

Bivariate Copula Model for selected stations in Peninsular Malaysia rainfall data

Siti Rohani Mohd Nor^{1*}, Lee Jun Heng², Rahmah Mohd Lokoman¹, Nurul Syuhada Samsudin¹, Siti Mariam Norrulashikin¹, Nur Arina Bazilah Kamisan¹, and Nur Syazwin Mansor¹

¹Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

²Emergent Computing, Faculty of Computing, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia

ARTICLE HISTORY

Received : 20 September 2025

Accepted : 27 November 2025

Online : 30 September 2026

KEYWORDS

copula,
bivariate,
rainfall,
stations,
Archimedean

✉ * CORRESPONDING AUTHOR

Dr. Siti Rohani Mohd Nor
Department of Mathematical
Sciences, Faculty of Science,
Universiti Teknologi Malaysia, 81310
Johor Bahru, Johor, Malaysia.
Email: sitirohani@utm.my

ABSTRACT

Malaysia faces a high risk of flooding nearly every year, driven by seasonal monsoons and sudden flash floods. These events result in significant economic damage; in 2021 alone, as recorded by Department of Statistics Malaysia, total losses reached approximately RM 6.1 billion. Hence, predicting extreme rainfall episodes is critical for the Malaysian governments to give early warning to authorities before the severe floods happen. Without such alerts, the government faces a massive financial burden to mitigate and repair flood-related damages. Developing reliable rainfall prediction models is therefore essential for the Malaysian government to issue timely alerts. Such systems rely on the analysis of historical rainfall data to provide authorities with a clear approximation of future rainfall intensity and volume, thereby reducing the financial burden of post-flood recovery. Thus, this study aims to predict rainfall data of the selected stations in Peninsular Malaysia from 2015 to 2021 by using bivariate copula model, specifically the Elliptical and Symmetric Archimedean Copula models. The selected monitoring stations are chosen from four different states which are Johor, Selangor, Kuala Terengganu and Kelantan. These four stations were chosen for their proximity to flood-prone zones, specifically regarding monsoon and flash flood risks. Bivariate copula model was employed in this study since it could capture the complex relationship between two different rainfall stations data with different distributions. Thus, the model can provide a more flexible and accurate representation of how heavy rainfall at one location relates to another. Before the copula model is selected, the distribution analysis is performed on the rainfall data and it is found that the lognormal distribution gives the best fit for the rainfall data for all stations. Then, according to Akaike Information Criterion, Clayton Copula model outperformed the other copula models, including Frank, Gumbel, Gaussian and *t* student copulas. The Clayton model demonstrated a better ability to capture the bivariate dependencies between stations, showing a substantial improvement in fit ranging from 25% to 230% over the other copula models. The results indicate that the Clayton Copula model can be effectively used by authorities, particularly government agencies, to predict reliable rainfall data by capturing inter-station dependencies. This prediction can be served as a tool to make better informed decision making particularly related to managing flood risks

© 2026 UMK Publisher. All rights reserved.

1. INTRODUCTION

In 2025, Southeast Asia was battered by devastating widespread flooding triggered by a trio of tropical cyclones: Senyar, Koto, and Ditwah. The inclusion of a rare Malacca Strait Cyclone further intensified the disaster, severely impacting Malaysia and its neighbours, including Indonesia, Thailand, and the Philippines (Levesque, 2025; Ahmad et al., 2025). As reported by Mai (2025), the catastrophe resulted in a death toll exceeding 1,350 individuals. Furthermore, the destruction of hundreds of thousands of residences led to the displacement of millions of people across the region. The development of Cyclone Senyar within the Strait of Malacca constitutes an unprecedented meteorological phenomenon, marking the first recorded instance of its kind in 135 years

(Sharma, 2025). More than that, according to Mohd Hisham Mohd Anip, the Director-General of Malaysia Meteorological Department, this is the first recorded instance of a tropical storm intensifying to such a high strength within this specific region (The Straits Times, 2025). This event represents a significant departure from historical climate patterns, as the region has traditionally been considered outside the typical trajectory for tropical cyclones. As mentioned by Buakamsri (2025), Cyclone Senyar serves as a critical indicator that the equatorial belt is no longer safe from the sudden severe cyclonic activity. The emergence of such storms implies that Southeast Asia's current meteorological risk assessments are outdated and fail to reflect the region's evolving climate reality. Hence, the current predictive risk assessments should be revised to better reflect the current climate volatility.

