International Journal of Multidisciplinary Research and Publications
ISSN (Online): 2581-6187

Modelling Lagos State, Nigeria Rainfall using Real Life Data: An application of SARIMA Model

Adetunji K. Ilori[1*], Adebisi Michael[2], Fatunsin L. Modupe[1], Olaiya O. O[1], Toyosi Adebambo[3]
[*1]Statistics Programme, National Mathematical Center, Abuja, Nigeria
[1]Mathematics Programme, National Mathematical Center, Abuja, Nigeria
[2]Nigeria Centre for Disease Control and prevention
[3]Department Public Health, UNICAF University, Zambia.
Email address: ilori_adek@yahoo.com

Abstract— *Accurate rainfall forecasting is crucial for effective climate monitoring, agricultural planning, and disaster risk management. This study develops and evaluates a predictive model for rainfall forecasting in Lagos State using the Seasonal Autoregressive Integrated Moving Average (SARIMA) approach. A monthly rainfall dataset spanning January 2016 to December 2022 was obtained from the relevant government agency in Nigeria. The dataset was systematically divided into a training subset (January 2016 – December 2021) for model development and a testing subset (January 2022 – December 2022) for performance evaluation. Time series analysis revealed significant seasonal patterns and long-term trends in Lagos State's rainfall data. Stationarity was assessed using the Augmented Dickey-Fuller (ADF) test, and necessary transformations, including first seasonal differencing, were applied to stabilize the series. Model selection was performed using the Akaike Information Criterion (AIC), with SARIMA (2,0,4)(2,2,1)[12] emerging as the optimal model due to its lowest AIC value. Model adequacy was further evaluated using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), confirming its superior predictive performance. The forecasting results demonstrated that the SARIMA model effectively captured the seasonal fluctuations in rainfall, with periodic peaks and troughs. The findings underscore the model's applicability for real-world decision-making and policy formulation in climate-sensitive sectors.*

Keywords— *Rainfall forecasting, SARIMA model, Time series analysis, Lagos State Rainfall, Stationarity test.*

I. INTRODUCTION

Rainfall plays a pivotal role in shaping the socio-economic and environmental landscape of Lagos State, Nigeria. As a coastal megacity located in the southwestern region of the country, Lagos experiences a tropical climate marked by alternating wet and dry seasons. Rainfall significantly influences agriculture, water resource management, urban development, and disaster mitigation efforts in the state. However, the unpredictable nature of rainfall, exacerbated by climate change, poses challenges for effective planning and sustainable development. Understanding the temporal and spatial dynamics of rainfall in Lagos is therefore essential to address these challenges and optimize resource allocation across various sectors.

Time series analysis provides a robust statistical framework for examining rainfall data over time, allowing researchers to uncover patterns, trends, and periodicities inherent in the data. In time series analysis, rainfall data are described as seasonal data due to the fact that the pick of rainfall is periodic and at regular spaced intervals of time (Usoro, A. E., & Awakessien, C. E., 2018). This analytical approach enables the decomposition of rainfall into key components, such as long-term trends, seasonal cycles, and irregular fluctuations. For Lagos State, where urban flooding and water scarcity have become recurring issues, time series analysis offers valuable insights for predicting rainfall behavior and informing proactive strategies to mitigate adverse impacts. Such insights are critical for enhancing urban resilience and safeguarding livelihoods in the face of growing climatic variability.

Lagos State’s susceptibility to the impacts of rainfall variability underscores the importance of accurate rainfall modeling and forecasting. The state's dense population, rapid urbanization, and extensive coastal infrastructure make it particularly vulnerable to extreme weather events, such as heavy downpours and prolonged dry spells. By leveraging time series methods to analyze historical rainfall data. World Bank Group (2021) provides a comprehensive dataset and analysis tools for understanding historical and projected climate data, including rainfall patterns, on a global and regional scale. For Nigeria, and particularly Lagos State, the portal offers detailed climatological data that spans decades, allowing for an in-depth examination of rainfall trends, variability, and their potential implications on socio-economic and environmental systems.

The aim of this study is to develop and evaluate a predictive model for rainfall forecasting in Lagos State by systematically analyzing historical rainfall data. By dividing the dataset into training (2016–2021) and testing (2022) subsets, the study seeks to identify underlying patterns, including seasonal variations and long-term trends, while ensuring the model's ability to generalize to future observations. The primary objective is to assess the accuracy and effectiveness of the SARIMA model in forecasting rainfall and its applicability to real-world decision-making in climate monitoring, agricultural planning, and disaster risk management.

