

A comparison of different weather forecasting models for the monthly forecast of Lahore city

SAJID ALI, ZAHEER ABBAS* and MUHAMMAD MOEEN BUTT**

Department of Statistics, Quaid-i-Azam University, Islamabad, 45320, Pakistan

**Department of Statistics, Government Post Graduate College Sahiwal, Sahiwal, Pakistan*

***Department of Economics and Statistics, Dr. Hasan Murad School of Management,
University of Management and Technology, Lahore, Pakistan*

(Received 7 May 2019, Accepted 5 October 2021)

e mail : sajidali.qau@hotmail.com

सार – इस पत्र में, हम मौसम की भविष्यवाणी में प्रयुक्त विभिन्न सांख्यिकीय मॉडलों के प्रदर्शन का अध्ययन करते हैं और उनकी पूर्वानुमान सटीकता की तुलना करते हैं। विशेष रूप से, हम टाइम सीरीज रिग्रेशन (TSR), मौसमी ऑटोरेग्रेसिव फ्रैक्शनल इंटीग्रेटेड मूविंग एवरेज (SARFIMA) और आर्टिफिशियल न्यूरल नेटवर्क (ANN) का उपयोग करते हैं। पूर्वानुमान सटीकता का अनुमान लगाने के लिए एक गतिशील गैर-रेखीय ऑटोरेग्रेसिव (एनएआर) बैक-प्रोपेगेशन एनएन एल्गोरिदम भी लागू किया जाता है। एनएन मॉडल के लिए, हम डेटा को प्री-प्रोसेस करने के लिए मूविंग एवरेज (एमए) और होल्ट-विंटर एक्सपोनेंशियल स्मूथिंग (एचडब्ल्यू-ईएस) ट्रांसफॉर्मेशन का उपयोग करते हैं। उपरोक्त मॉडलों को लागू करने के लिए विभिन्न मौसम मानकों का मासिक डेटा पाकिस्तान मौसम विज्ञान विभाग से प्राप्त किया जाता है। परिणाम बताते हैं कि डेटा के एमए परिवर्तन के साथ एनएन मॉडल में सबसे छोटी रूट माध्य चुकता त्रुटि है और विभिन्न मौसम मापदंडों के लिए उच्चतम सहसंबंध गुणांक है।

ABSTRACT. In this paper, we study the performance of different statistical models used in weather forecasting and compare their forecast accuracy. In particular, we use time series regression (TSR), seasonal autoregressive fractional integrated moving average (SARFIMA) and artificial neural network (ANN). A dynamic non-linear autoregressive (NAR) back-propagation ANN algorithm is also applied to estimate the forecasting accuracy. For ANN model, we use the moving average (MA) and Holt-Winter exponential smoothing (HW-ES) transformations for pre-processing the data. The monthly data of different weather parameters are obtained from the Pakistan Meteorological department to apply the aforementioned models. The results show that the ANN model with the MA transformation of the data has the smallest root mean squared error and the highest correlation coefficient for different weather parameters.

Key words – Artificial Neural Network, Seasonal autoregressive fractional integrated moving average, Time series regression, Root mean square error, Weather Forecasting.

1. Introduction

Weather forecasting is a critical issue in the meteorological literature and there are different models to forecast weather. However, several factors, for example, wind speed, humidity, evaporations, etc., significantly affect the accuracy of forecasting. Similarly, the time series models used in weather forecasting are not truly dynamic. The knowledge about the dynamics of weather parameters in a particular region is very helpful in planning, especially for the underdeveloped countries like Pakistan. A reliable prediction of Pakistan monsoon on seasonal and inter-seasonal time series is not just helpful in government planning but can also save people from losses. The use of statistical techniques for policy making depends on the understanding of the past behaviour of the weather data. However, the transient behaviour of weather

parameters over a particular period of time makes difficult to predict weather accurately and consistently. The Pakistan economy in general and agriculture and industrial sectors in particular significantly depends upon weather conditions. The frequent fluctuation in weather in Pakistan causes heavy losses to our economy. Therefore, some sophisticated statistical techniques must be utilized to forecast weather and policy making. Moreover, a comparison of different statistical models can give us an insight to choose the best model for weather forecasting.

In the literature, Goulden (1962) studied the relationship between the monthly average of weather parameters and crop yield by using multiple regression. Ramchandran (1967) conducted an analysis of the rainfall of 167 observatory stations distributed over India and the neighboring country. The author used the regression for

modeling the monthly and annually rainfall as a linear function of longitude, latitude and elevation above the sea level. Bali (1970) calculated the average yield and explained the inadequacy of currently employed methods for forecasting crop yield in India. Huda *et al.* (1975, 1976) used a second degree multiple regression for checking the relation between weather variables and rice and maize yields.

Baillie (1996) gave an overall analysis and review of the main econometric models on long memory methods, like fractional integration and their applications in finance and economics. For the quarterly UK inflation data, Franses *et al.* (1997) developed an extension known as the fractionally autoregressive integrated moving average (ARIMA) (0, d, 0), where d is a fractional integration parameter that is supposed to vary with the season "s". The proposed periodic model not only provides useful information for in-sample description, but also for out-of-sample forecasting.

Neural network is also a very popular technique in weather forecasting and many researchers used it. For example, to assess the accuracy of different weather models, Kihoro *et al.* (2004) compared artificial neural network (ANN) and ARIMA models used in the forecasting of monthly of time-series data. Abhishek *et al.* (2012) used ANN for reliable forecasting from non-linear weather forecasting models. Shrivastava *et al.* (2012) showed that ANN such as radial basis function network (RBFN) and back propagation network (BPN) performed better for forecasting monsoon-rainfall. Nayak *et al.* (2013) used ANN models for the estimation of rainfall. Khedhiri (2015) also compared the performance of the ANN and seasonal autoregressive fractional integrated moving average (SARFIMA) models and showed that the ANN model performed consistently as compared to the SARFIMA for forecasting. Valipour (2015) used the SARIMA and ARIMA models to study long-term runoff forecasting in the United States. Aftab *et al.* (2018) used data mining techniques, like a support vector machine and naïve Bayes classifier, to forecast the Lahore city weather. Shamshad *et al.* (2019) used ARIMA model for forecasting weather parameters of Lahore.

The objective of this paper is to compare different models to determine the best model for monthly weather forecasting. Monthly weather data of Lahore city are collected and statistical analysis is performed to assess the forecasting accuracy of time series multiple linear regression (TSMLR), SARFIMA models and compared to three ANN models. The methodology of the paper is to apply the ANN on weather data by using the moving average as a smoothing technique and compare the results obtained without smoothing. The root mean squared error

(RMSE) and correlation coefficient (r) are two statistical measures, which we use to compare and show the most accurate forecasting model of each considered weather parameters. The rest of the study is organized as follows : In Section 2, we discuss data and different statistical models which are used to assess the best model for weather forecasting. In Section 3, we estimate and compare different statistical models for weather forecasting. Finally, we conclude the article in Section 4.

2. Data and statistical models

The study is based on a time-series weather data collected at the Lahore station by the Pakistan Meteorological Department. The monthly time series data cover the period January 1951 to December 2015 for six parameters, *viz.*, rainfall recorded in millimeters (mm), surface maximum and minimum temperature (recorded in centigrade °C), surface relative humidity (8 am and 5 pm) (measured in percentage) and wind speed (miles per hours, mph). All these variable were recorded at 2 m height above the surface. Since the surface data are the observed data, these data sets are useful to evaluate the forecasting properties of different models. Table A1 lists the descriptive statistics of these variables. The results for the rainfall and maximum temperature are discussed here and for the sake of space, other results are provided in the supplementary text (Figs. S1-S12 and Tables S1-S12).

"Lahore has a semi-arid climate (Köppen climate classification *BSh*) and the hottest month is June, when average highs routinely exceed 40 °C (104.0 °F). The monsoon season starts in late June and the wettest month is July, with heavy rainfalls and evening thunderstorms with the possibility of cloud bursts. The coolest month is January with dense fog. The city's record high temperature was 48.3 °C (118.9 °F), recorded on 30 May 1944 while 48 °C (118 °F) was recorded on June 10, 2007. It is recorded lowest -1 °C (30 °F) on 13 January 1967. The highest rainfall in a 24-hour period is 221 millimeters (8.7 in), recorded on 13 August 2008. On 26 February 2011, Lahore received heavy rain and hail measuring 4.5 mm (0.18 in), which carpeted roads and sidewalks with measurable hail for the first time in the city's recorded history" (see https://en.wikipedia.org/wiki/Climate_of_Lahore). Therefore, it is necessary to quantify the weather parameters which causes such extreme weather conditions and help decision makers in planning to prevent hazard events.

2.1. Time series multiple regression model

A regression model is a mathematical model that determines the relationship between dependent and independent variables to predict the response value

(Anderson *et al.*, 2006; Zaw and Naing, 2008; Ghani and Ahmad, 2010; Shamshad *et al.*, 2019). In this study, we define the following time series regression model :

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \varepsilon_t \tag{1}$$

where β_0 = intercept, $\beta_i = i^{\text{th}}$ regression coefficient for $i = 1, 2, 3, 4, 5$, Y_t = Dependent variable, X_{it} = Independent variables, ε_t = error term

2.2. The SARFIMA model

The SARFIMA (p, d, q) (P, D, Q)_s model is defined as :

$$\begin{aligned} & (1 - \phi_1 B - \dots - \phi_p B^p) (1 - \beta_1 B^s - \dots - \beta_p B^{ps}) \\ & (1 - B)^d (1 - B^s)^D Y_t = \delta_0 + (1 - \theta_1 B - \dots - \theta_q B^q) \\ & (1 - \delta_1 B^s - \dots - \delta_l B^{Qs}) \varepsilon_t \end{aligned} \tag{2}$$

where, B = backshift operator, s = seasonal parameters and D = seasonal differencing. The first set of p -parameters shows that polynomial of AR with backshift operator (B). The second set of q -parameters defines the moving average polynomial function of IID error terms ε_t . The third set of d -parameter records the required demand of differencing of the time-series to reduce it to a stationary model. The ARFIMA models, generally called long memory processes, demonstrate a long-run dependence in their observations. It is worth mentioning that the parameter estimation of the ARFIMA model can be done by the method of maximum likelihood. An estimated value of D less than 0.5 indicates the stationarity of the considered series (Sowell, 1992; Olatayo and Adedotun, 2014; Li and Ye, 2015).

2.3. Artificial Neural Network (ANN) model

Neural networks have been used in a variety of fields, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games, medical diagnosis and in many other domains. The idea of neural network is derived from the neurons of the human brain, which are core processing elements for processing information as we perceive the real world. Following chart [Fig. 1] represents the model for a single neuron (Kumarasiri and Sonnadra, 2008).

Fig. 1 depicts a neuron along with synapses results in single/multi-layered neurons. A multi-layered ANN possibly includes three types of layers of neurons namely,

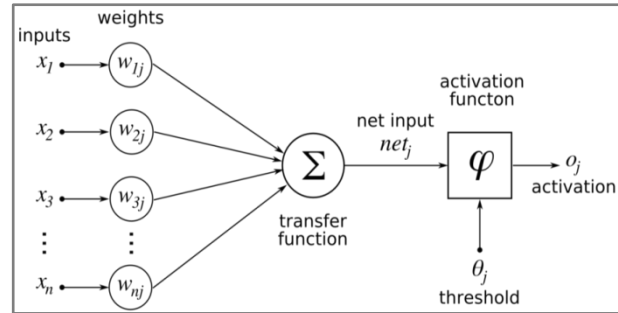


Fig. 1. Mathematical description of a single neuron

an input layer, an output layer and a set of hidden layers. Various types of ANN can be formed by considering the neuron type and its positioning. However, to determine which ANN model should be considered needs some important considerations, including how many layers and nodes in the network would suffice and which training algorithm would serve the purpose. For example, instead of learning, too much nodes in a hidden layer can help the network to memorize and adjust the fluctuations. For recent literature related to ANN in forecasting, we refer to Chattopadhyay (2007), Dibike and Solomatine (2001), El-Shafie *et al.*, (2011), Hayati and Mohebi (2007), Hung *et al.*, (2009), Kannan *et al.*, (2010), Kumar *et al.*, (2004, 2010), Kumarasiri and Sonnadra (2008), Litta *et al.*, (2013), Mesgari *et al.*, (2015), Nirmala (2015), Rahman and Matin (2015), Sahai *et al.*, (2000), Sohn *et al.*, (2005), Wu *et al.*, (2010) and references cited therein.

2.3.1. Training and testing of ANN

After the identification of the most appropriate network for rainfall forecasting, one can select inputs and the corresponding targets to train the network to provide a reasonable output. In practice, the network should be trained until the change in weights in a training cycle converges to a minimum value, because the performance of the network will be improved through the learning process. Once the network is sufficiently trained, the next step is to test its ability to produce accurate forecasts. In this study, back-propagation algorithm is used. In particular, 70% of the data (Jan 1951 - Dec 1995) is used for training and the remaining 30% (Jan 1996 - Dec 2015) for validation and testing the network. The weights of the network are calibrated using a training set, while the progress of the training process is monitored based on the cross-validation set. A MATLAB (2010) neural network toolbox is used in this article to implement ANN.

2.4. Performance assessment

To assess the performance of the aforementioned models, we use the root mean squared error (RMSE) and

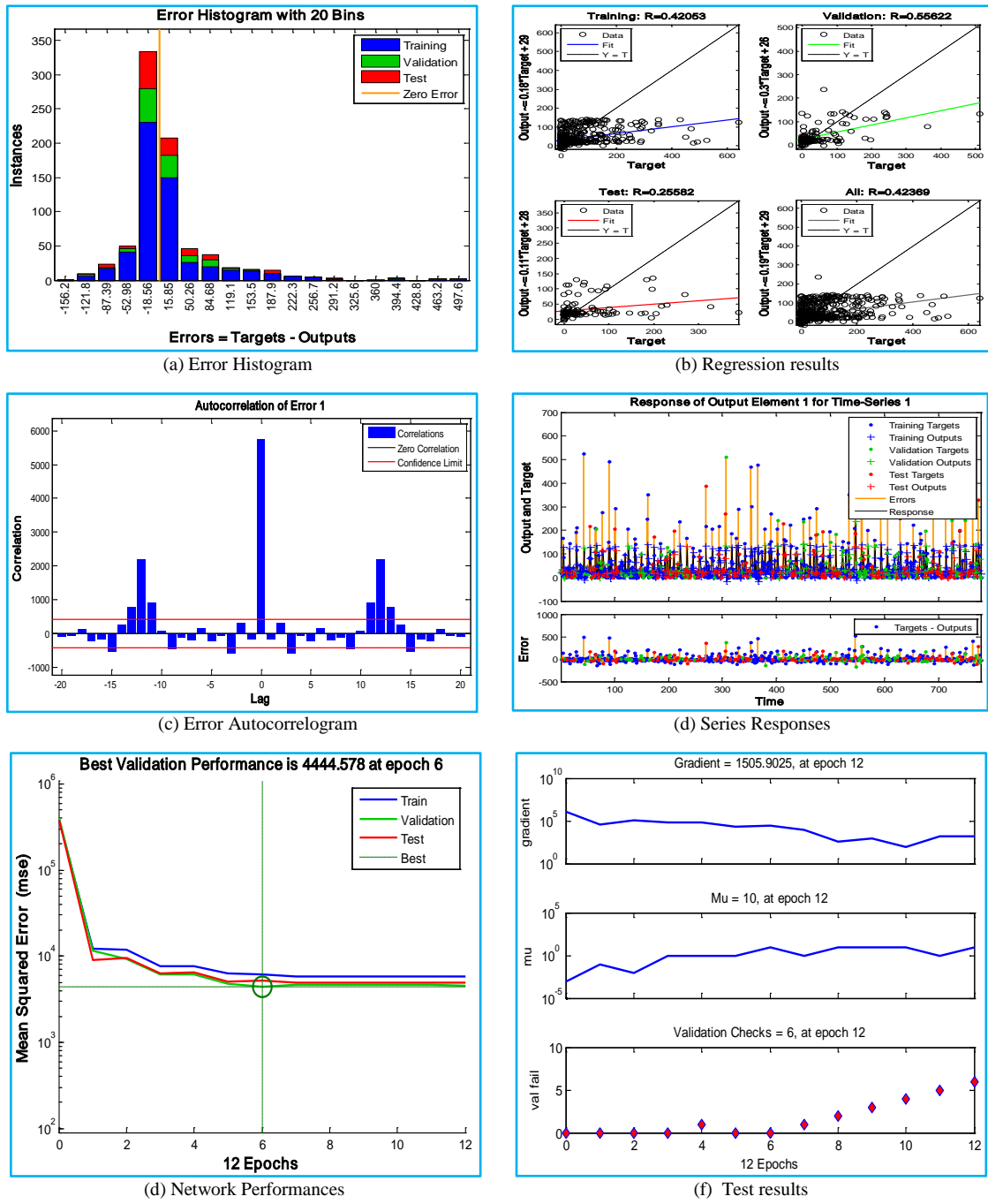
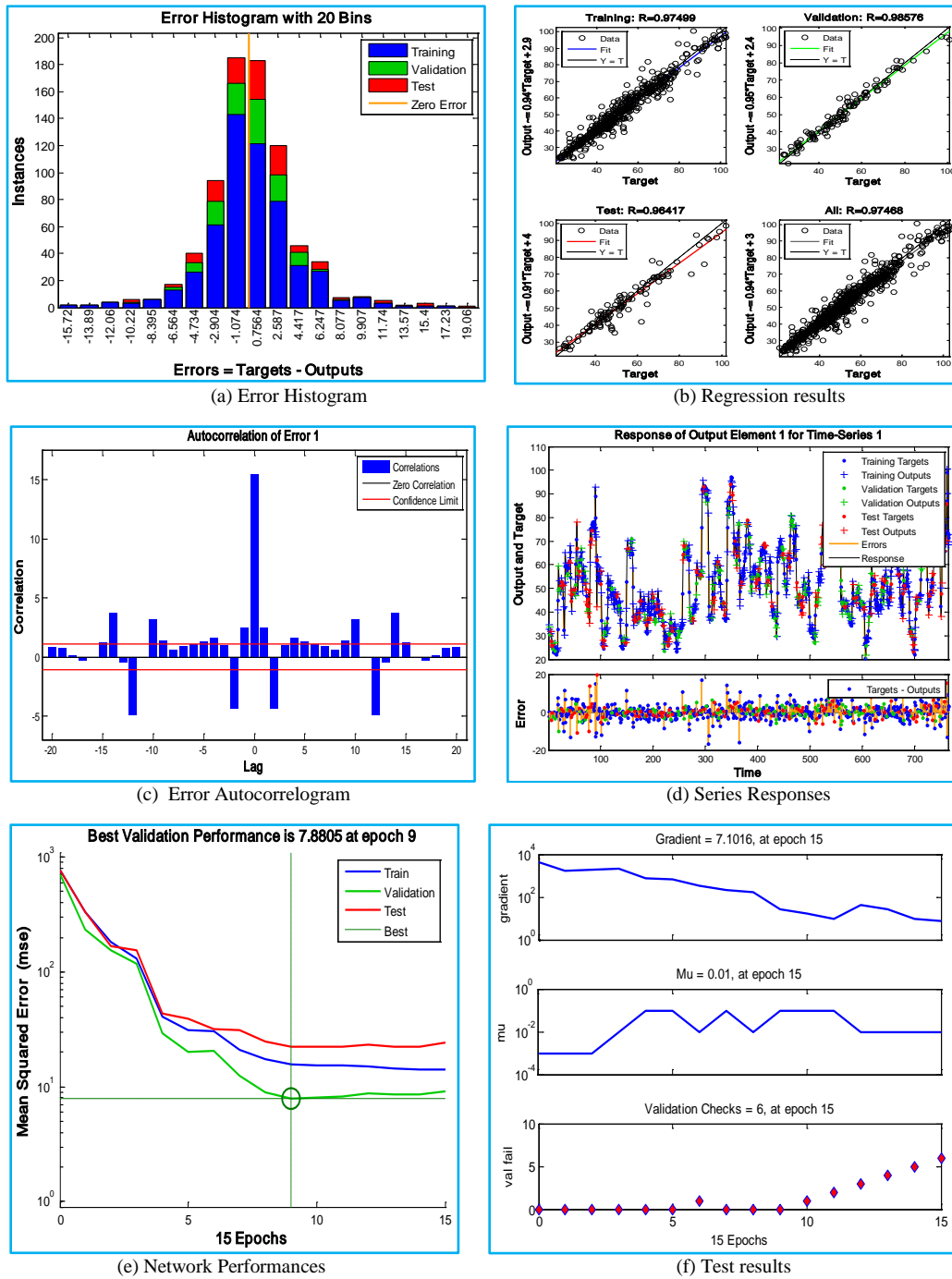


