

Comparative Assessment of Random Forest and Support Vector Regression Models for Rainfall Time-Series Forecasting and Flood Risk Implications in Maiduguri, Nigeria.

Abdullahi Maina^{1†}, Utpal Dhar Das², Mustapha Grema³, Raji Muhammed Raji⁴

^{1,4}Department of Statistics, University of Abuja, Abuja, FCT, Nigeria

²Department of Mathematics and Data Science & Analytics, Sharda University, Greater Noida, Uttar Pradesh, India.

³Department of Statistics, School of Science, Federal Polytechnic Monguno, Borno State, Nigeria

†Corresponding Author: Email: mainastat@gmail.com ORCID: <https://orcid.org/0009-00041276-9877> Phone: +2347033177145

ABSTRACT:

Rainfall variability is a major driver of hydrological extremes, particularly in semi-arid regions, which are remarkably sensitive to changes in the planetary climate. Precise rainfall forecasting is an imperative component in flood risk mapping. This study assessed the applicability of machine learning models to provide rainfall time series forecasting in Maiduguri, in northeastern Nigeria, using monthly rainfall data from 1981 to 2023, collected from the Nigerian Meteorological Agency. Two machine learning models, Random Forest (RF) and Support Vector Regression (SVR), were evaluated using both regression and classification frameworks. The performance metrics used in assessing the models in forecasting rainfall were Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2), in addition to classification using a confusion matrix after binning the rainfall into flood-classes based on a specific threshold. Results indicate a considerable outperformance of the RF algorithm in forecasting rainfall in Maiduguri, Nigeria, based on its smaller prediction errors (MSE = 1571.61, RMSE = 39.64 mm, MAE = 21.99 mm, $R^2 = 0.80$) than those of SVR (MSE = 1956.12, RMSE = 44.23 mm, MAE = 24.80 mm, $R^2 = 0.77$). Also, a confusion analysis on the classification results revealed a higher capability on the part of the RF algorithm in recognizing events of rainfall that are prone to flooding. The results demonstrate the superior predictive capability of the RF algorithm in forecasting rainfall, particularly in flood risk mapping in semi-arid areas sensitive to changes in planetary climate.

Keywords: Rainfall forecasting; Machine learning; Random Forest; Support Vector Regression; Flood risk assessment

1.0 Introduction

The Earth's climate varies on both temporal and spatial scales. Climate change refers to the continuous, long-term change to average weather conditions (e.g., average rainfall). Climate change has the potential to impact all natural systems, thereby threatening human development and survival across economic, social, and political dimensions (Karabulut, 2015). Furthermore, it is widely acknowledged that developing nations in tropical areas, like Nigeria, are likely to be more severely affected by climate change than their developed counterparts.

Rainfall is one of the climate parameters that affect the pattern and behaviour in which humans live. It impacts all elements of the ecosystem, including plants and animals. Therefore, the examination of rainfall cannot be overemphasized (Obot and Onyeukwu, 2010). In addition to its positive effects, rainfall can also have negative effects, such as producing natural disasters like floods and droughts, which represent opposite extremes of the hydrological cycle. The IPCC has documented a general decline in rainfall patterns from a global perspective but increasing trend in Africa (IPCC, 2023). In line with IPCC findings, empirical data indicates that rainfall trends have increased in some regions of Africa, primarily due to warming climatic conditions and related atmospheric dynamics (Yvonne et al., 2020). Rainfall trends in Nigeria exhibit significant temporal and spatial variability. An initial analysis of data from six national meteorological stations revealed a trend of declining rainfall (Nnodu & Magaji, 2024; Obot et al., 2025). The spatial sensitivity of trend outcomes to data coverage and region is highlighted by the fact that rainfall behaviour varies significantly depending on geographic location and the temporal scale of analysis. For example, recent trend analyses in Nigeria have shown statistically significant increasing rainfall trends at several stations while some locations exhibit decreasing rainfall trends (Nnodu & Magaji, 2024; Obot et al., 2025). Rising rainfall levels can result in flooding, although drought and flood are interlinked constituents of a single hydrological cycle, studies often focus on only one of these extremes because of the multi-complexities and unique spatial and temporal nature of each (Matano et al., 2022). Ever since the early 1970s, the Sahel drought has been a matter of considerable interest to researchers and policymakers worldwide (Nicholson et al., 2018). While African researchers have pleaded with the United Nations to assist in building preparations for future droughts with growing climate change (Urama & Ozor, 2010, p. 15), the problem of flooding in the Sahel has been mostly ignored. The floods take place more often and cause more economic losses across Africa, while the most deaths are caused by droughts (WMO, 2021). This disparity aggravated the susceptibility of both humans and the environment to the effects of flood. According to Elagib (2021), historical flood records lack comprehensiveness, standardization, quality control, and scientific rigor.

In Maiduguri, northeastern Nigeria, notable variability in climatic parameters, particularly rainfall, has been reported over recent decades (Bello et al., 2023; Abatcha et al., 2024). Further evidence from Bello et al. (2023) indicates that annual rainfall in Maiduguri is characterized by more positive anomalies than negative ones, alongside a modest upward trend of 0.0361 mm per year over the period 1974–2014. At the monthly scale, rainfall trends vary markedly across the year. July shows the strongest increasing trend, rising by about 2.14 mm per year, whereas April exhibits a declining trend of approximately -0.25 mm per year. In addition,

time-series modelling using ARIMA techniques has been applied to rainfall data in Maiduguri. Uba and Bakari (2015) identified the ARIMA (1,1,0) model as suitable for capturing the underlying dynamics of monthly rainfall and for generating reliable short-term forecasts.

Due to their intrinsic linear assumptions, classical time series models like ARIMA (1,1,0) perform poorly when used for long-term trend analysis, despite their success in short-term rainfall forecasting. Recent research highlights how machine learning (ML) approaches outperform conventional statistical models in identifying intricate, nonlinear relationships in climate data. Classical time series forecasting techniques frequently underperform in regions where rainfall patterns are influenced by complex inter-regional climatic interactions. These methods typically rely on simplified linear assumptions and are unable to effectively capture short-term variability and intricate spatial dependencies (Wani et al., 2024). Siami-Namini et al. (2018) verified that machine learning models significantly reduce prediction errors, outperforming traditional ARIMA models by approximately 84–87%. Kumar et al. (2025) showed that classical models such as ARIMA often fail to represent the nonlinear nature of rainfall pattern.

