Archives of Hydro-Engineering and Environmental Mechanics 72 (2025), pp. 29–49, doi: 10.2478/heem-2025-0003

sciendo

ARCHIVES OF HYDRO-ENGINEERING AND ENVIRONMENTAL MECHANICS

Published since 1954 ISSN 1231–3726

Univariate Monthly Rainfall Forecasting in Nigeria Using Multiple Statistical and Machine Learning Methods

Nsikan Ime Obot[1*], **Ibifubara Humphrey**[1], **Nkemdilim Maureen Ekpeni**[2], **Emmanuel Oluwatobiloba Tai-Ojuolape**[1]

[1]Department of Physics, Faculty of Science, University of Lagos, Akoka, Lagos State, Nigeria;
[2]Department of Geography, Faculty of Social Science, University of Lagos, Akoka, Lagos State, Nigeria;
[*]Corresponding author e-mail: nobot@unilag.edu.ng, phone: +2348077467463

(Received 31 March 2025; Revised 25 July 2025)

Abstract. Precise long-term rainfall prediction is important for agricultural planning, climate resilience, and reducing disaster risk, particularly for countries like Nigeria with diverse regimes of rainfall. In this research, the potential of machine learning (ML) and statistical models to predict monthly univariate rainfall in 24 Nigerian stations was evaluated. Model training employed historical rainfall data (1960–1999), while validation was carried out for 11 years (2000–2010). SARIMA $(p, d, q)\,(P, D, Q)_s$ models were used in Minitab®, R, and Python, and the most important parameters (p, d, q, P, D, Q) were tuned manually and by using auto.arima(). ML models such as feedforward neural networks, adaptive neuro-fuzzy inference systems, support vector regression and random forest were utilized in MATLAB® and R with hyperparameter-tuned models. Model performance was evaluated in using statistics such as root mean square error ($RMSE$) and coefficient of determination (r^2). SARIMA performed best in areas where rainfall variability was minimal. Nguru (12.03°N), the area with the lowest average monthly rainfall (35.71 mm), showed the highest SARIMA estimation with $RMSE$ of as low as 7.84 mm and r^2 of as high as 0.85. ML models underperformed in capturing seasonal dynamics. For instance, SVR failed to model temporal trends effectively, while random forest produced nearly constant outputs across all years. Adjustments to SARIMA parameters (e.g., setting seasonal differencing $D = 0$ or $Q = 1$) were essential in reducing unrealistic forecasts. The findings demonstrate that SARIMA, with proper tuning, is better suited for univariate rainfall forecasting in Nigeria than non-customized ML models. Forecast reliability strongly correlates with regional rainfall characteristics and model sensitivity to seasonality.

Key words: Rainfall forecasts, univariate time series, SARIMA models, machine learning algorithms, auto-ARIMA modifications, statistical modelling software

1. Introduction

Accurate prediction of rainfall – intensity, frequency, and timing – matters much to sectors such as agriculture, water resource planning, sports, recreation, and disaster management. Agriculture is the major activity in most developing nations, and rainfall variability determines crop output and food availability. Excessive rainfall is responsible for catastrophic floods, inundating agricultural fields, relocating individuals, and inducing erosion, loss of life, and damage to property. On the other hand, drought has the potential to cause food insecurity, migration, and stagnation of economic development. Public health interventions, particularly by seasonal diseases such as respiratory infections, skin diseases, and waterborne diseases, are also influenced by rainfall patterns (Ohba 2021, Gupta et al 2022). Rainfall also affects environmental visibility, radio wave propagation, hydropower production, vegetation dynamics, and freshwater supply in addition to agriculture and health. Precipitation pattern shifts can also indicate larger climate change effects, such as sea level rise, global warming, and increase in weather extremes (Bhaga et al 2020, Ibrahim et al 2021).

In order to model and interpret precipitation patterns well, knowledge of the mechanism of the various precipitation types must be acquired. Precipitation is generally of three major types: convectional, cyclonic (frontal), and orographic (relief). Convectional precipitation is caused by strong solar heating, producing evaporation and water vapor rise that condenses at higher altitudes to constitute rain. It occurs most frequently in the tropics (Al-Taai and Abbood 2020). Cyclonic or frontal precipitation takes place at convergence points between warm and cold air masses, for example, the Inter-Tropical Discontinuity (ITD), with less dense warm air rising above cold air masses to create clouds and rain. Relief or orographic precipitation takes place when moist air masses are pushed upwards by mountainous areas, cooling and condensing to create precipitation (Sun et al 2021). Evolution of rainfall is also controlled by many variables such as temperature, droplet collision frequency, atmospheric pressure, humidity, cloud composition, and even air particles like microorganisms or mineral dust. Wind speed, cloudiness, heat exchange mechanism, and adhesion at the molecular level also play their role in rainfall variability (Yin et al 2022).

In Nigeria, the biggest nation in Africa, West African Monsoon dominates precipitation. Though dew, hail, sleet, and snow may accompany this occasionally, the most common type of precipitation is rain. ITD, SST, and land-sea thermal exchange are the three significant climatic drivers of precipitation in this region (Igbawua et al 2024). These are most potent in the ITD form, accounting for approximately 70% of year-to-year rainfall variance. The ITD is a moving interface between two large air masses: wet maritime tropical air from the Atlantic and wet air and dry continental air from the Sahara Desert, which prevails in the dry season. North and southward displacements of the ITD control the initiation and retreat of the rainy season (Jackson et al 2020).

Rain forecasting in this active climatic setting is a very challenging process, requiring precision and computational capacity. Precise rainfall prediction is of the highest significance for prompt evacuation, irrigation scheme planning, allocation of water storage, and climate change adaptation (Nordin et al 2023). Various forecasting methods are being employed, starting from synoptic weather forecasting (SWF) and numerical weather prediction (NWP) to statistical and machine learning. NWP and

SWF typically rely on large-scale meteorological information to forecast weather conditions worldwide (Jansing et al 2022, Aminudin et al 2024). Statistical models like ARIMA, SARIMA, Fourier series, and linear regression have been widely applied in rainfall forecasting. Conversely, knowledge-intensive and intelligent techniques – i.e., artificial neural networks, support vector regression, adaptive neuro-fuzzy inference systems, and particle swarm optimization – proved successful by emulating biological learning processes for forecasting complex patterns. In addition, some organizations also constructed proprietary software packages in-house for forecasting rainfall, even though indigenous methods of forecasting remain relevant in rural cultures. These methods significantly differ from each other in accuracy and usability (Azad et al 2019, Isnaini et al 2025).

Different models have been used to forecast rainfall in Nigeria. These include ARIMA models, Fourier series, linear regression, and SARIMA models, which are typically developed from long-run rainfall records (Ikpang and Okon 2022). SARIMA models have been employed far-reaching, e.g., SARIMA models were created using 30 years of data by Akinbobola et al (2018), while Yaya and Fashae (2019) discovered that seasonal autoregressive fractional integrated moving average performed better than SARIMA when diagnostic tests were taken into account. Other discoveries include that SARIMA forecasts better in Southern Nigeria than fractional autoregressive integrated moving average models (Onyeka-Ubaka et al 2021).

Probabilistic models like Gumbel and Generalized Extreme Value distributions are useful in extreme rainfall event prediction but do not produce time-series forecasts (Wang and Gao 2023). Linear regression methods are also utilized in predicting the beginning or end of rainy seasons but explain only a fraction of the time series (Khastagir et al 2022). Parallelly, machine learning models like random forests and neural networks are increasingly being used and have been proven to perform better than conventional statistical models in certain studies (Rahaman et al 2023). Intelligent systems have been proven to exhibit better precision than traditional models like multiple regression and fuzzy logic (Ilaboya 2019, Rožman et al 2024). In spite of increasing interest in machine learning and artificial intelligence methods of rainfall prediction in Nigeria, few studies have directly compared these methods with statistical models for various locations and extended periods. The majority of studies have a spatial or temporal restriction, and no global standard method of forecasting is evident – especially under increasing pressures from climate change (Onyeka-Ubaka et al 2021, Zhang et al 2023, Kumar et al 2024). In addition, although multivariate models that account for various atmospheric factors are more detailed, univariate models accounting solely for rainfall data can also be very informative, particularly for specific applications (Necesito et al 2023, Othman et al 2024).

Nigeria's vegetal belts host various varieties of crops, and climatic unusualness in an area tends to spread into related weather systems. Climate change weakens these natural balance mechanisms, altering migration patterns, agriculture, and atmospheric conditions–particularly near the ITD (Tschora and Cherubini 2020, Akinbami 2021). It is therefore essential to analyze the performance of various forecasting methods over Nigeria's ecological regions for planning and policy purposes.

