International Journal of Statistics and Probability; Vol. 11, No. 4; July 2022
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Modeling Average Rainfall in Nigeria With Artificial Neural Network (ANN) Models and Seasonal Autoregressive Integrated Moving Average (SARIMA) Models

Ikpang Nkereuwem Ikpang[1], Ekom-obong Jackson Okon[2] & Ekemini Udoudo George [1]

1 Akwa Ibom State University, Ikot Akpaden, Mkpat Enin, Nigeria

2 University of Benin, Benin, Nigeria

Correspondence: Ikpang Nkereuwem Ikpang, Department of Statistics, Akwa Ibom State University, Ikot Akpaden, Nigeria

Received: May 18, 2022 Accepted: June 29, 2022 Online Published: June 30, 2022

doi:10.5539/ijsp.v11n4p53 URL: https://doi.org/10.5539/ijsp.v11n4p53

Abstract

Rainfall prediction is one of the most essential and challenging operational obligations undertaken by meteorological services globally. In this article we conduct a comparative study between the ANN models and the traditional SARIMA models to show the most suitable model for predicting rainfall in Nigeria. Average monthly rainfall data in Nigerian for the period Jan. 1991 to Dec.2020 were considered. The ACF and PACF plots clearly identify the SARIMA $(1,0,2)x(1,1,2)_{12}$ as an appropriate model for predicting average monthly rainfall. The performance of the trained Neural Network (NN) analysis clearly favours Levenberg-Marquardt (LM) over the Scaled Conjugate Gradient Descent (SCGD) algorithms and Bayesian Regularization (BR) method with Average Absolute Error 0.000525056. The forecasting performance metric using the RSME and MAE, showed that Neural Network trained by Levenberg-Marquadrt algorithm gives better predicted values of Nigerian rainfall than the SARIMA $(1,0,2)x(1,1,2)_{12}$.

Keywords: SARIMA, Rainfall, Neural Network, Forecasting

1. Introduction

Rainfall is a climatic parameter whose prediction is challenging and demanding as the world continues to experience climate change. It affects every component of the ecological system including fauna and flora. Therefore, rainfall investigation is vital and cannot be over accentuated. Climate events such as flood have been increasing recently all over the world and this trend has been attributed to climate change and global warming. One indicator of climate change is rainfall (Novotny & Stefan, 2007).

In Nigeria, rainfall has profound impact on agriculture, air and land transportation, hydroelectric power generation construction and water resources. Hence, normal rainfall is beneficial for agriculture and other economic activities. However, when it is excessive or above normal, it may result in flooding and its associated negative impacts (Obot & Onyekwu, 2010).

The complexity of the atmospheric processes that generate rainfall makes quantitative forecasting of rainfall an extremely and difficult task. However, water quality management and agriculture necessitate rainfall forecast temporally and spatially which contributes significantly in the long run to the economy of a nation but to mention a few (Ayodele & Precious, 2019).

In different studies, traditional methods like the autoregressive (AR), moving average (MA), autoregressive moving average (ARMA) and the autoregressive integrated moving average models (ARIMA) models have been used, as seen in (Okon and Ikpang, 2020). These models are based on assumptions that the seasonal component is deterministic and independent of other non-seasonal components with a belief that empirical knowledge of the data be known. More likely, the seasonal component may be stochastic and correlated with nonseasonal components.

Artificial neural network (ANN), on the other hand is based on the neural structure of the human brain and its complex pattern recognition hence, devoid of making initial assumptions. Data used is allowed to govern the process by itself through generation of input–output mapping for the set of data, however complex. Training the network with relevant data enables the network the ability to make predictions based on any input it encounters.

Globally, researches have been carried out on rainfall in terms of modeling and different models have been explored. In a study by Nwokike, Offorha, Obubu, Ugoala and Ukomah (2020), the use of artificial neural network (ANN) model was adopted to aid in comparing different methods of forecasting rainfall frequency in Umuahia, Abia State, Nigeria. The data used for the research covered the period 2006 to 2016 and was obtained from the National Root Crop Research Institute, Umudike, Abia State. Several forecast performance measures were employed namely, Forecast Error (FE), Mean Forecast Error (MFE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for basis of comparison and the fitted SARIMA model had lower error indicators across the forecast performance measures hence, was adjudged better than SANN model in forecasting rainfall frequency in Umuahia, Abia State. A t -test for significant difference showed that the forecasts obtained from both models were not significantly different and can be successfully used as substitutes. Amaefula (2021) in his study compared the SARIMA and adjusted SARIMA (ASARIMA) in a regular stationary series where the underlying variable is seasonally nonstationary. The Box-Jenkins iterative algorithm was adopted and based on the AIC, the ASARIMA $(2,1,1)_{12}$ was chosen out of the 11 sub-classes of SARIMA and 7 sub-classes of ASARIMA models. Diagnostic test indicates absence of autocorrelation up to the 48th lag while generated forecast values using the chosen model is not significantly different from the actual values. The ASARIMA model can therefore be recommended for regular stationary time series exhibiting seasonal characteristics with penalized redundant parameters and large sum of square errors.

