



<http://climacom.mudancasclimaticas.net.br/caudas-pesadas/>

A presença de caudas pesadas em variáveis hidrometeorológicas: conceitos, desafios de inferência e consequências

Veber Afonso Figueiredo Costa [1]

Dirceu Silveira Reis Junior [2]

RESUMO: O decaimento polinomial das caudas superiores, o qual é intrinsecamente relacionado à ocorrência frequente de eventos aleatórios de grande magnitude, é usualmente verificado em variáveis hidroclimáticas. Não obstante, e apesar de evidências empíricas e fundamentação conceitual para a adoção de modelos com caudas pesadas no contexto da Hidrologia, engenheiros e pesquisadores ainda se valem majoritariamente de distribuições com decaimento exponencial para dimensionamento de estruturas hidráulicas, previsão de processos hidroclimáticos e modelagem de resíduos de simulação hidrológica, o que pode ocasionar subestimação do risco e introduzir grande viés à tomada de decisão. Diante disso, neste artigo faz-se uma breve introdução às variáveis com caudas pesadas, contemplando seu formalismo matemático, aspectos físicos e estatísticos associados à sua gênese, e suas implicações no âmbito das Ciências Hidrológicas, com ênfase na análise de frequência e na construção de modelos de regressão. Em particular, abordam-se mecanismos físicos e argumentos estatísticos associados à ocorrência de caudas pesadas em variáveis como precipitação e vazão, e estratégias para estimação paramétrica amparadas em realismo físico. Tenciona-se com tal abordagem prover especialistas com elementos conceituais e práticos para a utilização de modelos dotados de cauda pesada no cotidiano das Ciências Hidrológicas, fornecendo-se assim um panorama mais realista de eventos hidroclimáticos extremos e diretrizes mais consistentes para a estimação do risco.

PALAVRAS-CHAVE: Eventos extremos. Cheias. Secas. Desastres. Risco.



Heavy tails in hydrometeorological variables: concepts, inference challenges and consequences

ABSTRACT: Polynomial decay of upper tails, which is inherently related to the frequent occurrence of large-magnitude random events, is commonly observed in hydroclimatic variables. However, despite empirical evidence and conceptual support for the adoption of heavy-tailed models in Hydrology, engineers and researchers still mostly rely on distributions with exponential decay for the design of hydraulic structures, forecasting of hydroclimatic processes, and modeling of residuals of rainfall-runoff simulation. This can lead to underestimation of risk and introduce significant bias to decision-making. Based on these arguments, in this paper we provide a brief introduction to heavy-tailed variates, which encompasses the mathematical formalism, physical and statistical aspects related to such variables, and the implications to Hydrological Sciences, with emphasis on frequency analysis and the identification of regression models. In particular, we address physical mechanisms and statistical arguments associated with the occurrence of heavy tails in variables such as precipitation and streamflow, and strategies for parametric estimation supported by physical realism. This approach aims to provide experts with conceptual and practical elements for the use of heavily tailed models in the daily practice of Hydrological Sciences, thus providing a more realistic perspective of extreme hydroclimatic events and more consistent guidelines for risk estimation.

KEYWORDS: Extreme events. Floods. Droughts. Disasters. Risk.



Introdução

Frequentemente, ao se analisar séries temporais de variáveis hidrometeorológicas, em especial aquelas mais curtas, percebe-se que algumas dessas observações assumem valores consideravelmente superiores à média (ignore-se, por ora, os efeitos dos erros amostragem na inferência dessa grandeza populacional). Nesses casos, é usual que tais pontos amostrais sejam interpretados como *outliers*, cuja probabilidade de superação seria bastante reduzida, ou mesmo como erros grosseiros, de medição ou de transcrição, o que justificaria sua exclusão do conjunto de dados antes da inferência estatística. Por outro lado, à medida que observações adicionais são agregadas à amostra, verifica-se que a ocorrência de eventos de maior magnitude (com relação à média) é, de fato, relativamente comum, o que coloca em perspectiva a hipótese de que esses reflitam condições “atípicas” do processo hidroclimático subjacente (Papalexiou e Koutsoyiannis, 2013; Koutsoyiannis, 2024).

Por certo, as afirmações acima são inteiramente calcadas em evidências empíricas – nenhuma fundamentação conceitual ou teórica foi empregada até aqui para sustentá-las. Não obstante, a situação descrita pode ter severas implicações nos contextos da inferência estatística e, principalmente, do projeto e dimensionamento de estruturas hidráulicas, os quais são, usualmente, baseados na análise de frequência de variáveis hidroclimáticas. Com efeito, a presença frequente de realizações de grande magnitude dessas variáveis pode afetar sobremaneira as estimativas de parâmetros de distribuições de probabilidade e de momentos, ainda que os tamanhos das amostras observadas compreendam centenas de observações (Vogel et al., 2025) – fato pouco usual nas ciências hidrológicas. Nesses casos, a confiabilidade para extrapolação a níveis não observados dos processos hidroclimáticos é bastante reduzida, ao passo que as incertezas decorrentes da inferência se tornam demasiadamente elevadas, o que dificulta a tomada de decisão relacionada ao *trade off* custo-risco.

O comportamento em questão decorre, via de regra, das denominadas caudas pesadas (ou, alternativamente, caudas polinomiais ou subexponenciais), para as quais as probabilidades de superação se reduzem de maneira relativamente mais lenta, com relação a distribuições com decaimento exponencial (ou caudas leves), à medida que os valores da variável aleatória aumentam (a definição formal será apresentada na próxima seção). A presença de caudas pesadas em variáveis hidroclimáticas pode decorrer de aspectos físicos ou de considerações puramente



estatísticas. Sob a primeira perspectiva, caudas pesadas usualmente se originam da interação, aditiva ou multiplicativa, de diferentes processos físicos, como, por exemplo, fenômenos atmosféricos associados à formação de chuvas intensas, ou da transformação não linear de uma variável aleatória, como a conversão de precipitação em escoamento em uma bacia hidrográfica (Merz et al., 2022; Laaha et al., 2025, Nguyen et al., 2025; Wang et al., 2025;) – note-se que os processos geradores “parentais” podem ser inteiramente descritos por caudas leves e, ainda assim, aqueles resultantes de sua interação podem apresentar caudas pesadas. Do ponto de vista estatístico, em especial no tocante a extremos, as caudas pesadas podem resultar do comportamento sub-assintótico de amostras de tamanho finito, tendo-se em vista que, diante da dependência temporal e da cicloestacionariedade de processos hidrológicos, a convergência para a verdadeira distribuição limite pode não ocorrer mesmo em blocos anuais (Serinaldi, Lombardo e Kilsby, 2020) – tal convergência é reconhecidamente lenta para processos parentais com caudas exponenciais. Nesse caso, e para os tempos de retorno usuais na hidrologia, distribuições com caudas pesadas tendem a apresentar melhor aderência às amostras observadas (Koutsoyiannis, 2024), provendo, assim, uma quantificação mais adequada do risco.

