MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Nurfarzana *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 22 (2026) 836-856

RESEARCH ARTICLE

On the Role of Copula based Dependence Modelling in ARMA-GARCH Rainfall Forecasting

Nurfarzana Syakirah Nor Afpidin[a], Wan Muhammad Haiqal Shah Mohd Sariff[a], Siti Meriam Zahari[a], Siti Roslindar Yaziz[b], Noor Fadhilah Ahmad Radi[a,*]

[a]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia; [b]Centre for Mathematical Sciences, Universiti Malaysia Pahang Al-Sultan Abdullah, 26300 Kuantan, Pahang, Malaysia

Abstract Accurate rainfall simulation is essential for hydrological risk assessment and water resources management. The ARMA-GARCH models effectively capture temporal dependence and volatility in rainfall series and copula models are frequently used to represent the joint dependence among rainfall stations. However, it is generally assumed that modelling cross-station dependence will improve forecasting performance, with limited attention given to whether sufficient residual dependence remains after temporal filtering to justify the additional modelling complexity. Hence, this study investigates whether incorporating copula-based dependence modelling provides additional benefits for rainfall simulation after ARMA-GARCH filtering. Daily rainfall data from three meteorological stations were first modelled using ARMA-GARCH, and the standardized residuals were transformed to uniform margins. Residual dependence was then evaluated using Kendall's tau and empirical upper tail dependence before fitting multivariate copula models. The dependence analysis revealed weak residual association among the stations, with Kendall's tau values ranging from 0.01 to 0.05. Although weak upper tail dependence was observed at the 90th percentile threshold, it decreased to approximately zero at higher thresholds, indicating limited co-occurrence of extreme rainfall events after temporal filtering. The forecasting comparison showed that the ARMA-GARCH model achieved lower MAE, MSE and RMSE values than the ARMA-GARCH-copula framework. These findings suggest that, for the selected rainfall stations, modelling the remaining cross-station dependence after temporal filtering contributed only modestly to forecasting performance. This study highlights the importance of evaluating residual dependence before introducing copula-based dependence structures in multisite rainfall modelling.

Keywords: ARMA; GARCH; Box-Jenkins; Copula; Forecast.

***For correspondence:**
noorfadhilah@uitm.edu.my

Received: 11 Dec. 2025
Accepted: 29 Jul 2026

Introduction

Heavy rainfall is a significant factor contributing to natural disasters. The seemingly erratic occurrence of heavy rainfall, as a direct consequence of climate change, has garnered concerns due to its direct causation to devastative flood disasters [1]. According to the World Health Organization (WHO), flooding is often triggered by heavy rainfall, rapid snowmelt, or the storm surge of tropical cyclones, disrupting health systems and denying communities access to essential services. It was reported that floods were the leading cause of natural disaster-related deaths in the 20th century, accounting for approximately 6.8 million fatalities [2]. This underscores the necessity of studying rainfall data to mitigate disaster risks effectively. Traditional models, such as the Autoregressive Moving Average (ARMA), are widely employed to capture the mean behavior of time series and predict short-term trends based on historical data. For instance, [3] used various ARMA models to identify the most suitable approach for accurate rainfall prediction. Besides that, [4] also used ARMA in forecasting the inflow of Dez dam reservoir. However, these models do not address time-varying volatility, which is critical in understanding rainfall dynamics.

To address the limitations of traditional models in capturing time-varying volatility advanced statistical frameworks such as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model have been increasingly adopted. Introduced by Tim Bollerslev in 1986 as an extension of Engle's ARCH model, GARCH incorporates lagged conditional variances and squared residuals, enabling it to model volatility clustering, a phenomenon where periods of high variability are followed by sustained instability. This characteristic is particularly relevant to rainfall patterns, where extreme precipitation episodes often cluster temporally, exacerbating flood risks and complicating predictive efforts. Several studies highlight GARCH's utility in modelling time series data [5, 6, 7]. Studies in rainfall data including [8] which proposed hybrid ARIMA-GARCH models for serial dependence and volatility in the daily rainfall series. Furthermore, [9] use hybrid of seasonal autoregressive integrated moving average (SARIMA)-(GARCH) to accounts the conditional variance of the SARIMA model of rainfall time series. The results of the study indicate that the hybrid SARIMA–GARCH models best fitted to the transformed rainfall across the humid and arid regions.

Univariate approaches are highly effective at capturing temporal dynamics at a single station, but real-world hydrological hazards are regional phenomena. Consequently, multisite rainfall modelling is required to capture daily rainfall variations across a network of spatially distributed gauge stations at the same time. The primary challenge in rainfall simulation is capturing the spatial dependency structure, which has driven the adoption of copula-based models. According to [10], the absence of an effective daily rainfall model at gauge stations limits the ability to accurately represent real-world dependencies. Copula models are highly effective in characterizing joint dependence, particularly when extreme values are distinctly observable [11]. An example of research that leverages copula includes [12] that exemplifies the application of copula-based methods in hydrological research. The research employs copula joint distribution models to investigate the spatial and temporal patterns of wet and dry conditions across three distinct river reaches. In another application, [13] presented a copula-based bivariate modeling framework to quantitatively assess meteorological and ecological droughts. The Gaussian copula provided the best fit for the Standardized Precipitation Index (SPI), while the Gumbel copula was optimal for the Ecological Index (EI). Furthermore, a study by [14] utilized trivariate copulas to analyze flood episodes at the Guillemard Bridge gauging station in the Kelantan River basin, Malaysia. Based on Cramer-von-Mises statistics (S_n) and p-values, the Gaussian copula was identified as the best-fit model for representing the trivariate dependence structure of the floods.

Recent advancements in the application of ARMA-GARCH-copula frameworks have demonstrated their immense versatility in modelling complex dependencies and volatility structures, offering critical insights for advanced risk management. In environmental and hydrological research, this integrated approach has steadily gained traction for its ability to simultaneously isolate temporal volatility and multivariate spatial dependencies. The foundational utility of Autoregressive Moving Average (ARMA) models to capture temporal autocorrelation, coupled with Generalized Autoregressive Conditional Heteroskedasticity (GARCH) structures to model time-varying volatility and clustering effects, is well documented in meteorological literature [8]. To scale these univariate predictions to regional catchments, modern researchers frequently pair GARCH processes with copula functions. This mathematical coupling enables the characterization of complex, non-linear, and non-Gaussian spatial tail dependencies between distinct monitoring stations without losing localized marginal traits or forcing the assumption of multivariate normality [15]. Globally, these frameworks have successfully mapped the joint probabilities of extreme precipitation events and subsequent severe drainage hazards [16].

Although the integration of copula models with time-series frameworks has gained increasing attention in hydrological forecasting, it is generally assumed that incorporating copula-based dependence modelling automatically improves rainfall simulation or forecasting. In many cases, the spatial or cross-station dependence structure is modelled immediately after the marginal time-series models are established, with limited investigation into whether sufficient residual dependence remains after temporal filtering [17]. Because ARMA-GARCH models are explicitly designed to filter out serial autocorrelation and conditional heteroscedasticity, the resulting standardized residuals often exhibit significantly weakened cross-station correlation. Under such circumstances, introducing a highly parameterized multivariate copula layer inflates model complexity without yielding meaningful gains in forecasting accuracy [18,19]. Despite the practical risk of over-parameterization, this diagnostics-driven question has received surprisingly little critical attention in the existing rainfall modelling literature.

Therefore, this study investigates whether incorporating copula-based dependence modelling provides additional benefits for rainfall simulation after ARMA-GARCH filtering. Although copula models have been widely applied in rainfall studies, it often incorporated without first evaluating whether sufficient residual dependence remains after temporal filtering. Consequently, it remains unclear whether modelling the residual dependence using copulas provides additional benefits when the temporal dynamics have already been effectively captured by ARMA-GARCH models. This study aims to evaluate the applicability of copula-based dependence modelling for rainfall simulation after temporal dependence has been accounted for using ARMA-GARCH models. Specifically, the study first models the temporal dependence and conditional heteroscedasticity of rainfall series using ARMA-GARCH. The residual dependence among the transformed standardized residuals is then assessed using Kendall's tau (τ) and empirical upper tail dependence before fitting multivariate copula models. Finally, the performance of the ARMA-GARCH and ARMA-GARCH-copula frameworks is compared to determine whether modelling the remaining cross-station dependence improves rainfall simulation accuracy.

Materials and Methods

Figure 1 illustrates the framework of this study. The initial phase focuses on partitioning the historical records and auditing the fundamental statistical properties of the raw data. The dataset is first divided into two subsets: a Training Dataset (90%) used for data pre-processing, model fitting and multivariate dependence modelling in Phases 1, 2 and 3, and a Testing Dataset (10%) reserved strictly for out-of-sample validation in Phase 4. Following this partition, descriptive statistics are calculated to profile the distribution, central tendency, and variance behavior of each series. A critical step in this phase is the evaluation of stationarity in mean. This study initially assesses variance stationarity, followed by mean stationarity. When variance is unstable over time, a Box-Cox transformation is utilized to stabilize it. Once stationarity is established, the process begins by fitting an ARMA model to capture the linear serial correlation in the mean of the series. Then, the residuals are evaluated to determine if they are serially correlated and if the residual variance is heterogeneous over time (heteroscedasticity).

Upon confirming the presence of ARCH effects or volatility clustering, a GARCH model is introduced and combined with the ARMA process to establish a joint ARMA-GARCH. The standardized residuals from the ARMA-GARCH models are subsequently tested for normality. If the residuals are non-normal, alternative distributions like the Student-t, skew-t, or Generalized Error Distribution (GED) are considered. To capture the spatial dependency, this study integrates copula functions into the time-series model. The standardized residuals obtained from the optimal ARMA-GARCH models in Phase 2 are transformed into uniform marginal distributions on the interval [0, 1]. Following this transformation, the cross-sectional rank correlation between stations is evaluated using non-parametric metrics such as Kendall's. Additionally, because concurrent extreme events represent the highest structural or environmental risk, joint tail behavior is assessed. This step focuses on upper tail dependence, which measures the probability that multiple stations experience simultaneous extreme values.