Sanni, Ahmed, Abubakar and Abdullahi (2019), obtained point rainfall field data from 1983 to 2017 of Kano State located in the Chad Basin from Nigeria Meteorology Agency, for prediction using Seasonal Autoregressive Integrated Moving Average (SARIMA) Model in R programming. Exploratory data analysis, time series decomposition, plots of Auto Correlation factor (ACF) and Partial Auto Correlation factor (PACF), fitting the model and diagnostic test were applied to obtain the prediction model. A seasonal cycle of appreciable rainfall that begins in April with an upward trend to August (the threshold month), then decline to March was revealed. The rainfall data has a mean higher than the median, which shows a positively skewed series and a heavy tailed distribution of kurtosis greater than 3. The best SARIMA Model parameters of the order (6,0,2,0,1,3,12) was selected at minimum AIC of 8.707182. The prediction shows insignificant decrease in rainfall amount from 2018 to 2030 with Mean Error (ME) of -8.65, Root Mean Square Error (RMSE) of 28.27 and Mean Absolute Error (MAE) of 18.88.

Several researchers have studied rainfall data using ANN or SARIMA comparatively or independently such as; Sekhar, Poola, Sekhar and Naidu (2020); Dabral and Murry (2017); Mishra, Soni, Sharma and Upadhyay (2018); Soumen and Debasis (2018); Esteves, de Souza Rolim and Ferraudo (2019); Darji, (2019); Ray, Das, Mishra and Al Khatib (2021); Aliyu, Auwal and Adenomon (2021).

In some cases, the ANNs are preferred to the SARIMA models based on efficiency. However, in this article, we evaluate the performance ability of these models in predicting rainfall in Nigeria so that it could be adopted for predictive purposes.

2. Method

The data for average monthly rainfall in Nigeria used for the study was obtained from www.climateknowledgeportal.worldbank.org .The data obtained was for the period January 1991 to December 2020. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model and the Artificial Neural Network (ANN) model were employed to model the average rainfall data. The software used for this analysis is Gretl and Matlab.

2.1 SARIMA Model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model fitted to the average monthly rainfall data is of the form

$$SARIMA(p,d,q)(P,D,Q)_s \tag{1}$$

where,

p and P = The order of non-seasonal and seasonal Autoregressive process respectively.

q and Q= The order of non-seasonal and seasonal Moving Average process respectively.

d and D = Non-seasonal and seasonal difference order.

s = Seasonal period.

Equation 1 can be expressed explicitly as follows;

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_q(B^s)\alpha_t \tag{2}$$

Where X_t is the time series at time t, $\phi_p(B)$ and $\theta_q(B)$ the regular autoregressive and moving average factors (polynomials) and $\Phi_P(B^s)$ and $\Theta_q(B^s)$ the seasonal autoregressive and moving average factors (or polynomials), respectively. α_t is the zero mean white noise process and the seasonal period $s=12$ for monthly data while B is the backward shift operator.

2.2 Artificial Neural Network

Artificial neural networks are made up of layers of input, process and output elements. It uses this information to perform a mathematical formulation in order to produce optimum result for any data set or problem segment.

The procedure for achieving a neural network model is as follows;

a. Configure or transform the series and set network architecture

The mapminmax transformation code that carries out a linear transformation on the data using the formula is employed

$$SV = TF_{\min} + \frac{(TF_{\max} - TF_{\min})(X_0 - X_{\min})}{\left(X_{\max} - X_{\min}\right)} \tag{3}$$

Where,

SV = Scaled value,

TFmin = Minimum value of transformation function

TFmax = Maximum value of transformation function

X0 = Value of the observation,

Xmax = Maximum value of original data set

Xmin = Minimum value of original data set.

Furthermore, in a time series forecasting problem, the number of input nodes corresponds to the number of lagged observations used to discover the underlying pattern in a time series and to make forecast.

b. Initialize all weights and bias and transform the training data

c. Run the network forward with the input data to get network output using

$$O_k = f_0\left\{B_0 + \sum_{j=1}^{n} w_{jk}\left(f_h\left(B_h + \sum_{i=1}^{n} w_{ij} x_i\right)\right)\right\} \tag{4}$$

where $w=$weight, $B=$bias, $f=$ Activation function, $x=$time series data

d. Update weight and biases until the error converges to an acceptable level.

2.3 Performance Measure

The performance of the Seasonal ARIMA and the ANN at time t will be checked using, the Root Mean Square Error and Mean Absolute error, given as

$$RMSE = \sqrt{\sum_{t=1}^{n} \frac{\left(A_t - F_t\right)^2}{n}} \tag{5}$$

$A_t =$ actual value of Nigerian Rainfall at time t

$F_t =$predicted value of Nigerian Rainfall at time t

It is calculated by taking the square root of the squared individual forecast deviation (error) for each period and then finding the average or mean value of the sum of squared errors. The smaller value of RMSE means better model.