Diante do exposto, é possível inferir que caudas pesadas são onipresentes em variáveis hidrometeorológicas: processos de interesse, como precipitações intensas, cheias e secas, decorrem das interações de um grande número de variáveis aleatórias, e, ao mesmo tempo, as transferências de água entre as diversas fases do ciclo hidrológico envolvem, com frequência, transformações não lineares dessas variáveis. Tal linha de raciocínio pode ser estendida a representações simplificadas dos sistemas hidrológicos por meio de modelos, sejam eles empíricos, conceituais ou fisicamente fundamentados, para os quais caudas pesadas emergiriam naturalmente na descrição dos resíduos (Costa e Fernandes, 2017). Não obstante, ainda prevalece o uso, por vezes não justificado, de distribuições com caudas exponenciais na modelagem de variáveis hidrometeorológicas (Koutsoyiannis, 2024). Em grande medida, tal fato pode ser atribuído à evolução histórica da Hidrologia Estatística, em especial com os trabalhos de Emil Gumbel (Gumbel, 1958), e a aspectos estatísticos bem “definidos” das distribuições com caudas leves, como, por exemplo, a existência de momentos de todas as ordens.

Cabe notar, porém, que a literatura acerca de variáveis com caudas pesadas é ainda restrita na Hidrologia (Merz et al., 2022) e as consequências da presença de caudas pesadas em expedientes



de modelagem, com relação à estimação paramétrica, quantificação de incertezas e interpretação física, ainda não são bem compreendidas por engenheiros e hidrológicos, o que pode limitar sua utilização. Com efeito, é importante entender que a presença de caudas pesadas, quando não devidamente identificada e modelada, pode resultar em (1) interpretações equivocadas sobre o nível de raridade (probabilidade de excedência) e identificação de possíveis causas de eventos extremos ocorridos no passado, tais como grandes enchentes, como aquela ocorrida no estado do Rio Grande do Sul em 2024, ou secas intensas, duradouras e espacialmente abrangentes, como aquela que atingiu o nordeste do Brasil no período 2012-2018; e (2) subestimativa da magnitude do perigo para um dado período de retorno pré-especificado, como no caso de precipitações intensas e vazões máximas, ou superestimativa de variáveis como vazões mínimas e intensidade de secas, com consequências relevantes no dimensionamento, operação, e efetividade de infraestrutura hídrica, na estimativa de risco nos processos de alocação e outorga de água, e na elaboração de planos de seca para redução do risco climático. Uma outra consequência da presença de caudas pesadas, que não está explicitamente presente no cotidiano dos profissionais que lidam com a água, mas que vale a pena explicitar aqui devido a sua importância em decisões de caráter técnico, normalmente tomadas por um grupo mais restrito de profissionais, é o aumento das incertezas na avaliação da capacidade de um modelo hidrológico em representar a natureza, realizada normalmente por meio de índices como os coeficientes de eficiência de Nash-Sutcliffe e Kling-Gupta, e aumento das incertezas nas simulações propriamente ditas. Desconsiderar essas consequências pode gerar um excesso de confiança depositada no modelo e em suas previsões, com potenciais consequências na efetividade das decisões tomadas com base nessas simulações.

Nesse contexto, este artigo tem por intuito prover uma breve introdução às variáveis com caudas pesadas, contemplando seu formalismo matemático, aspectos físicos e estatísticos associados à sua gênese, e suas implicações no âmbito das Ciências Hidrológicas, com ênfase na análise de frequência e na construção de modelos de regressão. Espera-se estabelecer um arcabouço teórico consistente para justificar a utilização de modelos distributivos dotados de caudas pesadas em aplicações na Hidrologia e, ao mesmo tempo, ilustrar os desafios práticos associados a essa escolha.

Fundamentação teórica – perspectiva estatística



Em termos formais, uma variável aleatória X , com função acumulada de probabilidades $F_X(x)$ e função de sobrevivência $\bar{F}_X(x) = 1 - F_X(x)$, possui cauda pesada, se para todo $t > 0$ (Koutsoyiannis, 2024),

$$e^{tx} \bar{F}_X(x) = \infty \quad (1)$$

Alternativamente, se uma variável aleatória X possui cauda pesada, pode-se demonstrar que (Koutsoyiannis, 2024):

$$x^{\frac{1}{\xi}} \bar{F}_X(x) = a \quad (2)$$

na qual a é uma constante positiva e finita e ξ é denominado índice de cauda, o qual controle a taxa de decaimento da cauda superior à medida que x aumenta – maiores valores de ξ implicam caudas gradativamente mais pesadas. Usualmente, um valor nulo para o índice de cauda é associado a caudas com decaimento exponencial (Merz et al., 2022; Koutsoyiannis, 2024), como no caso da distribuição Gumbel. Contudo, distribuições como a lognormal, correntemente empregada na análise de frequência de variáveis hidroclimáticas, satisfazem a Equação 1, ainda que apresente $\xi = 0$. Nesse sentido, a distribuição lognormal compreende um modelo com cauda polinomial.

A Figura 1 ilustra as distinções entre uma distribuição arbitrária com cauda leve ($\xi = 0$) e uma distribuição com cauda pesada ($\xi = 0,5$), em termos das funções de quantis (Figura 1a) e das funções densidade de probabilidade (Figura 1b). Supondo-se que a variável em questão expresse as vazões máximas anuais em uma dada seção fluvial, é possível perceber, de maneira clara, as implicações para o dimensionamento de uma estrutura hidráulica hipotética: ao se assumir um modelo com caudas exponenciais, os quantis de projeto, para quaisquer tempos de retorno, são consideravelmente inferiores àqueles estimados para a distribuição com caudas polinomiais. Nesse caso, se a variável de fato apresenta caudas pesadas e o modelo foi incorretamente prescrito, o risco associado será severamente subestimado e potenciais prejuízos econômicos e ambientais podem exceder as diretrizes originalmente estabelecidas na fase de projeto. Diante disso, é válido questionar se a emergência de caudas pesadas em variáveis hidroclimáticas pode ser estabelecida com base em raciocínio dedutivo (i.e., *a priori*), de maneira a se minimizar as consequências de definições incorretas do tipo de decaimento de suas caudas superiores.

