

Modelagem estocástica usando a distribuição generalizada de valores extremos e momentos LH: uma abordagem via software livre R

Stochastic modeling using the generalized distribution of extreme values and LH moments:
an approach through free software R

Wagner Alessandro Pansera¹

Universidade Tecnológica Federal do Paraná (UTFPR), Departamento de Engenharia Civil, Toledo,
PR, Brasil

 <https://orcid.org/0000-0002-6850-6775>,  <http://lattes.cnpq.br/9083466536550741>

Benedito Martins Gomes²

Universidade Estadual do Oeste do Paraná (UNIOESTE), Programa de Pós-Graduação em
Engenharia Agrícola, Cascavel, PR, Brasil

 <https://orcid.org/0000-0003-0223-4049>,  <http://lattes.cnpq.br/4355317240921602>

Resumo: A distribuição de probabilidade generalizada de valores extremos (GEV) é utilizada na modelagem de extremos naturais, tais como: chuvas intensas, vazões máximas, velocidade do vento e temperatura. Uma questão importante na utilização da GEV é a escolha da metodologia de estimativa de parâmetros. As metodologias comumente utilizadas são a máxima verossimilhança e os momentos convencionais. Contudo, estudos apontam que tais metodologias nem sempre produzem estimativas confiáveis dos parâmetros da GEV. Nesse sentido, é interessante a utilização dos momentos LH, pois caracterizam melhor o caudal superior da distribuição devido ênfase dada aos maiores valores observados. Porém, não existem rotinas computacionais desenvolvidas para utilização da GEV com momentos LH em software livre. Dessa forma, esta pesquisa teve por objetivo desenvolver uma rotina computacional no software livre R para modelagem estocástica por meio da GEV, utilizando momentos LH para estimar seus parâmetros e verificar a qualidade do ajuste. Foram utilizados dados de vazões máximas diárias anuais disponíveis na literatura para demonstrar a aplicabilidade da rotina computacional. Esta pesquisa contribui para disseminar o uso dos momentos LH e facilitar modelagem estocástica de eventos ambientais extremos.

Palavras-chave: GEV; valores extremos; software livre R; modelagem estocástica.

Abstract: Generalized Extreme Value (GEV) distribution is used to modeling extreme natural events, such as rainfall, floods, wind speed and temperature. An important issue for GEV use is the choice of parameter estimation methodology. The commonly used methodologies are maximum likelihood and conventional moments. However, studies indicate that such methodologies do not always produce a reliable estimate of GEV parameters. In this sense, it is interesting to use LH moments, as they better characterize the upper tail of the distribution due to the emphasis given to the highest observed values. Nevertheless, there are no computational routines developed for GEV use combined with LH moments in free software. Therefore, this research aimed at developing a computational routine in free software R for stochastic modeling through GEV, using LH moments to estimate its parameters and verify goodness-of-fit. Maximum annual flow data available in the literature was used to demonstrate the applicability of the computational routine. This research contributes to disseminate the use of LH moments and facilitate stochastic modeling of extreme environmental events.

Keywords: GEV; extreme values; free software R; stochastic modeling.

¹ **Currículo sucinto:** Bacharel em Engenharia Agrícola pela Universidade Estadual do Oeste do Paraná, mestre em Engenharia Agrícola pela Universidade Estadual do Oeste do Paraná, doutor em Engenharia Agrícola pela Universidade Estadual do Oeste do Paraná, docente da Universidade Tecnológica Federal do Paraná. **Contribuição de autoria:** Conceituação, Escrita – Primeira Redação, Escrita – Revisão e Edição, Investigação, Metodologia, Software. **Contato:** pansera@utfpr.edu.br.

² **Currículo sucinto:** Bacharel em Engenharia Agrícola pela Universidade Federal de Lavras, mestre em Engenharia Civil pela Escola de Engenharia de São Carlos (USP), doutor em Agronomia pela Universidade Estadual Paulista Júlio de Mesquita Filho, docente da Universidade Estadual do Oeste do Paraná. **Contribuição de autoria:** Conceituação, Supervisão, Validação e Visualização. **Contato:** benedito.gomes@unioeste.br.



Data de submissão: 13 de abril de 2021.

Data de aprovação: 22 de junho de 2021.

1. Introdução

A distribuição generalizada de valores extremos, comumente referida pela sigla GEV da terminologia inglesa *Generalized Extreme Value*, foi proposta por Jenkinson (1955) combinando os três possíveis tipos de distribuição de valores extremos (Fréchet, Gumbel e Weibull) em uma única equação. A distribuição GEV é utilizada para modelar uma extensa variedade de extremos naturais, tais como: vazões máximas, chuvas intensas, velocidade do vento e temperatura (KATZ; PARLANGE; NAVEAU, 2002; ABU EL-MAGD, 2010; PANSERA, 2013).

Uma questão importante na utilização da GEV é a escolha da metodologia de estimativa de parâmetros, pois interfere na precisão dos quantis estimados, principalmente nos caudais da distribuição, onde se encontram os valores desejados. As técnicas frequentemente utilizadas são máxima verossimilhança (COLES, 2001) e momentos convencionais (RAO; HAMED, 2000).