II. LITERATURE REVIEW

Understanding rainfall variability and its implications has been a key focus of numerous studies. Researchers have explored various factors contributing to changes in climatic patterns, the effectiveness of forecasting models, and the broader impact of rainfall variability on agriculture, water resources, and disaster management.

Kreibich et al. (2019) identify several key factors that contribute to changes in drought and flood impacts, including climatic variations, socio-economic developments, land-use changes, and alterations in vulnerability and exposure. The study emphasizes that isolating the effects of these factors is challenging due to their interconnected nature and the dynamic interactions between human and environmental systems. These findings highlight the complexity of rainfall variability and its wide-ranging consequences.

Nnaji (2011) investigates the temporal characteristics of Nigeria's monthly rainfall data over a 21-year period (1980–2000). Using statistical methods such as the rescaled range (R/S) statistic, standard fluctuation analysis (FA), and detrended fluctuation analysis (DFA), the study identifies self-organized criticality within rainfall patterns. These findings suggest inherent long-term dependencies and irregularities in Nigeria's rainfall data, which are essential for understanding regional climate behavior.

Odjugo (2005) examines rainfall data from 28 stations across Nigeria, covering the period from 1970 to 2002. The study reveals a general decrease in rainfall, with mean annual precipitation declining from 1,350 mm during 1941–1970 to 1,276 mm in 1970–2002. Notably, while most regions experienced reduced rainfall, the coastal areas saw a slight increase. These findings have significant implications for agriculture, water resources, and biodiversity conservation in Nigeria.

Several studies have employed time series models, particularly the Autoregressive Integrated Moving Average (ARIMA) model, to analyze and predict rainfall patterns. . Ette and Etuk (2013) conducted a study on modeling monthly rainfall data in Port Harcourt, Nigeria, using Seasonal Box-Jenkins methods. Their research aimed to identify the underlying seasonal and trend components in the rainfall data and develop an effective forecasting model. Their findings demonstrated that Seasonal Autoregressive Integrated Moving Average (SARIMA) models are well-suited for capturing periodic fluctuations in rainfall, allowing for improved predictions of future precipitation patterns.

Similarly, Emmanuel and Bakari (2015) applied the Box-Jenkins methodology to analyze rainfall data in Maiduguri. Their study focused on data pre-processing, stationarity testing, parameter estimation, and diagnostic checking. The results confirmed the effectiveness of ARIMA models in capturing seasonal variations and trends, making them suitable for future rainfall forecasting. The study further emphasized the strong seasonal component of rainfall data in the region, reinforcing the need for accurate prediction methods to support climate adaptation strategies.

Beyond rainfall forecasting, time series analysis has been employed to model broader climatic patterns. Chisimkwuo et al. (2014) applied ARIMA models to temperature data, successfully capturing trends, seasonality, and random variations. Their study revealed valuable insights into temperature dynamics in the region, providing evidence of long-term trends and seasonal cycles. These findings contribute to a broader understanding of climate variability, which is crucial for developing mitigation and adaptation strategies.

Ekwe et al. (2014) also explored long-term trends and seasonal variations in rainfall, emphasizing the implications of climatic variability on agriculture, water resources, and environmental management. Their results reinforce the need for reliable forecasting models to improve preparedness for extreme weather events and optimize water resource planning.

III. METHODOLOGY

Data Collection

The monthly rainfall dataset for Lagos State was derived from the relevant government agency in Nigeria from January 2016 to December 2022

Data Analysis

The Lagos State rainfall dataset was systematically divided into two distinct subsets to enhance the reliability of the study. The first portion, referred to as the Training dataset, spans the years January 2016 to December 2021. This dataset was used exclusively for model development, parameter estimation, and trend analysis. By focusing on this period, the study aimed to identify underlying patterns in rainfall, including seasonal variations, long-term trends, and fluctuations.

The second portion, known as the Testing dataset, comprises data from the year January 2022 to December 2022. Unlike the Training dataset, this portion was not utilized in the model-building process. Instead, it served as an independent benchmark to evaluate the accuracy and effectiveness of the forecasts generated from the trained model. By comparing the predicted values against the actual rainfall observed in 2022, the study will be able to assess the model's performance in capturing real-world trends.