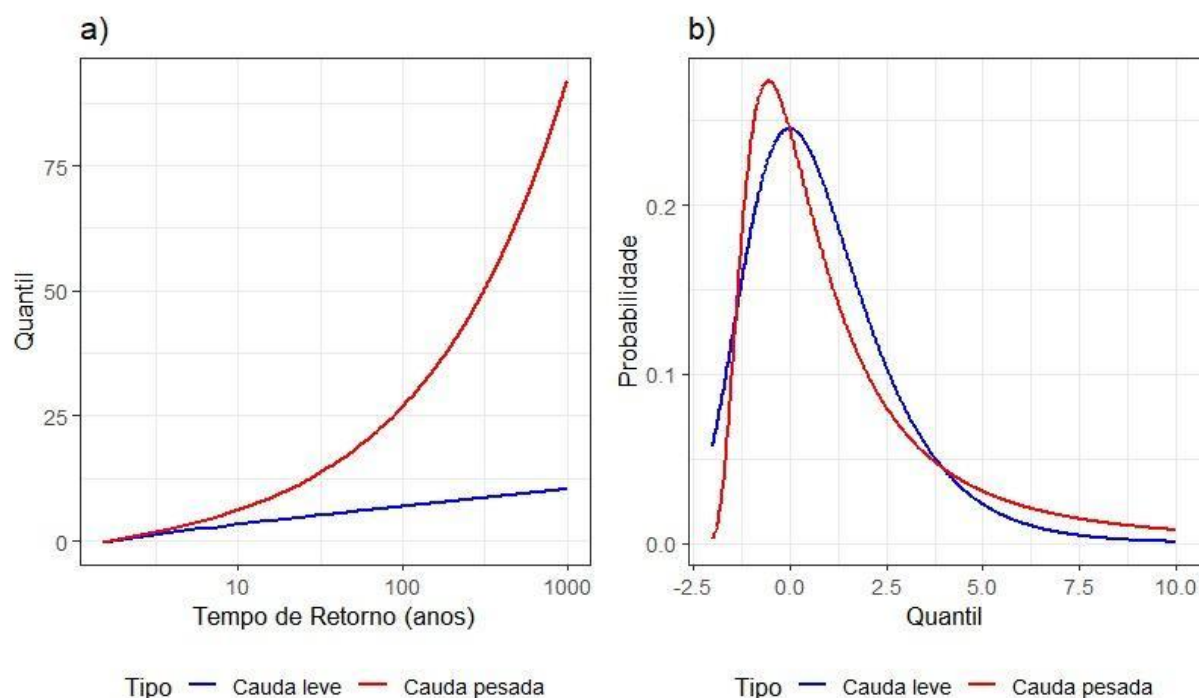


Figura 1 – Distinções entre distribuições com cauda leve e pesada. (a) curvas de quantis; (b) funções densidade de probabilidade.

De acordo com Koutsoyiannis (2024), o princípio de máxima entropia, que postula que a incerteza associada a uma variável aleatória deve sempre encontrar-se em seu máximo, considerando-se possíveis restrições físicas introduzidas na inferência (e.g., momentos da variável aleatória), oferece uma abordagem coerente para responder a tal questão. Com efeito, tal princípio fornece sustentação teórica para a definição do tipo de cauda somente com base no coeficiente de variação CV_X da variável aleatória em análise (cabe lembrar que $CV_X = \frac{\sigma_X}{\mu_X}$, em que μ_X é a média populacional e σ_X é o desvio padrão populacional). Naqueles casos em que $CV_X \leq 1$, as distribuições de máxima entropia resultam em caudas exponenciais. Por outro lado, quando $CV_X > 1$, caudas polinomiais emergem.

Em termos práticos, variáveis como precipitação em escalas diária ou inferiores tendem a apresentar valores elevados para o coeficiente de variação (Koutsoyiannis, 2024; note-se, novamente, que os efeitos dos erros de amostragem são ignorados aqui). Nesse contexto, tais variáveis meteorológicas seriam naturalmente descritas por caudas pesadas e o raciocínio indutivo somente seria empregado na estimação do índice de cauda. Por outro lado, ao se considerar



maiores escalas de tempo para as precipitações (e.g., mensal ou superior), a expressiva redução no coeficiente de variação levaria, também de maneira trivial, a distribuições com caudas leves. Essa linha de raciocínio encontra amparo em evidências empíricas, especialmente quando se dispõe de séries temporais mais longas. Como exemplo, a Figura 2 apresenta as curvas de quantis para precipitações diárias não nulas na estação pluviométrica de Bologna, na Itália, a qual dispõe de 206 anos de medições – cabe notar que a formulação usual para o cálculo do tempo de retorno é inadequada nesse caso uma vez que todos os valores não nulos são incluídos na inferência; tal formulação foi adaptada de maneira a se considerar o intervalo de amostragem de um dia, em lugar daquele anual, e o número de dias chuvosos ao longo do período de observação (para mais detalhes, sugere-se consultar Koutsoyiannis (2024)). Como modelos candidatos à descrição das precipitações diárias, empregou-se a distribuição exponencial, dotada de cauda leve, e a distribuição de Pareto com dois parâmetros, dotada de cauda pesada – o índice de cauda estimado, a partir do método dos mínimos quadrados, foi $\xi = 0,11$. É possível perceber que a distribuição de Pareto, obtida com base em considerações de máxima entropia, provê uma representação adequada das alturas de precipitação associadas a maiores tempos de retorno, as quais apresentam especial interesse para projetos. A distribuição exponencial, por sua vez, levou à subestimação de quantis em virtualmente toda a faixa de tempos de retorno considerada, refletindo a completa falta de aderência à amostra observada. Ademais, verifica-se uma diferença de aproximadamente 2,5 vezes nas estimativas do quantil associado ao tempo de retorno de 10.000 anos, usualmente empregado no dimensionamento de grandes vertedores. Claramente, ao se negligenciar a presença de caudas pesadas em precipitações diárias, pode-se incorrer em níveis inaceitáveis de risco para tais estruturas hidráulicas.