Next, the residuals are modelled using copula, specifically Gaussian, Student-*t* and Skewed-*t* copulas. The final phase of the framework focuses on validating the predictive and generative capacity of the integrated modelling approach. The validation process begins by generating out-of-sample forecasts on the reserved testing dataset, independent ARMA-GARCH models. Next, ARMA-GARCH-Copula model is utilized for synthetic data generation. By leveraging the joint dependence structure captured by the selected copula, the simulation produces synthetic datasets that replicate both the individual temporal properties of each station and the spatial dependencies observed across the network. Finally, a quantitative comparison is conducted between the forecasts generated by the ARMA-GARCH models and the simulated paths from the ARMA-GARCH-Copula model. Both models are evaluated using error metrics such as MAE, MSE, and RMSE.

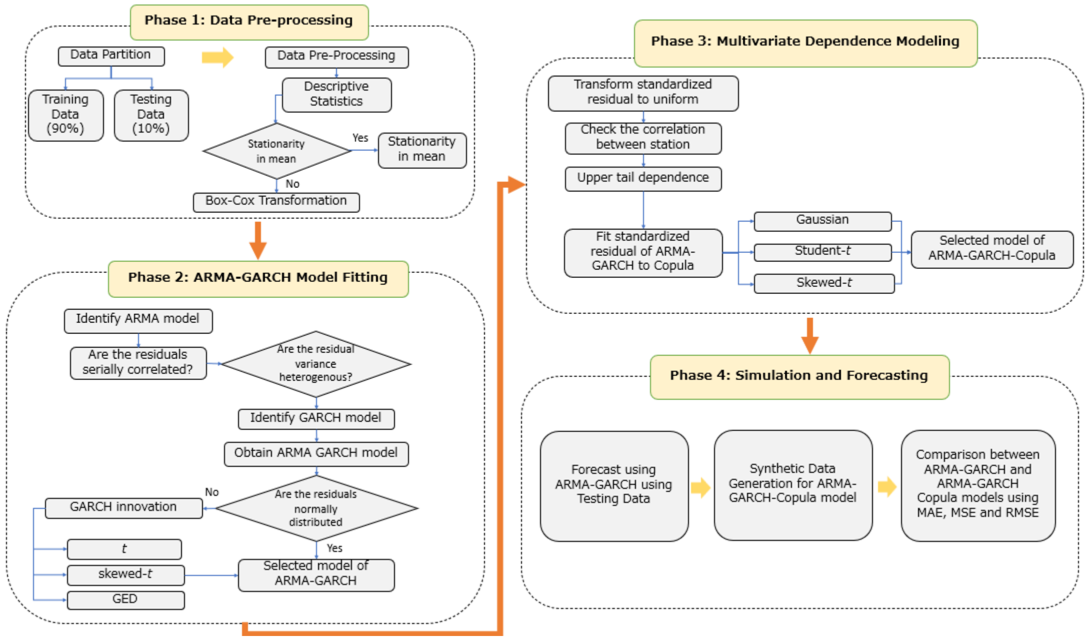

Figure 1. The overall research framework for ARMA-GARCH-Copula model

Study area and data

The study focuses on the East Coast region of Peninsular Malaysia, specifically the state of Kelantan, is unique due to weather conditions and geographical features, making it an ideal location for rainfall monitoring. Kelantan experiences significant rainfall, particularly due to its exposure to the Northeast Monsoon, which contributes to some of the highest daily rainfall records in the country [20]. According to the Malaysian Meteorological Department, Kelantan holds the record for the highest daily rainfall in Malaysia, with an extraordinary 698.7 mm recorded in 1981. A major flood event in Kelantan and Terengganu in December 2022 was also caused by exceptionally high daily rainfall. Based on a report from the National Disaster Management Agency (NADMA), a key trigger was the 340.6 mm of rain recorded in Tumpat, Kelantan, on 28 November 2024.

Given the consistently high annual rainfall in Kelantan, this study utilizes meteorological data obtained from the Malaysian Meteorological Department across three monitoring stations: Kota Bharu, Kuala Krai, and Pos Gob. Table 1 shows the latitude, longitude, and elevations for the stations involved. The daily rainfall data covers a 5-year period from 2018 to 2022. Although this period is relatively short for climatological trend analysis, this specific timeframe was selected to ensure maximum data completeness, minimize missing records across the chosen stations, and capture recent high-intensity monsoon behaviors. The rainfall data considered in this study corresponds to the Northeast Monsoon season, which occurs from November to March. The data has been split into two groups of training and testing data using common ratio of 90:10 [21]. However, data points containing zero values were excluded to avoid distortions in the analysis, ensure the validity of statistical transformations, and maintain result accuracy. This exclusion is necessary as the daily rainfall datasets often contain a high frequency of zero-rain days, which can heavily skew probability distributions and violate the normality assumptions required for parametric testing. Furthermore, removing zeros prevents mathematical errors during Box-Cox transformations, which strictly require positive data values. Figure 2 displays the map showing the geographic locations of the rainfall stations. The red bullets indicate the station use in Kelantan.

Table 1. The geographical coordinates for each station.

Stations	Elevations (m)	Latitude	Longitude
Kota Bharu	4.4	06° 10' N	102° 18' E
Kuala Krai	68.3	05° 32' N	102° 12' E
Pos Gob	457	05° 17' N	101° 38' E

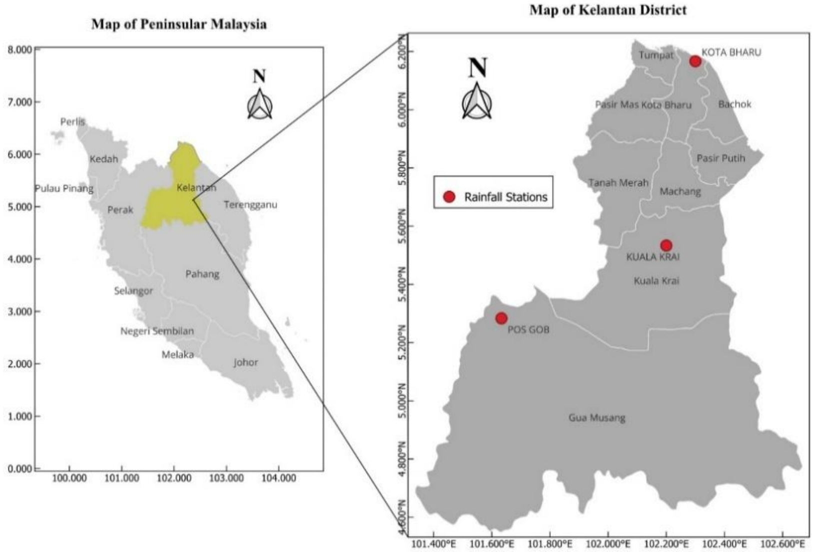

Figure 2. The location of rainfall stations

Exploratory Data Analysis on Rainfall Data

The initial stage of the modeling process entails conducting a descriptive statistical analysis to examine the fundamental characteristics of the dataset. This step is essential for assessing stationarity in both variance and mean, which are critical prerequisites for the robustness and validity of subsequent modeling techniques. In cases where the time series exhibits non-stationary in variance, a Box-Cox transformation is employed to stabilize the variance. According to [22], the Box-Cox transformation shown in Equation (1) is used to transform a skewed distribution into an approximately normal distribution by altering the original scale of the data

$$Y_{BC}(Y,\lambda) = X\beta + \sigma e \tag{1}$$

where X is covariates matrix, $\beta = (\beta_{0,}\ \beta_{1,}\ \beta_{2,} \ldots\ldots, \beta_k)$ is vector of regression coefficients, σ is variance of random error and e is random error follows standard normal distribution. Meanwhile, the Augmented Dickey Fuller (ADF), Phillips–Perron (PP) and Kwiatkowski Phillips Schmidt Shin (KPSS) tests are employed to verify the mean stationarity of the time series. For both the ADF and PP tests, the null hypothesis shows that the series is non-stationary due to the presence of a unit root, which is tested against the alternative hypothesis of stationarity [23,24]. A p-value less than the chosen significance level (typically 0.05) leads to the rejection of the null hypothesis, indicating that the series is stationary. For the KPSS test, the null hypothesis of time series stationarity is tested against the alternative hypothesis that the series is non-stationary due to the presence of a unit root [24]. In this case, a p-value greater than the chosen significance level (typically 0.05) fails to reject the null hypothesis, confirming that the series is stationary.

Box-Jenkins Model

The next part involves the ARMA-GARCH model. The Box-Jenkins methodology serves as the basis for ARMA modelling, which effectively addresses autocorrelation in time series data, and is subsequently enhanced by GARCH to account for volatility clustering. The Box-Jenkins model is a methodology developed by George Box and Gwilym Jenkins (1970) for analyzing and forecasting time series data. The methodology emphasizes model identification, estimation, diagnostic validation, and forecasting, ensuring models balance simplicity and accuracy. A foundational subset of the Box-Jenkins framework is the ARMA (Autoregressive Moving Average) model, which applies to stationary time series.

The ARMA model consists of two part which are Autoregressive (AR) and Moving Average (MA) [25]. The ARMA model is highly effective for analyzing time series data, particularly when the data series is stationary. These models are well-suited for applications where the time series exhibits stationarity without the need for differencing. According to [26] the equation for ARMA model is expressed in Equation (2),

$$y_t = \mu + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} + u_t \quad (2)$$

where *p* and *q* represent the number of autoregressive and moving average terms, $\varphi_1, \ldots, \varphi_p$ and $\theta_1, \ldots, \theta_q$ are the autoregressive and moving average coefficients, and u_t is the error term.

GARCH Model

GARCH models are specially created to capture and model heteroskedasticity in time series data, a task that ARMA models cannot accomplish. Based on research by [27], the combination of ARMA and GARCH allows for the incorporation of multiple types of information and data sources, resulting in more accurate and robust forecasts. Heteroskedasticity refers to the phenomenon where the variance of the residuals is not constant over time. The foundational ARCH model was developed first, followed by its widely used extension, the GARCH model [8]. The basic GARCH model extends the ARCH model by including lagged value of both squared residual and past variance. According to [27], the GARCH (m, s) model is expressed as:

$$\varepsilon_t = \sigma_t z_t \quad (3)$$

$$\sigma_t^2 = \omega + \sum_{k=1}^{m} \alpha_k \alpha_{t-k}^2 + \sum_{l=1}^{s} \beta_1 \sigma_{t-l}^2 \quad (4)$$

where ε_t is error at time *t*, σ_t is conditional standard deviation at time *t*, and z_t is the standardized residual. Besides that, σ_t^2 is the conditional variance at time *t*, ω is the constant term, α_k is the coefficient that measure the impact of past squared residuals on current variance and β_l coefficients that measure the impact of past variances on current variance.

