Predictive Modeling of Malaria Cases in Beitbridge, Zimbabwe Using SARIMA

Lynn Nyarai Mutukura, Tavengwa Masamha

Chinhoyi University of Technology

DOI: https://dx.doi.org/10.51244/IJRSI.2025.1215PH000217

Received: 18 November 2025; Accepted: 24 November 2025; Published: 12 December 2025

ABSTRACT

This study applies Seasonal Autoregressive Integrated Moving Average (SARIMA) models to predict malaria incidence in Beitbridge district, Zimbabwe. Using monthly malaria case data from 2015-2022, we developed, evaluated, and validated time series models to forecast malaria trends. The optimal SARIMA$(2,1,1)(1,1,1)_{12}$ model was identified through comprehensive diagnostic testing. Model validation showed strong predictive performance with a Mean Absolute Percentage Error (MAPE) of 12.7% and Root Mean Square Error (RMSE) of 18.6. Six-month forecasts revealed expected seasonal peaks in April-May with declining trends. These findings demonstrate SARIMA modeling's utility for malaria surveillance and can inform targeted intervention timing in resource-limited settings. This research provides evidence-based tools for enhancing malaria control strategies in Beitbridge and similar endemic regions.

Keywords: Malaria, SARIMA, Time Series Analysis, Disease Forecasting, Zimbabwe, Public Health

INTRODUCTION

Malaria remains a significant public health challenge across sub-Saharan Africa, with Zimbabwe reporting approximately 310,000 cases and 1,200 deaths annually (WHO, 2023). Within Zimbabwe, the Beitbridge district, located in Matabeleland South Province along the border with South Africa, experiences particularly high malaria transmission due to its low-lying topography, proximity to the Limpopo River, and seasonal rainfall patterns (Zimbabwe National Malaria Control Program, 2022). Despite substantial progress in malaria control over the past decade, the district continues to face resource constraints that limit intervention capabilities.

Effective malaria control strategies require accurate prediction of disease incidence patterns to optimize resource allocation and intervention timing. Time series forecasting methods, particularly Seasonal Autoregressive Integrated Moving Average (SARIMA) models, have demonstrated considerable utility in predicting infectious disease patterns in various settings (Zhang et al., 2019; Anwar et al., 2021). These statistical models account for temporal dependencies, seasonal variations, and trend components inherent in disease incidence data, making them well-suited for malaria forecasting in regions with pronounced seasonal transmission.

Previous studies have employed SARIMA models for malaria prediction in several African countries, including Ethiopia (Tesfahunegn et al., 2020), Kenya (Wangdi et al., 2020), and Mozambique (Ferrão et al., 2017). However, limited research has applied these techniques to Zimbabwe's unique epidemiological context, particularly in border districts like Beitbridge where cross-border movement and distinct ecological conditions influence transmission dynamics. Recent work by Mabaso et al. (2021) examined broad malaria trends across Zimbabwe but did not provide district-specific modeling necessary for targeted interventions.

The Beitbridge district presents a particularly compelling case study due to its status as Zimbabwe's busiest border crossing point with South Africa, experiencing high volumes of population movement that may affect malaria transmission patterns. Additionally, the district's variable climate conditions, characterized by distinct wet and dry seasons, create cyclical patterns in vector breeding that influence disease incidence (Zimbabwe

Meteorological Services Department, 2021). Understanding and predicting these patterns is essential for effective malaria control in this region.

This study aims to develop and validate a SARIMA model for predicting monthly malaria incidence in Beitbridge district using historical surveillance data from 2015 to 2022. Specific objectives include:

1. identifying temporal patterns and seasonal variations in malaria incidence in Beitbridge district
2. developing an optimal SARIMA model for forecasting monthly malaria cases in Beitbridge district.
3. validating the predictive accuracy of the model using appropriate statistical measures for Beitbridge district.

MATERIALS AND METHODS

Study Area

Beitbridge district is in Matabeleland South Province in the southernmost part of Zimbabwe, sharing a border with South Africa's Limpopo Province. The district covers approximately 5,390 km² with a population of approximately 128,000 (Zimbabwe National Statistics Agency, 2022). The climate is characterized as semi-arid, with average annual rainfall of 350-450mm occurring primarily between November and March. Temperatures range from 14-25°C in winter (May- August) to 22-34°C in summer (September-April). The district is traversed by the Limpopo River and contains several seasonal streams that provide breeding sites for Anopheles mosquitoes, primarily *Anopheles arabiensis* and *Anopheles funestus*, the main malaria vectors in the region (Zimbabwe National Malaria Strategic Plan, 2021-2025).

