

## Forecasting Air Temperatures Using Time Series Models and Neural-based Algorithms

<sup>1</sup>Mahmoud M. Smadi and <sup>2</sup>Farouq S. Mjalli

<sup>1</sup>Department of Mathematics and Statistics, Jordan University of Science and Technology  
P.O. Box 3030 Irbid, 22110, Jordan

<sup>2</sup>Department of Chemical Engineering, University of Qatar, P.O. Box 2713 Doha, Qatar

---

**Abstract:** Recently, neural network models have been popular and found useful for forecasting a wide variety of time series data in many disciplines. This is due to their favorable modeling properties of simplicity, fault and noise tolerance and their capability to adapt to process changes. Nevertheless, applications in climatology have been less widespread than other disciplines such as economics. In this study, feed-forward neural-network (FFNN) and autoregression (AR) time series models are used in forecasting the annual air temperature time series data in Jordan. The performance of the two predictors was compared using out-of-sample forecasts. The test period was shifted through the whole available time. As demonstrated by the forecasting experiments, the FFNN models gave better forecasts and were able to identify the dynamics of the temperature time series and gave more realistic forecasts. Both predictors showed a cooling trend in annual air temperatures for the coming 10 years.

**Key words:** Forecasting, Autoregression, Neural Network, Air Temperature

---

### INTRODUCTION

Recently, neural network models have been popular and found useful in forecasting a wide variety of historical data, although applications in climatology have been less widespread than other disciplines, even though climatic data often takes the form of series measures over time. Forecasting global temperature variations by neural networks was implemented using regularization network, multilayer perceptrons, linear autoregression and a local model known as the simplex projection method<sup>[1]</sup>. The forecasting results are consistent with the hypothesis that the climate dynamics is characterized by low-dimensional chaos and that it may have changes at some point after 1965, which is also consistent with the recent idea of climate change.

Different neural-network prediction algorithms have been proposed in the literature and used in many disciplines. Lapedes and Farber<sup>[2]</sup> reported that simple neural networks can outperform conventional methods for identifying historical data. Anderson *et al.*<sup>[3]</sup>, compared different recurrent training algorithms for the identification of time series data. Gómez-Ramírez *et al.*<sup>[4]</sup> implemented an adaptive algorithm of the architecture of Polynomial Artificial Neural Network (PANN) using Genetic Algorithm (GA) to improve the

learning process. The performance of this algorithm was compared to a multilayer perception network. Peter Zhang and Min Qi<sup>[5]</sup> considered the issue of how to effectively model time series with both seasonal and trend patterns. They concluded that detrending and preprocessing of data is essential for obtaining reliable predictions. Other applications of neural networks in this area can be found elsewhere<sup>[6-8]</sup>.

In the Middle East, investigations of long term variations in temperature data are not receiving enough attention even though, these countries suffer serious environmental, agricultural and water resources problems. In this work, feed-forward neural-network (FFNN) and autoregression (AR) time series models were used in forecasting the annual mean temperatures, annual mean minimum temperatures and annual mean maximum temperatures during the period 1923-2003 in Amman Airport station, Jordan. The Amman station is a strategic and historical station in the Middle East, due to its location, reliability and length of the record. The performance of the two predictors was compared by out-of-sample forecasts and examining the mean square errors. Finally, forecasts of the annual mean temperature mean minimum temperature and mean maximum temperature for the coming 10 years at Amman station in Jordan are given using the two prediction models.

---

**Corresponding Author:** Mahmoud M. Smadi, Department of Mathematics and Statistics, Jordan University of Science and Technology, P.O. Box 3030 Irbid, 22110, Jordan Fax: (+962) 7095014

**MATERIALS AND METHODS**

In the following, the feed-forward neural-network (FFNN) algorithms and Autoregression (AR) time series models were described.

**Feed-forward neural networks (FF-NN):** When used as a black-box modeling tool, the FF-NN artificial neural network structure approximates the behavior of certain mechanistic phenomena by the use of optimization search techniques. This is accomplished into two steps; the network is fed with the input-output data of the actual process to be modeled and then these data are used to train the network for emulating the training data.