Box-Jenkins-GARCH Model

Model Estimation

Parameter estimation determines the optimal parameter values required to accurately capture the underlying data-generating process. In the context of Box-Jenkins modeling, parameter estimation typically relies on Maximum Likelihood Estimation (MLE). Following the approach of [28], the estimator of δ in the MLE of the ARMA-GARCH model is obtained by solving Equation (5),

$$\hat{\delta}_{MLE} = \max_{\delta \epsilon S_\delta} f(\delta | y, x) \quad (5)$$

where $f(y|x, \delta)$ is the likelihood function of the ARMA-GARCH Model,

$$f(y|x, \delta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{\hat{\varepsilon}_t^2}{2\sigma_t^2}\right), \quad (6)$$

Since the parameter space S_δ for δ is typically a subset of $\mathbb{R}^{k+p+q+r+s+1}$ a constrained maximum likelihood problem must be solved to estimate δ. Here, the dimensions correspond to the *k* exogenous coefficients, *p* autoregressive terms, *q* moving average terms, *r* ARCH parameters, *s* GARCH parameters, and the baseline variance constant, respectively.

According to Naji et al. [21], Maximum Likelihood Estimation (MLE) identifies parameter values that maximize the likelihood of observing the empirical data. In the Box-Jenkins, preliminary model estimation must satisfy two diagnostic criteria to ensure parameter reliability: the absolute value of the model coefficient must exceed twice its standard error, and the p-value must be statistically significant (p-value $\leq \alpha$) [29]. Following parameter validation, optimal model selection is determined through information criteria, specifically the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The model yielding the lowest information criterion value is selected, balancing goodness-of-fit with model parsimony.

Model Selection

The Akaike Information Criterion (AIC) is employed as the model selection criterion and is computed using the following formula:

$$AIC = -2 \cdot \log L(\theta) + 2k \tag{7}$$

where log $L(\theta)$ is the log likelihood and *k* is the number of parameters. Additionally, other criteria such as the Bayesian Information Criterion (BIC) are also significant. The BIC is calculated using the equation:

$$BIC = -2 \cdot \log L(\theta) + k \cdot \log(n) \tag{8}$$

where log $L(\theta)$ is the log likelihood, *k* is the number of parameters, and *n* is the sample size.

Diagnostics Checking

Diagnostic checking is performed to assess the adequacy of the model by examining two key aspects, that are the serial correlation and heteroscedasticity in residuals. If the residuals exhibit serial correlation, it indicates that the ARMA model is insufficient, requiring modifications such as adjusting the model order or incorporating additional terms to improve its fit. The Ljung-Box statistics are used to test for serial correlation [30]. This statistic is formally defined in Equation (9),

$$Q_M = n(n+2)\sum_{k=1}^{M}\frac{r_k^2}{n-k} \tag{9}$$

where r_k is the sample autocorrelations, *n* is the sample size and *M* is the time lag.

Similarly, if the residuals are heteroscedastic, an ARMA-GARCH model is constructed to account for both the mean and variance structures of the data. After fitting the ARMA-GARCH model, the residuals are further analyzed for normality. The JB test statistics are used to evaluate the normality of the data [31]. This statistic is defined in Equation (10),

$$JB = n\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right) \tag{10}$$

where *S* is the squares of normalized skewness and *K* is kurtosis. If the residuals follow a normal distribution, the ARMA-GARCH model is finalized. However, if deviations from normality persist, alternative distributions such as the student-*t* distribution, skewed-*t* distribution, or Generalized Error Distribution (GED) are considered to better capture residual behavior.

GARCH Innovation

Standard normal distributions assume symmetric, average behavior, but real-world data frequently exhibits heavy tails and asymmetry. To capture these extremes, GARCH modeling relies on innovations, which are unpredictable shocks, residuals, or error terms that drive volatility over time. By modeling these innovations using non-normal distributions like the student-t, skewed-t or the Generalized Error Distribution (GED), GARCH models can far more accurately capture the sudden extreme shocks.

Student's t-Distribution

The Student-t distribution is a continuous probability distribution that arises when estimating the mean of a normally distributed population. According to [32], the probability density function is show as Equation (11),

$$f(z) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}}\left(1+\frac{z^2}{v}\right)^{-\frac{v+1}{2}} \tag{11}$$

where z represents the variable, v is for the degrees of freedom, and Γ is the function of gamma.

Skewed *t*-Distribution

The Skewed t-Distribution is probability distribution used to model data that exhibits both heavy tails (leptokurtosis) and asymmetry (skewness). The Skewed t-Distribution extends the traditional Student's t-Distribution by incorporating a shape parameter to capture asymmetry [32]. The probability density function is as Equation (12),

$$f(z) = \frac{2}{\sigma B\left(\frac{v}{2},\frac{1}{2}\right)}\left(1+\frac{\frac{x-\xi}{\sigma}}{v}\right)^{-\frac{v+1}{2}} \tag{12}$$

where ξ determines the skewness, v is for the degrees of freedom and $B\left(\frac{v}{2},\frac{1}{2}\right)$ denotes the Beta function.

Generalised Error Distribution (GED)
GED extends the normal distribution by introducing a shape parameter that allows for the flexible modeling of both lighter and heavier tails [32]. The probability density function is as in Equation (13),

$$f(z) = \frac{\kappa}{2^{1+1/\kappa}\,\Gamma(1/\kappa)\,\sigma} exp\left(-\frac{1}{2}\left|\frac{z-\mu}{\sigma}\right|^{\kappa}\right) \tag{13}$$

where μ represents the mean, σ represents the parameter scale and κ is the shape parameter that determines the tail distribution.

Dependence Measure

Before fitting the copula model, it is important to examine the dependence structure among the variables. Unlike linear correlation measures, copula-based dependence measures can capture nonlinear relationships and dependence in the joint distribution. In this study, Kendall's tau is employed to quantify the overall strength and direction of pairwise dependence, while tail dependence coefficients are used to assess the degree of association in the upper tails of the distribution. These measures provide valuable insights into the dependence characteristics of the rainfall data and assist in selecting an appropriate copula model for subsequent analysis.

Kendall’s correlation coefficient

Kendall’s tau serves as a non-parametric metric that quantifies the degree of concordance between two observed variables [33]. Kendall’s correlation coefficient (τ) can be calculated using Equation (14),

$$\tau = \frac{N_c - N_d}{\frac{n^2 - n}{2}} \tag{14}$$

where N_c is the number of corresponding ranks, N_d is the number of inconsistent ranks and *n* is the sample size.

Tail Dependence Measure

Tail dependence analysis evaluates the probability of a pair of random variables experiencing extreme events simultaneously. Generally, tail dependence is categorized into upper and lower tail dependence [34]. Since hydro-climatological risk assessments prioritize high-magnitude precipitation events, the primary focus is placed on the upper tail dependence. The upper tail dependence coefficient is defined as:

$$\lambda_{up} = \lim_{v \to 1^-} P\{X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v)\} > 0 \tag{15}$$

where v represents the probability threshold and F_1^{-1} and F_2^{-1} represent the generalized inverse distribution functions of X_1 and X_2 respectively. If the tail dependence coefficient equals zero, X_1 and X_2 are asymptotically independent, whereas a higher coefficient value indicates stronger tail dependence [35].

Copula

A copula is a mathematical function that links a multivariate distribution with its univariate marginal distributions [36]. Formally, it represents a multivariate cumulative distribution function (CDF) characterized by uniformly distributed margins on the interval [0,1]. According to [37], multivariate distribution $F\ (x_1, \ldots, x_n)$ can be expressed in terms of copula, *C* and its marginal distribution, $F_{X_n}\ (x_n)$.

$$\begin{gathered} F(x_1, \ldots, x_n) = C\left(F_{X_1}(x_1), \ldots, F_{X_n}(x_n)\right) \\ C(u_1, u_2, \ldots, u_n) = F(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)) \\ C : [0,1]^n \to [0,1] \end{gathered} \tag{16}$$

where $F_{X_1}(x_1),\ldots,F_{X_n}(x_n)$ represents the marginal distribution function for *n* variables and $F_1^{-1}(u_1),\ldots,F_n^{-1}(u_n)$ represents the marginal distribution inverse function for *n* variables.

Copulas are broadly categorized into families based on their mathematical structure and dependence properties. The two most prominent families are Archimedean and Elliptical copulas [38]. Elliptical copulas are derived from elliptical distributions, such as the Gaussian (normal) and Student-*t* distributions. Common Archimedean copula are Clayton, Frank and Gumbel. This work concentrates on the Elliptical copula family, particularly the Gaussian and Student-*t* and Skew-*t* copulas, owing to their appropriateness for modeling symmetric and linear dependency patterns typically found in standardized rainfall data.

Gaussian

The Gaussian copula, which originates from the multivariate normal distribution, serves as a mathematical function designed to model the dependence structure among random variables [39]. Gaussian copulas allow for arbitrary marginal distributions and use the copula function to describe the dependency structure. According to [40], the density function of Gaussian copula is defined by Equation (17),

$$c_{\mathbf{R}}(u_1,u_2,u_3)=\frac{1}{|\mathbf{P}|^{1/2}}\exp\left(\frac{1}{2}\mathbf{q}^T(\mathbf{I}-\mathbf{P})\mathbf{q}\right) \tag{17}$$

For this trivariate case, $\mathbf{P}$ represents the symmetric linear correlation matrix defined as:

$$\mathbf{P}=\begin{bmatrix}1 & \rho_{12} & \rho_{13}\\ \rho_{21} & 1 & \rho_{23}\\ \rho_{31} & \rho_{32} & 1\end{bmatrix}$$

where $\mathbf{q}=(q_1,q_2,q_3)^T$ is the vector of normal scores, $\mathbf{I}$ is the 3 × 3 identity matrix.