Based on the information obtained from Centre for Excellence in Disaster Management and Humanitarian Assistance, Malaysia experienced more frequent flooding compared to other Asian countries, particularly between July 2012 and January 2019 (Levesque, 2025). As 2025 approaches, floods in Malaysia have become more severe, largely due to Northeast Monsoon season and increased rainfall. While the rainy season typically occurs around November to March (Malaysian Meteorological Department, 2025), recent patterns have become less predictable, possibly as a result of climate change (Malay Mail, 2024). For instance, in 2021, Peninsular Malaysia experienced a 'once-in-100-years flood'. Although seasonal flooding is most frequent in East Coast states like Kelantan and Terengganu (Leeonis et al., 2025), experts observed that this time, the central and west regions were also severely affected such as Selangor, Negeri Sembilan Kuala Lumpur, Melaka and Pahang.

The severe flooding between late 2021 and early 2022 caused unprecedented disruption, impacting more than 125,000 people evacuated and resulting in financial losses of RM 6.5 billion (Ahmad et al., 2025). These figures represent one of the most significant flood-related catastrophes in Malaysia's recent history. The state of Johor remains highly susceptible to the recurring threat of flooding within Peninsular Malaysia. Data from the 2006 and 2007 disasters highlight this vulnerability, as these specific events claimed 18 lives and necessitated approximately RM 1.1 billion in state-funded relief and recovery efforts (Ahmad et al., 2025). Furthermore, Johor Bahru experienced its most severe flash flooding in a decade during 2022 (The Malaysian Reserve, 2022). This trend of escalating intensity continued into 2025, when torrential rains impacted 27 areas under the Johor Baru City Council; the Johor Bahru Mayor characterized this recent crisis as the most significant flooding the city has faced in 30 years (New Straits Times, 2025). These observations identify that there are two major factors that causes the floods in Malaysia which are monsoonal seasons and rapid urbanization with inadequate drainage system (Ahmad et al., 2025). Hence, this study aims to study the selected stations from the states that are prone to flooding due to these two factors which are Johor, Selangor, Kelantan and Terengganu. The objective is to evaluate the statistical dependency between these stations, which exhibit varying rainfall characteristics driven by these distinct factors. By examining these relationships, the study seeks to determine how different flood drivers interact across diverse geographical landscapes.

This study employs a bivariate copula model to improve rainfall prediction by capturing the dependency between the rainfall stations in Peninsular Malaysia. The

dependence structure is considered in this study by choosing regions that are prone to flooding, with the expectation that incorporating the correlation between the two stations will lead to more accurate predictions. With the incorporation of joint behaviour between stations, the underlying spatial relationship and tail dependencies can be captured, in which the traditional model might be overlooked. This would allow for more reliable prediction of extreme events, especially when rainfall from one station is strongly correlated to rainfall from another station. This is supported by a study conducted by Yee et al. (2014), which concluded that the traditional bivariate model is not suitable for modelling rainfall between stations due to the presence of inter-station dependency.

Other than that, unlike traditional standard bivariate model, copula model is more flexible in terms as they do not require the marginal distributions of each variable to be of the same type. This is important since the distributions of rainfall data from each station might be different due to their distinct geographical locations. There are many studies have used copulas in hydrology. Several studies have explored the suitability of different copula models in capturing the dependence structure of hydroclimatic data. These include studies by Yee et al. (2014), Ariff et al. (2012), Kamaruzaman et al. (2019), Dzupire et al. (2020), and Jafry et al. (2023). Yee et al. (2014) reported that the Gumbel copula provided the best fit for rainfall data in the southern region. Ariff et al. (2012) found that both Frank and Gaussian copulas were effective in capturing the dependence between rainfall intensity and duration. Similarly, Kamaruzaman et al. (2019) applied several copulas, including Gaussian, Clayton, Frank, Joe, and Gumbel, to data from the northern and eastern regions, and concluded that the Frank copula offered the best fit. Dzupire et al. (2020) identified a significant relationship between temperature and rainfall and successfully modelled their dependence using the Frank copula. In addition, Jafry et al. (2023) found that the Frank copula was the most suitable model for representing the dependence structure of rainfall data in the Johor River Basin. These articles highlight that copula models are effective in capturing complex dependencies between data with varying distributions, without the need to assume normality or linear relationships. This capability is particularly vital in the face of increasing global extreme weather events, as rainfall patterns are becoming more erratic and are less likely to follow normal distribution.