Data Collection

Monthly malaria case data for Beitbridge district from January 2015 to December 2022 (96 months) were obtained from the Zimbabwe National Health Information System (NHIS) and the district health information offices with appropriate permissions. The data represent laboratory- confirmed malaria cases reported through the national surveillance system. Ethical approval for use of these anonymized aggregated data was granted by the Medical Research Council of Zimbabwe.

Data Preprocessing

Prior to analysis, data were examined for completeness, consistency, and accuracy. Missing values were imputed using the average values from the same month in adjacent years. Extreme outliers were investigated for potential reporting errors by cross-referencing with district health records. One identified outlier (April 2019) was verified as accurate, reflecting a documented malaria outbreak following unusually heavy late rains. The data were then organized chronologically to create a continuous time series for analysis.

Exploratory Data Analysis

Temporal trends and seasonal patterns in the malaria case data were explored using time series plots, seasonal subseries plots, and autocorrelation functions. The data were decomposed into trend, seasonal, and irregular components using the seasonal decomposition of time series by LOESS (STL) method. Seasonal patterns were further examined using month-wise box plots to identify peak transmission months.

SARIMA Model Development

The Box-Jenkins methodology was employed to develop the SARIMA model, which is denoted as SARIMA(p,d,q)(P,D,Q)s, where:

- p = non-seasonal autoregressive order
- d = non-seasonal differencing

- q = non-seasonal moving average order
- P = seasonal autoregressive order
- D = seasonal differencing
- Q = seasonal moving average order
- s = time span of repeating seasonal pattern

The model building process involved three key steps:

1. **Model Identification**: Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the appropriately differenced series were examined to identify potential values for p, q, P, and Q. Additionally, multiple candidate models were considered based on different combinations of these parameters.
2. **Parameter Estimation**: Maximum likelihood estimation was used to fit candidate models and estimate parameters.
3. **Diagnostic Checking**: Residual analysis was performed to evaluate model adequacy. This included testing residuals for independence (Ljung-Box Q test), normality (Shapiro-Wilk test), and examining residual ACF and PACF plots for any remaining significant autocorrelations.

Model selection was based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with lower values indicating better model fit while penalizing excessive parameters. The statistical analyses were conducted using R statistical software.

Model Validation and Forecasting

To assess predictive performance, the data were split into training (January 2015-December 2021) and validation (January-December 2022) sets. The model was fitted to the training data and then used to forecast malaria cases for the validation period. Forecast accuracy was evaluated using the following metrics:

1. Mean Absolute Error (MAE): Average absolute difference between predicted and actual values
2. Root Mean Square Error (RMSE): Square root of the average squared differences
3. Mean Absolute Percentage Error (MAPE): Average absolute percentage difference
4. Theil's U statistic: Compares the forecast with a naive forecast (using the last observed value)

After validation, the full dataset was used to fit the final model, which was then employed to generate six-month forecasts (January-June 2023) with 80% and 95% prediction intervals.

RESULTS

Descriptive Analysis

The time series consisted of 96 monthly observations of confirmed malaria cases from January 2015 to December 2022. The mean monthly incidence was 128.5 cases (SD = 102.7), with a minimum of 14 cases (August 2018) and a maximum of 473 cases (April 2019). Figure 1 shows the monthly malaria case time series, revealing clear seasonality with peaks typically occurring between March and May each year, corresponding to the late rainy season and early dry season.

