**Jurnal Varian**
Vol. 8, No. 2, April 2025, pp 125–138
e-ISSN: 2581-2017

# Forecasting the Amount of Water Discharge Based on the VARIMA Model

**Hesti Meliyana, Mustika Hadijati, Lisa Harsyiah**
Universitas Mataram, Mataram, Indonesia

**Article Info**

**Article history:**

Received : 08-25-2023
Revised : 10-30-2024
Accepted : 11-26-2024

**Keywords:**

MAPE;
VARIMA;
Water Discharge.

**ABSTRACT**

Water is a necessary substance for every living thing. Clean water is the main requirement for ensuring human health and the environment PT. Air Minum Giri Menang (Perseroda). The purpose of this study is to determine the model and then predict the water discharge of PT. Air Minum Giri Menang uses the obtained model, which will be useful for the community and agencies to optimize the management, distribution, and use of clean water. The method used in this study is VARIMA (Vector Autoregressive Integrate Moving Average) which can process data for more than one variable. The data used in this study is water discharge data produced and distributed from January 2018 to December 2021. The results show that the best model obtained is VARIMA(0,1,1) with model accuracy for water discharge data that was produced and distributed based on the MAPE value of 4% and 5%, which states that the forecasting results can be categorized as very good. This means that the VARIMA (0,1,1) model has provided very accurate results in predicting water discharge with very small forecasting errors, thus indicating that the model is very effective. Suggestions for further research are to look for alternative forecasting methods that overcome non-stationarity data other than data transformation.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i2.3278

**Corresponding Author:**

Lisa Harsyiah,
Department of Statistics, Universitas Mataram,
Email: lisa_harsyiah@unram.ac.id



How to Cite:
Meliyana, H., Hadijati, M., & Harsyiah, L. (2025). Forecasting the Amount of Water Discharge Based on the VARIMA Model. *Jurnal Varian*, *8*(2), 125-138.


## A. INTRODUCTION

Clean water is healthy and raw water that can be used as the main material for drinking or cooking and for washing, bathing, or sanitation purposes. The clean water used must be free from disease-causing microorganisms and chemical substances that could affect its quality (Jannah & Itratip, 2016). The treatment and distribution of water to the communities in the city of Mataram and West Lombok Regency are directly handled by PT. Air Minum Giri Menang (Perseroda).

Based on data obtained from the website https://ntb.bps.go.id, the population of Mataram City and West Lombok Regency is unstable due to births and deaths every year. This affects the use of water provided by PT. Air Minum Giri Menang. This data will change over specific periods, indicating that it is a time series data. A method is needed to analyze and forecast two or more variables for several periods ahead, allowing the community to anticipate potential worst-case scenarios.

Forecasting that involves more than one-time series data is called a multiple variable time series. One way to forecast multiple variable time series data is the Vector Autoregressive Integrated Moving Average (VARIMA) model (Pertiwi et al., 2022). The VARIMA model is used to create a model of two objects that have a relationship or influence, and one of the objects can be used to predict the other object. Just like the ARIMA model, the VARIMA model also sometimes requires other explanatory variables to be included in the model as covariates, called the VARIMAX model (Sutthichaimethee, 2017).

Various studies on water flow forecasting or water usage by the PDAM (Public Drinking Water Company) have been extensively utilized as prior research materials. Previous research with similarities in theory, subjects, or research methods have been used as references in this study. Past studies discuss water flow forecasting or water usage by PDAM and/or utilize similar methods. Research by Amadea & Winarno (2016), titled "Prediction of Monthly Drinking Water Production Using Backpropagation Artificial Neural Network Method (Case Study: PDAM Tirta Moedal Semarang)," discusses the forecasting of drinking water production to balance the community's water needs with the water production by relevant institutions using the artificial neural network method. The variables used for prediction are the recapitulation of drinking water usage every month from 2009 to 2014 without distinguishing user groups. Applying the backpropagation artificial neural network method for prediction requires a considerable amount of time due to the need for numerous experiments in determining the number of hidden layers, the number of neurons in hidden layers, setting the learning rate and implementing learning techniques in the planned network. The resulting structure combination cannot yet be considered the best outcome of maximum performance from the neural network. Based on the issues at PDAM Tirta Moedal Semarang, applying the Backpropagation Artificial Neural Network method significantly assists in calculating the prediction of monthly drinking water production, as before its implementation, PDAM Tirta Moedal Semarang used only estimation techniques based on the previous month. This led to excessive production compared to the actual demand (Amadea & Winarno, 2016).

Research by Asfihani & Irhamah (2017), titled "Forecasting Water Usage Volume at PDAM Surabaya using the Time Series Method," discusses the forecasting of PDAM water usage using the ARIMA and transfer function methods. The researcher divides the data into two categories for two forecasting methods. The ARIMA method is used for the lower-middle-class households, while the transfer function method is employed for the middle-class and upper-middle-class households. The variables used in this study are Yt to define the water usage volume variable (output series) and Xt to define the number of customers variable (input series). The best forecasting model for lower-middle-class households is ARIMA (1,1,[12]), while for the middle-class households, the transfer function with b=5, r=0, s=0 and ARMA noise series ([1,4,12],[1,23]) is suitable. The best model for upper-middle-class households is the transfer function with b=5, r=0, s=0 and ARIMA noise series ([12],1) (Asfihani & Irhamah, 2017).

Research by Ikbal (2019), titled "Comparison of Vector Autoregressive Integrated Moving Average Model with Generalized Space-Time Autoregressive Integrated Moving Average for Forecasting Clean Water Usage Volume (Case Study: Surabaya City, Gresik Regency, and Sidoarjo Regency)," discusses the comparison of the two methods in terms of time series data forecasting. The GSTARIMA model has fewer parameter estimates than the VARIMA model and considers spatial and temporal effects. The variables used in this study are clean water usage in Surabaya City, Gresik Regency, and Sidoarjo Regency. The volume of clean water usage in these regions increases every year. The selected VARIMA model is based on the smallest AICC value, which is VARIMA (2,1,0). The time order of the GSTARIMA model is determined based on the VARIMA model order, while the spatial order is limited to one. Therefore, the GSTARIMA model is GSTARI (2,1)1, applying three location weights: uniform location weight, inverse distance weight, and normalized cross-correlation Ikbal (2019).

