SEASONAL RAINFALL PREDICTION IN LAGOS, NIGERIA USING ARTIFICIAL NEURAL NETWORK

ADIGUN PAUL AYODELE, *EBIENDELE EROMOSELE PRECIOUS

Dept. of meteorology, Federal University of Technology, Akure

Dept. of meteorology, Federal University of Technology, Akure

ABSTRACT

Deliberating the importance of rainfall in determining process such as agriculture, flood and water management, these study aim at evaluation of non-linear techniques on seasonal rainfall prediction (SRP). One of the non-linear method widely used is the Artificial Neural Networks (ANN) approach which has the ability of mapping between input and output patterns. The complexity of the atmospheric processes that generate rainfall makes quantitative forecasting of rainfall an extremely, difficult task. The research goal is to train/develop Artificial Neural Network model using backward propagation algorithm to predict seasonal Rainfall. Using some meteorological variables like, sea surface temperature (SST), U-wind at (surface, 700, 850 and 1000), air temperature, specific humidity, ITD and relative humidity. The study adopt monthly June-October (JJASO) rainfall data and January-May (JFMAM) monthly data of SST, U-wind at (surface, 700, 850 and 1000), air temperature, specific humidity and relative humidity for a period of 31 years (1986-2017) over Ikeja. The proposed ANN model architecture (9-4-1) in training the network using back-propagation algorithm indicated that the statistical performance of the model for predicting 2013 to 2017 (JJASO) rainfall amount indicated as follows; MSE, RMSE, and MAE were 7174, 84.7 and 18.6 respectively with a high statistical coefficient of variation of 94% when the ANN model prediction is validated with the observed rainfall. The result indicated that the propose ANN built network is reliable in prediction of seasonal rainfall amount in Ikeja with a minimal error

1. INTRODUCTION

Rainfall is a natural climatic phenomenon whose prediction is challenging and demanding as the world continues to witness an ever changing climate conditions. Its forecast plays an important role in water resource management and therefore, it is of particular relevance to agricultural sector, which contributes significantly to the economy of any nation.

The forecasting of the rainfall distribution spatially and temporally is important for water quality and quantity management. This quantitative forecasting can be readily used by the flood warning system to increase the lead time for warning. Besides, the qualitative forecasting of rainfall can be used to analyze many water quality problems

Rainfall is one of the most complex and difficult elements of the hydrological cycle to understand and model due to the tremendous range of variation over a wide range of scales both in space and time (French et al., 1992). The complexity of the atmospheric processes that generate rainfall makes quantitative forecasting of rainfall an extremely difficult task (Hung et al, 2008). Generally, rainfall has strong influence on the operation of dams and reservoirs, sewer systems, traffic and other human activities

A thorough statistical analysis of rainfall distribution over the study area will not only be imperative but useful to the agricultural, social, commercial and industrial sectors of the economy of the study area but at the same time be a stepping stone to sustainable development of the entire country. Thus, it is imperative therefore to find out if this condition is practically obtainable in Lagos considering the fact that agricultural practice in Nigeria is also rain fed. Also, not just for no other thing to study, but because of the importance of such knowledge for all planning schemes for which rainfall is widely used. Further to this is the fact that such assumption about rainfall condition in the area may have serious and delicate implications on agricultural production.

In agricultural production especially in the tropics, rainfall is without doubt a critical climatic factor. It is known fact that one of the two major limiting factors to agricultural production next to soil fertility is nothing but insignificant water supply (Oladipo, 1993). Rainfall is the main source of soil moisture in any given environment. Thus an assessment of its distribution be it monthly, weekly, and especially daily distribution is therefore of great importance in agricultural planning

Despite the recent advances made in science and technology, farmers and their crops are still left at the mercy of rainfall especially in Sub-Saharan Africa. Hence water supply for agricultural practices is highly dependent on precipitation. Moreover in the areas where the climate is greatly influenced by drought and desertification, the condition of precipitation in relation to yield, the rate of evapo-transpiration and soil moisture content may help promote or hinder crop production. This is subject to availability of moisture at the evaporating surface and the ability of the atmosphere to vaporize the water. When plants transpire at the maximum rate (i.e. when soil is completely saturated) the term potential evapo-transpiration is applied (Jackson, 1977; Oguntoyinbo, 1983). This condition has great implication on crop production.