Similarly, Absolute Error (MAE) measures the accuracy of the prediction by averaging the alleged error (the absolute value of each error). The value of MAE can be calculated using the following formula.

$$MAE = \sum_{t=1}^{n}\left|\frac{A_t - F_t}{n}\right| \tag{6}$$

$A_t = actual\ value\ of\ Nigerian\ Rainfall\ at\ time\ t$

$F_t = predicted\ value\ of\ Nigerian\ Rainfall\ at\ time\ t$

3. Results and Discussion

3.1 SARIMA Modeling

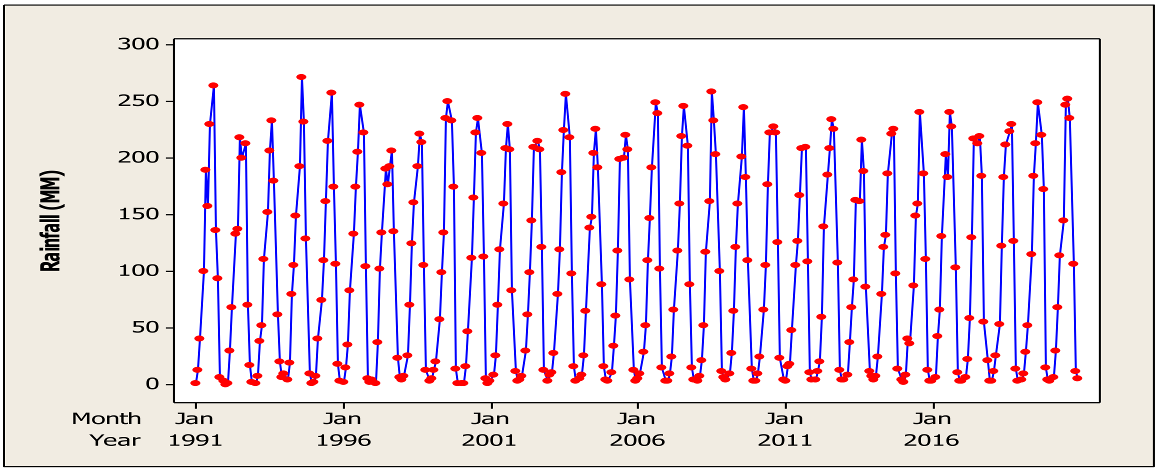

Figure 1. Time plot of Rainfall Data (Jan 1991-Dec. 2020)

Figure 1 presents monthly Rainfall and it indicates that the rainfall at level seems to be seasonal, reached its peak between July and August of every year as clearly be seen in seasonal graph depicted in figure 2.

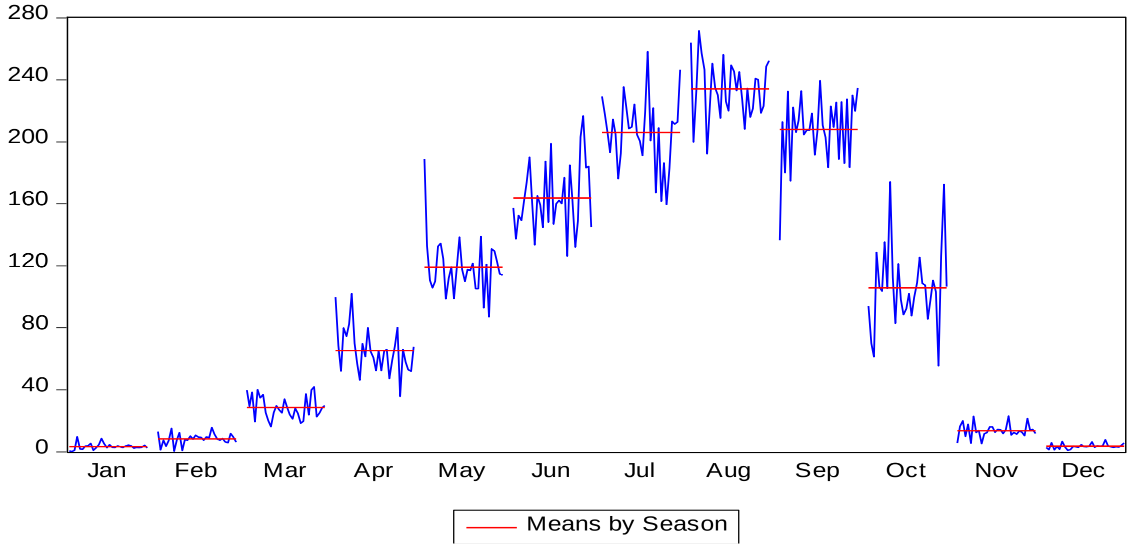

Figure 2. Seasonal Plot of Rainfall in Nigeria

Table 1. Unit root test of Nigerian Rainfall at level for trend stationary

Test	Test Statistic	Critical values	Decision
Augmented Dickey Fuller (ADF)	-4.719028	1% = −3.449053 5% = −2.869677 10% = 2.571174	Stationary
Kwiatkowski–Phillips–Schmidt–Shin (KPSS)	0.205985	1% = 0.739000 5% = 0.463000 10% = 0.347000	Stationary

