*International Journal of Science and Healthcare Research*
*[Indexed, Peer-Reviewed Journal]*
*Vol. 11; Issue: 1; Jan.-March 2026*
*Website: ijshr.com*
*Original Research Article* *ISSN: 2455-7587*

# Time Series Forecasting of Malaria Cases in Karnataka State, India Using Seasonal Autoregressive Integrated Moving Average Model for Prediction of Future Incidence

**S. R. Itagimath**

Assistant Professor in Biostatistics, Department of Community Medicine, KMCRI Hubballi, Karnataka, India

Corresponding Author: Dr. S. R. Itagimath

**DOI: https://doi.org/10.52403/ijshr.20260114**

## ABSTRACT

**Background:** Malaria remains a significant public health problem in India, particularly in endemic states such as Karnataka. Accurate forecasting of malaria incidence is essential for strengthening surveillance systems, optimizing resource allocation and enabling timely preventive interventions.

**Objective:** This study aimed to analyze the temporal pattern of malaria cases in Karnataka State, India and to develop an appropriate Seasonal Autoregressive Integrated Moving Average (SARIMA) model for forecasting future monthly malaria incidence.

**Methods:** A retrospective time series analysis was conducted using monthly malaria case data from January 2020 to August 2025 obtained from the National Centre for Vector Borne Diseases Control (NCVBDC), Government of India. The Box–Jenkins methodology was applied for model development. Data preprocessing included logarithmic transformation and seasonal and non-seasonal differencing to achieve stationary. Model identification was guided by autocorrelation function (ACF) and partial autocorrelation function (PACF) plots. Competing SARIMA models were estimated using maximum likelihood estimation and evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Bayesian Information Criterion (BIC). Residual diagnostics were assessed using the Ljung–Box Q test.

**Results:** The malaria time series demonstrated a clear increasing trend with pronounced seasonal variation. Among the candidate models, the SARIMA $(2,1,0)(0,1,1)_{12}$ model showed the best overall performance, with the lowest RMSE (29.413), MAE (18.279), MAPE (40.437) and normalized BIC (7.054). The Ljung–Box Q test indicated no residual autocorrelation ($p = 0.77$). Twelve-month forecasts (September 2025–August 2026) revealed marked seasonal peaks, with forecasted values ranging from 43 to 461 cases.

**Conclusion:** The SARIMA $(2,1,0)(0,1,1)_{12}$ model effectively captured the trend and seasonal dynamics of malaria transmission in Karnataka. SARIMA based forecasting offers a practical and reliable tool for malaria surveillance and can support evidence-based planning and timely public health interventions.

***Keywords:*** Malaria, Time series analysis, SARIMA, Forecasting, Karnataka, Public health

## INTRODUCTION

Malaria remains a major global public health challenge, particularly affecting tropical and subtropical regions where environmental conditions favour vector breeding and parasite transmission. In India, malaria transmission is heterogeneous, with certain states experiencing higher endemicity due to favourable climatic conditions, vector density and socio-demographic factors [2]. Karnataka, a southern state of India, has witnessed fluctuating malaria incidence over recent years, characterized by distinct seasonal patterns influenced by monsoon rainfall, temperature variations and human mobility. Understanding these temporal dynamics is essential for implementing targeted control measures and optimizing resource allocation.

Time series forecasting models, particularly Seasonal Autoregressive Integrated Moving Average (SARIMA) models, have proven to be effective tools for predicting infectious disease incidence, including malaria [3,4,5]. These models account for both trend and seasonal components inherent in epidemiological data, enabling reliable short-term forecasts that can guide public health decision-making [6,7,8]. SARIMA models have been successfully applied in various malaria-endemic settings, demonstrating their utility in surveillance systems and early warning mechanisms [9,10,11,12].

Accurate forecasting of malaria cases allows health authorities to anticipate disease outbreaks, strengthen vector control activities, ensure adequate drug and diagnostic supplies, and mobilize community awareness campaigns in a timely manner. Furthermore, predictive models facilitate evidence-based planning and contribute to the achievement of national and global malaria elimination targets.

Despite the availability of routine malaria surveillance data in Karnataka, there is limited published literature applying advanced time series forecasting methods to predict future malaria burden in this region. Addressing this gap, the present study aimed to systematically analyze the temporal patterns of malaria incidence in Karnataka and develop a robust SARIMA model for forecasting future monthly malaria cases.

