Hindawi
Journal of Environmental and Public Health
Volume 2023, Article ID 8907610, 8 pages
https://doi.org/10.1155/2023/8907610

Research Article

Epidemiological Forecasting Models Using ARIMA, SARIMA, and Holt–Winter Multiplicative Approach for Pakistan

Muhammad Riaz,[1,2] **Maqbool Hussain Sial**,[1] **Saira Sharif**,[1] **and Qaisar Mehmood**[1,3]

[1]*Department of Economics and Statistics, University of Management and Technology, Lahore, Pakistan*
[2]*Department of Statistics, Rahim Yar Khan Campus, Islamia University of Bahawalpur, Bahawalpur, Pakistan*
[3]*Government Graduate College, Bahawalnagar, Pakistan*

Correspondence should be addressed to Muhammad Riaz; muhammad.riaz@iub.edu.pk

Received 7 December 2022; Revised 7 February 2023; Accepted 22 March 2023; Published 29 May 2023

Academic Editor: Qiang Wang

Background of the Study. Statistical models have been extensively used in modeling and forecasting the different fields of agriculture, economics, social sciences, and medical sciences. The transmission of some diseases is a serious life threat around the globe; therefore, proper assessment and modeling need time. Malaria is one of the major life-threatening diseases in Pakistan, and some death cases due to this disease have been reported during the last decade. *Methodology*. The data have been collected from the Ministry of Health, Rahim Yar Khan, Pakistan, from January 2011 to March 2022. Data were analyzed by applying time series models for prediction purposes. Diagnostic measures such as RMSE, MAE, and MAPE were used to choose the best forecasting model. *Results and Discussion*. This study aims to forecast malaria cases by choosing the best forecast model. After comparison, it was concluded that the Holt–Winter multiplicative model outperformed the ARIMA and SARIMA models, with the lowest RMSE, MAPE, and MAE compared to other models. Malaria cases in the district Rahim Yar Khan were forecasted by the Holt–Winter multiplicative model, for the month of April 2022 to January 2023. From the forecasting results, the minimum number of cases was found to be 586.75 in June 2022 and the maximum number of cases was found to be 1281.93 in October 2022 among the next ten months. Based on the results, it is paramount for the GOP (Govt. of Pakistan) to enhance the vaccination policy to erase the impacts of malaria cases to flatten the curve.

1. Introduction

Malaria is a common and life-threatening infectious disease in many tropical and subtropical areas. The WHO reported that there were 214 million malaria cases in 2020 as compared to 227 million cases in 2019 which resulted in 627 thousand deaths in 2020. It is evident geographically that the district Rahim Yar Khan is an agriculture-based area located in south Punjab, Pakistan, where the growth of mosquitos and number of malaria patients are increased in rainy season; the number of malaria patients is also increased in the months of both cropping seasons when crops reach maturity level. An analytical understanding of malaria for rational planning of intervention programs needs to be supported by statistical and mathematical models. Time series techniques have great significance to forecast and predict the data set techniques; significant improvements have been occurring in various fields such as statistics, econometrics, earthquake forecasting, business, weather forecasting, and practical fields using the latest time series models. Researchers are developing new time series techniques that can improve better forecasting and accuracy measures, which may also be helpful for government officials and others. There are a lot of advantages of time series models that are used to search for the efficient potential for classification inference and study analysis. These methods might check the precision of the time series data more appropriately and accurately to provide significant forecasting. The well-known time series models are the autoregessive model (AR), model of moving average (MA), autoregressive of moving average model (ARMA), and ARIMA model (ARIMA). However, all these are useful when only effectually linear serial correlation exists between the data sets in the present. One of the most significant statistical techniques in the analysis of time series studies for forecasting is the ARIMA technique which has been broadly applied due to its importance and capacity in handling both the stationary and nonstationary time series data. This technique is useful for data with linear conditions as its assumption of linearity is associated with time series. The necessities for modeling which consider the effects of those specially related to malaria disease were firstly taken into consideration by [1] who recognized malaria as one of the first infectious diseases to be analyzed mathematically. The concept of R_0 in its simplified form when seasonality is considered was analyzed in [2]. There are a short number of nationally prevalence surveys in low budget countries. Much work has focused on methods for estimating high-resolution malaria risk from these data [3, 4]. Routine surveillance data of malaria case counts, often aggregated over administrative regions defined by geographic polygons, are becoming more reliable and widely available [4]. However, the collection of cases over space means that the data may be uninformative, especially if the case counts are aggregated over large areas since it is not clear where within and in which environments the malaria cases occurred. These data are therefore often underpowered for fitting flexible, nonlinear, mathematical, and statistical models as is required for accurate malaria mapping [5, 6]. However, mapping malaria in lower-budget countries has new risks as traditional mapping of prevalence from cluster-level surveys is often not effective [7–9]. High-dimensional maps of malaria risk are important for control and eradication [10–12]. ARIMA is one of the renowned techniques used for modeling the time series data, and it has been extensively studied throughout the literature for modeling and forecasting. Time series modeling of infectious diseases such as COVID-19 has been used and applied by several researchers [6, 13–18]. Rural Health Centers (RHCs) in Mozambique collect a large volume of time series case data from symptomatic malaria patients. Analyses of these data are retrospective, which generally only detect patterns after they have occurred. These data for mathematical modeling can be used to explain, describe, and predict malaria cases. Modeling not only can produce valid results but is also inexpensive. This can help plan malaria control and eradication efforts [19]. Malaria time series studies using weekly data are not common globally. In Asia, studies in Afghanistan and India were carried out to forecast malaria cases using the ARIMA model with monthly data [20, 21]. In Africa, the Box–Jenkins modeling was used in Zambia and Ghana to forecast malaria using monthly data [22, 23]. In Mozambique, malaria morbidity forecasting using the ARIMA model was performed weekly and intervention analysis for mortality monthly data was taken from Chimoio Municipality [24, 25]. Due to the transmissibility and seasonality of malaria, models with an ARIMA structure have more predictive power compared to other methods [26]; such models have been applied to predict numerous infectious diseases with similar periodic patterns over the past decades [27, 28].

