MAUSAM, **74**, 4 (October 2023), 1183-1190

MAUSAM

DOI : https://doi.org/10.54302/mausam.v74i4.5513
Homepage: https://mausamjournal.imd.gov.in/index.php/MAUSAM

UDC No. 551.509.313 : 551.524 : 004.032.26 : 519.233.5

Numerical modeling and forecasting temperature distribution by neural network and regression analysis

ADEEL TAHIR, MUHAMMAD ASHRAF, ZAHEER UDDIN*,

MUHAMMAD SARIM** and SYED NASEEM SHAH

Department of Physics, Federal Urdu University of Arts Science and Technology, Karachi

**Department of Physics, University of Karachi, Karachi,*

***Department of Computer Science, Federal Urdu University of Arts Science and Technology, Karachi*

(*Received 9 March 2022, Accepted 23 December 2022*)

e mail : adeel.tahir@fuuast.edu.pk

सार — पर्यावरणीय परिवर्तन विभिन्न मापदंडों के कारण होते हैं और भूमंडलीय उष्णता उन मापदंडों में से एक है। यह देखा गया है कि समय बीतने के साथ दैनिक औसत तापमान लगातार बढ़ रहा है। तापमान वितरण का ज्ञान हमें उस तथ्य को तय करने की अनुमति देता है जो तापमान परिवर्तन पर निर्भर करता है। 2018-2020 **के लिए तापमान परिवर्तन का पूर्वानुमान और मॉडल तैयार करने का प्रयास किया गया है। पाकिस्तान के सिंध (नवाबशाह) के शहरों में से एक के लिए दैनिक औसत तापमान का पूर्वानुमान करने के लिए कृत्रिम तंत्रिका नेटवर्क (एएनएन) और बहु समाश्रयण विश्लेषण का उपयोग किया गया है। दैनिक औसत तापमान का पूर्वानुमान करने के लिए** 2010 **से** 2020 **तक के पर्यावरणीय डेटा का उपयोग किया गया है। परिणामों की वैधता की जांच करने के लिए** RMSE, MABE **और** MAPE **जैसी सांख्यिकीय त्रुटियों और निर्धारण गुणांक** R^2 **की गणना की जाती है। दोनों मॉडल तापमान परिवर्तन के पूर्वानुमान के लिए उपयुक्त हैं; हालाँकि,** ANN **सर्वोत्तम परिणाम देता है। दो अलग-अलग समाश्रयण मॉडल (रैखिक और गैर-रैखिक) तापमान डेटा की संख्यात्मक फिटिंग के लिए कार्यरत हैं; गैर-रैखिक मॉडल बेहतर फिटिंग दर्शाता है।**

ABSTRACT. Environmental changes occur due to various parameters, and global warming is one of those parameters. It is observed that the daily mean temperature has constantly been increasing as time passes. The knowledge of temperature distribution allows us to decide the stuff that strongly depends upon temperature variation. An attempt has been made to model and forecast temperature distributions for 2018-2020. Artificial Neural Network (ANN) and multiple regression analyses have been used to forecast daily mean temperatures for one of Pakistan's cities of Sindh (Nawabshah). Environmental data from 2010 to 2020 has been used to predict daily mean temperature. The statistical errors such as RMSE, MABE and MAPE and coefficient of determination R^2 are calculated to check the results' validity. Both models are suitable for predicting temperature distribution; however, ANN gives the best result. Two different regression models (linear & non-linear) are employed for the numerical fitting of temperature data; the non-linear model shows the better fitting.

Key words – Artificial Neural Network (ANN), Multi regression analysis, Daily mean temperature, Nawabshah city.

