ARQII Publication
APPLICATIONS OF
MODELLING AND SIMULATION
http://arqiipubl.com/ams eISSN 2600-8084 VOL 9, 2025, 363-373

Evaluation of the Kagebunshin-Beta Distribution for Modelling Daily Rainfall in Indonesia

Mohamad Khoirun Najib*

Division of Computational Mathematics, School of Data Science, Mathematics, and Informatics, IPB University, 16680 Dramaga, Bogor Regency, Indonesia

*Corresponding author: mkhoirun@apps.ipb.ac.id

Submitted 06 August 2025; Revised 18 September 2025; Accepted 20 September 2025; Available online 25 September 2025.

Abstract: Accurately modelling daily rainfall data is essential for climate studies, hydrological modelling, and disaster risk management. Traditional probability distributions often struggle with the inherent complexities of rainfall data, such as skewness, excess zeros, and high variability. This study explores the potential of the Kagebunshin-Beta (KB) distribution as a novel approach to modelling daily precipitation in Indonesia. The KB distribution, derived from a transformation of beta distribution, offers flexibility in capturing diverse rainfall patterns. This research investigates its statistical properties, parameter estimation via the Nelder-Mead optimization algorithm, and performance compared to 17 well-known distributions, including gamma, Weibull, lognormal, and generalized extreme value distributions. Using two datasets (rainfall data from five major cities and a spatial dataset covering Indonesia), the KB distribution consistently outperformed other models based on the Akaike Information Criterion (AIC), Anderson-Darling, and Cramer-von Mises goodness-of-fit tests. The KB distribution was further applied to estimate the probability of dry and heavy rainy days across Indonesia, identifying regions prone to prolonged droughts and extreme rainfall events. These findings demonstrate the KB distribution's effectiveness as an alternative for rainfall modelling and its potential applications in bias correction, risk assessment, and climate adaptation strategies. By providing a more robust statistical representation of precipitation patterns, this study contributes to improved water resource management, disaster preparedness, and climate resilience.

Keywords: Bias correction; Climate risk assessment; Daily rainfall modelling; Extreme precipitation; Kagebunshin-Beta distribution.

1. INTRODUCTION

Rainfall, a fundamental component of Earth's climate system, plays a crucial role in various aspects of human life and environmental processes. It directly influences soil water movement, affecting groundwater recharge, soil moisture levels, and erosion dynamics, which are essential for maintaining ecosystem balance and agricultural productivity [1]. In the agricultural sector, rainfall is a key determinant of crop production, as it regulates water availability for plant growth, impacts yield variability, and influences food security [2]. Together with air temperature, rainfall patterns contribute to defining climate by shaping regional and global climatic conditions, influencing humidity levels, and atmospheric circulation [3], [4]. Furthermore, extreme rainfall events are closely linked to natural disasters such as floods, landslides, and droughts [5], [6], [7], which pose significant risks to human settlements, infrastructure, and ecosystems. Given its widespread impact, understanding rainfall variability and distribution is essential for effective water resource management, disaster preparedness, and climate adaptation strategies.

An effective approach to understanding and analyzing rainfall is using a probabilistic framework that incorporates both parametric and non-parametric distribution functions. Many studies utilize probabilistic statistical methods in their analyses, such as statistical bias correction [8], [9], risk assessment [10], return period analysis [11], and the impact of climate change [12]. However, the statistical analysis of daily rainfall data presents significant challenges due to its inherent characteristics, such as high variability [13], skewed distributions [14], and the presence of excess zero values, commonly referred to as zero-inflated or zero-excess data [15]. These characteristics lead to statistical complexities, making traditional parametric models inadequate for accurate representation [16]. The inability of conventional distributions to effectively handle zero-inflated rainfall data has substantial implications for hydrological modelling, climate change assessment, and risk analysis in water-related sectors.

In the field of statistical modelling, several probability distributions have been proposed to represent rainfall data. The most used include the gamma [17], Weibull [18], and lognormal [19] distributions, which have been widely applied in climatological and hydrological studies. However, these models have limitations in capturing zero-inflated patterns, leading to biased estimates and unreliable statistical inferences, especially for daily rainfall data. Ye [20] emphasized the importance of selecting appropriate probability distributions for analyzing daily rainfall intensity in the context of frequency analysis and climate trend assessment. Their research highlighted the limitations of standard models, advocating for more sophisticated parametric and non-parametric methods. Similarly, Gupta [21] introduced the exponentiated exponential distribution as an alternative to gamma and Weibull distributions, demonstrating its enhanced capability in handling skewed datasets. More recently, Ribeiro-Reis [22] introduced and demonstrated the superiority of the Kagebunshin-Beta (KB) distribution over traditional distributions by applying it to real-world datasets. The findings indicated that the KB distribution provided a more accurate representation of the underlying data structure, highlighting its potential as a valuable alternative for modelling lifetime data and reliability studies.

Given the demonstrated effectiveness of the KB distribution in modelling real-world datasets Ribeiro-Reis [22], an intriguing question arises regarding its applicability in climatological studies, particularly for daily rainfall modelling. The complexities associated with rainfall pose significant challenges for conventional probability distributions. While previous research has explored alternative models like the exponentiated Gumbel distributions [23], the log-skew-normal alpha-power [24], the transmuted Half-Normal distribution [25], the Gompertz-general extreme value (Go-GEV) [26], and the alpha power transformed XLindley (APTXL) [27], this new distribution called KB distribution remains relatively unexplored in this domain. Its theoretical advantages, including flexibility in capturing a wide range of data distributions, suggest that it could provide improved statistical representations of daily rainfall patterns.

This study aims to explore the potential of the KB distribution for modelling daily rainfall data. By examining both its theoretical properties and applied performance, this research evaluates the suitability of the KB distribution in representing daily rainfall patterns in Indonesia. Section 2 presents several properties of the KB distribution, including its probability density function, cumulative distribution function, survival function, hazard function, quantile function, and random generator. Section 3 proposes parameter estimation for the KB distribution using the Nelder-Mead simplex algorithm. In Sections 4 and 5, the applied performance of the KB distribution is assessed using daily rainfall data from Indonesia and compared with well-known distributions. Performance is evaluated using various statistical tests, such as the Akaike Information Criterion (AIC), the Anderson-Darling test, and the Cramer-von Mises test. Section 6 discusses the application of the KB distribution in analyzing the probability of dry days and extreme rainfall events in Indonesia. Furthermore, it explores potential developments of the KB distribution in climatological analysis.