The focus of this study was to find a new best suitable model for seasonal time series data for both linear and nonlinear conditions. This is a new study for modeling and forecasting malaria cases using seasonal and nonseasonal ARIMA. The proposed models would be conventional models, ARIMA model, SARIMA model, Exponential Smoothing model, and Holt–Winter model that can be helpful to increase the accuracy of the model by comparing all these models for better forecasting, reducing error, and following the tradeoff between bias in error and in variance reduction.

2. Materials and Methods

Monthly confirmed malaria cases in the district Rahim Yar Khan, Pakistan, from January 2011 to March 2022, were collected as secondary data provided through the source Ministry of Health, RYK District, Pakistan. The mean cases of malaria from January 2011 to March 2022 were 2817.11 with a standard deviation of 1203.111, which indicates that the malaria cases follow the nonnormal pattern and the number of malaria cases increased in rainy season and in the months of both cropping seasons when crops reach maturity level. The median cases of malaria from January 2011 to March 2022 were 2808. The maximum number of malaria cases was 6777 in September 2011 and the minimum number of cases was recorded as 533 in the month of March 2020.

2.1. Modeling of Malaria Cases Using Conventional Modeling Techniques. For initial analysis, without considering the time series data, conventional methods of modeling have been applied in this study to model our malaria data. There are several regression models which are present in the literature to model the data. The three most frequently used models have been selected.

The following are the three regression models that were compared for modeling and prediction purposes.

The functional form of the simple linear regression model (SLRM) is as follows:

$$y = a + \mathrm{bt} + e. \tag{1}$$

The functional form of the LRM is as follows:

$$y = a + b\ln(t) + e. \tag{2}$$

The functional form of the QRM is as follows:

$$Y_t = a + \mathrm{bt} + \mathrm{ct}^2 + e. \tag{3}$$

2.2. Modeling of Malaria Cases Using Nonseasonal ARIMA. The Box–Jenkins ARIMA (p, d, q) is given by

$$\begin{aligned}\widehat{Y}_t &= \mu + \alpha_1 \widehat{y}_{t-1} + \cdots + \alpha_p y_{t-p} + \cdots + \theta_1 e_{t-1}\\ &+ \cdots + \theta_q e_{t-q} + e_t,\end{aligned} \tag{4}$$

where $\alpha_1\hat{y}_{t-1},\ldots,\alpha_p\hat{y}_{t-p}$ are the lagged values and $\theta_1 e_{t-1},\ldots,\theta_q e_{t-q}$ are the lagged errors of the series $\hat{y}_t$. The constants $p, d,$ and q represent the order of the AR term, the degree of differencing series, and the order of the MA term, respectively. e_t is the white noise with mean 0 and variance σ^2. $\hat{y}_t$ can be differenced once or more. The model with the least root mean square error (RMSE) and mean absolute error (MAE) is chosen as the most suitable model for our data. The expressions RMSE and MAE are as follows:

$$\begin{aligned}
\text{RMSE} &= \sqrt{\frac{1}{T-N}\sum_{t=N+1}^{T}\left(y_t-\hat{y}_t\right)^2},\\
\text{MAE} &= \frac{1}{T-N}\sum_{t=N+1}^{T}\left|y_t-\hat{y}_t\right|,\\
\text{MAPE} &= \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t-\hat{y}_t\right|}{\left|Y_t\right|}\,x\,100,\\
\text{MSE} &= \frac{1}{n}\sum_{t=1}^{n}\frac{\left|y_t-\hat{y}_t\right|}{\left|Y_t\right|},
\end{aligned}\tag{5}$$

where $y_1,\ldots,y_N$ and $y_{N+1},\ldots,y_T$ are a partition of the data.

2.3. ARIMA with SARIMA and Exponential Smoothing. ARIMA models can also be used for modeling a wide range of seasonal data. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. The autoregressive moving average (ARIMA) model contains three combination models which are

(i) Autoregressive (AR) model
(ii) Moving average (MA) model
(iii) e_t white noise (WN) process.