Fig. 2(a-f). Neural network results for Y (rainfall-original data)

Next, we implement the neural network on the rainfall data.

In Figs. 2-4, the results of the ANN model for three rainfall series, *i.e.*, rain fall original data, smoothed series by the moving average and smoothed series by Holt and winter exponential smoothing are depicted. It is observed that the performance of the best validation occurred at

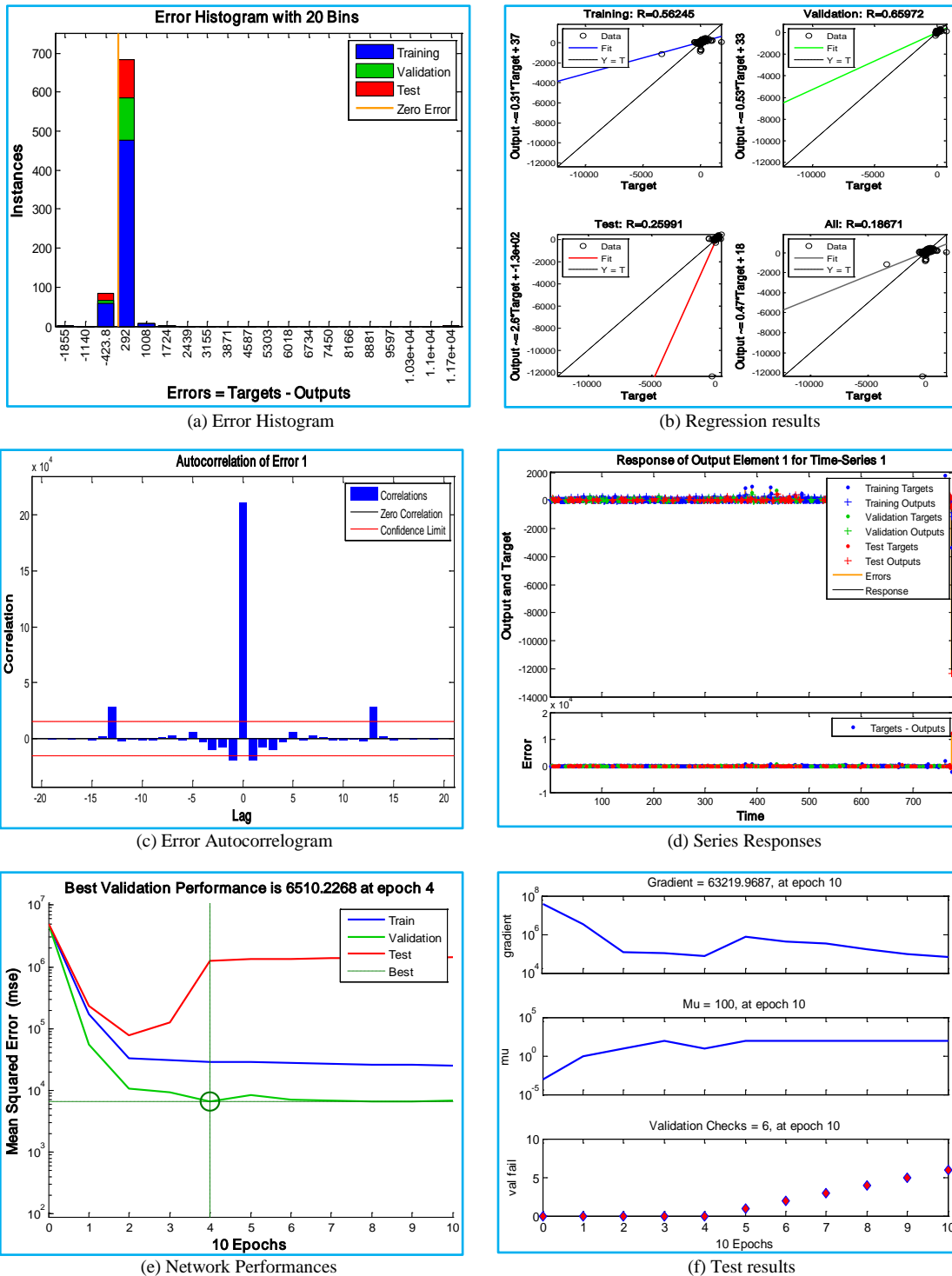
epoch 6th for rainfall original data while at 9th epoch for the moving average and 4th epoch for the Holt-Winter exponential pre-processed data. We also observed high correlation coefficients between estimated and target observations for the transformed rainfall data sets. The back propagation algorithm showed a very high-accuracy level with target values using moving average smoothed data. To be specific, a histogram of the errors is depicted



Figs. 3(a-f). Neural network for Y (rainfall data by moving average smoothing)

in Fig. 2(a) and it is clear that no outliers observed in the histogram. Fig. 2(b) depicts the regression line fit using the training and testing data while Fig. 2(c) depicts the autocorrelation, which indicates there is a seasonal component because 12th lag falls above the confidence limit. Fig. 2(d) presents the time series plot of the training and testing errors and it is clear that there is a close

agreement between them. Fig. 2(e) depicts the epoch at which the MSE is the minimum and for the rainfall series, it occurred at the 6th epoch. Fig. 2(f) depicts the gradient for 12 epochs and estimation of parameter at each epoch. Similarly, the results of other figures can be interpreted. Contrary to Fig. 2(b), Fig. 3(b) depicts the fitted regression line after smoothing the data with exponential



Figs. 4(a-f). Neural network results for YHW (HWE smoothed data)

smoothing and it is observed that the line is well fitted as compared to the Fig. 2(b). Also, the histogram [Fig. 3(a)] of errors is symmetric. However, the minimum MSE is observed at 9th epoch with exponential smoothing

[Fig. 2(e)] while at 4th epoch for the HW smoothing [Fig. 3(e)]. Comparing Figs. 2(d), 3(d) and 4(d), it is noticed that the variation in the response variable using HW smoothing is the minimum.

TABLE 3
Forecast performance of the rainfall models

Time series	RMSE	Correlation coefficient
Y	8.1650	0.5562
YMA	1.6754	0.9858
YHWE	80.6859	0.6597
Y (SARFIMA)	82.6351	0.7017
Time Series Regression	65.4982	0.6033

TABLE 4
Estimation of fractional differencing

Variable	Coefficient	Std. Error	t-Statistic	Prob.	95% [Confidence Interval]	
Y_cons	30.3539	363.8986	0.0834	0.9335	-683.9877	744.6949
ARFIMA d	0.4985	0.0046	108.9351	0.0000	0.4895	0.5075
SIGMA2	28.5990	1.8845	15.1755	0.0000	24.8996	32.2985

Log likelihood = -4532.404

F-statistic = 13.15942

Prob (F-statistic) = 0.000002

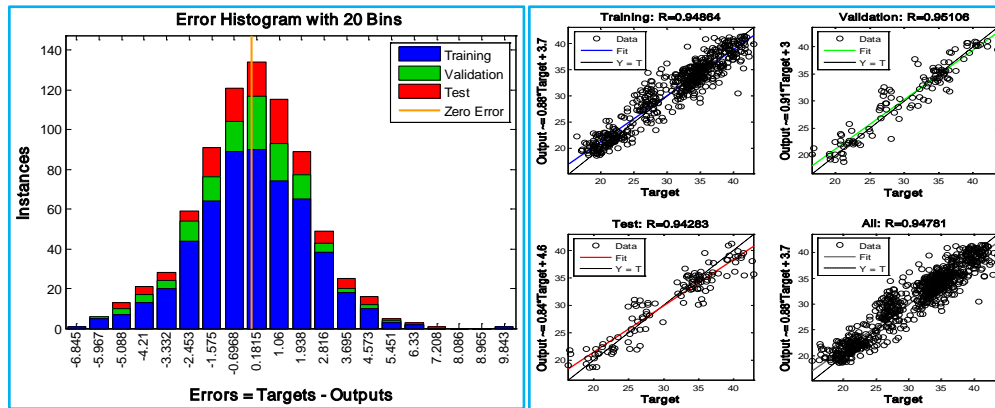
TABLE 5
Estimation of the SARFIMA model

Type	Coefficient	SE Coefficient	T	P
AR(1)	0.9421	0.0498	18.9166	0.0000
MA(1)	-0.6925	0.0634	-10.9106	0.0000
MA(2)	-0.1842	0.0413	-4.4557	0.0000
SAR(12)	0.9998	0.00005	18331.87	0.0000
SMA(12)	-0.9395	0.0131	-71.2818	0.0000
Constant	30.8078	4.4884	6.8638	0.0000

Next, the time-series multiple linear regression (MLR) model was fitted to predict the monthly rainfall as the response variable, while the other monthly variables such as maximum temperature, minimum temperature, relative humidity (RH) at 8 am and wind speed, are considered as the predictors. It is worth mentioning that all these variables were recorded at 2 m height above the surface. The stepwise regression analysis was used for the predictor selection. It is worth mentioning that we have 780 observations in-total of all the aforementioned variables. The estimated time series regression model is given below:

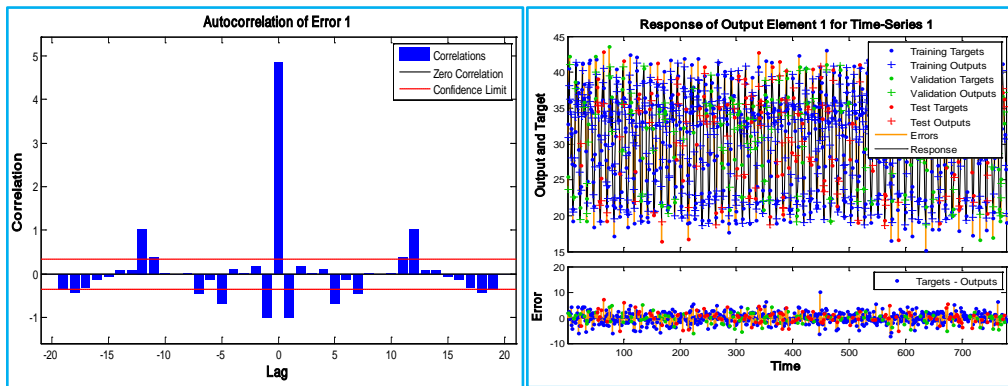
$$Y_t = -197.5494 + 5.0913X_{2t} + 0.9617X_{3t} + 1.8331X_{4t} + 8.5284X_{5t} \quad (3)$$

where minimum temperature (X_2), relative humidity at 8 am (X_3), relative humidity at 5 pm (X_4) and wind speed (X_5). Note that the coefficient of determination was 36.26% for the above model and the most contributing variables for predicting the rainfall (Y) was observed the wind speed. The variable maximum temperature (X_1) seemed to have less effect on the rainfall, because p-value was greater than 5% and hence did not include in the above model. The wind speed plays a key role on the surface temperature in situations where there is a strong temperature change with height in the boundary layer. The earth is heated and cooled from the ground and it is the wind that mixes this air at ground level with air higher aloft. During the day when wind is light and the sky is clear, heat will build on the surface and temperature in this case will be warmer than if the wind speed is stronger.



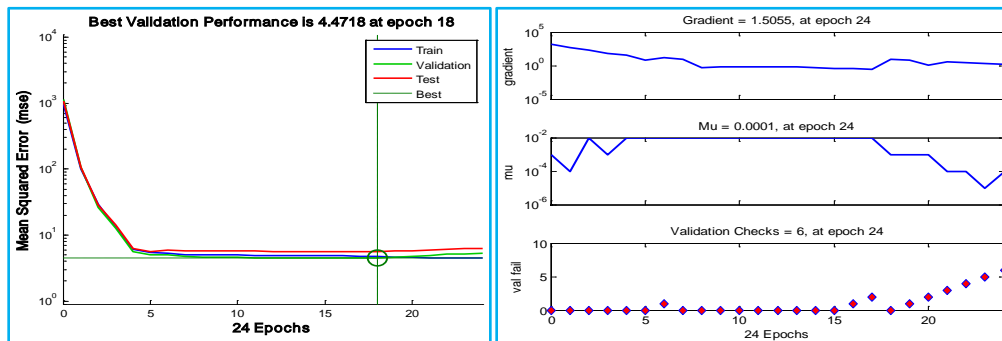
(a) Error Histogram

(b) Regression results



(c) Error Autocorrelogram

(d) Series Responses



(e) Network Performances

(f) Test results

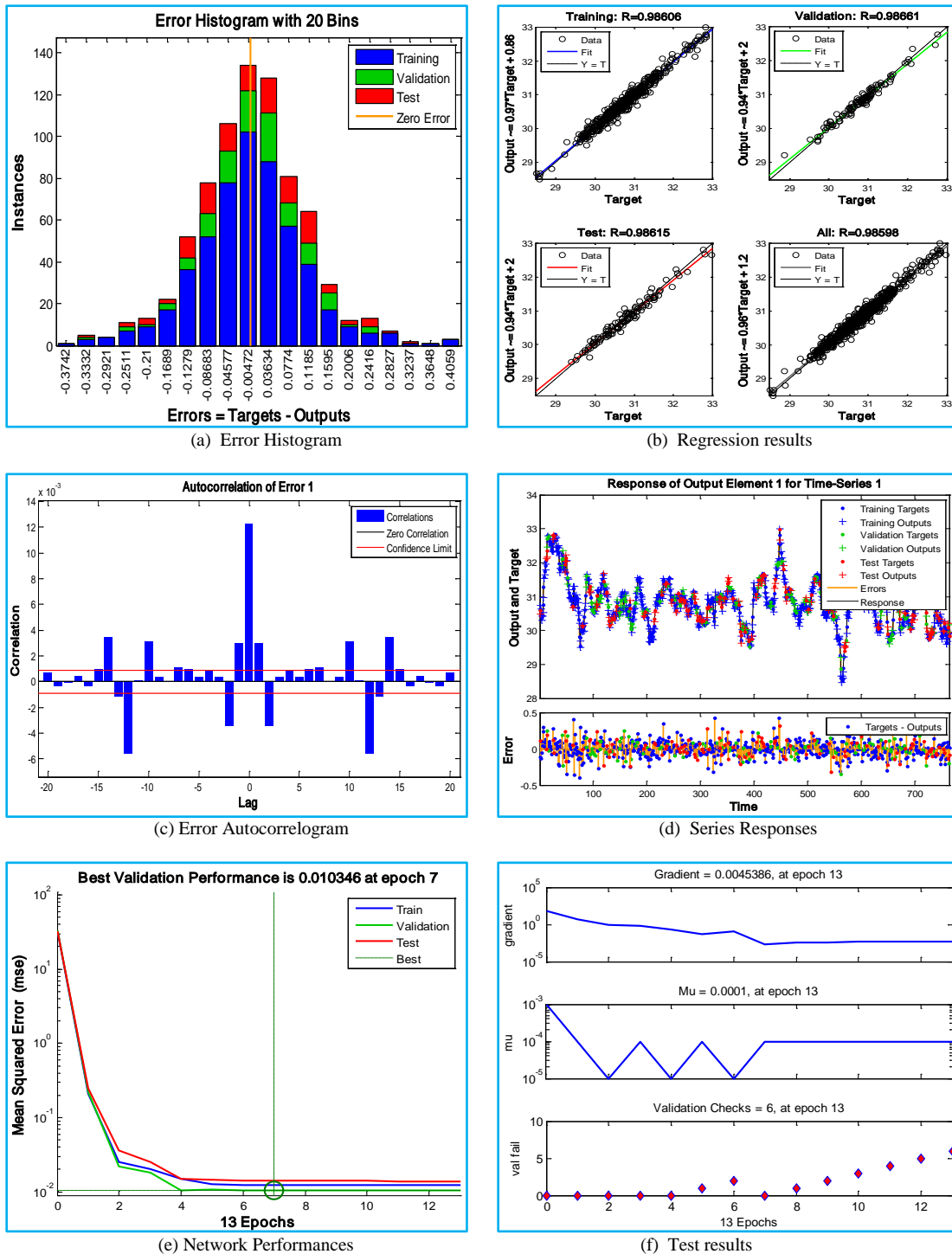
Figs. 5(a-f). Neural network results for Y (maximum temperature original data)

This is because stronger wind will mix the warm air near the surface with cooler air aloft.