Unlike Yee et al. (2014) and Jafry et al. (2023), who applied a copula model to rainfall data from different stations in Johor, this study utilizes the copula model on rainfall data from various flood-prone states across Peninsular Malaysia. Other researchers have applied copula models to different types of hydrological data, rather than rainfall, and found that

the simulated results closely matched the actual observation. Therefore, this study aims to incorporate the dependency structure between rainfall stations that are prone to flooding in Peninsular Malaysia by using bivariate copula. By leveraging the dependence structure between rainfall amounts in multiple states, these models have the potential to enhance rainfall prediction accuracy, support water resource management, inform flood prevention strategies, and facilitate climate adaptation efforts. This prediction can be served as a tool to make better informed decision making particularly related to managing flood risks.

2. MATERIALS AND METHODS

2.1 Study area

This study analyses rainfall data from 2015 to 2021 from several selected stations with flood prone areas in Peninsular Malaysia which are located in states like Johor, Selangor, Kelantan and Terengganu. According to Buslima et al (2018), the states of Kelantan and Terengganu are particularly vulnerable to flooding due to their location at the North side, which exposes them to Northeast Monsoon season. In comparison, states like Johor and Selangor are more susceptible to flash floods, largely driven by rapid urban growth and insufficient drainage infrastructure. Based on the nature of these four states that are flood-prone, 12 rainfall stations were chosen from them. There are 12 rainfall stations that were applied to bivariate copula model which were: Gong Kedak, Kota Bharu and Kuala Krai (Kelantan), Kerteh and Kuala Terengganu (Terengganu), Petaling Jaya, Subang, and KLIA (Selangor and Batu Pahat, Kluang and Mersing (Johor). The analysis was carried out using R software. The geographical details of these 12 rainfall stations are shown in Table 1 below:

Table 1: Geographical details of 12 rainfall stations.

Station	Latitude	Longitude	Elevation
Gong Kedak (GK)	5.8	102.5	6.0m
Kota Bharu (KB)	6.17	102.3	4.4m
Kuala Krai (KK)	5.53	102.2	68.3m
Kerteh (KRT)	4.54	103.43	3.7m
Kuala Terengganu (KT)	5.38	103.1	5.2m
Petaling Jaya (PJ)	3.1	101.65	60.8m
Subang (SBG)	3.12	101.55	16.5m
KLIA	2.73	101.7	16.1m
Batu Pahat (BP)	1.87	102.98	6.3m
Kluang (KLG)	2.02	103.32	88.1m
Mersing (MRSG)	2.45	103.83	43.6m
Senai	1.63	103.67	37.8m

2.2 Methods

2.2.1 Kendall's tau Correlation

Kendal's tau is used to capture the existence of correlation between data. Instead of Pearson, Kendal's tau is chosen since it is non-parametric and hence does not assume

specific distribution for the data. It is widely used in various fields, including social sciences, environmental studies, and epidemiology, where variables may not follow a normal distribution or may be ranked rather than continuous. The range of τ value from -1 to 1, in which as the value approaches -1, it indicates a strong negative correlation, and as the value approaches 1, it indicates a strong positive correlation. In comparison, if the value approaches 0, it indicates weak correlation between the variables (Willmott et al., 1982). The general Kendall's tau correlation model can be represented by the following equation:

$$\tau = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}$$

2.2.2 Marginal distribution

Before applying the bivariate copula model to the data, the distribution of the rainfall from each station must be determined first. In this study, four different distributions were considered which were gamma, lognormal, weibull and exponential distributions. The best selected distribution will be selected based on the lowest value of Akaike Information Criterion (AIC). The equations of these distributions are listed down as follows:

1. Gamma distribution

$$f(x) = \frac{1}{\Gamma(\gamma)\beta^\alpha} e^{-\frac{x}{\beta}} x^{\alpha-1}, \quad x > 0, \alpha > 0, \beta > 0$$

where α is denoted as the scale parameter and β is denoted as the shape parameter.

2. Lognormal distribution

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0$$

where μ is denoted as the mean and σ is denoted as standard deviation.

3. Weibull distribution

$$f(x) = \left(\frac{\alpha}{\beta}\right) \left(\frac{x}{\beta}\right)^{(\alpha-1)} e^{-\left(\frac{x}{\beta}\right)^\alpha}, \quad x > 0, \alpha > 0, \beta > 0$$

where α is denoted as the scale parameter and β is denoted as the shape parameter.

4. Exponential distribution

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0, \lambda > 0$$

where λ is denoted as the rate parameter.