Basically, the neural networks are constructed in layers namely; input layer, hidden layer(s) and an output layer. Each layer is composed of one or more neurons. The neuron is the building block of these networks. The neuron, has a scalar input  $p$  which is passed through a connection that multiplies its strength by the scalar weight  $w$ , to form the product  $wp$ . The product  $wp$  is fed as an argument to a transfer function  $f$ , which produces the scalar output  $a$ . A scalar bias,  $b$ , of unity value is added to the product  $wp$ . Thus, the neuron has two inputs  $p$  and  $b$  and one output  $a$ . This structure can be written mathematically as:

$$a = f(wp + b) \tag{1}$$

The transfer function which is some times called the activation function may take different mathematical forms. Some of these commonly used forms are; the *linear*, *log-sigmoid*, *tan-sigmoid* and the *radial basis* transfer functions. Selecting a proper activation function depends on the application used and network structure.

Neurons are stacked to form one layer as shown in Fig. 1. The  $R$  input vector  $p$  is fed to the first layer. These inputs are multiplied by an  $S \times R$  weight matrix  $W$  and an  $S$  bias vector  $b$  is added to the product to produce an output vector  $a$ . This can be expressed as:

$$a = f(Wp + b) \tag{2}$$

For  $n$  layers connected in series, the output vector at a certain time step  $k$  can be expressed as:

$$a^n = f^n \left[ W_{n,n-1} f^{n-1} \dots \left[ W_{1,1} f^1 (W_{1,1} p + b^1) + b^2 \right] \dots + b^n \right] \tag{3}$$

Thus, for a 3 layer network equation 12 becomes:

$$y = a^3 = f^3 \left[ W_{3,2} f^2 \left[ W_{2,1} f^1 (W_{1,1} p + b^1) + b^2 \right] + b^3 \right] \tag{4}$$

where  $y$  represents the network product vector. As shown in Fig. 1, the input vector  $p$  to the FF-NN network architecture consists of the recent input  $u(k)$  and the  $n$  past inputs  $u(n)$ ,  $n=k-1, \dots, k-n$ .

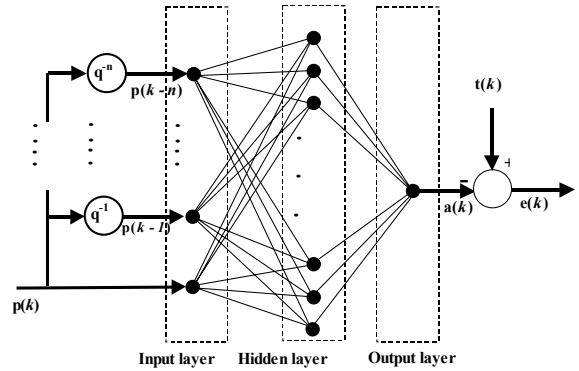


Fig. 1: Structure of the feedforward neural-network

The network is fed with a set of input-output pairs and trained to reproduce the outputs with a predefined degree of tolerance. For the case of time series data, the input is a vector of  $n+1$  values and the output is a single value at the  $k+1$  incidence of time. Network training is done by adjusting the neurons weights using an optimization algorithm to minimize the quadratic error between observed data and computed outputs.

Input-target training data are usually pretreated in order to improve the numerical condition of the optimization problem and to make better behavior of the training process. The input-target data is normally subdivided into three subsets namely; *training*, *validation* and *testing* subsets. The *training* subset data is used to accomplish the network learning and fit the network weights by minimizing an appropriate error function. *Backpropagation* is the training technique usually used for this purpose. It refers to the method for computing the gradient of the case-wise error function with respect to the weights for a feedforward network. The performance of the networks is then compared by evaluating the error function using the *validation* subset data, independently. The *testing* subset data is then used to measure the generalization of the network (i.e. how accurately the network predicts targets for inputs that are *not* in the training set) this is some times referred to as *hold-out* validation.

