GAUSSIAN COPULA MARGINAL REGRESSION FOR MODELING EXTREME DATA WITH APPLICATION

Sutikno, Heri Kuswanto and Iis Dewi Ratih

Department of Statistics, Faculty of Mathematics and Natural Sciences,
Institut Teknologi Sepuluh Nopember Surabaya, Kampus ITS Sukolilo, Surabaya, 60111, Indonesia

ABSTRACT

Regression is commonly used to determine the relationship between the response variable and the predictor variable, where the parameters are estimated by Ordinary Least Square (OLS). This method can be used with an assumption that residuals are normally distributed $(0, \sigma^2)$. However, the assumption of normality of the data is often violated due to extreme observations, which are often found in the climate data. Modeling of rice harvested area with rainfall predictor variables allows extreme observations. Therefore, another approximation is necessary to be applied in order to overcome the presence of extreme observations. The method used to solve this problem is a Gaussian Copula Marginal Regression (GCMR), the regression-based Copula. As a case study, the method is applied to model rice harvested area of rice production centers in East Java, Indonesia, covering District: Banyuwangi, Lamongan, Bojonegoro, Ngawi and Jember. Copula is chosen because this method is not strict against the assumption distribution, especially the normal distribution. Moreover, this method can describe dependency on extreme point clearly. The GCMR performance will be compared with OLS and Generalized Linear Models (GLM). The identification result of the dependencies structure between the Rice Harvest per period (RH) and monthly rainfall showed a dependency in all areas of research. It is shown that the real test copula type mostly follows the Gumbel distribution. While the comparison of the model goodness for rice harvested area in the modeling showed that the method used to model the exact GCMR in five districts RH1 and RH2 in Jember district since its lowest AICc. Looking at the data distribution pattern of response variables, it can be concluded that the GCMR good for modeling the response variable that is not normally distributed and tend to have a large skew.

Keywords: Copula, Archimedean, Gaussian, Dependencies, Extreme

1. INTRODUCTION

The method used to determine the pattern of the relationship between variables is correlation analysis and regression analysis. Correlation analysis that is frequently used is the Pearson correlation. Parameter estimation methods in regression analysis that commonly used is Ordinary Least Squares (OLS). Both of these methods (Pearson correlation and OLS) can be used well if it satisfies the assumption of normal distribution of data. Data normality assumption is often violated if the data there exists an extreme.

In the development, the method that can be used when the response is not normally distributed is Generalized Linear Model (GLM). Requirements that must be satisfied for this method are the relationship between the predictor variables are linear and the distribution of the response variable should be Exponential family members. Distributions that frequently used are the binomial, Poisson, negative binomial, normal, gamma, inverse Gaussian and lognormal (McCullagh and Nelder, 1989). In addition, GLM.

If the pattern or curve regression relationships between predictor variables and the response is not known, the method is used a nonparametric approach.

Corresponding Author: Sutikno, Department of Statistics, Faculty of Mathematics and Natural Sciences,
Institut Teknologi Sepuluh Nopember Surabaya, Kampus ITS Sukolilo, Surabaya, 60111, Indonesia

Embrechts *et al.* (2002) suggests using copula approach to address violations of the assumption of normality of the data. Copula is a statistical method that shows the relationship between variables, where the method is not too strict on the assumption of distribution, in particular the normal distribution. Copula excellence can also describe dependencies on extreme points clearly. In the recent years, the copula has been widely used to model the structure of the relationship at risk management (Villarini *et al.*, 2008; Embrechts *et al.*, 2001), climatology and meteorology (Vreac *et al.*, 2005; Schölzel and Friedrich, 2008; De Michele and Salvadori, 2003) and in other areas.

The results showed that copula method has better performance in conditions of normality assumptions violated. However, previous research is still limited to a correlation, not identified until the relationship causality. The method that can be used to model causality in extreme events is Gaussian Copula Marginal Regression (GCMR). This method has been used by Masarotto and Varin (2012).

Natural events such as climate often erratic over time, thus causing extreme climate (Bekti, 2009). Phenomenon of nature in the form of extreme weather events is one of the problems that are difficult to address in the agricultural sector. Currently, the rainfall pattern is erratic causing a significant drop in national rice production. Climate does not directly affect the production of rice, but the rice harvested area. Therefore, the need for information about the rice harvested area forecast the future as part of efforts to support food security.

Some researches on rice production involving climate indicator has been done in recent years. Regression modeling anomalous area harvested per period and weighted rainfall index estimated with OLS produces small R^2 due to outlier observations (Miconnet *et al.*, 2005; Sutikno *et al.*, 2010). Other studies using indicator variables El-Nino Southern Oscillation (ENSO) are analyzed by simple regression analysis (Bekti, 2009). This method has not been properly interpreted because the data analyzed do not meet the assumptions of normality of the data due to the extremes.

Therefore, modeling the effect rainfall for the area of the rice harvest using the Gaussian Copula Marginal Regression is necessarily to be carried out. This method is expected to be able to properly model the rice harvested area affected by extreme climate in central production in East Java rice, namely Lamongan, Bojonegoro, Ngawi, Jember and Banyuwangi.

