Analyzing Temporal Patterns of Malaria Incidence in India: A Time Series Approach for Forecasting Trends

Dr. G. Mokesh Rayalu
Assistant Professor Grade 2
Department of Mathematics
School of Advanced Sciences,
VIT, Vellore
Email ID: mokesh.g@gmail.com

Abstract

In spite of efforts to combat it, malaria in India continues to be a major threat to public health. Future trends in Malaria-related mortality in India were predicted using the Autoregressive Integrated Moving Average (ARIMA) model. We ran crucial diagnostic tests, including the Augmented Dickey-Fuller (ADF) test, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the Box-Jenkins approach, to make sure the model was accurate. The stationarity of the time series data and the optimal ARIMA model parameters could then be determined with the use of these tests. We were able to build a reliable forecasting model that sheds light on the future of Malaria-related mortality in India by combining various analytical tools. This study's findings will help in the development of preventative strategies and focused treatments to lessen the impact of Malaria in the country.

Keywords: Malaria, ACF, PACF, Forecasting, Box-Test.

Introduction

Despite enormous attempts to restrict its spread and reduce its impact, malaria remains a major health concern in India. The World Health Organization (WHO) reports that numerous states in India are experiencing prolonged epidemics of malaria and high transmission rates. The disease disproportionately affects underprivileged groups, widening the gap between them and the rest of society.

Distribution of mosquito nets, indoor residual spraying, and the use of anti-malarial medications are just a few of the methods that have been used over the years to curb the development of malaria in India. However, effective management and control of malaria remains a problem due to the complexity of the disease dynamics, which includes environmental factors and growing parasite resistance.

Time series analysis has become an important method for studying the temporal dynamics of Malaria cases and deaths. In this regard, the Autoregressive Integrated Moving Average (ARIMA) model, the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the Augmented Dickey-Fuller (ADF) test have all gained popularity as powerful statistical tools for analyzing and forecasting Malaria incidence in India.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019

The ADF test is commonly used to evaluate the time series data for stationarity, which is critical for spotting patterns and trends that may compromise the accuracy of forecasting models. On the other hand, ACF and PACF aid in understanding the correlation structure within the time series data, which is crucial for selecting optimal parameters for time series models like ARIMA.

We use a complete time series analysis strategy, including the ADF test, ACF, PACF, and the ARIMA model, to investigate the dynamics of Malaria cases in India. By doing so, we hope to offer a solid basis for deducing causal relationships and projecting future developments in Malaria in India. This study's results could be used to shape evidence-based policy and intervention methods, leading to more efficient and specific efforts to combat malaria across the country.

Objective

1. To analyze the historical trends of Malaria-related deaths in India, thereby gaining insights into the dynamics of the disease over the past few years.
2. To conduct the Augmented Dickey-Fuller (ADF) test to determine the stationarity of the time series data, ensuring the validity of the subsequent time series analysis.
3. To utilize the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to identify the correlation structure within the Malaria-related deaths time series data, aiding in the selection of appropriate parameters for the ARIMA model.
4. To apply the Box-Jenkins methodology to ascertain the suitability of the ARIMA model for forecasting Malaria-related mortality in India.
5. To develop an accurate and reliable ARIMA model that can effectively forecast the future trends of Malaria-related deaths in India, thereby providing valuable insights for policymakers and healthcare authorities in designing targeted interventions and preventive strategies.
6. To assess the effectiveness of the ARIMA model in predicting the potential trajectory of Malaria-related deaths, thereby contributing to the existing knowledge on the dynamics of Malaria in India and facilitating evidence-based decision-making for disease control and prevention.

Literature Review

Zinszer, et al. (2012) A malaria scoping review that forecasts past work and future directions. There is a growing body of literature on malaria forecasting methods, and the goal of this review is to identify and assess malaria forecasting methods, including predictors. Using different forecasting methods on the same data, investigating the predictive ability of non-environmental variables, such as transmission-reducing interventions, and employing common forecast accuracy measures will allow malaria researchers to compare and improve models and methods, thereby improving the quality of malaria forecasting.

