Pertanika J. Sci. & Technol. 29 (1): 263 - 284 (2021)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Modelling High Dimensional Paddy Production Data using Copulas

Nuranisyha Mohd Roslan[1], Wendy Ling Shinyie[1]* and Sim Siew Ling[2]

[1]Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[2]School of Business and Management, University College of Technology Sarawak, Jalan Universiti, 96000 Sibu Sarawak, Malaysia

ABSTRACT

As the climate change is likely to be adversely affecting the yield of paddy production, thence it has brought a limelight of the probable challenges on human particularly regional food security issues. This paper aims to fit multivariate time series of paddy production variables using copula functions and predicts the next year event based on the data of five countries in southeast Asia. In particular, the most appropriate marginal distribution for each univariate time series was first identified using maximum likelihood parameter estimation method. Next, we performed multivariate copula fitting using two types of copula families, namely, elliptical copula family and Archimedean copula family. Elliptical copula family studied are normal and *t* copula, while Archimedean copula family considered are Joe, Clayton and Gumbel copulas. The performance of marginal distribution and copula fitting was examined using Akaike information criterion (AIC) values. Finally, we used the best fitted copula model to forecast the succeeding event. In order to assess the performance of copula function, we computed the forecast means and estimation errors of copula function with a generalized autoregressive conditional heteroskedasticity model as reference group. Based on the smallest AIC, the majority of the data favoured the Gumbel copula, which belongs to Archimedean copula family as well as extreme value copula family. Likewise, applying the historical data to forecast the future trends may assist

ARTICLE INFO

Article history:
Received: 28 September 2020
Accepted: 17 November 2020
Published: 22 January 2021

DOI: DOI: https://doi.org/10.47836/pjst.29.1.15

E-mail addresses:
aliakisya9731@gmail.com (Nuranisyha Mohd Roslan)
sy_ling@upm.edu.my (Wendy Ling Shinyie)
simsiewling@ucts.edu.my (Sim Siew Ling)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

all relevant stakeholders, for instance government, NGO agencies, and professional practitioners in making informed decisions without compromising the environmental as well as economical sustainability in the region.

Keywords: Archimedean copula family, dependence structure, elliptical copula family, paddy production

INTRODUCTION

Paddy is an essential crop and a staple food for more than half of the universal population. Particularly nearly 90% of the world's paddy production likewise consumption takes place in Asia (Bandumula, 2017). Abreast of the heedfulness about global climate change whereby scrutinising the Intergovernmental Panel on Climate Change (IPCC) most recently released report has articulated that the agricultural products' yields are highly correlated with atmospheric indicators, wherein the extreme whether such as droughts and flooding would eventuate grievous repercussion on the livelihood of small scale farmers particularly in Southeast Asia region (IPCC, 2019). An accelerating temperature due to the shifting of climate pattern could resultant in the decline of crop yields which may eventually shed the limelight on the shortage of global food supply. Subsequently, it may trigger food security issue wherein on the grounds of the estimated world population which is envisioned to be 9.7 billion in the next three decades despite the projection indicating a stagnant growth ever since 1950 (United Nations, 2019).

The variability of climatic factors such as total rainfall and maximum temperature has direct impacts on paddy productivity, as the extreme weather such as flood and drought can retard normal growth and grain yield (Nyang'au et al., 2014). For instance, El Nino affects the components of grain production ranging from cropping area (area planted) as well as cropping intensity (volume of production per year) in Southern Asian regions. Furthermore, the importance of soil fertility on paddy production has been well validated in previous literature whereby the farming practices in maintaining adequate input such as fertilizer is important to ensure good quality of crop (Putri et al., 2019).

In ASEAN countries, there is approximately 46.171 million (M) ha of paddy planted area in 2019 (ASEAN Food Security Information System, 2019). The largest area is found in Indonesia (10.290 M ha), followed by Thailand (11.356 M ha), Vietnam (7.478 M ha), Myanmar (7.228 M ha) and Malaysia (0.700 M ha). In general, the cultivation of paddy in Southeast Asian consists of three main systems which are (i) upland or so called as aerobic rice that is planted in dry fields; (ii) lowland rice which farmed in irrigated field for the most part of the crop growing period; and (iii) floating rice that is grown in water depths between 0.5-4.5 m (Muhammad, & Abdullah, 2013). Although approximately 55 percent of the paddy production in Southeast Asia is cultivated using floating rice system, yet the rest (i.e. upland and lowland) is highly dependent on the timely and consistent rainfall particularly during the reproductive growth stage (USDA, 2015). In term of the fertilizer

usage, Vietnam recorded the highest increment in fertilizer consumption (15.1%), followed by Myanmar (12.2%), Thailand (6.2%), Indonesia (4.6%), and Malaysia (3.7%) in the interval period of 1990-1999 (Mutert & Fairhurst, 2002).

Agriculture is a crucial economic sector that contributed approximately $3 trillion real global gross domestic production (GDP) in 2017 (Food and Agriculture Organisation, 2019). However, Hsiang et al. (2017) had revealed the simulation findings about the likelihood of global GDP to shrink in between one to three percent every year if the global warming issue persisted beyond the 21st century. A recent publication of OECD divulges that the trading volume amongst ASEAN countries merely comprises 2% of their regional yielding despite the fact that greater extent of integration between the territorial rice market can improve the malnourishment plight by 1% and even up to 6% when there are constraints in production factors (OECD, 2018).

While contemplating the level of production of paddy in 2018, the top four ranking ASEAN countries comprise Indonesia (83,037,000 tons), Vietnam (44,046,250 tons), Thailand (32,190,090 tons), and Myanmar (25,418,140 tons) while Malaysia is merely producing 2,718,990 tons (Moore, 2020). Though an observation of a declining trend apropos the consumption of table rice in some countries such as Japan and South Korea due to the diversification of choice for caloric diets, yet such incident does not betide in Malaysia wherein rice ingestion has reached triple times or more in comparison with other sources of carbohydrate such as wheat over the past four decades (OECD, 2020; Khazanah Research Institute, 2019). Thenceforth, this study has included Malaysia as one of study countries on the ground that the adverse weather condition has further reduced the paddy production along with the restricted paddy harvested area due peninsular geographical landscape. However, based on past research, the paddy harvested area had declined by 2.7% on year over year (y-o-y) in 2019 as compared to year 2018. This has made Malaysia a net importer of rice in spite of improving in paddy yield ensuing of enhancement of seed variability (ASEAN Food Security Information System, 2019).

The extraction statistics from the similar report in year 2019 manifesting that even with an upsurge of paddy planted area in countries such as Indonesia (y-o-y increment of 6.90%) and Vietnam (y-o-y increment of 0.93%), after all the paddy yield has not improved accordingly i.e. Indonesia (y-o-y declined by 1.36%) and Vietnam (y-o-y fallen by 0.17%). Nearly 80% of destructed paddy planted plots in ASEAN have been affected by adverse climate conditions, for an illustration purpose, a shrinkage of harvested area (y-o-y decreased by 0.86%) has slumped Thailand's paddy production by 2.33%, wherewith approximately 187,118 hectares and 364,773 hectares of paddy area have been damaged by flood and drought respectively (ASEAN Food Security Information System, 2019). Almost all of the paddy cultivation grown in irrigated and rainfed lowland, the aquatic environment has been further impelled the importance of having proper nutrient management as inefficient utilisation of fertilizer may constraint the grain yield (Singh & Singh, 2017).

This study intends to furnish the policy maker with intuitiveness about how the variability of condition such as rice area, fertilizer usage, total annual average rainfall and highest average temperature could affect the paddy production in the key rice producing ASEAN nations expressly Indonesia, Vietnam, Thailand, Myanmar and Malaysia with the interval between 1961 and 2014. Likewise, applying the historical data to forecast the future trends may assist all relevant stakeholders, for instance government, NGO agencies, and professional practitioners in making informed decisions without compromising the environmental as well as economical sustainability in the region. On the whole, the database is retrieved from well-organised, international repositories namely World Bank climate data portal and *ricepedia*.