Table 1 presents unit root tests to determine whether the Nigerian rainfall series is stationary at level. The results from ADF test leads to the rejection of the null hypothesis of unit root since the test statistic value is smaller than the critical values at 1%, 5% and 10% level of significance. The Nigerian rainfall series is trend stationary. Similarly, the KPSS test leads to the acceptance of the null hypothesis of stationarity since the test statistic is less than critical value hence, the Nigerian rainfall series is trend stationary. However, almost by definition, examination of the differenced data is necessary if seasonality is evident. Seasonality contributes to nonstationary of a time series because the average values are different at different time periods within the calendar year. Seasonal differencing is defined as a difference between a value and a value with lag that is a multiple of **S**. With S=12, which may occur with monthly data, a seasonal difference is $(1-B^{12})x_t = x_t - x_{t-12}$. The difference from previous year may be about the same for each month of the year giving us a stationary series.

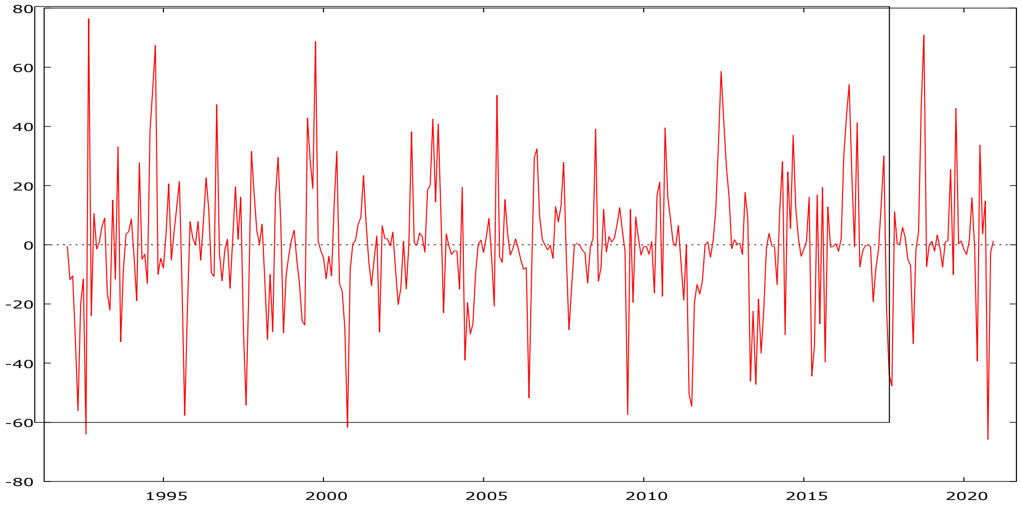

Figure 3. Plot of Nigerian Rainfall series after the seasonal difference

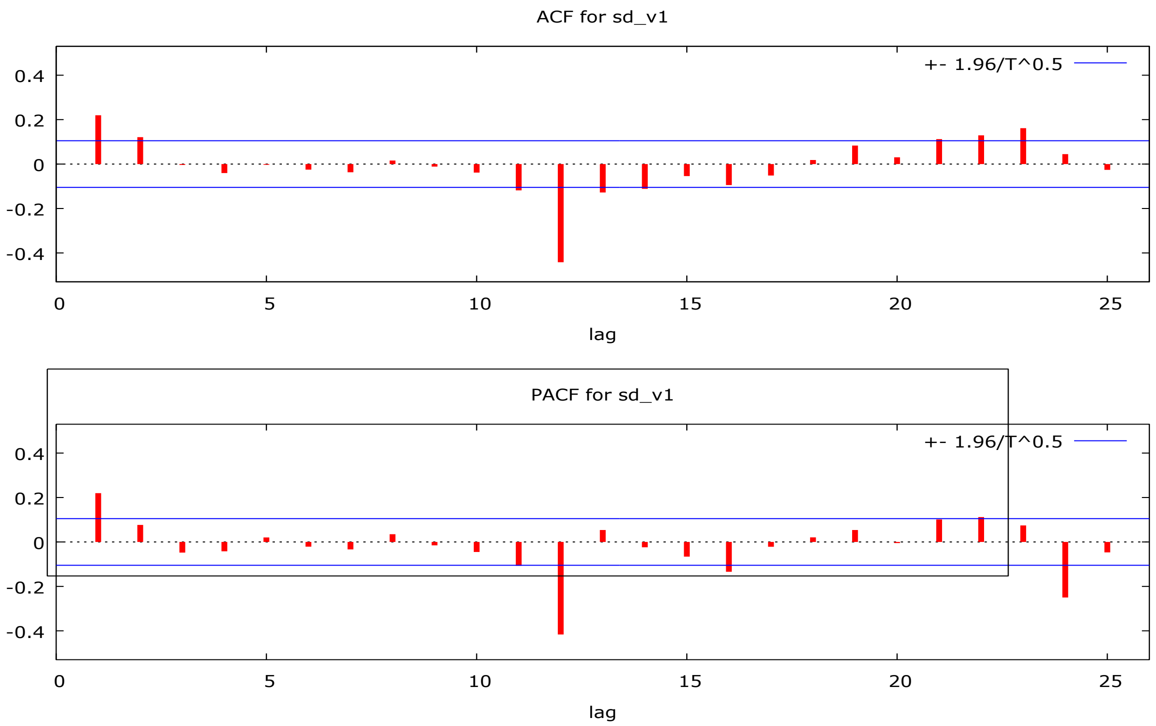

Figure 4. ACF and PACF for the first seasonal difference series

Figure 4 suggests non seasonal and seasonal MA(2) and AR(1) as there is significant spike for lag 1 and 2 at ACF and significant spike for lag 1 at PACF. Similarly, it indicates that there is monthly seasonal variation having observed a significant spike at S=12 and its multiples. Theoretically, figure 4 suggests SARIMA $(1,0,2)x(1,1,2)_{12}$.

