Forecasting Rainfall Patterns in the Naivasha Region Using ARIMA and SARIMA Models

Florence Awino Adem
Master Student, Department of Statistics and Actuarial Science
Maseno University, Kenya.
mflowad@gmail.com

Dr.Thomas Mawora
Lecturer, Department of Statistics and Actuarial Science
Maseno University, Kenya.

ARTICLE INFO

Key Words:

Seasonality
Rainfall variability
SARIMA
ARIMA

ABSTRACT

Rainfall variability poses significant challenges to agricultural planning and water-resource management in the Naivasha region of Kenya. This study models and forecasts monthly rainfall using the Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average approaches. Monthly data from 1990–2024 were analyzed, revealing strong seasonal patterns and high variability. Stationarity tests confirmed that the series was stationary in levels, while autocorrelation analysis indicated significant dependence at a 12-month lag. Model comparison showed that the SARIMA(1,0,0)(2,0,0)$_{12}$ model outperformed ARIMA(1,0,0), with lower AIC and improved forecast accuracy. Forecasts suggest stable mean rainfall with increasing uncertainty over time. The results highlight the importance of incorporating seasonality for reliable rainfall prediction in semi-arid regions.

INTRODUCTION

1.0 BACKGROUND

Rainfall variability is a key determinant of agricultural productivity, hydrological balance, and socio-economic stability in Kenya, particularly in the Naivasha region of Nakuru County, which supports floriculture, horticulture, and dairy farming and is closely linked to the ecology of Lake Naivasha [1]. Accurate understanding and forecasting of rainfall patterns are essential for irrigation planning, pasture management, and climate-resilient decision-making [2].

Recent studies indicate that East Africa has experienced significant shifts in rainfall patterns and increased variability, influenced by large-scale climate drivers such as ENSO and the Indian Ocean Dipole [3]. These changes, together with observed hydrological fluctuations in the Rift Valley, highlight the need for reliable localized rainfall models [4].

Time series approaches, particularly the Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average, provide effective and interpretable tools for modeling and forecasting rainfall [5]. However, accurate prediction remains challenging due to strong seasonal patterns and variability [6]. Locally developed models using long-term data from institutions such as KALRO are therefore essential for capturing site-specific climatic dynamics and supporting agricultural and water-resource management decisions [7].

.

.

1.2 PROBLEM STATEMENT

Accurate modelling and forecasting of rainfall are essential for understanding temporal climate dynamics. However, rainfall time series are often characterized by strong seasonality and variability, which pose challenges for traditional forecasting methods. Classical time series models such as the Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average are widely used due to their ability to capture temporal dependence and seasonal patterns.

Despite their usefulness, these models may produce unreliable forecasts if seasonal structures are not adequately specified or compared. In many localized contexts, including the Naivasha region, there is limited evidence on the relative performance of seasonal and non-seasonal models for rainfall forecasting. This creates a gap in identifying the most appropriate modelling framework for accurate prediction.

Therefore, this study seeks to evaluate and compare ARIMA and SARIMA models in modelling and forecasting rainfall patterns in the Naivasha region, with the aim of improving forecast accuracy and supporting informed decision-making in agriculture and water-resource management.

.

1.3. OBJECTIVE OF THE STUDY

The broad objective of this study was to analyze and forecast monthly rainfall patterns in the Naivasha region using the Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average, and to evaluate their forecasting performance using appropriate model diagnostics and accuracy measures.

2.0 LITERATURE REVIEW

Rainfall and precipitation have been modelled using a wide range of approaches, including classical time series models (ARIMA/SARIMA), state-space models, generalized linear models, stochastic weather generators, extreme-value models, and machine learning methods. Hybrid approaches that combine statistical and data-driven techniques have also gained increasing attention for improving predictive performance [8]. For medium-term forecasting of monthly rainfall totals, the Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average remain widely used due to their interpretability, computational efficiency, and strong theoretical foundation [9].