A time series $\{Y_t\}$ is said to follow the ARMA (p, q) model if

$$\begin{aligned}
Y_t &= \mu+\phi_1 Y_{t-1}+\phi_2 Y_{t-2}\ldots\ldots+\phi_p Y_{t-p}\\
&\quad+e_t-\Theta_1 e_{t-1}-\Theta_2 e_{t-2}\ \ldots\ldots-\Theta_q e_{t-q},
\end{aligned}\tag{6}$$

where p and q are greater than zero, p refers to the autoregressive part AR, q refers to the moving part MA, and e_t is the white noise term of the model.

By using the back shift operator, the ARMA (p, q) model can be defined as

$$\phi_1(B)Y_t=\Theta_1(B)e_t,\tag{7}$$

where $\phi_1(B)=1-\phi_1 B-\phi_2 B^2-,\ \ldots,\ \phi_p B^p$ and $\Theta(B)=1+\Theta_1 B+\Theta_2 B^2\ +,\ldots,+\ \Theta_q B^q$.

For the nonstationary time series, first, we convert it by taking the difference in the series where the difference in the series is denoted by d, and then the ARIMA (p, d, q) model is as follows:

$$\phi_1(B)\Delta^d Y_t=\Theta_1(B)e_t.\tag{8}$$

Seasonal ARIMA can be expressed as follows:

$$\text{SARIMA}\ (p,d,q)(P,D,Q)m.\tag{9}$$

The additional seasonal terms are simply multiplied by the nonseasonal terms. For the seasonal ARIMA model, $p=1$ and 2 and $q=1$ and 2. For nonseasonal portion, $P=1$ and 2 and $Q=1$ and 2. Table 1 presents the list of different seasonal ARIMA models. The seasonal length has been considered 12 since it is the monthly data.

2.4. Holt–Winter Nonseasonal Model (Two Parameters). Exponential smoothing is a time series forecasting method for univariate data that can be extended to support data with a systematic trend or seasonal component. We will be using mainly two types of linear exponential smoothing methods which consider both linear and seasonal trends in the data. This method is appropriate for series with a linear time trend and no seasonal variation. Being an adaptive method, this methodology allows the level, trend, and seasonality patterns to change over time.

The smoothed series can be expressed in the following functional form:

$$\hat{y}_t=a+bk,\tag{10}$$

where a and b are the permanent component and trends as defined above. The value of a = 1.000 and b = 0.000. The RMSE is 562.70; the coefficient value of b is zero which means that the trend component is estimated as fixed and not changing.

2.5. Holt–Winter Multiplicative Model (Three Parameters). This method is appropriate for series with a linear time trend and multiplicative seasonal variation. The smoothed series is given by

$$\hat{y}_t=(a+bk)\left(c_{t+k}\right),\tag{11}$$

where a = permanent component (intercept), b = trend, and c_t = multiplicative seasonal factor.

2.5.1. Packages Used for Analysis. The SARIMA model was selected using the ARIMA package of R which resulted in the final model of SARIMA $(2,1,2)(1,1,0)_{12}$. The Holt–Winter methodology has been applied using the R package of Holt–Winter and forecasting has been performed using the "forecast" package.

3. Results

Results of different models are presented in this section. Table 2 shows the coefficients with their AIC and BIC from the three regression models. R^2 for the LRM is 0.645, and the adjusted R^2 is 0.642. The R^2 for the QRM is 0.646, which is higher than the LRM (0.535) and SLRM (0.645); therefore, for comparison purposes, the AIC and BIC methodologies

Table 1: ARIMA (p, d, q) models with the diagnostics for malaria cases.

Model (p, d, q)	AIC	R-square	RMSE	MAE	MAPE	MSE	p value
ARIMA (1,1,0)	15.56*	0.79	560.92	357.46	13.48	567.28	0.001
ARIMA (1,1,1)	15.57	0.78	559.76	360.13	13.49	568.25	0.001
ARIMA (2,1,0)	16.29	0.54	801.33	552.69	21.02	812.95	0.001
ARIMA (2,1,1)	15.57	0.78	561.85	356.70	13.50	568.30	0.001

Table 2: Regression models with their coefficients and diagnostic checks.

Models	Coefficients with p value		R-square	Adjusted R-square	AIC	BIC
SLRM	$a = 4497.01337$ (0.001***)	$b = -24.704336$ (0.001***)	0.645	0.642	2163.28	2171.99
QRM	$a = 4436.8575$ (0.001***)	$b = -22.069775$ (0.001***), $c = -0.019372$ (0.001***)	0.646	0.640	2165.09	2176.72
LRM	$a = 6492.6574$ (0.001***)	$b = -935.191189$ (0.001***)	0.535	0.531	2199.81	2208.53

are more widely used and chosen. The best model was chosen based on the large value of R^2 and the small value of AIC and BIC. From Table 2, it can be clearly observed that the SLRM shows better outcomes for every model selection criterion. It has a small AIC and BIC as compared to all three models, which are 2163.28 and 2171.99, respectively. The value of the DW statistic is 1.92, which lies in no autocorrelation area. Therefore, there is no evidence of autocorrelation; hence, it will not impact the forecasting of our data. Moreover, there should be multicollinearity between t and t^2, but it has no impact as it is formed. This is known as structural multicollinearity as new predictors are created from the same predictor.