The objectives of this study were:

1. To analyze the trend of malaria cases in Karnataka state, India, using the Box-Jenkins (SARIMA) methodology
2. To develop an appropriate SARIMA model for accurate forecasting of monthly malaria cases in Karnataka state, India

## MATERIALS AND METHODS

**Study Design and Data Source:** This study employed a retrospective time series design using secondary data. Monthly malaria case counts for Karnataka State, India, from January 2020 to August 2025 (n = 68 observations) were obtained from the National Centre for Vector Borne Diseases Control (NCVBDC), Ministry of Health and Family Welfare, Government of India [1]. The statistical analyses were performed using IBM SPSS Statistics version 26 and $P<0.05$ is considered as significant.

**Data Pre-processing:** The dataset was examined for completeness, consistency and accuracy prior to analysis. Missing observations were handled using the series mean imputation method in SPSS to maintain continuity of the time series. Data were arranged chronologically to form a continuous monthly series suitable for seasonal time series modelling.

To stabilize variance and reduce the effect of extreme values, a natural logarithmic transformation was applied. Both non-seasonals first-order differencing (d = 1) and seasonal first-order differencing (D = 1) were used to stabilize mean to achieve stationary.

**Exploratory Data Analysis**: Exploratory analysis was conducted to assess underlying

trends, seasonal patterns and serial dependence in malaria incidence. Time series plots were used to visualize long-term trends and seasonal fluctuations. Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were examined to evaluate autocorrelation structure and to guide model identification.

**SARIMA Model Specification**: Seasonal Autoregressive Integrated Moving Average (SARIMA) models were used to account for both non-seasonal and seasonal patterns in the malaria time series. A SARIMA model is denoted as SARIMA(p,d,q) (P, D, Q) $_s$, where *p*, *d*, and *q* represent the non-seasonal autoregressive, differencing, and moving average orders, respectively, and *P*, *D*, and *Q* denote the corresponding seasonal components. The parameter *s* indicates the seasonal period, set to 12 months for monthly data. Non-seasonal and seasonal differencing were applied to achieve stationarity, while autoregressive and moving average terms were used to capture temporal dependence and seasonal fluctuations in malaria incidence.

**Model Building Procedure:**
Model development followed four standard box–Jenkins steps:

1. **Stationary Assessment:** Stationary was assessed visually using time series plots and statistically through ACF and PACF plots. The non-seasonal and seasonal differencing and logarithmic transformation were applied to stabilize the mean and variance.
2. **Model Identification:** ACF and PACF plots of the stationary series were examined to identify plausible values of p, q, P, and Q. Multiple candidate SARIMA models were specified accordingly.
3. **Parameter Estimation:** Model parameters were estimated using maximum likelihood estimation. Although individual parameter significance was examined using t-statistics, overall model adequacy and forecasting performance were prioritized, consistent with Box–Jenkins principles.
4. **Model** Validation **and Diagnostic Checking:** Residual diagnostics were conducted to evaluate model adequacy. The ACF of residuals was inspected and the Ljung–Box Q test was applied to confirm that residuals were independent and resembled white noise.

**Model Selection Criteria:** Candidate SARIMA models were compared using RMSE, MAE, MAPE, and Bayesian Information Criterion (BIC). Models with lower error measures and optimal parsimony were preferred. The Ljung–Box Q test was used to ensure absence of residual autocorrelation.

**Forecasting**: The selected SARIMA model was used to generate 12-month-ahead forecasts of malaria cases from September 2025 to August 2026. Forecasts were accompanied by 95% prediction confidence intervals to account for uncertainty.