The objective of this study is to compare and determine the best copula for modelling the paddy production variables, which are paddy production ('000 ton), planted area ('000 hectare), fertilizer used ('000 ton), total annual average rainfall (mm) and maximum average temperature (°C). This analysis was employed for five countries in South East Asia, i.e. Malaysia, Thailand, Indonesia, Vietnam and Myanmar. In particular, ten univariate distributions were fitted to each variable and the distribution with minimal Akaike information criterion (AIC) value would be selected for copula modelling. Two copulas from elliptical copula family and three copulas from Archimedean copula family were selected for statistical modelling of five variables. The copula which give consistent results, namely, the copula that provide smallest AIC values for all five countries, will be proposed as guidelines for practitioners in the paddy industry. Paddy planting has been the main economic activity of the rural community in ASEAN countries, hence the dependence modelling findings in this study aim to provide an efficient tool in order to increase the production and income generation for farmers and also to provide sufficient grains for the nations.

MATERIALS AND METHODS

Study Area and Data

This research focused on five variables, which were, paddy production, planted area, fertilizer usage, total annual average rainfall and maximum average temperature collected from Malaysia, Thailand, Indonesia, Vietnam, and Myanmar. The data variables used in this study were from year 1961 to 2013, which were collected from *ricepedia* and World Bank climate data portal. In this study, our main interest was paddy production. The other four variables were the factors that might affect the production. Paddy planted area is positively related with the production. Water and nutrients are important to improve aerobic conditions and support the yields. Although the optimum weather for paddy cultivation is in tropical countries, which is between 25-35°C, higher temperature will reduce the weight and quality of paddy produced. Hence, we would like to use copula functions to analyse the effects of these variables on paddy production.

Marginal Distributions

Copula is a multivariate probability distribution function for two or more variables, and their marginal probability distribution are uniformly distributed on the interval [0,1]. Most researchers have opted probability distribution functions as the marginal distributions, rather than using empirical distribution function. The plausible elucidation is the restrictions that may encountered when using empirical distributions i.e. they are relatively inefficient as they require actual values of probability to describe the full distribution and they are incapable to estimate the distribution of higher amplitudes in the long term (Sørensen, 2011). The main advantage of copula is that the marginal distribution can come from different distribution families and there is no need to assume a specific distribution to model the data. Therefore, for this research, we examined the most suitable probability distribution function for each variable using continuous probability distributions. Since there was no single suitable probability distribution for all countries and variables, we would examine the performance of fitting for ten probability distributions, namely, exponential, gamma, Weibull, Pareto, Gumbel, Laplace, normal, inverse Gaussian, log normal and logistic distributions. The marginal parameters would be estimated using maximum likelihood method. The Akaike information criterion (AIC) values would be identified and compared to select the most suitable univariate distribution for each variable. A smaller value of AIC indicates a better fit.

Copula Theory

Copulas are models for the dependence between two or more random variables when their joint distribution function is not explicitly known. The fundamental theorem of copulas states that, with F be a d-dimensional cumulative distribution function with marginal distributions F_1, $i = 1,\ldots,d$. This exists a unique decomposition $F(x_1,\ldots,x_d) = C(F_1(x_1),\ldots, Fd(x_d))$ and the copula

$$C(u_1,\ldots,u_d) = P(U_1 \leq u_1,\ldots,U_d \leq u_d), \qquad U_i \equiv F_i(X_i)$$

on $[0,1]^d$ which comprises the information on the underlying dependence structure.

Types of Copulas Used

In this study, we would focus on two families of parametric copulas, which are:

Elliptical Copula Family. Normal and *t*-distribution are the two most commonly used univariate distributions whereby through the incorporation of Sklar's theorem, the bivariate and multivariate elliptical copula family had been constructed (Fouque & Zhou, 2008; Luo & Shevchenko, 2012). Elliptical copula family has been substantially employed by keeping

the identical elliptical copula function and varying the marginal distributions (Okhrin et al., 2017). The main advantage of elliptical copula family is that different levels of correlation between the marginals can be specified, however, elliptical copula family does not have closed form expressions and are restricted to have radial symmetry.

(a) Normal copula

The normal copula with correlation matrix Σ is defined as

$$C(u_1,\ldots,u_d) = \Phi_\Sigma\left(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_d)\right)$$

With Φ is denoted as the cumulative distribution function of the standard normal variable and Φ^{-1} signifies as its inverse.

(b) *t* copula

According to Okhrin et al. (2017), the *t* copula with correlation matrix Σ is defined as

$$C(u_1,\ldots,u_d) = \int_0^1 \Phi_\Sigma(z_1(u_1,s),\ldots,z_d(u_d,s))\, ds$$

where Φ as the cumulative distribution function of the standard normal variable and $z_i(u_i,s) = t_{\upsilon_i}^{-1}(u_i)/G_{\upsilon_i}^{-1}(s)$, with t_υ^{-1} denotes the inverse for cumulative distribution function of a Student's *t* variable with degree of freedom υ and G_υ^{-1} denote the inverse for cumulative distribution function of $\sqrt{\upsilon/\chi_\upsilon^2}$.

Archimedean Copula Family. Archimedean copula family is one of the most popular copula family that has been extensively employed. This is mainly because most of the copulas in Archimedean copula family admit an explicit formula, while the elliptical copula family does not comply. Archimedean copula family has been studied in numerous research fields, for example, rainfall frequency analysis (Zhang & Singh, 2007; Zhang & Singh, 2012), intensity-duration-frequency relationship (Ariff et al., 2012), modelling wind speed dependence (Xie et al., 2012), modelling option pricing (Cherubini & Luciano, 2002) and probabilistic estimates of heat stress for rice (Zhang et al., 2018).

(a) Joe copula

According to Joe (1997) (Equation 1),

$$C(u_1,\ldots,u_d) = 1 - \left(\sum_{i=1}^{d}(1-u_i)^\theta - \prod_{i=1}^{d}(1-u_i)^\theta\right)^{1/\theta} \qquad (1)$$

for $\theta \geq 1$.

(b) Clayton copula

Clayton (1978), Cook and Johnson (1981) and Oakes (1982) had defined the copula as in Equation 2

$$C(u_1,\ldots,u_d)=\left(\sum_{i=1}^{d}u_i^{-\theta}-d+1\right)^{-1/\theta} \quad (2)$$

for $\theta > 0$. Independence will lead $\theta \rightarrow 0$. A complete dependence corresponds to $\theta \rightarrow \infty$.

(c) Gumbel copula

Gumbel (1960) had derived the Gumbel copula as in Equation (3)

$$(u_1,\ldots,u_d)=\exp\left(-\left(\sum_{i=1}^{d}(-\log u_i)^{\theta}\right)^{\frac{1}{\theta}}\right) \quad (3)$$

for $\theta \geq 1$, $\theta = 1$ if the structure is independent. Besides that, Gumbel copula is the only copula that was grouped as an Archimedean copula family as well as an extreme value copula family.

Estimating Copula Parameter

The parameter for the five selected copulas will be estimated using maximum likelihood estimator (Equation 4). Given a sample $u_i, i \in \{1,\ldots,d\}$,

$$\hat{\theta}=\underset{\theta\in\Theta}{\operatorname{argsup}}\sum_{i=1}^{d}\log c(u_i) \quad (4)$$

where $c(u_i)$ is the density function of *C*.

By using the derived likelihood value, we will compare the performance of copulas using Akaike Information Criterion (AIC) values. The best fit copula would be the copula with minimal AIC value.

Prediction Method using Best Fit Copula

The identification of best fit copula is useful for researchers and practitioners to predict the next year event as well as extreme quantiles. For each country, the *d*-dimensional time series for *n* years is denoted as $(x_1,\ldots,x_d)$ and the marginal distribution of x_i is F_i.

