Asian Research Journal of Mathematics

11(2): 1-10, 2018; Article no.ARJOM.44049
ISSN: 2456-477X

Skill Comparison of Multiple-Linear Regression Model and Artificial Neural Network Model in Seasonal Rainfall Prediction-North East Nigeria

Ebiendele Ebosele Peter[1*] and Ebiendele Eromosele Precious[2]

[1]Department of Basic Science, Auchi Polytechnic, Auchi, Nigeria.
[2]Department of Meteorology and Climate Science, Federal University of Technology, Akure, Nigeria.

Authors' contributions

This study was carried out in collaboration between the authors. Author Ebiendele Ebosele Peter designed the study, performed the statistical analysis, wrote the protocol and wrote the first draft of the manuscript. Author Ebiendele Eromosele precious managed the analyses and the literature searches of the study. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/ARJOM/2018/44049
Editor(s):
(1) Dr. Raducanu Razvan, Assistant Professor, Department of Applied Mathematics, Al. I. Cuza University, Romania.
(2) Dr. Nikolaos Dimitriou Bagis, Department of Informatics and Mathematics, Aristotelian University of Thessaloniki, Greece.
Reviewers:
(1) Isaac Mugume, Makerere University, Uganda.
(2) Nurul Nadrah Aqilah bt Tukimat, Universiti Malaysia Pahang, Malaysia.
(3) Nitin Mishra, Graphic Era University, India.
Complete Peer review History: http://www.sciencedomain.org/review-history/27073

Original Research Article

Received: 21 August 2018
Accepted: 03 October 2018
Published: 06 November 2018

Abstract

Seasonal Rainfall in Nigeria is usually complex due to the tremendous range of variation over a wide range of scales both in space and time. Forecasting techniques such as Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) have been used to study rainfall. This study is motivated by the need to compare MLR and ANN models to know which one is more reliable in predicting Seasonal rainfall. The rainfall datasets used in this study were collected from the achieving of Nigeria meteorological agency from (1986-2017). The model comparison was based on four criteria; the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), prediction error and Correlation coefficient (r). The error measures are comparable for the two models. The analysis of the models performance shows that, overall, the ANN model performs better than the MLR model in terms of PE, RMSE, MAE and correlation coefficient. ANN had the minimum MAE=56.66 mm, RMSE=74.84 and PE=0.11209 respectively and in terms of correlation coefficient between the observed rainfall and predicted rainfall amount, ANN model had high correlation coefficient (0.93) compared to MLR model whose correlation coefficient was (0.66).

Corresponding author: E-mail: peter.ebiendele@yahoo.com;

Keywords: Prediction; multiple regression; artificial neural network; seasonal rainfall.

1 Introduction

Rainfall is one of the most complex and difficult weather phenomena to model, due to its enormous range of variation over an extensive range of scales, both in space and time [1]. The complexity of the atmospheric processes that generate rainfall makes quantitative forecasting of rainfall an extremely difficult task [2]. Rainfall events depend on some precursors from other parameters such as the sea surface temperature for monthly to seasonal time scales, the surface pressure for weekly to daily time scale and other atmospheric parameters for daily to hourly time scale. The latter atmospheric parameters could be potential equivalent temperature, air temperature, relative humidity, specific humidity, zonal wind speed at a different pressure level and sunshine duration. Variability of these atmospheric parameters influences seasonal rainfall events in Nigeria. Fundamentally, two approaches are used for predicting rainfall. One is Empirical approach and the other is Dynamical approach. The empirical approach is based on analysis of historical data of the rainfall and its relationship to a variety of atmospheric and oceanic variables over different parts of the world. The most widely used empirical approaches, which are used for rainfall prediction, are regression, artificial neural network and fuzzy logic. On the other hand in dynamical approach, predictions are generated by physical models based on systems of equations that predict the evolution of the global climate system in response to initial atmospheric conditions. The Dynamical approach is implemented by using numerical rainfall forecasting method [3]. In statistical analysis, multiple linear regression models are often used for estimating future events or values based on the previous values and events but could not address the challenge posed by the non-linear nature of rainfall. Thus, the advent of digital computer neural simulation has made data-driven techniques substitute forecasting tools in time series, which is useful for seasonal rainfall prediction. The use of nonlinear models such as Artificial Neural Network (ANN), are capable of modeling complex nonlinear problems and can be suitable for real-world temporal data such as seasonal rainfall data in Nigeria.

