Climate modeling using vector moving average autoregressive

H Fitrianti¹, S M Belwawin¹, M Riyana¹ and R Amin²

¹Department of Early Childhood Teacher Education, Faculty of Teacher Training and Education, Universitas Musamus, Merauke, Indonesia
²Department of Agro-technology, Faculty of Agriculture, Universitas Hasanuddin, Makassar, Indonesia

E-mail: hariani@unmus.ac.id

Abstract. Information on weather and climate is a major need to support the smooth running of activities in various sectors, one of which is the agricultural sector. The agricultural sector is very dependent on climate change, where extreme climate change can cause storms, floods, droughts, resulting in crop failure. Therefore, climate information is very important, with information on climate change predictions, it can be used to determine when the right planting time is so that the quality and yield of agricultural production increases. The purpose of this study is to model climate data rainfall, humidity, maximum temperature, maximum temperature, and the length of solar radiation using a Vector Autoregressive Integrated Moving Average (VARIMA) approach. VARIMA is a time series model involving multivariate time series data, where the VARIMA model is a vector form of the ARIMA model. The time series analysis model requires data to meet stationary requirements. The results of the analysis obtained stationary climate data after being differed once, so the differencing data results that will be used in modelling. The time series climate data model was combined until the fourth lag based on SAS software assistance and the results were obtained namely the VARIMA model (1,1,1) with MSE values of 6045.33 and AICC namely 17.97823 and VARIMA (2,1,0) with MSE values namely 7432.434 and AICC which is 17.978023. Based on the two models, one of the best models was chosen with the smallest MSE and AIC value criteria. Thus the best VARIMA model that can be used to predict climate data in Merauke is the VARIMA (1,1,1) model.

1. Introduction
Information on weather and climate is a major need to support the smooth running of activities in various sectors, one of the sectors that are strongly influenced by climate, namely the agricultural sector. In the agricultural sector is very dependent on climate conditions[1]. Extreme climatic conditions can cause floods, droughts, and agricultural land to shift functions [2,3], so that the quality of harvests decreases[4,5], other adverse effects that can be caused by climate can result in crop failure [6]. Various methods have been used to get predictions of climate change that will occur. Climate change is a change in the statistical distribution of weather patterns for long periods (ie, decades to millions of years). Climate change can refer to changes in average weather conditions, or in variations in weather time around long-term average conditions. Climate change predictions are used to help determine when the right planting time is so that the quality and yield of agricultural production increases [7]. Merauke has great potential for the agricultural sector and can be used as a center for rice production in Indonesia[6]. A significant factor affecting the agricultural sector in Merauke, especially rice, is rainfall. Rainfall in Merauke is divided into two categories, namely low at 0-100 mm and currently at 100-300 [7]. Other factors that affect the agricultural sector are also influenced by temperature[7,8]. Several studies have been conducted to determine the impact of climate on the
agricultural sector, namely Hsiang and Burke [9] investigating the effects of climate change on the value of agricultural land, profitability, and production efficiency, Fitrianti, H [6] examined the effects of climate change on rice crop failure, Peng, S., Huang, et al. [10] investigating the impact of climate change on crop yields, Nurhayanti, Y [11] examined the impact of climate change in Indonesia on the agricultural sector namely maximum temperature rainfall, temperature average, minimum temperature, and others. In this study, the authors will research climate data modeling of rainfall, humidity, maximum temperature, average temperature, and duration of solar radiation using a time series model. The time series model used is VARIMA. VARIMA is a time series model involving multivariate time series data, where the VARIMA model is a vector form of the ARIMA model.

