Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bivariate Copula Modeling of Senegal River Floods: Mathematical Foundations and a Simulation-Based Empirical Study

Diam BA

diam.ba@ucad.edu.sn

ENSETP/UCAD

Research Article

Keywords: Senegal River, Bakel station, Peaks over threshold, Gumbel–Hougaard copula, Upper-tail dependence, Conditional distribution, Joint return period

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10705874/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Bivariate Copula Modeling of Senegal River Floods: Mathematical Foundations and a Simulation-Based Empirical Study

Diam BA[1,2*]

[1*]Industrial Techniques, ENSETP/UCAD, Cheikh Anta Diop, Dakar-Fann, 5004, Dakar, Senegal.
[2]UFR SAT/UGB, LERSTAD, Sanar, 234,Saint-Louis, Senegal.

Corresponding author(s). E-mail(s): diam.ba@ucad.edu.sn;

Abstract

This article develops a mathematical framework for the bivariate analysis of flood frequency in the Senegal River recorded at the Bakel station, using peak flow and flood volume as event characteristics. It clarifies the distinction between annual maxima and peaks over threshold (POT), states the assumptions required for marginal standardization and copula identification, derives the Gumbel–Hougaard conditional laws and tail-dependence coefficient, and proves the formulas and ordering properties of AND and OR return periods. A marked-Poisson formulation converts event probabilities into return periods in years. The empirical workflow is illustrated with a reproducible synthetic sample of 100 hydrologically plausible flood events distributed over 1990–2025. Peak excesses are fitted by a Generalized Pareto Distribution (GPD) and volumes by a shifted Gamma (Pearson III) distribution. The simulated sample yields Kendall's $\widehat{\tau} = \mathbf{0.604}$, a Gumbel–Hougaard parameter $\widehat{\theta} = \mathbf{2.523}$, and upper-tail dependence $\widehat{\lambda}_U = \mathbf{0.684}$. Among four one-parameter copulas, Gumbel–Hougaard has the lowest AIC and then consequently outperforms.

Keywords: Senegal River, Bakel station, Peaks over threshold, Gumbel–Hougaard copula, Upper-tail dependence, Conditional distribution, Joint return period

1 Introduction

Flood severity is intrinsically multivariate: the instantaneous hydraulic load is related to the peak discharge Q, whereas storage, inundation duration, and cumulative stress are strongly influenced by the flood volume V [1]. A univariate analysis cannot, by itself, specify the probability that both design variables exceed prescribed levels. Copulas separate marginal behaviour from dependence and therefore provide a natural framework for this problem ([1], [2]). They also enable the flexibility to model complex dependency structures while bypassing the constraints of traditional multivariate distributions [3]. Furthermore, by isolating the tail dependence, this approach facilitates a more accurate assessment of joint risks associated with extreme hydrological events that might be otherwise underestimated by simpler models [4]. This methodology proves particularly advantageous when dealing with non-independent variables, as it allows for the derivation of conditional return periods essential for robust hydrologic design [5]. In this context, the Gumbel–Hougaard copula serves as an effective instrument for capturing the strong upper-tail dependence characteristic of extreme peak–volume pairs [6]. This selection is justified by the copula's capacity to model joint extreme behavior, as it accounts for the upper tail dependence that is critical for flood frequency analysis [7]. This selection is further validated by empirical rank-plots, which demonstrate that extreme observations frequently cluster along the upper diagonal, reflecting a strong positive correlation between discharge and volume [8]. Moreover, the application of such copula functions addresses the traditional requirement that marginal distributions share a common family, allowing for more flexible, independent modeling of individual variable characteristics [9]. Consequently, this study explores the efficacy of the Gumbel–Hougaard copula in representing the dependence structure of extreme events at the Bakel station, consistent with its established utility in capturing strong positive association between hydrologic variables [10, 11]. In particular, Archimedean copula structures are frequently employed due to their ease of construction and their capacity to represent diverse dependency patterns [12]. By enabling the decoupling of marginal selection from the choice of dependence structure, these models overcome the limitations of classical multivariate distributions that often force unrealistic homogeneity across hydrological variables [13, 14]. Specifically, such frameworks allow researchers to select optimal marginals–such as the Generalized Pareto or Gamma distributions used here–while independently specifying a copula to encapsulate the specific tail-dependence characteristics observed in the Senegal River dataset [15, 16]. Furthermore, the study addresses the non-stationary challenges posed by climatic shifts in the Senegal basin, acknowledging the emerging need to move beyond static parameterization [17]. While stationary models provide a foundational baseline, time-varying copulas are increasingly adopted to assess how evolving climatic drivers modulate the joint probability of hydrological extremes [18]. Consequently, this research builds upon these methodological advancements to provide a robust statistical assessment of flood risk at the Bakel gauge, specifically addressing the non-linear interdependencies between discharge magnitudes and total volumes [19, 20]. To evaluate the robustness of these models, we employ formal Goodness-of-Fit tests alongside the Cramér-von-Mises statistic, acknowledging that selecting a single copula requires balancing the inherent

simplification of dependence structures with the need for accurate extreme value estimation [21]. Finally, by systematically evaluating both AND and OR return periods, this study provides a comprehensive probabilistic assessment of potential dike failures or agricultural disruptions that result from the concurrence of high peak discharge and large flood volumes [22, 23]. These joint return periods offer critical insights into the reliability of hydraulic structures under combined flood properties, a necessity for informed engineering decision-making in water resources management [24], [25]. However, practitioners must remain cognizant that uncertainties inherent in both the marginal parameter estimation and the selected dependence structure can significantly propagate into final risk predictions [26]. To mitigate these effects, a sensitivity evaluation framework is required to quantify how individual parameter uncertainties and the chosen dependence structure contribute to the overall variance in predicted joint return periods [27]. Such rigorous uncertainty quantification is essential, as the choice of AND versus OR exceedance definitions can fundamentally alter the perceived risk profile for infrastructure design [28, 29]. Ultimately, integrating these joint return periods into hydraulic design ensures that infrastructure resilience is evaluated against compound hazard scenarios rather than isolated flood attributes [30],[31].