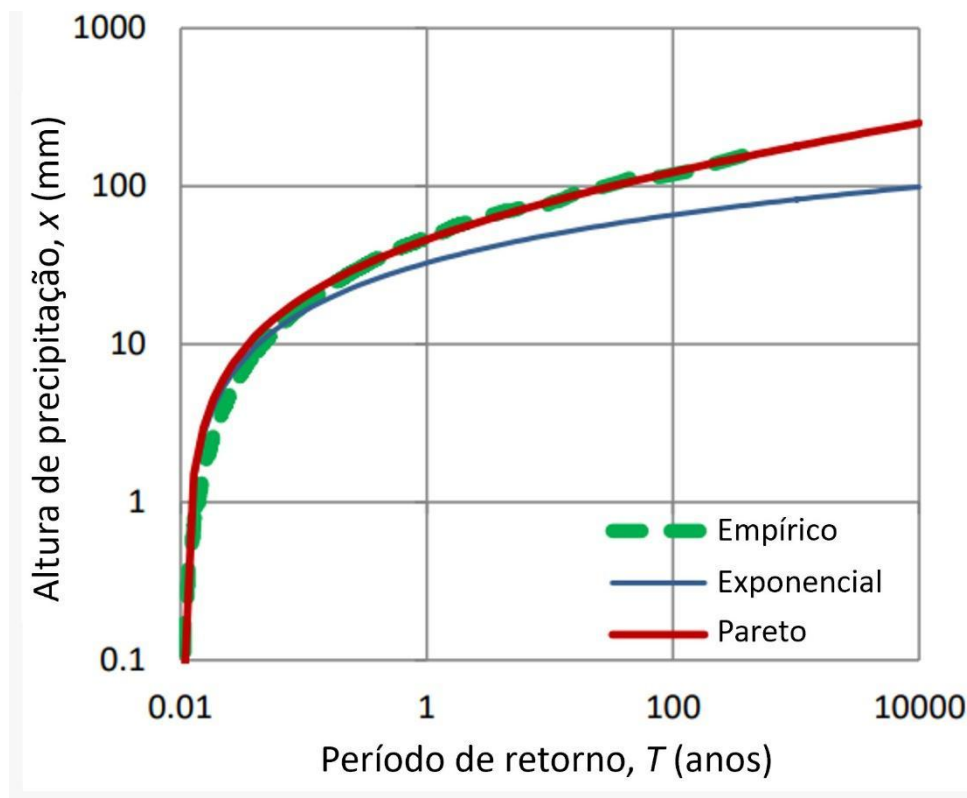


Figura 2 – Curvas de quantis para as alturas de precipitação diária na estação de Bologna, Itália.

Fonte: Koutsoyiannis (2024).

Por outro lado, quando o interesse se volta somente aos extremos de precipitação, como, por exemplo, máximos anuais, a argumentação acima pode ser insuficiente. De fato, nesses casos, os valores do coeficiente de variação podem ser inferiores à unidade, o que sugeriria a maior adequação de modelos com caudas leves. Ainda assim, é possível se advogar pela utilização de modelos com cauda pesada a partir de considerações assintóticas, uma vez que, para reduzidas probabilidades de superação, as caudas de distribuições parentais são equivalentes àquelas dos máximos (Serinaldi, Lombardo e Kilsby, 2020; Serinaldi et al., 2025). Sob tal perspectiva, a utilização de modelos com caudas polinomiais na modelagem de precipitações extremas também poderia ser estabelecida *a priori* e, novamente, o raciocínio indutivo se restringiria à estimação do peso dessas caudas.

Cabe ressaltar, porém, que o trato matemático de distribuições com caudas pesadas apresenta complexidade relativamente maior do que aquelas com caudas exponenciais. Com efeito, enquanto os momentos são sempre finitos para essas últimas, esses somente existem para ordens



$q < \frac{1}{\xi}$ para variáveis com caudas pesadas. Por exemplo, a variância de um processo hidrometeorológico diverge se $\xi > 0,5$, o que, de um ponto de vista físico, implicaria que um valor infinito de energia seria necessário para a materialização do referido processo, violando as leis da termodinâmica (Koutsoyiannis, 2024). Nesse sentido, é imperativo, ao se prescrever o uso de distribuições com caudas pesadas, observar a coerência entre aspectos físicos estabelecidos *a priori* e as medidas descritivas populacionais materializadas pelo índice de cauda, de maneira a assegurar que a extrapolação a níveis não observados do processo modelado reflita condições fisicamente plausíveis. É também necessário mencionar que a inferência confiável do índice de cauda, em geral, exige um número elevado (e.g., centenas ou milhares) de realizações dos processos hidroclimáticos em análise – os efeitos de erros de amostragem podem comprometer a estimação paramétrica e levar a valores pouco realistas para o referido parâmetro (alguns dos desafios associados à inferência na presença de caudas pesadas serão apresentados posteriormente). Assim, a utilização de modelos com caudas pesadas deve ser precedida pela construção cautelosa do quadro de inferência e pela seleção apropriada de métodos de estimação, de maneira a assegurar que o modelo estatístico seja apto a descrever, com realismo físico, o processo em condições extremas.

A gênese de caudas pesadas em variáveis hidrometeorológicas

Na seção anterior, discutiu-se, a partir de uma perspectiva estatística, condições sob as quais a maximização da entropia originária distribuições com caudas pesadas. É conveniente, porém, avaliar, sob uma perspectiva física, os mecanismos associados à gênese de caudas pesadas em variáveis hidrometeorológicas. Em síntese, tais mecanismos mais comumente compreendem interações, aditivas ou multiplicativas, entre variáveis aleatórias, ou transformações não lineares dessas variáveis, ainda que as caudas pesadas possam ser preservadas em transformações lineares (Merz et al., 2022).

Como exemplo de interações entre variáveis, alturas não nulas de precipitação diária podem ser expressas como o produto de três variáveis Gaussianas, quais sejam, o fluxo médio vertical de ar, a umidade específica e a eficiência de precipitação (Zorzetto, Canale e Marani, 2024) – cabe notar que a distribuição Normal apresenta caudas leves, mas o produto dessas variáveis origina caudas



pesadas, como demonstrado pela utilização do princípio de máxima entropia. Outro exemplo de interação entre variáveis aleatórias compreende a soma de vazões em uma confluência, a qual preserva possíveis caudas pesadas de uma ou ambas as componentes do processo (Merz et al., 2022). Por fim, processos expressos por meio de misturas de distribuições também podem apresentar decaimento polinomial se uma das componentes da mistura for dotada de caudas pesadas. Tal situação é comum, por exemplo, em picos de vazão em um curso d'água, os quais podem ser associados a diferentes mecanismos causais (e.g., precipitações ordinárias e rios atmosféricos; ver Costa e Sampaio (2021)), e tendem a “herdar” o comportamento daquela componente com cauda mais pesada. A heterogeneidade dos processos de precipitação (e.g. mecanismos convectivos e estratiformes) parece também resultar num aparente comportamento de cauda pesada, mesmo quando os mecanismos geradores podem ser bem representados por distribuições de cauda mais leve (Marra et al., 2026).

Com relação às transformações de variáveis hidrometeorológicas, a conversão de precipitações em vazões em uma bacia hidrográfica compreende o exemplo mais comum de propagação de caudas polinomiais entre distintas fases do ciclo hidrológico. Com efeito, as respostas usualmente não lineares impostas pelas bacias podem amplificar o peso das caudas das precipitações, resultando em cheias extremas mesmo para alturas (ou intensidades) de precipitação moderadas (Wang et al. (2025). Tal característica é ainda mais facilmente percebida em pequenas bacias, nas quais a variabilidade espacial das precipitações e, por conseguinte, os efeitos de agregação espacial dessa variável, são menos pronunciados (Merz et al., 2022; Wang et al., 2023). Ademais, há na literatura evidências de que condições climáticas mais secas podem amplificar o peso das caudas das vazões em decorrência da maior variabilidade espacial dos coeficientes de escoamento ao longo da bacia hidrográfica, quando comparados àqueles observados em áreas úmidas (Merz et al., 2022). Com efeito, a não linearidade das respostas hidrológicas tende ser aumentada nesses casos em função das diferentes condições de infiltração, com efeitos mais perceptíveis na formação de cheias mesmo diante de condições homogêneas de precipitação. Macdonald et al. (2022, 2024) realizaram análises aprofundadas com o intuito de entender quais são os principais fatores que resultam na presença de causa pesadas nas vazões extremas. Para isso, os autores utilizaram duas abordagens. A primeira é baseada em registros de precipitação e vazão durante eventos extremos



registrados em 480 bacias na Alemanha e Áustria, em conjunto com modelos estatísticos multivariados.