No entanto, segundo Hosking, Wallis e Wood (1985) os estimadores de máxima verossimilhança são instáveis quando aplicados em pequenas amostras, e podem gerar valores absurdos/irreais para o parâmetro de forma (MARTINS; STEDINGER, 2000). Por outro lado, os momentos convencionais são geralmente inferiores em qualidade e não são tão eficientes quanto a máxima verossimilhança, especialmente para distribuições com três ou mais parâmetros (RAO; HAMED, 2000).

Com intuito de evitar esses problemas recomenda-se a utilização dos momentos LH (WANG, 1997), uma generalização dos momentos L (HOSKING, 1990), como estimador dos parâmetros da GEV. Com a utilização dos momentos LH busca-se a caracterização da parte superior da distribuição e dos maiores eventos amostrais.

Os momentos LH são calculados usando combinações lineares das estatísticas de altas ordens, por isso é acrescido um "H" (da língua inglesa "*Higher*") em momentos L. Quando a ordem é igual a zero, os momentos LH têm comportamento idêntico aos momentos L. Para ordens maiores que zero é dado maior peso aos maiores valores observados, pois esses valores contêm informações relevantes sobre a cauda superior da distribuição GEV.

Wang (1997), por meio de simulações de Monte Carlo e análise de vazões máximas anuais, mostrou que os momentos LH, quando comparados com os momentos L, aumentam a precisão na estimativa de quantis com alto período de retorno, pois reduzem a influência dos menores valores observados.



Nesse sentido, os momentos LH possibilitam o ajuste da GEV aos valores observados, desde a forma feita pelos momentos L até ordens mais altas, com o propósito de obter estimativas mais precisas de eventos com alto período de retorno.

Gilleland, Ribatet e Stephenson (2013) apresentam diversos pacotes computacionais para análise de eventos extremos, porém nenhum deles utiliza momentos LH. Queiroz e Chaudhry (2006) apresentam uma rotina computacional em Matlab (software proprietário) para análise de eventos ambientais extremos utilizando a GEV e momentos LH. Contudo, não são encontradas rotinas computacionais em software livre, como a plataforma R.

Dessa forma, o objetivo deste trabalho é desenvolver e aplicar uma rotina computacional em linguagem R, para modelagem estocástica de eventos ambientais extremos, utilizando a distribuição generalizada de valores extremos e momentos LH.

2. Material e métodos

Nesta seção, serão abordados os tópicos relativos ao desenvolvimento do trabalho.

2.1. Distribuição generalizada de valores extremos (GEV)

Pertencente à família das distribuições de probabilidades contínuas, a GEV incorpora em uma expressão três distribuições de valores extremos: Gumbel, Fréchet e Weibull. Essas distribuições também são conhecidas como distribuições de valores extremos tipos I (Gumbel), II (Fréchet) e III (Weibull) (ABU EL-MAGD, 2010). A Equação 1 apresenta a função de probabilidade acumulada da distribuição GEV (JENKINSON, 1955):

$$F(x) = \begin{cases} \exp\left(-\left\{1 - k\left(\frac{x-u}{\alpha}\right)\right\}^{\frac{1}{k}}\right), & \text{se } k \neq 0 \\ \exp\left(-\exp\left(-\frac{x-u}{\alpha}\right)\right), & \text{se } k = 0 \end{cases} \quad (1)$$

em que:

$F(x)$ – função de probabilidade acumulada da distribuição GEV;

$k \in R$ – parâmetro de forma da distribuição;

$\alpha \in R_+^*$ – parâmetro de escala da distribuição;

$u \in R$ – parâmetro de posição da distribuição;

x – variável aleatória analisada.

Conforme apresentado por Naghettini e Pinto (2007), o valor e sinal de k determina a forma assintótica de valores extremos máximos. Quando $k < 0$, a GEV representa a distribuição do Tipo II (Fréchet), definida apenas para $x \in](u+\alpha)/k, +\infty[$. Quando $k > 0$, a GEV representa a distribuição do Tipo III (Weibull), definida apenas para $x \in]-\infty, (u+\alpha)/k]$. Já se $k = 0$, a GEV



corresponde à distribuição do tipo I (Gumbel) com parâmetro de escala α e parâmetro de posição u , definida para $x \in R$.

Os valores de x são os quantis da distribuição GEV, obtidos através da Equação 2 (NAGHETTINI; PINTO, 2007):

$$x(T) = u + \frac{\alpha}{k} \left[1 - \left(-\ln \left(1 - \frac{1}{T} \right) \right)^k \right] \tag{2}$$

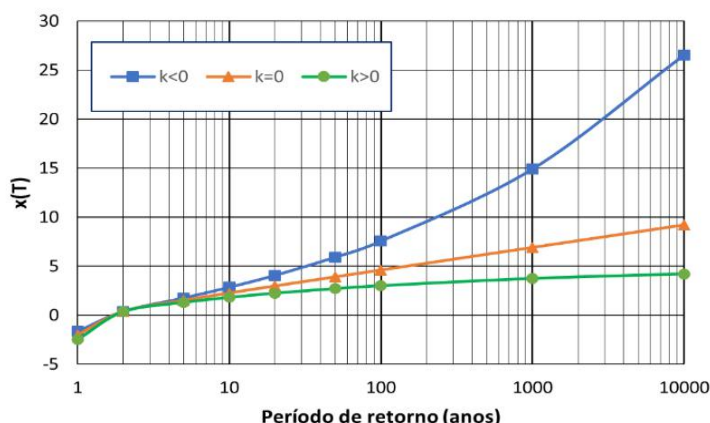
em que:

$x(T)$ – quantil da distribuição GEV;

T – período de retorno (anos).