2. Method
Research is quantitative research with research data used, namely climate data obtained from the Meteorology, Climatology and Geophysics Agency (BMKG) Merauke, where the data is monthly climate data taken at the last 10 years. Climate data that will be used in modeling are climate data of rainfall, humidity, maximum temperature, maximum temperature, and the long duration of solar radiation. The approach used in climate data modeling is the VARIMA model approach. The VARIMA model is a vector form of the ARIMA model. So that the application requires data that has been stationary before further analysis. The VARIMA model has advantages, one of which is that it is easy to apply in multivariate data types. The equation of the VARIMA model can be written as follows [12]:

\[ Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \]

From the above equation, it is obtained that observation \( Y_t \) is a linear combination of the values of the previous observation multiplied by the weight on each observation which is adjusted to the data. Also besides, for each \( t \) it is assumed that \( \varepsilon_t \) each other is free with \( Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-p} \). The steps in modeling can be seen in the following Figure:
1) The time-series data plot for rainfall, humidity, maximum temperature, maximum temperature, and the length of the sun irradiation data.
2) Stationary test of data on the mean using differencing and variance using transformation. Stationary testing uses the ADF test.
3) Modeling and Testing of the best VARIMA models using AICC testing.
4) Estimating the VARIMA model
5) Testing the parameters of the VARIMA model using the method *ordinary least square* (OLS)[13].
6) Diagnosis checking is white noise residual examination and normal distribution.
7) Climate data prediction of rainfall, humidity, maximum temperature, average temperature, and long duration of sun irradiation.
8) Calculates the mean square error (MSE) value to determine the best prediction result.

3. Result and discussion

3.1. Descriptive statistics

| Table 1. Statistical results description climate data |
|-----------------------------------------------|
| Rainfall | Humidity | Maximum | temperature | Sunlight duration |
|         |          | Average |
| Mean    | 186      | 80      | 31.0302     | 26.9583         | 158.1635 |
| Median  | 119      | 81      | 31.2000     | 27.3000         | 156.5500 |
| Std. Deviation | 179 | 3      | 1.20351     | 0.99680         | 45.05411 |
| Variance | 31991    | 9      | 1.448       | 0.994           | 2029.873 |
| Minimum | 1.6      | 72      | 28.20       | 24.50           | 46.00   |
| Maximum | 663      | 85      | 33.70       | 28.60           | 270.00  |
Research study has 5 climate data variables studied, namely rainfall, humidity, maximum temperature, average temperature, and duration of solar radiation. The following are the results of descriptive statistics on climate data:

In table 1 data obtained humidity, the maximum temperature, and the average temperature, mean and median are not significantly different, this indicates that the data tends to be normal and stable. This is shown in the minimum and maximum data showing the range value that is not significant and is supported by the data variance value of each data. Whereas the rainfall data shows that the different mean and median are very significantly different and the variation in data is quite large seen from the maximum and minimum data ranges that are very significant, besides that also from the variance values obtained. For the duration of solar irradiation data is very different from other data, if seen from the mean and median data tend to be not significant, but when viewed from the spread of data, the data is very varied. This can be seen from the minimum and maximum data range, besides that it can be seen in the value of the variance. Based on the results of the calculation of table 1, the climate data is significantly the most influential, namely rainfall. The results of descriptive statistics on climate data are in line with Fitrianti’s research [7] that rainfall is the most significant factor affecting the occurrence of crop failure in Merauke.

3.2. Climate data

Data VARIMA Modeling VARIMA modeling is a time series modeling which involves multivariate time series data.

3.2.1. Stationary data

![Figure 1. Time series plot climate data](image)

Based on climate data plots in Figure 1 it can be seen that rainfall data, humidity, average temperature, maximum temperature, and duration of solar irradiation are stationary to the mean, but the spread is quite significant so that data is not stationary to variance. To be more accurate, the results of the analysis will be carried out by data testing using the ADF test. The results of the stationary test data using the ADF can be seen in Table 2.

| Variable                  | Zero Mean p-value (prob) | Intercept p-value (prob) | Trend p-value (prob) |
|---------------------------|--------------------------|--------------------------|----------------------|
| Rainfall                  | 0.0072                   | <0.0001                  | <0.0001              |
| Humidity                  | 0.4971                   | 0.0005                   | 0.0035               |
| Temperature averages      | 0.6453                   | <0.0001                  | <0.0001              |
| Maximum temperature       | 0.6971                   | <0.0001                  | 0.0002               |
| Lam irradiation           | 0.5573                   | 0.0002                   | 0.0005               |