This approach of dividing the dataset into training and testing sets is essential in time series forecasting. It ensures that the model is not only well-fitted to historical data but also capable of generalizing to future observations. Such a strategy is particularly useful in applications such as climate monitoring, agricultural planning, and disaster risk management, where accurate rainfall predictions are crucial for decision-making (Jones, M., & Wang, P. (2021)).

Rainfall Forecasting Using the SARIMA Model

A Seasonal Autoregressive Integrated Moving Average (SARIMA) model was used to establish the best forecasting model for rainfall patterns in Nigeria. SARIMA is an extension of the Autoregressive (AR) and Moving Average (MA) models that accounts for seasonality. The model is mathematically represented as:

$$y_t = \beta_0 + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \emptyset_{12}\emptyset_{t-12} + \cdots \emptyset_{12p}\emptyset_{t-12p} + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} + \theta_{12}\theta_{t-12} + \cdots + \theta_{12p}\theta_{t-12p} + \varepsilon_t \quad (1)$$

where:
$\emptyset$ represents the coefficients of the autoregressive (AR) process,
θ represents the coefficients of the moving average (MA) process, and
ε_t *represents random errors (white noise).*

The SARIMA model assumes that a time series exhibits correlation with its past values (lags), meaning that historical data can help predict future values.

The seasonal autoregressive integrated moving average (SARIMA) model for rainfall can be written as follows:

$$RF_t = \beta_0 + \emptyset_1 RF_{t-1} + \emptyset_2 RF_{t-2} + \cdots + \emptyset_p RF_{t-p} + \emptyset_{12} RF_{t-12} + \cdots + \emptyset_{12p} RF_{t-12p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} + \theta_{12} \varepsilon_{t-12} + \cdots + \theta_{12p} \varepsilon_{t-12p} + \varepsilon_t \quad (2)$$

Where
RF_t is the rainfal at time t
β_0 is the intercept
$\emptyset_i$ are the autoregressive (AR) coefficients
$\emptyset_{12}, \emptyset_{12p}$ are the seasonal autoregressive effects
θ_j are the moving average (MA) coefficients
$\theta_{12}, \theta_{12p}$ are the seasonal moving average effects
ε_t is the white noise error term

Stationarity Conditions

For the SARIMA model to be stationary, it must satisfy two key conditions related to the autoregressive (AR) terms and the moving average (MA) terms.

First, in terms of stationarity in the autoregressive (AR) terms, the parameters of the autoregressive component, denoted as $\emptyset$, must meet the unit root condition. This condition requires that the roots of the characteristic equation lie outside the unit circle to ensure a stable process. If the rainfall data exhibits a trend, it may be necessary to apply differencing to transform the data into a stationary form. This is achieved using first-order differencing, expressed mathematically as:

$$\nabla RF_t = RF_t - RF_{t-1} \quad (3)$$

Similarly, if the data displays seasonal patterns, seasonal differencing can be applied to remove these effects. Seasonal differencing is represented as:

$$\nabla_{12} RF_t = RF_t - RF_{t-12} \quad (4)$$

This transformation helps to eliminate seasonal dependencies and makes the series more suitable for modeling with SARIMA.

The second condition is the invertibility of the moving average (MA) terms. For the SARIMA model to be properly specified, the moving average parameters, denoted as θ, must be invertible. This means that the moving average process should not diverge or explode over time, ensuring that past error terms do not disproportionately influence future values.

IV. Data Analysis and Discussion of Results

Table 1 presents key descriptive statistics for Lagos State rainfall dataset, which include measures of central tendency, dispersion, and distribution shape. The dataset exhibits a high degree of variability, as shown by the large standard deviation and variance**.** The right-skewed nature of the data suggests that a few extreme values may be influencing the mean.