Os resultados indicaram que as resposta da bacia tendem a dominar o processo, seguido por características do evento de precipitação, sazonalidade e das áreas de drenagem. A outra abordagem é baseada em uso de geradores estocásticos de variáveis climáticas que alimentam um modelo hidrológico, o qual representa os mecanismos de transporte de água na bacia. Com essa abordagem, os autores mostraram que o processo de geração de escoamento superficial tem forte influência na distribuição dos picos de vazão. Entretanto, a partir de períodos de retorno mais elevados, essa influência vai perdendo força, e a distribuição da precipitação passa a ter mais relevância. A faixa de período de retorno a partir da qual se observa tal mudança de relevância entre os fatores intensificadores de cauda pesada depende da relação entre o volume de armazenamento da bacia e a precipitação média anual.

É válido também mencionar que a interferência antrópica pode alterar o decaimento das caudas superiores de variáveis hidrológicas. De fato, a construção de reservatórios e a retificação de cursos d'água podem ter grandes implicações na formação de grandes cheias, em comparação às condições naturais, com consequente modificação do peso das caudas – note-se que o efeito dessas interferências é reduzido naquelas cheias mais frequentes, o que pode também alterar as condições de assimetria das distribuições dessa variável. Não obstante, os efeitos de atividades antrópicas no decaimento das caudas de variáveis hidrológicas são ainda pouco compreendidos (Merz et al., 2022), uma vez que esses dependem da localização das intervenções ao longo curso d'água ou de regras de operação específicas (e.g., práticas para a gestão de cheias), o que tende a reduzir as possibilidades de generalização das conclusões correlatas às modificações das caudas diante das alterações do regime natural.

Por fim, a gênese de caudas pesadas pode ser associada ao comportamento sub-assintótico de amostras finitas, sem ligações diretas com mecanismos físicos. De fato, a convergência de variáveis com cauda superior exponencial à distribuição limite é reconhecidamente lenta (Koutsoyiannis, 2024), mesmo em condições de independência entre realizações do processo. Além disso, a modulação dos parâmetros da distribuição de probabilidades oriunda da cicloestacionariedade do processo subjacente pode dificultar ainda mais a convergência, e, assim, blocos de amostragem com tamanho superior a 365 (um ano) podem ser insuficientes para a validade de propriedades



assintóticas. Nesses casos, a emergência de caudas pesadas pode compreender tão somente um efeito espúrio da amostragem. Ainda assim, Koutsoyiannis (2024) demonstra que, para os tempos de retorno mais comumente utilizados na prática da Hidrologia e os tamanhos de amostras usualmente disponíveis, o emprego de distribuições de caudas pesadas pode levar a estimativas mais confiáveis de quantis, mesmo que o processo parental apresente caudas exponenciais – tal situação refletiria a não validade do regime assintótico em muitas variáveis hidrometeorológicas na análise clássica de máximos anuais, por exemplo.

Diante do exposto, o emprego de distribuições dotadas de cauda pesada para a modelagem de variáveis hidrometeorológicas parece justificado a partir de argumentos teóricos, sejam eles físicos ou estatísticos. Não obstante, a componente indutiva da análise apresenta grandes desafios. Em especial, as dificuldades para se estimar o índice de cauda são, há muito, reportadas na literatura (Merz et al., 2022). Desse modo, a próxima seção tenciona sumariar potenciais problemas associados à inferência paramétrica e à quantificação de incertezas quando da utilização de modelos com caudas pesadas.

Desafios de inferência na presença de caudas pesadas

Até aqui, a argumentação teórica acerca da gênese de caudas pesadas em variáveis hidroclimáticas foi construída com base em medidas descritivas populacionais. Em termos práticos, porém, a inferência paramétrica é sempre realizada a partir de amostras finitas e, usualmente, de um conjunto de hipóteses acerca do comportamento assintótico dos estimadores correspondentes (e.g., normalidade assintótica decorrente do teorema do limite central). Sob caudas exponenciais, comumente são observados poucos valores extremos, e, assim, os desvios entre as estimativas e os verdadeiros valores das medidas descritivas tendem a ser menos pronunciados. Ademais, as condições de convergência são satisfeitas com relativa facilidade, ainda que as distribuições das variáveis em análise sejam assimétricas (Koutsoyiannis, 2024). A presença de caudas pesadas, por outro lado, pode introduzir grande complexidade à inferência. Com efeito, a estimação de momentos de ordem reduzida, como a média, pode requerer centenas de observações nos casos em que caudas muito pesadas são verificadas (Vogel et al., 2025). Além disso, as propriedades assintóticas dos estimadores podem não ser obtidas nessas situações, comprometendo a quantificação das incertezas.



Nesse sentido, Vogel et al. (2025) demonstraram, a partir de simulações, os severos efeitos decorrentes de caudas muito pesadas na inferência estatística: i) vieses de subestimação e inabilidade de reproduzir a variabilidade de estatísticas comuns, como a média, podem ser verificadas ainda que se tenha milhares de observações; ii) a normalidade assintótica dos estimadores, decorrente do teorema do limite central, não se materializa, mesmo para amostras de tamanho 10.000; e iii) a reamostragem via *bootstrap* é inapta a prover a convergência aos verdadeiros valores populacionais (i.e., o estimador torna-se enviesado) e origina intervalos de confiança muito estreitos. Esses resultados retratam as enormes dificuldades de se lidar com variáveis hidrometeorológicas com caudas pesadas, uma vez que as amostras disponíveis, em especial aquelas associadas a extremos, são frequentemente curtas e, por conseguinte, a inferência é mais fortemente afetada pelos erros de amostragem.

Grandes dificuldades são também reportadas na estimação dos índices de cauda, os quais se relacionam estatisticamente aos momentos centrais de terceira ordem das variáveis em análise. De fato, na presença de caudas pesadas, os maiores desvios com relação às médias populacionais se propagam mais pronunciadamente nas estimativas dos demais momentos amostrais, podendo ocasionar valores fisicamente inconsistentes para ξ (e.g., valores superiores a 0,5, para os quais a variância divergiria). Tal situação tem motivado a busca de alternativas para a estimação mais confiável dos referidos parâmetros. Por exemplo, a escola de inferência Bayesiana fornece meios para se restringir os valores de ξ a faixas fisicamente plausíveis a partir de conhecimento *a priori* (Martins e Stedinger, 2000; Costa e Sampaio, 2021). Estudos de larga escala, como, por exemplo, Papalexiou e Koutsoyiannis (2013), podem também fornecer fundamentação empírica para a correção das estimativas do índice de cauda em função do número de pontos amostrais, o que, em grande medida, pode atenuar os efeitos dos erros de amostragem na inferência de ξ . Por fim, na ausência de informação externa, o emprego de métodos de estimação menos sensíveis a eventos de maior magnitude, como o método dos momentos-L, pode ser conveniente no trato com variáveis com caudas pesadas (Vogel et al., 2025), com a vantagem adicional de que o coeficiente de curtose-L compreende uma medida trivial dos pesos das caudas nesse caso.

