AIMS Mathematics, 11(6): 16788–16810.
DOI: 10.3934/math.2026689
Received: 14 April 2026
Revised: 20 May 2026
Accepted: 26 May 2026
Published:11 June 2026

https://www.aimspress.com/journal/Math

Research article

Optimal distribution selection and spatial drought modeling in upper northern Thailand using an integrated actual precipitation index and regular vine copula framework

Kritdilada Luanmuang[1], Manad KhamKong[2,*], Nawapon Nakharutai[2] and Pimwarat Srikummoon[2]

[1] Ph.D. Program in Applied Statistics, Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
[2] Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

* **Correspondence:** Email: manad.k@cmu.ac.th.

Abstract. The standardized precipitation index (SPI) is a foundational tool for drought assessment. However, its application is constrained by complex probability distribution selection and mandatory normal transformation. To overcome these limitations, this study introduces a systematic framework that uses the Anderson–Darling test to optimize distribution selection. This approach refines drought evaluation through the actual precipitation index (API). By using untransformed precipitation data, the API enables a more direct climatic assessment. Results demonstrated strong consistency between the API and SPI across upper northern Thailand. Both indices successfully detected the 2011 regional floods and the 2015-2016 El Niño drought. Comparatively, the API demonstrated superior sensitivity to moisture saturation, particularly in Chiang Mai, Nan, and Phayao. Furthermore, spatial dynamic analysis using regular vine (R-vine) copulas identified Lampang as the primary regional hub. Lampang governs the central cluster (Chiang Mai, Lamphun, and Phrae) and mediates dependencies between Phayao and Chiang Rai. Nevertheless, localized geographic interactions create substantial concurrent extreme rainfall risks for the Chiang Mai–Lamphun pair. Because it preserves physical rainfall units, the API facilitates more actionable risk management than the SPI. Consequently, integrating R-vine copulas within the API framework is strongly recommended. This integration enhances spatial rainfall modeling, early warning systems, and adaptive water resource management, ultimately supporting climate resilience and sustainable development in vulnerable regions.

Keywords: actual precipitation index; climate resilience; drought classification; extreme rainfall modeling; regular vine copula; standardized precipitation index; sustainable water management
Mathematics Subject Classification: 62F03, 60E05, 62H05

1. Introduction

Drought is recognized as a complex and devastating natural hazard, stemming primarily from prolonged precipitation deficits that trigger water scarcity across ecological, agricultural, and socio-economic systems. In Thailand, drought is a recurrent phenomenon that imposes severe constraints on water resources and agricultural productivity, particularly in rainfed regions lacking adequate storage infrastructure [1,2]. Given the agricultural sector's pivotal role in the national economy, implementing robust drought monitoring and early warning systems is imperative to minimize adverse impacts and foster sustainable development. Upper northern Thailand, comprising the Chiang Mai, Chiang Rai, Lamphun, Phayao, Lampang, Mae Hong Son, Phrae, and Nan provinces, is among the most drought-prone regions in the country. The region is characterized by complex mountainous terrain, strong seasonal rainfall variability, and frequent influence from large-scale climate phenomena such as the El Niño–Southern Oscillation (ENSO), which can exacerbate drought severity by altering rainfall timing and intensity [1]. Similar challenges associated with monsoon-dominated rainfall regimes have been reported across Southeast Asia, where strong seasonality and skewed precipitation distributions complicate drought characterization and monitoring [3]. Quantitative drought indices are fundamental for assessing the severity, duration, and spatial distribution of drought. While the standardized precipitation index (SPI) [4] is favored for its simplicity and multiscale capability, extensions like the standardized precipitation evapotranspiration index (SPEI) [5] account for thermodynamic effects on water balance. Despite their widespread utility, these indices depend on normalizing transformations that may yield inaccuracies when applied to non-normal or highly skewed precipitation data [6]. Given the significant spatiotemporal variability of rainfall, studies suggest that alternative probability functions, including lognormal, Weibull, and generalized exponential distributions, often provide a better fit for specific local climatology [7–9]. Failure to select an appropriate distribution can significantly distort drought severity estimation and impair the detection of extreme dry and wet events, especially in analyses sensitive to distribution tails [10].

To circumvent these limitations, Sen and Almazroui [11] proposed the actual precipitation index (API) as an alternative rainfall-based metric that avoids the standard normal transformation. However, their original framework was exclusively formulated using the gamma distribution. Instead of relying on a single function, the API approach should ideally utilize quantiles derived from the best-fitted precipitation distribution, allowing drought thresholds to be expressed in terms of tangible rainfall magnitudes. Therefore, the accurate identification of suitable probability distributions beyond just the gamma function is a fundamental prerequisite for robust drought analysis. While model selection protocols based on information criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are standard in hydrological frequency analysis [12,13], simulation studies highlight the superior performance of tail sensitive metrics, particularly the Anderson–Darling test, in characterizing skewed hydro-climatic variables [14,15]. Despite these advances, there remains a paucity of research systematically examining how distribution selection

criteria influence API-based drought classification in monsoon-dominated regions like northern Thailand [16]. Specifically, comparative literature elucidating the nexus between statistical distribution choice, drought index formulation, and the sensitivity to extreme dry and wet events is currently lacking. Moreover, specific studies investigating the inter-station rainfall relationships within the upper northern region of Thailand remain unavailable. Notably, precipitation patterns at individual stations are characterized by distinct right-skewed distributions.

This study aims to implement the API for drought characterization in the upper northern region of Thailand through the identification of optimal probability distributions. Furthermore, it investigates rainfall correlations derived from the selected distribution models. By improving drought monitoring precision and understanding spatial rainfall dynamics, this study provides a framework for adaptive water management and early warning systems under the challenges of contemporary climate variability.

2. Materials and methods

2.1. Statistical distribution and model selection

In this study, the selected statistical distributions are two-parameter, right-skewed models [16], as previous studies have shown that they adequately fit rainfall data in the upper northern region of Thailand. These distributions are summarized in Table 1. Goodness-of-fit tests are standard procedures for quantifying the agreement between statistical distributions and empirical observations [12]. In conjunction with these tests, information criteria, specifically the AIC and BIC, are employed to compare and select the best-fitting distribution. This study utilizes the following statistical measures for model assessment. The goodness-of-fit tests, including the Anderson–Darling (AD) and Kolmogorov–Smirnov (KS) tests, are defined as follows in Eqs (1) and (2):

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2n-1)\left[\ln F(x_i) + \ln(1 - F(x_{n+1-i}))\right], \tag{1}$$

$$KS = \max_{1\le i\le n}\left(\left|F(x_i) - \frac{i-1}{n}\right|, \left|\frac{i}{n} - F(x_i)\right|\right), \tag{2}$$

where x_{i} represents ordered observations, F is the cumulative distribution function (cdf) of the specified distribution, and n is the sample size. The model selection criteria, including AIC and BIC, are given in Eqs (3) and (4) as follows:

$$AIC = 2k - 2\ln(L), \tag{3}$$

$$BIC = k\ln(n) - 2\ln(L), \tag{4}$$

where k is the number of parameters in the model, L is the maximized value of the likelihood function for the estimated model, and n is the sample size.

Table 1. Types of probability distributions.

Distribution	Probability density function (pdf)	Cumulative distribution function (cdf)
Gamma	$f(x;\alpha,\beta) = x^{(\alpha-1)}\frac{e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$, where $x>0,$ and $\alpha,\beta>0$, α is the shape parameter, and β is the scale parameter.	$F(x;\alpha,\beta) = \frac{\gamma\left(\alpha,\frac{x}{\beta}\right)}{\Gamma(\alpha)}$, where $\Gamma(.)$ is the gamma function, and $\gamma(.)$ is the lower incomplete gamma function.
Weibull	$f(x;\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^{\beta}}$, where $x>0,$ and $\alpha,\beta>0$.	$F(x;\alpha,\beta) = 1-e^{-\left(\frac{x}{\beta}\right)^{\alpha}}$
Lognormal	$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}$, where $x>0, \mu\in R$ and $\sigma^2>0$.	$F(x;\mu,\sigma) = \frac{1}{2}erfc\left[\frac{\ln x-\mu}{\sigma\sqrt{2}}\right]$, where $erfc(x)=\frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\,dt$ is the complementary error function.
Generalized exponential	$f(x;\alpha,\lambda) = \alpha\lambda(1-e^{-\lambda x})^{\alpha-1}e^{-\lambda\alpha}$, where x,α and $\lambda>0$.	$F(x;\alpha,\lambda) = (1-e^{-\lambda x})^{\alpha}$

The AD test is a widely used hypothesis test, augmented by information-theoretic criteria such as the AIC and BIC. To ensure reliable model selection, a combination of these techniques is applied [17,18]; specifically, the distribution exhibiting the lowest AD statistic, and AIC and BIC values are deemed the most suitable.

2.2. Standardized precipitation index and actual precipitation index

The SPI, introduced by McKee et al. [8], is a multi-scalar index that requires fitting rainfall data to a probability distribution before transforming it to a standard normal distribution. However, summer rainfall in northern Thailand is characterized by a right-skewed distribution rather than normality. SPI values reflect deviations from the long-term mean; positive values indicate wet conditions, while negative values signal drought. Thus, the index is effective for monitoring the full spectrum of hydro-climatic variability [19]. Because the summer rainfall data contains months with zero rainfall, an adjusted cumulative distribution function was applied to handle these records, as follows:

$$H(x)=q+(1-q)G(x), \tag{5}$$

where q is the probability of zero precipitation, $G(x)$ is the cumulative distribution function of non-zero rainfall amounts, and $H(x)$ is the adjusted cumulative probability. Following Edwards and McKee [4], the final SPI value is determined by transforming $H(x)$ into a standard normal variable. This transformation is effectively computed using the approximation provided by Abramowitz and Stegun [19]:

$$SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right), \quad t = \sqrt{\ln\left(\frac{1}{(H(x))^2} \right)}, \text{ for } 0 < H(x) \le 0.5, \tag{6}$$

$$SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right), \quad t = \sqrt{\ln\left(\frac{1}{(1 - H(x))^2} \right)}, \text{ for } 0.5 < H(x) \le 1, \tag{7}$$

where $c_0 = 2.515517$, $c_1 = 0.802853$, $c_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, $d_3 = 0.001308$.