1. Introduction

The daily mean temperature generally depends on Global solar radiation and other earth-based thermal processes. As one of the most important aspects of the natural world, climate influences natural and human life. Kaymaz (2005) and Ustaoglu *et al*. (2008) analyzed temperature as the main significant climatic parameter directly affecting evaporation, snow melting and frost and indirectly affecting atmospheric stability and precipitation conditions. Based on new climatic change studies, agriculture, vegetation, water management, and tourism are directly impacted by temperature changes. Effective weather forecasting is needed to avoid unanticipated temperature-related hazards such as frost and drought, resulting in financial and human losses. Various scientists have predicted daily mean temperatures for several parts of the world. The daily air temperature of 5 different stations in Italy for 20 years (1858 - 1977) was studied by Cuomo *et al*. (1986). Amato *et al*. (1989) designed a

Fig. 1. Geographical location of Nawabshah, Pakistan

mathematical model based on a composite series of 2 periodic variables, mean and standard deviation, and one stochastic parameter, stationary, using a fulfilled first-order autocorrelation model. Stochastic-dynamic models have discussed the air temperature and solar irradiance based on a 20-year daily time series for five stations to predict the Italian climate. Islam *et al*. (2009) observed that the Simple periodic models fit well the seasonality in both the mean and standard deviations of measured data. Also, first-order autoregressive processes with constant parameters accurately describe the short-term statistical variability of the studied variables. Nadia and Islam (2006) studied the temperature pattern and suggested an increase in Pakistan's maximum and minimum temperatures. Significantly changes are observed at minimum temperature as compared to maximum temperature. During summer, 21 Percentile represented the daily minimum temperature becoming warmer than the increase in daily maximum temperature. During winter, the change in maximum temperature is greater than the threshold value. Neural networks (NN) are now widely used in meteorological applications, including Weather Forecasting.

Artificial Neural Network (ANN) came into existence in 1986 and considerable attention is drawn to research workers because it handles and solves complex and non-linearity problems better than traditional or conventional techniques. Though the weather is illogical, highly complex and non-linear behavior, ANN results are relatively accurate, as suggested by Abhishek *et al*. (2012) and Sallehuddin *et al*. (2007); Narvekar *et al.* (2015); Dombayc *et al*. (2009) used the ANN model to predict the daily mean temperature of Denizli, south-western Turkey. Three years of meteorological data from 2003 to 2005 were used for training and the 2006 data testing. ANN model based on 3 to 30 neurons in its hidden layers was designed; it has been observed that the Levenberg Marquardt algorithm, having six neurons in the hidden layer, produced the best results. Behrang *et al.* (2010) used several parameters, such as daily mean air temperature, relative humidity, sunshine hours, evaporation and wind speed, topredict global solar radiation by using Multi-layer perceptron (MLP) and radial basis function (RBF) neural networks. It has been analyzed that the ANN model provided more stable and prescribed results than other prediction models.

2. Study area Nawabshah / Shaheed Benazirabad

Shaheed Benazirabad, historically known as Nawabshah, is a district in the Sindh province and the 27^{th} largest city in Pakistan (Fig. 1). According to the Pakistani Census-2017, its population and area are 2,79,688 and 4,239 square km. It is located in a fertile agricultural area. The city is known for its sugarcane, mango and cotton production centers one of Pakistan's largest banana producers. Geographically, Shaheed Benazirabad is located in Asia's northern hemisphere at 26° 15' 00" N latitude and 68° 25' 25" E longitude (Leghari, 2011; Statistics, 2018; Govt., 2012). The climate of Shaheed Benazirabad is hot and desert-like. The city is regarded as one of Pakistan's hottest. Summer temperatures can reach 53 °C. Temperatures of over 45 °C are typical in late May and early June. Winters are mild, dry, and short, with mostly clear skies all year. Winters begin late, around mid-November and last until mid-February, with nighttime temperatures approaching 4 °C and temperatures below 0 °C occurring on average two or three times in January (Mahessar *et al*., 2020).

3. Method

Our study aims to predict the daily mean temperature of the city of Nawabshah based on relative humidity and dew point using an Artificial neural network and multi-regression models.

August-Roche-Magnus approximation enables us to calculate one of the three variables, dew point, humidity, and temperature if two are known. The temperature can be calculated using this approximation variation in vapor pressure. The vapor pressure is related to the dew point, too. A simple formula calculates the dew point if both temperatures and humidity are known. It shows that humidity, temperature and dew points are interrelated.