Several studies in Kenya and East Africa have applied ARIMA and SARIMA models for rainfall forecasting at seasonal and monthly scales. SARIMA models have been successfully implemented in counties such as Nyeri, Uasin Gishu, and Nairobi, demonstrating reasonable short-term predictive performance [10]. Most studies follow the Box–Jenkins methodology involving stationarity testing, identification using ACF and PACF, parameter estimation, and diagnostic checking [11]. However, limitations have been observed, particularly in capturing abrupt changes and extreme rainfall events, which can reduce forecast accuracy [12].

Rainfall in East Africa is strongly influenced by climate variability drivers such as ENSO and the Indian Ocean Dipole, resulting in high inter-annual variability and irregular rainfall distribution [13]. Such variability introduces complexity into forecasting, especially in semi-arid regions where rainfall is both seasonal and highly unpredictable.

Hydrological and climatological studies in the Lake Naivasha basin highlight complex rainfall–runoff interactions and strong sensitivity of lake levels to rainfall variability and human water use [14]. Recent evidence shows significant fluctuations in Rift Valley lake levels, including Naivasha, with important ecological and socio-economic implications for surrounding communities [15].

Although ARIMA and SARIMA models are widely used in Kenyan rainfall studies, most applications focus on regional or county-level forecasting rather than localized basin-scale modelling. Many studies emphasize hydrological impacts and land-use effects rather than robust statistical forecasting frameworks based on long-term meteorological records [14]. Increasing land-use change and water abstraction in the Naivasha basin further highlight the need for improved localized forecasting models.

Despite extensive use of classical time series methods in rainfall analysis, there remains limited application of basin-specific SARIMA modelling for Naivasha rainfall data. This study therefore focuses on applying and evaluating ARIMA and SARIMA models for monthly rainfall in the Naivasha region to improve forecasting accuracy and support agricultural and water resource management.

3.0MATERIALS AND METHODS

3.1 Area of the study

The study was conducted in Naivasha, Nakuru County, Kenya, located within the Rift Valley region. The area experiences a semi-arid to semi-humid climate characterized by strong seasonal rainfall patterns influenced by the Intertropical Convergence Zone (ITCZ) and large-scale climate systems such as the El Niño–Southern Oscillation (ENSO) [8]. The region supports agriculture, including horticulture, floriculture, and dairy farming, which are highly dependent on rainfall variability [14]. The dataset was obtained from the Kenya Agricultural and Livestock Research Organization (KALRO) Dairy Research Institute meteorological station in Naivasha [7].

3.2 Ethical considerations

The rainfall data were obtained from KALRO with formal authorization. The dataset contained no personal or confidential information, and all analyses complied with institutional guidelines for secondary data use.

3.3 Data collection and mode of analysis

The study used monthly total rainfall data recorded at the KALRO Naivasha meteorological station from January 1990 to December 2024 [7]. The data were measured in millimetres, providing a long-term series suitable for capturing rainfall variability, seasonality, and long-run climatic patterns [14].

The main objective of this study was to analyze and forecast monthly rainfall patterns in Naivasha using time series models. This was operationalized through four specific objectives.

First, the structure and behaviour of rainfall were examined using exploratory data analysis, including time-series plots, descriptive statistics, and seasonal decomposition to identify trend, seasonal, and irregular components.

Second, the statistical properties of the series were assessed using Augmented Dickey–Fuller and KPSS tests for stationarity, supported by ACF and PACF analysis. Differencing was applied where necessary to achieve stationarity [12].

Third, forecasting models were developed using the Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average frameworks, estimated through the Box–Jenkins approach and maximum likelihood methods [10].

Finally, model performance was evaluated using AIC, BIC, Ljung–Box diagnostics, and forecast accuracy measures including RMSE, MAE, and MAPE [9].

4.0 RESULTS

4.1 Descriptive Statistics

Table 4.1 presents the summary statistics of monthly rainfall in the Naivasha region for the period January 1990 to December 2024. The minimum recorded rainfall was 0.00 mm, indicating the occurrence of completely dry months, while the maximum was 94.57 mm, reflecting periods of intense precipitation.