Student-*t*

The Student-*t* copula is derived from the multivariate Student-*t* distribution and is particularly useful for capturing tail dependence. According to [40], the density function of Student-t copula is then defined as follows:

$$c_{v,\mathbf{P}}^{t}(u_1,u_2,u_3)=\frac{f_{v,\mathbf{P}}(t_v^{-1}(u_1),t_v^{-1}(u_2),t_v^{-1}(u_3))}{\prod_{i=1}^{3}f_v\,(t_v^{-1}(u_i))} \tag{18}$$

where $f_{v,\mathbf{P}}$ is the joint density of a $t_3(v,\mathbf{0},\mathbf{P})$-distributed random vector, f_v is the density of the univariate standard t-distribution with v degrees of freedom and $\mathbf{P}$ is the correlation matrix. The degrees of freedom parameter, denoted as v, plays an important role in determining the strength of dependence in the tails of the distribution. Specifically, lower values of v correspond to stronger tail dependence, while higher values result in weaker tail dependence. In the limiting case, as v approaches infinity, the Student-*t* copula converges to the normal copula, which lacks tail dependence [38].

Skew -*t*

The skew-*t* distribution is a flexible probability distribution that generalizes the Student-*t* distribution by incorporating a skewness parameter, enabling the modeling of asymmetric, heavy-tailed data. As described in Yusof and Syed Jamaludin [40], it is constructed by skewing the density of the Student-*t* distribution. According to [41], a copula is categorized as a skew $t_{3,v}$-copula if:

$$C_{3,v}(u_1,u_2,u_3;\boldsymbol{\mu},\mathbf{P},\boldsymbol{\gamma})=G_{3,v}\,(G_{1,v}^{-1}(u_1;\mu_1,\sigma_{11};\alpha_1),G_{1,v}^{-1}(u_2;\mu_2,\sigma_{22};\alpha_2),G_{1,v}^{-1}(u_3;\mu_3,\sigma_{33};\alpha_3);\boldsymbol{\mu},\mathbf{P},\boldsymbol{\gamma}) \tag{19}$$

Where $\boldsymbol{\mu}$ is the location vector, $\mathbf{P}$ is the matrix, $\boldsymbol{\gamma}$ is the skewness vector, $G_{1,v}^{-1}$ denotes the inverse of the univariate $t_{1,v}$ distribution function and $G_{3,v}$ is the distribution function of the trivariate skew $t_{p,v}$ distribution with the density.

The density function of the associated copula as defined in [42], is given by:

$$g_{3,v}(\mathbf{x};\boldsymbol{\mu},\mathbf{P},\boldsymbol{\gamma})=2\cdot t_{3,v}(\mathbf{x};\mathbf{P})\cdot T_{1,v+3}\left[\boldsymbol{\gamma}^T\mathbf{x}\left(\frac{v+3}{\mathbf{x}^T\mathbf{P}^{-1}\mathbf{x}+v}\right)^{1/2}\right] \tag{20}$$

where $t_{3,v}(\mathbf{x};\mathbf{P})$ is the probability density function (PDF) of the trivariate skew-t distribution, evaluated at the vector of transformed variables $\mathbf{x}$, with correlation matrix $\mathbf{P}$, skewness vector $\boldsymbol{\gamma}$, and v degrees of freedom.

The selection of the optimal copula model follows criteria that prioritize choosing models with the smallest AIC and BIC values. This approach not only aligns with best practices in dependence modeling but also strengthens the results by finding the most suitable copula.

Forecasting

ARMA-GARCH Model

In this study, model evaluation is conducted using testing datasets to assess final forecasting accuracy. This process helps avoid overfitting training data and ensures reliability in real-world scenarios. According to [43], for a fitted ARMA(*p*,*q*)-GARCH(*m*,*s*) model, the equations used for forecasting are as follows:

Conditional Mean Equation (ARMA):
For the *l*-step-ahead forecast, the formula is derived from the ARMA model structure:

$$\hat{r}_h(\ell) = E(r_{h+\ell}|F_h) = \varphi_0 + \sum_{i=1}^{p} \varphi_i \hat{r}_h(\ell - i) - \sum_{i=1}^{q} \theta_i a_h(\ell - i) \tag{21}$$

where $\hat{r}_h(\ell)$ is the ℓ-step-ahead forecast of the return $r_{h+\ell}$ and $E(r_{h+\ell}|F_h)$ is represents the conditional expectation. $\sum \varphi_i \hat{r}_h(\ell - i)$ (AR component) is the forecast depends recursively on its own previous forecasts. While $\sum \theta_i\, a_h(\ell - i)$ (MA component) is the part that incorporates the forecasted shocks.

Conditional Variance Equation (GARCH):
In the GARCH(1,1) model, multistep-ahead forecasts are generated by using the relation $a_t^2 = \sigma_t^2 \varepsilon_t^2$ and rewriting the volatility equation as:

$$\sigma_{t+1}^2 = \alpha_0 + (\alpha_1 + \beta_1)\sigma_t^2 + \alpha_1 \sigma_t^2(\varepsilon_t^2 - 1) \tag{22}$$

where α_1 is the ARCH parameter that measures how much volatility is influenced by past squared shocks and β_1 is the GARCH parameter that measures volatility persistence from past conditional variance.

ARMA-GARCH-Copula Model

This paper outlines an integrated methodological framework for generating synthetic multivariate rainfall data by combining univariate time-series models with multivariate copula functions. The implementation process is executed through the following step-by-step approach:

Step 1: Apply an ARMA-GARCH model to the daily rainfall data of each individual meteorological station to filter out localized temporal trends, conditional means, and time-varying volatility.

Step 2: Following the estimation of the ARMA-GARCH model, the standardized residuals were modelled using four alternative conditional distributions, namely the Normal, Student-t, Skewed Student-t, and Generalized Error Distribution (GED). These distributions were considered to accommodate different residual characteristics, including skewness and heavy-tailed behaviour, which are commonly observed in rainfall time series. The most appropriate residual distribution was selected based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The selected model was subsequently used to generate standardized residuals for copula modelling.

Step 3: Extract the residuals from the fitted ARMA-GARCH with innovation models for all three locations and standardize these residuals.

Step 4: Convert the standardized residuals into uniform marginal distributions constrained to the $[0,1]$ interval. This is achieved by applying the cumulative distribution function (CDF) corresponding to the parametric distribution of the ARMA-GARCH residuals.

Step 5: Assess the dependence structure of the uniformly transformed data using Kendall's tau and tail dependence coefficients. Kendall's tau is employed to quantify the strength and direction of pairwise dependence between rainfall series, while the upper tail dependence coefficients is used to evaluate the extent of dependence during extreme rainfall events.

Step 6: Fit the trivariate copula models (Gaussian, Student-t, and Skew-t) to the uniformly transformed data to capture the joint spatial and tail-dependence structures across the three stations.

Step 7: Utilize the best-fitting copula model (selected via AIC/BIC criteria) to simulate a large, dependent vector of uniform random variables $(\hat{u}_1, \hat{u}_2, \hat{u}_3)$.

Step 8: Transform the simulated uniform variables back into the standardized residual scale by applying the inverse CDF (quantile function) with the previously estimated location-specific parameters.

Step 9: Reintroduce the ARMA-GARCH temporal structure to the simulated residuals. Finally, convert these values back to the original physical rainfall scale (mm) using an inverse Box-Cox transformation with location-specific shape parameters (λ). The final output yields synthetic daily rainfall series that accurately preserve both individual temporal dynamics and joint spatial dependence structures.

Model Evaluation

The forecast results are used to evaluate the accuracy of the model predictions. The error metrics considered are Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), which are standard measures for assessing the performance and predictive accuracy of time series models. The MAE, MSE, and RMSE are defined in Equations (23), (24), and (25), respectively.

(a) Mean Absolute Error (MAE)
According to [44], the mean absolute error (MAE) is the average of the absolute distance between the observed and predicted value. The MAE is calculated by using:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - y_t'| \tag{23}$$

where *n* is the sample size, y_t is the actual value, and y_t' is the predicted value.

(b) Mean Squared Error (MSE)
According to [46], the Mean Squared Error (MSE) is a risk metric that calculates the average of the squared residuals. This squaring operation ensures that more substantial prediction errors contribute more significantly to the total error. The equation of MSE is given by:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - y_t')^2 \tag{24}$$

where *n* is the sample size, y_t is the actual value, and y_t' is the predicted value.

(c) Root Mean Squared Error (RMSE)
RMSE is a measure of the dispersion of residuals, indicating how closely the data points cluster around the best-fit line [46]. The formula for RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - y_t')^2} \tag{25}$$

where *n* is the sample size, y_t is the actual value, and y_t' is the predicted value.

Results and Discussion

This section presents a comprehensive analysis of the study's findings. It begins with a detailed exploration of descriptive statistics, which serve to characterize the core properties of the data, including central tendencies, dispersion, and distributional shapes. Subsequently, the results of the ARMA model are discussed, and the analysis is extended with the GARCH model. Finally, a copula model is used to examine the dependency structure providing deeper insights into their relationships.

Exploratory Data Analysis on Rainfall Data

Table 2 provides a suite of descriptive statistics, presented to summarize the central tendencies, variability, and distributional shapes of the rainfall data for each location. The measures include the mean, median, minimum, maximum, standard deviation, skewness, and kurtosis. Kota Bharu records the highest mean daily rainfall at 21.83 mm, closely followed by Kuala Krai at 21.33 mm, while Pos Gob exhibits a substantially lower average of 15.58 mm. Additionally, the kurtosis values for all stations exceed 3, implying the presence of heavier tails than in a normal distribution. In practice, this characteristic indicates a greater risk of extreme rainfall events, emphasizing the significance of using advanced statistical models capable of capturing such non-normal behavior.

Table 2. The descriptive statistics of the rainfall data for each station.

Stations	Mean	Standard Deviation	Skewness	Kurtosis	Coefficient of Variation (%)
Kota Bharu	21.83	43.09	5.22	47.22	197.38
Kuala Krai	21.33	35.73	2.88	12.14	167.53
Pos Gob	15.58	16.26	1.69	5.24	104.35

To resolve distributional asymmetries and stabilize conditional variance prior to time-series modelling, a Box-Cox transformation was deployed. The procedure yielded optimal λ (lambda) values of 0.1 for both Kota Bharu and Kuala Krai, and 0.2 for Pos Gob. Since these values deviate significantly from 1, the transformation was necessary to address the non-stationarity in variance across all stations. Furthermore, Table 3 reveals the result of stationarity tests. All stations exhibit stationarity in the mean, as confirmed by the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) test results, which yielded *p*-values below 0.05 for each station. This indicates that the null hypothesis of non-stationarity is rejected at the 5% significance level, validating the stationarity in the mean for all stations. Additionally, the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test results, with values greater than 0.05, support the conclusion of stationarity, as the null hypothesis of stationarity is not rejected at the 5% significance level.