**Autoregressive models:** Assuming  $\{X_t, t=1, \dots, n\}$  is a zero mean stationary time series, a popular and wide class of time series models is the autoregressive models,  $AR(p)$ , where  $p$  is nonnegative integer and the order of the autoregressive model. The model is given in a form of difference equation. This model can be represented as<sup>[9]</sup>

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t, \quad t = 1, \dots, n \tag{5}$$

where  $Z_t$  is a sequence of uncorrelated random variables, with zero mean and variance  $\sigma^2$  and is written as  $Z_t \sim WN(0, \sigma^2)$ , ( $WN$  stands for white noise).

The model could be written in the following form:

$$\Phi(B)X_t = Z_t, \tag{6}$$

where  $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ , and  $B^k X_t = X_{t-k}$ .

Important properties of the AR(p) models like stationarity and causality depend on the location of the roots of the characteristic polynomial  $\Phi(B)$ <sup>[9]</sup>. The following four steps are required to build an AR model:

- (i) Model identification.
- (ii) Estimation of the model parameters  $\varphi_1, \dots, \varphi_p$  and  $\sigma^2$ .
- (iii) Checking the Model.
- (iv) Forecasting.

Model parameters estimation can be achieved using different methods. Among these methods, Yule-Walker estimates, which are moment type estimates, least squares estimates, conditional maximum likelihood estimates, and maximum likelihood estimates<sup>[9,10]</sup>. The maximum likelihood estimator which is used in this work is one of the popular estimation methods. For the AR(p) model given in (5), assuming independent identically normally distributed white noise  $\{Z_t\}$ , the likelihood function is given by Brockwell and Davis<sup>[9]</sup>.

$$L(\varphi_1, \dots, \varphi_p, \sigma^2, x_1, \dots, x_n) = (2\pi)^{-n/2} |\det \Gamma_n|^{-1/2} \exp[-\underline{x}' \Gamma_n^{-1} \underline{x}] \tag{7}$$

Where  $\underline{x} = (x_1, \dots, x_n)'$  and  $\Gamma_n$  is the matrix of covariance's;

$$\Gamma_n = [\gamma(i-j)]_{i,j=1}^n \tag{8}$$

This can be computed from (1). The maximum likelihood estimators  $\hat{\varphi}_1, \dots, \hat{\varphi}_p, \hat{\sigma}^2$  are the values of  $\varphi_1, \dots, \varphi_p, \sigma^2$  which maximize (7) for given observed data. Different optimization algorithms are available in the time series literature to numerically estimate the parameters  $\varphi_1, \dots, \varphi_p, \sigma^2$ . The ITSM software package was used in this study to estimate these parameters.

The AICC (a bias-corrected form of Akaike's information criterion) is chosen as a selection criterion of the order of AR(p) model by selecting the value of p that minimize

$$AICC = -2 \ln L(\varphi_1, \dots, \varphi_p, \sigma^2; x_1, \dots, x_n) + \frac{2(p+1)n}{n-p-2} \tag{9}$$

Before accepting the fitted model and developing forecasts of future values, the model must be validated by reconciliation its predictions with actual time series data.

## RESULTS AND DISCUSSION

The fitted autoregression models of annual mean minimum, annual mean and annual mean maximum temperatures based on the minimum AICC criterion are given as follows:

The fitted model of annual mean minimum temperature is

$$X_t = 0.4457X_{t-1} + 0.1662X_{t-2} + 0.2266X_{t-3},$$

$$\hat{\sigma}^2 = 0.3029.$$

The fitted model of annual mean temperature is

$$X_t = 0.2303X_{t-1} + 0.0809X_{t-2} + 0.2572X_{t-3},$$

$$\hat{\sigma}^2 = 0.3203.$$

The fitted model of annual mean maximum temperature is

$$X_t = 0.2107X_{t-1} + 0.0990X_{t-2} + 0.2187X_{t-3},$$

$$\hat{\sigma}^2 = 0.4985.$$

Where  $\hat{\sigma}^2$  is the estimated white noise variance.

The fitted models are of the same order, AR(3). This indicates that there was a fairly consistent behavior in the annual mean, annual mean minimum and annual mean maximum temperature data. None of the two *portmanteau* tests (*Ljung-Box* and *McLead-Li*) rejects the hypothesis of i.i.d residuals at level 0.05 and the minimum-AICC Yule Walker AR model for the residuals is of order zero.