Consequently, ML algorithms have emerged as robust alternatives for improving rainfall prediction accuracy and reliability, gaining increasing attention in recent climatological research (Truong et al., 2023). This study employed machine learning approaches, notably Support Vector Regression SVR and Random Forest RF, with the RF model demonstrating superior predictive accuracy compared to the other methods considered. The results further indicate that focusing on the most effective algorithms improves prediction accuracy while also reducing computational demands and supporting more efficient practical implementation.

Despite the substantial body of research on rainfall variability and trends in Nigeria and across the Sahel, many existing studies still depend on classical linear time-series models that are not well suited to capturing the nonlinear, non-stationary, and extreme-event behaviour of rainfall in semi-arid regions. Although machine-learning methods have shown promise for rainfall forecasting in other climatic contexts, their application in northeastern Nigeria, particularly in Maiduguri, remains limited, and their usefulness for flood-related decision making is seldom examined. In addition, most previous studies focus primarily on regression accuracy, with little attention given to classification-based evaluations that are directly relevant to flood risk. This study bridges these gaps by comparing Random Forest and Support Vector Regression for long-term monthly rainfall forecasting in Maiduguri and by combining conventional regression metrics with confusion-matrix-based flood classification to improve the practical value of rainfall predictions for flood risk assessment.

2.0 METHODOLOGY

2.1 Study Area

Maiduguri is the state capital of Borno State, North Eastern Nigeria. Maiduguri lies approximately at latitude 11°51'N and longitude 13°40'E with an Altitude of 300 m above mean sea level and borders Chad, Cameroun and Niger (Olofin, 1997). The land has an area of 543km² and has a population of about 357,104 people. It is within the semi-arid climatic zone referred to as the SAHEL zone. The city practically experiences two distinct climatic seasons yearly. These are; a short rainy season usually from the month of June to September and a long

dry season from October to May. The hottest months in the year are March, April and May having temperatures ranging between 30°C-43°C (Maina and Grema, 2026).

2.2 Data Description

Monthly average rainfall data used in this study were obtained from ground-based observations recorded by the Nigeria Meteorological Agency (NiMet).

The analysis is based on a univariate time series of monthly rainfall observations spanning approximately forty-two years, corresponding to about 480 consecutive records. Monthly aggregation is adopted to capture seasonal cycles and inter-annual rainfall variability, which are fundamental characteristics of long-term hydro-climatic processes. The dataset is arranged chronologically to preserve temporal dependence.

2.2 Data Preprocessing

Prior to model development, the rainfall series undergoes standard quality control procedures. Extreme rainfall values are carefully examined to distinguish genuine hydrological extremes from potential measurement errors; verified extremes are retained to preserve physical realism.

To ensure numerical stability, rainfall values are normalized using Min-Max scaling, defined as follows

$$R_t^* = \frac{R_t - R_{min}}{R_{max} - R_{min}}$$

Where R_t represents observed rainfall at time t , and R_{max} and R_{min} denote the minimum and maximum rainfall values in the training dataset.

2.3 Models

In this study machine learning methods are used to examine rainfall trends. Specifically, the analysis uses two machine learning models: RF and SVR. Details on model configurations and parameters are provided below.

2.3.1 Random Forest (RF)

3 RF is an ensemble learning method that combines multiple decision trees trained on bootstrap samples with random feature selection. This dual randomness improves generalization and reduces overfitting. In regression, predictions are obtained by averaging outputs from all trees. RF was used to capture nonlinear relationships between lagged rainfall inputs and future rainfall values.

$$\hat{R}_t = \frac{1}{B} \sum_{b=1}^B T_b(X_t) \quad (1)$$

Here, T_b denotes the prediction produced by the b_{th} decision tree, which is fitted using a bootstrap sample drawn from the training dataset. The predictive capability of the Random Forest model depends on appropriate selection of several hyperparameters, notably the number of trees in the ensemble, the maximum allowable depth of individual trees, the minimum number of observations required at terminal nodes, and the subset of predictor variables evaluated at each split.

For multi-step forecasting, a recursive strategy is adopted:

$$\hat{R}_{t+h} = f(\hat{R}_{t+h-1}, \hat{R}_{t+h-2}, \dots), \quad h > 1 \quad (2)$$

Feature importance is assessed using the mean decrease in impurity, enabling identification of dominant lagged predictors

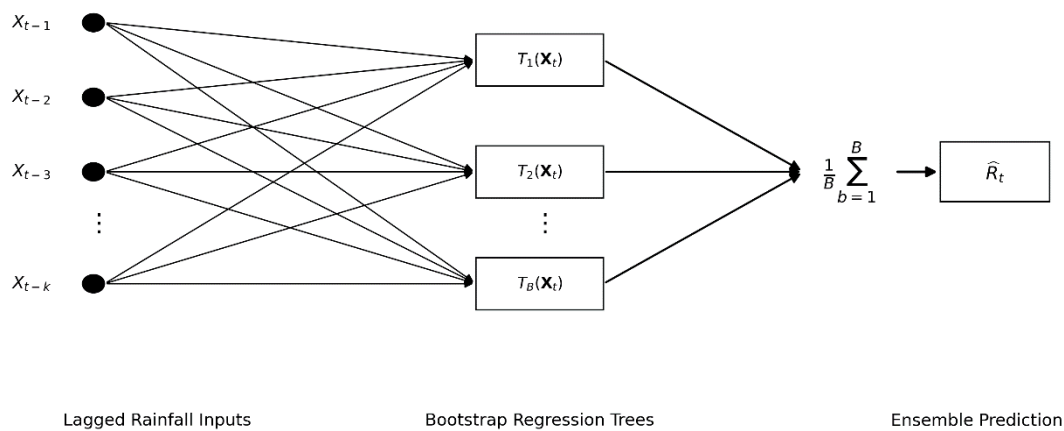


Figure 1: Random Forest Ensemble Learning Framework Using Lagged Rainfall Inputs

3.2.1 Support Vector Regression (SVR)

SVR was developed by Vapnik (1997), a machine learning method used for time series forecasting. The SVR is based on several key components. The input data is mapped into a higher-dimensional feature space using a Radial Basis Function (RBF) kernel. The accuracy of the model can be improved by modifying key hyperparameters such as the regularisation

parameter (C) and the radial basis function kernel. One of SVR's advantages is that, with the right kernel, it can identify complex, nonlinear correlations between inputs and outputs.