The following prediction algorithm was proposed by Simard and Remillard (2015) to forecast the x_{n+1}:

1. Let $u = (u_1, \ldots, u_d)$ be the copula data of d-dimensional for the best fit copula. Then the Rosenblatt's transformation of u, will be denoted as $y = (y_1, \ldots, y_d)$. As mentioned in Schepsmeier (2015), $y_1 := u_1, y_2 := C(u_2|u_1), \ldots, y_d := C(u_d|u_1, \ldots, u_{d-1})$.
2. Simulate k realizations for conditional copula, which is $C_{u_n|u_{n-1}}(y)$. Set the simulation result as $U^{(j)}$, $j \in \{1, \ldots, k\}$.
3. Set $x_{n+1}^{(j)} = F^{-1}(U^{(j)})$ and determine the predicted value using $\hat{x}_{n+1} = \frac{1}{k}\sum_{j=1}^{k} x_{n+1}^{(j)}$. Hence, we will use $\hat{x}_{i,n+1}$ as the predictor for $x_{i,n+1}$.
4. Determine the mean and standard error of the prediction value.

RESULTS AND DISCUSSIONS

Exploratory Data Analysis

The annual time series for five countries in southeast Asia, namely Malaysia, Thailand, Indonesia, Vietnam and Myanmar, were used in this study. Five adopted variables were paddy production, planted area, fertilizer used, total annual average rainfall and maximum average temperature. The summary statistics for the five countries are shown in Table 1. Among the five countries, Malaysia has the smallest paddy planted area as well least outputs of paddy production, while Indonesia ranks the highest for both aforementioned indicators. For the climatological variables, Thailand has the lowest total annual rainfall and the highest maximum temperature. By observing the maximum paddy production of Thailand (in year 2012) and Indonesia (in year 2013), we found that although the difference for paddy planted area was only 1880 ('000 hectare), the paddy production harvested in Indonesia was almost twice that of Thailand. This might be due to the variation in climatology factors (amount of annual rainfall difference is almost twice) or other unperceived factors.

Figures 1 to 5 present the plot of five variables studied for Malaysia, Thailand, Indonesia, Vietnam, and Myanmar. These plots can be used to provide preliminary insights about the trend and relationship of the variables. In general, a clear linear pattern is visible in paddy production, planted area and fertilizer usage variable. For total annual rainfall and maximum temperature, only slight linear pattern can be observed. Multivariate Mann-Kendall trend test were performed to analyse data collected over time for consistently increasing or decreasing trends. All five countries are having small p-value which are approximately zero. This indicates that the multivariate data are all having monotonic trends which also means that the data are showing a trend, that can be either positive, negative or non-null.

Figures 6 to10 provide the boxplot for each country. We can observe the shape of distribution for each time series and have an initial understanding of the data. For paddy production variable, Malaysia and Indonesia are showing left-skewed distribution, while the rest are right-skewed. For paddy planted area, only Malaysia and Thailand are showing negative skewness. Based on the fertilizer usage data, only Indonesia's data is left-skewed,

Table 1
Summary statistics of variables used for five countries

Country	Variables used	Min.	Q1	Median	Mean	Q3	Max
Malaysia	Paddy Production	1089	1696	1995	1913	2141	2604
	Planted Area	516.5	659	676.2	669.6	693.4	766.2
	Fertilizer Usage	93.71	228.5	723.7	825	1270	2241
	Annual Rainfall	2498	2830	3008	3036	3250	3733
	Maximum Temperature	25.36	25.76	26.13	26.11	26.42	27.32
Thailand	Paddy Production	10150	13920	19550	20720	25840	38000
	Planted Area	6120	7743	9147	8971	9913	11960
	Fertilizer Usage	1.72	12.16	35.13	59.41	108.1	167.7
	Annual Rainfall	1268	1472	1565	1564	1651	1974
	Maximum Temperature	27.84	28.55	29.09	29.1	29.63	30.49
Indonesia	Paddy Production	11600	22340	40080	38280	51100	71280
	Planted Area	6731	8369	9988	10030	11570	13840
	Fertilizer Usage	5.25	27.17	105.5	93.57	142.7	205.4
	Annual Rainfall	2163	2647	2934	2861	3073	3581
	Maximum Temperature	25.83	26.23	26.46	26.45	26.65	27.44
Vietnam	Paddy Production	8366	10600	16000	21000	32110	44040
	Planted Area	4497	5030	5718	6116	7329	7903
	Fertilizer Usage	8.29	45.93	94.08	153.6	292.3	403.9
	Annual Rainfall	1525	1638	1834	1825	1978	2146
	Maximum Temperature	26.79	27.17	27.41	27.43	27.7	28.17
Myanmar	Paddy Production	6636	8602	14150	15990	21320	32680
	Planted Area	4254	4672	4884	5519	6302	8078
	Fertilizer Usage	0.61	4.71	8.95	9.743	15.82	20.76
	Annual Rainfall	1527	1868	1988	1992	2100	2483
	Maximum Temperature	24.88	25.43	25.77	25.87	26.24	27.24