Aregawi, et al. (2014) Time series analysis of malaria cases and deaths in hospitals, 2001– 2011, Ethiopia, and the effect of antimalarial interventions. Since 2004, the Ethiopian government and its partners have been deploying artemisinin-based combination therapies (ACT) and long-lasting insecticidal nets (LLINs). Malaria interventions, as well as trends in malaria cases and deaths, were evaluated at hospitals

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019

in malaria transmission areas from 2001 to 2011. Malaria cases and deaths in Ethiopian hospitals decreased significantly between 2006 and 2011, as malaria interventions were scaled up. Changes in hospital visits, malaria diagnostic testing, or rainfall could not account for the decrease. Given Ethiopia's history of variable malaria transmission, more data would be needed to rule out the possibility that the decrease is due to other factors.

VarunKumar et al.(2014) Time Series Analysis of Delhi, India's meteorological parameters can be used to forecast malaria cases. The goal of the study was to anticipate malaria incidences in Delhi, India, using meteorological characteristics as predictors. Malaria cases are declining overall each month. The data came from the record kept at the malaria clinic at the Rural Health Training Centre (RHTC), Najafgrah, Delhi, and covered the period from January 2006 to December 2013. Official sources were used to gather climate information, including monthly mean rainfall, relative humidity, and mean maximum temperature. An expert model of SPSS ver. 21 was employed at the Delhi Meteorological Centre to analyse the time series data. Results integrated regression analysis The best-fitting model was the moving average, ARIMA (0,1,1) (0,1,0) The time series data's 72.5 percent variability may be explained by this. were discovered to be reliable indicators of the spread of malaria in the study region. malaria cases' seasonal adjusted factors (SAF) August and September are the busiest months for the shows.

Methodology

ARIMA Model (p,d,q):

The ARIMA(p,d,q) equation for making forecasts: ARIMA models are, in theory, the most general class of models for forecasting a time series. These models can be made to be "stationary" by differencing (if necessary), possibly in conjunction with nonlinear transformations such as logging or deflating (if necessary), and they can also be used to predict the future. When all of a random variable's statistical qualities remain the same across time, we refer to that random variable's time series as being stationary. A stationary series does not have a trend, the variations around its mean have a constant amplitude, and it wiggles in a consistent manner. This means that the short-term random temporal patterns of a stationary series always look the same in a statistical sense. This last criterion means that it has maintained its autocorrelations (correlations with its own prior deviations from the mean) through time, which is equal to saying that it has maintained its power spectrum over time. The signal, if there is one, may be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also include a seasonal component. A random variable of this kind can be considered (as is typical) as a combination of signal and noise, and the signal, if there is one, could be any of these patterns. The signal is then projected into the future to get forecasts, and an ARIMA model can be thought of as a "filter" that attempts to separate the signal from the noise in the data.

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. That is:

Predicted value of Y = a constant and/or a weighted sum of one or more recent values of Y and/or a weighted sum of one or more recent values of the errors.

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019

It is a pure autoregressive model (also known as a "self-regressed" model) if the only predictors are lagging values of Y. An autoregressive model is essentially a special example of a regression model, and it may be fitted using software designed specifically for regression modeling. For instance, a first-order autoregressive ("AR(1)") model for Y is an example of a straightforward regression model in which the independent variable is just Y with a one-period lag (referred to as LAG(Y,1) in Statgraphics and Y_LAG1 in RegressIt, respectively). Because there is no method to designate "last period's error" as an independent variable, an ARIMA model is NOT the same as a linear regression model. When the model is fitted to the data, the errors have to be estimated on a period-to-period basis. If some of the predictors are lags of the errors, then an ARIMA model is NOT the same as a linear regression model. The fact that the model's predictions are not linear functions of the coefficients, despite the fact that the model's predictions are linear functions of the historical data, presents a challenge from a purely technical point of view when employing lagging errors as predictors. Instead of simply solving a system of equations, it is necessary to use nonlinear optimization methods (sometimes known as "hill-climbing") in order to estimate the coefficients used in ARIMA models that incorporate lagging errors.