Na Figura 1, é possível visualizar o comportamento das três possíveis formas da GEV, considerando $\alpha = 1$, $u = 0$ e utilizando a Equação 2 para diferentes períodos de retorno. Nota-se na Figura 1 que: (i) quando $k < 0$, a GEV tem comportamento exponencial; (ii) quando $k = 0$, a GEV tem comportamento linear; e (iii) quando $k > 0$, tem comportamento assintótico.

Figura 1 – Comportamento da GEV em função do parâmetro de forma



Fonte: Elaboração dos autores (2021)

2.2. Momentos LH

Para a utilização dos momentos LH, tem-se uma amostra aleatória ordenada, representada por $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$. A estimativa dos momentos LH amostrais é efetuada de acordo com as Equações 3 a 6 (WANG, 1997):

$$\hat{\lambda}_1^\eta = \frac{1}{n C_{\eta+1}} \sum_{i=1}^n i^{-1} C_\eta x_{(i)} \tag{3}$$

$$\hat{\lambda}_2^\eta = \frac{1}{2} \frac{1}{n C_{\eta+2}} \sum_{i=1}^n \left(i^{-1} C_{\eta+1} - i^{-1} C_\eta {}^{n-1} C_1 \right) x_{(i)} \tag{4}$$



$$\hat{\lambda}_3^\eta = \frac{1}{3} \frac{1}{n C_{\eta+3}} \sum_{i=1}^n \left(i^{-1} C_{\eta+2} - 2^{i-1} C_{\eta+1} n^{-1} C_1 + i^{-1} C_{\eta} n^{-i} C_2 \right) x_{(i)} \tag{5}$$

$$\hat{\lambda}_4^\eta = \frac{1}{4} \frac{1}{n C_{\eta+4}} \sum_{i=1}^n \left(i^{-1} C_{\eta+3} - 3^{i-1} C_{\eta+2} n^{-i} C_1 + 3^{i-1} C_{\eta+1} n^{-i} C_2 - i^{-1} C_{\eta} n^{-i} C_3 \right) x_{(i)} \tag{6}$$

$${}^m C_j = \binom{m}{j} = \frac{m!}{j!(m-j)!} \tag{7}$$

em que:

$\hat{\lambda}_1^\eta, \hat{\lambda}_2^\eta, \hat{\lambda}_3^\eta, \hat{\lambda}_4^\eta$ – momentos LH amostrais;

η – ordem dos momentos LH.

Wang (1997) recomenda a utilização de η no intervalo $[0,4]$. Para $\eta = 1, 2, 3$ e 4 , os momentos LH são denominados de L1, L2, L3 e L4. Para $\eta = 0$, os momentos LH são iguais aos momentos L de Hosking (1990). A utilização de $\eta > 0$ faz com que os momentos LH representem melhor os maiores valores observados, possibilitando um melhor ajuste do caudal superior da GEV (WANG, 1997, QUEIROZ; CHAUDHRY, 2006).

Quando se normaliza os momentos LH, obtém-se os coeficientes de variação, assimetria e de curtose, conforme as Equações 8 a 10, respectivamente:

$$\tau_2^\eta = \frac{\lambda_2^\eta}{\lambda_1^\eta} \tag{8}$$

$$\tau_3^\eta = \frac{\lambda_3^\eta}{\lambda_2^\eta} \tag{9}$$

$$\tau_4^\eta = \frac{\lambda_4^\eta}{\lambda_2^\eta} \tag{10}$$

em que:

τ_2^η – coeficiente de variação-LH;

τ_3^η – coeficiente de assimetria-LH;

τ_4^η – coeficiente de curtose-LH.

Ainda de acordo com Wang (1997), os momentos LH populacionais da GEV são definidos pelas Equações 11 a 14:

$$\lambda_1^\eta = u + \frac{\alpha}{k} [1 - \Gamma(1+k)(\eta+1)^{-k}] \tag{11}$$

$$\lambda_2^\eta = \frac{(\eta+2)\alpha\Gamma(1+k)}{2!k} [-(\eta+2)^{-k} + (\eta+1)^{-k}] \tag{12}$$

$$\lambda_3^\eta = \frac{(\eta+3)\alpha\Gamma(1+k)}{3!k} [-(\eta+4)(\eta+3)^{-k} + 2(\eta+3)(\eta+2)^{-k} - (\eta+2)(\eta+1)^{-k}] \tag{13}$$

$$\lambda_4^\eta = \frac{(\eta+4)\alpha\Gamma(1+k)}{4!k} \left[-(\eta+6)(\eta+5)(\eta+4)^{-k} + 3(\eta+5)(\eta+4)(\eta+3)^{-k} \right. \\ \left. - 3(\eta+4)(\eta+3)(\eta+2)^{-k} + (\eta+3)(\eta+2)(\eta+1)^{-k} \right] \tag{14}$$



em que Γ é função gama completa, definida como $\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$.