TABLE 1: Descriptive Statistics of Lagos State Rainfall (mm) (January, 2016 – December, 2022)

Statistic	Value
Minimum	0.00
First Quartile	28.27
Median	106.85
Mean	131.72
Third Quartile	206.03
Maximum	439.90
Variance	13503.75
Standard Deviation	116.21

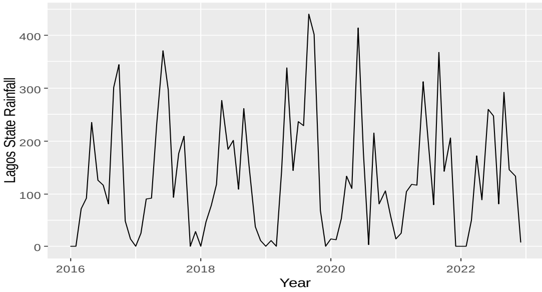

Figure 1: Time series plot of Lagos State rainfall (mm)

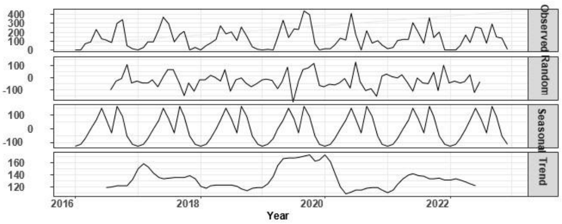

Figure 2: Decomposition Time Plot of Lagos State Rainfall

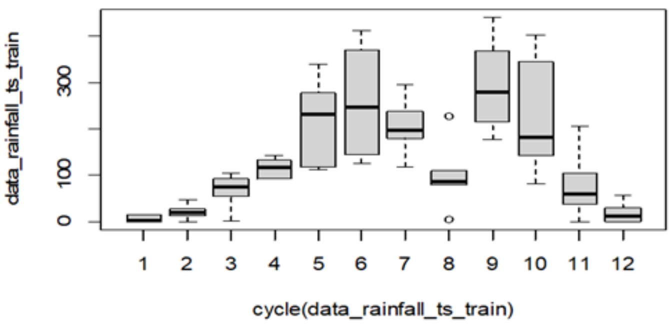

Figure 3: Box plot of Lagos State rainfall

The time series plot in Figure 1 depicts the rainfall pattern in Lagos State from 2016 to 2022**.** The graph exhibits clear seasonal variations, with repeated peaks and troughs occurring annually, indicating a cyclical pattern in rainfall distribution over time. The variations in rainfall intensity over the years suggest climatic fluctuations. Figure 2 represents the time series decomposition of rainfall data in Lagos State. Time series decomposition breaks down the data into four key components as observed series, random component, seasonal component and trend component. The presence of seasonality and trend components suggests that seasonal models like SARIMA.

Figure 3 further provide the visual distribution of monthly rainfall in Lagos State over multiple years. Each box represents the distribution of rainfall values for a given month, highlighting seasonal patterns, variations, and potential outliers. The box plot effectively illustrates the seasonal nature of rainfall in Lagos State, confirming a wet season from April to October and a dry season from November to March. The presence of outliers and variations in peak rainfall months suggests occasional extreme weather events, which should be considered in climate modeling and flood risk assessment.

Stationarity Test

Before using a time series for modeling, it is essential to test for stationarity to ensure that statistical properties such as mean and variance remain constant over time. If a time series is non-stationary, differencing techniques must be applied to transform it into a stationary form, which is crucial for effective time series modeling, including ARIMA and SARIMA models.

To assess the stationarity of the series, unit root tests are commonly used. In this analysis, the Augmented Dickey-Fuller (ADF) test was employed to examine the stationarity of the Lagos State Rainfall series. The ADF test helps determine whether a unit root is present in the data, indicating non-stationarity.

The test results showed that the initial IR series was non-stationary, requiring transformation. First seasonal differencing was applied to eliminate trends and seasonality. After applying this transformation, the Dickey-Fuller test statistic was -3.8786, with a p-value of 0.02101, indicating that the series became stationary at a 5% significance level.

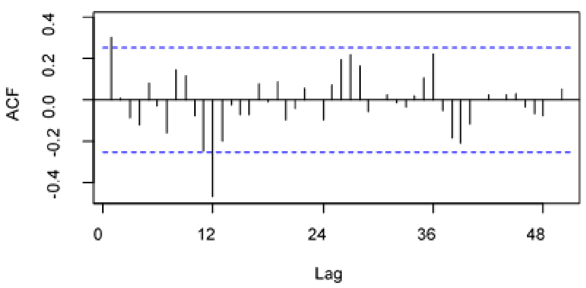

Figure 4: ACF Plot of seasonal differenced series

Additionally, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots (as shown in Figures 4 and 5) further confirmed that the Lagos State rainfall series became stationary after the first seasonal differencing. The ACF plot showed rapid decay, while the PACF plot exhibited a significant cutoff after a few lags, both of which are characteristics of a stationary time series.

