Rainfall forecasting with climate change detection and its pattern relationship to rice production

Endah R. M. Putri¹, I Gusti Ayu Riska Astari², Nuri Wahyuningsih³

¹,²Department of Mathematics, Faculty of Mathematics, Computing, and Data Sciences, Institut Teknologi Sepuluh Nopember, Jl. Raya ITS, Sukolilo, Surabaya, 60111, Indonesia
³Department of Actuarial Sciences, Faculty of Mathematics, Computing, and Data Sciences, Institut Teknologi Sepuluh Nopember, Jl. Raya ITS, Sukolilo, Surabaya, 60111, Indonesia

E-mail: ¹ endahrmp@matematika.its.ac.id

Abstract. Rainfall as one of main factor of climate change, has a significant effect to agriculture. Prediction modelling of rainfall intensity is used to identify and to provide rainfall future description as a reference of agriculture planning regarding to food security. This study aims to provide the modelling of rainfall intensity using Seasonal Autoregressive Integrated Moving Average. Climate change in the rainfall intensity is detected using the existence of heavy-tail phenomena and therefore a Generalized Pareto Distribution is used to model it. To know the relationship between the rainfall intensity and its impact to rice production, Copula method is used as the extension. The study found: SARIMA ([1, 2, 6]; [1, 0])_{12} is the best model of the rainfall forecasting, climate change indication is detected but it occurs not significantly in area of Tabanan, Bali, and the relationship of the rainfall and the rice production indicates that a decreasing of rainfall intensity will lead to a decreasing rice production.

1. Introduction

Climate change is becoming a significant issue almost in this decade as its impact affects the population in many ways such as a high intense of rainfall or a less intense of rainfall, an increasing temperature, and a rise in sea-level. Rainfall as one of main factor in climate, gives major effect to agriculture [1]. The high increasing rainfall intense will affect the water balance in soil and farms but the very low rainfall intense will also lead to drought and lack of water supply to the farms area. In Indonesia, the high intense of rainfall increases the severity of flood in Jakarta. The increasing of flood severity in Jakarta, clearly leads to the more worse value of damage. In more general, it is suggested that in South East Asia, the extreme rainfall is increasing in this twenty-first century, both in severity and frequency, and potentially causing more severe natural disaster such as flood [2].

Variations in rainfall during this period will cause effect on agricultural production. The influence of climate on agricultural production in every region is different based on geography, topography, and also material in a region [3]. Climate data becomes a useful information to determine the level of influence on rice production so that we can set the right step in terms of agricultural management as an effort to increase production. Rainfall intensity is used by researchers as a sign of identifying climate change in a region where climate change is a long-term change in the distribution of weather patterns statistically [4].
The study of modelling and forecasting the rainfall have been a considerable interest to the scientist using various methods such as machine learning and time series [5, 6, 7, 8]. Some classical time series methods are commonly used to forecast the rainfall. Rainfall forecast using ARMA and ARIMA model for water allocation planning of agriculture in Thailand is proposed by Weesakul and Lowanichchai [9]. Subsequently, similar studies using ARIMA for forecasting the rainfall in Sri Lanka is proposed by Partheepan and Jeyakumar [10]. Considering the seasonal factor of the rainfall, some studies proposed a seasonal ARIMA (SARIMA). Seneviratna and Rathnayaka studied the forecasting of rainfall in Sri Lanka using SARIMA and compare it with the results obtained by using back propagation neural network method [11]. Another implementation of SARIMA in rainfall forecasting in India is proposed by Nirmala and Sundaram which is extended into a multivariate model of SARIMA [12]. Seasonal ARIMA for rainfall model with outlier detection and missing value is studied by Arumugam and Saranya [13]. Similar methods of ARIMA to forecast rainfall is presented by Patel et al [14], Delleur and Kavvas [15] and some others. The ARIMA method has been widely used and proven to be accurate significantly in rainfall forecasting [16, 11, 17, 9]. The method can be considered to obtain good prediction of rainfall and subsequently can provide an overview of future rainfall conditions as a reference in anticipation to food insecurity.

The existence of extreme values in rainfall data causes the need of Extreme Value Theory (EVT) to describe the behavior of the rainfall distribution [18]. The theory is widely used in many fields, such as finance [19], climate change [20], risk management [21], engineering [22], and so on. The studies aim to detect the existence of extreme values and heavy tail phenomena in the case. The main statistical distribution which shows properties such as the heavy tail phenomena is a Generalized Pareto Distribution [23]. Some studies about a Generalized Pareto Distribution and its parameters estimations are Hosking and Wallis [24] and Choulakian and Stephens [25].