The series is not stationary. This means that mean, variance, and covariance are not stationary over some time. The ACF and PACF also show that the series is not stationary. Using the ADF test, it has been observed that the series is not stationary with the Dickey–Fuller test statistic $= -2.7809$ and p value $= 0.2517$. This violation has been removed by taking the first difference and it was found that the series is stationary at first difference and it is stationary with test statistic $= -6.5213$ and p value $= -0.01^{***}$. Figure 1 shows the time series plots of ACF and PACF of malaria cases from January 2011 to March 2022. After making the series stationary by taking the first difference in the observed series, Figure 2 shows the graph of stationary series with ACF and PACF.

In Box–Jenkins ARIMA methodology, after having established the stationary of the series at 1st difference, the next step is to look for the appropriate autoregressive and moving average terms to be included in the model which depends upon the behavior of the correlogram of autocorrelations and partial autocorrelations. From the behavior of the correlogram of the autocorrelation function, we determine the number of moving average terms (q), and from the behavior of the partial autocorrelation function, we decide the number of autoregressive terms (p) to be included in the model. Different nonseasonal ARIMA models will be applied on the series and that model will be selected which has the smallest value of AIC, RMSE, and MAE. ACF has declining trend and PACF has one significant spike which suggests that the initial model will be ARIMA (1,1,0). Tentative autoregressive integrated moving average (ARIMA) model is ARIMA (1, 1, 0). Starting from this model, the best forecast ARIMA (p, d, q) model is chosen by diagnostically comparing all possible fitted models. The first step of ARIMA model, i.e., identification, was achieved by making it stationary using the ADF test and selecting the appropriate model using ACF and PACF. The next step would be the estimation of parameters of the model with their diagnostic checks [29].

From Table 1, it can be seen that all the models selected are significant at 5% level of significance since the p value is less than 0.05. However, it can be concluded that the ARIMA (1,1,0) model showed the lowest values of AIC, RMSE, and MAE and a high value of R^2 for malaria case series among all other models for modeling and forecasting. Figure 3 shows the decomposition of the series for malaria cases and it is showing the seasonal trend in the model. Therefore, the SARIMA model has been applied to look for the seasonal pattern in the series.

From the different models of SARIMA evaluated, the best is the SARIMA $(2,1,2)(1,1,0)_{12}$ model since it has the smallest value of AIC 15.25, RMSE = 476.90, MAE = 370.88, MSE = 562.21, and MAPE = 13.21. The SARIMA model was selected using the ARIMA() package of the R under forecast () library which resulted in the final model of SARIMA $(2,1,2)(1,1,0)_{12}$. To check the validity of the model, we move on to the third step of the Box–Jenkins methodology which is diagnostic checking and applying the normality tests to the residual of the selected model. To check the normality of the model's residuals, we apply the Shapiro–Wilk normality test. The value of the Shapiro–Wilk normality test for full sample data is $W = 0.98633$ and p value =0.2334 which means we do not reject our null hypothesis. The test statistics of Ljung box test $Q^* = 12.998$ with p value =0.477 showing that the model does not show any lack and residuals are uncorrelated. The time plot of the residuals in Figure 4 depicts that the variation of the residuals stays much the same across the historical data, apart from the one outlier, and therefore, the residual variance can be treated as constant. This can also be seen on the histogram of the residuals. The histogram suggests that the residuals are normal.

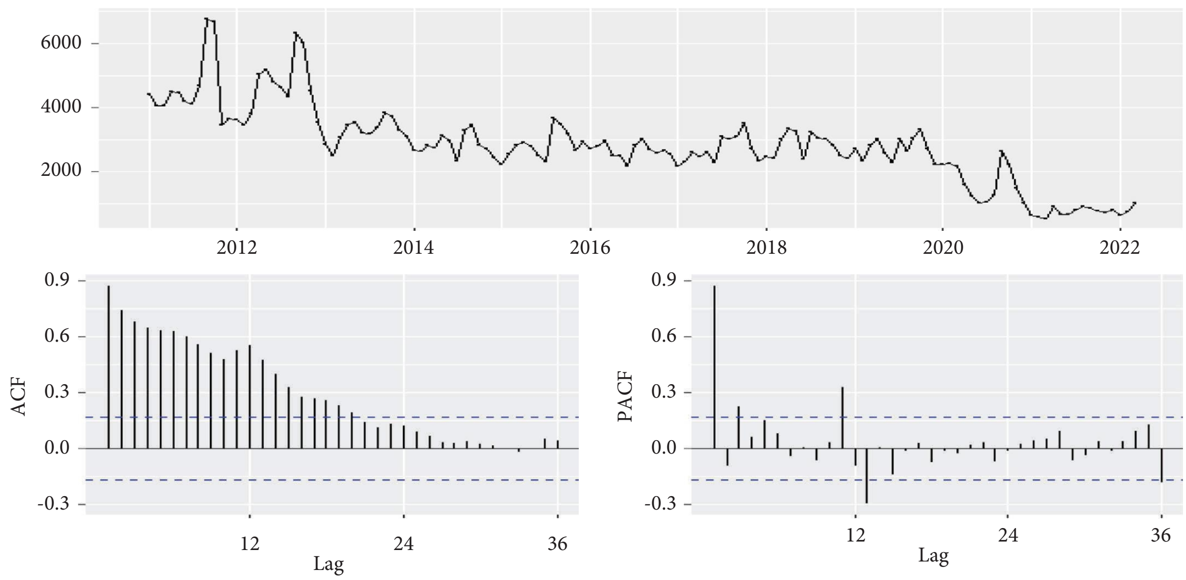

Figure 1: Plot of observed malaria cases, ACF and PACF, from January 2011 to March 2022.