The interpretation of SPI values and the corresponding cumulative probabilities is presented in Table 2.

Table 2. SPI classification, cumulative probability values, and drought severity classes [11].

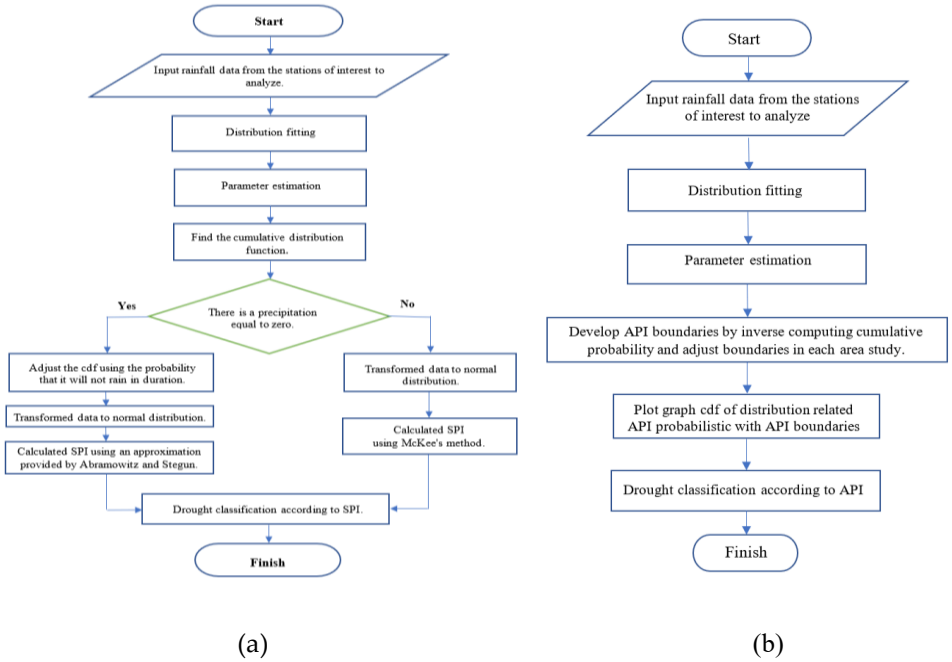
SPI values	Cumulative probability	Drought classes
2.00 and above	≥0.9772	Extremely wet
1.50 to 1.99	0.9332 to 0.9772	Severely wet
1.00 to 1.49	0.8413 to 0.9332	Moderately wet
0.00 to 0.99	0.5000 to 0.8413	Mildly wet
0.00 to -0.99	0.1587 to 0.5000	Mildly dry
-1.00 to -1.49	0.0668 to 0.1587	Moderately dry
-1.50 to -1.99	0.0228 to 0.0668	Severely dry
-2.00 and less	<0.0228	Extremely dry

Another method used to evaluate drought conditions in this study is the API, developed by Sen and Almazroui [11]. The main difference between the SPI and the API is that the SPI standardizes rainfall data into a normal distribution, whereas the API retains the original rainfall scale. To establish API thresholds, the optimal distribution for the dataset is selected, and its quantiles are derived by mapping the cumulative probabilities from the standard normal cdf (corresponding to SPI thresholds) back to the original scale. A distinct advantage of the API is that it bypasses the transformation of data to a standard normal distribution. This approach is particularly effective for managing right-skewed distributions, thereby mitigating the limitations of normalization, the performance of which depends significantly on the chosen transformation method [20-21]. Consequently, the API proposed in this study is developed by identifying the most suitable probability distributions for summer rainfall. The API values are subsequently calculated using the quantiles of the cdf for the distributions identified in Table 1. Specifically, the API thresholds are determined by the cumulative probability (p) values outlined in Table 2, which involves solving for the quantile x_p in Eq (8):

$$F_X(x_p) = P(X \le x_p) = p. \tag{8}$$

These thresholds serve as the criteria for classifying precipitation levels into their respective drought severity classes. Subsequently, the most effective criteria were applied to the summer rainfall data from the eight observation stations to determine the optimal distribution for each site. Finally, a comparative analysis was conducted between the SPI (as shown in Figure 1a) and the API (as shown

in Figure 1b).



(a) (b)

Figure 1. Workflow for the determination of drought indices: (a) SPI and (b) API.

2.3. Simulation methodology and study area

2.3.1. Numerical methodology

A Monte Carlo simulation evaluated the distributional transformation reliability for summer rainfall using four right-skewed models [9,21-22]: gamma, Weibull, generalized exponential (shape = 0.2, 0.8, 1.2; scale = 1), and log-normal (location = -1.5, 1.5, 2.5; scale = 1). The scale parameter was fixed at 1 across all models due to its scale-invariant nature (Supplementary Figure S1). For sample sizes $n \in \{30, 50, 100, 300\}$, the data generation algorithm was performed over 1,000 iterations following these sequential steps:

1) Generate synthetic random samples from the specified distributions.

2) Estimate model parameters using maximum likelihood estimation (MLE)

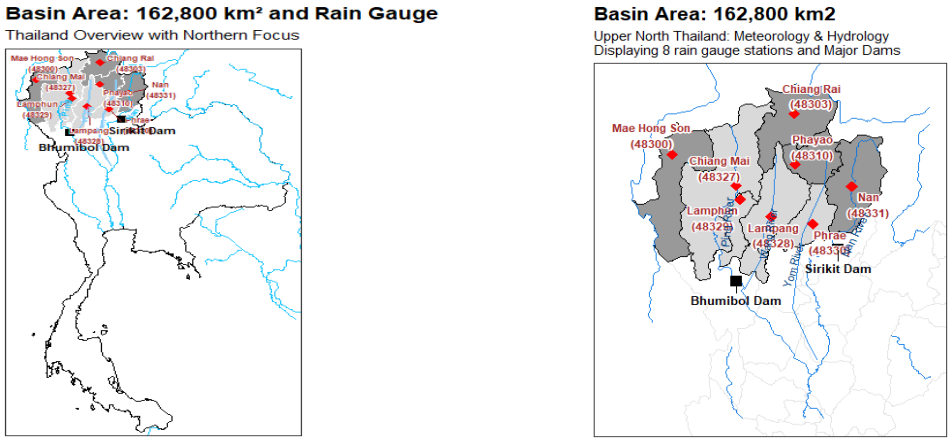
3) Evaluate the empirical Type I error rates at $\alpha = 0.05$ and empirical power, utilizing the AD, and KS tests, along with the AIC and BIC criteria.

4) Apply the AD tests to verify whether the SPI-transformed series follow a standard normal distribution.

2.3.2. Study area and rainfall data

The upper northern region of Thailand comprises eight provinces: Mae Hong Son, Chiang Mai, Chiang Rai, Lampang, Lamphun, Phayao, Phrae, and Nan (Figure 2b). Some of these provinces share borders with Myanmar and Laos. This region features four main rivers: the Ping, Wang, Yom, and Nan. These rivers are tributaries of the Chao Phraya River in central Thailand (Figure 2a). The

region experiences three distinct seasons: summer (February to May), the rainy season (June to September), and winter (October to January). Monthly summer rainfall data from eight rain gauge stations from 1991 to 2023 were selected to analyze drought conditions [23].



(**a**) (**b**)

Figure 2. Map of (a) Thailand overview and (b) eight provinces in upper northern Thailand.

2.4. Investigating inter-station rainfall dependence structures using regular vine copula

To examine the heterogeneous and right-skewed rainfall distributions in upper northern Thailand [16], the regular vine copula (R-vine) approach by Dißmann et al. [24] is adopted. This framework is specifically employed to scrutinize nonlinear dependence structures and tail characteristics between stations. The findings confirm that inter-station dependence is not only spatially driven but also exhibits significant tail dependence, particularly during extreme rainfall events. The analytical framework consists of the following four stages:

1) Marginal distribution fitting: The optimal distribution for each station is selected based on the criteria specified in Section 2.1.

2) Probability integral transform (PIT): To standardize the data and eliminate the influence of varying units and distribution scales, raw rainfall data (X_i, in mm) are transformed into uniform (0,1) variables (U_i) using the fitted cdf in Eq (9):

$$U_i = F_i(X_i;\hat{\theta}_i), \tag{9}$$

where $\hat{\theta}_i$ represents the estimated parameters from the MLE process. This step prepares the data for subsequent dependence structure analysis.

3) R-vine structure selection: An R-vine copula is employed to model the spatial and conditional dependencies among the 8 monitoring stations ($d = 8$), comprising 7 hierarchical trees (T_1, T_2, …, T_7). According to Sklar's theorem, the joint density $f(x_1, x_2, \ldots, x_8)$ is formulated as in Eq (10):

$$f(x_1,x_2,...,x_8) = \prod_{k=1}^{8} f_k(x_k) \prod_{i=1}^{7} \prod_{e\in E_i} c_{j(e),k(e)|D(e)} \left(F(x_{j(e)} \mid x_{D(e)}), F(x_{k(e)} \mid x_{D(e)}); \boldsymbol{\theta}_e \right), \tag{10}$$

where E_i is the edge set in tree i, and $c_{j(e),k(e)|D(e)}$ denotes the bivariate pair-copula density with parameter vector $\boldsymbol{\theta}_e$ for edge e. Here, $j(e)$ and $k(e)$ denote the conditioned nodes connected by edge e, while $D(e)$ represents the conditioning set of nodes. Furthermore, $F(x_{j(e)} \mid x_{D(e)})$ and $F(x_{k(e)} \mid x_{D(e)})$ are the conditional cumulative distribution functions (often referred to as h-functions). This decomposition fundamentally expresses joint density as a product of marginal densities and sequential bivariate conditional copulas. The vine hierarchy is constructed sequentially. At each level, a maximum spanning tree (MST) algorithm is applied to select pairs with the strongest dependence by maximizing the absolute empirical Kendall's Tau ($|\hat{\tau}|$) as in Eq (11):

$$T_l = \arg\max_{T=(N,E)} \sum_{e=\{i,j\}\in E} |\hat{\tau}_{i,j}|. \tag{11}$$

For each selected edge across the 7 trees, various copula families are evaluated. The optimal copula family for each pair is selected based on the AIC. Subsequently, the parameters $\boldsymbol{\theta}$ are estimated sequentially, tree-by-tree, using maximum pseudo-likelihood estimation (MPLE) via the *VineCopula* package in R.