The extreme points resulted in rainfall data will lead to a non-normal distribution which dominates most distributions in statistics. Therefore, an approach with not too strict distribution assumptions is needed. A non-parametric approach which is able to describe the dependencies of on extreme points without normality assumption is Copula [26]. Some works on Copula in describing relationship between variables in various areas are: relationship between incident of disease and death on two people by Clayton [27], a regression-based of Copula applied in rice-harvested areas by Sutikno et al [28], measuring a financial contagion by Rodriguez [29], and measuring the reliability characteristics on a complex engineering system by Ram and Singh [30].

As the rainfall intensity has effects to agricultural production including rice, it is necessary to describe the relationship pattern between the rainfall and rice production [1]. The method is applied to rice production data in Tabanan area, Bali, as the largest rice supplier area in Bali. The reason to choose this area to be observed is that Bali has the agricultural sector as main contributor to its regional income. In recent years, Bali suffers a significant decrease on its rice production and the highest decrease is in Tabanan. The hypothesis raised due to this trend is lead to the climate change effect.

Based on the background discussed, this paper aims to forecast the rainfall intensity using Seasonal ARIMA method, identifying the climate change, and investigated the relationship pattern of rainfall and rice production to provide empirical study for contributing in efforts of increasing the rice productivity using Copula method. This paper is organized as follows: Mathematical Models including Seasonal ARIMA, Maximum Likelihood Estimation, and Copula, are discussed in Section 2, Results and Discussion in Section 3, and last Conclusion is in Section 4.
2. Mathematical Models

2.1. Seasonal Autoregressive Integrated Moving Average (SARIMA)

The existence of seasonal factor in some forecasting objects lead to the modification of ARIMA model. For time series data which has seasonal pattern, the Seasonal ARIMA (SARIMA) model is written as follows,

\[ \phi_p \Phi_p (B)^S (1-B)^d (1-B^S)Z_t = \theta_o + \theta_q (B) \Theta_Q (B)^S a_t, \]

where:
- \( p, d, q \): order of AR, differencing, and non-seasonal MA
- \( P, D, Q \): order of AR, differencing, and seasonal MA
- \( (1-B)^d \): operator for differencing non-seasonal
- \( (1-B)^D \): operator for differencing seasonal
- \( Z_t \): data at t period
- \( a_t \): residual value at time \( t \)
- \( \phi_0 \): a constant
- \( B \): backward operator.

A Box-Cox transformation is used to test the stationarity on variance and Autocorrelation Function (ACF) plot to test the stationarity on mean. However, if the data is not stationary then differencing is needed to transform the data close to stationarity. Accordingly, the significance of parameters of autoregressive and moving average are tested.

The best Seasonal ARIMA model is selected based on the model which has all parameters significant, satisfy the assumptions that residual is normally distributed and white noise. Kolmogorov-Smirnov test is used for normality test and Ljung-Box test is used for a white noise test.

2.2. Generalized Pareto Distribution (GPD)

If a random variable \( X \) follows a generalized Pareto distribution with parameters \( (k, \sigma) \) then the probability density function is

\[ f(x) = \begin{cases} 
\frac{1}{\sigma} \left( 1 + \frac{kx}{\sigma} \right)^{-\left(\frac{1}{k}+1\right)}, & k \neq 0 \\
\frac{1}{\sigma e^{\frac{x}{k}}}, & k = 0
\end{cases}, \quad 0 \leq x < \infty \text{ if } k \leq 0 \text{ and } 0 \leq x < -\frac{\sigma}{k} \text{ if } k < 0. \]

There are three types of distribution in a generalized Pareto distribution (GPD): exponential distribution if \( k = 0 \), Pareto distribution if \( k > 0 \), and beta distribution if \( k < 0 \). The greater the value of \( k \), the heavier tail that the distribution has. Therefore the probability of extreme values occurrence becomes greater. According to Kotz and Nadarajah [6], if \( k < 0 \) the distribution has short tail and if \( k > 0 \) so the distribution has long tail. Peak over threshold method is used to identify the extreme value based on a threshold value. The threshold is determined as a 10% value of \( N \) number of data after sorting.

The parameters of the distribution is estimated using Maximum Likelihood Estimation (MLE) method which maximizes likelihood function or joint probability distribution function \( x_1, x_2, ..., x_n \). In the following Eq. (3), we present the likelihood function of a generalized Pareto
distribution (GPD) pdf for $k \neq 0$.

$$L(k) = \prod_{i=1}^{n} f(k, \sigma|x)$$

$$= \sigma^{-n} \prod_{i=1}^{n} \left(1 + \frac{kx_i}{\sigma}\right)^{-\left(\frac{1}{k} + 1\right)}$$

(3)

