ANÁLISE MULTIVARIADA DO RISCO DE CHEIAS: UM ESTUDO COMPARATIVO UTILIZANDO CÓPULAS C-VINE COM DIFERENTES DISTRIBUIÇÕES MARGINAIS NO RESERVATÓRIO DE FURNAS

Fernanda da Serra Costa [1] ; Ricardo de Melo e Silva Accioly [2]

Palavras-Chave – C-Vine, Distribuição Multivarada, Eventos de Cheia.

INTRODUÇÃO

Eventos hidrológicos extremos causam desastres com grandes impactos sociais e econômicos. Reservatórios de múltiplos usos são uma das estratégias estruturais para mitigar estes impactos. Em relação aos eventos de cheia, no Brasil, os reservatórios de regularização de algumas usinas hidrelétricas (UHE), além de armazenar água nos períodos de abundância para utilizá-la nos períodos hidrológicos menos favoráveis, de forma a garantir a continuidade da produção de energia elétrica, também são utilizados para minimizar os efeitos de grandes cheias. Estes dois usos são conflitantes, uma vez que, para amortecer as ondas de cheia é necessário manter espaço vazio nos reservatórios durante a estação chuvosa, ONS (2025) e Costa *et al.* (2014). O que torna a gestão integrada dos reservatórios de UHEs um desafio para a otimização da operação energética e do controle de cheia.

Devido ao caráter estocástico dos eventos de cheia, o estudo desses eventos baseia-se na teoria estatística. Tradicionalmente, a análise univariada predominou nos critérios de dimensionamento e nos planos de ação emergencial na engenharia de recursos hídricos. No entanto, eventos de cheia são fenômenos inerentemente multivariados. Para sua descrição, é necessário considerar o pico da vazão, o volume total afluente e a duração do evento. Ignorar a interdependência e a estrutura de correlação entre essas três variáveis pode levar à subestimação do risco de cheia ou ao superdimensionamento dos espaços de amortecimento da cheia, aumentando desnecessariamente o conflito com a operação energética e de controle de cheias nos reservatórios das UHEs.

A questão que se coloca no uso de distribuições multivariadas tradicionais é que elas assumem as distribuições marginais das variáveis envolvidas como iguais, o que não é o caso nos eventos de cheia. Para superar esta limitação, a aplicação do Teorema de Sklar, Sklar (1959), por meio de funções de cópula, consolidou-se como o estado da arte na hidrologia estatística, permitindo modelar a estrutura de dependência conjunta independentemente das distribuições marginais. Contudo, quando a análise se estende a três ou mais dimensões, as cópulas multivariadas clássicas (as famílias de Arquimedes ou elípticas tradicionais) revelam restrições teóricas, como pressupostos de simetria e homogeneidade na dependência de cauda entre todos os pares de variáveis, Grimaldi e Serinaldi (2006).

Neste trabalho, para superar essas questões, utilizaram-se as cópulas Vines, Kurowicka e Joe (2011), especificamente as C-Vines (Canonical Vines), devido à sua flexibilidade, que permite, além de considerar distribuições marginais distintas, captar a assimetria da dependência das caudas, Ganguli e Reddy (2012). Comparou-se o desempenho de dois modelos de copulas C-Vine, acoplados a diferentes distribuições marginais (empírica e paramétrica), aplicados à série histórica de cheias do Reservatório da UHE Furnas. Os resultados mostraram não só a importância de considerar as três variáveis que definem o evento de cheia por meio de uma abordagem multivariada para avaliar o risco de cheia, como também a influência das distribuições marginais adotadas para representá-las.

1) Universidade do Estado do Rio de Janeiro - UERJ, fcosta@ime.uerj.br
2) Universidade do Estado do Rio de Janeiro – UERJ, raccioly@ime.uerj.br

METODOLOGIA

O teorema de Sklar demonstrou que o problema de obter uma distribuição multivariada pode ser decomposto em duas partes. A função cópula, $C(u_1, \cdots, u_n)$, contém toda a estrutura de dependência da distribuição conjunta, e as distribuições marginais podem ser definidas de forma independente, não sendo necessário impor que sejam identicamente distribuídas.

$$H(x_1, \cdots, x_n) = C\big(F_1(x_1), \cdots, F_n(x_n)\big) = C(u_1, \cdots, u_n) \quad (1)$$