Next, to decide which model is more appropriate for rainfall forecasting, we tabulated the RMSEs and correlation coefficients different models in Table 3.

In Table 3, a comparison of different measures for rainfall (mm) parameter assuming different models is given and we observed that the moving average

smoothing has the lowest RMSE (RMSE = 1.6754 with $r = 0.9858$). Similarly, the moving average smoothed data has the highest correlation coefficient using the ANN algorithm. However, the performance of YHWE and SARFIMA model is observed very poor (RMSE = 82.6351) as compared to the YMA ANN model (RMSE = 1.6754). Thus, we conclude that the ANN method is the most efficient for producing accurate predictions as compared to the ARIMA and time series multiple regression models.



Figs. 6(a-f). Neural network results for YMA (moving average transformation data)

3.2. Maximum temperature (°C)

Similar to the previous section, in this section we first estimate the fractional differencing parameter for the maximum temperature. Table 4 lists the estimated results

of a SARFIMA model for the maximum temperature (°C) weather parameter. From the table, it is observed that the data were generated by a long-period time series with fractional parameter $d = 0.4985$. Moreover, the model was observed stationary as $d < 0.5$.

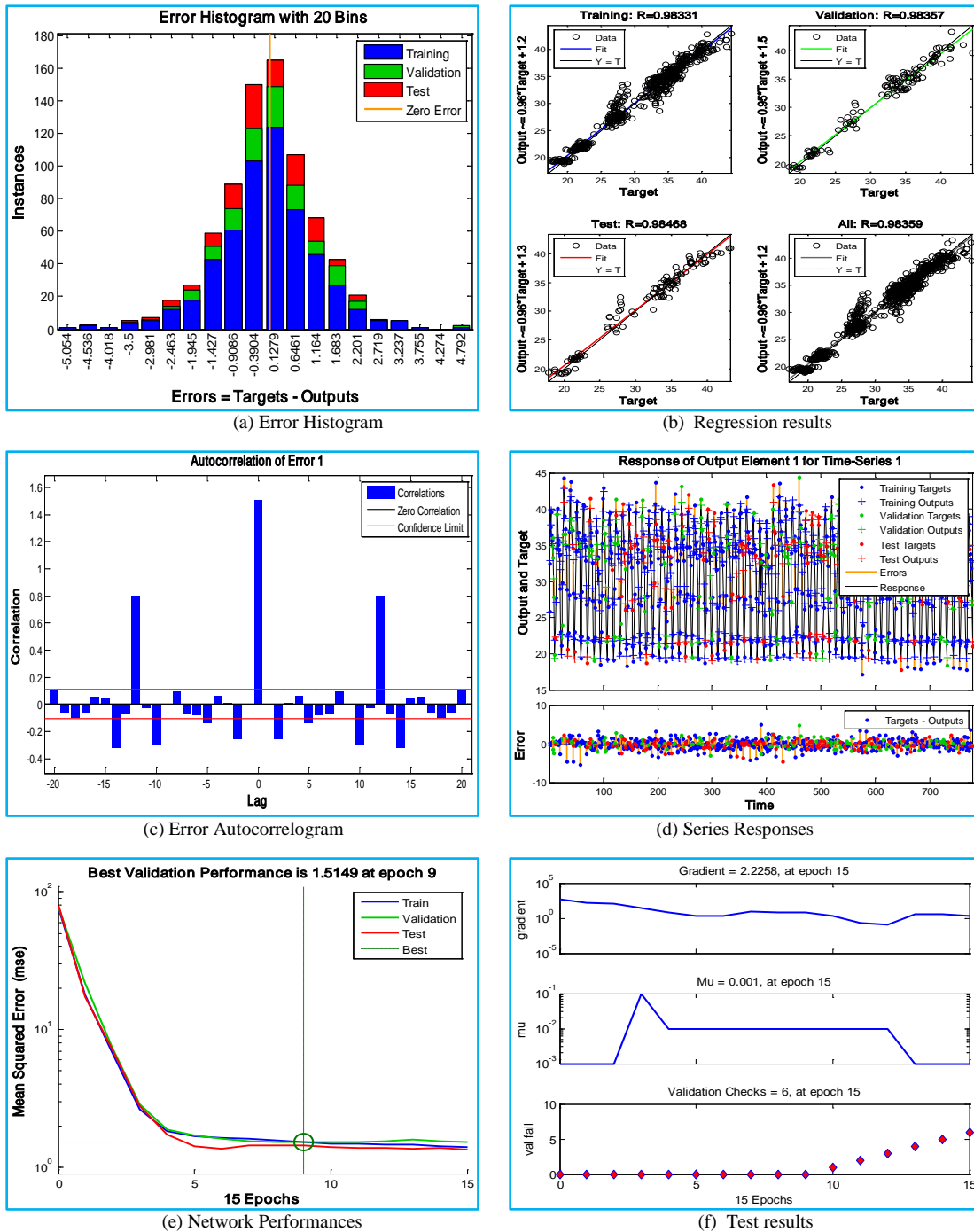


Fig. 7(a-f). Neural network results for Y (smoothed data by Holt-Winter transformation)

In Table 5, we have tabulated the parameter estimates of the SARFIMA model. Moreover, the AIC and BIC are observed 11.6292 and 11.6472, respectively. It is noticed that the maximum temperature ($^{\circ}\text{C}$) data can be modelled by SARFIAM (1, 0.4985, 2) (1, 0, 1) with $s = 12$ (Table 4).

Figs. 5 to 7 depict the results of ANN modeling for three smoothed and unsmoothed series for the maximum temperature. It is observed that the performance of the best validation occurred at 18th epoch for the maximum temperature original data while at 7th epoch for the moving average and 9th epoch for the Holt-Winter

TABLE 6

Forecasting performance of the maximum temperature models

Time series	RMSE	Correlation coefficient
Y	2.1146	0.9511
YMA	0.1017	0.9866
YHWE	1.2308	0.9835
Y (SARFIMA)	7.8569	0.9746
Time Series Regression	1.5150	0.9756

exponential smoothing. A high correlation coefficient is calculated between the estimated and the target observations for the transformed maximum temperature data set. The back propagation algorithm has shown a very high-accuracy level with target values for the smoothed data by moving average.

To be specific, a histogram of the errors is depicted in Fig 5(a) and it is clear that no outliers observed in the histogram. Also, the histogram is more symmetric than the histogram depicted in Fig. 2(a). Fig. 5(b) depicts the regression line fitted on the training and testing data while Fig. 5(c) depicts the autocorrelation, which indicates there is a seasonal component because 12th lag falls above the confidence limit. Fig. 5(d) presents the time series plot of the training and testing errors and a close agreement between them is noticed. However, the variations are minimum for the exponential smoothing [Fig. 6(d)] than the HW smoothing [Fig. 7(d)]. Fig 5(e) depicts the epoch at which the MSE is the minimum and for the temperature series, it is at the 18th epoch. Fig. 5(f) depicts the gradient for 24 epochs and estimation of parameter at each epoch. Similarly, the results of other figures can be interpreted. Contrary to Fig. 5(b), Fig. 6(b) depicts the fitted regression line after smoothing the data with exponential smoothing and it is observed that the line is well fitted as compared to the Fig. 5(b). The minimum MSE is observed at 7th epoch with exponential smoothing [Fig. 6(e)] while at 9th epoch for the HW smoothing [Fig. 7(e)].

Next, the MLR model is fitted to predict the monthly maximum temperature (°C) as the response variable, while the other monthly variables such as rainfall, the minimum temperature (°C), relative humidity (RH) at 8 am and wind speed (Km/H), considered as the predictors. The stepwise regression analysis was used for selecting the variables. The estimated time series regression model is given below :

$$Y_t = 23.9995 + 0.8096X_{2t} - 0.0426X_{3t} - 0.3569X_{5t} \quad (4)$$

where minimum temperature (X_2), relative humidity (RH) at 8 am (X_3), relative humidity at 5 pm (X_4) and wind speed (X_5). Note that the coefficient of determination was 95.19% for the above model and the most contributing variable for predicting the maximum temperature (Y) is the minimum temperature. The variable rainfall (X_1) has the least effect on the maximum temperature, because the p -value is greater than 5% and hence removed from the model.

In Table 6, a comparison of different models is tabulated and it is observed that the minimum RMSE (0.1017) is obtained from the moving average smoothed data. Moreover, the moving average smoothed data has the highest correlation coefficient ($r = 0.9866$) using the ANN algorithm. However, the performance of the MLR and SARFIMA models is observed very poor as compared to the ANN models. Thus, it is safe to conclude that the ANN method is the most efficient with accurate predictions as compared to the ARIMA and time series multiple regression models.

4. Conclusion

The aim of this study is to evaluate the accuracy of different weather models through the ANN algorithm. For this purpose, we evaluated the out-of-sample RMSE and correlation coefficient of different weather forecasting models. Moreover, two types of transformations, namely the moving average (MA) and the Holt-Winter exponential smoothing, are also compared in this study. We considered the weather data from Lahore station collected by the Pakistan meteorological department from January 1951 to December 2015 on different weather parameters. Multiple Linear Regression (MLR), Seasonal Autoregressive Fractional Integrated Moving Average (SARFIMA) and dynamic Artificial Neural Network (ANN) models are used to estimate the accuracy of the weather parameters. Initially, MLR, SARFIMA and ANN are applied to the original monthly weather parameter data. Then, the ANN model is fitted to the transformed data obtained by different smoothing techniques. The performances of these three ANN (original, smoothed by MA and Holt-Winter smoothing operators) models are compared to the SARFIMA and multiple linear regression models.

The results tabulated in previous sections suggest that the ANN model applied using MA transformation yield the best results for the weather forecasting (Table 3 and Table 6). We also observed that the ANN algorithm applied to the smoothed data results into a smaller RMSE and relatively large correlation coefficient as compared to the MLR and SARFIMA models.

Supplementary Material : The supplementary material associated with this article can be downloaded from <https://drive.google.com/file/d/1szXV0hWk55etnld36sPpy5hynBVJmIn/view?usp=sharing>.

Acknowledgements

The authors would like to thank the anonymous referee for his constructive comments to improve the presentation of the work.

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., Singh, M. P., Ghosh, S. and Anand, A., 2012, "Weather forecasting model using artificial neural networks", *Procedia Technology*, **4**, 311-318.
- Aftab, S., Ahmad, M., Hameed, N., Bashir, M. S., Ali, I. and Nawaz, Z., 2018, "Rainfall prediction in Lahore city using data mining techniques", *International Journal of Advanced Computer Science and Applications*, **9**, 4, 254-260.
- Anderson, M. D., Sharfi, K. and Gholston, S. E., 2006, "Direct demand forecasting model for small urban communities using multiple linear regression", *Journal of the Transportation Research Board*, 1981, 114-117.
- Baillie, T. R., 1996, "Long memory processes and fractional integration in econometrics", *Journal of Econometrics*, **73**, 5-59.
- Bali, Y.N., 1970, "On objective estimation of crop yield at pre-harvest stage", *Agricultural Situations in India*, **25**, 3, 267-271.
- Chattopadhyay, S., 2007, "Feed forward artificial neural network model to predict the average summer monsoon rainfall in India", *Acta Geophysica*, **55**, 369-382.
- Dibike, Y. B. and Solomatine, D. P., 2001, "River flow forecasting using artificial neural networks, Physics and Chemistry of the Earth, Part B: Hydrology", *Oceans and Atmosphere*, **26**, 1-7.
- El-Shafie, A.H., El-Shafie, A., El-Mazoghi, H.G., Shehata, A. and Taha, M. R., 2011, "Artificial neural network technique for rainfall forecasting applied to Alexandria, Egypt", *International Journal of the Physical Science*, **6**, 1306-1316.
- Franses, P. H. and Ooms, M., 1997, "Periodic long-memory model for quarterly UK inflation", *International Journal of Forecasting*, **13**, 117-126.
- Ghani, I. M. M. and Ahmad, S., 2010, "Stepwise Multiple Regression Method to forecast fish landing", *Procedia-Social and Behavioral Sciences*, **8**, 549-554.
- Goulden, C. H., 1962, "Methods of statistical analysis", John Wiley & Sons, New York.
- Hayati, M. and Mohebi, Z., 2007, "Temperature forecasting based on neural network approach", *World Applied Science Journal*, **2**, 613-620.
- Huda, A. K. S., Ghildyal, B. P. and Tomar, V. S., 1976, "Contribution of climatic variables in predicting maize yield under monsoon conditions", *Agricultural Meteorology*, **17**, 33-47.
- Huda, A. K. S., Ghildyal, B. P., Tomar, V. S. and Jain, R. C., 1975, "Contribution of climatic variables in predicting rice yield", *Agricultural Meteorology*, **15**, 71-86.
- Hung, N. Q., Babel, M. S., Weesakul, S. and Tripathi, N., 2009, "An artificial neural network model for rainfall forecasting in Bangkok, Thailand", *Hydrology and Earth System Sciences*, **13**, 1413-1425.
- Kannan, M., Prabhakaran, S. and Ramachandran, P., 2010, "Rainfall forecasting using data mining technique", *International Journal of Engineering and Technology*, **2**, 397-401.
- Khedhiri, S., 2015, "Artificial neural network for forecasting rainfall pattern in Prince Edward Island, Canada", *International Journal of Environmental Studies*, **72**, 331-340.
- Kihoro, J. M., Otieno, R. O. and Wafula, C., 2004, "Seasonal time series forecasting: A comparative study of ARIMA and ANN models", *African Journal of Science and Technology*, **5**, 41-49.
- Kumar, D. N., Raju, K. S. and Sathish, T., 2004, "River flow forecasting using recurrent neural networks", *Water resources Management*, **18**, 143-161.
- Kumarasiri, A. D. and Sonnadra, U. J., 2008, "Performance of an artificial neural network on forecasting the daily occurrence and annual depth of rainfall at a tropical site", *Hydrological Processes*, **22**, 3535-3542.
- Li, H. and Ye, X., 2015, "Forecasting high-frequency long memory series with long periods using the SARFIMA model", *Open Journal of Statistics*, **5**, 66-74.
- Litta, A. J., Idicula, S. M. and Mohanty, U. C., 2013, "Artificial neural network model in prediction of meteorological parameters during premonsoon thunderstorms", *International Journal of Atmospheric Sciences*, 2013, Article ID 525383, 14 pages, <http://dx.doi.org/10.1155/2013/525383>.
- MATLAB, 2010, "The MathWorks", Inc., Natick, Massachusetts, United States.
- Mesgari, E., Asheri, A., Hooshyar, M. and Hemmesy, M. S., 2015, "Rainfall modeling and forecasting using neural networks: A case study of Zeb watershed", *International Bulletin of Water Resources and Development*, **3**, 24-31.
- Nayak, R. D., Mahapatra, A. and Mishra, P., 2013, "A survey on rainfall prediction using artificial neural network", *International Journal of Computer Applications*, **72**, 32-40.
- Nirmala, M., 2015, "Computational models for forecasting annual rainfall in Tamilnadu", *Applied Mathematical Sciences*, **9**, 617-621.
- Olatayo, O. T. and Adedotun, F. A., 2014, "On the Test and Estimation of Fractional Parameter in ARFIMA Model : Bootstrap approach", *Applied Mathematical Sciences*, **8**, 4783-4792.
- Rahman, M. H. and Matin, M. A., 2015, "On the prediction of average monsoon rainfall in Bangladesh with artificial neural network", *International Journal of Computer Applications*, **127**, 45-52.
- Ramchandran, G., 1967, "Rainfall distribution in India in relation to longitude-latitude and elevation", *Indian Journal of Meteorological Geophysics*, **18**, 227-232.
- Sahai, A. K., Soman, M. K. and Satyan, V., 2000, "All India summer monsoon rainfall prediction using an artificial neural network", *Climate Dynamics*, **16**, 291-302.
- Shamshad, B., Khan, M. Z. and Omer, Z., 2019, "Modeling and forecasting weather parameters using ANN-MLP, ARIMA and ETS model : A case study for Lahore, Pakistan", *International Journal of Scientific & Engineering Research*, **10**, 4, 351-366.