Therefore, this paper attempt to find out what model is suitable, with a high level of performance in predicting seasonal rainfall.

An ANN model was employed using radar, satellite and weather-station data together with numerical products generated by the Japan Meteorological Agency (JMA) Asian Spectral Model and the model was trained using 1-year data. The Author found that the ANN skills were better than the persistence forecast (after 3 h), the linear regression forecasts, and the numerical model precipitation prediction. As the ANN model was trained with only 1 year data, the results were limited. The author concluded that the performance of the neural network would be improved when more training data are introduced. It is still unclear to what extent each predictor contributed to the forecast and to what extent recent observations might improve the forecast [4].

An attempt to predict rainfall in São Paulo using observed daily rainfall, the author [5], found that ANNs outperformed MLR, which showed a high bias for days without rain. The author [6], in forecasting rainfall in Brazil found that the ANN method tended to predict moderate rainfall with greater accuracy during austral summer. In his research, potential temperature, vertical component of the wind, specific humidity, air temperature, perceptible water, relative vortices and moisture divergence flux were used as input data for training of networks. Results of ANN were superior to the ones obtained by the linear regression model, weighted average, harmonic average and other linear statistical methods which is revealing a great potential for suitable performance. David and Marengo [7] reported that the daily rainfall in the Amazon Basin is better represented by ANNs than autocorrelation models. Rainfall forecasting in South-Western Nigeria was attempted by Agboola [8]. The author developed an Artificial Neural Network Model for which performance evaluation of the model was done by calculating Prediction Error, Root Mean Square Error; Mean Absolute Error, and Prediction Accuracy to know how efficient the model was. The author concluded that as the PE, RMSE, MAE values on data were comparatively less, the ANN prediction model is reliable and efficient and can be used for rainfall prediction. ANNs have been reported [9,10] to provide best performance among the data-driven models despite considerable influence by the problems of over fitting and memorization when subjected to small amounts of datasets.

2 Materials and Methods

2.1 Data

The study adopted total monthly rainfall from June to October as the dependent variable and eight (8) independent variables/predictors of monthly means of Sea Surface Temperature (°C), Air Temperature (°C), Specific Humidity, Relative Humidity (%) and Uwind (m/s) at surface different pressure levels, 750hpa, 800hpa, 1000hpa) from January to May for a period of 32 years (1986-2017) over Bauchi. Data were obtained from the archives of Nigerian Meteorological Agency (NIMET).

2.3 Choice of predictors

As noted in [11], and the new method (probably, most common) of predicting rainfall characteristics as in the tropics relating to this study is the use of sea surface temperature (SST) see [12,13,14,15,16,17]. Studies have established correlation coefficients of 0.70 [13,14] using SST as a predictor. The fact is that SST has become a more promising parameter for the prediction of rainfall in West Africa. Other synoptic predictors suggested for this study (air temperature, specific humidity, relative humidity and Wind speed) were chosen to improve on the model results of previous studies. See [14,15,16].

2.4 Mathematical formulation of multiple linear regression model (MLR)

Multiple regression models describe how a single response variable Y(seasonal rainfall) depends linearly on a number of predictor variables. A multiple linear regression model with k predictors' variables $X_1, X_2, \ldots\ X_k$ can be represented by the following equation

$$Y = a + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k + C \ \rightarrow \tag{1}$$

2.5 Artificial neural network

The Artificial Neural Network (ANN) is an engineering concept of knowledge in the field of artificial intelligence designed by adopting the human nervous system. Wherein, the main processing of the human nervous system is composed of the brain nerve cells as the basic unit of information processing. In the concept of ANN, the basic units of information processing are called neurons which serve process information in parallel and immediately. Furthermore, the processes of training the ANN have many types and uses, including Back propagation (BPNN), Resilient back propagation (RBPNN) with or without weight backtracking or the Modified globally convergent type, Self-Organizing Map (SOM), and Delta [18]. Therefore, this study proposes BPNN algorithm to predict seasonal rainfall by analyzing the patterns of non-linear nature of seasonal rainfall with some several inputs/predictors.