Equation (4) is the log-likelihood function of a generalized Pareto distribution (GPD),

$$ln L(k) = ln(\sigma^{-n} \prod_{i=1}^{n} \left(1 + \frac{kx_i}{\sigma}\right)^{-\left(\frac{1}{k} + 1\right)})$$

$$= -n ln \sigma - \left(\frac{1}{k} + 1\right) \sum_{i=1}^{n} \left(1 + \frac{kx_i}{\sigma}\right).$$

(4)

The next step is to maximize the log-likelihood function with respect to the parameter $\hat{k}$ and the scale parameter $\hat{\sigma}$ as follows,

$$\hat{k} = \frac{1}{1 + \hat{k}} \frac{1}{\sum_{i=1}^{n} \frac{ln(1 + \frac{kx_i}{\sigma})}{x_i} + \hat{k}}$$

(5)

$$\hat{\sigma} = \frac{(1 + \hat{k} - n\hat{k})}{n^2} \sum_{i=1}^{n} x_i.$$  

(6)

Based on MLE, the parameters estimated are not in closed-form formulas, therefore a numerical method is needed to obtain the parameters’ value.

2.3. Copula Approach

Copula approach provides some useful features such as deriving a joint distribution when variables are not normal, defining a non-parametric measures of dependence for pairs of random variables, and modelling joint distribution and dependencies [31].

If there are some random variables $(X_1, X_2, ..., X_m)$ and a cumulative density function (CDF) $F_{x_1}, F_{x_2}, ..., F_{x_m}$ of the random variables with domain $R$. The cumulative density function (CDF) is a non-decreasing function $F_x(-\infty) = 0$ and $F_x(+\infty) = 1$, then the joint density function is written as follows:

$$F(x_1, x_2, ..., x_m) = C_{X_1, X_2, ..., X_m} F_{x_1}(x_1), F_{x_2}(x_2), ..., F_{x_m}(x_m),$$

where $C_{X_1, X_2, ..., X_m}$ is a Copula with $C_x[0, 1]^m \rightarrow [0, 1]$.

The steps in applying Copula approach is starting with transformation process of the random variables $X_i$ to a Uniform distribution $UNIF[0,1]$ to obtain the scatter plot of the random variables $X_i$. The observation of scatter plot will result to the identification of the dependency pattern and the analysis of tail dependencies of the random variables. However, the scatter plot analysis should be supported by a further method, a Tau-Kendall method is used to estimate the Copula parameters. To assure that the parameters are statistically significant, a hypothesis test is conducted.
2.3.1. Copula parameter estimation with Tau-Kendall

Suppose that \((X_1, Y_1)\) and \((X_2, Y_2)\) are continuous random variable which are independent to a joint probability function \(H_1\) and \(H_2\) a marginal function \(F\) for \((X_1, Y_1)\) and \(G\) for \((X_2, Y_2)\). If \(C_1\) and \(C_2\) are notation of a Copula cumulative density function of \((X_1, Y_1)\) and \((X_2, Y_2)\) so \(H_1(x, y) = C_1(F(x), G(y))\) and \(H_2(x, y) = C_2(F(x), G(y))\). If Q is denoted as the difference between concordance and discordant probability of \((X_1, Y_1)\) and \((X_2, Y_2)\) where:

\[
Q = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0)
\]

Next, for \(C_1 = C_2 = C\), the value of \(Q\) is formulated as,

\[
Q(C.C) = 4 \int \int_I C(u, v)dC(u, v) - 1
\]

The value of \(Q\) shows the value of the Copula parameter estimation function with Tau-Kendall approach and it is written as \(\tau_{XY}\). In the next steps of this study, the calculation is conducted based on Eq. (8) to find Tau-Kendall correlation equation which relates to the generator function of the Copula Archimedean family. There are three types of Copula Archimedean family: Copula Clayton, Copula Gumbel and Copula Frank. Each type has a different generator function. Let \(u\) and \(v\) be the random variable with Uniform distribution \(UNIF(0, 1)\) which have joint distribution function \(C\). \(K_c(t)\) denotes the joint probability distribution function of \(C(u, v)\) [15]. Value of the joint probability distribution function \(K_c(t)\) is written as follows:

\[
K_c(t) = t - \frac{\phi(t)}{\phi'(t)}
\]

where \(K_c(t)\) is joint probability distribution function of \(C(u, v)\), \(\phi(t)\) is a Copula generator function, \(\phi'(t)\) is the first derivative of Copula generator function.

The value of Tau-Kendall’s approach for estimation of Copula parameter based on Equation (8) is:

\[
\tau = 4(\text{E}(C(u, v))) - 1
\]
\[
= 4(1 - \int_0^1 t - \frac{\phi(t)}{\phi'(t)}dt) - 1
\]
\[
= 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)}dt.
\]

Eq.(10) is the Copula Archimedean parameter with Tau-Kendall’s approach where \(\tau\) is a Tau-Kendall value and \(\phi(t)\) is a Copula Archimedean generator function.