2. MATERIALS AND METHODS

The data used in this study is a secondary data obtained from the Central Statistical Agency and the Board Meteorology, Climatology and Geophsycs. Central Statistical Agency data including data per subround rice harvested area (area harvested first period (January-April), II (May-August) and III (September-December) as the response variable. Board Meteorology,Climatology and Geophsycs data is the data of rainfall per month (January-December) used as the predictor variable spanning from 1990 to 2010. The data are collected from center of rice production in East Java, namely: Banyuwangi, Lamongan, Bojonegoro, Ngawi and Jember, as **Fig. 1**.

Stages of the data analysis are described as follows:

- Create a scatterplot between variables X and Y to identify patterns of linear and nolinear two variables
- Calculate the correlation between variables X and Y with Pearson correlation and Kendall-Tau
- Test the normality of the data by Anderson Darling test
- Create Scatter Plot transformation results [0,1] between the variables X and Y
- Estimate the Copula parameter using Tau Kendall approach
- Fitting Copula with Maximum Likelihood Ratio
- Create a scatterplot rank Copula and Copula choose the best
- Create a regression model based Copula selected by OLS, GLM and GCMR
- Test models using AIC_C goodness in each model
- Getting the best regression model with the smallest AIC_C

The wide of Rice Harvested period 1 (RH1) was harvested area for rice harvested in late April, wide of Rice Harvested period 2 (RH2) was harvested area of rice harvested in late August. While the vast Rice Harvest period 3 (RH3) were harvested area of rice harvested in late December.

Formulation of the regression model is:

$$RH_{p(i)} = \beta_0 + \beta_1 \, RF_{1i} + \beta_2 \, RF_{2i} + \beta_3 \, RF_{3i} + \beta_4 \, RF_{4i} + \varepsilon_i$$

With i = 1, 2, 3, ... n (size of observations) and p = 1, 2, 3 (p is the period). RF1, RF2, RF3 and RF4 denote the rainfall in the first to forth in each period. The data is analyzed by software R.

2.1. Copula

If there is a random vector $(X_1, X_2, …X_m, Y)$ has the distribution cumulative marginal function $F_{X1}, F_{X2},..,F_{Xm},$ F_Y with the R domain which is not going down, $F_{Xi} \, (-\infty)$ $= F_Y \, (-\infty) = 0$ and $F_{Xi} \, (\infty) = F_Y \, (\infty) = 1$, then its distribution could be shown at first Equation 1:

$$F(X_1, X_2,.., X_m, Y)(x_1, x_2,..., y) =$$
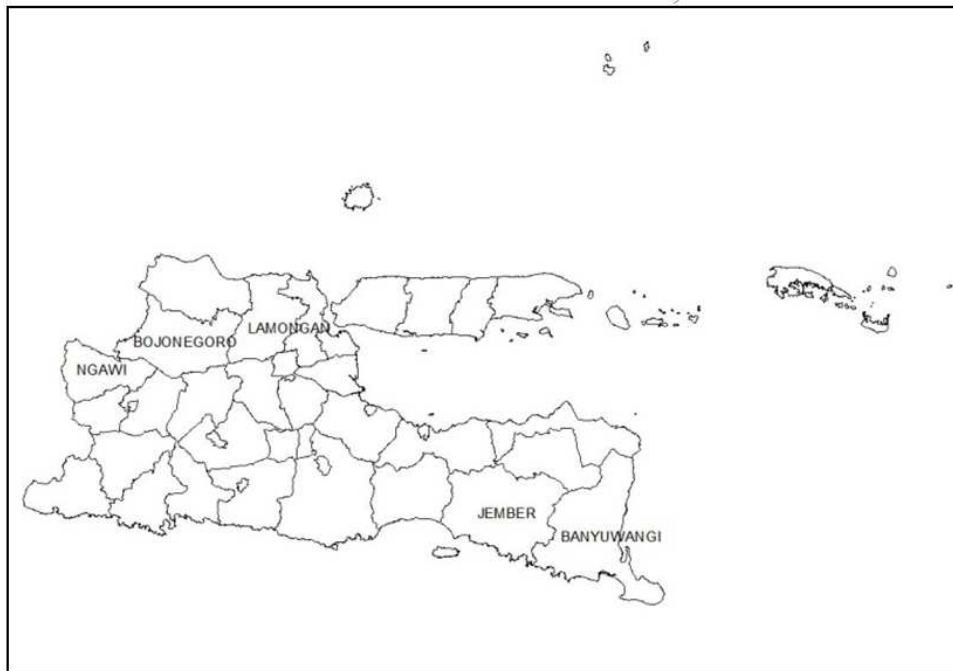$$C(X_1, X_2,.., X_m, Y)(F_{X1}(x_1), F_{X2}(x_2),..., F_{Xm}(x_m), F_Y(y)) \quad (1)$$

Fig. 1. Five locations of the research regency centers rice: Jember, Banyuwangi, Ngawi, Lamongan and Bojonegoro

C_x : $[0,1]x...x[0,1] \rightarrow [0,1]$ is Copula. If the marginal function marginal from F_{Xj} (X_j) is continue, then C_x is unique (Sklar, 1959) and could be shown as in Equation 2:

$$C(X_1, X_2,..,X_m, Y)(u_1,..,u_m, w) =$$
$$\int_0^{u_1} ... \int_0^{u_m} \int_0^{w} c(X_1, X_2,..,X_m, Y)(u_1,..,u_m, w)du_1..du_m dw \quad (2)$$

2.2. Copula Family

Two of the most popular Copula family is Archimedean Copula and Copula ellipse. Elliptical Copula consists Normal Copula and Copula-t. While Archimedean Copula consists of Clayton, Gumbel and Frank.

