Software development to predict the level of paddy production using Gaussian Copula Marginal Regression

Gandes Goldestan¹, Sutikno², Budi Setiyono¹, Imam Mukhlash¹

Department of Mathematics¹, Department of Statistics², Faculty of Mathematics, Computing, and Data Sciences, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

e-mail: imamm@matematika.its.ac.id

Abstract. Appropriate decision making in rice production management is needed to support national food security efforts, especially in East Java which is the largest rice production province in Indonesia. This study aims to develop a web-based decision support system to predict rice production levels in five districts of rice production centers in East Java. A web-based decision support system is constructed to make information accessible and understandable. The method used in this study is Gaussian Copula Marginal Regression (GCMR), which is regression based on Copula. Predictor variables (rainfall) and response variables (harvested area) are identified their correlation using Copula correlation. Estimation of harvested area is constructed using GCMR model. The results showed that the GCMR model used was able to model the area of rice harvest in five districts of rice production centers in East Java. Furthermore, the system is also able to predict the level of rice production in the short term as decision support to help the authorities give consideration in taking policy related to agriculture and food security in East Java.

1. Introduction

Rice is the main source of food for the people of Indonesia. Based on the Agency for the Assessment and Application of Technology (BPPT) in 2016, Indonesia became the country with the highest rice consumption per capita in the world, followed by China, Japan and Korea. On the other hand, the number of Indonesian people reached 252.17 million people with a growth rate of 1.31% and the level of rice consumption reached 132.98 kg / capita / year making rice a major commodity in the agricultural sector. Therefore, agriculture is a very important sector and increasing rice production is really a priority to reduce supply shortages [1].

According to Figures Prediction II (ARAM II) in 2017, East Java is the largest rice producing province in Indonesia, followed by West Java, Central Java, South Sumatra, and Lampung. The need for food will continue to increase in line with population growth and increased per capita consumption. Therefore, proper management of rice production is needed to support national food security efforts.

Rice production is obtained based on harvested area which is influenced by several factors, one of which is rainfall. Therefore, we need a method that can model the relationship between rainfall and harvest area so that it can predict the level of rice production.
Research on the relationship of climate change to agricultural production was previously carried out by Sutikno et al (2013) using the extreme value theory approach. In this study, Ordinary Least Square (OLS) models were used to predict rainfall and then Copula models were obtained to predict harvest area in an area with rainfall data [2]. Then the research was developed by Maulidiyah and Mukhlash (2014) by integrating it into web-based decision support software so that the results can be represented in the form of information that is easily accessible and understood [3]. From this study, Miftachurohmah and Mukhlash (2015) developed it again by adding ENSO indicator and robust regression model with an estimate of M to predict its production area in Decision Support Systems [4]. Furthermore, research on the same thing was developed again by Maziyah and Mukhlash (2017) by using another method, namely robust regression with S-estimation and MM-estimation to connect between harvested area and rainfall in predicting rice production levels that have been successfully integrated into in the existing web-based Decision Support System software [5].

Copula is a statistical method that shows the relationship between variables, where this method is not too strict on distribution assumptions, especially normal distribution. Besides that Copula also has the advantage of being able to clearly describe dependencies at extreme points. Several studies using the Copula approach in the field of climatology, such as those carried out by Scholzel and Friederichs (2008) and Oktaviana (2012) show that the Copula method has a better performance under conditions of normal violations [6] [7]. However, these studies are still limited to correlation and do not identify the causal relationship. The method that can be used to model the causal relationship in extreme events is the Gaussian Copula Marginal Regression (GCMR).

In Indonesia, Dewi Ratih (2013) has previously conducted research on the relationship between harvested area and rainfall as well as modelling of rice harvest area using the GCMR method, where this method is very well used to model causal relationships in extreme events such as climate and rainfall [8]. The calculation process can be done manually, but it takes a long time and allows many human errors to occur so that it is not effective. In addition, the model will certainly be more useful if integrated into a web-based software so that the results can be represented in the form of information that is easily accessible and understood. Therefore, based on this background the author conducted this study entitled "Software Development to Predict the Level of Paddy Production Using Gaussian Copula Marginal Regression".