4) Model diagnostics and interpretation: Model adequacy is evaluated via goodness-of-Fit (GOF) tests comparing simulated and observed dependencies. Graphical tools (contour and scatter plots) are employed to verify the accurate capture of both bulk and tail dependencies. Finally, a baseline comparison against an independent copula confirms the statistical significance of the proposed model.

3. Results and discussion

3.1. Simulation study

Evaluating Type I error control (α = 0.05) and empirical power revealed that under the gamma (Figure 3), Weibull (Figure 4), and GE (Figure 5) null hypotheses, only the AD and KS tests maintained the Type I error rate at zero; the AIC and BIC criteria failed to control α. Regarding power, the AD test proved superior to the KS test against a log-normal alternative hypothesis. The tests struggled to differentiate among the gamma, Weibull, and GE distributions due to highly similar PDFs (Figure S1). Conversely, the log-normal distribution (Figure 6) was easily distinguishable. Ultimately, while AIC and BIC effectively discriminate between distributions, they are statistically inadequate due to their failure to control Type I errors. Additionally, none of the studied distributions can be successfully transformed into a standard normal distribution when converted to the SPI (Table 3).

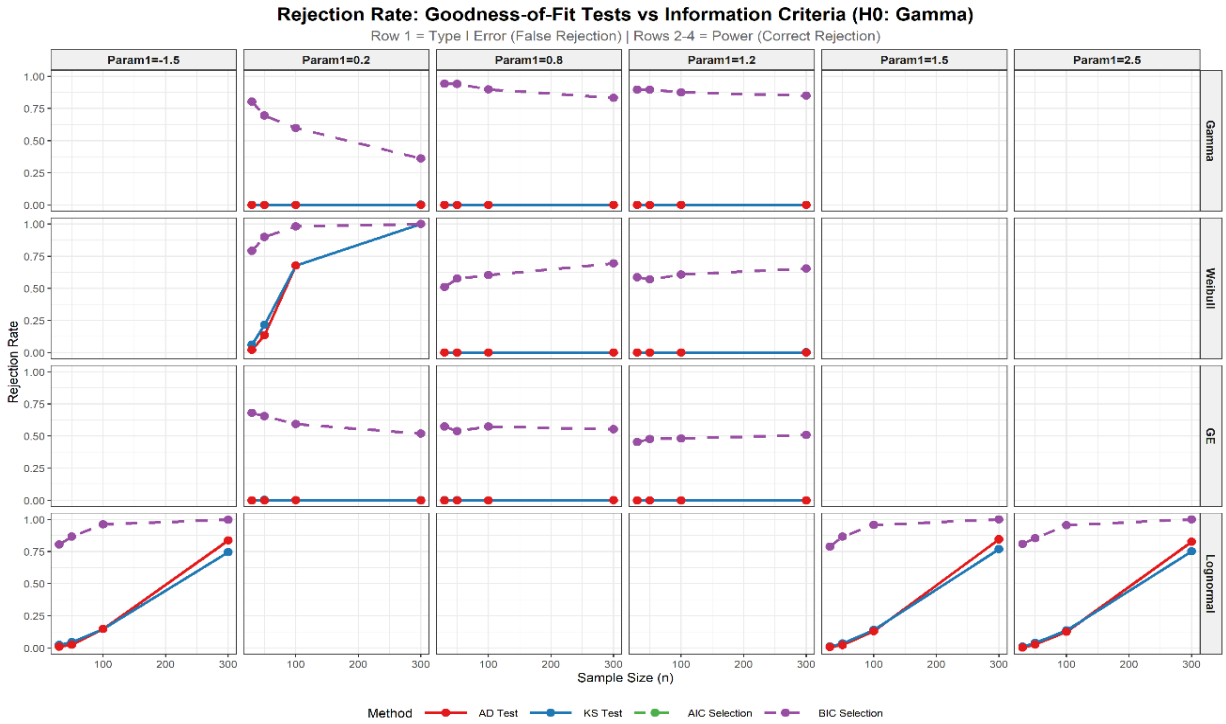


Figure 3. Empirical Type I error rates (α = 0.05) and statistical power of the AD and KS tests, compared with AIC and BIC criteria, under the gamma null distribution.

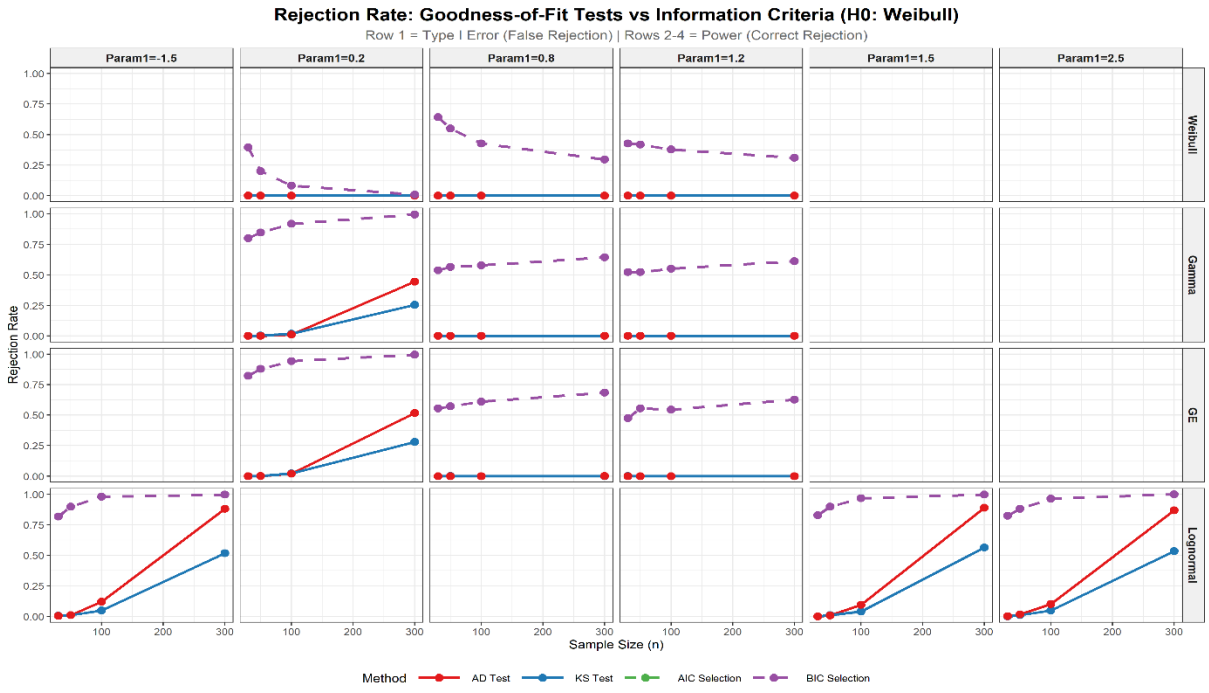


Figure 4. Empirical Type I error rates (α = 0.05) and statistical power of the AD and KS tests, compared with AIC and BIC criteria, under the Weibull null distribution.

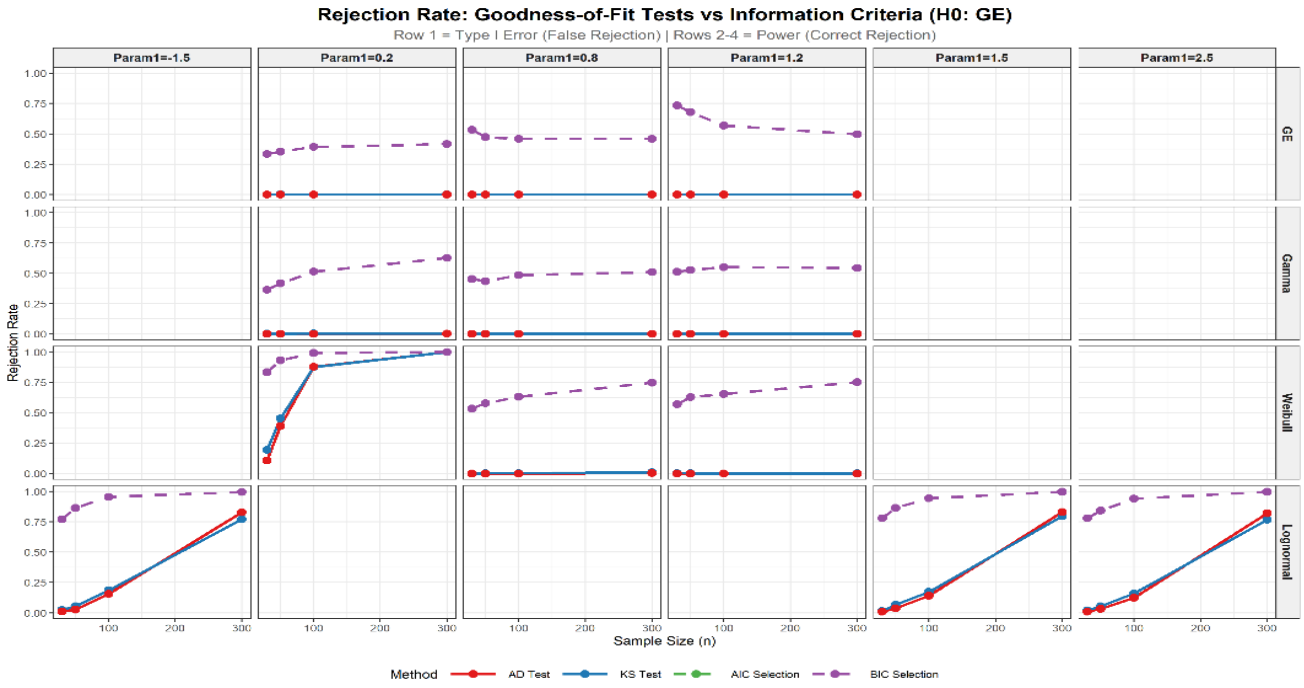


Figure 5. Empirical Type I error rates (α = 0.05) and statistical power of the AD and KS tests, compared with AIC and BIC criteria, under the GE null distribution.