## RESULTS

**Table No.1: Descriptive statistics of malaria cases in Karnataka state India**

| | |
|---|---|
| Mean | 63.72 |
| Median | 40.5 |
| Mode | 28 |
| Minimum value | 5.0 |
| Maximum value | 244.0 |
| Range | 239.0 |
| Standard deviation | 54.84 |
| Sample variance | 3008.22 |
| Standard error | 6.65 |
| Skewness | 1.39 |
| Kurtosis | 1,82 |
| Counts | 68 |
| Sum | 4333.3 |

Analysis of 68 monthly observations from January 2020 to August 2025 revealed considerable variation in malaria incidence across Karnataka state. The mean monthly malaria cases were 63.72 (SD = 54.84), with values ranging from a minimum of 5.0 to a maximum of 244.0 cases. The wide standard

deviation relative to the mean indicates substantial dispersion in the data,

The distribution of malaria cases exhibited positive skewness (1.39), indicating a right-skewed distribution with a longer tail toward higher case counts. This asymmetry is characteristic of infectious disease count data, The kurtosis value of 1.82 suggests a platykurtic distribution, flatter than the normal curve, indicating fewer extreme outliers than would be expected under normality.

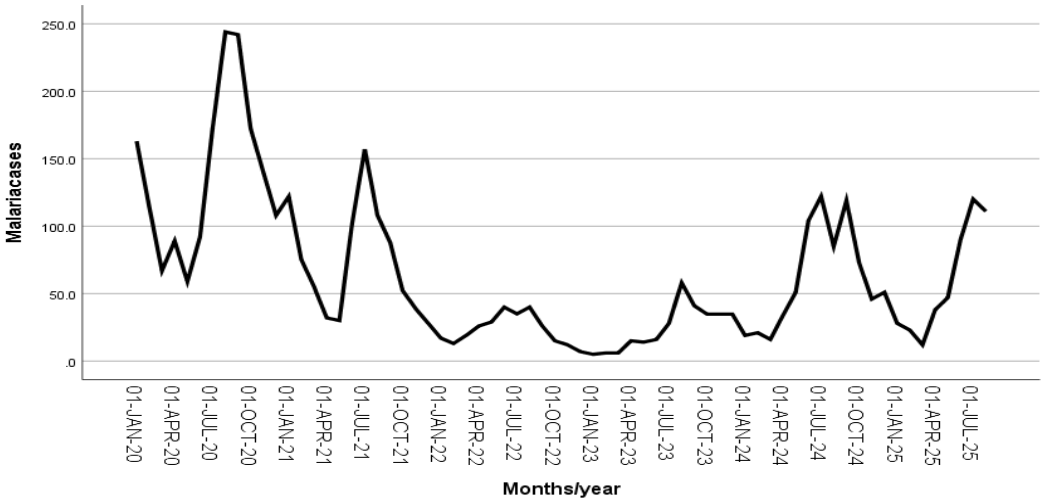


**Figure 1: Time series plot of malaria cases in Karnataka state, India**

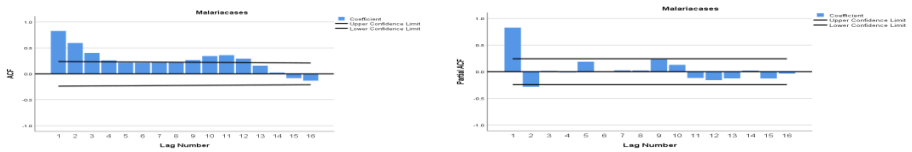


**Figure 2: ACF and PACF plots of malaria cases in Karnataka state, India**

**Temporal Patterns:** The time series plot (Figure 1) revealed distinct patterns in malaria incidence over the study period. A general increasing trend was observed, with notable seasonal fluctuations superimposed on this underlying trend. The seasonality pattern was sharp and consistent. The presence of both upward and downward trends at different time points, coupled with seasonal variations, indicated non-stationarity in the original series. The gradually changing mean and unstable variance over time confirmed the need for differencing transformations before model fitting.

Initial ACF and PACF plots of the non-stationary data (Figure 2) showed slowly decaying autocorrelations, a characteristic pattern indicating non-stationarity

