Int. J. Nonlinear Anal. Appl. 13 (2022) No. 1, 2911–2920
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.6023

Using ARIMA model and neuro-fuzzy approach to forecast the climatic temperature in Mosul-Iraq

Naam Salem Fadhil[a,*]

[a] *Department of Statistics and Informatics, Faculty of Computer Sciences and Mathematics, University of Mosul, Mosul, Iraq.*

(Communicated by Madjid Eshaghi Gordji)

Abstract

The accuracy of temperature forecasting in maximum and minimum cases is important to control the environmental effects. In this study, integrated autoregressive and moving average (ARIMA) model is used to forecast climatic temperature variable in maximum and minimum cases in Mosul, Iraq as traditional method. Neuro-Fuzzy (NF) is also proposed as modern approach to improve the forecasting results. The results in this study reflect outperforming in forecasting for NF approach comparing to ARIMA model. In conclusion, NF approach can be used for more accuracy to forecast climatic temperature datasets in maximum and minimum cases.

Keywords: ARIMA model, Neuro-Fuzzy (NF), Forecasting, climatic temperature.

1. Introduction

The forecasting of climatic data is a complicated process. That may belong to the seasonality changes and the non-homogeneity nature through different seasons. High quality forecasting results of maximum and minimum temperature may depend on suitable choice of methods those will be used for forecasting. ARIMA models is used in this study as a classical statistical approach for temperature forecasting. Following the methodology of Box-Jenkins will be the optimal to find the appropriate ARIMA model for maximum and minimum temperature. ARIMA model is a common traditional statistical method can be used for acceptable forecasting result with univariate time series data [1]. Ksiksi and Al-Blooshi [2] suggested ARIMA model as valid model to forecast the extreme events temperature in in the United Arab Emirates (UAE)..

*Corresponding author
Email address: naamsalem@uomosul.edu.iq, naamsalem87@gmail.com (Naam Salem Fadhil)

Received: May 2021 *Accepted:* October 2021

Some intelligent methods such neuro-fuzzy may be suggested to improve the forecasting results by handling the non-homogeneity nature in meteorological data which may lead to nonlinearity. Adaptive neuro-fuzzy inference system (ANFIS) is adaption of neuro-fuzzy for high quality forecasting results such as in [3] for daily mean temperature forecasting.

Several previous studies such as in [4, 5] for weather forecasting present good comparisons between ARIMA and ANFIS as forecasting approaches.

In this study, maximum and minimum temperature data have been taken from two types of months: hot months and cold months. November, December, January, February, and March are assumed as cold months, while May, June, July, August, and September are assumed as hot months. Commonly, dataset is divided into two periods: training and testing for forecasting purposes. The dataset of cold season includes 243 daily observations for (November\ 2018, December\ 2018, January\ 2019, February\ 2019, March\ 2019, November\ 2019, December\ 2019, and January\ 2020) for training period and 60 daily observations for (February\ 2020, and March\ 2020) for testing period. The dataset of hot season includes 292 daily observations for (May – September\ 2018 and 2019) for training period and 80 daily observations (1\5\2020 – 19\7\2020) for testing period.

2. Material and Method

2.1. Data and framework of the study

In this study, Iraqi daily datasets of maximum and minimum temperature will be studied for 243 daily observations for cold training period and 60 daily observations for cold testing period. Hot data includes 292 daily observations for training period and 80 daily observations for testing period. The framework of this study includes the following:

a. Dividing the full period into two groups for cold and hot months.

b. Dividing each group into two periods for training and testing.

c. Modeling training data by using ARIMA model.

d. Simulating testing data by using ARIMA model.

e. Modeling training data by using ANFIS method.

f. Simulating testing data by using ANFIS methods.

g. Comparing ARIMA model as traditional approach and ANFIS as intelligent method to determine the highest accuracy of forecasting.