1.1 Harvest Area and Productivity
The land area is the area of rice fields that will be planted with rice in certain seasons. The harvest area is the area of the crop that is taken after the plant is old enough. While rice productivity is rice production per unit area of land used in rice farming and is measured in tons per hectare (ton / ha)

1.2 Rainfall
Rain is a symptom of meteorology and also elements of climatology. Rain is a hydrometeor that falls in the form of water particles that have a diameter of 0.5 mm or more. Hydrometeors that fall to the ground are called rain while those that do not reach the ground are called Virga [9]. Rainfall is one of the elements of the weather whose data is obtained by measuring it using a rain gauge, so that it can be known in millimetres (mm). Rainfall of 1 mm is the amount of rainwater that falls on the surface per unit area (m2) with the note that nothing evaporates, seeps or flows.
1.3 Decision Support System

Decision Support System (DSS) is a system that is able to provide problem solving abilities and communication capabilities for problems with semi-structured and unstructured conditions where no one knows for certain how a decision should be made [10].

![Figure 1. DSS Component Schema.](image)

To be able to implement a decision support system there are four subsystems that must be provided, namely data management subsystems, model management subsystems, knowledge management subsystems and user interface subsystems. As shown in Figure 1.

1.4 Copula’s Family

The most popular Copula families used are the Gaussian Copula and Archimedean Copula. Normal Gaussian or Copula is obtained from the transformation of random variables into standard normal distributions. Normal Copula function is:

\[ C_{(X_1,X_2,...,X_m)}(u_1,u_2,...,u_m) = (F_{N(0,1)}^{-1}(u_1), F_{N(0,1)}^{-1}(u_2), ..., F_{N(0,1)}^{-1}(u_m)) \]  

(1)

With

\[ \Sigma = \left[ \begin{array}{cccc} \sigma_{11} & \cdots & \sigma_{1,m+1} \\ \vdots & \ddots & \vdots \\ \sigma_{m+1,1} & \cdots & \sigma_{m+1,m+1} \end{array} \right] = \left[ \begin{array}{cccc} 1 & \cdots & \sigma_{1,m+1} \\ \vdots & \ddots & \vdots \\ \sigma_{m+1,1} & \cdots & 1 \end{array} \right] \]

If the Normal Copula is used in a multivariate normal distribution, then the linear relationship is assumed.

The Archimedean Copula family has tail dependencies that are different from each other, Copula Clayton has tail dependencies at the bottom, Copula Frank does not have tail dependencies, and Copula Gumbel has tail dependencies at the top. Generators of each Copula are presented in Table 1.

| Family       | Generator $\phi(u)$                          | Copula Bivariate $C(u_1, u_2)$                      |
|--------------|----------------------------------------------|----------------------------------------------------|
| Clayton (1978)| $u^{-\theta} - 1, \theta \in (0, \infty)$    | $(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$ |
| Gumbel (1960) | $(-\log(u))^\theta, \theta \in [1, \infty)$   | $\exp\left\{ -[(-\log(u_1))^\theta + (-\log(u_2))^\theta]^{-\frac{1}{\theta}} \right\}$ |
| Frank (1979)  | $\log\left(\frac{e^{\theta u_1} - 1}{e^{\theta} - 1}\right), \theta \in R[0]$ | $\frac{1}{\theta} \log\left(1 + \frac{(e^{\theta u_1} - 1)(e^{\theta u_2} - 1)}{e^\theta - 1}\right)$ |
1.5 Copula Parameter Estimation

Copula parameter estimation is obtained by Maximum Likelihood Estimation (MLE) [11]. The likelihood function of the Archimedean Copula parameter estimation with MLE cannot produce a closed form. As an alternative parameter estimation procedure for Archimedean Copula can use the Tau Kendall approach. Parameter estimates for Archimedean Copula with Tau Kendall approach can be written as follows.

\[ \hat{\tau}_C = 1 + 4 \int_0^1 \frac{\phi(u)}{\phi'(u)} du \]  

Based on the Tau Kendall approach equation for each Copula Clayton, Frank, and Gumbel shown in Table 2.