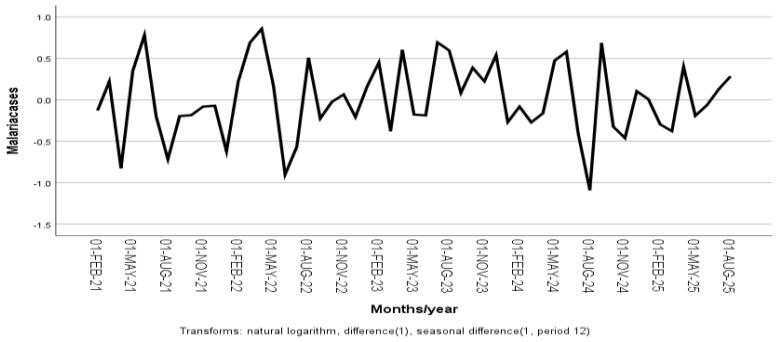


**Figure 3: Stationarity Assessment of time series plot of malaria cases in Karnataka state, India**

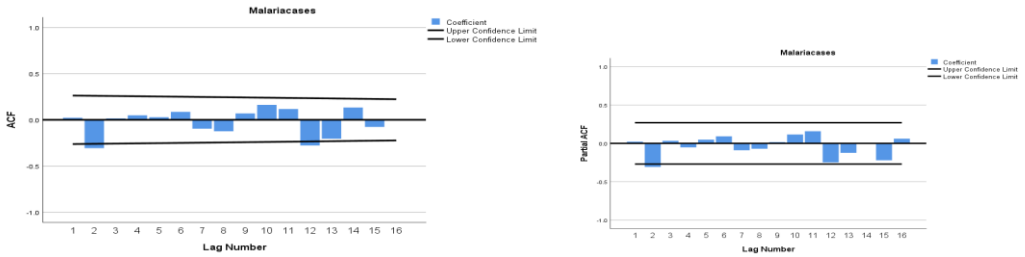


**Figure 4: Stationarity assessment of ACF and PACF plots of malaria cases in Karnataka state, India**

**Stationarity Assessment:** Taking the first-order non-seasonal differencing (d=1) and first-order seasonal differencing (D=1), combined with natural logarithmic transformation (Figure 3), the series achieved stationarity. The ACF and PACF plots of the differenced series (Figure 4) displayed rapid decay to insignificance, confirming successful stationarity transformation and enabling reliable model identification.

**Table 2: Comparison and Selection of Seasonal Arima models:**

| Model (p,d,q) (P, D, Q)$_{12}$ | RMSE | MAPE | MAE | Normalized BIC | Lung Box Q statistics |
|---|---|---|---|---|---|
| SARIMA (2,1,2) (0,1,1)$_{12}$ | 29.845 | 40.919 | 18.339 | 7.229 | 10.651<br>df=13, p= 0.64 |
| SARIMA (1,1,1) (0,1,1)$_{12}$ | 33.184 | 41.105 | 19.349 | 7.296 | 13.16<br>df =15 p= 0.590 |
| SARIMA (2,1,0) (0,1,1)$_{12}$ | 29.413 | 40.437 | 18.279 | 7.054 | 10.64<br>df=15 p= 0.77 |
| SARIMA (0,1,2) (0,1,1)$_{12}$ | 29.476 | 40.515 | 18.239 | 7.059 | 9.837<br>df=15 p=0.83 |
| SARIMA (0,1,1) (0,1,1)$_{12}$ | 33.372 | 40.681 | 19.287 | 7.234 | 13.14<br>df=16 p=0.662 |

**Model Selection Based on Forecast Accuracy:** Several Seasonal ARIMA models were evaluated to identify the most accurate forecasting model. The selection was based on forecast accuracy measures, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Bayesian Information Criterion (BIC), and the Ljung–Box Q statistic to assess residual independence.

Among the competing models, the SARIMA (2,1,0) (0,1,1)$_{12}$ model demonstrated the best overall performance. This model recorded the lowest RMSE (29.413), lowest MAE (18.279), and lowest MAPE (40.437), indicating superior forecast accuracy compared to the other models. Additionally, it produced the lowest normalized BIC value (7.054), suggesting a better balance between model fit and parsimony. The Ljung–Box Q test for this model was statistically insignificant (p = 0.77), confirming that the residuals are free from autocorrelation and resemble white noise.

**Diagnostic Checking:** All candidate models underwent rigorous diagnostic evaluation. The Ljung-Box Q test results for each model indicated no significant residual autocorrelation (all p-values > 0.05), confirming adequate model specification. The selected SARIMA (2,1,0) (0,1,1)$_{12}$ model yielded a Ljung-Box Q statistic of 10.64 with 15 degrees of freedom (p=0.77), strongly supporting the hypothesis that residuals resemble white noise. This finding indicates that the model has successfully captured all systematic patterns in the data, leaving only random variation in the residuals.