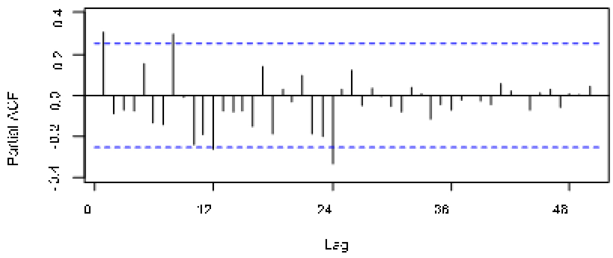

Figure 5: PACF Plot of seasonal differenced series


Figure 6: Seasonal differenced data

The Model

This study analyzed the models using Akaike Information Criterion (AIC) as penalty function statistics. Table 2 presents the Akaike Information Criterion (AIC) values for different ARIMA models fitted to the dataset. The AIC is a crucial measure used in model selection, where a lower AIC value indicates a better model fit while also considering model complexity. Among the models evaluated, the ARIMA (2,0,4) (2,2,1) (12) model has the lowest AIC value of 607.09**,** making it the most optimal choice for this dataset. Closely following this is the ARIMA (2,0,2) (2,2,0) (12) model**,** with an AIC of 607.13, suggesting that it is also a strong candidate Other models, such as ARIMA (2,0,3) (2,2,0) (12) and ARIMA (2,1,5) (2,2,0) (12)**,** have slightly higher AIC values **(**608.91 and 608.55, respectively**),** indicating that they may not fit the data as well as the top two models. The ARIMA (3,0,3) (2,2,0) (12) model**,** with the highest AIC of 610.04, is the least optimal among those tested.'

TABLE 2: Model fitting

Model	AIC
ARIMA (2, 0,2) (2, 2, 0) (12)	607.13
ARIMA (2, 0,3) (2, 2, 0) (12)	608.91
ARIMA (2, 0,4) (2, 2, 1) (12)	607.09
ARIMA (2, 1,5) (2, 2, 0) (12)	608.55
ARIMA (3, 0,3) (2, 2, 0) (12)	610.04

Table 2 presents the estimated parameters for the ARIMA (2,0,4) (2,2,1) (12) model, which was selected based on its optimal fit criteria, particularly its low Akaike Information Criterion (AIC) value. These parameter estimates provide insights into the contributions of the autoregressive (AR), moving average (MA), and seasonal components to the overall model performance. The AR terms capture the dependence of the current rainfall values on past observations, while the MA terms account for short-term fluctuations by modeling the relationship between past forecast errors. Additionally, the seasonal parameters help in addressing annual variations in rainfall patterns, ensuring that the model effectively captures long-term trends and periodic fluctuations. These estimates play a crucial role in determining the predictive capability of the model and its suitability for forecasting future rainfall trends in Lagos State.

TABLE 3: Parameter Estimation

Variable	Coefficients
AR1	-0.7199
AR2	-0.7795
MA1	1.0769
MA2	1.2073
MA3	0.4210
MA4	-0.0776
SAR1	-0.7199
SAR2	-0.6059
SMA1	-0.9954

$$y_t = -0.7199y_{t-1} - 0.7795y_{t-2} - 0.7199Y_{t-12} - 0.6059Y_{t-24} + 1.0769e_{t-1} + 1.2073e_{t-2} + 0.4210e_{t-3} - 0.0776e_{t-4} - 0.9954E_{t-12} + e_t$$

Where:

Y_{t-s} is the seasonally lagged terms (t-s)

e_t is the white noise error at time t

E_{t-s} is the seasonal error terms (t-s)

Diagnostic Check

Table 4 presents the adequacy assessment of different Seasonal Autoregressive Integrated Moving Average (SARIMA) models for predicting rainfall in Lagos State. The models are evaluated using two key statistical error measures: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). RMSE measures the average magnitude of prediction errors, with lower values indicating better performance, while MAE represents the average absolute difference between actual and predicted values, where smaller values signify greater accuracy.

Among the models tested, the SARIMA (2,0,4) (2,2,1) [12] model demonstrates the best predictive performance, achieving the lowest RMSE of 54.20841 and the lowest MAE of 32.48127. This suggests that it is the most suitable model for capturing the rainfall pattern in Lagos State. Other models, such as SARIMA (3,0,3) (2,2,0) [12], also perform reasonably well, with an RMSE of 62.66010 and an MAE of 35.39133, though not as effectively as the best-performing model.

