BAREKENG: Journal of Mathematics and Its Applications
September 2026 Volume 20 Issue 3 Page 1923-1936
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss3pp1923-1936

MODEL HYBRID MARS ARIMA FOR TRIBAL-BASED MALARIA PREDICTION IN TANAH BUMBU DISTRICT, SOUTH KALIMANTAN

Abdul Khair[1], **Bambang Widjanarko Otok**[2*], **Noraida**[3],
Angga Dwi Mulyanto[4], **Cindy Cahyaning Astuti**[5]

[1,3]*Departement of Environmental Sanitation, Health Polytechnic Banjarmasin*
Jln. Mistar Cokrokusumo No.1A, Banjarbaru, 70714, Indonesia

[2,4,5]*Departement of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember*
Jln. Raya ITS, Sukolilo, Surabaya, 60111, Indonesia

[5]*Departement of Information Technology Education, Faculty of Psychology and Education,*
Universitas Muhammadiyah Sidoarjo
Jln. Mojopahit No. 666 B, Sidoarjo, 61271, Indonesia

*Corresponding author's e-mail: ** **dr.otok.bw@gmail.com**

Article Info

Article History:
Received: 8th July 2025
Revised: 26th January 2026
Accepted: 16th March 2026
Available online: 8th April 2026

***Keywords**:*
ARIMA;
Hybrid;
Malaria;
MARS;
Tanah Bumbu.

ABSTRACT

Tanah Bumbu Regency has the highest rate of malaria in South Kalimantan Province. Due to the non-linear fluctuations in malaria cases by ethnicity, a hybrid model combining Autoregressive Integrated Moving Average (MARS ARIMA) and Multivariate Adaptive Regression Splines was proposed for time series forecasting. The purpose of this study is to use the MARS ARIMA hybrid model to predict malaria cases by ethnicity in Tanah Bumbu Regency. The findings demonstrate that the best inputs for MARS modeling are significant lags found using ACF and PACF. The hybrid MARS ARIMA model performs better than standalone ARIMA or MARS models, according to predictions. Key findings show that the number of patients over 35 during the preceding two periods influences increases in malaria cases for the Banjar ethnic group. Cases exceeding 13 in two prior periods and 19 in one prior period are associated with increases for the Javanese group. Cases of more than two or fewer than two in the preceding two periods and more than eleven in one preceding period have an impact on increases among the Bugis. Prior cases below 26 have an impact on Banjar case declines, whereas prior cases below 13 and above 3 have a significant impact on Javanese case declines. This study demonstrates how well the MARS ARIMA hybrid model predicts malaria cases according to ethnicity.

How to cite this article:

A. Khair, B. W. Otok, Noraida, A. D. Mulyanto and C. C. Astuti., "MODEL HYBRID MARS ARIMA FOR TRIBAL-BASED MALARIA PREDICTION IN TANAH BUMBU DISTRICT, SOUTH KALIMANTAN," *BAREKENG: J. Math. & App.*, vol. 20, no. 3, pp. 1823-1836, Sep, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Plasmodium is a protozoan parasite that causes malaria, an infectious disease. However, a variety of factors influence the spread of malaria, just like they do for other infectious diseases. The most common topics in the context of malaria transmission in a given area are usually those related to the environment, behavior, and transmission of mosquitoes. A particular model can be developed in order to conceptualize the temporal occurrence of a disease. Future events can be predicted using this model. Both land and sea borders define the boundaries of Tanah Bumbu Regency. Both oil palm plantations and coal mining operations are quite prevalent in this region. For immigrants, this condition is an alluring offer. In Tanah Bumbu Regency, community groups are diverse gatherings that might display unique ethnic customs [1].

With a steady number of malaria cases documented between 2011 and 2013, Tanah Bumbu Regency has been designated as an endemic area. Historically, the prevalence of malaria in this region has been higher than in other South Kalimantan Province cities and regencies. Out of 29,212 cases of malaria in South Kalimantan, 6,044 cases were found in Tanah Bumbu Regency iFn 2012 [2].

Usually, time series data modeling is used in situations where the data exhibits linear and stationary fluctuations. If the two requirements listed above are satisfied, a very successful model called Autoregressive Integrated Moving Average (ARIMA) can be produced. The resulting model is insufficient in describing the system's behavior pattern, making it unsuitable for forecasting if the data conditions turn out to be inconsistent with the model's assumptions [3]. To obtain the value of one period ahead of time from data exhibiting non-linear fluctuation conditions and time series modeling, a novel methodology called MARS is proposed. It is demonstrated that this method outperforms the stochastic model.

Research by [4] includes time series data modeling with ARIMA and MARS. The results show that MARS modeling outperforms ARIMA modeling with optimal input on significant lag in HIV-AIDS prediction. The ARIMA-FFNN ensemble method was used to forecast rainfall [4]. The researchers used the ANFIS-ARIMA ensemble method to forecast rainfall in their study. The superiority of the MARS method over soft computing models in rain prediction is demonstrated in the existing literature by Taylor's work [5], which uses multiple seasonal ARIMA models, and Abraham's work [3]. Results from Septon's research [6] on recession prediction using the MARS method are better than those from ARIMA.

Resolving the main problem in the application of forecasting techniques requires choosing a time series forecasting model that generates a valid and accurate forecast value. The creation of a forecasting model that combines the benefits of Autoregressive Integrated Moving Average (ARIMA) and Multivariate Adaptive Regression Splines (MARS) must be investigated to achieve this goal. It is predicted that the combination of these two techniques, known as MARS ARIMA, will improve forecasting accuracy, especially when estimating the number of cases of malaria based on ethnicity [7], [8].

Because each forecasting technique has unique benefits and drawbacks when it comes to data analysis, the multivariate adaptive regression spline (MARS) hybrid model with autoregressive integrated moving average (ARIMA) is used in combination with other techniques. This method makes it possible to determine the best model form, which reduces prediction error [9]. Rather than choosing a single model based on how well it performs, integrating several linear combination models can improve forecasting accuracy. This combination approach is known as "hybrid." Numerous studies, both theoretical and empirical, have shown that combining several models is a useful tactic to improve prediction abilities [10], [11].

The following study findings relate to the prevalence of malaria: Humidity and Temperature [12], [13]. [14] studied modeling for disease incidence forecasting. Ethnic behavior patterns are used to forecast the prevalence of malaria in Tanah Bumbu Regency. The description above makes it clear that a study predicting malaria cases in Tanah Bumbu Regency, with an emphasis on ethnicity, is necessary. The forecasting model used in this study combines the advantages of Autoregressive Integrated Moving Average (ARIMA) and Multivariate Adaptive Regression Splines (MARS) [15], [16].

2. RESEARCH METHODS

The number of monthly malaria events by sub-district, as documented in the Tanah Bumbu District Health Office's recording and reporting system from 2014 to 2016 [17], serves as the study's unit of analysis. H.L. Bloom's theory can be used to model the incidence of malaria. 4 (four) factors (hereditary, health services, behavior, and environment) can affect the incidence of a disease. A particular model can be formed based on the frequency or occurrence of a disease over time. Future events can be predicted using this model. Fig. 1 displays the research concept framework.