Rainfall is one of the key climatic elements of Lagos, because crops, animals and indeed humans derive their water resources largely from rainfall. It is considered as the main determinant of the types of crops that can be grown in the area and also the determinant of social activities of such crops and the farming systems practiced. Rainfall over Nigeria has been found to be controlled by the advection of moisture from the Gulf of Guinea in the low levels of the atmosphere. With the seasonal excursion of the Sun, the monsoon develops over this part of the African continent during the northern summer, bringing the Inter- Tropical Convergence Zone (ITCZ) and the associated rainfall maxima to their northernmost location in mid- July and early part of August, and this determines the time for the rainy season in this region. There are also both vertical and horizontal motions in the atmosphere that carry water vapor across the water bodies from the continents and especially the oceans to produce significant precipitation over the area of study. This water content of the atmosphere varies in response to the loss of water as precipitation and the gain from evaporation along a streamline (Kunstmann and Jung, 2005).

Over the last few decades, several statistical models have been developed, attempting the successful forecasting of rainfall in Nigeria. But some of this model failed by overestimating and underestimating the rainfall amount, majorly due to selected Predictors used in formulating the models. Approximately a dozen predictors have emerged from different studies as being most important for predicting Nigeria rainfall. These predictors are location of the inter-tropical discontinuity zone, sea surface temperature, Synoptic observation data such as Relative humidity, Land temperature and the surface pressure and land/sea thermal contrast. A very basic question, which does not seem to have been addressed systematically, is which of these predictors should be utilized at a given time?

STUDY AREA

Ikeja, being state capital of Lagos State, Nigeria, is the location for this study. The city is located on latitude 6.59'N and longitude 3.34' and it is situated 41m above sea level. 17km northwest of Lagos, southwestern Nigeria. Across Nigeria, rainfalls mostly when an area is overlain by the maritime air mass, and there is drought when the area is overlain by the continental air mass. The specific location Ikeja

($6.60°N, 3.35°E$) was chosen because of its ITD influence and its close proximity to the Sea (Gulf of Guinea $3.735°N, 3.7435°E$). Rainfall in coastal climatic zone is greatly influenced by sea surface temperatures, as noted by (Omotosho *et al*. 2002 and Odekunle 2005) the correlation coefficient of the prediction model generated by earlier studies, using SST for the above rainfall parameters, are rather weak (mostly 0.30–0.40) due to farthest distance to the sea.

2. DATA AND METHODOLOGY

Ground Observation data

The research work adopted total monthly rainfall from June to October and monthly means of Sea Surface Temperature (SST), Air Temperature, Specific Humidity, Relative Humidity (RH), and u wind at different pressure levels (surface, 750hpa, 800hpa, 1000hpa) from January to May for a period of 31 years (1986-2016) over Lagos. Rainfall and inter-tropical discontinuity (ITD) data were obtained from the archives of Nigerian Meteorological Agency (NIMET) Ikeja.

Era-Interim Reanalysis Data

These data are global reanalysis dataset produce by European Center for Medium-Range Weather Forecast (ECMWF) covering the period from 1979 to present. Data are gridded and available at 2.5° x 2.5°, 1° x 1° and 0.75° x 0.75° and 37 vertical pressure levels. Compared to the previous ERA-40 data, several problems concerning the humidity and hydrological cycle over the tropical region were improved for the ERA-interim to better agree with observations provide a detailed description of the ERA-interim product archive. The monthly means of Sea Surface Temperature, U-wind at three different pressure (surface, 1000hpa, 800hpa, and 750hpa) levels from January to May for a period of 32years (1986-2017) over Lagos were also used for this study. The data was obtained at 0.75° x 075° grid points.

Experimental Set-up of Artificial Neural Network Model (ANN)

Artificial Neural Network model (ANN) algorithm was developed using open-source software called Rstudio, the ANN model was trained using the neuralnet package in Rstudio. In construction of an artificial neural network the selected inputs or predictors (u-wind at different pressure levels, SST, air temperature, relative and ITD) at the summing junction can be written as;

$$U_K = \sum_{j=1}^{m} w_{kj} x_j \quad and\ V_k = U_k + b_k$$

The net input is then applied to an activation function, whose main objective is to limit the amplitude of the neuron to some finite value, and helps in achieving exact output. The output of the k^{th} neuron is;