The research by Jusmawati et al. (2020), titled "Application of Vector Autoregressive Integrated Moving Average Model in Forecasting Inflation Rate and Interest Rate in Indonesia," presents that in their study, they successfully obtained predictive results for the inflation rate and interest rate for the next 12 months. However, in practice, there were challenges regarding the non-fulfillment of the normality test requirement for residuals. Although this didn't affect the forecasting results, it indicated the presence of outlier data in the study (Jusmawati et al., 2020).

Previous studies have discussed the same methods and case studies as those that will be investigated. Still, there are several differences and shortcomings in the previous research that can be used as references to enhance the quality of this study. The research conducted by Asfihani & Irhamah (2017) addressed the time series method for forecasting water usage volume using only a single-variable dataset. This study will compare this method with the VARIMA method, which can forecast time series data involving multiple variables. The VARIMA method is an advancement from the previous method (Nugroho, 2022).

Additionally, the research by Putri & Oktaviarina (2024) focused on the VARIMA method, suitable for forecasting multivariate time series data. Then, the study conducted by Amadea & Winarno (2016) employed an artificial neural network method, which cannot be considered the optimal method for forecasting when compared to the VARIMA method. VARIMA is suitable for this study because it can process multivariate data more effectively. Lastly, the research by Ikbal (2019) compared the VARIMA method with the

GSTARIMA method. The results showed that the VARIMA model outperformed the GSTARIMA model. Consequently, this study aims to develop the processing and forecasting of water flow data produced and distributed using a multivariate method, specifically VARIMA. The gap analysis of this research is previous research with the same case study but using artificial neural network methods, so the novelty of this research is using VARIMA to predict the amount of water produced and distributed. Although VARIMA has been tested in several multivariate time series contexts, specifics for forecasting water discharge in the context of clean water distribution by water supply companies are still rare.

The research titled ”Forecasting the Amount of Produced and Distributed Water by PT. Air Minum Giri Menang (Perseroda) Based on the VARIMA Model” is conducted to determine a forecasting model using the VARIMA method and present its forecasting results. The contribution of this research is to obtain a pattern of forecasting results for water discharge in the future so that it can become a reference for the community and PT leadership. Giri Menang Drinking Water (Perseroda) so that clean water on earth can continue to be maintained to meet primary human needs.

## B. RESEARCH METHOD

The tools used to assist the data processing process in this research are SAS, Microsoft Excel, Minitab 16, and RStudio. The data used in this study are secondary, namely the water discharge data of PT. Air Minum Giri Menang (Perseroda), with its factors being the produced and distributed water discharge over four years from January 2018 to December 2021, obtained from the Planning and Development Division of PT. Air Minum Giri Menang (Perseroda).

### 1. Stationarity

Stationarity means that there are no significant changes in the data, data fluctuations are around a mean value, and the constant variance remains unchanged over time (Pratiwi & Herlina, 2023). Data that does not meet the stationary condition for variance can be addressed by using the Box-Cox transformation like Equation (1) (Putri & Oktaviarina, 2024).

$$T\left(Y_t\right) = \frac{Y_t^{(\lambda)} - 1}{\lambda} \tag{1}$$

with $Y_t$ is the value of the variable at time-t and $\lambda$ are transformation parameters.

The Box-Cox transformation is performed to address the nonstationarity of data for variance, depending on the lambda value it holds, as shown in the following Table 1.

Table 1. Box-Cox Transformation

| $\lambda$ | Transformation |
|---|---|
| -1 | $\frac{1}{Y_t}$ |
| -0.5 | $\frac{1}{\sqrt{Y_t}}$ |
| 0 | $\ln\left(Y_t\right)$ |
| 0.5 | $\sqrt{Y_t}$ |
| 1 | $Y_t$ (without transformation) |

Source: (Wilujeng, 2018)

Another method used to address nonstationarity with respect to the mean is by differencing the time series data. In general, if differencing for a period-d is used to achieve stationarity, it can be expressed in Equation (2).

$$Y_t^d = \left(1 - B\right)^d Y_t \tag{2}$$

with, d is differencing ordo and B is backshift operator.

### 2. Matrix Autocorrelation Function (MACF)

If there is a time series vector with n observations, namely $Y_1,\ Y_2, \ldots,\ Y_n$, then the equation for the sample correlation matrix is in Equation (3).

$$\hat{\rho}(k) = [\hat{\rho}_{ij}(k)] \tag{3}$$

Where $\hat{\rho}_{ij}(k)$ represents the sample cross-correlation for the $i$-th and $j$-th data components as expressed in Equation (4).

$$\hat{\rho}_{ij}(k) = \frac{\sum_{t=1}^{n-k}(Y_{i,t} - \bar{Y}_i)(Y_{j,t+k} - \bar{Y}_j)}{\left[\sum_{t=1}^{n}(Y_{i,t} - \bar{Y}_i)^2 \sum_{t=1}^{n}(Y_{j,t} - \bar{Y}_j)^2\right]^{\frac{1}{2}}} \tag{4}$$

The symbols denoted by (+), (-), and (.) in the sample correlation matrix at $(i,\ j)$ can be interpreted in Table 2.