While previous studies have commonly employed distributions such as Gamma, Weibull, Lognormal, and GEV to model rainfall data, these models struggle to represent the zero-inflated and highly variable nature of daily rainfall. Recent alternative models (e.g., Exponentiated Gumbel [23], Log-Skew-Normal Alpha-Power [24], Transmuted Half-Normal [25], Go-GEV [26], and APTXL [27]) improved performance but still exhibit limitations in fitting extreme events and dry-day occurrences. In contrast, this study introduces the KB distribution [22] for the first time in the context of climatology. Our contribution lies in: (i) formulating its statistical properties and parameter estimation approach, (ii) rigorously comparing its performance with 17 well-known distributions across two large rainfall datasets in Indonesia, and (iii) demonstrating its ability to map the spatial probability of dry and heavy rainy days. These contributions provide a new, more flexible probabilistic tool for rainfall modelling that addresses gaps left by existing approaches

2. KAGEBUNSHIN-BETA DISTRIBUTION

The KB distribution is a newly introduced two-parameter lifetime distribution derived through a logarithmic transformation of the Beta distribution. This transformation allows the KB distribution to inherit the flexibility of the Beta family while extending its support from (0, 1) to (0, ∞), which is suitable for non-negative variables such as daily rainfall. Unlike conventional distributions (e.g., Gamma, Weibull, Lognormal) that often fail to represent zero-inflated or highly skewed precipitation data, the KB distribution can accommodate heavy tails, high variance, and an accumulation of near-zero values. These characteristics make the KB distribution particularly promising for modelling daily rainfall data, which typically exhibit these complex features.

2.1 General Properties

The KB distribution is obtained from a transformation of a random variable with a beta distribution [22]. A random variable Y with a beta distribution, has a cumulative distribution function (cdf) and a probability density function (pdf) given by

$$H(y|a,b) = \frac{1}{B(a,b)}\int_0^y z^{a-1}(1-z)^{b-1}\,dz, \qquad y \in (0,1) \tag{1}$$

and

$$h(y|a,b) = \frac{1}{B(a,b)}y^{a-1}(1-y)^{b-1}, \qquad y \in (0,1) \tag{2}$$

respectively, where $B(a,b)$ represents the beta function, i.e., $B(a,b) = \int_0^1 z^{a-1}(1-z)^{b-1}\,dz$ and $a > 0$ and $b > 0$ are shape parameters.

With the transformation of $X = -\log Y$, the cdf and pdf of X are given by

$$F(y|a,b) = 1 - \frac{1}{B(a,b)}\int_0^{e^{-x}} z^{a-1}(1-z)^{b-1}\,dz, \qquad x \in (0,\infty) \tag{3}$$

and

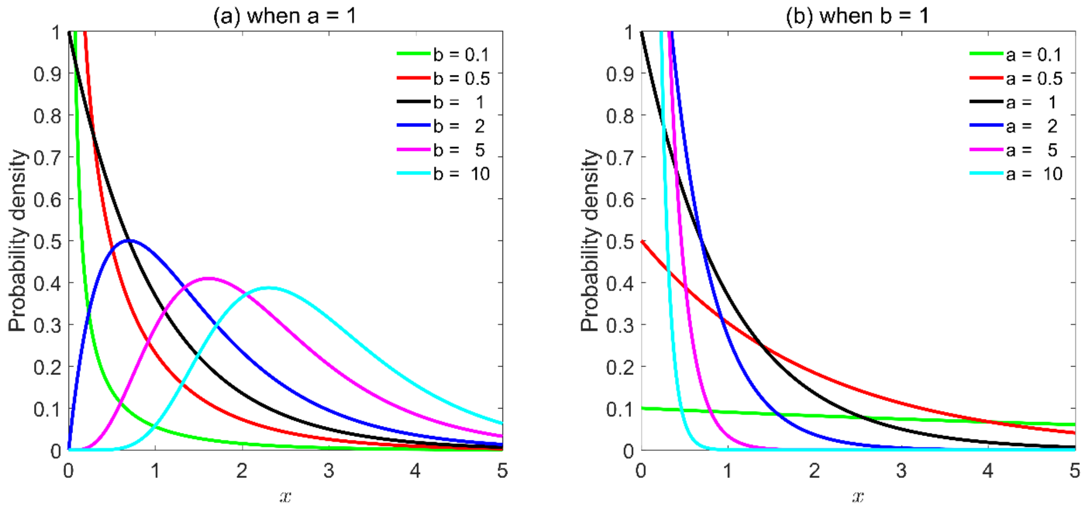

Figure 1. Some of the pdfs of the KB distribution.

$$f(y|a,b) = \frac{1}{B(a,b)} e^{-ax}(1-e^{-x})^{b-1}, \qquad x \in (0,\infty) \tag{4}$$

respectively. Here, a is the scale parameter and b is the shape parameter. The random variable X with pdf $f(x)$ is said to have a KB distribution, denoted by $X \sim KB(a,b)$. Some plots of the probability density function of random variables with kagebunshin-beta distribution for some parameter values are given in Figure 1.

Figure 1(a) shows the pdfs of the KB distribution when parameter a is set equal to 1 with several variations of parameter b. When $b \leq 1$, the peak of pdfs (or pdf's modes) is zero. Meanwhile, for $b > 1$, the pdf mode is given by

$$\text{mode}(X) = -\log\left(\frac{a}{a+b-1}\right), \qquad \text{for } b > 1 \tag{5}$$

Furthermore, Figure 1(b) shows the pdfs of the KB distribution when parameter b is set equal to 1. When b equal to 1, the pdf of KB distribution (Equation (4)) becomes

$$f(x|a,1) = ae^{-ax}, \qquad x \in (0,\infty) \tag{6}$$

which corresponds to the pdf of the exponential distribution. Therefore, the KB distribution has the exponential distribution as special case. When $b = 1$, the higher the value of parameter a the higher the peak of the pdf at zero.

In practical terms, applying the KB distribution for rainfall modelling involves a systematic procedure aimed at capturing the probabilistic behavior of daily precipitation. The first step is to fit the two parameters (a,b) of the KB distribution to the observed rainfall dataset using the Maximum Likelihood Estimation (MLE) method. This fitting process ensures that the theoretical probability density function (pdf) closely aligns with the empirical distribution of the rainfall data, especially in terms of skewness, variance, and the occurrence of near-zero values. Once the model parameters are estimated, the fitted KB distribution can be used for several climatological applications: (i) estimating the probability of rainfall exceeding or falling below specific thresholds (e.g., dry days <1 mm, heavy rainfall >50 mm), (ii) generating synthetic rainfall series through random sampling from the fitted distribution to support stochastic weather generators or climate risk models, and (iii) mapping spatial rainfall risk by fitting the KB model to multiple grid locations and visualizing the spatial variation of event probabilities across regions. This modelling approach is consistent with standard probabilistic frameworks widely used in climatology, yet the KB distribution offers a distinct advantage in its flexibility to represent highly skewed and zero-inflated rainfall data; features that often challenge conventional distributions such as Gamma or Lognormal.