2.3. Gaussian Copula

Gaussian copula or Normal Copula can be obtained from transformation from the random variable to standard normal distribution. Copula Gaussian's function can be written as follows Equation 3:

$$C(u_0, u_1...u_m) =$$
$$F_{N(0,\Sigma)}\left(F_{N(0,1)}^{-1}(u_0), F_{N(0,1)}^{-1}(u_1),..., F_{N(0,1)}^{-1}(u_m)\right) \quad (3)$$

If normal copula is used to the multivariate normal distribution then it assumes linear relationship (Schölzel and Friederichs, 2008).

2.4. Archimedean Copula

Copula Archimedean family has tail dependency which is different each other, Clayton Copula has tail dependency in lower area, Frank Copula has not tail dependency and Gumbel Copula has tail dependency in upper area (**Fig. 2**). The generator of each copula is shown on **Table 1**.

2.5. Transformation of Random Variables

Marginal distribution from the random variables X and Y which is unknown is shown as in the Equation 4 respectively:

$$F_{Xi}(x_i) = \frac{1}{n+1}\sum_{j=1}^{n} 1\left(X_i^{(j)} \leq x\right); x_i \in R$$
$$F_{yi}(y) = \frac{1}{n+1}\sum_{j=1}^{n} 1\left(Y_i^{(j)} \leq y\right); y \in R \quad (4)$$

The data transformation to the uniform domain can be done with making the scatter plot [0,1] and form the rank plot for X and Y as shown in Equation 5:

$$\left(\left(\frac{R_{1i}^{(j)}}{n+1} \right), \left(\frac{R_2^{(j)}}{n+1} \right) \right), 1 \le j \le \text{and } i = 1, 2, ..., m \qquad (5)$$

Referring to the transformation, the Copula equation can be given as follows (Genest and Nešlehová, 2010) Equation 6:

$$C_n(u_i, w) = \frac{1}{n} \sum_{j=1}^{n} 1 \left(\frac{R_{1i}^{(j)}}{n+1} \le u_i, \frac{R_2^{(j)}}{n+1} \le w \right), u_i, w \in (0,1) \qquad (6)$$

2.6. Parameter Estimate

Parameter Estimation for Archimedean Copula can be done by using Tau Kendall's approach, can be written as follows Equation 7:

$$\hat{\tau}_C = 1 + 4 \int_0^1 \frac{\phi(\mu)}{\phi'(\mu)} du \qquad (7)$$

Tau Kendall's approach for each Copula Clayton, Frank and Gumbel being shown on **Table 2**.

2.7. Ordinary Least Square

One of the estimation procedure for linear regression models is the ordinary least squares procedure. The concept of this method is to estimate the regression coefficient (β) to minimize sum square error, so that the estimators for β can be formulated as follows (Draper and Smith, 1981) Equation 8:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \qquad (8)$$

2.8. Generalized Linear Model (GLM)

Development of classical linear models with response variables is not normally distributed. GLM has three components, namely: Random component, Systematic component and Link function (McCullagh and Nelder, 1989; Agresti, 2007) Equation 9:

$$g(\mu) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k \qquad (9)$$

Function g (.) is a link function that links the random and systematic components. The parameter estimation of β used the maximum likelihood method.
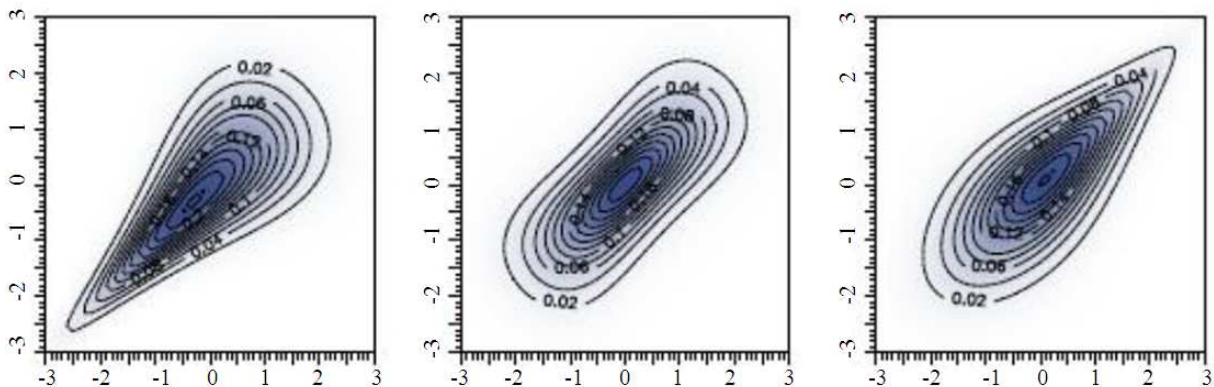