In this paper, the application is designed to represent the post-Manantali dam period $1990-2025$ corresponding L= 36 years. To this end, and in the absence of an authenticated event observational series, we construct $N = 100$ synthetic events consistent with the climatic conditions of Senegal. This gives $\eta = N/L = 2.7778$ events per year. The contribution is threefold: (i) a complete mathematical foundation; (ii) a reproducible simulation calibrated to plausible ranges of Bakel station ; and (iii) a fully consistent estimation of margins, dependence, and joint return periods. By focusing on the stochastic generation of these synthetic series, this study addresses the scarcity of continuous observational records and provides a framework to quantify environmental uncertainty in regional water management.

The remainder of the article is organized as follows. Section 2 presents the mathematical framework and statistical methods. Section 3 provides empirical results obtained through simulation and finally, Section 4 establishes the conclusion and future directions.

2 Mathematical framework and statistical methods

2.1 Flood events, POT sampling, and assumptions

Let $X(t)$, $t \in [0,T]$, denote the discharge process at the gauging station: a non-negative, continuous-time stochastic process (in practice observed at a fixed sampling interval δt, e.g. daily or sub-daily).

Let q_0 denote a fixed high threshold, expressed in the same units as $X(t)$ (m^3/s), used to define exceedances in the peaks-over-threshold (POT) framework.

A **flood event** is a maximal time interval $\mathcal{E}_i = [t_i^{start}, t_i^{end}] \subset [0,T]$ during which $X(t) > q_0$ continuously, indexed $i = 1, \ldots, n$ over the observation period, and treated

as a single hydrologically coherent episode after declustering — the merging of successive excursions above q_0 separated by less than a minimum inter-event time Δ (in the same time units as t), so that each $\mathcal{E}_i$ corresponds to one physically distinct flood.

Each event $\mathcal{E}_i$ is characterized by a bivariate mark (Q_i, V_i), where:

- $Q_i = \max_{t \in \mathcal{E}_i} X(t)$ is the **peak discharge** of event i, in m^3/s;
- $V_i = \int_{\mathcal{E}_i} \big(X(t) - q_0\big)_+ \, dt$ is the **flood volume** of event i: the volume of discharge exceeding q_0 over the event window, with $(x)_+ := \max(x, 0)$ denoting the positive part.

Units. With $X(t)$ in m^3/s and t in seconds, V_i as defined above is in m^3. To report V_i in hm^3 ($1\ \mathrm{hm}^3 = 10^6\ \mathrm{m}^3$), the conversion

$$V_i^{[hm^3]} = V_i^{[m^3]} / 10^6$$

must be applied and stated explicitly. If $X(t)$ is observed at discrete times with sampling interval δt (seconds) rather than continuously, V_i is computed as the Riemann-sum approximation:

$$V_i \approx \delta t \sum_{t \in \mathcal{E}_i} \big(X(t) - q_0\big)_+.$$

Assumption 1 (Event process) After declustering, the event times form a stationary homogeneous Poisson process with rate $\eta > 0$ events per year, and the marks (Q_i, V_i) are independent and identically distributed, independent of the event times, with continuous marginal distribution functions F_Q and F_V.

Remark 1 Assumption 1 requires diagnostics for seasonality and trend, a documented run length or hydrograph-separation rule, and tests of residual event dependence. If non-stationarity is detected, η, the margins, or the copula parameter must be allowed to depend on covariates.

Definition 1 (Conditional event-scale margins) Let Q and V denote the random variables representing, respectively, the peak discharge and the flood volume of a generic flood event extracted by the POT procedure of subsection 2.1 — i.e. Q and V are the (unindexed) random-variable versions of the event marks Q_i and V_i introduced there. Since only events satisfying $Q_i > q_0$ are retained in the extracted sample (by construction of the POT scheme), Q and V are not observed unconditionally: their distributions must be understood as conditional on membership in this extracted event population, i.e. conditional on the event $\{Q > q_0\}$, where q_0 is the same peak threshold used in subsection 2.1 to define event membership.

The **conditional event-scale margins** are the cumulative distribution functions of Q and V given this conditioning event:

$$F_Q(q) = \mathbb{P}(Q \le q \mid Q > q_0), \qquad q \ge q_0,$$
$$F_V(v) = \mathbb{P}(V \le v \mid Q > q_0), \qquad v \ge 0,$$

where F_Q is the conditional cdf of the peak discharge and F_V the conditional cdf of the flood volume — both conditioned on the *same* selection event $\{Q > q_0\}$ (peak-based membership), rather than on separate thresholding of Q and V individually. F_Q is supported on $[q_0, \infty)$ since $Q > q_0$ is guaranteed by conditioning; F_V is supported on $[0, \infty)$, as an event's volume can in principle be arbitrarily small even when its peak exceeds q_0.