Onde $F_j(x_j) = u_j$, $j = 1, \ldots, n$, $U_j \sim U(0,1)$ se F_j é contínua, H é a distribuição conjunta, C é a cópula e F a função de distribuição acumulada das marginais. Os modelos clássicos de cópulas Archimedianas e Elípticas a partir de três variáveis se mostram limitados, por impor o mesmo padrão de dependência entre os pares de variáveis. As cópulas C-Vines superam esta limitação, decompondo distribuições multivariadas complexas em cascatas ordenadas de cópulas bivariadas condicionais (Construção por Cópulas aos Pares – CCP). Considerando uma cópula trivariada, a abordagem C-Vine decompõe a densidade da cópula em um produto de densidades bivariadas condicionais, Aas et. al. (2009) e Kurowicka e Joe (2011).

$$f_{123}(x_1, x_2, x_3) = f_1(x_1).f_2(x_2).f_3(x_3).c_{12}\big(F_1(x_1), F_2(x_2)\big).c_{13}\big(F_2(x_1), F_3(x_3)\big).c_{23|1}\left(F_{2|1}(x_2|x_1), F_{3|1}(x_3|x_1)\right) \quad (2)$$

Onde f_{123} é a densidade conjunta, f_1, f_2 *e* f_3 são densidades marginais, c_{12}, c_{13} são densidades das cópulas bivariadas diretas e $c_{23|1}$ é a densidade da cópula condicional e x_1, x_2 *e* x_3 são as variáveis analisadas. Generalizadamente a construção da C-Vine segue os seguintes passos: (i) obter as densidades marginais de cada variável, (ii) determinar a variável principal da estrutura, (iii) modelar as relações bivariadas entre as variáveis restantes com a principal, (iv) modelar as variáveis secundárias condicionadas a variável principal.

Apesar dos avanços na modelagem da estrutura de dependência via C-Vines, persiste a questão de qual estratégia de modelagem de distribuições marginais univariadas oferece o melhor compromisso entre a aderência aos dados históricos e a confiabilidade na extrapolação de eventos extremos. Consideraram-se duas abordagens: (i) as marginais empíricas, que replicam perfeitamente as assimetrias amostrais originais, mas são incapazes de extrapolar para tempos de retorno superiores ao máximo observado; (ii) as marginais paramétricas, que viabilizam a extrapolação para tempos de retorno elevados, podendo introduzir vieses e apresentar menor aderência aos dados observados. A escolha entre as duas abordagens pode afetar o cálculo dos tempos de retorno.

A seleção das distribuições univariadas paramétricas foi feita com base no teste de aderência de Cramer-von Mises (CVM), conforme Karian e Dudewicz (2016). A função de distribuição empírica foi construída utilizando o critério $i/(n+1)$ e a média dos postos para tratar casos de empate. Na modelagem da cópula C-Vine as melhores estruturas bivariadas foram validadas por meio de critérios de verossimilhança penalizada dentro da amostra, AIC e mBICv, conforme Nagler e Vatter (2025), e validação cruzada fora da amostra. Com base nos modelos estatisticamente superiores, foram estimados e comparados os tempos de retorno univariados e multivariados, neste último considerando a probabilidade de algum exceder um valor $P(X_1 \geq x_1, \vee\ X_2 \geq x_2, \vee\ X_3 \geq x_3)$, "OU", e a probabilidade de as três variáveis simultaneamente excederem um valor $P(X_1 \geq x_1, \wedge\ X_2 \geq x_2, \wedge\ X_3 \geq x_3)$, "E", Ganguli e Reddy (2012), constituindo um arcabouço metodológico para a gestão de riscos de eventos extremos de cheias. Os tempos de retornos (TR) trivariados "OU" e "E" são:

$$TR^{"OU"}_{X_1X_2X_3}(x_1, x_2, x_3) = \frac{1}{1 - P(X_1 \leq x_1, \vee\ X_2 \leq x_2, \vee\ X_3 \leq x_3)} = \frac{1}{1 - C(u_1, u_2, u_3)} \quad (3)$$

$$TR^{"E"}_{X_1X_2X_3}(x_1,x_2,x_3) = \frac{1}{1-P(X_1 \le x_1, \wedge\ X_2 \le x_2, \wedge\ X_3 \le x_3)}$$
$$= \frac{1}{1-F_1(x_1)-F_2(x_2)-F_3(x_3)+C(u_1,u_2)+C(u_2,u_3)+C(u_1,u_3)-C(u_1,u_2,u_3)} \quad (4)$$