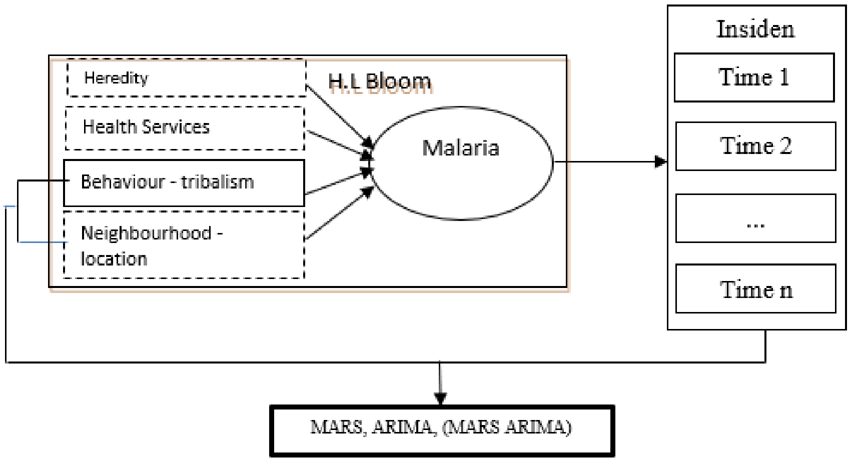


Figure 1. Conceptual Framework for Ethnic Malaria Incidence

The research framework is presented as follows:

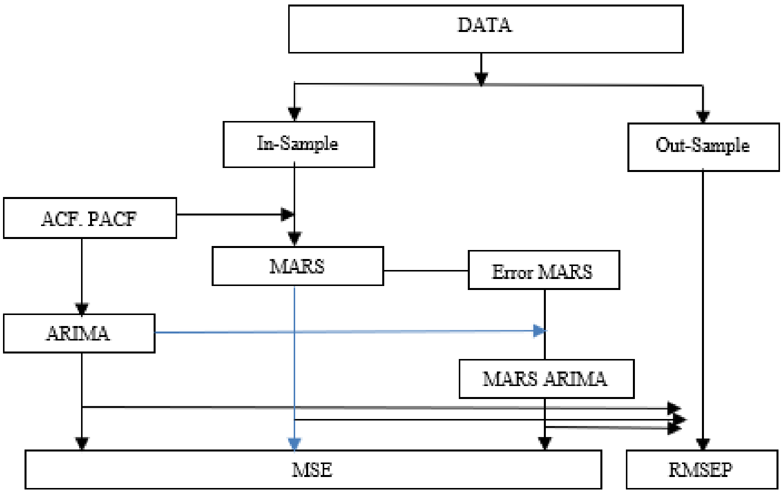


Figure 2. Research Framework ARIMA, MARS, Hybrid MARS ARIMA

Fig. 2 illustrates how response and predictor variables are used in MARS modeling. The response variable is the number of ethnic-based Malaria cases, while the predictor variable is the number of ethnic-based Malaria cases in the previous period. The research steps are as follows.

1. Secondary data, such as the monthly number of malaria events from the Tanah Bumbu District Health Office's recording and reporting system, are gathered through observations and surveys. After the data is obtained, pre-processing is carried out, then the data is divided into two parts, namely: in-sample data and out-sample data.
2. The data used to create the best model is known as in-sample data. The data used to validate and test the best model that was derived from the in-sample data is known as the out-of-sample data.
3. The predictor variable (independent variable) is the number of ethnic-based Malaria cases in the previous period obtained from determining the optimal input through the ACF and PACF plots. The response variable is the number of ethnic-based Malaria cases in 2014 - 2016. The in-sample data is the number of ethnic-based Malaria cases from 2014 to 2016 (36 months). The out-sample data is the number of ethnic-based Malaria cases in 2017.
4. In-sample data is used to construct models in three methods, namely, ARIMA, MARS and the MARS-ARIMA hybrid.
5. ARIMA model formation procedure [18].

This is the ARIMA(p,d q) model:

$$\phi(B) = \varphi_p(B)(1-B)^d = \varphi_p(B)\nabla^d. \tag{1}$$

6. MARS model formation procedure [19].
   Finding the maximum interaction (MI), the upper bound of basis functions (BF), and the minimum number of observations (MO) between knots constitute the first step. The general basis function that is employed is known to be two to four times the number of predictor variables. Each of the two predictor variables used in this study has a different basis function. Four, six, and eight are the two basic function predictor variables. The number of MO used ranges from 0 to 5, and the maximum interaction (MI) used ranges from 1 to 2. By applying a predetermined combination of BF, MI, and MO values, the MARS model is formed. Determining which predictor variables to include in the model is made easier by the creation of GCV values from each combination. This is the MARS model.

   $$y_i = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{K_m} \left[s_{km} \cdot \left(x_{v(k,m)} - t_{km}\right)\right] + \varepsilon_i = \alpha_0 + \sum_{m=1}^{M} \alpha_m B_m(x) + \varepsilon_i, \tag{2}$$

   and GCV as follows.

   $$GCV(M) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i,\alpha)\right)^2}{\left\{1 - \frac{C(\widetilde{M})}{n}\right\}^2}, \tag{3}$$

   with,

   $\tilde{C}(M) = C(M) + d.M$, $C(M) = \text{Trace}(\boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}^T) + \mathbf{1}$, $2 \le d \le 4$.

   K : The number of basis functions (nonconstant basis functions) determined in the forward stage
7. The procedure for establishing the MARS-ARIMA hybrid model [20].

   a. Determination of optimal input through ACF and PACF plots (significant lag).
   b. MARS modeling, the MARS model and the MARS error model are formed.
   c. MARS error modeling with ARIMA

      i. Develop the ARIMA model through checking Assumptions.
      ii. The MARS error model selects the optimal ARIMA model.

   d. Hybrid MARS ARIMA

      $$Y(t) = M(t) + error,$$

      with,

      $M(t) = MARS + ARIMA(error\ MARS)$,

      $M = \left[1,\ \left(y_{t,(k,m)} - \xi_{km}\right)_1^K\right]$,

      where, ARIMA (*error* MARS) are independent, identical, normally distributed).

8. Selection of the best ARIMA, MARS and ARIMA MARS Hybrid models on in-sample data based on Mean of Square Error (MSE)
9. Selection of the best ARIMA, MARS and ARIMA MARS hybrid models on out-sample data based on Root Mean Square Error Prediction (RMSEP).

3. RESULTS AND DISCUSSION

The incidence of malaria according to each ethnic group (Banjar, Javanese, Bugis) in Tanah Bumbu Regency from January 2014 (period 1) to May 2017 (period 42) is presented in a time series plot as follows [4].