Remark 2 (POT versus block maxima) The Gumbel distribution is the GEV model with shape parameter $\xi = 0$ and has an exponentially decaying right tail. Classical POT theory, however, motivates a generalized Pareto distribution (GPD) for sufficiently high threshold excesses, not a Gumbel law. A Gumbel margin may still be treated as a candidate empirical model, but its use must be justified by diagnostics and model comparison.

For the peak excess $Y = Q - q_0 > 0$, a theoretically motivated candidate is the GPD

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - (1 + \xi y/\sigma)^{-1/\xi}, & \xi \neq 0, \\ 1 - e^{-y/\sigma}, & \xi = 0, \end{cases} \qquad 1 + \xi y/\sigma > 0,\ \sigma > 0,$$

where ξ is the shape parameter and σ the scale parameter.

Alternatively, if the block-maxima framework is adopted instead — i.e. modeling the peak Q itself rather than the excess Y over q_0 — and a Gumbel candidate is retained, its CDF is

$$F_Q^{Gum}(q) = \exp\left[-\exp\left(-\frac{q - a_Q}{b_Q}\right)\right], \qquad b_Q > 0, \tag{1}$$

with location parameter a_Q and scale parameter b_Q. We write F_Q^{Gum} (rather than F_Q) to distinguish this specific parametric candidate from the general conditional margin $F_Q(q) = \mathbb{P}(Q \leq q \mid Q > q_0)$ of Definition 1.

For the companion margin F_V of Definition 1, a candidate parametric form is the shifted Gamma (Pearson type III) distribution

$$F_V^{P3}(v) = \frac{1}{\Gamma(k)} \int_0^{(v - a_V)/b_V} t^{k-1} e^{-t}\, dt, \qquad v > a_V, \quad k, b_V > 0, \tag{2}$$

with shape parameter k, scale parameter b_V, and location (shift) parameter a_V; equivalently, the unscaled integral form is

$$F_V^{P3}(v) = \left[b_V^k\, \Gamma(k)\right]^{-1} \int_{a_V}^{v} (x - a_V)^{k-1} e^{-(x - a_V)/b_V}\, dx.$$

Since Definition 1 requires $v \geq 0$, this candidate is only admissible when $a_V \geq 0$; if the fitted location parameter is negative, the model should be truncated at zero and renormalized, or an alternative candidate (e.g. GPD on V, by analogy with Q) considered instead.

2.2 Copulas and probability integral transforms

Definition 2 (Bivariate copula) A bivariate copula is a distribution function $C : [0,1]^2 \to [0,1]$ with uniform margins. In particular, $C(u, 0) = C(0, v) = 0$, $C(u, 1) = u$, $C(1, v) = v$, and C is two-increasing.
(See Nelsen (2006) [32], for more details.)

Theorem 2 (Sklar) *For any bivariate distribution function H with margins F_1 and F_2, there exists a copula C such that*

$$H(u, v) = C\big(F_1(u), F_2(v)\big). \tag{3}$$

If both margins are continuous, C is unique on $[0,1]^2$. Conversely, any copula and any two univariate distribution functions define a bivariate distribution through (3).

Proof (See Sklar (1959), or Nelsen (2006) [32]). □

In the application below, H is taken to be the joint conditional distribution of (Q, V) given $Q > q_0$, with margins F_Q and F_V as given in Definition 1; Theorem 2 then guarantees that the dependence between peak and volume can be represented separately from their marginal behavior, via a copula C. Let's adopt the proof of the theorem in our context.

Let $F_Q^{-1}(u) = \inf\{q : F_Q(q) \geq u\}$ and define F_V^{-1} similarly. On $\mathrm{Range}(F_Q) \times \mathrm{Range}(F_V)$, set $C(u, v) = H\big(F_Q^{-1}(u), F_V^{-1}(v)\big)$; this is then extended to all of $[0, 1]^2$ by the standard boundary convention $C(u, 0) = C(0, v) = 0$, $C(u, 1) = u$, $C(1, v) = v$ at the endpoints outside the margins' ranges. The distribution properties of H imply groundedness, uniform margins, and two-increasingness of C on this domain, and substitution recovers (3). When both margins are continuous, $\mathrm{Range}(F_Q)$ and $\mathrm{Range}(F_V)$ are dense in $[0, 1]$, and continuity of C forces the extension to be unique on the closure $[0, 1]^2$. Conversely, given any copula C and univariate CDFs F_Q, F_V, the composition $H(q, v) = C(F_Q(q), F_V(v))$ inherits groundedness and two-increasingness from C, and has margins F_Q, F_V since C has uniform margins.

Remark 3 (Probability integral transform.) Under continuity of the margins, $U = F_Q(Q)$ and $W = F_V(V)$ are uniform on $(0, 1)$ and have joint CDF C. Indeed, for $u \in (0, 1)$, continuity gives $\mathbb{P}(U \leq u) = \mathbb{P}(Q \leq F_Q^{-1}(u)) = u$; similarly for W. Since

$$\mathbb{P}(U \leq u, W \leq w) = \mathbb{P}\big(Q \leq F_Q^{-1}(u), V \leq F_V^{-1}(w)\big) = H\big(F_Q^{-1}(u), F_V^{-1}(w)\big),$$

equation (3) gives $\mathbb{P}(U \leq u, W \leq w) = C\big(F_Q(F_Q^{-1}(u)), F_V(F_V^{-1}(w))\big) = C(u, w)$.