The median rainfall was 8.36 mm, whereas the mean was slightly higher at 11.31 mm, indicating a positively skewed distribution driven by occasional high rainfall events. The first and third quartiles were 4.09 mm and 15.32 mm, respectively, showing moderate variability in monthly rainfall amounts.

The difference between the mean and median suggests the presence of extreme rainfall observations, a common feature of climatic time series. Overall, the results indicate that rainfall in the Naivasha region is highly variable and skewed, characteristics that are important for time series modelling and justify the use of methods capable of capturing non-normal distributional behavior.

Figure 4.1 illustrates the monthly rainfall time series for the Naivasha region from 1990 to 2024. The plot reveals substantial variability in rainfall amounts across the study period, character ized by frequent fluctuations and intermittent extreme rainfall events. Periods of low rainfall, including several near-zero observations, indicate the occurrence of dry months, while sharp spikes reflect episodes of intense precipitation. The absence of a clear long-term upward or downward trend suggests that the rainfall series may be weakly stationary in mean; however, the presence of volatility clustering is evident, with periods of heightened variability followed by relatively calmer intervals.

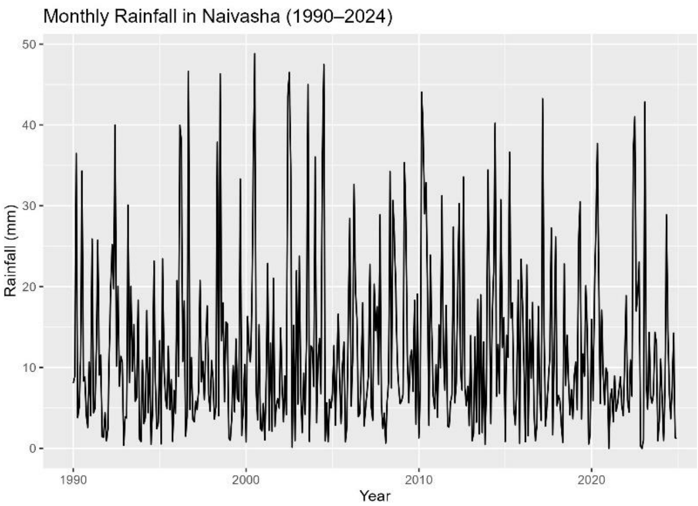

Figure 4. 1: Monthly Rainfall in Naivasha Region from 1990 to 2024

Table 4. 1: Summary Statistics of Monthly Rainfall in Naivasha Region (1990–2024)

Statistic	Rainfall (mm)
Minimum	0.00
First Quartile (Q1)	4.09
Median	8.36
Mean	11.31
Third Quartile (Q3)	15.32
Maximum	94.50

4.2 Decomposition of Monthly Rainfall

Figure 4.2 presents the seasonal-trend decomposition of the monthly rainfall series for the Naivasha region over the period 1990 to 2024. The decomposition separates the observed series into four components: observed data, trend, seasonal, and remainder.

The trend component shows gradual fluctuations over time, indicating long-term variability in rainfall without a clear upward or downward trend. This suggests that rainfall in the region is characterized more by variability than by persistent long-term change.

The seasonal component reveals a clear and stable annual cycle with a 12-month periodicity, reflecting the strong and consistent seasonal rainfall pattern typical of the region. The persistence of this pattern across the study period confirms pronounced seasonality in the data, thereby supporting the use of seasonal time series models such as the Seasonal Autoregressive Integrated Moving Average for effective rainfall forecasting.

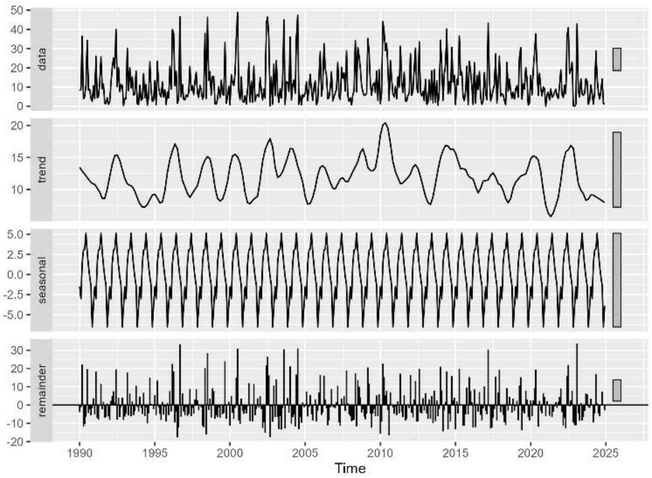

Figure 4. 2: Decomposition of monthly rainfall in the Naivasha region (1990–2024)