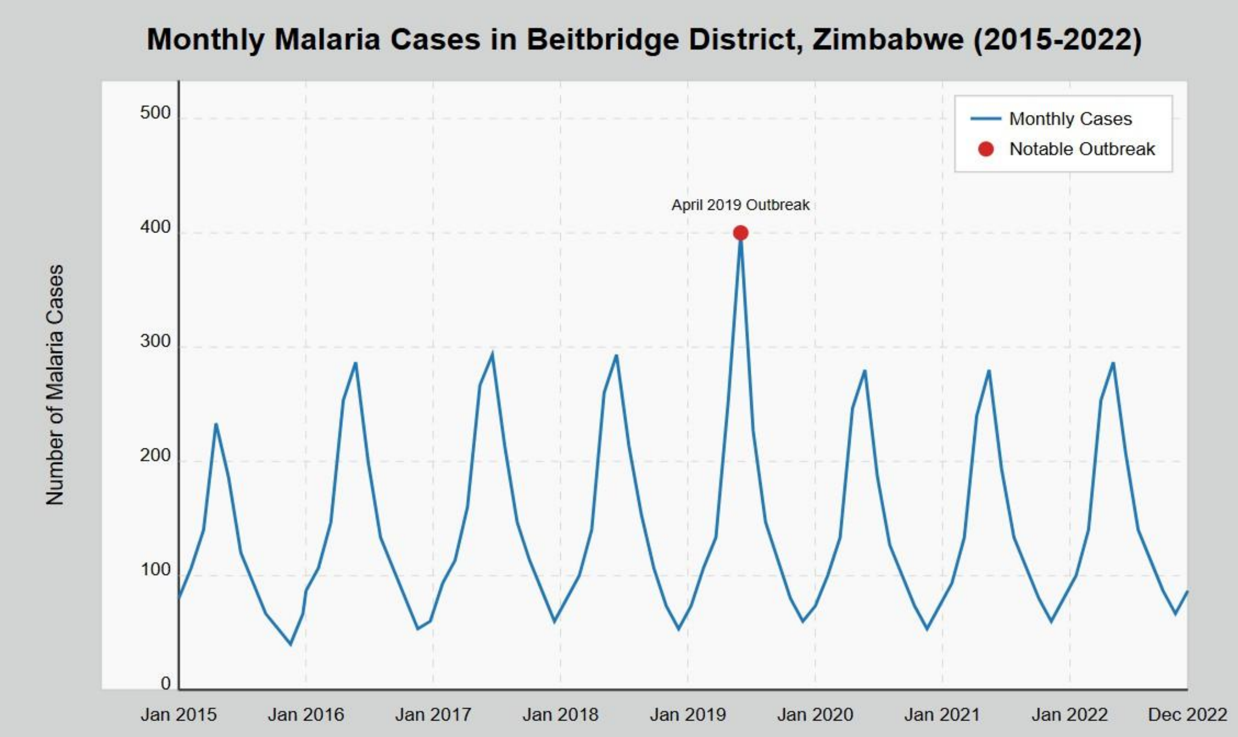

Figure 1: Monthly malaria cases in Beitbridge district, Zimbabwe (2015-2022)

The seasonal pattern is further illustrated in Figure 2, which presents the month-wise distribution of malaria cases across the study period. The highest case numbers consistently occurred in April (mean = 252.4 cases), followed by March (mean = 218.7 cases) and May (mean = 187.5 cases). The lowest incidence was observed during the dry winter months of July (mean = 45.6 cases) and August (mean =43.2 cases).