Table 2. Estimated SARIMA $(1,0,2)x(1,1,2)_{12}$

	Coefficient	**Standard error**	**Z**	**P-value**
phi_1	0.394744	0.22157	1.7816	0.07482^{*}
Phi_1	0.426991	0.212311	2.0112	0.04431^{**}
$theta\_1$	−1.18508	0.238556	−4.9677	$< 0.00001^{***}$
$theta\_2$	0.185084	0.237882	0.7781	0.43654
$Theta\_1$	−1.35732	0.200892	−6.7565	$< 0.00001^{***}$
$Theta\_2$	0.499635	0.17583	2.8416	0.00449^{***}

0.01*, 0.05**, 0.10***

Table 2 presents parameter estimates for the SARIMA $(1,0,2)x(1,1,2)_{12}$ model, it indicates that both seasonal and non-seasonal coefficients are all significant except the non-seasonal second components of Moving average because its probability value (0.43654) is greater than common choices of 1%, 5% and 10% level of significance. This could be as result of the fact that the second spike of ACF in figure 4 is much away from the interval.

The SARIMA $(1,0,2)x(1,1,2)_{12}$ can be expressed using the backshift operator as

$$(1-B^{12})(1-\phi B)(1-\Phi_1 B^{12})X_t = (1-\theta_1 B-\theta_2 B^2)(1-\Theta_1 B-\Theta_2 B^{24})\alpha_t \tag{6}$$

$$(1-B^{12})(1-0.394744B)(1-0.426991B^{12})X_t = (1+1.18508B-0.185084B^2)(1+1.35732B$$
$$-0.499635B^{24})\alpha_t$$

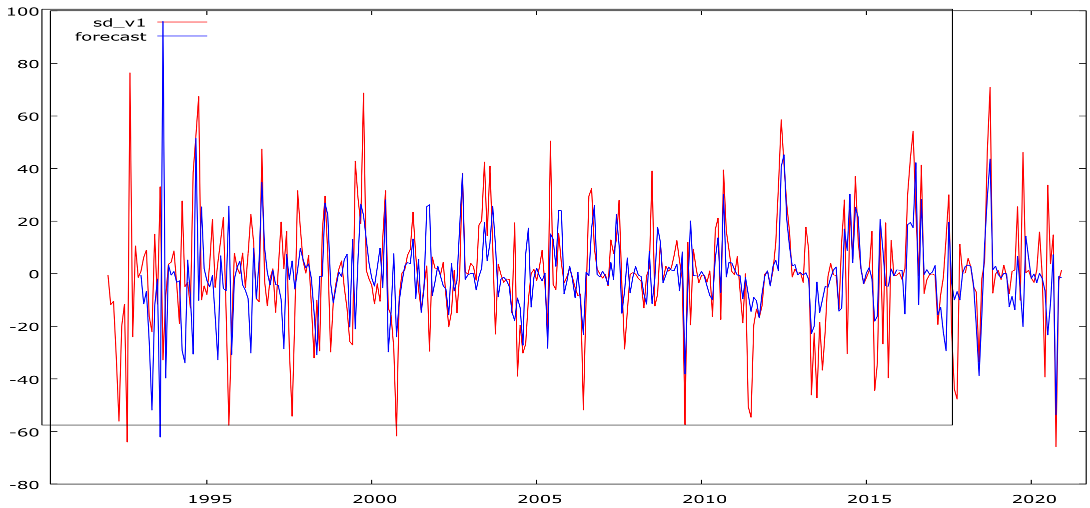

Figure 5. Rainfall In-Sample forecast plot for SARIMA $(1,0,2)x(1,1,2)_{12}$

Figure 5 provides the fitted values for in-sample one-step forecasts. Assuming normally distributed errors with 95% prediction intervals. The plot shows the comparisons between the actual values and predicted values for SARIMA $(1,0,2)x(1,1,2)_{12}$.