$$y_k = \varphi(u_k + b_k)$$

The ANN model investigated the effects of January to May (JFMAM) U-wind at four different levels (surface, 700hpa,800hpa, 1000hpa) levels, Relative Humidity, ITD, SST and Air Temperature on JJASO seasonal rainfall using ANN model. The effect of the number of hidden neurons on ANN model was also examined. The predictors (u-wind, SST, air temperature, relative and ITD) from January to May, which were chosen as inputs in the network architecture in predicting June to October rainfall (network output/target), were divided into two parts each, one part (84% of data which is equivalent to 30years (1986-2012)) for training the network model and the other part (16% of data which is equivalent to 4years (2013-2017)) for validating the network model. Both input and dependent data must be normalized to avoiding over-fitting of model results. Normalization of data is done mathematically below;

$$x' = \frac{(x-a)}{(b-a)} where\ a = minimum\ value\ and\ b = maximum\ value$$

Statistical Evaluation of Model Performance

Statistical assessment and performance of ANN model results was carried out using an open-source software called Rstudio, The model comparison will be carried out using the following criteria:

(a) Mean square error; given as

$$MSE = \frac{\sum_{i=1}^{n}(yi - \hat{y}i)2}{n} \quad (3.1)$$

Where $y_i = Observed\ Rainfall(mm)$,

$$\hat{y}_i = Predicted\ Rainfall(mm)$$

$$n = number\ of\ row$$

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(yi - \hat{y}i)2}{n}} \quad (3.2)$$

Where $y_i = Observed\ Rainfall(mm)$,

$$\hat{y}_i = Predicted\ Rainfall(mm)$$

$$n = number\ of\ row$$

(c) Prediction Error (PE); given as

$$\frac{(|y_{Predicted} - y_{Observed}|)}{(y_{Observed})} \quad (3.3)$$

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

(d) Correlation Coefficient (r)

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (3.4)$$

3. RESULT AND DISCUSSION

The rainfall amounts during the summer season over Ikeja have been analyzed using artificial neural network (ANN), a nonlinear model to predict the seasonal rainfall over the study area. A seasonal rainfall ANN model was built, trained and validated using observed data sets. In order the results are presented and discussed below.

3.1 Training and Construction of Artificial Neural Network Algorithm

Annual rainfall amount during the rainfall season of June to October in Ikeja was used as dependent variable with 9 parameters as inputs. The parameters are air Temperature, SST, U-wind at surface, 750hpa, 800hpa, 1000hpa, relative humidity, the surface location of ITD 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. The data was divided into a training data set (1986-2012) and test data test (2013-2017) for cross validation/model performance. Both input and dependent data must be normalized to avoiding over-fitting of model results. Normalization of data is done mathematically below;

$$x' = \frac{(x - a)}{(b - a)} where\ a = minimum\ value\ and\ b = maximum\ value$$

Figure 1 depicts a graph showing the ANN model architecture with 9-4-1 (9 inputs, 4 hidden layers and 1 output), with the eight inputs parameters. The relationship between the nine input and the four hidden layers are also shown as values on the lines joining each of the input parameters and the hidden layers as well as those of the hidden layers and the final output are also shown in Figure 1, the four hidden layers are responsible for mapping a nonlinear relationship between the nine input and output (rainfall amount). The significance of these values between the nine 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.

The artificial neural network algorithm was developed using R-programming language, this is shown in appendix (a). The model architecture was seen to be 9-4-1 (8 inputs, 4 hidden layers and 1 output) in Figure 1. The training stopped when the error on the validation set reached the minimum.