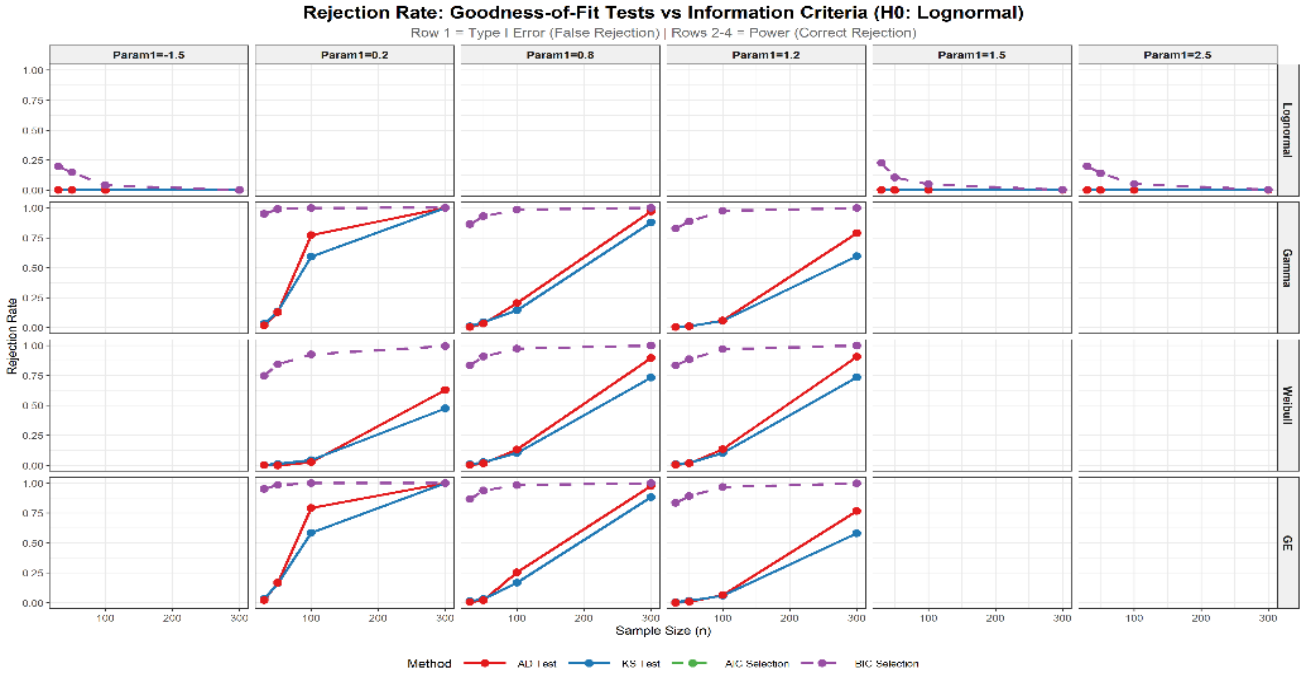


Figure 6. Empirical Type I error rates (α = 0.05) and statistical power of the AD and KS tests, compared with AIC and BIC criteria, under the log-normal null distribution.

Table 3. AD test rejection rates (α = 0.05) for standard normality of the SPI-transformed data.

n	*Shape*	Gamma	Weibull	GE	*Location*	Log-normal
30	0.2	0.99	1.00	0.99	-1.5	0.54
30	0.8	0.25	0.40	0.22	1.5	0.51
30	1.2	0.08	0.03	0.08	2.5	0.49
50	0.2	1.00	1.00	1.00	-1.5	0.89
50	0.8	0.72	0.89	0.67	1.5	0.90
50	1.2	0.35	0.19	0.35	2.5	0.89
100	0.2	1.00	1.00	1.00	-1.5	1.00
100	0.8	1.00	1.00	1.00	1.5	1.00
100	1.2	0.94	0.78	0.94	2.5	1.00
300	0.2	1.00	1.00	1.00	-1.5	1.00
300	0.8	1.00	1.00	1.00	1.5	1.00
300	1.2	1.00	1.00	1.00	2.5	1.00

Note: All distributions were simulated with a *scale parameter* of 1.

3.2. Rainfall data study

3.2.1. Optimal probability distribution selection and comparative analysis of SPI and API for drought assessment

Descriptive statistics (Table 4) confirm that summer rainfall series at all locations are universally characterized by right-skewed distributions (skewness: 0.35–0.84; Supplementary Figure S2). Notably, leptokurtic behavior (kurtosis > 3) was identified at Chiang Mai, Chiang Rai, Nan, and Mae Hong Son. Regarding precipitation magnitude, Chiang Rai recorded the highest mean (357.6 mm) and maximum (787.0 mm) rainfall, whereas Lamphun registered the lowest minimum (32.5 mm). Furthermore, the Ljung–Box test indicates that the rainfall data at each station are independent and identically distributed (*i.i.d.*) (*p*-value > 0.05) at the 0.05 significance level.

Table 4. Statistics of summer rainfall at eight meteorological stations.

Station	**Min**	**Max**	**Median**	**Mean**	**SD**	**Skewness**	**Kurtosis**	**Ljung–Box test**	***p*-value**
Chiang Mai	42.3	673.6	243.5	262.8	139.6	0.84	3.56	3.06	0.216
Chiang Rai	120.7	787.0	321.3	357.6	164.3	0.81	3.14	2.68	0.261
Lampang	65.7	563.2	246.8	279.6	133.0	0.52	2.36	2.6	0.265
Lamphun	32.5	527.5	220.7	245.7	121.1	0.45	2.45	1.3	0.520
Mae Hong Son	100.1	536.3	224.6	249.9	100.3	0.74	3.34	1.03	0.596
Nan	92.3	646.8	281.9	312.5	122.6	0.77	3.55	2.82	0.243
Phrae	69.5	547.0	275.9	298.9	116.5	0.35	2.53	3.79	0.151
Phayao	64.7	598.1	295.3	310.2	133.8	0.35	2.46	1.88	0.391

Note: mm denotes millimeters and SD denotes standard deviation.

For station-specific model selection, Dey and Kundu [22] recommend identifying the distribution with the maximum log-likelihood alongside goodness-of-fit (GOF) testing. Given equal parameter counts among candidate distributions, AIC and BIC are also applicable. While models are ideally selected when GOF and information criteria converge, the GOF statistic serves as the primary criterion during divergence due to its superior statistical power and lower sensitivity to skewness and outliers. Consequently, the AD test is employed due to its proven Type I error control and higher statistical power than the KS test (Section 3.1). Table 5 indicates that the gamma distribution provides the best fit for the Chiang Mai, Lampang, Mae Hong Son, Nan, and Phrae stations. The Weibull distribution is optimal for Lamphun and Phayao, while the generalized exponential distribution is most suitable for Chiang Rai. These findings are visually corroborated in Figure 7, which overlays the theoretical densities onto the rainfall histograms.

The comparison of SPI and API for detecting drought conditions is shown in Figure 8, and the criteria of API for each rain gauge station are summarized in Table 6. The results show that extremely dry conditions were detected at Chiang Mai in 1991 and 2010 by both API and SPI, and at the Chiang Rai station in 1992 and 2019 by API. Lampang, Lamphun, and Phayao experienced extremely dry conditions in 1992 by both methods. Nan recorded extremely dry conditions in 2012 and 2023, while Phrae experienced extremely dry conditions in 2010, as indicated by both API and SPI. Furthermore, severely dry conditions were observed at Chiang Rai in 1992 and 2019 based on SPI and in 2020 based on API. Lamphun experienced severely dry conditions in 2010, while Mae Hong Son recorded such conditions in 1992 and 2016 by both API and SPI. Phrae and Phayao exhibited severely dry conditions in 2016 and 2019, respectively, by each method. However, the API demonstrated greater sensitivity in detecting these extremes. The comparison of the number of years classified into each drought condition by SPI and API is shown in Figure 9.