2.3 The Gumbel–Hougaard model

For $\theta \geq 1$, define

$$C_\theta(u, w) = \exp\Big[-\big\{(-\log u)^\theta + (-\log w)^\theta\big\}^{1/\theta}\Big]. \tag{4}$$

At $\theta = 1$ this is the independence copula uw; as $\theta \to \infty$, it converges to the comonotonic copula $\min(u, w)$.

The Gumbel–Hougaard family is **Archimedean**: it can be written $C_\theta(u, w) = \varphi^{-1}\big(\varphi(u) + \varphi(w)\big)$ for a generator $\varphi : [0, 1] \to [0, \infty)$ that is continuous, strictly decreasing, convex, and satisfies $\varphi(1) = 0$. Here $\varphi(u) = (-\log u)^\theta$, with inverse $\varphi^{-1}(t) = \exp(-t^{1/\theta})$; substituting recovers (4).

The dependence between U and W under C_θ is summarized by three standard coefficients: **Kendall's tau**,

$$\tau = \mathbb{P}\big[(U_1 - U_2)(W_1 - W_2) > 0\big] - \mathbb{P}\big[(U_1 - U_2)(W_1 - W_2) < 0\big]$$

for independent copies $(U_1, W_1), (U_2, W_2) \sim C_\theta$; the **upper tail dependence coefficient**,

$$\lambda_U = \lim_{t\uparrow 1} \mathbb{P}(W > t \mid U > t) = \lim_{t\uparrow 1} \frac{1 - 2t + C_\theta(t,t)}{1-t};$$

and the **lower tail dependence coefficient**,

$$\lambda_L = \lim_{t\downarrow 0} \mathbb{P}(W \le t \mid U \le t) = \lim_{t\downarrow 0} \frac{C_\theta(t,t)}{t}.$$

Proposition 3 (Dependence coefficients) *For the Gumbel–Hougaard copula,*

$$\tau = 1 - \frac{1}{\theta}, \qquad \lambda_U = 2 - 2^{1/\theta}, \qquad \lambda_L = 0.$$

Consequently, inversion of Kendall's tau gives $\hat{\theta} = 1/(1-\hat{\tau})$ *whenever* $0 \le \hat{\tau} < 1$.

Proof (See Genest & MacKay (1986), [33]). □

The Archimedean generator is $\varphi(u) = (-\log u)^\theta$. The standard identity for Archimedean copulas ,

$$\tau = 1 + 4\int_0^1 \frac{\varphi(u)}{\varphi'(u)}\, du$$

yields $\varphi(u)/\varphi'(u) = u\log(u)/\theta$, hence $\tau = 1 + (4/\theta)\int_0^1 u\log u\, du = 1 - 1/\theta$, using $\int_0^1 u\log u\, du = -1/4$.

Furthermore, $C_\theta(t,t) = t^{2^{1/\theta}}$. Therefore

$$\lambda_U = \lim_{t\uparrow 1} \frac{1 - 2t + C_\theta(t,t)}{1-t} = 2 - 2^{1/\theta}$$

by L'Hôpital's rule. Finally, $C_\theta(t,t)/t = t^{2^{1/\theta}-1} \to 0$ as $t \downarrow 0$ for finite $\theta \ge 1$, giving $\lambda_L = 0$; the independence boundary $\theta = 1$ gives the same result.

For the synthetic empirical estimate $\widehat{\tau} = 0.6036$, Proposition 3 gives $\widehat{\theta} = 2.5229$ and $\widehat{\lambda}_U \approx 0.6838$. The estimate of θ should not be used as the sole model-selection criterion: Gumbel–Hougaard must be compared with plausible alternatives using likelihood criteria, goodness-of-fit tests, and upper-tail diagnostics.

2.4 Conditional distributions: Conditioning on a volume

Suppose C is continuously differentiable and V has a positive density at v. With $u = F_Q(q)$ and $w = F_V(v)$,

$$\mathbb{P}(Q \le q \mid V = v) = \partial_2 C(u, w), \tag{5}$$

where $\partial_2 C(u,w) := \partial C(u,w)/\partial w$ denotes the partial derivative of C with respect to its second argument, evaluated at (u, w).

For (4), putting $A = (-\log u)^\theta + (-\log w)^\theta$,

$$\partial_2 C_\theta(u,w) = C_\theta(u,w)\, A^{1/\theta-1}\, \frac{(-\log w)^{\theta-1}}{w}. \tag{6}$$

Indeed, for small $h > 0$, the ratio $\mathbb{P}(Q \le q, v < V \le v+h)/\mathbb{P}(v < V \le v+h)$ equals

$$\frac{C(u, F_V(v+h)) - C(u, F_V(v))}{F_V(v+h) - F_V(v)}.$$

Letting $h \downarrow 0$ gives (5). Differentiating (4) gives (6).
We call

$$\mathbb{P}(Q > q \mid V > v) = \frac{1 - u - w + C(u,w)}{1-w}, \qquad w < 1, \tag{7}$$

the **conditional exceedance probability** of Q given $V > v$.