Table 2. Box-Cox Transformation

| Symbol | Correlation | Value $\hat{\rho}_{ij}(k)$ |
|---|---|---|
| (+) | Positive | $>$ 2SE |
| (-) | Negative | $<$ 2SE |
| (.) | No correlation | $\pm$ 2SE |

### 3. Matrix Partial Autocorrelation Function (MPACF)

The equation for MPACF is in Equation (5) (Putri & Oktaviarina, 2024).

$$\rho(k) = \begin{cases} \Gamma'(1)[\Gamma(0)]^{-1} & , k = 1 \\ \dfrac{\left(\Gamma'(k) - c'(k)[\mathbf{A}(k)]^{-1}b(k)\right)}{\left(\Gamma(0) - b'(k)[\mathbf{A}(k)]^{-1}b(k)\right)}, k > 1 & \end{cases} \tag{5}$$

for k $\geq$ 2, the values of $\mathbf{A}(k)$, $\mathbf{b}(k)$, and $c(k)$ are in Equations (6), (7), (8):

$$\mathbf{A}(k) = \begin{bmatrix} \Gamma(0) & \Gamma'(1) & \cdots & \Gamma'(k-2) \\ \Gamma(1) & \Gamma(0) & & \Gamma'(k-3) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(k-2) & \Gamma(k-3) & \cdots & \Gamma(0) \end{bmatrix} \tag{6}$$

$$\mathbf{b}(k) = \begin{bmatrix} \Gamma'(k-1) \\ \Gamma'(k-2) \\ \vdots \\ \Gamma'(1) \end{bmatrix} \tag{7}$$

$$\mathbf{c}(k) = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(k-1) \end{bmatrix} \tag{8}$$

with,

$$\hat{\Gamma} = \frac{1}{n}\sum_{t=1}^{n-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y}) \tag{9}$$

If $0 \leq k \leq (n-1)$
$\hat{\Gamma} = [\hat{\gamma}_{i,j}(k)]$

where,

$$\hat{\gamma}_{i,j}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(Y_{i,t} - \bar{Y}_i)(Y_{j,t+k} - Y_j) \tag{10}$$

with, $\hat{\gamma}_{i,j}(k)$ are coefficient autocovariance at lag- $k$, $n$ is the number of observations, $k$ is the time lag, $Y_{i,t}$ is value of the $i$-th series variable at time-$t$, $i = 1, 2, \ldots, n$, and $Y_{j,t+k}$ is value of the i-th series variable at time-$t + k$, $j = 1, 2, \ldots, n$. Then $\bar{Y}_i$ and $\bar{Y}_j$ Represents the sample average of corresponding series components. Identifying data based on the value of MPACF is also denoted in the form of (+), (-), and (.) like MACF.

### 4. Vector Autoregressive Integrated Moving Average (VARIMA) Model

This is a time series method used to determine the relationships among several time series variables at time $t$ with $p$ periods in the past. Thus, it can be understood that a variable is influenced by itself and other variables at a specific lag. In general, the form of the VARIMA model is in Equation (11) (Hermayani et al., 2014).

$$\phi_P(B)\, D(B)\boldsymbol{Y_t} = \theta_q(B)\varepsilon_t \tag{11}$$

with

$$\begin{aligned}
\phi_p(B) &= \phi_0 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p \\
\theta_q(B) &= \theta_0 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \\
\phi_0 &= \theta_0 = 1
\end{aligned}$$

where, $\boldsymbol{Y_t}$ is an observation vector with $\boldsymbol{Y_t} = [Y_{1,t}, Y_{2,t}, \ldots, Y_{n,t}]$ of size $n\ x\ 1$, $\varepsilon_t$ is error value at time $-t$, $p$ is AR order, $d$ is differencing order, $q$ is MA order, $B$ is backshift operator, $D(B)$ is a differencing operator, $\phi_p(B)$ is parameter matrix of an autoregressive vector of order- $p$ with size $n \times n$, and $\theta_p(B)$ is parameter matrix of moving average vector of order- $q$ with size $n \times n$.

### 5. Residual Testing

The white noise test is the first residual test carried out in time series analysis. The test statistic for the multivariate used is the Ljung Box-Pierce test statistic with Equation (12) (Panjaitan et al., 2018).

$$\mathcal{Q}_k(m) = N^2 \sum_{k=1}^{m} \frac{1}{N-k} tr\left(\boldsymbol{\hat{T}_1 \hat{T}_0^{-1} \hat{T}_1 \hat{T}_0^{-1}}\right) \tag{12}$$

where $N$ is the amount of data, $m$ is the number of lags tested, $\hat{T}_l$ is the autocorrelation estimate for the period $m$, and $tr$ is the sum of the main diagonals.

The hypothesis used for testing residual white noise is as follows:

$H_0$ : $\rho_1 = \rho_2 = \rho_3 = \ldots = \rho_k = 0$ (Residual meets white noise requirements)
$H_1$ : There is at least one $\rho_i \neq 0$ with $i = 1, 2, \ldots k$ (Residuals do not meet the white noise requirements)

The decision-making criteria for removing the white noise assumption is to accept $H_0$ if $\mathcal{Q}_k(m) < \chi^2 orp\ value > \alpha$.

Next, residual normality testing was carried out. The Kolmogorov-Smirnov test is a popular normality test based on the $D$ value defined by Equation (13).

$$D = SUP_x\, \lfloor |F(y) - F_0(y)| \rfloor \tag{13}$$

where $F(y)$ is the cumulative probability function of sample data and $F_0(y)$ is the normal distribution cumulative probability function.

The hypothesis for testing the normal assumption of multivariate residuals is as follows:

$H_0$ : $F(y) = F_0(y)$ for all values of $y$ (Residuals are normally distributed)
$H_1$ : $F(y) \neq F_0(y)$ for at least a value of$y$(Residuals are not normally distributed)

The decision-making criteria for multivariate normality testing is to reject or accept $H_0$ if $D > D_{(1-\alpha,n)}$.