2.4. Fitting for Copula Archimedean with Log-Likelihood value

The value of the probability density function (pdf) of the Copula Bivariate function is \(c(u, v)\). By applying the Maximum Likelihood Estimation (MLE), the likelihood function can be written
in Eq. (11),

$$L = \prod_{j=1}^{n} c(u_j, v_j)$$

$$\ln L(k) = \sum_{j=1}^{n} \ln c(u_j, v_j). \quad (11)$$

The likelihood functions for each type Archimedean copula are:

(i) **Log-Likelihood Value of Clayton**

$$\ln L(\theta) = \sum_{j=1}^{n} \ln(1 + \theta)((u_j^{-\theta} + v_j^{-\theta} - 1)^{-\frac{1}{\theta}} - 2v_j^{-\theta-1}u_j^{-\theta-1}). \quad (12)$$

(ii) **Log-Likelihood Value of Gumbel**

$$\ln L(\theta) = \sum_{j=1}^{n} \ln \left( \frac{(\ln u_j)^{\theta-1}}{u_j} \frac{(\ln v_j)^{\theta-1}}{v_j} \left( e^{-(\ln u_j)^{\theta} + (\ln v_j)^{\theta}} \right)^{\frac{1}{\theta}} \frac{1}{(\ln u_j)^{\theta} + (\ln v_j)^{\theta}} + (\theta - 1) \right) \quad (13)$$

(iii) **Log-Likelihood Value of Frank**

$$\ln L(\theta) = \sum_{j=1}^{n} \ln \left( \frac{\theta e^{\theta(u_j + v_j)}(e^\theta - 1)}{(e^\theta - 1 + (e^{\theta u_j} - 1)(e^{\theta v_j} - 1))^2} \right). \quad (14)$$

3. Results and discussions

3.1. Rainfall forecasting

The monthly rainfall is obtained from monthly data of year 2007-2016 and the monthly rice harvest data is of year 2013-2016 in Tabanan, Bali province. The rainfall data is collected from the Board of Meteorology, Climatology and Geophysics and Central Statistical Agency, Bali province, Indonesia. The pattern of rainfall data in Tabanan Bali during year 2007-2016, shown by Figure 1, is shaped almost resembles the U curve. This typical pattern is called monsoon region which means that the area of Tabanan Bali has a clear distinction between the rainy season period and the dry season period. However, it is also seen in the pattern of rainfall data that the peak rainfall in each year is generally in different months, but mostly occurs in February and March, so that data plot shows an indication of seasonal influence on the rainfall data. Time series plot of rainfall data in Figure 2 indicates that data has not been stationary in variance and mean. It means that the mean and variance can not be guaranteed to be unchanged significantly for a period of time. The trend is also not clearly shown in the figure. To be more accurate in capturing the stationarity, a Box-Cox plot and an ACF plot are applied. In Figure 3, the summary of the Box-Cox plot has been shown and the round value ($\lambda$) is 0.28 and it means that the data is not stationary. A transformation is applied to the data using a transformation function $Y_t = (Z_t)^\lambda$ where $Z_t$ is the actual rainfall, $\lambda$ is the round value, and $Y_t$ is the value of Box-Cox transformation. After the transformation, the data has been stationary to variance indicated by the round value is equal to 1 (see Figure 4). The stationarity to mean is shown by plot ACF in Figure 5. The plot has a decreasing pattern similar to sine wave pattern and many
lags out of the significance limit which indicates that the lag is significant. Thus, it is still being suspected that the data is not stationary on the mean and is also still difficult to determine the model. Therefore process of differencing in one time is applied to the data. The result of differencing on the rainfall transformation data is shown in Figure 6 as an ACF plot. The lag out of significant limit are less than before and are slowly decreasing (see lag 12, lag 24, lag 36, lag 48 and lag 60). It is also identified as the presence of seasonal patterns in the data over 12 months observations. So a 12-lag differencing is required to be able to see the pattern of seasonal ACF and to determine the forecasting model for this rainfall data. In Figure 7, lag 1 and lag 12 (seasonal) are out of the significant limit and this means that the lag is significant. PACF is observed before applying the differencing process on lag 12 of rainfall data. In Figure 8, it can be seen that out of the significant lags limit are lag 1, lag 2, lag 6 and lag 12 (seasonal). Therefore, based on Figure 7 and Figure 8, the temporary model is SARIMA \((1, 2, 6, 1, 1)(1, 1, 1)^{12}\).