- Shrivastava, G., Karmakar, S., Kowar, M. K. and Guhathakurta, P., 2012, "Application of artificial neural networks in weather forecasting: A comprehensive literature review", *International Journal of Computer Applications*, **51**, 17-29.
- Sohn, T., Lee, J. H., Lee, S. H. and Ryu, C. S., 2005, "Statistical prediction of heavy rain in South Korea", *Advances in Atmospheric Sciences*, **22**, 703-710.
- Sowell, F., 1992, "Maximum likelihood estimation of stationary univariate fractionally integrated time series models", *Journal of Econometrics*, **53**, 165-188.
- Valipour, M., 2015, "Long-term runoff study using SARIMA and ARIMA models in the United States", *Meteorological Applications*, **22**, 592-598.
- Wu, C. L., Chau, K. L. and Fan, C., 2010, "Prediction of rainfall time series using modular artificial neural networks coupled with data pre-processing techniques", *Journal of Hydrology*, **389**, 146-167.
- Zaw, W. T. and Naing, T. T., 2008, "Empirical Statistical Modelling of Rainfall prediction over Myanmar", *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, **46**, 3418-3421.

APPENDIX

TABLE A1
Summary of the Data

Summary	Temperature (Minimum)	Temperature (Maximum)	Humidity (8 AM)	Humidity (5 PM)	Wind Speed	Rainfall
Minimum	15.2	2.1	27	13	0	0
Q1	24.9	10.7	58	34	0.5	4.4
Median	32.9	19.2	72	45	1.3	20.55
Mean	30.78	18.26	68.08	43.65	1.374	52.6
Q3	35.8	26	79	54	2	62.05
Maximum	43.6	29.6	95	76	9	64

A comparison of different weather forecasting models for the monthly forecast of Lahore city

SAJID ALI, ZAHEER ABBAS* and MUHAMMAD MOEEN BUTT**

Department of Statistics, Quaid-i-Azam University, Islamabad, 45320, Pakistan

**Department of Statistics, Government Post Graduate College Sahiwal, Sahiwal, Pakistan*

***Department of Economics and Statistics, Dr. Hasan Murad School of Management, University of Management and Technology, Lahore, Pakistan*

e mail : sajidali.qau@hotmail.com

S1. Relative Humidity at 8am

For relative humidity level at 8 am, we first calculated the difference parameter.

In Table S1, we tabulated the results of a SARFIMA model for relative humidity at 8am. From the table, we observed that data were generated by a long-period time series with fractional parameter $d = 0.4858$. Moreover, the

data were stationary as $d < 0.5$. Next, we estimated the parameters of the SARFIMA model in Table S2.

In Table S2, we tabulated the parameter estimates of the ARFIMA model. Moreover, the AIC and BIC were observed 6.3884 and 6.4369, respectively. It is observed from the table that the relative humidity at 8am data can be modeled by SARFIAM (2, 0.4858, 2) (1, 0, 1) with $s = 12$. Next, we applied the ANN model.

TABLE S1

Estimation of fractional differencing

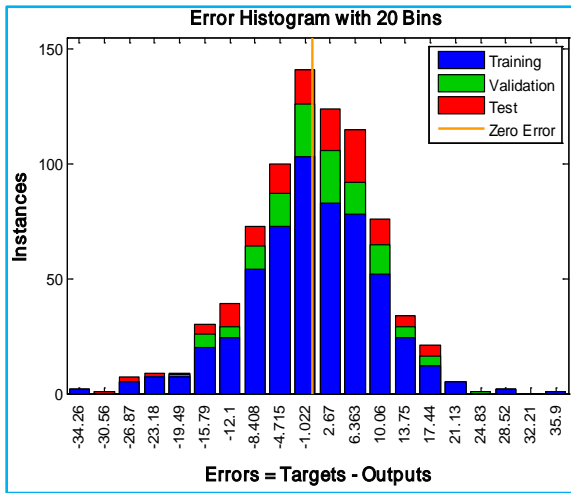
Variable	Coefficient	Std. Error	<i>t</i> -Statistic	Prob.	95% [Confidence Interval]	
Y_cons	68.54277	311.9646	0.219713	0.8262	-543.8681	680.9536
ARFIMA_d	0.485889	0.008050	61.59902	0.0000	0.480085	0.511692
SIGMA2	165.6234	9.974184	16.60521	0.0000	146.0433	185.2035

Log likelihood = -3046.644 F-statistic = 121.2680 Prob (F-statistic) = 0.0000

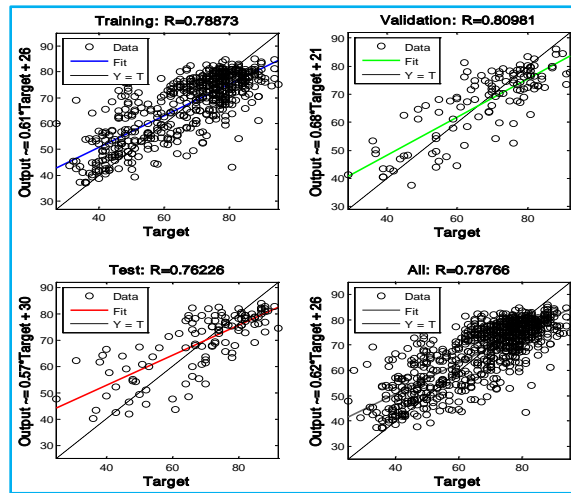
TABLE S2

Estimation of the SARFIMA model

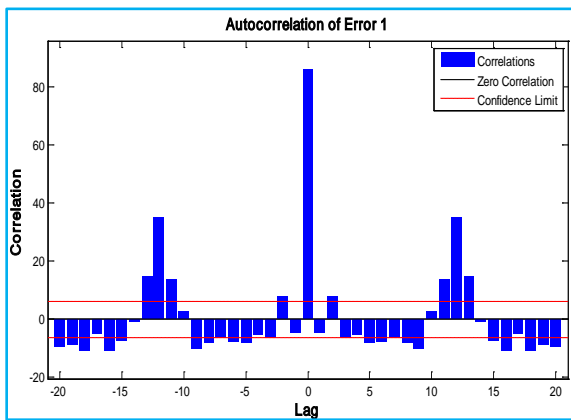
Type	Coefficient	SE Coefficient	T	P
AR(1)	AR(1)	-0.628959	0.110082	-5.713543
AR(2)	AR(2)	0.352358	0.105953	3.325606
MA(1)	MA(1)	0.931499	0.114934	8.104662
MA(2)	MA(2)	-0.065586	0.114150	-0.574559
SAR(12)	SAR(12)	0.998552	0.000885	1127.893
SMA(12)	SMA(12)	-0.887298	0.021426	-41.41167
Constant	Constant	68.41038	5.091542	13.51465



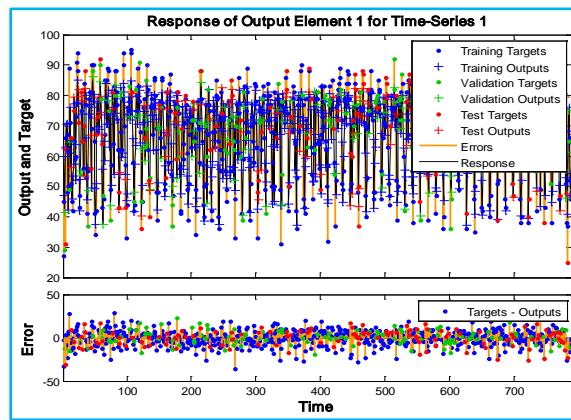
(a) Error Histogram



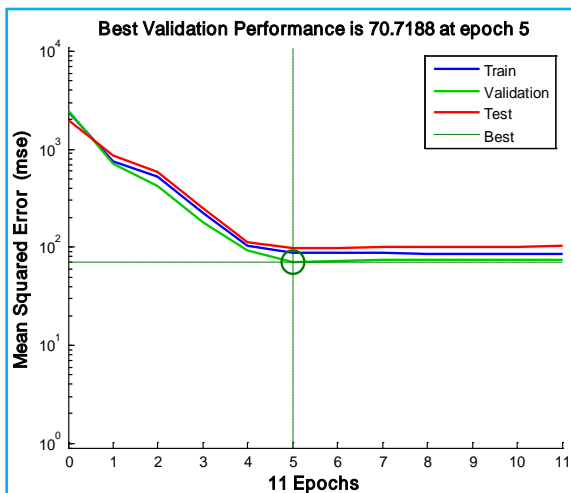
(b) Regression results



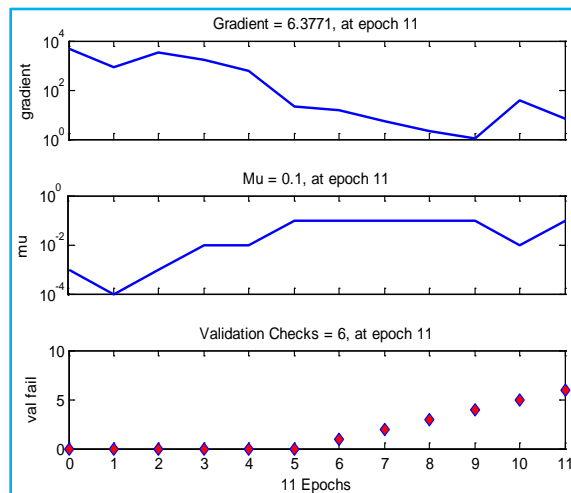
(c) Error Autocorrelation



(d) Series Responses

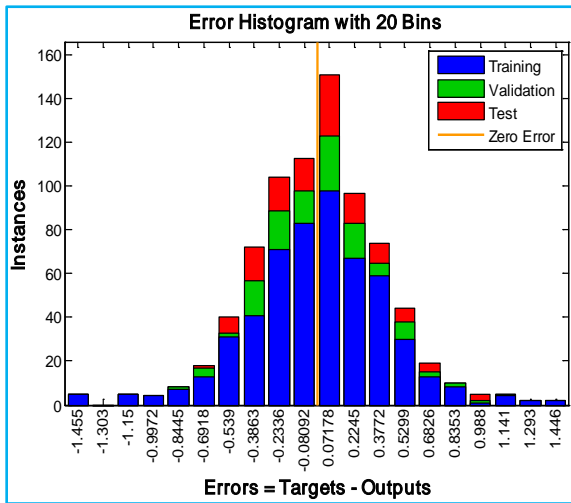


(e) Network Performances

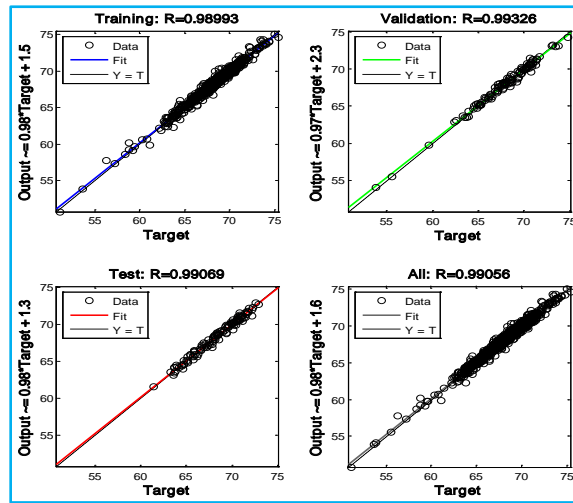


(f) Test results

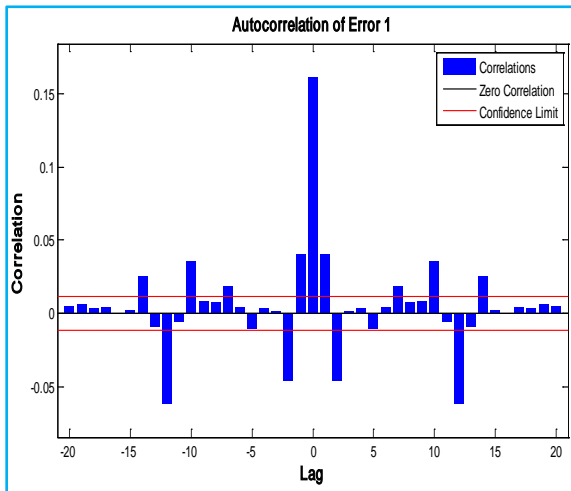
Figs. S1(a-f). Neural network results for Y (original Humidity level data recorded at 8am)



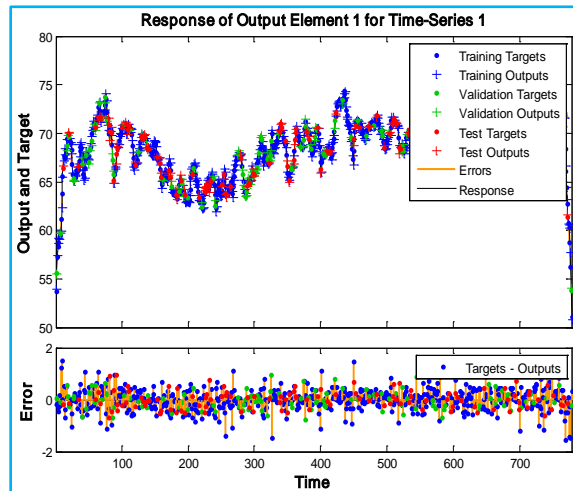
(a) Error Histogram



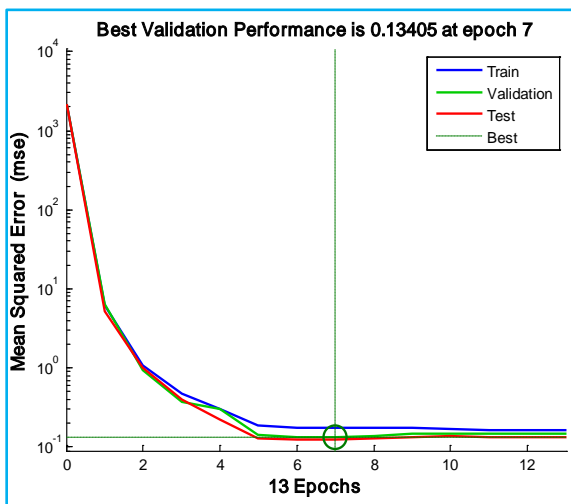
(b) Regression results



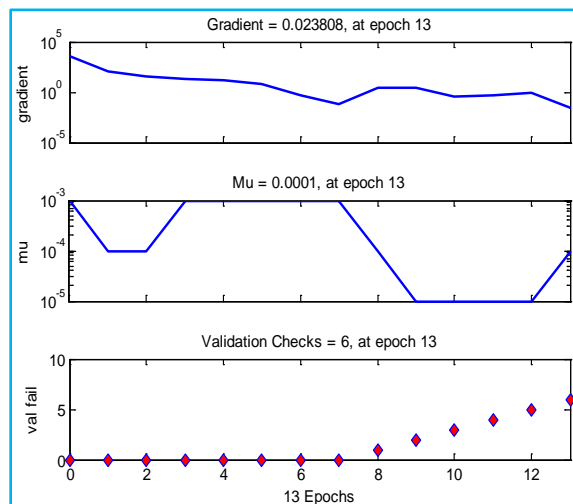
(c) Error Autocorrelogram



(d) Series Responses

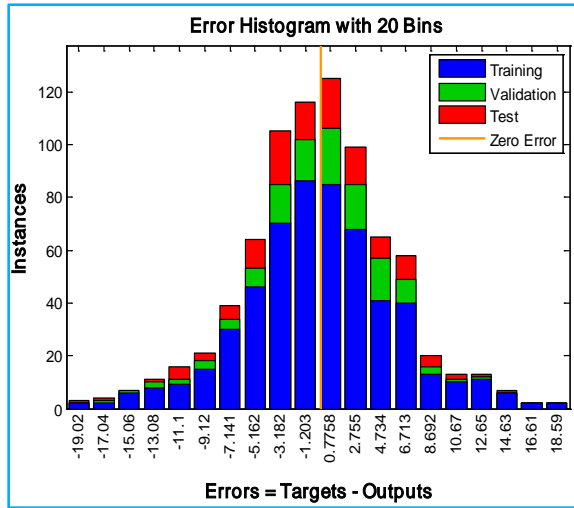


(e) Network Performances

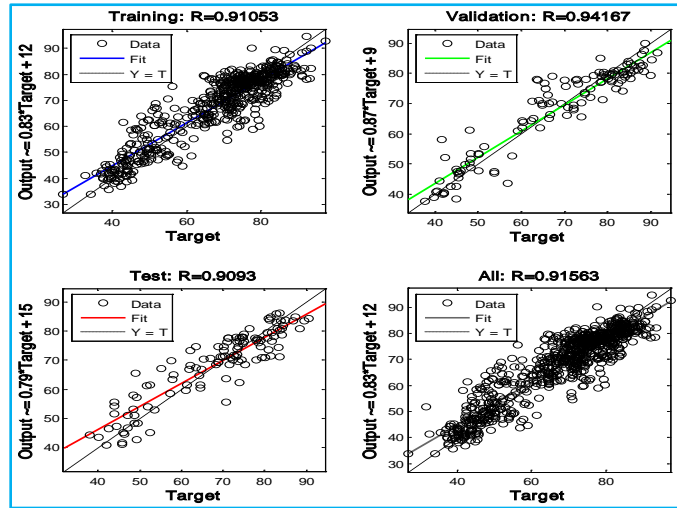


(f) Test results

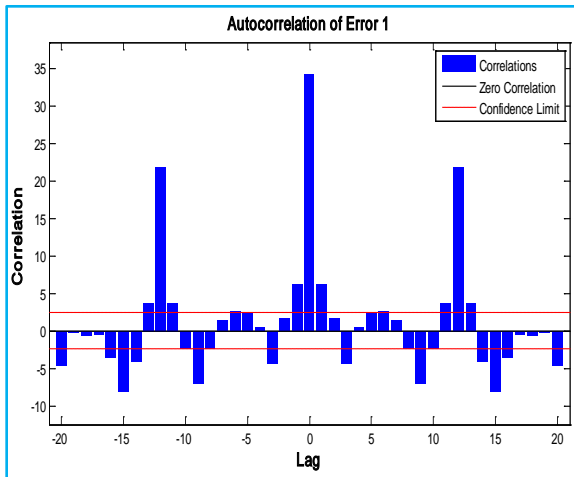
Figs. S2(a-f). Neural network results for Y (data smoothed by moving average)