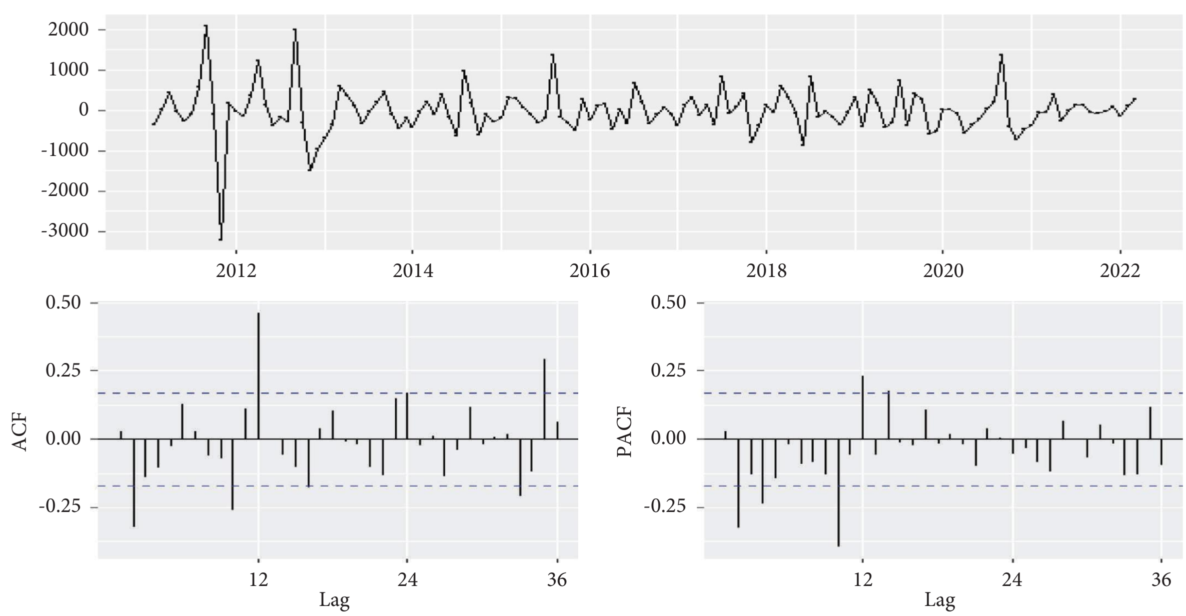

Figure 2: Plot of stationary series with ACF and PACF.

Applying the Holt–Winters multiplicative model (three parameters) to our malaria data, the estimated permanent component (intercept) of the proposed model is 0.48 and the trend and multiplicative seasonal factor are zero which means they have been estimated as fixed and are not changing. The RMSE of the model is 417.42, MAE = 350.34, MAPE = 12.10, and MSE = 558.91 which are the lowest among all the earlier estimated models. This model will be plotted against the original data. We have included the forecasted values of the Holt–Winter model ARIMA (1,1,0) and a SARIMA $(2,1,2)(1,1,0)_{12}$ term for comparison purposes between them. The forecasts from the multiplicative exponential smoothing method are doing a great job in detecting the seasonal movements in the actual series of malaria cases. The figure shows the fit of all three proposed models. The RMSE of ARIMA (1,1,0) is 560.92, SARIMA (2,1,2)(1,1,0) is 476.92, and Holt–Winter multiplicative model is 417.42. Therefore, from the RMSE perspective, the Holt–Winter model is the best forecasting model among the ARIMA and SARIMA models.

The short-term forecasting ahead by Holt–Winter multiplicative best selected model is mentioned in Table 3. There will be 738 cases of malaria in RYK till Jan 2023.

From Table 3, the maximum number of forecast malaria cases in the district Raheem Yar khan was found to be 1281 in the month of October and the minimum number of cases was found to be 586.75 in the month of Jun 2022 by applying the best selected Holt–Winter multiplicative forecast model.

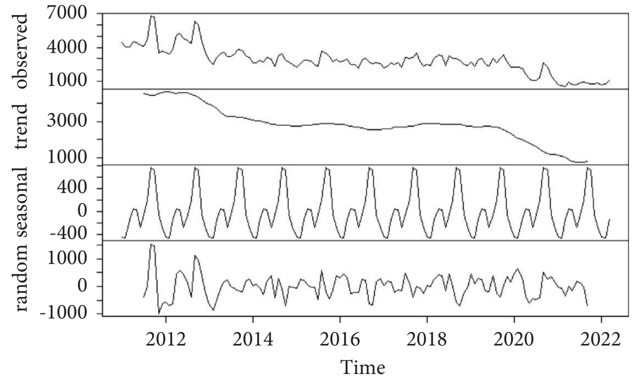

Figure 3: Decomposition of the malaria case series into different trends.