2.2 Survival and Hazard Functions

Here, we focus on discussing additional properties of the KB distribution, i.e., its survival and hazard functions. It is difficult to formulate survival and hazard functions in closed form. According to cdf and pdf of the KB distribution (Equations (3) and (4)), the corresponding survival function is

$$S(y|a,b) = \frac{1}{B(a,b)} \int_0^{e^{-x}} z^{a-1}(1-z)^{b-1}\,dz, \qquad x \in (0,\infty) \tag{7}$$

and the hazard function is

$$h(y|a,b) = \frac{e^{-ax}(1-e^{-x})^{b-1}}{\int_0^{e^{-x}} z^{a-1}(1-z)^{b-1}}dz, \qquad x \in (0,\infty) \tag{8}$$

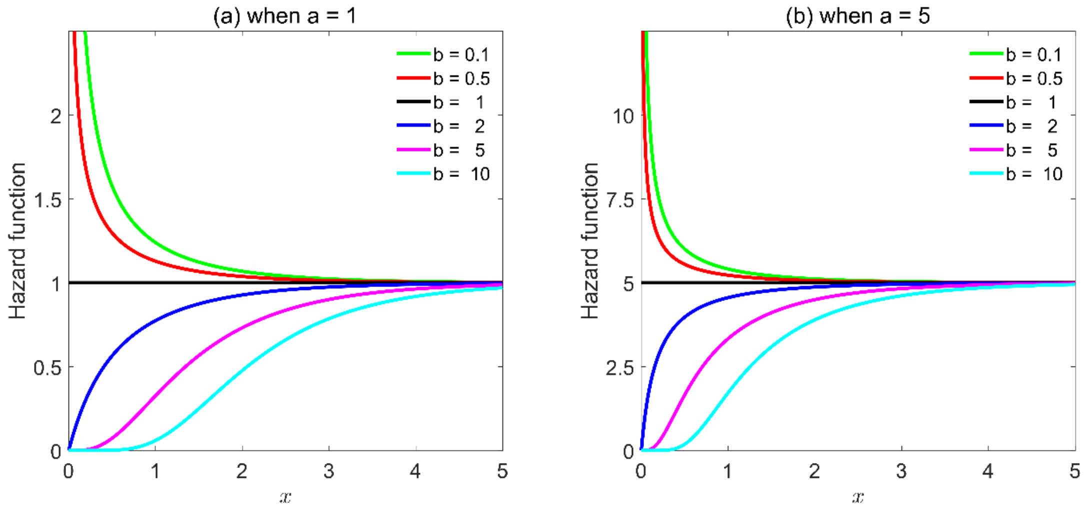

Figure 2. Some of the hazard functions of the KB distribution.

Figure 2 shows some of the hazard functions of the KB distribution when parameter a is set equal to 1 and 5, respectively, with several variations of parameter b. For any parameter a, the hazard function is increasing from 0 to a, if $b > 1$ and is decreasing from ∞ to a, if $b < 1$. Meanwhile, it is constant for $b = 1$. The behavior of this hazard function resembles that of the gamma and exponentiated exponential distributions, discussed by [21]. The closer parameter b is to 1, the faster the rate of the hazard function is towards a. Moreover, the higher the parameter a, the hazard function is relatively faster towards a, when parameter b is set equal.

2.3 Quantile Function and Generating Random Variables

By inverting $F(x|a,b) = u$, the quantile function of X is

$$F^{-1}(u|a,b) = -\log Q_1(1-u|a,b) \tag{9}$$

where $Q_1(\cdot\,|a,b)$ is the inverse cdf of the beta distribution (Equation (1)). Using the quantile function, the random variable

$$X = -\log Q_1(1-V|a,b) = -\log Q_1(V|a,b) \tag{10}$$

has density function (Equation (4)), where V is a uniform random variable over the interval $(0,1)$.

3. PARAMETER ESTIMATION

Accurate estimation of the parameters (a, b) is crucial for the practical application of the KB distribution in rainfall modelling. In this study, we employed the MLE framework because of its well-known properties of consistency, efficiency, and asymptotic unbiasedness. The estimation procedure was refined and described in detail as follows:

3.1 Maximum Likelihood Estimator

Let the random variables $X_1, X_2, \ldots, X_n \sim KB(a,b)$ with observed values $x_1, x_2, \ldots, x_n$. From Equation (4), the log-likelihood function can be written as

$$\log L(a,b) = -n\log B(a,b) - a\sum_{i=1}^{n} x_i + (b-a)\sum_{i=1}^{n}\log(1-e^{x_i}) \tag{11}$$

Therefore, to find the MLE of a and b, we can maximize Equation (11) directly with respect to a and b. We can view the functions in Equation (11) as a nonlinear function and find the maximum value of $\log L$ using numerical methods. One of them is the Nelder-Mead simplex algorithm which will be discussed next.

3.2 Optimization Algorithm

To enable the use of generic numerical solvers, the maximization problem is reformulated as the minimization of the negative log-likelihood:

$$\hat{a}, \hat{b} = \min_{a>0,b>0}\{-\log L(a,b)\} \tag{12}$$

This transformation does not change the optimum solution but aligns with the standard structure required by unconstrained optimization algorithms.

In this study, we use the *fminsearch* algorithm to estimate the parameter values of the KB distribution. The *fminsearch* algorithm uses the Nelder-Mead simplex algorithm as described in [28]. Since parameters a and b must be greater than 0, bound constraints are required but the *fminsearch* algorithm does not admit these constraints. The *fminsearchbnd* algorithm is used exactly like *fminsearch*, except that bounds are applied to the variables. The bounds are applied internally, using a transformation of the variables. Convergence was considered achieved when the change in both parameter estimates, and the objective function was below 10^{-6} between successive iterations. This criterion ensured the stability and reproducibility of the parameter estimation.

3.3 Simulation

To assess the statistical accuracy and reliability of the parameter estimation framework, we designed a Monte Carlo simulation experiment following Ribeiro-Reis [22]. Two scenarios are considered, and the six different sample sizes were chosen, i.e., $n = \{25,50,75,100,200,400\}$. The random numbers are generated using Equation (10) and the true parameters are: $a = 1.9$ and $b = 1.5$ in scenario 1 representing moderate skew and $a = 4.5$ and $b = 2.5$ in scenario 2 representing less skew. However, we add an additional scenario where parameters a and b result in a zero-inflated distribution. We choose parameters $a = 0.2$ and $b = 0.5$ in this situation.

For each scenario and sample size, we generated 15,000 synthetic samples and estimated the parameters using the MLE–Nelder-Mead framework. The performance of the estimators was evaluated using the average estimates (AE) across replications, the bias (difference between AE and true values), and the mean squared error (MSE) as a measure of estimator variability. Table 1 shows the results of all scenarios including the AE, the bias of the AE towards the actual parameter values, and the MSE of each replicate. Obviously, the MLE converges to the true parameters while the bias and MSE decrease as the sample size n increases.