Table 3. The result of stationarity test.

Stations	ADF (p-value)	PP (p-value)	KPSS (p-value)	Stationarity
Kota Bharu	0.01	0.01	0.1	✓
Kuala Krai	0.01	0.01	0.1	✓
Pos Gob	0.01	0.01	0.1	✓

Box-Jenkins-GARCH Model Identification

Based on the previous analysis, the daily rainfall series has shown stationary patterns over time. Therefore, an ARMA(p,q) model serves as the appropriate Box-Jenkins for modeling the daily rainfall data.

Model estimation and Diagnostic Checking

This study utilizes the auto.arima() function in R, which automatically identifies the optimal parameters for the Autoregressive Integrated Moving Average (ARIMA) model. Based on the AIC, BIC and parameter significance value, the ARMA(2,1) model is selected as the most suitable for Kota Bharu station, the ARMA(2,3) model for Kuala Krai station, and the ARMA(2,0) model for Pos Gob station. These models are chosen as they are statistically significant at the 5% significance level. Among all models for Kota Bharu station, the ARMA(2,1) model has the lowest AIC and BIC values of 1763.73 and 1783.457, respectively. Similarly, for Kuala Krai station, the ARMA(2,3) model has the lowest AIC and BIC values of 1557.87 and 1585.146, respectively. On the other hand, the ARMA(2,0) model is selected for Pos Gob station, as it has the lowest AIC and BIC values of 1453.61 and 1469.62, respectively. Table 4 presents the subsequent parameter estimates and their associated standard errors (S.E) for the selected best-fit ARMA models at the three stations. The distinct model structures for Kota Bharu, Kuala Krai, and Pos Gob reflect the unique temporal patterns and dynamics inherent to each location's data. These differences likely arise from a confluence of localized factors, including topography, land use, and microclimatic conditions.

\

Table 4. Result for estimated ARMA model parameters

	Kota Bharu ARMA (2,1)		Kuala Krai ARMA (2,3)		Pos Gob ARMA (2,0)	
Coefficients	Estimates	S.E	Estimates	S.E	Estimates	S.E
φ_1	1.1259	0.0960	0.9842	0.0614	0.1361	0.0492
φ_2	-0.2055	0.0657	-0.8424	0.0644	0.1535	0.0492
θ_1	-0.8061	0.0783	-0.6216	0.0678	-	-
θ_2	-	-	0.6092	0.0649	-	-
θ_3	-	-	0.3104	0.0620	-	-

To mathematically validate the adequacy of these selected ARMA configurations, diagnostic checks were systematically performed on their estimated residuals. The linear dependencies were evaluated using the Ljung-Box test, which operates under the null hypothesis (H_0) that the residuals are independently and identically distributed with no serial correlation. As presented in Table 5, the empirical p-values for all three stations comfortably exceed the 0.05 significance threshold, for example Pos Gob p-values $= 0.9971$. Consequently, H_0 cannot be rejected, confirming that the chosen ARMA frameworks successfully and exhaustively captured the linear autocorrelation structure within the daily rainfall time series. However, ensuring linear adequacy is insufficient for volatile hydrometeorological data. Thus, the raw residuals were subsequently tested for time-varying volatility clustering using the Engle ARCH Lagrange Multiplier (ARCH-LM) test. This test evaluates the null hypothesis (H_0) that no conditional heteroskedasticity exists in the residuals. In contrast to the linear diagnostic results, the empirical p-values for Kota Bharu (0.0440), Kuala Krai (0.0360), and Pos Gob (0.0010) all fall strictly below the 0.05 significance level as shown in Table 6. As a result, the null hypothesis of homoskedasticity is decisively rejected across all locations, providing robust empirical evidence of significant ARCH effects (volatility clustering) and formal justification for extending the framework to a conditional variance (GARCH) model.

Table 5. The Ljung-Box test

Stations	Test Statistics	p-value	Decision (at 5% level)	Autocorrelation
Kota Bharu	23.138	0.2821	Fail to Reject H_0	×
Kuala Krai	7.6573	0.9939	Fail to Reject H_0	×
Pos Gob	6.8641	0.9971	Fail to Reject H_0	×

Table 6. The ARCH-LM test

Stations	Test Statistics	p-value	Decision (at 5% level)	ARCH Effect
Kota Bharu	11.401	0.0440	Reject H_0	✓
Kuala Krai	45.326	0.0360	Reject H_0	✓
Pos Gob	29.467	0.0010	Reject H_0	✓

The GARCH model captures the time-varying volatility (heteroskedasticity) in the residuals of the ARMA model. The aim of this GARCH model was to improve modeling by taking into account past conditional variance (volatility) in such a way that it is dependent on previous return values [47]. The order of the GARCH model is determined based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the squared residuals of the ARMA model as shown in Figure 3. The diagnostic plots in Kota Bharu (Figure 3a) exhibited a significant spike at lag 1 in the PACF alongside a gradual geometric decay up to lag 2 in the ACF. Kuala Krai (Figure 3b) displayed prominent, highly isolated spikes strictly at lag 1 for both functions and Pos Gob (Figure 3c) revealed significant spikes at the first lag with minor lingering decay into the second lag. Given that the significant conditional heteroskedasticity across all stations was concentrated within these initial lags, a comprehensive suite of low order specifications, specifically GARCH(1,1), GARCH(1,2), GARCH(2,1), and GARCH(2,2), was evaluated to select the most parsimonious and statistically optimal framework for each location.

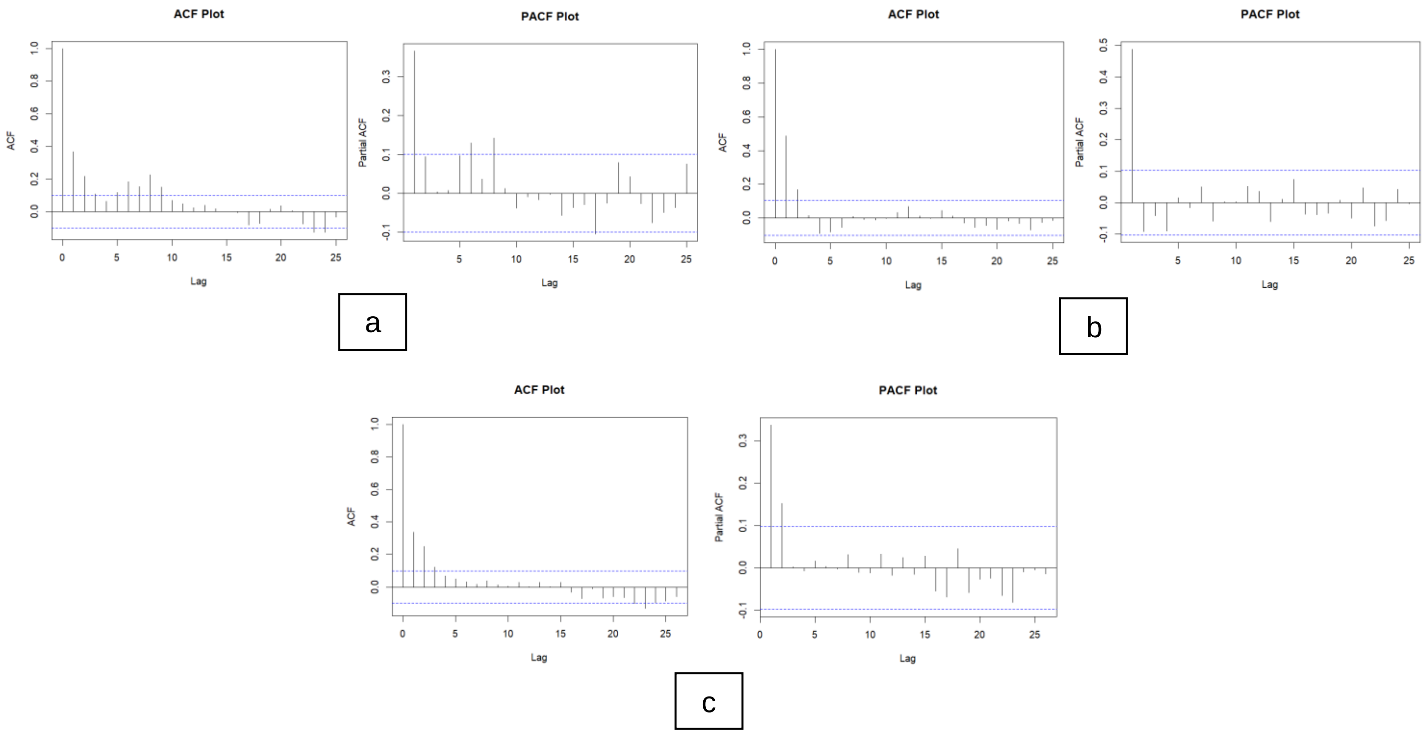

Figure 3. The squared residuals of ACF and PACF plots of the (a) Kota Bharu, (b) Kuala Krai, (c) Pos Gob.

Subsequent analysis guided by these diagnostics led to distinct ARMA-GARCH frameworks for each monitoring station, as summarized in Table 7. Specifically, an ARMA(2,1)-GARCH(1,1) model was selected for Kota Bharu, an ARMA(2,3)-GARCH(2,2) model for Kuala Krai, and an ARMA(2,0)-GARCH(1,1) model for Pos Gob. The estimated parameters reveal varying degrees of persistence in both the conditional mean and conditional variance across the stations. For instance, the autoregressive parameter φ_1 is prominent in Kota Bharu (0.5928) and Kuala Krai (0.2494), indicating notable short-term memory in the mean process, whereas Pos Gob's mean structure relies entirely on its second-order lag ($\varphi_2 = 0.0094$). In terms of volatility dynamics, all stations exhibit high volatility persistence, as evidenced by the GARCH coefficients (β). This persistence is especially strong in Kota Bharu ($\beta_1 = 0.9810$) and Pos Gob ($\beta_1 = 0.9402$), suggesting that volatility shocks (periods of highly fluctuating or extreme data) take a considerable amount of time to decay. In contrast, Kuala Krai distributes its volatility memory across two lags ($\beta_1 = 0.7034$ and $\beta_2 = 0.2410$).