### 6. Akaike's Information Criterion (AIC)

The Akaike's Information Criterion (AIC) criteria can be formulated in Equation (14).

$$AIC_{(p,\ q)} = \ln|\mathbf{\Sigma}| + \frac{2k^2(p+q)}{N} \tag{14}$$

where, $\ln$ is the natural logarithm, $N$ is a number of observations, $k$ is the number of variables in the model, $p$ is an order of AR, $q$ in the order of MA, and $\Sigma$ is the residual covariance matrix.

**7. Mean Absolute Percentage Error (MAPE)**

An evaluation is performed to measure the model estimation error using the Mean Absolute Percentage Error (MAPE) criteria. The MAPE equation used is in Equation (15) (Afriliani et al., 2022).

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - F_t}{Y_t}\right| \times 100\% \tag{15}$$

with $Y_t$ is actual data at period $t$, $F_t$ is forecast at period $t$, and $n$ is number of data.

A smaller MAPE value can identify a better forecasting model, indicating better forecasting accuracy. In general, the MAPE criteria can be observed in the following Table 3.

Table 3. MAPE Criteria

| MAPE | Description |
|---|---|
| $<10\%$ | Very Good Forecasting Accuracy |
| 10% - 20% | Good Forecasting Accuracy |
| 20% - 50% | Adequate Forecasting Accuracy |
| $>50\%$ | Poor Forecasting Accuracy |

Figure 1 illustrates the steps in the research procedure.

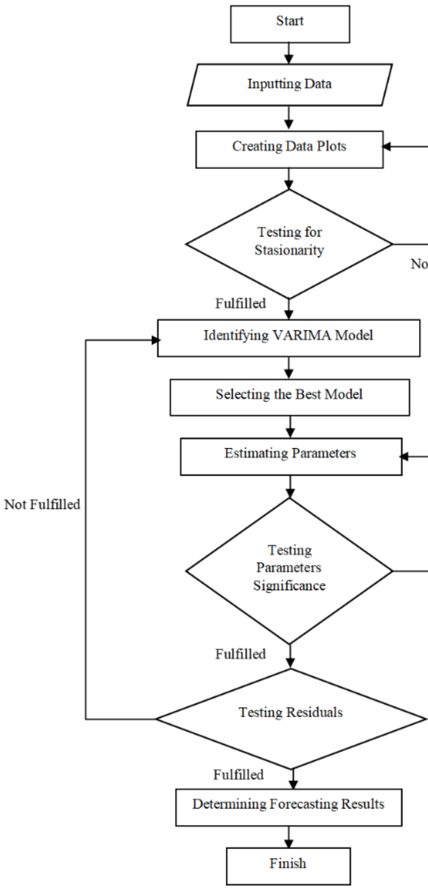


Figure 1. Research Flowchart


DOI: https://doi.org/10.30812/varian.v8i2.3278

## C. RESULT AND DISCUSSION

### 1. Stationary

The initial step in forecasting using the VARIMA method is to test the stationarity of each data. Stationarity is a fundamental requirement that must be fulfilled in time series data modeling. The resulting model will yield incorrect conclusions if this stationarity assumption cannot be met. Time series plots of each data are as follows:

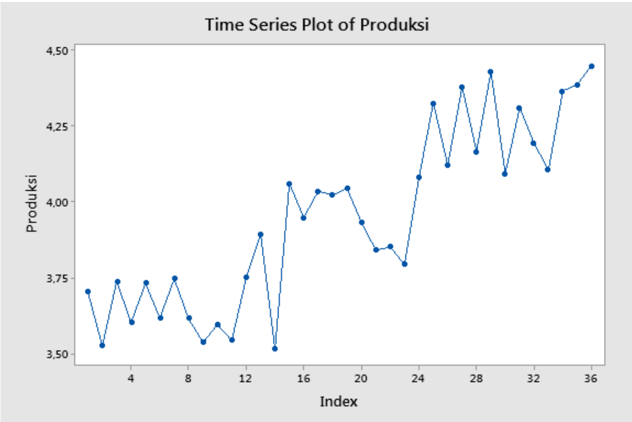


Figure 2. Time Series Plot of Produced Water Flow Data

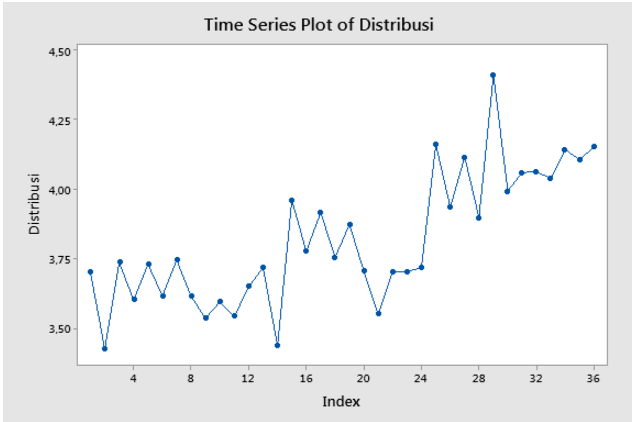


Figure 3. Time Series Plot of Distributed Water Flow Data

Based on Figures 2 and 3, it can be seen that the pattern of produced water flow data and distributed water flow data has an upward and downward trend. A process is said to be stationary if there is no change in trend in either the mean or variance. Stationarity can be seen by paying attention to the time series plot results in the Minitab output. One of the characteristics of stationarity is marked by plot results whose graphs tend to be parallel. For the data to be stationary for the average, differencing is carried out, while transformation needs to be carried out for it to be stationary for the variance.

Furthermore, the results of the variance stationarity test obtained a $\lambda$ value for the production water discharge data of 0.67 and the distributed water discharge data of -0.52. Each variable $\lambda$, so it is not stationary regarding the variance due to outliers in the data. The following is an ACF and time series plot of transformed and differentiated data.