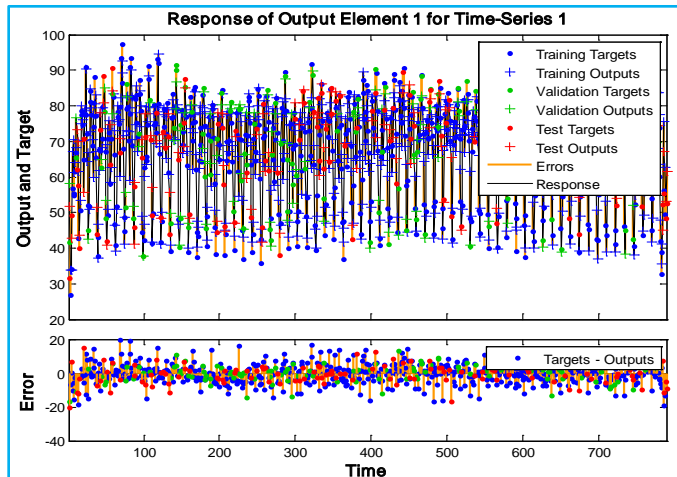
(a) Error Histogram



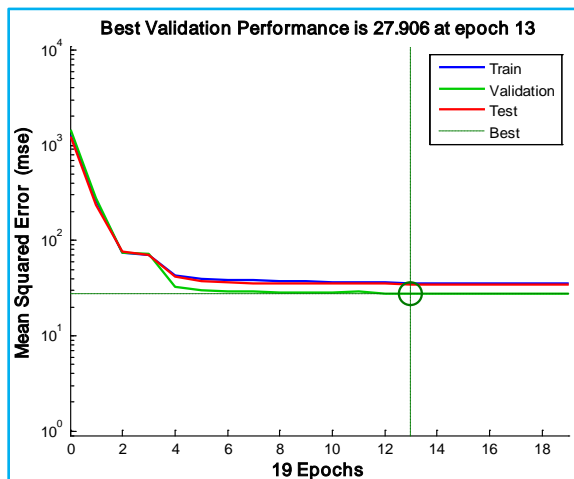
(b) Regression results



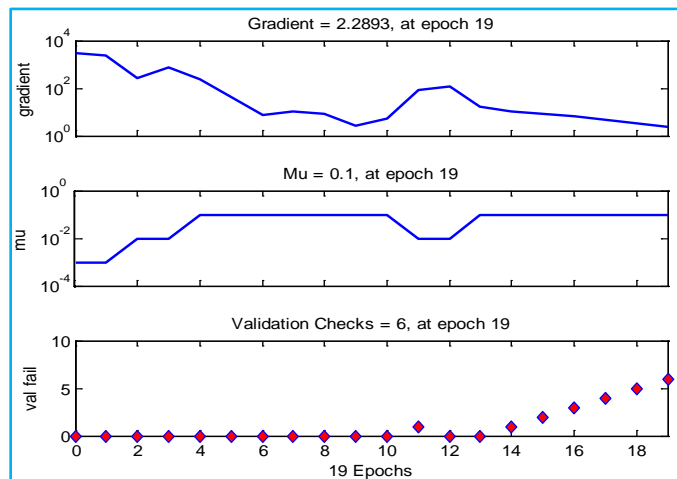
(c) Error Autocorrelogram



(d) Series Responses



(e) Network Performances



(f) Test results

Figs. S3(a-f). Neural network results for Y (data smoothed by Holt-Winter smoothing)

TABLE S3

Forecasting performance of the relative humidity recorded at 8am of different models

Time series	RMSE	Correlation coefficient
Y	8.4094	0.8098
YMA	0.3661	0.9933
YHWE	5.2826	0.9417
Y (SARFIMA)	14.9468	0.9223
Time Series Regression	5.3023	0.9351

TABLE S4

Estimation of fractional differencing

Variable	Coefficient	Std. Error	t-Statistic	Prob.	95% [Confidence Interval]	
Y_cons	43.85638	90.13159	0.486582	0.6267	-133.0790	220.7917
ARFIMA_d	0.486042	0.022236	21.85841	0.0000	0.442391	0.529693
SIGMA2	146.7108	8.509548	17.24073	0.0000	130.0059	163.4157

Log likelihood = -2999.569 F-statistic = 78.89917 Prob (F-statistic) = 0.00000

TABLE S5

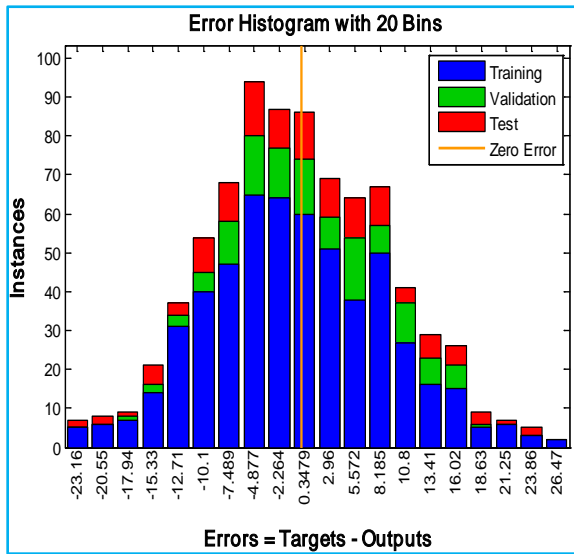
Estimation of the SARFIMA model

Type	Coefficient	SE Coefficient	T	P
AR(1)	-1.214235	0.036899	1.324868	0.0000
AR(2)	-0.460166	0.056110	-8.201145	0.0000
AR(3)	0.325172	0.034886	9.321069	0.0000
MA(1)	0.532690	0.110891	4.803706	0.0000
MA(2)	-0.550046	0.186047	-2.956495	0.0032
MA(3)	-0.982644	0.536247	-1.832447	0.0073
SAR(12)	0.999190	0.000656	1523.790	0.0000
SMA(12)	-0.935777	0.019160	-48.83969	0.0000
Constant	0.007184	0.005422	1.324868	0.1856

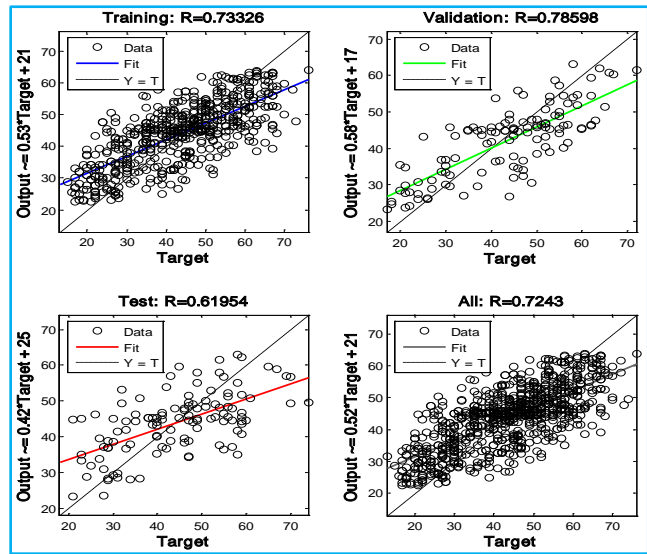
In Figs. S1-S3, we have shown the results of ANN modeling for three relative humidity series recorded at 8 am. It was observed that the performance of the best validation occurred at epoch 6th for the maximum temperature original series while at 7th epoch for the moving average and 13th epoch for the Holt-Winter exponential pre-processed data. We also observed high correlation coefficient between estimated and target. The back propagation algorithm showed a very high-accuracy

level with the target values for data smoothed by the moving average.

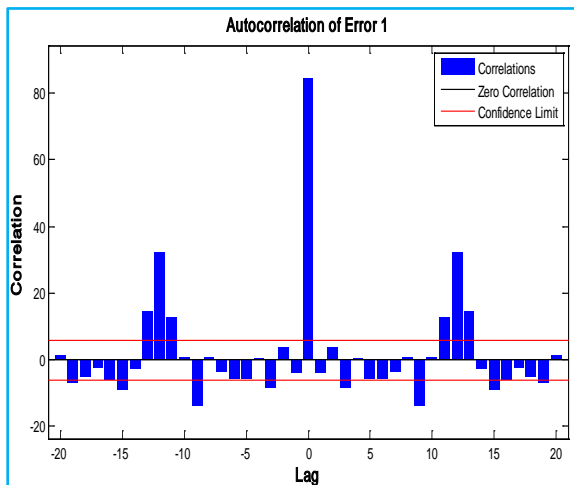
Next, we fitted the time-series multiple linear regression (MLR) model to predict the monthly relative humidity (RH) recorded at 8am as a response variable, while the other monthly variables such as rainfall, maximum temperature, minimum temperature relative humidity (RH) recorded at 5 pm and wind speed



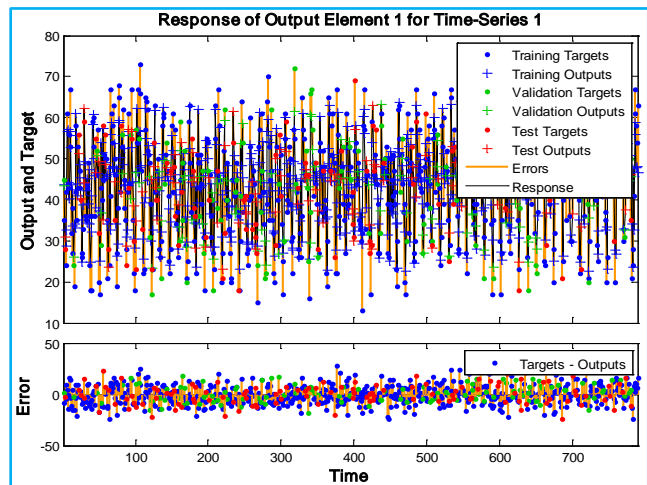
(a) Error Histogram



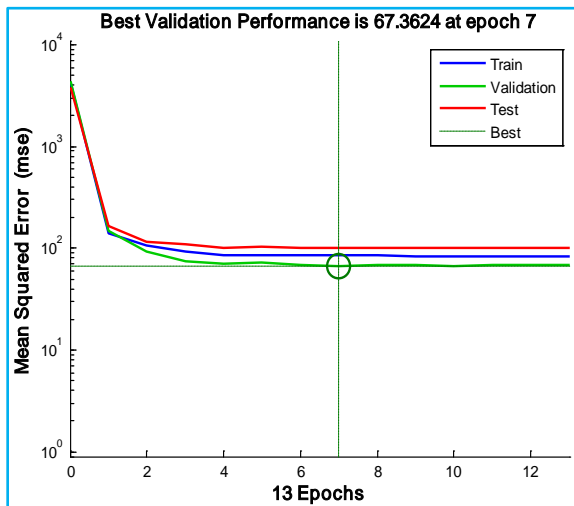
(b) Regression results



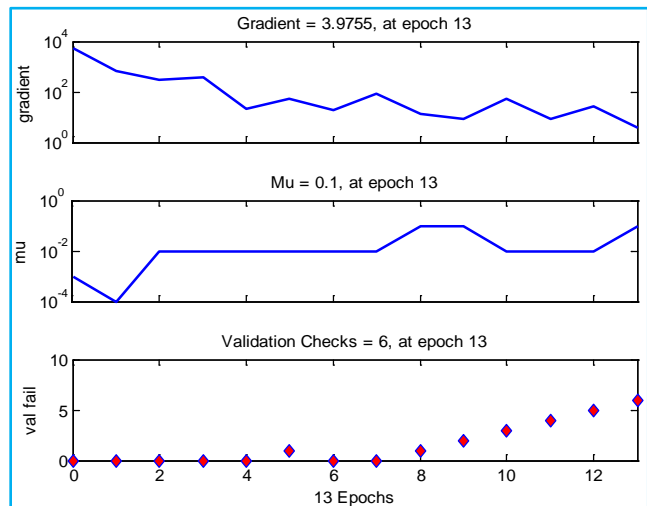
(c) Error Autocorrelation



(d) Series Responses

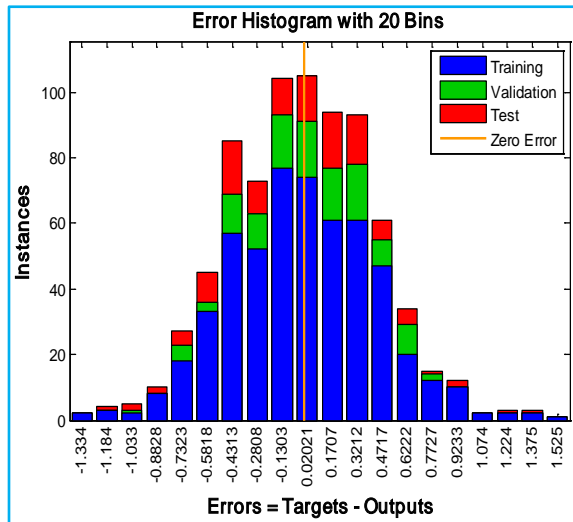


(e) Network Performances

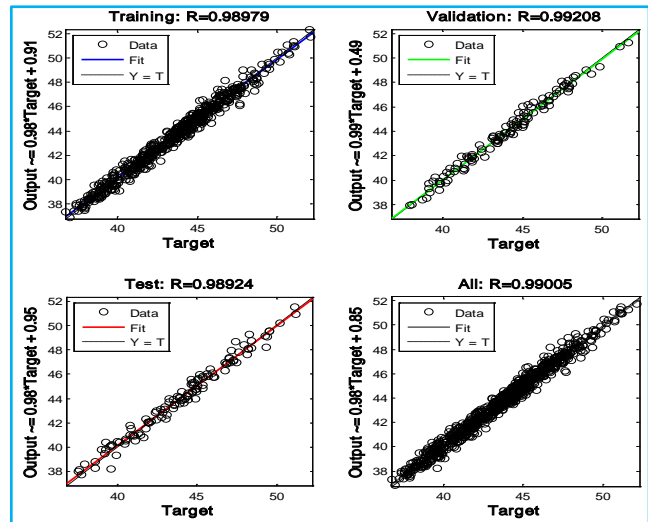


(f) Test results

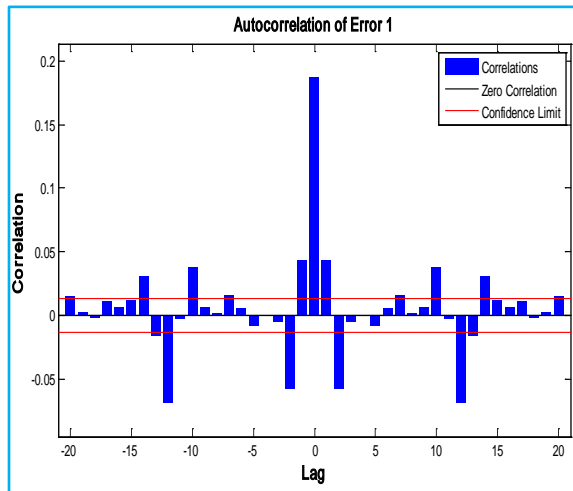
Figs. S4(a-f). Neural network results for Y (original series)