4.2 Stationarity test

KPSS test results presented in Table 4.2 show a test statistic of 0.0821 with a p-value of 0.10. Since the p-value exceeds the 5% significance level, the null hypothesis of level stationarity is not rejected, further confirming that the series is stationary.

Table 4.2: KPSS Test Results for Level Stationarity of Monthly Rainfall

Statistic	Value
KPSS Statistic	0.0821
Truncation Lag	5
p-value	0.10
Decision (5% level)	Fail to reject H_0
Conclusion	Level stationary

4.3 Autocorrelation Function and Partial Autocorrelation Function

Figure 4.3 and Figure 4.4 present the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the monthly rainfall series for the Naivasha region, with the dashed horizontal lines indicating the 95% confidence limits for statistical significance. The ACF results show that most autocorrelation coefficients lie within the confidence bounds, indicating weak serial dependence and no clear slow decay pattern, which is consistent with the stationarity results obtained from the KPSS tests in Table 4.2 [12]. Although most correlations are insignificant, small spikes at selected lags suggest limited short-term dependence and possible seasonal effects. Similarly, the PACF plot shows that most partial autocorrelations fall within the

Table 4.4: SARIMA(1,0,0)(2,0,0)$_{12}$ Model Results

Parameter	Estimate	Standard Error	Interpretation
AR(1)	0.0852	0.0490	Weak non-seasonal persistence
SAR(1)	0.0440	0.0479	Mild annual dependence
SAR(2)	0.1434	0.0488	Significant biannual seasonal effect
Mean	11.9172	0.6812	Average monthly rainfall (mm)

confidence intervals, indicating weak direct autoregressive structure, with no clear cut-off at early lags [11]. Minor seasonal spikes further support the presence of weak seasonal dependence, consistent with previous decomposition results [10]. Overall, the combined ACF and PACF analysis indicates limited persistence in the rainfall series and supports the use of parsimonious Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average models with low-order parameters for modelling monthly rainfall dynamics in the Naivasha region [9].