**Table 3: SARIMA (2,1,0) (0,1,1 )$_{12}$ model parameter estimation and model adequacy with natural log transformation**

| SARIMA (2,1,0) (0,1,1 )$_{12}$ | | Estimate | SE | T -test | P - values |
|---|---|---|---|---|---|
| Constant | .026 | .028 | .923 | .360 | |
| AR | Lag 1 | -.017 | .138 | -.120 | .905 |
| | Lag 2 | -.208 | .136 | -1.527 | .133 |
| Difference | 1 | | | | |
| Seasonal Difference | 1 | | | | |
| MA, Seasonal | Lag 1 | .520 | .177 | 2.937 | .005 |

Table 3 presents the estimated parameters for the selected SARIMA (2,1,0) $(0,1,1)_{12}$ model applied to the logarithmically transformed series. The constant term (0.026, SE=0.028, p=0.360) was statistically insignificant, which is commonly observed in differenced models and does not indicate inadequacy. The non-seasonal autoregressive parameters at lag 1 (-0.017, p=0.905) and lag 2 (-0.208, p=0.133) were also not statistically significant at the conventional 0.05 level.

However, consistent with Box-Jenkins principles, individual parameter insignificance does not necessarily invalidate the model, as parameters may work jointly to capture temporal dependencies and enhance forecast accuracy. The key finding was the statistically significant seasonal moving average parameter (0.520, SE=0.177, p=0.005), confirming a strong seasonal component in Karnataka's malaria incidence pattern.

## Forecast Performance:

**Table 4: Twelve months malaria case Forecasting values with 95% prediction confidence intervals for SARIMA (2,1,0) (0,1,1)$_{12}$**

| Months/ Years | Forecast values | 95% LCL | 95% UCL |
|---|---|---|---|
| Sep-2025 | 128 | 53.1 | 261.9 |
| Oct 2025 | 99 | 27.8 | 260.1 |
| Nov 2025 | 82 | 18.7 | 240.7 |
| Dec 2025 | 85 | 15.9 | 273.3 |
| Jan 2026 | 56 | 8.6 | 198.4 |
| Feb 2026 | 57 | 7.3 | 217.7 |
| Mar 2026 | 43 | 4.6 | 176.6 |
| Apr 2026 | 115 | 10.5 | 498.0 |
| May 2026 | 154 | 12.0 | 705.3 |
| Jun 2026 | 309 | 20.6 | 1488.0 |
| Jul 2026 | 453 | 26.0 | 2279.9 |
| Aug 2026 | 489 | 24.3 | 2563.2 |

Table 4 presents the selected model generated 12-month-ahead forecasts from September 2025 to August 2026. The forecasts reveal expected seasonal patterns with projected increases during traditional high-transmission months. Forecast values ranged from 43 cases (March 2026) to 489 cases (August 2026), reflecting the seasonal cycle typical of malaria transmission in Karnataka.

The 95% prediction intervals appropriately widened for longer forecast horizons, acknowledging increased uncertainty in distant predictions. The relatively wide intervals (e.g., August 2026: 24.3 to 2563.2) reflect the inherent variability in malaria incidence and the limitations of long-range forecasting.