The FFNN modeling was performed using the following procedure; *data preprocessing, network structure selection, network training, early stopping, out of sample testing*. A series of exploratory experiments were performed to select the best NN structure. A three layers network was finally selected with 4 sigmoidal neurons in the input layer, 30 sigmoidal neurons in the hidden layer and 1 linear neuron in the output layer. Due to the time dependence of the data, the network output was fed back to the input as shown in Fig. 1. Three time delays were selected, making the total number of inputs four in any time prediction instance. In order to guarantee the generalization of the trained neural network and confirm the acceptance of the network performance over a wide range of process operating conditions, the network needs to be trained with data which covers the entire range of possible network inputs. For the process under consideration, the input-target range spans all the experiments data (the period 1923-2003). The original historical data set of inputs was subdivided into three subsets for network *training, validation* and *testing* in a ratio of 4:1:1, respectively. The inputs and the targets were normalized so that they have zero mean and unity

standard deviation. This would make the neural network training more efficient. Network training was accomplished by changing its weights and biases to achieve certain performance criteria. This is accomplished by using an optimization algorithm that searches for network parameters which minimizes the performance index.

The networks' training was stopped before reaching the performance criterion of  $1 \times 10^{-3}$ . This is because of the increase of the validation error. The normalized mean square error (MSE) attained for each network is less than 0.5, but the search gradients indicate that the training goal was not achieved completely. This is because the time series data is so complex and it can not be modeled with a high degree of efficiency such as the case with other function approximation applications. A certain degree of model deviation must be accepted in this case. The attained MSE for the three series give an indication of the good accuracy of prediction.

In forecasting, out-of-sample forecasts was implemented by a *leave-three-out* technique (need at least 3 years in an AR(3)). In this case we fit/learn with the total interval minus three years ( $81-3=78$  years) and test in the remaining three years. The test period was shifted through the whole available time. As demonstrated by the forecasting experiments, FFNN models gave the best forecasts. The FFNN and AR predictors for test period (78) years along with the time series plots of annual mean temperatures, annual mean minimum temperatures and annual mean maximum temperatures are shown in Fig. 2.

Since temperature memory decays, one expects that in a AR model the higher lag coefficient should be smaller than the lower lag regression coefficient. However, in the calculated models it is noticed that the third coefficient is larger than the second one for the three different time series. Also it is larger than the first coefficient for the mean temperature and slightly higher for the maximum temperature (0.2107 and 0.2187). One good reason for this contradicted trend in the values of the parameters can be attributed to the well known abrupt climatic changes in the time series behavior at this station. Major statistically significant change points in the mean minimum (night-time) and mean maximum (day-time) temperatures occurred in 1957 and 1967, respectively<sup>[11]</sup>. The presence of these abrupt changes might affect the natural behavior of the time series structure and a general conclusion about the comparative values of models parameters can not be drawn.