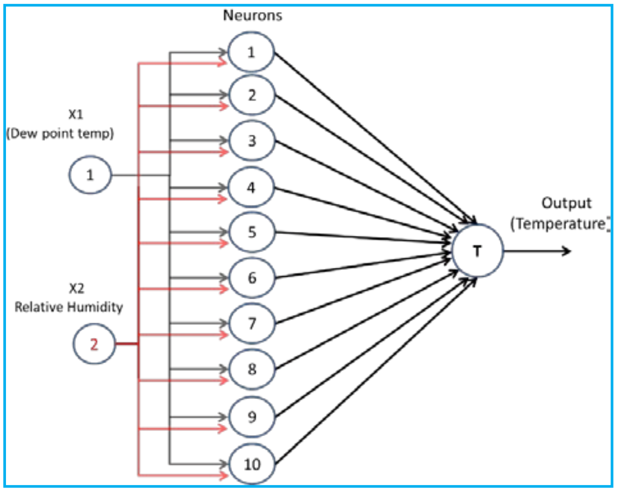


Fig. 2. Feed-forward neural network with 10 Neurons

That was the basis of using humidity and dewpoints as input parameters to predict the temperature.

3.1. *Artificial Neural Network*

In the Artificial Neural Network, to train the machine, the computer learns the behavior of the input data and performs numerous tasks. The output data is predicted once the machine has been properly trained (Agatonovic *et al*., 2000; Malone, 1955; Maqsood *et al*., 2004). This study uses ANN to predict the average daily temperature of Nawabshah city. Fig. 2 shows the architecture of ANN.

ANN model was developed on two inputs, dew point, and relative humidity, to estimate the temperature on MATLAB, a three-layered feed-forward network. The Input layer contains two neurons without any hidden layer. Ten neurons are taken in the hidden layers to ensure suitable fitting of the input to output variables. Hidden layer neurons have a sigmoidal transfer function, while neurons in the output layer have a linear transfer function. The Output layer represents the predicted temperature T, and it is given by :

$$T = \sum_{i=1}^{10} w_i H_i + B \tag{1}$$

where bias B = -0.198 and weights w_i are given in Table 1.

Here, H_i can be calculated by the following equation.

$$H_i = \frac{1}{1+e^{-E_i}} \tag{2}$$

TABLE 1

Weights of the input variable and neurons and bias associated with the neurons

I	w_i	W_{1i}	W_{2i}	b_i
1	3.0569	2.1023	0.0138	-3.1048
2	3.7068	-3.9415	0.1215	-3.8332
3	3.2423	-3.5097	0.0567	-1.8492
4	1.0574	-0.9549	0.3140	-0.1963
5	-1.4193	-3.8597	-0.0260	-0.0183
6	1.1818	2.4144	-0.0815	0.1264
7	0.6319	-0.2417	1.2435	0.4155
8	1.9449	3.6015	-0.0404	2.1441
9	-4.8280	1.1750	-0.0193	-3.5692
10	0.7819	2.9949	-0.2688	3.3276

E_i can be calculated using the following formula,

$$E_i = W_{1i} X_1 + W_{2i} X_2 + b_i \tag{3}$$

where X_1 and X_2 represent input dew point and humidity, respectively, b_i is the bias associated with the i^{th} neuron of the hidden layer, and W_{ij} is the branch's weight connecting the i^{th} neuron and j^{th} input.

3.2. *Multiple Regression Models*

Two multiple regression models (see equations 4 and 5) are generated to predict the daily mean temperature based on two independent variables: dew point and relative humidity.

The first regression model is linear, whereas the second is non-linear functions of these independent variables. In equations (4), a, b, c, and (5), a_0, a_1, a_2 and a_3 are regression coefficients; T is the average daily temperature, T_d is the average dew point, and Rh is the relative humidity.

