

# MODELAÇÃO DE PRECIPITAÇÃO MÁXIMA NA BACIA HIDROGRÁFICA DO RIO VEZ

## MAXIMUM PRECIPITATION MODELING IN THE VEZ RIVER WATERSHED

Maria da Glória Gonçalves<sup>1</sup>

### Introdução

A Teoria dos Valores Extremos (TVE) é um ramo da estatística que investiga os valores usualmente denominados de *outliers* e que são, muitas vezes, ignorados ou mesmo eliminados ao ajustar um modelo estatístico aos dados em análise. Deste modo, o estudo dos valores extremos investiga as observações, de um conjunto de dados, que se localizam longe do seu centro, contrariando o que é usual na maioria das aplicações da estatística, cujas metodologias analisam o comportamento dos dados em torno da sua média e/ou mediana. Contudo, estes acontecimentos, mesmo não sendo os que mais frequentemente ocorrem, podem ter um impacto negativo, ou mesmo catastrófico, com consequências desastrosas para a sociedade e, como tal, é fundamental ter ferramentas metodológicas disponíveis para os quantificar. A TVE possibilita a modelação e análise de eventos raros (baixa frequência) mas de grande impacto, eventos extremos, como as cheias, que têm grande repercussão na sociedade. Podemos, então, mencionar que é a teoria de modelação de eventos que ocorrem com probabilidade muito pequena, consideradas situações atípicas, mas que podem ter impactos significativos sociais e económicos.

Um valor extremo é aquele que não se encontra incluído no intervalo de valores previamente definido como “normal”, ou seja, é um valor que só raramente é atingido, um acontecimento raro. O estudo do comportamento destes valores, refere-se ao estudo do comportamento da cauda da distribuição de interesse. Segundo o Teorema de Fisher-Tippett (lei limite para extremos), os valores extremos (após normalização, caso existam as sucessões de constantes

---

<sup>1</sup> Centro de Estudos Geográfico, Universidade de Lisboa, R. Branca Edmée Marques, 1600-276, Lisboa, Portugal. ORCID: <https://orcid.org/0000-0003-3627-5404>. E-mail: [goncalves.gloria@gmail.com](mailto:goncalves.gloria@gmail.com).

normalizadoras e não convirja para uma variável aleatória degenerada) só convergem para uma das três distribuições limite possíveis: *Gumbel*, *Fréchet* e *Weibull*, que apresentam formas de comportamento distintas da cauda da distribuição de interesse. Estas três famílias de distribuição podem, no entanto, ser representadas numa única expressão, designada por distribuição Generalizada de Valores Extremos (GEV).

Sendo as excedências do nível  $u$  as observações com valor superior a este nível  $u$  e os excessos a diferença entre o valor dessas observações e nível  $u$ , a distribuição dos excessos, condicionada à ocorrência de uma observação acima do nível elevado  $u$ , é caracterizada (se pertencer ao domínio de atração da GEV) por uma distribuição de Pareto Generalizada (GP), a denominada Teorema de *Pickands-Balkema-de Haan* (Alves; Rosário, 2015). Tal como a GEV, a GP pode ser separada em três famílias de acordo com o parâmetro da forma (*Exponencial*, *Pareto* e *Beta*). Um resultado crucial para a aplicação da metodologia POT, para a qual a escolha do valor do nível  $u$  é determinante.

Ora, os parâmetros de maior interesse relacionados com a cauda da distribuição são o índice de cauda ( $\alpha$ ), que estão associados à frequência com que ocorrem os eventos extremos, e os quantis elevados ( $Q_{1-p}$ ), níveis que são excedidos com probabilidade reduzida ( $p$ ).

A inferência estatística também se baseia em observações extremas para modelar acontecimentos raros (estimadores de índice de cauda e de quantis de ordem elevada ou reduzida), sendo que as principais formas de abordar o problema são a modelação por blocos (máximo e/ou mínimo por blocos) e a modelação das excedências de um nível elevado.

Atualmente, são muitas as áreas de aplicação da TVE, como por exemplo nas apólices de seguros multirriscos, na hidrologia ou hidráulica [consulte-se, por exemplo, Beirlant *et al.* (2004), Castillo *et al.* (2005), Embrechts *et al.* (2001), Gomes (2019), Reiss e Thomas (2007)]. Assim, entre outros exemplos, a TVE é aplicada na modelação de dados meteorológicos extremos ou catastróficos com consequências para a sociedade, tais como a ocorrência de cheias devido a níveis de precipitação intensos [citando alguns autores, Alves e Rosário (2015), Neves (2010), Reis (2014), Rosário (2013)]. Por exemplo extremos pode ser aqui apontado, na estimação da probabilidade de o volume de precipitação ultrapassar um determinado limite que coloque em risco um reservatório de água de uma barragem.