SVR is applied to capture nonlinear rainfall dynamics within a regularized regression framework. SVR estimates a function of the form:

$$f(X) = \mathbf{w}^T \phi(X) + b \quad (3)$$

Where $\phi(\cdot)$ maps the input vector into a high-dimensional feature space, \mathbf{W} is the weight vector, and b is the bias term.

The model is obtained by solving the optimization problem:

$$\min_{\mathbf{w}, b, \xi, \xi^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (4)$$

Subject to:

$$R_i - f(X_i) \leq \varepsilon + \xi_i$$

$$f(X_i) - R_i \leq \varepsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

Where c is the regularization parameter, ε defines the ε -insensitive loss and ξ_i, ξ_i^* are slack variables.

A radial basis function kernel is employed

$$K(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2) \quad (5)$$

Hyperparameters C , γ and ε are optimized using time-series-aware validation. Multi-step forecasting is performed using a recursive prediction framework consistent with the Random Forest approach.

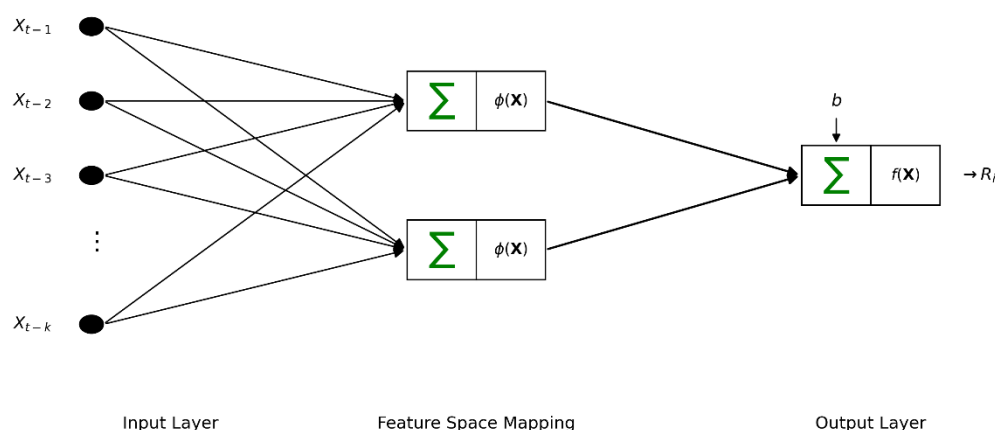


Figure 2: Architecture of the Support Vector Regression (SVR) Model for Rainfall Forecasting

3.3 Transformation to Supervised Learning Framework

Machine learning models require a supervised input-output structure. Accordingly, the rainfall time series was transformed using a lag-based embedding approach widely used in time series forecasting.

$$X_t = [R_{t-1}, R_{t-2}, \dots, R_{t-p}]$$

Where p represents the number of lagged observations. Given the strong annual seasonality of monthly rainfall, lag lengths of up to 12 months are considered to capture annual seasonality persistence and multi-year dependencies. To explicitly represent seasonality, a categorical month indicator $M_t = \{1, 2, 3, \dots, 12\}$ is included as additional predictor. Rolling statistical descriptors, such as moving averages and rolling standard deviations, are also captured:

$$\mu_t^{(k)} = \frac{1}{k} \sum_{i=1}^k R_{t-i} \tag{6}$$

$$\sigma_t^{(k)} = \sqrt{\frac{1}{k} \sum_{i=1}^k (R_{t-i} - \mu_t^{(k)})^2} \tag{7}$$

R_{t-i} : Rainfall observation recorded at time interval $t - i$ where $i = 1, 2, \dots, k$, t the current time index at which rolling statistic is evaluated, k is the lag-size

3.4 Statistical Evaluation

Evaluating the performance of a predictive model is a critical component of the modelling process, particularly when its effectiveness is assessed relative to alternative modelling approaches. This process typically depends on the use of particular statistical measures. Nevertheless, models can have almost identical results for a particular measure, using only one measure might not give a comprehensive view of the performance of the model. Each measure identifies only a specific feature of the model's capacity to mimic actual data. Hence, it is advisable to employ a set of statistical measures to achieve a more complete and accurate assessment of model performance so that more meaningful comparisons between various modelling techniques can be made. Among the most frequently used are MSE, RMSE, MAE and R^2 .

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

Where;

- y_i : is the actual value, \hat{y}_i : is the fitted value, n : is the number of observation.

4 Result

4.2 Statistical Analysis of Rainfall Variability and Seasonality

Figure 3 shows the average rainfall in Maiduguri per month, and it is clear that there is very prominent seasonality in the data. There is substantial rainfall between May and September, with peaks in July and August, while the other months get mild or no rainfall. Such prominent seasonality makes it reasonable to include seasonal lag features in machine learning models.

Figure 4 presents the yearly variability in total rainfall per year from 1981 to 2023, which exhibits high variability and at times extreme rainfall years. The existence of high variability

and non-stationarity in the time series suggests that traditional time-series modelling techniques may be inadequate in identifying the underlying patterns. The applicability of machine learning algorithms like Random Forest and Support Vector Machines in this case comes in handy because they are flexible methods in relation to the assumption of stationarity in modelling complex patterns. The increase in extreme rainfall occurrences in current years adds to the rationale of applying data modelling techniques in identifying irregular patterns. In Figure 5 the decomposition of the monthly rain data into the observation, trend, seasonal, and residual components using the additive model is presented.

- Observed Component: There are evident peaks and a growing level of variability in the original series that indicate nonlinear behaviour well adapted for machine learning modelling.
- Trend component: This component exhibits slow growth along with occasional fluctuations. This trend can be inferred via lag structures in the RF model and the Support Vector Machine model.
- Seasonal Component: The fact that there is a strong seasonal component is identified, meaning that this component is necessary in the development of input features for both RF and SVM models. • The use of a third independent variable with a strong relation with one of the target variables indicates that there are variables that are
- Residual Component: The residuals follow chaotic variations and outliers that are difficult to characterize in a statistical manner but can easily be characterized in a Kernel modelling technique.