2.2.3 Copula Theory

Copula model is based on theorem of Sklar (1959), which is introduced to incorporate the complex dependency structure between multivariate data. The distributions of each data are not strictly assigned to only one kind of distribution

and it might be different, making the copula model more flexible as compared to traditional bivariate model. The cumulative distribution function (CDF) of the variable is denoted as below:

$$F(x_1, \dots, u_d) = C[F_1(X_1), \dots, F_d(x_d)]$$

The joint probability density function of the copula model is as follows:

$$F(x_1, \dots, u_d) = \left[\prod_{i=1}^d f_i(x_i) \right] C[F_1(X_1), \dots, F_d(x_d)]$$

This study aims to predict rainfall data of the selected stations in Peninsular Malaysia from 2015 to 2021 by using bivariate copula model, specifically the Elliptical and Symmetric Archimedean Copula models.

1. Elliptical copulas

Elliptical copulas are known for their ability to capture the correlation between any two pairs of marginal distributions. They are specifically designed to correlate with elliptical distributions. The general representation of Elliptical copula is as follows:

$$C_{\Sigma}(u_1, \dots, u_n) = \Phi_{\Sigma}[\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)]$$

2. Archimedean copulas

The general form of the Archimedean copula family for symmetric (or exchangeable) copulas can be represented as:

$$C(u) = C(u_1, \dots, u_d; \psi) = \psi(\psi^{-1}(u_d)), u \in [0, 1]^d$$

where C is the copula function $u_1, \dots, u_d$ are marginal cumulative distribution functions (CDFs) corresponding to the variables, ψ is the Archimedean copula generator function and ψ^{-1} is the inverse of the generator function. The Archimedean copula that are used in this study are Clayton, Frank and Gumbel.

Table 2 presents the equations of the Elliptical and Archimedean copulas:

Table 2: Bivariate copula families of Elliptical and Archimedean copulas.

Copula	$C_{\theta}(u, v)$	θ range
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\theta \geq -1$
Frank	$-\theta^{-1} \log \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right]$	$\theta \neq 0$
Gumbel	$\exp \left[-((-\log u)^{\theta} + (-\log v)^{\theta})^{\frac{1}{\theta}} \right]$	$\theta \geq 1$
Gaussian	$\Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v))$	$\theta \in [-1, 1]$
t Student	$t_{\rho, \delta}(t_{\delta}^{-1}(u), t_{\delta}^{-1}(v))$	$\theta \in [-1, 1]$

2.2.4 Akaike Information Criterion (AIC)

The Akaike information criterion (AIC) method was used in this study to select the best fitted distribution and best copula model to the selected rainfall stations in Peninsular Malaysia. The equation of the AIC method is as follows:

$$AIC = 2p - 2\ln(L)$$

where p is denoted as the number of estimated parameters in the statistical model, and parameter L is denoted as the likelihood function's value based on these parameters.

3. RESULT AND DISCUSSION

The descriptive statistics for each selected stations for all states are shown in Table 3. The descriptive statistics are used to understand the behaviours of the 12 selected rainfall stations in Johor, Selangor, Kelantan and Terengganu. An analysis of data characteristics—specifically marginal distributions and dependence structures—is essential for identifying the most appropriate Copula models. Understanding these properties ensures that the selected model accurately captures the relationship between variables (Schwitter et al., 2025).

According to Table 3, the average daily rainfall for all stations over the course of ten years ranges from 5.80 mm to 9.90 mm, with Petaling Jaya reporting the highest mean value of 9.89 mm per day and KLIA reporting the lowest mean of 5.81 mm per day. This shows that generally all stations in Johor have received lower value of rainfall as compared to other selected states, while stations situated in the North and East Coast regions - specifically those in Kelantan and Terengganu, exhibited higher rainfall intensities, likely due to their direct exposure to the Northeast Monsoon (NEM).

The variance across the twelve rainfall stations spans from 173.30 mm² to 587.10 mm², while the standard deviation ranges from 13.16 mm at Batu Pahat to 24.23 mm at Gong Kedak. The notably elevated variation observed in Gong Kedak and other Northern and East Coast stations reflects high volatility and a broader distribution within their daily rainfall series. This is primarily attributed to their direct exposure to the NEM, which brings intense and sporadic precipitation surges. In contrast, stations in the Southern region exhibit the lowest dispersion, indicating a more stable and concentrated data range due to their relatively sheltered geographical position. A critical observation across all locations is that the standard deviation consistently exceeds the mean.

This phenomenon - where the spread of the data outweighs the average - is a direct consequence of extreme precipitation events that create a heavy right-hand tail in the

data. This significantly inflating the variance, and shifting the data distributions away from standard normal curve.

