SARIMA model for compiling the planting calendar

Irma Yahya¹, Mukhsar²*, Agusrawati¹, Lilis Laome², Andi Tenriawaru³, Ida Usman⁴ and Sitti Wirdhana Ahmad⁵

¹Department of Mathematics, FMIPA, Halu Oleo University, Kendari Indonesia
²Department of Statistics, FMIPA, Halu Oleo University, Kendari Indonesia
³Department of Computer Sciences, FMIPA, Halu Oleo University, Kendari Indonesia
⁴Department of Physics, FMIPA, Halu Oleo University, Kendari Indonesia
⁵Department of Biology, FMIPA, Halu Oleo University, Kendari Indonesia

mukhsarmuhu@gmail.com, mukhsar.mtmk@uho.ac.id (corresponding author)

Abstract. Rainfall patterns uncertainty in Southeast Sulawesi-Indonesia is causing many paddy crop failures. There needs to be an adjustment when the appropriate farming time. One of the adaptation efforts in dealing with the influence of climate shift is to establish a planting calendar. Paddy planting calendar is to determine the success of the harvest. The SARIMA model (0,0,0)(0,1,3)¹² is the best model for predicting rainfall data. Prediction results show that the highest rainfall occurs in January-July and November-December. Planting time can be done in January-July and November-December by avoiding August, September, and October.

1. Introduction
Indonesia has very large land for paddy plants. It should be enough for domestic needs. This plant can grow well in 34 provinces in Indonesia, including in Southeast Sulawesi Province. Generally, paddy fields still rely on rainfall patterns. Paddy plants in Southeast Sulawesi-Indonesia are relying on rainfall. Irrigation has not been able to cover the entire paddy field, only a small portion uses irrigation, less than 50% [1, 11]. Therefore, it is necessary to know the pattern of shifts and the distribution of rainfall in order to avoid crop failure by taking into account seasonal factors [2, 7, 8, 9, 10, 12].

Supplementary distribution of rainfall, especially in Southeast Sulawesi-Indonesia, greatly affects the productivity of paddy [11]. In recent years, rainfall in Southeast Sulawesi-Indonesia has experienced a shift and uncertainties. In general, farmers rely solely on rainfall to determine planting time. The harvest failed because it did not match the pattern of the rainy season. There needs to be an adjustment when the appropriate farming time so that farmers can adjust paddy varieties that are more adaptive to weather changes. One of the adaptation efforts in dealing with the influence of climate shift is to establish a planting calendar. Paddy planting calendar is to determine the success of the harvest using SARIMA model [3, 4, 5, 6].

2. Data Description
Monthly rainfall data used in this study are secondary data collected from the meteorology and geophysics agency of Kendari-Indonesia, period 2017-2018. The monthly rainfall data in Southeast Sulawesi-Indonesia for January-December period indicates that the data contain seasonal patterns (Figure 1a). This is confirmed by autocorrelation function or ACF (Figure 1b) and the plot partial
autocorrelation function or PACF (Figure 1c). Data transformation uses Box-Cox (Figure 1d), to guarantee the data to be stationary against variance, parameter $\lambda = 1$.

Rainfall data also has a seasonal effect, a seasonal differencing is needed ($S = 12$) and the results are presented in Figure 2b. Meanwhile in Figure 2a is the plot of transformation data. Based on the rainfall data description, the appropriate method to use is the Seasonal Autoregressive Integrated Moving Average (SARIMA).

![Figure 1. Plot of rainfall data description](image1)

![Figure 2. Data transformation](image2)

3. SARIMA Model
Seasonal Autoregressive Integrated Moving Average (SARIMA) is a tool for time series analysis by estimating the value of future variables based on current values. The SARIMA model consists of coordinates $(p, d, q)$, $p$ stands for the number of autoregressive terms, $d$ denotes the number of...
differences needed to make a stationary time series, and q represents moving averages from previously estimated errors in our model, or values that lag behind error terms.