The aim of the present research is to predict long-run univariate time-series rainfall in Nigeria from statistical and machine learning models without the use of hybrid systems or external variables. It analyzes and compares the forecasting potential of classical statistical methods (e.g., ARIMA models) and intelligent models (e.g., decision trees, neural networks) for northern and southern sites. Innovation resides in its sole emphasis on autonomous models, regional analyses, and extended forecasting horizons, with a realistic estimate of model performance. Outcomes are designed to guide climate adaptation planning and make decisions for agriculture, engineering, public health, and disaster risk reduction more robust.

2. Materials and Methods

2.1. Data, Site, and Assessment Measures

Monthly cumulative rainfall records for 1960–2010 for 24 Nigerian stations were utilized in the research, which are accessible at the Nigerian Meteorological Agency (NIMET). The specific locations are listed in Table 1 and depicted in Fig. 1. Training was performed using about 78% of the data (1960–1999), and the remaining 22% (2000–2010) was stored for testing purposes. Missing values for the years 1967 at Katsina and Yola, which might have been missing because of the civil war, were filled by utilizing the average of surrounding months of the neighboring years (Obot et al 2024). The data mirrored regional rainfall pattern variations with altitude, proximity to bodies of water, and location. The southern states, though at lower altitudes, received greater levels of rainfall by the impinging force of the Atlantic Ocean, with monthly averages varying from 35.71 mm at Nguru to 241.10 mm at Calabar. An interesting anomaly was a record peak rainfall of 1034.3 mm in Benin in July 1965 and low rainfall during dry months in Maiduguri. Seasonality in rain was also different: northern peaks were in July–August, and parts of the south, such as Calabar, did not receive the usual August break (Figs 2 and 3). The research locations cover Nigeria's eight vegetation zones from Sahel Savanna to Mangrove Forest, latitudes 4.51°N to 13.01°N and elevation from 6.1 m to 1242.8 m.

The statistical measures, such as mean bias error, root-mean-square error, and coefficient of determination, play a crucial role in evaluating the forecasting proficiency of the models. These measures provide a quantitative assessment of accuracy and reliability. Additional measures, like the percentage error at maximum points and the absolute difference between minimum values, further enhance the evaluation process.

The expression for mean bias error is:

$$MBE = \frac{\sum \left(V_o - V_f\right)}{n}. \tag{1}$$

Table 1. Descriptive statistics of rainfall data used for forecasting models at locations across Nigeria

Location	Region	Latitude (°)	Longitude (°)	Elevation (m)	Vegetation	Median (mm)	Mean (mm)	Range (mm)
Katsina	north	13.01	7.41	517.6	Sudan Savanna	0.55	44.88	410.20
Sokoto	north	13.01	5.15	350.8	Sudan Savanna	2.45	52.80	374.90
Nguru	north	12.53	10.28	343.1	Sahel Savanna	0.00	35.71	335.30
Kano	north	12.03	8.12	475.2	Sudan Savanna	3.25	67.52	573.00
Gusau	north	12.03	6.42	463.9	Sudan Savanna	12.55	76.19	530.10
Maiduguri	north	11.51	13.05	354.8	Sudan Savanna	2.25	47.97	467.70
Yelwa	north	10.53	4.45	244.0	Sudan Savanna	37.45	83.10	609.90
Kaduna	north	10.35	7.27	645.4	Sudan Savanna	44.60	101.70	548.50
Bauchi	north	10.17	9.49	609.7	Sudan Savanna	25.25	84.24	520.70
Jos	north	9.52	8.45	1242.8	Montane Region	63.45	106.62	426.00
Minna	north	9.37	6.32	256.4	Guinea Savanna	72.70	101.70	533.90
Yola	north	9.14	12.28	186.1	Sudan Savanna	37.00	74.16	387.00
Ilorin	north	8.29	4.35	307.4	Guinea Savanna	86.65	102.12	382.50
Ibi	north	8.11	9.45	110.7	Guinea Savanna	74.05	90.52	453.20
Lokoja	north	7.47	6.44	62.5	Derived Savanna	86.35	101.35	448.60
Makurdi	north	7.44	8.32	112.8	Derived Savanna	81.05	102.97	884.20
Oshogbo	south	7.41	4.29	302.0	Lowland Rain Forest	101.40	110.42	449.60
Ibadan	south	7.26	3.54	227.2	Lowland Rain Forest	97.45	107.23	605.11
Ikeja	south	6.35	3.20	39.4	Fresh Water Swam Forest	98.15	128.34	832.00
Enugu	south	6.28	7.33	141.8	Derived Savanna	135.80	142.88	594.60
Benin	south	6.19	5.06	77.8	Lowland Rain Forest	152.70	178.45	1034.34
Warri	south	5.31	5.44	6.1	Lowland Rain Forest	206.00	232.43	941.70
Calabar	south	4.58	8.21	61.9	Lowland Rain Forest	227.60	241.10	796.60
Port Harcourt	south	4.51	7.01	19.5	Mangrove Forest	183.95	196.91	776.70

V_o represents the actual value, V_f represents the forecasted value, and n is the total number of instances.

The root-mean-square error of a model can be obtained as:

$$RMSE = \frac{\sqrt{\sum (V_o - V_f)^2}}{n}. \tag{2}$$

The coefficient of determination is a statistical tool that assesses the trend disparity between the actual and forecasted values. It is given as:

$$r^2 = \frac{n\left(\sum V_o V_f\right) - (\sum V_o)\left(\sum V_f\right)}{\sqrt{\left[n \sum V_o^2 - (\sum V_o)^2\right]\left[n \sum V_f^2 - \left(\sum V_f\right)^2\right]}}. \tag{3}$$

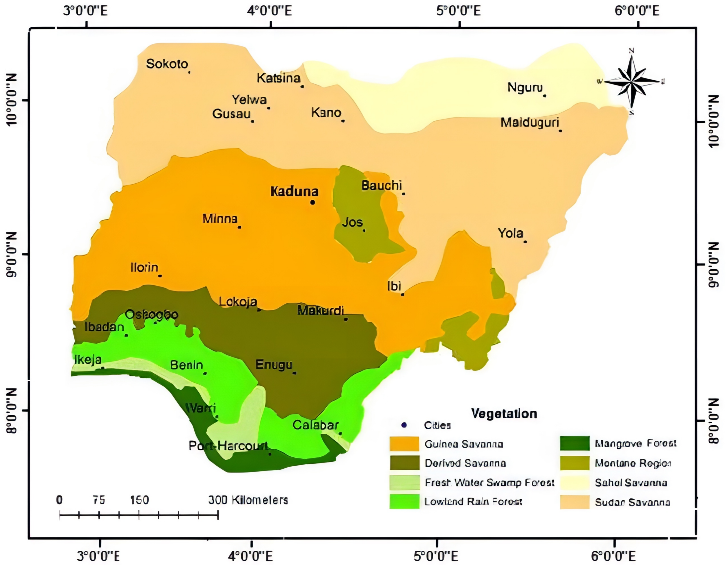

Fig. 1. Locations in their vegetation zones across Nigeria

To evaluate the extent of discrepancy between the maximum values of both the actual and forecasted rains, the percentage error at maximums ($PE_{\max}$) is proposed and adopted. It can be calculated as:

$$PE_{\max} = \frac{V_{o-\max} - V_{f-\max}}{V_{o-\max}} \times 100. \tag{4}$$

Here, $V_{o-\max}$ represents the maximum actual value, and $V_{f-\max}$ represents the maximum forecasted values.

Furthermore, the absolute difference between the minimums of the actual values and the forecasted values can be determined using the following:

$$|\Delta_{\min}| = \left|V_{o-\min} - V_{f-\min}\right|, \tag{5}$$

where $V_{o-\min}$ is the minimum actual value, and $V_{f-\min}$ is the minimum forecasted value.

2.2. Seasonal Autoregressive Integrated Moving Average Model

The seasonal autoregressive integrated moving average (SARIMA), along with its variants such as seasonal autoregressive fractional integrated moving average (SAFRIMA) and seasonal autoregressive fractional integrated moving average with exogenous factor (SARIMAX), are all derived from the autoregressive integrated moving average (ARIMA) model (Yaya and Fashae 2015, Cowpertwait and Metclafe 2009). ARIMA is a stochastic mathematical procedure developed by Box-Jenkins for forecasting, utilizing best-fit and deviation terms from past time series data. It combines statistical techniques such as autoregression, differentiation, and moving average. Unlike the basic ARIMA model, SARIMA incorporates seasonal and non-seasonal properties suitable for stationary and nonstationary data with a time-series cycle. As described by Box and Jenkins (1976), the expression for the seasonal ARIMA(p, d, q) model is SARIMA $(p, d, q)(P, D, Q)_s$, where p represents the non-seasonal autoregressive order, d represents the nonseasonal differencing order, and p represents the nonseasonal moving average order. P, D, and Q respectively indicate the seasonal autoregressive order, seasonal differencing order, and seasonal moving average order, while the parameter s represents the periodicity of the time series.