2.6 Experimental set-up of back propagation neural network model (BPNN)

The Back Propagation Neural network (BPNN) is an ANN supervised learning method. The BPNN was first introduced by Paul Werbos in 1974, and then popularised by Rumelhart and McCelland in 1986. In general, the BPNN works by forwarding the output layer to the input layer in changing the weights [19]. Furthermore, ANN works like humans in which learning is performed through examples or cases and exercises in each layer of ANN. The layer in BPNN consists of three parts, namely input layer, hidden layer and output layer, Fig. 1.

In general, the steps to build the BPNN algorithm for the prediction of rainfall data were adopted from previous research (Haviluddin & Alfred. R.) which is outlined in [19]:

The model building process consists of four sequential steps:

1. Selection of the input and the output data for the supervised learning algorithm.
2. Normalization of the input and the output data is done using an asymptotic function called sigmoid function in order to get the value of both input and output data in smaller intervals which are [0,1] using the following equation;

$$X' = \frac{(X-a)}{(b-a)} \rightarrow \tag{2}$$

Where a is the minimum value, b is the maximum, X' is the data that have been transformed.

3. Constructing a BPNN design to determine the amount of data input, hidden, and output layers and parameters to be used.
4. Training of the normalized data from (1986 to 2013).
5. Testing the goodness of fit of the model using a set of test data (2014 to 2017), different from those employed in training the model, were used to assess the level of skill that the model is likely to achieve in real time prediction.
6. Comparing the predicted output with the desired output.

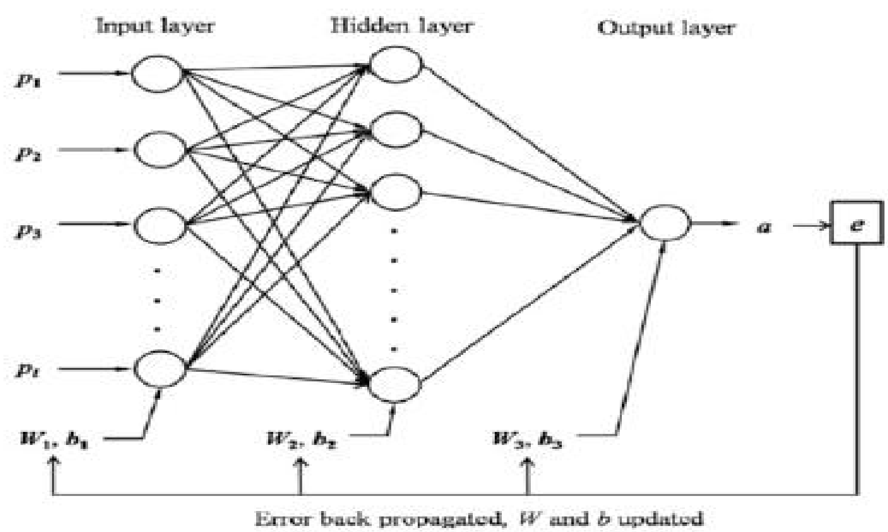

Fig. 1. Mathematical model of back propagation neural network (BPNN)

2.7 Model comparison

In this paper, a comparative analysis of the Multiple Linear Regression Model (MLR) and Artificial Neural Network model (ANN) was carried out using an open-source software called R-studio, the ANN model was trained using the neuralnet package in Rstudio. The model comparison will be carried out using the following criteria:

(a) Mean Absolute Error; given as

$$MSE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}\mathrm{i}| \tag{3}$$

Where $y_i = Observed\ Rainfall(mm),$
$\hat{y}_i = Predicted\ Rainfall(\ mm)$
$n = number\ of\ events$

(b) Root Mean Square Error (RMSE): is a good measure of prediction accuracy. It is frequently used to measure the differences between values predicted by a model and the values observed from the thing being modeled. These individual differences are also called residuals. The RMSE is given as;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(yi-\hat{y}i)2}{n}} \tag{4}$$

Where $y_i = Observed\ Rainfall(mm),$
$\hat{y}_i = Predicted\ Rainfall(mm)$
$n = number\ of\ events$

(c) Prediction Error (PE); given as

$$\frac{(|y_{\ Predicted} - y_{\ Observed}|)}{(y_{\ Observed})} \tag{5}$$

The predictive model is identified as a good one if the PE is sufficiently small i.e. close to 0

(d) Correlation Coefficient (r)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}} \tag{6}$$

Where $x = Observed\ or\ actual\ rainfall,\ \ y = Predicted\ seasonal\ rainfall$

Equations 3 to 6 have been widely used to measured forecast errors, see [20,9,21,22].

3 Results and Discussion

3.1 Overview of seasonal rainfall trend

Seasonal rainfall in Nigeria is usually total rainfall amount from June to October (JJASO). Fig. 2 shows the seasonal rainfall variation from June to October 1986-2017 in Bauchi.

3.2 Multiple linear regression model (MLR)

The MLR model equation is presented below, the model equation was deployed to predict the test data (June to October 2014-2017).

Multiple Linear Regression Model (MLR)

$$Y = -1143.96 - 20.33(X_1) + 459.76(X_2) - 218.61(X_3) - 31.23(X_4) + 34.97(X_5)$$
$$-8161.15(X_6) - 3..49(X_7) + 84.36(X_8) \quad \rightarrow (7)$$

Where

Y =Predicted Seasonal Rainfall (mm) (June to October)

X_1 = January to May Sea Surface Temperature (°C)
X_2 = January to May Uwind at Surface (m/s)
X_3 = January to May Uwind at 1000 mb Pressure level (m/s)
X_4 = January to May Uwind at 800 mb Pressure level (m/s)
X_5 = January to May Uwind at 750 mb Pressure level (m/s)
X_6 = January to May Specific Humidity
X_7 = January to May Relative Humidity (%)
X_8 = January to May Air Temperature (°C)

From this equation we can calculate Seasonal Rainfall which is June to October total for future years knowing the sea surface temperature, uwind at surface, uwind at 100mb,800mb and 750mb pressure level, specific humidity, relative humidity and air temperature from January to May.