Data is said to be stationary if \( p-value < \alpha = 0.05 \). Based on Table 2, data on humidity, average temperature, maximum temperature, and duration of solar irradiation are obtained from the mean, \( p-value < \alpha = 0.05 \) so that \( H_0 \) is rejected, meaning that climate data has no unit-roots. So it can be concluded that the variable humidity, average temperature, maximum temperature, and duration of solar irradiation are not stationary. But the variance of the data is stationary. While rainfall data is obtained \( p-value < \alpha = 0.05 \) for the mean and variance, it can be concluded that rainfall data has a unit root or stationary to the mean and variance.
The results of stationary test data in Table 2 obtained data that were not stationary to the mean, namely data on humidity, average temperature, maximum temperature, and duration of solar radiation. Therefore, differencing will be carried out for all climate data and then the re-test for the data is carried out. Stationary testing uses the same method that is using the ADF method. The results of the stationer test are presented in Table 3.

### Table 3. Results of testing climate data after one time differencing

| Variable                        | Zero Mean | Intercept | Trend     |
|---------------------------------|-----------|-----------|-----------|
| Climate                         | <0.0001   | <0.0001   | <0.0001   |
| Rainfall                        | <0.0001   | 0.0005    | 0.0035    |
| Humidity                        | <0.0001   | <0.0001   | <0.0001   |
| average temperature             | <0.0001   | <0.0001   | 0.0002    |
| maximum temperature of          | <0.0001   | 0.0002    | 0.0005    |
| Lamnya solar radiation          | <0.0001   |           |           |

Table 4 obtained climate data for testing the mean and variance that $p-value < \alpha = 0.05$ is so that it can be concluded the data had been thinking about root unit or data stationary and ready to be used in modeling. Thus the data VARIMA modeling that will be used is climate data that has been differentiated once because the data that meets the modeling criteria is stationary data.

#### 3.2.2. Determination of AICC based models

Determination of the initial climate data model is determined by the AICC value of the model, with the help of SAS 9.2 software combined 25 possible models, namely models with maximum order 4 and q-order maximum 4. Based on AICC criteria a good model is a model which has the smallest AICC value. Following are the AICC results from a combination of models built by SAS:

### Table 4. Minimum information criterion based on AICC (output results using SAS)

| Lag | MA 0     | MA 1     | MA 2     | MA 3     | MA4     |
|-----|----------|----------|----------|----------|---------|
| AR 0| 18.41999 | 18.405267| 18.245001| 18.43389 | 18.605452|
| AR 1| 18.047967| 17.978023| 18.355903| 18.645407| 18.838473|
| AR 2| 17.990675| 18.063033| 18.505116| 18.868923| 18.692016|
| AR 3| 18.25238 | 18.304312| 18.708119| 19.324077| 19.573657|
| AR 4| 18.412602| 18.541531| 18.675589| 19.861826| 20.906301|

From Figure 2, there are 2 smallest AICC models, namely at number 17, but the difference in AICC between the two models is only 0.02. Therefore, both models will be proposed as prospective models that meet data that has been *differenced* 1 time. The model is VARIMA (1,1,1) with AICC = 17.978023 and VARIMA (2,1, 0) with an AICC value of 17.990675.

#### 3.2.3. Estimating Parameter

##### 3.2.3.1. VARIMA Model (2,1,0)

Estimating parameters for the first model VARIMA (2,1,0) using the Least Squared Error method with the help of the quasi-Newton optimization method (maximum iteration: 500). Each parameter for each variable will be tested for significance with a t-test where the decision is based on the $p-value$, if $p-value < \alpha = 0.05$, then $H_0$ is rejected, so the parameter cannot be considered 0 or a significant parameter. The results of testing the parameters of the VARIMA model (2,1,0) can be seen in Table 5 and Table 6.
The result of testing the parameters of the VARIMA (2,1,0), namely using the Least Squared Error method with the help of quasi-Newton optimization method (maximum iteration: 500). The results of testing the parameters of rainfall, humidity, and average temperature using OLS, the estimated results of these parameters will be used to test the significance of the parameters of rainfall, humidity, and average temperature using the test t. The result of testing the significance of these parameters obtained means that the parameters of the significant variable so that it cannot be ignored. The parameter that cannot be ignored for rainfall is the variable t-1. Parameters that cannot be ignored in humidity are at t-1 and t-2. In the bulk data, humidity, and average temperature the accuracy of the predicted results of the data in the next period is strongly influenced by the data in the previous two periods.