2.5 Joint return periods under a marked-Poisson model

For thresholds (q, v), define the sets of risks

$$D_{AND} = \{Q > q, V > v\}, \qquad D_{OR} = \{Q > q \text{ or } V > v\}.$$

Theorem 4 (Return period of a set of risks) *Under Assumption 1 — recalling that η denotes the mean annual rate of the marked-Poisson event process — if $p_D = \mathbb{P}((Q,V) \in D) > 0$, then dangerous events form a Poisson process with rate ηp_D. Their mean waiting time is*

$$T_D = \frac{1}{\eta p_D} = \frac{\mu}{p_D}, \qquad \mu = \frac{1}{\eta}, \tag{8}$$

where μ denotes the mean inter-event time of the underlying (unmarked) Poisson process — i.e. the mean time between successive flood events regardless of their marks.

The probability of at least one dangerous event during a horizon of h years is

$$R_D(h) = 1 - \exp(-h/T_D). \tag{9}$$

Proof Independent marking and Poisson thinning imply that events whose marks lie in D form a Poisson process of rate ηp_D. Its waiting time is exponential with mean $(\eta p_D)^{-1}$, proving (8). The probability of zero events in h years is $\exp(-\eta p_D h)$, which gives (9). □

Corollary 5 (AND and OR formulas) With $u = F_Q(q)$ and $w = F_V(v)$,

$$T_{\text{AND}}(q,v) = \frac{1}{\eta\{1 - u - w + C(u,w)\}}, \tag{10}$$

$$T_{\text{OR}}(q,v) = \frac{1}{\eta\{1 - C(u,w)\}}. \tag{11}$$

Proof Inclusion–exclusion gives $\mathbb{P}(Q > q, V > v) = 1 - u - w + C(u,w)$. Also, $\mathbb{P}(Q \le q, V \le v) = C(u,w)$, so the complement has probability $1 - C(u,w)$. Apply Theorem 4. □

Define event-scale univariate return periods $T_Q = [\eta(1-u)]^{-1}$ and $T_V = [\eta(1-w)]^{-1}$. Correspondingly, write $p_Q = \mathbb{P}(Q > q) = 1-u$, $p_V = \mathbb{P}(V > v) = 1-w$, $p_{AND} = \mathbb{P}((Q,V) \in D_{AND})$, and $p_{OR} = \mathbb{P}((Q,V) \in D_{OR})$ for the corresponding exceedance probabilities, so that each return period above is $T = 1/(\eta p)$ for its respective p.

Corollary 6 (Ordering and diagnostic inequalities) For all nondegenerate thresholds,

$$T_{\mathrm{OR}} \le \min(T_Q, T_V) \le \max(T_Q, T_V) \le T_{\mathrm{AND}}. \tag{12}$$

Proof Because $D_{AND} \subseteq \{Q > q\}, \{V > v\} \subseteq D_{OR}$, their probabilities satisfy $p_{AND} \le \min(p_Q, p_V) \le \max(p_Q, p_V) \le p_{OR}$. Since $p \mapsto 1/(\eta p)$ is strictly decreasing in p, taking reciprocals reverses each inequality; multiplying by $1/\eta$ then proves (12). □

These return periods correspond to different sets of risks; neither is intrinsically "the" bivariate return period. The set must be chosen from the failure mechanism of the structure being studied.

3 Simulation-based empirical results

3.1 Reproducible simulation design

The empirical illustration uses exactly 100 synthetic POT events distributed across the 36 hydrological years 1990–2025. Twenty-eight years contain three events and eight years contain two. Event dates fall within the broad flood-season window from late July to early November. Dependence is generated from a Gumbel–Hougaard copula with simulation parameter $\theta_0 = 2.5$. Conditional on event membership,

$$Q - q_0 \sim \mathrm{GPD}(\xi_0 = -0.12, \sigma_0 = 850), \qquad V - 1150 \sim \mathrm{Gamma}(k_0 = 7.2, b_0 = 410).$$

Durations are obtained from the volume–peak relation

$$D_i \approx \frac{10^6 V_i}{86400\, S_i Q_i}, \qquad S_i \in [0.58, 0.86],$$

where S_i is a dimensionless hydrograph shape factor. This construction preserves the threshold, plausible physical ranges, positive peak–volume dependence, and the approximate relation between peak, volume, and duration as shown in Figure 1 . It does not reconstruct historical floods.

3.2 Descriptive statistics and marginal fitting

Table 1 summarises the simulated event marks. Every peak exceeds $q_0 = 2400\ \mathrm{m^3\,s^{-1}}$. The range of durations, 13–29 days, is consistent with the imposed hydrograph relation.

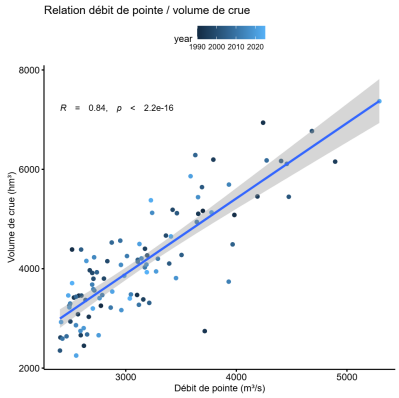


Fig. 1: Peak discharges and flood volumes.

Table 1: Descriptive statistics for the 100 synthetic events.