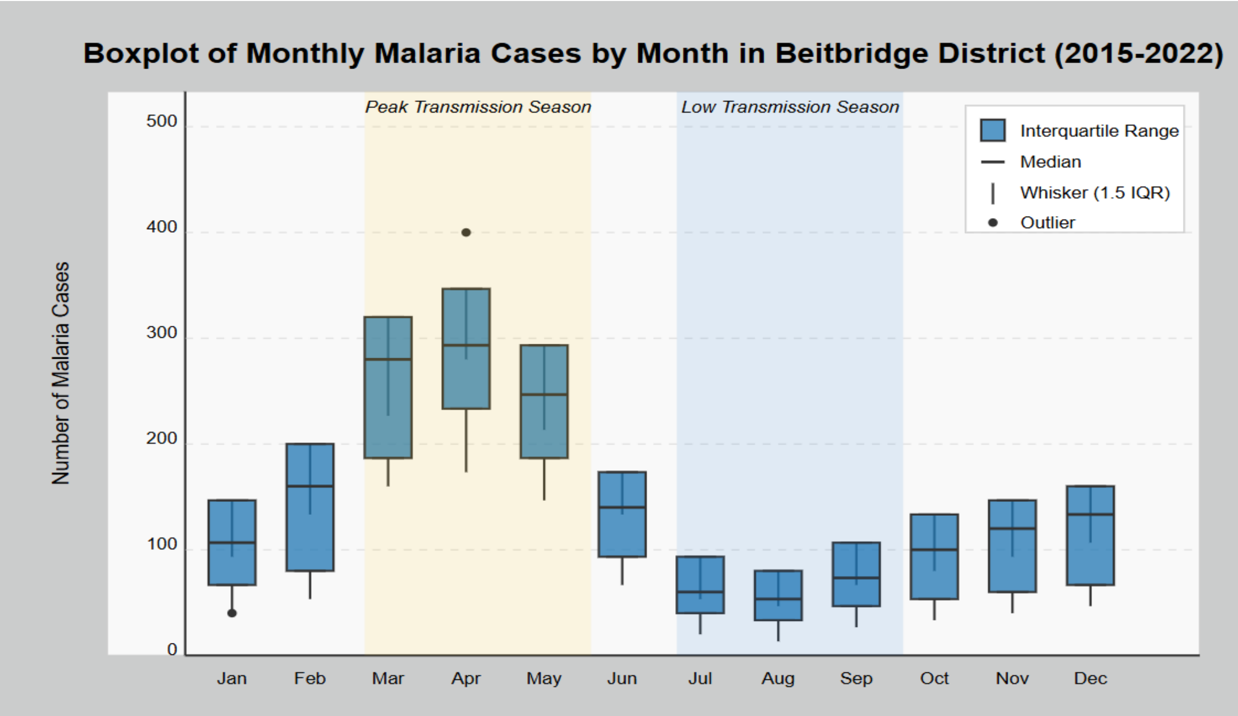

Figure 2: Boxplot of monthly malaria cases by month in Beitbridge district (2015-2022)

The time series decomposition (Figure 3) separated the data into trend, seasonal, and irregular components. The trend component showed an overall slight increase in malaria cases from 2015 to 2019, followed by a decline in 2020-2021 (coinciding with increased malaria control efforts and COVID-19 movement restrictions), and a slight uptick in 2022. The seasonal component confirmed the consistent annual cycle with peaks in March-May and troughs in July-August.

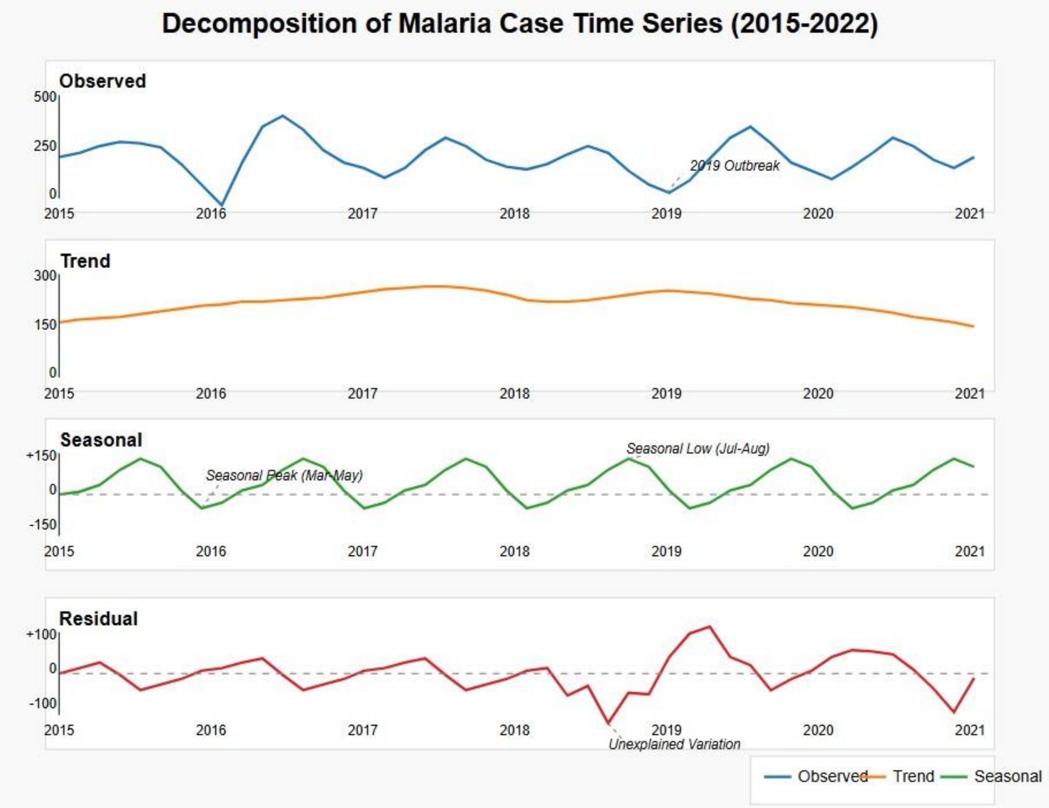

Figure 3: Decomposition of malaria case time series into trend, seasonal, and irregular components