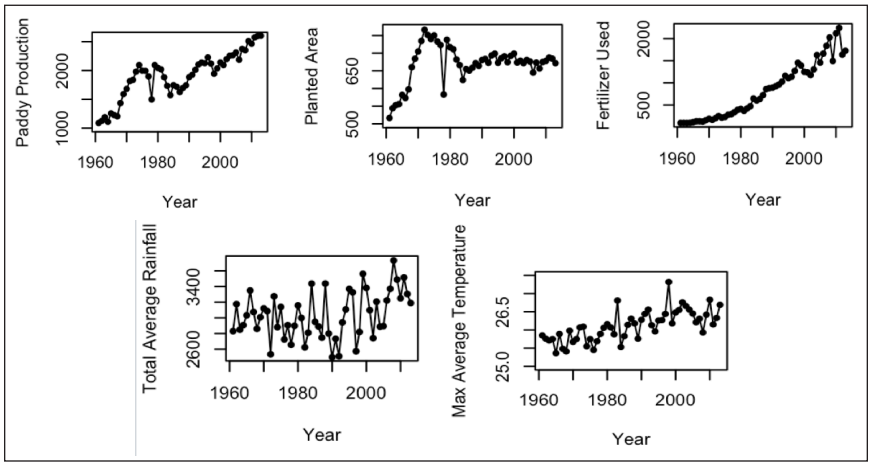

Figure 1. Scatterplots of five variables used for Malaysia

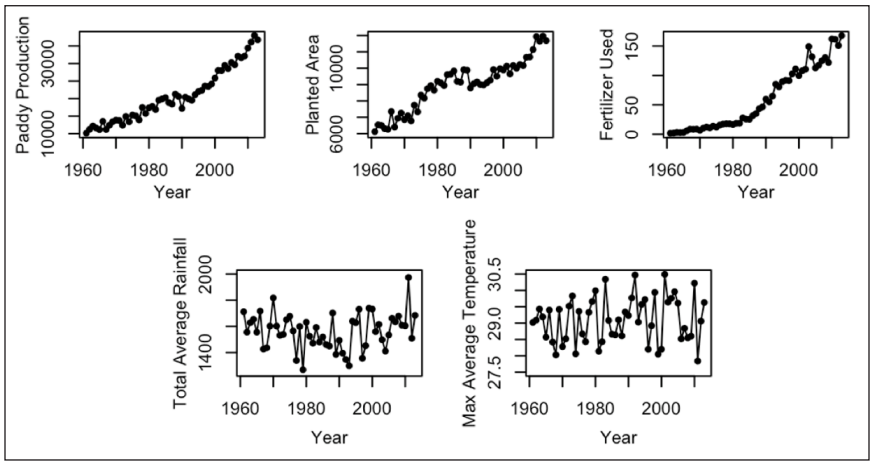

Figure 2. Scatterplots of five variables used for Thailand

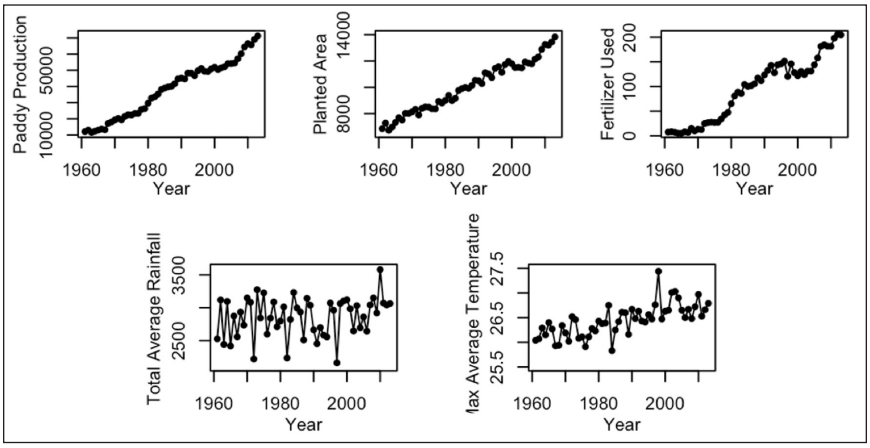

Figure 3. Scatterplots of five variables used for Indonesia

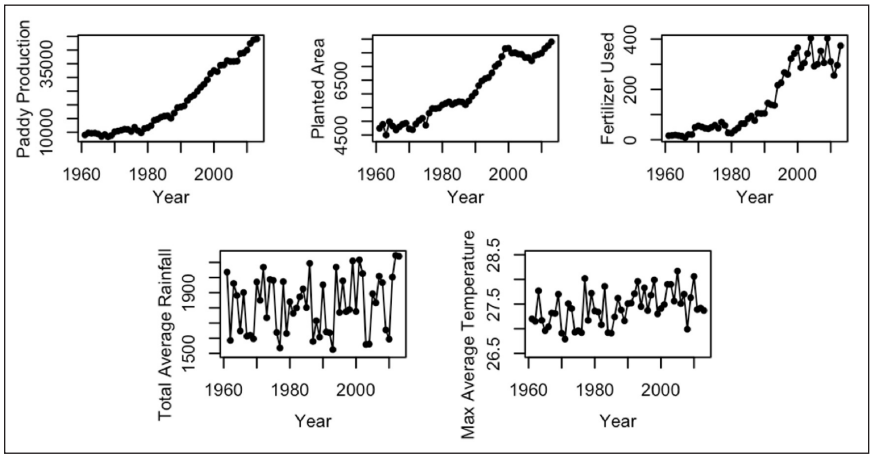

Figure 4. Scatterplots of five variables used for Vietnam

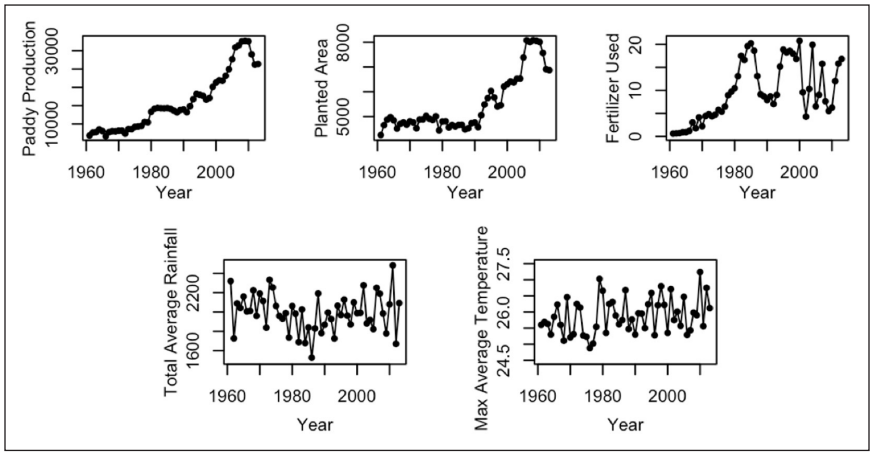

Figure 5. Scatterplots of five variables used for Myanmar

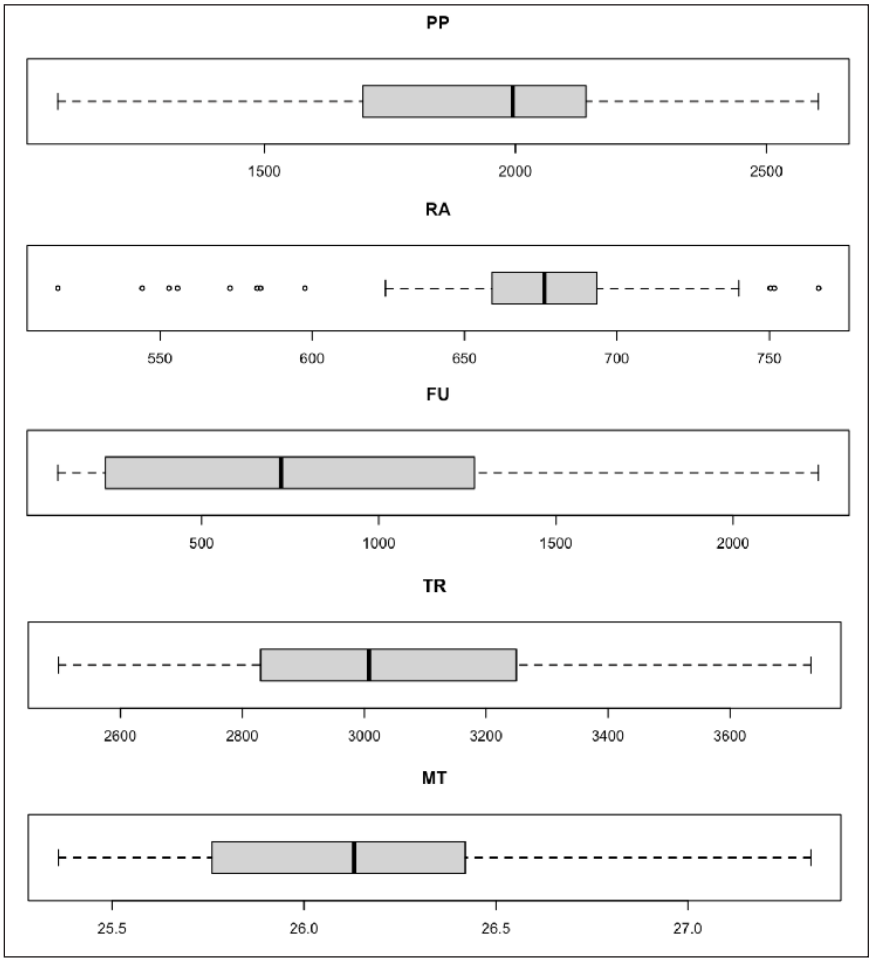

Figure 6. Boxplot of variables used for Malaysia

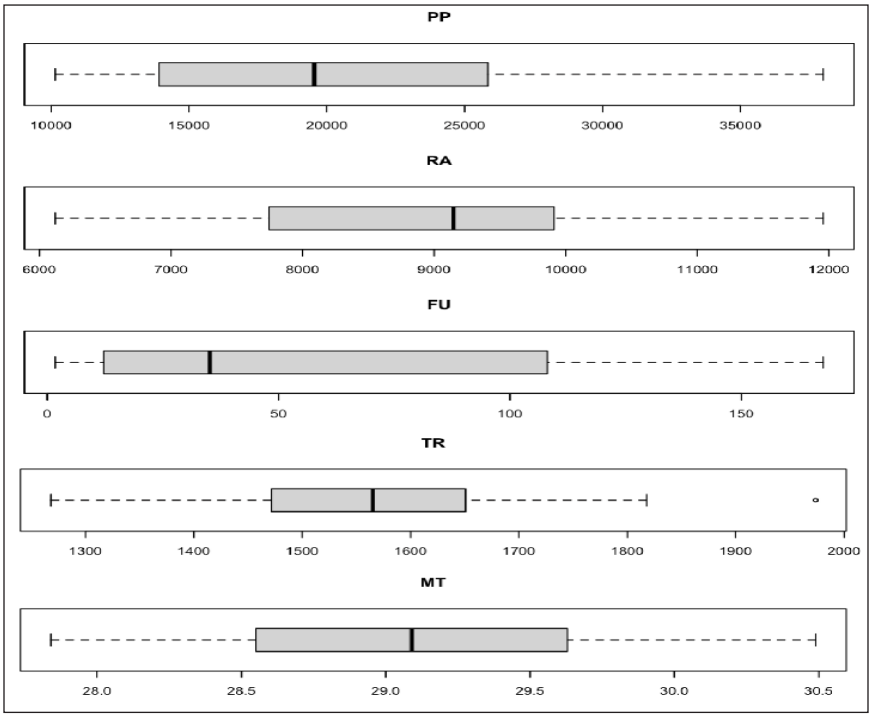