Table 3: Descriptive statistics of the daily rainfall amount for selected stations.

No.	Station	Mean	Standard Deviation	Variance	Skewness	Kurtosis
1	Kerteh	7.8710	19.8008	392.0716	5.1369	44.1080
2	Kuala Terengganu	7.0197	18.6472	347.7172	5.4392	45.7938
3	Gong Kedak	9.1783	24.2291	587.0480	6.3336	65.8676
4	Kota Bharu	7.2163	18.8137	353.9564	5.0367	38.2143
5	Kuala Krai	7.2078	18.4337	339.8026	5.2478	40.6483
6	KLIA	5.8138	13.5367	183.2430	4.9603	49.9279
7	Petaling Jaya	9.8919	18.5023	342.3355	3.0466	16.9539
8	Subang	8.6740	17.4572	304.7525	4.1858	39.0847
9	Batu Pahat	5.8230	13.1646	173.3065	3.6516	19.7038
10	Kluang	6.0371	13.9544	194.7240	4.0189	25.2056
11	Mersing	5.9551	16.2239	263.2139	6.5947	76.1570
12	Senai	6.8301	14.4557	208.9673	3.6891	22.9001

The coefficient of skewness serves as a primary measure for assessing the distributional asymmetry of rainfall data relative to the mean. Analysis of the 12 selected stations reveals that all locations exhibit positive skewness, with values ranging significantly from 3.05 at Petaling Jaya to 6.59 at Mersing. These results indicate a distribution that is heavily right-skewed. Complementing this, the kurtosis values were calculated to evaluate the relative peakness and the tail weight of the distributions. With values spanning from 16.95 to 76.16, all stations are categorized as leptokurtic. Mersing's exceptionally high kurtosis of 76.16 indicates a sharp, distinct peak near the mean coupled with heavy tails, confirming that the dataset is prone to extreme outliers that deviate far from the average.

These descriptive statistics provide the necessary justification for the subsequent modelling phases. First, the high skewness and over-dispersion (where standard deviation exceeds the mean) necessitate the use of Log-normal distributions for marginal fitting, as they are mathematically designed to handle non-negative, right-skewed hydrological data. The analysis on the best fitted distribution will be further analysis and confirm in Table 5.

Second, the initial analysis of the descriptive statistics in Table 3 suggests the presence of underlying spatial dependencies among the twelve rainfall stations; however, these relationships appear to be non-uniform across the different states. While some stations—particularly those within the same geographic region—exhibit synchronized patterns, others demonstrate a lack of significant association.

To rigorously quantify these interactions, Kendall's tau is employed as a robust, non-parametric measure. Unlike traditional linear correlation, Kendall's tau is better suited to these datasets because it is more robust (Croux et al., 2010) when extreme monsoonal outliers identified in the descriptive results. Finally, to move beyond simple correlation and model the complex, asymmetric relationships between these variables, a bivariate copula framework is utilized. This approach allows for a deeper examination of the joint probability and dependence structures, providing a more comprehensive understanding of how rainfall behavior at one station relates to another across Johor, Selangor, Kelantan, and Terengganu.

The first step in developing the bivariate copula model for rainfall data is to examine the degree of correlation between the selected rainfall stations. Since most correlation values presented in Table 4 were below 0.5, indicating weak inter-station relationship, only stations with the highest correlations numbers were retained to ensure a more meaningful dependence structure. This step is essential because the purpose of a bivariate copula is to capture the dependency structure between pairs of stations. Hence, identifying station pairs with the strongest correlations ensures that the model represents meaningful spatial dependence. High correlation between rainfall stations indicates that rainfall events tend to occur simultaneously or exhibit similar patterns across locations, reflecting spatial coherence. Therefore, only station pairs with sufficiently strong correlations are appropriate for this type of modelling. The correlation coefficients for all stations are presented in Table 4, while the corresponding station abbreviations are listed in Table 4.

In Table 4, green highlights indicate the highest correlation value for each station within the same state. For example, for Kerteh (KRT), the correlation with Kuala Krai (KK) is 0.3220, which is slightly higher than its correlation with Kota Bharu (KB) at 0.3061. Thus, the value 0.3220 is highlighted to indicate that KK has the strongest association with Kerteh among the Kelantan stations. Based on these highlighted values, Kuala Terengganu, Kota Bharu, KLIA, and Mersing were selected for the subsequent bivariate copula analysis. Although both Kota Bharu and Kuala Krai exhibited similarly high correlation levels within their respective state groupings, Kota Bharu was ultimately selected to represent Kelantan.