2.3. Estimativa de parâmetros

Os parâmetros forma (k), escala (α) e posição (u) da GEV são estimados igualando-se os três primeiros momentos LH amostrais (Equações 3-5) aos seus respectivos populacionais (Equações 11-13), para cada valor de η selecionado (WANG, 1997). Para resolver esse sistema de equações não lineares, deve-se recorrer a métodos numéricos.

No entanto, para facilitar o procedimento computacional, foi utilizada a sugestão de Wang (1997). Dessa forma, o fator de forma foi calculado utilizando a aproximação dada pela Equação 15:

$$k = \alpha_0 + \alpha_1[\tau_3^\eta] + \alpha_2[\tau_3^\eta]^2 + \alpha_3[\tau_3^\eta]^3 \tag{15}$$

Em que: $\alpha_0, \alpha_1, \alpha_2, \alpha_3$ são coeficientes para estimativa do parâmetro k da GEV.

Os coeficientes da Equação 15 variam em função de η , conforme a Tabela 1. Obtidos os valores de k (Equação 15), encontram-se os valores de u e α , através das Equações 11 e 12, respectivamente (QUEIROZ; CHAUDHRY, 2006).

Tabela 1 – Coeficientes para estimativa do parâmetro k da GEV

η	α_0	α_1	α_2	α_3
0	0,2849	-1,8213	0,8140	-0,2835
1	0,4823	-2,1494	0,7269	-0,2103
2	0,5914	-2,3351	0,6442	-0,1616
3	0,6618	-2,4548	0,5733	-0,1273
4	0,7113	-2,5383	0,5142	-0,1027

Fonte: Wang (1997, p. 2843)

2.4. Teste de qualidade de ajuste

O valor de τ_4^η (curtose-LH) e τ_3^η (assimetria-LH) são funções de k (parâmetro de forma), observe as Equações (9-10) e Equações (13-14), além disso, τ_4^η é função de τ_3^η . Devido a essas interdependências, Wang (1998) propôs um teste de qualidade de ajuste baseado no valor estimado da curtose-LH amostral.

Conforme procedimento descrito na seção 2.3, a curtose-LH amostral não é utilizada na estimativa de parâmetros da GEV. Portanto, pode-se realizar um teste de hipótese, comparando a curtose-LH amostral ($\hat{\tau}_4^\eta$) com a populacional (τ_4^η), conforme apresenta a Equação 16, com valores críticos de uma distribuição normal padrão (WANG, 1998).



$$Z_W^\eta = \frac{\hat{\tau}_4^\eta - \tau_4^\eta}{\sigma(\hat{\tau}_4^\eta | \hat{\tau}_3^\eta = \tau_3^\eta)} \tag{16}$$

em que:

Z_W^η – valor para o teste estatístico de Wang (1998);

$\sigma(\hat{\tau}_4^\eta | \hat{\tau}_3^\eta = \tau_3^\eta)$ – desvio padrão de $\hat{\tau}_4^\eta$ condicionado a $\hat{\tau}_3^\eta = \tau_3^\eta$.

O desvio padrão da Equação 16 é função de τ_3^η e do tamanho amostral, sendo obtido por meio de simulações de Monte Carlo (QUEIROZ; CHAUDHRY, 2006, PANSERA, 2013). Como as simulações de Monte Carlo exigem um grande esforço computacional, Wang (1998) propôs a aproximação apresentada na Equação 17.

$$\sigma^2(\hat{\tau}_4^\eta | \tau_3^\eta = \hat{\tau}_3^\eta) = \frac{b}{n} + \frac{c}{n^2} \tag{17}$$

$$b = b_0 + b_1[\tau_3^\eta] + b_2[\tau_3^\eta]^2 + b_3[\tau_3^\eta]^3 + b_4[\tau_3^\eta]^4 \tag{18}$$

$$c = c_0 + c_1[\tau_3^\eta] + c_2[\tau_3^\eta]^2 + c_3[\tau_3^\eta]^3 + c_4[\tau_3^\eta]^4 \tag{19}$$

em que:

n – tamanho amostral;

b, c – coeficientes da Equação (17);

b_0, b_1, b_2, b_3, b_4 – coeficientes em função de η ;

c_0, c_1, c_2, c_3, c_4 – coeficientes em função de η .

Os coeficientes das Equações 19 e 20 variam de acordo com os valores de η , e foram obtidos por Wang (1998). Estes coeficientes são apresentados nas Tabelas 2 e 3.

