BAREKENG: Journal of Mathematics and Its Applications
March 2024 Volume 18 Issue 1 Page 0117–0128
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss1pp0117-0128

VECTOR AUTOREGRESSIVE WITH OUTLIER DETECTION ON RAINFALL AND WIND SPEED DATA

Lisa Lestari [1*], Evy Sulistianingsih[2], Hendra Perdana[3]

[1,2,3]*Statistic Study Program, Faculty of Mathematics and Natural Sciences, Tanjungpura University*
Prof. Dr. H. Hadari Nawawi Street, Pontianak, 78124, Indonesia

Corresponding author's e-mail: [1*] ***lisalestari@student.untan.ac.id***

Article History:

Received: 26th July 2023
Revised: date, month year
Accepted: date, month year

Keywords:

VAR;
Outliers;
Rainfall;
Wind Speed.

ABSTRACT

Vector Autoregressive (VAR) is a multivariate time series model that analyzes more than one variable where each variable in the model is endogenous. VAR is one of the models used in forecasting rainfall and wind speed. In observations of rainfall and wind speed, there are usually a series of events whose values are far from other observations or can be said to be outliers. The purpose of this study is to compare the VAR model on rainfall and wind speed data before and after outlier detection. This study uses secondary data, namely monthly data on rainfall and wind speed from 2019 to 2021. From the analysis results, the smallest AIC value obtained in the VAR model before outlier detection was 4.94, then the smallest AIC value in the VAR model after outlier detection was 0.25. Thus, it can be concluded that the best model is obtained in the VAR model after outlier detection seen from the smallest AIC value of the two VAR models.

How to cite this article:

L. Lestari, E. Sulistianingsih and H. Perdana., "VECTOR AUTOREGRESSIVE WITH OUTLIER DETECTION ON RAINFALL AND WIND SPEED DATA," *BAREKENG: J. Math. & App.*, vol. 18, iss. 1, pp. 0117-0128, March, 2024.

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

Climate change is a change in the physical conditions of the Earth's atmosphere such as temperature, precipitation patterns, and several other climatic variables. Climate change and weather conditions are factors that affect activities in various fields of human life, specifically in the transportation sector. Air and sea transportation activities are carried out by considering weather conditions such as rainfall and wind speed. Therefore, rainfall and wind speed forecasts are important for the success of all planned transportation activities [1].

Forecasting is an activity carried out to measure future uncertainty based on past conditions and to assist decision-making. Forecasting is generally done on time series data. Forecasting time series data can be done for univariate and multivariate time series models. A univariate time series model is a model that contains only one observed variable, and a multivariate time series model is a model that contains several observed variables simultaneously [2].

Rainfall is one of the dynamic physical properties for which data is collected at regular time intervals, so time series analysis is necessary [3]. Rainfall is related to other weather factors such as temperature, humidity, and wind speed. Wind speed has an erratic nature. Therefore, a model is needed to predict accurate wind speed [4]. Rainfall and wind speed variables are interconnected variables [4] so, the determination of endogenous and exogenous variables is unknown for forecasting modeling. Therefore, simultaneous forecasting of rainfall and wind speed is carried out by viewing both variables as endogenous variables, using the Vector Autoregressive (VAR) method. VAR is simultaneous equation modeling that has multiple endogenous variables simultaneously [5].

Some previous studies that use VAR models include Hardani, Hoyyi, and Sudarno [6] on inflation rate data, interest rates, and IHSG. Ibnas, Salam, and Agustina [7] used the VAR model in forecasting the population in the Gowa Regency. This study aims to model rainfall and wind speed using the VAR model. In observations of rainfall and wind speed, there are usually unexpected events, which can produce a series of events whose values are far from other observations or can be said to be outliers. Outliers can cause conclusions from data analysis to be invalid, so it is necessary to detect and remove the effects of outliers in data analysis [8]. Therefore, in this study, outlier detection is performed on the VAR model to compare the VAR model before and after outlier detection to show the impact of outliers on the VAR model.

2. RESEARCH METHODS

2.1 Data Source

This study uses monthly data of rainfall in millimeters (mm) and wind speed in knots. The data used is secondary data obtained from the Pontianak City Meteorological Station. The data period used in the formation of the Vector Autoregressive (VAR) model was three years, from January 2019 to December 2021. The sample size used in this study was 36 samples.

2.2 Stationarity

The method of testing stationarity against the average in this study was carried out by the Augmented Dickey-Fuller (ADF) test. If the data is not stationary on average, then a differencing process is performed [9]. Suppose the equation estimated by the ADF test is as follows.

$$\Delta Y_t = \beta_0 + \phi Y_{t-1} + \sum_{i=1}^{p} a_i \Delta Y_t + e_{t.}$$

The hypothesis used for ADF testing is as follows.

$H_0: \phi = 0$ (there is a unit root or the data is non-stationary).
$H_1: \phi < 0$ (no unit root or the data is stationary).