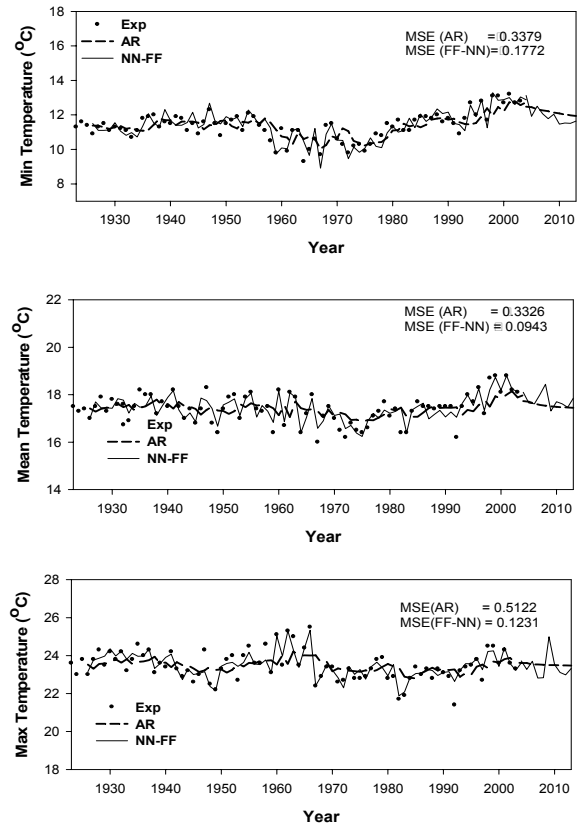


Fig. 2: FFNN versus Autoregression forecasts of annual air temperatures

The MSE for the two predictors are shown in Fig. 2. The MSE of the AR predictors are 0.3379, 0.3326 and 0.5122 for the annual mean, annual mean minimum and annual mean maximum temperatures respectively whereas the corresponding FFNN predictors MSE for the three time series are 0.1772, 0.0943 and 0.1231 respectively. It is clear that the MSE using the FFNN predictor are all through significantly smaller than the corresponding AR predictor, from which we conclude that the FFNN models gave the better forecasts than AR predictors. The AR models forecasts showed steady temperature profiles, on the other hand, the FFNN models were able to identify the dynamics of the temperature time series and gave more realistic forecasts. Generally, predicted annual temperatures shows steady rise from 1923 to 1950, then a decreasing (cooling) trend until late 1970's, followed by a further clear rise in temperature from late 1970's to the end of the record in 2003. These results agree generally with the trend patterns of the global annual mean temperatures by Smith<sup>[12]</sup>.

The forecasts for the future 10 years (2004-2014) are also shown in Fig. 2 for the three temperatures data

sets. The AR and FFNN models show a decreasing pattern in annual air temperatures for the coming 10 years, especially, the annual mean minimum temperature.

### CONCLUSION

As demonstrated by the forecasting experiments, FFNN models gave better forecasts than AR models. The neural network based models exhibited lower MSE as compared to the AR models. The AR models forecasts showed steady temperature profiles. On the other hand, the FFNN models were able to identify the dynamics of the temperature time series and gave more realistic forecasts. Finally, both predictors revealed a cooling trend in the annual air temperatures for the coming 10 years.

### REFERENCES

1. Miyano, T. and F. Girosi, 1994. Forecasting global Temperature variations by neural networks. The MIT Computer Science and Artificial Intelligence Laboratory, Report no. AIM-1447, CBCL-101.
2. Lapedes, A. and R. Farber, 1987. Nonlinear signal processing using neural network prediction and system modeling. Technical Report LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.
3. Anderson, J.S., I.G. Kevrekidis and R. Rico-Martinez, 1996. A comparison of recurrent training algorithms for time series analysis and system identification. *Computers Chem. Engg.*, 20 (Suppl.): S751-S756.
4. Gómez-Ramírez, E., A. Poznyak, A. González-Yunes and M. Avila-Alvarez, 1999. Adaptive architecture of polynomial artificial neural network to forecast nonlinear time series. Congress on Evolutionary Computation CEC99. Washington D.C. Jul. 6-9.
5. Peter, Z.G. and Q. Min, Neural network forecasting for seasonal and trend time series. *Eur. J. Oper. Res.*, 160: 501-514.
6. Windsor, C.G. and A.H. Harker, 1990. Multivariate financial index prediction-a neural network study. Proc. of Intl. Neural Network Conf., Paris, France, pp: 357-360.
7. Jagielska, I., 1993. The application of neural networks to business information systems. Proc. of 4th Australian Conf. on Information Systems, Brisbane, pp: 565-574.
8. Tangang, F.T., B. Tang, A.H. Monahan and W.W. Hsieh, 1998. Forecasting ENSO events: A neural network-extended EOF approach. *J. Climate*, 11: 29-41.
9. Brockwell, P.J. and R.A. Davis, 1991. *Time Series: Theory and Methods*, Springer, New York.
10. Box, G.E.P. and G.M. Jenkins, 1976. *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.
11. Smadi, M.M., 2006. Observed abrupt changes in minimum and maximum temperatures in Jordan in the 20<sup>th</sup> century. *Am. J. Environ. Sci.*, 2(3): 114-120.
12. Smith, T.M. and W.R. Reynolds, 2005. A global merged land-air-sea surface temperature reconstruction based on historical observations (1980-1997). *J. Climate*, 18: 2021-2036.