MATEMATIKA, MJIAM, Volume 42, Number 2, 205–219

Comparative Analysis of Multivariate Copula Models for Extreme Rainfall Simulation in Pahang, Malaysia

Wan Muhammad Haiqal Shah Mohd Sariff, Noor Fadhilah Ahmad Radi*, Sayang Mohd Deni, Nor Azura Md Ghani and Ahmad Zia Ul-Saufie

Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA,
40450, Shah Alam, Selangor, Malaysia.
*Corresponding author: noorfadhilah@uitm.edu.my

Article history
Received: 11 April 2025
Received in revised form: 22 July 2025
Accepted: 24 September 2025
Published on line: 20 July 2026

Abstract Extreme rainfall events, often occurring at a regional scale, imply a degree of interdependence among several variables within the same locality. To replicate this dependency structure, a joint probability analysis utilizing multivariate approach is required. Conventional multivariate probabilistic distributions inherently imply that all variables adhere to the same distribution, which is realistically untrue. Alternatively, this study utilized the copula theory; a robust statistical method that enables the separate modelling of joint distributions and univariate marginals, thereby allowing for the simultaneous modelling and simulation of annual maximum daily rainfall data obtained from three rainfall stations in Pahang. From the modelling perspective, the efficacy of four distinct copula models was illustrated and the goodness-of-fit test indicates that the Gumbel copula is the superior model, followed by the Frank, Clayton, and Skew-t copulas, in that sequence. Moreover, the simulated data was able to be reverted back into meaningful value, allowing direct comparison to be done in terms of mean and variance of simulated and empirical values. The finding suggests all copula models successfully produced simulation outcomes that closely mimic the empirical distributions and characteristics of the observed extreme rainfall. These findings establish the basis for subsequent research on spatial and risk analysis of extreme events, which is essential for effective decision-making and catastrophe mitigation.

Keywords Extreme rainfall; Copula; Archimedean; Elliptical; Skew-t.

Mathematics Subject Classification 46N30, 62G32, 60G10.

1 Introduction

The seemingly significant shift in climate condition on the global scale is garnering the attention and concern of many governmental bodies across many countries in the world, including Malaysia. As a matter of fact, the Sustainable Development Goals (SDG) Roadmap of Malaysia for 2021 to 2025 highlights many climate-related issues pertaining to extreme weather events as obstacles to the country's developmental goals, since the occurrence of these natural disasters impacts both the population and economic wellbeing [1]. Annually, flood events are expected to happen in different part of the country primarily due to the monsoonal effect. In the area of Pahang, thousands of people were affected or displaced in the last few decades due to these inevitable disasters [2]. Montes-Pajuelo et al. [3] and Guo et al. [4] in their works emphasized extreme precipitation as a crucial variable in flood risk analysis and how its modelling is vital in hydrological engineering designs. Consequently, this present study will be focusing on the modelling and simulation of annual extreme rainfall events in the state of Pahang, Malaysia.

Seminal studies focusing on frequency analysis of extreme rainfall events have been shown to work with, but not limited to, annual daily maximum rainfall [3,5]. Typically, the stochastic modelling and simulation of such events can be conducted using various univariate probability distributions, as illustrated in a prior local study by Win and Win [5], which compared eight distinct probability distributions to fit the annual maximum daily rainfall series in the Kuantan River basin, Pahang. This research revolving on analysis of extreme precipitation has identified the Generalized Extreme Value (GEV) distribution as the most suitable model for all rainfall stations within the study area. Twumasiankrah et al. [6] also concludes a similar finding in their study that the GEV model offers the optimal fit for extreme rainfall data in northern Ghana. A recent study by Montes-Pajuelo et al. [3], which also focuses on extreme rainfall; has demonstrated the efficacy of 10 different probability distributions in modelling annual maximum daily rainfall data. The findings indicate that the gamma and Gumbel distributions are among the most suitable models based on goodness-of-fit results. According to prior research, the normal, lognormal, gamma, Gumbel and GEV distributions are recognized as appropriate options for univariate modelling of the annual maximum daily rainfall series [3–6].

While its less complex and easy-to-implement features are appealing, univariate stochastic models suffer from a particular drawback; it cannot portray the dependency between multiple variables. More precisely, Brunner et al. [7] stated that such univariate analysis may be valid on the local scale, but for the relevancy to hold on the regional scale, a multivariate analysis is needed instead. For example, moderate precipitation amount may be observed loosely in different stations across a certain state area due to local conditions, however heavy downpour occurs on a regional scale, therefore all of the stations in the area are highly likely to record extreme observations of similar magnitude and severity. Traditionally, the derivation of the multivariate distribution for hydrologic variables is made under either of these assumptions; (i) the variables follow the same distribution, (ii) the variables are assumed to follow the joint normal distribution or (iii) the variables independent of each other, which may not be the case in real-world application [8]. Therefore, an alternative solution would be to implement the copula theory, a robust statistical tool introduced by Sklar [9] which allows the multivariate model and univariate marginal to be modelled separately, while preserving the dependence structure. Copula can be described as a function that links the multiple instances of univariate distribution functions to form a multivariate distribution function [8].

The application of copulas in hydrology has garnered more attention due to their efficacy in delineating the dependent structure among variables [10]. The Archimedean family of copulas is extensively utilized in hydrology because it effectively captures diverse dependence structures, can be easily generated, and possesses favourable properties of symmetry and associativity [11]. Conversely, rainfall data tend to have positive skewness, indicating that they may be more accurately modelled by a copula from the elliptical family, such as the Skew-t copula [12]. Regrettably, the discourse concerning the utilization of copula for modelling rainfall in the

Malaysian setting remains constrained.