The significance test of parameter estimation is done by t test and $\alpha = 5\%$. The parameter values $\phi_1$ and $SE(\phi_1)$ for the SARIMA model $(1,0,0)(0,1,1)^{12}$. The significant models are SARIMA $(0,0,0)(0,1,1)^{12}$, SARIMA $(0,0,0)(0,1,2)^{12}$ and SARIMA $(0,0,0)(0,1,3)^{12}$, that all the parameters of models are significant (for more detail, see Table 1). The significance test for residual inspection is sufficient to white noise and normal distribution assumptions. Based on Table 1, the SARIMA $(0,0,0)(0,1,1)^{12}$, SARIMA $(0,0,0)(0,1,2)^{12}$ and SARIMA $(0,0,0)(0,1,3)^{12}$ for all lags according to white noise and normal distribution assumptions. Furthermore, it has a value of $Q < X^2_{\alpha,df}$. There have fulfilled the requirement that residuals are white noise (see Table 2).

### Table 1. Parameter significant test

| Model                             | Parameter | $t_{hit}$ | p-value | Explanation     |
|-----------------------------------|-----------|-----------|---------|-----------------|
| SARIMA $(1,0,0)(0,1,1)^{12}$      | $\phi_1$  | 1.89      | 0.062   | Not significant |
|                                   | $\theta_1$ | 9.05      | 0.000   | Significant     |
| SARIMA $(0,0,1)(0,1,1)^{12}$      | $\theta_1$ | -1.61     | 0.110   | Not significant |
|                                   | $\theta_2$ | 8.97      | 0.000   | Significant     |
| SARIMA $(1,0,1)(0,1,1)^{12}$      | $\phi_1$  | 0.68      | 0.500   | Not significant |
|                                   | $\theta_1$ | 0.29      | 0.771   | Not significant |
|                                   | $\theta_2$ | 8.92      | 0.000   | Significant     |
| SARIMA $(0,0,0)(0,1,1)^{12}$      | $\theta_1$ | 9.11      | 0.000   | Significant     |
| SARIMA $(0,0,0)(0,1,2)^{12}$      | $\theta_1$ | 12.35     | 0.000   | Significant     |
|                                   | $\theta_2$ | -3.71     | 0.000   | Significant     |
| SARIMA $(0,0,0)(0,1,3)^{12}$      | $\theta_1$ | 7.11      | 0.000   | Significant     |
|                                   | $\theta_2$ | 2.78      | 0.007   | Significant     |
|                                   | $\theta_3$ | -3.59     | 0.001   | Significant     |

The best model is chosen using MSE. The model which has the smallest MSE is the best model. We obtained the SARIMA model $(0,0,0)(0,1,3)^{12}$ is the best one, see Table 3 for detail.

### Table 2. Residual inspection results for white noise assumptions

| Model                             | Lag (K) | df | Statistik test | $\alpha = 0.05$ | $X^2_{\alpha,df}$ | Explanation   |
|-----------------------------------|---------|----|----------------|-----------------|-------------------|---------------|
| SARIMA $(0,0,0)(0,1,1)^{12}$      | 12      | 11 | 6.5            | 0.05            | 19.68            | white noise   |
|                                   | 24      | 23 | 19.1           | 0.05            | 35.17            | white noise   |
|                                   | 36      | 35 | 32.4           | 0.05            | 49.80            | white noise   |
|                                   | 48      | 47 | 55.8           | 0.05            | 64.00            | white noise   |
|                                   | 12      | 10 | 8.3            | 0.05            | 18.31            | white noise   |
| SARIMA $(0,0,0)(0,1,2)^{12}$      | 24      | 22 | 19.6           | 0.05            | 33.92            | white noise   |
|                                   | 36      | 34 | 34.5           | 0.05            | 48.60            | white noise   |
|                                   | 48      | 46 | 49.5           | 0.05            | 62.83            | white noise   |
|                                   | 12      | 9  | 6.4            | 0.05            | 16.92            | white noise   |
| SARIMA $(0,0,0)(0,1,3)^{12}$      | 24      | 21 | 11.8           | 0.05            | 32.67            | white noise   |
|                                   | 36      | 33 | 26.5           | 0.05            | 47.40            | white noise   |
|                                   | 48      | 45 | 44.6           | 0.05            | 61.66            | white noise   |
### Table 3. Model selection

| Model            | MSE  |
|------------------|------|
| SARIMA (0,0,0)(0,1,1)$^{12}$ | 16541 |
| SARIMA (0,0,0)(0,1,2)$^{12}$ | 15756 |
| SARIMA (0,0,0)(0,1,3)$^{12}$ | 12972 |

#### 4. Forecasting

The model used for forecasting is the model with the smallest MSE, SARIMA (0,0,0)(0,1,3)$^{12}$. The forecast results of rainfall forecasting in Southeast Sulawesi-Indonesia from January-December 2018, are presented in Figure 3.