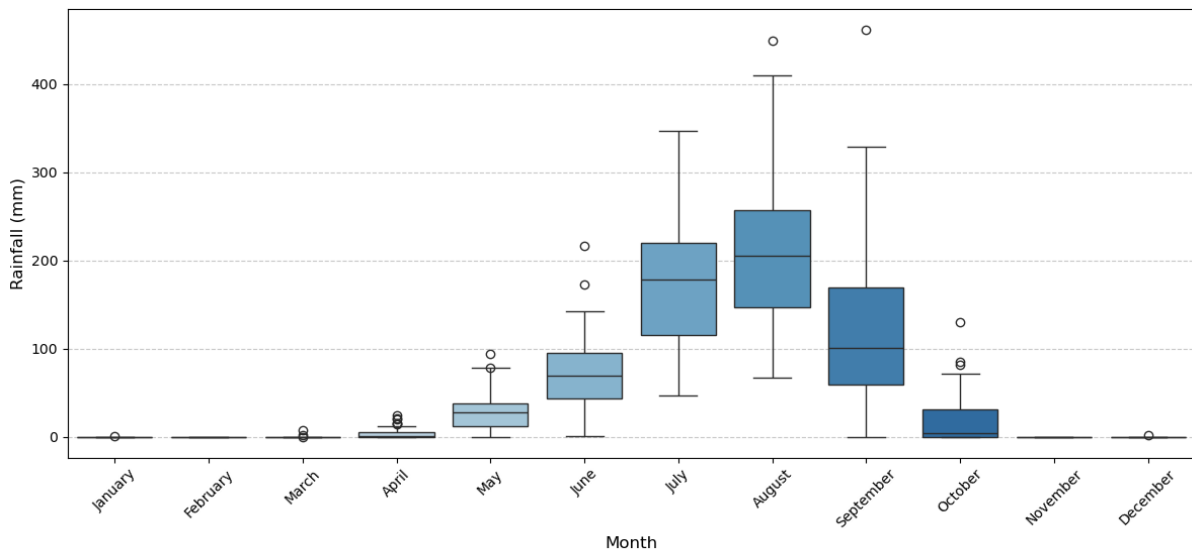


Figure 3: Monthly Rainfall Distribution – Maiduguri (1981-2023)