The greater variability in rainfall at Kota Bharu makes it more suitable for capturing the state-level dependence structure in the copula model, especially when comparing with other states, as it allows for a clearer analysis of differences in rainfall behaviour across regions with varying levels of variability. Kuala Terengganu was selected as it exhibits the.

Table 4: Kendall's tau correlation coefficient for each station

		Terengganu			Kelantan		Selangor			Johor			
		KRTH	KT	GK	KB	KK	KLIA	PJ	SBG	BP	KLG	MRSNG	SENAI
Terengganu	KRTH	1.0000	0.4010	0.2483	0.3061	0.3220	0.1189	0.0671	0.0837	0.1505	0.1465	0.2108	0.1321
	KT	0.4010	1.0000	0.3373	0.4675	0.3626	0.1385	0.0827	0.0987	0.1570	0.1206	0.2042	0.0995
	GK	0.2483	0.3373	1.0000	0.3473	0.2663	0.0908	0.0196	0.0382	0.1004	0.0699	0.1593	0.0640
Kelantan	KB	0.3061	0.4675	0.3473	1.0000	0.3702	0.1345	0.0841	0.0982	0.1406	0.1061	0.1738	0.1023
	KK	0.3220	0.3626	0.2663	0.3702	1.0000	0.1276	0.0724	0.0815	0.1549	0.1671	0.1938	0.1488
Selangor	KLIA	0.1189	0.1385	0.0908	0.1345	0.1276	1.0000	0.3245	0.3556	0.2515	0.1888	0.0924	0.1648
	PJ	0.0671	0.0827	0.0196	0.0841	0.0724	0.3245	1.0000	0.5702	0.1965	0.1530	0.0610	0.1488
	SBG	0.0837	0.0987	0.0382	0.0982	0.0815	0.3556	0.5702	1.0000	0.2147	0.1596	0.0548	0.1523
Johor	BP	0.1505	0.1570	0.1004	0.1406	0.1549	0.2515	0.1965	0.2147	1.0000	0.3125	0.0710	0.2643
	KLG	0.1465	0.1206	0.0699	0.1061	0.1671	0.1888	0.1530	0.1596	0.3125	1.0000	0.1244	0.3860
	MRSNG	0.2108	0.2042	0.1593	0.1738	0.1938	0.0924	0.0610	0.0548	0.0710	0.1244	1.0000	0.1008
	SENAI	0.1321	0.0995	0.0640	0.1023	0.1488	0.1648	0.1488	0.1523	0.2643	0.3860	0.1008	1.0000

highest correlation among the stations and is also one of the most prominent flood-affected areas in Terengganu, having experienced significant flooding events between 1985 and 2013 (Abu Bakar et al., 2016). KLIA and Mersing is chosen since the data pattern is more synchronized with other stations as compared to the other stations.

In summary, the station selection process was guided by both correlation strength and statistical representativeness to ensure a present of dependency structure between stations. The analysis focuses on stations that not only exhibit relatively stronger inter-station relationships but also capture diverse rainfall characteristics across different states. With these representative stations identified, the next step involves determining the most appropriate marginal distributions for each station, followed by the selection of the best-fitted copula model to accurately characterize the dependence structure.

Each station was then assessed for their best fit distribution by using Akaike Information Criterion (AIC). Distribution with the lowest AIC will be selected. Based on the results in Table 5, the lognormal distribution was identified as the best fit for all stations, as it consistently recorded lower AIC values compared to the exponential, gamma, and Weibull distributions. This finding is consistent with the preliminary analysis in Table 3, where the descriptive statistics indicated high skewness and standard deviations exceeding the mean, characteristics commonly associated with lognormal behaviour. Therefore, the lognormal distribution was selected as the marginal distribution for modelling the rainfall data

within the bivariate copula framework.

Table 5: The best fitted marginal distribution for each station.