Percebe-se do exposto que, apesar da argumentação teórica coerente para a especificação de modelos dotados de cauda pesada na análise de frequência de variáveis hidrometeorológicas, severas distorções podem ser introduzidas à inferência estatística pelo mero expediente de



amostragem. Nesse sentido, Vogel et al. (2025) recomendam cautela ao se advogar que “os dados falem por si mesmos” – uma concepção fisicamente embasada do modelo distributivo pode ser necessária antes da inferência para se assegurar a consistência das estimativas e previsões. Além disso, o emprego da teoria sub-assintótica para valores extremos (Serinaldi, Lombardo e Kilsby, 2020) pode ser útil para se estimar o índice de caudas por agregar maior evidência empírica (i.e., observações ordinárias e extremos) à inferência paramétrica.

Caudas pesadas e modelos de simulação hidrológica

A presença de caudas pesadas também exerce significativo impacto na simulação chuva-vazão e em modelos de regionalização hidrológica. De fato, tais modelos usualmente materializam transformações não lineares de variáveis hidrometeorológicas, as quais, assim como nos sistemas físicos, tendem a amplificar o peso das caudas das variáveis de resposta. Nesses casos, estimadores usuais, baseados no método dos mínimos quadrados, podem levar à inferência inconsistente de parâmetros e insuficiente capacidade preditiva dos modelos (Vogel et al., 2025). Ademais, os intervalos de confiança podem ser demasiadamente estreitos, uma vez que esses dependem da hipótese de normalidade assintótica, a qual pode não se verificar no caso de caudas pesadas.

Para modelos de regressão, uma solução trivial para o problema descrito é a transformação dos preditores e das variáveis de resposta (e.g., logarítmica, Box-Cox etc.) para aproximar sua distribuição a um modelo Gaussiano, o qual é descrito por caudas leves e, em grande medida, contorna os problemas de inferência mencionados na seção anterior. Nesses casos, as previsões pontuais e limites de confiança devem ser estimados em espaço transformado e, em seguida, expressos no espaço original mediante transformação inversa.

Por outro lado, ao se lidar com modelos de transformação chuva-vazão, pode ser necessário proceder à estimação de parâmetros com auxílio de um modelo de erros dotado de caudas pesadas – somente a transformação das variáveis de resposta em variáveis com caudas leves, como usualmente preconizado, pode ser insuficiente para produzir estimativas plausíveis para tais parâmetros e previsões confiáveis de vazões. Um exemplo interessante nesse contexto é o modelo de resíduos proposto por Schoups e Vrugt (2010), o qual é baseado na distribuição “*Skew Exponential Power*” (SEP), que possui a distribuição Normal e a distribuição de Laplace como



formas limite. Esse modelo permite reproduzir distintos graus de assimetria e, por óbvio, caudas pesadas, bem como a incerteza preditiva das vazões – note-se que a inferência paramétrica é realizada a partir de preceitos da escola Bayesiana.

A Figura 3 ilustra as funções densidade de probabilidade teóricas (linhas contínuas) e empíricas (pontos) dos resíduos de modelagem, assumindo-se um modelo Gaussiano (Figura 3a) e um modelo SEP dotado de caudas pesadas (Figura 3b). Percebe-se, de maneira clara, que, ao se ignorar a natureza de caudas pesadas dos resíduos, a distribuição empírica não apresenta aderência ao modelo teórico, o que coloca em perspectiva a confiabilidade das estimativas dos parâmetros do modelo hidrológico. Ademais, a distribuição empírica apresenta caudas muito mais leves, em comparação à distribuição Gaussiana, o que compromete a estimação da incerteza preditiva das vazões. Por outro lado, ao se considerar as caudas pesadas dos resíduos de modelagem, as densidades teórica e empírica apresentam maior similaridade e as frequências nas caudas são adequadamente reproduzidas. Cabe notar ainda que as estimativas de parâmetros do modelo hidrológico são distintas daquelas obtidas com a distribuição Normal, o que sugere que o comportamento dos erros não afeta somente a incerteza preditiva, mas também o processo de simulação de vazões.

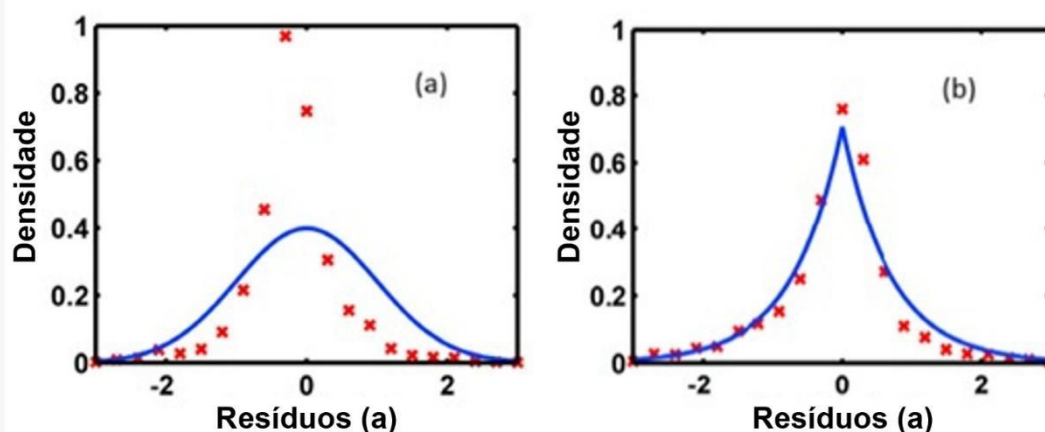


Figura 3 – funções densidade de probabilidade teóricas (linhas contínuas) e empíricas (pontos) dos resíduos de modelagem considerando um modelo Gaussiano (a) e um modelo SEP (b). Fonte:

Schoups a Vrugt (2010)

Mais recentemente, outra linha de modelagem hidrológica, sugerida por Vogel (2017), vem sendo explorada para uso na gestão de risco na área de recursos hídricos. Essa linha, ao invés de tentar



descrever as incertezas de forma separada, em cada fonte de erro envolvida no processo de modelagem hidrológica (e.g. dados de entrada como precipitação, estrutura do modelo, estimativa dos parâmetros, e dados de saída como a vazão), foca na modelagem dos resíduos. De fato, conceitualmente, a ideia é a mesma da apresentada por Shoups e Vrugt (2010), com diferenças na modelagem propriamente dita e na ênfase dada às aplicações práticas dessa abordagem. Tendo um modelo adequado para os resíduos da simulação hidrológica, é possível criar um conjunto grande de simulações hidrológicas que expressem o grau de incerteza sobre o conhecimento que se tem do sistema modelado, permitindo que analistas e tomadores de decisão avaliem, de forma mais ampla e com visão de gestão de risco, o desempenho de sistemas hídricos, bem como as possíveis repercussões de um conjunto de ações a serem adotadas na bacia.