Table 7. Result for estimated ARMA-GARCH model parameters

	Kota Bharu ARMA (2,1)-GARCH(1,1)		Kuala Krai ARMA (2,3)-GARCH(2,2)		Pos Gob ARMA (2,0)-GARCH(1,1)	
Coefficients	Estimates	S.E	Estimates	S.E	Estimates	S.E
μ	0.0070	0.1201	0.0011	0.1146	-0.0512	0.0761
φ_1	0.5928	0.5815	0.2494	0.0274	0.0007	0.0502
φ_2	-0.0085	0.0517	-0.9325	0.0672	0.0094	0.0508
θ_1	-0.5948	0.5794	-0.2215	0.0552	-	-
θ_2	-	-	0.9531	0.0552	-	-
θ_3	-	-	0.0105	0.0399	-	-
ω	0.1082	0.0934	0.1335	0.1632	0.0407	0.0579
α_1	0.0000	0.0160	0.0230	0.0773	0.0416	0.0295
α_2	-	-	0.0000	0.0770	-	-
β_1	0.9810	0.0107	0.7034	0.0574	0.9402	0.0524
β_2	-	-	0.2410	0.0557	-	-

The Jarque-Bera (JB) test is a test that measures how much the skewness and kurtosis of the dataset differs from those of a standard normal distribution. In this study, a significance level of α = 0.05 was used for the normality test. The decision to accept or reject the normality assumption is based on the *p*-value derived from the JB test. A *p*-value > 0.05 indicated normality, while a *p*-value ⩽ 0.05 led to its rejection. Notably, for the ARMA-GARCH models in Kota Bharu and Kuala Krai, the *p*-values were 0.0086 and 0.0414 as shown in Table 8, respectively, both of which are significantly lower than the 0.05 threshold. This indicates that the assumption of normality is not met by these two locations. In contrast, Pos Gob has a *p*-value of 0.0686, which exceeds the 0.05 threshold, suggesting that the normality assumption is satisfied for this location.

Table 8. The Jarque-Bera test

Stations	Test Statistics	*p*-value	Decision (at 5% level)	Normality
Kota Bharu	9.5138	0.0086	Reject H_0	×
Kuala Krai	6.3690	0.0414	Reject H_0	×
Pos Gob	5.3589	0.0686	Accept H_0	✓

Since normality assumption is not satisfied for Kota Bharu and Kuala Krai, additional distributions of GARCH such as the *t* -distribution, skew-*t* distribution, and Generalized Error Distribution (GED), are considered. Table 9 and Table 10 present the results of the ARMA-GARCH model for Kota Bharu and Kuala Krai using these alternative distributions. As shown in Table 9, the ARMA-GARCH model with Generalized Error Distribution (GED) innovations achieves the lowest information criteria (AIC = 4.5823; BIC = 4.6649). This indicates that capturing heavy-tailed behavior via the GED framework provides a superior statistical fit over standard symmetric or skewed distributions. To verify model adequacy, Ljung-Box diagnostic tests (Q and Q^2) were conducted on the standardized (ε_t) and squared standardized (ε_t^2) residuals at lags 10 and 15. At lag 10, the model exhibits minor residual dependence, with p-values falling slightly below the standard threshold ($p = 0.0354$ for Q; $p = 0.0452$ for Q^2), refer (Table 9, LBQ (10) and LBQ2 (10)). However, at the extended lag 15, the p-values increase substantially and comfortably exceed the 0.05 significance level ($p = 0.1752$ for Q; $p = 0.1129$ for Q^2), refer (Table 9, LBQ (15) and LBQ2 (15)). This lack of serial correlation and conditional heteroscedasticity at higher lags confirms that the GED-specified model successfully filters out the underlying temporal dependencies, establishing its overall statistical robustness for empirical analysis.

Similarly, Table 10 outlines the parameter estimations, information criteria, and diagnostic checks for the ARMA-GARCH model applied to the Kuala Krai station across the four innovation distributions. Consistent with the findings from Kota Bharu, the model utilizing the Generalized Error Distribution (GED) innovations yields the most optimal performance, achieving the lowest information criteria with an AIC of 4.2628 and a BIC of 4.3913. For the optimal GED specification, the diagnostic tests yield highly favorable results. The p-values for the standardized residuals are 0.9842 at lag 10 and 0.9981 at lag 15, while the p-values for the squared standardized residuals are 0.5635 at lag 10 and 0.2661 at lag 15. These diagnostic outcomes confirm that the ARMA-GARCH model with GED innovations successfully captures the underlying dynamics of the series, ensuring its statistical validity and robustness. However, Pos Gob station exhibits normal behavior. Therefore, no additional innovation distribution of GARCH is required, and the estimated parameters from Table 7 are applied.

Table 9. Result for estimated ARMA-GARCH model parameters with GARCH innovation for Kota Bharu station

Parameter	Innovations			
	Normal	t	Skew-t	GED
	Parameter Estimations (Standard Error)			
μ	0.0070 (0.1201)	0.0100 (0.1251)	0.0255 (0.1182)	-0.0956 (0.0884)
φ_1	0.5928 (0.5815)	-0.8126 (0.2671)	0.5958 (0.7012)	0.6120 (0.2085)
φ_2	-0.0085 (0.0517)	0.0227 (0.0534)	-0.0187 (0.0523)	-0.0335 (0.0497)
θ_1	-0.5948 (0.5794)	0.8134 (0.2629)	-0.5955 (0.6959)	-0.6580 (0.2027)
ω	0.1082 (0.0934)	0.0060 (0.0651)	0.0067 (0.0632)	0.0062 (0.0508)
α_1	0 (0.0160)	0 (0.0118)	0 (0.0114)	0 (0.0089)
β_1	0.9810 (0.0107)	0.9990 (0.0001)	0.999 (0.0001)	0.9990 (0.0000)
ζ	-	-	0.7842 (0.0747)	-
ν	-	99.9992 (84.8732)	59.9998 (41.0067)	4.0428 (0.5773)
	Information Criterion			
AIC	4.6268	4.6341	4.6262	**4.5823**
BIC	4.6991	4.7168	4.7192	**4.6649**
	Diagnostics Checking (*p*-values)			
LBQ (10)	15.4660 (0.1160)	14.9780 (0.1329)	15.4360 (0.1169)	19.4040 (0.0354)
LBQ (15)	15.9920 (0.3826)	15.6260 (0.4073)	15.9630 (0.3845)	19.9150 (0.1752)
LBQ2 (10)	15.0870 (0.1289)	13.8530 (0.1798)	15.0460 (0.1304)	18.6260 (0.0452)
LBQ2 (15)	17.7090 (0.2783)	16.4900 (0.3502)	17.7260 (0.2773)	21.8110 (0.1129)

Table 10. Result for estimated ARMA-GARCH model parameters with GARCH innovation for Kuala Krai station

Parameter	Innovations			
	Normal	t	Skew-t	GED
	Parameter Estimations (Standard Error)			
μ	0.0011 (0.1146)	-0.0030 (0.1094)	-0.0016 (0.1119)	-0.0535 (0.1096)
φ_1	0.2494 (0.0274)	0.3435 (0.3053)	0.2551 (0.0790)	0.2390 (0.0844)
φ_2	-0.9325 (0.0672)	-0.8107 (0.1776)	-0.9074 (0.1971)	-0.9443 (0.0553)
θ_1	-0.2215 (0.0552)	-0.3243 (0.3200)	-0.2237 (0.0912)	-0.2399 (0.1169)
θ_2	0.9531 (0.0552)	0.8409 (0.1692)	0.9332 (0.1563)	0.9645 (0.0522)
θ_3	0.0105 (0.0399)	-0.0255 (0.0694)	0.0069 (0.0671)	-0.0145 (0.0754)
ω	0.1335 (0.1632)	0.0062 (0.0574)	0.0068 (0.0566)	0.1062 (0.1291)
α_1	0.0230 (0.0773)	0 (0.0154)	0 (0.0166)	0.0251 (0.0704)
α_2	0 (0.0770)	0 (0.0159)	0 (0.0178)	0 (0.0742)
β_1	0.7034 (0.0574)	0.0235 (0.0014)	0.0289 (0.0028)	0.6482 (0.0858)
β_2	0.2410 (0.0557)	0.9755 (0.0001)	0.9701 (0.0004)	0.3011 (0.0831)
ζ	-	-	0.8752 (0.0940)	-
ν	-	99.9999 (97.0102)	59.9999 (45.3090)	3.4107 (0.5808)
	Information Criterion			
AIC	4.2892	4.2995	4.3027	**4.2628**
BIC	4.4069	4.428	4.4419	**4.3913**
	Diagnostics Checking (*p*-values)			
LBQ (10)	2.4286 (0.9919)	2.0763 (0.9957)	2.1252 (0.9953)	2.8756 (0.9842)
LBQ (15)	3.5109 (0.9990)	3.0752 (0.9995)	3.3686 (0.9992)	3.8895 (0.9981)
LBQ2 (10)	7.1184 (0.7142)	8.1790 (0.6114)	7.0527 (0.7205)	8.6716 (0.5635)
LBQ2 (15)	15.9900 (0.3827)	15.9970 (0.3822)	15.6480 (0.4058)	17.935 (0.2661)

Fitting Copula

Before selecting and fitting a copula model, it is essential to conduct a preliminary non-parametric dependence analysis. Following the ARMA–GARCH modelling, the standardized residuals are transformed into uniform variables using their corresponding cumulative distribution functions, satisfying the marginal uniformity requirement for copula modelling. This transformation ensures that the dependence structure can be modelled independently of the marginal distributions, thereby improving the validity of the fitted copula. Subsequently, Kendall's tau is employed to assess the overall pairwise dependence among the rainfall stations. The results from the Kendall tau test presented in Table 11 provide foundational insights into both the overall correlation structure and the extreme tail behavior of the precipitation or hydrological variables between the stations of Kota Bharu, Kuala Krai, and Pos Gob.