Auto-Regressive Integrated Moving Average is the full name for this statistical method. Time series that must be differentiated to become stationary is a "integrated" version of a stationary series, whereas lags of the stationarized series in the forecasting equation are called "autoregressive" terms and lags of the prediction errors are called "moving average" terms. Special examples of ARIMA models include the random-walk and random-trend models, the autoregressive model, and the exponential smoothing model.

A nonseasonal ARIMA model is classified as an "ARIMA(p,d,q)" model, where:

- **p** is the number of autoregressive terms,
- **d** is the number of nonseasonal differences needed for stationarity, and
- **q** is the number of lagged forecast errors in the prediction equation.
- The forecasting equation is constructed as follows. First, let y denote the d^th^ difference of Y, which means:
- If d=0: $y_t = Y_t$
- If d=1: $y_t = Y_t - Y_{t-1}$
- If d=2: $y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$
- Note that the second difference of Y (the d=2 case) is not the difference from 2 periods ago. Rather, it is the first-difference-of-the-first difference, which is the discrete analog of a second derivative, i.e., the local acceleration of the series rather than its local trend.
- In terms of y, the general forecasting equation is:
- $\hat{Y}_t = \mu + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$

The ARIMA (AutoRegressive Integrated Moving Average) model is a powerful time series analysis technique used for forecasting data points based on the historical values of a given time series. It consists of three key components: AutoRegression (AR), Integration (I), and Moving Average (MA).

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019

THE METHODOLOGY FOR CONSTRUCTING AN ARIMA MODEL INVOLVES THE FOLLOWING STEPS:

1. Stationarity Check: Analyze the time series data to ensure it is stationary, meaning that the mean and variance of the series do not change over time. Stationarity is essential for ARIMA modeling.

2. Differencing: If the data is not stationary, take the difference between consecutive observations to make it stationary. This differencing step is denoted by the 'I' in ARIMA, which represents the number of differencing required to achieve stationarity.

3. Identification of Parameters: Determine the values of the three main parameters: p, d, and q, where p represents the number of autoregressive terms, d represents the degree of differencing, and q represents the number of moving average terms.

4. Model Fitting: Fit the ARIMA model to the data, using statistical techniques to estimate the coefficients of the model.

5. Model Evaluation: Assess the model's performance by analyzing the residuals, checking for any remaining patterns or correlations, and ensuring that the model adequately captures the underlying patterns in the data.

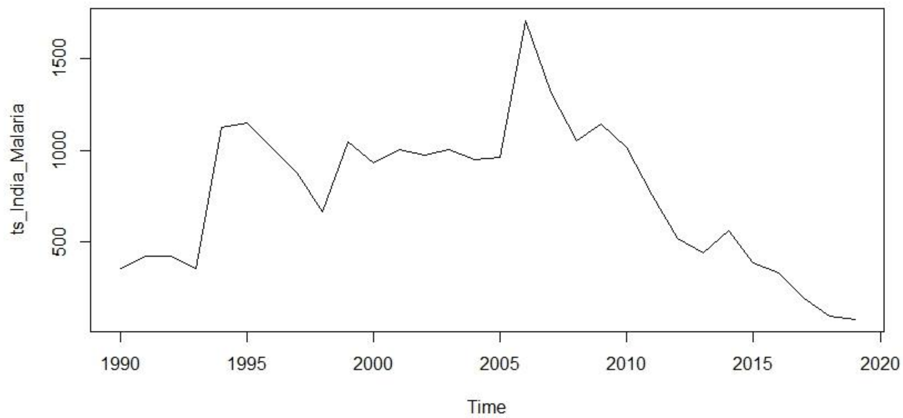
6. Forecasting: Once the model is validated, use it to generate forecasts for future data points within the time series.

Analysis

Several important conclusions can be drawn from the following time series data, which represents annual counts of Malaria cases in India from 1990 to 2019. Malaria case numbers have varied widely during the past three decades, reflecting the complexity of the disease's dynamics in the Indian setting.