Given the sequence $\{Y_t : t\in (0, \pm 1, \pm 2, \ldots)\}$, the autoregressive (AR) and the moving average (MA) components for an event Y_t at time t can be expressed as:

$$Y_t + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_p Y_{t-p} = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \ldots + b_q e_{t-q}. \quad (6)$$

The combination of both AR and MA results in:

$$Y_t = e_t + (b_1 e_{t-1} - a_1 Y_{t-1}) + (b_2 e_{t-2} - a_2 Y_{t-2}) + \ldots + \left(b_q e_{t-q} - a_p Y_{t-p}\right). \quad (7)$$

In Equation 6, a_1, a_2, …, a_p represent the coefficients associated with the AR component, while $b_1, b_2, \ldots, b_q$ represent the coefficients of the MA component. Additionally, e_t represents white noise, which follows a normal distribution with a zero mean and variance σ^2, such that $\mathrm{cov}\,(\mathrm{e_t}, \mathrm{e_{t-k}}) = 0\forall \mathrm{k}\neq 0$, where $k = 1, 2, 3, \ldots$.

By applying differencing, which transforms the time series model from the autoregressive moving average (ARMA) to the autoregressive integrated moving average (ARIMA), any stationary component (if present) in the sequence is removed. Equation 7 can then be rewritten as:

$$Y_t^d = e_t + \left(a_1 Y_{t-1}^d + b_1 e_{t-1}\right) + \left(a_2 Y_{t-2}^d + b_2 e_{t-2}\right) + \ldots + \left(a_p Y_{t-p}^d + b_q e_{t-q}\right). \quad (8)$$

Furthermore, the general form of the ARIMA equation can be expressed as:

$$Y_t^d = e_t + \sum_{i=1}^{p} \varphi Y_{t-p}^d + \sum_{i=1}^{q} \theta e_{t-p}, \quad (9)$$

where $\varphi = \left\{a_1, a_2, \ldots, a_p\right\}$ and $\theta = \left\{b_1, b_2, \ldots, b_q\right\}$.

To obtain the general seasonal ARIMA (SARIMA) model, we apply the backshift operator B and the difference operator ∇ to the ARIMA model. Similar to Equation 6 we have:

$$\varphi\,(B)\,\Phi\,(B)\,(1 - B^s)^D\,(1 - B)^d\,Y_t = \theta\,(B)\,\Theta\,(B)\,e_t, \quad (10)$$

where

$$\varphi\,(B) = \left(1 - \varphi_1 B - \varphi_2 B^2 - \ldots, \varphi_p B^p\right)\ \text{(non−seasonalAP)}, \quad (11)$$

$$\Phi\,(B) = \left(1 - \Phi_1 B - \Phi_2 B^2 - \ldots, \Phi_p B^p\right)\ \text{(seasonalAP)}, \quad (12)$$

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \ldots, \theta_q B^q\right) \text{ (non−seasonalAP)}, \tag{13}$$

$$\Theta(B) = \left(1 - \Theta_1 B - \Theta_2 B^2 - \ldots, \Theta_q B^q\right) \text{ (seasonalAP)}. \tag{14}$$

Both $(1 - B^s)^D$ and $(1 - B)^d$ on the left-hand side of Equation 10 represent the orders of seasonal differencing and non-seasonal differencing, respectively.

For realistic forecasts, two fundamental requirements were set for this study in choosing SARIMA models: predictions must be nonnegative and have lowest predicted values near zero, indicating dry-season rainfall behavior.

The seasonal autoregressive integrated moving average (SARIMA) model was implemented using Minitab®, Python, and R to forecast monthly rainfall across Nigeria. While SARIMA follows a well-defined algorithm and should theoretically yield identical results across platforms, observed variations were primarily attributed to differences in default parameter initialization, optimization techniques, and the handling of seasonal differencing within software implementations. For instance, the auto.arima() function in R and Python employs maximum likelihood estimation but uses distinct search strategies and convergence criteria, which can result in different model selections if parameters are not manually constrained. To ensure consistency, we manually specified all model parameters and tuned optimization settings across platforms. This approach minimized discrepancies and ensured that any differences in outputs were related to software-level algorithmic implementations rather than the SARIMA methodology itself.

2.3. Machine Learning Algorithms

The next four machine learning models were utilized in this research for univariate rainfall prediction: adaptive neuro-fuzzy inference system (ANFIS), feedforward neural networks (FNN), support vector regression (SVR) and random forest (RF).

A Sugeno-type ANFIS was implemented in MATLAB®, with five layers: fuzzification, rule, normalization, defuzzification, and output layers. Gaussian membership functions (MFs) were used, three MFs per input variable, which lead to nine fuzzy inference rules. A hybrid optimization scheme, the least-squares estimation--backpropagation gradient descent, was used during training. The model was trained for 100 epochs with a training error tolerance of $1e-5$. Similarly, a three-layer feedforward neural network implemented on MATLAB using the backpropagation algorithm. The network architecture comprised an input layer (serial number, month, and year as features), a hidden layer of 10 units, and a single output neuron representing the predicted rainfall. Levenberg-Marquardt was utilized as the algorithm for optimization with a learning rate of 0.01. Training was done for 200 epochs using mean squared error as the performance measure.

SVR was applied in R with the aid of the e1071 package. A linear basic function kernel was employed, while hyperparameters were optimized with grid search and 5-fold cross-validation. The most important parameters were cost (C = 1.0), epsilon

($\varepsilon = 0.1$), and gamma ($\gamma = 0.01$). Input data were preprocessed to standardize prior to training to enhance model stability. The RF model was also implemented in R. It used 200 decision trees, with 2/3 of the input data reserved for validation during training, and considered 3 randomly selected variables at each split. In addition to the three input features, all the models were trained and validated on the split dataset (78% for training, 1960–1999 and 22% for testing, 2000–2010).

3. Results

3.1. Model Formats

3.1.1. Minitab SARIMA Models

Various SARIMA models were used across Nigeria's regions, considering rainfall seasonality and also software variability. The northern places like Katsina, Sokoto, Kano, Yelwa, and others needed longer periodicities – essentially 24 months, although Nguru needed a 36-month periodicity – because of long dry seasons. Alternatively, all the southern places and part of the northern places were near perfection with a 12-month periodicity (Table 2). The north Kaduna and south Oshogbo alone used identical SARIMA models in Minitab, but the order of differencing (d) for nonseasonal was usually 0, although in a few instances $d = 1$.

Some sites, such as Yelwa, Jos, Minna, and Ikeja, needed 24-month models with $d = 0$ or 12-month models with $d = 1$, depending on the local rainfall regime. Ikeja, for example, has an August break and Jos is a mountainous area. Model performance was validated using performance measures. In Calabar, the $(2, 0, 1)\,(1, 1, 0)_{12}$ model forecasted low at 7.3 mm and was eliminated based on low r^2 (0.48) and high *RMSE* (134.4 mm). An improved model, $(4, 0, 1)\,(1, 1, 1)_{12}$, had a higher minimum (31.3 mm) but with better performance statistics ($r^2 = 0.67$; $RMSE = 104.9$ mm), highlighting that realism in forecasting must be weighed against accuracy.

3.1.2. Python SARIMA Models

It was found that the majority of the stations provided unique SARIMA models in Python, indicating the inefficiency of the automated process, which was mainly using $d = 0$ and $D = 0$ and not providing consistent forecasts. A trial-and-error process was thus used to make the model more appropriate (Table 2). In some instances, the auto process resulted in errant warnings of nonstationary seasonal autoregressive parameters. This was offset by specifying the seasonal differencing order (D) as 1, a move that dominated in Lokoja, Makurdi, and Enugu – the three towns with the same latitudes (7.47°N, 7.44°N, and 6.28°N) and the same models. So too was the same model, $(1, 1, 1)\,(1, 1, 1)_{12}$, used by Ilorin and Ibi (8.29°N and 8.11°N), though their results were the outcomes of manual adjustment and not automation. A strange but working solution was seen for Calabar, with $d = 2$ being superior to normal values of 0 or 1.