In Figure 9, the time series plot shows that the data is stationary in variance and mean. The conclusion is based on the data range which is ranging from \(-1.5\) to 1.5 and the mean of the data is close to 0. As the data has been transformed by Box-Cox transformation and differencing on lag 1 and lag 12, they are stationary in the mean. The occurrence of changes in variance and mean values significantly do not happen and this means that the data is stationary. Accordingly, 14 temporary SARIMA models are found and the significance test is applied. Table 1 described the significant test results of 14 temporary SARIMA models and it can be seen that among the 14 models after overfitting process, only 5 temporary SARIMA models have significant parameters, provided that \(P_{value} < 0.05\). Then the parameters in the model are significant and the 5 significant models are continued on diagnostic test to determine the best model for rainfall forecasting.
Figure 5. ACF plot for rainfall data transformation

Figure 6. ACF plot for differencing 1 lag

Figure 7. ACF plot for differencing 12 lag

Figure 8. PACF plot for differencing 12 lag

Figure 9. Time Series Plot for Rainfall Data after Differencing 12 lag

The diagnostic test needs normality assumption on residual which is shown in Table 2. The five SARIMA temporary models are satisfying the normal assumption on the residual, shown by the value $D_{count} > 0.05$, thus a white noise test on the residual is applied for 5 SARIMA temporary models.

In Table 3, a white noise test is conducted to find the model with independent residue. From five SARIMA temporary models, only one model can satisfy the white noise test, the model is $((1, 2, 6), 1, 0)(0, 1, 1)^{12}$. 
The model can be written as follows:

\[
\begin{align*}
&\phi_1 = -0.40527, \\
&\phi_2 = -0.23365, \\
&\phi_3 = -0.28001, \\
&\phi_4 = -0.12346, \\
&\phi_5 = 0.08892, \\
&\phi_6 = 0.1392, \\
&\phi_7 = 0.9874, \\
&\phi_8 = 0.9991.
\end{align*}
\]

Based on Table 1, Table 2, and Table 3, we can determine that the best model is SARIMA [(1, 2, 6), (1, 1, 1)] that satisfies the assumptions of significant parameters, normal distributed residual and white noise. The model can be written as follows:

\[
\begin{align*}
&\phi_1 = -0.40527, \\
&\phi_2 = -0.23365, \\
&\phi_3 = -0.28001, \\
&\phi_4 = -0.12346, \\
&\phi_5 = 0.08892, \\
&\phi_6 = 0.1392, \\
&\phi_7 = 0.9874, \\
&\phi_8 = 0.9991.
\end{align*}
\]
### Table 3. White Noise Test

| SARIMA Model | Lag | P-value | Decision       |
|--------------|-----|---------|----------------|
| $([1, 2, 6], 1, 0)(0, 1, 1)_{12}$ | 6   | 0.2480  | White Noise    |
|             | 12  | 0.3650  |                |
|             | 18  | 0.7766  |                |
|             | 24  | 0.8807  |                |
|             | 30  | 0.9635  |                |
|             | 36  | 0.9799  |                |
|             | 42  | 0.8491  |                |
|             | 48  | 0.8188  |                |
| $([1, 6], 1, 0)(0, 1, 1)_{12}$ | 6   | 0.0045  | Not White Noise|
|             | 12  | 0.0089  |                |
|             | 18  | 0.0652  |                |
|             | 24  | 0.1112  |                |
|             | 30  | 0.2617  |                |
|             | 36  | 0.4336  |                |
|             | 42  | 0.2717  |                |
|             | 48  | 0.3484  |                |
| $([1, 2], 1, 0)(0, 1, 1)_{12}$ | 6   | 0.0193  | Not White Noise|
|             | 12  | 0.0258  |                |
|             | 18  | 0.0829  |                |
|             | 24  | 0.11156 |                |
|             | 30  | 0.2337  |                |
|             | 36  | 0.3736  |                |
|             | 42  | 0.1905  |                |
|             | 48  | 0.1479  |                |
| $(1, 1, 0)(0, 1, 1)_{12}$ | 6   | 0.0001  | Not White Noise|
|             | 12  | < 0.0001 |                |
|             | 18  | < 0.0001 |                |
|             | 24  | < 0.0001 |                |
|             | 30  | < 0.0001 |                |
|             | 36  | 0.0030  |                |
|             | 42  | 0.0013  |                |
|             | 48  | 0.0017  |                |
| $(6, 1, 0)(0, 1, 1)_{12}$ | 6   | 0.0005  | Not White Noise|
|             | 12  | 0.0022  |                |
|             | 18  | 0.0158  |                |
|             | 24  | 0.0297  |                |
|             | 30  | 0.0786  |                |
|             | 36  | 0.1533  |                |
|             | 42  | 0.0203  |                |
|             | 48  | 0.0482  |                |