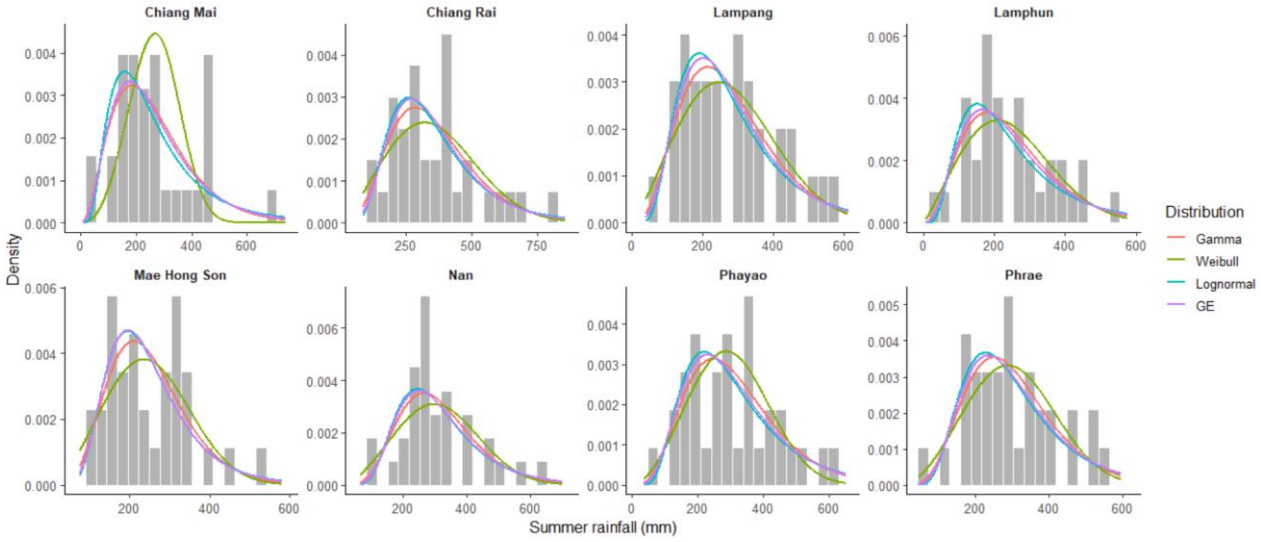


Figure 7. Histograms and theoretical density curves of summer rainfall for eight rain gauge stations.

Table 5. Parameter estimates and model selection criteria for the selected distributions.

Station	Distribution	Parameter estimate					AD Test	*p*-value	AIC	BIC	Selected
		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}$					
Chiang Mai	Gamma	3.4337	0.0131				**0.3609**	**0.8857**	**417.761**	**420.754**	**Gamma**
	Weibull	2.0173	297.0892				0.4306	0.8168	417.992	420.985	
	log-normal			5.4186	0.5919		0.5325	0.7126	420.671	423.664	
	GE	3.9993				0.0079	0.3806	0.8669	418.321	421.314	
Chiang Rai	Gamma	5.0248	0.0141				0.2235	0.9826	427.805	430.798	**GE**
	Weibull	2.3627	404.8109				0.3800	0.8674	429.873	432.866	
	log-normal			5.7766	0.4614		0.2219	0.9832	427.860	430.853	
	GE	7.2484				0.0073	**0.2067**	**0.9885**	**427.740**	**430.733**	
Lampang	Gamma	4.3589	0.0156				**0.2333**	**0.9785**	**415.523**	**418.516**	**Gamma**
	Weibull	2.2958	316.6560				0.2978	0.9394	416.021	419.014	
	log-normal			5.5141	0.5068		0.2702	0.9585	416.731	419.724	
	GE	5.6218				0.0085	0.2341	0.9754	416.004	418.997	
Lamphun	Gamma	3.6616	0.0149				0.2048	0.9891	411.664	414.657	**Weibull**
	Weibull	2.1820	277.6150				**0.2002**	**0.9904**	**410.516**	**413.509**	
	log-normal			5.3613	0.5826		0.4319	0.8155	415.845	418.838	
	GE	4.1857				0.0086	0.2572	0.9663	412.748	415.741	
Mae Hong Son	Gamma	6.5524	0.0262				**0.3720**	**0.8752**	396.508	399.501	**Gamma**
	Weibull	2.6959	281.4758				0.4496	0.7973	398.880	401.873	
	log-normal			5.4427	0.4002		0.3848	0.8628	**396.434**	**399.427**	
	GE	10.9041				0.0122	0.3795	0.8679	396.494	399.487	
Nan	Gamma	6.5888	0.0211				**0.4615**	**0.7852**	**411.130**	**414.123**	**Gamma**
	Weibull	2.7316	351.2099				0.6865	0.569	412.787	415.780	
	log-normal			5.6666	0.4094		0.5467	0.6984	412.706	415.699	
	GE	9.8651				0.0093	0.5183	0.7269	412.085	415.078	
Phrae	Gamma	6.1023	0.0204				**0.2216**	**0.9834**	410.403	413.396	**Gamma**
	Weibull	2.8166	335.8544				0.2759	0.9548	**409.619**	**412.612**	
	log-normal			5.6158	0.4348		0.3724	0.8747	413.328	416.321	
	GE	8.6037				0.0091	0.318	0.9235	412.148	415.141	
Phayao	Gamma	4.9377	0.0159				0.1621	0.9976	418.920	421.913	**Weibull**
	Weibull	2.5394	350.0715				**0.1335**	**0.9995**	**418.022**	**421.015**	
	log-normal			5.6326	0.4861		0.3331	0.9108	421.797	424.790	
	GE	6.2346				0.0079	0.2359	0.9773	420.268	423.261	

Table 6. API classification thresholds (mm) for each rain gauge station.

Station	Extremely dry	Severely dry	Moderately dry	Mildly dry	Mildly wet	Moderately wet	Severely wet	Extremely wet
Chiang Mai	0.00–60.00	60.01–89.13	89.14–127.55	127.56–236.88	236.89–396.45	396.46–497.21	497.23–612.53	≥ 612.54
Chiang Rai	0.00–125.14	125.15–161.00	161.01–205.18	205.19–325.65	325.66–506.35	506.36–627.59	627.60–773.56	≥ 773.57
Lampang	0.00–79.58	79.59–111.11	111.12–150.89	150.90–258.41	258.42–408.25	408.26–500.64	500.65–605.05	≥ 605.06
Lamphun	0.00–49.25	49.26–81.50	81.51–124.07	124.08–234.65	234.66–367.28	367.29–438.29	438.30–510.97	≥ 510.98
Mae Hong Son	0.00–94.78	94.79–122.19	122.20–154.87	154.88–237.40	237.41–345.22	345.23–409.40	409.41–480.56	≥ 480.57
Nan	0.00–118.54	118.55–152.70	152.71–193.41	193.42–296.17	296.18–430.32	430.33–510.14	510.15–598.63	≥ 598.64
Phrae	0.00–108.49	108.50–141.57	141.58–181.34	181.35–282.84	282.85–416.80	416.81–496.99	497.00–586.19	≥ 586.20
Phayao	0.00–88.22	88.23–130.21	130.22–180.20	180.21– 294.91	294.92–416.99	416.98–478.02	478.03–538.22	≥ 538.23

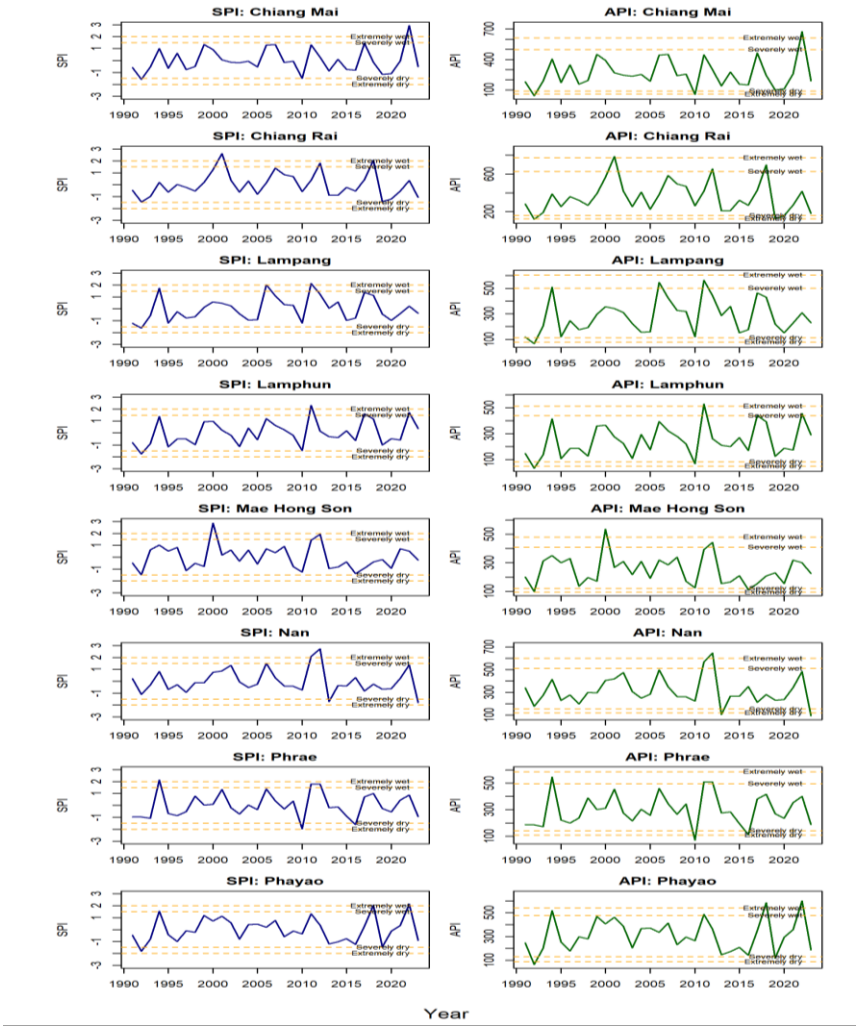


Figure 8. Comparison of SPI and API for eight rain gauge stations.

