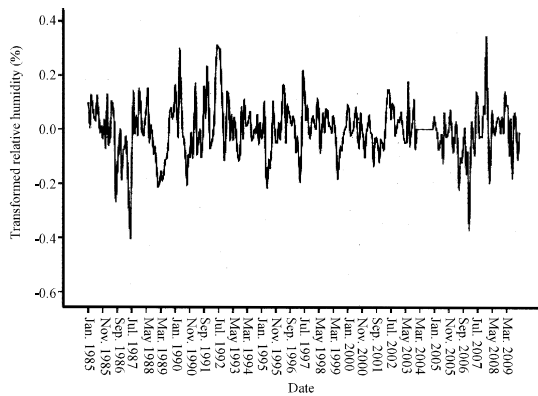

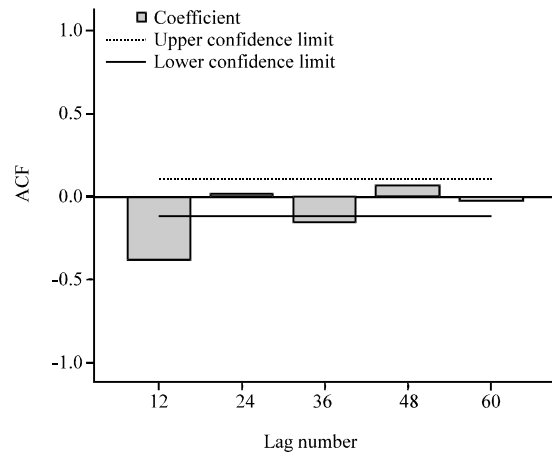Fig. 5: Time series plot of relative humidity of Pars Abad-e-Moghan station with once seasonal differncing

Fig. 8: Autocorrelation function plot of relative humidity of Pars Abad-e-Moghan station for multiplesalternation period

autocorrelationin multiples 12 were shown at Fig. 8 and 9, respectively. It had some significant coefficients. Generally, studying the past graphs revealed that relative
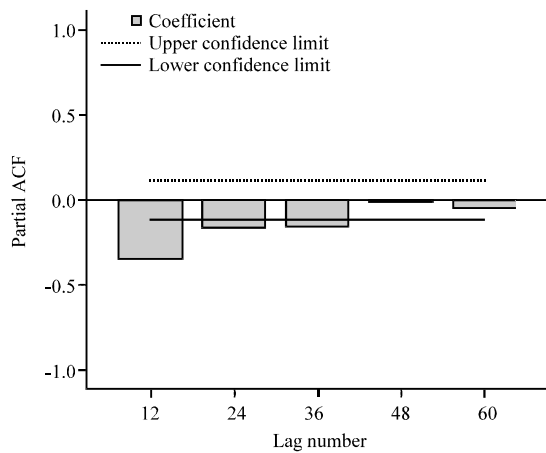
Fig. 9: Partial autocorrelation function plot of relative humidity of Pars Abad-e-Moghan station for multiplesalternation period
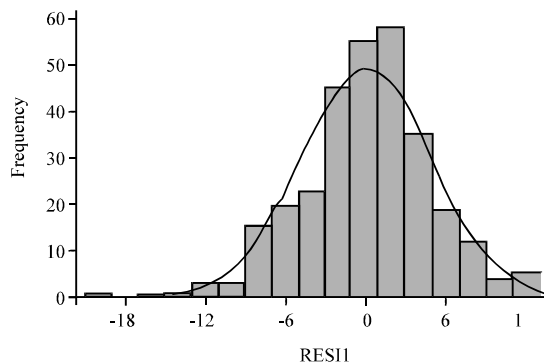

Fig. 10: Residual histogram gragh of relative humidity of Pars Abad-e-Moghan station

humidity series of Pars Abad-e-Moghan station experiences seasonal changes. Therefore, the chosen model would be SARIMA. Moreover, the mentioned series became stationary by once seasonal differencing (D = 1). Additionally, to choose different orders of D and d; first, different orders of seasonal and non-seasonal differencing (D and d) were considered for D and d according to Table 1.

Then the variance of each case was calculated separately. According to Table 1, orders of D = 1 and d = 0 having the lowest variance are selected for the model because it is just under these conditions (D = 1 and d = 0) that the mentioned series experiences the lowest variance and the most stationary surface. This conclusion is also verified in autocorrelation and partial autocorrelation time series figures with seasonal differencing (D = 1).