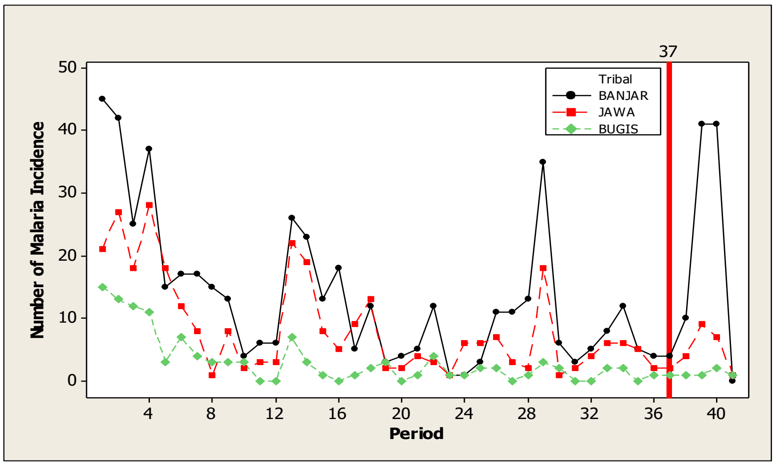
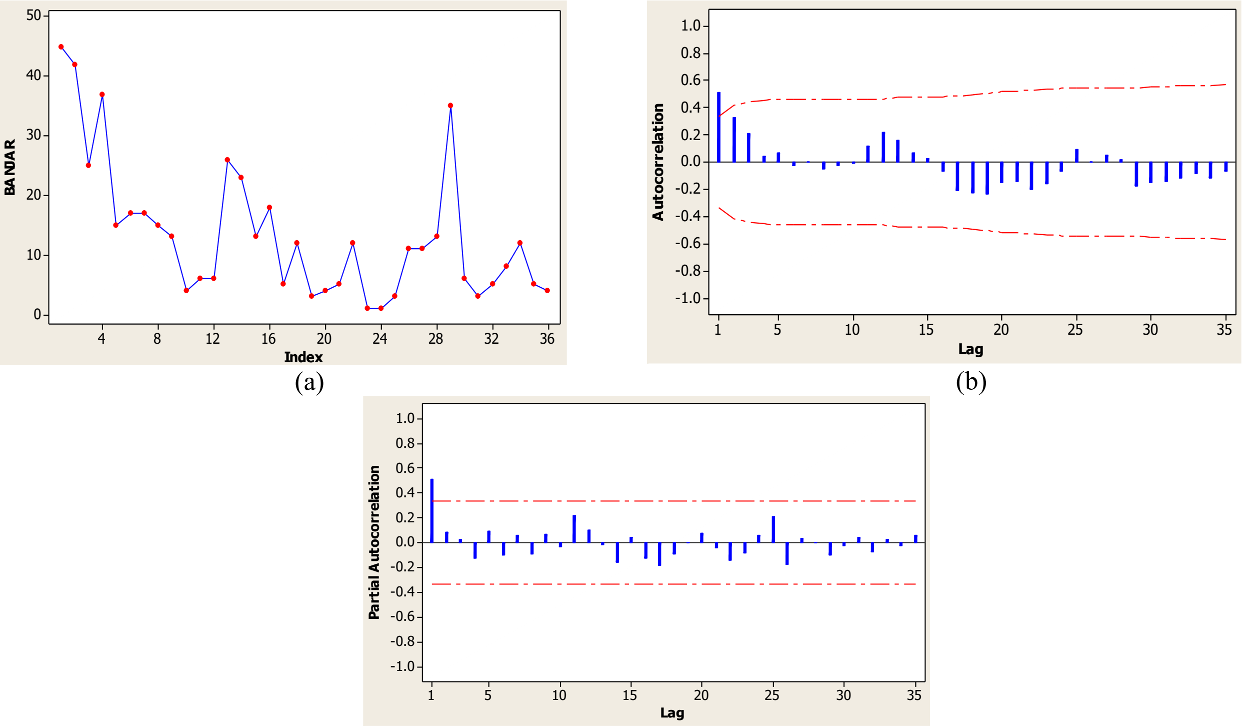


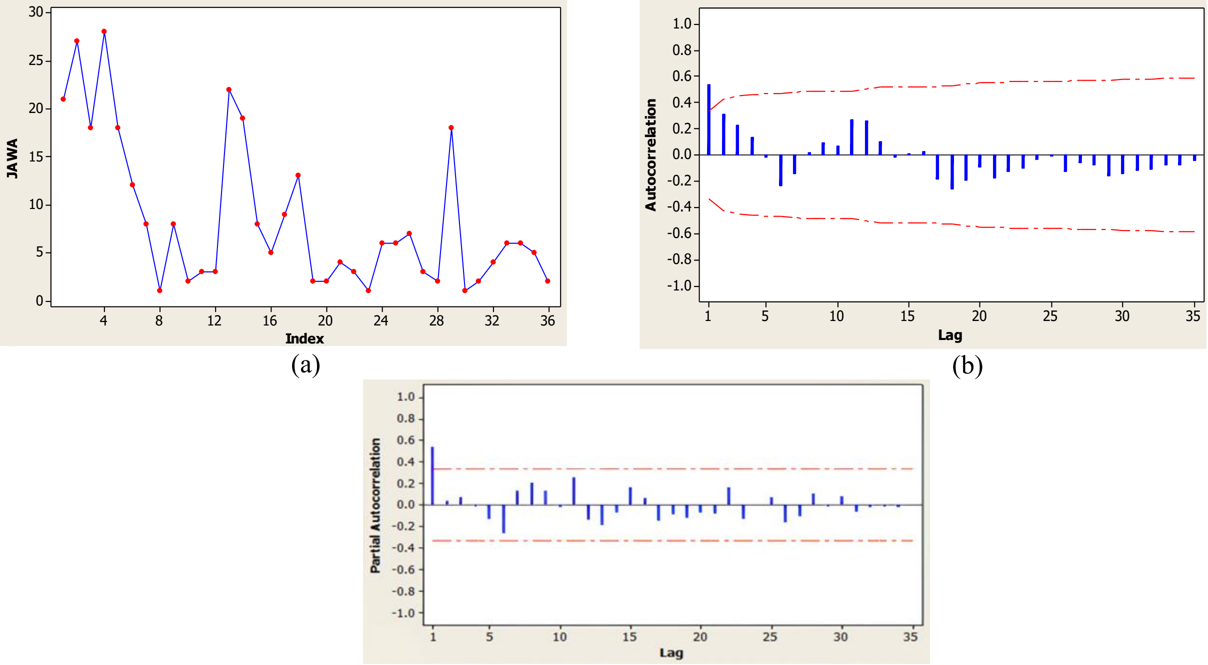
Figure 3. Timeseries Plot of Tribal-Based Malaria Incident
(Source: Analysis using Minitab)

Furthermore, tribal-based malaria incidence forecasting was carried out using ARIMA, MARS, and hybrid MARS - ARIMA methods in each tribe. Data on the number of Malaria patients in Tanah Bumbu Regency are data series, often timeseries plots, ACF and PACF are used as input for MARS modelling. The timeseries plot, ACF and PACF of data on the number of Malaria patients in Banjar tribe, Java tribe, and Bugis tribe are as follows.