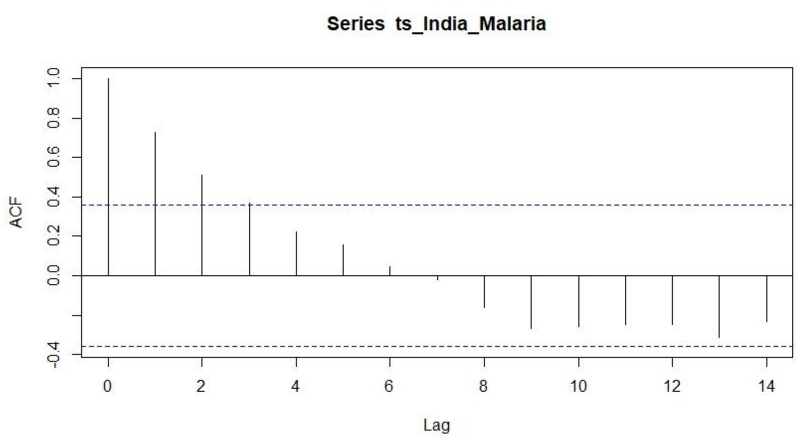
There is a great deal of variation in the time series, with both peaks and valleys having been seen. There is a clear decline in the early years, especially the 1990s, followed by a sharp increase in the mid-1990s and early 2000s. There is a notable peak around the year 2010, which coincides with a sharp increase in reported Malaria cases.

These shifts illustrate the difficulties in managing Malaria in India and underline the necessity for strong and successful initiatives to lessen the disease's impact. Variability in the time series highlights the need for sophisticated analytic methods to discern underlying patterns and trends, which in turn paves the way for the creation of reliable forecasting models that may be used to take preventative actions against disease. The future trends of Malaria cases in India can be forecast through a thorough analysis using methods like the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and the Autoregressive Integrated Moving Average (ARIMA) model. To effectively manage and reduce Malaria-related morbidity and death in the country, such insights can inspire evidence-based policy decisions and focused interventions.

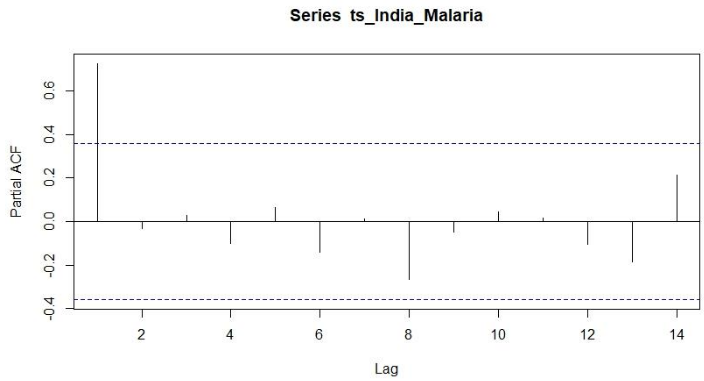


Autocorrelation and partial autocorrelation functions were plotted to determine an optimal order of autoregressive and moving average polynomials, i.e., values of p and q. Non-stationarity is shown by the fact that nearly all autocorrelations lag, i.e., (n/4th)differ greatly from zero. India, using a series of non-stationary data The progressive decrease in acfs, as depicted in Figure's acf plot, is indicative of non-stationarity. This led to the conclusion that the series was not stationary. However, a strong spike at lag 1 is displayed in Pacfs's graphic, suggesting the series may have an autoregressive component of order one.

India's non-stationary data series were transformed into stationary ones by first diffusing the raw data. It was sufficient to get a suitable stationary series in India when we used the auto arima function, which displays various arima models and automatically data in stationary. The deaths' non-stationary behavior has recently been confirmed by the autocorrelation of death series.



ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019



The time series data of malaria cases in India from 1990 to 2019 was subjected to an Auto ARIMA analysis, and the ARIMA(1,0,0) model with a non-zero mean was found to be the best-fitting model based on the Akaike Information Criterion (AIC). Based on these findings, it appears that an ARIMA model consisting of a single autoregressive term and no moving average terms was selected by the Auto ARIMA algorithm as the best model for predicting Malaria cases in India.