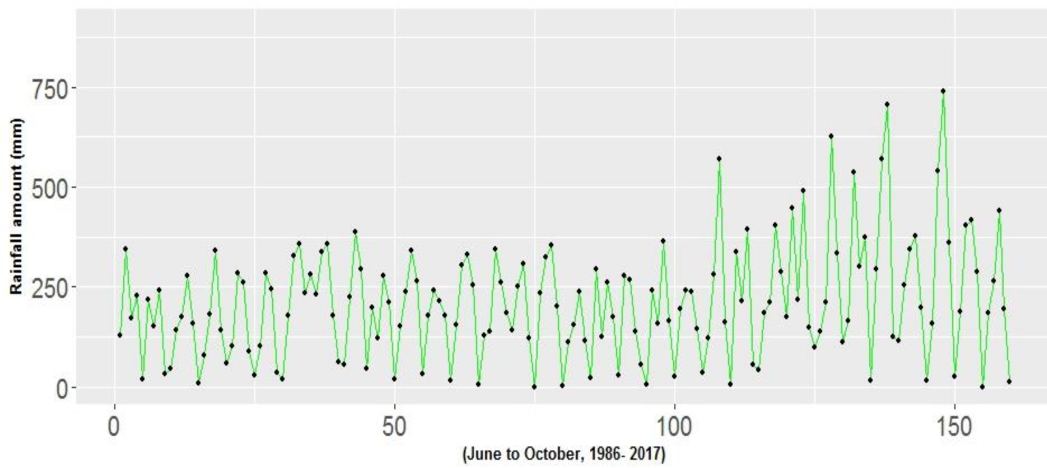

Fig. 2. Seasonal rainfall amount from June to October 1986-2017

3.3 Artificial neural network model

In construction of the Artificial neural network architecture, seasonal rainfall amount during the rainfall season of June to October in Bauchi was used as dependent variable with 8 parameters as inputs. The parameters used was, air Temperature, SST, U-wind at surface, 750hpa, 800hpa, 1000hpa, relative humidity, and specific humidity from January to May (JFMAM), apart from the input variables, the number of hidden layers also determines the performance of the model as shown in (Fig. 3). The data was divided into a training data set (1986-2013) and test data test (2014-2017) for cross-validation/model performance. Both input and dependent data must be normalised to avoiding over-fitting of model results. Normalization of data is done mathematically below;

$$x' = \frac{(x-a)}{(b-a)} \rightarrow \quad (8)$$

$where\ a = minimum\ value\ and\ b = maximum\ value$

In Fig. 3, the four hidden layers are responsible for mapping a nonlinear relationship between the eight inputs and output (rainfall amount). The significance of these values between the eight input parameters and the four hidden layer, accounts for capturing nonlinear and complex underlying characteristics of rainfall amount with a high degree of accuracy. However, this computation cannot deal with uncertainties.

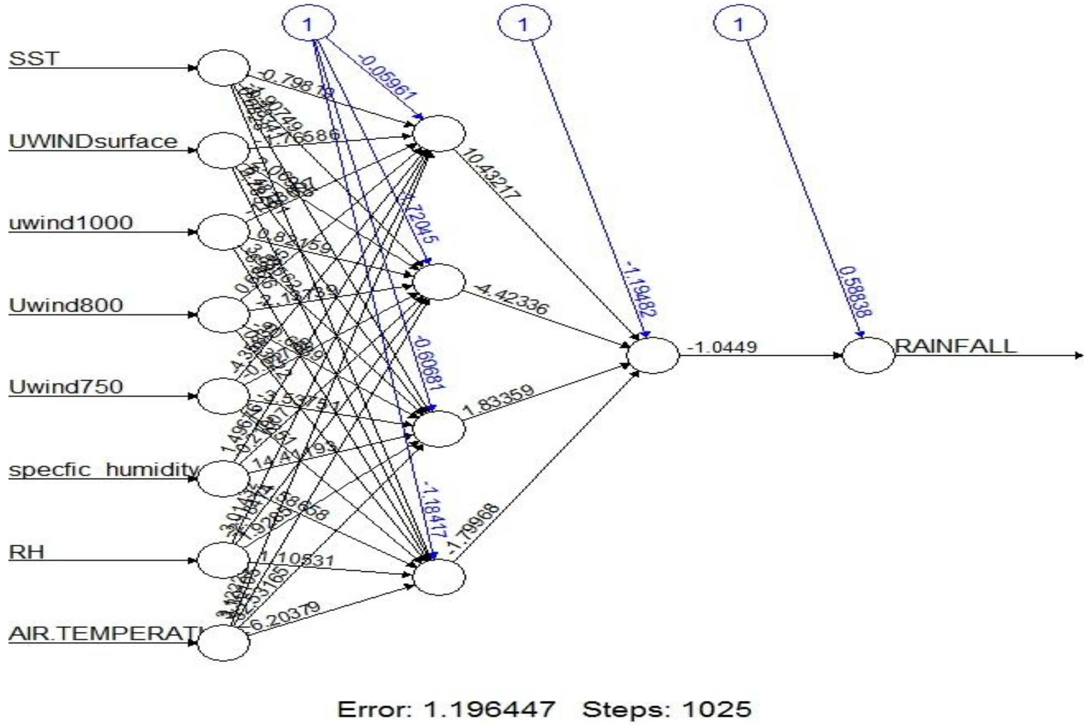

Fig. 3. BPNN Model architecture structure (8-4-1)

3.4 Comparison of MLR and ANN model results

The graphical comparison between observed (Red Bar) and predicted value (Blue Bar) of Rainfall for both MLR and ANN model results (June to October 2014-2017) is represented in Fig. 4.