Table 1. Accuracy of the estimation method using optimization algorithm.

n	Param	Scenario 1			Scenario 2			Scenario 3		
		AE	Bias	MSE	AE	Bias	MSE	AE	Bias	MSE
25	a	2.1386	0.2386	0.4778	5.0995	0.5995	2.8737	0.2132	0.0132	0.0027
	b	1.6792	0.1792	0.2748	2.8141	0.3141	0.8102	0.5721	0.0721	0.0473
50	a	2.0096	0.1096	0.1712	4.7901	0.2901	1.0763	0.2066	0.0066	0.0012
	b	1.5817	0.0817	0.1019	2.6533	0.1533	0.3081	0.5348	0.0348	0.017
75	a	1.9721	0.0721	0.1052	4.6811	0.1811	0.6526	0.2047	0.0047	0.0007
	b	1.5544	0.0544	0.0605	2.5936	0.0936	0.1829	0.5254	0.0254	0.0103
100	a	1.9537	0.0537	0.0767	4.6304	0.1304	0.4477	0.204	0.004	0.0006
	b	1.5399	0.0399	0.0456	2.5706	0.0706	0.1284	0.5164	0.0164	0.0068
200	a	1.9268	0.0268	0.0367	4.5671	0.0671	0.218	0.2015	0.0015	0.0003
	b	1.5206	0.0206	0.0216	2.5361	0.0361	0.0617	0.5052	0.0052	0.0031
400	a	1.9115	0.0115	0.0179	4.5297	0.0297	0.1044	0.2011	0.0011	0.0001
	b	1.509	0.009	0.0105	2.5165	0.0165	0.0291	0.5041	0.0041	0.0015

Table 1 presents the results of a Monte Carlo simulation assessing the statistical performance of the MLE procedure for estimating the two parameters (a, b) of the KB distribution under three different parameter settings. Across all scenarios, the results reveal a clear convergence pattern: as the sample size increases from $n = 25$ to $n = 400$, the AE approach the true parameter values, while both the bias and mean squared error (MSE) decline sharply. This behavior reflects the desirable consistency property of MLE, in which larger samples yield more accurate and stable estimates. For example, in Scenario 1 ($a = 1.9, b = 1.5$), the bias of a decreases from 0.2386 at $n = 25$ to 0.0115 at $n = 400$, and the MSE from 0.4778 to 0.0179. A similar trend occurs in Scenario 2 ($a = 4.5, b = 2.5$), though with higher initial bias (e.g., 0.5995 for a at $n = 25$), indicating that less skewed distributions require larger samples for stable estimation. In contrast, Scenario 3 ($a = 0.2, b = 0.5$) representing zero-inflated data shows comparatively low MSE even at small samples, suggesting that the estimator converges more rapidly for distributions concentrated near zero. Overall, these results provide strong empirical evidence that the proposed MLE–Nelder–Mead estimation framework is statistically reliable and well-suited for modelling daily rainfall data, which often exhibit high skewness and zero-inflation.

4. APPLICATION TO DAILY PRECIPITATION DATA

Having established the theoretical formulation of the KB distribution and demonstrated the accuracy and stability of its parameter estimation through a comprehensive simulation study, the next step is to evaluate its practical applicability. In Section 4, we apply the KB distribution to real-world daily precipitation datasets from Indonesia. This analysis aims to examine how well the KB model fits observed rainfall data compared to 17 well-known probability distributions. By assessing its empirical performance using standard goodness-of-fit criteria, we seek to verify whether the theoretical advantages of the KB distribution translate into improved modelling accuracy in practical climatological contexts.

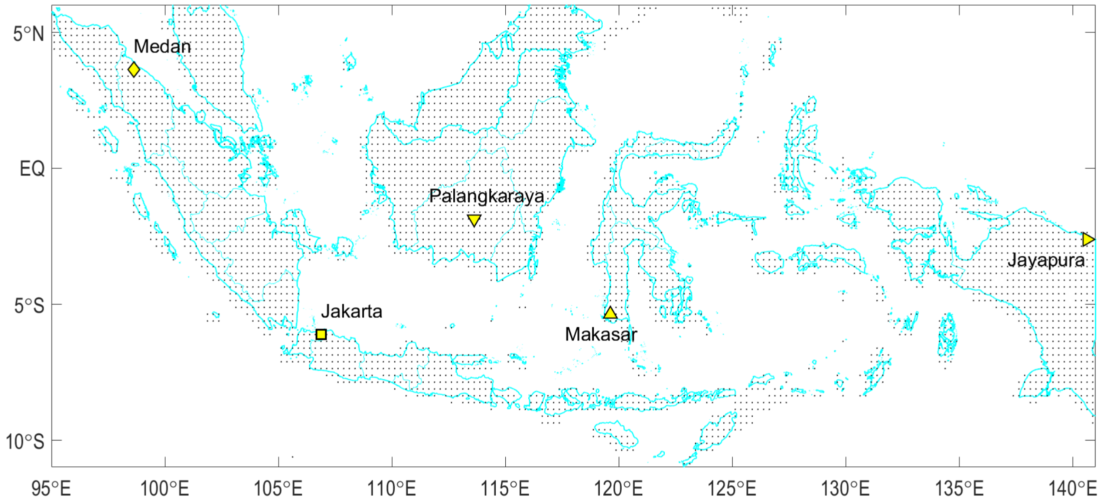

Figure 3. Location of the datasets.

4.1 Datasets

We used data from the Climate Prediction Center morphing technique with corrected bias (CMORPH-CRT) product [29], [30]. It is produced by The National Oceanic and Atmospheric Administration Climate Prediction Center (NOAA CPC), which incorporates precipitation estimates from multiple passive microwave satellites and geostationary satellite infrared data [31]. The reprocessing and bias-correction of CMORPH original product makes CMORPH-CRT better at representing the magnitude and spatial-temporal variations of precipitation over the quasi-global domain [32]. Some experts have used CMORPH-CRT products to estimate actual precipitation. Trinh and Molkenthin [33] show that this product performs better in terms of statistical error than the GSMaP_GNRT6 product. Adding soil moisture with CMORPH-CRT can improve the prediction performance of direct SWDI [34]. Guo et al., [35] also showed that CRT products had better performance than GSMaP_MVK and PERSIANN.

Using the CMORPH-CRT product, we generate two types of datasets: precipitation data for five cities in Indonesia, and precipitation data over Indonesia region (95°E-141°E and 11°S-6°N). We use data with a spatial resolution of 0.25° × 0.25° from 2000-2022 or for 23 years. Figure 3 shows the location of the datasets. The grid with yellow markers shows the locations where the precipitation from five cities was taken for dataset 1, while the black dots show the grid points taken for dataset 2 where there are 3939 grid points in it.