Table 2. The most suitable SARIMA $(p, d, q)\,(P, D, Q)_s$ models for rainfall forecasting in Nigeria

Locations	Minitab Models	Python Models	R Models
Katsina	$(1,0,1)\,(2,0,1)_{24}$	$(2,1,0)\,(2,0,1)_{12}$	$(0,0,1)\,(1,0,1)_{12}\,*$
Sokoto	$(1,1,0)\,(1,1,0)_{24}$	$(2,1,0)\,(3,1,1)_{12}$	$(0,0,0)\,(1,0,1)_{12}\,*$
Nguru	$(1,1,0)\,(1,1,1)_{36}$	$(2,0,1)\,(0,1,1)_{12}$	$(0,0,1)\,(1,0,1)_{12}\,*$
Kano	$(0,1,0)\,(0,1,1)_{24}$	$(0,1,0)\,(1,1,1)_{12}$	$(2,0,0)\,(0,1,1)_{12}$
Gusua	$(1,0,1)\,(1,1,0)_{24}$	$(1,1,1)\,(1,1,2)_{12}$	$(0,0,1)\,(2,0,1)_{12}$
Maiduguri	$(1,0,1)\,(1,1,0)_{12}$	$(2,1,0)\,(1,1,1)_{12}$	$(0,0,0)\,(1,0,1)_{12}\,*$
Yelwa	$(1,0,1)\,(1,1,1)_{24}$	$(0,0,1)\,(3,0,1)_{12}$	$(0,0,0)\,(0,1,1)_{12}$
Kaduna	$(2,0,1)\,(1,1,0)_{12}$	$(0,0,1)\,(1,0,1)_{12}\,\blacktriangledown$	$(0,0,1)\,(2,1,0)_{12}$
Bauchi	$(1,1,0)\,(2,1,0)_{24}$	$(1,0,1)\,(1,0,1)_{12}$	$(0,0,1)\,(1,1,0)_{12}$
Jos	$(0,1,0)\,(3,1,2)_{24}$	$(1,0,0)\,(1,0,1)_{12}\,\blacktriangledown$	$(0,0,0)\,(1,0,1)_{12}\,*$
Minna	$(0,1,2)\,(0,1,1)_{12}$	$(0,0,1)\,(1,0,1)_{12}\,\blacktriangledown$	$(0,0,1)\,(1,0,1)_{12}\,*$
Yola	$(2,0,2)\,(1,1,0)_{24}$	$(2,1,1)\,(2,1,2)_{12}$	$(0,0,1)\,(2,0,1)_{12}\,*$
Ilorin	$(2,0,2)\,(1,1,1)_{12}$	$(1,1,1)\,(1,1,1)_{12}$	$(0,0,2)\,(2,0,1)_{12}\,*$
Ibi	$(1,0,1)\,(1,1,1)_{24}$	$(1,1,1)\,(1,1,1)_{12}$	$(0,0,0)\,(2,0,1)_{12}\,*$
Lokoja	$(2,0,2)\,(3,1,2)_{24}$	$(1,0,0)\,(2,1,1)_{12}\,+\blacktriangledown$	$(0,0,2)\,(2,1,0)_{12}$
Makurdi	$(1,0,1)\,(0,1,1)_{12}$	$(1,0,0)\,(2,1,1)_{12}\,+$	$(0,0,1)\,(2,1,0)_{12}$
Osogbo	$(2,0,1)\,(1,1,0)_{12}$	$(1,0,0)\,(1,0,1)_{12}\,\blacktriangledown$	$(0,0,1)\,(2,0,1)_{12}\,*$
Ibadan	$(3,0,0)\,(1,1,0)_{12}$	$(3,1,1)\,(3,1,0)_{12}$	$(0,0,1)\,(2,0,1)_{12}\,**$
Ikeja	$(1,1,1)\,(0,1,1)_{12}$	$(1,0,0)\,(1,0,2)_{12}$	$(1,0,0)\,(2,0,1)_{12}\,*$
Enugu	$(3,0,2)\,(1,1,1)_{12}$	$(1,0,0)\,(2,1,1)_{12}\,+\blacktriangledown$	$(0,0,0)\,(2,1,0)_{12}$
Benin	$(1,0,0)\,(0,1,1)_{12}$	$(1,0,0)\,(1,0,2)_{12}$	$(0,0,1)\,(2,0,1)_{12}\,*$
Warri	$(2,0,1)\,(0,1,1)_{12}$	$(1,0,1)\,(1,1,1)_{12}$	$(1,0,0)\,(1,1,0)_{12}$
Calabar	$(4,0,1)\,(1,1,1)_{12}$	$(2,2,2)\,(1,1,1)_{12}$	$(0,0,0)\,(2,0,1)_{12}\,*$
Port Harcourt	$(1,0,1)\,(2,1,1)_{12}$	$(1,0,0)\,(1,0,1)_{12}\,\blacktriangledown$	$(0,0,1)\,(2,0,1)_{12}\,*$

Notes:

▼: Model obtained using the auto.arima() procedure in Python.

+▼: Non-stationarity warning raised by the automated model was resolved by setting $D = 1$ in Python.

∗: Model obtained using the modified auto.arima() in R with D set to 0.

∗∗: The auto.arima() in R initially led to rapid convergence, which was mitigated by adjusting Q to 1.

3.1.3. R SARIMA Models

Despite cases where the automatic SARIMA procedure in R produced low minimum values and wide ranges, modified models often yielded better statistical performance. This modified approach was not restricted to any region, although two southern locations – Ibadan and Warri – displayed unique behaviors. For Ibadan, only the seasonal moving average order (Q) was adjusted, while Warri's modified model resulted in negative forecasts, prompting the use of the original auto.arima() output, albeit with high forecast errors. The automatic procedure performed effectively at 9 of the 24 sites – Kano, Gusau, Bauchi, Yelwa, Kaduna, Lokoja, Makurdi, Warri, and Enugu – with Kaduna and Makurdi sharing the same model configuration (see Table 2). Additionally, two modified models were successfully applied across multiple locations: $(0, 0, 1)\,(1, 0, 2)_{12}$ for Katsina, Nguru, and Minna (north), and $(0, 0, 1)\,(2, 0, 1)_{12}$ for Ibadan, Benin, and Port Harcourt (south). Parameter ranges were generally between 0 and 2, with $d + D$ rarely exceeding 1 in R, unlike Minitab or Python. R was the most user-friendly platform, avoiding trial-and-error and offering reliable, straightforward modelling.

3.2. Forecast Proficiency

Table 3 presents the results of standard statistical measures, such as the coefficient of variation, mean bias error, and root-mean-square error. In contrast, Table 4 shows the outputs of less common statistics, including the percentage error at maximum values ($PE_{\max}$) and the absolute difference between the minimums of the actual and forecasted values ($|\Delta_{\min}|$). Some results exhibit a systematic pattern, while others do not follow any discernible rhythm.

Table 3. Statistical results of SARIMA modelling using Minitab, Python, and R software

Location	Minitab			Python			R		
	r^2	RMSE (mm)	MBE (mm)	r^2	RMSE (mm)	MBE (mm)	r^2	RMSE (mm)	MBE (mm)
Katsina	0.73	54.56	23.20	0.72	52.79	19.88	0.71	54.76	21.85
Sokoto	0.50	66.68	-1.81	0.67	51.94	5.16	0.67	52.58	7.17
Nguru	0.81	25.42	-0.37	0.83	23.01	-0.39	0.83	23.20	2.00
Kano	0.73	88.60	33.39	0.74	91.71	38.00	0.72	81.28	-7.40
Gusau	0.79	50.81	0.35	0.85	42.02	-1.36	0.83	45.20	1.65
Maiduguri	0.64	62.42	-10.95	0.76	42.94	3.74	0.74	44.28	4.53
Yelwa	0.79	48.72	4.15	0.79	48.51	3.09	0.79	48.93	-3.93
Kaduna	0.69	65.88	-5.94	0.76	55.98	-4.23	0.70	63.57	-6.31
Bauchi	0.66	73.63	-3.67	0.70	66.42	8.15	0.62	86.88	-19.19
Jos	0.77	54.95	-0.66	0.79	51.53	4.92	0.79	51.50	3.99
Minna	0.80	52.42	-4.11	0.80	52.70	7.11	0.80	52.75	6.11
Yola	0.65	53.75	-9.89	0.76	40.28	-3.46	0.76	40.16	-3.87
Ilorin	0.59	60.48	-4.18	0.59	60.33	-3.22	0.59	60.38	-5.97
Ibi	0.69	50.75	-6.56	0.70	50.00	-5.66	0.71	49.62	-6.01
Lokoja	0.62	61.60	0.26	0.63	60.83	0.85	0.54	81.53	-21.81
Makurdi	0.72	52.19	-9.66	0.72	51.59	-1.36	0.67	79.63	-38.33
Osogbo	0.46	72.56	-9.79	0.65	50.23	-0.11	0.65	50.06	-0.81
Ibadan	0.45	81.79	-20.54	0.51	67.28	-8.56	0.58	60.09	4.92
Ikeja	0.44	87.28	3.15	0.44	88.00	2.63	0.44	87.75	0.77
Enugu	0.78	68.34	4.34	0.78	64.02	7.37	0.71	72.93	12.42
Benin	0.61	97.30	20.46	0.61	97.03	19.34	0.62	96.84	17.56
Warri	0.60	114.60	-8.38	0.61	113.61	-9.50	0.28	182.72	-8.36
Calabar	0.67	104.88	3.30	0.67	105.00	-0.80	0.67	104.17	2.95
Port Harcourt	0.70	79.00	2.50	0.70	78.78	-3.59	0.70	78.86	-5.89