Station	Lognormal	Exponential	Gamma	Weibull
Mersing	8205.72	14146.84	9543.32	8931.38
Kuala Terengganu	8587.91	14938.42	9939.45	9316.05
KLIA	8618.42	13991.69	9788.65	9276.47
Kota Bharu	8600.52	15077.80	9951.68	9332.00

Table 6 presents the AIC values for the bivariate copula model applied to the four rainfall stations in Peninsular Malaysia. Similar to Table 5, the best copula model will be selected based on the lowest AIC values obtained. According to the results obtained in Table 6, Clayton copula outperformed the others by having the lowest AIC values for all stations, which indicate that Clayton gave the best fit to the dependence structure between rainfall stations across several selected states in Malaysia. Clayton outperformed the other copula models because of its valuable invariant property, which enhance the robustness of the model when dealing with nonlinear, truncated or missing data (Louzada et al., 2016). Thus, Clayton copula model can be used as a reference to enhance the accuracy of rainfall data prediction in Peninsular Malaysia since it can be used to generate synthetic rainfall data. The synthetic rainfall data can be simulated in multiple scenarios to estimate probabilities of extreme events more accurately. Consequently, this could improve water resource management by providing more precise water availability

information. On the other hand, the simulated rainfall data can assist the stakeholders in adopting effective flood prevention

strategies and facilitate climate adaptation efforts through the simulated changes of rainfall pattern.

Table 6: AIC values for the bivariate copula model at the selected stations.

Mersing – Kuala Terengganu			KLIA – Kuala Terengganu		Kota Bharu – Kuala Terengganu	
	$\hat{\rho}$ or $\hat{\theta}$	AIC	$\hat{\rho}$ or $\hat{\theta}$	AIC	$\hat{\rho}$ or $\hat{\theta}$	AIC
Gaussian	0.1811	-85.6859	0.1513	-59.1989	0.3772	-400.3268
t Student	0.2342	-129.0438	0.1856	-106.6954	0.4625	-587.2753
Clayton	0.7446	-206.5296	0.6342	-142.4368	1.4106	-737.8501
Gumbel	1.0971	-89.1829	1.0829	-67.1979	1.2551	-426.6159
Frank	0.9759	-82.6197	0.8078	-57.3647	2.2388	-392.8998
Mersing - KLIA			Kota Bharu - Mersing		Kota Bharu - KLIA	
	$\hat{\rho}$ or $\hat{\theta}$	AIC	$\hat{\rho}$ or $\hat{\theta}$	AIC	$\hat{\rho}$ or $\hat{\theta}$	AIC
Gaussian	0.2962	-229.1981	0.1696	-71.4787	0.1615	-64.7975
t Student	0.3849	-327.2689	0.2247	-116.7895	0.2097	-113.0802
Clayton	1.1423	-512.5678	0.7951	-221.2432	0.7540	-197.6346
Gumbel	1.1788	-214.8305	1.0086	-65.3691	1.0848	-62.9407
Frank	1.7260	-247.6987	0.9723	-80.6099	0.9202	-73.3670

4. CONCLUSION

Accurate rainfall prediction is essential for effective risk management, particularly in addressing extreme weather events. This study employs a bivariate copula model to analyse rainfall data from 12 selected stations across four states in Peninsular Malaysia: Kelantan, Terengganu, Johor, and Selangor. The strength of the relationship between stations was evaluated using Kendall's tau correlation coefficient. Based on this analysis, one representative station was selected from each state—Kota Bharu, Kuala Terengganu, Mersing, and KLIA—due to their relatively stronger correlations within their respective states. Although some station pairs across different states exhibited weaker correlations, this is expected as inter-state relationships are generally less dependent than intra-state ones. Nonetheless, since the primary objective of this study is to examine rainfall dependency patterns in flood-prone regions across different states, analysing inter-state relationships remains both relevant and valuable. Subsequently, the best-fitting marginal distribution and copula model were identified using the Akaike Information Criterion (AIC). The results indicate that the lognormal distribution and the Clayton copula provide the best fit for the data. This model can be used as a reference for the stakeholders in managing flood risk by providing accurate synthetic rainfall through simulation. Further studies could explore the application multivariate copula models to rainfall data, enabling a more comprehensive understanding of dependency across more than one station.

ACKNOWLEDGEMENT

The authors would like to express their appreciation

to Universiti Teknologi Malaysia for their financial support under the UTM Fundamental Research, Reference No: PY/2024/00951, with vote number Q.J130000.3854.23H62.