With the following test statistics:

$$ADF_{hitung} = \frac{\hat{\emptyset}}{SE(\hat{\emptyset})}, \quad (1)$$

with $\hat{\emptyset} = \frac{\sum_{t=2}^{n}(\Delta Y_t - \Delta\bar{Y}_t)(Y_{t-1} - \bar{Y}_{t-1})}{\sum_{t=2}^{n}(Y_{t-1} - \bar{Y}_{t-1})^2}$, $\hat{\sigma}^2 = \frac{\sum_{t=2}^{n}\hat{e}_t^2}{n-m}$, n is the number of observations, and m is the number of parameter estimates. The null hypothesis is refused if the absolute value of $ADF_{calculate}$ is greater than the ADF critical value or the ADF probability value is smaller than the specified significance level, so the data can be said to be stationary [10].

The data is considered stationary in variance if the data is constant over time using Box-Cox. When the value of λ is close to one, it can be said that the data is stationary in variance. A Box-Cox transformation is performed if the data is non-stationary in variance. The following is given as a transformation for several values of λ [11].

Table 1. Box Cox Transformation

Value λ	-2.0	-1.0	-0.5	0.0	0.5	1.0	2.0
Transformation	$\frac{1}{Z_t^2}$	$\frac{1}{Z_t}$	$\frac{1}{\sqrt{Z_t}}$	In Z_t	$\sqrt{Z_t}$	Z_t	Z_t^2

Matrix Partial Autoregressive Function (MPACF) is a generalization of PACF into time series vector form built by Tiao and Box in 1981 [12], which is used to identify the order of the tentative model. MPACF has cut off properties after lag p in the VAR(p) model [13].

2.3 Akaike Information Criteria (AIC)

The best model is chosen based on the smallest Akaike Information Criteria (AIC) value. According to Hermayani (2014), AIC is a criterion introduced by Akaike in 1973 to select the best model by considering many parameters in the model [2]. AIC is formulated as follows [12]:

$$AIC(M) = n\text{In}(\hat{\sigma}_{\alpha}^2) + 2M \quad (2)$$

where M is the number of parameters in the model. The optimal order in a model is chosen based on the value of M, which is a function of p and q, so that the AIC(M) value is minimum.

2.4 Autoregressive Vector Model

Vector Autoregressive (VAR) is a system of equations that shows every variable as a linear function of constants and lag values of the variable itself, as well as lags of other variables in the model. The definition of the VAR model is that all variables in the model are endogenous variables, or it can be said that the VAR model does not need to distinguish between endogenous and exogenous variables [14]. The VAR model with order p for k independent variables at time t can be modeled as Equation (3) [12].

$$\boldsymbol{Y_t} = \boldsymbol{\phi_0} + \boldsymbol{\phi_{1,}Y_{1,t-1}} + \boldsymbol{\phi_{2,}Y_{2,t-2}} + \cdots + \boldsymbol{\phi_{p,}Y_{p,t-p}} + \boldsymbol{a_t} \quad (3)$$

with $\boldsymbol{Y_t}$ is a vector of independent variables of size $m \times 1$, $\boldsymbol{\phi_0}$ is a vector of intercepts of size $m \times 1$, $\boldsymbol{\phi_i}$ is a parameter matrix of size $m \times m$ for each $i = 1,2,\ldots,p$ and $\boldsymbol{a_t}$ is a vector of residuals of size $m \times 1$.

2.5 White Noise Residual Assumption Test

The White Noise residual assumption test on multivariate time series data aims to see whether the residuals of the model are independent of each other. This assumption test can be done using the Ljung-Box test statistic for multivariate cases with the following hypothesis [11].

$H_0 : \rho_1 = \rho_2 = \rho_3 = \cdots = \rho_k = 0$ (residuals qualified by the White Noise condition)
H_1 : There is at least one $\rho_i \neq 0$ with i=1, 2, ..., k (residuals do not qualify the White Noise condition)

The test statistic used is as follows:

$$Q_h = n\sum_{i=1}^{h} tr(\boldsymbol{\hat{C}_i'}\boldsymbol{\hat{C}_0^{-1}}\boldsymbol{\hat{C}_i}\boldsymbol{\hat{C}_0^{-1}}) \quad (4)$$

with the assumption of rejecting H_0 if $Q_h \geq \chi^2$ or *p-value* $< \alpha$. The value of C_i is obtained from $\boldsymbol{\hat{C}_i} = n^{-1}\sum_{t=i+1}^{n}\hat{a}_t\hat{a}_{t-1}'$, where $\boldsymbol{\hat{C}_i}$ is the autocovariance estimation matrix of the residuals a_t, $\boldsymbol{\hat{C}_0}$ is the $\boldsymbol{\hat{C}_i}$ matrix when $i = 0$ and n is the number of samples.

