

# A Bayesian Copula Approach for Flood Analysis

Izzat Fakhruddin Kamaruzaman<sup>a,b,\*</sup>, Wan Zawiah Wan Zin<sup>a</sup>,  
Noratiqah Mohd Ariff<sup>a</sup>

<sup>a</sup> School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia; <sup>b</sup> Faculty of Business, Multimedia University, Jalan Ayer Keroh Lama, 75450 Bukit Beruang, Melaka, Malaysia

---

**Abstract** This study aims to provide joint modelling of rainfall characteristics in Peninsular Malaysia using two-dimensional copula. Two commonly regarded as important variables in the field of hydrology, namely rainfall severity and duration derived using the Standard Precipitation Index (SPI) and their univariate marginal distributions are further identified by fitting into several distributions. The paper uses a Bayesian framework to estimate the parameter values in the marginal and copula model. The approximation of the posterior distribution by random sampling has been done by Monte Carlo Markov Chain (MCMC). Next, the authors compared these findings with those based on the classical procedure. The results indicated that the Bayesian approach can be substantially more reliable in parameter estimation for small samples.

**Keywords:** Bayesian analysis, copula, MCMC, rainfall modelling.

---

## Introduction

Floods are one of the most common natural disasters that frequently occur around the world. These climatic phenomena are difficult to comprehend and continue to haunt modern society due to their big impacts on a variety of domains, including the economic, environmental and social sectors. The current floods in Malaysia appear to be one of the worst in recent decades. In 2017, the number of flood incidents recorded was 498 cases [1]. In terms of monetary, a typical flood cost RM1.2 billion in 2012 [2], more than damages incurred because of other disasters. Due to this, flood protection has a high priority in countries that are in danger. As a measure of flood mitigation, it is very significant to determine the probabilistic rainfalls characteristics.

There are many major rainfall characteristics in floodplain management and hydrologic design such as severity, duration, intensity and depth. These rainfall characteristics are normally employed by various researchers in their hydrological studies [3-7]. As these characteristics are random in nature and taking into accounts some of them are associated with flood, the appropriate method to discover hydrologic events usually by using multivariate probabilistic analysis. However, in traditional multivariate analysis, each univariate marginal distribution must come from the same family and this will lead to a very limited number of available models. Copulas, on the other hand, present a general framework for creating multivariate distributions which make use of any given univariate marginals and a copula function  $C$  that ties these marginals [8]. The importance of copulas is rooted in Sklar's theorem [9] that states that any multivariate distribution can be represented as a copula function of its marginal distributions, making

**\*For correspondence:**  
izzat.kamaruzaman@mmu.  
edu.my

**Received:** 12 Nov 2020

**Accepted:** 15 August 2021

© Copyright Kamaruzaman.  
This article is distributed  
under the terms of the  
[Creative Commons  
Attribution License](#), which  
permits unrestricted use  
and redistribution provided  
that the original author and  
source are credited.

copulas an attractive tool in multivariate statistical modelling. In order to construct the copula modelling, both marginal and copula parameter needs to be estimated. There are several technics to obtain parametric estimators and the most popular for classical statistics is the maximum likelihood estimation (MLE). However, for small sample sizes, MLEs do not always behave well and maybe substantially inferior to the estimators obtained by other methods. An alternative method to classical approach in estimating the copula parameter is Bayesian approach.

Bayesian parameter estimation is a universal method for fitting models to data. It can be more reliable for studies with small sample sizes, while it will also lead to the same conclusions as MLE when the sample size approaches infinity. The advantage of a Bayesian approach compared to MLE is the use of whole posterior distribution instead of one representative of a family, giving more robust processes and completely exploiting the benefits of integration versus optimization. Bayesian analysis provides better results than the classical when there exists sufficient amount of prior information which should not be ignored and sometimes it is still performing better even with very weak prior information is used [10]. Importantly, when Bayesian analyses have been shown to outperform classical analyses, informative priors are frequently derived from subject-specific expertise or directly from previous studies [11]. However because of specific prior information is not always available and very limited, the main objective is to discover prior distributions that are not only informative but also general enough to apply to many situations. This class of prior distributions is classified as weakly informative prior [12,13]. Specification of a proper prior distribution for a parameter is the most important aspect of a Bayesian analysis that distinguishes it from classical analysis. The prior distribution also plays an important role in small sample sizes and in this case a careful study is necessary [11].

Another problem with Bayesian is the posterior distribution rarely has closed form. As copula functions and densities have usually nonstandard forms, copula-based bivariate distributions lead in most cases to analytically intractable posterior distributions. Therefore, customized Markov chain Monte Carlo (MCMC) scheme are necessary [14] in order to obtain a sample from the joint posterior distribution which allows the development of Bayesian inference. MCMC enables many researchers to solve a wide range of Bayesian problems which cannot be solved using analytical methods.