3.3 Neural Network Analysis

The neural network model was trained using Variants of Back propagation for weights and biases updates. An Intel (R) Core (TM) i3-2310M CPU @ 2.10GHz processor was used to train neural network models. Figures 6 and 7 are examples of the trained NAR configuration using Nigerian Rainfall data, Figure 6 is used for training purposes which are generally referred to as (Open loop) and Figure 7 is used for multi-step ahead prediction and are hence referred to as (Closed loop)

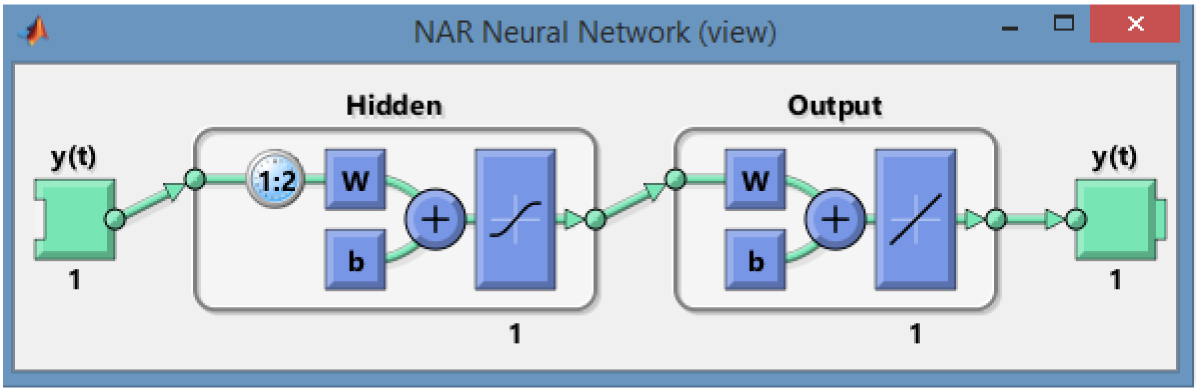

Figure 6. Open-Loop Architecture for Nigerian Rainfall forecaster

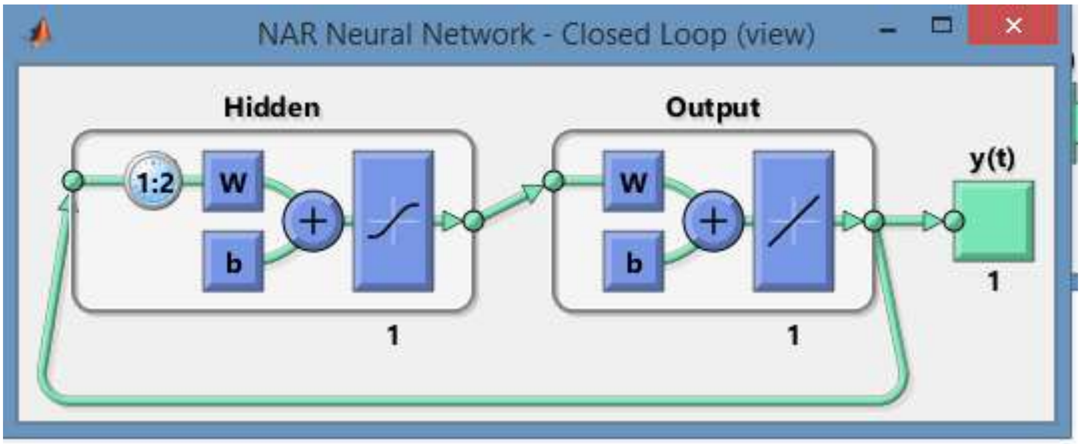

Figure 7. Closed-Loop for Nigerian Rainfall Forecaster

The data was divided to enable training, validating and testing the neural network developed using the Scaled Conjugate Gradient Descent (SCGD) and Levenberg Marquardt (LM) algorithms. The training process, validation process and the testing process received 70%, 15% and 15% of the data respectively. The neural networks developed using the BR algorithm on the other hand utilized 70% of the data for the training process, 15% for the testing process and the remaining 15% was not allocated (in order to obtain a relevant comparison of the final results achieved by the three algorithms). In all the cases, the samples have been randomly chosen as to cover the specified percentages. The mean square error (MSE) was used as an objective function in training the network. However, accuracy is optimized depending on the range of values of the output element. Thus, the network will learn to fit the first output element very well, while the second output element is not fit as accurate as the first hence, normalization of errors. This is achieved by setting the normalization performance parameter to its 'standard' value with the errors being computed as if both of the output elements had values ranging from -1 to 1 and consequently, the two output elements have been fitted very well.