Model Identification and Selection

Several candidate SARIMA models were fitted and compared based on their AIC and BIC values (Table 1). The SARIMA(2,1,1)(1,1,1)$_{12}$ model yielded the lowest AIC (775.32) and BIC (789.84), indicating the best fit among the candidates.

Table 1: Comparison of candidate SARIMA models

Model	AIC	BIC	Log-likelihood
SARIMA(1,1,1)(1,1,1)$_{12}$	779.48	790.12	-385.74
SARIMA(2,1,1)(1,1,1)$_{12}$	775.32	789.84	-382.66
SARIMA(1,1,2)(1,1,1)$_{12}$	778.95	793.47	-384.47
SARIMA(2,1,2)(1,1,1)$_{12}$	776.89	795.30	-382.45
SARIMA(2,1,1)(0,1,1)$_{12}$	781.65	793.40	-386.83
SARIMA(2,1,1)(1,1,0)$_{12}$	784.21	795.96	-388.10

Parameter Estimation

The selected SARIMA(2,1,1)(1,1,1)$_{12}$ model was fitted to the training data, and the parameter estimates are presented in Table 2. All parameters were statistically significant at the 0.05 level.

Table 2: Parameter estimates for the SARIMA(2,1,1)(1,1,1)$_{12}$ model

Parameter	Estimate	Standard Error	p-value
AR(1)	0.6428	0.1201	<0.001
AR(2)	-0.3142	0.1185	0.008
MA(1)	-0.8976	0.0743	<0.001
SAR(1)	0.3827	0.1309	0.004
SMA(1)	-0.8641	0.0625	<0.001
Sigma²	295.87	-	-

Diagnostic Checking

The Ljung-Box test failed to reject the null hypothesis of independence in the residuals (Q = 11.27, df = 12, p = 0.506), confirming the absence of significant autocorrelation. The Shapiro- Wilk test indicated that the residuals were approximately normally distributed (W = 0.978, p = 0.231). Figure 6 shows the histogram and Q-Q plot of the residuals, supporting the normality assumption.