The use of Bayesian method is widely known for its effectiveness in obtaining inferences based on likelihood. Some of researchers who have successfully applied the Bayesian approach to copula modelling are such as Huard *et al.* [15] who proposed a method for selecting the best bivariate copula, Hoff [16], Silva and Lopes [17], Craiu and Craiu [18] and Atique and Attoh-Okine [19] who have used Monte Carlo Markov Chain (MCMC) simulation to estimate copula parameter and Smith [20] has discussed Bayesian correlations for discrete and continuous data. Other important studies such as Wu *et al.* [21] and Grazian and Liseo [22] applied Bayesian copula approach in non-parametric context.

Although the impact of copula modelling in multivariate analysis has been significant in many fields. Yet, Bayesian inferential approaches have been employed by only a few empirical analysts to date. Nevertheless, they show great potential for computing competent likelihood-based inference in a number of contexts.

In this study, the authors derived bivariate rainfall distribution between rainfall severity and duration using the Bayesian copula method. Prior to copula fitting, the Standardized Precipitation Index (SPI), developed by Mckee *et al.* [23], was employed to defined the rainfall characteristics. Each flood event was characterized by firstly fitted rainfall duration and severity, separately using probability distributions, namely gamma, log normal, weibull, exponential and log logistic. Later on, univariate marginal distributions were linked by Frank Copula to create the joint distribution of rainfall duration and severity.

## Study area and data

This study was focused on the area of Peninsular Malaysia located in the Northern latitude zone between 1 and 6° N and the Eastern longitude zone from 100 to 103° E. This area is experiencing a warm and

humid climate throughout the year. The level of temperatures and rainfall is strongly influenced by Southwest Monsoon Wind, blowing from May to September, and Northeast Monsoon Wind that blows from November to March. The transitional period between the two monsoon events is known as the intermonsoon period which brings constant rainfalls to almost all areas of the peninsula. The annual rainfall is eventually tabulated to be 80% per year, ranging from 2000mm to 2500mm.

As to make reliable statistical modelling, the authors have taken into account and reviewed 51 years records of data during the years 1965-2015. These data involved 48 rainfall stations and they have been obtained with collaboration with the Department of Irrigation and Drainage Malaysia (DID). All selected rainfall stations were located at flood-prone areas in Peninsular Malaysia [24].

## Methodology

In this section, the Bayesian inferential scheme, selection of prior distribution and MCMC simulation techniques are discussed, followed by the description of the function, density and Kendall Tau dependence measure for Frank Copula. This copula was selected because of its capability in generalizing the best copula for flood analysis study in Peninsular Malaysia [25].

### Bayesian parameter estimation

Bayesian parameter estimation involves inferring the posterior probability density function (PDF) over model parameters  $\theta$  given the data  $D$  for some model  $M$ . The posterior PDF is given by Bayes' theorem

$$P(\theta|D, M) = \frac{P(D|\theta, M)P(\theta|M)}{P(D|M)}$$

where the terms in the numerator are called the likelihood,  $P(D|\theta, M)$ , and the prior,  $P(\theta|M)$ , and the denominator is the evidence. As the evidence does not depend on the model parameters, it can consider as a normalization constant for the posterior. In that case, the unnormalized posterior is just the product of the likelihood and prior.

### Bayesian copula inference

Let  $(X_1, X_2)$  be a bivariate random variable with joint probability function given by

$$f(x_1, x_2|\Psi) = c(F_1(x_1|\Psi), F_2(x_2|\Psi))f_1(x_1|\Psi)f_2(x_2|\Psi) \tag{2}$$

where  $\Psi$  is the parameter vector comprising copula and marginal distribution parameters and  $f_i$  and  $F_j$ ,  $j = 1, 2$ , represent the probability and cumulative marginal distribution functions, respectively.  $C$  is a cumulative distribution function and  $c$  is a copula density, that is the probability density function calculated as the mixed derivative in  $x_1$  and  $x_2$ .

Now, let  $\mathbf{x} = ((x_{11}, x_{21}), \dots, (x_{1n}, x_{2n}))$  be a sample of size  $n$  from independent and identically distributed data from the probability function in Equation 2, then the likelihood function is given by

$$L(\mathbf{x}|\Psi) = \prod_{i=1}^n c(F_1(x_{1i}|\Psi), F_2(x_{2i}|\Psi))f_1(x_{1i}|\Psi)f_2(x_{2i}|\Psi) \tag{3}$$

leading to the posterior distribution  $g(\Psi|x) \propto L(x|\Psi)g(\Psi)$  for prior distribution  $g(\Psi)$ .

### Prior distribution

The type of prior used in this study was the weak informative prior (WIP). WIP distributions use prior information for regularization and stabilization. It aims to prevent a problem such as overfitting by

introducing some additional information. The use of this type of prior will perform more logical and better estimates than the maximum likelihood, which is still the standard approach in theoretical and applied statistics.

**Markov chain Monte Carlo (MCMC)**

It is impossible to obtain a direct posterior distribution due to its intractability form. Therefore, the authors characterize it by sampling, from which the best model parameters and their uncertainties can be found. The authors will use the Gibbs sampling technique for marginal parameters estimation and random walk Metropolis algorithm (RWMA) for copula parameter estimation.