Fig. 2. Plot Copula Archimedean for Clayton, Frank, Gumbel [left to right] (Source: Schölzel and Friederichs, 2008)

Table 1. Generator archimedean copula

Family	Generator $\phi(u)$	Bivariate copula $C(u_1, u_2)$
Clayton	$\dfrac{u^{-\theta} - 1}{\theta}, \theta \in (0, \infty)$	$\left(u_1^{-\theta} + u_2^{-\theta} - 1 \right) \dfrac{1}{\theta}$
Gumbel	$\left(-\log(u) \right)^\theta, \theta \in [1, \infty)$	$\exp\left\{ -\left[\left(-\log(u_1) \right)^\theta + \left(-\log(u_2) \right)^\theta \right]^{\frac{1}{\theta}} \right\}$
Frank	$\log\left(\dfrac{e^{\theta t} - 1}{e^\theta - 1} \right), \theta \in R / \{0\}$	$\dfrac{1}{\theta} \log\left(\dfrac{e^{\theta u_1} - 1 \left(e^{\theta u_2} - 1 \right)}{e^\theta - 1} \right)$

Table 2. Estimation parameter archimedean copula

Family	Generator ϕ (u)
Clayton	$\hat{\tau} = \dfrac{\theta_C}{\theta_C + 2}$ then $\hat{\theta}_C = \dfrac{2\tau}{1-\tau}$
Gumbel	$\hat{\tau} = 1 - \dfrac{1}{\theta_G}$ then $\hat{\theta}_G = \dfrac{1}{1-\tau}$
Frank	$\hat{\tau} = 1 - 4(1 - D_1(\theta_F)) / \theta_F$
	where $D_K(x)$ = the Debye function
	$D_k(x) = \dfrac{k}{x^k} \displaystyle\int_0^x \dfrac{u^k}{e^u - 1} du$

2.9. Gaussian Copula Marginal Regression

Common form of Gaussian copula marginal regression models is as follows Equation 10:

$$Y_i \, g(x_i, e_i; \vec{\lambda}), i = 1, .., n \qquad (10)$$

where, g (.) is the corresponding function, i.e., the error models and $\bar{\lambda}$ is a parameter. Among the many possible g (.), the selection of the model is as follows Equation 11-13:

$$Y_i = F_i^{-1}\left\{\phi(e_i); \vec{\lambda}\right\}, i = 1, ...n \qquad (11)$$

where, Φ (.) is the cumulative distribution function of Yi given xi. When the model using Weibull distribution, then $\mu_i = \exp(x_i^{\top}\vec{\beta})$ the $\bar{\lambda} = \vec{\beta}$ (Masarotto and Varin, 2012).

2.10. Akaike Information Criteria Corrected

One of the frequently used information criteria are AIC:

$$AIC = -2\log L(k) + 2k \qquad (12)$$

Where:
L(k) = The likelihood function and
k   = The number of sample is relatively small parameters

If n/k <40, the criteria used is the AIC_C (Hu, 2007):

$$AIC_C = AIC + \frac{2k(k+1)}{n-k-1} \qquad (13)$$

3. RESULTS

3.1. Data Exploration and Patterns Identification of Relationships Between the Response Variable and the Predictor

In the following discussion we present in detail the fact observed in one regency, namely Banyuwangi regency. The discussion for other areas is presented in the summary.

The relationships between rainfall and rice harvested area in Banyuwangi does not show a clear pattern, though there are several adjacent points that indicate a relationship between rice harvested area and rainfall, as presented in **Fig. 3**. In addition, pearson correlation and Kendall Tau cannot explain the relationship well because each test gives different results. The correlation results concluded that most of the rice harvested area do not have a close relationship with rainfall (**Table 3**). The unclear relationship between the two variables is alleged absence of data extreme observations (outliers).

In addition, the result of normality distribution test using Anderson Darling shows that most of the data do not follow normal distribution, such as RH1, RH3, April rainfall, rainfall from June to December. Only the RH2, rainfall in January, February, March and April follows normal distribution. Therefore, further analyzes on the dependencies is carried out using copula approach in order to specifically look at the model dependencies.

Table 4 presents the Copula parameter estimation by Tau-Kendall approach. There are several variables that one Copula (Gumbel) can not estimate. This is due to the value of $\theta < 1$. Copula has selected value based on the value of the type of copula that yields on largest loglikelihood. Visually we can see the dependencies between two variables as shown on each rankplot Copula in **Fig. 4**.

Figure 4 shows that the dependencies between variables in particular of rice harvested area and rainfall. For instance, the relationship between RH3 with rainfall in October shows that the tail dependencies exists, which is the characteristic of the Clayton Copula. We can see that there are extreme points in minimum area and when rainfall fell in October, the rice harvested area will also decrease. As for the plot which has a tail dependencies above, for example, is the relationship between rainfall and January with RH1 (following the Gumbel copula) showed that the higher rainfall in January, RH1 will be growing as well. While the relationship between rainfall in the month of May with RH2 follows the Normal Copula and the type of the relationship between both variables is linear.