(a)
(b)
(c)

Figure 4. Timeseries Plots of the Number of Malaria Patients of Banjar Tribe (a), ACF (b) and PACF (c)
(Source: Analysis using Minitab)

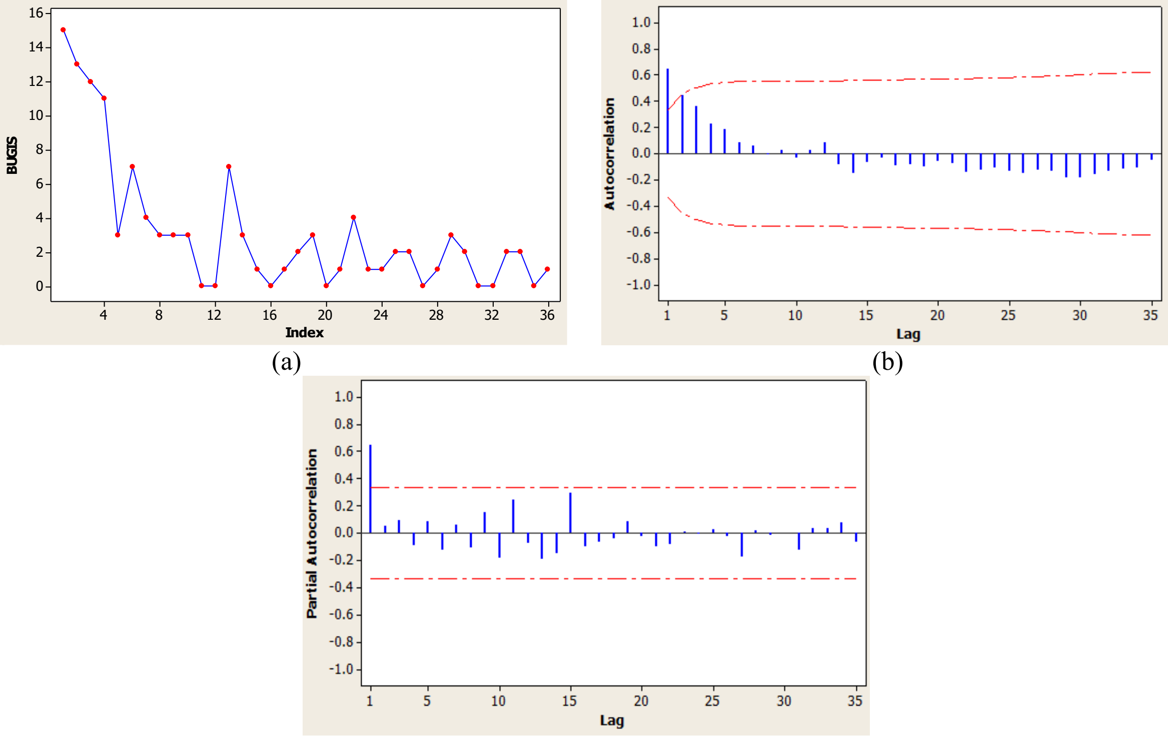


(a) (b)

(c)

Figure 5. Timeseries Plots of the Number of Malaria Patients of Javanese Tribe (a), ACF (b) and PACF (c)

(Source: Analysis using Minitab)



(a) (b)

(c)

Figure 6. Timeseries Plots of the Number of Malaria Patients of Bugis Tribe (a), ACF (b) and PACF (b)

(Source: Analysis using Minitab)

Based on Fig. 4, lag 1 is out of the ACF and PACF boundaries, while lag 2 in the ACF is almost out of the boundary, so that as input in MARS modelling is lag 1 and lag 2. Identification of ARIMA order based on ACF and PACF plots shows that for non-seasonal the PACF plot in Fig. 5 is significant at lag-1 and tends to cut off, and for seasonal the ACF plot in Fig. 5 is significant at lag-6 and tends to approach the boundary or has a pattern that dies down slowly. The PACF plot for non-seasonal order ARIMA is significant at lag-1 and for seasonal is significant at lag-6, both significant lags show a tendency to cut off. Thus, it can be predicted that the ARIMA model that is created is ARIMA (1,1,0)(0,0,0) and ARIMA $(0,0,0)(1,1,0)^6$.

Fig. 5, the real data timeseries plot shows that there is a downward trend, in addition to a sharp decline shown in the second to eighth observation, then increased again in the 13th observation, and so on until the 36th. In the ACF plot, there are 2 lags that are out of bounds, while in the PACF there is 1 lag that is out of bounds, this indicates that the lag is thought to be significantly influential in determining as MARS input prediction. The ACF and PACF plots are used to identify the ARIMA order. For non-seasonal data, the ACF plot in Fig. 5 is significant at lag-1 and tends to cut off, while for seasonal data, the ACF plot in Fig. 5 is significant at lag-6 and tends to approach the limit or exhibit a dies down pattern. The PACF plot for the non-seasonal ARIMA order is significant at lag-1 and for the seasonal is significant at lag-6, both significant lags show the tendency of cuts off. So it can be predicted that the ARIMA model formed is ARIMA (1,1,0)(0,0,0) and ARIMA $(0,0,0)(1,1,0)^6$.

It is evident from Fig. 6 that the differenced data is stationary and that the shape of the ACF and PACF patterns of the difference result data resembles the AR(1) model pattern after the level 1 differencing process is completed and the time series data plot is produced. Based on the ACF and PACF plots, the ARIMA order can be identified. For non-seasonal data, the ACF plot in Fig. 6 is significant at lag-1 and tends to cut off, while for seasonal data, the ACF plot in Fig. 6 is significant at lag-6 and tends to approach the limit or has a dying-down pattern. The PACF plot for the non-seasonal ARIMA order is significant at lag-1 and for the seasonal is significant at lag-6, both significant lags show the tendency of cuts off. Thus, it can be predicted that the ARIMA model formed is ARIMA (1,1,0)(0,0,0) and ARIMA $(0,0,0)(1,1,0)^6$.

A good model is a model whose parameters are significant, testing the significance of parameters ϕ is as follows.

Table 1. Number of Malaria Patients of Banjar, Javanese and Bugis Tribes Model Parameters Significance Test

Tribe	ARIMA Model	Parameter	Estimation	P-value	Decision	MSE	Residual (White Noise)
Banjar	(1,1,0)(0,0,0)	const ϕ_1	-1.619 -0.3983	0.316 0.018	Not Significant Significant	**88.340**	lag 12, lag 24 (pvalue=0.061; 0.104)
	(1,1,0)(0,0,0)	ϕ_1	-0.3808	0.022	Significant	**88.350**	lag 12, lag 24 (pvalue=0.093; 0.135)
	$(0,0,0)(1,1,0)^6$	const Φ_1	-5.207 -0.9765	0.000 0.008	Significant Significant	**98.640**	lag 12, lag 24 (pvalue=0.290; 0.590)
	$(0,0,0)(1,1,0)^6$	Φ_1	-0.9765	0.000	Significant	**117.24**	lag 12, lag 24 (pvalue=0.200; 0.361)
Javanese	(1,1,0)(0,0,0)	const ϕ_1	-0.7660 -0.3375	0.515 0.048	Not Significant Significant	**47.370**	lag 12, lag 24 (pvalue=0.061; 0.104)
	(1,1,0)(0,0,0)	ϕ_1	-0.3275	0.052	Significant	**46.590**	lag 12, lag 24 (pvalue=0.093; 0.135)
	$(0,0,0)(1,1,0)^6$	const Φ_1	-3.5900 -0.9816	0.006 0.000	Significant Significant		lag 12, lag 24 (pvalue=0.290; 0.590)
	$(0,0,0)(1,1,0)^6$	Φ_1	-0.9806	0.000	Significant		lag 12, lag 24 (pvalue=0.200; 0.361)
Bugis	(1,1,0)(0,0,0)	const ϕ_1	-0.5190 -0.3018	0.226 0.079	Not Significant Significant	**6.187**	lag 12, lag 24 (pvalue=0.349; 0.434)
	(1,1,0)(0,0,0)	ϕ_1	-0.2711	0.019	Significant	**5.976**	lag 12, lag 24 (pvalue=0.422; 0.490)
	$(0,0,0)(1,1,0)^6$	const Φ_1	-2.3702 -0.7929	0.000 0.000	Significant Significant	**10.080**	lag 12, lag 24 (pvalue=0.290; 0.590)
	$(0,0,0)(1,1,0)^6$	Φ_1	-0.5956	0.000	Significant	**13.637**	lag 12, lag 24 (pvalue=0.200; 0.361)