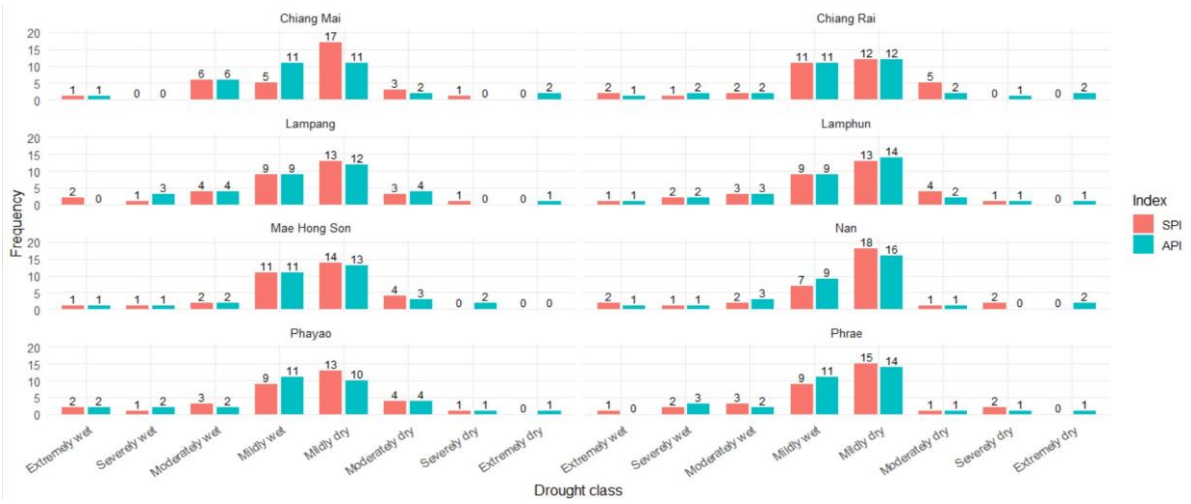


Figure 9. Comparison of the number of years detected in each drought class between SPI and API for eight rain gauge stations.

Although Cohen's Kappa analysis indicates significant agreement between the SPI and API at the 0.05 level (Table S1), the SPI is less effective at capturing tail-end drought events. As illustrated in Figure 10, its normal transformation is sensitive to the skewness and outliers inherent in the rainfall data. Consequently, the API is more appropriate than the SPI because it maintains the same units as the original data for easier interpretation and utilizes quantile transformations derived from the selected optimal distributions.

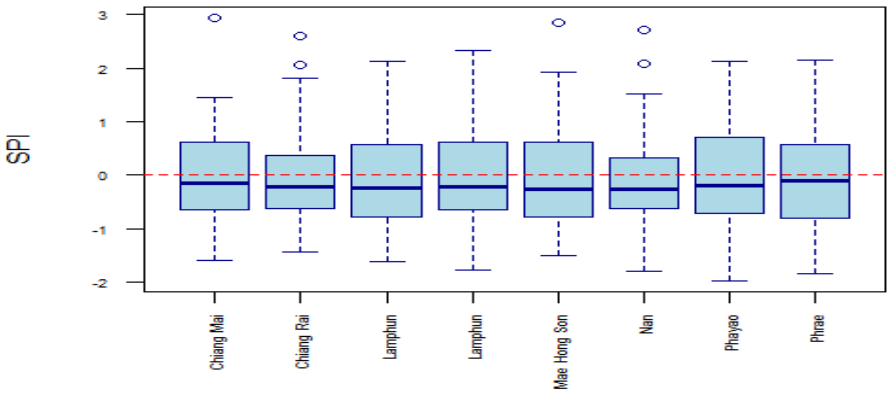


Figure 10. Boxplot of SPI for eight rain gauge stations.

3.2.2. Modeling complex inter-station rainfall dependencies via a flexible R-vine copula framework

For clarity and relevance, this study presents only the primary and first conditional tree levels (Trees 1 and 2) of the 7-tree R-vine framework (Table 7 and Figure 11). The remaining higher-order trees generally exhibit negligible dependence ($\tau < 0.15$) or utilize independence copulas, which reflects conditional independence among the locations after accounting for underlying geographical influences.

1) Primary spatial dependence (Figure 11a): The results identify Lampang (Station 3) as the central "Hub" of the region, directly connecting to Chiang Mai (1), Lamphun (4), and Phrae (7). This position indicates that rainfall in Lampang serves as a primary proxy for the region; heavy precipitation in Lampang strongly suggests concurrent heavy rainfall in these neighboring provinces. Analysis of the strongest edges reveals that the Lampang–Lamphun pair (3,4) exhibits the highest correlation ($\tau = 0.58$), reflecting their geographical proximity and shared monsoonal influence within the same river basin. This is followed by the Lampang–Chiang Mai pair (3,1) ($\tau = 0.55$), highlighting the synchronized climate patterns of these key economic centers. Critically, the tail dependence analysis shows that the Phrae–Lampang pair (7,3) reaches a critical threshold with an upper tail dependence of 0.73. This value implies a 73% probability that extreme rainfall will occur simultaneously in both provinces, which is vital for basin-level flood management. Furthermore, the high lower tail dependence for pairs (3,4) and (3,1), ranging from 0.63 to 0.65, indicates a strong tendency for these provinces to experience synchronized drought conditions.

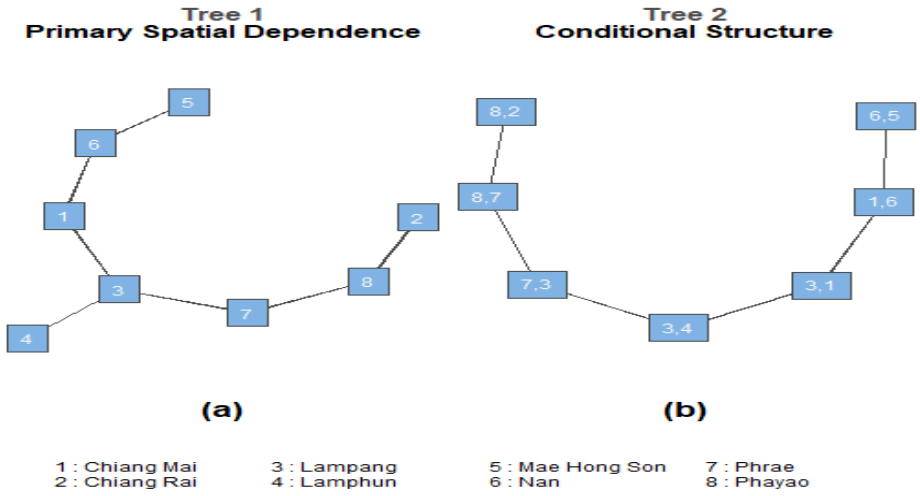


Figure 11. R-vine copula structure for rainfall dependence (a) Tree 1: Primary spatial dependence. (b) Tree 2: Conditional dependence structure after filtering primary influences.

2) Conditional dependence structure (Figure 11b): This stage involves analyzing the deep-seated relationships (conditional dependence) remaining after removing the common influences captured in Tree 1. The key findings regarding the latent connections are as follows:

2.1) *Residual localized factors*: The analysis of the Lamphun–Chiang Mai given Lampang (4,1|3) pair reveals a persistent correlation ($\tau = 0.34$), even after accounting for rainfall patterns in Lampang. This suggests that Chiang Mai and Lamphun share localized climatic factors such as the influence of the Thanon Thong Chai Range on cloud formation that operate independently of the region's primary monsoonal system.

2.2) *Dependence exhaustion*: Regarding the saturation of relationships, the Lampang–Nan given Chiang Mai (3,6|1) pair transitions to an independence copula (Family 0) with $\tau = 0$. This indicates that the initial correlation between Lampang and Nan is entirely explained through their mutual relationship with Chiang Mai. Consequently, no direct or unique latent

dependency exists between these two stations once the influence of Chiang Mai is partitioned out.

2.3) *Asymmetric conditional risk*: Examining conditional risk through the Joe Copula for the (4,1|3) pair reveals a significant upper tail dependence (0.57). This implies that even when controlling for Lampang, Chiang Mai and Lamphun maintain a high concurrent risk of extreme burst rainfall due to their specific localized interactions.

Overall, a comparison between the R-vine model and the independence model (which assumes no inter-station correlation) via the Vuong test (V = 8.427, p-value < 0.001) demonstrates that the R-vine model provides a superior representation of spatial rainfall dependence. Furthermore, the presence of asymmetric copula families within the structure confirms that the R-vine model is the most appropriate framework for capturing the complex, nonlinear relationships inherent in the data.

Table 7. Summary of spatial dependence structures and pair-copula parameters for the primary and first conditional tree levels (Tree 1 and Tree 2).

						Tail dependence	
Tree	**Edge**	**Copula family**	**Kendall's τ**	**Parameter 1**	**Parameter 2**	**Lower**	**Upper**
1	(6,5)	Tawn (type 2)	0.37	3.94	0.43	-	0.42
	(1,6)	Frank	0.45	4.95	0	-	-
	(3,1)	Survival Gumbel	0.55	2.2	0	0.63	-
	(3,4)	**Survival BB7**	**0.58**	**2.29**	**1.6**	**0.65**	**0.65**
	(7,3)	Survival Clayton	0.52	2.18	0	-	0.73
	(8,2)	Survival Gumbel	0.52	2.07	0	0.60	-
	(8,7)	Frank	0.52	6.04	0	-	-
2	(1,5|6)	Tawn (type 2, 180°)	0.23	3.46	0.26	0.26	-
	(3,6|1)	Independence	0	-	-	-	-
	(4,1|3)	**Joe**	**0.34**	**1.92**	**0**	**-**	**0.57**
	(7,4|3)	Tawn	0.07	20	0.07	-	0.07
	(8,3|7)	Independence	0	-	-	-	-
	(7,2|8)	Tawn	0.12	7.4	0.12	-	0.12

Model fit: Overall log-likelihood = 154.61, AIC = -239.23, BIC = -186.85.

Vuong test: V = 8.427 (p-value < 0.001), comparing R-vine against the independence model.