Model 1: (linear model)

$$T = a + bT_d + cRh \tag{4}$$

Model 2: (Non-linear model)

$$T = a_0 + a_1 T_d + a_2 Rh + a_3 \left(\frac{T_d}{Rh}\right) \tag{5}$$

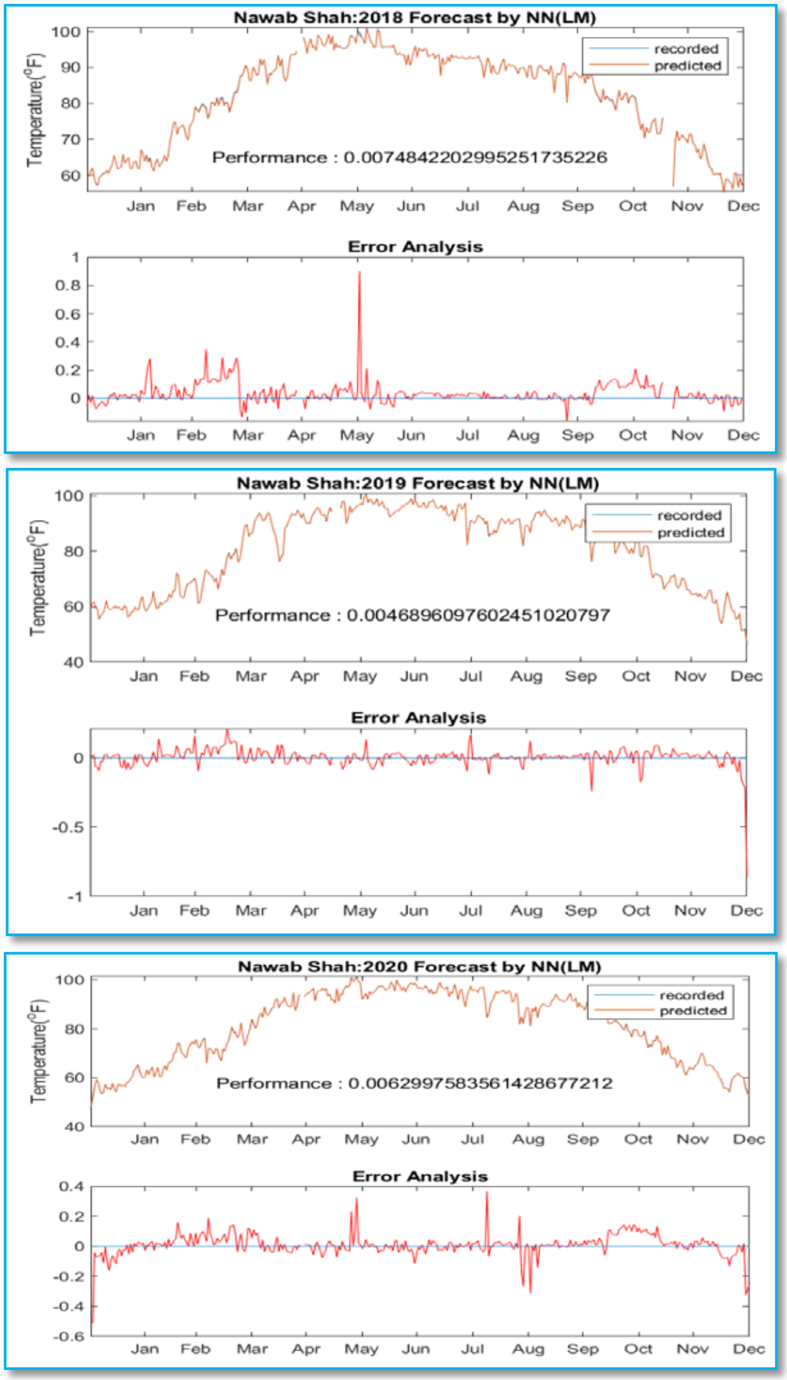


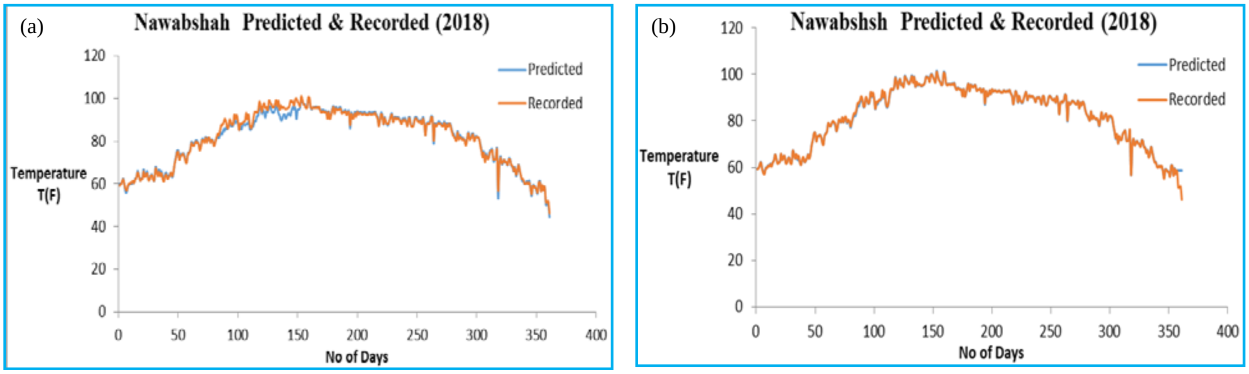
Fig. 3. Comparison of the predicted and actual average daily temperature of Nawabshah for the years 2018-2020