An exemplary study by Yee et al. [13] employed different copula models, including the Clayton and Gumbel copulas from the Archimedean family, as well as the Skew-t copula, to model extreme rainfall in Johor; however, this research strictly addresses the application of bivariate copulas only. In the trivariate context, a couple of studies have highlighted the use of both Archimedean and elliptical copula families for modelling hydrological data, although not exclusively limited to extreme rainfall applications [12, 14, 15]. Furthermore, Radi et al. [12] highlighted that the variance of simulated rainfall events are often lower compared to the observed variance, and it is of particular interest to better replicate the statistical properties of the extreme rainfall events. However, the basis of comparison found in past studies seems to be limited to information criterions of the fitted copula models [13] or correlation coefficient between the simulated and observed value [12], yet, no direct comparison was reported between the generated synthetic data and actual extreme rainfall events.

Therefore, the primary aim of this study is twofold; (i) to investigate the feasibility and practicality of trivariate copula for modelling of extreme rainfall data in the region of Pahang, Malaysia and (ii) to provide insight of each models' performance through direct comparison of the mean and variance of the simulated and observed extreme rainfall data. To achieve this, we will evaluate the efficacy of four copula models; Gumbel, Clayton, Frank and Skew-t. Considering how the properties of copula allow for the joint and marginal distribution to be modelled separately, we shall also discuss on the most suitable probability distribution to fit the rainfall data in each of the observatory stations from the following candidates: the normal, lognormal, gamma, Gumbel and GEV distributions. Most importantly, to illustrate the performance of each model under better clarity, we will also demonstrate the simulation of synthetic extreme rainfall values and observe whether the simulated data mirror or deviate from the empirical distribution. The findings of this study are crucial as an impetus to catalyze future studies to be done by utilizing trivariate copula models for extreme rainfall analysis, which can be extended to the realm of spatial analysis and risk quantification, which is undoubtedly beneficial for better preparedness against the current erratic climate change.

2 Methodology

As previously stated, the most notable characteristic of copula theory is its ability to separately model the joint and marginal distributions of rainfall data. This section will be structured to first address the modelling of the marginal distribution, followed by a discussion of the employed copula models. Subsequently, we will explore on the method for parameter estimation and the evaluation metrics for both the copula and marginal models. However, it is crucial to comprehend the expected characteristics of the data utilized in this study, which will be addressed foremost in this section.

2.1 Study Area and Data

In Peninsular Malaysia, the state of Pahang is not only the largest state, it also accommodates the largest river basin known as the Pahang River Basin [2]. The whole state spans the total area of 35 960 km^2, housing many species of endangered faunas in its rich tropical rainforest [16]. Due to the shielding of the Titiwangsa Range, the state of Pahang (inland region) are less influenced

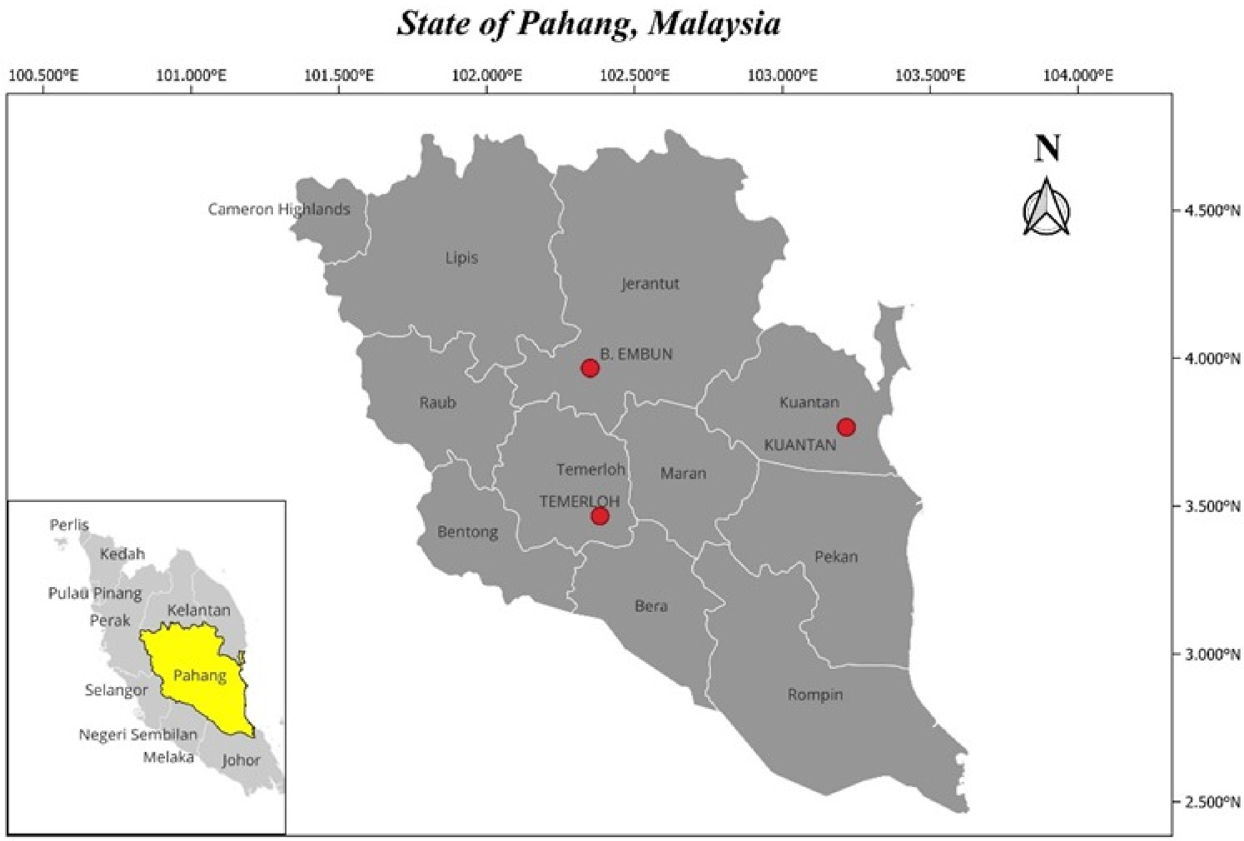

Figure 1: Location of Selected Rainfall Stations in the State of Pahang, Malaysia