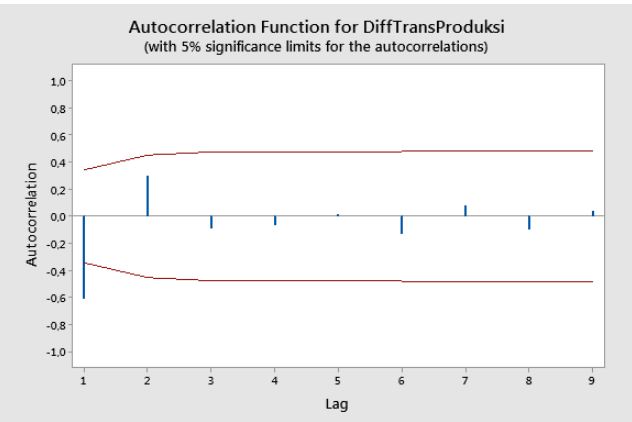


Figure 4. ACF Plot of Stationary Produced Water Flow Data

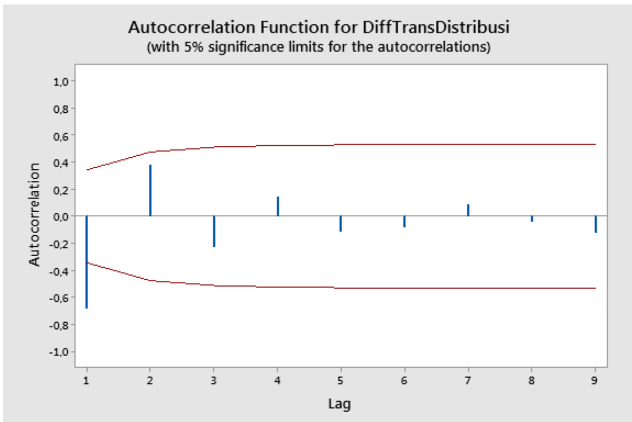


Figure 5. ACF Plot of Stationary Distributed Water Flow Data

After the first difference, it can be seen in Figures 4 and 5 that the data on produced water discharge and distributed water discharge have reached data stationarity for the average. For greater clarity, time series plots are provided for the following data:


DOI: https://doi.org/10.30812/varian.v8i2.3278

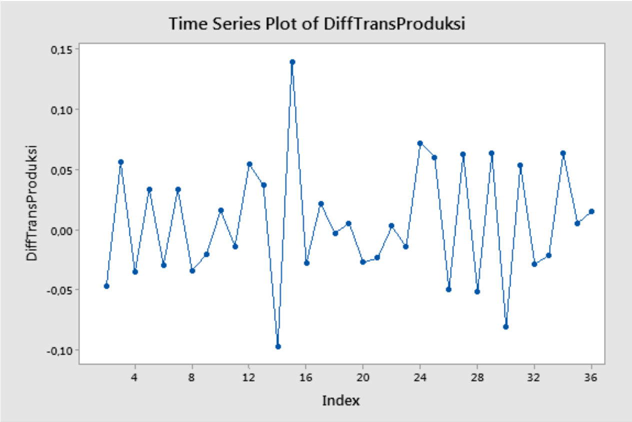


Figure 6. Time Series Plot of Stationary Produced Water Flow Data

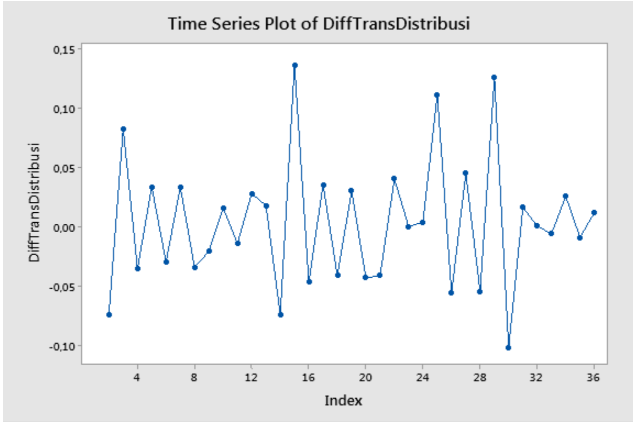


Figure 7. Time Series Plot of Stationary Distributed Water Flow Data

Based on Figures 6 and 7, it is clear that the autocorrelation value at each lag is not significantly different from zero. This can strengthen the assumption that the time series data of water flow produced after the first difference has fulfilled the data requirements that are stationary with respect to the average but not yet stationary regarding the variance because there are still outliers in the plot.

## 2. Identifying the VARIMA Model

Table 4. MACF Data After Stationarity

| Lag/Variable | Produced | Distributed |
|---|---|---|
| 0 | ++ | ++ |
| 1 | – | – |
| 2 | .. | ++ |
| 3 | .. | .. |
| 4 | .. | .. |
| 5 | .. | .. |
| 6 | .. | .. |
| 7 | .. | .. |
| 8 | .. | .. |
| 9 | .. | .. |
| 10 | .+ | .+ |
| 11 | – | – |
| 12 | ++ | ++ |

Table 5. MPACF Data and Differencing

| Lag/Variable | Produced | Distributed |
|---|---|---|
| 1 | .. | .- |
| 2 | .. | .. |
| 3 | .. | .. |
| 4 | .. | .. |
| 5 | .. | .. |
| 6 | .. | .. |
| 7 | .. | .. |
| 8 | .. | .. |
| 9 | .. | .. |
| 10 | .. | .. |
| 11 | .. | .. |
| 12 | .. | .. |

Based on Tables 4 and Table 5, after performing one differencing (d=2), the MACF plot cuts off after lag two, indicating a possible order of MA as MA(2). In contrast, the MPACF plot cuts off after the first lag, suggesting a possible order of AR as AR(1). The simultaneous appearance of dots (.) signifies the absence of correlation between each variable and the variable indicated at the lag. **The findings of this research are** that a temporary combination of the VARIMA(1,1,0) and VARIMA(0,1,1) models was formed.