Figure 7. Boxplot of variables used for Thailand

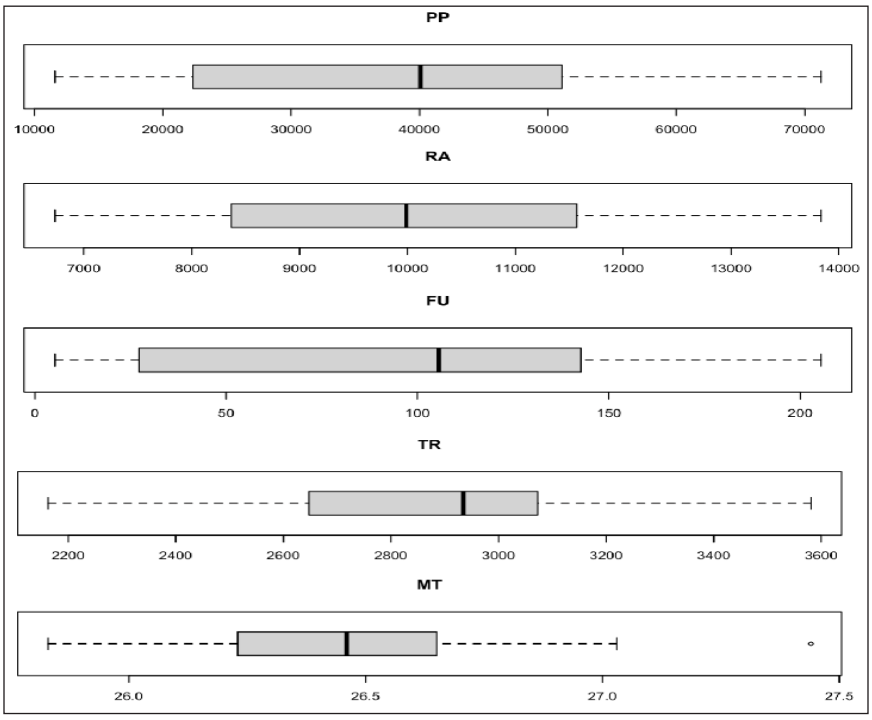

Figure 8. Boxplot of variables used for Indonesia

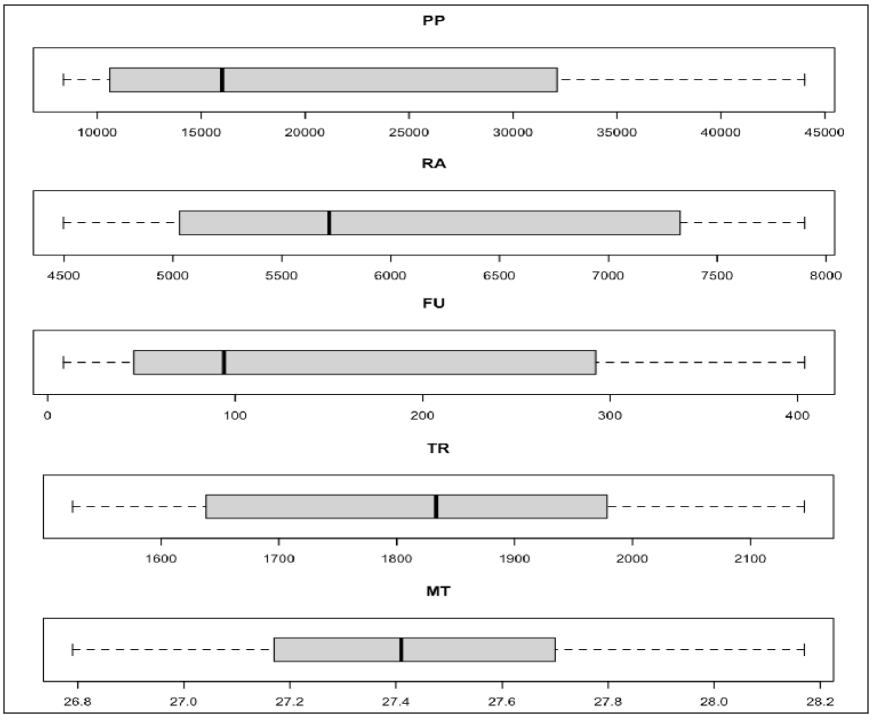

Figure 9. Boxplot of variables used for Vietnam

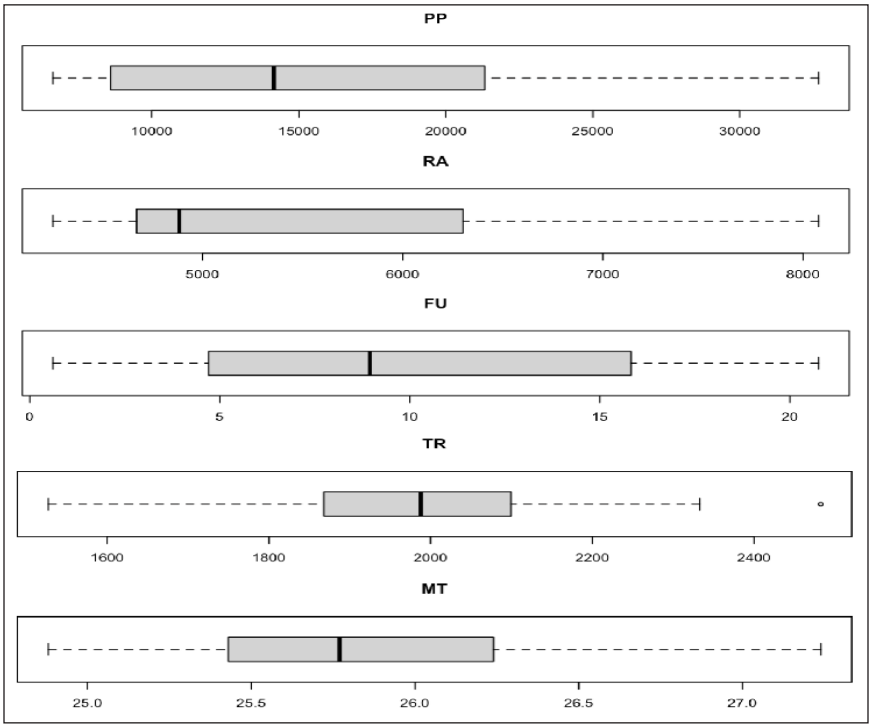

Figure 10. Boxplot of variables used for Myanmar

the other four countries are right-skewed. The average of total annual rainfall time series in Malaysia and Myanmar are having positive skewness behaviour, while Thailand, Indonesia and Vietnam have longer tail at the left. For the fifth adopted variable, which is maximum average temperature, only Malaysia and Indonesia are indicating negative skewness.