Com base em Alves e Rosário (2015), aplicamos estas abordagens aos dados de precipitação recolhidos na estação meteorológica de Casal Soeiro (centro urbano da Bacia Hidrográfica portuguesa do rio Vez [BHRVez]), para estimar o período de retorno para determinados níveis de precipitação, bem como os níveis de retorno a  $T$ -anos ou associados a uma probabilidade de excedência diminuta.

Portanto, com base em observações diárias entre 1961 e 1989, para a estimação dos parâmetros, utilizamos as observações mais elevadas anuais, as cinco e as dez mais elevadas observações de cada ano e ainda os valores máximos mensais acima de determinado limiar. Para os cálculos recorreremos ao *software RStudio*, sobretudo aos *packages evd, evir, fExtremes e ismev*, obtendo-se as explicações dos algoritmos em: Coles (2001), Gilleland (2018), Gomes *et al.* (2013), Pfaff *et al.* (2018), Rizzo (2019), Wuertz *et al.* (2017) e Stephenson (2018).

## Utilização e explicação da Teoria dos Valores Extremos

A teoria de valores extremos é fundamental para a modelagem de eventos raros de precipitação, sendo estes os responsáveis por desencadear as cheias. Assim, sendo, passamos a uma breve explicação da teoria. Seja  $X_1, \dots, X_n$  uma sequência de variáveis aleatórias (va) contínuas, independentes e identicamente distribuídas (iid) com função de distribuição (fd). Assim, se  $X_{i:n}$  representar a  $i$ -ésima estatística ordinal, com  $X_{1:n} = \min\{X_1, \dots, X_n\}$  e  $X_{n:n} = \max\{X_1, \dots, X_n\}$ , então pretendemos analisar o comportamento de  $X_{i:n}$  que seja próximo de zero ou da unidade, em particular do mínimo  $X_{1:n}$  e do máximo  $X_{n:n}$ . Por este motivo, a distribuição GEV para os máximos<sup>2</sup> desempenha um papel central nesta teoria. A GEV tem fd, conforme a Eq. 1 com parâmetros de localização  $\mu \in \mathbb{R}$ , de escala  $\sigma \in \mathbb{R}^+$  e de forma  $\gamma \in \mathbb{R}$ .

$$G_\gamma(x) = \exp\left(-\left(1 + \gamma \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\gamma}}\right), \quad 1 + \gamma \frac{x - \mu}{\sigma} > 0. \quad \text{Eq. 1}$$

Esta distribuição desempenha um papel central na TVE (comparável, em termos de relevância, ao Teorema Limite Central da estatística clássica), uma vez que se o máximo ( $X_{n:n}$ ) de uma sequência de  $n$  variáveis aleatórias iid  $X_1, \dots, X_n$ , após normalização, convergir para uma distribuição não degenerada quando  $n$  aumenta ( $n \rightarrow \infty$ ), então terá de convergir para a distribuição GEV (Teorema do Valor Extremo ou *Teorema de Fisher-Tippett*). Neste caso dizemos que a função  $F$  pertence ao domínio de atração (DA) de  $G$  (isto é,  $F \in DA(G)$ ). Além disso, esta distribuição pode ser dividida em três famílias consoante o valor do parâmetro (índice de valores extremos) que mede o peso da cauda-direita da função de sobrevivência  $G$ , ou seja,  $G(x) := 1 - G(x)$ .<sup>2</sup> Deste modo, obtemos as seguintes distribuições de valores ditos de extremos:

a) se  $\gamma = 0$  (no sentido  $\gamma \rightarrow 0$ ) obtemos uma distribuição de Gumbel (Tipo I) com fd ilustrada na Eq. 2.

$$\Lambda(x) = \exp \left\{ -e^{-\frac{x-\mu}{\sigma}} \right\}, \quad x \in \mathbb{R} \quad \text{Eq. 2}$$

b) se  $\gamma > 0$  obtemos uma distribuição de Fréchet (Tipo II) com fd (Eq. 3),

$$\Phi_{\alpha}(x) = \begin{cases} 0 & , \text{ se } x \leq \mu \\ \exp \left\{ -\left(\frac{x-\mu}{\sigma}\right)^{-\alpha} \right\} & , \text{ se } x > \mu \end{cases}, \quad \alpha \in \mathbb{R}^+ \quad \text{Eq. 3}$$

considerando  $\alpha = \frac{1}{\gamma}$ .

c) se  $\gamma < 0$  obtemos uma distribuição de Weibull (Tipo III) com fd (Eq. 4).

$$\Psi_{\alpha}(x) = \begin{cases} \exp \left\{ -\left(-\left(\frac{x-\mu}{\sigma}\right)\right)^{\alpha} \right\} & , \text{ se } x < \mu \\ 1 & , \text{ se } x \geq \mu \end{cases}, \quad \alpha \in \mathbb{R}^+ \quad \text{Eq. 4}$$

considerando  $\alpha = \frac{1}{\gamma}$ .