Table 2. Archimedean Copula Parameter Estimation

| No. | Family  | Estimation \( \hat{\theta} \) |
|-----|---------|-------------------------------|
| 1.  | Clayton | \( \hat{\tau} = \frac{\theta_C}{\theta_C + 2} \) maka \( \hat{\theta}_C = \frac{2\tau}{1-\tau} \) |
| 2.  | Gumbel | \( \hat{\tau} = 1 - \frac{1}{\theta_G} \) maka \( \hat{\theta}_G = \frac{1}{1-\tau} \) |
| 3.  | Frank  | \( \hat{\tau} = 1 - 4(1 - D_1(\theta_F)) / \theta_F \) di mana \( D_k(x) = \text{fungsi Debye} \) |

\[ D_k(x) = \frac{k}{x^k} \int_0^x \frac{u^k}{e^u - 1} du \]

1.6 Gaussian Copula Marginal Regression

The general form of the Gaussian Copula Marginal Regression model is as follows:

\[ Y_i = g(x_i, e_i; \lambda), \quad i = 1, ..., n \]  

Where \( g(.) \) is the appropriate regression function, \( e_i \) is an error of the model, and \( \lambda \) is a parameter. Among the many possibilities \( g(.) \). The model selection is as follows.

\[ Y_i = F_i^{-1}(\phi(e_i); \lambda), \quad i = 1, ..., n \]  

Where \( \phi(.) \) is the cumulative distribution function of \( Y_i \) given \( x_i \), based on the integral transformation theorem, the regression model in equation 2 ensures the marginal distribution of \( Y_i \). For example for Gaussian Linear Models \( Y_i = x_i^T \beta + \sigma e_i \) according to \( Y_i = (Y_i; \lambda) = \phi\left(Y_i - \frac{x_i^T \beta}{\sigma}\right) \) with \( \lambda = (\beta^T \sigma)^T \). When the model uses the Weibull distribution, then \( \mu_i = \exp\left(x_i^T \beta\right) \), with \( \lambda = \tilde{\beta} \) [12].

In a research journal written by Masarotto and Varin (2012), it was discussed about the use of the Gaussian Copula model for marginal regression analysis on data that is not normally correlated. Appropriate model specifications produce simple interpretations of marginal parameters and good flexibility in dependent structures, and have been demonstrated in time-series data, longitudinal studies, spatial data, and survival analysis [11]. In addition, in the research conducted by Dewi Ratih et al (2013) about the relationship between harvested area and rainfall and modelling of rice harvested area using the GCMR method which is also compared with OLS and GLM methods. The results show that the GCMR method is best used to model response variables that are not normally distributed with a large skew. GCMR is also better when compared to the GLM method in handling abnormal response variables [8].

1.7 Previous Researches

Various studies have been conducted in the field of Decision Support Systems (SPK). SPK is very helpful in solving problems in various sectors because it is very easy for the parties to take a decision.
without having to analyse it manually, because the manual analysis process takes a very long time and often causes human error.

Research on the relationship of climate change to agricultural production was previously carried out by Sutikno et al (2013) using the extreme value theory approach. In this study, Ordinary Least Square (OLS) models were used to predict rainfall and then Copula models were obtained to predict harvest area in an area with rainfall data [2]. Then the research was developed by Maulidiyah and Mukhlash (2014) by integrating it into web-based decision support software so that the results can be represented in the form of information that is easily accessible and understood [3]. From this study, Miftachurohmah and Mukhlash (2015) developed it again by adding ENSO indicator and robust regression model with an estimate of M to predict its production area in Decision Support Systems [4]. Furthermore, research on the same thing was developed again by Maziyah and Mukhlash (2017) by using another method, namely robust regression with S-estimation and MM-estimation to connect between harvested area and rainfall in predicting rice production levels that have been successfully integrated into in the existing web-based Decision Support System software [5].

Copula is a statistical method that shows the relationship between variables, where this method is not too strict on distribution assumptions, especially normal distribution. Besides that Copula also has the advantage of being able to clearly describe dependencies at extreme points. Several studies using the Copula approach in the field of climatology, such as those carried out by Scholzel and Friederichs (2008) and Oktaviana (2012) show that the Copula method has a better performance under conditions of normal violations [6][7]. However, these studies are still limited to correlation and not to identifying the causal relationship. The method that can be used to model the causal relationship in extreme events is the Gaussian Copula Marginal Regression (GCMR).

In Indonesia, Dewi Ratih et al. (2013) have done previously about the relationship between harvested area and rainfall and modelling of rice harvested area using the GCMR method which is also compared with OLS and GLM methods. The results show that the GCMR method is best used to model response variables that are not normally distributed with a large skew. GCMR is also better when compared to the GLM method in handling abnormal response variables [8].