3.3. Statistical errors for validation of regression analysis

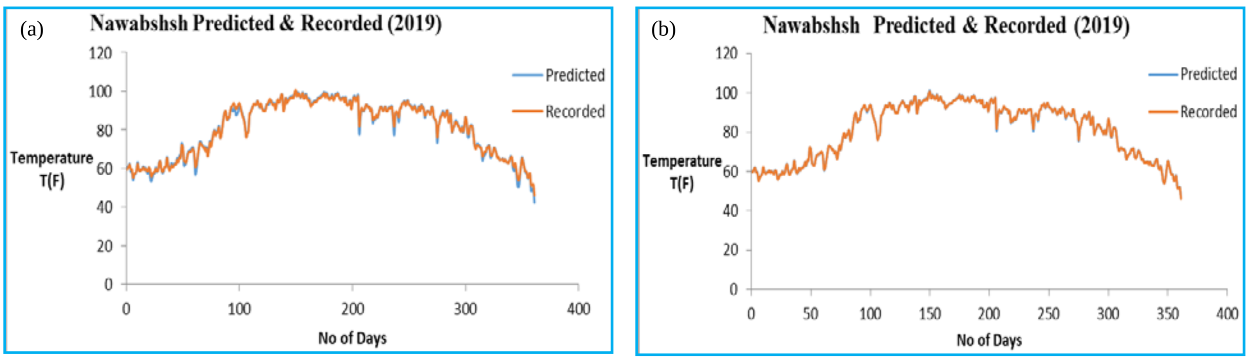
The Mean Square Error (MSE), Mean Absolute Error (MABE), Mean Absolute Percent Error (MAPE), and R-squared are calculated to check the validity of regression analysis results.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} \left(T_{c.i} - T_{m.i}\right)^2 \tag{6}$$