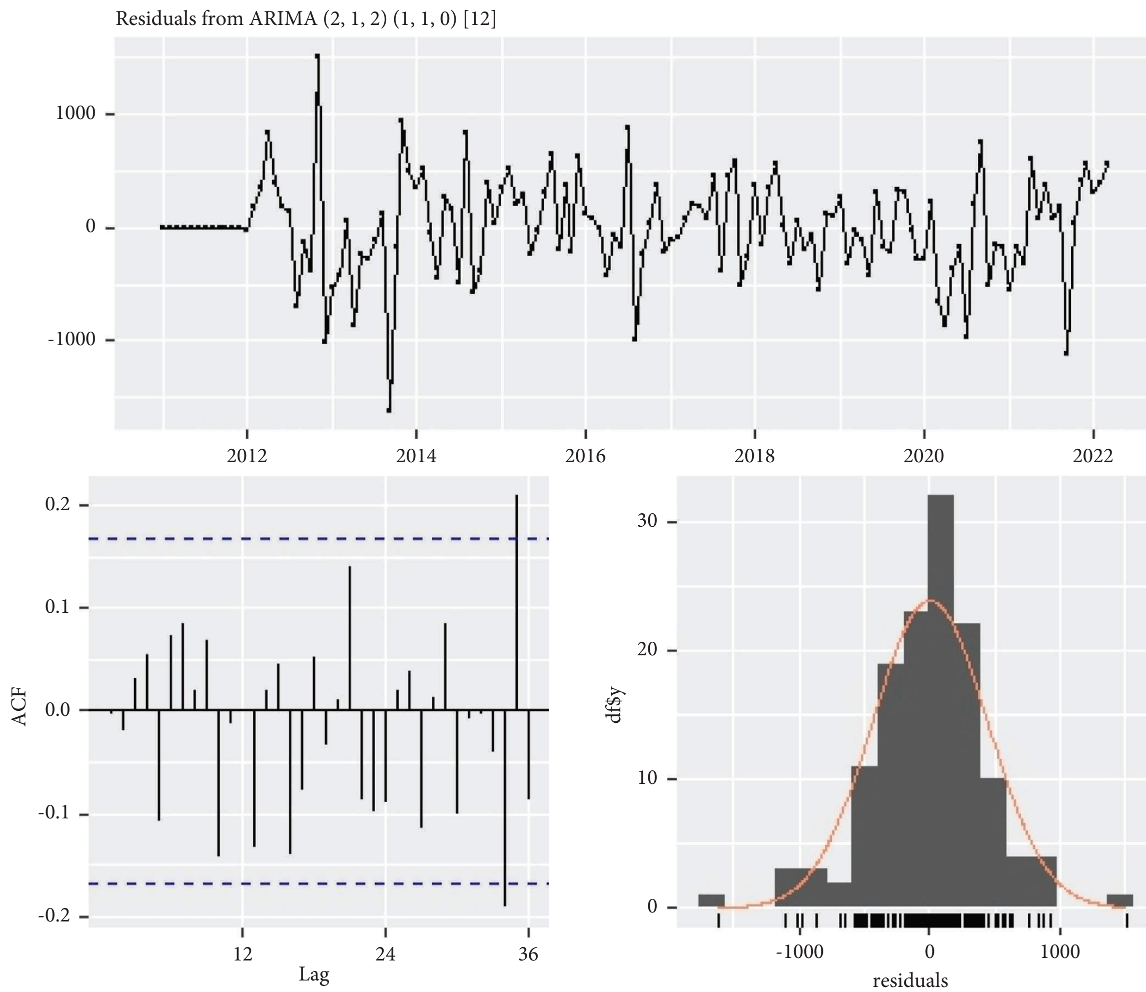

Figure 4: Residual diagnostic diagram for the SARIMA (2,1,2)(1,1,0).

Table 3: 10 month short-term forecasting by the Holt–Winter multiplicative model.

Time	Forecast	Lo 80	HI 80	LO 90	Hi 90
Apr 2022	912.85	238.22	1587.50	−118.91	1944.63
May 2022	633.11	−64.36	1330.57	−433.58	1699.79
Jun 2022	586.75	−133.29	1306.79	−514.45	1687.95
Jul 2022	742.80	0.42	1485.18	−392.57	1878.17
Aug 2022	912.62	148.11	1677.13	−256.59	2081.84
Sep 2022	1215.39	428.94	2001.85	12.61	2418.18
Oct 2022	1281.93	473.68	2090.17	45.83	2518.03
Nov 2022	1166.63	336.74	1996.51	−102.57	2435.82
Dec 2022	1051.74	200.35	1903.12	−250.34	2353.82
Jan 2023	738.97	−133.81	1611.74	−595.82	2073.76