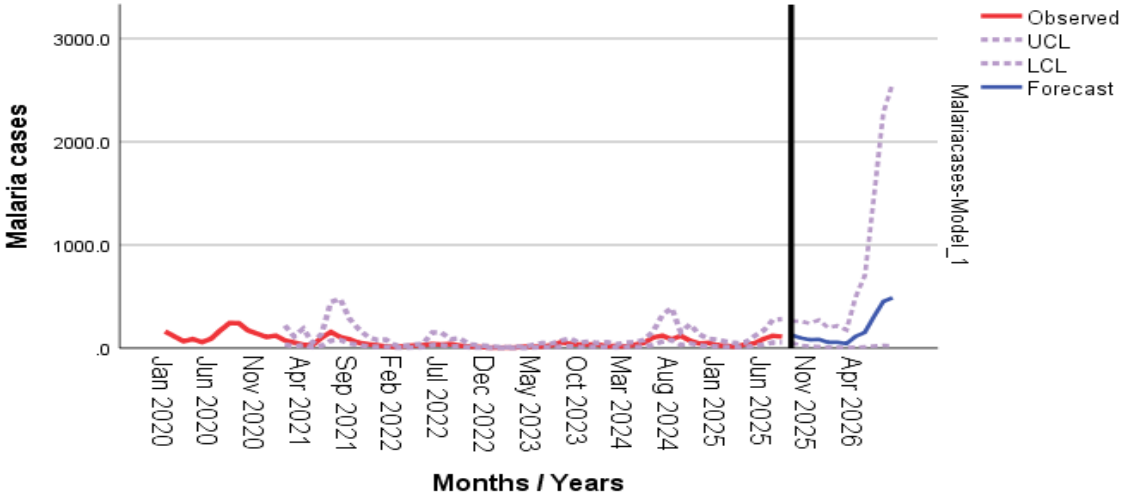


**Figure 5: Observed and forecasted malaria cases for the September 2025 to August 2025 with 95% confidence interval**

Figure 5 displays the observed values and forecasted values with confidence intervals, demonstrating good visual agreement between the model and historical data. The model successfully captured both the trend and seasonal components evident in the observed series.

## DISCUSSION

This study successfully applied the Seasonal Autoregressive Integrated Moving Average (SARIMA) methodology to model and forecast monthly malaria cases in Karnataka state, India. The analysis of 68 monthly observations from January 2020 to August 2025 revealed a clear increasing trend in malaria incidence, superimposed with pronounced seasonal fluctuations. These temporal patterns reflect the influence of climatic factors, particularly monsoon-related rainfall and temperature variations, which create favourable conditions for Anopheles mosquito breeding and malaria transmission [8,9].

The selected SARIMA $(2,1,0)\ (0,1,1)_{12}$ model demonstrated superior forecasting performance compared to alternative candidate models, as evidenced by the lowest values of RMSE (29.413), MAE (18.279), MAPE (40.437), and BIC (7.054). These metrics indicate that the model effectively captured both the non-seasonal and seasonal components of the malaria time series. The Ljung-Box Q test confirmed the adequacy of the model, with residuals exhibiting no significant autocorrelation ($p = 0.77$), suggesting that all systematic patterns in the data were successfully incorporated into the model structure [3].

The statistically significant seasonal moving average parameter (0.520, $p = 0.005$) underscores the importance of seasonality in Karnataka's malaria transmission dynamics. This finding is consistent with previous studies demonstrating strong seasonal patterns in malaria incidence driven by monsoon rains and associated environmental changes [6,9,11]. The presence of a seasonal component in the model highlights the need for temporally targeted interventions, such as intensified vector control measures and community awareness campaigns during high-transmission months.

While individual non-seasonal autoregressive parameters were not statistically significant at the conventional alpha level of 0.05, their inclusion contributed to the overall predictive accuracy of the model. In accordance with Box-Jenkins principles, parameters may work jointly to improve model performance even if individually non-significant [3]. Therefore, model selection was based on overall fit, forecast accuracy, and residual

diagnostics rather than solely on individual parameter significance.

The 12-month forecasts generated by the SARIMA (2,1,0) (0,1,1) $_{12}$ model revealed expected seasonal peaks, with projected malaria cases ranging from 43 in March 2026 to 489 in August 2026. These forecasts can serve as an early warning signal for public health authorities, enabling proactive resource mobilization, including the procurement of antimalarial drugs, distribution of insecticide-treated bed nets, and deployment of indoor residual spraying programs. The relatively wide 95% prediction intervals reflect inherent uncertainty in long-range forecasting and the stochastic nature of disease transmission, emphasizing the need for continuous model updating with incoming surveillance data [4,5].

Our findings are consistent with previous SARIMA-based malaria forecasting studies conducted in other endemic regions and successfully applied SARIMA models to predict monthly malaria cases demonstrated the utility of SARIMA forecasting [6.7]. Similarly, utilized time series analysis to explore the effects of climatic factors and population movement on malaria transmission [8.9]. These studies collectively reinforce the value of SARIMA models as practical and reliable tools for malaria surveillance and forecasting in diverse epidemiological contexts.