Using the similar procedure we identify the dependency structure between rice harvested area with rainfall in the four districts of rice production centers in the other East Java in **Table 5**.

The result of the identification of the dependency structure between the rice harvest per period and monthly rainfall showed a dependency in all areas of research. It is shown that the real test copula type mostly follow the Gumbel distribution. This phenomenon illustrates that the rice harvested area is very dependent on rainfall, especially in the 3rd period.
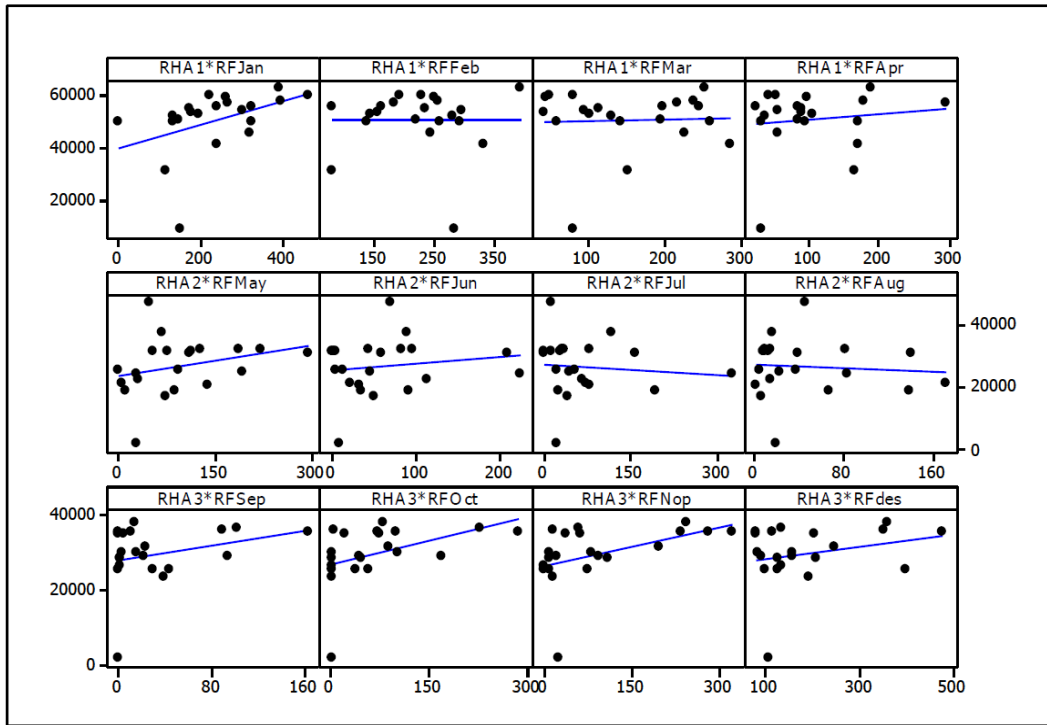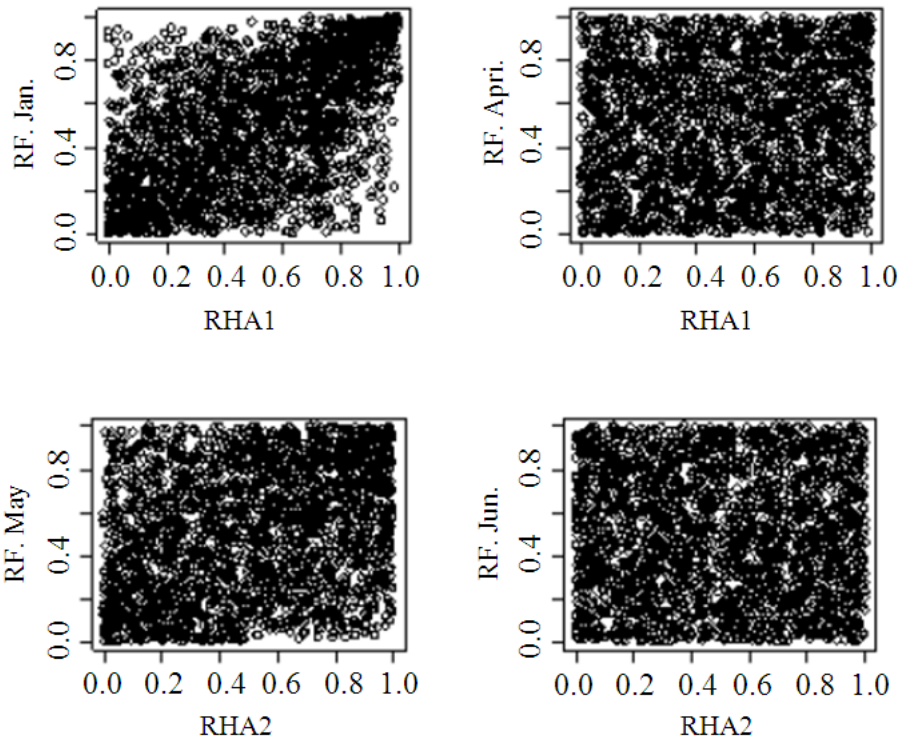
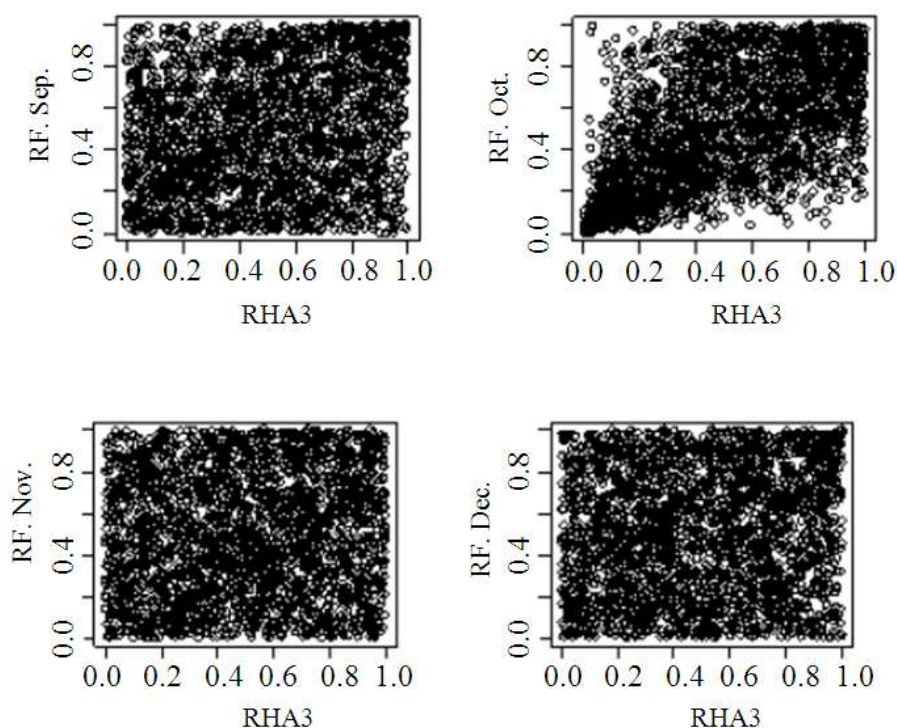**Fig. 3.** Scatterplot between the area of the rice harvest and rainfall