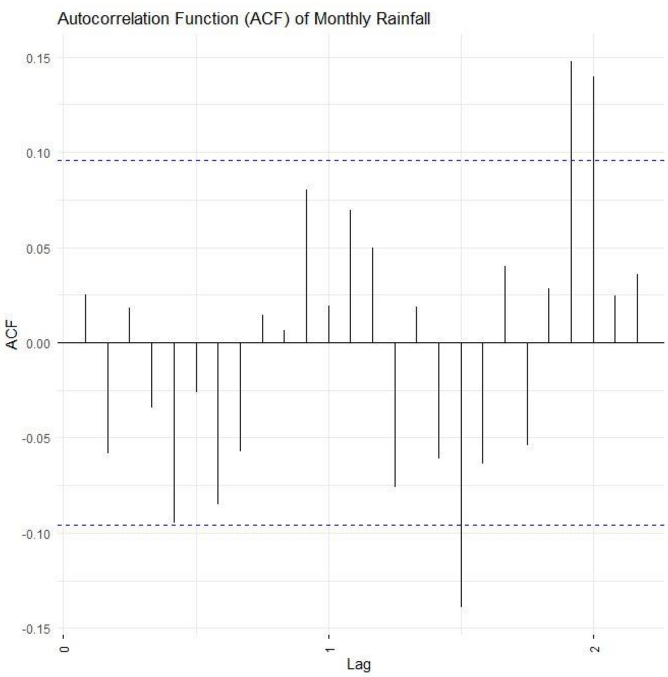

Figure 4. 3: Autocorrelation function of monthly rainfall in the Naivasha region

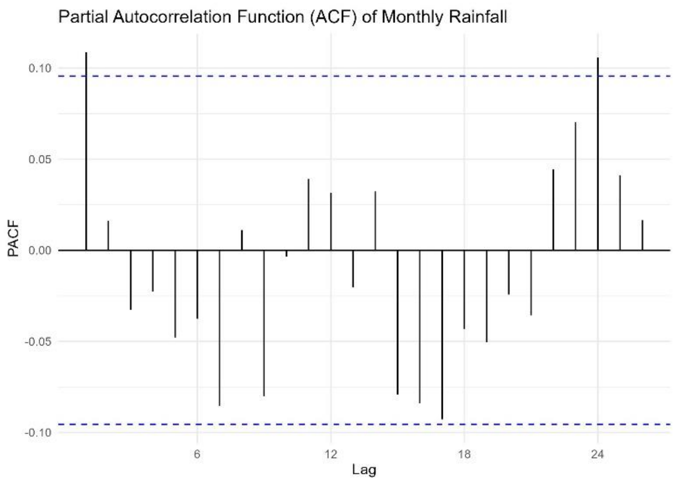

Figure 4. 4:Partial Autocorrelation function of monthly rainfall in the Naivasha region

4.3 ARMA (1,0,0) and SARIMA(1,0,0)(2,0,0)$_{12}$ Model Results

Table 4.3: ARMA (1,0,0) Model Results

Parameter	Estimate	Standard Error	Interpretation
AR(1)	0.1088	0.0485	Weak positive persistence
Mean	11.9624	0.5813	Average monthly rainfall (mm)

Table 4.3 presents the estimated parameters of the Autoregressive Integrated Moving Average (1,0,0) fitted to the monthly rainfall series. The autoregressive coefficient is positive and relatively small ($\phi_1 = 0.1088$), indicating weak persistence in rainfall dynamics. This suggests that current rainfall levels are only minimally influenced by rainfall in the previous month.The estimated mean monthly rainfall is 11.96 mm, which is consistent with the descriptive statistics reported earlier. The relatively small standard error associated with the mean indicates a stable long-term average rainfall level over the study period.

Tables 4.4 summarize the estimation results and performance metrics of the Seasonal Autoregressive Integrated Moving Average (1,0,0)(2,0,0)$_{12}$ model fitted to the monthly rainfall series. The seasonal period (s = 12) was selected to reflect the annual rainfall cycle inherent in monthly climatic data. In the Naivasha region, rainfall is strongly influenced by large-scale atmospheric systems such as the Intertropical Convergence Zone (ITCZ), which generates recurring seasonal patterns within each year. Therefore, a 12-month seasonal structure appropriately captures the observed annual cycle. The seasonal autoregressive order P = 2 was determined from inspection of the seasonal ACF and PACF plots, which revealed significant spikes at lags 12 and 24 months. This indicates dependence not only on rainfall from the same month in the previous year but also on rainfall two years prior, justifying the inclusion of a second-order seasonal autoregressive component.

4.4 Goodness-of-Fit Statistics for ARIMA and SARIMA Models

Table 4.5 compares the goodness-of-fit statistics for the ARIMA and SARIMA models fitted to monthly rainfall data in the

Naivasha region. The SARIMA model shows slightly better performance, indicated by lower AIC, AICc, and error variance (σ^2), as well as a higher log-likelihood, suggesting improved model fit. Although the BIC is slightly higher for SARIMA, the overall results indicate that incorporating seasonality improves model adequacy for capturing rainfall dynamics.

Table 4.5: Goodness of fit

Statistic	ARIMA(1,0)	SARIMA(1,0,0)(2,0,0)$_{12}$
Log-Likelihood	-1588.33	-1583.72
AIC	3182.65	3177.44
AICc	3182.71	3177.59
BIC	3194.77	3197.65
σ^2	113.30	111.30

Table 4.6 compares the Ljung–Box test results for residuals of the ARIMA and SARIMA models. The ARIMA model shows evidence of remaining autocorrelation in the residuals ($p = 0.0485$), indicating model inadequacy. In contrast, the SARIMA model exhibits no significant residual autocorrelation ($p = 0.6175$), suggesting that it adequately captures the dependence structure in the rainfall series and provides a better overall model fit.