**Random walk metropolis algorithm (RWMA)**

There are many choices of distribution which can be used as symmetric proposals, but normal distribution with varying  $\sigma$  usually provides sufficient flexibility. An important calibration or scaling issue is the choice of parameter  $\sigma$  [26], which has a great impact on the effectiveness of the RWMA algorithm. Too large  $\sigma$  can bring about excessive overdispersion while too small  $\sigma$  leads to a chain with a very high acceptance rate, exhibiting a substantial bias.

The RWMA method has the following algorithm. Suppose the current state of the Markov chain is  $X_n$  and assume the density of the equilibrium distribution as  $h$ . Then the next Markov chain steps are simulated as follows:

- Simulate  $Y_n$  with a distribution of  $\mathcal{N}(\mu, \sigma)$ .
- Calculate  $r = \frac{h(X_n+Y_n)}{h(X_n)}$
- Simulate  $U_n$  with a distribution of  $\mathcal{U}(0,1)$ .
- If  $U_n < r$ , then set  $X_{n+1} = X_n + Y_n$ .
- Otherwise set  $X_{n+1} = X_n$ .

There are no firm rules dictating how to choose the scaling parameter. Therefore, it can be adjusted to achieve a reasonable acceptance rate. Roberts et al. [27] recommend acceptance rates close to 0.25 for high-dimensional models and 0.5 for one or two-dimensional models. A high acceptance rate may indicate a weak convergence pattern.

**Gibbs sampling**

Gibbs sampling acts as an alternative approach to multi-parameter problems by sampling one parameter given the value of another parameter [28]. Gibbs sampling which was first used by Geman and Geman [29] can reduce the problem related to dimensionality. The general steps for Gibbs sampling are as follows:

Suppose  $f(\theta, \phi) \propto \pi(\theta, \phi|x)$  is the target density with data  $x$  and two parameters  $\theta$  and  $\phi$ . Let  $f_1(\theta)$  and  $f_2(\phi)$  be two marginal densities and

$$g_1(\theta|\phi) = \frac{f(\theta, \phi)}{f_2(\phi)}, g_2(\phi|\theta) = \frac{f(\theta, \phi)}{f_1(\theta)} \tag{4}$$

are two conditional densities. If direct sampling from conditional is feasible, the algorithm is as follows:

For  $t = 0$  choose  $\theta_0$  arbitrarily. Then draw  $\phi_0$  from  $g_2(\phi|\theta)$ .

For  $t = 1, 2, \dots$

- Generate a random value  $\theta_t \sim g_1(\theta|\phi_{t-1})$ ;
- Generate a random value  $\phi_t \sim g_2(\phi|\theta_t)$ ;
- Use pairs  $(\theta_t, \phi_t)$  as a sample from  $f(\theta, \phi)$ .

**Frank Copula**

Bivariate Frank Copula function can be defined as:

$$C_{\theta}(u, v) = -\theta^{-1} \log \left( 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \tag{5}$$

and its density function:

$$c_{\theta}(u, v) = \frac{\partial^2 C_{\theta}(u, v)}{\partial u \partial v} \tag{6}$$

$$= \frac{e^{-\theta u}(-e^{-\theta v - \theta} + e^{-\theta v} \theta)}{(e^{-\theta u - \theta v} - e^{-\theta u} - e^{-\theta v} + e^{-\theta})^2} \tag{7}$$

The value of theta for Frank Copula is closely related to Kendall's tau value, which is the strength of dependence between the component of two variables as in the following equation:

$$\tau = 1 + \frac{4(D(\theta) - 1)}{\theta} \tag{8}$$

through Debye's integral

$$D(\theta) = \frac{1}{\theta} \int_0^{\theta} \frac{t}{e^t - 1} dt \tag{9}$$

**Results and discussion**

This study used the same marginal distribution in the previous research [25] for the severity and duration variables. Next, these two marginal distributions were combined to become a bivariate distribution using Frank Copula. Copula allows the marginal parameter to be estimated first and then the estimation of the copula parameter is performed given the marginal parameter value.

**Marginal parameter estimation**

In Bayesian marginal estimation, the conjugate prior is used for easy analytical computation. For Kuala Brang station for example, the severity and duration variables were modelled using log normal and gamma distribution respectively with the densities of the two distributions as follows:

Log-Normal Distribution

$$f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad \mu, \sigma > 0 \tag{10}$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation of  $\ln(x)$  respectively.

Gamma Distribution

$$f(x|\alpha, \gamma) = \frac{\gamma^{\alpha}}{\Gamma(\alpha)} e^{-\gamma x} x^{\alpha-1}, \quad \alpha, \gamma > 0 \tag{11}$$

where  $\alpha$  is the shape parameter and  $\gamma$  is the rate parameter.