À medida que essa abordagem vem sendo mais utilizada (e.g. Shabestanipour et al., 2023, 2024; Brodeur, Gupta e Steinschneider, 2023; Brodeur et al., 2024;), percebe-se fortemente a presença de resíduos com cauda pesada, especialmente se a escala temporal de simulação for diária ou sub-diária. Nesses casos, além de ter de lidar com a presença de sazonalidade, correlação temporal, e assimetria, é essencial que a modelagem dos resíduos leve em consideração a questão de cauda pesada. Caso contrário, o conjunto de simulações de um sistema pode ficar comprometido.

Outro aspecto da modelagem hidrológica que é influenciado pela presença de distribuições com cauda pesada, e que é muito pouco discutido na comunidade de recursos hídricos, é o desempenho de estimadores de eficiência da simulação, que procuram avaliar a qualidade das simulações provenientes do modelo hidrológico frente à série observada da variável de interesse, normalmente as vazões. Os estimadores mais empregados são o coeficiente de eficiência de Nash-Sutcliffe (NSE) e o coeficiente de eficiência de Kling-Gupta (KGE). Esses estimadores, por serem baseados numa amostra, são variáveis aleatórias, cujo desempenho depende de um conjunto de características do processo estocástico, e consequentemente dos estimadores dessas características.

Lamontagne, Barber e Vogel (2020) apresentaram uma abrangente discussão acerca desses estimadores, e recomendaram alternativas que seguem a mesma lógica dos estimadores originais, porém são mais adequados a séries hidrológicas, levando em consideração a presença de sazonalidade, correlação temporal, e assimetria, bem como a própria distribuição de



probabilidades das vazões diárias, que certamente afetam o desempenho de tais estimadores. Os autores demonstraram que esses novos estimadores são mais eficientes em estimar a qualidade das simulações do que os estimadores originais, e procuraram deixar clara a ideia de que qualquer estimador de eficiência de modelagem possui incertezas. A questão das incertezas amostrais dos estimadores de eficiência como o NSE e o KGE foi discutida com profundidade por Clark et al. (2021), que mostraram, por meio de uma análise com um conjunto grande amostras, que tais estimadores possuem incertezas amostrais substancialmente grandes. Isso ocorre porque a distribuição de probabilidade dos erros quadráticos entre as simulações e observações possui cauda pesada, o que significa dizer que o desempenho de tais métricas é fortemente afetado por um conjunto pequeno de observações extremas.

Considerações finais

A presença de caudas pesadas em variáveis hidrometeorológicas é amplamente embasada por fundamentos físicos e estatísticos, em especial o princípio de máxima entropia. Contudo, é ainda prática corrente a utilização de distribuições com caudas exponenciais para a modelagem dessas variáveis, tanto no contexto de pesquisa quanto de projetos. Nesses casos, a ocorrência relativamente frequente de eventos extremos, aqui entendidos como aquelas observações com valores consideravelmente superiores à média, tem sido, em grande medida, atribuída erroneamente à intensificação de forçantes externas (e.g., mudanças climáticas), em lugar da especificação incorreta do modelo distributivo. Tal interpretação pode incorrer na subestimação dos riscos inerentes ao dimensionamento e operação de estruturas hidráulicas, com severas consequências sociais, econômicas e ambientais.

Diante desse cenário, a literatura recente tem buscado ampliar a compreensão acerca da gênese e da propagação do decaimento polinomial das caudas em diferentes componentes dos sistemas hidrológicos, como a atmosfera e as redes fluviais, de maneira a reconciliar princípios físicos e aspectos estatísticos, especialmente no trato com os extremos de variáveis hidroclimáticas (Merz et al., 2022; Koutsoyiannis, 2024; Vogel et al., 2025). Em paralelo, os grandes desafios associados à estimação paramétrica e à quantificação de incertezas de modelos distributivos dotados de caudas pesadas têm sido tratados sob diferentes paradigmas, como abordagens empíricas (e.g., Papalexiou e Koutsoyiannis, 2013) e a introdução de conhecimento físico *a priori* à inferência (e.g.,



Costa e Sampaio, 2021), o que tem conferido maior confiabilidade às estimativas de quantis e à extrapolação. Em conjunto com o aumento das amostras disponíveis, tais estratégias têm tornado mais natural a utilização de distribuições com caudas polinomiais na prática da Hidrologia.

Por outro lado, o emprego de distribuições com caudas pesadas para a descrição dos resíduos de modelos de transformação chuva-vazão e em expedientes de previsão é ainda insipiente, em comparação à análise de frequência, e desenvolvimentos de pesquisa são necessários nesse contexto. Com efeito, iniciativas recentes, incluindo-se aquelas baseadas em modelos estruturalmente complexos, têm demonstrado que, ao se negligenciar o decaimento polinomial das caudas dos resíduos, estimativas e predições pouco acuradas dos eventos mais extremos são obtidas.

Ademais, a quantificação de incertezas nesses modelos permanece pouco explorada, o que, em grande medida, oculta os efeitos de propagação das caudas pesadas nas estimativas dos resíduos. Abordar tais limitações em investigações futuras parece necessário para se comunicar com transparência as reais habilidades descritivas e preditivas de modelos correntemente em uso.

A discussão introduzida neste artigo buscou elucidar aspectos teóricos e práticos associados a variáveis com caudas pesadas, com vistas a facilitar o trânsito entre desenvolvimentos acadêmicos e aplicações de engenharia. Espera-se que, ao se alterar o foco de projetos e modelos às variáveis com caudas pesadas, seja possível contribuir para a melhor compreensão de estados mais extremos dos processos hidroclimáticos e para a quantificação mais rigorosa dos riscos associados.