![Figure 3. Forecasting of SARIMA (0,0,0)(0,1,3)$^{12}$](image)

Validation model of rainfall forecasting data is to determine the accuracy of the predicting rainfall data for future period. Validation model uses monthly rainfall data, for January 2009-December 2017 period as training data to predict the rainfall data for January-December 2018 period. Comparison data uses rainfall data, January-December 2018 period.

Figure 3 is the result of the SARIMA model (0,0,0)(0,1,3)$^{12}$ and gives a fairly accurate prediction result. This model is able to predict rainfall in Southeast Sulawesi-Indonesia with a 95% confidence interval. This model is used to compile the planting time calendar, as illustrated in Figure 4.

**Figure 4. Planting time calendar of Southeast Sulawesi-Indonesia**

The appropriate planting time to plant rice is January-July and November-December each year. Because the planting cycle is 4 months, planting time can be done in January-April, March-June, May-August and June-September are not recommended for planting because the last month did not get rainfall.

#### 5. Conclusion

The SARIMA model (0,0,0)(0,1,3)$^{12}$ is the best model for predicting rainfall data. Prediction results show that the highest rainfall occurs in January-July and November-December. Planting time can be done in January-July and November-December by avoiding August, September, and October.
Acknowledgement

The authors would like to thank the Directorate of Research and Community Service (or DPRM) Ristekdikti Jakarta and Faculty of Mathematics and Sciences, Halo Oleo University. We also would like to thank the International Conference on Science and Mathematics (ICSM) 2019 and anonymous reviewers of ICSM 2019 for their valuable suggestions.

References

[1] Agriculture Ministry of Republic of Indonesia 2018 Paddy production of Indonesia, Kementan RI

[2] Adhikary S.K, Rahman M.D.M., Gupta A.D. 2012. A stochastic modelling technique for predicting groundwater table fluctuations with Time Series Analysis. International Journal of Applied Sciences and Engineering Research. Vol. 1. Iss. 2 p. 238–249.

[3] Box G.E.P, Jenkins G.M., Reinsel G.C. 2008. Time series analysis, forecasting and control. 4 th ed. New Jersey. John Wiley and Sons, Inc. ISBN 978-0470272848 pp. 784.

[4] Brockwell P.J., Davis R.A. 2002. Introduction to time series and forecasting. 2 nd ed. New York. Springer-Verlag. ISBN 978-1475777505 pp. 437.

[5] Chang X., Gao M., Wang Y., Hou X. 2012. Seasonal autoregressive integrated moving average model for precipitation of time series. Journal of Mathematics and Statistics. Vol. 8. Iss. 4 p. 500–505.

[6] Dickey D.A., Fuller W.A. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association. Vol. 74 Iss. 366 p. 427–431.

[7] Durbin J., Watson G.S. 1951. Testing for serial correlation in least squares regression, II. Biometrika. Vol. 38. Iss. 1–2 p. 159–179.

[8] Gautam R., Sinha A.K. 2016. Time series analysis of reference crop evapotranspiration for Bokaro District, Jharkhand, India. Journal of Water and Land Development. No. 30 p. 51–56. DOI 10.1515/jwld-2016-0021.

[9] Gerretsadikan A., Sharma M.K. 2011. Modelling and forecasting of rainfall data of Mekele for Tigray region (Ethiopia). Statistics and Applications. Vol. 9. Iss. 1–2 p. 31–53.

[10] Janhabi M., Jha R. 2013. Time-series analysis of monthly rainfall data for the Mahanadi River Basin, India. Sciences in Cold and Arid Regions. Vol. 5. Iss. 1 p. 73–84.

[11] Statistics Central Bureau 2018 Sulawesi Tenggara in numbers, BPS Sulawesi Tenggara.

[12] Wei, W. W.S. 2006. Time Series Analysis. Department of Statistics, Temple University, Philadelphia, Pennsylvania, USA.