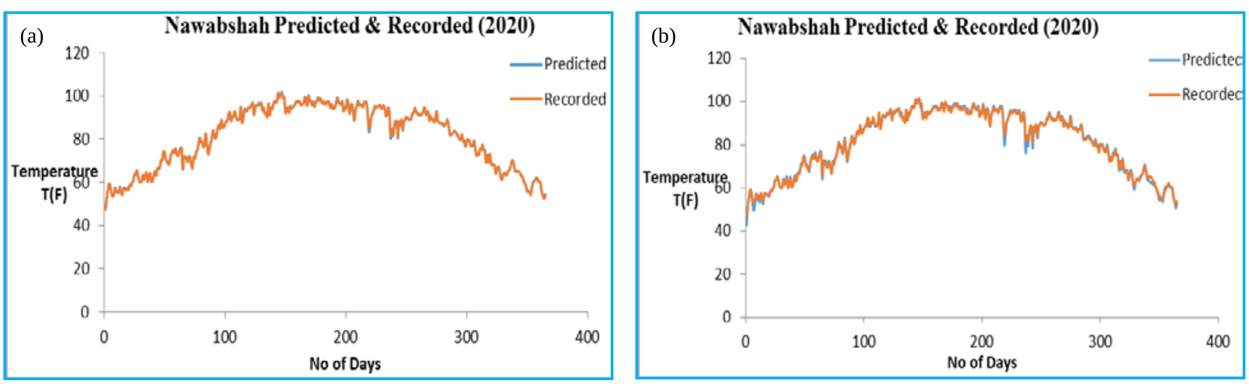
$$\text{MABE} = \frac{1}{n}\sum_{i=1}^{n} \left|T_{c.i} - T_{m.i}\right| \tag{7}$$



Figs. 4(a&b). The recorded and predicted temperature distribution (by (a) model 2 and (b) model 3) for 2018



Figs. 5(a&b). The recorded and predicted temperature distribution (by (a) model 2 and (b) model 3) for 2019



Figs. 6(a&b). The recorded and predicted temperature distribution (by (a) model 2 and (b) model 3) for 2020

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(T_{c.i} - T_{m.i})}{T_{m.i}}\right| \times 100 \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(T_{c.i} - T_{m.i})^2}{\sum_{i=1}^{n}(T_{c.i} - \overline{T_m})^2} \quad (9)$$

4. Result and discussion

In several meteorological parameters, one of the most important parameters is temperature, which widely affects different sectors, the industry of renewable energy, agriculture, and the daily life of a common person. The temperature also affects the water contents in the soil.

TABLE 2

Statistical Errors in the predicted temperature distribution of 2018-2020 for ANN (model 1), Linear Regression model 2 and non-linear Regression model 3

	Year	RMSE	MABE	MAPE	R^2
Model 1	2018	0.0074842	0.0503135	0.0626279	0.999954
ANN	2019	0.0046896	0.0378067	0.05238	0.999976
	2020	0.0062997	0.047664	0.065467	0.999969
Model 2	2018	2.3457925	0.98174177	1.14558	0.986962
Linear Regression	2019	1.0067889	0.6990821	0.918204	0.995284
	2020	0.915536	0.685373	0.90306	0.996256
Model 3	2018	0.0790401	0.194316	0.243275	0.999496
Non-Linear Regression	2019	0.046926	0.15758069	0.201875	0.999768
	2020	0.0499764	0.1551319	0.199333	0.999759

It influences several processes like seed germination, plant growth, etc. The variation in daily temperature is also considered in designing houses and hospitals. Scientists and researchers designed numerous temperature forecasting methods. One of them is Artificial Neural Network.

In this study, ANN is used to estimate the daily mean temperature. ANN consists of two input variables (dew point and relative humidity), ten neurons in a hidden layer, and one output variable (daily mean temperature). From 2010 to 2017, data points (average daily temperature, dew point and relative humidity) were used to train the network by selecting the Levenberg-Marquardt algorithm. 70% of data samples are used to train the prescribed network, whereas 15% are used for validation and 15% for testing. Once the network is trained, it is further used to forecast temperatures from 2018 to 2020. The Predicted and actual values, along with errors, are plotted in Fig. 3. The Weights W_1 and W_2 of the neuron that connects different nodes are determined and presented in Table 1. The table also gives the bias values in equation (3). The bias 'B' in equation (1) is -0.198. The weights w_i nodes connecting neurons in the hidden layer and out are also given in Table 1.