By clarification of orders of D and d, it is necessary to determine the orders of P, p, Q, q to choose the final

Table 1: The different orders of seasonal and non-seasonal differencing (D and d) and their variance of relative humidity (%) of Pars Abad-e-Moghan

D	d	Mean	Varince	No of time series data
0	0	71.7200	99.600	312
1	0	0.0199	54.943	311
0	1	-0.1737	47.817	300
1	1	-0.0308	53.785	299
2	0	0.0539	94.819	310
0	2	-1.3020	135.404	288
1	2	0.0317	155.922	287
2	1	0.0510	137.530	298
1	2	-0.0797	389.744	286

Table 2: The values of RMSE, AIC and MAD for different patterns of relative humidity (%) of Pars Abad-e-Moghan

Pattern	Orders of modal	RMSE	MAD	AIC
A_1	(0, 0, 1) (0, 1, 1) 12	5.473	4.120	1019.8
A_2	(1, 0, 1) (2, 1, 1) 12	5.138	3.893	985.6
A_3	(1, 0, 1) (1, 1, 1) 12	5.133	3.895	984.1
A_4	(1, 0, 1) (0, 1, 1) 12	5.148	3.947	986.8

SARMIA model. To select the appropriate model among different ones, the values for RMSE, AIC, AIC MAD and MAD are calculated for over 100 patterns. To consider brevity, just values for four patterns, A_1, A_2, A_3 and A_4 having the lowest values of AIC, RMSE and MAD are shown in Table 2.

The results revealed that pattern A_3 with the lowest values of AIC, RMSE and MAD could be selected as an appropriate model for prediction (Table 2).

Finally, Pars Abad-e-Moghan station general monthly relative humidity series model with a combination of two seasonal and non-seasonal components by observing parsimony principle will be SARIMA (1, 0, 1) (1, 1, 1) 12. After selection of the final model, it should be assessed accurately. One can point out checking normality and independence of the model residuals as the most important assessments noticed.

Residuals histogram of the main model for Pars Abad-e-Moghan monthly relative humidity and its normal curve show the degree of closeness of normal curve to frequency distribution of residual observations then normality of the residual observations series.

The graph of autocorrelation coefficients for residuals of Pars Abad-e-Moghan main monthly average relative humidity series model also shows that the coefficients are within the confidence range and the residuals of the mentioned series are completely independent of each other.

In addition, it could be concluded of residuals normal line graph that no distinguished deviation from normal line is observed. Then the hypothesis of the residuals are normal is confirmed (Fig. 10-12). So, the relative humidity of Pars Abad-e-Moghan station was forecasted based on
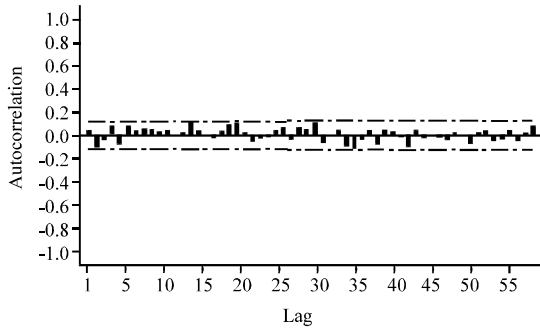
Fig. 11: Residual autocorrelation function plot of relative humidity of Pars Abad-e-Moghan station
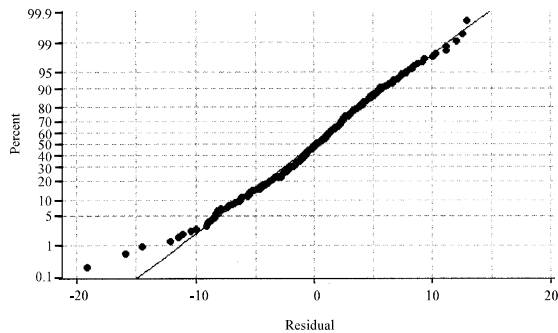

Fig. 12: Residual normal line graph of relative humidity of Pars Abad-e-Moghan station

Table 3: Predicted values of relative humidity (%) of Pars Abad-e-Moghan station based on SARIMA (1,0,1) (1,1,1) 12 model

Months	2010	2011	2012	2013	2014	Mean of 26 years
Jan.	78.9	78.9	78.9	79.0	79.0	78.7
Feb.	74.9	74.7	74.5	74.4	74.2	77.1
Mar.	75.4	75.4	75.4	75.3	75.3	75.6
Apr.	76.5	76.7	76.9	77.2	77.4	73.5
May	72.0	72.2	72.5	72.7	72.9	68.8
June	59.6	59.6	59.6	59.6	59.5	59.8
July	58.9	58.9	58.9	58.9	58.9	58.4
Aug.	61.8	61.9	62.0	62.2	62.3	60.3
Sep.	71.6	72.0	72.3	72.6	72.9	67.9
Oct.	79.8	80.0	80.2	80.3	80.5	77.5
Nov.	80.0	79.8	79.7	79.6	79.5	81.3
Dec.	83.3	83.4	83.5	83.6	83.7	82.5