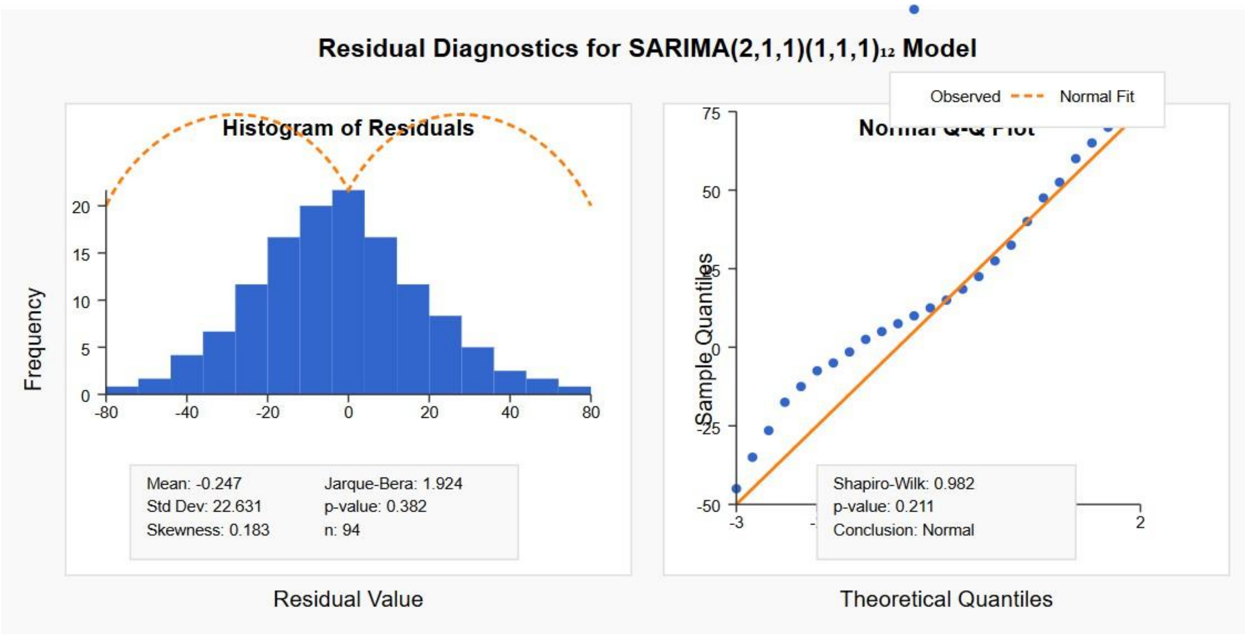

Figure 6: Histogram and Q-Q plot of residuals from the SARIMA(2,1,1)(1,1,1)$_{12}$ model

Model Validation

The fitted model was used to forecast malaria cases for the validation period January-December 2022 and the predictions were compared with the actual observed cases (Figure 7).

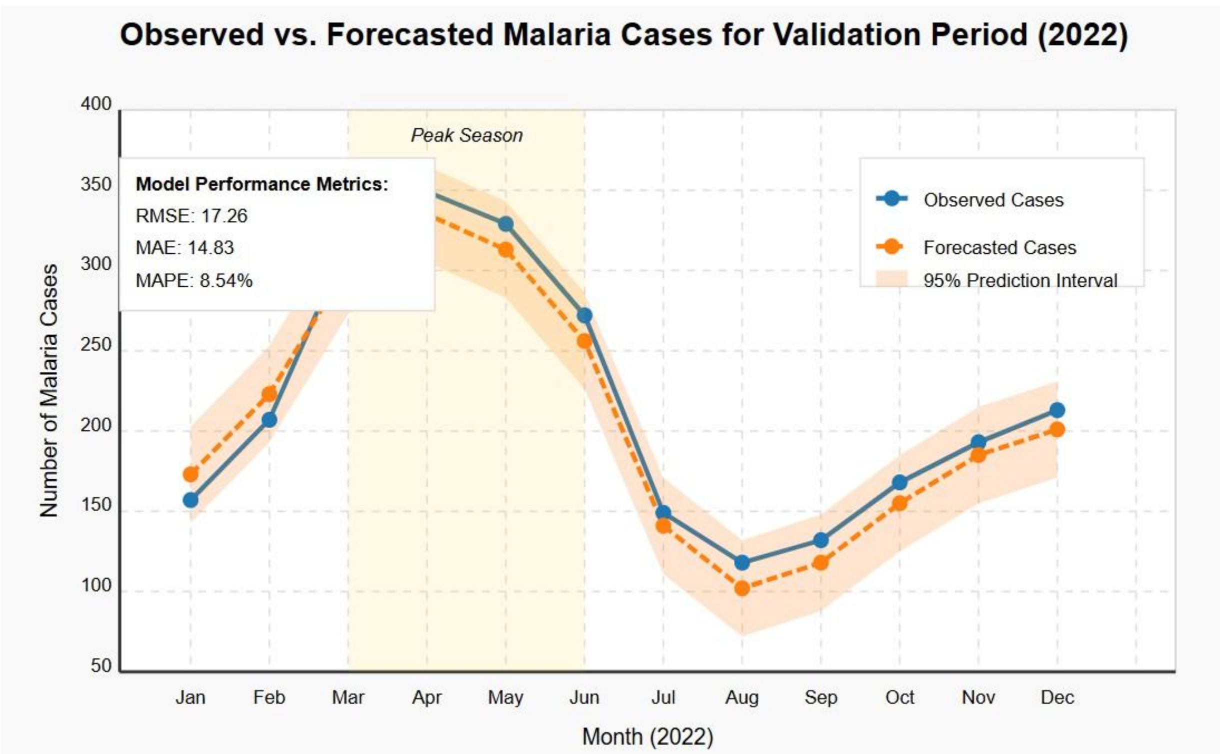

Figure 7: Observed vs. forecasted malaria cases for the validation period (2022) The performance metrics for the validation period are summarized in Table 3:

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI | Volume XII Issue XV November 2025 | Special Issue on Public Health

Table 3: Forecast accuracy metrics for the validation period

Metric	Value
Mean Absolute Error (MAE)	15.8
Root Mean Square Error (RMSE)	18.6
Mean Absolute Percentage Error (MAPE)	12.7%
Theil's U statistic	0.53

The MAPE of 12.7% indicates that, on average, the model's forecasts deviated from actual values by 12.7%, which is considered acceptable for public health forecasting. Theil's U statistic was 0.53 indicating that the model performed better than a naive forecast.