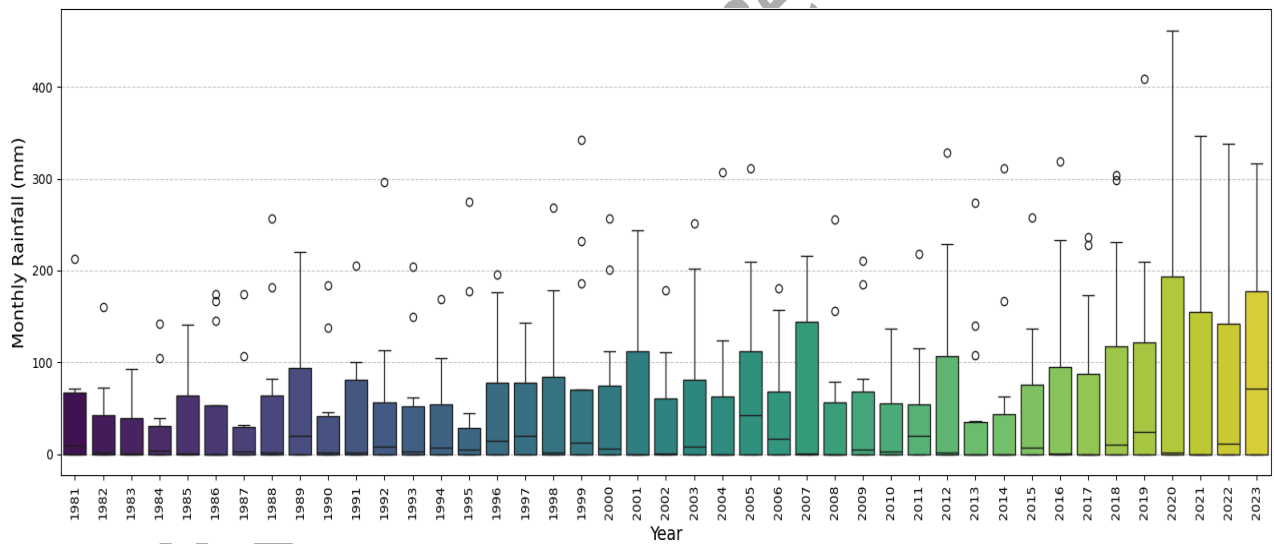


Figure 4: Rainfall Distribution by Year – Maiduguri

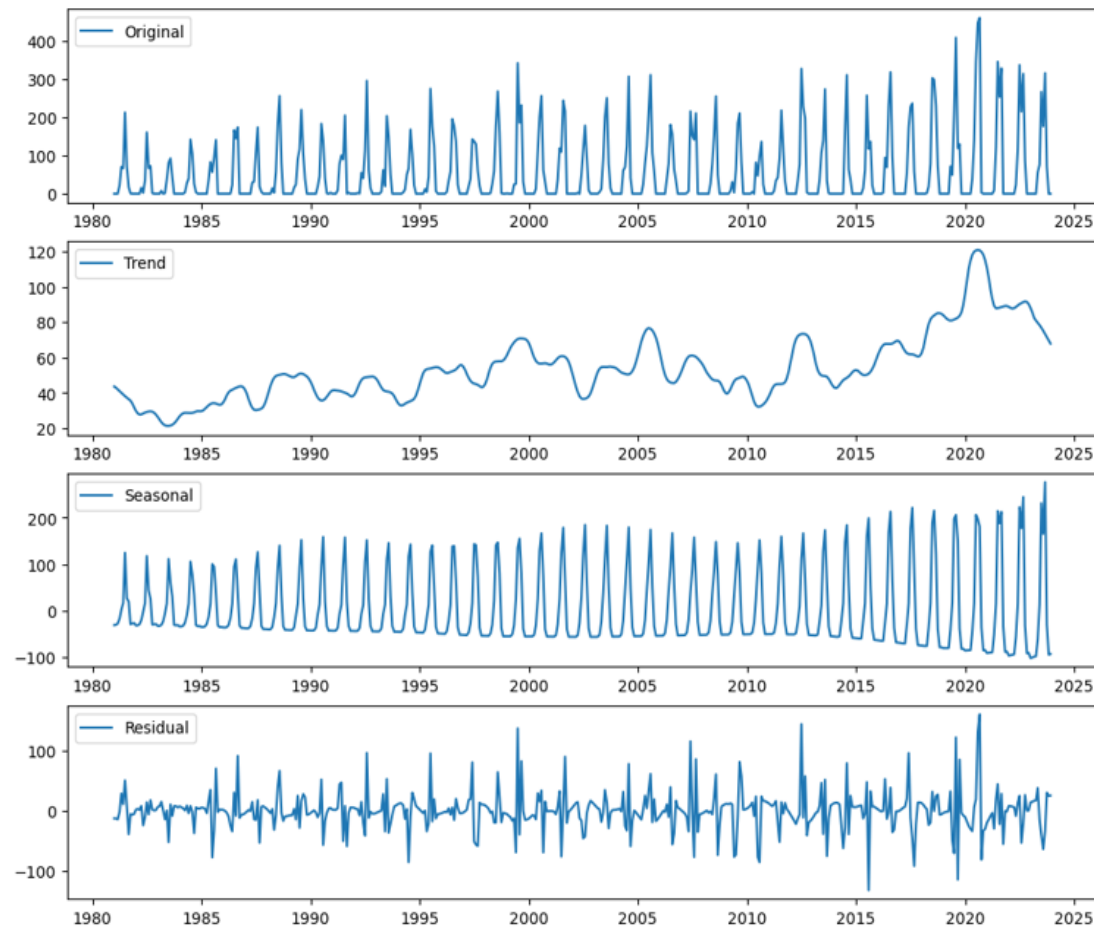


Figure 5: Decomposition of Monthly Rainfall Time Series for Machine Learning

3.2 Application of Confusion Matrix for RF and SVM

The rainfall is inherently a continuous variable measured in millimetres, but in this case, the rainfall forecasting was reframed as a classification problem by discretizing the rainfall amounts into different classes using a predefined threshold. The past rainfall observations were considered as predictors to capture the temporal dependence in the rainfall series, and the target variable was considered to represent the rainfall occurrence or intensity class in the future time step. Using this framework, the application of a confusion matrix is relevant for performance analysis of a classification model. The Random Forest Classifier demonstrated high predictive performance with high accuracy and a good compromise between true positive and true negative instances. The performance of the Random Forest Classifier suggests that it is a suitable and robust tool for rainfall time-series classification to capture the nonlinear rainfall patterns and is also capable of providing useful predictive information for rainfall occurrence.

Figure 6 indicates that the confusion matrix assumes significant implications in estimating both models' suitability toward flood-based decisions. Based on observations in the Random Forest

model, results suggest that with the large number of true positives and low values of false positives, it has significant suitability in accurately forecasting and preventing floods with considerable low values of false alarms. On the other hand, since it has values of false negatives, it assumes considerable risks since it might overlook significant flood-based rainfall amounts.

Figure 7 shows that the Support Vector Machine model also has a relatively high level of false negatives, suggesting a higher possibility of under-predicting rainfall instances which might culminate in flooding. From a flood-risk management strategy, it is essential to note the shortcoming presented by the high incidence of false negatives. Even though the SVM model displays a decent level of accuracy and a sound mechanism of handling false positives, the lower sensitivity of the model to rainfall instances makes it less reliable when it comes to the implementation of early warning systems for floods.

On the comprehensive comparison of the two confusion matrices, it is clear that the Random Forest model is more appropriate for rainfall forecasting data related to flooding, as it strikes a good balance between event identification and the regulation of false alarm, making it more useful from a hydrological risk perspective.