by the monsoonal effects brought forth during the Southwest Monsoon (SWM) lasting from May to August, and the Northeast Monsoon (NEM) which typically occurs from November to February each year, resulting in generally less amount of precipitation received [17, 18]. Unfortunately, the state has witnessed several episodes of major flash floods which affected thousands of people, and previously unaffected areas in the state are also experiencing the disaster in more recent cases [2]. Hence, further analysis shall be conducted to better understand the behavior of extreme rainfall in the area.

For this study, we considered three rainfall stations located in this state which are the Kuantan (103.2167°E, 3.7667 °N), Temerloh (102.3833°E, 3.4667 °N) and Batu Embun (102.3500°E, 3.9667 °N) stations as shown in Figure 1 above. Due to limited availability, 40 years (1983 - 2022) length of daily rainfall series recorded in each station were requested from the Malaysian Meteorological Department. The observations are almost complete except for the 0.05% of missing data, which was simply estimated using the inverse distance weighting method [19]. The annual maximum daily precipitation series were then extracted from each station, resulting in a total of three distinct extreme rainfall series with length of 40 elements. Table 1 provides a brief summary of the extreme rainfall series, and it can be seen that the Kuantan station generally recorded the highest daily rainfall amount, followed by the Temerloh and Batu Embun stations.

Table 1: Descriptive Statistics of the Annual Maximum Daily Rainfall Series

Station	Elevation (m a.s.l.)	Average (mm)	Max. (mm)	Min. (mm)
Kuantan	15.2	211.6	407.0	63.1
Temerloh	39.1	108.0	200.1	61.0
Batu Embun	59.5	106.3	175.6	58.0

2.2 Modelling of Marginal Distribution

Prior to the construction of the copula model, the first step is to determine the marginal distribution that best represent each of the extreme rainfall series. The appropriate choice of suitable probability distributions is directly linked to the performance and accuracy of the copula model [4]. From past literatures, a number of possible candidates have been identified namely the normal, lognormal, gamma, Gumbel, and GEV distributions. The probability density functions (PDF) and parameters for each model are provided below in Table 2. Each of the extreme rainfall series will be fitted to every PDF and the best model based on the evaluation metrics will be chosen as the marginal assumption.

Table 2: Summary of Probability Distributions for Marginal Assumptions

Distribution	PDF	Parameter(s)
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),\ x \in \mathbb{R}$	μ = location σ = scale
Lognormal	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right),\ x > 0$	μ = location σ = scale
Gamma	$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}},\ x > 0$	α = shape β = scale
Gumbel	$f(x) = \frac{1}{\sigma} \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right) - \frac{x-\mu}{\sigma}\right),\ x \in \mathbb{R}$	μ = location σ = scale
GEV	$f(x) = \begin{cases} \frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{\xi}} \times \exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right], & \xi \neq 0, \\ \frac{1}{\sigma}\exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)-\frac{x-\mu}{\sigma}\right), & \xi = 0. \end{cases}$ $1+\xi\frac{x-\mu}{\sigma} > 0\ (\xi \neq 0), \qquad x \in \mathbb{R}\ (\xi = 0)$	μ = location σ = scale ξ = shape

2.3 Copula

The work of Sklar [9] introduces the concept of copula, a powerful tool that allows the dependence structure of multiple variable to be modelled without any marginal restriction, where the Sklar's theorem states that for one d-dimensional multivariate distribution function, F, of a random variables $\mathbf{X}_i$ with marginal distribution F_i $(i = 1, 2, \ldots, d)$ can be generally expressed as

$$F(\mathbf{X}_1, \ldots, \mathbf{X}_d) = C(F_1(\mathbf{X}_1), \ldots, F_d(\mathbf{X}_d); \Theta) = C(u_1, \ldots, u_d; \Theta), \tag{1}$$

where, Θ is the copula parameter vector and $F(\mathbf{X}_i) = F(X \leq x)$ is the marginal distribution of $\mathbf{X}_i$ and u_i follows the uniform distribution $u_i \sim U(0, 1)$. Moreover, the joint distribution function of a d-dimensional copula can be represented as a multivariate uniform distribution with the cumulative distribution function C that maps $C[0, 1]$ to $[0, 1]$. The copula technique provides flexibility in selecting marginal distributions while maintaining the dependence structure among numerous variables; yet, no singular copula model can accommodate all data scenarios situation [8]. Therefore, this study will demonstrate the applicability of four distinct trivariate copulas ($d = 3$) from the Archimedean and elliptical families.