REFERENCES

- Abu Bakar, M. F., & Alias, N. E. (2016). Extreme rainfall analysis on the December 2014 flood, Terengganu. *Journal of Environmental Engineering, Hydraulics and Hydrology*, 3, 265-282.
- Ahmad, N., Ng, C. P., Ismail, N., Wong, M. M. R., Abbas, F., Law, T. H., Zul Bahari, A. N. A., & Mohd Rofi, M. F. (2025). A review of major flood events in Malaysia between 1970-2024. *Jurnal Kejuruteraan*, 37(2), 909–921. [https://doi.org/10.17576/jkukm-2025-37\(2\)-28](https://doi.org/10.17576/jkukm-2025-37(2)-28)
- Ariff, N. M., Jemain, A. A., Ibrahim, K., & Zin, W. W. (2012). IDF relationships using bivariate copula for storm events in Peninsular Malaysia. *Journal of Hydrology*, 470, 158–171.
- Buakamsri, T. (2025, December 1). *Cyclone Senyar shows the equator is no longer a safe zone—Southeast Asia's risk maps are already out of date*. Climate Connectors. <https://climate-connectors.org/cyclone-senyar-shows-the-equator-is-no-longer-a-safe-zone-southeast-asias-risk-maps-are-already-out-of-date/>
- Buslima, F. S., Omar, R. C., Jamaluddin, T. A., & Taha, H. (2018). Flood and flash flood geo-hazards in Malaysia. *International Journal of Engineering and Technology*, 7(4), 760–764.
- Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, 19(4), 497-515.
- Dzupire, N. C., Ngare, P., & Odongo, L. (2020). A copula based bi-variate model for temperature and rainfall processes. *Scientific African*, 8, e00365.
- Jafray, N. A., Suhaila, J., Yusof, F., Nor, S. R. M., & Alias, N. E. (2023). Bivariate copula for flood frequency analysis in Johor river basin. *IOP Conference Series: Earth and Environmental Science*, 1167, 012018.
- Kamaruzaman, I. F., Wan Zin, W. Z., & Mohd Ariff, N. (2019). A generalized bivariate copula for flood analysis in Peninsular Malaysia. *Malaysian Journal of Fundamental and Applied Sciences*, 15, 38–49.
- Leeonis, A. N., Ahmed, M. F., Mokhtar, M. B., Halder, B., Lim, C. K., Majid, N. A., & Scholz, M. (2025). Importance of geographic information system (GIS) application to reduce the impact of flood disasters in Malaysia: A Meta-Analysis. *Water*, 17(2), 181.
- Levesque, F. (2025, December 1). *Southeast Asia hit by deadly widespread flooding with three tropical cyclones, including rare Malacca Strait cyclone*. DTN APAC. <https://apac.dtn.com/international/southeast-asia-hit-by-deadly-widespread-flooding-with-three-tropical-cyclones-including-rare-malacca-strait-cyclone/>

- Louzada, F., & Ferreira, P. H. (2016). Modified inference function for margins for the bivariate clayton copula-based SUN Tobit Model. *Journal of Applied Statistics*, 43(16), 2956–2976.
- Mai, L. (2025, December 11). *The latest on Southeast Asia: Severe flooding in Southeast Asia*. Center for Strategic and International Studies. <https://www.csis.org/blogs/latest-southeast-asia/latest-southeast-asia-severe-flooding-southeast-asia>
- Malay Mail. (2024, September 16). *Monsoon season shifts, intensity complicate flood management, says deputy minister*. <https://www.malaymail.com/news/malaysia/2024/09/16/monsoon-season-shifts-intensity-complicate-flood-management-says-deputy-minister/150608>
- Malaysian Meteorological Department. (2025, May). *Fenomena cuaca*. <https://www.met.gov.my/en/pendidikan/fenomena-cuaca/>
- New Straits Times. (2025, March 28). *Floods in Johor Baru worst in 30 years, says Mayor*. <https://www.nst.com.my/news/nation/2025/03/1194225/floods-johor-baru-worst-30-years-says-mayor>
- Schwitzer, N., & Liebe, U. (2025). Sometimes, a descriptive figure is worth more than a thousand model coefficients: the importance of data description in social research. *International Journal of Social Research Methodology*, 29(2), 289–294.
- Sharma, M. (2025, December 2). *What made Cyclone Senyar a once-in-a-century weather anomaly in Malacca Strait*. India Today.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8 (3), 229–231.
- The Malaysian Reserve. (2022, August 3). *Johor Bahru hit by worst flash floods in 10 years*. <https://themalaysianreserve.com/2022/08/03/johor-bahru-hit-by-worst-flash-floods-in-10-years/>
- The Straits Times. (2025, November 28). *Storm Senyar set to hit 10 locations in Malaysia*. <https://www.straitstimes.com/asia/se-asia/storm-senyar-set-to-hit-10-locations-in-malaysia>
- Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63, 1309–1313.
- Yee, K. C., Suhaila, J., Yusof, F., & Mean, F. H. (2014). Bivariate copula in fitting rainfall data. *AIP Conference Proceedings*, 1605, 986–990.