After understanding the data variables, we proceeded with determining best fit marginal distribution for each variables. Variables used were paddy production (PP), paddy planted area (RA), fertilizer used (FU), total annual rainfall (TR) and maximum temperature (MT). The distributions tested were exponential, gamma, Weibull, Pareto, Gumbel, Laplace, normal, inverse Gaussian, log normal and logistic. Table 2 provides the best fit distributions and their respective parameter estimates. Rate parameter shown is also the inverse of scale parameter, it is one of the parameters for exponential and gamma distributions. Generally,

Table 2
Parameter estimates of best fit distributions for each univariate time series

			Parameters			
		Best fit distribution	Location	Scale	Shape	Rate
Malaysia	PP	Weibull		2069.768	5.752	
	RA	Laplace	676.2	36.33		
	FU	Weibull		886.54	1.26	
	TR	Inverse Gaussian	3036		0.0003049	
	MT	Inverse Gaussian	26.11		0.0009491	
Thailand	PP	Inverse Gaussian	20722		0.00000654	
	RA	Weibull		9613.993	6.672	
	FU	Exponential				0.01683
	TR	Gamma			133.08559	0.08511
	MT	Inverse Gaussian	29.1		0.0001882	
Indonesia	PP	Weibull		43290.981	2.416	
	RA	Gamma			26.927714	0.002684
	FU	Weibull		99.922	1.269	
	TR	Weibull		2989.7	11.21	
	MT	Normal	26.4466	0.3179		
Vietnam	PP	Inverse Gaussian	20997		0.0000163	
	RA	Inverse Gaussian	6116		0.0000053	
	FU	Exponential				0.006509
	TR	Gamma			96.59	0.05292
	MT	Inverse Gaussian	27.43		0.0005807	
Myanmar	PP	Inverse Gaussian	15993		0.0000159	
	RA	Pareto		4253.7	4.148	
	FU	Weibull		10.701	1.475	
	TR	Normal	1991.7	191.4		
	MT	Gumbel	25.611	0.4616		

Weibull and inverse Gaussian were most suitable distribution for the variables studied. These two distributions have similar shape as both are positively skewed and having long tail. Besides, the gamma distribution which is a family of right-skewed probability distributions is the third most suitable distribution. This indicates that most of the data studied in this research are skewed to the right and exhibiting heavy tail.

Table 3 provides the further details of model fitting. The models listed in second column are used to represent the different variables used for model fitting. Variables studied for their respective models are as listed below:

Model 1 : Paddy production and planted area
Model 2 : Paddy production and fertilizer usage
Model 3 : Paddy production and total annual rainfall
Model 4 : Paddy production and maximum temperature
Model 5 : Paddy production, planted area and fertilizer usage
Model 6 : Paddy production, planted area and total annual rainfall
Model 7 : Paddy production, planted area and maximum temperature
Model 8 : Paddy production, fertilizer usage and total annual rainfall
Model 9 : Paddy production, fertilizer usage and maximum temperature
Model 10 : Paddy production, total annual rainfall and maximum temperature
Model 11 : Paddy production, planted area, fertilizer usage and total annual rainfall
Model 12 : Paddy production, planted area, fertilizer usage and maximum temperature
Model 13 : Paddy production, planted area, total annual rainfall and maximum temperature
Model 14 : Paddy production, fertilizer usage, total annual rainfall and maximum temperature
Model 15 : Paddy production, planted area, fertilizer usage, total annual rainfall and maximum temperature

The purpose of studying different combinations are to measure the relationship for these variables and to compare the performance of different variables combination for three types of model fitting approaches. Multiple correlation (Corr) for the models are shown in third column which signifying as model 3 and 4 whereby these models with only paddy production and climatological variables have low correlation. However, when we included more variables, the correlation magnitude had increased and became strongly correlated. In addition, model 15 exhibited the highest correlation values for all five countries although for clarity only the nearest thousandth are shown (as presented in bold font in Table 3).

Apart from modelling using high dimensional copulas, we have included the AIC values for multiple regression model (MRM) and multivariate normal distribution (MVN). Multiple regression model is the simplest method to distinguish relationship between multiple independent variables and one dependent variable. From the results shown in

Table 3
Results for all methods

					AIC				
	Model	Corr	MRM	MVN	Normal	t	Joe	Clayton	Gumbel
Malaysia	1	0.647	761.88	-15.81	-19.20	-17.19	2.00	-41.14	-6.11
	2	0.800	736.46	8.97	-60.39	-58.28	-51.78	-48.23	-58.47
	3	0.298	785.67	-32.16	-3.51	-1.33	-7.67	7.55	-5.64
	4	0.569	769.82	-9.90	-18.61	-16.62	-7.21	-18.09	-13.87
	5	0.952	667.60	33.96	-99.63	-94.43	-103.22	-101.86	-110.12
	6	0.716	754.42	-29.24	-24.82	-17.87	-48.50	-48.38	-37.34
	7	0.788	741.27	-6.16	-38.60	-32.18	-68.38	-70.24	-57.07
	8	0.801	738.32	-7.13	-62.47	-55.78	-60.73	-61.28	-63.90
	9	0.802	738.02	33.44	-92.75	-86.73	-84.10	-87.03	-92.47
	10	0.645	764.12	-23.43	-23.44	-16.60	-21.48	-24.51	-23.06
	11	0.952	669.38	21.00	-103.53	-92.57	-107.63	-107.40	-113.80
	12	0.952	669.59	59.05	-130.51	-118.74	-133.54	-134.12	-142.12
	13	0.847	727.73	-14.34	-48.29	-34.37	-74.40	-76.55	-53.71
	14	0.802	740.01	22.96	-99.68	-86.60	-94.05	-96.43	-100.26
	15	**0.952**	671.37	51.15	-138.74	-118.54	-138.66	-141.01	**-147.78**
Thailand	1	0.929	997.41	95.28	-131.77	-129.65	-87.93	-131.53	-112.39
	2	0.955	973.36	79.88	-128.83	-126.62	-101.77	-105.38	-120.35
	3	0.192	1100.60	-37.89	1.04	3.08	-0.03	2.94	0.47
	4	0.123	1101.80	-43.60	0.85	2.86	1.01	1.47	0.90
	5	0.982	928.62	205.93	-264.23	-257.47	-233.96	-235.21	-261.60
	6	0.940	990.48	80.68	-130.31	-123.71	-130.95	-132.41	-138.05
	7	0.929	999.41	72.96	-129.24	-123.00	-128.15	-128.98	-135.44
	8	0.956	974.58	65.10	-131.17	-124.45	-104.40	-107.38	-124.28
	9	0.955	975.31	56.74	-126.02	-119.31	-102.66	-103.91	-119.02
	10	0.270	1100.59	-58.09	-1.79	4.67	-7.29	-6.68	-5.42
	11	0.985	921.64	198.96	-269.56	-256.38	-234.77	-237.66	-267.01
	12	0.982	930.24	183.62	-259.70	-246.61	-230.32	-231.97	-257.45
	13	0.941	991.50	62.58	-132.21	-120.01	-137.57	-137.79	-143.86
	14	0.956	976.57	44.90	-132.37	-119.29	-114.98	-117.81	-133.20
	15	**0.985**	923.50	181.72	-270.70	-249.42	-244.04	-248.81	**-275.33**
Indonesia	1	0.992	970.25	172.35	-214.97	-212.72	-204.44	-73.69	-221.12
	2	0.987	998.13	124.24	-160.81	-158.43	-125.14	-122.94	-148.88
	3	0.065	1190.50	-39.54	-0.99	1.03	0.36	0.31	-0.14
	4	0.127	1189.86	5.39	-31.45	-29.39	-15.72	-27.36	-24.31
	5	0.997	917.49	349.71	-392.55	-384.97	-345.17	-344.85	-386.04
	6	0.993	962.88	165.14	-222.04	-215.12	-211.12	-211.20	-229.30
	7	0.992	972.15	201.58	-244.92	-238.88	-233.44	-233.55	-252.92
	8	0.987	997.96	110.09	-162.79	-155.64	-126.44	-125.50	-151.67
	9	0.987	997.20	154.01	-190.54	-183.18	-149.87	-153.81	-178.42
	10	0.719	1154.19	-7.55	-31.57	-25.28	-23.61	-26.90	-29.46

Table 3 (*continue*)