2. System Development

2.1. Planning Phase

Based on the research methodology, user requirements on the system made include:

1. Display and change data into predictor variables and response variables.
2. Displaying graphs of variable transformations to uniform domains, calculation results of Pearson & Tau-Kendall correlations, normality tests, and the results of Copula correlation calculations with the help of R software stored in the database
3. Calculate rice production (dry milled grain) and convert it to rice from the model obtained
4. Perform calculations and display the results of estimating the amount of rice inventory and rice consumption needs of each district

2.2. Process Design

The course of the program from the initial data that is ready to be processed until the predicted results of the level of rice production in East Java are illustrated in the following flow diagram.
2.3 Data Preparation
The data used in this study are secondary data from the period 1990-2011 obtained from the Central Statistics Agency (BPS) and the Meteorology, Climatology and Geophysics Agency (BMKG). BPS data is in the form of rice harvest area data and rice productivity data per sub-round. While BMKG data is in the form of rainfall data per month. The selected research locations are districts which are the centers of rice production in East Java, namely Lamongan, Ngawi, Bojonegoro, Banyuwangi, and Jember.

2.4 Identification of Relationships Inter Variables
Data that will be estimated using the regression model must have response variables and predictor variables. In this study rainfall data were used as predictors and harvested area data as a response. Predictor variables and response variables must have a good correlation in order to have a good estimation result. In addition the data also needs to be tested for its normality by using Kolmogorov-Smirnov and Anderson Darling. If one or both methods indicate that the data is normally distributed, the correlation will be calculated using the Gaussian Copula and Archimedean Copula. If the two methods do not show normal distribution, the correlation will be calculated using Archimedean Copula. Estimation of parameters in the Gaussian Copula uses the Pearson approach, while in Archimedean Copula uses the Tau-Kendall approach. Maximum Likelihood Estimation (MLE) is also used for fitting and selecting copula with the best correlation results.

The steps taken to identify the relationship between the two variables are shown by the algorithm as follows.

**Figure 2. Process Model Diagram**
2.5 Correlation Analysis

Calculation of coefficients and p-values is done using R software and the results are stored in the database. Furthermore, a significant p-value is sought, namely p-value < α, with α = 0.05. A significant p-value result with α = 0.05 for 2012 data out sample in each district is presented in Table 3

Table 3. Significant Correlation Result with α = 0.05

| District  | Month | Sub-round | Correlation | Coefficient | p-value |
|-----------|-------|-----------|-------------|-------------|---------|
| Lamongan  | July  | 2         | Tau Kendall | -0.349643   | 0.02493 |
|           | August| 2         | Tau Kendall | -0.327872   | 0.03977 |
| Ngawi     | May   | 2         | Pearson     | 0.512190    | 0.01481 |
|           |       | 2         | Tau Kendall | 0.451194    | 0.00335 |
| Banyuwangi| April | 1         | Pearson     | 0           | -0.04356|
|           | October| 3        | Pearson     | 0.472880    | 0.02624 |
|           | November| 3       | Tau Kendall | 0.411605    | 0.00829 |
|           | November| 3       | Tau Kendall | 0.587793    | 0.00402 |
|           | November| 3       | Tau Kendall | 0.381780    | 0.01305 |
| Bojonegoro| September| 3      | Pearson     | 0.426147    | 0.04797 |
| Jember    | April  | 1         | Pearson     | 0.479861    | 0.02382 |
|           | May    | 2         | Pearson     | 0.425420    | 0.04840 |
|           | September| 3       | Pearson     | 0.498863    | 0.01811 |
|           | October| 3         | Tau Kendall | 0.526092    | 0.0063  |
|           |         | 3         | Tau Kendall | 0.355749    | 0.02071 |
Based on the coefficients in Table 3 shows that the pattern of the relationship between harvested area and rainfall in Lamongan, Ngawi, Banyuwangi, Bojonegoro, and Jember cannot be explained properly using Pearson and Tau Kendall correlations, because in each test it gives different results and the correlation concluded that most of the rice harvested area did not have a close relationship with rainfall.

2.6 Data Normality Test
Data normality test is carried out for decision making using the Copula method which is used to identify correlation of harvest area and rainfall. The method used to test data normality is Kolmogorov-Smirnov and Anderson Darling. Testing is done by using software R.