RESULTADOS

Como aplicação, considerou-se a série de vazões diárias afluentes à usina hidrelétrica de Furnas, localizada no rio Grande, na bacia do rio Paraná, ONS (2025). O período considerado foi de 1933 a 2023. A cada ano, foi selecionada a maior vazão acima do limite de 4000 m³/s, que corresponde ao limite de vazão que não causa danos no vale a jusante. Para cada um destes eventos, calcularam-se o volume e a duração correspondentes, obtendo-se as três variáveis que descrevem os eventos de cheia: pico de vazão (Qp), volume (V) e duração (D) da cheia, totalizando 66 observações. Os cálculos foram realizados no ambiente R v.4.5.3. A Tabela 1 apresenta as estatísticas dos eventos de cheia. Na Tabela 2, apresentam-se os resultados dos ajustes das distribuições marginais paramétricas.

Tabela 1 – Resumo estatístico dos eventos

Variável	Min.	1°Q	Mediana	Média	3°Q	Max.
Duração (dias)	1	5	9	11,3	15,8	35
Volume (km³)	0,0008	0,3300	0,7319	1,2940	1,6262	6,8312
Pico de vazão (m³/s)	4008	5077	5535	5980	6521	13440

Tabela 2 – Ajuste das melhores distribuições marginais paramétricas

Variável	Distribuição	Parâmetros	Estatística CVM
Duração (D)	Lognormal	μ=2,13 σ=0,82	p-value=0,83
Volume (V)	Weibull	Shape=0,81 Scale=1,16	p-value=0,93
Pico de vazão (Qp)	Gumbel	Loc=5341,93 Scale=1047,80	p-value=0,85

Todos os modelos ótimos para as C-Vine tiveram como nó raiz o Volume (1), considerando os nós Duração (2) e Pico de Vazão (3) adjacentes. No entanto, a depender do tipo de distribuição marginal, as relações bivariadas obtidas resultaram em padrões de dependência distintos. No caso das marginais empíricas, a cópula que modelou o par (2,1) foi uma Gaussiana com θ=0,94, o par (1,3) foi BB7 (Clayton-Gumbel), com θ=4 e δ=7 e o par condicionado ao Volume (2,3|1) foi uma Clayton, com θ=1,5 com rotação de 90°. No caso das distribuições paramétricas, a cópula que modelou o par (2,1) foi uma Frank com θ=16; o par (1,3) foi uma Gaussiana, com θ=0,93 e o par (2,3|1) foi uma BB8 com θ=4,38 e δ=0,87 com rotação de 270°. A Tabela 3 apresenta os valores de Qp, D e V dos eventos de cheia de alguns anos da amostra e seus respectivos TRs, obtidos de forma univariada e trivariada, considerando as modelagens empírica (E_{mp}) e paramétrica (P).

Tabela 3 – Tempos de retorno univariados e trivariados (“OU”/”E”)

Ano	Qp (m³/s)	Duração (dias)	Volume (km³)	TR Marginal Qp		TR Marginal D		TR Marginal V		TR “OU”		TR “E”	
				E_{mp}	P	E_{mp}	P	E_{mp}	P	E_{mp}	P	E_{mp}	P
1981	6.534	14	1,61	4,1	3,7	3,6	3,7	3,8	3,7	2,8	2,7	5,7	5,6
1991	13.440	26	6,83	∞	2.404	13,0	11,9	∞	68,6	13,0	11,3	∞	6.098
1996	8.336	35	4,98	10,8	17,8	∞	24,7	32,4	26,3	10,8	11,3	∞	110,4
2006	6.284	35	3,73	3,1	3,0	∞	24,7	10,8	13,3	3,1	2,9	∞	39,8
2020	8.368	21	3,47	12,9	18,4	9,2	7,6	9,3	11,5	7,1	6,3	18,1	31,5