4.2 Comparison With Other Well-Known Distributions

Using the two datasets, we compared the KB distribution with 17 other distributions, such as Birnbaum-Saunders, exponential, extreme value, gamma, generalized extreme value, generalized pareto, half normal, inverse gaussian, logistic, log logistic, lognormal, Nakagami, normal, Rayleigh, Rician, t-location scale, and Weibull distributions. This is intended to measure the performance of the KB distribution with other distributions that are often used.

4.3 Goodness-of-fits Measurement

In this study, three statistical indices were used to evaluate the goodness-of-fit of each candidate probability distribution: Akaike Information Criterion (AIC), Anderson–Darling statistic (A*), and Cramér–von Mises statistic (T*). AIC is an information-theoretic measure that balances model fit and complexity, where lower AIC values indicate a better trade-off between accuracy and parsimony. The Anderson–Darling (A*) and Cramér–von Mises (T*) statistics are empirical distribution function (EDF)-based measures that quantify the distance between the empirical and fitted cumulative distribution functions. Both A* and T* assign larger penalties to deviations, particularly in the tails (for A*). Consequently, lower values of AIC, A*, and T* consistently indicate a better fit of the distribution to the observed data. In our analysis, the distribution achieving the lowest value across these indices is considered the best-fitting model for the given dataset. The formula of each statistic is given by

$$\text{AIC} = 2k - 2\log L \tag{13}$$

$$A^* = -n - \sum_{i=1}^{n} \frac{2i-1}{n}\left[\log F(x_i) + \log\left(1 - F(x_{n+1-i})\right)\right] \tag{14}$$

$$T^* = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - F(x_i)\right]^2 \tag{15}$$

where k is the number of parameters, L is the likelihood function, and n is the sample size. The Anderson-Darling and Cramer-von Mises statistics belong to the class of quadratic EDF statistics (tests based on the empirical distribution function). The Anderson-Darling statistic is a special case of the Cramer-von Mises statistic by placing more weight in the tails of the distribution.

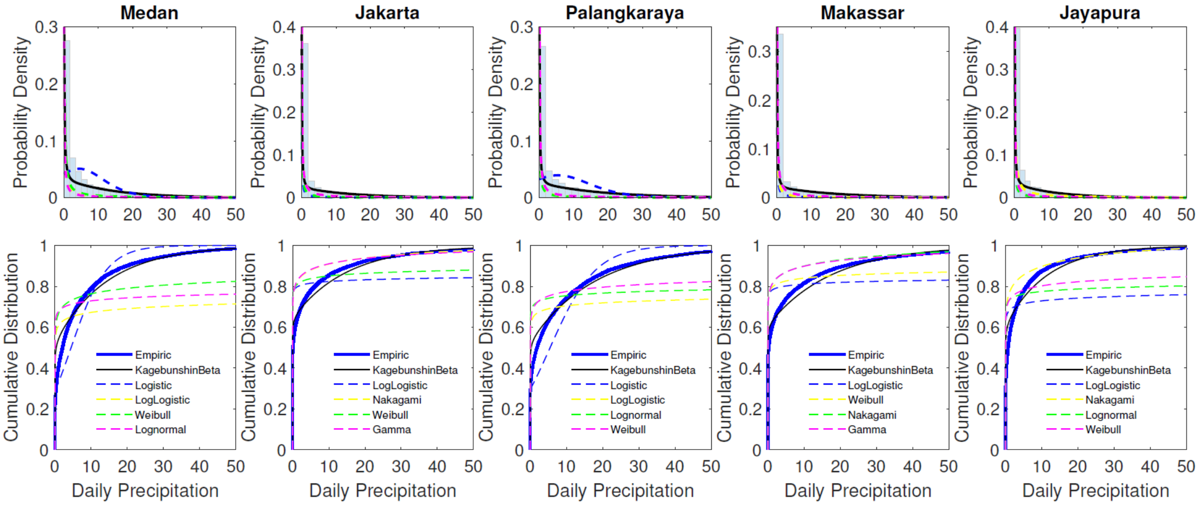

Figure 4. Graphical comparison of the pdfs and cdfs the KB distribution and other distributions with the empirical distribution.

5. PERFORMANCE OF THE KB DISTRIBUTION

This section discusses the performance of the KB distribution compared to other well-known distributions in two datasets.

5.1 Dataset 1

In dataset 1, precipitation data in five cities is used. Table 2 shows the comparison of the KB distribution with top four other distributions sorted by the Anderson-Darling statistic. Meanwhile, Figure 4 shows the graphical comparison of the pdfs and cdfs these parametric distributions with the empirical distribution.

Table 2. Comparison of the KB distribution with top four other distributions sorted by the Anderson-Darling statistic.

City	Distribution	AIC	A*	T*
Medan	Kagebunshin Beta	-67650	466	80.5
	Logistic	61934	737.9	115
	Log Logistic	-52082	1569	240
	Weibull	-56395	1657	260
	Lognormal	-51813	1663	305
Jakarta	Kagebunshin Beta	-212720	638.4	104
	Log Logistic	-204660	1162	194
	Nakagami	-210630	1238	206
	Weibull	-206300	1261	208
	Gamma	-210320	1266	210
Palangkaraya	Kagebunshin Beta	-102200	373.9	57.2
	Logistic	66103	796.7	128
	Log Logistic	-87099	1386	220
	Lognormal	-87622	1490	264
	Weibull	-90757	1519	241
Makassar	Kagebunshin Beta	-201090	578.8	92.1
	Log Logistic	-192110	1157	192
	Weibull	-193940	1262	207
	Nakagami	-198600	1280	211
	Gamma	-198230	1285	212
Jayapura	Kagebunshin Beta	-117060	541.1	90.5
	Log Logistic	-104830	1347	215
	Nakagami	-114730	1413	253
	Lognormal	-105470	1455	256
	Weibull	-108330	1477	236

Table 2 clearly shows that the KB distribution outperformed other distributions indicated by the lowest values of AIC, A* and T*. In the daily precipitation data for Palangkaraya, Makassar and Jayapura, the goodness-of-fit value of the KB distribution far outperforms the second place with more than half of the criteria. Meanwhile, other well-known distributions that have fairly good information criteria include Logistic, Log logistics, Weibull, Gamma, Nakagami, and Lognormal.

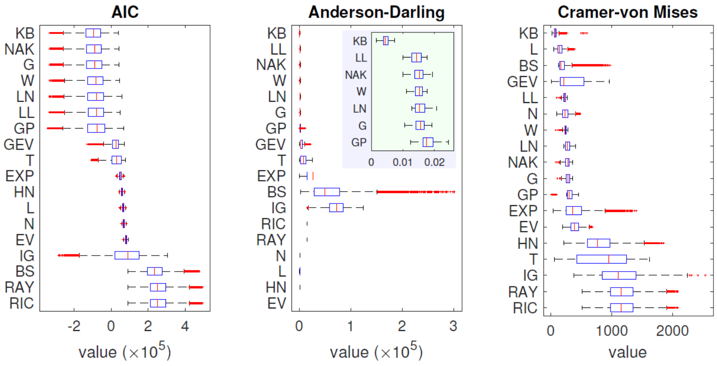

Figure 5. Boxplot of goodness-of-fits for each distribution using dataset 2.