If at least one variable has normal distribution results, the Copula decision that will be used is the Normal Copula and Archimedean Copula. If there are no variables that have normal distribution results, the Copula decision that will be used is only Archimedean Copula. As an example, the results of the normality of the harvested area and rainfall for normality for Lamongan (2012 data out sample) are obtained.

| Variable | Kolmogorov Smirnov | Anderson Darling |
|----------|-------------------|-----------------|
|          | D     | p-value | Decision | p-value | Decision |
| HA.1     | 0.099161 | >0.150 | Normal | 0.8944 | Normal |
| HA.2     | 0.12932 | >0.150 | Normal | 0.2216 | Normal |
| HA.3     | 0.15023 | >0.150 | Normal | 0.1027 | Normal |
| RF.Jan   | 0.22928 | <0.010 | Non-normal | <0.010 | Non-normal |
| RF.Feb   | 0.17426 | 0.087 | Normal | 0.0111 | Non-normal |
| RF.Mar   | 0.10012 | >0.150 | Normal | 0.6404 | Normal |
| RF.Apr   | 0.10462 | >0.150 | Normal | 0.5744 | Normal |
| RF.May   | 0.1599  | >0.150 | Normal | 0.0439 | Non-normal |
| RF.Jun   | 0.1698  | 0.098 | Normal | 0.0682 | Normal |
| RF.Jul   | 0.27143 | <0.010 | Non-normal | <0.010 | Non-normal |
| RF.Ags   | 0.31407 | <0.010 | Non-normal | <0.010 | Non-normal |
| RF.Sep   | 0.23244 | <0.010 | Non-normal | <0.010 | Non-normal |
| RF.Oct   | 0.21919 | <0.010 | Non-normal | <0.010 | Non-normal |

With:
RF = Rain Fall
HA = Harvest Area

Based on Table 4 most of the harvest area and rainfall variables in Lamongan are declared to be normally distributed. The same thing happened in Ngawi, Banyuwangi, Bojonegoro and Jember. Therefore, all harvested area variables that were correlated with rainfall were decided using the Normal Copula and Archimedean Copula.

2.7 Copula Parameter Estimation
If there is one variable used following the normal distribution, then the Copula Gaussian is also used. In this Copula analysis used are Frank, Clayton, Gumbel, and Gaussian Copula. Parameter calculation θ is done first before calculating Copula parameter estimates.

The results of the calculation of Copula parameter estimates in each harvested sub-round area in five districts have significant results or in other words |Z_hitang| > Z_α with Z_α = Z_{table} = 1.96 or p-value < α with α = 0.05 in sub-round. However, the pattern of relationship between rice harvested areas in five districts and rainfall has a relationship that follows more than one type of Copula. Therefore, Copula fitting or selection of the best Copula using Maximum Likelihood Estimation (MLE) is calculated using R software. The decision on the best relationship pattern of each variable pair is chosen based on fitting results with a significant p-value and the largest likelihood value.
Table 5 shows that most of the correlation results between rice harvest area and rainfall for 2012 sample data out followed Copula Gumbel. The results of the relationship between rice harvest area and rainfall are identified well in sub-round 1, sub-round 2, and sub-round 3 or between January and December.

**Table 5. Selected Copula Parameter in Each District for 2012 Out Sample Data**

| District  | Month  | Copula  |
|-----------|--------|---------|
| Lamongan  | January| Normal  |
|           | April  | Frank   |
|           | May    | Gumbel  |
|           | November| Frank |
|           | December| Gumbel |
| Ngawi     | January| Gumbel  |
|           | April  | Gumbel  |
|           | May    | Normal  |
|           | July   | Frank   |
|           | August | Gumbel  |
|           | September| Gumbel |
|           | October| Gumbel  |
|           | November| Frank |
| Banyuwangi| January| Gumbel  |
|           | May    | Frank   |
|           | September| Gumbel |
|           | October| Normal  |
|           | November| Normal |
|           | December| Gumbel |
| Bojonegoro| January| Frank   |
|           | May    | Gumbel  |
|           | June   | Normal  |
|           | July   | Gumbel  |
|           | September| Gumbel |
|           | October| Normal  |
|           | November| Frank |
| Jember    | January| Frank   |
|           | February| Frank |
|           | March  | Gumbel  |
|           | April  | Gumbel  |
|           | May    | Normal  |
|           | June   | Frank   |
|           | July   | Frank   |
|           | August | Gumbel  |
|           | September| Frank |
|           | October| Gumbel  |
|           | November| Frank |