2.3.1 Archimedean copula

Application of copulas in the area of hydrology has also been gaining more interest due to its efficiency in describing the dependence structure among variables [10]. For instance, the Archimedean family of copula is widely implemented in the field of hydrology due to its capability of capturing a wide range of dependence structure and the copulas can be easily generated, as well as their desirable properties of symmetricity and associativity [11]. According to Orcel et al. [20], the Archimedean family of copula holds an advantageous property, whereby its bivariate form can be easily extended to the d-dimensional version, given by

$$C(u_1, u_2, \cdots, u_d) = \phi^{-1}(\phi(u_1), \phi(u_2), \cdots, \phi(u_d));\ u_1, u_2, \cdots, u_d \in [0, 1], \tag{2}$$

where, $u_i = F_i(x_i)$ is the marginal cumulative distribution function of the variable represented by $X_i (i = 1, 2, \ldots, d)$, and $\phi(\cdot)$ is identified as the generator function of the copula; $\phi^{-1}(\cdot)$ is inverse of such function. In the context of this study, the trivariate form of Clayton (3), Gumbel (4) and Frank (5) copula are given by [15, 20] as

$$C(u_1, u_2, u_3) = \left[u_1^{-\frac{1}{\theta}} + u_2^{-\frac{1}{\theta}} + u_3^{-\frac{1}{\theta}} - 2\right]^{-\theta}, \tag{3}$$

$$C(u_1, u_2, u_3) = \exp\left(-\left[(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta} + (-\ln u_3)^{\theta}\right]^{\frac{1}{\theta}}\right), \tag{4}$$

$$C(u_1, u_2, u_3) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)(e^{-\theta u_3} - 1)}{e^{-\theta} - 1}\right). \tag{5}$$

2.3.2 Elliptical copula

Similar to Archimedean copula, elliptical copula is also capable of preserving the dependence structure of related variables; but the main difference is that they are constructed from the multivariate of the marginal distribution, with the following relationship [8].

$$\begin{aligned} H(y_1, y_2, y_3) &= C\left(F_1(y_1), F_2(y_2), F_3(y_3)\right), \\ &= F\left(F_1^{-1}(v_1), F_2^{-1}(v_2), F_3^{-1}(v_3)\right), \\ &= C(v_1, v_2, v_3), \end{aligned} \tag{6}$$

where, $F_i(y_i)$ are the cumulative distribution function (CDF) of the marginal random variable, y_i and $F^{-1}(\cdot)$ are the quantile functions, or the inverse of CDF. Zakaria et al. [21] proposed that rainfall data is skewedly distributed, therefore the Skew-t distribution is suitable to be employed for analysing the asymmetric tail dependence. The probability density function of multiple variables, $x \in \Re^d$, governed by the multivariate Skew-t distribution can be shown as

$$f_{MST}(x; \mu, \Omega, \gamma, \upsilon) = 2t_d(x-\mu; \Omega, \upsilon) \cdot T\left\{\gamma^T \Omega^{-1}(x-\mu)\sqrt{\frac{\upsilon+d}{Q(x)+\upsilon}}; \upsilon+d\right\}, \tag{7}$$

where, $Q(x) = (x-\mu)^T \Omega^{-1}(x-\mu)$, γ is the skewness parameter and μ is the location parameter. The d-dimensional standard student's t distribution with $\mu = 0$, Ω scale matrix and υ degrees of freedom is defined as [22]

$$t_d(x; \Omega, \upsilon) = \frac{\Gamma\left(\frac{\upsilon+d}{2}\right)}{\sqrt{|\Omega|}(\upsilon\pi)^{\frac{d}{2}}\Gamma\left(\frac{\upsilon}{2}\right)}\left(1+\frac{Q(x)}{\upsilon}\right)^{-\frac{\upsilon+d}{2}}, \ x \in \Re^d, \tag{8}$$

where, $T(\cdot)$ is the cumulative distribution function of Student's t distribution with $\upsilon+d$ degrees of freedom. An extensive discussion on the construction and application of the trivariate Skew-t copula can be found in the work of Radi et al. [12].

2.3.3 Maximum likelihood estimation

The Maximum Likelihood Estimation (MLE) approach is frequently utilized for parameter estimation of the PDF [23]. Let a random variable X has the PDF $f(x; a_1, a_2, \cdots, a_m)$ with unknown parameters a_i, $i = 1, 2, \ldots, m$. The likelihood of the sample can be determined if n random observations chosen from $f(x; a_1, a_2, \cdots, a_m)$ is maximized, and the determined estimation of the parameters are considered as the MLE. The loglikelihood function is given by

$$L^*(a_1, a_2, \cdots, a_m) = \ln \prod_{i=1}^{n} f(x_i; a_1, a_2, \cdots, a_m) = \sum_{i=1}^{n} \ln f(x_i; a_1, a_2, \cdots, a_m). \tag{9}$$

Orcel et al. [20] in their work demonstrated how the same concept can be extended for copula parameter estimation, while Yoshiba [22] in their work extensively discusses on MLE for Skew-t copula specifically.

2.3.4 Model validation and selection

To reiterate, the choice of marginal assumption has a substantial impact towards the copula performance, hence, the appropriateness of the estimated univariate marginal will first be assessed using the Kolmogorov-Smirnov (KS) test [24]. The KS test measures the vertical difference between the empirical and theoretical distribution, with the null hypothesis that states both distributions are similar. The alternative hypothesis contrarily suggests that there exists significant difference between the empirical and fitted model. Furthermore, to choose the best model, the Akaike Information Criterion (AIC) will be employed [13]. AIC measures the trade-off between the model complexity (number of parameters) and the estimated likelihood, and the best model is chosen based on lowest AIC value. Both measures of the KS test statistics, D_n and AIC are given as (10) and (11), respectively, by

$$D_n = \max_{1\leq i\leq n}(\hat{\delta}_i),\ \hat{\delta}_i = \max\left[\frac{i}{n} - F_0(x_i;\hat{\delta}_i) - \frac{i-1}{n}\right], \tag{10}$$

$$AIC = -2\ln L + 2k. \tag{11}$$

Further error analysis on the copula model will also be assessed through the Root Mean Squared Error (RMSE) and graphical cumulative distribution. Song and Singh [24] defines RMSE as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[P_c(i) - P_o(i)\right]^2}, \tag{12}$$

where, n is the sample size, $P_c(i)$ and $P_o(i)$ is the computed and observed probability, respectively. The result of this analysis will help to illustrate each of the copula model's performance.