Table 5 is the results of estimating the parameters of rainfall, humidity, and average temperature using OLS, the estimated results of these parameters will be used to test the significance of the parameters of rainfall, humidity, and average temperature using the test t. The result of testing the significance of these parameters obtained means that the parameters of the significant variable so that it cannot be ignored. The parameter that cannot be ignored for rainfall is the variable t-1. Parameters that cannot be ignored in humidity are at t-1 and t-2. In the bulk data, humidity, and average temperature the accuracy of the predicted results of the data in the next period is strongly influenced by the data in the previous two periods.

Table 6 obtained significance test results based on the calculated parameters obtained in the maximum temperature data only influenced by rainfall data at t-1. Whereas the duration of solar radiation is significant in the data of solar radiation duration at t-1, and the average humidity and temperature at t-2. Thus for the whole VARIMA (2,1,0) it is influenced by the data of the previous 2 periods.

3.2.3.2. VARIMA Model (1,1,1).
Parameter estimation for the VARIMA (1,1,1) model uses the same method to estimate the parameters VARIMA (2,1,0), namely using the Least Squared Error method with the help of quasi-Newton optimization method (maximum iteration: 500). The results of testing the parameters of the VARIMA model (1,1,1) can be seen in Table 7 and Table 8.
Table 7. Results of parameter estimation and testing of significance of rainfall, humidity, and average temperature data of VARIMA (1,1,1)

| Parameter | Estimation | Variable | Parameter | Estimation | Variable | Parameter | Estimation | Variable |
|-----------|------------|----------|-----------|------------|----------|-----------|------------|----------|
| AR1_1_1   | 0          | X1 (t-1) | AR1_2_1   | 0          | X1 (t-1) | AR1_3_1   | 0          | X1 (t-1) |
| AR1_1_2   | 0          | X2 (t-1) | AR1_2_2   | 0          | X2 (t-1) | AR1_3_2   | 0          | X2 (t-1) |
| AR1_1_3   | 64.540     | X3 (t-1) | AR1_2_3   | 1.225      | X3 (t-1) | AR1_3_3   | -1.142     | X3 (t-1) |
| AR1_1_4   | 34.970     | X4 (t-1) | AR1_2_4   | -3.244     | X4 (t-1) | AR1_3_4   | 3.259      | X4 (t-1) |
| AR1_1_5   | 0          | X5 (t-1) | AR1_2_5   | 0          | X5 (t-1) | AR1_3_5   | -0.032     | X5 (t-1) |
| AR2_1_1   | 0          | e1 (t-1) | AR2_2_1   | 0          | e1 (t-1) | AR2_3_1   | 0          | e1 (t-1) |
| AR2_1_2   | 0          | e2 (t-1) | AR2_2_2   | 2.140      | e2 (t-1) | AR2_3_2   | 0          | e2 (t-1) |
| AR2_1_3   | 25.862     | e3 (t-1) | AR2_2_3   | 1.039      | e3 (t-1) | AR2_3_3   | 0          | e3 (t-1) |
| AR2_1_4   | 32.460     | e4 (t-1) | AR2_2_4   | -3.101     | e4 (t-1) | AR2_3_4   | 3.012      | e4 (t-1) |
| AR2_1_5   | 0          | e5 (t-1) | AR2_2_5   | 0          | e5 (t-1) | AR2_3_5   | -0.037     | e5 (t-1) |

Table 7 results obtained overall rainfall, humidity, and the average temperature is affected at t-1. In detail, it can be seen that rainfall and humidity data are significantly affected by the average and maximum temperatures, while the mean temperature is significantly affected by the average temperature, maximum temperature, and the duration of sun exposure. The overall results of the test results obtained data of rainfall, humidity, and the length of the sun’s irradiation are only influenced by the data of one previous period.