3.5 Statistical performance of ANN and MLR models

The performance of both models (ANN and MLR) are based on some error measures, which was presented in Table 1, ANN had the minimum MAE=56.66mm, RMSE=74.84 and PE=0.11209 respectively. In Fig. 5, in terms of correlation coefficient between the desired observed rainfall and predicted rainfall amount, ANN model had high correlation coefficient (0.93) compared to MLR model whose correlation coefficient was low (0.66).

The analysis of the model accuracy, shows that overall, the artificial neural network model (ANN) outperformed the multiple linear regression model (MLR) in terms of MAE, RMSE, PE and correlation coefficient (r) in prediction of seasonal rainfall for the test data (June to October 2014-2017).

Table 1. Statistical comparison of ANN and MLR model performance

Model	MAE (Mean Absolute error)	RMSE (Root mean square error)	PE (Prediction error)	R (correlation coefficient)
ANN	56.66	74.84	0.11209	0.93
MLR	1067.5	144.4	1.60564	0.66

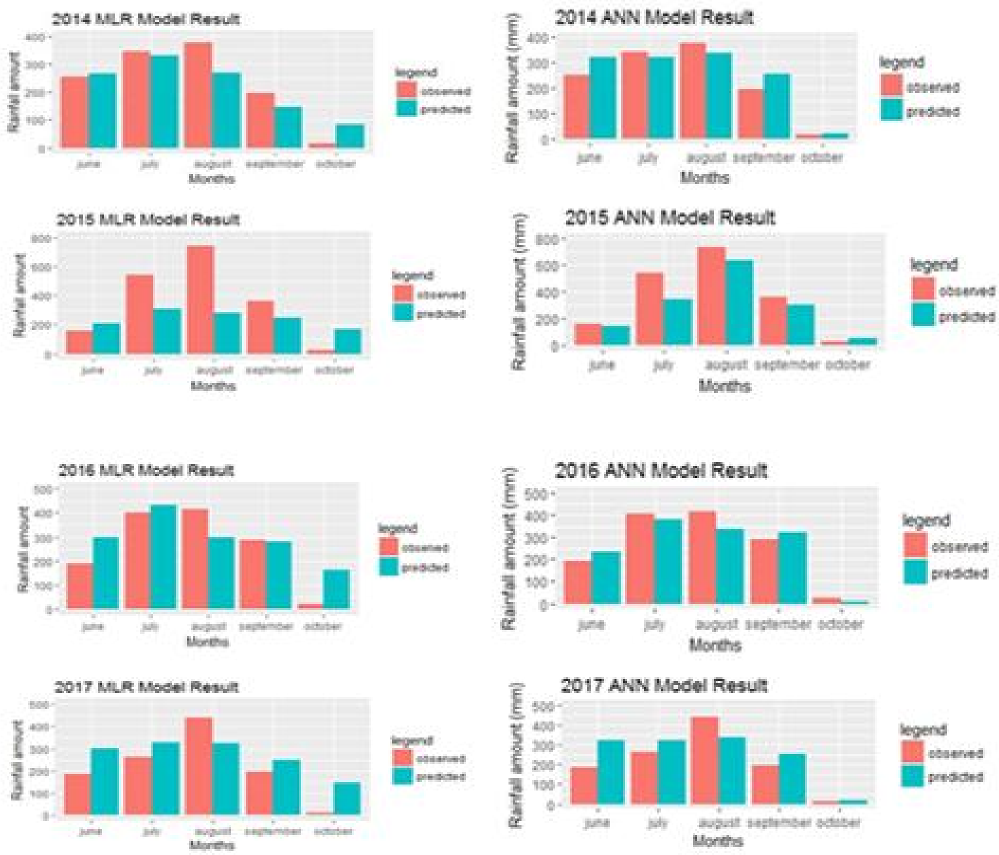

Fig. 4. Comparison of both MLR and ANN model results for the test data (2014 to 2017)