3 Results and Discussions

Foremost, this section shall present the relevant findings pertaining the fitting of extreme rainfall data to various univariate marginal models. To construct a proper copula model, a reliable choice needs to be made for the marginal assumption of the variables. As shown in Table 3, the MLE approach was able to provide an estimation for all PDF parameters, within a certain level of tolerance as measured by the corresponding standard error (denoted S.E). Further analysis was conducted through regular bootstrapping approach to measure the upper and lower bound of the estimates, given by the 95% confidence interval. The result shows that all point estimation of the parameters lies in the 95% confidence interval, indicating that the MLE point estimator is not significantly different from the actual parameter for the population distribution.

Moreover, based on 95% confidence level for the KS test portrayed in Table 4, all PDF was considered valid in representing the annual extreme rainfall series in every station. This premise is drawn from the fact that all recorded p-value are greater than 0.05, indicating no significant difference between the empirical and fitted theoretical distribution. However, we still need an evaluation metric to choose the best model, and the AIC value will be the basis comparison for this purpose. Measuring the trade-off between number of parameters and the realized likelihood, a lower AIC value indicates that a particular model is able to provide a good estimate of the empirical distribution despite having less numbers of parameter.

Table 3: MLE Point Estimator, Standard Error and 95% Confidence Interval of the Parameters for Fitted Distribution of Annual Maximum Rainfall

Station	PDF	Estimated Parameters					
		Location		Scale		Shape	
		MLE (S.E)	95% C.I	MLE (S.E)	95% C.I	MLE (S.E)	95% C.I
Kuantan	Normal	211.55 (13.58)	(186.40,237.46)	85.91 (9.61)	(66.43,103.4)	-	-
	Lognormal	5.27 (0.07)	(5.14,5.40)	0.43 (0.05)	(0.32,0.51)	-	-
	Gamma	-	-	0.03 (0.01)	(0.02,0.05)	5.93 (1.28)	(4.15,10.03)
	Gumbel	171.09 (11.83)	(148.83,196.15)	70.92 (8.77)	(53.54,87.4)	-	-
	GEV	173.77 (13.26)	(147.88,201.3)	72.57 (9.7)	(53.02,90.61)	-0.07 (0.14)	(-0.37,0.19)
Temerloh	Normal	107.96 (5.68)	(97.1,118.39)	35.9 (4.01)	(27.75,43.88)	-	-
	Lognormal	4.63 (0.05)	(4.53,4.72)	0.32 (0.04)	(0.25,0.38)	-	-
	Gamma	-	-	0.09 (0.02)	(0.06,0.15)	9.92 (2.18)	(6.74,16.52)
	Gumbel	91.72 (4.45)	(83.58,100.95)	26.78 (3.45)	(19.93,33.19)	-	-
	GEV	89.86 (4.75)	(81.42,99.46)	25.18 (3.74)	(18.2,31.55)	0.13 (0.17)	(-0.15,0.42)
Batu Embun	Normal	106.28 (4.26)	(97.97,114.38)	26.97 (3.01)	(20.79,32.53)	-	-
	Lognormal	4.63 (0.04)	(4.56,4.72)	0.26 (0.03)	(0.19,0.31)	-	-
	Gamma	-	-	0.15 (0.03)	(0.10,0.25)	15.65 (3.46)	(11.12,27.23)
	Gumbel	93.39 (3.86)	(86.66,101.7)	23.11 (2.82)	(17.55,28.38)	-	-
	GEV	95.01 (4.31)	(87.21,104.3)	24.01 (3.1)	(17.68,29.76)	-0.13 (0.12)	(-0.40,0.08)

Table 4: Summary of Goodness-of-fit Test for Marginal Distribution

Station	PDF	KS Test		AIC
		Test Statistics, D_n	p-value	
Kuantan	Normal	0.1202	0.569	473.7792
	Lognormal	0.0755	0.9635	471.1028
	Gamma	0.0895	0.8773	470.0105*
	Gumbel	0.0882	0.8877	470.5789
	GEV	0.0961	0.8197	472.3354
Temerloh	Normal	0.1497	0.3009	403.9701
	Lognormal	0.0919	0.8573	395.7571
	Gamma	0.1127	0.6483	397.5024
	Gumbel	0.0793	0.9456	395.4870*
	GEV	0.0719	0.9765	396.7537
Batu Embun	Normal	0.0701	0.9893	381.0815
	Lognormal	0.0765	0.9735	379.1009
	Gamma	0.0674	0.9933	379.0480*
	Gumbel	0.092	0.8872	379.8646
	GEV	0.0688	0.9915	380.9839

The model with the least AIC value was marked with an asterisk, and the result suggests that the extreme rainfall series in Kuantan and Batu Embun stations are best modelled with the gamma distribution, and Gumbel PDF for the Temerloh station. This result is parallel to past literatures whereby the gamma distribution was widely used as the marginal distribution for copula modelling of hydrological variables and the Gumbel distribution was found to perform well in modelling precipitation data [3–6, 12]. Based on this findings, the extreme rainfall series are then transformed into uniform unit by using the suitable marginal distribution before initiating the copula modelling stage.