TABLE 4: Model Adequacy

Model	RMSE	MAE
SARIMA (2, 0,2) (2, 2, 0) [12]	68.26932	39.66393
SARIMA (2, 0,3) (2, 2, 0) [12]	67.72430	39.65817
SARIMA (2, 0,4) (2, 2, 1) [12]	54.20841	32.48127
SARIMA (2, 1,5) (2, 2, 0) [12]	67.72430	39.65817
SARIMA (3, 0,3) (2, 2, 0) [12]	62.66010	35.39133

Foresting

Figure 9 presents the forecasted monthly rainfall in Lagos State using the SARIMA (2,0,4)(2,2,1)[12] model. The black line represents historical rainfall data, while the blue line signifies the predicted values. The shaded gray region represents the confidence interval, which captures the uncertainty in the forecast. From the figure, it is evident that the SARIMA model effectively captures the seasonal patterns in rainfall, with periodic peaks and troughs. The forecast suggests an upward trend in rainfall towards the later months, with significant fluctuations. Figure 10 further shows how forecasted rainfall exhibits a similar pattern to the observed rainfall in the year 2022. The peak rainfall months (e.g., September and October) show significant increases in both observed and forecasted values.

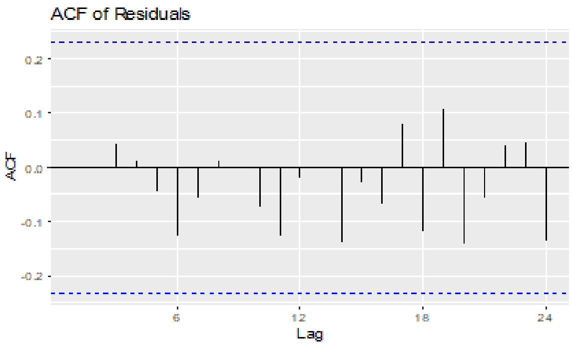

Figure 7: ACF plot of SARIMA (2,0,4) (2,2,1) Residuals

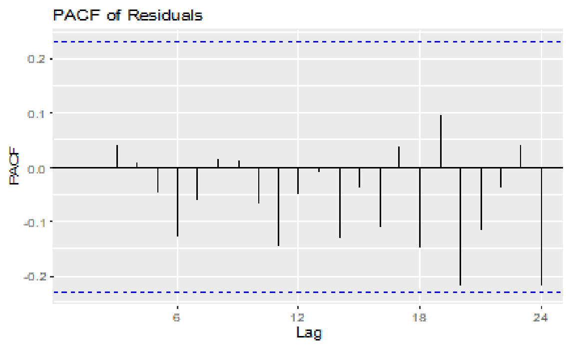

Figure 8: PACF plot of SARIMA (2,0,4) (2,2,1) Residuals

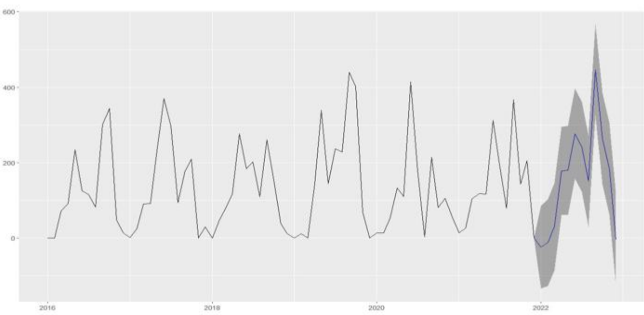

Figure 9: Forecasted monthly rainfall (mm) with SARIMA (2, 0,4) (2, 2, 1) [12]

Adetunji K. Ilori, Adebisi Michael, Fatunsin L. Modupe, Olaiya O. O, and Toyosi Adebambo, "Modelling Lagos State, Nigeria Rainfall using Real Life Data: An application of SARIMA Model," *International Journal of Multidisciplinary Research and Publications (IJMRAP)*, Volume 7, Issue 10, pp. 122-127, 2025.