Table 8. Results of the estimation parameters and significance testing the maximum temperature data and the length of solar radiation

| Parameter | Estimation | Variable | Parameter | Estimation | Variable |
|-----------|------------|----------|-----------|------------|----------|
| AR1_4_1   | 0          | X1 (t-1) | AR1_5_1   | 0          | X1 (t-1) |
| AR1_4_2   | 0.174      | X2 (t-1) | AR1_5_2   | 0          | X2 (t-1) |
| AR1_4_3   | -0.559     | X3 (t-1) | AR1_5_3   | -16.92     | X3 (t-1) |
| AR1_4_4   | 1.405      | X4 (t-1) | AR1_5_4   | 13.009     | X4 (t-1) |
| AR1_4_5   | -0.051     | X5 (t-1) | AR1_5_5   | 0          | X5 (t-1) |
| AR2_4_1   | 0          | e1 (t-1) | AR2_5_1   | 0          | e1 (t-1) |
| AR2_4_2   | 0.159      | e2 (t-1) | AR2_5_2   | 0          | e2 (t-1) |
| AR2_4_3   | 0          | e3 (t-1) | AR2_5_3   | -12.297    | e3 (t-1) |
| AR2_4_4   | 1.315      | e4 (t-1) | AR2_5_4   | 13.364     | e4 (t-1) |
| AR2_4_5   | -0.053     | e5 (t-1) | AR2_5_5   | 0          | e5 (t-1) |

Table 8 obtained significant variables for the maximum temperature data that is the data humidity, average temperature, maximum temperature, and the duration of solar radiation at t-1. As for the duration of a significant variable solar irradiation at the mean temperature and maximum temperature at t-1. So it can be concluded that VARIMA (1,1,1) is only influenced by data from one previous period.

3.2.4. Diagnostic

Test The diagnostic test used to test the VARIMA (2,1,0) and VARIMA (1,1,1) models meet the white noise criteria. White noise is the process between successive random variables which does not occur in correlation and follows a certain distribution. White noise criteria are used to determine the best model that will be used to predict the climate in the next period. The white noise test consists of two tests, namely the correlation test and the normal distribution for the residuals of each model.
3.2.4.1. Residual correlation test residual

Testing is done for the VARIMA (2,1,0) and VARIMA (1,1,1) models. The results can be seen in the following figure:

![Figure 2](image-url) Results of testing residual correlation

| Variable/Lag | VARIMA (2,1,0) | VARIMA (1,1,1) |
|--------------|---------------|---------------|
| Curah        | ![](image-url) | ![](image-url) |
| Lembap       | ![](image-url) | ![](image-url) |
| Rata         | ![](image-url) | ![](image-url) |
| Maks         | ![](image-url) | ![](image-url) |
| Sinar        | ![](image-url) | ![](image-url) |

Model VARIMA (2,1,0) of Schematic Representation of Cross Correlations of Residuals above we can see that for lags 1 and 2 there is no correlation between residuals, new correlations occur in lag-3, namely the humidity-\(t\) variable and the humidity variable \(\{t-3\}\) and max- \(\{t-3\}\). Whereas for the VARIMA (1,1,1) model for lag-1, 2, and 3 there is no correlation between residuals, new correlations occur in 1 variable in lag-4, 2 variables in lag-5, 1 variable in lag-6, and 1 variable in lag-7. Based on the cross-correlation table above, because the correlation occurs in a large lag, the correlation can be ignored.

3.2.4.2. Chi-squared

Residuals from the model must be normally distributed to meet the white-noise assumptions, the normal test, chi-squared, will be used to test the normality of the residuals. If \( p-value < \alpha = 0.05 \), it can be concluded that the residual is not normally distributed. The test results can be in Figure 4.