Next, the uniformized data was then modelled using four distinct copula models: Frank, Clayton, Gumbel and Skew-t. The AIC result in Table 5 suggests the Gumbel copula to be the best model as it produces the least AIC value which measures at 1.4727, followed by the Frank, Clayton and Skew-t copula with the largest AIC of 9.6720. This result is also parallel to the findings reported in prior work, where the Skew-t copula was found to be the least performing model [13]. In general, the performance measure of the Archimedean copula portrays minute difference of RMSE value; 0.0017 between the best and worst model in the family. This indicates that the Archimedean copula is indeed suitable for modelling the extreme rainfall data as suggested in past studies [13–15]. Furthermore, the Skew-t copula were found to produce smaller error as compared to the Clayton copula. While this may seem counterintuitive, it can be explained through the Skew-t model's higher loglikelihood;indicating that the model better fits the empirical data, yet it is penalized due to its complexity ($k = 7$), thus resulting to a higher AIC value.

Table 5: Summary of Parameter Estimate and Error Measure of Trivariate Copula Models

Copula	**Estimated Parameter**	**Loglikelihood**	**AIC**	**RMSE**
Gumbel	$\theta = 1.0920$	0.2636	1.4727*	0.0227*
Frank	$\theta = 0.6839$	-0.0421	2.0842	0.0233
Clayton	$\theta = 0.1104$	-0.4458	2.8915	0.0244
Skew-t	$\rho_1 = -0.1353$, $\rho_2 = 0.3164$, $\rho_3 = -0.2319$ $\gamma_1 = -0.6352$, $\gamma_2 = 0.7050$, $\gamma_3 = -0.6184$ $\upsilon = 40.8614$	2.1640	9.6720	0.0236

Thus, so far, we have established that all four copula models performs satisfactorily for modelling the dependency structures of extreme rainfall data in the area of Pahang, Malaysia. For a deeper insight on their applicability and suitability in representing the empirical extreme rainfall distribution, a simulation study is necessary using the fitted copula models. Towards this end, the property of the simulated data will be illustrated under two conditions; where the number of generated observations, n, is (i) sufficiently large to represent the population and (ii) sampled at smaller scale for practical application such as imputations or replacement. Without loss of generality, we first simulated 10,000 synthetic extreme precipitation data to satisfy the first condition using each copula model and calculated the mean and variance, as presented in Table 6. To highlight, the simulated values are initially in uniform unit, thus to revert back to a meaningful value, the respective quantile function of the marginal distribution was applied. The finding from this simulation study is well within expectation, as the extreme rainfall series simulated from the fitted Gumbel copula consistently produces mean value close to the observed data for all rainfall stations. It is evident that the copula models are able to closely mimic the empirical distribution of the annual maximum daily rainfall data in the state of Pahang.

Table 6: Comparison Between Mean and Variance of the Observed and Simulated Annual Maximum Rainfall Series

Station	Statistics	Observed	Simulated			
			Gumbel	Frank	Clayton	Skew-t
Kuantan	Mean	211.55	211.50	210.83	211.60	210.88
	Variance	7569.80	7659.86	7363.20	7658.52	7487.06
Temerloh	Mean	107.96	107.48	107.05	106.56	106.62
	Variance	1321.73	1243.32	1172.97	1187.02	1144.51
Batu Embun	Mean	106.28	106.09	105.73	106.08	106.94
	Variance	745.82	721.75	728.09	711.09	739.42

Further visualization of the results can be seen in Figure 2, which portrays the comparison of the fit between measured probability density of the empirical and various theoretical copula. The graphical finding also shows no contradiction as the copula models all seem to be closely fitted to the observed empirical copula. This further solidifies the premise that all of the copula model can be considered appropriate for modelling and simulation of extreme rainfall. However, for application purpose, the sample size of n = 10 000 may not be contextually suitable, considering the data used in this study is sampled annually. Therefore, to better understand the property of the simulated data on a smaller scale, we repeated the sampling of simulated data with equal size to the observed data (n =40), half its size (n =20) and twice its size (n = 80) through regular bootstrapping process.

The 95% confidence interval of mean and variance for the simulated sample of size n was constructed based on 1000 bootstrapped sample, presented in Table 7. Foremost, it is evident that for all sample sizes, the variance of simulated sample from the Skew-t model is significantly different from the observed variance of annual maximum rainfall in Kuantan. The upper bound for the variance interval of simulated sample from Skew-t model is consistently smaller than the observed variance in Kuantan, (refer observed variance in Table 6). For n = 80, the Gumbel model recorded significant difference for simulated mean in Kuantan, while the Clayton model produced simulated variance that is significantly different from the observed value in Batu Embun. In this context, only the Frank model is capable of generating simulated annual maximum rainfall series with similar statistical properties to the empirical data at all sample sizes. From this result, it can be inferred that the Frank copula may be preferred for simulating smaller samples. Ultimately, we can reaffirm that these findings are similar to the work of Yee et al. [13] in which: (i) the Gumbel copula is found to be the best for modelling extreme rainfall data and (ii) copulas of the Archimedean family generally performs better as compared to the Skew-t copula in this context. Most importantly, the findings presented provides enough evident to vouch for the feasibility of applying trivariate copula approach for modelling and simulation of annual extreme rainfall series in Pahang, Malaysia.