\[ Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \phi_6 Y_{t-6} - Y_{t-1} \\
+ \phi_1 Y_{t-2} + \phi_2 Y_{t-3} + \phi_6 Y_{t-7} \\
- Y_{t-12} + \phi_1 Y_{t-13} + \phi_2 Y_{t-14} + \phi_6 Y_{t-18} \\
+ Y_{t-13} - \phi_1 Y_{t-14} - \phi_2 Y_{t-15} - \phi_6 Y_{t-19} = a_t - \Theta_{12} a_{t-12} \] (15)
The decomposition equation of SARIMA \(([1, 2, 6], 1, 0)(0, 1, 1)^{12}\) is written as follows:

\[
Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_6 Y_{t-6} + Y_{t-1} - \phi_1 Y_{t-2} - \phi_2 Y_{t-3} - \phi_6 Y_{t-7} + Y_{t-6} - \phi_1 Y_{t-7} - \phi_2 Y_{t-12} - \phi_6 Y_{t-13} - \phi_1 Y_{t-13} + \phi_2 Y_{t-14} + \phi_6 Y_{t-19} + a_t - \Theta_{12} a_{t-12},
\]

where \(Y_t\) is transformed rainfall data, \(\phi\) is an AR coefficient value, and \(\Theta\) is an MA seasonal coefficient value. The forecasting result using SARIMA \(([1, 2, 6], 1, 0)(0, 1, 1)^{12}\) is presented in Table 4.

| No. | Month       | Forecasting (mm/day) |
|-----|-------------|-----------------------|
| 1   | January     | 31.00593051           |
| 2   | February    | 31.01863238           |
| 3   | March       | 23.07059878           |
| 4   | April       | 20.43871533           |
| 5   | May         | 15.67197823           |
| 6   | June        | 5.843018774           |
| 7   | July        | 5.449416359           |
| 8   | August      | 3.690993773           |
| 9   | September   | 1.493669625           |
| 10  | October     | 11.15543705           |
| 11  | November    | 20.63065345           |
| 12  | December    | 30.27560874           |

3.2. Characteristics of rainfall

The following is a general overview of the characteristics of rainfall using the data collected in Tabanan Bali. The rainfall data generated by SARIMA \(([1, 2, 6], 1, 0)(0, 1, 1)^{12}\) is plotted with the training data during 2007-2016 in Figure 10. The data is likely to be in the likelihood of the training data. It indicates that the model is satisfying. Simultaneously, the data is roughly divided into two periods: 2008-2012 and 2013-2016 (plus the forecasting as 2017 data). The tails distribution of data is found in both periods which decrease very slowly, presented in Figure 10.
11 and 12. Therefore, the data is concluded to have the occurrence of heavy tail. Heavy tail shows the existence of extreme data in rainfall data. Extreme rainfall (maximum) is identified by a very high rainfall value in certain periods of time. Extreme data in each period can be seen from the determination of the value that exceeds the threshold value of the data. The threshold value of rainfall data of the first period and the second period is used to obtain the extreme data (see Table 5). The determination of threshold value is obtained by 10% percentage of the sorted data \((b = 10\% \times N)\) where \(N\) is the amount of data from any period.

| Period I | Period II |
|----------|-----------|
| \(N\)    | 60        |
| \(b\)    | 6         |
| \(threshold\) (cm/day) | 2.8473 | 2.5214 |

### 3.3. Parameters estimation of Generalized Pareto Distribution

Climate change is a long-term change in the distribution of weather patterns statistically from a decade (10 years) until millions of years. Climate change is characterized by changes in weather patterns, especially rainfall within a period of at least 10 years, where in the case of Tabanan Bali, rainfall samples are taken to detect the occurrence of climate change. The existence of heavy tail distribution and the threshold have been investigated in the previous subsection. Therefore the proper statistical distribution to show the heavy tail phenomena is the Generalized Pareto Distribution (GPD). The following is the result of parameter estimation from Generalized Pareto Distribution (GPD) using Maximum Likelihood Estimation (MLE) as shown in Table 6. The results in Table 6 is tested using Kolmogorov-Smirnov Test to check the significance of the data following GPD. The relationship between rice production and rainfall follows Copula based on significant parameters, by testing \(Z_{\text{count}} > Z_{\text{table}}\). The Kolmogorov-Smirnov test results is presented in Table 7 and the conclusion is that the data is following GPD. As the data is following the GPD then the parameters should be estimated. The result of shape parameter from period I is \(\hat{k} = 0.1650\), where in GPD if value \(k > 0\) then the distribution type is Pareto. While in period II the value of the shape parameter is \(\hat{k} = -0.2849\), where if the value of \(\hat{k} < 0\) then the type of distribution is Beta. So that in period I and II, the climate change is
detected in the area of Tabanan Bali since the distributions changes from period I to period II. However, when assessing the confidence interval, the parameter estimation value of period I is within the confidence interval of period II and otherwise. Also it can be seen from the value of their distributions through Kolmogorov-Smirnov test, the results show that the extreme data of rainfall in period I and period II has followed GPD. So that it can be said that the occurrence of climate change is not very significant in both periods.