2.2. Autoregressive integrated moving average (ARIMA) model

An ARIMA model is used in this study to forecast maximum and minimum temperature datasets. Box-Jenkins methodology includes the identification, estimation, diagnostic checking, and forecasting stages followed to obtain optimal ARIMA model. A general expression of the seasonal model ARIMA(p,d,q)(P,D,Q)s is as follows.

$$\phi(B)\Phi(B)(1-B)^D(1-B)^d Z_t = \theta(B)\Theta(B)a_t \tag{2.1}$$

$$\phi(B)\Phi(B)W_t = \theta(B)\Theta(B)a_t \tag{2.2}$$

where

$$W_t = (1-B)^D(1-B)^d Z_t,$$
$$\phi(B) = (1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p),$$
$$\Phi(B) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}),$$
$$\theta(B) = (1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q),$$

and

$$\Theta(B) = (1-\Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}).$$

where Z_t is a time series variable, t is an integer value as index of event's time, B is a backshift operator, s is a seasonal period, ϕ_k and $\theta_k$are the k th parameter of autoregressive and moving average respectively, Φ_k and $\Theta_k$are the k th parameter of seasonal autoregressive and seasonal moving average respectively, p, q, P and $Q$are the number of parameters or the ranks of autoregressive, moving average, seasonal autoregressive and seasonal moving average respectively, d and $D$are the number of successive differences and seasonal differences respectively, and a_{t-k} is the independent and identically distributed random error $a_t \sim i.i.d.N(0, \sigma_a^2)$.

Box-Jenkins methodology has four successive steps: identification, estimation diagnostic checking, and forecasting. The accurate forecasting has an important role in different fields. The first step in Box-Jenkins methodology is the identification which includes satisfying the mean and variance stationarity. The stationarity can also be detected via time series data plot, and the figures of ACF and PACF. Slow dying out pattern of ACF or PACF plots indicate to non-stationary series. Power transformation, successive differencing, and seasonal differencing can be used to satisfy the stationarity. After satisfying the stationarity, the next sted to complete the identification is to identify the ranks of ARIMA model p, q, P, Q to obtain the empirical ARIMA model such as in Table 1. Several empirical time series models can be identified by observing the plots of ACF and PACF.

Table 1: Identifying the suitable rank of time series

Model	ACF	PACF
AR(p)	Dies out	Cuts off after lag p
MA(q)	Cuts off after lag q	Dies out
ARMA(p,q)	Dies out but goes to zero after lag q	Dies out but goes to zero after lag p
SAR(P)	Dies out	Cuts off after lag sP
SMA(Q)	Cuts off after lag sQ	Dies out
SARMA(P,Q)	Dies out but goes to zero after lag sQ	Dies out but goes to zero after lag sP

After identifying several empirical time series models, the parameters of time series model can be estimated using maximum likelihood estimation (MLE) method. To check the validation of the estimated model, the significance of parameters and the insignificance of residual series in ACF plot can be used to ensure satisfying model assumptions and determine the most fitted ARIMA model. The residual series in ACF plot should be insignificant for all lags, while the marginal lags can be acceptable. The first step-ahead point forecast for ARIMA model is depended in this study by using minimum mean square error (MMSE) approach.

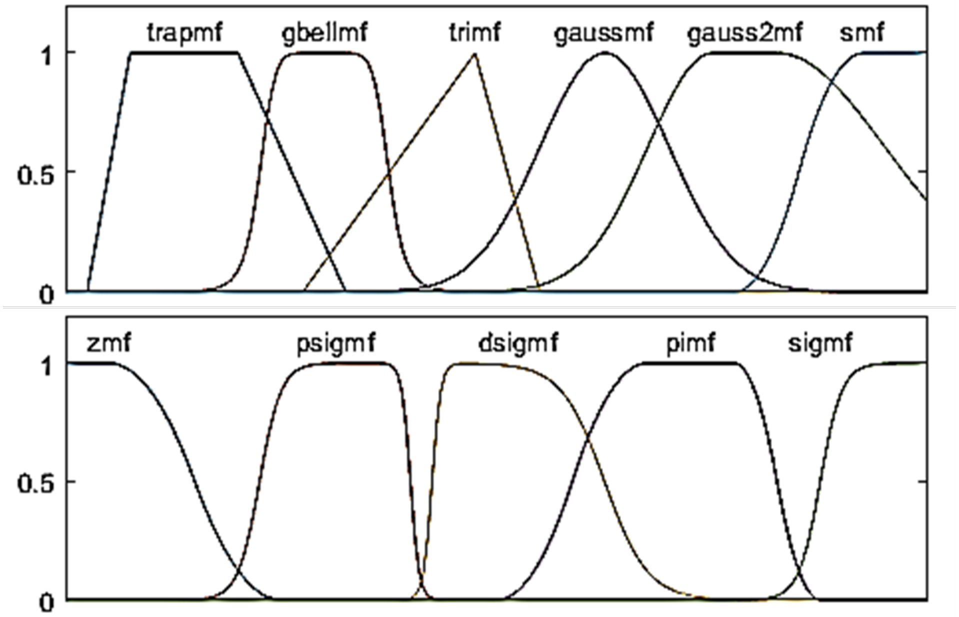

Figure 1: MF carves for ANFIS [9].