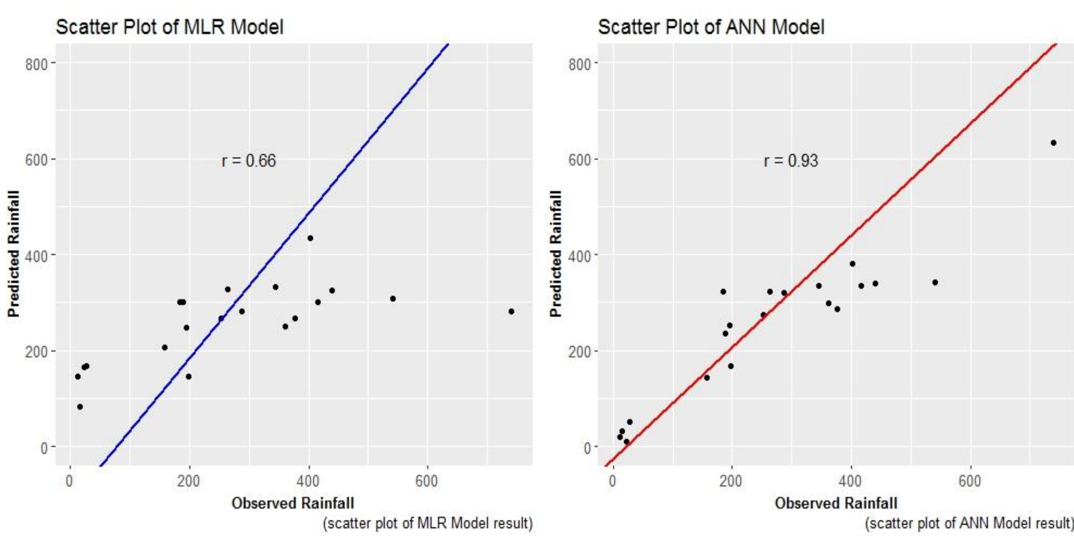

Fig. 5. Scatter plot of observed and predicted MLR and ANN model output

4 Conclusion and Recommendations

This study has contributed to knowledge by establishing the best and reliable model for predicting seasonal rainfall in Bauchi, Nigeria. In this study, Artificial Neural Network (ANN) and Multiple Linear Regression models were developed for rainfall prediction. The performance analysis of the two models is done using mean absolute error, root mean square error, prediction error and correlation coefficient. The prediction skill of artificial neural network was 86.49% and that of multiple linear regression was 43.56%. The results show that the Artificial Neural Network model was better than the Multiple Linear Regression in prediction of rainfall in Bauchi.

The result of this study has shown that ANN model is efficient and reliable in modeling complex non-linear nature of rainfall in Bauchi and has the ability of accurately predicting seasonal rainfall. The challenge of sparse meteorological data has been an issue that reduces the representativeness of a system and can, therefore, have a significant effect on the conclusion that can be drawn. This should be looked into by related institutions as it will bring a lasting solution to errors in the analysis. Also, for further studies, more predictors should be introduced.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] French MN, Krajewski WF, Cuykendall RR. Rainfall forecasting in space and time using a neural network. Journal of Hydrology. 1992;137:1-31.

[2] Hung NQ, Babel MS, Weesakul S, Tripathi NK. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Journal of Hydrology and Earth System Sciences. 2008;13:1413-1425.

[3] Ismail Z, et al. Forecasting gold prices using multiple linear regression method. American Journal of Applied Sciences. 2009;6(8):1509-1514.