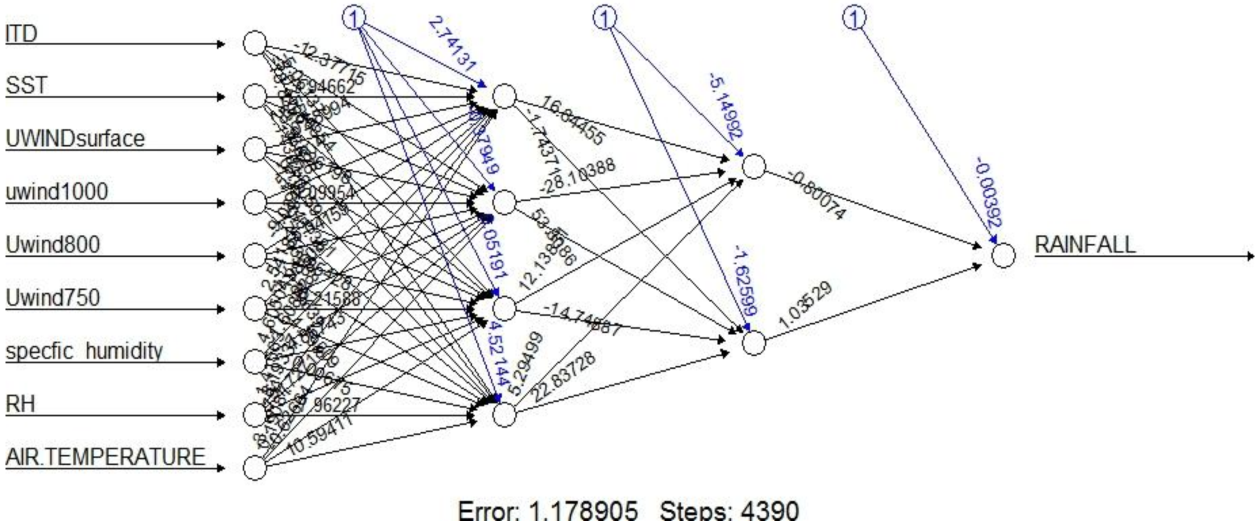

Figure 1: ANN model architecture for the nine input (9) parameters, 4 hidden layers and the final output (rainfall) parameter

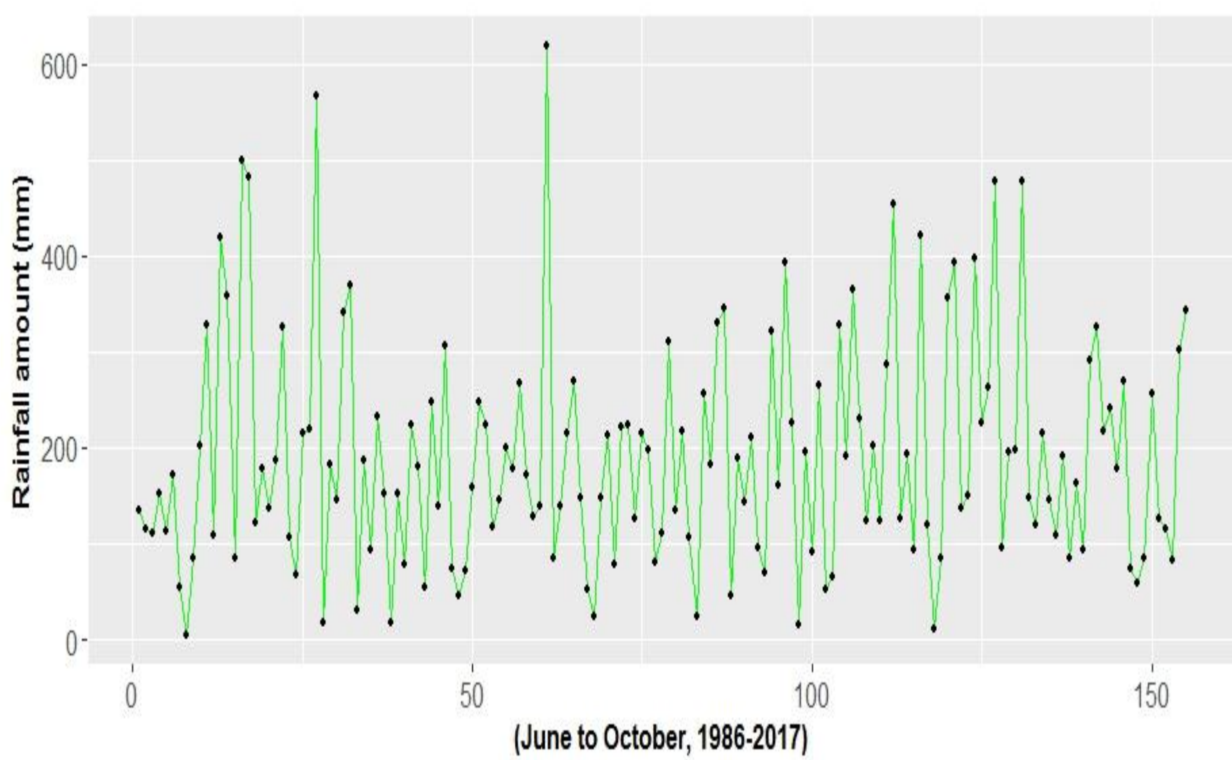

Figure 2: seasonal rainfall amount for the period of June to October from 1986 to 2017 in Ikeja

3.2 Comparison of ANN Model Output and Observed

Figure (3) shows the monthly predicted and observed rainfall amounts for Ikeja for the periods of June to October of 2013, 2014, 2015, 2016 and 2017 respectively