Tabela 2 – Valores dos coeficientes b para realizar o teste de qualidade de ajuste da GEV

η	b_0	b_1	b_2	b_3	b_4
0	0,0745	0,0555	0,0067	-0,3090	0,2240
1	0,0579	-0,0328	0,1524	-0,4102	0,2672
2	0,0488	-0,0527	0,1620	-0,3856	0,2566
3	0,0380	-0,0309	0,0354	-0,1233	0,0878
4	0,0241	0,0024	-0,0813	0,0733	-0,0210

Fonte: Wang (1998, p. 3499)



Tabela 3 – Valores dos coeficientes c para realizar o teste de qualidade de ajuste da GEV

η	c_0	c_1	c_2	c_3	c_4
0	1,0100	-0,0282	-2,9336	4,0801	-1,0874
1	1,3403	-0,8291	-3,8777	9,5371	-5,7866
2	1,8800	-2,2233	-2,5825	10,4350	-7,3887
3	2,6784	-4,8418	3,5255	2,3736	-3,2076
4	3,7793	-8,3485	11,5170	-7,9095	1,9459

Fonte: Wang (1998, p. 3499)

2.5. Rotina computacional

A rotina computacional desenvolvida em linguagem R (R Core Team, 2021) teve como base a rotina apresentada por Queiroz e Chaudhry (2006), escrita em Matlab. A rotina desenvolvida pode ser utilizada na versão 3.5.2 ou superior. A escolha da linguagem R se deu pelas seguintes razões (WICKHAM, 2015):

- I. O software é gratuito, com código aberto e disponível em diversas plataformas. Por isso, as análises realizadas em R podem ser facilmente replicadas;
- II. O R não é somente uma linguagem de programação, mas também, um ambiente para análise iterativa de dados;
- III. Os recursos de metaprogramação do R permitem que se escrevam funções resumidas rapidamente.

Para transcrever o procedimento de análise de eventos extremos para uma rotina em linguagem R, optou-se por criar duas funções, cujos objetivos são: (i) estimar os momentos LH amostrais; (ii) gerar conjuntos de parâmetros da GEV e realizar teste de qualidade do ajuste proposto por Wang (1998). Nas seções 2.5.1 e 2.5.2 são apresentadas as rotinas das funções citadas.

2.5.1. Função para cálculo dos momentos LH amostrais

O script abaixo apresenta a função “samlh” para o cálculo dos momentos LH amostrais.

```
samlh=function(x, lh){
m=sort(x)
n=length(x)
i=c(1:n)
l1=(1/choose(n, lh+1))*sum(choose(i-1, lh)*m)
l2=.5*(1/choose(n, lh+2))*sum((choose(i-1, lh+1)-choose(i-1, lh))*choose(n-
i, 1))*m)
```




```
l3=(1/3)*(1/choose(n,lh+3))*sum((choose(i-1,lh+2)-2*choose(i-
1,lh+1)*choose(n-i,1)+choose(i-1,lh)*choose(n-i,2))*m)
l4=.25*(1/choose(n,lh+4))*sum((choose(i-1,lh+3)-3*choose(i-
1,lh+2)*choose(n-i,1)+3*choose(i-1,lh+1)*choose(n-i,2)-choose(i-
1,lh)*choose(n-i,3))*m)
t=l2/l1
t3=l3/l2
t4=l4/l2
resp=c(l1,l2,t,t3,t4)
resp}
```

2.5.2. Função para cálculo de parâmetros e teste de qualidade de ajuste da GEV

O script abaixo apresenta a função “pargev” para o cálculo de parâmetros e teste de qualidade de ajuste da GEV. Ressalta-se que a função “pargev” depende da função “samlh”.

```
pargev=function(x){
k=rep(0,5); a=rep(0,5); u=rep(0,5); z=rep(0,5)
N=length(x)
momlh=matrix(0,5,5)
for(i in 1:5){momlh[i,]=samlh(x,i-1)}
Ao=c(0.2849,0.4823,0.5914,0.6618,0.7113)
A1=c(-1.8213,-2.1494,-2.3351,-2.4548,-2.5383)
A2=c(0.8140,0.7269,0.6442,0.5733,0.5142)
A3=c(-0.2835,-0.2103,-0.1616,-0.1273,-0.1027)
for (i in 1:5) {
aux1=i-1
k[i]=Ao[i]+A1[i]*momlh[i,4]+A2[i]*momlh[i,4]^2+A3[i]*momlh[i,4]^3
a[i]=(momlh[i,2]*2*k[i])/((aux1+2)*(-(aux1+2)^(-k[i]))+(aux1+1)^(-
k[i]))*gamma(1+k[i]))
u[i]=momlh[i,1]-a[i]*(1-gamma(1+k[i])*(aux1+1)^(-k[i]))/k[i]}
b=c(0.0745,0.0555,0.0067,-0.3090,0.2240,0.0579,-0.0328,0.1524,-
0.4102,0.2672,0.0488,-0.0527,0.1620,-0.3856,0.2566,0.0380,-
0.0309,0.0354,-0.1233,0.0878,0.0241,0.0024,-0.0813,0.0733,-0.0210)
c=c(1.0100,-0.0282,-2.9336,4.0801,-1.0874,1.3403,-0.8291,-3.8777,9.5371,-
5.7866,1.8800,-2.2233,-2.5825,10.435,-7.3887,2.6784,-
4.8418,3.5255,2.3736,-3.2076,3.7793,-8.3485,11.517, -7.9095,1.9459)
b=matrix(b,5,5,byrow=T)
c=matrix(c,5,5,byrow=T)
for (j in 1:5) {
bo=b[j,1]+b[j,2]*momlh[j,4]+b[j,3]*momlh[j,4]^2+b[j,4]*momlh[j,4]^3+b[j,5]
*momlh[j,4]^4
co=c[j,1]+c[j,2]*momlh[j,4]+c[j,3]*momlh[j,4]^2+c[j,4]*momlh[j,4]^3+c[j,5]
*momlh[j,4]^4
```



```

sig=(bo/N)+(co/(N^2))
t2=(j+1)*a[j]*gamma(1+k[j])*(-(j+1)^(-k[j])+(j^(-k[j]))) / (2*k[j])
t4=(j+3)*a[j]*gamma(1+k[j])*(-(j+5)*(j+4)*(j+3)^(-
    k[j])+3*(j+4)*(j+3)*(j+2)^(-k[j])-3*(j+3)*(j+2)*(j+1)^(-
    k[j])+(j+2)*(j+1)*(j)^(-k[j])) / (4*3*2*k[j])
cct=t4/t2
z[j]=abs(momlh[j, 5]-cct)/(sig^0.5)
mat=cbind(z,0:4,momlh[, 3],momlh[, 4],momlh[, 5],u,a,k)
o=order(z,0:4,momlh[, 3],momlh[, 4],momlh[, 5],u,a,k)
resp=mat[o,]
aux2=c("z","n","cv","ca","cc","u","a","k")
colnames(resp)=aux2
resp}
    
```