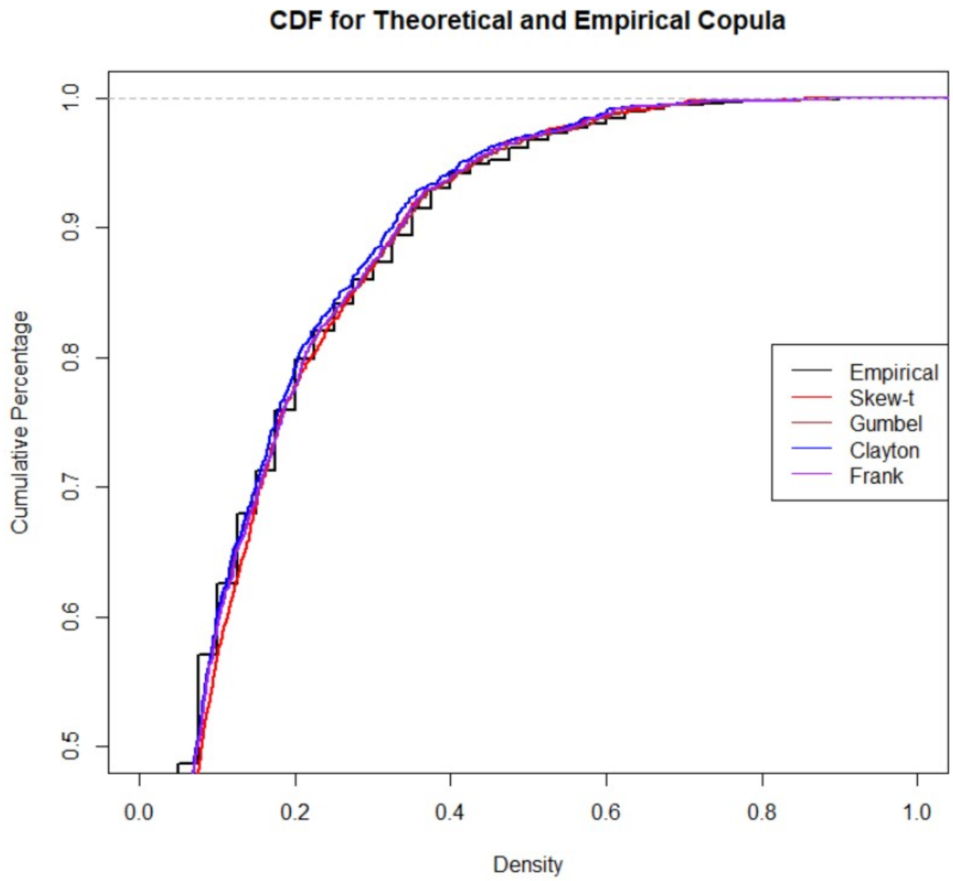

Figure 2: Graphical Representation of the Theoretical and Empirical Copula Function

4 Conclusion

The erratic and unpredictable behavior of climactic disasters is a growing concern in the state of Pahang, Malaysia. Uncovering this pattern will require a thorough statistical analysis to be done relating to the extreme rainfall events; which can be achieved through modelling and simulation of the data. This study focused in illustrating the efficacy of distinct trivariate copula models from the Archimedean and elliptical families, which allows for simultaneous simulation of three variables. The findings suggest that the extreme rainfall distributions in Kuantan and Batu Embun station are most suited to be modelled with gamma marginal distribution, while in Temerloh, the extreme rainfall is better represented by the Gumbel marginal. Among the deployed copula models, the Gumbel copula boasts the best performance as it produces the least error, followed by Frank, Clayton and Skew-t copula. Through simulation and graphical representation, we have provided sufficient evident that all copula models are able to mirror the actual distribution of observed extreme rainfall events in the study region; yet for simulating smaller sample sizes, the Frank copula emerges as the best model. We can conclude that the findings of this study are parallel to those of the preceding studies, with the major difference that higher dimensional copulas were applied instead. One limitation that can be highlighted would be the seemingly rigorous process in determining the suitable marginal distribution for

Table 7: Mean and Variance of Simulated Sample with Variying Sizes at 95% Confidence Interval

Sample Size	Station	Statistics	Gumbel	Frank	Clayton	Skew-t
n = 20	Kuantan	Mean	(197.91,288.17)	(173.62,252.42)	(185.23,257.97)	(162.69,221.74)
		Variance	(4004.65,19427.22)	(3939.17,12327.44)	(3305.17,11420.88)	(2309.66,6770.34)
	Temerloh	Mean	(86.65,111.07)	(96.35,121.33)	(88.79,113.91)	(97.21,147.19)
		Variance	(296.97,1348.01)	(379.68,1501.48)	(310.17,1501.79)	(525.23,7739.57)
	Batu Embun	Mean	(96.43,117.47)	(95.2,114.48)	(92.7,115.07)	(92.47,112.82)
		Variance	(321.05,789.88)	(215.75,817.72)	(325.12,881.74)	(206.42,919.37)
n = 40	Kuantan	Mean	(210.96,266.4)	(174.12,232.07)	(191.17,243.18)	(174.32,216.01)
		Variance	(4362.99,12957.42)	(5344.31,10417.9)	(3862.57,9639.18)	(3091.06,6333.72)
	Temerloh	Mean	(92.89,114.41)	(95.57,114.3)	(99.00,122.63)	(99.72,126.98)
		Variance	(586.75,1994.63)	(547.43,1373.68)	(719.72,2192.86)	(665.79,4647.45)
	Batu Embun	Mean	(93.72,108.56)	(101.07,117.74)	(94.07,110.99)	(100.19,115.6)
		Variance	(378.78,796.98)	(450.02,1059.03)	(476.35,858.34)	(365.17,898.14)
n = 80	Kuantan	Mean	(222.34,264.1)	(183.77,219.15)	(203.28,243.76)	(187.1,216.94)
		Variance	(5832.87,12188.56)	(4760.7,8254.6)	(6175.61,12278.29)	(3802.02,6797.41)
	Temerloh	Mean	(99.46,112.39)	(102.08,117.36)	(99.16,113.86)	(100.74,119.64)
		Variance	(636.48,1368.64)	(719.53,1763.98)	(700.96,1716.58)	(894.45,3070.51)
	Batu Embun	Mean	(97.41,108.15)	(101.26,112.88)	(96.18,106.71)	(99.93,110.07)
		Variance	(454.09,824.13)	(530.79,951.65)	(443.27,703.94)	(417.00,761.73)

each variable. Furthermore, the fitted copula models can also be deployed for spatial analysis or quantification of risk.