Se o objetivo for analisar os excessos  $Y_i = X_i - u$ , onde  $u$  representa um nível elevado da  $va X$ , então o *Teorema de Pickands-Balkema-de Haan* garante que a distribuição de  $X$ , condicionada a  $X > u$ , isto é,  $F_u(y) = P(Y \leq y | X > u)$ , converge para a distribuição GP se  $F \in DA(G)$ . Por conseguinte, outra distribuição central na TVE é a distribuição de GP com fd dada pela Eq. 5.

$$G_{\gamma, \sigma_u}(x) = 1 - \left(1 + \gamma \frac{x-\mu}{\sigma_u}\right)^{-\frac{1}{\gamma}} = \begin{cases} 1 - \left(1 + \gamma \frac{x}{\sigma_u}\right)^{-\frac{1}{\gamma}} & , \text{ se } \gamma > 0, \quad x \in \mathbb{R}_0^+ \\ 1 - e^{-\frac{x}{\sigma_u}} & , \text{ se } \gamma = 0, \quad x \in \mathbb{R}_0^+ \\ 1 - \left(1 + \gamma \frac{x}{\sigma_u}\right)^{-\frac{1}{\gamma}} & , \text{ se } \gamma < 0, \quad 0 \leq x \leq -\frac{\sigma_u}{\gamma} \end{cases} \quad \text{Eq. 5}$$

Onde, os parâmetros de escala  $\sigma_u \in \mathbb{R}^+$  (que depende do nível  $u$  considerado) e de forma  $\gamma \in \mathbb{R}$ . Esta distribuição, consoante o valor do parâmetro de forma, inclui três famílias: a distribuição exponencial (Tipo I) quando  $\gamma = 0$  (no sentido  $\gamma \rightarrow 0$ ), a distribuição de Pareto (Tipo II) quando  $\gamma > 0$  e a distribuição Beta (Tipo III) quando  $\gamma < 0$ .

Na aplicação a dados de precipitação, consideremos  $X_1, \dots, X_n$  as observações diárias de chuva e  $Y_1, \dots, Y_k$  o máximo anual com distribuição  $F_Y$ . Dois conceitos centrais na análise pretendida são o período de retorno para o nível  $u$  (período  $T$  médio, em anos, em que o máximo anual ultrapassa o nível  $u$ , ou seja, o valor esperado de tempo entre duas ocorrências de valores superiores ao nível  $u$ ), dado pela Eq. 6.

$$T = \frac{1}{P(Y > u)} = \frac{1}{1 - F_Y(u)} \quad \text{Eq. 6}$$

Onde, o nível de retorno a T-anos (nível de precipitação que o máximo ultrapassa, em média, todos os  $T$  anos, que representa o valor que  $Y$  excede com probabilidade  $\frac{1}{T}$  em cada ano) determinado pela Eq. 7.

$$U(T) = F_Y^{\leftarrow} \left( 1 - \frac{1}{T} \right) = Q_{1-\frac{1}{T}} \quad \text{Eq. 7}$$