The monthly rainfall amounts predicted in June, July, August, September and October were very close to the observed in 2013 (Figure 3a). The error was somehow large in June which is the start of seasonal rainfall in Ikeja. The differences between the observed and predicted in percentage are 36%, 8.7%, 4%, 4.3% and 3.8% for June to October respectively

The monthly rainfall amounts predicted in June, July, August, September and October were very close to the observed in 2014 (Figure 3b). The difference between the predicted and observed are 4%, 1.8%, 53.5%, 4.7% and 7.8% for June to October respectively. The error was somehow large in August which is during the period of the little dry season phenomena. The ANN model output underestimated rainfall in July, August and September except June and October that was overestimated.

The monthly rainfall amounts predicted in June, July, August, September and October were very close to the observed in 2015 (Figure 3c). The difference between the predicted and observed are 3.9%, 11.1%, 1.5%, 3.4% and 10.7% for June to October respectively. The error was somehow large in July which is the peak of the rainy season. The ANN model output underestimated rainfall in June, September and October except July and August that was overestimated.

The monthly rainfall amounts predicted in June, July, August, September and October were very close to the observed in 2016 (Figure 3d). The difference between the predicted and observed in percentages are 0.2%, 34.8%, 6.8%, 11.3% and 8.7% for June to October respectively. The error was somehow large in July which is the peak of the rainy season. The ANN model output underestimated rainfall in July, September and October except June and August that was overestimated.

The monthly rainfall amounts predicted in June, July, August, September and October were very close to the observed in 2017 (Figure 3e). The difference between the predicted and observed are 3.6%, 4.2%, 10.9%, 4.9% and 9.2% for June to October respectively. The error was somehow large in August which is the period of the little dry season phenomena. The ANN model output underestimated rainfall in June, August September and October except July that was overestimated.