With a p-value below $\alpha=0.05$, Table 1 on the Banjar Tribe demonstrates that all parameters in the model ARIMA(1,1,0) and ARIMA $(0,0,0)(1,1,0)^6$ are significant. The ARIMA model's viability is assessed using a residual diagnosis check. The white noise assumption must be met by the input sequence used in modeling. The white noise assumption has been satisfied by the results of white noise residual testing for the models ARIMA(1,1,0) and ARIMA $(0,1,0)(1,0,0)^6$ with a significance level of $\alpha=0.05$, as indicated in Table 1, since the p-value of the residuals on the tested lags is greater than $\alpha=0.05$. However, the Java tribe demonstrates that the white noise assumption has been satisfied by the test results of white noise residuals for the models ARIMA(1,1,0)(0,0,0) and ARIMA $(0,0,0)(1,1,0)^6$ with a significance level of $\alpha=0.05$, as the p-value of the residuals on the tested lags has a value greater than $\alpha=0.05$. With a p-value below $\alpha = 0.05$, Bugis Tribe demonstrates that all parameters in the models ARIMA(1,1,0)(0,0,0) and ARIMA $(0,0,0)(1,1,0)^6$ are significant. The ARIMA model's viability is assessed using a residual diagnosis check. The white noise assumption must be met by the input sequence used in modeling. As the p-value of the residuals on the tested lags is greater than $\alpha=0.05$, the test results demonstrate that the white noise residuals for the models ARIMA(1,1,0)(0,0,0) and ARIMA $(0,0,0)(1,1,0)^6$ with a significance level of $\alpha=0.05$ have satisfied the white noise assumption. Accordingly, ARIMA (1,1,0)(0,0,0) is the best model for predicting the number of Bugis tribe malaria patients based on the smallest MSE.

Furthermore, from the best ARIMA model in each tribe, it shows that the lag is thought to be significantly influential in determining as MARS prediction input. The following Table 2 displays the MARS model with two predictor variables based on the values of GCV and MSE and O-R^2 is not centered.

Table 2. MARS Model of the Number of Malaria Patients with Banjar, Java and Bugis tribes BF, MO, MI Variations Based on GCV, O-R^2

Tribe	Model					Prediction Model
	BF	MI	MO	GCV	O-R^2	
Banjar	**4**	**1**	**1**	**70.562**	**0.761**	**Y = 16.352 + 1.593 * BF1 - 0.394 * BF4** **BF1 = max(0, B_YT2 - 35);** **BF4 = max(0, 26 - B_YT1);**
Javanese	6	1	1	40.811	0.801	Y = 53.318 + 6.809 * BF1 - 4.295 * BF2 - 5.474 * BF3 + 1.367 * BF5 BF1 = max(0, J_YT2 - 13); BF2 = max(0, 13 - J_YT2); BF3 = max(0, J_YT2 - 3); BF5 = max(0, J_YT1 - 19);
Bugis	**4**	**1**	**1**	2.912	0.872	Y = 0.511 + 0.479 * BF1 + 1.376 * BF2 + 3.147 * BF3 BF1 = max(0, BUG_YT2 - 2); BF2 = max(0, 2 - BUG_YT2); BF3 = max(0, BUG_YT1 - 11);

Source: Processed Data

The second model, which has BF = 4, MI = 1, and MO = 1 and has a GCV value of 70.562 and an O-R^2 value of 0.761, is the best model based on the criterion of the model with the smallest GCV value, as shown in Table 2 for Banjar Tribe. The most effective MARS model is:

$$Y = 16.352 + 1.593 * BF1 - 0.394 * BF4, \tag{4}$$

with,

BF1 = max(0, B_YT2 - 35.000);
BF4 = max(0, 26.000 - B_YT1);

Eq. (4) provides the following interpretation of the MARS model.

BF1 = max(0, B_YT2 - 35.0000).

Accordingly, the coefficient of BF1 indicates that the number of Banjar tribe malaria patients will rise by 1.593 ~ 2 patients if there were more than 35 patients from the Banjar tribe during the preceding two periods.

BF4 = max(0, 26.000 - B_YT1).

Accordingly, the BF4 coefficient indicates that the number of Banjar tribe malaria patients will drop by 0.394 ~ 1 patient if there were fewer than 26 patients in a prior period.

The Java tribe displays every model that has been attempted in order to determine the best model using the criterion of the model with the lowest GCV value. This model is the 12th model, which has the numbers BF = 6, MI = 1, and MO = 1 and has a GCV value of 40.811 and an O-R2 value of 0.801. The most effective MARS model is:

$$Y = 53.318 + 6.809 * BF1 - 4.295 * BF2 - 5.474 * BF3 + 1.367 * BF5, \tag{5}$$

with

BF1 = max(0, J_YT2 - 13.000);
BF2 = max(0, 13.000 - J_YT2);
BF3 = max(0, J_YT2 - 3.000);

BF5 = max(0, J_YT1 - 19.000);

Eq. (5) provides the following interpretation of the MARS model.

BF1 = max(0, J_YT2 - 13.000).

According to the coefficient of BF1, the number of malaria patients from the Javanese tribe will rise by 6.809 ~ 7 individuals if there were more than 13 cases in the preceding two periods.

BF2 = max(0, 13.000 - J_YT2).

According to the BF2 coefficient, the number of malaria patients from the Javanese tribe will drop by 4.295 to 5 individuals if there were fewer than 13 cases during the preceding two periods.

BF3 = max(0, J_YT2 - 3.000).

According to the coefficient of BF3, the number of malaria patients from the Javanese tribe will drop by 5.474 to 6 individuals if there were more than three cases during the preceding two periods.

BF5 = max(0, J_YT1 - 19.000).

According to the coefficient of BF5, this means that the number of malaria cases among Javanese tribes will increase by 1.367 to 2 individuals if there were more than 19 cases in a previous period.