2.8 Harvest Area Model Estimation
The method that will be used to estimate the parameters of the harvest area model based on rainfall in this study is the Gaussian Copula Marginal Regression (GCMR). Calculation of harvest area model parameter estimates is done using R software and then the results are stored in the database. The model formed consists of three models according to the sub-round that has been arranged as follows.
\[ H_A^1 = (RF_{Jan}, RF_{Feb}, RF_{Mar}, RF_{Apr}) = b_0 + b_1 RF_{Jan} + b_2 RF_{Feb} + b_3 RF_{Mar} + b_4 RF_{Apr} \] (5)

\[ H_A^2 = (RF_{Mei}, RF_{Jun}, RF_{Jul}, RF_{Agst}) = b_0 + b_1 RF_{Mei} + b_2 RF_{Jun} + b_3 RF_{Jul} + b_4 RF_{Agst} \] (6)

\[ H_A^3 = (RF_{Sept}, RF_{Okt}, RF_{Nov}, RF_{Des}) = b_0 + b_1 RF_{Sept} + b_2 RF_{Okt} + b_3 RF_{Nov} + b_4 RF_{Des} \] (7)

**Figure 4.** Process Diagram of Best Harvest Area Model Selection

The steps to obtain a harvest area model are depicted on Figure 4 as follows:

a. Take the data from the estimation of the harvest area model parameters in the database and arrange a temporary harvest area model

b. Perform a partial test of the harvest area model by getting the standard error value and t test value

c. Select significant parameter estimation values with the provisions of t count > t table

d. Reorder the rice harvest area model using significant parameter values to get the best harvest area model

Modelling of rice harvest area in the production centres in East Java was carried out using the GCMR method. The results of the estimation of harvest area model parameters using GCMR for five districts (2012 data out sample) can be seen in Table 6.

**Table 6. Marginal Coefficient Estimation Model**

| District   | Subround I | Subround II |
|------------|------------|-------------|
|            | Intercept  | RF.Jan      | RF.Feb      | RF.Mar     | RF.Apr     | Intercept  | RF.May    | RF.Jun    | RF.Jul    | RF.Ags     |
| Lamongan   | 64019,094  | 5,680       | -           | -          | 6,909      | 67062,390  | -3,952    | -         | -         | -          |
| Ngawi      | 39313,023  | 2,200       | -           | -          | 6,780      | 36140,729  | 17,844    | -11,93    | -32,741   | -          |
| Banyuwangi | 46014,43   | 23,60       | -           | -          | -          | 27376,037  | 9,646     | -         | -         | -          |
| Bojonegoro | 61873,06   | 12,88       | -           | -          | -          | 15599,26   | 59,75     | -11,93    | -32,741   | -          |
| Jember     | 64792,805  | 4,781       | 3,594       | 1,069      | 19,237     | 45314,56   | 37,79     | 30,14     | -30,07    | 45,67      |
Based on the Copula parameters that have been selected from each district, then the estimated harvest area estimation model for each sub-round in each district is a function of rainfall \( f(CH) \). For the following example the results of the best harvest area estimation model for Lamongan:

\[
\hat{H}_1 = 64019.094 + 5.680CH.Jan + 6.909CH.Apr
\]

\[
\hat{H}_2 = 67062.390 - 3.952CH.May
\]

\[
\hat{H}_3 = 12415.746 - 10.032CH.Nov + 4.576CH.Dec
\]

2.9 Best Model Selection

After modelling the rice harvest area in each district, then immediately followed by partial testing of the model of rice harvest area to determine the variables that significantly influence using the t test. The following is an example of a partial test of the harvest area model for Lamongan (2012 out sample data).