Table 4. Maximum forecasted rainfall, absolute differences between minimum values, and percentage errors at maximum points for SARIMA models in Nigeria

Location	Minitab			Python			R		
	Maximum forecasted (mm) rainfall (mm)	$\lvert\Delta_{\min}\rvert$	$PE_{\max}$	Maximum forecasted (mm) rainfall (mm)	$\lvert\Delta_{\min}\rvert$	$PE_{\max}$	Maximum forecasted (mm) rainfall (mm)	$\lvert\Delta_{\min}\rvert$	$PE_{\max}$
Katsina	129.7	0.8	58.8	123.0	1.6	60.9	119.4	1.3	62.1
Sokoto	283.5	0.0	21.6	214.3	1.4	40.7	201.7	0.4	44.2
Nguru	197.9	0.0	23.6	162.3	2.0	37.3	158.1	0.6	38.9
Kano	308.8	0.0	48.9	292.5	0.0	51.6	433.6	3.4	28.2
Gusau	379.6	0.0	17.3	301.7	1.8	34.3	324.2	0.2	29.4
Maiduguri	303.5	0.2	18.3	199.8	2.1	46.2	193.6	1.2	47.9
Yelwa	251.3	0.7	36.7	285.8	0.0	28.0	295.3	1.0	25.6
Kaduna	345.1	0.0	25.4	301.4	0.0	45.1	327.1	0.1	29.3
Bauchi	347.0	0.0	23.9	288.8	0.0	44.6	407.0	0.8	21.9
Jos	296.8	0.0	33.4	254.0	0.1	37.4	253.3	1.2	37.6
Minna	281.5	0.0	43.4	264.8	0.2	46.8	264.3	1.6	46.8
Yola	298.4	0.0	-11.5	204.5	0.0	23.6	207.2	0.1	22.5
Ilorin	234.6	0.5	36.6	230.1	3.5	37.9	230.0	4.8	37.9
Ibi	216.8	0.3	36.0	208.4	0.1	38.5	204.7	1.5	39.6
Lokoja	239.5	0.1	47.9	220.8	0.1	51.9	343.2	1.0	25.3
Makurdi	231.4	0.3	34.4	224.7	0.8	36.3	359.8	1.3	-0.02
Osogbo	304.2	1.3	13.6	207.5	7.8	41.0	203.8	10.3	42.1
Ibadan	319.6	2.7	2.7	227.4	2.7	30.8	174.2	2.2	47.0
Ikeja	299.4	10.6	35.4	306.6	12.7	33.8	303.0	17.9	34.6
Enugu	284.8	5.9	35.1	289.5	4.3	34.1	319.0	0.6	27.3
Benin	359.5	12.7	51.8	361.9	14.5	51.5	357.8	20.7	52.0
Warri	482.5	2.3	39.4	489.2	26.8	38.5	622.9	13.9	21.7
Calabar	456.2	31.1	43.6	476.5	20.0	41.0	450.8	32.1	44.2
Port Harcourt	356.7	14.4	36.0	358.6	19.4	35.7	365.0	26.1	34.5

3.2.1. Correlating Measures

Coefficient of determination for SARIMA predictions shows higher precision in northern Nigeria compared to the south on all three platforms – Minitab, Python, and R. Theoretically, r^2 should be over 70%; but whereas the majority of northern sites can achieve this value (e.g., Gusau: 0.85 in Python; Nguru and Gusau: 0.83 in R), few southern sites can manage to do so. Ikeja, for example, possesses a lowest of r^2 levels (Python 0.44, ~ 40% on all of them) and Enugu and Port Harcourt are in the south, concurring at a consistent 70%. Ilorin, which is in the middle, has a consistent r^2 of 0.59 on all platforms (Table 3).

RMSE is also in the opposite direction – lower in the north, larger in the south. Warri always displays the largest *RMSE* (as high as 183 mm in R), while Nguru in the north displays the smallest (~ 23 mm). Calabar is yet another town in the south displaying *RMSE* greater than 100 mm (Table 3). The differences clearly indicate that

SARIMA models are not quite efficient in predicting extremes in rainfall, especially in the south.

Table 4 shows the way minimum forecast values ($|\Delta_{\min}|$) increase as latitude decreases, indicating more extreme southern rainfalls. Though no model-predicted rain had negative values, more zero errors were provided by Minitab, while R was less so. The maximum $||\Delta_{\min}|$ values were 31.1 mm (Minitab, Calabar), 26.8 mm (Python, Warri), and 32.1 mm (R, Calabar), which show room for model improvement in the detection of seasonal extremes.

3.2.2. Uncorrelated Measures

MBE was applied in the determination of the direction of the forecasting error by station. Over-forecasting is indicated by positive *MBE* values and under-forecasting by negative values. From the Table 3 results via Minitab, there was under-forecasting at 14 stations and over-forecasting at 10, ranging from −20.54 mm at Ibadan to 33.39 mm at Kano. Python outputs also performed with an equal split: 12 over-forecasting and 12 under-forecasting sites, from −9.50 mm (Ibadan) to 38.00 mm (Kano). R software also gave 12 positive and 12 negative *MBE* values, ranging from −38.33 mm (Makurdi) to 21.85 mm (Katsina). The geographical trend shows increasing over-forecasting in the north and increasing under-forecasting in the south.

In comparing models' performance in predicting extreme rainfall events, $PE_{\max}$ was examined (Table 4). $PE_{\max}$ values were mainly positive, which reflects over--prediction of extreme rainfall, with only two instances of negative values (Yola from Minitab and Makurdi from R). Katsina consistently recorded the highest $PE_{\max}$ on each platform: 58.8 (Minitab33), 60.9 (Python), and 62.1 (R), and it reflects over-prediction in this northern state.

3.3. Intelligent Systems Failure

Univariate rainfall prediction with a range of different neural network (NN) configurations was unsatisfactory. Models generated implausible or constant negative outputs throughout the 11-year forecast period even with more complex configurations like a [4 6 1] FNN with log-sigmoid, radial basis, and linear transfer functions. Splitting into 70% training, 15% validation, and 15% test, but no suitable nonnegative outputs were given by any of the time-series-based networks (e.g., Elman, NAR, NARX). All these findings indicate that the neural network may be more appropriate for year or multivariate rainfall modelling. RF gave more accurate seasonal predictions but created the same result annually and is therefore only applicable to yearly research. SVR also failed as it was unable to identify seasonal rain patterns. Amongst its kernels, the linear function came closest but the predicted curve was in a linear upward trend from January to December, unrealistic when compared to the typical rainfall curve of Nigeria. In the same way, ANFIS models could not deliver realistic results, mostly producing negative rainfall values (Table 5). Such outcomes are consistent with previous studies that postulate that intelligent systems may perform better under multivariate, hybrid, or yearly forecasting scenarios compared to monthly univariate time-series use (Darji et al 2015, Wu et al 2015).

Table 5. Limitations observed in using ML models for univariate rainfall forecasting

ML model	Software package	Remarks
ANFIS	MATLAB	Produced negative rainfall forecasts.
Neural networks (FNN, time delay, distributed delay, layer recurrent, Elman, NAR, NARX, etc.)	MATLAB	Generated negative values forecasts and large prediction errors in certain cases.
SVR	R, MATLAB, Python	Failed to capture the seasonal variability of rainfall effectively.
RF	R	Returned identical yearly values over an 11-year period.

4. Discussion

Based on results from Minitab, vegetation and geography exert strong influences on SARIMA model parameters for Nigerian locations. Nguru (Sahel Savanna) in Minitab contained 36-month models, whereas the majority of the Sudan Savanna locations utilized 24-month models. Ibi and Lokoja, situated mid-distance between Guinea and Derived Savanna locations, utilized 24-month models. There were some areas within the Lowland Rain Forest zone such as Calabar and Benin that had high non-seasonal autoregressive orders, whereas areas such as Osogbo and Warri had lower orders although they were in the same zone, indicating there was erratic vegetation influence on model structure.

Model differences existed among different software packages as well. In Python, for example, areas within the Derived Savanna – Lokoja, Makurdi, and Enugu – had the same SARIMA models. The pattern took place in places within Guinea Savanna, with the exception of Minna, which was fitted with auto.arima(). In R, height was a viable variable influencing model performance, especially in Warri (6.1 m), where worse model precision and the occurrence of negative forecasts corresponded with little altitude, whereas in Ibadan, which had more altitude, results were more reliable.