Table 4.6: Ljung–Box Test Results for ARIMA and SARIMA Model Residuals

Statistic	ARIMA(1,0,0)	SARIMA(1,0,0)(2,0,0)$_{12}$
Q-statistic	35.303	18.496
Degrees of Freedom	23	21
Number of Lags	24	24
Model Degrees of Freedom	—	3
p-value	0.0485	0.6175
Decision (5% level)	Reject H_0	Fail to reject H_0

4.4 Forecasting

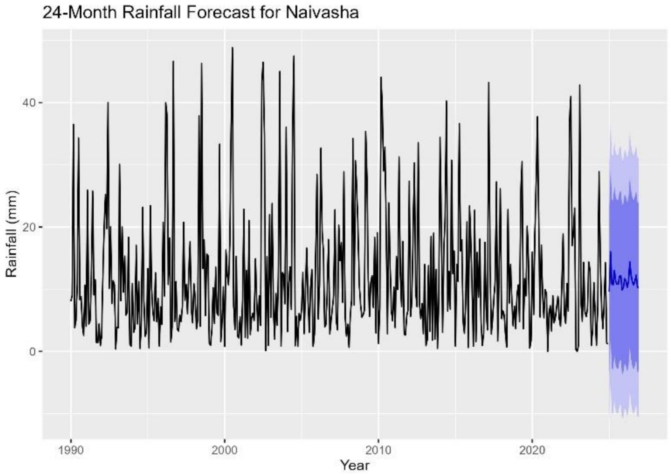

Figure 4.5: Twenty-four-month rainfall forecast for Naivasha based on the selected SARIMA model

Figure 4.5 presents the 24-month ahead rainfall forecasts for the Naivasha region based on the fitted Seasonal Autoregressive Integrated Moving Average (1,0,0)(2,0,0)$_{12}$ model. The observed rainfall series is shown in black, while the forecasted mean values are represented by the blue line. The shaded bands indicate the 80% and 95% prediction intervals, reflecting uncertainty in future rainfall estimates.

From the figure, the forecasted rainfall exhibits a relatively stable pattern, fluctuating around the historical monthly average of approximately 10–12 mm. This suggests that the model does not indicate any strong upward or downward trend over the forecast horizon, but rather maintains a stable seasonal structure consistent with past behavior.

However, the widening of the prediction intervals as the forecast horizon increases is clearly evident in the figure. This indicates increasing uncertainty in long-term forecasts, which is typical of rainfall processes due to their inherent variability and dependence on complex climatic drivers. The broader intervals in later months also highlight the possibility of extreme rainfall fluctuations, even when mean values remain stable.

Overall, Figure 4.5 shows that the SARIMA model effectively captures the seasonal dynamics of rainfall in Naivasha while producing stable mean forecasts. At the same time, the expanding prediction bands emphasize the importance of accounting for uncertainty when using rainfall forecasts for agricultural planning, water resource management, and climate risk assessment.

DISCUSSION

5.0.1 Analysis and Forecasting of Monthly Rainfall Using ARIMA and SARIMA Models.

Guided by the study objective of analyzing and forecasting monthly rainfall patterns in Naivasha using time series models, the results provide important insights into the behaviour and predictability of rainfall in the region. The descriptive analysis showed that rainfall is highly variable and positively skewed, indicating the occurrence of occasional extreme rainfall events, a common feature in East African climatic series [8]. This variability underscores the need for robust modelling approaches capable of capturing both central tendency and fluctuations.

The stationarity analysis using the Augmented Dickey–Fuller and KPSS tests confirmed that the rainfall series is stationary in levels, implying stable statistical properties over time. This finding supports the direct application of time series models without the need for differencing and is consistent with established time series theory [11]. The agreement between the two tests further strengthens confidence in the reliability of the results.