Table 3. Descriptive statistics of different training algorithms for Nigerian Rainfall Data

Training Algorithms	**Average Time(s)**	**Maximum Time(s)**	**Minimum Time(s)**	**Standard Deviation**
Scaled Conjugate Gradient Descent	1.154545	3	0	0.934199
Levenberg Marquardt	1.00667	2	0	0.752773
Bayesian Regularization	13.0000	17	9	3.391165

Table 3 summarizes the results of training the proposed networks using the three training algorithms namely Scaled Conjugate Gradient Descent (SCGD), Levenberg Marquardt (LM) and Bayesian Regularization (BR). Each entry in the table represents 20 different trials, with random weights taken for each trial to rule out the weight sensitivity of the performance of the different training algorithms. The network was trained in each case till the value of the error was 0.001 or less. The average time required for training the network using the Levenberg-Marquardt algorithm was generally the least, whereas, maximum time is required for training the network using Scaled Conjugate Gradient Descent algorithm. The training algorithm employing Bayesian Regularization continuously modifies its performance function and hence, takes more time on average to train compared to the Levenberg-Marquardt and Scaled Conjugate Gradient Descent algorithms.

Table 4. Performance of trained Neural Network for Nigerian Rainfall Data

Training Methods	**Average Absolute Error**	**Minimum**	**Maximum**	**Average no. of Epoch**
Scaled Conjugate Gradient Descent	0.00193506	0.001215	0.003741	18
Levenberg Marquardt	**0.000525056**	0.000347	0.00087132	11
Bayesian Regularization	0.000767934	0.00064785	0.00082544	90

Similarly, Table 4 presents an average best validation of a mean squared error from each of the three training algorithms of twenty trials. It can be seen from this table that Levenberg-Marquadrt algorithm has least average mean squared error. Levenberg-Marquadrt algorithm is subsequently chosen and used for the prediction of Nigerian Rainfall.

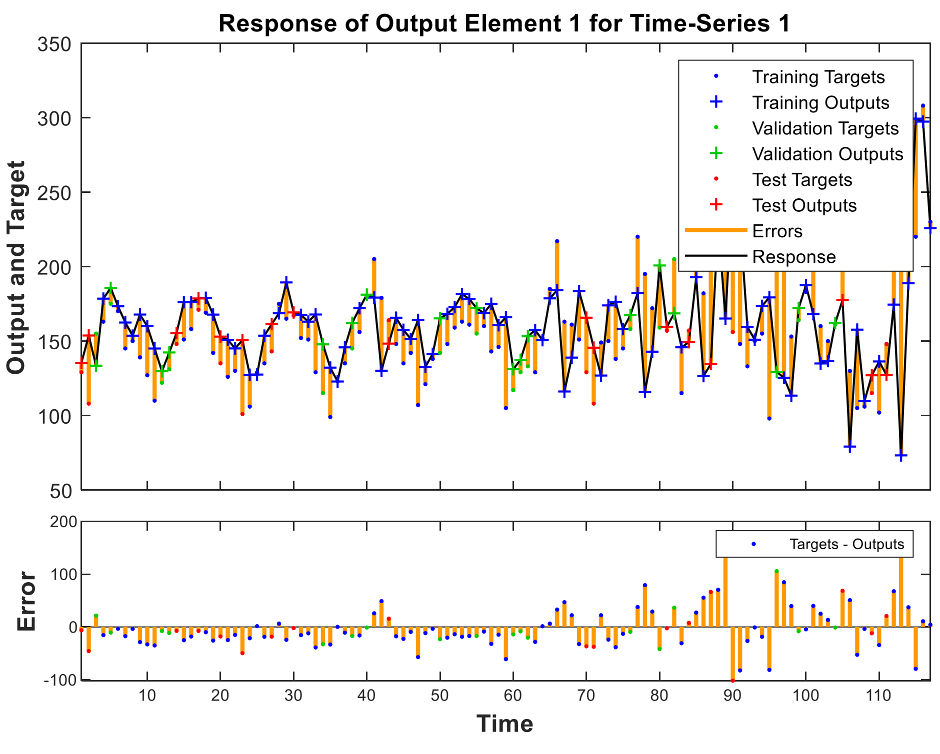

Figure 8. Some Actual and Predicted Values by Levenberg-Marquadrt

Figure 8 is the plot of actual and predicted values from Neural Network trained by Levenberg-Marquadrt algorithm of Nigerian rainfall. It indicates that LM_NN predicted values mimics the behaviour of the data very well.

Table 5. Prediction performance of the selected models

Forecasting performance metric	SARIMA(1,0,2)x(1,1,2)$_{12}$	LM_NN model
RMSE	0.01631	0.003359
MAE	0.01238	0.00285

Table 5 shows the performances of the two selected models in terms of Root mean square error and mean absolute error.

It is observed from Table 5 that the Neural Network trained by Levenberg-Marquadrt algorithm exhibits the least values in terms of the forecast performance metric.