For the log normal parameter estimation, normal prior is assumed for the mean,  $\mu$  and gamma

prior for the precision parameter,  $\tau$  or  $\left(\frac{1}{\sigma}\right)^2$ . While for the gamma parameter estimation, exponential prior is assumed for the shape parameter,  $\alpha$  and gamma prior for the rate parameter,  $\gamma$ . The posterior distribution was generated by using the Gibbs sampling technique. The results obtained for the Kuala Brang station are shown in the following diagram. Figure 1 and Figure 2 show the trace plots of the four parameters: mean, precision, shape and rate generated by the MCMC algorithm, whereas Figure 3 and Figure 4 show a histogram of the posterior mean for the log normal and gamma distribution.

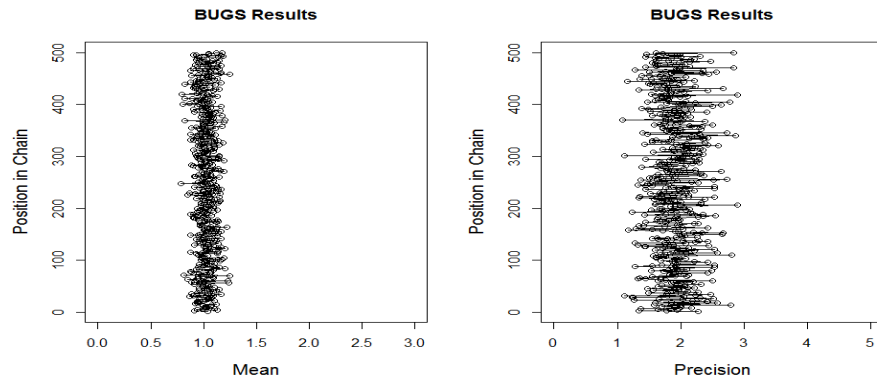


Figure 1. Trace plot for log normal distribution

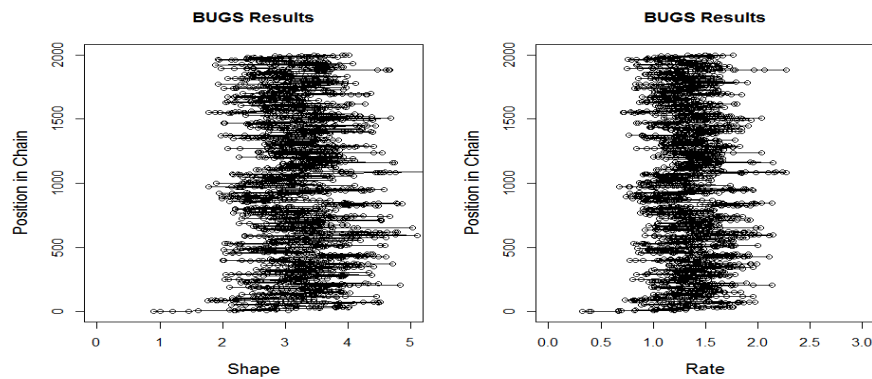


Figure 2. Trace plot for gamma distribution

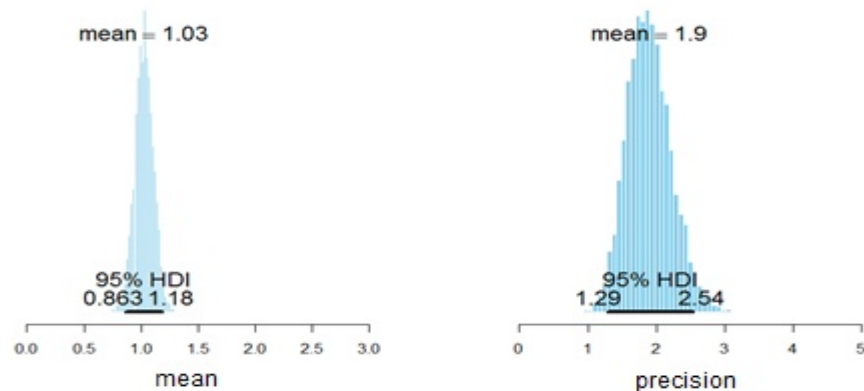


Figure 3. Histograms of the posterior distributions for log normal

**Table 1.** Marginal parameter estimation for severity and duration

Station	Severity			Duration		
	Distribution	Parameter 1	Parameter 2	Distribution	Parameter 1	Parameter 2
Meranti	Log Logistic	$\alpha = 2.16$	$\beta = 2.52$	Weibull	$\alpha = 1.50$	$\lambda = 0.23$
Kuala Jambu	Log Logistic	$\alpha = 2.18$	$\beta = 2.45$	Weibull	$\alpha = 1.52$	$\lambda = 0.24$
Stesen Keretapi Tumpat	Log Normal	$\mu = 1.01$	$\tau = 1.82$	Weibull	$\alpha = 1.62$	$\lambda = 0.22$
Kampung Ibok	Log Logistic	$\alpha = 2.15$	$\beta = 2.22$	Log Logistic	$\alpha = 2.76$	$\beta = 1.59$
Dungun	Weibull	$\alpha = 1.25$	$\lambda = 0.17$	Weibull	$\alpha = 1.50$	$\lambda = 0.23$
Kuala Brang	Log Normal	$\mu = 1.03$	$\tau = 1.90$	Gamma	$\alpha = 3.19$	$\gamma = 1.34$
Kuala Telemong	Weibull	$\alpha = 1.40$	$\lambda = 0.18$	Weibull	$\alpha = 1.81$	$\lambda = 0.24$
Marang	Weibull	$\alpha = 1.39$	$\lambda = 0.16$	Weibull	$\alpha = 1.77$	$\lambda = 0.22$
Kuala Terengganu	Log Logistic	$\alpha = 2.06$	$\beta = 2.55$	Weibull	$\alpha = 1.75$	$\lambda = 0.23$
Kampung Rahmat	Weibull	$\alpha = 1.19$	$\lambda = 0.18$	Weibull	$\alpha = 1.62$	$\lambda = 0.23$
Banggol	Exponential	$\gamma = 0.24$	-	Weibull	$\alpha = 1.45$	$\lambda = 0.23$
Setiu	Log Normal	$\mu = 0.96$	$\tau = 1.90$	Weibull	$\alpha = 1.67$	$\lambda = 0.25$
Pelanggi Kampung Jawi 2	Log Logistic	$\alpha = 2.21$	$\beta = 2.45$	Weibull	$\alpha = 1.51$	$\lambda = 0.23$
Bentong	Log Logistic	$\alpha = 2.15$	$\beta = 2.50$	Log Logistic	$\alpha = 2.60$	$\beta = 1.74$
Paya Membang	Weibull	$\alpha = 1.42$	$\lambda = 0.17$	Weibull	$\alpha = 1.79$	$\lambda = 0.23$
Kampung Serambi	Gamma	$\alpha = 1.96$	$\gamma = 0.51$	Weibull	$\alpha = 1.69$	$\lambda = 0.21$
Kerdau	Weibull	$\alpha = 1.34$	$\lambda = 0.17$	Weibull	$\alpha = 1.64$	$\lambda = 0.22$
Sanggang	Weibull	$\alpha = 1.46$	$\lambda = 0.16$	Weibull	$\alpha = 1.72$	$\lambda = 0.23$