					AIC				
	Model	Corr	MRM	MVN	Normal	t	Joe	Clayton	Gumbel
	11	0.997	915.19	343.20	-397.69	-383.79	-351.97	-352.03	-392.22
	12	0.997	919.21	379.73	-420.53	-410.27	-368.02	-370.89	-413.83
	13	0.993	964.71	197.81	-252.06	-240.20	-238.62	-240.94	-261.05
	14	0.988	996.00	143.23	-192.02	-178.88	-149.84	-154.50	-181.70
	15	**0.997**	917.19	376.37	**-425.77**	-408.19	-374.16	-378.19	-420.09
Vietnam	1	0.964	1007.36	89.40	-42.63	-79.83	-106.82	-7.41	-97.85
	2	0.951	1022.93	79.61	-40.07	-79.97	-101.16	-5.90	-94.68
	3	0.212	1145.00	-41.28	0.03	3.17	-2.12	1.25	-1.05
	4	0.408	1137.81	-26.18	-3.87	-1.34	-5.23	1.47	-5.89
	5	0.970	999.75	214.66	-162.97	-195.56	-229.15	-231.66	-232.91
	6	0.965	1007.37	62.60	-40.60	-72.30	-107.46	-106.80	-97.14
	7	0.965	1008.17	80.97	-51.83	-84.11	-110.82	-112.82	-105.12
	8	0.953	1023.36	53.25	-38.07	-72.70	-101.28	-101.00	-93.73
	9	0.951	1024.93	69.54	-47.86	-82.03	-104.26	-106.83	-99.80
	10	0.524	1132.46	-45.72	-8.72	-0.64	-7.68	-10.38	-14.83
	11	0.971	999.64	188.75	-159.13	-184.15	-230.50	-231.79	-231.96
	12	0.971	1000.54	206.24	-170.18	-195.73	-231.15	-235.00	-238.17
	13	0.965	1009.01	61.85	-55.63	-85.28	-120.44	-120.83	-113.94
	14	0.953	1025.13	51.51	-51.95	-82.00	-109.69	-111.47	-107.45
	15	**0.971**	1001.28	188.25	-172.29	-193.06	-238.76	-241.00	**-244.99**
Myanmar	1	0.946	987.50	36.35	-31.69	-38.20	-86.55	-1.83	-65.60
	2	0.422	1096.38	-17.90	-24.01	-21.88	-4.50	-33.25	-14.55
	3	0.009	1106.75	-33.49	1.77	2.94	2.00	1.55	1.91
	4	0.250	1103.33	-39.95	-2.37	-0.26	-1.02	-2.39	-2.01
	5	0.973	953.38	55.95	-54.23	-63.54	-118.84	-120.36	-92.26
	6	0.958	976.68	33.09	-27.93	-35.18	-88.14	-88.63	-64.51
	7	0.952	983.43	22.01	-33.25	-34.92	-89.26	-90.47	-67.78
	8	0.448	1096.90	-23.46	-27.03	-22.20	-34.06	-36.19	-31.31
	9	0.450	1096.75	-35.54	-24.93	-18.17	-32.39	-34.12	-29.03
	10	0.265	1104.89	-45.75	-3.40	2.83	-0.83	-1.91	-2.13
	11	0.976	949.92	54.34	-55.26	-60.74	-122.46	-124.29	-94.38
	12	0.975	951.45	41.64	-54.22	-55.97	-119.55	-122.45	-92.50
	13	0.960	976.31	22.28	-32.29	-30.64	-86.49	-87.98	-65.89
	14	0.491	1096.15	-35.71	-30.20	-18.40	-34.18	-36.51	-33.49
	15	**0.977**	949.84	-42.36	-57.57	-53.09	-126.48	**-127.63**	-96.52

Table 3, multiple regression method is the least fitted model. As for multivariate normal distribution, it is an approach to generalize univariate normal distribution to higher dimensions and is derived from the multivariate central limit theorem. Since the sample sizes of variables used are sufficiently large, the central limit theorem assumption is fulfilled. Based upon the AIC values shown, we can conclude that the multivariate normal

distribution fits the data relatively better than multiple regression model. Most of the AIC values are negative, this indicates that there is less information loss as compared to MRM.

Subsequently, we discuss the performance of the copulas fitted to different combination of variables. Generally, for all five countries, model 15 performed the best as the AIC values produced were generally lower than other models (as shown as bold font in Table 3). Apart from that, we also found that the Gumbel copula performed best for Malaysia, Thailand and Vietnam. Since Gumbel copula typically signifies as an extreme value copula, we can further presume that the variables for Malaysia, Thailand and Vietnam exhibit a heavy tail behaviour. But for Myanmar, both Clayton and Joe copulas for model 15 were performing identically well, with Clayton copula having a relatively lower AIC value. Besides that, the best fit copula for Indonesia was the normal copula, with AIC value equal to -425.77. Therefore, the best fit marginal distributions and copula for each countries would be utilised for further prediction.

Finally, we predicted the next year event using best fit marginal distributions for each variables and best fit copula function for the countries. For each country, one thousand simulations were performed and the average together with estimation error of the predictions were computed. In order to identify the performance of prediction results, we also forecasted the next year event using univariate generalized autoregressive conditional heteroskedasticity (GARCH) time series model. The GARCH model is selected as the reference group for prediction using copula modelling as it is a most common forecasting method for time series and GARCH model is known as an effective model that aims to minimize errors in forecasting. For this study, the number of autocorrelation term used for the GARCH model was 1 (also known as AR(1)), and GARCH(1,1) was used to model the variance term. This AR(1)-GARCH(1,1) model was selected as it is one of the most prevalent GARCH model and the error approximation is relatively smaller as compared to other models that examined using our existing data.

Table 4 shows the mean and standard error of predicted values for copula and GARCH model. Based on the values of standard error, copula model behaved reasonably well compared to GARCH model since fifteen out of the total of twenty-five predictions of copula showed lower figure. Other than that, the predicted means are provided in the same table for comparative purpose. It can be seen that copula model produces higher mean than the GARCH model in nineteen predictions.

The result of prediction has indicated that the forecasting method based on copula models are also capable in detecting autocorrelation and volatility of multivariate time series. The findings in this research can be useful for practitioners and other related stakeholders in monitoring the variables that will affect paddy production and to predict the future trend.

Table 4
Mean and standard error for predicted values

		Copula		GARCH	
	Variables	Mean	SE	Mean	SE
Malaysia	PP	2610.1	58.4	2626.3	80.3
	RA	719.4	20.0	671.0	13.1
	FU	1859.6	362.5	1732.8	395.1
	TR	3414.9	180.3	3078.2	305.3
	MT	27.94	0.57	26.49	0.38
Thailand	PP	31267.0	3312.0	31498.6	3665.8
	RA	10967.0	522.5	11660.9	442.6
	FU	133.5	20.2	125.0	54.8
	TR	1716.4	65.1	1571.2	146.5
	MT	30.77	0.85	29.07	0.74
Indonesia	PP	59006.0	7345.9	58789.3	7243.0
	RA	12601.8	1668.9	12351.3	809.6
	FU	191.9	25.6	162.7	31.6
	TR	3186.6	148.5	2831.6	298.0
	MT	26.72	0.221	26.65	0.26
Vietnam	PP	35584.8	3651.0	37520.7	3892.9
	RA	7444.8	334.4	7504.9	185.4
	FU	385.3	109.1	374.8	81.5
	TR	2066.4	55.9	1812.7	202.6
	MT	27.76	0.17	27.42	0.34
Myanmar	PP	26936.2	4417.2	27330.1	4277.2
	RA	6997.2	267.5	6874.2	272.4
	FU	19.2	3.205	17.1	2.4
	TR	2221.8	100.5	1979.1	199.1
	MT	26.38	0.36	25.88	0.56

CONCLUSION

Rice is important for human consumption as well as for economic growth particularly for countries that produce rice in tropical region. Therefore, the objective of this study was to identify the best fit model and perform prediction using the model. We had compared the high dimensional copulas with multiple regression model and multivariate normal distribution by using several variables which were paddy production, paddy planted area, fertilizer usage, total annual rainfall and maximum temperature for five countries in southeast Asia. The five countries were Malaysia, Thailand, Indonesia, Vietnam, and Myanmar located in the tropical region in southeast Asia. Prior to multivariate analysis,

we had to determine the best fit univariate marginal distributions for all variables using maximum likelihood estimation method. The results indicate that Weibull and inverse Gaussian probability distributions fitted well to most of the variables.