4. Conclusion

The SARIMA$(1,0,2)x(1,1,2)_{12}$ model and the Neural network model were developed to compare their performances based on the root mean square error (RMSE) and the mean absolute error (MAE). The neural network trained by Levenberg-Marquadrt algorithm performed better than other training algorithms used in the study while the SARIMA$(1,0,2)x(1,1,2)_{12}$ model was chosen out of the subclasses of the SARIMA models. Comparison between the SARIMA$(1,0,2)x(1,1,2)_{12}$ model and the neural network trained by Levenberg-Marquadrt algorithm revealed that the neural network trained by Levenberg-Marquadrt algorithm exhibits the least RMSE and MAE values. It is concluded based on this information that the neural network trained by Levenberg-Marquadrt algorithm will predict values of Nigerian rainfall series better than the SARIMA $(1,0,2)x(1,1,2)_{12}$.

References

Aliyu, A. S., Auwal, A. M., & Adenomon, M. O (2021). Application of Sarima Models in Modelling and Forecasting Monthly Rainfall in Nigeria. *Asian Journal of Probability and Statistics*, *13*(3), 30-43. https://doi.org/10.9734/ajpas/2021/v13i330310

Amaefula, C. G. (2021). A SARIMA and Adjusted SARIMA Models in a Seasonal Nonstationary Time Series; Evidence of Enugu Monthly Rainfall. *European Journal of Mathematics and Statistics*, *2*(1), 13-18. https://doi.org/10.24018/ejmath.2021.2.1.15

Ayodele, A. P., & Precious, E. E. (2019). Seasonal Rainfall Prediction in Lagos, Nigeria using Artificial Neural Network. *Asian Journal of Research in Computer Science*, *3*(4), 1-10. https://doi.org/10.9734/ajrcos/2019/v3i430100

Dabral, P. P., & Murry, M. Z. (2017). Modelling and forecasting of rainfall time series using SARIMA. *Environmental Processes*, *4*(2), 399-419. https://doi.org/10.1007/s40710-017-0226-y

Darji, M. (2019). Rainfall Forecasting Using Neural Networks. *International Journal of Research and Analytical Reviews*. Retrieved July 30, 2021 from https://ssrn.com/abstract=3659043

Esteves, J. T., de Souza Rolim, G., & Ferraudo, A. S. (2019). Rainfall prediction methodology with binary multilayer perceptron neural networks. *Clim Dyn*, *52*, 2319–2331. https://doi.org/10.1007/s00382-018-4252-x

Mishra, N., Soni, H. K., Sharma, S., & Upadhyay, A. K. (2018). Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data. *International Journal of Intelligent Systems & Applications*, *10*(1), 16-23. https://doi.org/10.5815/ijisa.2018.01.03

Novotny, E. V., & Stefan, H. G. (2007). Stream flow in Minnesota: Indicator of climate change. *Journal of Hydrology*, *334*, 319-333. https://doi.org/10.1016/j.jhydrol.2006.10.011

Nwokike, C. C., Offorha, B. C., Obubu, M., Chukwuma, B., Ugoala, C. B., & Ukomah, H. I. (2020). Comparing SANN and SARIMA for forecasting frequency of monthly rainfall in Umuahia, *Scientific African*, *10*, 1-15. https://doi.org/10.1016/j.sciaf.2020.e00621

Obot, N., & Onyeukwu, N. O. (2010). Trend of rainfall in Abeokuta, Ogun State Nigeria: A 2-year experience (2006 – 2007), *Journal of Environmental Issues and Agriculture in Developing Countries*, *2*(1), 70-81.

Okon, E. J., & Ikpang, I. N. (2020). Modeling and Forecasting Exchange Rate Values between the Naira and US Dollar to Assess the Effect of COVID-19 Pandemic Period on the Rate. *Asian Journal of Probability and Statistics*, *8*(1), 55-65. https://doi.org/10.9734/ajpas/2020/v8i130200

Ray, S., Das, S. S., Mishra, P., & Al Khatib, A. M. G. (2021). Time series SARIMA Modeling and forecasting of monthly rainfall and temperature in the south Asian countries. *Earth Systems and Environment*, 1-16. https://doi.org/10.1007/s41748-021-00205-w

Sanni, I., Ahmed, H., Abubakar, U., & Abdullahi, S. (2019) Kano State Point Rainfall Seasonal Arima Model Prediction in R. *Proceedings on Big Data Analytics & Innovation (Peer-Reviewed), 4, 24-35*

Sekhar, H., Poola, K., Sekhar, K. R., & Naidu, M. B. (2020), Modelling and prediction of coastal Andhra rainfall using ARIMA and ANN models. *International Journal of Statistics and Applied Mathematics*, *5*(6), 104-110.

Soumen, P., & Debasis, M. (2018). Forecasting monthly rainfall using artificial neural network. *RASHI: Journal of the Society for Application of Statistics in Agriculture and Allied Sciences*, *3*(2), 65-73.