![Figure 3](image-url) Residual test results are normally distributed

| Variable | VARIMA Model (2,1,0) | VARIMA Model (1,1,1) |
|----------|-----------------|-----------------|
| Durbin Watson | Normality | Durbin Watson | Normality |
| Curah    | 2.11738 | 0.83 | 0.6612 | 2.12218 | 0.95 | 0.6206 |
| Lembap   | 2.35130 | 1.93 | 0.3602 | 2.07751 | 5.09 | 0.0785 |
| Rata     | 2.15231 | 0.75 | 0.6887 | 1.94487 | 1.19 | 0.5527 |
| Maks     | 2.04583 | 1.71 | 0.5501 | 2.10470 | 2.06 | 0.5001 |
| Sinar    | 2.20067 | 0.13 | 0.9350 | 2.29368 | 0.91 | 0.6333 |
In Figure 4 for VARIMA models (2,1,0) and VARIMA models (1,1,1) obtained all climate variables have a value of \( p\)-value > 0.05 so that H0 is not rejected, it can be concluded that the residuals meet the criteria for normal distribution.

3.3. Climate Data Prediction Doing

Forecasting for January 2019-December 2019 data using the two suitable models earlier, to obtain:

![Figure 4](image)

The picture shows the results of forecasting on average for climate data, namely VARIMA model (1,1,1) the closest to the original data. Furthermore, the forecasting results are used to determine the best VARIMA model that will be used to predict climate data based on the Mean squared error (MSE) criteria. The model chosen is a model that has the smallest MSE value. The results obtained by the VARIMA model (2,1,0) have MSE value 7432.4 and VARIMA model (1,1,1) have MSE value 6045.3, so the best model for predicting climate data is VARIMA model (1,1,1) with parameters, namely:

![Figure 5](image)

4. Conclusion

Based on the results of this study, it can be concluded that the VARIMA (1,1,1) model has residuals that are more uncorrelated than the VARIMA model (2,1,0) and the best model for modeling Merauke district climate factors is the VARIMA (1,1,1) model based on the AICC and MSE values.
References

[1] Rasmikayati E, Pertanian F, Padjadjaran U and Km J R B 2015 Dampak Perubahan Iklim terhadap Perilaku dan Pendapatan Petani 22 372–9

[2] Hukom E, Limantara L M and Andawayanti U 2012 Ketersediaan Air Di Irigasi Way Mital Propinsi Maluku J. Tek. Pengair. 3 24–32

[3] Surmaini E, Runtuwu E and Las I 2011 Efforts of Agricultural Sector in Dealing with Climate Change J. Litbang Pertan. 30 1–7

[4] Hidayati I N and Suryanto 2015 Pengaruh Perubahan Iklim Terhadap Produksi Pertanian Dan Strategi Adaptasi Pada Lahan Rawan Kekeringan J. Ekon. Stud. Pembang. 16 42–52

[5] Angles S, Chinnadurai M and Sundar A 2011 Awareness on Impact of Climate Change on dryland agriculture and Coping Mechanisms of Dryland Farmers Indian J. Agric. Econ. 66 365–72

[6] Keguruan F 2016 Menentukan Premi Murni menggunakan Generalized Linear Models dan Model Copula Magistra 3 69–81

[7] Pasaribu Y P, Fitrianti H and Suryani D R 2018 Rainfall Forecast of Merauke Using Autoregressive Integrated Moving Average Model E3S Web Conf. 73 12010

[8] Welch J R, Vincent J R, Auffhammer M, Moya P F, Dobermann A and Dawe D 2010 Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures Proc. Natl. Acad. Sci. 107 14562–7

[9] Hsiang S M and Burke M 2014 Climate, conflict, and social stability: What does the evidence say? Clim. Change 123 39–55

[10] Peng S, Huang J, Sheehy J E, Laza R C, Visperas R M, Zhong X, Centeno G S, Khush G S and Cassman K G 2004 Rice yields decline with higher night temperature from global warming Proc. Natl. Acad. Sci. 101 9971–5

[11] Nurhayanti Y, Nugroho M, Pascasarjana P, Ekonomi I, Indonesia U and Djohadikusumo S 2016 Sensitivitas Produksi Padi terhadap Perubahan Iklim di Indonesia 1974-2015 Agro Ekon. 27

[12] Cryer J D and Chan K-S 2008 Time Series Analysis With Applications in R (New York: Springer)

[13] Fitrianti H, Pasaribu Y P and Betaubun P 2019 Modeling factor as the cause of traffic accident losses using multiple linear regression approach and generalized linear models IOP Conf. Ser. Earth Environ. Sci. 235