Bugis Tribe displays every model that has been attempted in order to determine the best model based on the criterion of the model with the lowest GCV value. This is the second model, which has a GCV value of 2.912 and an O-R2 value of 0.872 and has the numbers BF = 4, MI = 1, and MO = 1. It's the best MARS model.

$$Y = 0.511 + 0.479 * BF1 + 1.376 * BF2 + 3.147 * BF3, \tag{6}$$

with

BF1 = max(0, BUG_YT2 - 2.000);
BF2 = max(0, 2.000 - BUG_YT2);
BF3 = max(0, BUG_YT1 - 11.000);

Eq. (6) provides the following interpretation of the MARS model.

BF1 = max(0, BUG_YT2 - 2.000).

This indicates that the coefficient of BF1 will indicate that the number of Bugis people with malaria will rise by 0.479 ~ 1 person if the number of Bugis people with malaria in the preceding two periods was greater than 2.

BF2 = max(0, 2.000 - BUG_YT2).

Accordingly, the BF2 coefficient indicates that if there were fewer than two cases of Bugis malaria during the preceding two periods, there will be 1.376 ~ 2 more cases.

BF3 = max(0, BUG_YT1 - 11.000).

According to the coefficient of BF3, this means that the number of malaria cases among Bugis tribes will increase by 3.147 to 4 individuals if there were more than 11 cases in a previous period.

The error value, in this case referred to as MARS error, is modelled using the ARIMA method. ARIMA modelling starts from identification through time series plots to checking residual diagnoses to obtain the following model.

Table 3. Malaria Patients of Banjar, Java and Bugis Tribes MARS Error Model Number Parameter Significance Test

Tribe	**Model ARIMA**	**Parameter**	**Estimation**	**P-value**	**Decision**	**MSE**	**Residual (White Noise)**
Banjar	(1,1,0)(0,0,0)	ϕ_1	-0.5696	0.000	Significant	81.36	lag 12, lag 24 (pvalue=0.061; 0.104)
Javanese	(2,1,0)(0,0,0)	ϕ_1	-0.5199	0.001	Significant	25.916	lag 12, lag 24 (pvalue=0.061; 0.104)
		ϕ_2	-0.6063	0.000	Significant		
Bugis	(1,1,0)(0,0,0)	ϕ_1	-0.7759	0.226	Significant	14.327	lag 12, lag 24 (pvalue=0.349; 0.434)
		ϕ_2	-0.6932	0.079	Significant		

From the MARS model, two results are obtained, namely the MARS model and the MARS error. Hybrid MARS ARIMA is a combination of MARS model with MARS error modelled with ARIMA. Moderate error MARS on Banjar tribe is.

$$(\varepsilon M_t - \varepsilon M_{t-1}) = -0.5696M_{t-1} + a_t \Rightarrow \varepsilon M = 0.4304\varepsilon M_{t-1} + a_t,$$

So that the Hybrid MARS ARIMA model for predicting the number of Malaria disease sufferers in the Banjar tribe is:

$$Y = 16.352 + 1.593 * BF1 - 0.394 * BF4 + \varepsilon M_t, \tag{7}$$

$$Y_t = 16.352 + 1.593 * [\max(0, Y_{t-2} - 35)] - 0.394 * [\max(0, 26 - Y_{t-1})] + \varepsilon M_t,$$

with eM_t = 0.4304M_t-1 + a_t

then.

$$Y_t = 16.352 + 1.593 * [\max(0, Y_{t-2} - 35)] - 0.394 * [\max(0, 26 - Y_{t-1})] + 0.4304\text{Mt-1} + a_t.$$

The error value, in this case referred to as MARS error, is modelled using the ARIMA method. ARIMA modelling starts from identification through time series plots to checking residual diagnoses on each term, the following model is obtained.

ARIMA 2 1 0 'er_jawa';

$$(\varepsilon M_t - \varepsilon M_{t-1}) = -0.5199\varepsilon M_{t-1} - 0.6063\varepsilon M_{t-2} + a_t \Rightarrow \varepsilon M = 0.4801\varepsilon M_{t-1} - 0.6063\varepsilon M_{t-2} + a_t.$$

From the MARS modl, two results are obtained, namely the MARS model and the MARS error. Hybrid MARS ARIMA is a combination of MARS model with MARS error modelled with ARIMA. The form of the Hybrid MARS ARIMA model for predicting the number of Malaria patients is:

$$\mathrm{Y} = 53.318 + 6.809\mathrm{BF1} - 4.295\mathrm{BF2} - 5.474\mathrm{BF3} + 1.367\mathrm{BF5} + \varepsilon M_t, \tag{8}$$

or

$$\begin{aligned}\mathrm{Y} = 53.318 &+ 6.809(\max(0, \mathrm{Y}_{t-2} - 13)) - 4.295(\max(0,13 - \mathrm{Y}_{t-2})) - 5.474(\max(0, \mathrm{Y}_{t-2} - 3)) \\ &+1.367(\max(0, \mathrm{Y}_{t-1} - 19)) + \varepsilon M_t,\end{aligned} \tag{9}$$

with $\varepsilon M_t = 0.4801\ \varepsilon M_{t-1} - 0.6063\ \varepsilon M_{t-2} + a_t$, then

$$\begin{aligned}\mathrm{Y} = 53.318 &+ 6.809(\max(0, \mathrm{Y}_{t-2} - 13)) - 4.295(\max(0,13 - \mathrm{Y}_{t-2})) - 5.474(\max(0, \mathrm{Y}_{t-2} - 3)) \\ &+1.367(\max(0, \mathrm{Y}_{t-1} - 19)) + 0.4801\varepsilon M_{t-1} - 0.6063\varepsilon M_{t-2} + a_t.\end{aligned} \tag{10}$$

Bugis Tribe:

The error value, in this case referred to as MARS error, is modelled using the ARIMA method. ARIMA modelling starts from identification through time series plots to checking residual diagnoses on each term, the following model is obtained.

ARIMA 2 1 0 'er_Bugis';

$$(\varepsilon M_t - \varepsilon M_{t-1}) = -0.7759\varepsilon M_{t-1} - 0.6932\varepsilon M_{t-2} + a_t \Rightarrow \varepsilon M = 0.2241\varepsilon M_{t-1} - 0.6932\varepsilon M_{t-2} + a_t.$$

From the MARS model, two results are obtained, namely the MARS model and the MARS error. Hybrid MARS ARIMA is a combination of MARS model with MARS error modelled with ARIMA. The form of the Hybrid MARS ARIMA model for predicting the number of Malaria patients is:

$$\mathrm{Y} = 0.511 + 0.479\mathrm{BF1} + 1.376\mathrm{BF2} + 3.147\mathrm{BF3} + \varepsilon M_t, \tag{11}$$

or,

$$\begin{aligned}\mathrm{Y} = 53.318 &+ 6.809(\max(0, \mathrm{Y}_{t-2} - 13)) - 4.295(\max(0,13 - \mathrm{Y}_{t-2})) - 5.474(\max(0, \mathrm{Y}_{t-2} - 3)) \\ &+1.367(\max(0, \mathrm{Y}_{t-1} - 19)) + \varepsilon M_t\end{aligned} \tag{12}$$

Comparison of GCV and R^2 values as well as RMSEP on each ARIMA, MARS and HYBRID MARS ARIMA models results are presented in the following Table.