2.3. Adaptive neuro-fuzzy inference system (ANFIS).

ANFIS had been presented by [6] and can be expounded as a multilayer feed forward neural network through inputs to outputs with fuzzy inference system (FIS) [7]. The main stage in ANFIS is to build FIS structure based on Sugeno-type which presented by [8]. Building FIS includes determining the method for generating FIS, optimization method, tolerance of error, the type of membership functions (MF) and parameters, rules, and others. First order Sugeno-type in consists of a set of two if–then rules such as follows.

Rule 1 includes: If x is equal to A_1 and y is equal to B_2 then $p_1x + q_1y + r_1 = f_1$.

Rule 2 includes: If x is equal to A_2 and y is equal to B_2, then $p_2x + q_2y + r_2 = f_2$.

where A_1, B_1, A_2 and B_2 are the MF for the input data, while p_1, q_1, r_1, f_1,p_2, q_2, r_2, and f_2 are the parameters of MF of output. The number of MF can be determined via trial and error principle. The most common MF types are trapezoidal, generalized bell, triangular, sigmoidal, *S*, *Z,* and *Pi* curves, and two different curves of Gaussian in addition to the product of two sigmoidal membership functions. The MF are presented in Figure 1 such as follows.

After building FIS, ANFIS will be accomplished when the input layer will be structured such in neural network. Input variables can be determined based on autoregressive (AR) principles for time series in the structure of neuro-fuzzy such as in neural network which will be a simple input structure according to [10, 11, 12].

Mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) are used in this study as criteria of forecasting error. MAE, MAPE, RMSE can be written such as follows [13].

$$RMSE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \,. \tag{2.3}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{e_i}{\mathrm{Z}_i}\right| \times 100 \tag{2.4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i)^2}. \tag{2.5}$$

where e_i is the random forecasting error, n is the number of observations, and y_i is the original time series variable [13].

3. Results and Discussion.

In this study, maximum and minimum temperature data have been taken from two types of months: hot and cold months. The dataset of cold season includes 243 daily observations for training period and 60 daily observations for testing. The dataset of hot season includes 292 daily observations for training and 80 daily observations for testing.

3.1. ARIMA model

The best empirical ARIMA models of maximum and minimum Iraqi temperature are such as in Table 2 below.

Table 2: Optimal empirical ARIMA model.

Dataset	Maximum	Minimum
Hot months	$ARIMA(0,2,2)(0,1,1)_5$	$ARIMA(0,1,1)(0,1,1)_5$
Cold months	$ARIMA(1,1,1)(0,1,1)_5$	$ARIMA(0,1,1)(0,1,1)_5$

All the parameters of empirical ARIMA models in Table 2 are significant and the ACF plots of the residuals of these models indicate to insignificants residual autocorrelation and satisfy the assumptions. Table 3 clears that all parameters of ARIMA models are significant with p values smaller than the significant level of 5% for maximum and minimum Iraqi temperature data.

Table 3: The significance of parameters in ARIMA models.

Dataset	Maximum			Minimum		
	Type	**Parameter**	**p-value**	**Type**	**Parameter**	**p-value**
Hot months	MA 1	0.7501	0.000			
	MA 2	0.2226	0.000	MA 1	0.6478	0.000
	SMA 1	0.9445	0.000	SMA 1	0.9580	0.000
Cold months	AR 1	0.6099	0.000			
	MA 1	0.8525	0.000	MA 1	0.2382	0.000
	SMA 1	0.9566	0.000	SMA 1	0.9461	0.000

For the ARIMA models, the measurements of error (MAPE, RMSE, and MAE) are used to reflect the forecasting accuracy. Table 4 and Table 5 contain the measurements of error for training and testing temperature respectively for maximum and minimum data.