Fig. 4. Rank scatter plot Copula generation of data (n = 2500) to Banyuwangi

Table 3. Pearson correlation, Tau Kendall between The Area of the rice harvest and Rainfall in Banyuwangi

Variables	Coefisien correlation	
	Pearson	Tau-Kendall
RH1 and rainfall January	0,423	0,400
RH1and rainfall February	0,006	-0,029
RH1and rainfall March	0,042	-0,076
RH1and Rainfall April	0,136	0,057
RH2 and rainfall May	0,282	0,248
RH2 and rainfall June	0,158	0,010
RH2 and rainfall July	-0,101	-0,138
RH2 and rainfall August	-0,087	-0,043
RH3 and rainfall September	0,273	0,150
RH3 and rainfall October	0,423	0,410
RH3 and rainfall November	0,451	0,364
RH3 and rainfall December	0,252	0,076

3.2. Selection of the Best Model

Lowest AIC_C between three methods for modeling RH1 in five regency are GCMR method. While modeling RH1 and RH3 in Banyuwangi and Ngawi is more appropriate using OLS, because its AIC_C lowest value, nor the modeling RH2 in Lamongan. While the GLM method is more appropriate to model in Bojonegoro RH2 and RH3 in Bojonegoro, Jember and Lamongan as presented in **Table 6.**

If rainfall happens in January (x+1) mm, then the RH1 tended to rise by 1.0005 ha times than when rainfall x mm. Meanwhile, the rainfall in May rose by 1 mm, then RH2 tend to increase by 30.09 ha and if the rainfall in June rose 1 mm, then the rice harvested area tends to increase by 11.51 ha. Based on this it can be concluded that most affect rainfall RH2 increase in Banyuwangi is rainfall in May because it provides the most substantial change to the RH2 in Banyuwangi.

Similar as in Banyuwangi, RH3 models indicates that precipitation increased by 1 mm 2 in September, will likely reduce the amount of rice harvested area of 9.029 ha. Meanwhile, if the rainfall in October rose 1 mm, then the rice harvested area tends to increase by 36.33 ha. If the rainfall in November rose 1 mm, it tends to grow rice harvested area of 16.486 ha and if the rainfall in December rose by 1 mm, the harvested area of 13.179 ha of rice tends to increase. More regression models for each district is presented in **Table 7**.