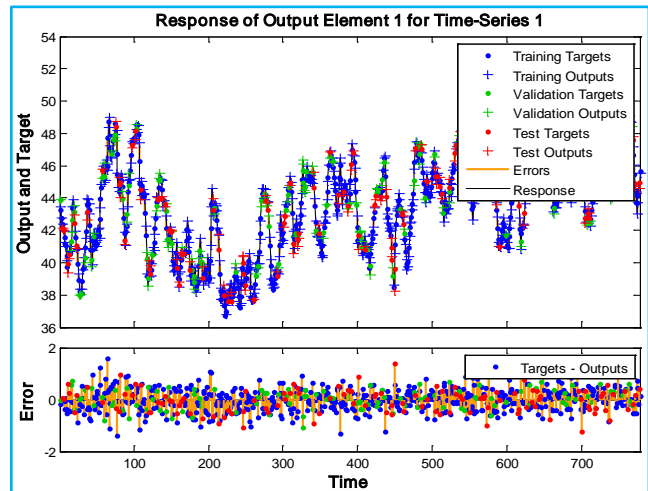
(a) Error Histogram



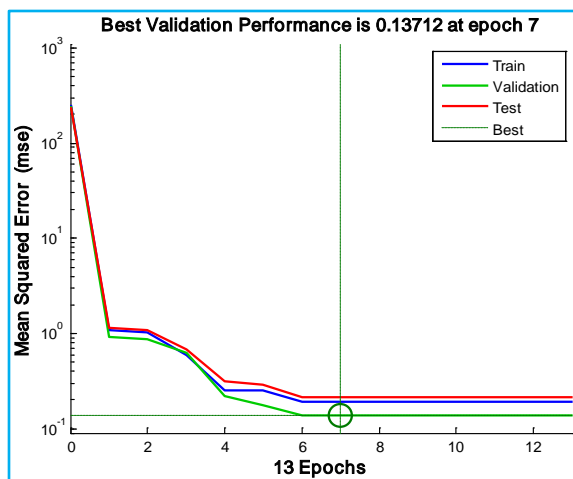
(b) Regression results



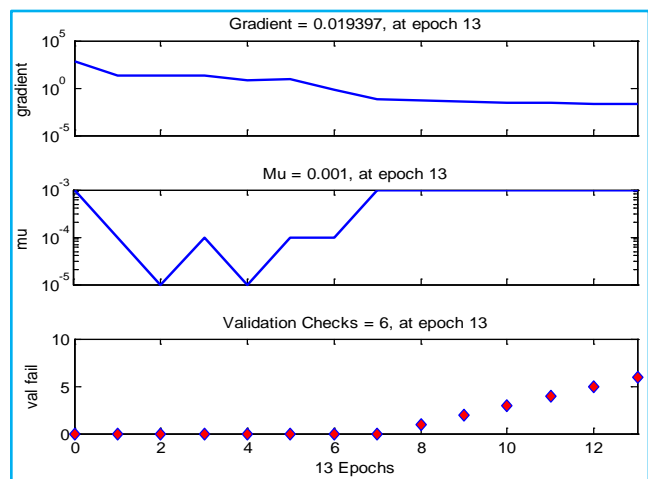
(c) Error Autocorrelogram



(d) Series Responses

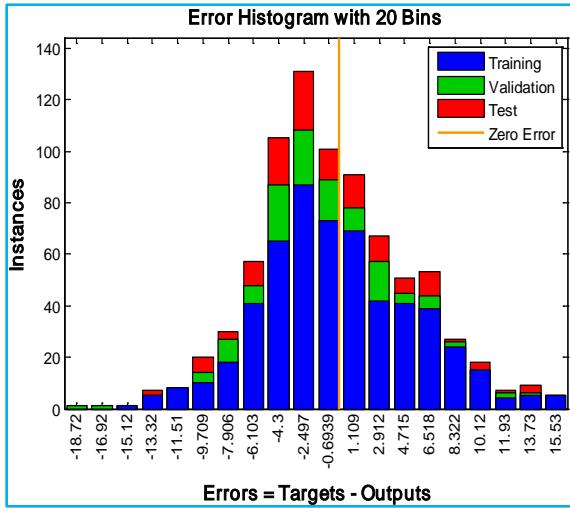


(e) Network Performances

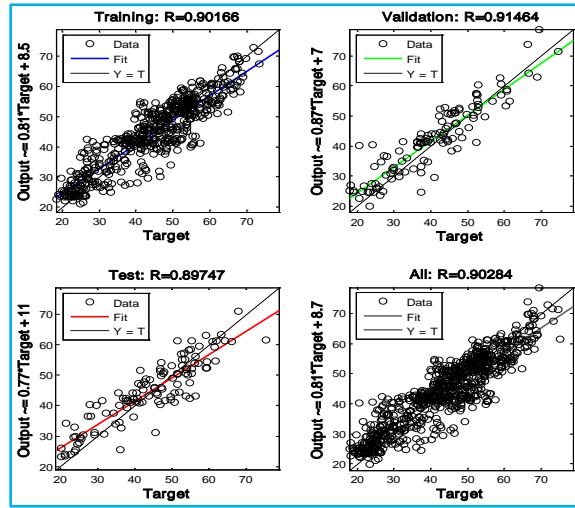


(f) Test results

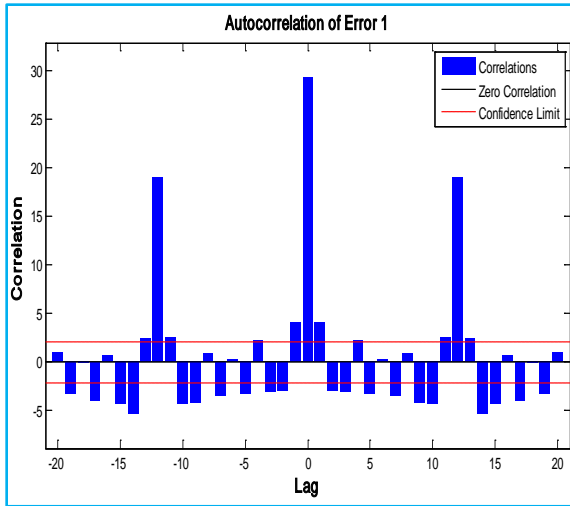
Figs. S5(a-f). Neural network results for Y (smoothed by Moving average)



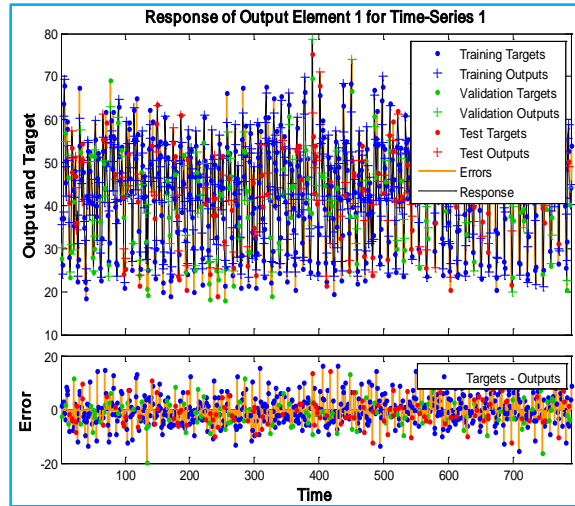
(a) Error Histogram



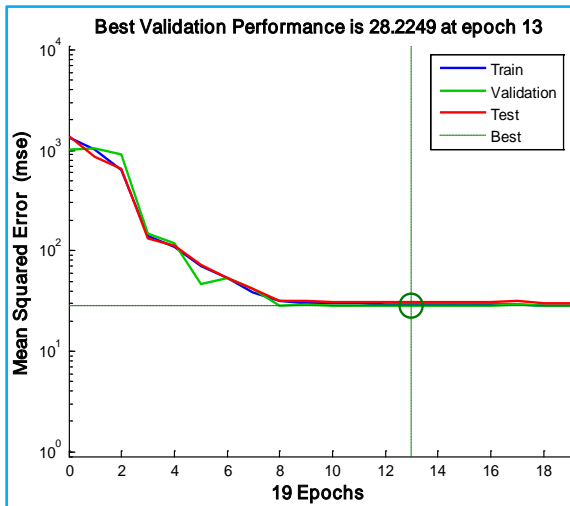
(b) Regression results



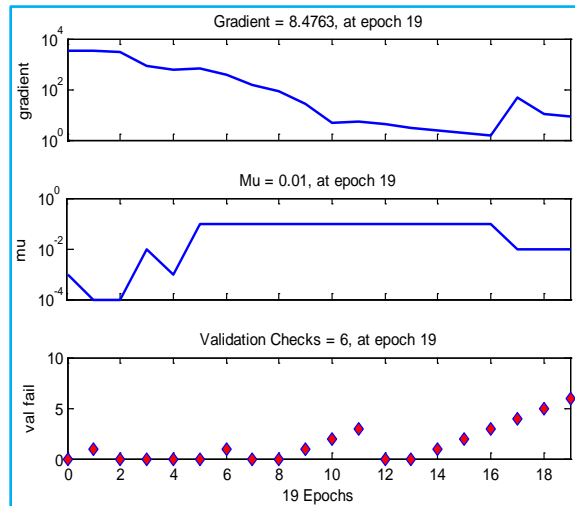
(c) Error Autocorrelation



(d) Series Responses



(e) Network Performances



(f) Test results

Figs. S6(a-f).Neural network results for Y (smoothed by Holt-Winter smoothing)

considered as the predictors. The stepwise regression analysis was used for selecting an appropriate set of predictors. The estimated time series regression model is given below:

$$Y_t = 55.2191 + 0.0057X_{1t} - 0.5212X_{2t} - 0.2456X_{3t} + 0.7985X_{4t} - 1.2881X_{5t} \quad (1)$$

where rainfall (X_1), maximum temperature (X_2), minimum temperature (X_3), relative humidity at 5 pm (X_4) and wind speed (X_5). Note that the coefficient of determination was 85.08% for the above model and the most contributing variables for predicting the humidity at 8 am (Y), was relative humidity recorded at 5 pm.

In Table S3, the performance of different models for forecasting relative humidity at 8am has been listed and we observed that the moving average process had the minimum RMSE. Thus, the moving average smoothed data has the highest correlation coefficient using the ANN algorithm. Also, the performance of MLR and SARFIMA models was observed very poor as compared to the ANN models. Thus, we conclude that the ANN method was the most efficient for producing accurate predictions as compared to the ARIMA and time series multiple regression models.

S2. Relative Humidity level recorded at 5 pm

In Table S4, we tabulated the estimated results of a SARFIMA model of relative humidity recorded at 5 pm. From the table, it is observed that data were generated by a long-period time series with fractional parameter $d = 0.4860$ and the fitted model has been observed stationary as the value of d was less than 0.5.

In Table S5, we tabulated the parameter estimates of the ARFIMA model. Moreover, the AIC and BIC were observed 6.5925 and 6.6532, respectively. It is observed from the table that the relative humidity level recorded at 5 pm can be modeled by SARFIAM (3, 0.4860, 3) (1, 0, 1) with $s = 12$.

In Figs. S4-S6, we have depicted the results of ANN modeling for three types of series, *i.e.*, the first is the original series; the second is smoothed series by moving average while the last series is smoothed by Holt-Winter smoothing. It was observed that the performance of the best validation occurred at the epoch 7th for the original as moving average data and 13th epoch for the Holt-Winter exponential smoothing. We also observed a high correlation coefficient between the estimated and the target observations for the transformed relative humidity level recorded at 5 pm. The back propagation algorithm showed a very high-accuracy level with the target values

TABLE S6

Forecasting performance of the humidity at 5pm models

Time series	RMSE	Correlation coefficient
Y	8.2074	0.7860
YMA	0.3703	0.9921
YHWE	5.3127	0.9146
Y (SARFIMA)	14.0440	0.8633
Time Series Regression	5.0583	0.9248

for the smoothed data by moving average. Figs. S4-S6 depict the results of ANN modeling for three types of series, *i.e.*, the first is the original series; the second is smoothed series by moving average while the last series is smoothed by Holt-Winter smoothing. It was observed that the performance of the best validation occurred at epoch 7th for the original data while at 7th epoch for the moving average and 13th epoch for the Holt-Winter exponential smoothing. We also observed a high correlation coefficient between the estimated and the target observations for the transformed relative humidity level recorded at 5pm. The back propagation algorithm showed a very high accuracy level with the target values for the smoothed data by moving average.

Next, we fitted the time series multiple linear regression (MLR) model to predict the monthly relative humidity (RH) level recorded at 5pm as the response variable, while the other monthly variables such as rainfall, maximum and minimum temperature relative humidity (RH) level recorded at 8am and wind speed are considered as the predictors. The stepwise regression analysis was used for the selection of variables. The estimated time series regression model is given below:

$$Y_t = 3.9151 + 0.0097X_{1t} - 1.1603X_{2t} + 1.4337X_{3t} + 0.7267X_{4t} \quad (2)$$

where rainfall (X_1), maximum temperature (X_2), minimum temperature (X_3), relative humidity at 8am (X_4) and wind speed (X_5). Note that the coefficient of determination was 85.53% for the above model and the most contributing variable for predicting the humidity level (Y) was the minimum temperature. The variable wind speed (X_5) seemed to have less effect on the relative humidity, because the p -value was greater than 5% and hence did not include in the above model. In Table S6, a comparison of different models for the relative humidity level has been given and we observed that the moving average process had the minimum RMSE. Moreover, the moving average smoothed data has the highest correlation

TABLE S7

Estimation of fractional differencing

Variable	Coefficient	Std. Error	t-Statistic	Prob.	95% [Confidence Interval]	
Y_cons	1.438290	14.24254	0.100985	0.9196	-26.52093	29.39751
ARFIMA_d	0.494273	0.010005	49.40234	0.0000	0.474632	0.513914
SIGMA2	0.687668	0.014581	47.16176	0.0000	0.659045	0.716292

Log likelihood = -946.0387 F-statistic = 1476.9668 Prob (F-statistic) = 0.00000

TABLE S8

Estimation of the SARFIMA model

Type	Coefficient	SE Coefficient	T	P
AR(1)	1.069903	0.054916	19.48252	0.0000
AR(2)	-0.098465	0.038791	-2.538367	0.0113
MA(1)	-0.912846	0.045846	-19.91112	0.0000
SAR(12)	0.999426	0.000626	1595.261	0.0000
SMA(12)	-0.956456	0.022159	-43.16288	0.0000
Constant	1.421077	0.799907	1.776553	0.0760

TABLE S9

Forecasting performance of the wind-speed models

Time series	RMSE	Correlation coefficient
Y	0.6532	0.6681
YMA	0.0321	0.9919
YHWE	0.3087	0.9494
Y (SARFIMA)	1.1390	0.8084
Time Series Regression	0.6878	0.7100

TABLE S10

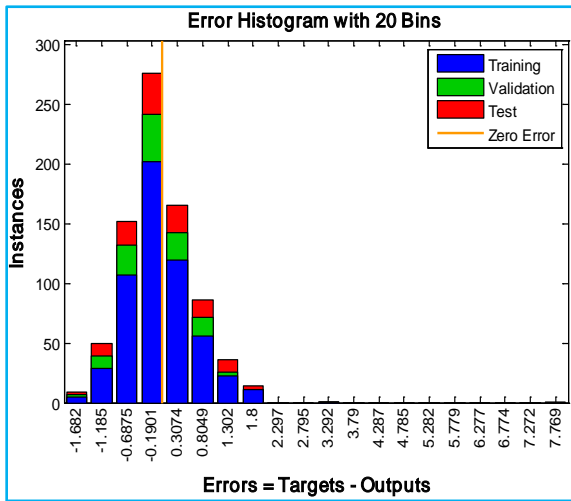
Estimation of fractional differencing

Variable	Coefficient	Std. Error	t-Statistic	Prob.	95% [Confidence Interval]	
Y_cons	18.27999	563.5407	0.032438	0.9741	-1087.997	1124.557
ARFIMA_d	0.498890	0.002265	220.2143	0.0000	0.494443	0.503338
SIGMA2	31.79289	2.741507	11.59687	0.0000	26.41109	37.17469

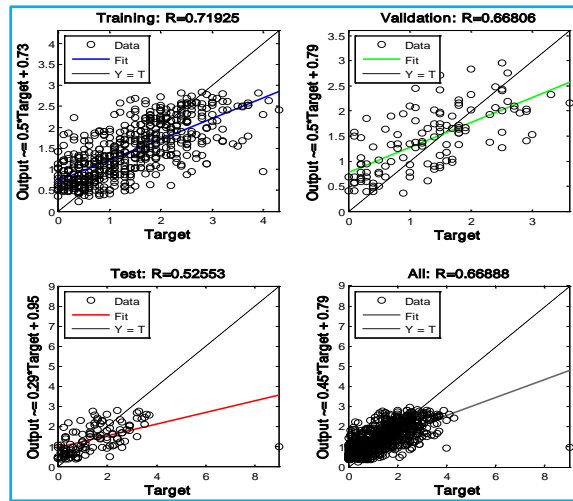
Log likelihood = -2412.025 F-statistic = 330.9707 Prob (F-statistic) = 0.00000

coefficient using the ANN algorithm. However, the performance of MLR and SARFIMA models was observed very poor as compared to the ANN models.

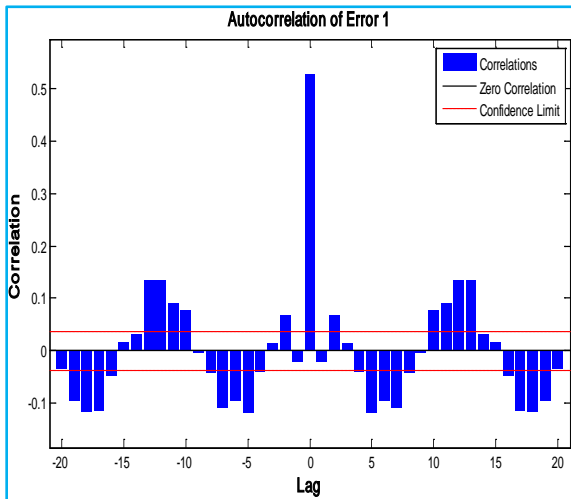
Thus, we conclude that the ANN method is the most efficient for producing accurate predictions as compared to the ARIMA and time series multiple regression models.