Variable	Mean	Std. dev.	Minimum	Maximum
Q ($\mathrm{m^3\,s^{-1}}$)	3133.9	620.8	2405.2	5292.8
V ($\mathrm{hm^3}$)	4101.1	1120.7	2252.6	7370.5
D (days)	20.9	3.5	13	29

Maximum-likelihood estimation gives

$$\widehat{\xi} = -0.2215, \qquad \widehat{\sigma} = 898.06$$

for $Q - q_0$, and

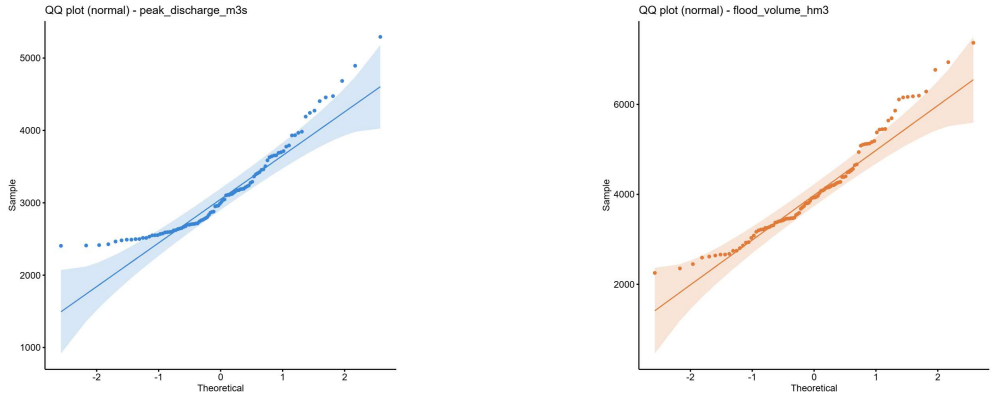
$$\widehat{k} = 3.5367, \qquad \widehat{a}_V = 1953.18, \qquad \widehat{b}_V = 607.34$$

for the shifted Gamma/Pearson III volume margin. Because parameters are estimated, goodness-of-fit is evaluated by a 400-replicate parametric bootstrap rather than by ordinary Kolmogorov–Smirnov critical values. The resulting statistics are $D_Q = 0.0534$ with $p_Q = 0.741$ and $D_V = 0.0531$ with $p_V = 0.653$. Thus neither marginal model is rejected at the 5% level. The Q–Q plots in Figure 2 show no systematic departure over the simulated range.

3.3 Dependence estimation and copula selection

The sample Kendall coefficient is

$$\widehat{\tau} = 0.6036,$$



(a) Fitted GPD quantiles

(b) Fitted Pearson III quantiles

Fig. 2: Q–Q diagnostics for the fitted GPD peak-excess and Pearson III volume margins.

with nonparametric bootstrap interval $[0.5051, 0.6867]$. Inversion gives

$$\widehat{\theta} = \frac{1}{1-\widehat{\tau}} = 2.5229, \qquad \widehat{\theta}_{95\%} \in [2.0205, 3.1914],$$

and therefore

$$\widehat{\lambda}_U = 2 - 2^{1/\widehat{\theta}} = 0.6838.$$

The Gumbel parameter estimated by maximum pseudo-likelihood is 2.5016, close to both the inversion estimate and the simulation value 2.5.

Table 2: One-parameter copula comparison based on pseudo-likelihood.

Copula	Parameter	Log-likelihood	AIC
Gumbel–Hougaard	2.5016	56.362	−110.724
Gaussian	0.8043	51.828	−101.656
Frank	8.0569	49.787	−97.574
Clayton	1.5052	31.352	−60.703

Gumbel–Hougaard has the smallest AIC as shown by the results presented in Table 2. Relative to the Gaussian alternative, $\Delta\text{AIC} = 9.07$, providing substantial model-selection support for explicit upper-tail dependence within this synthetic experiment.

3.4 Joint return periods

Table 3 reports design pairs defined by common marginal non-exceedance probabilities p, so that $F_Q(q_p) = F_V(v_p) = p$. The annual event rate is $\widehat{\eta} = 100/36$. All values obey Corollary 6.

Table 3: Return periods derived from the fitted synthetic model.

p	q_p (m^3/s)	v_p (hm^3)	$T_Q = T_V$ (yr)	T_{OR} (yr)	T_{AND} (yr)
0.80	3615.80	4957.62	1.80	1.41	2.47
0.90	4019.80	5632.59	3.60	2.78	5.11
0.95	4366.30	6257.35	7.20	5.51	10.37
0.98	4749.85	7035.74	18.00	13.72	26.16

For example, the pair $(4366.3\ \mathrm{m}^3/\mathrm{s}, 6257.4\ \mathrm{hm}^3)$ has marginal return period 7.2 years for either component, OR return period 5.51 years, and AND return period 10.37 years. The shorter OR period reflects the larger union event; the longer AND period reflects the smaller intersection. Figure 3 and Figure 4 display this geometry over a continuous range of thresholds.

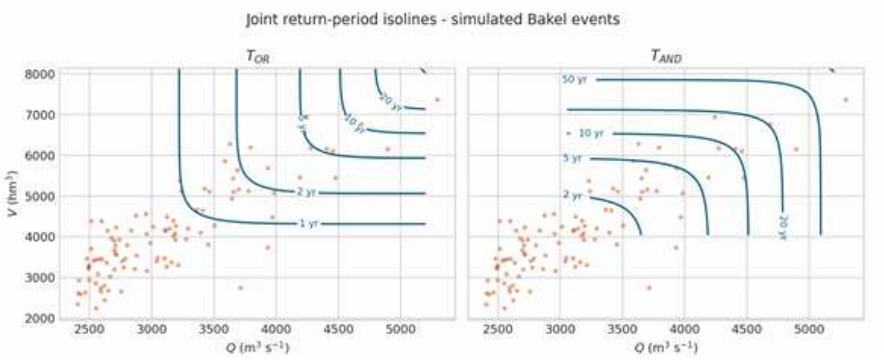


Fig. 3: OR and AND joint return-period isolines under the fitted Gumbel–Hougaard model.