Pekan	Log Logistic	$\alpha = 2.22$	$\beta = 2.33$	Weibull	$\alpha = 1.73$	$\lambda = 0.26$
Penor	Log Logistic	$\alpha = 2.19$	$\beta = 2.35$	Weibull	$\alpha = 1.63$	$\lambda = 0.26$
Kuala Krau	Weibull	$\alpha = 1.30$	$\lambda = 0.17$	Weibull	$\alpha = 1.61$	$\lambda = 0.24$
Paya Kangsar	Weibull	$\alpha = 1.18$	$\lambda = 0.19$	Weibull	$\alpha = 1.53$	$\lambda = 0.24$
Ladang Kuala Reman	Log Logistic	$\alpha = 2.20$	$\beta = 2.37$	Weibull	$\alpha = 1.40$	$\lambda = 0.26$
Kuala Lipis	Log Logistic	$\alpha = 2.06$	$\beta = 2.61$	Weibull	$\alpha = 1.47$	$\lambda = 0.22$
Kota Tinggi	Log Normal	$\mu = 0.91$	$\tau = 2.14$	Weibull	$\alpha = 1.78$	$\lambda = 0.24$
Sembrong	Log Logistic	$\alpha = 2.33$	$\beta = 2.57$	Weibull	$\alpha = 1.80$	$\lambda = 0.21$
Ladang Lambak	Log Logistic	$\alpha = 2.31$	$\beta = 2.13$	Log Logistic	$\alpha = 2.82$	$\beta = 1.51$
Yong Peng	Weibull	$\alpha = 1.46$	$\lambda = 0.18$	Weibull	$\alpha = 1.83$	$\lambda = 0.22$
Ladang Ulu Paloh	Log Logistic	$\alpha = 2.21$	$\beta = 2.24$	Weibull	$\alpha = 1.60$	$\lambda = 0.26$
Jementah	Log Normal	$\mu = 1.04$	$\tau = 1.67$	Gamma	$\alpha = 2.58$	$\gamma = 1.02$
Segamat	Weibull	$\alpha = 1.37$	$\lambda = 0.15$	Weibull	$\alpha = 1.84$	$\lambda = 0.20$
Empangan Labong	Log Logistic	$\alpha = 2.19$	$\beta = 2.43$	Weibull	$\alpha = 1.56$	$\lambda = 0.24$
Pusat Pertanian Endau	Log Logistic	$\alpha = 2.24$	$\beta = 2.17$	Weibull	$\alpha = 1.60$	$\lambda = 0.29$
Stor Jps Endau	Log Normal	$\mu = 0.97$	$\tau = 1.86$	Weibull	$\alpha = 1.60$	$\lambda = 0.22$
Parit Nibong	Log Logistic	$\alpha = 2.31$	$\beta = 2.22$	Weibull	$\alpha = 1.57$	$\lambda = 0.28$
Rantau Panjang	Log Normal	$\mu = 1.02$	$\tau = 1.44$	Log Logistic	$\alpha = 2.49$	$\beta = 1.86$
Jeniang	Weibull	$\alpha = 1.45$	$\lambda = 0.17$	Weibull	$\alpha = 1.82$	$\lambda = 0.22$
Telok Rimba	Log Logistic	$\alpha = 2.30$	$\beta = 2.26$	Weibull	$\alpha = 1.59$	$\lambda = 0.26$
Jasin	Log Normal	$\mu = 0.87$	$\tau = 2.00$	Gamma	$\alpha = 2.93$	$\gamma = 1.48$
Jalan Empat	Log Logistic	$\alpha = 2.07$	$\beta = 2.53$	Log Logistic	$\alpha = 2.58$	$\beta = 1.85$
Ladang Bukit Bertam	Gamma	$\alpha = 2.17$	$\gamma = 0.62$	Weibull	$\alpha = 1.66$	$\lambda = 0.20$
Batu Kurau	Log Logistic	$\alpha = 2.48$	$\beta = 2.15$	Weibull	$\alpha = 1.78$	$\lambda = 0.27$
Ladang Sepang	Log Normal	$\mu = 0.87$	$\tau = 2.58$	Weibull	$\alpha = 2.01$	$\lambda = 0.23$
Sungai Mangg	Log Logistic	$\alpha = 2.39$	$\beta = 2.55$	Weibull	$\alpha = 1.67$	$\lambda = 0.21$
Ladang Bukit Kerayong	Log Logistic	$\alpha = 2.07$	$\beta = 2.55$	Weibull	$\alpha = 1.42$	$\lambda = 0.22$
Ladang Tuan Mee	Weibull	$\alpha = 1.39$	$\lambda = 0.15$	Weibull	$\alpha = 1.75$	$\lambda = 0.20$
Tanjung Karang	Log Normal	$\mu = 0.97$	$\tau = 1.68$	Weibull	$\alpha = 1.57$	$\lambda = 0.23$
Sungai Bernam	Log Normal	$\mu = 0.88$	$\tau = 2.19$	Weibull	$\alpha = 1.84$	$\lambda = 0.24$