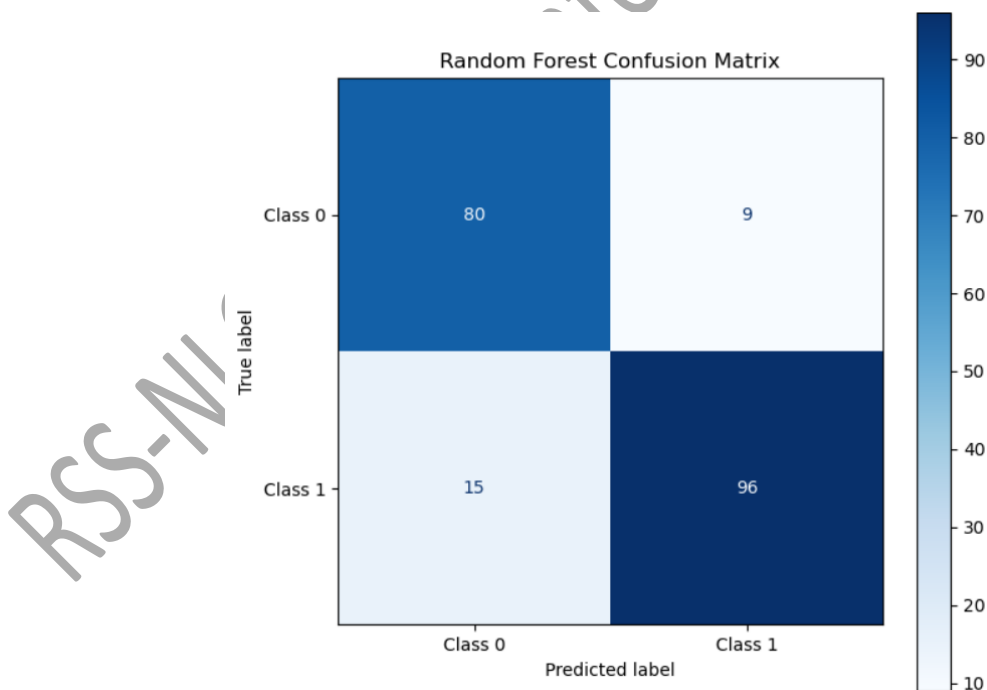


Figure 6: Random Forest Confusion Matrix

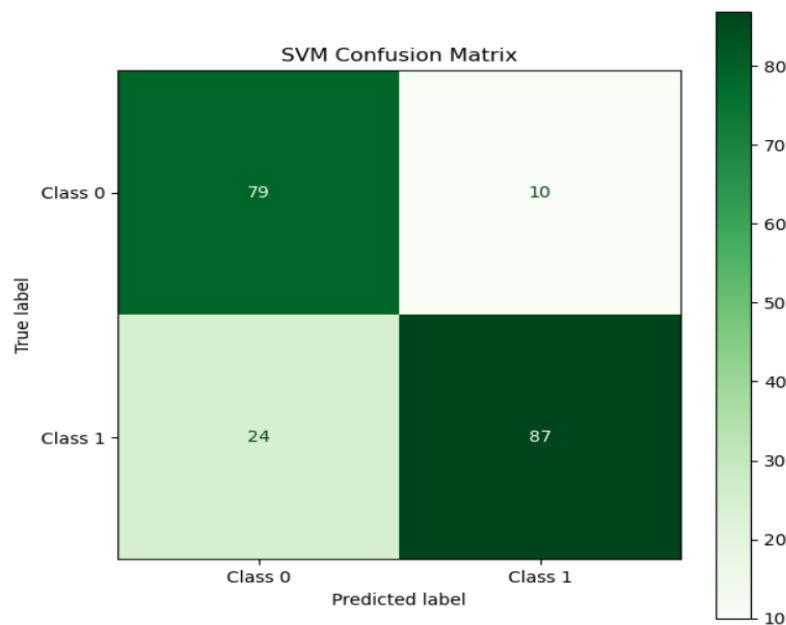


Figure 7: SVM Confusion Matrix

4.3 Rainfall Time Series Forecasting Using SVR and Random Forest

Fig. 8 and fig. 9 Depict the observed monthly rainfall series for Maiduguri (1982–2023) along with in-sample model fits and out-of-sample forecasts (2024–2030) developed using Support Vector Regression and Random Forest, respectively. Both models successfully portray the marked seasonality in rainfall that typifies the Sahelian climate, where rainfall occurs within a few months of the year while it is practically zero in the dry season. This suggests that the models can learn the prevailing seasonal structure inherent in the historical rainfall record. However, there are noticeable deviations in the two models' performance in generating the magnitude and variability of rainfall. The fit of the SVR model is rather smoothed, especially over periods of high rainfall intensity, where peak values are significantly underestimated. Such a smoothing effect becomes more pronounced during years with extreme events in rainfall, which indicates that SVR cannot capture the full nonlinear dynamics and sudden fluctuations of the rainfall time series. Thus, the forecasts of SVR for the 2024-2030 period have moderate peaks in rainfall with limited variability and may result in an underestimate of flood risk events in association with extreme rainfall.

In contrast, the Random Forest model is in better agreement with the observed rainfall series, especially for the reproduction of high-intensity rainfall peaks and interannual variability. The RF in-sample fit better captured the track of extreme rainfall episodes, which also supports its higher capability for nonlinear relationships and complex interactions among lagged rainfall inputs. RF forecasts up to 2030 also retain higher variability and peak magnitudes relative to

SVR, which can be interpreted as more realistic future rainfall behaviour in the study area. The underestimation of peak rainfall by the SVR model suggests a higher likelihood of failing to detect extreme rainfall events that may lead to flooding and limits its suitability for early warning and disaster preparedness applications. By contrast, the capability of Random Forest to capture higher intensities of rainfall and temporal variability increases its reliability to identify potential flood-inducing rainfall conditions. This observation agrees with the superior quantitative performance of RF, as evident from the lower prediction errors and higher coefficient of determination reported in earlier sections. Overall, the fit of the model and forecast comparison shows that though both are good at capturing the seasonal pattern of rainfall, Random Forest provides a robust and realistic representation of rainfall dynamics in Maiduguri. Its better handling of extremes and nonlinear behaviour makes it more suitable for long-term rainfall forecasting and the valuation of flood risk in semi-arid regions facing increasing climate variability.