2.6. Estudo de caso

Para demonstrar a aplicabilidade das funções “samlh” e “pargev”, realizou-se um estudo de caso utilizando os dados disponibilizados em Naghettini e Pinto (2007), apresentados na Tabela 4, que mostra as vazões médias diárias máximas anuais observadas no rio Hercílio em Ibirama/SC.

Além disso, após aplicar as rotinas computacionais e obter os resultados, foram realizadas as seguintes comparações: (i) teste de Wang (1998) com o teste de Kolmogorov, e (ii) Quantis GEV com quantis Gumbel. Optou-se pela utilização do teste de Kolmogorov, pois ele avalia a diferença máxima entre as funções de probabilidade acumuladas empírica e teórica (NAGHETTINI; PINTO, 2007).

Tabela 4 – Vazões médias diárias máximas anuais observadas no rio Hercílio em Ibirama/SC

Ano	Q (m³/s)	Ano	Q (m³/s)	Ano	Q (m³/s)	Ano	Q (m³/s)	Ano	Q (m³/s)
1935	1342	1945	474	1955	969	1965	708	1975	1406
1936	625	1946	763	1956	566	1966	998	1976	801
1937	619	1947	592	1957	1300	1967	477	1977	741
1938	797	1948	981	1958	526	1968	298	1978	1002
1939	1250	1949	438	1959	520	1969	872	1979	1090

Fonte: Naghettini e Pinto (2007, p. 347)

3. Resultados e discussão

Nesta seção são apresentados os resultados da aplicação da rotina computacional desenvolvida em relação aos dados apresentados na Tabela 4, e também as devidas discussões e comparações.



3.1. Análise preliminar dos dados

Conforme observado na Tabela 4, a amostra estudada conta com dados do período de 1935 a 1984, no entanto, para a realização das análises, os anos com falhas foram desconsiderados. Definiu-se as estatísticas descritivas apresentadas na Tabela 5, as quais sintetizam a distribuição amostral dos dados.

Tabela 5 – Estatísticas amostrais convencionais

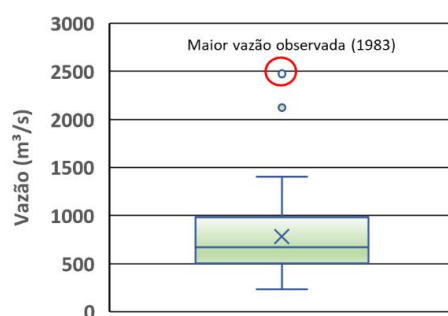
<i>Estatística Amostral</i>	<i>Valor</i>
Média (m ³ /s)	783,7083
Desvio Padrão (m ³ /s)	439,3854
Coefficiente de Variação	0,5606
Coefficiente de Assimetria	1,8456
Coefficiente de Curtose	4,7832

Fonte: Elaboração dos autores (2021)

Ao observar a Tabela 5, percebe-se que o coeficiente de variação obtido é de 0,5606, o que indica alta variabilidade dos dados em relação à média, e que pode aumentar o erro de previsão das séries de vazões. Além disso, conforme mencionado por Naghettini e Pinto (2007), geralmente eventos máximos possuem coeficiente de assimetria positivo, fato observado para a série hidrológica estudada.

Na Figura 2, que apresenta o diagrama do tipo *boxplot*, é possível visualizar, de maneira geral, a dispersão dos dados de vazão apresentados na Tabela 3.

Figura 2 – Boxplot dos dados de vazão



Fonte: Elaboração dos autores (2021)

Em relação à Figura 2, são observadas duas vazões que destoam do restante, principalmente a vazão referente ao ano de 1983, sendo a maior vazão observada. Ou seja, tais valores se afastam muito da média e, conseqüentemente, são responsáveis pela assimetria e



curtose observadas na Tabela 4. Isso demonstra a aplicabilidade dos momentos LH, pois a metodologia dá maior peso aos maiores valores observados.