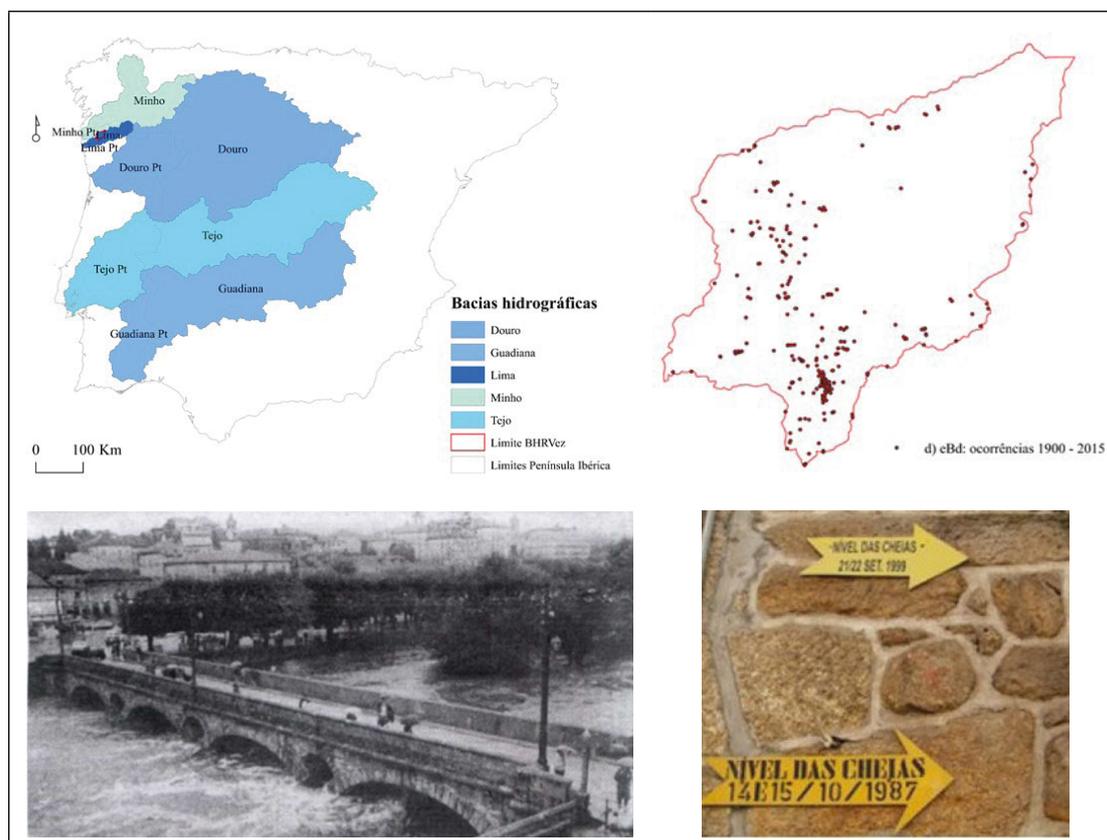
Onde,  $F^{\leftarrow}$  representa a função inversa generalizada da função  $F$  ( $F^{\leftarrow}(p) = \inf\{x: F(x) \geq p\}$ ) e  $Q_p$  o quantil com probabilidade  $p$ . Ora, analisamos estas características associadas a probabilidades de excedência pequenas, isto é, a valores de  $u$  e de  $p = P(X > u)$  tais que  $u > x_{n,n}$  (valores superiores  $u$  ao máximo da amostra).

### Fonte de dados e metodologia

Com vista à aplicação das metodologias da TVE, a amostra em análise compreende níveis de precipitação diários (*mm*) da BHRVez, mais propriamente dados da estação meteorológica de Casal Soeiro, localizada na Vila de Arcos de Valdevez, e referentes ao período de 01/10/1960 a 30/09/1990 (Figura 1), sendo conhecidas diversas ocorrências de cheias. Estes dados são disponibilizados pelo Sistema Nacional de Informação de Recursos Hídricos, e encontram-se disponíveis em <http://snirh.pt>. Sendo que, esta amostra tem uma dimensão de 10957 níveis pluviométricos diários, sem valores omissos.

A estação de monitorização de Casal Soeiro, segundo o histórico de dados, é a estação com os mais elevados registos de níveis de precipitação em Portugal. Por este motivo, esta região é uma ótima candidata à aplicação metodológica de uma análise dos valores extremos. Outro dos fatores determinantes do clima é a orografia da região, Noroeste (Minho) de Portugal, a ultrapassarem os 1000 metros de altitude. Esta região regista valores mais elevados de precipitação,

atingindo uma média de precipitação anual acumulada, em alguns locais, de valores superiores a 3000 mm, o que, por vezes, leva à ocorrência de fenómenos extremos associados aos excessos de precipitação (cheias), com bastantes consequências nefastas às populações que aí residem. Além disso, esta bacia hidrográfica é conhecida pela sua camada litológica granítica que contribui para o escoamento superficial mais rápido.



Fonte: Fotografia obtida do Jornal de Notícias, 1998. Mapa criado a partir de dados do Instituto Geográfico do Exército (1996, 1997); concebido a partir da Imprensa Periódica (1900/2015).

**Figura 1.** Localização da BHRVez (à direita); as ocorrências dos centroides das cheias entre 1900 e 2015 (à direita), rio Vez (em baixo do lado esquerdo) observando-se a cheia de 11/11/1987; níveis máximos dos eventos fixos numa das paredes de um edifício (lado direito).