[4] Koizumi K. An objective method to modify numerical model forecasts with newly given weather data using an artificial neural network. Weather Forecast. 1999;14:109–18.

[5] Ramírez, de Campos Velho HF, Ferreira NJ. Artificial neural network technique for rainfall forecasting applied to the São Paulo region. J. Hydrol. 2005;301:146–162.
DOI: 10.1016/j.jhydrol.2004.06.028, 2005

[6] Ramírez MC, Ferreira NJ, Velho HFC. Linear and nonlinear statistical downscaling for rainfall forecasting over southeastern Brazil. Weather Forecast. 2006;21:969–989.
DOI: 10.1175/WAF981.1, 2006

[7] David Mendes, Jose A. Marengo. Temporal downscaling: A comparison between artificial neural network and autocorrelation techniques over the Amazon Basin in present and future climate change scenarios. Journal of Theoretical Applied Climatology. 2010;100:413-421.

[8] Agboola A. H. Iyare. An artificial neural network model for rainfall forecasting in South-Western Nigeria. Canadian Journal on Computing in Mathematics, Natural Sciences, Engineering and Medicine. 2012;3(6):188-196.

[9] Barua S, Perera B, Ng A, Tran D. Drought forecasting using an aggregated drought index and artificial neural network. Journal of Water and Climate Change. 2010;1(3):193-206.

[10] Dorofki M, Elshafie AH, Jaafar O, Karim OA, Mastura S. Comparison of artificial neural network transfer functions abilities to simulate extreme runoff data. International Proceedings of Chemical, Biological and Environmental Engineering. 2012;33:39-44.

[11] Omotosho JB, Balogun AA, Ogunjobi K. Predicting monthly and seasonal rainfall onset and cessation of the rainy season in West Africa using only surface data. Int J Climatol. 2000;20:865–880.

[12] Chiang JCH, Kushnir Y, Giannin A. Deconstructing Atlantic intertropical convergence zone variability: influence of the local cross – equatorial sea surface temperature gradient and remote forcing from the Eastern Equatorial Pacific. J Geophy Res. 107(D1): 1029=2000 JD000307. 2002;3-1–3-19.

[13] Colman AW, Davey MK, Graham R, Clark R. Experimental forecast of 2000 seasonal rainfall in the Sahel and other regions of tropical North Africa, using dynamical and statistical methods. Experimental Long-Lead Forecast Bulletin; 2000.

[14] Folland C, Owen J, Ward MN, Colman A. Prediction of seasonal rainfall in the Sahel region, using empirical and dynamic methods. J Forecasting. 1991;10:21–56.

[15] Fontaine G, Janicot S. Sea surface temperature fields associated with West Africa rainfall anomaly types. J Climate. 1996;9:2935–2940.

[16] Janicot S, Harzallah A, Fontaine B, Moron V. West African monsoon dynamics and the eastern equatorial Atlantic and Pacific SST anomalies (1970–80). Journal of Climate. 1998;8:1874–1882.

[17] Odekunle TO, Balogun EE. On the prediction of rainfall onset and retreat dates in Nigeria. Theor. Appl. Climatology. 2005;81:101-112.

[18] Shrivastava G, Karmakar S, Kowar MK. Application of artificial neural networks in weather forecasting: A comprehensive literature review. International Journal of Computer Applications. 2012;51(18):17-29.

[19] Haviluddin, Alfred R. Daily network traffic prediction based on backpropagation neural network. Australian Journal of Basic and Applied Sciences. 2014;8(24 Special):164-169.

[20] Antonopoulos VZ, Papamichail DM, Mitsiou KA. Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. Hydrology and Earth System Sciences Discussions. 2001;5(4):679-692.

[21] Ogaji S, Singh R. Artificial neural networks in fault diagnosis: A gas turbine scenario. Computational Intelligence in Fault Diagnosis, Springer. 2006;179-207.

[22] Wonham WM. Linear multivariable control, optimal control theory and its applications. Springer. 1974;392-424.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sciencedomain.org/review-history/27073