3.2. Estimativa de parâmetros e avaliação do ajuste da GEV

A Tabela 6 apresenta os resultados da aplicação das rotinas “samLh” e “pargev” à série de vazões (Tabela 4). Ainda nessa tabela, são apresentados os valores para o teste de Kolmogorov (KS).

Tabela 6 – Ajuste da distribuição GEV à série de vazões

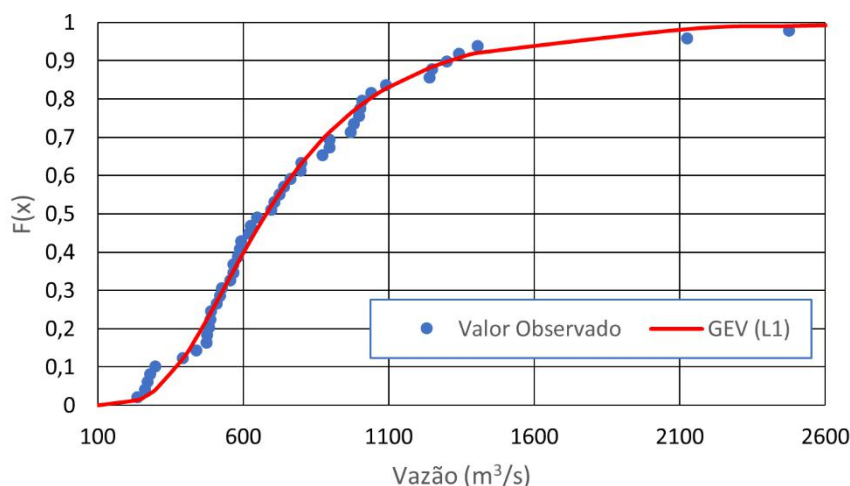
η	$\hat{\lambda}_1^\eta$	$\hat{\lambda}_2^\eta$	τ_2^η	τ_3^η	τ_4^η	u	α	k	Zw	KS
0	1008,1826	214,5910	0,2864	0,2746	0,2131	575,3920	272,967	-0,1598	0,2505	0,0849
1	783,7083	224,4743	0,2128	0,3401	0,2131	577,6823	266,1557	-0,1729	0,1604	0,0780
2	1151,2433	215,4647	0,1872	0,3632	0,2244	579,9418	262,3082	-0,1796	0,3749	0,0727
3	1258,9756	218,6876	0,1737	0,3790	0,2392	586,6140	253,7913	-0,1931	0,6409	0,0771
4	1346,4507	222,9655	0,1656	0,3934	0,2506	599,8966	240,0492	-0,2139	0,7393	0,0910

Fonte: Elaboração dos autores (2021)

Em relação ao teste de Wang (1998), Tabela 6, é possível visualizar que para os cinco conjuntos de parâmetros obtidos o resultado é inferior ao valor teórico de 1,96. Já em relação ao teste de Kolmogorov, valor teórico de 0,1963, percebe-se que todos os valores obtidos estão abaixo. Dessa forma, os testes de Wang (1998) e de Kolmogorov indicam que os ajustes da GEV aos dados de vazões médias diárias máximas anuais não foram rejeitados ao nível de 5% de significância.

Na Figura 3 é apresentado o ajuste da GEV aos dados amostrais para $\eta = 1$ ou L1.

Figura 3 – Melhor ajuste da distribuição GEV à série de vazões



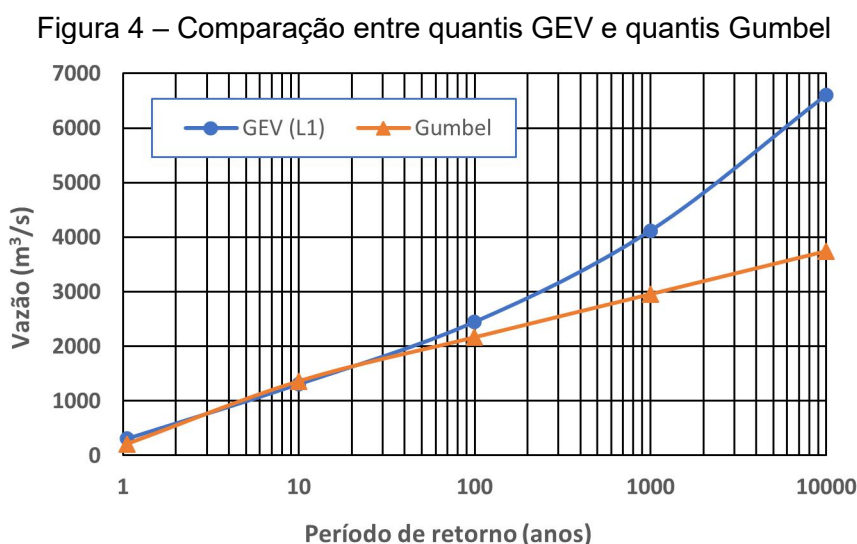
Fonte: Elaboração dos autores (2021)



A escolha do conjunto de parâmetros de “melhor” ajuste foi realizada em função do menor Z_w , conforme sugestão de Queiroz e Chaudhry (2006), o que é verificado para $\eta = 1$ ou L1. Na Figura 3, é possível verificar que os dados amostrais se ajustaram muito bem ao conjunto de parâmetros selecionado.