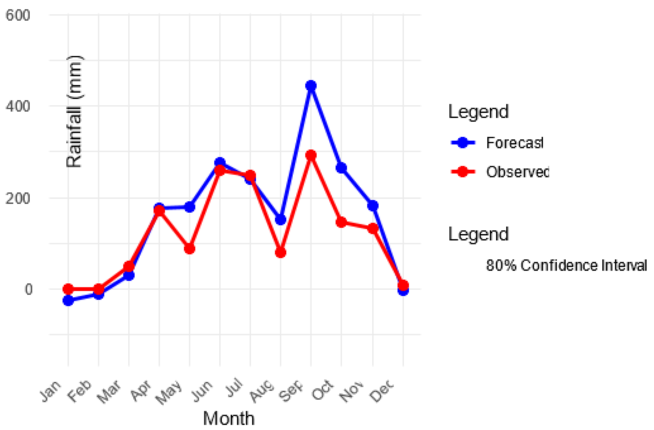

Figure 10: Forecasted vs Observed Rainfall in Lagos State (2022)

V. Conclusion and Recommendation

This study successfully developed and evaluated a SARIMA-based predictive model for rainfall forecasting in Lagos State. By systematically analyzing historical rainfall data and dividing it into training (2016–2021) and testing (2022) subsets, the study was able to identify seasonal patterns, long-term trends, and fluctuations. The selection of the SARIMA (2,0,4) (2,2,1) [12] model was based on statistical criteria, including the Akaike Information Criterion (AIC), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE), all of which indicated its superior performance in capturing rainfall variability.

The stationarity of the dataset was confirmed through unit root testing, and the first seasonal differencing transformation ensured that the time series met the necessary modeling assumptions. Further validation was conducted using ACF and PACF plots, which reinforced the appropriateness of the SARIMA model for forecasting Lagos State's rainfall patterns.

The model's forecasts for the year 2022, reveal the effectiveness of SARIMA (2,0,4) (2,2,1) [12] in predicting seasonal trends and rainfall distribution. The model's ability to capture major trends while allowing for uncertainty through confidence intervals makes it a valuable tool for climate monitoring and decision-making in sectors such as agriculture, disaster risk management, and urban planning.

Conflict of interest: The authors declare that there is no conflict of interest

References

[1] Ette and Etuk (2013). Modelling Monthly Rainfall Data of Port Harcourt, Nigeria by Seasonal Box-Jenkins Methods. International Journal of Sciences.

[2] Ekwe, M. C., Nwosu, K. I., & Okoye, B. C. (2014). Analysis of rainfall patterns in Nasarawa State, Nigeria. Journal of Environmental Management, 134, 50–56. https://doi.org/10.1016/j.jenvman.2013.12.021

[3] John Chisimkwuo, George Uchechukwu and Chukwuemeka Okezie Sampson, (2014). Time series analysis and forecasting of monthly maximum temperature in South Eastern Nigeria, Internat. J. Innovative Res. Develop. 3(1).

[4] Jones, M., & Wang, P. (2021). Deep Learning for Seasonal Rainfall Prediction: A Review. Climate Informatics Journal, 9(2), 112-129.

[5] Kreibich, H., Blauhut, V., Aerts, J. C. J. H., Bouwer, L. M., Van Lanen, H. A. J., Mejia, A., Mens, M., & Van Loon, A. F. (2019). How to improve attribution of changes in drought and flood impacts. Hydrological Sciences Journal, 64(1), 1–18. https://doi.org/10.1080/02626667.2018.1558367

[6] Michael Chibuike Ekwe, Johnah Kunda Joshua, Johnson Eze Igwe and Aekunle Ayodotun Osinowo, (2014). Mathematical study of monthly and annual rainfall trends in Nasarawa State, Nigeria, IOSR J. Math. 10(1), 56- 62.

[7] Nnaji, C. C. (2011). Time series analysis of monthly rainfall in Nigeria with emphasis on self-organized criticality. Journal of Science and Technology (Ghana), 31(1), 75–86. https://doi.org/10.4314/just.v31i1.64900

[8] Sambo Uba Emmanuel and H. R. Bakari, (2015). An application of time series analysis in modelling monthly rainfall data for Maiduguri, North East Nigeria, Math. Theo. Modell. 5(11),.

[9] Usoro, A. E., & Awakessien, C. E. (2018). Time series modelling of rainfall data in different locations in Nigeria. JP Journal of Fundamental and Applied Statistics, 14(1 & 2), 1–22.

[10] World Bank Group. (2021). Comprehensive dataset and analysis tools for Understanding historical and projected climate data. Retrieved from https://climateknowledgeportal.worldbank.org