Rainfall variables greatly influenced forecasting performance. For example, Ikeja's peculiar "August break" behavior resulted in poor performance on all three SARIMA model platforms (Fig. 2). Conversely, arid regions north of the nation such as Nguru, Gusau, and Minna had consistent performance. Nguru was remarkable, with best forecast performance in both Minitab and R, and second-best in Python – indicating its consistently low values for rainfall (Table 1). Conversely, Ilorin alone among northern stations received non-zero 1st quantile rainfall and performed worst among northern stations. Warri and Calabar both recorded high $|\Delta_{min}|$ errors, indicative of their high

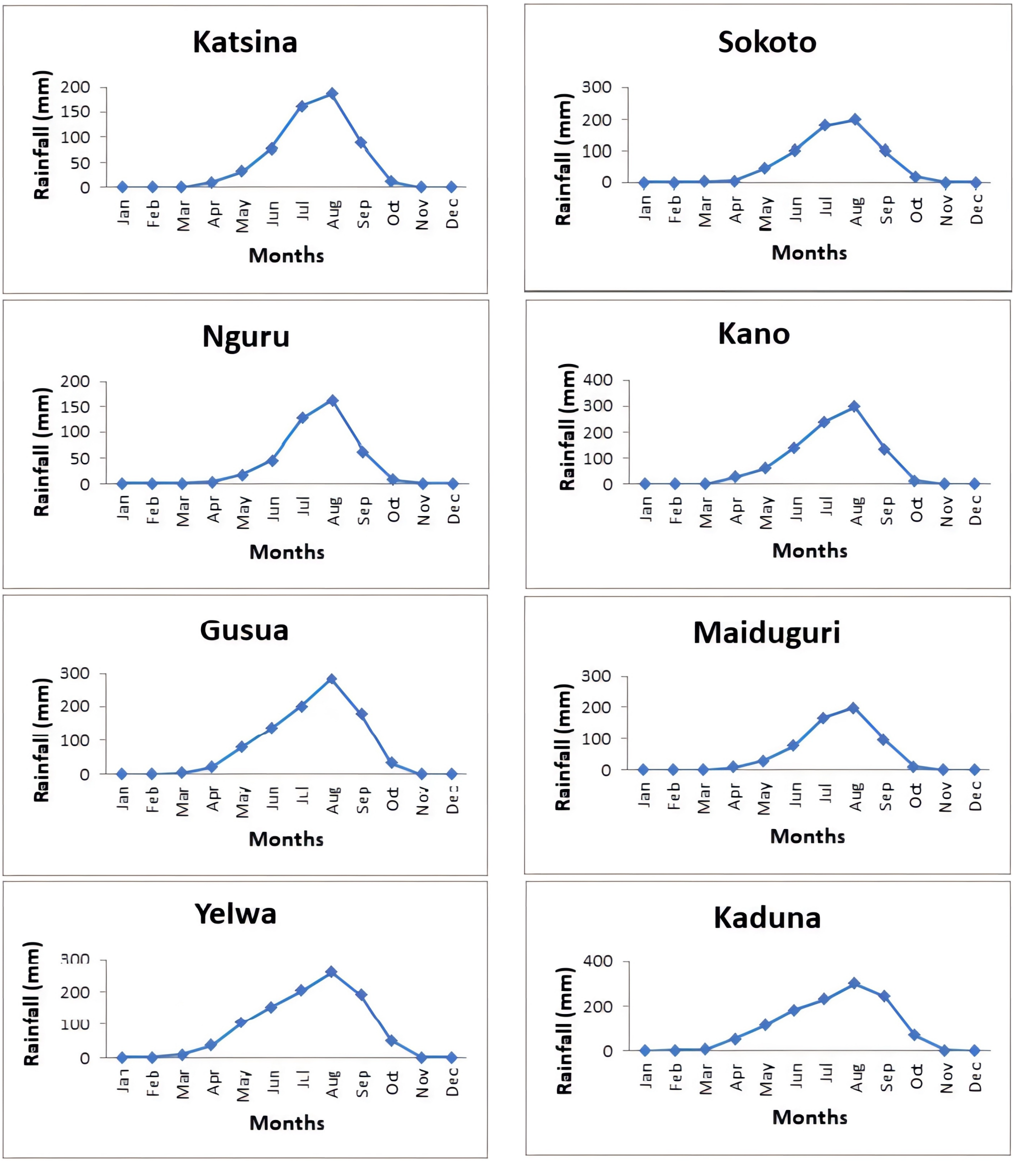

Fig. 2. Rainfall pattern of northern sites considered in this study

rainfall even in lowest quartiles. Yola also recorded the lowest $PE_{\max}$ in Minitab and Python and fourth lowest in R, indicative of its susceptibility to spasmodic extreme rainfall events (Table 4). In contrast, Katsina always had the highest $PE_{\max}$ values on all platforms, reflecting a trend towards extreme dryness. Furthermore, this research affirms previous studies on the inability of conventional models of forecasting to accurately capture sudden changes in time-series data. For instance, estimated rainfall maxima at two points in the Savanna were higher than actual, a demonstration of how

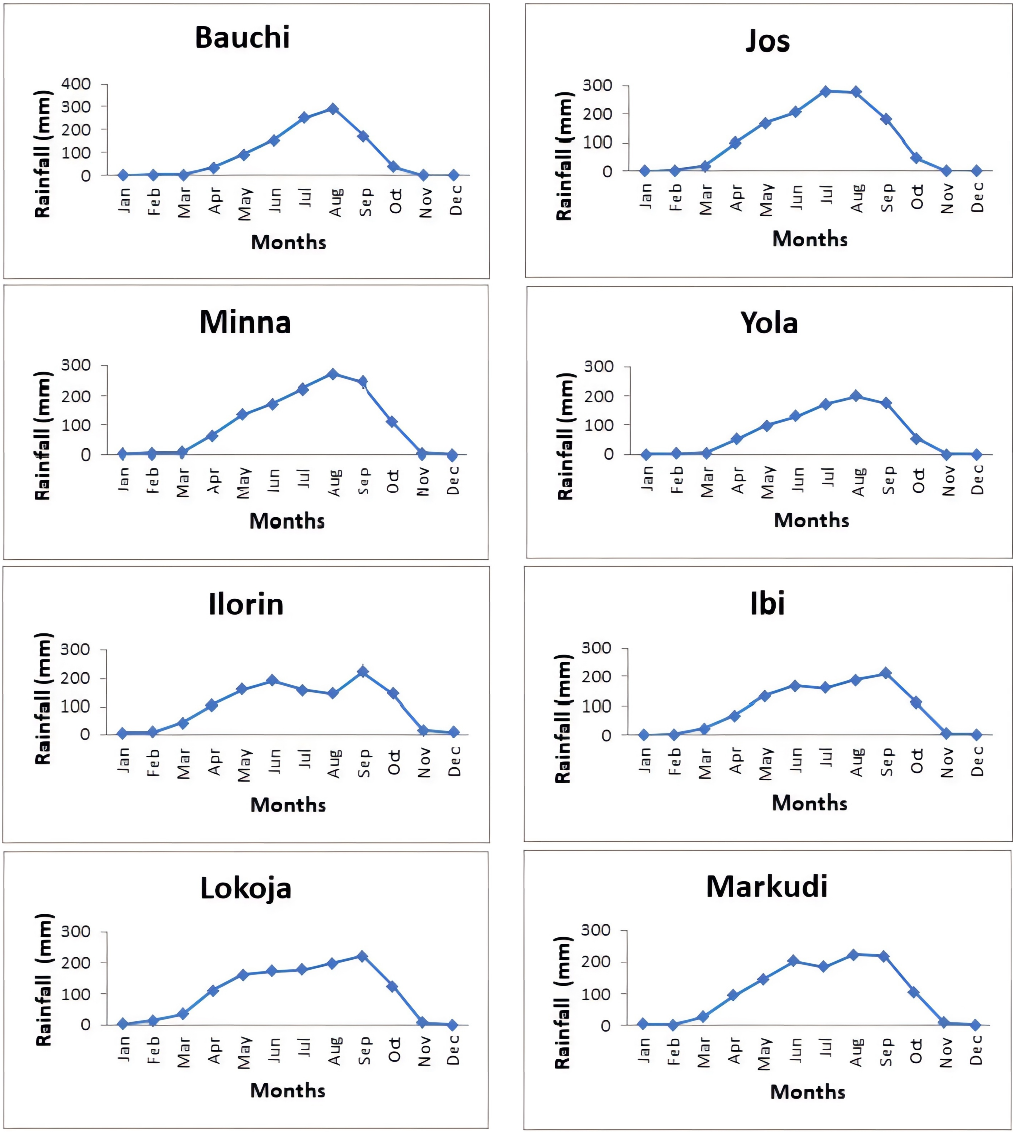

Fig. 2. (continued)

prone the areas are to rainfall extremes. Moreover, the interconnectedness of climate systems also means that such extremes have more regional effects.