Table 4: MAPE, RMSE, and MAE for training maximum and minimum temperature by using ARIMA model.

Dataset	MAPE	RMSE	MAE
Maximum hot	3.4826	1.7847	1.3832
Minimum hot	4.2341	2.1209	1.7128
Maximum cold	10.2824	2.1393	1.6441
Minimum cold	9.7458	2.0982	1.6109

Table 5: MAPE, RMSE, and MAE for testing maximum and minimum temperature by using ARIMA model.

Dataset	MAPE	RMSE	MAE
Maximum hot	38.4626	18.5607	15.9733
Minimum hot	13.6578	6.6398	5.5391
Maximum cold	45.1769	7.9787	7.1976
Minimum cold	32.4178	5.4109	4.8522

From Table 4 and Table 5, the training forecasts of ARIMA model indicated that the forecasting results of hot months outperform the forecasting results of cold. In general, the results of training forecasts outperform the results of testing forecasts because ARIMA models were built by training datasets.

3.2. Neuro-fuzzy Approach

In this study, ANFIS is applied to forecast maximum and minimum temperature time series based on the principle of autoregressive of time series as mentioned in previous section. Autoregressive of 3^{rd} and 4^{th} ranks have been depended to build the inputs structures of ANFIS. Therefore, two types of ANFIS model have been created. First one has 3 lags of original series as inputs, while second one has 4 lags of original series as inputs. If z_t assumed the original time series variable of maximum or minimum temperature, the inputs of 3 lags and 4 lags will be represented as in (3.1) and (3.2) equations bellow respectively.

$$\{z_{t-1}, z_{t-2}, z_{t-3}\} \tag{3.1}$$

$$\{z_{t-1}, z_{t-2}, z_{t-3}, z_{t-4}\} \tag{3.2}$$

The target of ANFIS is the original time series variable of maximum or minimum temperature z_t. The output of ANFIS model $\hat{z}_t$will be compare to the original time series to obtain the residual or the forecasting error such as in equation (3.3) below.

$$e_t = z_t - \hat{z}_t \tag{3.3}$$

MATLAB is used to forecast maximum and minimum temperature by using ANFIS. The framework for forecasting by using ANFIS in MATLAB can be detailed practically such as follows.

1. The inputs and target variables should be entered for training and testing separately into workspace of MATLAB as successive columns in one variable, last column should be specified for target variable.

2. In ANFIS training process, error tolerance, optimization method, and epochs should be appointed.

3. The output variables of training and testing processes represent ANFIS forecast.

Figure 2 till Figure 5 represent the fitness between original series and forecast series by using ANFIS for training and testing where inputs are 3 lags series.

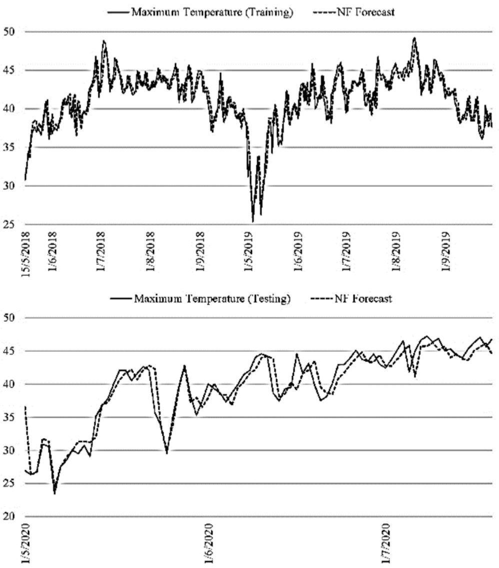

Figure 2: The fitness between original and forecasted series by using (ANFIS/3 inputs) for hot maximum training and testing series.

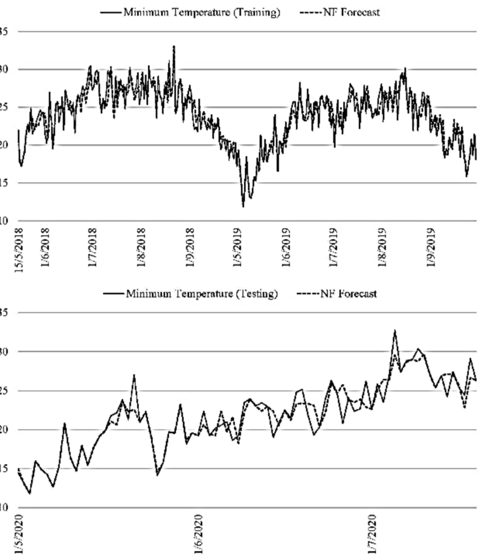

Figure 3: The fitness between original and forecasted series by using (ANFIS/3 inputs) for hot minimum training and testing series.