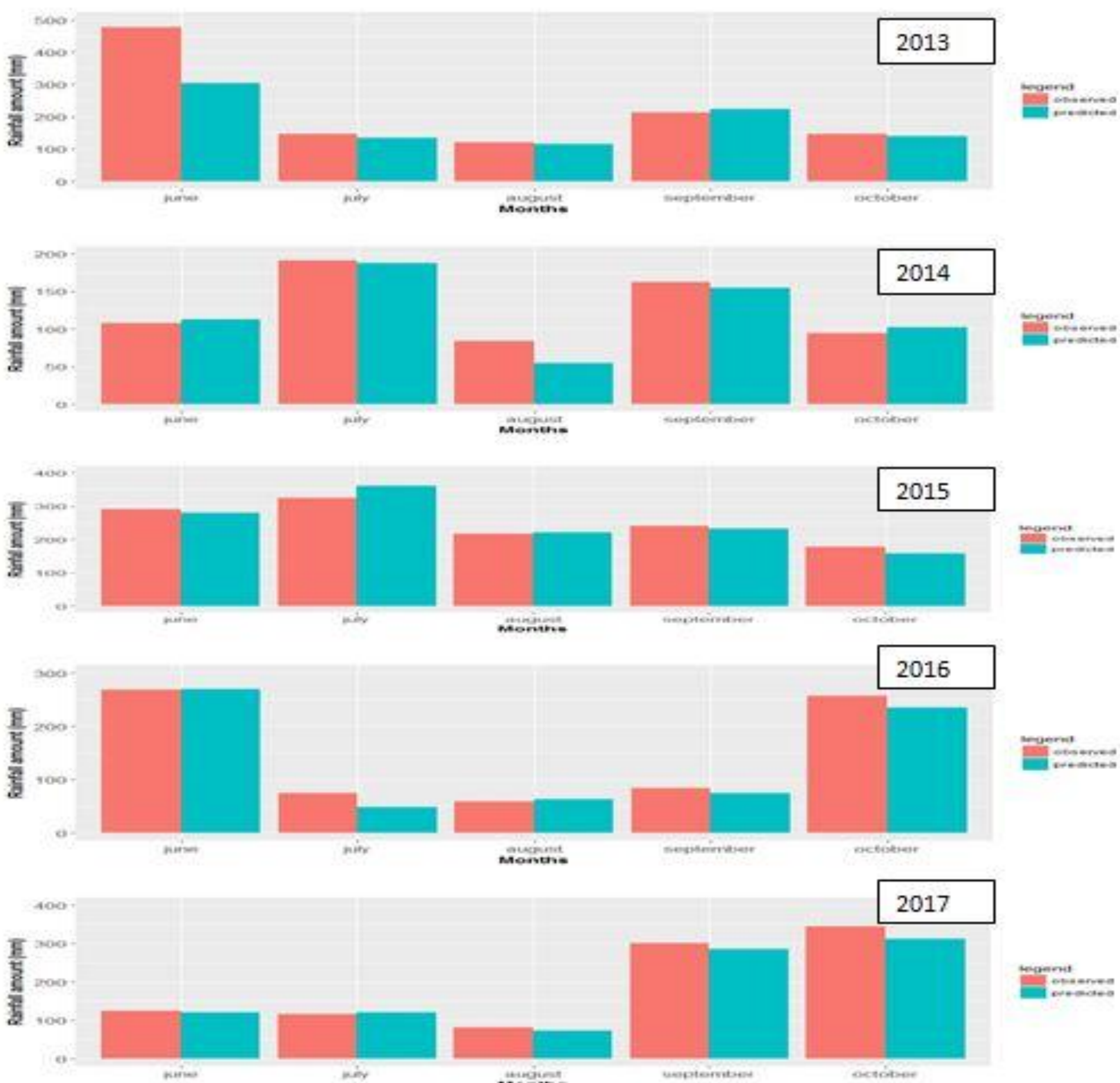

Figure 3: monthly predicted and observed rainfall amounts for 2013 to 2017 over Ikeja

Table 1: Predicted and Observed monthly and seasonal rainfall amounts

Year	Month	Observed rainfall amount (mm)	Predicted rainfall amount (mm)	Differences $PRE-OBS$ (mm)	Percentage error (%) $\frac{DIF}{OBS} \times 100$
2013	June	477.6	304.2	173.4	36.3
2013	July	147.5	134.6	12.9	8.7
2013	August	120	115.2	4.8	4.0
2013	September	214.1	223.4	9.3	4.3
2013	October	146.3 Total= 1105.5	140.7 Total=918.1	5.6	3.8
2014	June	108	112.6	4.6	4.3
2014	July	190.8	187.4	3.4	1.8
2014	August	84	54.7	29.3	34.8
2014	September	162	154.6	7.4	4.6
2014	October	94.2 Total=639.0	102.2 Total=611.5	8	8.4
2015	June	291.7	280.3	11.4	3.9
2015	July	325.2	361.2	36	11.1
2015	August	217.4	220.8	3.4	1.5
2015	September	240.8	232.5	8.3	3.4
2015	October	177.6 Total= 1252.7	158.5 Total=1253.3	19.1	10.7
2016	June	268.6	269.2	0.6	0.2
2016	July	74.4	48.5	25.9	34.8

2016	August	58.9	62.9	4	6.8
2016	September	84.4	74.8	9.6	11.3
2016	October	257.1 Total= 743.4	234.8 Total=690.2	22.3	8.7
2017	June	124.9	120.3	4.6	3.6
2017	July	115.6	120.5	4.9	4.2
2017	August	81.5	72.6	8.9	10.9
2017	September	301.1	286.1	15	4.9
2017	October	343.9 Total= 967	312.4 Total= 911.9	31.5	9.2