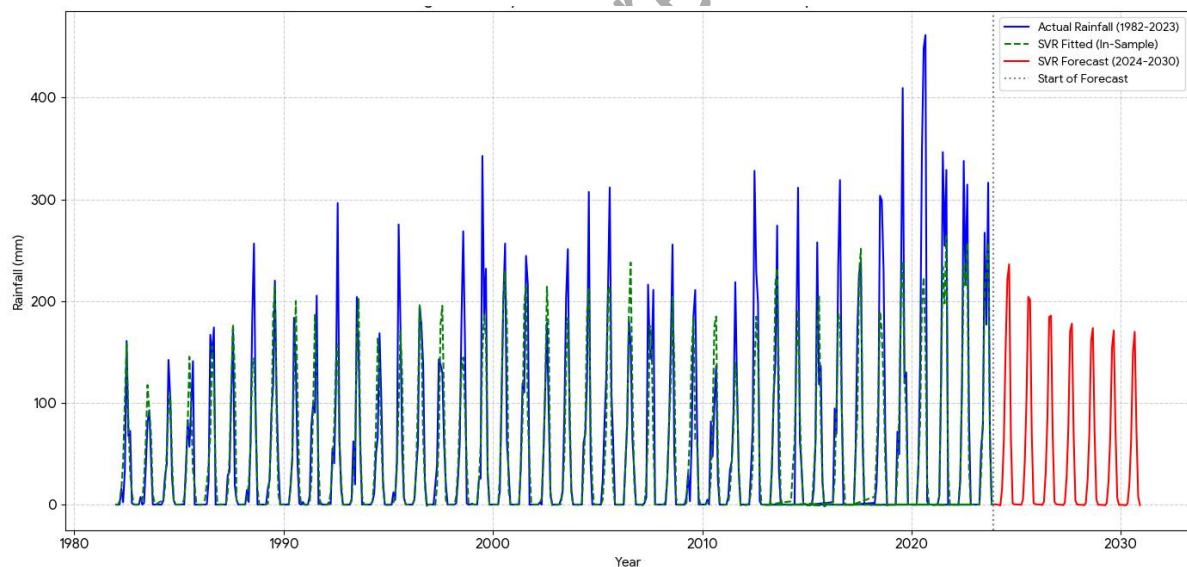


Figure 8: Maiduguri Monthly Rainfall: SVR Model Fit and Forecast up to 2030

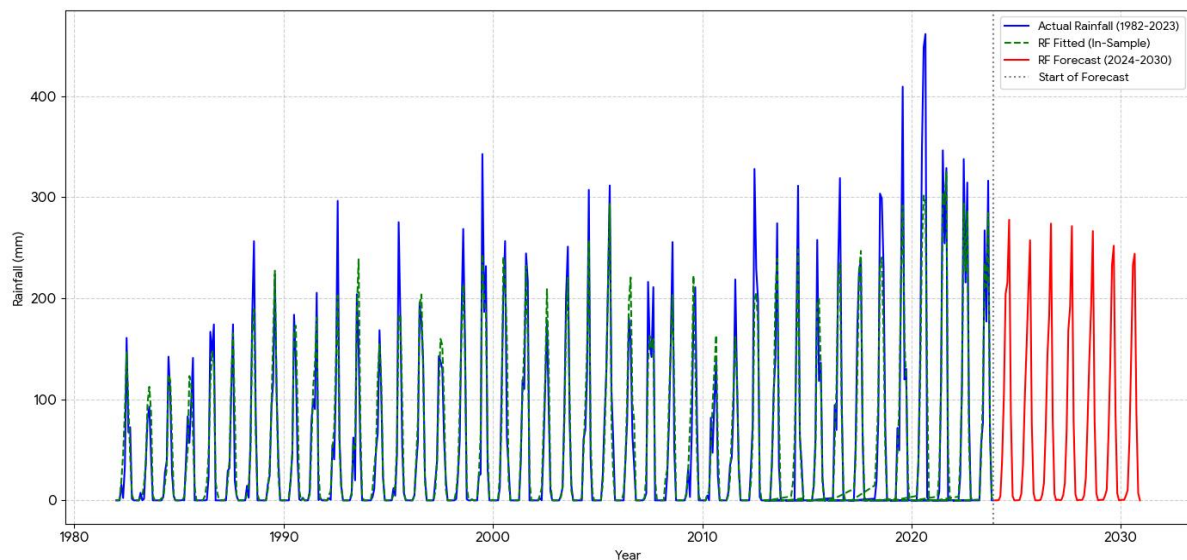


Figure 9: Maiduguri Monthly Rainfall: Random Forest Model Fit and Forecast up to 2030

4.4 Evaluation Metrics

Table 1: Shows the quantitative performance measures further accentuate the relative superiority of the Random Forest approach over the Support Vector Regression (SVR) approach for rainfall modeling. This is evidenced by the significantly lower error measures in the Random Forest approach with an MSE of 1571.61, RMSE of 39.64 mm, and MAE of 21.99 mm recorded in the table above, when contrasted with the higher error values for the SVR approach of MSE = 1956.12, RMSE = 44.23 mm, and MAE = 24.80 mm, respectively. These measures indicate the increased accuracy and predictability of Random Forest for realistic rainfall estimation with less variation from the actual values for better flood risk mitigation and warning strategies. Furthermore, the increased values of the coefficient of determination (R^2) for the Random Forest approach at $R^2 = 0.80$ also support the understanding that Random Forest can explain a larger variation in the rainfall magnitude measures under investigation when compared with the $R^2 = 0.77$ for the SVR approach. These improvement in accuracy and explanatory power highlight the practical advantage of Random Forest approach from the perspective of flooding when inundation projections beyond actual values can potentially cause missed warnings with serious socio-economic repercussions for the spell of life with severe implications for the communities at risk at the time of catastrophic flooding events.

Table 1: Performance Metrics for Machine Learning Models

Model	MSE	RMSE	MAE	R ² SCORE
SVR	1956.12	44.23	24.80	0.77
Random Forest	1571.61	39.64	21.99	0.80

4.0 Discussion

This study assessed machine learning methods for forecasting Monthly rainfall in Maiduguri, a semi-arid area facing growing flood risks (Obot et al., 2025). By comparing RF and SVR using both standard regression metrics and flood-threshold classification, the analysis offers practical evidence regarding the reliability of data-driven models for hydrological early warning systems in the Sahel.

4.1 Rainfall Variability, Nonlinearity, and Model Suitability

The rainfall record of Maiduguri follows the typical Sahelian pattern, concentrated between May and September with peaks in July and August. It also shows significant variation from year to year, along with sporadic extreme events that make the data non-linear and non-stationary. Classical time-series models often struggle with this kind of irregularity. When we decompose the data, the contrast becomes clear: the seasonal component is steady, but the residuals are sharp and random. This randomness likely comes from complex local atmospheric shifts that are hard to pin down with a formula. Because the underlying structure is so messy, flexible data-driven models—especially those combining multiple predictors—simply do a better job of handling the complexity than rigid traditional methods.

4.2 Comparative Performance of Random Forest and Support Vector Regression

The quantitative performance measures further accentuate the relative superiority of the RF to SVM model across all regression metrics. This is evidenced by its significantly lower MSE, RMSE, and MAE. These demonstrate the ability of RF to generate predictions that are consistently closer to observed rainfall values. Moreover, the value R² indicates that the RF model explains a large proportion of the variance in monthly rainfall, highlighting its robustness in capturing both seasonal and interannual dynamics. The superior performance of RF can be attributed to its ensemble structure, which combines multiple decision trees trained on bootstrap samples and randomly selected predictors. This design enables RF to capture complex nonlinear relationships, threshold effects, and interactions among lagged rainfall

variables without requiring strong assumptions about the underlying data distribution. Additionally, the inclusion of seasonal lags and rolling statistical features further enhances the model's ability to represent persistence and delayed rainfall responses.

4.3 Flood Risk Assessment and Confusion Matrix Analysis

Transforming rainfall forecasting into a classification problem based on flood-relevant rainfall thresholds is particularly valuable from a disaster risk management perspective (Kumar et al., 2025). When the forecasting task was transformed into a classification problem, the RF model offered a safer balance between sensitivity and specificity. It minimized false alarms (false positives) while successfully flagging flood-prone events. Conversely, SVR produced a higher rate of false negatives. In an operational early warning system, missing an extreme event (false negative) is far more dangerous than a false alarm; therefore, despite SVR's acceptable general accuracy, its lower sensitivity to extremes limits its viability for safety-critical applications.

4.4 Forecasting Behaviour and Representation of Extremes

The comparison of in-sample performance and out-of-sample forecasts highlights distinct differences in the behaviour of the two models. Although both RF and SVR are able to reproduce the seasonal occurrence of rainfall, they differ markedly in their representation of rainfall magnitude and variability. The SVR model produces smoother forecasts with attenuated peaks, particularly during periods of intense rainfall, indicating a limited capacity to adequately capture extreme events.

By comparison, the Random Forest model more accurately reproduces observed peak rainfall values and retains interannual variability across both the calibration period and projected horizon. Preserving extreme rainfall events is especially important for flood risk assessment, as reliable identification of high-impact episodes underpins effective hazard evaluation. Recent extreme value inferences for Maiduguri quantify this threat, estimating significant 50-year and 100-year flood return levels of 469.44 mm and 516.21 mm of rainfall, respectively (Maina and Bakari, 2025). Because these extreme return levels dictate critical design values for urban drainage and flood resilience mechanisms, the SVR model's tendency to attenuate peak rainfall during intense episodes severely limits its practical application. Furthermore, the persistence of variability in RF projections through 2030 indicates that the model provides a more plausible representation of future rainfall behaviour under conditions of heightened climate variability.