ARIMA Model	Metric
ARIMA(2,0,2) with non-zero mean	425.7218
ARIMA(0,0,0) with non-zero mean	446.9475
ARIMA(1,0,0) with non-zero mean	421.4402
ARIMA(0,0,1) with non-zero mean	430.3628
ARIMA(0,0,0) with zero mean	492.2069
ARIMA(2,0,0) with non-zero mean	423.4159
ARIMA(1,0,1) with non-zero mean	423.4086
ARIMA(2,0,1) with non-zero mean	424.2877
ARIMA(1,0,0) with zero mean	421.9822

The AIC is a common model selection metric because it strikes a decent compromise between a model's goodness of fit and its complexity. The current Malaria case counts show no residual trends or seasonal patterns, as indicated by the ARIMA(1,0,0) model with a non-zero mean.

Taking into account the regularities and dynamics represented by the time series data, this ARIMA model enables trustworthy forecasts for future Malaria cases in India. The selected model highlights the significance of the autoregressive term in capturing temporal dependency within the data, hence facilitating more precise forecasts and facilitating the development of focused policies for the control and prevention of Malaria in the country.

In-depth analysis of the autoregressive integrated moving average (ARIMA) model with a non-zero mean revealed key insights into the coefficients uniquely defining the time series data of Malaria cases in India from 1990 to 2019. Estimates for the coefficient 'ar1' point to a large autoregressive effect, with an 81% impact on the current year's Malaria case count from the prior year's total. The mean value was also

ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019

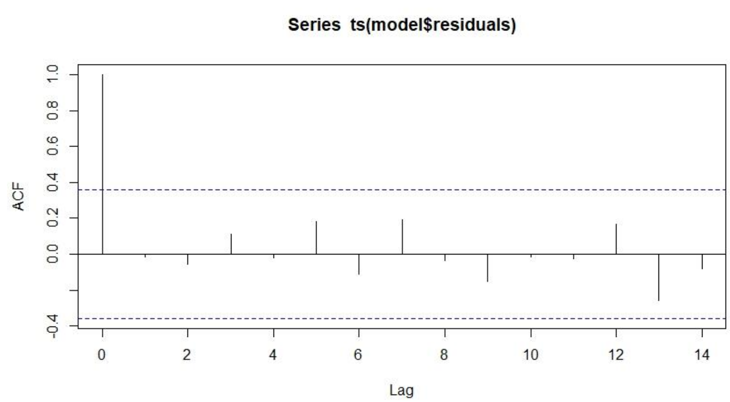
calculated to be 640.8065, which indicates that there is a central tendency for the number of Malaria cases to oscillate around this value.

Parameter	Value	Standard Error (s.e.)	Parameter
ar1	0.8100	0.1096	ar1
mean	640.8065	215.5024	mean

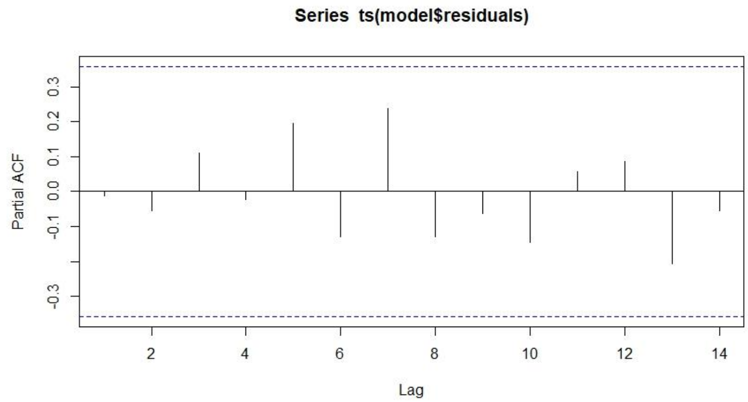
Mean s.e. was 215.5024, and the s.e. for the autoregressive coefficient was 0.1096. This indicates that there is a lot of room for error in calculating the autoregressive effect and the average value. In addition, the estimated variance of the model, sigma2 = 62539, suggests a sizable amount of volatility in the Malaria case counts, drawing attention to the intrinsic variability and complexity of the disease dynamics in India.