Analisando os resultados das tabelas 1 e 3, de acordo com a severidade dos eventos, percebe-se que o evento de 1981 situa-se próximo à média para as três variáveis, resultando em TRs semelhantes na

modelagem univariada, com pequeno aumento na modelagem conjunta "E" e redução na "OU", como esperado. Em 2020, as três variáveis estão acima do terceiro quartil, observam-se diferenças entre os TRs empíricos e paramétricos nas abordagens univariadas e mais significativas em relação à multivariada "E". Nos eventos de 1996 e 2006, a duração (D) corresponde ao máximo observado, o que levou à diferença significativa entre os TRs empíricos e paramétricos, devido à diferença nas caudas das distribuições. Neste caso, os TRs da modelagem multivariada são fundamentais para representar os eventos de forma completa. Finalmente, em 1991, as variáveis Qp e V correspondem aos valores máximos observados, o que torna a modelagem paramétrica multivariada mais relevante. A comparação dos TRs "OU" e "E" evidencia a diferença entre considerar a ocorrência de apenas uma das variáveis ou de todas as três variáveis simultaneamente, para representar o evento.

CONCLUSÕES

Este trabalho demonstrou que, na avaliação do risco de eventos de cheia, a consideração simultânea das três variáveis que descrevem o fenômeno (V, D e Qp) é fundamental, o que torna importante o uso de distribuições multivariadas. Entretanto, como as distribuições univariadas pertencem a diferentes famílias, a modelagem tradicional não é viável. Neste trabalho utilizou-se as cópulas C-Vine, que não apresentam esta limitação e permitem diferentes estruturas de dependência entre as variáveis. Além disso, ficou evidente que as distribuições marginais adotadas têm um efeito significativo na estimativa do risco. As distribuições paramétricas são mais indicadas, pois representam melhor as caudas superiores, permitindo a avaliação de eventos superiores aos observados. Uma alternativa à distribuição empírica é a semiparamétrica.

A avaliação do tempo de retorno de alguns eventos observados, mostrou que a avaliação de risco considerando apenas uma variável, pode levar à subestimação do risco, em especial quando pelo menos uma das variáveis se encontra próxima do limite superior da distribuição. Os tempos de retorno "OU" e "E" servem como balizas na análise de risco, em que o "OU" fornece o limite inferior e o "E" o limite superior.

REFERÊNCIAS

AAS, K.; CZADO, C.; FRIGESSI, A.; BAKKEN, H. (2009). "*Pair-copula construction of multiple dependence*". Insurance: Mathematics and Economics, v.44, p.182-198.

COSTA, F.S.; RAUPP, I.P.; DAMÁZIO, J. M.; OLIVEIRA, P.D.; GUILHON, G.F. (2014). *"The Methodologies for the Flood Control Planning Using Hydropower Reservoirs in Brazil"* in Anais do 6th International Conference on Flood Management, São Paulo, SP, Brasil.

GANGULI, P.; REDDY, M.J. (2012). "Probabilistic Assessment of flood risk using trivariate copulas". Theoretical and Applied Climatology, v.111, p.341-360.

GRIMALDI, S.; SERINALDI, F. (2006). "*Asymmetric copula in multivariate flood frequency analysis*". Advances in Water Resources, v.29, p.1155-1167.

KARIAN, Z.A.; DUDEWICZ, E.J. (2016). *Handbook of Fitting Statistical Distributions with R*. 1st Edition. CRC Press, Boca Raton/FL, 1718p.

KUROWICKA, D.; JOE, H. (2011). "*Dependence Modeling: vine copula handbook*". World Scientific, Singapore, p.368.

NAGLER, T.; VATTER, T. (2025). "*rvinecopulib: High Performance Algorithms for Vine Copula Modeling*". Versão 0.7.3.1.0. Disponível em: https://vinecopulib.github.io/rvinecopulib/. Acesso em: 23 jun. 2026.

ONS (2025). Plano Anual de Prevenção de Cheias – Ciclo 2025/2026. Rio de Janeiro/RJ, 156 p.

ONS (2025). "SINTEGRE". https://sintegre.ons.org.br/sites/9/13/63 acessado em 01/04/2025.

SKLAR, A. (1959). "Fonctions de répartition à n dimensions et leurs marges". Publ. Inst. Statist. Univ. Paris, 8(1):11.