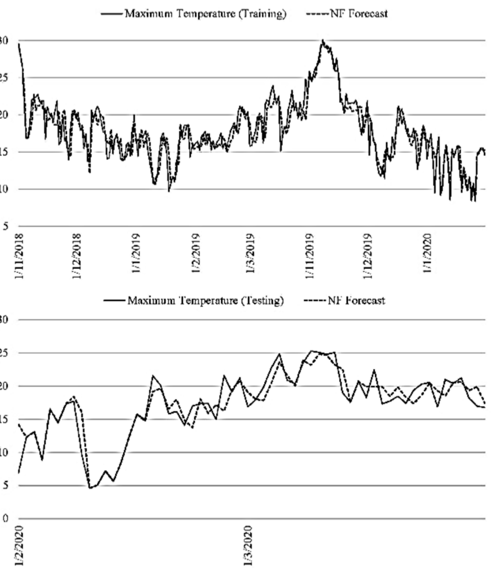

Figure 4: The fitness between original and forecasted series by using (ANFIS/3 inputs) for cold maximum training and testing series.

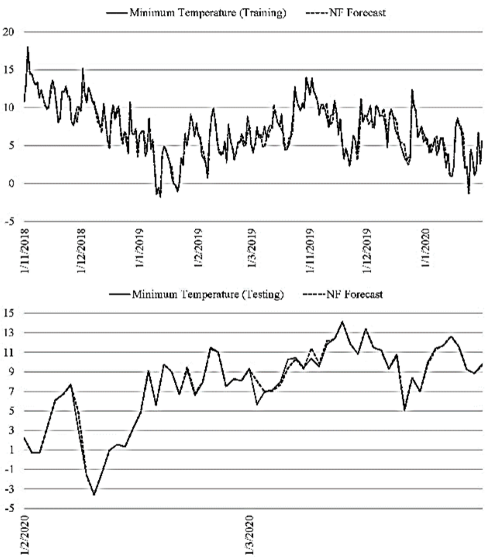

Figure 5: The fitness between original and forecasted series by using (ANFIS/3 inputs) for cold minimum training and testing series.

Figure 6 till Figure 9 represent the fitness between original series and forecast series by using ANFIS for training and testing where inputs are 3 lags series.

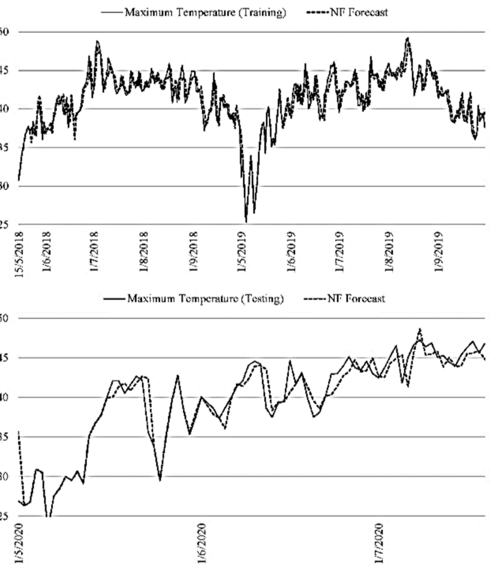

Figure 6: The fitness between original and forecasted series by using (ANFIS/3 inputs) for cold maximum training and testing series.