Table 4. Best Model Criteria Based on MSE

Tribes	**Methods**	**Data In Sample**	
		MSE	**RMSE**
BANJAR	ARIMA (1,1,0)(0,0,0)	88.350	9.399
	MARS (4 1 1)	52.894	7.272
	HYBRID MARS ARIMA	**50.076**	**7.076**
JAVANESE	ARIMA (1,1,0)(0,0,0)	46.590	6.826
	MARS (6 1 1)	21.293	4.614
	HYBRID MARS ARIMA	**20.684**	**4.548**
BUGIS	ARIMA (1,1,0)(0,0,0)	5.976	2.444
	MARS (4 1 1)	1.893	1.378
	HYBRID MARS ARIMA	**1.481**	**1.217**

The MSE value in Table 4 shows that for in-sample data that has the smallest MSE value is the HYBRID MARS ARIMA model. Thus, the model is the best model to predict the number of Malaria patients in Tanah Bumbu Regency, and is presented in the following Fig. 7.

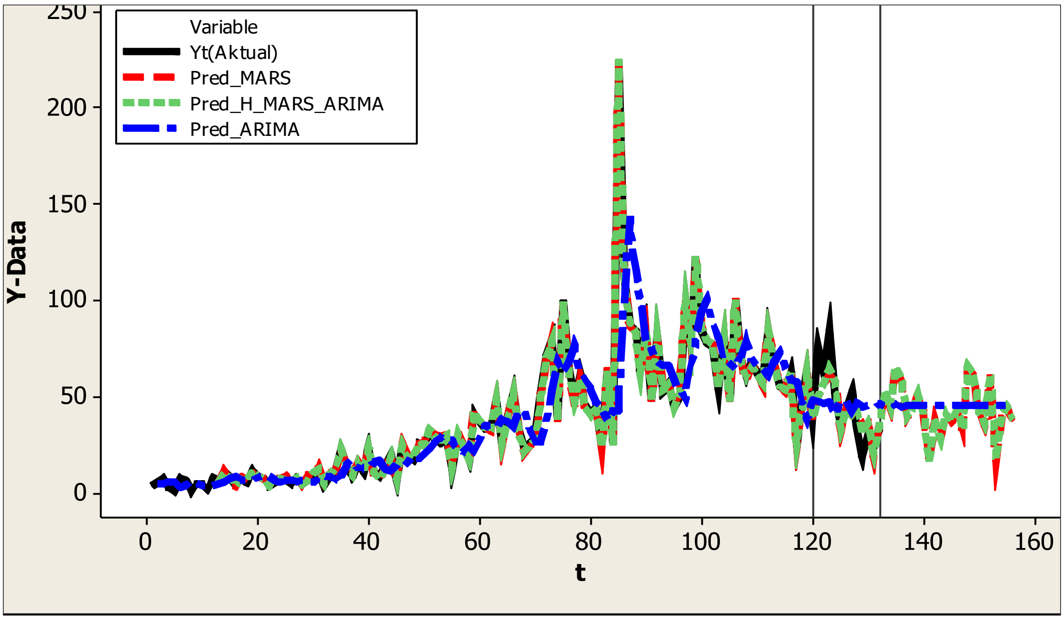


Figure 7. Plot of in-sample and Out-Sample Data on The Number of Malaria Patients in Tanah Bumbu District *(Source: Analysis using Minitab)*

The MSE values in Table 5 and Fig. 7 show that for in-sample data that has the smallest RMSE value is the HYBRID MARS ARIMA model, both for the Banjar Tribe, Javanese Tribe, and Bugis Tribe. Therefore, it is the best model to predict the number of Malaria patients in Tanah Bumbu Regency.

Table 5. Best Model Criteria Based on RMSEP

Malaria Incidence Forecast	**Banjar Tribe**				**Javanese Tribe**				**Bugis Tribe**			
	Actual	**ARIMA**	**MARS**	**Hybrid**	**Actual**	**ARIMA**	**MARS**	**Hybrid**	**Actual**	**ARIMA**	**MARS**	**Hybrid**
2017 January	4	4.235	6.11	4.353	2	2.660	2.52	1.311	1	0.802	1.26	1.023
2017 February	10	4.291	6.11	5.201	4	2.766	2.52	5.199	1	0.782	1.26	0.862
2017 March	41	4.270	6.11	7.784	9	2.731	2.52	3.546	1	0.787	1.26	1.081
2017 April	30	4.278	6.11	21.126	7	2.742	2.52	4.734	2	0.786	1.26	1.381
2017 May	13	4.274	6.11	16.392	1	2.739	2.52	0.742	1	0.786	1.26	1.006
RMSEP		423.90	371.0	243.342		12.483	13.36	7.372		0.330	0.164	0.082

The RMSEP value in Table 5 shows that for the out-sample data that has the smallest RMSE value is the HYBRID MARS ARIMA model, both for the Banjar Tribe, Javanese Tribe, and Bugis Tribe. Therefore, this model is the best model to predict the number of Malaria patients in Tanah Bumbu Regency.

4. CONCLUSION

4 (four) steps make up the process of creating the Hybrid MARS ARIMA model: determining the best input, MARS modeling, MARS error modeling with ARIMA, and Hybrid MARS ARIMA. Using RMSE, the best ARIMA, MARS, and Hybrid MARS ARIMA models are chosen for out-of-sample data. The best input for MARS modeling is significant lag in PACF since it yields the lowest GCV value. The Hybrid MARS ARIMA model is the prediction model that performs well on both in-sample and out-of-sample data. The number of patients in the Banjar tribe who were exposed to more than 35 people during the previous two periods, the number of patients in the Java tribe who were exposed to more than 13 people and the number of patients who were exposed to more than 19 people during the previous two periods, and the number of patients in the Bugis tribe who were exposed to more than two people, fewer than two people, and more than eleven people during the previous period all influenced the prediction of an increase in the number of malaria patients in Tanah Bumbu Regency. The number of patients in the previous period who were exposed to fewer than 26 people influences the Banjar tribe's prediction of a decline in the number of malaria patients in Tanah Bumbu Regency, while the Java tribe's prediction is influenced by the number of patients in the previous two periods who were exposed to fewer than 13 people and more than three people.

Author Contributions

Abdul Khair: Conceptualization, Methodology, Software. Bambang Widjanarko Otok: Conceptualization, Resources, Draft Preparation. Noraida: Formal Analysis, Validation. Angga Dwi Mulyanto: Validation, Writing-Review and Editing. Cindy Cahyaning Astuti: Validation, Writing-Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The research team would like to express our deepest gratitude to the individuals and institutions that have made important contributions to the success of our research. The support, guidance, and collaboration provided have been invaluable to the smooth progress and success of this research.

Declarations

The authors declare that there are no conflicts of interest associated with this study. The authors have equally contributed to the design, analysis, and interpretation of the study, and the final manuscript was reviewed and approved by all authors.

Declaration of Generative AI and AI-assisted Technologies

The authors declare that no generative AI or AI-assisted technologies were used in the preparation of this manuscript, including for writing, editing, data analysis, or the creation of tables and figures.