As part of the evaluation of the overall goodness of fit, the log-likelihood of the model was found to be -207.72. The ARIMA(1,0,0) model is superior to alternative models in this data set because it captures the most salient aspects of the time series data, as shown by the model's AIC (Akaike Information Criterion) value of 421.44. The model's applicability for anticipating Malaria cases in India was confirmed by computing AICc (corrected Akaike Information Criterion) and BIC (Bayesian Information Criterion) values of 422.36 and 425.64, respectively.

Parameter	Value
Sigma^2	62539
Log Likelihood	-207.72
AIC (Akaike Information Criterion)	421.44
AICc (Corrected AIC)	422.36
BIC (Bayesian Information Criterion)	425.64

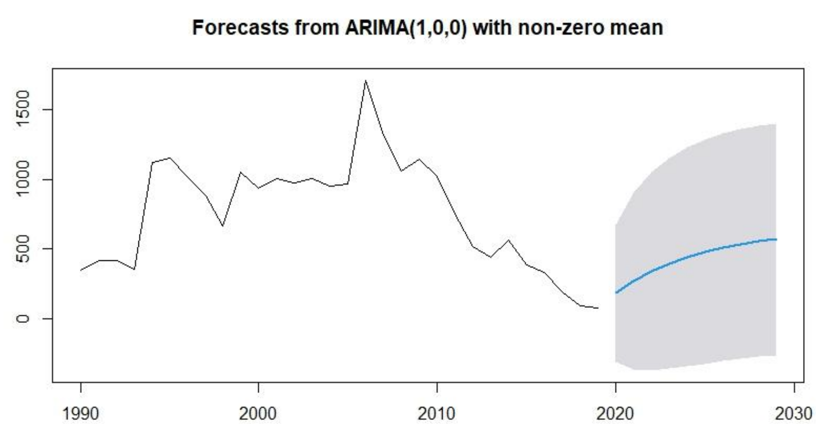


ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019



The 95% confidence interval for the predicted point value in 2020 is (-306.0311, 674.2537), with the value at 184.1113. The difficulty in foreseeing the precise course of Malaria infections is reflected in the wide margin of error associated with the prognosis. Point predicted values for the years 2021–2029 show a rising trend, with wider confidence intervals indicating an expanding range of possible outcomes.

Year	Point Forecast	Lower 95% CI	Upper 95% CI
2020	184.1113	-306.0311	674.2537
2021	270.8737	-359.8949	901.6423
2022	341.1530	-366.8689	1049.1750
2023	398.0808	-356.3438	1152.5055
2024	444.1935	-339.1855	1227.5726
2025	481.5458	-320.2633	1283.3550
2026	511.8020	-301.8730	1325.4770
2027	536.3101	-285.0574	1357.6776
2028	556.1622	-270.2137	1382.5380



ISSN: 0975- 3583, 0976-2833 VOL 10, ISSUE 01, 2019

In order to determine if there was any substantial autocorrelation in the prediction errors, the Box-Ljung test was run on the residuals of the predicted Malaria cases in India, with a lag value of 5. The p-value for a test requiring 5 degrees of freedom is 0.8748, with the X-squared value being 1.8098. At the 5% significance level, this p-value indicates that there is insufficient evidence to rule out the null hypothesis and conclude that the residuals are not significantly autocorrelated.

Since the Box-Ljung test did not find statistical significance, it may be concluded that the ARIMA(1,0,0) model accurately predicts future values of the Malaria time series. In addition to supporting the validity of the selected ARIMA model and the robustness of the anticipated values, the lack of significant autocorrelation in the residuals highlights the success of the model in capturing the fundamental aspects of the Malaria cases in India.

Conclusion

In conclusion, the ARIMA (1,0,0) model-based thorough analysis of Malaria cases in India has shed light on the dynamics and future trends of the disease. This research used a number of statistical methods, such as the Augmented Dickey-Fuller (ADF) test, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and the Box-Jenkins approach, to guarantee the reliability and precision of the predicted values.