The empirical values across all three pairs are remarkably close to zero, signalling weak overall association. Specifically, the pair comprising Kota Bharu and Kuala Krai exhibits a slight positive dependence ($\tau = 0.043$), whereas the pairs involving Pos Gob reveal a minor negative dependence ($\tau = -0.025$ for Kota Bharu - Pos Gob, and $\tau = -0.039$ for Kuala Krai - Pos Gob). Beyond global correlation, understanding how these variables behave during extreme events is vital for risk assessment, particularly when modelling concurrent floods or heavy rainfall. The empirical upper tail dependence coefficients reveal how joint exceedance probabilities change as events become more extreme. Generally, the upper tail behavior reveals that while there is a slight probability of concurrent heavy events at the 90th percentile, while this joint probability vanishes completely at the 95th percentile. This rapid decay signifies limited concurrent extreme rainfall behaviour among the locations. However, the multivariate copula models were subsequently incorporated to evaluate whether the remaining dependence structure could provide additional information for multisite rainfall forecasting.

Table 11. Kendall tau test

Pair	Kendall tau (τ)	Upper Tail (u = 0.90)	Upper Tail (u = 0.95)
Kota Bharu - Kuala Krai	0.043	0.125	0
Kota Bharu - Pos Gob	-0.025	0.031	0
Kuala Krai - Pos Gob	-0.039	0.091	0

Following ARMA–GARCH modelling, the residuals are standardized, resulting in time-independent series for each rainfall station. Nevertheless, dependence may still exist across stations due to the underlying spatial relationship between rainfall observations. The use of copula modeling in this work is critical for describing the spatial dependency structure among the three rainfall locations. The optimal copula parameters are estimated, and the AIC and BIC values are presented in Table 12. The Gaussian copula is selected as it exhibits the smallest AIC and BIC values.

Table 12. Result for fitting Copula

Copula	Estimated Parameter	Loglikelihood	AIC	BIC
Gaussian	$\rho_1 = 0.05675$, $\rho_2 = -0.04119$, $\rho_3 = -0.07442$	1.8337	2.3325	14.0239
Student-t	$\rho_1 = 0.05741$, $\rho_2 = -0.04141$, $\rho_3 = -0.07326$ $\nu = 147.36796$	1.6318	4.7362	20.3248
Skew-t	$\rho_1 = 0.06074$, $\rho_2 = -0.22427$, $\rho_3 = -0.06873$ $\gamma_1 = -0.32533$, $\gamma_2 = -0.03296$, $\gamma_3 = 0.89448$ $\nu = 398.30118$	2.62690	12.7461	47.8205

Forecasting

The subsequent phase of the research involves the implementation of forecasting techniques to model prospective outcomes. For Kota Bharu, the optimal statistical representation was achieved through an ARMA(2,1)-GARCH(1,1) specification with Generalized Error Distribution (GED) innovations. The parameter estimates for this model, as summarized in Table 9, are used for further analysis. The ARMA(2,1)-GARCH(1,1) model with GED innovations can be expressed as:

$$y_t = -0.0956 + 0.6120y_{t-1} - 0.0335y_{t-2} + u_t - 0.6580u_{t-1}$$

$$\sigma_t = 0.0062 + 0\alpha_{t-1}^2 + 0.9990\sigma_{t-1}^2, \quad \varepsilon_t \sim \mathrm{GED}_{4.0428}^*$$

In contrast to the other stations, Kuala Krai station exhibits a more complex dynamic structure, which is most accurately captured by an ARMA(2,3)-GARCH(2,2) model employing Generalized Error Distribution (GED) innovations. The use of GED innovations is particularly critical, as it provides the flexibility to model the leptokurtic properties of the series that is, the higher likelihood of extreme observations compared to a normal distribution. The parameter estimates for this model, summarized in

Table 10, are used for further analysis. The ARMA(2,3)-GARCH(2,2) model with GED innovations can be expressed as:

$$y_t = -0.0535 + 0.2390 y_{t-1} - 0.9443 y_{t-2} + u_t - 0.2399\, u_{t-1} + 0.9645\, u_{t-2}$$

$$\sigma_t = 0.1062 + 0.0251\alpha_{t-1}^2 + 0\alpha_{t-2}^2 + 0.6482\sigma_{t-1}^2 + 0.3011\sigma_{t-2}^2, \quad \varepsilon_t \sim \mathrm{GED}_{3.1407}^*$$

Furthermore, for the Pos Gob station, the data is best fitted by an ARMA(2,0)-GARCH(1,1) model with normal distribution. Additionally, the parameter estimates for this model, summarized in Table 7, are used for further analysis. The selection of a normal distribution for the innovations indicates that extreme values in the Pos Gob rainfall series are less pronounced compared to those observed at Kuala Krai, suggesting a relatively moderate variability in daily rainfall. Specifically, from Table 7, the ARMA(2,0)-GARCH(1,1) model with normal distribution can be expressed as:

$$y_t = -0.0512 + 0.0007 y_{t-1} - 0.0094 y_{t-2}$$

Table 13 shows the MAE, MSE and RMSE values for the ARMA (2, 1)-GARCH (1,1), ARMA (2,3)-GARCH (2,2) and ARMA (2,0)-GARCH (1,1) models for Kota Bharu, Kuala Krai and Pos Gob station, respectively. The relatively low values of these error metrics indicate that the models fit the data well, capturing both the mean structure and the volatility of daily rainfall effectively. This strong model performance suggests that the ARMA-GARCH framework is suitable for forecasting rainfall at these stations, as it can accommodate autocorrelation in the mean process and conditional heteroskedasticity in the variance.

Table 13. The results of MAE, MSE, RMSE for each station

Station	Model	Forecast Evaluation		
		MAE	MSE	RMSE
Kota Bharu	ARMA (2,1) - GARCH (1,1)	0.9992	1.8950	1.3766
Kuala Krai	ARMA (2,3) - GARCH (2,2)	1.5132	5.6084	2.3682
Pos Gob	ARMA (2,0) - GARCH (1,1)	5.4416	49.3204	7.0228

Comparison between ARMA-GARCH and ARMA-GARCH-Copula model

To further evaluate the performance of the modeling approaches, synthetic rainfall data were generated from both the ARMA-GARCH and ARMA-GARCH-Copula frameworks. For each station, the ARMA-GARCH model was first used to simulate future rainfall series based on the fitted parameters, while the ARMA-GARCH-Copula model additionally captured the dependence structure between stations through the copula function. The accuracy of the synthetic data was then assessed by comparing it against the observed training data using standard error metrics, including MAE, MSE and RMSE. The corresponding error measures are presented in Table 14.

The results presented in Table 14 indicate that the inclusion of the copula component does not consistently improve simulation accuracy across all stations. For Kota Bharu, the ARMA-GARCH model achieved lower MAE (19.7970 vs. 21.6136), MSE (1495.9409 vs. 1790.4564), and RMSE (38.6774 vs. 42.3138) compared with the ARMA-GARCH-Copula model. A similar pattern was observed for Kuala Krai, where the ARMA-GARCH model produced slightly lower errors in terms of MAE (20.3171 vs. 21.4138), MSE (1640.7060 vs. 1672.4087), and RMSE (40.5056 vs. 40.8950). These findings suggest that the temporal dynamics captured by the ARMA-GARCH model were sufficient to represent the rainfall variability at these stations, while the additional spatial dependence structure introduced by the copula did not provide substantial improvement. In contrast, for Pos Gob, the ARMA-GARCH-Copula model showed marginal improvement, with slightly lower MAE (14.0848 vs. 13.8460) and RMSE (21.4987 vs. 21.0091) compared with the ARMA-GARCH model, although the MSE increased slightly (462.1932 vs. 441.3856). This indicates that incorporating inter-station dependence may provide additional predictive information for certain locations where rainfall variability is more strongly influenced by spatial interactions.

Overall, the results demonstrate that the effectiveness of the ARMA-GARCH-Copula framework depends on the strength of the dependence structure among the rainfall stations. When the spatial dependence is relatively weak, the copula component may introduce additional complexity without significant improvement in simulation accuracy. This finding is consistent with the dependence analysis, where the rainfall stations exhibited moderate to weak dependence based on Kendall's tau values. Therefore, while the copula approach provides a flexible framework for modelling multivariate rainfall dependence, its contribution to prediction accuracy is highly dependent on the underlying spatial dependence characteristics of the study area.

Table 14. Error Comparison between ARMA-GARCH and ARMA-GARCH-COPULA

Station	ARMA-GARCH			ARMA-GARCH-COPULA		
	MAE	MSE	RMSE	MAE	MSE	RMSE
Kota Bharu	19.7970	1495.9409	38.6774	21.6386	1790.4564	42.3138
Kuala Krai	20.3171	1640.7060	40.5056	21.4138	1672.4087	40.8950
Pos Gob	13.8460	441.3856	21.0091	14.0848	462.1932	21.4987

Conclusions

This study investigated the role of copula-based dependence modelling within an ARMA-GARCH framework for rainfall forecasting. The ARMA-GARCH models were first employed to capture the temporal dependence and conditional variability of individual rainfall series. Next, the standardized residuals were transformed into uniform margins to examine the remaining cross-station dependence. The dependence analysis revealed weak residual association among the rainfall stations, with Kendall's tau values ranging from 0.01 to 0.05. Although some upper tail association was observed at the 90th percentile threshold, the dependence gradually diminished as more extreme thresholds were considered, indicating limited concurrent extreme rainfall behaviour among the investigated locations. Multivariate copula models were subsequently incorporated to evaluate whether the remaining dependence structure could provide additional information for multisite rainfall forecasting. The comparative assessment between the ARMA-GARCH and ARMA-GARCH-copula frameworks showed that the ARMA-GARCH model achieved lower RMSE and MAE values, indicating better forecasting performance for the selected rainfall stations. This finding suggests that the temporal dynamics and conditional variability captured by the ARMA-GARCH model represent the dominant characteristics, while the residual cross-station dependence contributed only marginally to the forecasting process.

The findings of this study highlight that the incorporation of copula-based dependence structures should be evaluated based on the characteristics of the residual dependence after temporal modelling rather than being applied solely due to the nature of rainfall data. In cases where the remaining dependence among stations is limited, additional copula modelling may introduce complexity without substantial improvement in forecasting accuracy. Therefore, residual dependence diagnostics provide an important step in determining the suitability of multivariate dependence modelling approaches for rainfall forecasting applications. Future research may extend this framework by considering a larger number of rainfall stations with different climatic and geographical characteristics, as well as investigating whether copula-based approaches provide greater benefits in regions exhibiting stronger spatial dependence or more pronounced joint extreme rainfall behaviour.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

The authors are grateful to the Malaysian Meteorological Department for providing the rainfall data. This research is fully supported by Fundamental Research Grant Scheme (FRGS) grant, FRGS/1/2022/STG06/UITM/02/5. The authors fully acknowledged Ministry of Higher Education (MOHE) and Universiti Teknologi MARA for the approved fund, ensuring this important study viable and effective.