2.6 Multivariate Normality Test

Multivariate normality tests can be performed using inferential tests, one of which is the *Henze-Zirkler* test. The *Henze-Zirkler* test tests data normality based on the distance between two distribution functions. Multivariate normality testing with the *Henze-Zirkler* test uses the following hypothesis [15]:

H_0: multivariate residuals are normally distributed
H_1: multivariate residuals are not normally distributed

The test statistic used is the *Henze-Zirkler* test statistic with **Equation (5)**:

$$HZ = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} e^{\frac{\beta^2}{2}\boldsymbol{D_{ij}}} - 2(1+\beta^2)^{\frac{k}{2}}\frac{1}{n}\sum_{i=1}^{n} e^{\frac{\beta^2}{2(1+\beta^2)}\boldsymbol{D_i}} + (1+1\beta^2)^{-\frac{k}{2}} \qquad (5)$$

where $\beta = \frac{1}{\sqrt{2}}\left(\frac{n(2p+1)}{4}\right)^{\frac{1}{k+4}}$, $\boldsymbol{D_{ij}} = \left(x_i - x_j\right)^T \boldsymbol{S^{-1}}\left(x_i - x_j\right)$, $\boldsymbol{D_i} = (x_i - \bar{x})^T \boldsymbol{S^{-1}}(x_i - \bar{x})$, k is the number of variables, and $\boldsymbol{S^{-1}}$ is the variance covariance matrix. The testing criterion is that H_0 is rejected if the $p-value < \alpha$, thus the residuals are not normally distributed.

2.7 Outlier Detection

Time series observations are usually affected by unexpected events. The result of the disturbance of the unexpected event causes observations whose values are far from other observations or can be said to be outliers [16]. Outliers are observations that are not consistent in the series. For this reason, there are two treatments for them, namely removing them on the condition that they do not damage the model or retaining them because they are not disturbed. Outliers are removed in the sense that they replace the outliers with new values. If the outliers are due to errors in the data, the outlier value can be substituted using the mean or median of the variable based on the data distribution [17].

2.8 Vector Autoregressive Process with the Outlier Detection

The process of collecting and analyzing data in this study was carried out with the following steps.

a. The first step is to collect monthly data on rainfall and wind speed for the observation point of Pontianak Maritime Meteorological Station.
b. The data stationarity test is carried out in variance and average. If the data is not stationary in variance, data transformation is carried out. If the data is not stationary in average, data differencing and ADF test are required again.
c. If the data tested for stationarity has been stationary in variance and average, then identify a temporary VAR model by looking at the cut-off in the MPACF function and determine the best model based on the smallest AIC value.
d. After obtaining the best model from the smallest AIC value, model parameter estimation and diagnostic tests are carried out to test the White Noise residual assumption and the multivariate normal distribution residual assumption.
e. If the model that has been obtained has met the residual assumptions, then identify outliers in the residual data of the multivariate model.
f. If there are outliers in the multivariate model, the model will be subjected to outlier detection by replacing it with the mean of the initial data so that "New Data" will be obtained.
g. Then the "New Data" without outliers obtained after removing outliers and replaced with the mean of this initial data is modeled again, so that a new model will be obtained.
h. After obtaining the VAR model before and after outlier detection, the next step is to compare the two models based on the AIC value.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

This study uses monthly data on rainfall in millimeters (mm) and wind speed in knots. Table 2 presents descriptive statistics of rainfall and wind speed data.

Table 2. Descriptive Statistics of Rainfall and Wind Speed Data

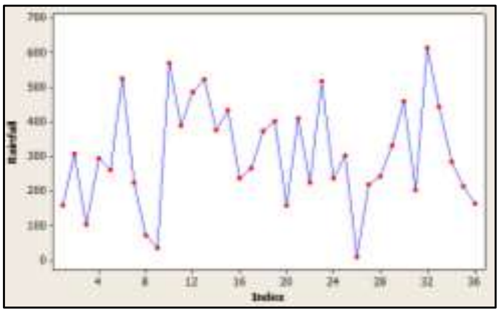
Variable	N	Mean	Median	Mode	Min	Max
Rainfall (mm)	36	308	290	224	8	616
Wind Speed (knots)	36	25	25	25	19	37

Based on Table 2, rainfall data has a right-skewed pattern where the mean value is greater than the median and mode. While the wind speed data has a symmetrical data pattern where the mean, median, and mode values have the same value.