### Public Health Implications

The developed forecasting model offers several practical benefits for malaria control in Karnataka:

1. **Early Warning System:** The ability to forecast malaria cases 3-12 months in advance provides valuable lead time for health authorities to prepare interventions, allocate resources, and intensify surveillance during predicted high-transmission periods.
2. **Resource Optimization:** Accurate forecasts enable efficient allocation of diagnostic supplies, antimalarial drugs, and personnel, potentially reducing both costs and treatment delays during epidemic periods.
3. **Targeted Interventions:** Understanding seasonal patterns allows for strategic timing of vector control activities, such as insecticide spraying and distribution of insecticide-treated bed nets, maximizing their impact when transmission risk is highest.
4. **Evaluation Framework:** The model provides a baseline for evaluating the impact of control interventions by comparing actual cases against predicted values under business-as-usual scenarios.

These applications align with World Health Organization recommendations for strengthening malaria surveillance systems [2] and support India's commitment to malaria elimination.

## CONCLUSION

The SARIMA (2,1,0) $(0,1,1)_{12}$ was the most effective for forecasting monthly malaria cases in Karnataka state, India. It adequately captured both trend and seasonality, with diagnostic test confirming by white noise residuals (Ljung-Box Q test p=0.77).

Time series forecasting using SARIMA models provides a valuable, accessible tool for malaria surveillance and public health decision-making. The forecasts generated can support proactive resource allocation, targeted interventions and early warning systems for Karnataka's malaria control program.

***Declaration by Author***

**Ethical Approval:** Not applicable

**Acknowledgement:** None

**Source of Funding:** None

**Conflict of Interest:** The author declares no conflict of interest.

## REFERENCES

1. Ministry of Health & Family Welfare, Government of India. National Centre for Vector Borne Diseases Control (NCVBDC). Available at: https://ncvbdc.mohfw.gov.in/
2. World Health Organization. (2025). *World malaria report 2025*. World Health

Organization. Geneva, Switzerland. https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2025
3. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. 5th ed. Hoboken (NJ): John Wiley & Sons; 2015. p. 712. doi:10.1111/jtsa.12194.
4. Chatfield C. The analysis of time series: an introduction**.** 6th ed. Boca Raton (FL): CRC Press; 2016. doi:10.4324/9780203491683.
5. Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice. 3rd ed. Melbourne (Australia): OTexts; 2021. Available from: https://otexts.com/fpp3/
6. Ebhuoma O, Gebreslasie M, Magubane L. A Seasonal Autoregressive Integrated Moving Average (SARIMA) forecasting model to predict monthly malaria cases in KwaZulu-Natal, South Africa. S Afr Med J. 2018 Jun 26;108(7):573-578. doi: 10.7196/SAMJ. 2018.v108i7.12885. PMID: 30004345.
7. Lynn Nyarai Mutukura, Tavengwa Masamha. "Predictive Modelling of Malaria Cases in Beitbridge, Zimbabwe Using SARIMA," vol. 12, no. 15, 2025, doi:10.51244/IJRSI.2025.1215PH000217.
8. Anwar MY, Lewnard JA, Parikh S, Pitzer VE. Time series analysis of malaria in Bangladesh: Exploring the effects of climatic factors and population movement. Malaria Journal. 2021;20(1):175.
9. Ferrão JL, Mendes JM, Painho M, João SZ. Spatio-temporal variation and socio-demographic characteristics of malaria in Chimoio municipality, Mozambique. Malaria Journal. 2017;16(1):423.
10. Twumasi-Ankrah S, Pels WA, Nyantakyi K, Addo DK. Comparison of Statistical Techniques for Forecasting Malaria Cases in Ghana. J Biostat Biometric App. 2019;4(1):102.
11. Ferrão JL, Earland D, Novela A, Mendes R, Tungadza A, Searle KM. Malaria Temporal Variation and Modelling Using Time-Series in Sussundenga District, Mozambique. Int J Environ Res Public Health. 2021 May 26;18(11):5692. doi: 10.3390/ijerph18115692.
12. Alhassan EA, Isaac AM, Emmanuel A. Time Series Analysis of Malaria Cases in Kasena Nankana Municipality. International Journal of Statistics and Applications. 2017;7(2):43-56. doi: 10.5923/j.statistics.20170702.01.

How to cite this article: S. R. Itagimath. Time series forecasting of malaria cases in Karnataka State, India using seasonal autoregressive integrated moving average model for prediction of future incidence. *Int. J. Sci. Healthc. Res.* 2026; 11(1): 124-132. DOI: *https://doi.org/10.52403/ijshr.20260114*

******