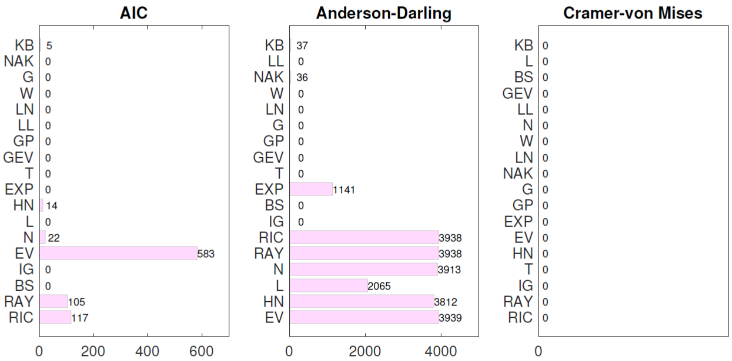

Figure 6. Number of trials that assign infinite values to each of the AIC, Anderson-Darling, and Cramer-von Mises statistics.

The Weibull and Lognormal distributions are the most used and widely applicable distributions for life data analysis. Furthermore, Logistics, Log logistics, and Gamma distribution also have various applications and can be found in many statistical references. Meanwhile, the Nakagami distribution is related to the gamma distribution. In particular, given a random variable $Y \sim \text{gamma}(k, \theta)$, it is possible to obtain a random variable $Y \sim \text{Nakagami}(m, \Omega)$ by setting $k = m$, $\theta = \Omega/m$, and and taking the square root of Y: $X = \sqrt{Y}$.

Based on Figure 4, the KB distribution produces satisfactory results compared to other distributions. The KB distribution provides pdf and cdf graphs closer to the empirical distribution of daily precipitation data than other distribution functions. However, the cdf of the KB distribution has higher values than the empirical cdf at low rainfall intensities.

5.2 Dataset 2

Dataset 2 has a much larger number of trials than dataset 1. If dataset 1 only consists of five trial points, i.e., five cities, then dataset 2 has 3939 points which will be fitted using the 18 distribution functions. It takes approximately seven hours to fit all distribution functions and measure their goodness-of-fit. Figure 5 shows the boxplot of goodness-of-fit for each distribution using dataset 2 and we have sorted it based on the median.

The KB distribution outperforms other distributions based on three information criteria. This corroborates the results in dataset 1 by involving a much larger number of trials. Based on AIC and Anderson-Darling statistics, the distribution functions that can be considered good enough to represent daily precipitation data are the KB, Nakagami, gamma, Weibull, lognormal, log logistic, and generalized Pareto distributions. However, referring to the Anderson-Darling statistics, the KB distribution far outperforms the log logistic distribution below it. Meanwhile, there are additional distributions that can be considered good based on Cramer-von Mises statistics, such as logistic, Birnbaum-Saunders, generalized extreme value, and normal distributions. However, some of these distributions have low AIC and Anderson-Darling values, because many of the trials produce infinite values for these distributions. Figure 6 shows the number of trials that assign infinite values to each of the AIC, Anderson-Darling, and Cramer-von Mises statistics.

Although the KB distribution gives satisfactory results, there are a few of the AIC and Anderson-Darling criteria that provide infinite values, which are 5 (0.12%) and 37 (0.94%) of trials, respectively. Less than one percent. Distributions that provide the most infinite value to AIC criteria are Eigen Value, Rician, and Rayleigh Distributions. Furthermore, the logistic distribution of more than 50\% of trials gives infinite value to the Anderson-Darling criteria. Thus, logistic distribution is not recommended even though it provides satisfactory value to the criteria of Cramer-Von Mises. Overall, the KB distribution is the most selected distribution based on the three criteria as can be seen in Figure 7.

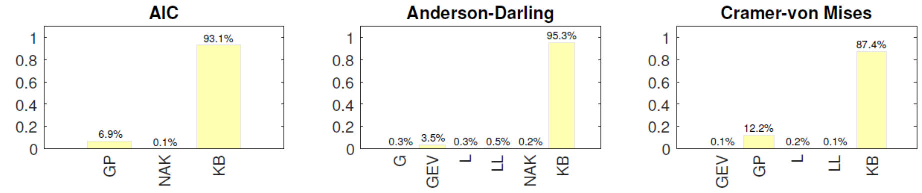

Figure 7. Percentage of selected distribution by the AIC, Anderson-Darling, and Cramer-von Mises statistics.

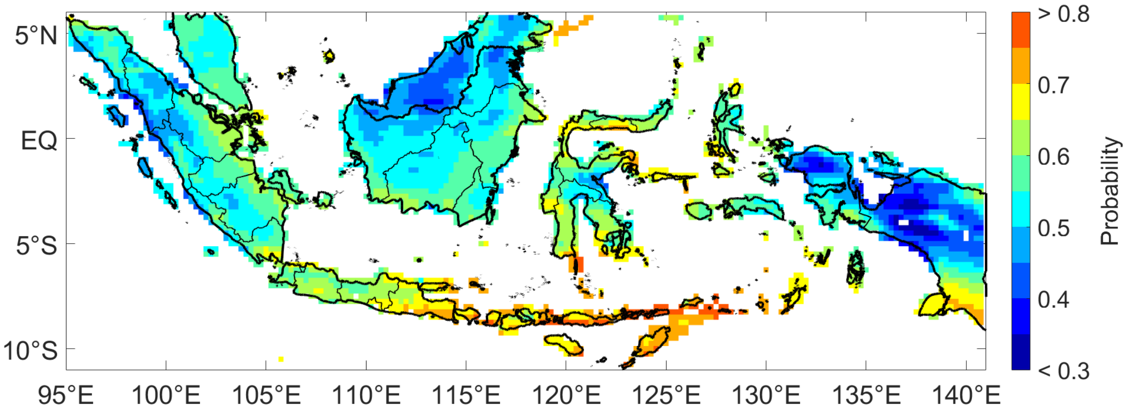

Figure 8. Probability of dry days in Indonesia using the KB distribution.