## 3. Parameter Estimation

Table 6. Estimated Parameters VARIMA(0,1,1)

| Variable | Parameter | Estimated Parameters | Standard Error | T-test | Prob. Value |
|---|---|---|---|---|---|
| Data of Produced Water Debit | $\theta_{01}$ | 0.004532 | 1.293E-07 | 35040 | <2e-16 |
| | $\theta_{11}$ | -1.346 | 0.000019 | 70838 | <2e-16 |
| | $\theta_{12}$ | 0.1197 | 0.000002751 | 43535 | <2e-16 |


DOI: https://doi.org/10.30812/varian.v8i2.3278

| Variable | Parameter | Estimated Parameters | Standard Error | T-test | Prob. Value |
|---|---|---|---|---|---|
| Data of Distributed Water Debit | $\theta_{02}$ | 0.004763 | 1.442E-07 | 33035 | <2e-16 |
| | $\theta_{21}$ | 0.3854 | 0.000006047 | 63732 | <2e-16 |
| | $\theta_{22}$ | -0.9402 | 0.00001418 | 66314 | <2e-16 |

Based on Table 6, the estimated parameters of the VARIMA(0,1,1) model can be observed, indicating that the model has six parameters. Looking at the probability values of each parameter, they are all less than the significance level of $\alpha$=5%, which means the parameters have a significant influence on the model. The following is the VARIMA(0,1,1) model obtained for the produced and distributed water debit:

$$\begin{bmatrix} Y_{1(t)} \\ Y_{2(t)} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} - \begin{bmatrix} -1.346 & 0.1197 \\ 0.3854 & -0.9402 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} \tag{16}$$

This model interprets that the water debit variable produced at time t is influenced by the water debit variable produced at time t-1 and by the water debit variable distributed at time t-1. Similarly, the water debit variable distributed at time t is influenced by the water debit variable produced at time t-1 and by the water debit variable distributed at time t-1.

Table 7. Estimated Parameters VARIMA(1,1,0)

| Variable | Parameter | Estimated Parameters | Standard Error | T-test | Prob. Value |
|---|---|---|---|---|---|
| Data of Produced Water Debit | $\phi_{01}$ | 0.008957 | 0.0006262 | 1.430 | 0.152648 |
| | $\phi_{11}$ | -0.149324 | 0.261568 | -0.571 | 0.568081 |
| | $\phi_{12}$ | 0.204815 | 0.272228 | 0.752 | 0.451831 |
| Data of Distributed Water Debit | $\phi_{02}$ | 0.007008 | 0.006517 | 1.075 | 0.282241 |
| | $\phi_{21}$ | -0.471037 | 0.236059 | -1.995 | 0.045997 |
| | $\phi_{22}$ | -0.840619 | 0.245679 | -3.422 | 0.000623 |

Based on Table 7, the estimated parameters of the VARIMA(1,1,0) model can be observed, indicating that the model has six parameters. Looking at the probability values of the parameter $\varphi_{21}$ and $\varphi_{22}$ which are less than the significance level of $\alpha$=5%, which means the parameters have a significant influence on the model. The following is the VARIMA(1,1,0) model obtained for the produced and distributed water debit:

$$\begin{bmatrix} Y_{1(t)} \\ Y_{2(t)} \end{bmatrix} = \begin{bmatrix} -0.149 & 0.205 \\ -0.471 & -0.841 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{17}$$

This model interprets that the water debit variable produced at time t is influenced by the water debit variable produced at time t-1 and by the water debit variable distributed at time t-1. Similarly, the water debit variable distributed at time t is influenced by the water debit variable produced at time t-1 and by the water debit variable distributed at time t-1.

## 4. Residual Assumption Test

a. Residual White Noise

$H_0$ : Residuals satisfy White's Noise conditions

$H_1$ : Residuals do not satisfy White's Noise conditions

Table 8. Ljung-Box Test VARIMA(0,1,1)

| $m$ | $Q_k(m)$ | $df$ | $p-value$ |
|---|---|---|---|
| 12 | 59.61 | 48.00 | 0.12 |
| 24 | 91,92 | 96.00 | 0.60 |

Based on Table 8, it can be seen that the value of $Q_k(m) < \chi^2_{\alpha,df}$ and the $p-value > \alpha\ (0.05)$ then $H_0$ is accepted, which means the residual from the VARIMA(0,1,1) model meets the white noise assumption.

Table 9. Ljung-Box Test VARIMA(1,1,0)

| $m$ | $Q_k(m)$ | $df$ | $p-value$ |
|---|---|---|---|
| 12 | 11.72 | 10.00 | 0,30 |
| 24 | 14.36 | 22.00 | 0.89 |

Based on Table 9, it can be seen that the value of $Q_k(m) < \chi^2_{\alpha,df}$ and the $p - value > \alpha\ (0.05)$ then $H_0$ is accepted, which means the residual from the VARIMA(1,1,0) model meets the white noise assumption.

b. Residual Normality

$H_0$ : Residuals are normally distributed

$H_1$ : Residuals are not normally distributed

Table 10. Residual Normality Test

| Model | Kolmogorov-Smirnov (KS) | $p - value$ |
|---|---|---|
| VARIMA(0,1,1) | 0.107 | 0.051 |
| VARIMA(1,1,0) | 0.433 | 0.000 |

Based on Table 10, it is known that the residual data value from the VARIMA(0,1,1) model obtained $p{-}value\ (0.051) > \alpha\ (0.05)$, which means $H_0$ is accepted so that the conclusion is that the residuals are normally distributed. Meanwhile, the residual data value from the VARIMA(1,1,0) model obtained $p - value\ (0.000) < \alpha\ (0.05)$ which means $H_0$ is rejected, so the conclusion is that the residuals are not normally distributed.