REFERENCES

[1] K. Alemu, A. Worku, Y. Berhane, and A. Kumie, "MEN TRAVELING AWAY FROM HOME ARE MORE LIKELY TO BRING MALARIA INTO HIGH ALTITUDE VILLAGES, NORTHWEST ETHIOPIA," *PLoS One*, vol. 9, no. 4, Apr. 2014, doi: https://doi.org/10.1371/journal.pone.0095341 .

[2] J. C. Berthélemy, J. Thuilliez, O. Doumbo, and J. Gaudart, "MALARIA AND PROTECTIVE BEHAVIOURS: IS THERE A MALARIA TRAP?," *Malar J*, vol. 12, no. 1, 2019, doi: https://doi.org/10.1186/1475-2875-12-200.

[3] K. Bashar, H. M. Al-Amin, M. S. Reza, M. Islam, Asaduzzaman, and T. U. Ahmed, "SOCIO-DEMOGRAPHIC FACTORS INFLUENCING KNOWLEDGE, ATTITUDE AND PRACTICE (KAP) REGARDING MALARIA IN BANGLADESH," *BMC Public Health*, vol. 12, no. 1, 2018, doi: https://doi.org/10.1186/1471-2458-12-1084.

[4] G. T. Wilson, "TIME SERIES ANALYSIS: FORECASTING AND CONTROL, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1," *J Time Ser Anal*, vol. 37, no. 5, pp. 709–711, Sep. 2016, doi: https://doi.org/10.1111/jtsa.12194.

[5] S. C. Hillmer and W. W. S. Wei, “TIME SERIES ANALYSIS: UNIVARIATE AND MULTIVARIATE METHODS.,” *J Am Stat Assoc*, vol. 86, no. 413, p. 245, Mar. 2019, doi: https://doi.org/10.2307/2289741
[6] Helmut. Lütkepohl, *NEW INTRODUCTION TO MULTIPLE TIME SERIES ANALYSIS*. New York: Springer, 2015.
[7] P. E. Pfeifer and S. J. Deutsch, “A THREE-STAGE ITERATIVE PROCEDURE FOR SPACE-TIME MODELING,” 2018.
[8] F. Nida Aryani, S. Sulistijowati Handajani, E. Zukhronah Program Studi Statistika, and F. Matematika dan IlmuPengetahuan Alam Universitas Sebelas Maret, “PENERAPAN MODEL GENERALIZED SPACE TIME AUTOREGRESSIVE PADA DATA NILAI TUKAR PETANI DI TIGA PROVINSI PULAU SUMATERA APPLICATION OF GENERALIZED SPACE TIME AUTOREGRESSIVE MODEL ON FARMER EXCHANGE RATE DATA IN THREE PROVINCES OF THE SUMATERA ISLAND,” 2020. Doi: https://doi.org/10.19184/jid.v21i2.17226
[9] D. N. Gujarati, B. Burr Ridge, I. Dubuque, I. Madison, and W. New York San Francisco St Louis Bangkok Bogota Caracas Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal New Delhi Santiago Seoul Singapore Sydney Taipei Toronto, *BASIC ECONOMETRICS FOURTH EDITION*. 2017. [Online]. Available: www.mhhe.com
[10] A. Khair, S. Sarmanu, S. Martini, and B. W. Otok, “VECTOR AUTOREGRESSIVE MODELING ON CASES OF MALARIA BASED ON THE TRIBAL IN TANAH BUMBU DISTRICT,” *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, vol. 5, no. 3, pp. 150–160, Dec. 2018, doi: https://doi.org/10.18860/ca.v5i3.5880.
[11] R. Faulina, D. Ayu Lusia, B. W. Otok, and H. Kuswanto, “ENSEMBLE METHOD BASED ON ANFIS-ARIMA FOR RAINFALL PREDICTION*M 2017, International Conference on Statistics in Science, Business and Engineering (ICSSBE),* doi: https://doi.org/10.1109/ICSSBE.2012.6396564 .
[12] S. Borovkova, H. P. Lopuhaä, and B. N. Ruchjana, “CONSISTENCY AND ASYMPTOTIC NORMALITY OF LEAST SQUARES ESTIMATORS IN GENERALIZED STAR MODELS,” *Stat Neerl*, vol. 62, no. 4, pp. 482–508, Nov. 2018, doi: https://doi.org/10.1111/j.1467-9574.2008.00391.x
[13] W. Enders (2018), “STUDENTS’ RESOURCE GUIDE TO ACCOMPANY APPLIED ECONOMETRIC TIME SERIES (4TH EDITION) THIS VERSION OF THE GUIDE IS FOR STUDENT USERS OF RATS AND EVIEWS.” [Online]. Available: www.time-series.net.
[14] B. W. Otok, D. A. Lusia, Suhartono, R. Faulina, Sutikno, and H. Kuswanto, “ENSEMBLE METHOD BASED ON ARIMA-FFNN FOR CLIMATE FORECASTING,” in *ICSSBE 2012 - Proceedings, 2012 International Conference on Statistics in Science, Business and Engineering: “Empowering Decision Making with Statistical Sciences,”* 2017, pp. 244–247. doi: https://doi.org/10.1109/ICSSBE.2012.6396565.
[15] Suhartono, Ria Faulina, Dwi Ayu Lusia, Bambang W. Otok, Sutikno, Heri Kuswanto, (2017). ENSEMBLE METHOD BASED ON ANFIS-ARIMA FOR RAINFALL PREDICTION, *International Conference on Statistics in Science, Business and Engineering (ICSSBE)*. 10.1109/ICSSBE.2012.6396564, Langkawi, Malaysia.
[16] B. W. Otok, D. A. Lusia, Suhartono, R. Faulina, Sutikno, and H. Kuswanto, “ENSEMBLE METHOD BASED ON ARIMA-FFNN FOR CLIMATE FORECASTING,” in *ICSSBE 2012 - Proceedings, 2015 International Conference on Statistics in Science, Business and Engineering: “Empowering Decision Making with Statistical Sciences,”* 2012, pp. 244–247. doi: https://doi.org/10.1109/ICSSBE.2012.6396565.
[17] Biro Pusat Statistik (BPS) Kabupaten Tanah Bumbu, (2014). *TANAH BUMBU DALAM ANGKA*. www.time-series.net.
[18] Pfeifer, P.E dan Deutsch, S.J. (2021a). A THREE STAGE ITERATIVE PROCEDURE FOR SPACE TIME MODELLING. *Technometrics*, 22 (1), 35-47.doi: https://doi.org/10.2307/1268381
[19] W. Enders (2019), “STUDENTS’ RESOURCE GUIDE TO ACCOMPANY APPLIED ECONOMETRIC TIME SERIES (4TH EDITION) THIS VERSION OF THE GUIDE IS FOR STUDENT USERS OF RATS AND EVIEWS.” [Online]. Available: www.time-series.net.
[20] Wei, W.W.S. (2016). *TIME SERIES ANALYSIS: UNIVARIATE AND MULTIVARIATE METHODS*. Addison-Wesley Publishing Co., USA.

.