4.5 Broader Implications for Climate-Sensitive Semi-Arid Regions

Climate change significantly enhance rainfall variability and increases the frequent of extreme events, the shortcomings of conventional linear modelling approaches become increasingly evident (Martel et al., 2025). The results of this study therefore carry broader relevance for rainfall prediction and flood risk management in semi-arid regions of sub-Saharan Africa. The stronger performance of the RF model underscores the value of ensemble-based ML methods that are better in capturing the nonlinear behaviour characteristic of hydro-climatic processes.

Moreover, integrating regression-based metrics with classification-oriented analyses provides a more comprehensive evaluation of model performance by linking quantitative accuracy to operational usefulness. This approach improves the practical value of rainfall forecasting models for applications including early warning systems, urban drainage infrastructure design, and long-term climate adaptation strategies.

4.6 Limitations and Future Research Directions

Although the Random Forest model exhibits strong predictive skill, its performance is inherently limited by a reliance on data-driven learning, without explicitly accounting for the structured temporal features that are typically represented in classical time-series models. In rainfall records marked by strong seasonality, long-term non-stationarity, and increasing climatic variability, this shortcoming may reduce both interpretability and robustness over longer forecasting horizons. Future studies could therefore improve rainfall prediction by developing hybrid modelling approaches that integrate conventional statistical methods with machine learning techniques. Specifically, the use of seasonal-trend decomposition and ARIMA models to represent systematic components of the series, combined with machine learning algorithms to capture remaining nonlinear behaviour, presents a promising direction. Additional improvements may be realised by incorporating deep learning models and broader climate information, particularly for extended-range forecasting and flood risk assessment in semi-arid environments.

Reference

Abatcha, I., Mustapha, A., & Barkindo, A. (2024). Comprehensive analysis of rainfall variability in urban Maiduguri, Nigeria: Implications for climate resilience and sustainable development. *International Journal of Environment and Climate Change*, 14(3), 149-159.

- Amin, D., Keshta, E., Ragab, R., Chakraborty, R., & Elaksher, A. F. (2025). Impact of climate change on hydrological extremes (Floods and Droughts) in the upper blue Nile basin. *Earth Systems and Environment*, 1-20.
- Bello, Y., Msheliza, D. S., & Nyikun, P. R. (2023). Anomalies and trend analysis of temperature and rainfall in Maiduguri, Borno State, Nigeria. *Dutse Journal of Pure and Applied Sciences*, 9(4b), 414-422.
- Elagib, N. A., Khalifa, M., Babker, Z., Musa, A. A., & Fink, A. H. (2021). Demarcating the rainfed unproductive zones in the African Sahel and Great Green Wall regions. *Land Degradation & Development*, 32(3), 1400-1411.
- Elagib, N. A., Khalifa, M., Babker, Z., Musa, A. A., & Fink, A. H. (2021). Demarcating the rainfed unproductive zones in the African". *International Journal of Environment and Climate Change* 14 (3):149–159. <https://doi.org/10.9734/ijec/2024/v14i34027>.
- Intergovernmental Panel on Climate Change. (2023). Summary for policymakers. In H. Lee & J. Romero (Eds.), *Climate change 2023: Synthesis report* (pp. 1–34). IPCC. <https://doi.org/10.59327/IPCC/AR6-9789291691647/001>
- Karabulut, M. (2015). Drought analysis in Antakya-Kahramanmaraş Graben, Turkey. *Journal of Arid Land*, 7(6), 741-754. <https://doi.org/10.1007/s40333-015-0011-6>
- Kumar, V., Kedam, N., Kisi, O., Alsulamy, S., Khedher, K. M., & Salem, M. A. (2025). A comparative study of machine learning models for daily and weekly rainfall forecasting. *Water Resources Management*, 39(1), 271-290. <https://doi.org/10.1007/s11269-024-03969-8>
- Kumar, V., Sharma, K. V., Mangukiya, N. K., Tiwari, D. K., Ramkar, P. V., & Rathnayake, U. (2025). Machine learning applications in flood forecasting and predictions, challenges, and way-out in the perspective of changing environment. *AIMS Environmental Science*, 12(1), 72-105.
- Matano, A., de Ruiter, M. C., Koehler, J., Ward, P. J., & Van Loon, A. F. (2022). Caught between extremes: Understanding human-water interactions during drought-to-flood events in the Horn of Africa. *Earth's Future*, 10(9), e2022EF002747.
- Nicholson, S. (2000). Land surface processes and Sahel climate. *Reviews of Geophysics*, 38(1), 117-139. <https://doi.org/10.1029/1999RG900014>
- Nicholson, S. E. (2018). Climate of the Sahel and West Africa. In *Oxford research encyclopedia of climate science*.

- Obot, N. I., & 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.
- Obot, N. I., Humphrey, I., Ekpeni, N. M., & Tai-Ojuolape, E. O. (2025). Univariate Monthly Rainfall Forecasting in Nigeria Using Multiple Statistical and Machine Learning Methods. *Archives of Hydro-Engineering and Environmental Mechanics*, 72(1), 29-49. <https://doi.org/10.2478/heem-2025-0003>
- Siame-Namini, S., & Namin, A. S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. *arXiv preprint arXiv:1803.06386*.
- Uba, E. S., & Bakari, H. R. (2015). *An application of time series analysis in modeling monthly rainfall data for Maiduguri, North Eastern Nigeria*. *Mediterranean Journal of Social Sciences*, 6(4), 1–9.
- Urama, K. C., & Ozor, N. (2010). Impacts of climate change on water resources in Africa: the role of adaptation. *African Technology Policy Studies Network*, 29(1), 1-29.
- Wani OA, Mahdi SS, Yeasin M, Kumar SS, Gagnon AS, Danish F, Al-Ansari N, El-Hendawy S, Mattar MA. Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the North-Western Himalayas. *Sci Rep*. 2024 Nov 13;14(1):27876. <https://doi.org/10.1038/s41598-024-77687-x>
- Yvonne, M., Ouma, G., Olago, D., & Opondo, M. (2020). Trends in climate variables (temperature and rainfall) and local perceptions of climate change in Lamu, Kenya. *Geography, Environment, Sustainability*, 13(3), 102-109. <https://doi.org/10.37284/ajccrs.3.1.2129>