**5. Selection of the Best Model**

Determining the best model can be seen through a comparison of the AIC values presented in Table 11 below.

Table 11. Comparison of AIC Values

| Model | AIC Values |
|---|---|
| VARIMA(0,1,1) | -14.62962 |
| VARIMA(1,1,0) | -13.81726 |

Based on Table 11, the AIC value of the VARIMA(0,1,1) model is -14.62962, and that of the VARIMA(1,1,0) model is -13.81726. Thus, the best model is VARIMA(0,1,1) because it obtains the smallest AIC value. Then, the Mean Percentage Error (MAPE) calculation is carried out to measure the level of accuracy of the forecast results of the produced and distributed water discharge data.

**6. Forecasting**

Based on the VARIMA(0,1,1) model, the next step is forecasting the water debit data produced and distributed by PT. Air Minum Giri Menang for the next twelve months (January 2021 to December 2021). The results of the water debit data forecasting for production and distribution can be seen in Table 12.

Table 12. Forecast Results of Produced Water Debit Data using VARIMA(0,1,1) Model

| Month- | Produced Water | | Distributed Water | |
|---|---|---|---|---|
| | Actual Data | Predicted Data | Actual Data | Predicted Data |
| 1 | 4.370973 | 4.426926 | 4.098912 | 4.151969 |
| 2 | 3.95551 | 4.374385 | 3.708789 | 4.101031 |
| 3 | 4.369403 | 3.958922 | 4.320682 | 3.710908 |
| 4 | 4.285596 | 4.372815 | 4.103468 | 4.322801 |
| 5 | 4.350571 | 4.289008 | 3.989398 | 4.105587 |
| 6 | 4.367596 | 4.353983 | 3.857638 | 3.991517 |
| 7 | 4.607647 | 4.371008 | 3.941496 | 3.859757 |
| 8 | 4.473838 | 4.611059 | 3.894518 | 3.943615 |
| 9 | 4.258493 | 4.47725 | 3.754976 | 3.896637 |
| 10 | 4.428551 | 4.261905 | 3.858568 | 3.757095 |
| 11 | 4.20911 | 4.431963 | 3.68932 | 3.860687 |
| 12 | 4.245478 | 4.212522 | 3.859933 | 3.691439 |

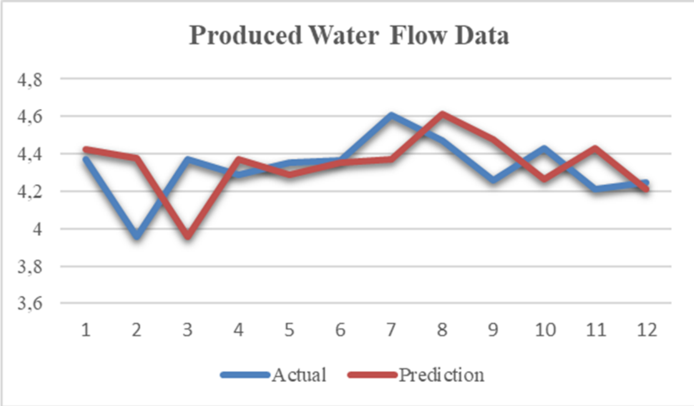


Figure 8. Actual and Predicted Plot of Produced Water Flow

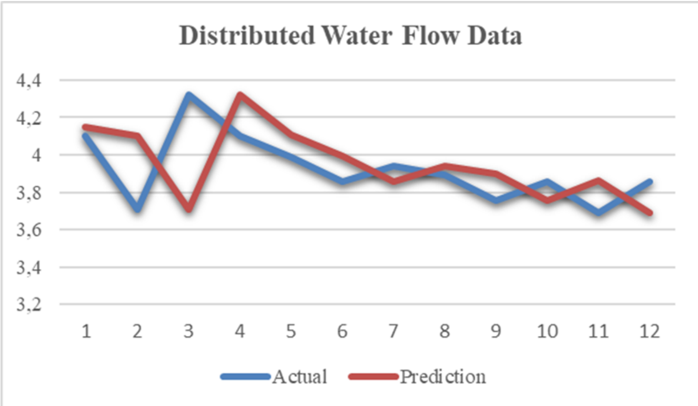


Figure 9. Actual and Predicted Plot of Distributed Water Flow

Based on Table 10, it can be observed that the water debit forecasting results based on the VARIMA(0,1,1) model in the year 2021 are not significantly different from the actual data, which is further illustrated by the graph shown in Figures 8 and 9 above. Furthermore, the Mean Absolute Percentage Error (MAPE) calculation is conducted to measure the water debit data forecasting accuracy for production and distribution. If the MAPE value is relatively small, it can be considered that the forecasting results are good.

Table 13. Model Accuracy Testing.

| Produced Water | Distributed Water |
|---|---|
| 4% | 5% |

The calculated MAPE for the produced and distributed water debit data is 4% and 5%, indicating that the model's accuracy is very good as it has MAPE values $< 10\%$. Compared with research conducted by Asfihani & Irhamah (2017) the novelty of this research can predict time series data involving many variables. So, the VARIMA method can be used in the data forecasting process for a certain period with a very good level of forecasting accuracy (Putri & Oktaviarina, 2024; Suharsono & Anggraeni, 2014). Forecasting results can have important implications for water resources management. Accurate predictions regarding future water discharge enable PT. Air Minum Giri Menang (PERSERODA) needs to plan water distribution more effectively, as maintaining water availability for community needs is important.