N

Note:  $\alpha$  = shape;  $\beta$  = scalar;  $\lambda$  = lambda;  $\mu$  = mean;  $\tau$  = precision;  $\gamma$  = rate parameter

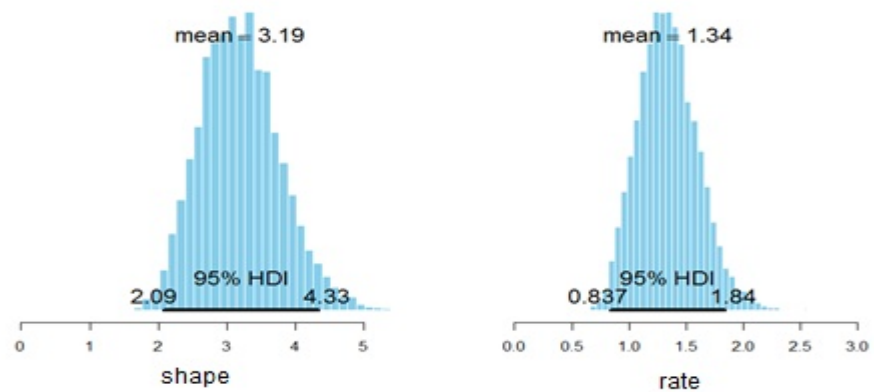


Figure 4. Histograms of the posterior distributions for gamma

Based on the trace trajectories in Figure 1 and Figure 2, it shows a good sampling property, which means good mixing. The complete results of marginal parameter estimation for all 48 stations for severity and duration can be referenced in Table 1.

**Frank Copula parameter estimation**

For estimating the Frank Copula parameter, gamma prior is assumed for theta,  $\theta$ . The copula parameter is estimated given the marginal parameter value using RWMA technique. When dealing with RWMA, the standard deviation parameter value should be tuned carefully in order to gain a reasonable acceptance rate. The results of the estimated parameters and their acceptance rates for all stations are shown in Table 2. In general, all stations have a decent acceptance rate, ranging from 0.46 to 0.54 percent.

**Performance of Bayesian approach**

This section describes the effectiveness of the Bayesian method in estimating the parameters of a distribution. When the studies involved large sample sizes, there were no significant differences were noted between Bayesian and frequentist. However, the Bayesian method can be more reliable for studies with small sample sizes. To test the performance of this method with one obtained from frequentist method [25], comparisons were made between simulation data from Bayesian approach and the frequentist approach towards the observation data using mean absolute error (MAE) goodness of fit. The results obtained are shown in Figure 5.