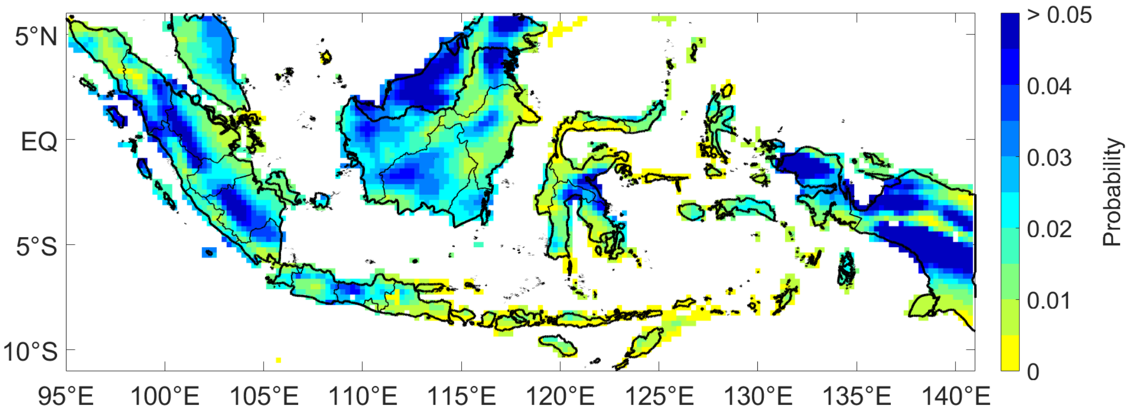

Figure 9. Probability of heavy rainy days in Indonesia using the KB distribution.

6. PROBABILITY OF DRY DAYS AND HEAVY RAINY DAYS IN INDONESIA

Knowing that KB Distribution is the most recommended parametric distribution for daily precipiration data in Indonesia, this section will analyze the probability of dry days and heavy rainy days using the distribution. Figure 8 shows the probability of dry days in Indonesia, i.e., days with less than one millimeter of precipitation. Meanwhile, Figure 9 shows the probability of heavy rainy days in Indonesia using the KB distribution, i.e., daily probability of more than 50 mm per day.

Based on Figure 8, areas in Indonesia that have a probability of dry days of less than 50% include the Lake Toba area of North Sumatra, the west coast of West and North Sumatra, the heart of Borneo (covering parts of East Kalimantan, North Kalimantan, and Sabah and Sarawak, Malaysia), the mountains of Papua, and a small part of Central Sulawesi. The probability of a dry day that is less than 50% indicates that the probability of a day without rain is less than 50%, meaning that in a year the days without rain are less than six months. Thus, on the contrary, days with rain are more than six months a year without considering other variables such as the influence of global climate conditions, i.e., El Nino-Southern Oscillation and Indian Ocean Dipole cycles. Meanwhile, several areas detected have a high probability of dry days (i.e., more than 70%) spreading from the eastern part of East Java, Bali, Nusa Tenggara, to Merauke. A probability of more than 70% means that there will be dry days more than 8.5 months a year without considering other climatic factors.

Regions that have a relatively high probability (i.e., more than 3%) include the highlands of Sumatra, Kalimantan, Central Sulawesi, to Papua according to Figure 9. Also, several areas on the island of Java, i.e., in Central Java province and Bogor regency in West Java province. A probability of more than 3% means that there will be heavy rainy days with more than 50 mm of rain per day more than 10 times a year regardless of other climatic factors.

7. CONCLUSION

A lifetime distribution has been proposed previously as the KB distribution by [22] obtained from the transformation of the beta distribution. This article discusses the behavior of the hazard function of the KB distribution which resembles the gamma and exponentiated exponential distributions. Using the maximum likelihood estimator, this article simulates the parameter estimation of the KB distribution using the *fminsearchbnd* algorithm which is based on the Nelder-Mead simplex algorithm. The simulation results show satisfactory accuracy of the several scenarios used in the parameter estimation process.

We compare the performance of the KB distribution with several other better-known distributions on daily precipitation data in Indonesia. Of the two types of datasets used, the KB distribution outperforms other distributions based on several statistical criteria, such as AIC, Anderson-Darling, and Cramer-von Mises statistics. Knowing the performance of the KB distribution is better than other distributions, the KB distribution is applied to estimate the probability of dry days and heavy rainy days in Indonesia. Therefore, this study found several areas in Indonesia that have a relatively high probability of dry days and heavy rainy days.

The potential applications of the KB distribution in climatological analysis extend beyond its current use in modelling daily precipitation. Future developments could explore its utility in bias correction methodologies, enhancing the accuracy of satellite-based rainfall estimations and climate model outputs. Otherwise, future studies could further evaluate the predictive capability of the KB distribution by fitting it to station-based observational rainfall data and comparing the simulated rainfall values directly with observed measurements to assess its accuracy in reproducing actual daily rainfall patterns. Additionally, the KB distribution holds promise for risk assessment in hydrology, particularly in evaluating the probability of extreme weather events such as floods and droughts. Its flexibility in capturing complex data distributions makes it a strong candidate for return period estimation, providing more reliable statistical foundations for infrastructure planning and disaster mitigation strategies. Furthermore, integrating the KB distribution with machine learning techniques could open new avenues for advanced modelling approaches, improving long-term climate predictions and decision-making processes related to water resource management and adaptation to climate change. These advancements align with global efforts in climate resilience and sustainability, contributing to data-driven policymaking for environmental risk reduction.

ACKNOWLEDGEMENT AND FUNDING

The author receives no financial support for the research, authorship, and publication of this article.

DECLARATION OF CONFLICTING INTERESTS

The author declares no potential conflicts of interest with respect to the research and publication of this article.

REFERENCES

[1] M. R. Karim, F. Sultana, M. S. R. Saimun, S. A. Mukul and M. A. S. Arfin-Khan, Plant diversity and local rainfall regime mediate soil ecosystem functions in tropical forests of north-east Bangladesh, *Environmental Advances*, 2, 2020, 100022.

[2] H. R. Bedane, K. T. Beketie, E. E. Fantahun, G. L. Feyisa and F. A. Anose, The impact of rainfall variability and crop production on vertisols in the central highlands of Ethiopia, *Environmental Systems Research*, 11, 2022, 26.

[3] M. Z. Islam and C. Wang, Cost of high-level flooding as a consequence of climate change driver?: A case study of China's flood-prone regions, *Ecological Indicators*, 160, 2024, 111944.

[4] J. Chen, H. Yang, T. Jin and K. Wu, Assessment of terrestrial ecosystem sensitivity to climate change in arid, semi-arid, sub-humid, and humid regions using EVI, LAI, and SIF products, *Ecological Indicators*, 158, 2024, 111511.

[5] Z. Fang, A. B. Morales, Y. Wang and L. Lombardo, Climate change has increased rainfall-induced landslide damages in central China, *International Journal of Disaster Risk Reduction*, 119, 2025, 105320.

[6] K. Khosravi, A. A. Farooque, S. M. Bateni, C. Jun and J. Dhiman, Prediction of three vital rainfall characteristics: Advanced hybrid tree- or lazy-based learner?, *Results in Engineering*, 25, 2025, 103840.