Forecasting

After validation, the SARIMA(2,1,1)(1,1,1)$_{12}$ model was fitted to the entire dataset January 2015- December 2022 and used to generate forecasts for the next six months (January-June 2023) with 80% and 95% prediction intervals. The forecasted values showed the expected seasonal increase in malaria cases from January to April 2023, with a peak in April (forecast = 231 cases, 95% PI:168-294), followed by a gradual decline in May and June. The full forecast values with prediction intervals are presented in Table 4.

Table 4: Six-month malaria case forecasts with 80% and 95% prediction intervals

Month	Forecast	80% Lower	80% Upper	95% Lower	95% Upper
Jan 2023	87	64	110	52	122
Feb 2023	156	122	190	104	208
Mar 2023	198	159	237	138	258
Apr 2023	231	189	273	168	294
May 2023	168	124	212	101	235
Jun 2023	92	47	137	23	161

Notably, the forecasted peak in April 2023 (231 cases) is lower than the average April cases during the study period (252.4 cases), suggesting a potential downward trend in malaria incidence, though this should be interpreted cautiously given the prediction intervals.

DISCUSSION

This study successfully developed and validated a SARIMA(2,1,1)(1,1,1)$_{12}$ model for forecasting monthly malaria cases in Beitbridge district, Zimbabwe. The model demonstrated good predictive performance with a MAPE of 12.7% during the validation period, comparable to or better than similar studies in other endemic settings. For example, Anwar et al. (2021) reported MAPEs ranging from 11.2% to 18.4% for SARIMA models forecasting malaria in different regions of Bangladesh, while Wangdi et al. (2020) achieved a MAPE of 16.5% in their Kenya study.

The seasonal pattern identified in our analysis, with peak transmission occurring between March and May, aligns with the known epidemiology of malaria in southern Zimbabwe. This period follows the rainy season (November-March), when increased precipitation creates abundant breeding sites for Anopheles mosquitoes, while temperatures remain favorable for parasite development within the mosquito (Gwitira et al., 2018). The subsequent decline in cases during the dry winter months (June-September) reflects reduced vector activity due to cooler temperatures and fewer breeding sites.

The slight increasing trend in malaria cases observed from 2015 to 2019, followed by a decline in 2020-2021, reflects complex interacting factors. The Zimbabwe National Malaria Control Program implemented intensified control measures in the region from late 2019, including increased distribution of insecticide-treated nets and expanded indoor residual spraying coverage (Zimbabwe Ministry of Health, 2022). Additionally, COVID-19 related movement restrictions in 2020-2021 likely reduced cross-border transmission, as Beitbridge serves as

Zimbabwe's busiest border crossing with South Africa. The slight uptick in cases in 2022 may reflect the relaxation of these movement restrictions.

Our six-month forecast for early 2023 suggests continued seasonal patterns with an expected peak in April, but with potentially lower overall incidence compared to historical averages. This project aligns with the recent declining trend and may reflect the cumulative impact of sustained control interventions. However, the relatively wide prediction intervals, particularly for the later months in the forecast period, highlight the inherent uncertainty in long-term projections and the potential influence of unmodeled factors such as climate anomalies or changes in intervention coverage.

Public Health Implications

The findings from this study have several important implications for malaria control in Beitbridge district:

1. **Targeted Timing of Interventions**: The clear seasonal pattern with predictable peaks provides evidence for optimizing the timing of preventive measures. Indoor residual spraying campaigns should be completed before the onset of the transmission season (ideally in October-November), while community awareness and early diagnosis/treatment efforts should be intensified during the February-May peak period.
2. **Resource Allocation**: The monthly forecasts can guide efficient allocation of limited resources, including diagnostic supplies, antimalarial medications, and healthcare worker deployment. By anticipating caseload fluctuations, health facilities can better prepare for seasonal increases in demand.
3. **Cross-Border Collaboration**: The border location of Beitbridge necessitates coordinated malaria control efforts with neighboring South African authorities. Sharing forecasting results can facilitate synchronized interventions that address population movement as a driver of transmission.
4. **Early Warning System**: The validated SARIMA model provides a foundation for an early warning system that could alert health authorities to unexpected deviations from predicted patterns, potentially signaling outbreaks that require rapid response.