The ACF and PACF analyses revealed weak serial dependence, with most coefficients falling within the confidence bounds. However, the presence of seasonal spikes indicated a clear annual pattern in rainfall behaviour. This finding aligns with the study objective of identifying the underlying structure of the rainfall series and justifies the inclusion of seasonal components in the modelling process [10].

In line with the objective of developing appropriate forecasting models, both the Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average were fitted to the data. The results showed that while both models provided reasonable representations of the rainfall series, the SARIMA model performed better in capturing seasonal dynamics. This was evidenced by lower AIC and AICc values and improved residual behaviour compared to the ARIMA model [9].

Further, the evaluation of model adequacy demonstrated that the SARIMA model successfully eliminated residual autocorrelation, unlike the ARIMA model, which showed signs of model misspecification. This finding directly addresses the study objective of assessing model performance and confirms that incorporating seasonality improves model reliability.

The forecasting results also align with the study objective, showing that rainfall is expected to remain relatively stable over the forecast horizon, with no clear long-term increasing or decreasing trend. However, the widening prediction intervals highlight increasing uncertainty, reflecting the inherently stochastic nature of rainfall processes influenced by complex climatic systems [8].

5.0 CONCLUSION

5.0.1 Model Performance and Forecasting Implications for Rainfall in Naivasha.

This study set out to analyze and forecast monthly rainfall patterns in the Naivasha region using time series modelling techniques. The results demonstrate that rainfall in the region is characterized by variability, weak persistence, and strong seasonal behaviour.

Among the models considered, the Seasonal Autoregressive Integrated Moving Average $(1,0,0)(2,0,0)_{12}$ was found to be the most appropriate for capturing the underlying rainfall dynamics. The model provided a better fit and more reliable forecasts compared to the non-seasonal ARIMA model, primarily due to its ability to account for annual seasonal patterns.

The study further established that future rainfall is likely to remain stable around historical averages, although considerable uncertainty exists in long-term forecasts. This emphasizes the need for caution when using rainfall predictions for planning purposes.

Overall, the findings highlight the importance of incorporating seasonal structures in rainfall modelling and demonstrate the usefulness of SARIMA models in environmental time series analysis. Future research should consider integrating additional climatic variables or advanced modelling approaches to enhance predictive accuracy and better capture the complexity of rainfall variability.

REFERENCES

[1] Becht, R., Odada, E. O., & Higgins, S. (2006). *Lake Naivasha: Experience and lessons learned brief*. International Lake Environment Committee Foundation.

[2] Muthawatta, L. P., et al. (2004). *Lake Naivasha: Water balance and impacts of land use changes*. Hydrological Sciences Journal.

[3] Nicholson, S. E. (2017). *Climate and climatic variability of rainfall over Eastern Africa*. Reviews of Geophysics.

[4] Onywere, S. M., et al. (2012). *Recent lake level changes in the Rift Valley lakes of Kenya*. Hydrological Sciences Journal.

[5] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

[6] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

[7] Olang, L. O., & Fürst, J. (2011). *Effects of land cover change on flood peak discharges in the Nyando River Basin*. Hydrology and Earth System Sciences.

[8] Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.

[9] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

[10] Ogallo, L. A. (2010). *Climate variability and rainfall forecasting in East Africa*. African Journal of Meteorological Studies.

[11] Chatfield, C. (2004). *The Analysis of Time Series: An Introduction* (6th ed.). Chapman & Hall/CRC.

[12] Mwangi, J. W., & Wambua, R. M. (2018). *Limitations of ARIMA models in rainfall forecasting in Kenya*. Journal of Climate Research.

[13] Nicholson, S. E. (2017). *Climate and climatic variability of rainfall over Eastern Africa*. Reviews of Geophysics.

[14] Becht, R., Odada, E. O., & Higgins, S. (2006). *Lake Naivasha: Experience and lessons learned*. International Lake Environment Committee Foundation.

[15] Onywere, S. M., et al. (2012). *Hydrological changes in the Rift Valley lakes of Kenya*. Hydrological Sciences Journal.