Table 6. Result of GPD Parameter Estimation for Rainfall Data in Tabanan Bali

| Value                  | Period I (2008-2012) | Period II (2013-2017) |
|------------------------|-----------------------|------------------------|
| $\hat{\sigma}$ (scale parameters) | 0.5192                | 1.1062                 |
| CI 95% for $\hat{\sigma}$     | 0.1234 $\leq \hat{\sigma} \leq$ 2.1845 | 0.3095 $\leq \hat{\sigma} \leq$ 3.9538 |
| $\hat{k}$ (shape parameters) | 0.1650                | -0.2849                |
| CI 95% for $\hat{k}$         | -1.0356 $\leq \hat{k} \leq$ 1.3656 | -1.3018 $\leq \hat{k} \leq$ 0.7321 |
| Distribution Type         | Pareto Distribution   | Beta Distribution      |

Table 7. Result for Kolmogorov Smirnov Test

|                      | Period I | Period II |
|----------------------|----------|-----------|
| Mean                 | 3.4640   | 3.3744    |
| Standard Deviation   | 0.7326   | 0.7646    |
| $D_{count}$          | 0.3163   | 0.3166    |
| $P_{value}$          | 0.4874   | 0.4863    |
| Conclusion           | Following GPD | Following GPD |

Table 8. The result of Copula Archimedean Parameter Estimation with Tau Kendall

| Copula       | Estimasi $\hat{\theta}$ | SE($\hat{\theta}$) | $|Z_{hitung}|$ | Conclusion |
|--------------|--------------------------|---------------------|--------------|------------|
| Clayton      | 0.1145                   | 0.0402              | 2.8480       | Significant |
| Gumbel       | 1.0573                   | 0.0423              | 24.9672      | Significant |
| Frank        | -0.4885                  | 0.0378              | 12.9372      | Significant |

In each type of Archimedian Copula (see Table 8), all estimation values are within the boundary range of each Copula Archimedean. Firstly, Clayton parameter estimation is $\hat{\theta} = 0.1145$ where boundary for Clayton parameter values is $\theta \in [-1, \infty) - (0)$. Secondly, the estimated Gumbel parameter is $\hat{\theta} = 1.0573$ where boundary for Gumbel parameter value is $\theta \in [1, \infty)$ and lastly the value of the parameter of Frank is $\hat{\theta} = -0.4885$ where boundary for Frank parameter value is $\theta \in R - (0)$. Since all parameter values lie within the specified limits and their parameter values are significant, it is said that the dependency structure of rice production and rainfall has a relationship following more than one type of Copula. Consequently the Copula fitting process is conducted by considering the largest Log-Likelihood value among every Copula Archimedean family type to get the best model.
3.4. Relations identification of rainfall and rice production

The scatterplots of rainfall and rice production are presented in Figure 13 and 14 and are aimed to describe the pattern relationship between rice production and rainfall. However, the scatterplots should be supported by other statistical tool. Tau-Kendall correlation coefficient is used and calculated resulting the value 0.0542 which indicates a very weak relationship between rice production and rainfall. This conclusion is strengthen by the \( p \)-value > \( \alpha \) which shows that there is no-close relationship between the two variables. A further analysis of dependencies using Copula Archimedean to look at the dependency model in particular.

![Figure 13](image-url)  
**Figure 13.** Scatterplot for Rice Production and Rainfall in Tabanan Bali

![Figure 14](image-url)  
**Figure 14.** Scatterplot for Rice Production and Rainfall in Tabanan Bali (Transformation to UNIFORM \([0, 1]\))

### Table 9. Fitting Copula with MLE

| No. | Copula Archimedean | Log-Likelihood |
|-----|--------------------|----------------|
| 1   | Clayton            | 0.0995         |
| 2   | Gumbel             | 0.0248         |
| 3   | Frank              | 0.0017         |

The best model was chosen based on the results of MLE fitting by looking at the largest log-likelihood value and estimated value is significant (see Table 9), so that the dependency structure between rice production and rainfall follows Clayton copula since Clayton has the largest log-likelihood value. The pattern of relationship between rice production and rainfall in Tabanan Bali follows the Copula Clayton type where Copula has below tail dependencies, so it is concluded that extreme events occur when rainfall and rice production are low. The lower the rainfall and rice production, the relationship is stronger, meaning that if rainfall intensity decreases then rice production in Tabanan Bali will decrease.