4 Conclusion and perspectives

This work supplies the definitions, assumptions, proofs, and diagnostics required for a rigorous bivariate POT analysis. It corrects the Pearson III CDF, the characterization of the Gumbel margin, the uniqueness condition in Sklar's theorem, the conditional-copula notation, and the annualisation of return periods. It also proves a simple ordering theorem that serves as a numerical quality check.

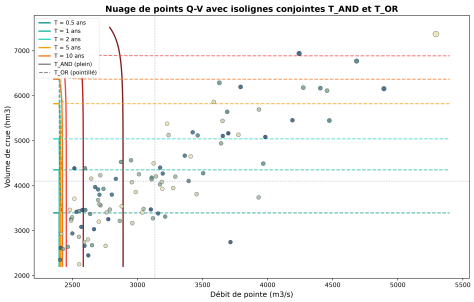


Fig. 4: Q-V scatter plot with conjoined contour lines T_{AND} and T_{OR}.

The synthetic experiment produces a coherent empirical chain: plausible event marks, acceptable marginal fits, strong upper-tail dependence, selection of the generating Gumbel–Hougaard family, and joint return periods satisfying the theoretical ordering constraints. It therefore demonstrates that the proposed methodology is computationally and mathematically consistent. However, no operational conclusion concerning Senegal River infrastructure should be drawn from simulated values. Consequently, authenticated discharge records, event-extraction rules, units, and uncertainty analysis remain necessary for a site specific risk assessment.

Declarations

- Funding : **There was no funding for this work.**
- Conflict of interest/Competing interests : **Not applicable.**
- Ethics approval and consent to participate : **Not applicable.**
- Consent for publication : **Not applicable.**

References

[1] Favre, A., Adlouni, S.E., Perreault, L., Thiémonge, N., Bobée, B.: Multivariate hydrological frequency analysis using copulas. Water Resources Research **40** (2004) https://doi.org/10.1029/2003wr002456

[2] Balistrocchi, M., Orlandini, S., Ranzi, R., Bacchi, B.: Copula-based modeling of flood control reservoirs. Water Resources Research **53**, 9883–9900 (2017) https://doi.org/10.1002/2017wr021345

[3] Chowdhary, H., Escobar, L.A., Singh, V.P.: Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. Hydrology research **42**, 193–216 (2011) https://doi.org/10.2166/nh.2011.065

[4] Poulin, A., Huard, D., Favre, A., Pugin, S.: Importance of tail dependence in bivariate frequency analysis. Journal of Hydrologic Engineering **12**, 394–403 (2007) https://doi.org/10.1061/(asce)1084-0699(2007)12:4(394)

[5] Zhang, L., Singh, V.P.: Bivariate flood frequency analysis using the copula method. Journal of Hydrologic Engineering **11**, 150–164 (2006) https://doi.org/10.1061/(asce)1084-0699(2006)11:2(150)

[6] Mohanty, M.P., Hari, V., Yadav, V., Ghosh, S., Rao, G.S., Karmakar, S.: A new bivariate risk classifier for flood management considering hazard and socio-economic dimensions. Journal of Environmental Management **255**, 109733–109733 (2019) https://doi.org/10.1016/j.jenvman.2019.109733

[7] Jiang, C., Xiong, L., Yan, L., Dong, J., Xu, C.: Multivariate hydrologic design methods under nonstationary conditions and application to engineering practice. Hydrology and earth system sciences **23**, 1683–1704 (2019) https://doi.org/10.5194/hess-23-1683-2019

[8] Salvadori, G., Michele, C.D., Durante, F.: On the return period and design in a multivariate framework. Hydrology and earth system sciences **15**, 3293–3305 (2011) https://doi.org/10.5194/hess-15-3293-2011

[9] Zhang, L., Singh, V.P.: Trivariate flood frequency analysis using the gumbel–hougaard copula. Journal of Hydrologic Engineering **12**, 431–439 (2007) https://doi.org/10.1061/(asce)1084-0699(2007)12:4(431)

[10] Mitková, V.B., Halmová, D.: Joint modeling of flood peak discharges, volume and duration: a case study of the danube river in bratislava. Journal of Hydrology and Hydromechanics **62**, 186–196 (2014) https://doi.org/10.2478/johh-2014-0026

[11] Michailidi, E.M., Bacchi, B.: Dealing with uncertainty in the probability of overtopping of a flood mitigation dam. Hydrology and earth system sciences **21**, 2497–2507 (2017) https://doi.org/10.5194/hess-21-2497-2017

[12] Requena, A.I., Mediero, L., Garrote, L.: A bivariate return period based on copulas for hydrologic dam design: accounting for reservoir routing in risk estimation. Hydrology and earth system sciences **17**, 3023–3038 (2013) https://doi.org/10.5194/hess-17-3023-2013

[13] Fan, Y., Huang, W., Huang, G., Li, Y., Huang, K., Li, Z.: Hydrologic risk analysis in the yangtze river basin through coupling gaussian mixtures into copulas. Advances in Water Resources **88**, 170–185 (2015) https://doi.org/10.1016/j.advwatres.2015.12.017