Two multiple regression models are suggested for the daily mean temperature distribution. Eleven years of weather data are used to establish and validate the two models. The data points are bifurcated into two parts. The data points from 2010-2017 are used to establish models, whereas data points from 2018-2020 are used to validate the models. The models are established by regression analysis. The first part of the dataset is also used to find the regression coefficients; these coefficients are substituted in equations 4 and 5 (see equations 10 and 11). For further validation, statistical errors are calculated and the performance of the models is compared. These errors include Mean Square Error (MSE), Mean Absolute Error (MABE), Mean Absolute Percent Error (MAPE), and R-R-square. The best model was identified by predicting temperature distribution for 2018-2020 and comparing the associated errors with the models.

$$T = 47.06728 + 1.0797067 T_d - 0.614788\ Rh \tag{10}$$

$$T = 28.55959 + 0.8352719 T_d - 0.282572\ Rh + 13.187799\left(\frac{T_d}{Rh}\right) \tag{11}$$

Multiple regression analysis determines the regression coefficients based on the years 2010 -2017; these coefficients are then employed to predict temperature distribution for 2018-2020 (Figs. 4-6) for linear and non-linear regression models.

Table 2 shows statistical errors in ANN and both regression models. The statistical errors RMSE, MABE, and MAPE are the lowest in the case of ANN; the coefficient of determination between predicted and actual daily mean temperatures is highest for ANN. The maximum absolute difference between measured and calculated temperature by model 2 was greater than 2.34°F in 2018 and 1.006°F in 2019. The corresponding difference for model 3 was 0.079°F for 2018 and 0.0499°F for 2020. The statistical errors corresponding to models are listed in Table 2. The table shows that all four statistical errors for the non-linear regression model 3 are less than those for linear regression model 2. Both temperature distribution and statistical error predictions indicate that the ANN model 1 is better than non-linear regression model 3 and predicts more reliable temperature values for Nawabshah city.

5. Conclusion

The daily mean temperature distribution modeling for Nawabshah, Pakistan, for 2001-2020 has been carried out. Two methods are employed: (*i*) Artificial Neural Network and (*ii*) Multiple regression models. Two models, linear and non-linear, were tested in multiple regression. The daily mean temperatures for 2010-217 were used in ANN for training and validation. The weights and bias of training and validation were then used to forecast the daily mean temperatures for the three years (2018-2020). The forecasted results were compared with the recorded daily mean temperature distribution for 2018-2020. An excellent agreement was found between forecasted and recorded distribution. Table 2 shows four different types of errors in forecasted data. The lower error values show the good quality of daily mean temperature distribution estimation for 2018-2020. The coefficient of determination values supports this quality as it is approximately near one.

The regression coefficients were calculated from 2010 to 2017 based on daily mean temperature, dew point, and relative humidity in multi-regression models. The daily mean temperature distribution for 2018 to 2020 is forecasted using regression coefficients with the help of two independent variables, dew point, and relative humidity. In linear regression and non-linear models, the estimated values are better represented by recorded daily mean temperature distribution. However, the errors are more than the corresponding ones in ANN; nevertheless, the recorded values are in the acceptance range. Figs. (3-6) shows the plots comparing recorded and estimated values; the two graphs almost overlap, so the coefficient of determination is near 0.99 or close to 1.

The statistical errors such as RMSE, MABE, and MAPE and the coefficient of determination R^2 are calculated for all models to check the validity. The errors are given in Table II; comparing the errors in three models shows that all three models give temperature values close to the recorded temperature. The ANN gives the best results; non-linear regression models take second place.

Acknowledgement

We are thankful to the Meteorological Department of Pakistan, Karachi, for providing the data for this study. The data is also available on the website https://en.tutiempo.net/climate/ws-417490.html.

Author statement (Disclaimer)

Adeel Tahir (adeel.tahir@fuuast.edu.pk)

Collected weather data and processed data in a format of the computer program, wrote a literature review, designed and analyzed the graphs and results.

Dr. Muhammad Ashraf (m.ashraf@fuuast.edu.pk)

Helped in designing the ANN model and also contributed to the literature review

Dr. ZaheerUddin (zuddin@uok.edu.pk)

Supervised and designed the Regression model and helped in analyzing the results.

Dr. Muhammad Sarim (msarim@fuuast.edu.pk)

Provided the idea of ANN and supervised the research work.