[7] C. Wang, H. Liu, Y. Xu and F. Zhang, A forest fire prediction framework based on multiple machine learning models, *Forests*, 16, 2025, 329.

[8] X. Li and Z. Li, Assessment of bias correction methods for high resolution daily precipitation projections with CMIP6 models: A Canadian case study, *Journal of Hydrology: Regional Studies*, 58, 2025, 102223.

[9] M. Taie Semiromi and M. Koch, Statistical downscaling of precipitation in northwestern Iran using a hybrid model of discrete wavelet transform, artificial neural networks, and quantile mapping, *Theoretical and Applied Climatology*, 155, 2024, 6591-6621.

[10] A. Bezdan, J. Bezdan, M. Marković, D. Mirčetić, A. Baumgertel, A. Salvai and B. Blagojević, An objective methodology for waterlogging risk assessment based on the entropy weighting method and machine learning, *Catena*, 249, 2025, 108618.

[11] M. K. Mohamed Salleh, N. A. Mohd Amin, N. F. Ahmad Radi, S. A. Zakaria, W. N. M. Ariffin and M. H. Masran, The comparison of extreme rainfall prediction for Northern region of Peninsular Malaysia based on GEV and GPD models, *AIP Conference Proceedings*, 3189, 2024, 100008.

[12] M. ALRakathi and A. Alodah, Assessing the impact of climate change on intensity-duration-frequency (IDF) curves for the Qassim region, Saudi Arabia, *Atmosphere*, 16, 2025, 59.

[13] Z. Sa'adi, Z. Yusop, N. E. Alias, M. F. Chow, M. K. I. Muhammad, M. W. A. Ramli, Z. Iqbal, M. S. Shiru, F. I. W. Rohmat, N. A. Mohamad and M. F. Ahmad, Evaluating imputation methods for rainfall data under high variability in Johor river basin, Malaysia, *Applied Computing and Geosciences*, 20, 2023, 100145.

[14] F. Salgado-Castillo, M. Barrios and J. Velez Upegui, Skew-normal distribution model for rainfall uncertainty estimation in a distributed hydrological model, *Hydrological Sciences Journal*, 68, 2023, 542-551.

[15] L. Liu, Y. -Chen, T. Shih, R. L. Strawderman, D. Zhang, B. A. Johnson and H. Chai, Statistical analysis of zero-inflated nonnegative continuous data, *Statistical Science*, 34, 2019, 253-279.

[16] M. B. Ben Khedher and D. Yun, An interpretable machine learning-based hurdle model for zero-inflated road crash frequency data analysis: Real-world assessment and validation, *Applied Sciences*, 14(23), 2024, 10790.

[17] C. M. McBride, A. C. Kruger and L. Dyson, Changes in extreme daily rainfall characteristics in South Africa: 1921–2020, *Weather and Climate Extremes*, 38, 2022, 100517.

[18] W. Zheng, S. Liu, Z. Zhou, G. Zhong and Q. Zhuang, A weibull-based framework for uncertainty evaluation in rainfall frequency analysis, *SSRN*, 2023, doi: 10.2139/ssrn.4677800

[19] D. Ozonur, I. Pobocikova and A. de Souza, Statistical analysis of monthly rainfall in Central West Brazil using probability distributions, *Modeling Earth Systems and Environment*, 7, 2021, 1979-1989.

[20] L. Ye, L. S. Hanson, P. Ding, D. Wang and R. M. Vogel, The probability distribution of daily precipitation at the point and catchment scales in the United States, *Hydrology and Earth System Sciences*, 22, 2018, 6519-6531.

[21] R. D. Gupta and D. Kundu, Exponentiated exponential family: An alternative to gamma and Weibull distributions, *Biometrical Journal*, 43, 2001, 117-130.

[22] L. D. Ribeiro-Reis, The kagebushin-beta distribution: An alternative for gamma, Weibull and exponentiated exponential distributions, *Journal of the Egyptian Mathematical Society*, 30, 2022, 24.

[23] S. Nadarajah, The exponentiated Gumbel distribution with climate application, *Environmetrics*, 17, 2006, 13-23.

[24] S. Vergara-cardozo, The Family of log-skew-normal alpha-power distributions using precipitation data, *Revista Colombiana de Estadistica*, 36, 2013, 43-57.

[25] A. Menezes, J. Mazucheli, J. Cardoso and S. Chakraborty, The transmuted half-normal distribution with application to precipitation data, *Pesquisa Operacional*, 40, 2020, 1-30.

[26] E. Tanprayoon, U. Tonggumnead and S. Aryuyuen, A new extension of generalized extreme value distribution: extreme value analysis and return level estimation of the rainfall data, *Trends in Sciences*, 20(1), 2023, 4034.

[27] N. Alsadat, A new extension of XLindley distribution with mathematical properties, estimation, and application on the rainfall data, *Heliyon*, 10(9), 2024, e38143.

[28] J. C. Lagarias, J. A. Reeds, M. H. Wright and P. E. Wright, Convergence properties of the Nelder-Mead simplex method in low dimensions, *SIAM Journal on Optimization*, 9, 1998, 112-147.

[29] M. K. Najib, S. Nurdiati and A. Sopaheluwakan, Copula-based joint distribution analysis of the ENSO effect on the drought indicators over Borneo fire-prone areas, *Modeling Earth Systems and Environment*, 8, 2022, 2817-2826.

[30] M. K. Najib, S. Nurdiati and A. Sopaheluwakan, Multivariate fire risk models using copula regression in Kalimantan, Indonesia, *Natural Hazards*, 113, 2022, 1263-1283.

[31] R. Satyaningsih, V. Jetten, J. Ettema, A. Sopaheluwakan, L. Lombardo and D. E. Nuryanto, Dynamic rainfall thresholds for landslide early warning in Progo Catchment, Java, Indonesia, *Natural Hazards*, 119, 2023, 2133-2158.

[32] P. Xie, R. Joyce, S. Wu, S. H. Yoo, Y. Yarosh, F. Sun and R. Lin, Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998, *Journal of Hydrometeorology*, 18, 2017, 1617-1641.

[33] M. X. Trinh and F. Molkenthin, Flood hazard mapping for data-scarce and ungauged coastal river basins using advanced hydrodynamic models, high temporal-spatial resolution remote sensing precipitation data, and satellite imageries, *Natural Hazards*, 109, 2021, 441-469.

[34] Q. Zhu, Y. Luo, D. Zhou, Y. P. Xu, G. Wang and Y. Tian, Drought prediction using in situ and remote sensing products with SVM over the Xiang River Basin, China, *Natural Hazards*, 105, 2021, 2161-2185.

[35] H. Guo, S. Chen, A. Bao, J. Hu, B. Yang and P. M. Stepanian, Comprehensive evaluation of high-resolution satellite-based precipitation products over China, *Atmosphere*, 7(1), 2016, 6.