The study incorporated the potential influence of solar activity on rainfall. The 2000–2010 period was chosen as the forecast interval to encompass a full solar cycle, covering the peak of Solar Cycle 23 (around 2001) and the trough of Solar Cycle 24 (around 2008). Solar cycles, which average about 11 years in duration, affect the Earth's atmosphere through sunspots, solar flares, and variations in magnetic flux

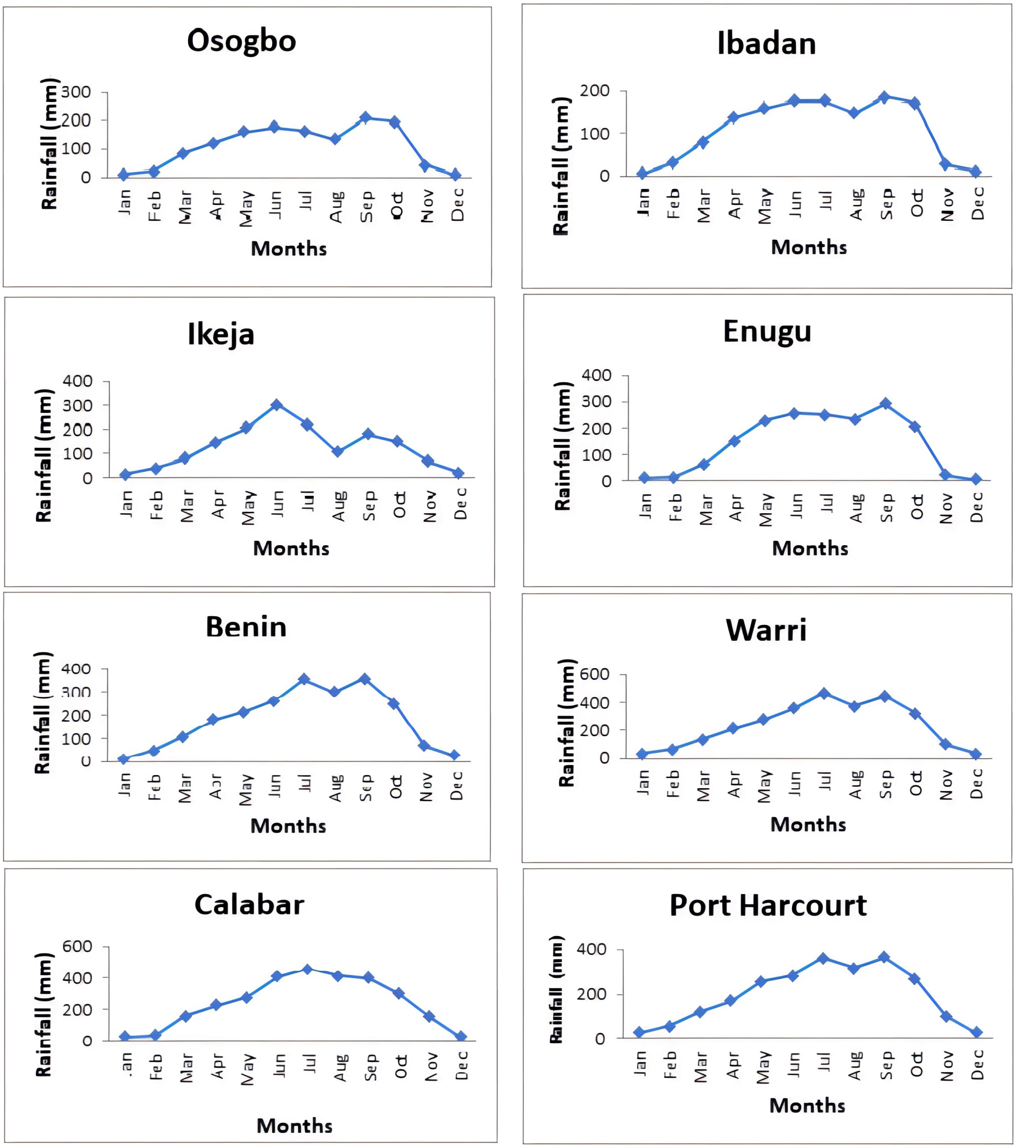

Fig. 3. Rainfall pattern of southern sites considered in this study

– factors that can influence rainfall patterns, heatwaves, and sea-level fluctuations (Somaila et al 2022, Usoskin et al 2021, Budikova 2009, Schaack et al 1990). These effects are particularly significant in tropical regions, where solar radiation is most intense and interactions between tropical heat and polar cold help shape global climate dynamics.

Due to differences in statistical software, experimental procedures, and methodologies, direct comparison with other studies may be inappropriate. A somewhat similar study by Ojo and Ogunjo (2022) used ML models (ANN, SVR, ANFIS) with geographic variables to predict rainfall in 16 Nigerian locations (1983–2013), but their work lacked time-series inputs, used fewer sites, and did not split data into training and testing sets – an essential step in ML evaluation. Despite these limitations, their study supports ML's superiority over linear regression, identifying ANFIS as most effective. Unlike ML models, regression models show consistent performance, eliminating the need to test SARIMA models on the development dataset (Obot 2024, Obot et al 2019). Changes in rainfall measurement tools have improved coverage but introduced data consistency issues. Though this study uses older data, the findings remain valid and suggest the need for future work using broader, more recent datasets.

5. Conclusion

Seasonal rainfall forecasting was carried out over 11 years across 24 locations in Nigeria using SARIMA models and machine learning algorithms. The analysis used 40 years of monthly data, with missing values estimated by averaging nearby years. Negative rainfall predictions were considered invalid, even during the dry season. Alongside standard evaluation metrics such as r^2, $RMSE$, and MBE, additional measures like percentage error at maximum rainfall and the absolute difference at minimums were used to assess prediction accuracy. SARIMA models were compared across platforms including Minitab, Python, and R. Machine learning methods, such as FNN, time series models, ANFIS, SVR, and RF, were primarily implemented using MATLAB and R.

Knowledge based intelligent systems such as NN, RF, ANFIS, and SVR show limitations in univariate rainfall forecasting, especially in capturing seasonal variations. Custom trained models like subset-based RF model performed reasonably but often returned identical yearly values, while SVR failed to reflect seasonal patterns. Automated SARIMA models in R and Python occasionally produced pessimistic outputs due to limitations in default settings such as auto.arima(). Forecast accuracy improved by adjusting the seasonal differencing order, for example by modifying d or D in Python or d and Q in R, highlighting the importance of fine tuning for nonstationary data. Platform specific differences were observed as R, Python, and Minitab behaved uniquely across vegetation zones and elevation profiles, especially in southern regions. Rainfall characteristics also influenced model performance; areas with lower rainfall variability such as Nguru yielded better forecasts than highly variable regions like Ikeja, where extreme events such as the August break reduced accuracy. Occasionally, forecasted values exceeded measured ranges, underscoring the challenges of modeling highly variable rainfall patterns.

Declaration of competing interest

The authors declare that there are no conflicts of interest, financial or otherwise, associated with this manuscript.

Acknowledgement

The authors are extremely grateful to NIMET for the data used in this study.

References

Akinbobola A., Okogbue E.C., Ayansola A.K. (2018) Statistical modeling of monthly rainfall in selected stations in forest and savannah eco-climatic regions of Nigeria, *Journal of Climatology & Weather Forecasting*, **6**, 226. https://doi:10.4172/2332-2594.1000226.

Al-Taai O.T., Abbood Z.M. (2020).Analysis of the convective available potential energy by precipitation over Iraq using ECMWF data for the period of 1989–2018, *Scientific Review Engineering and Environmental Sciences*, **29**(2), 196–11. https://doi.org/10.22630/pniks.2020.29.2.17.

Aminudin N.A.I., Yunus N.H.M., Basarudin H., Ramli A.F., Nadzir M.S.M., Sampe J., Hasan N. (2024) Analysis of rainfall distribution in Malaysia through the employment of hydro-estimator data, *Engineering, Technology & Applied Science Research*, **14**(5), 16680–16685. https://doi.org/10.48084/etasr.7601.

Azad A., Manoochehri M., Kashi H., Farzin S., Karami H., Nourani V., Shiri J. (2019) Comparative evaluation of intelligent algorithms to improve adaptive neuro-fuzzy inference system performance in precipitation modelling, *Journal of Hydrology*, **571**, 214–224. https://doi.org/10.1016/j.jhydrol.2019.01.062

Bhaga T.D., Dube T., Shekede M.D., Shoko C. (2020) Impacts of climate variability and droughts on surface water resources in sub-Saharan Africa using remote sensing: a review, *Remote Sensing*, **12**, 4184. https://doi:10.3390/re/2244184.

Box G.E.P., Jenkins G.M. (1976) *Time series analysis: forecasting and control*, Holden-Day, San Francisco.

Budikova D. (2009) Role of arctic sea ice in global atmospheric circulation: a review, *Global and Planetary Change*, **68**, 149–163. https://doi.org/10.1016/j.gloplacha.2009.04.001.

Cowpertwait P.S.P., Metclafe A.V. (2009) *Introductory time series*, R. Springer.