[14] Jane, R., Cadavid, L., Obeysekera, J., Wahl, T.: Multivariate statistical modelling of the drivers of compound flood events in south florida. Natural hazards and earth system sciences **20**, 2681–2699 (2020) https://doi.org/10.5194/nhess-20-2681-2020

[15] Ahn, K.: Streamflow estimation at partially gaged sites using multiple-dependence conditions via vine copulas. Hydrology and earth system sciences **25**, 4319–4333

(2021) https://doi.org/10.5194/hess-25-4319-2021

[16] Haque, N., Rahman, K., Hossain, M.S., Alam, S.: Copula-based joint distribution of flood peak and volume for improved flood risk assessment: a case study of the surma-meghna river basin, bangladesh. Journal of Water and Climate Change **17**, 1–18 (2025) https://doi.org/10.2166/wcc.2025.803

[17] Feng, Y., Shi, P., Qu, S., Mou, S., Chen, C., Dong, F.: Nonstationary flood coincidence risk analysis using time-varying copula functions. Scientific Reports **10**, 3395–3395 (2020) https://doi.org/10.1038/s41598-020-60264-3

[18] Song, M., Zhang, J., Liu, Y., Liu, C., Bao, Z., Jin, J., He, R., Bian, G., Wang, G.: Time-varying copula-based compound flood risk assessment of extreme rainfall and high water level under a non-stationary environment. Journal of Flood Risk Management **17** (2024) https://doi.org/10.1111/jfr3.13032

[19] Zhou, Y., Guo, S., Xu, C., Xiong, L., Chen, H., Ngongondo, C., Li, L.: Probabilistic interval estimation of design floods under non-stationary conditions by an integrated approach. Hydrology research **53**, 259–278 (2021) https://doi.org/10.2166/nh.2021.007

[20] Gilja, G., Ocvirk, E., Kuspilić, N.: Joint probability analysis of flood hazard at river confluences using bivariate copulas. Journal of the Croatian Association of Civil Engineers **70**, 267–275 (2018) https://doi.org/10.14256/jce.2173.2017

[21] Couasnon, A., Sebastian, A., Morales-Nápoles, O.: A copula-based bayesian network for modeling compound flood hazard from riverine and coastal interactions at the catchment scale: An application to the houston ship channel, texas. Water **10**, 1190–1190 (2018) https://doi.org/10.3390/w10091190

[22] Méndez, D.A.U., Rave, D.V.G., Jesús, M.: Exploring Multivariate Return Periods: Enhancing Accuracy in Hydrological Analysis for Flood Prediction (2024). https://doi.org/10.5194/egusphere-egu24-5685 . https://doi.org/10.5194/egusphere-egu24-5685

[23] Tootoonchi, F., Sadegh, M., Haerter, J.O., Räty, O., Grabs, T., Teutschbein, C.: Copulas for hydroclimatic analysis: A practice-oriented overview. Wiley Interdisciplinary Reviews Water **9** (2022) https://doi.org/10.1002/wat2.1579

[24] Ganguli, P., Reddy, M.J.: Probabilistic assessment of flood risks using trivariate copulas. Theoretical and Applied Climatology **111**, 341–360 (2012) https://doi.org/10.1007/s00704-012-0664-4

[25] Soltaninia, S., Eskandaripour, M.: Trivariate flood frequency analysis using copula functions: assessing non-stationary hydrological risks under climate change. Journal of Water and Climate Change (2025) https://doi.org/10.2166/wcc.2025.068

[26] Fan, Y., Huang, K., Huang, G., Li, Y.P.: A factorial bayesian copula framework for partitioning uncertainties in multivariate risk inference. Environmental Research **183**, 109215–109215 (2020) https://doi.org/10.1016/j.envres.2020.109215

[27] Huang, K., Fan, Y.: Parameter uncertainty and sensitivity evaluation of copula-based multivariate hydroclimatic risk assessment. Journal of Environmental Informatics (2021) https://doi.org/10.3808/jei.202100462

[28] Kim, W., Kang, S.-K., Woo, S., Choi, S.: Bivariate copula framework for drought risk assessment under climate change and practical water supply systems. Natural Hazards **121**, 22067–22094 (2025) https://doi.org/10.1007/s11069-025-07677-3

[29] Latif, S., Simonović, S.P.: Compounding joint impact of rainfall, storm surge and river discharge on coastal flood risk: an approach based on 3d fully nested archimedean copulas. Environmental Earth Sciences **82** (2023) https://doi.org/10.1007/s12665-022-10719-9

[30] Latif, S., Simonović, S.P.: Trivariate probabilistic assessments of the compound flooding events using the 3-d fully nested archimedean (fna) copula in the semi-parametric distribution setting. Water Resources Management **37**, 1641–1693 (2023) https://doi.org/10.1007/s11269-023-03448-6

[31] Carvalho Studart, T.M., Araújo Pontes Filho, J.D., Gomez, G.R., Portela, M.M., Souza Filho, F.: A bivariate return period copula application of flood peaks and volumes for climate adaptation in semi-arid regions. Water **17**, 2963–2963 (2025) https://doi.org/10.3390/w17202963

[32] Nelson, J.S.: Fishes of the World, 4th edn., p. 601. John Wiley & Sons, Hoboken, New Jersey (2006)

[33] Genest, C., MacKay, R.J.: The joy of copulas: Bivariate distributions with uniform marginals. The American Statistician **40**(4), 280–283 (1986) https://doi.org/10.1080/00031305.1986.10475322