Dr. S. Naseem Shah (syednaseemshah@fuuast.edu.pk)

Contributed to writing and proofreading the manuscript.

Disclaimer : The contents and views expressed in this study are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

References

Abhishek, K., Kumar, A., Ranjan, R. and Kumar, S., 2012, “A Rainfall Prediction Model using Artificial Neural Network”, *IEEE Control and System Graduate Research Colloquium (ICSGRC 2012)*, 82-87.

Agatonovic-Kustrin, S.and Beresford, R., 2000, “Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research”, *Journal of Pharmaceutical and Biomedical Analysis*, **22**, 5, 717-727.

Amato, U., Cuomo, V., Fontana, F. and Serio, C., 1989, “Statistical predictability and parametric models of daily ambient temperature and solar irradiance: An analysis in the Italian climate”, *Journal of Applied Meteorology and Climatology*, **28**, 8,711-721.

Behrang, M., Assareh, E., Ghanbarzadeh, A. and Noghrehabadi, A. R., 2010, “The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data”, *Solar Energy*, **84**, 1468-1480.

Cuomo, V., Fontana, F.and C. Serio., 1986, “Behaviour of ambient temperature on daily basis in Italian climate”, *Rev. Phys. Appl.*, 211-218.

Dombayc, O. and Golcu., M., 2009, "Daily means ambient temperature prediction using artificial neural network method : A case study of Turkey", *Renewable Energy*, **34,** 1158-1161.

Govt, S., 2012, April 07, *Goverment of Sindh*. Retrieved April 07, 2012, from Population Welfare Development, https://web.archive.org/web/20120407025659/http://www.sindh.gov.pk/dpt/PopulationDev/nawabshah.htm.

Islam, S. U., Rehman, N. and Sheikh, M. M., 2009, "Future change in the frequency of warm and cold spells over Pakistan simulated by the PRECIS regional climate model", *Climatic Change*, **94**, 1, 35-45.

Kaymaz, B., 2005, "Hazards and their impact on human", 29.IMISE (International Movement for Interdisciplinary Study of Estrangement) Conference. Paris: The American University of Paris.

Leghari, M., 2011, Jan 07, ref-7. Retrieved Jan 07, 2011, from wikipedia : https://en.wikipedia.org/wiki/Nawabshah#cite_ref-7.

Mahessar, A., Qureshi, A. L., Sadiqu, B., Kori, S., Mukwana, K. and Leghari, K., 2020, "Rainfall Analysis for Hyderabad and Nawabshah, Sindh, Pakistan", *Engineering, Technology & Applied Science Research*, **10**, 6, 6597-6602.

Malone, T. F., 1955, "Application of statistical methods in weather prediction", *Proceedings of the National Academy of Sciences of the United States of America*, **41**, 11, 806.

Maqsood, I., R. Khan, M. and Abraham, A., 2004, "An ensemble of neural networks for weather forecasting", *Neural Computing & Applications*, **13**, 2, 112-122.

Nadia, R. and Islam, S., 2006, Future change in the frequency of warm and cold spells durations over Pakistan simulated by the PRECIS regional climate model, Young Scientist Conference, Beijing, China

Narvekar, M. and Fargose, P., 2015, "Daily weather forecasting using artificial neural network", *International Journal of Computer Applications*, **(121) 22**, 0975-8887.

Sallehuddin, R., Shamsuddin, S. M., Hashim, S. Z. Mohd and Abraham, A., 2007, "Forecasting time series data using hybrid grey relational artificial neural network and auto regressive integrated moving", *Neural Network World*, **17**, 6, 573.

Statistics, P. B., 2018, Jan 20, "PAKISTAN: Provinces and Major Cities", Retrieved Jan 20, 2018, from PAKISTAN: Provinces and Major Cities: https://www.citypopulation.de/en/pakistan/cities/.

Ustaoglu, B., Cigizoglu, H. K., andKaraca, M., 2008, "Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods", *Meteorological Applications*, Meteorological Applications : A Journal of forecasting, practical applications, training techniques and modelling, **15**, 4, 431-445.