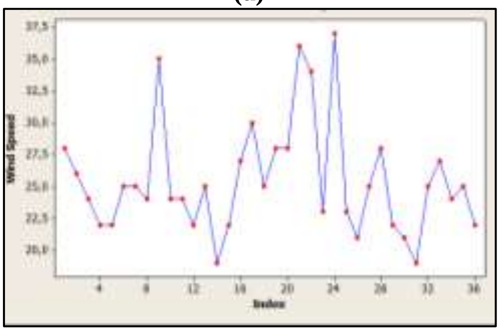
3.2 Vector Autoregressive Model

3.2.1 Stationarity Test

The first step in modeling time series data is to test the stationarity of the data. If the data stationarity requirements cannot be met, it will produce an inaccurate model for prediction. The following is an initial plot of each rainfall and wind speed data:



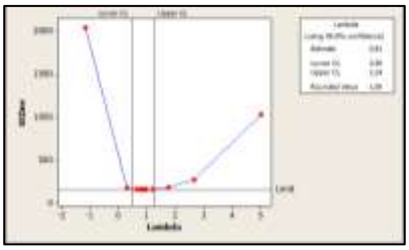
(a)

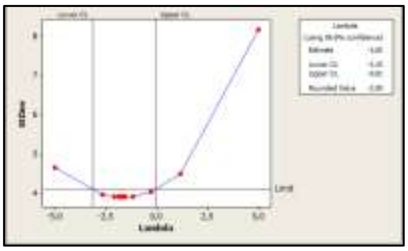


(b)

Figure 1. Time Series Plot (a) Rainfall and (b) Wind Speed

Based on Figure 1 (a) and Figure 1 (b), rainfall and wind speed data are not stationary in a mean or variance. Testing for stationary data in variance is done by looking at the λ value of the Box-Cox plot. The following are the results of the Box-Cox transformation of rainfall and wind speed data.





(a) **(b)**

Figure 2. *Box-Cox* (a) Rainfall and (b) Wind Speed

Based on **Figure 2** (a) the rainfall data is stationary with respect to variance as seen from the value of $\lambda = 1$, while in **Figure 2** (b) the wind speed data is not stationary with respect to variance because the value of $\lambda = 2$, so it is necessary to transform the data by $1/x^2$ so that the wind speed data is stationary on variance. The results of the wind speed data transformation can be seen in **Figure 3**.

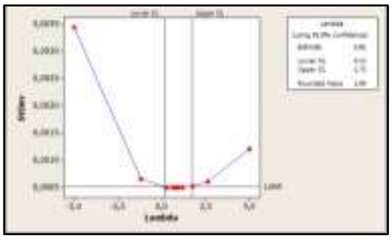


Figure 3. *Box-Cox* Transformation of Wind Speed

Figure 3 is the result of the Box-Cox test of wind speed from the transformed "New Data", resulting in a value of $\lambda = 1$, which means that the wind speed data is stationary concerning variance. Next, the data was tested for stationarity concerning the mean.

Furthermore, the stationarity test of the mean was conducted using the Augmented Dickey-Fuller (ADF) test. The ADF test results for rainfall and wind speed data are not stationary in the mean. Therefore, a data differencing step is required and the ADF test is performed again. **Table 3** presents the ADF test results after differencing the data once.

Table 3. Differencing Data Stationarity Test

Variable	ADF Count	Probability	Information
Rainfall (mm)	-4.16	0.01	Stationary
Wind Speed (knots)	-3.60	0.03	Stationary

Table 3 shows that the probability value of the two variables is smaller than the significance value of 0.05, which means that the two variables have been stationary in mean.

3.2.2 Identification of The VAR Model

Temporary VAR model identification can be done by looking at the cut off on the MPACF function shown in **Table 4**.

Table 4. Stationary MPACF Results

Variablel/Lag	1	2	3	4	5	6
Rainfall	-.	.+	..	..	..	..
Wind Speed	.-	..	..	..	..	..

Based on **Table 4**, the MPACF function cuts off after lag 1 and lag 2, so the temporary VAR models formed are VAR(1) and VAR(2) models. The next step will be the model feasibility test process by estimating the model parameters and determining the best model by looking the smallest AIC value. The following is given the AIC value to determine the feasibility of the model in **Table 5**.

Table 5. AIC Value of Temporary Model

Temporary Model	AIC
VAR (1)	4.94
VAR (2)	5.42

From the results of the model feasibility test analysis, the best model with the smallest AIC value is the VAR(1) model resulting from differencing once with an AIC value of 4.94.

3.2.3 Parameter Estimation of VAR Model

The parameter estimation results are tested for parameter significance to determine whether the parameters have an effect on the model. If the p-value < 0.05 then the parameter has a significant effect on the model. The results of the parameter significance test are presented in **Table 6**.

Table 6. Model Estimation and Significance Test

Variable	Parameter	Parameter Estimation	t - value	Probability	Information
Δ Rainfall	ϕ_{10}	-3.256e+00	-0.103	0.9185	Insignificant
	ϕ_{11}	-5.208e-01	-3.436	0.0017**	Significant
	ϕ_{12}	-1.417e+04	-0.271	0.7882	Insignificant
$\Delta \frac{1}{(wind\ speed)^2}$	ϕ_{20}	2.097e-05	0.206	0.8378	Insignificant
	ϕ_{21}	-1.232e-07	-0.253	0.8022	Insignificant
	ϕ_{22}	-3.668e-01	-2.181	0.0369*	Significant