To further evaluate goodness-of-fit, Figure S3 illustrates the bivariate copula density for the Lampang–Lamphun pair (Stations 3 and 4). The overlay of empirical pseudo-observations on the theoretical Survival BB7 contours clearly demonstrates that the chosen asymmetric copula successfully captures the true dependence structure and the specific asymmetric tail behaviors of the extreme rainfall events. Tail dependence metrics [Upper Tail Dependence (UTD), Lower Tail Dependence (LTD)] translate directly into quantitative conditional probabilities for risk management. A high UTD (e.g., Lampang–Lamphun) quantifies the risk of concurrent extreme rainfall, enabling "Coordinated Reservoir Operations" such as preemptive water releases. Likewise, high LTD empowers authorities to enforce synchronized, basin-wide restrictions on water-intensive crops during droughts.

4. Conclusions and recommendations

This study employed the Actual Precipitation Index (API) to characterize rainfall patterns and spatial dependencies in upper northern Thailand. Based on the AD test, and AIC and BIC criteria [22], the gamma distribution was identified as the optimal fit for most stations (Chiang Mai, Lampang, Mae Hong Son, Nan, and Phrae), while the Weibull (Lamphun and Phayao) and generalized exponential (Chiang Rai) distributions provided the best fit for the remaining areas. Overall, the API and SPI indices exhibited high consistency across all stations, particularly during extreme events. In 2011, a major regional flood triggered "Extremely Wet" classifications at all stations. Conversely, during 2015-2016, all stations recorded a sharp decline to "Severely/Extremely Dry" levels, consistent with the widespread El Niño phenomenon. Regarding threshold response, API outperformed SPI in identifying wetness (upper thresholds), specifically in Chiang Mai, Nan, and Phayao, where it exhibited sharper peaks beyond the "Extremely Wet" level. Conversely, SPI showed greater sensitivity in detecting statistical drought, notably in Chiang Rai and Mae Hong Son, where it dropped below the "Extremely Dry" threshold more frequently than API. In closed basins or valleys, such as Lamphun and Phayao, API captured finer fluctuations during "Near Normal" periods. In contrast, the SPI was constrained by its normal transformation process, which is susceptible to distortions caused by right-skewed data and extreme outliers–both characteristic features of this region [20,21].

The R-vine copula analysis further elucidated the region's complex spatial dynamics: Primary Dependence (Tree 1), Lampang emerged as the regional hub, directly influencing Chiang Mai, Lamphun, and Phrae. The Lampang–Lamphun pair exhibited the strongest correlation, reflecting shared monsoonal and river basin influences. Conditional Structure (Tree 2), after controlling for primary hub influences, the Lamphun–Chiang Mai pair maintained persistent correlations driven by localized geography. Conversely, the Lampang–Nan relationship was found to be entirely mediated by Chiang Mai, transitioning to an independence copula. Notably, when controlling for Lampang, Chiang Mai and Lamphun still retain a high concurrent risk of extreme burst rainfall due to specific localized interactions. A fundamental advantage of the API over the SPI is its practical utility for local stakeholders, such as farmers and water managers. Unlike the SPI, which relies on abstract, dimensionless Z-scores, the API avoids normal transformation to directly reflect regional rainfall characteristics. Consequently, API-derived thresholds can be readily translated into tangible physical units (e.g., millimeters). This interpretability empowers decision-makers to optimize reservoir operations and irrigation schedules, thereby enhancing localized drought and flood risk management. For future spatial rainfall analysis, integrating R-vine copulas with the API is highly recommended to enhance the efficiency and accuracy of water resource management.

Author contributions

Kritdilada Luanmuang: Methodology, Formal analysis, Data curation, and Writing – original draft; Manad KhamKong: Conceptualization, Investigation, Supervision, and Writing – review and editing; Nawapon Nakharutai: Writing – review and editing; Pimwarat Srikummoon: Writing – review and editing. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

During the preparation of this work, the authors used Gemini 1.5 Flash (Google) in order to im-prove the English language, academic phrasing, and clarity of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the final version of the publication. The authors declare that the original ideas, research findings, and data analysis presented in this manuscript were generated entirely by the human authors without AI assistance.

Acknowledgments

This research was partially supported by Chiang Mai University. One of the authors, Kritdilada Luanmuang, would like to express her deepest gratitude to the Ministry of Higher Education, Science, Research and Innovation (MHESI), Thailand, for the financial support provided for this study. The authors gratefully acknowledge all institutional and financial assistance received.

Conflict of interest

The authors declare no conflicts of interest.

References

1. Office of Agricultural Economics (OAE), *Agricultural Statistics of Thailand*, Bangkok, Thailand: Ministry of Agriculture and Cooperatives, 2020. Available from: https://oae.go.th/home/article/386.
2. T. A. Räsänen, P. Someth, H. Lauri, J. Koponen, J. Sarkkula, M. Kummu, Observed river discharge changes due to hydropower operations and climate variability in the Mekong Basin, *J. Hydrol.*, **545** (2017), 28–41. https://doi.org/10.1016/j.jhydrol.2016.12.023
3. M. Khamkong, P. Bookkamana, Development of statistical models for maximum daily rainfall in upper northern region of Thailand, *Chiang Mai J. Sci.*, **42** (2015), 1044–1053.
4. T. B. McKee, N. J. Doesken, J. Kleist, The relationship of drought frequency and duration to time scales, *Proceedings of the 8th Conference on Applied Climatology*, Anaheim, CA, USA: American Meteorological Society, 1993, 179–184.
5. S. M. Vicente-Serrano, S. Beguería, J. I. López-Moreno, A multiscalar drought index sensitive to global warming: the SPEI, *J. Climate*, **23** (2010), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1
6. J. H. Stagge, L. M. Tallaksen, L. Gudmundsson, A. F. Van Loon, K. Stahl, Candidate distributions for climatological drought indices, *Int. J. Climatol.*, **35** (2015), 4027–4040. https://doi.org/10.1002/joc.4267
7. T. Chaito, M. Khamkong, Analyzing runoff variability index in northern Thailand using length-biased Weibull-Rayleigh distribution, *AIMS Math.*, **10** (2025), 14539–14559. https://doi.org/10.3934/math.2025655

8. Q. Zhang, C. Y. Xu, Z. Zhang, Observed changes of drought/wetness episodes in the Pearl River basin, China, using the standardized precipitation index and aridity index, *Theor. Appl. Climatol.*, **98** (2009), 89–99. https://doi.org/10.1007/s00704-008-0095-4
9. F. Yusof, F. Hui-Mean, J. Suhaila, Z. Yusop, K. Ching-Yee, Rainfall characterisation by application of standardised precipitation index (SPI) in Peninsular Malaysia, *Theor. Appl. Climatol.*, **115** (2014), 503–516. https://doi.org/10.1007/s00704-013-0918-9
10. J. Almedeij, Drought analysis for Kuwait using standardized precipitation index, *Sci. World J.*, **2014** (2014), 451841. https://doi.org/10.1155/2014/451841
11. Z. Şen, M. Almazroui, Actual precipitation index (API) for drought classification, *Earth Syst. Environ.*, **5** (2021), 59–70. https://doi.org/10.1007/s41748-021-00201-0
12. F. Laio, G. Di Baldassarre, A. Montanari, Model selection techniques for the frequency analysis of hydrological extremes, *Water Resour. Res.*, **45** (2009), W07416. https://doi.org/10.1029/2007WR006666
13. H. Akaike, A new look at the statistical model identification, *IEEE Trans. Autom. Control.*, **19** (1974), 716–723. https://doi.org/10.1109/TAC.1974.1100705
14. H. Zhang, L. Chen, V. P. Singh, Flood frequency analysis using generalized distributions and entropy-based model selection method, *J. Hydrol.*, **595** (2021), 125610. https://doi.org/10.1016/j.jhydrol.2020.125610
15. H. Wu, M. D. Svoboda, M. J. Hayes, D. A. Wilhite, F. Wen, Appropriate application of the standardized precipitation index in arid locations and dry seasons, *Int. J. Climatol.*, **27** (2007), 65–79. https://doi.org/10.1002/joc.1371
16. K. Luanmuang, M. Kamkhong, N. Nakharutai, P. Srikummoon, An analysis of drought in the upper northern of Thailand using actual precipitation index, *AIP Conf. Proc.*, **3123** (2024), 020012. https://doi.org/10.1063/5.0223852
17. X. Chen, Q. Shao, C. Y. Xu, J. Zhang, L. Zhang, C. Ye, Comparative study on the selection criteria for fitting flood frequency distribution models with emphasis on upper-tail behavior, *Water*, **9** (2017), 320. https://doi.org/10.3390/w9050320
18. S. Morid, V. Smakhtin, M. Moghaddasi, Comparison of seven meteorological indices for drought monitoring in Iran, *Int. J. Climatol.*, **26** (2006), 971–985. https://doi.org/10.1002/joc.1264
19. M. Abramowitz, I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, New York, NY, USA: Dover Publications, 1965.
20. T. Chaito, M. Khamkong, A modified Box and Cox power transformation to determine the standardized precipitation index, *Songklanakarin J. Sci. Technol.*, **40** (2018), 867–877. https://doi.org/10.14456/sjst-psu.2018.91
21. T. Chaito, M. Khamkong, P. Murnta, Appropriate transformation techniques to determine a modified standardized precipitation index for the Ping River in northern Thailand, *EnvironmentAsia*, **12** (2019), 32–42. https://doi.org/10.14456/ea.2019.43
22. A. K. Dey, D. Kundu, Discriminating among the log-normal, Weibull, and generalized exponential distributions, *IEEE Trans. Reliab.*, **58** (2009), 416–424. https://doi.org/10.1109/TR.2009.2019494
23. Upper Northern Region Irrigation Hydrology Center, Rainfall data. Available from: http://www.hydro-1.net/.

24. J. Dißmann, E. C. Brechmann, C. Czado, D. Kurowicka, Selecting and estimating regular vine copulae and application to financial returns, *Comput. Stat. Data Anal.*, **59** (2013), 52–69. https://doi.org/10.1016/j.csda.2012.08.010

Supplementary

Table S1. Cohen's Kappa statistics for agreement between the SPI and API based on the optimal probability distributions for each station.