## D. CONCLUSION AND SUGGESTION

Based on the research that has been carried out, it can be concluded that the lagged error variable of water discharge produced and distributed at time t-1 significantly affects the amount of water discharge produced and distributed by PT. Air Minum Giri Menang (PERSERODA). The novelty of this research compared to previous ones lies in the progress of the method used; the VARIMA(0,1,1) model obtained shows forecasting results with a very good level of accuracy for the production and distribution of water discharge data at PT. Air Minum Giri Menang (PERSERODA) with values of 4% and 5%, respectively. The results of the MAPE calculation are below 10%, which indicates very good forecasting capabilities. Accurate predictions regarding future water discharge enable PT. Air Minum Giri Menang (PERSERODA) needs to plan water distribution more effectively, as it is important to maintain water availability for community needs. Based on the conclusions obtained, further research is recommended to find an alternative forecasting method that overcomes non-stationary data other than data transformation.

## ACKNOWLEDGEMENT

We would like to express our gratitude to Universitas Mataram and all parties involved in this research.

## DECLARATIONS

AUTHOR CONTIBUTION

All authors contributed to this manuscript, from exploring ideas to writing this article.

FUNDING STATEMENT

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

COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

## REFERENCES

Afriliani, K., Martha, S., & Imro'ah, N. (2022). Model Vector Autoregresive Integrated Moving Average dalam Meramalkan Indeks Harga Konsumen Kota Pontianak. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, *11*(4), 641–648. https://doi.org/10.26418/bbimst.v11i4.57332

Amadea, M., & Winarno, A. (2016). Prediksi Produksi Debit Air Minum Per Bulan Dengan Metode Jaringan Syaraf Tiruan Back-propagation (Studi Kasus: PDAM Tirta Moedal Semarang). *JOINS (Journal of Information System)*, *1*(01), 18–26. https://doi.org/10.33633/joins.v1i01.1169

Asfihani, M. A., & Irhamah, I. (2017). Peramalan Volume Pemakaian Air Di PDAM Kota Surabaya dengan Menggunakan Metode Time Series. *Jurnal Sains dan Seni ITS*, *6*(1), 150–156. https://doi.org/10.12962/j23373520.v6i1.22978

Hermayani, H., Nohe, D. A., & Fathurahman, M. (2014). Mengatasi Heteroskedastisitas pada Model ARIMA dengan Menggunakan ARCH-GARCH (Studi Kasus: Indeks Harga Konsumen Provinsi Kalimantan Timur Tahun 2005-2012). *EKSPONENSIAL*, *5*(1), 73–80.

Ikbal, M. (2019). *Perbandingan Model Vector Autoregressive Integrated Moving Average dengan Generalized Space Time Autoregressive Integrated Moving Average untuk Peramalan Volume Pemakaian Air Bersih (Studi Kasus : Kota Surabaya, Kab. Gresik Dan Kab. Sidoarjo)* [Bachelor Thesis]. Universitas Airlangga.

Jannah, W., & Itratip, I. (2016). Kajian Pengolahan dan Distribusi Air Minum PDAM Giri Menang. *Jurnal Ilmiah Mandala Education*, *2*(2), 351. https://doi.org/10.58258/jime.v2i2.699

Jusmawati, J., Hadijati, M., & Fitriyani, N. (2020). Penerapan Model Vector Autoregressive Integrate Moving Average dalam Peramalan Laju Inflasi dan Suku Bunga di Indonesia. *EIGEN MATHEMATICS JOURNAL*, *3*(2), 73–82. https://doi.org/10.29303/emj.v3i2.62

Nugroho, A. A.-Z. (2022). Pemodelan Multivariate Time Series dengan Vector Autoregressive Integrated Moving Average (VARIMA). *Jurnal Riset Statistika*, *2*(2), 93–102. https://doi.org/10.29313/jrs.v2i2.1150

Panjaitan, H., Prahutama, A., & Sudarno, S. (2018). Peramalan Jumlah Penumpang Kereta Api Menggunakan Metode Arima, Intervensi dan Arfima (Studi Kasus: Penumpang Kereta Api Kelas Lokal EkonomiDAOP IV Semarang). *Jurnal Gaussian*, *7*(1), 96–109. https://doi.org/10.14710/j.gauss.7.1.96-109

Pertiwi, A., Dewi, L. F., Toharudin, T., & Ruchjana, B. N. (2022). 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*, *2*(1), 431–442. https://doi.org/10.30598/PattimuraSci.2021.KNMXX.431-442

Pratiwi, S., & Herlina, M. (2023). Pengaruh Harga Pangan terhadap Inflasi dengan Metode Vector Autoregressive Integrated Moving Average. *Jurnal Riset Statistika*, *3*(2), 87–96. https://doi.org/10.29313/jrs.v3i2.2690

Putri, I. D., & Oktaviarina, A. (2024). Penerapan Vector Autoregressive Integrated Moving Average (VARIMA) Pada Prediksi Indeks Standar Pencemaran Udara. *MATHunesa: Jurnal Ilmiah Matematika*, *12*(2), 364–373. https://doi.org/10.26740/mathunesa.v12n2.p364-373

Suharsono, A., & Anggraeni, A. D. (2014). Peramalan Penjualan Sepeda Motor Tiap Jenis di Wilayah Surabaya dan Blitar dengan Model ARIMA Box-Jenkins dan Vector Autoregressive (VAR). *Jurnal Sains dan Seni ITS*, *3*(2), 326–331. https://doi.org/10.12962/j23373520.v3i2.8071

Sutthichaimethee, P. (2017). Varimax Model to Forecast the Emission of Carbon Dioxide From Energy Consumption in Rubber and Petroleum Industries Sectors in Thailand. *Journal of Ecological Engineering*, *18*(3), 112–117. https://doi.org/10.12911/22998993/70200

Wilujeng, F. R. (2018). Faktor–Faktor yang Memengaruhi Optimalisasi Penangkapan Ikan dengan Metode Transformasi Box Cox pada Regresi Linier Berganda. *JIEMS (Journal of Industrial Engineering and Management Systems)*, *11*(1). https://doi.org/10.30813/jiems.v11i1.1011

**[This page intentionally left blank.]**