From Table 6, significant parameters can be seen from the statistical probability value that is less than 0.05. Based on Table 6, the one-time differencing VAR(1) model formed is:

$$\begin{bmatrix}\Delta Y_{1t}\\ \Delta Y_{2t}\end{bmatrix} = \begin{bmatrix}-3.256\\ 0.000021\end{bmatrix} + \begin{bmatrix}-0.5208 & -14170\\ -0.00000012 & -0.3668\end{bmatrix}\begin{bmatrix}Y_{1,t-1}\\ Y_{2,t-1}\end{bmatrix} + \begin{bmatrix}a_{1t}\\ a_{2t}\end{bmatrix}.$$

or with the following equations.

$$\Delta Y_{1t} = -3.256 - 0.5208Y_{1,t-1} - 14170Y_{2,t-1} + a_{1t}$$

$$\Delta Y_{2t} = 0.000021 - 0.00000012Y_{1,t-1} - 0.3668Y_{2,t-1} + a_{2t}$$

3.2.4 White Noise Residual Assumption Test

This residual assumption test aims to determine the correlation between residual vectors from the VAR (1) model of the differencing results once formed. If the p-value < 0.05 then H_0 is rejected or the residuals do not meet the White Noise assumption. The Ljung-Box test results are presented in Table 7.

Table 7. White Noise Residuals

Lag	Qm	Probability Value
1	1.22	0.87
2	6.98	0.54
3	13.16	0.36
4	14.52	0.56
5	16.61	0.68
6	18.10	0.80

Based on Table 7, the probability value for each lag is more than 0.05, so it can be concluded that the residuals qualified the White Noise residual assumption.

3.2.5 Multivariate Normal Residual Assumption Test

The multivariate normal residual assumption test in this study uses the *Henze-Zirkler* test. Table 8 below provides the results of testing the multivariate normal assumption using the *Henze-Zirkler* test.

Table 8. Multivariate Normality test results

Test	Test Statistics	p-value
Henze-Zirkler	0.32	0.74

From Table 8, it can be concluded that the residuals of the VAR(1) model resulting from one-time differencing are multivariate normally distributed.

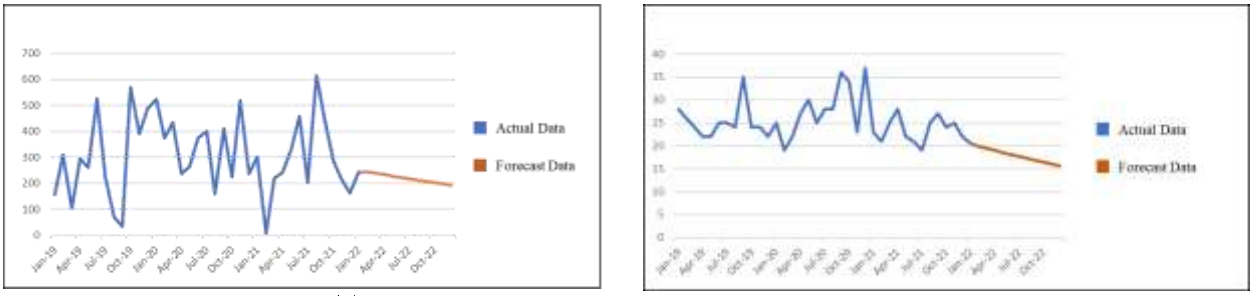
3.2.6 VAR Model Forecasting Before *Outlier* Detection

In the VAR(1) model before outlier detection, the results of forecasting rainfall and wind speed data for January 2022 to December 2022 are presented in Table 9.

Table 9. Forecasting Result Data

Month	Forecasting Results	
	Rainfall	Wind Speed
Jan-22	244	20
Feb-22	244	20
Mar-22	239	19
Apr-22	234	19
May-22	228	19
Jun-22	223	18
Jul-22	217	18
Aug-22	212	17
Sep-22	207	17
Oct-22	202	16
Nov-22	198	16
Dec-22	193	16

The forecast data obtained can be presented in graphs of actual data and forecast data as in Figure 4 (a) and Figure 4 (b).



(a) **(b)**

Figure 4. Graph of Actual and Forecast Data (a) Rainfall and (b) Wind Speed

3.3 Outlier Detection in VAR Model

To determine the presence or absence of *outliers*, *outlier* detection is performed on the multivariate VAR(1) model using the Mahalanobis distance. The results of outlier detection are presented in Figure 4.

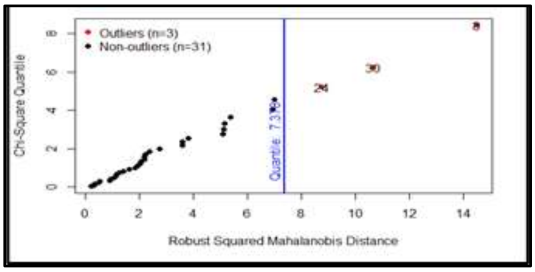


Figure 5. Outlier Points in VAR(1) Model

Figure 5, shows that the number of outliers detected is 3 points with different data distances. The data in the 8th observation has a data distance of 14.504. The data in the 30th observation has a data distance of 10.646 and the data in the 24th observation has a distance of 8.742.