STRENGTHS AND LIMITATIONS

This study has several strengths, including the use of an eight-year dataset that captures multiple seasonal cycles, comprehensive model diagnostics, and rigorous validation procedures. The relatively good forecasting performance demonstrates the utility of SARIMA modeling in this setting.

However, some limitations should be acknowledged. First, the analysis relied solely on passive surveillance data reported through health facilities, which may underestimate the true malaria burden due to cases that do not seek formal healthcare. Second, the model does not explicitly incorporate important covariates such as climate variables (rainfall, temperature, humidity), intervention coverage, or population movement patterns, which could enhance predictive accuracy. Third, while the model performs well for short-term forecasts (1-6 months), its reliability for longer-term projections may be limited.

CONCLUSION

This study demonstrates the effectiveness of SARIMA modeling for predicting malaria incidence in Beitbridge district, Zimbabwe. The SARIMA$(2,1,1)(1,1,1)_{12}$ model successfully captured the temporal patterns in monthly malaria cases and provided reliable short-term forecasts that can inform public health planning. The clear seasonal pattern, with peak transmission occurring between March and May, offers a window of opportunity for timely implementation of preventive measures.

The forecasting approach developed in this study represents a valuable tool for enhancing malaria surveillance in resource-limited settings. By anticipating seasonal increases in malaria transmission, health authorities can optimize intervention timing, allocate resources efficiently, and potentially improve the effectiveness of control strategies. The methodology could be adapted for other districts in Zimbabwe and similar endemic settings across sub-Saharan Africa.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Zimbabwe Ministry of Health and Child Care and The Provincial Medical Director for Matabeleland South Province Dr Andrew F Muza. We also thank the technical staff at the National Malaria Control Program for their insights and contextual information that enriched this analysis.

REFERENCES

1. Anwar, M. Y., Lewnard, J. A., Parikh, S., & Pitzer, V. E. (2021). Time series analysis of malaria in Bangladesh: Exploring the effects of climatic factors and population movement. Malaria Journal, 20(1), 175.
2. Ferrão, J. L., Mendes, J. M., Painho, M., & João, S. Z. (2017). Spatio-temporal variation and socio-demographic characters of malaria in Chimoio municipality, Mozambique. Malaria Journal, 16(1), 423.
3. Gwitira, I., Murwira, A., Mberikunashe, J., & Masocha, M. (2018). Spatial and spatio-temporal analysis of malaria cases in Zimbabwe. Infectious Diseases of Poverty, 7(1), 91.
4. Mabaso, M. L., Ngwenya, N. P., Mberikunashe, J., Gwitira, I., Saili, K., & Mudzengi, D. (2021). Malaria trends and correlates of malaria testing, treatment, and insecticide-treated nets utilization in Zimbabwe: Analysis of successive demographic and health surveys. Malaria Journal, 20(1), 356.
5. Tesfahunegn, A., Berhe, G., & Gebregziabher, E. (2020). Seasonal autoregressive integrated moving average (SARIMA) model for predicting monthly malaria cases in Adi-Yakeni hospital, Tigray, Ethiopia. International Journal of Infectious Diseases, 101, 242-251.
6. Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J., & Kaewkungwal, J. (2020). Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic districts of Bhutan. Malaria Journal, 19(1), 60.
7. World Health Organization. (2023). World Malaria Report 2022. Geneva: WHO.
8. Zimbabwe Meteorological Services Department. (2021). Climate Data Report for Matabeleland South Province 2015-2021.
9. Zimbabwe Ministry of Health and Child Care. (2022). Annual Malaria Control Program Report 2021-2022. Harare: Government of Zimbabwe.
10. Zimbabwe National Malaria Control Program. (2022). Malaria Strategic Plan 2021-2025. Harare: Ministry of Health and Child Care.
11. Zimbabwe National Statistics Agency. (2022). Population Projections Thematic Report. Harare: ZIMSTAT.