References

[1] Sariff, W. M. H. S. M., Deni, S. M., Ghani, N. a. M., Ul-Saufie, A. Z., & Radi, N. F. A. (2025). OUTLIER DETECTION IN RAINFALL DATA USING EXTREME VALUE THEORY. Jurnal Teknologi, 87(5), 887–898.
[2] Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review. PLoS Currents.

[3] Khan, A., Khan, H. U., & Ismail, M. (2025). Forecasting of monthly rainfall in DIR (L) KP Pakistan with ARMA models. Journal of Asian Development Studies, 14(1), 734–750.
[4] Valipour, M., Banihabib, M. E., & Behbahani, S. M. R. (2012). Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of Hydrology, 476, 433–441.
[5] Izati, P. P., Prastyo, D. D., & Akbar, M. S. (2024). Modeling the volatility of world energy commodity prices using the GARCH-Fractional Cointegration Model. Procedia Computer Science, 234, 412–419.
[6] Ampadu, S., Mensah, E. T., Aidoo, E. N., Boateng, A., & Maposa, D. (2024). A comparative study of error distributions in the GARCH model through a Monte Carlo simulation approach. Scientific African, 23, e01988.
[7] Lu, L., Lei, Y., Yang, Y., Zheng, H., Wang, W., Meng, Y., Meng, C., & Zha, L. (2023). Assessing nickel sector index volatility based on quantile regression for Garch and Egarch mod-els: Evidence from the Chinese stock market 2018–2022. Resources Policy, 82, 103563.
[8] Yusof, F., Kane, I. L., & Yusop, Z. (2013). Hybrid of ARIMA-GARCH modeling in rainfall Time Series. Jurnal Teknologi/Jurnal Teknologi, 63(2).
[9] Pandey, P. K., Tripura, H., & Pandey, V. (2018). Improving prediction accuracy of rainfall time series by hybrid SARIMA–GARCH modeling. Natural Resources Research, 28(3), 1125–1138.
[10] Bárdossy, A., & Pegram, G. G. S. (2009). Copula based multisite model for daily precipitation simulation. Hydrology and Earth System Sciences, 13(12), 2299–2314.
[11] Bhatti, M. I., & Do, H. Q. (2019). Recent development in copula and its applications to the energy, forestry and environmental sciences. International Journal of Hydrogen Energy, 44(36), 19453–19473.
[12] Han, Y., Li, J., Zhao, M., Guo, H., Wang, C., Huang, H., & Cao, R. (2025). Analysis of rainfall abundance and drought occurrence and probability of flood and drought.
[13] Bak, M., Kim, Y., Jan, S. S. S., Nam, W., Hong, E., & Kim, S. (2026). Assessment of environmental ecological drought using environmental flow-based indices and copula analysis in an urban tributary of South Korea. Ecohydrology & Hydrobiology, 26(3), 100781.
[14] Latif, S., & Mustafa, F. (2020). Trivariate distribution modelling of flood characteristics using copula function—A case study for Kelantan River basin in Malaysia. AIMS Geosciences, 6(1), 92–130.
[15] Vernieuwe, H., Vandenberghe, S., De Baets, B., & Verhoest, N. E. C. (2015). A continuous rainfall model based on vine copulas. Hydrology and Earth System Sciences, 19(6), 2685–2699.
[16] Lazoglou, G., Anagnostopoulou, C., Skoulikaris, C., & Tolika, K. (2019). Bias Correction of Climate Model's Precipitation Using the Copula Method and Its Application in River Basin Simulation. Water, 11(3), 600.
[17] Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. International Economic Review.
[18] Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. Econometrica.
[19] Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics.
[20] Ismail, W. R., & Haghroosta, T. (2018). Extreme weather and floods in Kelantan state, Malaysia in December 2014. In Research in Marine Sciences (Vols. 3–1, Issue 1, pp. 231–244).
[21] Naji, A. S. M., Yaziz, S. R., Zakaria, R., Mohamad, N. N., & Radi, N. F. A. (2024). Gold price forecasting using ARIMA-GARCH model during COVID-19 pandemic outbreak. AIP Conference Proceedings, 3080, 090018.
[22] Marimuthu, S., Mani, T., Sudarsanam, T. D., George, S., & Jeyaseelan, L. (2022). Preferring Box-Cox transformation, instead of log transformation to convert skewed distribution of outcomes to normal in medical research. Clinical Epidemiology and Global Health, 15, 101043.
[23] Roza, A., Violita, E. S., & Aktivani, S. (2022). Study of Inflation using Stationary Test with Augmented Dickey Fuller & Phillips-Peron Unit Root Test (Case in Bukittinggi City Inflation for 2014-2019). EKSAKTA Berkala Ilmiah Bidang MIPA, 23(02), 106–116.
[24] Sathyanarayana, S., & Mohanasundaram, T. (2025). Stationarity and Unit Roots in Time Series: Theoretical Insights and Practical Considerations. IRA-International Journal of Management & Social Sciences (ISSN 2455-2267), 21(2), 46.
[25] Molina, J., Zazo, S., Rodr´ıguez-Gonza´lvez, P., & Gonza´lez-Aguilera, D. (2016). Innovative Analysis of Runoff Temporal Behavior through Bayesian Networks. Water, 8(11), 484.
[26] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time Series analysis: Forecasting and Control. John Wiley & Sons.
[27] Alaminos, D., Salas, M. B., &Partal-Uren˜a, A. (2024). Hybrid ARMA-GARCH-Neural Networks for intraday strategy exploration in high-frequency trading. Pattern Recognition, 148, 110139.
[28] Nakatsuma, T., Tsurumi, H. (1996). ARMA-GARCH models: Bayes Estimation versus MLE, and Bayes Non-stationarity Test. In Department of Economics, Rutgers University, Department of Eco-nomics, Rutgers University (Working Paper No. 1996-19). Department of Economics, Rutgers University.
[29] Radzi, N. N., & Yaziz, N. S. (2021). Forecasting Malaysian overnight islamic interbank rate us-ing the Box-Jenkins model. Data Analytics and Applied Mathematics (DAAM), 2(1), 38–51.
[30] Burns, P. (2003). Robustness of the Ljung-Box Test and its Rank Equivalent. SSRN Electronic Journal.
[31] Glinskiy, V., Ismayilova, Y., Khrushchev, S., Logachov, A., Logachova, O., Serga, L., Yambart-sev, A., Zaykov, K. (2024). Modifications to the Jarque–Bera test. Mathematics, 12(16), 2523.
[32] Mashavhela, D., Ravele, T., & Sigauke, C. (2026). Exploring the dynamics of ZAR/USD Exchange RateVolatility using the FGARCH and First-Order Beta-Skew-T-EGARCH models. Econometrics, 14(3), 37.
[33] Chumnaul, J., Taowato, A., & Lupboon, C. (2025). Performance of Pearson, Spearman, and Kendall correlation coefficients under bivariate count distributions. European Journal of Statistics, 5, 14.
[34] Echaust, K., & Just, M. (2021). Tail Dependence between Crude Oil Volatility Index and WTI Oil Price Movements during the COVID-19 Pandemic. Energies, 14(14), 4147.
[35] Fauzi, M. F. M., Rauf, U. F. A., Zeephongsekul, P., & Fataf, N. A. A. (2018). Tail dependence between rainfall characteristics using Copula: The case of Pahang state (Version 1). RMIT University.
[36] Laux, P., Vogl, S., Qiu, W., Knoche, H. R., & Kunstmann, H. (2011). Copula-based statistical refinement of

precipitation in RCM simulations over complex terrain. Hydrology and Earth System Sciences, 15(7), 2401–2419.
[37] Sklar, A. (1959). Fonctions de r´epartition a` n dimensions et leurs marges. Publications de l'Institut de Statistique de l'Universit´e de Paris, 8, 229-31.
[38] Alokley, S. A., Araichi, S., & Alomair, G. (2024). Exploring the relationship and predictive accuracy for the Tadawul All share index, oil prices, and Bitcoin using copulas and machine learning. Energies, 17(13), 3241.
[39] Lian, Y., Yasmin, S., & Haque, M. M. (2024). Influence of road safety policies on the long-term trends in fatal Crashes: A Gaussian Copula-based time series count model with an autoregressive moving average process. Accident Analysis & Prevention, 211, 107795.
[40] Yusof, F. & Syed Jamaludin, S. S. (2023). Copula Modelling and Its Application. PENERBIT UTM PRESS.
[41] Jaworski, P., Durante, F., Ha¨rdle, W. K., & Rychlik, T. (2010). Copula Theory and its applications: Proceedings of the Workshop Held in Warsaw, 25-26 September 2009. Springer Science & Business Media.
[42] Kollo, T., Pettere, G., & Valge, M. (2015). Tail dependence of skew t-copulas. Communications in Statistics - Simulation and Computation, 46(2), 1024–1034.
[43] Tsay, R. S. (2010). Analysis of financial time series. John Wiley Sons.
[44] Jiang, Q., Jin, X., Lee, S., & Yao, S. (2017). Protein secondary structure prediction: A sur-vey of the state of the art. Journal of Molecular Graphics and Modelling, 76, 379–402.
[45] Khojasteh, M., Bahreinian, S., Riasi, A. (2025). Predictive modeling for savonius hy-drokinetic turbine performance: a machine learning investigation. Energy, 340, 139109.
[46] Ali, F., Sarwar, A., Bakhsh, F. I., Ahmad, S., Shah, A. A., & Ahmed, H. (2022). Param-eter extraction of photovoltaic models using atomic orbital search algorithm on a decent basis for novel accurate RMSE calculation. Energy Conversion and Management, 277, 116613.
[47] Ma¨ınassara, Y. B., Kadmiri, O., & Saussereau, B. (2022). Estimation of multivariate asymmetric power GARCH models. Journal of Multivariate Analysis/Journal of Multivariate Analysis, 192, 105073.