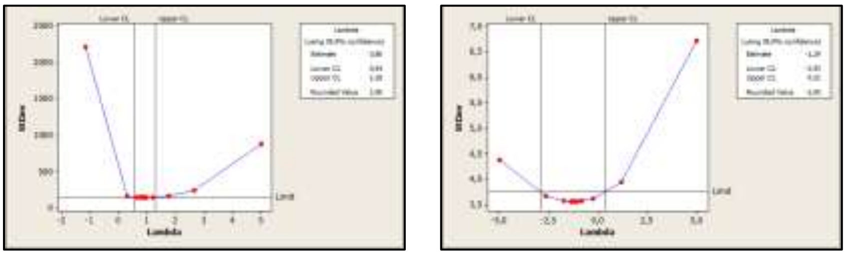
After the outlier point is known, the outlier data is replaced with the mean value of the initial data so that "New Data" will be obtained, then the "New Data" will be modeled again [17].

3.4 Vector Autoregressive Model After Outlier Detection

The data to be re-identified is rainfall and wind speed data that has undergone the process of replacing outlier values. So that the data can be said to be "New Data".

3.4.1 Stationarity Test

The first step to identify the model with "New Data" is to test whether the "New Data" is stationary in variance using Box-Cox transformation as shown in **Figure 6** (a) and **Figure 6** (b).



(a) **(b)**

Figure 6. *Box-Cox* (a) Rainfall and (b) Wind Speed

Based on the Box-Cox transformation results in **Figure 6** (a) the rainfall data has been stationary in variance because $\lambda = 1$, while in **Figure 6** (b) it can be seen that the wind speed data is not stationary in variance because the value of $\lambda \neq 1$. Therefore, the wind speed data needs to be transformed to 1/x to make the wind speed data stationary in variance. The results of the wind speed data transformation can be seen in **Figure 7**.

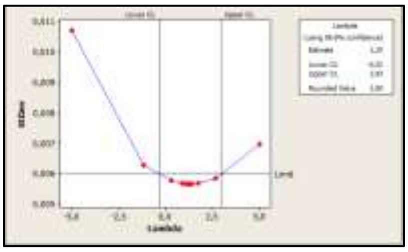


Figure 7. *Box-Cox* Transformation of Wind Speed

Figure 7, is the result of the Box-Cox transformation of wind speed data which produces a value of $\lambda = 1$, which means the data is stationary in variance.

The stationarity test for the mean in "New Data" is not stationary because the probability value is more than the significance value of 0.05 so, it is necessary to differencing the data and ADF test again. The ADF test results after differencing once are summarized in **Table 10**.

Table 10. Stationarity Test of New Data after Differencing

Variable	ADF Count	Probability	Information
Rainfall (mm)	-4.42	0.01	Stationary
Wind Speed (knots)	-3.91	0.02	Stationary

Table 10 explains that both variables are stationary to the mean because they have probability values that are smaller than the significance value of 0.05.

3.4.2 Identification of The VAR Model

The identification of a temporary VAR model based on "New Data" is carried out by looking at the cut-off on the MPACF function.

Table 11. Stationary MPACF Result

Variable/Lag	1	2	3	4	5	6
Rainfall	-.	..	-.	..	..	..
Wind Speed	.-	..	..	..	..	..

Based on **Table 11**, the MPACF function cuts off after lag 1 and lag 3, so the temporary VAR models formed are VAR (1) and VAR (3) models. The following is given the AIC value to determine the feasibility of the model in **Table 12**.

Table 12. Stationary MPACF Result

Temporary Model	AIC Value
VAR (1)	0.25
VAR (3)	1.24

From the results of the model feasibility test analysis, the best model with the smallest AIC value is the VAR(1) model resulting differencing once with an AIC value of 0.25.

3.4.3 Parameter Estimation of VAR Model

Then the parameter estimation of the VAR(1) model can be retrieved in **Table 13**.

Table 13. VAR Model Estimation

Variable	Parameter	Parameter Estimation	t - value	Probability	Information
Δ Rainfall	ϕ_{10}	-3.6427	-0.126	0.900265	Insignificant
	ϕ_{11}	-0.6072	-4.307	0.00015***	Significant
	ϕ_{12}	3141.8477	0.752	0.457846	Insignificant
$\Delta \frac{1}{(windspeed)^2}$	ϕ_{20}	2.594e-04	0.226	0.8226	Insignificant
	ϕ_{21}	-3.731e-06	-0665	0.5110	Insignificant
	ϕ_{22}	-3.767e-01	-2.266	0.0306*	Significant

From **Table 13**, significant parameters can be seen from the probability value that is less than 0.05. Based on **Table 13**, the one-time differencing VAR(1) model formed is:

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \end{bmatrix} = \begin{bmatrix} -3.6427 \\ 0.0002594 \end{bmatrix} + \begin{bmatrix} -0.6072 & 3141.8477 \\ 0.000003731 & -0.3767 \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}.$$

or with the following equations.

$$\Delta Y_{1t} = -3.6427 - 0.6072 Y_{1,t-1} + 3141.8477 Y_{2,t-1} + a_{1t}$$
$$\Delta Y_{2t} = 0.0002594 - 0.000003731 Y_{1,t-1} - 0.3767 Y_{2,t-1} + a_{2t}$$

3.4.4 White Noise Residual Assumption Test

In testing the suitability of the model includes a residual assumption test with Ljung-Box to see if the residuals have met the White Noise assumption. The Ljung-Box test results are presented in **Table 14**.

Table 14. White Noise Residual

Lag	Qm	Probability Value
1	1.75	0.78
2	7.43	0.48
3	14.34	0.27
4	15.45	0.49
5	16.85	0.66
6	17.12	0.84

From **Table 14**, it is known that the probability value for each lag is more than 0.05, so it can be concluded that the residuals qualified the White Noise assumption.

3.4.5 Multivariate Normal Residual Assumption Test

The conclusion of the normal multivariate distribution assumption test in this study was carried out using the *Henze-Zirkler* test. **Table 15** below provides the results of testing the normal multivariate assumption using the *Henze-Zirkler* test.

Table 15. Multivariate Normality Test Results

Test	Test Statistics	*p-value*
Henze-Zirkler	0.323096	0.78

Table 15 is the result of testing multivariate normality using the *Henze-Zirkler* test. Judging from the p-value that is more than 0.05, it can be concluded that the residuals have qualified the multivariate normal assumption.

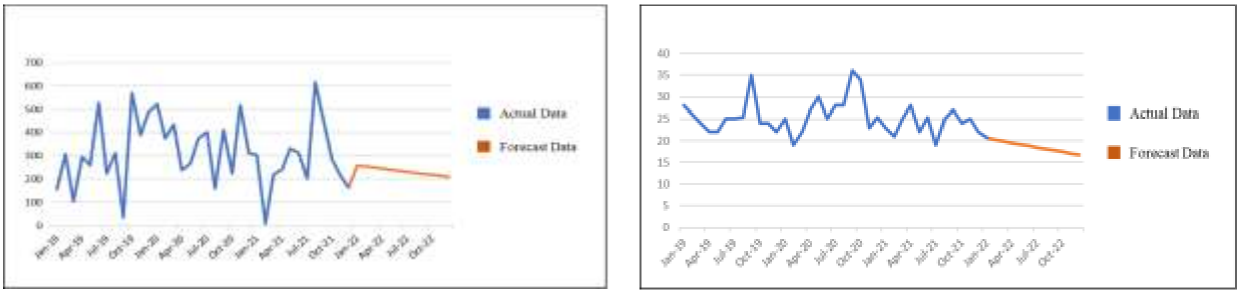
3.4.6 VAR Model Forecasting After Outlier Detection

In the VAR (1) model after outlier detection, the results of forecasting rainfall and wind speed data for January 2022 to December 2022 are presented in **Table 16**.

Table 16. Forecasting Result Data

Month	Forecasting results	
	Rainfall	Wind Speed
Jan-22	257	21
Feb-22	254	20
Mar-22	249	20
Apr-22	245	20
May-22	240	19
Jun-22	236	19
Jul-22	231	18
Aug-22	227	18
Sep-22	223	18
Oct-22	218	17
Nov-22	214	17
Dec-22	210	17

The forecast data obtained can be presented in graphs of actual data and forecast data as in **Figure 8** (a) and **Figure 8** (b).



(a) (b)

Figure 8. Graph of actual and Forecast Data of (a) Rainfall and (b) Wind Speed

3.5 Comparison of VAR Models Before and After Outlier Detection

After identifying the VAR model with outliers and the VAR model without outliers, a comparison between the two models will be made in **Table 17**.

Table 17. Comparison of VAR Models with Outliers and Without Outliers

Model	AIC
There are Outliers	4.94
There are not Outlier	0.25

Based on **Table 17**, after performing outlier detection on the model by replacing the observation data where outliers occur with the mean of the initial data so that "New Data" is obtained, it is known that the smallest AIC value is found in the model that does not have outliers or after outlier detection is 0.25. Thus, it can be concluded that the best Vector Autoregressive model is the model after outlier detection.

4. CONCLUSIONS

From the analysis that has been done in comparing the Vector Autoregressive model before and after outlier detection, the best model before outlier detection is the VAR(1) model which has the smallest AIC value of 4.94. While the best model after outlier detection is the VAR(1) model with the smallest AIC value of 0.25. Judging from the smallest AIC value of the two models, it can be concluded that the VAR(1) model after outlier detection is better than the model before outlier detection.