The result of forecasting was shown in Table 3 and Fig. 13. Previously, the ARIMA modal was applied in investigation of weather parameters by Al-Awadhi and Jolliffe (1998), Jahanbakhsh and Torabi (2004), Jahanbakhsh and Baser (2003), Jameei (2002) and Taheri (1998). The t-test is employed to investigate if any change will happen to trends of relative humidity of different months in next 5 years or not. The results show an increasing trend in relative humidity of May, June and September months for Pars Abad-e-Moghan station with no remarkable change in other months in next 5 years (Table 4).
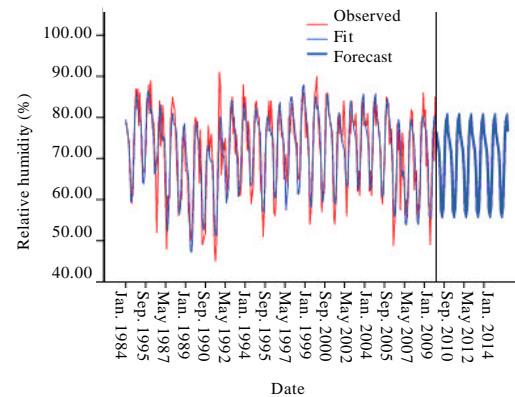

Fig. 13: The observed and fitted value and forcasted value of relative humidity for Pars Abad-e-Moghan station upto 2014

Table 4: Evaluation of relative humidity values in future years based on average predicted and average values of the past 5 years

Months	Predicted values of 5 coming years	Actual values of 5 past years	t
Jan.	80.4	77.0	0.69[NS]
Feb.	77.4	74.5	1.03[NS]
Mar.	77.2	71.6	2.05[NS]
Apr.	75.5	72.3	1.50[NS]
May	71.9	66.2	2.71[*]
June	61.7	56.0	2.83[*]
July	58.7	56.1	1.23[NS]
Aug.	62.1	59.1	1.79[NS]
Sep.	71.1	67.2	2.74[*]
Oct.	78.7	78.7	1.69[NS]
Nov.	82.3	76.6	2.07[NS]
Dec.	83.6	82.9	0.48[NS]

CONCLUSION

Relative humidity is of important parameters in agriculture which could cause increase in fungal diseases in cultivated plants. Therefore, considering the obtained results here about increasing trend of relative humidity in May, June and September months for this station in coming years, it is better for farmers in this region to get prepared to fight against probable outbreak of plant diseases. Since, the predicted values for far years are very close to each other in monthly scale then it is better to repeat the forecast every few years.

ACKNOWLEDGEMENTS

REFERENCES

Al-Awadhi, S. and I. Jolliffe, 1998. Time series modeling of surface pressure data. Int. J. Climatol., 18: 443-455.

Azar, A. and M. Momeny, 2001. Statistics and its Application in Management. Vol. 2, SAMT Publications, Tehran, Iran, pp: 390.

Box, G.E.P. and G.M. Jenkins, 1976. Time Series Analysis: Forecasting and Control. 3rd Edn., Holden Day, San Francisco.

Jahanbakhsh, S. and A.A.B. Baser, 2003. The investigation and forecasting of average monthly temperature of Tabriz with ARIMA modal. Geog. Res., 18: 34-46.

Jahanbakhsh, S. and S. Torabi, 2004. The investigation and forecasting of temperature and precipitation of Iran. Geog. Res., 19: 104-125.

Jameei, J., 2002. Analysis of weather characteristics of Sanandaj and Miyandoab stations and ARMA application in forecasting drought. M.A. Thesis, Tabriz University, Iran.

Leite, S.M. and J.P. Peixoto, 1996. The Autoregressive model of climatological time series an application to the longest time series in Portugal. Int. J. Climatol., 16: 1165-1173.

Modares, R. and S. Eslamiyan, 2004. The parameter estimation and detection of model accuracy for time series models. First Congress of Civil Engineering, Sanat Sharif University, Tehran, Iran.

Moshkani, M.R., 1994. Time Series Analysis: Forecasting and Control. Vol. 1 and 2, Shahid Beheshti University Publications, Tehran, Iran, pp: 424.

Ramezani, N., 2001. Analysis and prediction of drought and wet years in Mazandaran province. M.A. Thesis, Trbiyat Moelm University, Tehran, Iran.

Taheri, M., 1998. The Modeling and Forecasting of Temperature and Rainfall for Eleven Weather Stations in Iran up to 2000. Deputy Education and Research Publications, Meteorological Organization, Tehran, Iran.