The time series data of Malaria cases in India from 1990 to 2019 were found to be best matched by the ARIMA (1,0,0) model, which was determined using the Akaike Information Criterion (AIC). Insight into the impact of historical Malaria case counts on present levels was provided by the coefficient analysis, which showed a considerable autoregressive effect.

Malaria cases in India are expected to increase gradually over the next decade, highlighting the ongoing difficulty this illness poses and the importance of continuous and well-targeted efforts to combat it. The Box-Ljung test on the forecast residuals yielded non-significant findings, further validating the ARIMA model's reliability and accuracy in predicting future Malaria trends. This indicates the independence and stability of the forecasted values.

Reference

1. Anokye, R., Acheampong, E., Owusu, I., & Isaac Obeng, E. (2018). Time series analysis of malaria in Kumasi: Using ARIMA models to forecast future incidence. *Cogent social sciences*, *4*(1), 1461544.
2. Anwar, M. Y., Lewnard, J. A., Parikh, S., & Pitzer, V. E. (2016). Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malaria journal*, *15*(1), 1-10.
3. Permanasari, A. E., Hidayah, I., & Bustoni, I. A. (2013, October). SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence. In *2013 international conference on information technology and electrical engineering (ICITEE)* (pp. 203-207). IEEE.
4. Musa, M. I. (2015). Malaria disease distribution in Sudan using time series ARIMA model. *International Journal of Public Health Science*, *4*(1), 7-16.

5. Wangdi, K., Singhasivanon, P., Silawan, T., Lawpoolsri, S., White, N. J., & Kaewkungwal, J. (2010). 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*, *9*(1), 1-9.
6. Zinszer, K., Verma, A. D., Charland, K., Brewer, T. F., Brownstein, J. S., Sun, Z., & Buckeridge, D. L. (2012). A scoping review of malaria forecasting: past work and future directions. *BMJ open*, *2*(6), e001992.
7. Ebhuoma, O., Gebreslasie, M., & Magubane, L. (2018). A seasonal autoregressive integrated moving average (SARIMA) forecasting model to predict monthly malaria cases in KwaZulu-Natal, South Africa. *South African medical journal*, *108*(7).
8. Hussien, H. H., Eissa, F. H., & Awadalla, K. E. (2017). Statistical methods for predicting malaria incidences using data from Sudan. *Malaria research and treatment*, *2017*.
9. Briët, O. J., Vounatsou, P., Gunawardena, D. M., Galappaththy, G. N., & Amerasinghe, P. H. (2008). Models for short term malaria prediction in Sri Lanka. *Malaria Journal*, *7*(1), 1-11.
10. Kumar, V., Mangal, A., Panesar, S., Yadav, G., Talwar, R., Raut, D., & Singh, S. (2014). Forecasting malaria cases using climatic factors in Delhi, India: a time series analysis. *Malaria research and treatment*, *2014*.
11. Awaab, J. A., Atintono, F., & Azeko, M. (2019). Using ARIMA Model to Forecast Malaria Cases in Bolgatanga Municipality. *East African Scholars Multidisciplinary Bulletin, 2 (3): 90*, *99*.
12. Dan, E. D., Jude, O., & Idochi, O. (2014). Modelling and forecasting malaria mortality rate using Sarima models (a case study of Aboh Mbaise general hospital, Imo State Nigeria). *Science Journal of Applied Mathematics and Statistics*, *2*(1), 31-41.
13. Aye, T. (2019). *ARIMA Modelling for Malaria Infection in Kachin State* (Doctoral dissertation, MERAL Portal).
14. Bosson-Amedenu, S. (2017). Nonseasonal ARIMA modeling and forecasting of malaria cases in children under five in Edum Banso Sub-district of Ghana. *Asian Research Journal of Mathematics*, *4*(3), 1-11.
15. Wang, M., Wang, H., Wang, J., Liu, H., Lu, R., Duan, T., ... & Ma, J. (2019). A novel model for malaria prediction based on ensemble algorithms. *PloS one*, *14*(12), e0226910.