REFERENCES

[1] D. A. H. Panggabean, F. M. Sihombing, and N. M. Aruan, “Prediksi Tinggi Curah Hujan Dan Kecepatan Angin Berdasarkan Data Cuaca Dengan Penerapan Algoritma Artificial Neural Network (ANN),” *SEMINASTIKA*, vol. 3, no. 1, pp. 1–7, Nov. 2021, doi: 10.47002/seminastika.v3i1.237.

[2] M. P. Ayudhiah, S. Bahri, and N. Fitriyani, “Peramalan Indeks Harga Konsumen Kota Mataram Menggunakan Vector Autoregressive Integrated Moving Average,” *EIGEN MATHEMATICS JOURNAL*, pp. 1–8, Jun. 2020, doi: 10.29303/emj.v3i1.61.

[3] “Saputro dan kutipan trisasongko. - Model Vektor Autoregressive Untuk Peramalan Curah”.

[4] Y. W. A. Nanlohy, Dr. B. S. S. U., M.Si, and S. W. P., M.Si., Ph.D, “Model Fungsi Transfer Multi Input Untuk Peramalan Curah Hujan Di Kota Surabaya,” *VARIANCE: Journal of Statistics and Its Applications*, vol. 1, no. 2, pp. 82–92, Feb. 2020, doi: 10.30598/variancevol1iss2page82-92.

[5] F. Fariz Ichsandi, R. Rahmawati, and Y. Wilandari, “Peramalan Laju Inflasi dan Nilai Tukar Rupiah Terhadap Dolar Amerika Menggunakan Model Vector Autoregressive (VAR),” vol. 3, no. 4, pp. 673–682, 2014, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[6] P. Rialita Hardani and A. Hoyyi, “Peramalan Laju Inflasi, Suku Bunga Indonesia Dan Indeks Harga Saham Gabungan Menggunakan Metode Vector Autoregressive (VAR),” *JURNAL GAUSSIAN*, vol. 6, no. 1, pp. 101–110, 2016, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[7] U. I. Negeri, A. Makassar, and R. Ibnas, “dalam Meramalkan Jumlah Penduduk (Studi Kasus : Kabupaten Gowa) Nurwahyu Agustin.”

[8] F. D. Islami, A. Hoyyi, and D. Ispriyanti, “Pemodelan Fungsi Transfer Dengan Deteksi Outlier Untuk Memprediksi Nilai Inflasi Berdasarkan Bi Rate (Studi Kasus BI Rate dan Inflasi Periode Januari 2006 sampai Juli 2016),” *JURNAL GAUSSIAN*, vol. 6, no. 3, pp. 323–332, 2017, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[9] A. R. Putri, M. Usman, Warsono, Widiarti, and E. Virginia, “Application of Vector Autoregressive with Exogenous Variable: Case Study of Closing Stock Price of PT INDF.Tbk and PT ICBP.Tbk,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jan. 2021. doi: 10.1088/1742-6596/1751/1/012012.

[10] Juanda B and Junaidi, *Ekonometrika Deret Waktu: Teori dan Aplikasi*. Bogor: IPB Press Tahun 2012, 2012.

[11] A. Pertiwi, L. F. Dewi, T. Toharudin, and B. N. Ruchjana, “Penerapan Model Vector Autoregressive Integrated Moving Average (Varima) Untuk Prakiraan Indeks Harga Saham Gabungan Dan Kurs Rupiah Terhadap USD,” *Pattimura Proceeding: Conference of Science and Technology*, pp. 431–442, Apr. 2022, doi: 10.30598/pattimurasci.2021.knmxx.431-442.

[12] W. W. S. Wei, *Time Series Analysis Univariate and Multivariate Methods*, Second. USA: Greg Tobin, 2006.

[13] A. Hoyyi, Tarno, D. A. I Maruddani, and R. Rahmawati, “Vector autoregressive model approach for forecasting outflow cash in Central Java,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, May 2018. doi: 10.1088/1742-6596/1025/1/012105.

[14] Sugito, Mustafid, D. Safitri, D. Ispriyanti, A. R. Hakim, and H. Yasin, “Rainfall and Wave Height Prediction in Semarang City Using Vector Autoregressive Neural Network (VAR-NN) Methods,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Nov. 2019. doi: 10.1088/1742-6596/1320/1/012017.

[15] Henze N and Zirkler B, *A Class of Invariant Consistent Tests for Multivariate Normality*, vol. 19. Communications in Statistics - Theory and Methods, 1990.

[16] L. Budiarti, B. Warsito, M. Jurusan Statistika FSM UNDIP, and S. Pengajar Jurusan Statistika, “Analisis Intervensi Dan Deteksi Outlier Pada Data Wisatawan Domestik (Studi Kasus di Daerah Istimewa Yogyakarta),” 2013. [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian

[17] Soemartini, *OUTLIER (pencilan)*. Bandung: UNPAD, 2007.