| Parameter | Estimation | s.e | t  | t_{19,0.1/2} |
|-----------|------------|-----|----|--------------|
| Intercep  | 64019.094  | 2621.628 | 24,420 | 1.729 |
| January   | 5.680      | 6.567 | 0.865 |         |
| April     | 6.909      | 14.933 | 0.463 |         |
| Intercep  | 67062.390  | 1701.373 | 39,541 | 1.725 |
| Mei       | -3.952     | 13.839 | -0.283 |         |
| Intercep  | 12415.746  | 3500.369 | 3,547  | 1.729 |
| November  | -10.032    | 15.570 | -0.644 |         |
| Desember  | 4.576      | 9.292  | 0.492  |         |

Table 7 explains that rainfall does not significantly affect the rice harvest area. This is indicated by the absence of a significant rainfall variable in several sub-rounds in each district, because \( t_{\text{count}} < t \) table. That is, only intercepts enter the model. Following is the best rice harvest area model in Lamongan (Ha):

1. \( \hat{H}_1 = 64019.094 \)
2. \( \hat{H}_2 = 67062.390 \)
3. \( \hat{H}_3 = 12415.746 \)

Table 8. Modelling Error in Data from Sample Harvesting Area Based on the Best Model of Lamongan 2012

| e_1  | e_2  | e_3  |
|------|------|------|
| 5203.9 | 22653.4 | 281.3 |

Based on the prediction results, it can be calculated that the estimated yield of rice harvest area in Lamongan in 2012 is shown in Table 8. Errors from the assessment are still very high (thousands and
even tens of thousands of hectares). This is because the observation data in the study are few, therefore the model obtained for the sample data out gives a fairly large error.

2.10 Prediction of Rice Production

Calculation of rice production is carried out based on the results of the calculation of the prediction of harvest area in accordance with the best harvest area model that has been prepared previously. As explained, the yield of rice is equivalent to the yield of dry milled grain (MPD).

The prediction of the need or consumption of rice in Lamongan, Ngawi, Banyuwangi, Bojonegoro, and Jember has mostly increased every year. This is because the rate of population growth in the five districts is positive, resulting in an increase in the population in each year.

If the number of predictions of rice production or rice supply is less than the need for a deficit, it can be said that the production level is low. If the amount of rice supplies is more than the need then there is a surplus so that the level of production is high. The results of the prediction of rice production in each district mostly produce high levels of production, so it is still safe to meet the consumption needs of the community's rice. This is because the five districts are the centres of rice production in East Java.

2.11 Accuracy of Method Measurement

In reality there are no predictions that have a 100% accuracy rate, because every prediction must contain errors. In this study, the MAPE (Mean Absolute Percentage Error) method is used to measure the accuracy of the harvest area estimation results using GCMR. MAPE is calculated using absolute errors in each period divided by real observation values for that period, then averaging the absolute percentage errors. MAPE indicates how much errors in fortune telling are compared to real values.

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \times 100\%
\]

Where:

- \(A_t\) = actual value at time \(t\)
- \(F_t\) = predictive value at \(t\)
- \(n\) = amount of data

The following measures of accuracy for each district (out sample data 2012):

| District    | Sub-round | HA Actual (Ha) | HA Estimation (Ha) | Error   |
|-------------|-----------|----------------|--------------------|---------|
| Lamongan    | 1         | 69223          | 64019.1            | 0.075176|
|             | 2         | 61229          | 38575.6            | 0.369978|
|             | 3         | 12697          | 12415.7            | 0.022155|
| Ngawi       | 1         | 46093          | 39313              | 0.147094|
|             | 2         | 45377          | 9059.19            | 0.800357|
|             | 3         | 24791          | 13649.5            | 0.449417|
| Banyuwangi  | 1         | 49435          | 52522.1            | 0.06245 |
|             | 2         | 37962          | 27376              | 0.278858|
|             | 3         | 30789          | 27888.5            | 0.094206|
| Bojonegoro  | 1         | 75828          | 61873.1            | 0.184034|
|             | 2         | 48912          | 41519.5            | 0.151139|
|             | 3         | 9094           | 6017.4             | 0.338311|
| Jember      | 1         | 76160          | 74531.5            | 0.021383|
|             | 2         | 61873          | 52927.3            | 0.144582|
|             | 3         | 20535          | 18187.9            | 0.114298|
| Error Total |           |                |                    | 3.128538|
| MAPE        |           |                |                    | 20.8569% |

\(\text{MAPE} = \frac{\text{Error Total}}{n} \times 100\%\)
Based on the above calculation, it was found that the error of using the GCMR method for modeling the rice harvest area of the five districts in each period for the 2012 sample data out was 20.8569%.