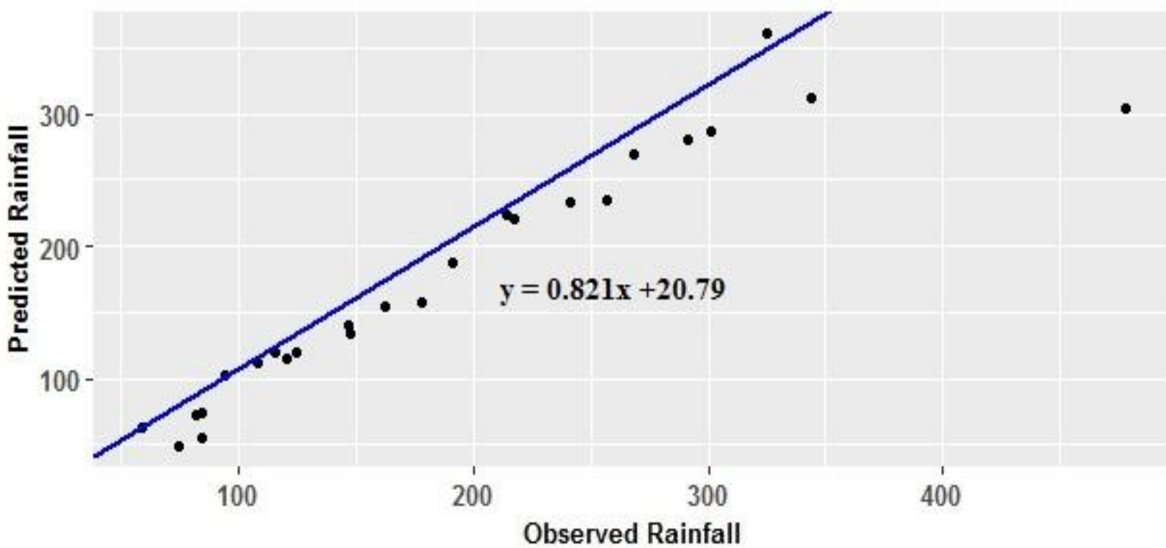

Figure 4: Scatter diagram of observed and predicted rainfall amounts

Figure 4: depicts a scatter plot of the predicted and the observed rainfall amounts (JJASO) from the ANN model output. This Figure showed that there is a linear relationship between the observed and predicted rainfall values.

3.3 EVALUATING PERFORMANCE OF BUILT MODEL

The accuracy of the ANN model has been evaluated using the Root Mean Square Error (RMSE), Mean Square Error (MSE), Mean Absolute Error (MAE) and correlation coefficient. This was to further evaluate the performance of the network output with the validation data (2013 to 2017) over Ikeja.

The results of testing performance evaluation for Ikeja station are presented below

Table 2: Evaluation of ANN model using error analysis

NETWORK	Year	MSE	RMSE	MAE	R^2
9-4-1	**2013 (JJASO)**	6074	77.9	41.2	0.94
9-4-1	**2014 (JJASO)**	202.0	14.2	8.9	0.98
9-4-1	**2015 (JJASO)**	374.2	19.3	11.8	0.95
9-4-1	**2016 (JJASO)**	255.3	15.9	8.0	0.96
9-4-1	**2017 (JJASO)**	268.3	16.8	6.7	0.97
9-4-1	**2013-2017 (JJASO)**	7174	84.7	18.6	0.94

The result for testing of model presented in table 2 indicates the built ANN model with network architecture of 9-4-1 the following results:

In 2013 seasonal rainfall, the RMSE, MSE and MAE were, 77.9, 6074 and 41.2 respectively with a statistical coefficient of variation of 94% when the ANN model prediction is compared with the observed rainfall.

In 2014 seasonal rainfall, the MSE, RMSE, and MAE were 202.0, 14.2 and 8.9 respectively with a statistical coefficient of variation of 98% when the ANN model prediction is compared with the observed rainfall.

In 2015 seasonal rainfall, the MSE, RMSE, and MAPE were 374.2, 19.3 and 11.8 respectively with a statistical coefficient of variation of 95% when the ANN model prediction is compared with the observed rainfall.

In 2016 seasonal rainfall, the MSE, RMSE, and MAPE were 268.3, 15.9 and 8.0 respectively with a statistical coefficient of variation of 96% when the ANN model prediction is compared with the observed rainfall.

In 2017 seasonal rainfall, the MSE, RMSE, and MAPE were 255.3, 16.8 and 6.7 respectively with a statistical coefficient of variation of 97% when the ANN model prediction is compared with the observed rainfall.

ANN model statistical performance for 2014 to 2017 (JJASO) rainfall amount indicated as follows; MSE, RMSE, and MAPE were 7174, 84.7 and 18.6 respectively with a high statistical coefficient of variation of 94% when the ANN model prediction is compared with the observed rainfall.

It is hereby discovered that the ANN model performs best in 2014 (having lowest RMSE,

MSE and MAPE with the highest r^2) which is the year with the lowest seasonal rainfall amount and performs badly in 2013 (having the highest RMSE, MSE and MAPE with the lowest r^2) which happens to be a year with high seasonal rainfall amount among the years for validation. It can therefore be concluded that ANN is good for predicting monthly seasonal rainfall amounts over West Africa if historical data is available but its efficiency or performance decreases with degree of wetness of the year or months considered.