Station	Model	p_0	p_e	Cohen's Kappa	SE	95% CI	z	p-value	Interpretation
Chiang Mai	Gamma	0.7575	0.2617	0.6715	0.1011	0.47-0.87	6.65	3.03×10^{-11}	Substantial
Chiang Rai	GE	0.8787	0.2599	0.8361	0.0768	0.69-0.99	10.88	1.31×10^{-27}	Almost Perfect
Lampang	Gamma	0.8787	0.2461	0.8391	0.0754	0.69-0.99	11.13	8.86×10^{-29}	Almost Perfect
Lamphun	Weibull	0.9091	0.2626	0.8766	0.0679	0.74-1.00	12.92	3.59×10^{-38}	Almost Perfect
Mae Hong Son	Gamma	0.9091	0.2948	0.8711	0.071	0.73-1.00	12.28	1.22×10^{-34}	Almost Perfect
Nan	Gamma	0.8181	0.3315	0.7279	0.1005	0.53-0.93	7.25	4.29×10^{-13}	Substantial
Phrae	Gamma	0.8787	0.2975	0.8273	0.0809	0.67-0.99	10.22	1.51×10^{-24}	Almost Perfect
Phayao	Weibull	0.8181	0.2369	0.7616	0.088	0.59-0.93	8.65	4.95×10^{-18}	Substantial

Note: Cohen's Kappa (κ) is derived from the formula $(p_0 - p_e)/(1 - p_e)$, where p_0 represents the observed agreement and p_e represents the expected chance agreement. Statistical significance is determined by the z-test.

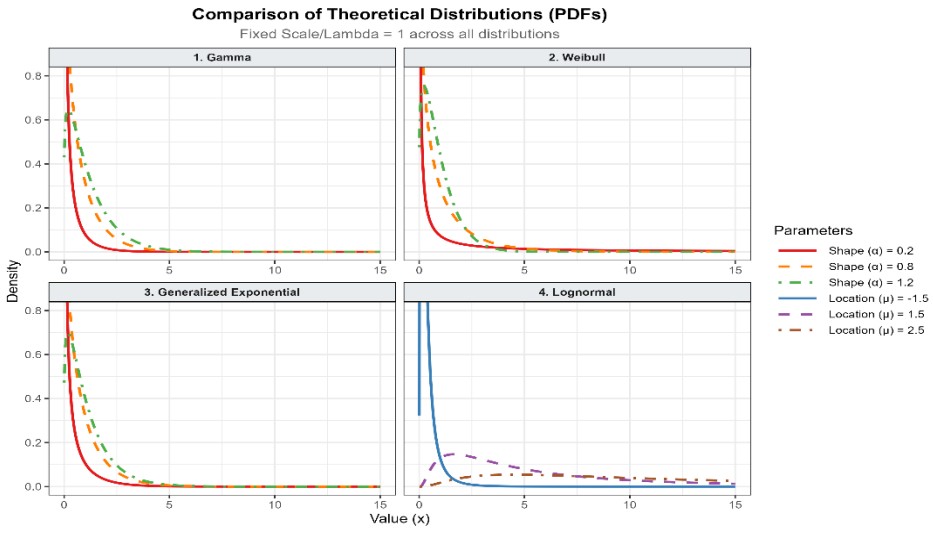


Figure S1. Theoretical probability density functions (PDFs) of the four right-skewed distributions evaluated in the simulation study.

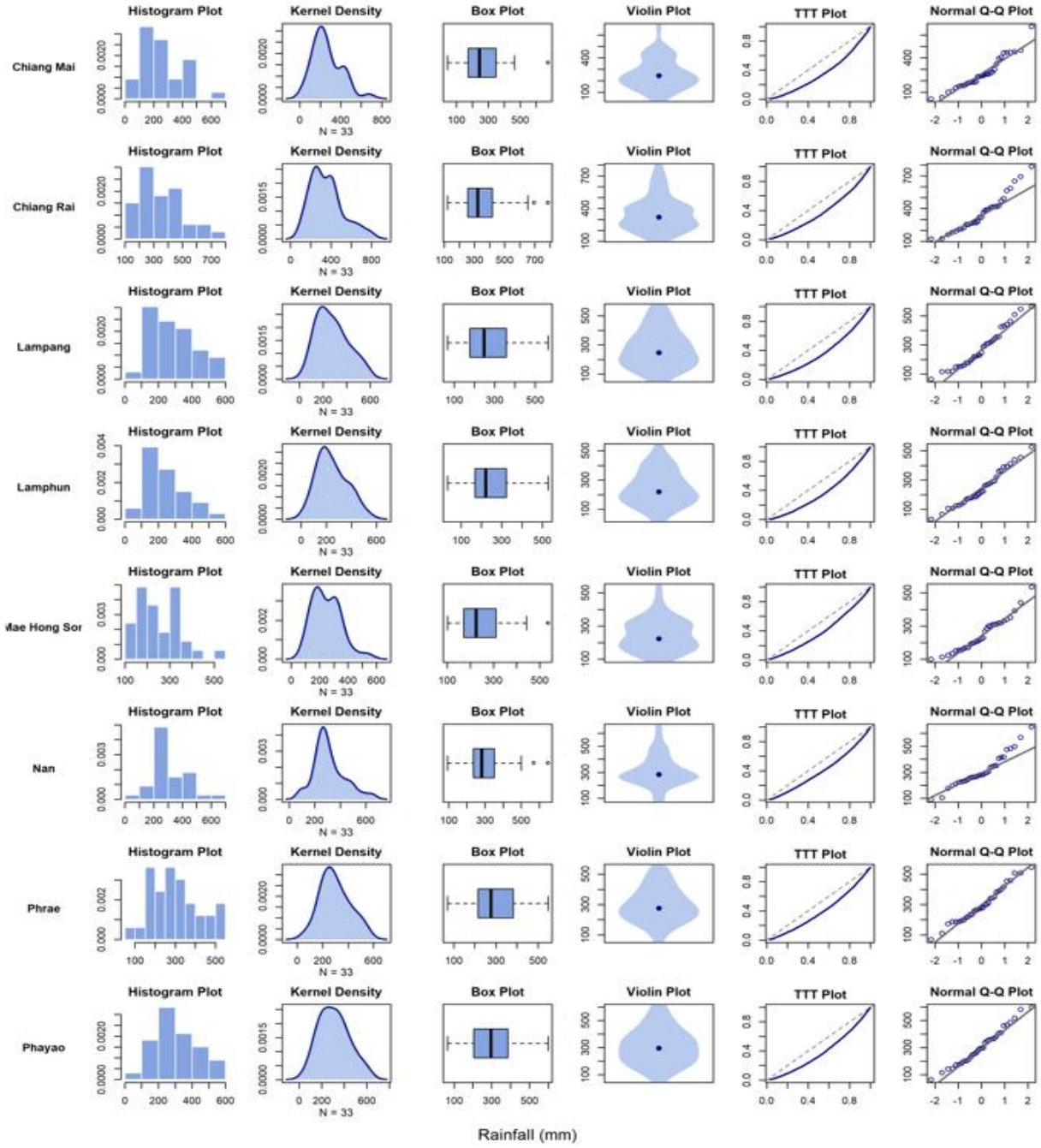


Figure S2. Nonparametric plots of summer rainfall for eight rain gauge stations.

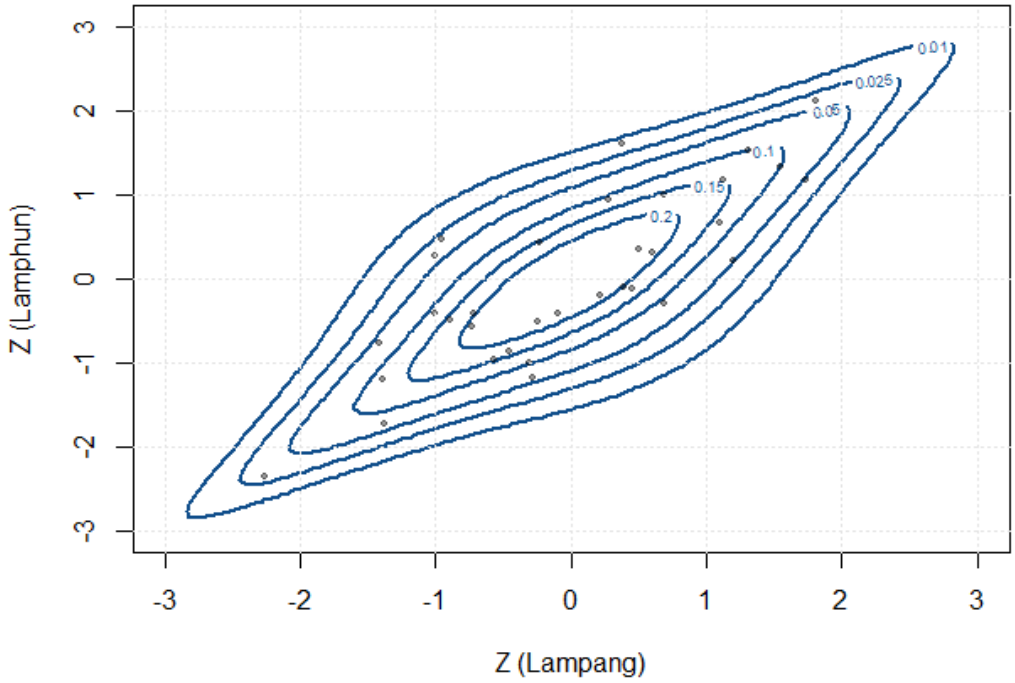


Figure S3. Goodness-of-fit diagnostic for the Lampang–Lamphun pair: Survival BB7 copula density contours overlaid with empirical observations on standard normal margins.