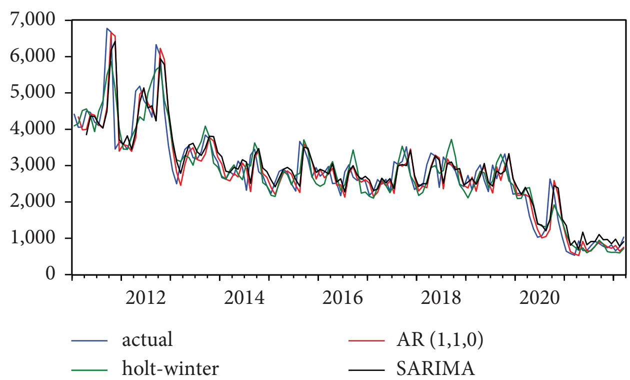

Figure 5: Graph of observed, ARIMA, SARIMA, and Holt–Winter models.

Table 3 also shows malaria cases are higher in the months of August–November 2022 than the rest of the months.

Figure 5 shows the plotting of actual data against all the selected models, and it was concluded that the Holt–Winter multiplicative model best fitted among all the other models. The graph of actual versus fitted shows that malaria cases are decreasing gradually due to better management.

4. Discussion

ARIMA, SARIMA, and Holt–Winter techniques have been applied which considered seasonal, linear, and nonseasonal factors for modeling the malaria cases in the district RYK, Pakistan. SARIMA $(2,1,2)(1,1,0)_{12}$ showed the best results since the series contained a seasonal component. A model with small RMSE was proposed by comparing the models. After comparison, the Holt–Winter multiplicative model outperformed ARIMA and SARIMA models in modeling and forecasting the malaria cases in Pakistan. It was noted from the forecasting results that malaria cases in the district Raheem Yar khan will be 1281 in the month of October and minimum cases will be 586.75 in the month of Jun 2022 by applying the best selected Holt–Winter multiplicative forecast model. Forecast estimates also showed malaria cases are higher in the months of August–November 2022 than in the rest of the months.

5. Conclusion and Way Forward

Several climate and environmental variables have been associated with malaria incidence. Forecast estimates of malaria cases were high in the month of October and minimum in the month of Jun 2022. The forecasting models developed in the study provide stakeholders of RYK district with expected malaria cases in advance, which would be a useful guidance for timely prevention and control measures to be effectively planned. The knowledge of the forecasting results of the malaria at the subdistricts would also greatly aid in targeting the control measures, even though the forecasting is feasible at the district-level data due to the small number of cases at the subdistrict level. It is paramount for the GOP (Govt. of Pakistan) to enhance the vaccination policy to erase the effects of malaria cases to flatten the curve.

The main limitation of models based on time series is that they provide forecasting for a longer period, and this results in the negative confidence intervals. Further research is recommended to evaluate the effectiveness of integrating the forecasting model into the existing malaria control programmed in terms of its impact in reducing the disease occurrence and also the cost of control interventions.

Data Availability

Monthly confirmed malaria cases in the Rahim Yar Khan district, Pakistan, from January 2011 to March 2022, were collected as secondary data provided through the source Ministry of Health, RYK district, Pakistan.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Muhammad Riaz conceptualized the study, wrote the manuscript, and collected the data; Saira Sharif developed the methodology and did analysis; Maqbool Hussain Sial supervised the study; Qaisar Mehmood did literature review and discussed the study. All authors have read and agreed to publish the manuscript.

Acknowledgments

The authors are thankful to the University of Management and Technology, Lahore, for providing research environment and they are also thankful to the Ministry of Health, Raheem Yar Khan, Pakistan, for providing data on malaria cases.

References

[1] R. Ross, *The Prevention of Malaria*, John Murray, London, UK, 1911.

[2] O. Diekmann, J. A. Heesterbeek, and J. A. Metz, "On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous

populations," *Journal of Mathematical Biology*, vol. 28, pp. 365–382, 1990.
[3] H. J. Sturrock, J. M. Cohen, P. Keil et al., "Fine-scale malaria risk mapping from routine aggregated case data," *Malaria Journal*, vol. 13, no. 1, p. 421, 2014.
[4] H. J. Sturrock, A. F. Bennett, A. Midekisa, R. D. Gosling, P. W. Gething, and B. Greenhouse, "Mapping malaria risk in low transmission settings: challenges and opportunities," *Trends in Parasitology*, vol. 32, no. 8, pp. 635–645, 2016.
[5] T. C. D. Lucas, A. K. Nandi, S. H. Keddie et al., "Improving disaggregation models of malaria incidence by ensembling non-linear models of prevalence," *Spat Spatiotemporal Epidemiol*, vol. 41, Article ID 100357, 2022.
[6] S. Bhatt, E. Cameron, S. R. Flaxman, D. J. Weiss, D. L. Smith, and P. W. Gething, "Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization," *Journal of The Royal Society Interface*, vol. 14, no. 134, Article ID 20170520, 2017.
[7] Y. Wang, Z. Shen, and Y. Jiang, "Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China," *PLoS One*, vol. 13, no. 9, pp. 1–11, 2018.
[8] H. C. Law, D. Sejdinovic, E. Cameron et al., "Variational learning on aggregate outputs with Gaussian processes," *Advances in Neural Information Processing Systems*, vol. 31, pp. 6081–6091, 2018.
[9] O. Johnson, P. Diggle, and E. Giorgi, "A spatially discrete approximation to log Gaussian Cox processes for modelling aggregated disease count data," *Statistics in Medicine*, vol. 38, no. 24, pp. 4871–4887, 2019.
[10] K. E. Battle, T. C. D. Lucas, M. Nguyen et al., "Mapping the global endemicity and clinical burden of Plasmodium vivax, 2000–17: a spatial and temporal modelling study," *Lancet*, vol. 394, 2019.
[11] D. J. Weiss, T. C. D. Lucas, M. Nguyen et al., "Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study," *The Lancet*, vol. 394, no. July, pp. 322–331, 2019.
[12] Q. Mehmood, M. H. Sial, M. Riaz, and N. Shaheen, "Forecasting the production of sugarcane in Pakistan for the year 2018-2030, using box-jenkin's methodology," *J. Anim. Plant Sci*, vol. 29, no. 5, pp. 1396–1401, 2019.
[13] M. Daniyal, K. Tawiah, S. Muhammadullah, and K. O. Ameyaw, "Comparison of conventional modeling techniques with the neural network autoregressive model (NNAR): application to COVID-19 data"," *Journal of Healthcare Engineering*, vol. 2022, Article ID 4802743, 9 pages, 2022.
[14] S. Ghosal, S. Sengupta, M. Majumder, and B. Sinha, "Prediction of the number of deaths in India due to SARS-CoV-2 at 5–6 weeks," *Diabetes & Metabolic Syndrome: Clinical Research Reviews*, vol. 14, no. 4, pp. 311–315, 2020.
[15] A. Roy and S. Kar, "Nature of transmission of Covid19 in India," *medRxiv*, pp. 57–66, 2020.
[16] A. Tiwari, "Modelling and analysis of COVID-19 epidemic in India," *medRxiv*, pp. 77–89, 2020.
[17] J. Bhola, V. R. Venkateswaran, and M. Koul, "Corona epidemic in Indian context: predictive mathematical modelling," *medRxiv*, 2020.
[18] A. A. H. Ahmadini, M. Naeem, M. Aamir, R. Dewan SS, and W. K. Mashwani, "Analysis and forecast of the number of deaths, recovered cases, and confirmed cases from COVID-19 for the top four affected countries using kalman filter," *Frontiers in Physiology*, vol. 9, Article ID 629320, 2021.
[19] S. Mandal, R. R. Sarkar, and S. Sinha, "Mathematical models of malaria - a review," *Malaria Journal*, vol. 10, p. 202, 2011.
[20] M. Y. Anwar, J. A. Lewnard, and S. Parikh, "Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence," *Malaria Journal*, vol. 15, p. 566, 2016.
[21] V. Kumar, A. Mangal, S. Panesar, G. Yadav, and R. Talwar, "Singh forecasting malaria cases using climatic factors in delhi, india: a time series analysis," *Malaria research and treatment*, vol. 2014, Article ID 482851, 6 pages, 2014.
[22] R. Anokye, E. Acheampong, I. Owusu, and E. Obeng, "Time series analysis of malaria in Kumasi: using ARIMA models to forecast future incidence," *Cogent Social Sciences*, vol. 4, no. 1, Article ID 1461544, 2018.
[23] S. Jere and E. Moyo, "Modelling epidemiological data using box-jenkins procedure," *Open Journal of Statistics*, vol. 6, pp. 295–302, 2016.
[24] J. L. Ferrao, J. M. Mendes, and M. Painho, "Modelling the influence of climate on malaria occurrence in chimoio municipality, mozambique," *Parasites & Vectors*, vol. 260, pp. 1–12, 2017.
[25] J. L. Ferrao, J. M. Mendes, M. Painho, and S. Zacarias, "Malaria mortality characterization and the relationship between malaria mortality and climate in Chimoio, Mozambique," *Malaria Journal*, vol. 16, p. 212, 2017.
[26] F. Nobre, A. Monteiro, P. Telles, and G. Williamson, "Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology," *Statistics in Medicine*, vol. 20, pp. 3051–3069, 2001.
[27] M. Ture and I. Kurt, "Comparison of four different time series methods to forecast hepatitis A virus infection," *Expert Systems with Applications*, vol. 31, pp. 41–46, 2006.
[28] P. M. Luz, B. V. Mendes, C. T. Codeço, C. J. Struchiner, and A. P. Galvani, "Time series analysis of dengue incidence in Rio de Janeiro. Brazil," *The American Journal of Tropical Medicine and Hygiene*, vol. 79, pp. 933–939, 2008.
[29] A. S. Ahmar, M. Botto-Tobar, A. Rahman, and R. Hidayat, "Forecasting the value of oil and gas exports in Indonesia using ARIMA box-jenkins," *JINAV: Journal of Information and Visualization*, vol. 3, no. 1, pp. 35–42, 2022.