Based on the results of model fitting, copulas produced the lowest AIC values while multivariate normal distribution produced a moderate AIC and multiple regression model has the highest AIC. For Malaysia, Thailand and Vietnam, Gumbel copula is the most suitable copula for the model that consists of all five variables. On the contrary, although the model contains of all five variables performs best for Myanmar and Indonesia too, the best fit copula for Myanmar is Clayton copula whilst for Indonesia is normal copula. In general, we can conclude that copulas are able to reduce the information loss in model fitting. Besides that, planted area, fertilizer usage, rainfall and temperature do play an important role in paddy production.

The forecasted values of the following year event were computed based on best fit marginal distribution and copula functions. In order to compare the effectiveness of copula, we have also computed the mean and standard error of forecasted values using AR(1)-GARCH(1,1) model. GARCH model is treated as reference group due to its ability to minimize errors in forecasting and to enhance the accuracy of further predictions. Based on the results, we found that the performance of the prediction is relatively similar with GARCH model. Hence, this proves that the effectiveness of multivariate copula model is comparable to univariate GARCH model.

ACKNOWLEDGEMENT

We acknowledge the financial support of Universiti Putra Malaysia (UPM-GP-IPM-9587800).

REFERENCES

Ariff, N. M., Jemain, A. A., Ibrahim, K., & Zin, W. Z. W. (2012). IDF relationships using bivariate copula for storm events in peninsular Malaysia. *Journal of Hydrology, 470*, 158-171. doi: https://doi.org/10.1016/j.jhydrol.2012.08.045

ASEAN Food Security Information System. (2019). *ASEAN Agricultural Commodity Outlook, No. 22, June 2019.* Retrieved July 11, 2020, from http://www.aptfsis.org/uploads/normal/ ACO%20Report%2022/ ACO%20Report22.pdf

Bandumula, N. (2017). Rice production in Asia: Key to global food security. In *Proceedings of the Natural Academy of Sciences, India, Section B: Biological Sciences, 88*(4), 1323-1328. doi: https://doi.org/10.1007/s40011-017-0867-7

Cherubini, U., & Luciano, E. (2002). Bivariate option pricing with copulas. *Applied Mathematical Finance, 9*(2), 69-86.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, *65*(1), 141-151. doi: https://doi.org/10.1093/biomet/65.1.141

Cook, R. D., & Johnson, M. E. (1981). A family of distributions for modeling nonelliptically symmetric multivariate data. *Journal of the Royal Statistical Society: Series B,* 4(2)3, 210-218. doi: https://doi.org/10.1111/j.2517-6161.1981.tb01173.x

Food and Agriculture Organization of the United Nations. (2019). *Macroeconomic statistics: Global trends in GDP, agriculture value added, and food-processing value added (1970-2017).* Retrieved June 28, 2020, from http://www.fao.org/economic/ess/ess-economic/ gdpagriculture/en/

Fouque, J. P., & Zhou, X. (2008). Perturbed gaussian copula. In J. P. Fouque, T. B. Fomby & K. Solna (Eds.), *Econometrics and risk management* (pp. 103-121). Bingley, England: Emerald Group Publishing Limited. doi: https://doi.org/10.1016/S0731-9053(08)22005-0

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association, 55*(292), 698-707.

Hsiang, S., Kopp, R., Jina, A., Rising, J., Delgado, M., Mohan, S., … & Oppenheimer, M. (2017). Estimating economic damage from climate change in the United States. *Science, 356*(6345), 1362-1369. doi: 10.1126/science.aal4369

IPCC. (2019). Summary for policymakers. *In Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty.* Intergovernmental Panel on Climate Change (IPCC), Geneva, Switzerland. Retrieved June 28, 2020, from https://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/ SR15_Full_Report _High_Res.pdf

Joe, H. (1997). *Multivariate models and dependence concepts*. London, UK: Chapman and Hall.

Khazanah Research Institute. (2019). *The status of the paddy and rice industry in Malaysia.* Retrieved July 11, 2020, from http://www.krinstitute.org/assets/contentMS/img/template/ editor/20190409_RiceReport_Full%20Report_Final.pdf

Luo, X., & Shevchenko, P. V. (2012). Bayesian model choice of grouped t-copula. *Methodology and Computing in Applied Probability, 14*(4), 1097-1119. doi: https://doi.org/10.1007/s11009-011-9220-4

Muhammad, M., & Abdullah, M. H. H. (2013). Modelling and forecasting on paddy production in Kelantan under the implementation of system of rice intensification (SRI). *Academic Journal of Agricultural Research, 1*(7), 106-113. doi: http://dx.doi.org/10.15413/ajar.2013.0112

Moore, M. (2020). *Rice paddy production in the Asia Pacific region in 2018, by country.* Retrieved July 11, 2020, from https://www.statista.com/statistics/681740/asia-pacific-rice-paddy-production-by-country/#statisticContainer

Mutert, E., & Fairhurst, T. H. (2002). Developments in rice production in Southeast Asia. *Better Crops International, 15*(Suppl), 12-17.

Nyang'au, W., Mati, B., Kalamwa, K., Wanjogu, R., & Kiplagat, L. (2014). Estimating rice yield under changing weather conditions in Kenya using CERES rice model. *International Journal of Agronomy, 2014*, 1-12. doi: https://doi.org/10.1155/2014/849496

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society: Series B, 44*(3), 414-422. doi: https://doi.org/10.1111/j.2517-6161.1982.tb01222.x

OECD. (2018). *Joint working party on agriculture and trade. ASEAN rice market integration: Findings from a feasible study.* Organisation for Economic Co-operation and Development. Retrieved July 11, 2020, from http://www.oecd.org/officialdocuments/publicdisplay documentpdf/? cote=TAD/TC/CA/WP(2018)7/FINAL&docLanguage=En

OECD. (2020). *OECD-FAO agricultural outlook 2019 – 2028: OECD-FAO agricultural outlook 1990-2028, by country*. Retrieved July 7, 2020, from https://stats.oecd.org/

Okhrin, O., Ristig, A., & Xu, X. F. (2017). Copulae in high dimensions: An introduction. In W. Härdle, C. H. Chen & L. Overbeck (Eds.), *Applied quantitative finance, statistics and computing*. Heidelberg, Germany: Springer. doi: https://doi.org/10.1007/978-3-662-54486-0_13

Putri, R. E., Yahya, A., Adam, N. M., & Aziz, S. A. (2019). Rice yield prediction model with respect to crop healthiness and soil fertility. *Food Research, 3*(2), 174-180. doi: http://doi.org/10.26656/fr.2017.3(2).117

Simard, C., & Rémillard, B. (2015). Forecasting time series with multivariate copulas. *Dependence Modeling, 3*, 59-82.

Singh, B., & Singh, V. K. (2017). Fertilizer management in rice. In B. Chauhan, K. Jabran & G. Mahajan (Eds.), *Rice production worldwide* (pp. 217-253). Cham, Switzerland: Springer. doi: https://doi.org/10.1007/978-3-319-47516-5_10

Sørensen, M. (2011). Estimating functions for diffusion-type processes. In M. Kessler, A. Lindner & M. Sørensen (Eds.), *Statistical methods for stochastic differential equations*. London, UK: Chapman & Hall.

United Nations. (2019). *World Population Prospects 2019: Highlights.* Department of Economic and Social Affairs, Population Division. Retrieved June 28, 2020, from https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf

USDA. (2015). *Southeast Asia: 2015/16 rice production outlook at record levels*. Commodity Intelligence Report. United State Department of Agriculture.

Xie, K., Li, Y., & Li, W. (2012). Modelling wind speed dependence in system reliability assessment using copulas. *IET Renewable Power Generation, 6*(6), 392-399.

Zhang, L., & Singh, V. P. (2007). Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology, 332*, 93-109. doi: https://doi.org/10.1016/j.jhydrol.2006.06.033

Zhang, L., & Singh, V. P. (2012). Bivariate rainfall and runoff analysis using entropy and copula theories. *Entropy, 14*, 1784-1812. doi: https://doi.org/10.3390/e14091784

Zhang, L., Yang, B., Guo, A., Huang, D., & Huo, Z. (2018). Multivariate probabilistic estimates of heat stress for rice across China. *Stochastic Environmental Research and Risk Assessment, 32*, 3137-3150. doi: https://doi.org/10.1007/s00477-018-1572-7