3.3. Comparativo entre quantis das distribuições GEV e Gumbel

Após a escolha do melhor ajuste da GEV, gerou-se os quantis da distribuição. Foram obtidos também os quantis através da distribuição de Gumbel, conforme método dos momentos convencionais apresentado por Naghettini e Pinto (2007). A Figura 4 apresenta os quantis de cada uma das distribuições. Observa-se que para até aproximadamente 100 anos de tempo de retorno os valores obtidos para vazão são próximos para as duas distribuições. No entanto, conforme aumenta o tempo de retorno, os quantis ficam com valores mais distantes, pois a GEV assume a forma da distribuição Fréchet que possui um caudal mais pesado que a Gumbel.

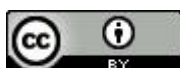


Fonte: Elaboração dos autores (2021)

4. Considerações finais

As rotinas computacionais em linguagem R fornecem, como resultados, as estimativas dos três parâmetros da distribuição GEV, as taxas de momentos LH, além dos valores do teste de Wang (1998), que avalia a qualidade do ajuste para cada ordem (η).

A distribuição GEV, combinada com momentos LH, forneceu ajustes adequados para os dados de vazão apresentados no estudo de caso, conforme indicaram os testes de aderência de Wang e de Kolmogorov. Isso também foi observado por Queiroz e Chaudhry (2006).



Para períodos de retornos maiores que 100 anos, a GEV apresenta valores muito diferentes da Gumbel. Isso conduz a duas considerações no estudo de vazões: (i) para baixos períodos de retorno, pode-se utilizar tanto Gumbel como GEV; (ii) para elevados períodos de retorno deve-se dar preferência na utilização da GEV.

Referências

- ABU EL-MAGD, N. A. T. TL-moments of the exponentiated generalized extreme value distribution. **Journal of Advanced Research**, v. 1, n. 4, p. 351-359, 2010. DOI: <https://doi.org/10.1016/j.jare.2010.06.003>.
- COLES, S. **An introduction to statistical modeling of extreme values**. London: Springer, 2001. 209 p.
- GILLELAND, E.; RIBATET, M.; STEPHENSON, A. G. A software review for extreme value analysis. **Extremes**, v. 16, n. 1, p. 103-119, 2013. DOI: <https://doi.org/10.1007/s10687-012-0155-0>.
- HOSKING, J. R. M. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 52, n. 1, p. 105-124, 1990. Disponível em: <https://www.jstor.org/stable/2345653>. Acesso em: 9 ago. 2021.
- HOSKING, J. R. M.; WALLIS, J. R.; WOOD, E. F. Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. **Technometrics**, v. 27, n. 3, p. 251-261, 1985. DOI: <https://doi.org/10.1080/00401706.1985.10488049>.
- JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, v. 81, n. 348, p. 158-171, abr. 1955. DOI: <https://doi.org/10.1002/qj.49708134804>.
- KATZ, R. W.; PARLANGE, M. B.; NAVEAU, P. Statistics of extremes in hydrology. **Advances in Water Resources**, v. 25, n. 8-12, p. 1287-1304, ago./dez. 2002. DOI: [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8).
- MARTINS, E. S.; STEDINGER, J. R. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. **Water Resources Research**, v. 36, n. 3, p. 737-744, 1 mar. 2000. DOI: <https://doi.org/10.1029/1999WR900330>.
- NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia estatística**. Belo Horizonte: CPRM. 2007. 552 p.
- PANSERA, W. A. **Distribuição generalizada de chuvas máximas no Estado do Paraná**. Orientador: Benedito Martins Gomes. Coorientadores: Marcio Antonio Vilas Boas e Miguel Angel Uribe-Opazo. 2013. 94 f. Tese (Doutorado em Engenharia Agrícola) - Programa de Pós-Graduação em Engenharia Agrícola, Universidade Estadual do Oeste do Paraná, Cascavel, 2013. Disponível em: <http://tede.unioeste.br/handle/tede/2626>. Acesso em: 9 ago. 2021.
- QUEIROZ, M. M. F.; CHAUDHRY, F. H. Analysis of extreme hydrological events using GEV distribution and LH moments. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 10, n. 2, p. 381-389, 2006. DOI: <https://doi.org/10.1590/S1415-43662006000200020>.



R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2021. Disponível em: <http://www.R-project.org>. Acesso em: 27 abr. 2021.

RAO, A. R.; HAMED, K. H. **Flood Frequency Analysis**. Boca Raton, Flórida: CRC Press, 2000. 350 p.

WANG, Q. J. Approximate goodness-of-fit tests of fitted generalized extreme value distributions using LH moments. **Water Resources Research**, v. 34, n. 12, p. 3497-3502, 1 dez. 1998. DOI: <https://doi.org/10.1029/98WR02364>.

WANG, Q. J. LH moments for statistical analysis of extreme events. **Water Resources Research**, v. 33, n. 12, p. 2841-2848, 1997. Disponível em: <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/97WR02134>. Acesso em: 9 ago. 2021.

WICKHAM, Hadley. **Advanced R**. Boca Raton, Florida: CRC, 2015. 456 p.