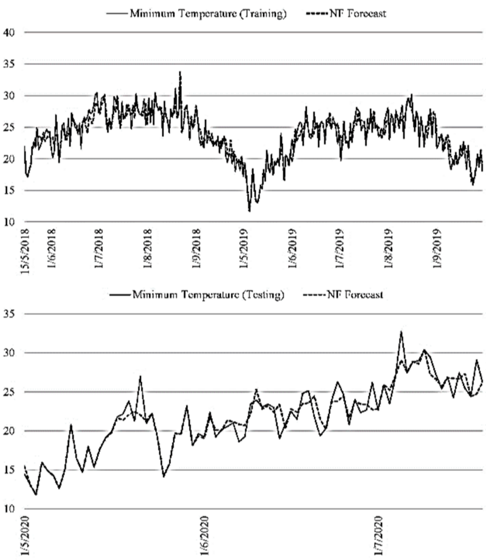

Figure 7: The fitness between original and forecasted series by using (ANFIS/4 inputs) for hot minimum training and testing series.

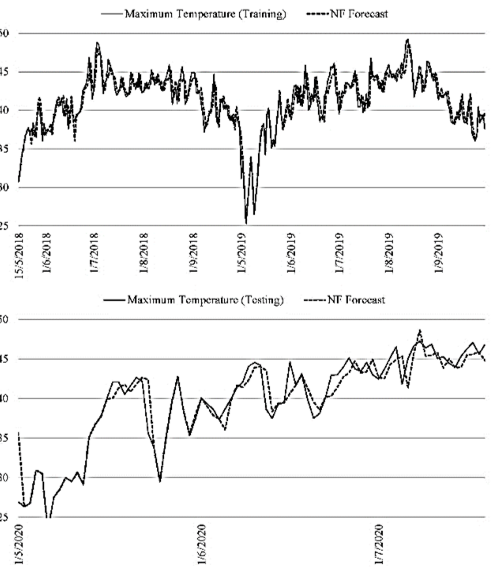

Figure 8: The fitness between original and forecasted series by using (ANFIS/4 inputs) for cold maximum training and testing series.

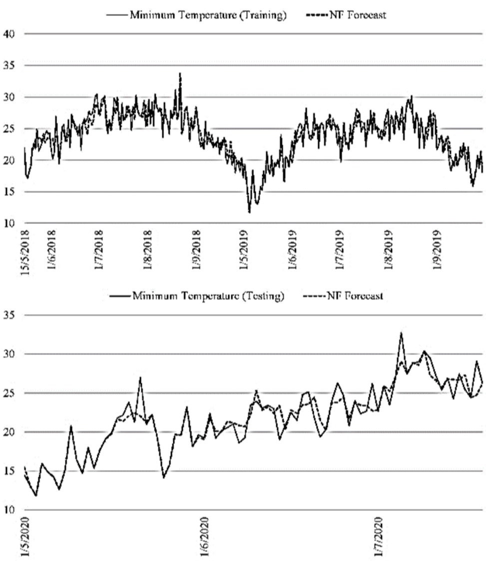

Figure 9: The fitness between original and forecasted series by using (ANFIS/4 inputs) for cold minimum training and testing series.

For Figure 2 till Figure 9, ANFIS outputs which represent the forecast series have high fitness with original series. This also reflect high quality forecasting by using ANFIS for training and testing periods. The fitness between original and forecasts by using ARIMA are unacceptable, therefore this fitness can't be plotted. For ANFIS forecasting results, (MAPE, RMSE, and MAE) are also used to reflect the forecasting accuracy. Table 6 and Table 7 include the measurements of error for training and testing temperature respectively for maximum and minimum data by using ANFIS where inputs are 3 lags series.

Table 6: MAPE, RMSE, and MAE for training maximum and minimum temperature by using (ANFIS/3 inputs).

Dataset	**MAPE**	**RMSE**	**MAE**
Maximum hot	2.7743	1.4845	1.1285
Minimum hot	3.6806	1.2735	0.8873
Maximum cold	6.9987	1.6214	1.1551
Minimum cold	6.577	0.6692	0.3513

Table 7: MAPE, RMSE, and MAE for testing maximum and minimum temperature by using (ANFIS/3 inputs).

Table 7: MAPE, RMSE, and MAE for training maximum and minimum temperature by using (ANFIS/3 inputs).

Dataset	**MAPE**	**RMSE**	**MAE**
Maximum hot	3.5972	2.0695	1.3677
Minimum hot	3.5799	1.4117	0.8164
Maximum cold	8.2823	1.9872	1.2617
Minimum cold	2.7336	0.4412	0.16

Table 8 and Table 9 include the measurements of error for training and testing temperature respectively for maximum and minimum data by using ANFIS where inputs are 4 lags series.