Table 4. Copula Parameter using Tau Kendall approach and log likelihood value for Banyuwangi

	Type of copula			
Variables	Clayton	Gumbel	Frank	Normal
RH1 and rainfall	1,333	1,667	4,161	0,588
January	0,913	4,308*	3,549	4,127
RH1and rainfall	-0,056	-	-0,257	-0,045
February	-0,071	-	-0,502	0,015
RH1and rainfall	-0,142	-	-0,689	-0,119
March	-0,214	-	-0,934	-0,179
RH1and rainfall	0,121	1,061	0,516	-
April	0,199	0,618*	0,857	-
RH2 and rainfall	0,658	1,329	2,347	0,379
May	0,642	0,861	2,374	1,305*
RH2 and rainfall	0,019	1,010	0,086	0,015
June	-0,103	0,001*	1,010	0,366
RH2 and rainfall	-0,243	-	-1,266	-0,216
July	-0,231	-	-1,315	-0,261
RH2 and rainfall	-0,082	-	-0,387	-0,067
August	0,082	-	-0,603	-0,064
RH3 and rainfall	0,353	1,177	1,377	-
September	0,501	1,148*	1,707	-
RH3 and rainfall	1,389	1,695	4,299	-
October	4,619*	4,072	4,353	-
RH3 and rainfall	1,143	1,571	3,680	-
November	0,823	2,808	2,918*	-
RH3 and rainfall	0,165	1,082	0,689	-
December	-0,206	0,354*	0,731	-

Notes: Bold indicates that significant at α = 0.05, figures in italics are the likelihood
* = Copula elected with the biggest Loglikelihood
- = Can not estimate

Table 5. Copula Parameter between the area of the rice harvest and rainfall in five regency of Centra production in East Java

	Period I			
District	January	February	March	April
Banyuwangi	1,667[A]	-	-	1,061[A]
Bojonegoro	1,221[A]	1,061[A]	-	-
Ngawi	-	1,019[A]	1,082[A]	1,470[A]
Jember	1,221[A]	-	-	1,419[A]
Lamongan	1,154[A]	-	-	1,050[A]
Period II				
District	May	June	July	August
Banyuwangi	0,379[D]	1,010[A]	-	-
Bojonegoro	1,438[A]	3,087[C]	1,374[A]	1,079[A]
Ngawi	0,669[D]	-	1,975[C]	1,051[A]
Jember	0,588[D]	0,414[D]	1,222[A]	-
Lamongan	1,094[A]	-	-	1,030[A]
Period III				
District	Sept	October	November	December
Banyuwangi	1,177[A]	1,389[B]	3,680[C]	1,082[A]
Bojonegoro	1,093[A]	0,434[D]	1,129[A]	1,061[A]
Ngawi	1,093[A]	0,546[D]	1,296[A]	-
Jember	6,898[C]	1,346[A]	1,296[A]	-
Lamongan	-	1,129[A]	1,130[A]	-

Note: A = Following Gumbel Copula
B = Following Clayton Copula
C = Following Frank Copula
D = Following Normal Copula
- = Not following Copula

Table 6. AIC$_C$ of OLS, GLM and GCMR Model

Regency	Variables	OLS	GLM	GCMR
Banyuwangi	RH1	457,20	472,97	455,44*
	RH2	447,71*	459,15	452,42
	RH3	441,11*	466,20	448,70
Bojonegoro	RH1	458,39	471,50	454,65*
	RH2	464,84	459,20*	460,89
	RH3	415,00	406,17*	406,34
Ngawi	RH1	440,53	459,14	432,01*
	RH2	447,32*	479,78	460,55
	RH3	422,88*	435,11	430,37
Jember	RH1	459,37	472,72	445,95*
	RH2	448,41	461,98	441,15*
	RH3	401,70	397,43*	404,17
Lamongan	RH1	454,46	463,97	443,97*
	RH2	453,71*	455,24	454,03
	RH3	419,44	414,09*	414,66

Note: * = Smallest AIC$_C$

Table 7. Models of harvested rice in rice production center in East Java

Banyuwangi

$\hat{R}$ H1 = exp (10, 7680+0,0005Rf. Jan.) = 47476, 9680 $(1,0005^{Rf.Jan.})$

$\hat{R}$ H2 = 23078, 4800+30, 0900Rf. Jun. = 2340, 1350-9,

$\hat{R}$ H3 0290Rf. Sep+36, 3330 Rf.Oct. +16, 4860Rf. Nov. +13, 1790Rf.Dec.

Bojonegoro

$\hat{R}$ H1 = exp (11, 0210+0,0002Rf. Jan.) (1,0001Rf.Feb) = 61144,7938 $(1,0002Rf.Jan.)(1,0001^{Rf.Feb.})$

$\hat{R}$ H2 = exp (9, 6169+0, 0041Rf. May+0, 0034Rf.

Jun+0, 0015Rf. Jul.

-0, 0024Rf. Aug.) = 15015, 9012$(1,0045^{Rf.May})$ $(1,0034^{Rf.Jun.})$ (1, 0015Rf.July) $(0, 9976^{Rf.\ Aug.})$

$\hat{R}$ H3 = exp (8, 7187+0, 0033Rf. Oct. +0,0006 Rf. Nov. -0,0002Rf.Dec.) = 6116, 0970 (1, $0033^{Rf.Oct.}$) (1, $0006^{Rf.Nov.}$)(0, $9998^{Rf.Dec.}$)

Ngawi

$\hat{R}$ H1 = exp (10,5858+0,0001Rf.Feb.+0, 0001Rf. Mar. +0, 0002Rf.Apr.) =39568, 9500(1, $0001^{Rf.Feb.}$) ($0,9999^{Rf.\ Mar.}$) (1, $0002^{Rf.Apr.}$)

$\hat{R}$ H2 =32242,0880+20,2060Rf. May+42, 0900Rf.July -8,0630Rf.Aug.