Darji M.P., Dabhi V.K., Prajapati H.B. (2015) Rainfall forecasting using neural network: a survey, *2015 International Conference on Advances in Computer Engineering and Applications*, IEEE, 706–713. DOI: 10.1109/ICACEA.2015.7164782.

Gupta M., Kour S., Bharat R., Sharma B.C., Kachroo D. (2022) Impact of agri-horti landuse system on farm profitability and sustainability under rainfed conditions, *Journal of Soil and Water Conservation*, **21**, 107–113.

Ibrahim I.D., Hamam Y., Alayli Y., Jamiru T., Sadiku E.R., Kupolati W.K., Ndambuki J.M., Eze A.A. (2021) A review on Africa energy supply through renewable energy production: Nigeria, Cameroon, Ghana and South Africa as a case study, *Energy Strategy Reviews*, **38**, 100740. https://doi.org/10.1016/j.esr.2021.100740.

Igbawua T., Ujoh F., Mkighirga S.K., Adagba G. (2024) Assessment of relationship between sea surface temperature (SST) changes and precipitation types in Nigeria from 2000 to 2022, *Results in Earth Sciences*, **2**, 100031. https://doi.org/10.1016/j.rines.2024.100031.

Ikpang I.N., Okon E.O.J., George E.U. (2022) Modeling average rainfall in Nigeria with artificial neural network (ANN) models and seasonal autoregressive integrated moving average (SARIMA) models, *International Journal of Statistics and Probability*, **1**1(4), 53–62. https://doi.org/10.5539/ijsp.v11n4p53.

Ilaboya I. (2019) Performance of multiple linear regression (MLR) and artificial neural network (ANN) for the prediction of monthly maximum rainfall in Benin City, Nigeria, *International Journal of Engineering Science and Application*, **3**, 21–37.

Isnaini M.N., Sari J., Wardhani K.D.K., Wahyuni R.T. (2025) Recommendation system for mobile-based oil palm fertilization period with rainfall prediction using ANN, *JOIV: International Journal on Informatics Visualization*, **9**(2), 831–837. http://dx.doi.org/10.62527/joiv.9.2.2883.

Jackson L.S., Finney D.L., Kendon E.J., Marsham J.H., Parker D.J., Stratton R.A., Tomassini L., Tucker S. (2020) The effect of explicit convection on couplings between rainfall, humidity, and ascent over Africa under climate change, *Journal of Climate*, **33**, 8315–8337. https://doi.org/10.1175/JCLI-D-19-0322.1.

Jansing L., Papritz L., Dürr B., Gerstgrasser D., Sprenger M. (2022) Classification of Alpine south foehn based on 5 years of kilometre-scale analysis data, *Weather and Climate Dynamics*, **3**, 1113–1138. https://doi.org/10.5194/wcd-3-1113-2022.

Khastagir A., Hossain I., Anwar A.F. (2022) Efficacy of linear multiple regression and artificial neural network for long-term rainfall forecasting in Western Australia, *Meteorology and Atmospheric Physics*, **134**(4), 69. https://doi.org/10.1007/s00703-022-00907-4.

Kumar R., Bhanu M., Mendes-Moreira J., Chandra J. (2024) Spatio-temporal predictive modeling techniques for different domains: a survey, *ACM Computing Surveys*, **57**(2), 1–42. https://doi.org/10.1145/3696661.

Necesito I.V., Kim D., Bae Y.H., Kim K., Kim S., Kim H.S. (2023). Deep learning-based univariate prediction of daily rainfall: application to a flood-prone, data-deficient country, *Atmosphere*, **14**(4), 632. https://doi.org/10.3390/atmos14040632.

Nordin N.Z.A., Yunus M.M., Dahalan A.N. (2023), Implementation of rainfall forecasting using ANN, *5th International Conference on Applied Engineering and Natural Sciences*, Turkey. https://doi.org/10.59287/icaens.1071.

Obot N.I. (2024) Regression models and hybrid intelligent systems for estimating clear-sky downward longwave radiation in Equatorial Africa, *Journal of the Earth and Space Physics*, **49**(4), 143–159. http//doi.org/10.22059/jesphys.2023.352877.1007484.

Obot N.I., Humphrey I., Jolayemi O.R. (2024) Significant trends and structural shifts in rainfall patterns in Nigeria – Pomembni trendi in strukturni premiki v vzorcih padavin v Nigeriji, *Acta hydrotechnica*, **37**/67, 103–120. https://doi: 10.15292/acta.hydro.2024.06.

Obot N.I., Humphrey I., Chendo M.A.C., Udo S.O. (2019) Deep learning and regression modelling of cloudless downward longwave radiation, *Beni-Suef University Journal of Basic and Applied Sciences*, **8**, 23. https://doi.org/10.1186/s43088-019-0018-8.

Ohba M. (2021) Precipitation under climate change, In: *Precipitation* (pp. 21–51), Elsevier. https://doi.org/10.1016/b978-0-12-822699-5.00002-1.

Ojo O.S., Ogunjo S.T. (2022) Machine learning models for prediction of rainfall over Nigeria, *Scientific African*, *16*, e01246. https://doi.org/10.1016/j.sciaf.2022.e01246

Onyeka-Ubaka J.N., Halid M.A., Ogundeji R.K. (2021) Optimal stochastic forecast models of rainfall in South-West region of Nigeria, *International Journal of Mathematical Sciences and Optimization: Theory and Application*, **7**, 1–20.

Othman S.A., Jameel H.H., Abdulazeez S.T. (2024) Comparative analysis of univariate SARIMA and multivariate VAR models for time series forecasting: a case study of climate variables in Ninahvah City, Iraq, *Mathematical Problems of Computer Science*, **61**, 27–49. https://doi.org/10.51408/1963-0113.

Rahaman M.M., Rani S., Islam M.R., Bhuiyan M.M.R. (2023) Machine learning in business analytics: advancing statistical methods for data-driven innovation, *Journal of Computer Science and Technology Studies*, **5**(3), 104–111. https://doi.org/10.32996/jcsts.2023.5.3.8.

Rožman M., Kišić A., Oreški D. (2024) Comparative analysis of nonlinear models developed using machine learning algorithms, *WSEAS Transactions on Information Science and Applications*, **21**, 303–307. https://doi.org/10.37394/23209.2024.21.29.

Schaack T.K., Johnson D.R., Wei M.-Y. (1990) The three-dimensional distribution of atmospheric heating during the GWE, *Tellus*, **42A**, 305–327. https://doi.org/10.3402/tellusa.v42i3.11880.

Somaila K., Yacoula S., Louis Z.J. (2022) Solar wind and geomagnetic activity during two antagonist solar cycles: comparative studies between solar cycles 23 and 24, *International Journal of Physical Sciences*, **17**, 57–66. https://doi.org/10.5897/IJPS2022.4998.

Sun Q., Zhang X., Zwiers F., Westra S., Alexander L.V. (2021) A global, continental, and regional analysis of changes in extreme precipitation, *Journal of Climate*, **34**(1), 243–258. https://doi.org/10.1175/jcli-d-19-0892.1.

Tschora H., Cherubini F. (2020) Co-benefits and trade-offs of agroforestry for climate change mitigation and other sustainability goals in West Africa, *Global Ecology and Conservation*, **22**, e00919. https://doi.org/10.1016/j.gecco.2020.e00919.

Usoskin I.G., Solanki S.K., Krivova N., Hofer B., Kovaltsov G.A., Wacker L., Brehm N., Kromer B. (2021) Solar cyclic activity over the last millennium reconstructed from annual 14C data, *Astronomy & Astrophysics*, **649**, A141. https://doi.org/10.1051/0004-6361/202140711.

Wang J., Gao Y. (2023) Generalized mixture model for extreme events forecasting in time series data, *China Automation Congress (CAC)* (pp. 7817–7822). IEEE. https://doi.org/10.1109/CAC59555.2023.10450893.

Wu J., Long J., Liu M. (2015) Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm, *Neurocomputing*, **148**, 136–142. https://doi.org/10.1016/j.neucom.2012.10.043.

Yaya O.S., Fashae O.A. (2015) Seasonal fractional integrated moving average models for rainfall data in Nigeria, *Theoretical and Applied Climatology*, **120**, 99–108. https://doi.org/10.1007/s00704-014-1153-8.

Yin J., Gu H., Liang X., Yu M., Sun J., Xie Y., Li F., Wu C. (2022) A possible mechanism for rapid production of the extreme hourly rainfall in Zhengzhou city on 29 July 2021, *Journal of Meteorological Research*, **36**, 6–25. https://doi.org/10.1007/s13351-022-1166-7.

Zhang J., Wen X., Zhang Z., Zheng S., Li J., Bian J. (2024) ProbTS: Benchmarking point and distributional forecasting across diverse prediction horizons, *Advances in Neural Information Processing Systems*, **37**, 48045–48082. https://github.com/microsoft/ProbTS.