Table 8: MAPE, RMSE, and MAE for training maximum and minimum temperature by using (ANFIS/4 inputs).

Dataset	**MAPE**	**RMSE**	**MAE**
Maximum hot	2.3288	1.2924	0.9668
Minimum hot	3.6734	1.2311	0.8772
Maximum cold	5.8111	1.3547	0.9602
Minimum cold	6.5944	0.6552	0.3364

Table 9: MAPE, RMSE, and MAE for testing maximum and minimum temperature by using (ANFIS/4 inputs).

Dataset	**MAPE**	**RMSE**	**MAE**
Maximum hot	2.5836	1.7957	1.0187
Minimum hot	3.3862	1.3803	0.7952
Maximum cold	6.1211	1.7186	0.8585
Minimum cold	2.6859	0.43	0.1997

By comparing (Table 4 and Table 5) to (Table 6 till Table 9), the training and testing forecasting results by using ANFIS outperform the training and testing forecasting results by using ARIMA model. Table 4, Table 6, and Table 8 for training period indicated that the forecasting results of hot months outperform the forecasting results of cold. Table 4 Table 6, and Table 8 for training period also indicated that the forecasting results of the maximum temperature of hot months outperform the forecasting results of minimum temperature of hot months.

4. Conclusions

A neuro-fuzzy or ANFIS approach was proposed to improve the accuracy forecasting of maximum and minimum temperature in Mosul city\ Iraq. Datasets divided into two season: season of hot month and season of cold months. The results showed that ARIMA and ANFIS were effective. However, the MAPE, RMSE, and MAE results indicated that ANFIS approach was the most effective tool for improving the accuracy of maximum and minimum temperature forecasts. The advantages of the ANFIS were improving the forecasting results and handling the non-homogeneity and non-linearity. ANFIS can be used for high accuracy of forecasting results for maximum and minimum temperature comparing to traditional statistical models such as ARIMA.

References

[1] O.B.Shukur, and M.H. Lee, *A Review of using Neural Network and Kalman Filter based on ARIMA for Wind Speed Forecasting*, Iraqi J. Stat. Sci., 16(29),(2019).

[2] T.S. Ksiksi, and L.S. Al-Blooshi, *Climate change in the UAE: Modeling air temperature using ARIMA and STI across four bio-climatic zones*, F1000Research, 8(973) ,(2019)973.

[3] M. Göçken, and A. Boru, *Integrating metaheuristics and ANFIS for daily mean temperature forecasting*, Int. J. Global Warming, 9(1), (2016)110-128.

[4] N.R. Babu, et al, *Comparison of ANFIS and ARIMA model for weather forecasting*, Indian J.Sci.Technol., 8(S2), (2015) 70-73.

[5] P.Srikanth, D.R. Rao, and P. Vidyullatha, *Comparative analysis of ANFIS, ARIMA and polynomial curve fitting for weather forecasting*, Indian J. Sci. Technol., 9(15), (2016)1-6.

[6] J.Jang, *ANFIS: adaptive network-based fuzzy inference system*, IEEE T. Syst. Man Cyb., 23 (3), (1993)665-685.

[7] M.Firat, and M. Güngör, *River flow estimation using adaptive neuro fuzzy inference system*, Math. Comput.Simul., 75(3-4), (2007) 87-96.

[8] M.Sugeno, and T. Takagi, *Fuzzy identification of systems and its applications to modelling and control*, Readings Fuzzy Sets Intell. Syst., 15(1), (1993)387-403.

[9] F.Cavallaro, *A Takagi-Sugeno fuzzy inference system for developing a sustainability index of biomass*, Sustainability, 7(9), (2015)12359-12371.

[10] H.Liu, , H.-q. Tian, and Y.-f. Li, *Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction*, Appl. Energy, 98, (2012)415–424.

[11] S.Tatinati, and K.C. Veluvolu, *A hybrid approach for short-term forecasting of wind speed*, Sci. World J., 2013,(2013).

[12] K.Chen, and J. Yu, *Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach*, Appl. Energy., 113, (2014)690-705.

[13] R.J.Hyndman, and A.B. Koehler, *Another look at measures of forecast accuracy*, Int. J. Forecasting, 22(4), (2006) 679-688.