4. Conclusion

Based on the analysis and discussion on chapters before, we can conclude that the best forecasting model for rainfall in Tabanan Bali in 2017 is SARIMA \((1, 2, 6, 1, 0)(0, 1, 1)_{12}\), with the result of forecasting process having maximum value in February is 31.018632 mm/day, the minimum value in September is 1.493696 mm/day and mean of the result of forecasting process is 16.595388 mm/day. Based on the estimation value of GPD for shape parameters in period I (2008-2012) and period II (2013-2017), there is a change of distribution type between two periods, so that it is indicating the detection of climate change in Tabanan Bali. However,
climate change does not occur too significantly based on the value of confidence interval in both periods. The identification of the relationship pattern between rice production and rainfall with Copula is resulted in a more specific relationship between the two variables. The pattern of relationship between rice production and rainfall follows Copula Clayton with the parameter estimation of \( \hat{\theta} = 0.1145 \) and the log-likelihood value is 0.0995. Copula Clayton has a down tail dependency, which means the incidence of intensity decrease in rainfall can affect the decrease of rice production.

References
[1] Levine D I and Yang D 2014 The impact of rainfall on rice output in Indonesia Working Paper 20302 National Bureau of Economic Research
[2] Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averty K, Tignor M and Miller H 2007 IPCC (intergovernmental Panel on Climate Change). Climate change 2007: The Physical Science Basis: Summary for Policymakers Technical report Cambridge University Press, Cambridge, UK
[3] Iizumi T and Ramankutty N 2015 Global Food Security 4 46 – 50 ISSN 2211-9124 URL http://www.sciencedirect.com/science/article/pii/S2211912414000583
[4] Thornton P K, Erickson P J, Herrero M and Challinor A J 2014 Global Change Biology 20 3313328
[5] Tapoglu E, Karatzas G P, Trichakis I C and Varouchakis E A 2014 Journal of Hydrology 519 3193 – 3203 ISSN 0022-1694
[6] Onof C, Chandler R E, Kakou A, Northrop P, Wheeler H S and Isham V 2000 Stochastic Environmental Research and Risk Assessment 14 384–411
[7] Nasserri M, Asghari K and Abedini M 2008 Expert Systems with Applications 35 1415 – 1421
[8] French M N, Krajewski W F and Cuykendall R R
[9] Weesakul U and Lowanichchai S 2005 Thammasat Int. J. Sci. Technol. 10
[10] Partheepan K and Jayakumar P 2005 Water Resources Research in Sri Lanka, Water Professionals’ Day Symposium
[11] DMKN K S and Rathnayaka R K T 2018 Indian Journal of Mathematics 13 57–63
[12] M N and SM S 2010 National Journal on Advances in Building Sciences and Mechanics 1 43–47
[13] Arumugam P and Saranya R 2018 Materials Today: Proceedings 5 1791 – 1799
[14] Patel D P, Patel M M and Patel D R 2014 IOSR Journal of Electronics and Communication Engineering 9 83–87
[15] Delleur J W and Kavvas M L 1978 Journal of Applied Meteorology 17 1528–1536
[16] Kaushik I and Singh S M 2008 Journal of Environmental Research and Development 3
[17] S S, R M and S E S International Journal of Climatology 27 819–829
[18] de Haan L and Ferreira A 2007 Extreme Value Theory : An Introduction (Springer Science & Business Media, 2007)
[19] Genay R and Seluk F 2004 International Journal of Forecasting 20 287 – 303 ISSN 0169-2070 forecasting Economic and Financial Time Series Using Nonlinear Methods
[20] Gilleland E, Katz and W R 2006 Analyzing seasonal to interannual extreme weather and climate variability with the extremes toolkit 18th Conference on Climate Variability and Change, 86th American Meteorological Society (AMS) Annual meeting vol 29 (Citeser)
[21] Gilli M et al. 2006 Computational Economics 27 207–228
[22] Castillo E 2012 Extreme value theory in engineering (Elsevier)
[23] Akinsete A, Fumoye F and Lee C 2008 Statistics 42 547–563
[24] Hosking J R and Wallis J R 1987 Technometrics 29 339–349
[25] Choulakian V and Stephens M A 2001 Technometrics 43 478–484
[26] Embrechts P, Lindskog F and McNeil A 2001 Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich
[27] Clayton D 1982 Communications in Statistics-Theory and Methods 11 2129–2155
[28] Sutikno, Kuswanto H and Ratih I D 2014 Journal of Mathematics and Statistics 10 192
[29] Rodriguez J C 2007 Journal of empirical finance 14 401–423
[30] Ram M and Singh S 2009 Journal of Reliability and Statistical Studies 2 91–102
[31] Trivedi P K, Zimmer D M et al. 2007 Foundations and Trends® in Econometrics 1 1–111