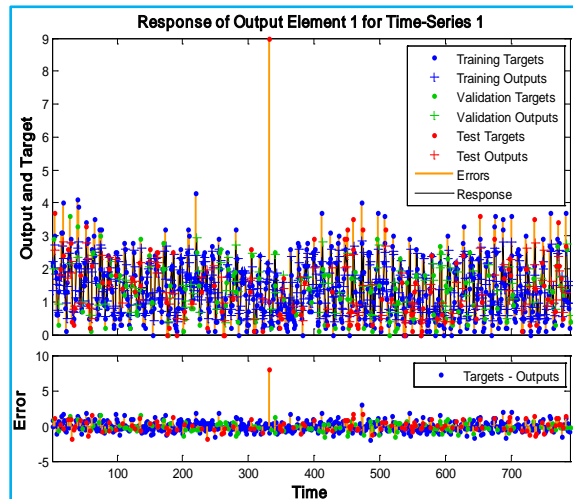
(a) Error Histogram



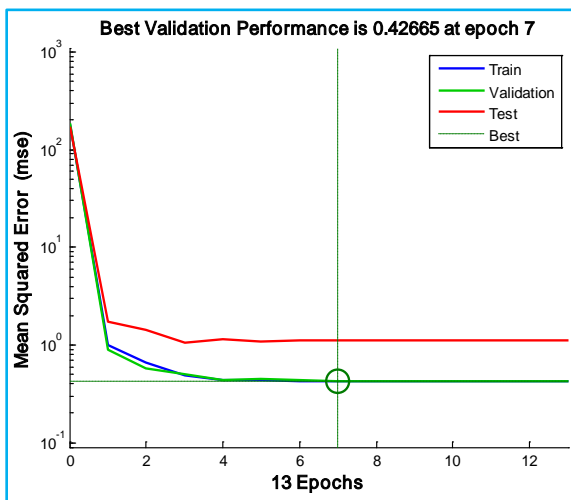
(b) Regression results



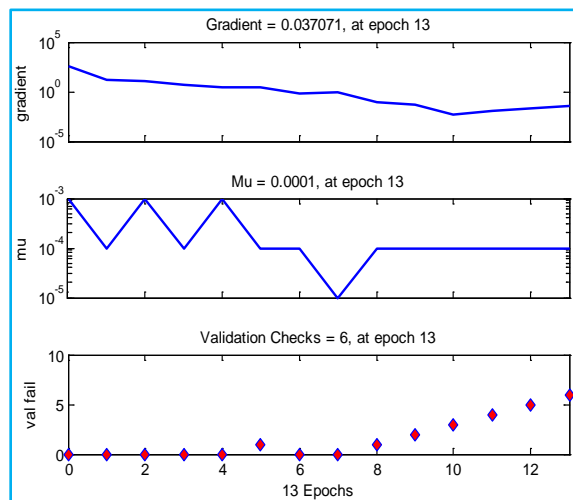
(c) Error Autocorrelation



(d) Series Responses

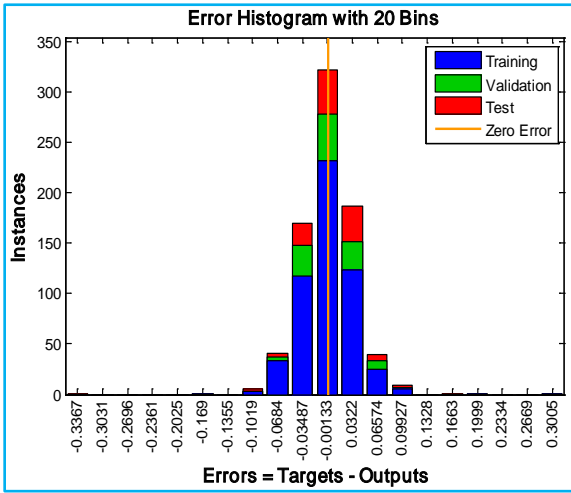


(e) Network Performances

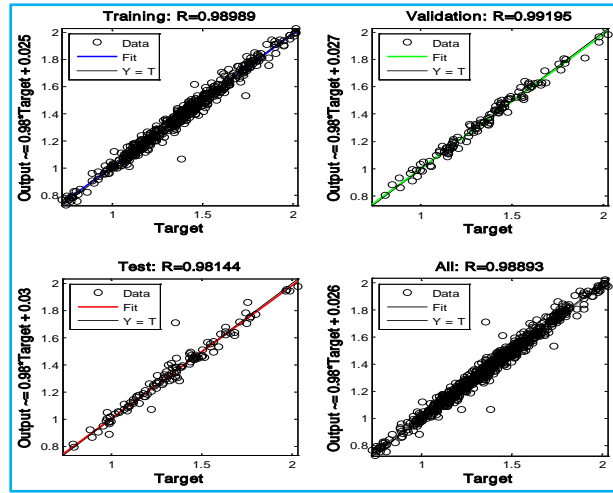


(f) Test results

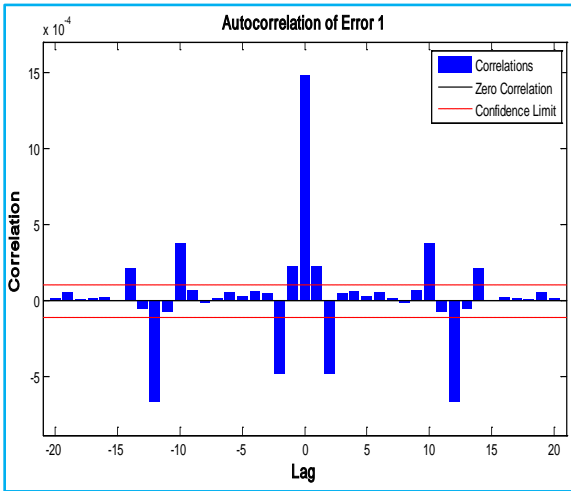
Figs. S7(a-f). Neural network results for Y (Wind-speed original data)



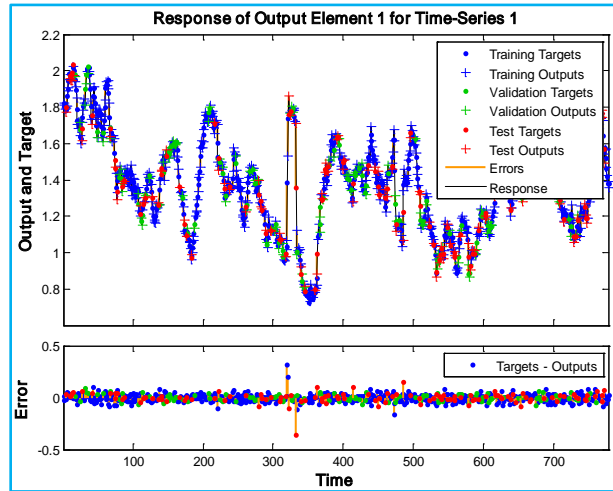
(a) Error Histogram



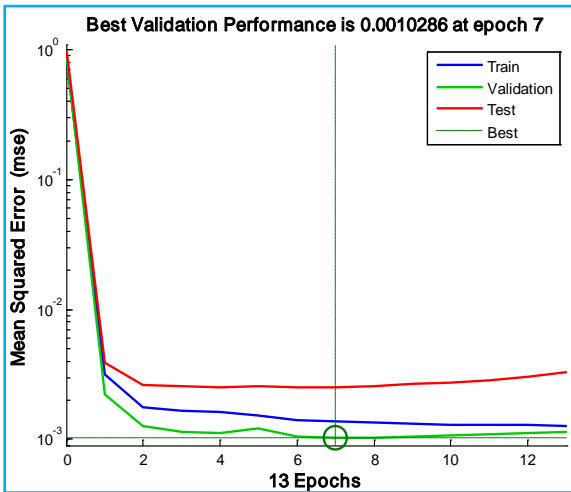
(b) Regression results



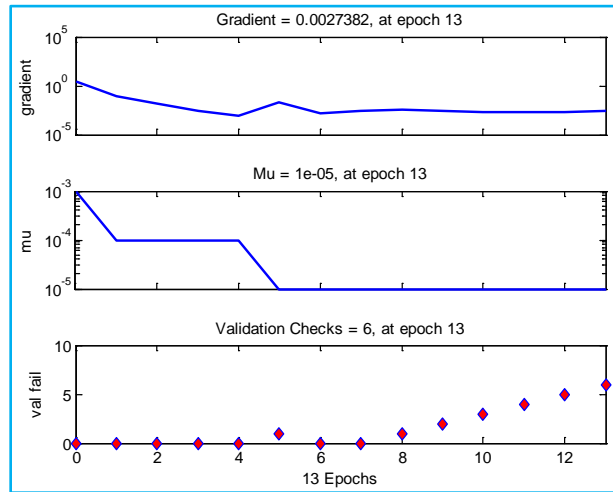
(c) Error Autocorrelogram



(d) Series Responses

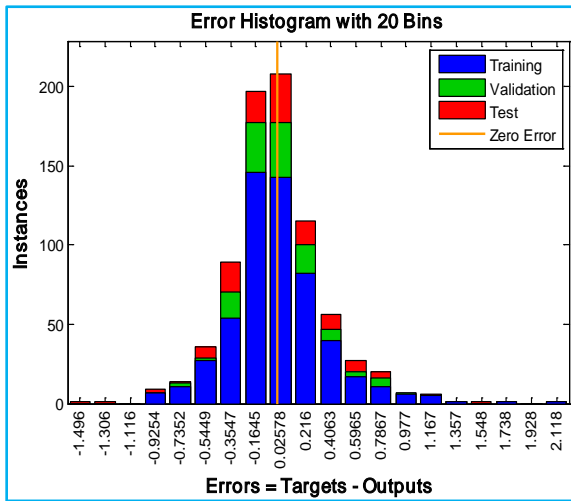


(e) Network Performances

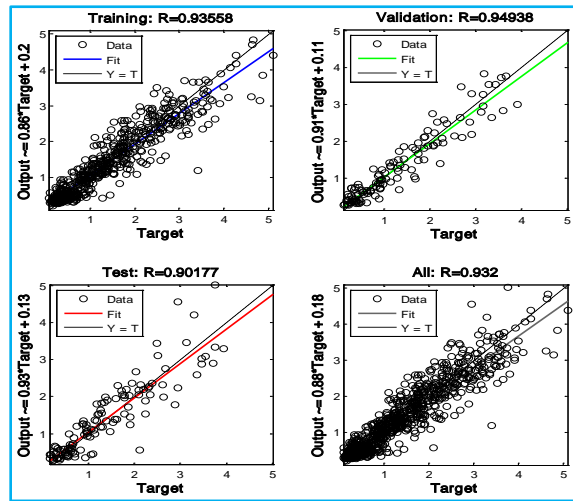


(f) Test results

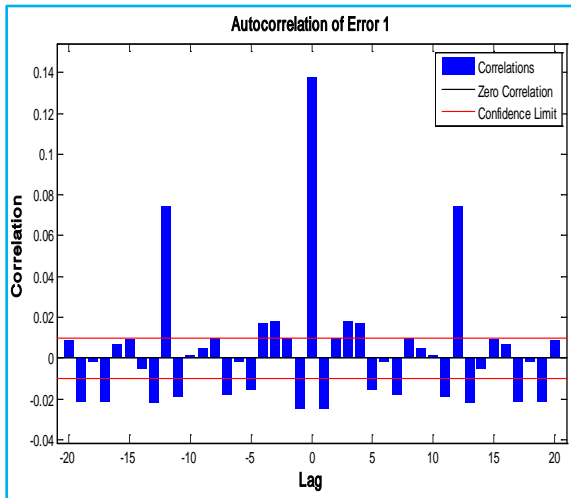
Figs. S8(a-f), Neural network results for Y (data smoothed by moving average)



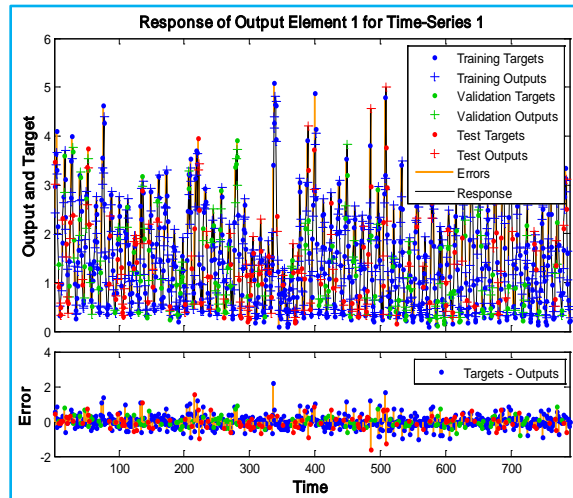
(a) Error Histogram



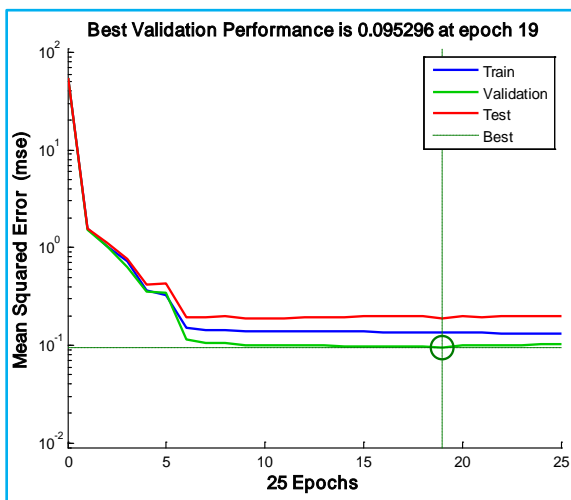
(b) Regression results



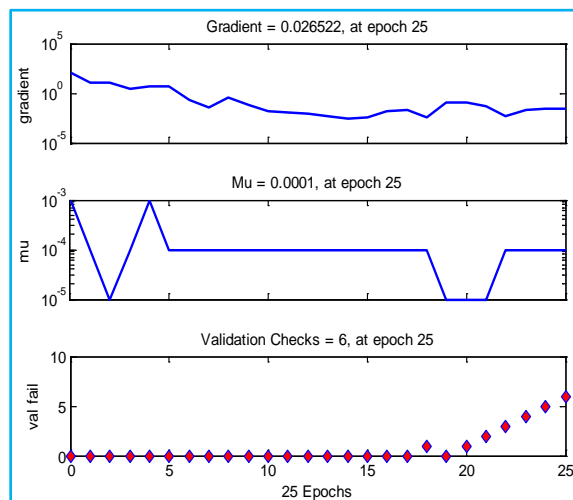
(c) Error Autocorrelation



(d) Series Responses



(e) Network Performances



(f) Test results

Figs. S9(a-f). Neural network results for Y (data smoothed by Holt-Winter smoothing)

TABLE S11
Estimation of the SARFIMA model

Type	Coefficient	SE Coefficient	T	P
AR (1)	0.256684	0.050029	5.130585	0.0000
AR (2)	0.976862	0.014957	65.31228	0.0000
AR (3)	-0.252867	0.041311	-6.121008	0.0000
MA (1)	-0.944757	0.026480	-35.67834	0.0000
MA (2)	-0.999990	0.021714	-46.05287	0.0000
MA (3)	0.944747	0.034595	27.30856	0.0000
SAR (12)	0.999802	0.000134	7433.475	0.0000
SMA (12)	-0.881317	0.021953	-40.14581	0.0000
Constant	0.001384	0.002655	0.521264	0.6023

S.3. Wind-Speed

In Table S7, we tabulated the estimated results of a SARFIMA model of the wind-speed weather parameter. From the table, we observed that data were generated by a long-period time series with fractional parameter $d = 0.4942$ and the data were stationary because $d < 0.5$. Next, in Table S8, we tabulated the parameter estimates of the SARFIMA model. Moreover, the AIC and BIC were observed 1.7948 and 1.8372, respectively. It is observed from the table that the wind-speed data can be modeled by SARFIAM (2, 0.4942, 1) (1, 0, 1) with $s = 12$.

In Figs. S7-S9, we have depicted the results of ANN modeling for the original series, smoothed by moving average and Holt-Winter methods. It is observed that the performance of the best validation occurred at 7th epoch for the wind speed original data while at 7th epoch for the moving average and 19th epoch for the Holt-Winter exponential smoothing. A high correlation coefficient between the estimated and the target observations for the transformed wind-speed data sets is also noticed. The back propagation algorithm showed a very high-accuracy level with the target values for the smoothed data by moving average smoothing.

Next, we fitted the time series multiple linear regression (MLR) model to predict the monthly wind-speed as the response variable, while the other monthly variables such as rainfall, maximum and minimum temperature, relative humidity (RH) recorded at 8am and at 5pm, respectively, are considered as the predictors. The

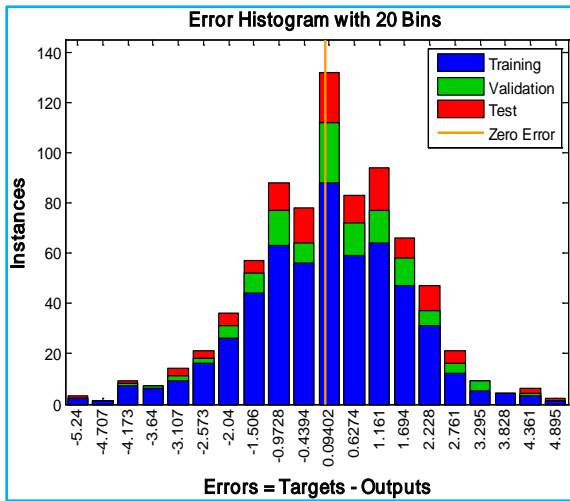
stepwise regression analysis was used for selecting the variables. The estimated time series regression model is given below:

$$Y_t = 3.2472 + 0.00086X_{1t} - 0.0736X_{2t} + 0.1190X_{3t} - 0.0217X_{4t} \quad (3)$$

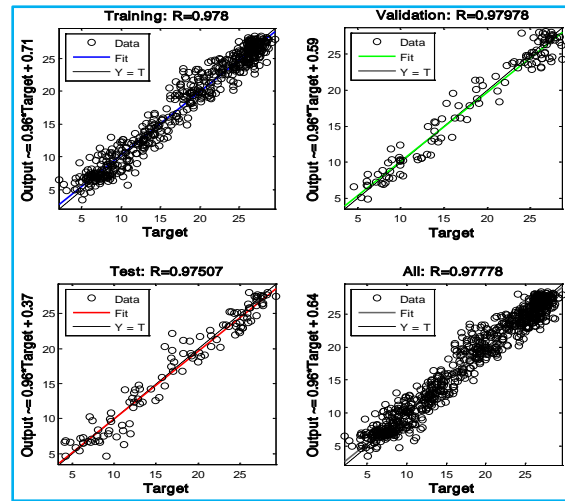
where, rainfall (X_1), maximum temperature (X_2), minimum temperature (X_3) and relative humidity at 8am (X_4). Note that the coefficient of determination was 50.42% for the above model and the most contributing variable for predicting the wind speed (Y) was the minimum temperature. The humidity at 5pm (X_5) seemed to have the minimum effect on the wind-speed, because p-value was observed greater than 5% and hence did not appear in the above model. In Table S9, we listed a comparison of different models and observed that the moving average process had the minimum RMSE. Thus, the moving average smoothed data has the highest correlation coefficient using the ANN algorithm. However, the performance of MLR and SARFIMA models was observed very poor as compared to the ANN models. Thus, we conclude that the ANN method is the most efficient for producing accurate predictions as compared to the ARIMA and time series multiple regression models.