Bibliografia

Brodeur, Z. P., Gupta, R. S., Steinschneider, S. A multisite stochastic watershed model (SWM) with intermittency for regional low flow and flood risk analysis. ESS Open Archive, 2023. <https://doi.org/10.22541/essoar.169168911.11270226/v1>

Brodeur, Z., Wi, S., Shabestanipour, G., Lamontagne, J., Steinschneider, S. A Hybrid, Non-Stationary Stochastic Watershed Model (SWM) for Uncertain hydrologic simulations under climate change. Water Resources Research, 60(5), e2023WR035042, 2024. <https://doi.org/10.1029/2023wr035042>

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., Papalexiou, S. M. The Abuse of popular performance metrics in hydrologic modeling. Water Resources Research, 57(9), e2020WR029001, 2021. <https://doi.org/10.1029/2020wr029001>



Costa, V., Fernandes, W. Bayesian estimation of extreme flood quantiles using a rainfall-runoff model and a stochastic daily rainfall generator. *Journal of Hydrology*, 554, 137-154, 2017. <https://doi.org/10.1016/j.jhydrol.2017.09.003>

Costa, V., Sampaio, J. Bayesian approach for estimating the distribution of annual maximum floods with a mixture model. *Journal of Hydrologic Engineering*, 26(6), 04021017, 2021. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.000209](https://doi.org/10.1061/(ASCE)HE.1943-5584.000209)

Gumbel, E. J. *Statistics of Extremes*, Columbia University Press, New York, 375 p., 1958.

Koutsoyiannis, D. *Stochastics of Hydroclimatic Extremes - A Cool Look at Risk*, Edition 4, ISBN: 978-618-85370-0-2, 400 p., Kallipos Open Academic Editions, Athens, 2024. 10.57713/kallipos-1

Laaha, G., Laimighofer, J., Özcelik, N. B., & Fischer, S. Exploring Process Heterogeneity in Environmental Statistics: Examples and Methodological Advances. *Austrian Journal of Statistics*, 54(3), 124–149, 2025. <https://doi.org/10.17713/ajs.v54i3.2101>

Lamontagne, J. R., Barber, C. A., Vogel, R. M. Improved estimators of model performance efficiency for skewed hydrologic data. *Water Resources Research*, 56(9), e2020WR027101, 2020. <https://doi.org/10.1029/2020wr027101>

MacDonald, E., Merz, B., Guse, B., Nguyen, V. D., Guan, X., Vorogushyn, S. What controls the tail behaviour of flood series: rainfall or runoff generation? *Hydrology and Earth System Sciences*, 28(4), 833–850, 2024. <https://doi.org/10.5194/hess-28-833-2024>

MacDonald, E., Merz, B., Guse, B., Wietzke, L., Ullrich, S., Kemter, M., Ahrens, B., Vorogushyn, S. Event and Catchment Controls of Heavy Tail Behavior of Floods. *Water Resources Research*, 58(6), e2021WR031260, 2022. <https://doi.org/10.1029/2021wr031260>

Marra, F., Dallan, E., Canale, A., Prosdocimi, I., Papacharalampous, G., Borga, M., & Papalexiou, S. M. Apparent Heavy Tails of Sub-Daily Precipitation Explained by the Coexistence of Lighter-Tailed Processes. *Geophysical Research Letters*, 53(2), 2026. <https://doi.org/10.1029/2025gl119705>

Martins, E. S., Stedinger, J. R. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737–744, 2000. <https://doi.org/10.1029/1999wr900330>

Merz, B., Basso, S., Fischer, S., Lun, D., Blöschl, G., Merz, R., Guse, B., Viglione, A., Vorogushyn, S., MacDonald, E., Wietzke, L., Schumann, A. Understanding heavy tails of flood peak distributions. *Water Resources Research*, 58(6), e2021WR030506, 2022. <https://doi.org/10.1029/2021WR030506>

Nguyen, T., Kravitz, B., O'Brien, T. A., Ficklin, D. L., Rasmussen, K. L., Kruczkiewicz, A., et al.. Future Intensity-Duration-Frequency Curves of Extreme Precipitation in the Midwest United States From Convection-Permitting Modeling. *Journal of Geophysical Research: Atmospheres*, 130(18), 2025. <https://doi.org/10.1029/2024jd042798>



Papalexiou, S. M., Koutsoyiannis, D. Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, 49(1), 187-201, 2013. <https://doi.org/10.1029/2012WR012557>

Serinaldi, F., Lombardo, F., KILSBY, C. G. All in order: Distribution of serially correlated order statistics with applications to hydrological extremes. *Advances in Water Resources*, 144, 103686, 2020. <https://doi.org/10.1016/j.advwatres.2020.103686>

Serinaldi, F., Lombardo, F., & Kilsby, C. G. Non-asymptotic distributions of water extremes: much ado about what? *Hydrology and Earth System Sciences*, 29(4), 1159–1181, 2025. <https://doi.org/10.5194/hess-29-1159-2025>

Shabestanipour, G., Brodeur, Z., Farmer, W. H., Steinschneider, S., Vogel, R. M., Lamontagne, J. R. Stochastic watershed model ensembles for long-range planning: verification and validation. *Water Resources Research*, 59(2), e2022WR032201, 2023. <https://doi.org/10.1029/2022wr032201>

Shabestanipour, G., Brodeur, Z., Manoli, B., Birnbaum, A., Steinschneider, S., Lamontagne, J. R. Risk-based hydrologic design under climate change using stochastic weather and watershed modeling. *Frontiers in Water*, 6, 1310590, 2024. <https://doi.org/10.3389/frwa.2024.1310590>

Schoups, G., Vrugt, J. A. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10), W10531, 2010. <https://doi.org/10.1029/2009wr008933>

Vogel, R. M. Stochastic watershed models for hydrologic risk management. *Water Security*, 1, 28–35, 2017. <https://doi.org/10.1016/j.wasec.2017.06.001>

Vogel, R. M., Papalexiou, S. M., Lamontagne, J. R., Dolan, F. C. When heavy tails disrupt statistical inference. *The American Statistician*, 79(2), 221-235, 2025. <https://doi.org/10.1080/00031305.2024.2402898>

Wang, H.-J., Merz, R., Yang, S., Tarasova, L., Basso, S. Emergence of heavy tails in streamflow distributions: the role of spatial rainfall variability. *Advances in Water Resources*, 171, 104359, 2023. <https://doi.org/10.1016/j.advwatres.2022.104359>

Wang, H.-J., Merz, R., & Basso, S. Constructing a geography of heavy-tailed flood distributions: insights from common streamflow dynamics. *Hydrology and Earth System Sciences*, 29(6), 1525–1548, 2024. <https://doi.org/10.5194/hess-29-1525-2025>

Zorzetto, E., Canale, A., Marani, M. A Bayesian non-asymptotic extreme value model for daily rainfall data. *Journal of Hydrology*, 628, 130378, 2024. <https://doi.org/10.1016/j.jhydrol.2023.130378>



Recebido em: 15/04/2026

Aceito em: 15/06/2026

[1] Departamento de Engenharia Hidráulica e Recursos Hídricos da Universidade Federal de Minas Gerais. Grupo de pesquisa hydro_stat. Email: veber@ehr.ufmg.br

[2] Departamento de Engenharia Civil e Ambiental da Universidade de Brasília. Grupo de pesquisa hydro_stat. Email: dirceureis@unb.br