4 CONCLUSION

Rainfall is one of the key entities of hydrological cycle that strongly influence the operation of dams and reservoirs, flood control, drought mitigation, operation of sewer systems, agricultural practice, traffic conditions and other human activities. As a result, accurate modeling of rainfall plays an important role in the management of water resources. The modeling of seasonal rainfall amount of Ikeja using ANN model developed by R-programming, showed that ANN is a good method of optimization, since error observed in the comparison of observed and model output is minimal. The neural network summary yielded 94% of good forecasts for seasonal rainfall amount in Ikeja. This showed that the trained network is reliable and fit to be used for the subsequent quantitative prediction of rainfall. Therefore it can be concluded that ANN model with eight (9) input parameters considered in this study will perform well in predicting seasonal rainfall amount in Ikeja. It was concluded that ANN is good for predicting monthly seasonal rainfall amounts over West Africa if historical data is available but its efficiency or performance decreases with degree of wetness of the year or months considered. The results from this study will provide information that will aid accurate seasonal rainfall prediction using the required parameters. In other words, agriculture, especially farmers can adopt this as a reliable forecast tool, also in water resource and other related sectors of the economy will benefit tremendously from the output of the study.

REFERENCES

[1] Bodri, L., & ermàk, V. (2000). Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia. Advances in Engineering Software, 31(5), 311-321.

[2] Chakraverty, S., & Gupta, P. (2008). Comparison of neural network configurations in the long-range forecast of southwest monsoon rainfall over India. Neural Computing prediction

[3] Chattopadhyay, S. (2007). Feed forward Artificial Neural Network model to predict the average summer monsoon rainfall in India. Acta Geophysica, 55(3), 369-382.

[4] Chattopadhyay, S., & Chattopadhyay, G. (2010). Univariate modelling of summer-monsoon rainfall time series: Comparison between ARIMA and ARNN. Comptes Rendus Geosciences, 342(2), 100-107.

[5] French MN, WF Cuykendall RR, Rainfall forecasting in space and time using a neural network , journal of hydrology. 1992 ;1-31

[6] Guhathakurta, P. (2008). Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model.

[7] Hornik, K(1991). Approximation capabilities of multilayer feedforward networks. Neural Networks, 4(2), 251-257. Islamic republic of Iran meteorological organization, veiwd Dec 2008, www.weather.ir

[8] Hung NQ, Babel MS, Weesakul S,Tripathi. An artificial neural network model for Rainfall forecasting in Bangkok, Thailand journal of hydrology and earth system sciences 2008; 13:1413-1425

[9] Kumar, D. N., & Sathish, T. (2008). Forecasting Hydrologic Time Series Using Artificial Neural Networks. The MathWorks, Inc 1984-2009, www.mathworks.com

[10] Mekanik, F. (2011). Rainfall Time Series Modeling for a Mountainous Region in West Iran. Master Thesis, Universitiy Putra Malaysia.

[11] Mekanik, F., Lee, T. S., & Shitan, M., (2009) Time Series Modelling of Rainfall in Hamedan. Iran. Proceedings of the Sains Matematik Ke-17, Melaka, Malaysia, Dec. 15-17

[12] Mishra, A., & Desai, V. (2005). Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment, 19(5), 326-339.

[13] Nourani, V., Alami, M. T., & Aminfar, M. H. (2009). A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Engineering Applications of Artificial Intelligence, 22(3), 466-472.

[14] Oguntunde, P. G.; Abiodun, BJ,; lischeid, G.2011, Rainfall trends in Nigeria, 1901-2000. J. Hydrol, 411, 207-218.

[15] Omotosho JB, Balogun AA, and Ogunjobi K. 2000. Predicting monthly and seasonal rainfall, onset and cessation of the rainy season in West Africa using only surface data. International Journal of Climatology 20:865-880.

[16] Omotosho J. A. and Abiodun B. J. (2007). A numerical study of moisture build-up and rainfall over West Africa. Meteorol. Appl. 14: 209-225.

[17] Olaniran, O. J. (2002). The Problems in the Measurement of Rainfall: An Experiment at Ilorin, Nigeria, Weather, 37 (7), pp. 201 – 204.

[18] Shamseldin, A. Y. (1997). Application of a Neural Network Technique to Rainfall-Runoff Modeling, J. of Hydrol., 199, pp. 272–294.

[19] Smith, J and Eli, R.N. „"Neural Network Models of Rainfall-Runoff Process"", Journal of Water Resources Planning and Management, pp. 499-508, 1995.

[20] Shamseldin AY. 1997. Application of neural network technique to rainfall-runoff modeling. Journal of Hydrology **199**: 272–294.

[21] Sivakumar, B., Jayawardena, A.W., and Fernando, T.M.K.G., „"River flow Forecasting: Use of Phase Space Reconstruction and Artificial Neural Networks Approaches"", Journal of Hydrology, 265, pp. 225-245, 2002.