3. Conclusion and Future Work

3.1 Conclusion
Based on the series of processes carried out as described in the previous chapter, the following conclusions are obtained:

1. The steps in modelling the rice harvest area using GCMR consist of several stages. The first stage is to identify the relationship between predictor variables (rainfall) and response variables (harvest area) by knowing the correlation value, test for normality, and then calculate and determine a significant Copula parameter estimate. After obtaining the selected Copula parameters, the second stage is followed by compiling a model of the estimated area of harvest area for each sub-round in each district which is a function of rainfall \( f(CH) \). After modelling the rice harvest area, continued with partial testing of the model of rice harvest area to determine the variables that significantly influence using the t test. After obtaining significant variables, the harvest area estimation model was rearranged so that the best results could be obtained.

2. Development of Decision Support Systems made has successfully predicted the level of rice production in 5 districts in each sub-round. The results show that most predictions of rice production levels in each district in each sub-round belong to the high level, some in the medium level, and none are classified as low levels. This is because the five districts are rice producing centres in East Java.

3. Development of Decision Support Systems that have been successfully integrated into the existing web-based decision support system software by adding several tables in the database and adding a new model, namely Gaussian Copula Marginal Regression (GCMR) and combining menus on the software.

3.2 Future Work

1. The harvest area model can be developed by adding other predictor variables that influence response variables, such as cropping patterns, irrigation systems, etc. in order to obtain a better harvest area model.

2. For further research it is recommended that the number of samples used be even greater so that the resulting pattern is clearer in describing the actual situation.

3. Can determine the estimated harvest area parameters automatically so that when the user updates the input data, the system can immediately adjust without having to update the estimated harvest area parameters.

Acknowledgments
We would like to thank the Ministry of Research, Technology and Higher Education of Republic of Indonesia for supporting this research.

References
[1] Suwandi et al 2015 Outlook Komoditas Pertanian Subsektor Tanaman Pangan : Padi. Jakarta : Data Centre and Ministry of Agriculture's agricultural information system
[2] Sutikno, Setiawan, and Imam M 2013 Pengembangan Teknologi Adaptasi Produksi Pertanian terhadap Perubahan Iklim Melalui Pendekatan Extreme Value Theory. 2nd Year National Strategic Research Final Report of the 3-Year Plan. Surabaya: Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) ITS.
[3] Imam Mukhlash et al 2017 J. Phys.: Conf. Ser. 890 012143
[4] Nisa M, Imam M, and Sutikno 2015 Design to Risk Analysis of Paddy Production With ENSO Indicator (Case-Study: East Java). Proceeding of International Conference On Research, Implementation and Education of Mathematics and Science 2015. Yogyakarta State University.
[5] Ais M M 2017 Penggunaan Regresi Robust dengan Estimasi-S dan Estimasi-MM dalam Pengembangan Sistem Pendukung Keputusan Guna Memprediksi Tingkat Produksi Padi. Undergraduate Theses : Institut Teknologi Sepuluh Nopember (ITS).
[6] Scholzel C and Friederichs P 2008 Multivariate Non-Normally Distributed Random Variables in Climate Research – Introduction to the Copula Approach. Nonlin. Processes Geophys., 15, 761-772
[7] Oktaviana PP 2012 Pendekatan Copula untuk Penyusunan Peta Kerawanan Pusat Produksi Padi di Jawa Timur dengan Indikator El-Nino Southern Oscillation (ENSO). Undergraduate Theses. Surabaya : Institut Teknologi Sepuluh Nopember (ITS)
[8] Sutikno, Heri K, Iis D R 2014 Gaussian Copula Marginal Regression For Modeling Extreme Data With Application, Journal of Mathematics and Statistics 10 (2) : 192-200
[9] Tjasyono, Bayong HK 2006 Klimatologi, ITB Publisher, Bandung
[10] Turban E and Aronson J E 2001 Decision Support Systems and Intelligent Systems. 6th edition. Prentice Hall: Upper Saddle River, NJ.
[11] Mikosch T 2008 Copula: Tales & Facts. Copenhagen Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken S, DK 2100
[12] Masarotto G and Varin C 2012 Gaussian Copula Marginal Regression. Electronic Journal of Statistics, Vol. 6 (1517 – 1549