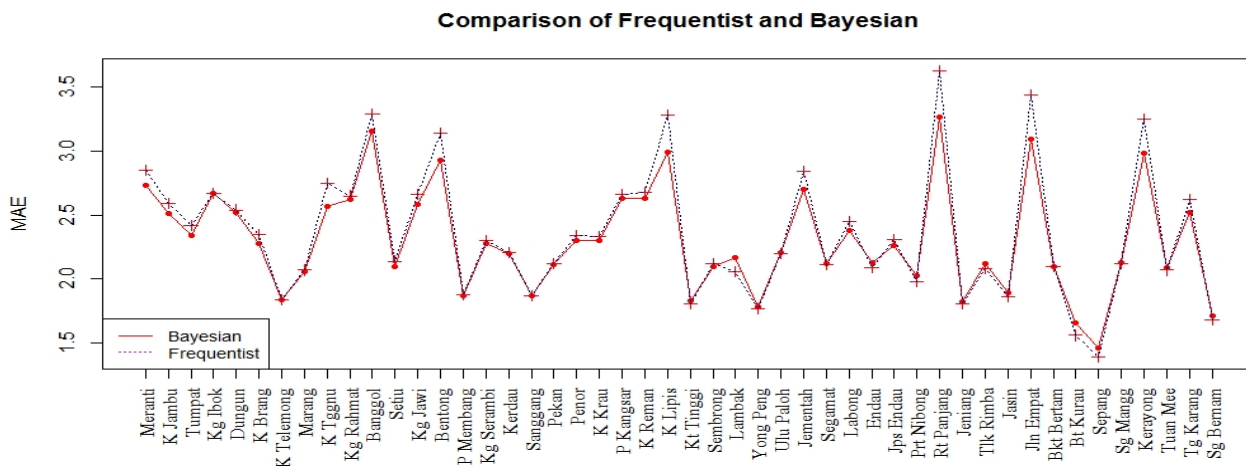


Figure 5. Comparison of mean absolute error (MAE) between Frequentist and Bayesian approach



**Table 2.** Frank Copula parameter estimation values

<b>Station</b>	<b>Theta</b>	<b>Acceptance Rate</b>
Meranti	19.76	0.48
Kuala Jambu	19.12	0.49
Stesen Keretapi Tumpat	19.17	0.47
Kampung Ibok	22.96	0.54
Dungun	23.16	0.48
Kuala Brang	18.84	0.47
Kuala Telemong	23.32	0.50
Marang	20.64	0.49
Kuala Terengganu	21.99	0.49
Kampung Rahmat	21.64	0.51
Banggol	21.54	0.52
Setiu	15.37	0.47
Pelangi Kampung Jawi 2	21.95	0.53
Bentong	26.66	0.54
Paya Membang	21.71	0.47
Kampung Serambi	22.33	0.50
Kerdau	26.08	0.54
Sanggang	22.92	0.53
Pekan	19.43	0.50
Penor	21.29	0.49
Kuala Krau	24.49	0.52
Paya Kangsar	28.17	0.52
Ladang Kuala Reman	18.97	0.48
Kuala Lipis	20.63	0.52
Kota Tinggi	20.47	0.49
Sembrong	22.20	0.53
Ladang Lambak	18.29	0.49
Yong Peng	23.74	0.51
Ladang Ulu Paloh	21.17	0.47
Jementah	19.35	0.49
Segamat	22.33	0.54
Empangan Labong	21.19	0.52
Pusat Pertanian Endau	20.19	0.49
Stor Jps Endau	20.48	0.52
Parit Nibong	19.06	0.48
Rantau Panjang	20.64	0.54
Jeniang	25.84	0.48
Telok Rimba	22.72	0.54
Jasin	23.84	0.51
Jalan Empat	23.88	0.53
Ladang Bukit Bertam	20.25	0.49
Batu Kurau	22.00	0.48
Ladang Sepang	19.85	0.49
Sungai Mangg	21.81	0.52
Ladang Bukit Kerayong	21.32	0.53
Ladang Tuan Mee	17.88	0.50
Tanjung Karang	22.16	0.51
Sungai Bernam	20.24	0.46

Based on the graph in Figure 5, it is clear that the Bayesian parameter estimation gives better results. Lower MAE values were successfully recorded for almost all stations. Factors such as the type of prior use have succeeded to reduce the gap in data estimation errors.

## Conclusions

Two types of distributions, namely marginal distribution and copula are estimated using Bayesian analysis which is an alternative method for estimating the distribution parameters. The Bayesian parameter estimation was carried out in two stages using MCMC simulation based on Gibbs and RWMA sampling techniques. Bayesian inference through MCMC simulation is fair and suitable to be used for various sample sizes. In contrast to the frequentist analysis of [25], it is more likely to produce biased results due to the small sample size. In this situation, Bayesian analysis is able to provide better results because of prior distributions can dominate the likelihood. This was supported by the result in the goodness of fit tests that the Bayesian analysis has a smaller MAE value compare to frequentist.

Another advantage of using Bayesian is the use of entire posterior distribution instead of one representative of a family, providing more stable procedures and fully utilizing the benefits of integration versus optimization.

## Data availability

The daily precipitation data that has been used in this study was provided by Department of Irrigation and Drainage Malaysia (DID).

## Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

## Funding statement

The work is financed by MyBrain15 Scholarship provided by the Ministry of Higher Education of Malaysia

## References

- [1] Department of Statistics Malaysia. Compendium of Environment Statistics 2018. *Department of Statistics Malaysia* 2018, pp. 1–4.
- [2] Department of Irrigation and Drainage Malaysia. Flood Management - Programme and Activities: Available: <https://www.water.gov.my/index.php/pages/view/419?mid=244> (12 November 2017).
- [3] Wazneh, H., Arain, M. A., Coulibaly, P. & Gachon, P. Evaluating the Dependence between Temperature and Precipitation to Better Estimate the Risks of Concurrent Extreme Weather Events. *Advances in Meteorology* 2020, 1–16.
- [4] Zhang, L., Singh, V. P. Bivariate Rainfall Frequency Distributions Using Archimedean Copulas. *Journal of Hydrology* 2007, 332(1–2), 93–109.
- [5] Abdul Rauf, U. F., Zeephongsekul, P. Modelling Rainfall Severity and Duration in North-Eastern Victoria Using Copulas. *19th International Congress on Modelling and Simulation*, Perth, Australia, 12–16 December 2011, 3462–3468.
- [6] Daneshkhah, A., Remesan, R., Chatrabgoun, O., Holman, I. P. Probabilistic Modeling of Flood Characterizations with Parametric and Minimum Information Pair-Copula Model. *Journal of Hydrology* 2016, 540, 469–487.
- [7] Ozga-Zielinski, B., Ciupak, M., Adamowski, J., Khalil, B., Malard, J. Snow-Melt Flood Frequency Analysis by Means of Copula Based 2D Probability Distributions for The Narew River in Poland. *Journal of Hydrology: Regional Studies* 2016, 6, 26–51.
- [8] Nelsen, R. B. *An Introduction to Copulas*, 2nd ed.; Springer: New York, 2006; pp. 1-272, 9780387286785
- [9] Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de statistique de l'Université de Paris* 1959, 8, 229–231.

- [10] Shemyakin, A. & Kniyazev, A. *Introduction to Bayesian Estimation and Copula Models of Dependence*, First Edition. New Jersey, 2017: John Wiley & Sons, Inc.
- [11] McNeish, D. On Using Bayesian Methods to Address Small Sample Problems. *Structural Equation Modeling* 2016, 23(5): 750–773.
- [12] Gelman, A. Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). *Bayesian Analysis* 2006,1(3): 515–534.
- [13] Gelman, A., Simpson, D. & Betancourt, M. The prior can often only be understood in the context of the likelihood. *Entropy* 2017,19(10): 1–13.
- [14] Gamerman, D. & Lopes, H. F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd Edition. London, 2006: Chapman and Hall.
- [15] Huard, D., Évin, G. & Favre, A. C. Bayesian copula selection. *Computational Statistics and Data Analysis* 2006, 51(2): 809–822.
- [16] Hoff, P. D. Extending The Rank Likelihood for Semiparametric Copula Estimation. *The Annals of Applied Statistics* 2007, 1(1): 265–283.
- [17] Silva, R. d S. & Lopes, H. F. Copula, marginal distributions and model selection: A Bayesian note. *Statistics and Computing* 2008, 18(3): 313–320.
- [18] Craiu, M. & Craiu, R. V. Bayesian inference for copula models. *UPB Scientific Bulletin, Series A: Applied Mathematics and Physics* 2008, 70(3): 3–10.
- [19] Atique, F. & Attoh-Okine, N. Copula Parameter Estimation Using Bayesian Inference for Pipe Data Analysis. *Canadian Journal of Civil Engineering* 2018, 45(1): 61–70.
- [20] Smith, M. S. Bayesian Approaches to Copula Modelling. *Journal of the American Statistical Association* 2011 forthcoming.
- [21] Wu, J., Wang, X. & Walker, S. G. Bayesian Nonparametric Inference for a Multivariate Copula Function. *Methodology and Computing in Applied Probability* 2014, 16(3): 747–763.
- [22] Grazian, C. & Liseo, B. Approximate Bayesian inference in semiparametric copula models. *Bayesian Analysis* 2017, 12(4): 991–1016.
- [23] Mckee, T. B., Doesken, N. J., Kleist, J. The Relationship of Drought Frequency and Duration to Time Scales. *Eighth Conference on Applied Climatology, Anaheim, California, 17-22 January 1993*; 179–184.
- [24] Department of Irrigation and Drainage Malaysia. *Updating of Condition of Flooding in Malaysia. National Register of River Basins* 2003, Vol. 2, 1–209.
- [25] Kamaruzaman, I. F., Wan Zin, W. Z. & Mohd Ariff, N. A Generalized Bivariate Copula for Flood Analysis in Peninsular Malaysia. *Malaysian Journal of Fundamental and Applied Sciences* 2019, 15(1): 38–49.
- [26] Knyazev, A., Lepekhin, O. & Shemyakin, A. Joint distribution of stock indices: Methodological aspects of construction and selection of copula models. *Applied econometrics* 2016, 42: 30–53.
- [27] Roberts, G. O., Gelman, A. & Gilks, W. R. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability* 1997, 7(1): 110–120.
- [28] Tierney, L. Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics* 1994, 22(4): 1701–1762.
- [29] Geman, S. & Geman, D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984, PAMI-6(6): 721–741.