S.4. Minimum Temperature (°C)

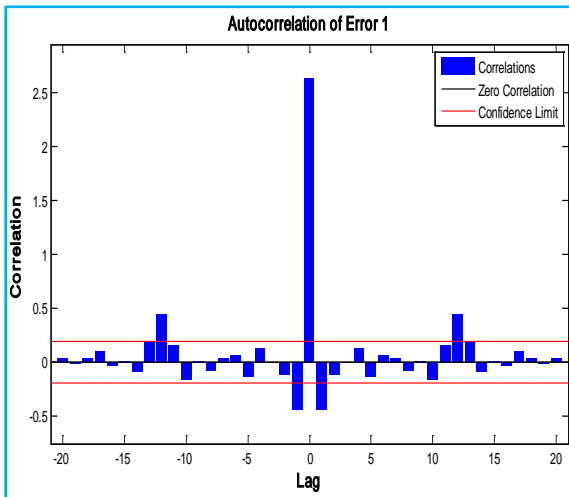
For minimum temperature, we first estimated the SARFIMA model and results are listed in Table S10.



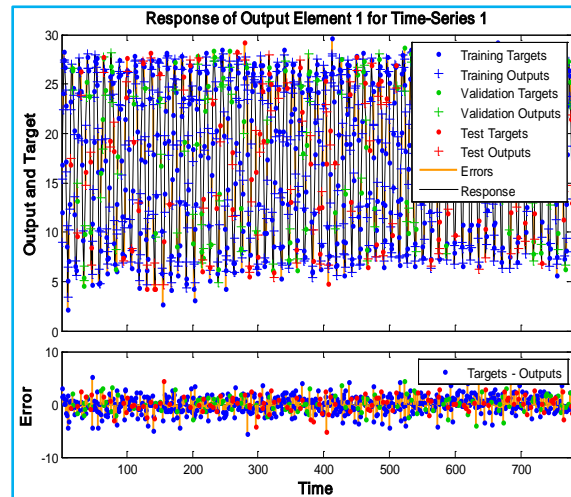
(a) Error Histogram



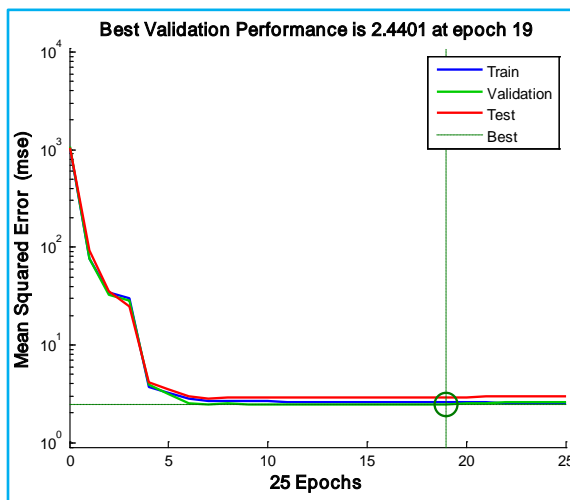
(b) Regression results



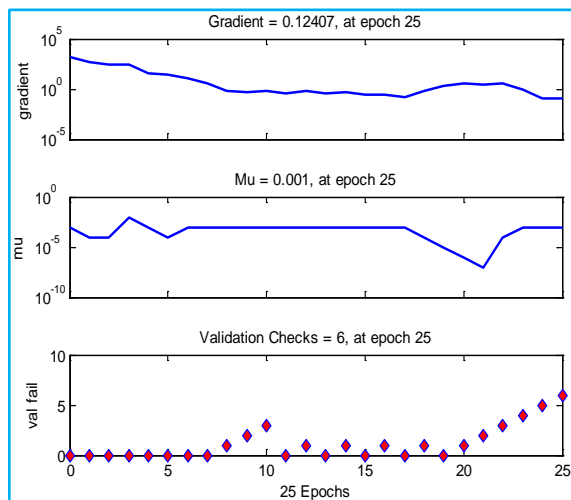
(c) Error Autocorrelation



(d) Series Responses

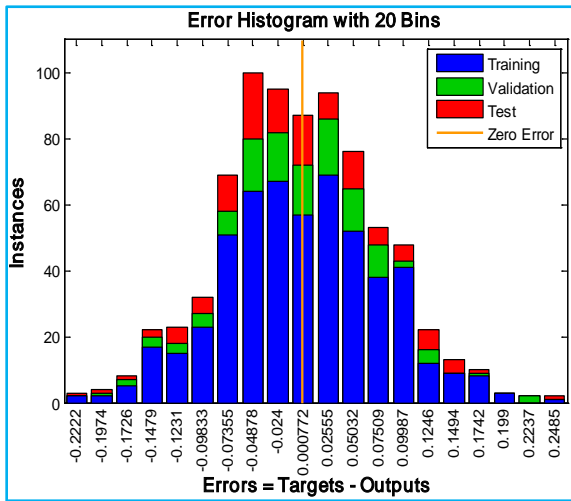


(e) Network Performances

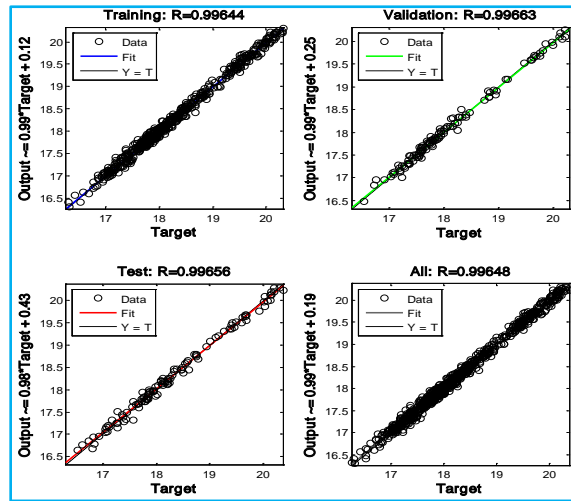


(f) Test results

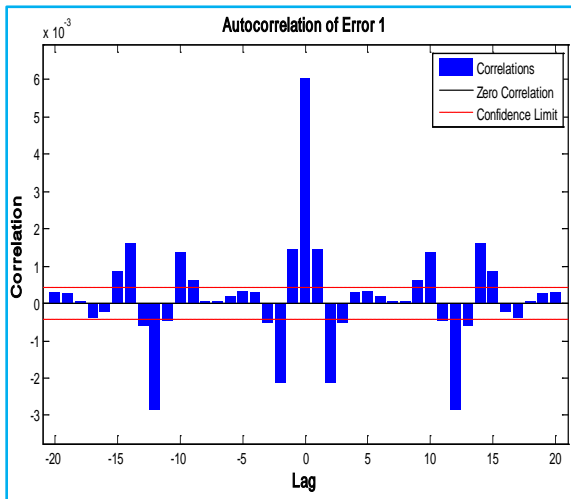
Fig. S10(a-f). Neural network results for Y (Minimum Temperature - original data)



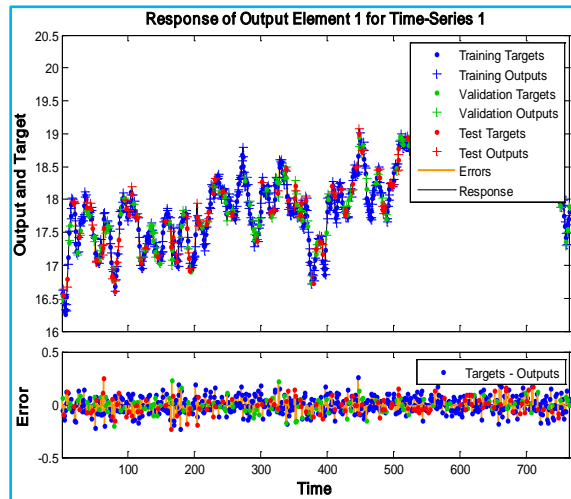
(a) Error Histogram



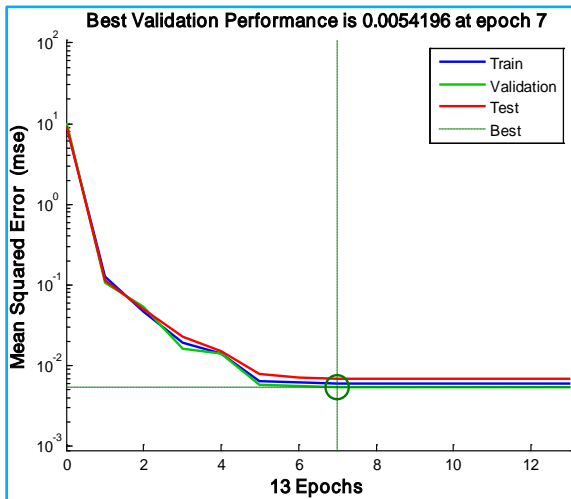
(b) Regression results



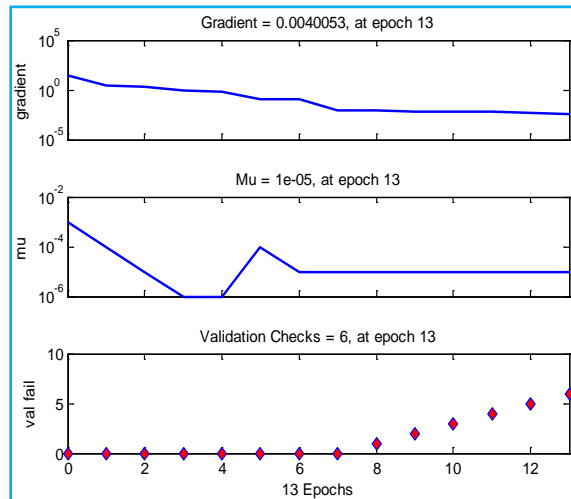
(c) Error Autocorrelation



(d) Series Responses

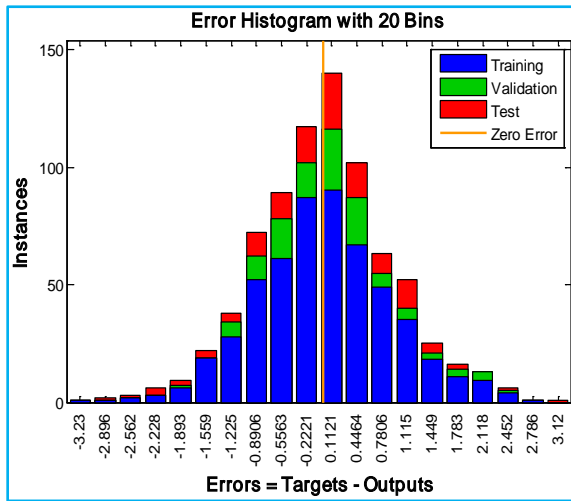


(e) Network Performances

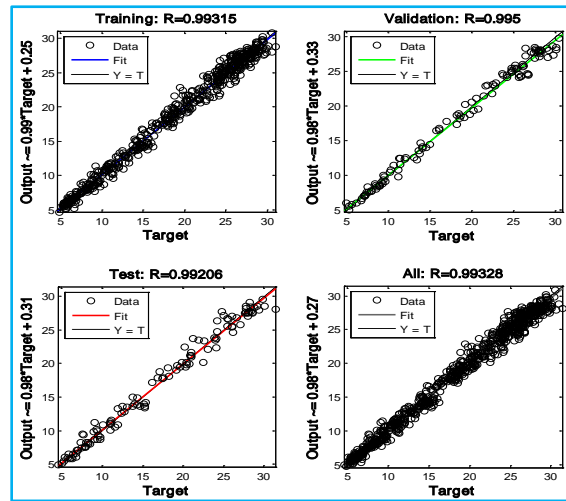


(f) Test results

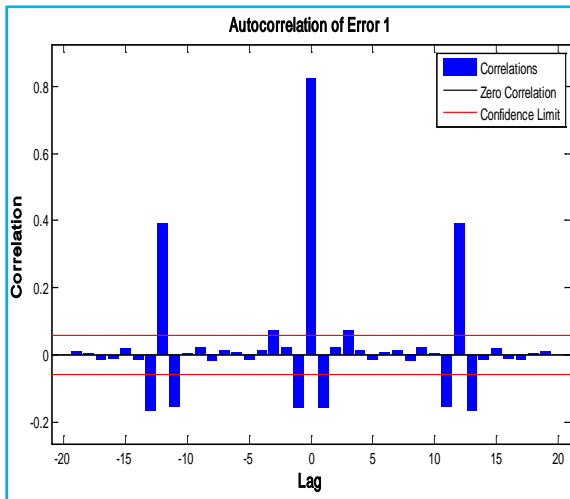
Fig. (a-f) S11. Neural network results for Y (data smoothed data by moving average)



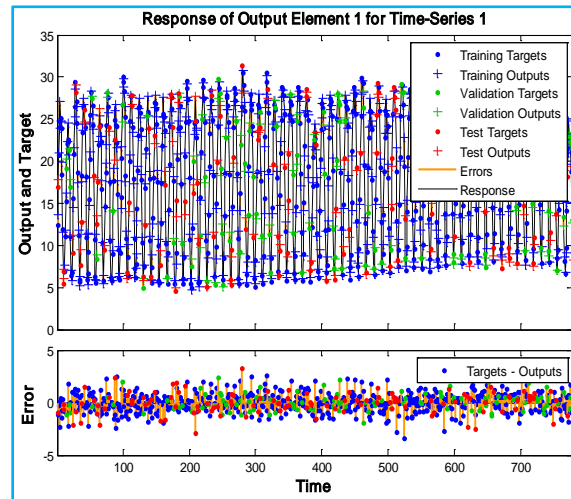
(a) Error Histogram



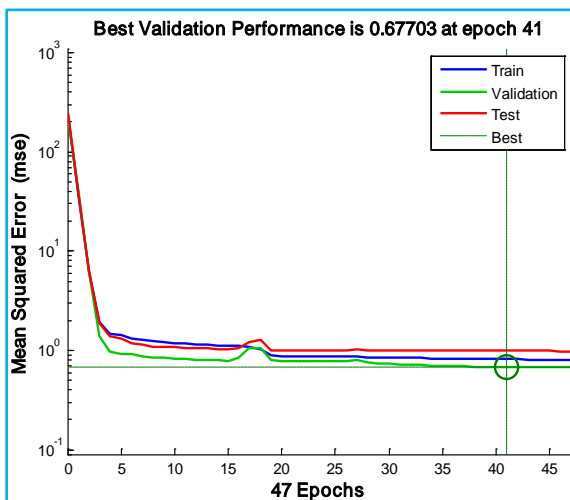
(b) Regression results



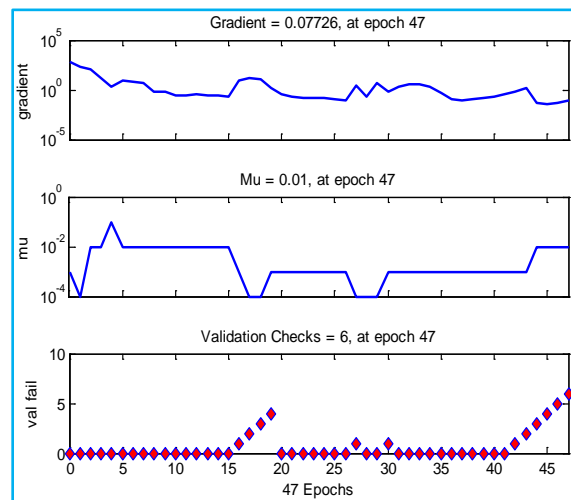
(c) Error Autocorrelation



(d) Series Responses



(e) Network Performances



(f) Test results

Fig. (a-f) S12. Neural network results for Y (data smoothed by Holt-Winter smoothing)

TABLE S12

Forecasting performance of the minimum temperature models

Time series	RMSE	Correlation coefficient
Y	1.5621	0.9798
YMA	0.0736	0.9966
YHWE	0.8228	0.9950
Y (SARFIMA)	7.7079	0.9653
Time Series Regression	1.7249	0.9749

From the table, we observed that data were generated by a long-period time series with fractional parameter $d = 0.4988$. Moreover, the data were observed stationary as $d < 0.5$.

In Table S11, we tabulated the parameter estimates of the ARFIMA model. Moreover, the AIC and BIC were observed 3.1568 and 3.2176, respectively. It is observed from the table that the minimum temperature ($^{\circ}\text{C}$) data can be modeled by SARFIAM (3, 0.4988, 3) (1, 0, 1) with $s = 12$. Next, we implemented the ANN model.

In Figs. S10-12, we have shown the results of ANN modeling for three types of series. The first series is the original data while the second series is the smoothed data by moving average smoother. The third series is the smoothed series by Holt-Winter smoothing. It is observed that the performance of the best validation occurred at 19th epoch for the maximum temperature original data while at 7th epoch for the moving average and 41st epoch for the Holt-Winter exponential pre-processed data. We also observed high correlation coefficient between estimated and target observations for the transformed minimum temperature data set. The back propagation algorithm showed a very high-accuracy level with target values for the smoothed data obtained by moving average. Next, we fitted the time-series MLR model to predict the monthly minimum temperature ($^{\circ}\text{C}$) as the response variable, while the other monthly variables such as rainfall, maximum temperature, relative humidity (RH) at 8 am and 5 pm and

wind speed, are considered as the predictors. The stepwise regression analysis was used for selecting the variables. The estimated time series regression model is given below :

$$Y_t = -20.8226 + 0.0046I_t + 1.0494X_{2t} - 0.0259X_{3t} + 0.1667X_{4t} + 0.7486X_{5t} \quad (4)$$

where rainfall (X_1), maximum temperature (X_2), RH at 8 am (X_3), RH at 5 pm (X_4) and wind speed (X_5). Note that the coefficient of determination was observed 95.04% for the above model and the most contributing variables for predicting the minimum temperature (Y) was the wind-speed. To assess the forecasting performance of different models, we computed RMSE and the correlation coefficient.

In Table S12, a comparison of performance of different models for minimum temperature has been given and we observed again that the moving average process has the minimum RMSE. Further, the moving average smoothed data has the highest correlation coefficient using the ANN algorithm. As the performance of MLR and SARFIMA models is observed very poor as compared to the ANN models, we conclude that the ANN method is the most efficient for producing accurate predictions as compared to the ARIMA and time series multiple regression models.