5 Acknowledgment

Authors acknowledge the Ministry of Higher Education (MOHE) for funding under the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2022/STG06/UITM/02/5). Authors would also like to extend our gratitude to the Malaysian Meteorological Department for the daily rainfall record provided which made this study possible.

References

[1] Malaysia Ministry of Economy. *SDG Roadmap for Malaysia*. Malaysia: Ministry of Economy. 2024.

[2] Kamarudin, M. K. A., Toriman, M. E., Wahab, N. A., Abu Samah, M. A., Abdul Maulud, K. N., Hamzah, F. M., Mohd Saudi, A. S. and Sunardi. Hydrological and climate impacts on river characteristics of Pahang River Basin, Malaysia. *Heliyon*. 2023. 9(11): e21573.

[3] Montes-Pajuelo, R., Rodríguez-Pérez, A. M., López, R. and Rodríguez, C. A. Analysis of probability distributions for modelling extreme rainfall events and detecting climate change: Insights from mathematical and statistical methods. *Mathematics*. 2024. 12(7): 1093.

[4] Guo, A., Chang, J., Wang, Y., Huang, Q. and Guo, Z. Maximum entropy-copula method for hydrological risk analysis under uncertainty: A case study on the Loess Plateau, China. *Entropy*. 2017. 19(11): 609.

[5] Win, N. L. and Win, K. M. The probability distributions of daily rainfall for Kuantan River Basin in Malaysia. *International Journal of Science and Research*. 2014. 3(8): 977–983.

[6] Twumasiankrah, S., Pels, W. A. and Nadarajah, S. Modeling rainfall extremes along the coastal and northern parts of Ghana. *Mausam*. 2024. 75(2): 433–442.

[7] Brunner, M. I., Furrer, R. and Favre, A. C. Modeling the spatial dependence of floods using the Fisher copula. *Hydrology and Earth System Sciences*. 2019. 23(1): 107–124.

[8] Yusof, F. and Jamaludin, S. S. S. *Copula Modelling and Its Application*. Johor Bahru: Universiti Teknologi Malaysia Press. 2023.

[9] Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*. 1959. 8: 229–231.

[10] AghaKouchak, A., Bárdossy, A. and Habib, E. Copula-based uncertainty modeling: Application to multisensor precipitation estimates. *Hydrological Processes*. 2010. 24(15): 2111–2124.

[11] Ganguli, P. and Reddy, M. J. Probabilistic assessment of flood risks using trivariate copulas. *Theoretical and Applied Climatology*. 2013. 111(1–2): 341–360.

[12] Radi, N. F. A., Zakaria, R. and Satari, S. Z. Generating monthly rainfall amount using multivariate skew-t copula. *Journal of Physics: Conference Series*. 2017. 890(1): 012133.

[13] Yee, K. C., Suhaila, J., Yusof, F. and Mean, F. H. Bivariate copula in Johor rainfall data. *AIP Conference Proceedings*. 2016. 1750: 020038.

[14] Latif, S. and Mustafa, F. Bivariate joint distribution analysis of the flood characteristics under semiparametric copula distribution framework for the Kelantan River Basin in Malaysia. *Journal of Ocean Engineering and Science*. 2021. 6(2): 128–145.

[15] Latif, S. and Mustafa, F. Trivariate distribution modelling of flood characteristics using copula function—A case study for Kelantan River Basin in Malaysia. *AIMS Geosciences*. 2020. 6(1): 92–130.

[16] Ghani, N. A. A., Mohamad, N. A. and Teo, W. H. Rainfall analysis to determine the potential of rainwater harvesting site in Kuantan, Pahang. *ARPN Journal of Engineering and Applied Sciences*. 2016. 11(11): 7264–7268.

[17] Wong, C. L., Venneker, R., Uhlenbrook, S., Jamil, A. B. M. and Zhou, Y. Variability of rainfall in Peninsular Malaysia. *Hydrology and Earth System Sciences Discussions*. 2009. 6(4): 5471–5503.

[18] Suhaila, J. Tweedie models for Malaysia rainfall simulations with seasonal variabilities. *Journal of Water and Climate Change*. 2023. 14(10): 3648–3670.

[19] Fung, K. F., Chew, K. S., Huang, Y. F., Ahmed, A. N., Teo, F. Y., Ng, J. L. and Elshafie, A. Evaluation of spatial interpolation methods and spatiotemporal modeling of rainfall distribution in Peninsular Malaysia. *Ain Shams Engineering Journal*. 2022. 13(2): 101571.

[20] Orcel, O. and Sergent, P. Trivariate copula to design coastal structures. *Natural Hazards and Earth System Sciences*. 2021. 21(1): 239–260.

[21] Zakaria, R., Metcalfe, A., Piantadosi, J., Boland, J. and Howlett, P. Using the skew-t copula to model bivariate rainfall distribution. *The ANZIAM Journal*. 2010. 51: 231.

[22] Yoshiba, T. Maximum likelihood estimation of skew-t copulas with its applications to stock returns. *Journal of Statistical Computation and Simulation*. 2018. 88(13): 2489–2506.

[23] Nashwan, M. S., Ismail, T. and Ahmed, K. Non-stationary analysis of extreme rainfall in Peninsular Malaysia. *Journal of Sustainability Science and Management*. 2019. 14(3): 17–34.

[24] Song, S. and Singh, V. P. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data. *Stochastic Environmental Research and Risk Assessment*. 2010. 24(3): 425–444.