$\hat{R}$ H3 = 11527, 1500+37, 4800Rf.Sep. +28, 4000Rf.Oct. -12, 5700Rf.Nov.

Jember

$\hat{R}$ H1 = exp (11, 1300+0, 0002Rf.Apr.) =68186,3700 $(1,0002^{Rf.Apr.})$

$\hat{R}$ H2 = exp (10, 7039+0,0001Rf.May+0,0008Rf.Jun -0,0012Rf.Jul.) =44529, 181807(1,$0011^{Rf.\ May}$) $(1,0008^{Rf.\ Jun.})(0,9988^{Rf.\ July})$

$\hat{R}$ H3 =exp (9,0739+0,0001Rf.Sep.+0,0009Rf.Oct. +0,0002$^{Rf.\ Nov.}$) = 8724, 9500 (1,$0001^{Rf.Sept.}$) $(1,0009^{Rf.\ Oct.})$ $(1,0002^{Rf.Nov.})$

Lamongan

$\hat{R}$ H1 =exp (11, 0731+ 0,0001Rf.Jan.+0,0001Rf.Apr.) =64414, 8838 (1,$0001^{Rf.\ Jan.}$) (1,$0001^{Rf.\ Apr.}$)

$\hat{R}$ H2 = 36685, 6780+13,6024Rf. May-9, 1330Rf. Aug.

$\hat{R}$ H3 =exp (9,3516+0,0015Rf.Oct.-0, 0009Rf.Nop) =11516, 7388(1,$0015^{Rf.Oct.}$)(0,9991Rf. Nov.)

4. CONCLUSION

Identification of the dependence structure between the results of harvested area of rice per subround and monthly rainfall showed a dependency in all areas of research. It is shown that the real test copula type mostly follows the Gumbel distribution. Copula type is characterized by the upper tail dependencies. This phenomenon illustrates that the rice harvested area is very dependent on rainfall (especially in subround III).

Modeling of rice harvested area by comparing the three methods (OLS, GLM and GCMR) showed that GCMR better to model response variables that are not normally distributed with a large skew trend. GCMR is better when compared with the GLM method in dealing with the response variable that is not normal.

5. ACKNOWLEDGEMENT

6. REFERENCES

Agresti, A., 2007. An Introduction to Categorical Data Analysis. 2nd Edn., John Willey and Sons, Inc., New York, ISBN-10: 0470114746, pp: 400.

Bekti, R.D., 2009. Relations model for paddy's harvest area anomaly and weighted rainfall index with robust regression. Department of Statistics, ITS.

De Michele, C. and G. Salvadori, 2003. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas. J. Geophys. Res., 108: 4067-4067. DOI: 10.1029/2002JD002534

Draper, N.R. and H. Smith, 1981. Applied Regression Analysis. 2nd Edn., John Wiley and Son's, New York, ISBN-10: 0471029955, pp: 709.

Embrechts, P., F. Lindskog and A. dan McNeil, 2001. Modelling dependence with copulas and applications to risk management. Pennsylvania State University.

Embrechts, P., A. McNeil and D. Straumann, 2002. Correlation and Dependence in Risk Management: Properties and Pitfalls. In: Risk Management: Value at Risk and Beyond, Dempster, M.A.H. (Ed.), Cambridge University Press, Cambridge, ISBN-10: 1139437496, pp: 176-223.

Genest, C. and J. dan Nešlehová, 2010. Copulas: introduction to the theory and implementation in 8 R with applications in finance and insurance. Universite Laval and McGill University.

Hu, S., 2007. Akaike information criteria. North Carolina State University.

Masarotto, G. and C.D. Varin, 2012. Gaussian copula marginal regression. Elec. J. Stat., 6: 1517-1549. DOI: 10.1214/12-EJS721

McCullagh, P. and J.A. Nelder, 1989. Generalized Linear Models. 1st Edn., Chapman and Hall, London.

Miconnet, N., A.H. Geeraerd, J.F. Van Impe, L. Rosso and M. dan Cornu, 2005. Reflections on the use of robust and least-squares non-linear regression to model challenge tests conducted in/on food products. Int. J. Food Microbiol., 104: 161-177. DOI: 10.1016/j.ijfoodmicro.2005.02.014

Schölzel, C. and P.D. Friederichs, 2008. Multivariate non-normally distributed random variables in climate research-introduction to the copula approach. Nonlin. Processes Geophys., 15: 761-772. DOI: 10.5194/npg-15-761-2008

Sklar, A., 1959. Fonctions de répartition à n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris, 8: 229-231.

Villarini, G., F. Serinaldi and W.F. dan Krajewski, 2008. Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. Adv. Water Resour., 31: 1674-1686. DOI: 10.1016/j.advwatres.2008.08.002

Sutikno, B.R., A. Bey, K.A. Notodiputro and I. Las, 2010. Forecasting model of rice production using weighted rainfall index in subang, karawang and indramayu regency. Ind. Soil Climate J.

Vreac, M., A. Chedin and E. Diday, 2005. Clusteringa global field of atmospheric profiles by mixture decomposition of copulas, J. Atmos. Ocean. Tech., 22: 1445-1459. DOI: 10.1175/JTECH1795.1