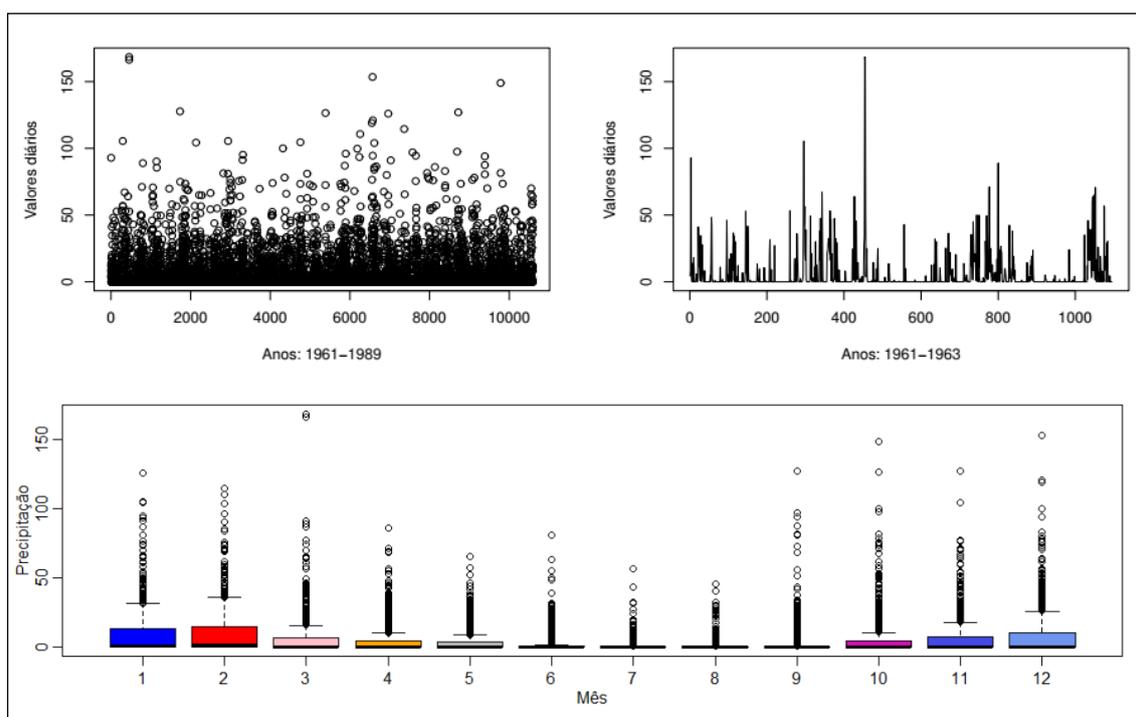
Para a aplicação da TVE não foram consideradas todas as observações, pois dados ambientais diários não são, geralmente, independentes. Isto significa que, um dia de chuva tem uma maior probabilidade de ser seguido por outro dia chuvoso, havendo igualmente sazonalidade (épocas do ano nas quais todos os anos se observam valores elevados de precipitação e, noutras épocas, se observam valores reduzidos). Por este motivo, de forma a evitarmos a forte correlação cronológica da série em análise, analisamos subconjuntos da amostra original como critério. Deste modo, criou-se uma subamostra de dimensão 29 anos, constituída pela observação máxima em cada um dos anos mencionados.

Para tal, os anos que tinham dados mensais em falta foram desconsiderados (o primeiro ano, que só tínhamos dados a partir de outubro, e o último ano, que termina em setembro) pois pretende-se que o máximo anual seja caracterizado pela mesma distribuição em todos os anos. Foram ainda criadas outras duas subamostras. Numa são considerados os 5 maiores níveis pluviométricos diários por ano e, na outra, as 10 maiores observações anuais. Estas amostras permitem a aplicação do método das maiores observações.

Por fim, foi ainda criada uma quarta subamostra, com os máximos mensais. Assim, cada mês foi considerado como um bloco. Repare-se que, desta forma, cada bloco poderá ter uma dimensão a variar entre 28 e 31 elementos. Além disso, nesta subamostra foram considerados unicamente os valores superiores a 54.5 mm (valor mínimo entre os máximos anuais calculado).

### Aplicação à precipitação de Casal Soeiro

A distribuição do conjunto de dados diários da precipitação em Casal Soeiro para todos os anos completos (1961-1989) está representada na Figura 2. Na figura do lado direito representamos unicamente os valores da precipitação nos três primeiros anos (1961-1989), evidenciando-se a sazonalidade ao longo do período estudado, devido às estações do ano com meses de maior quantidade de precipitação - outono-inverno (*boxplots* dos dados mensais, ilustrado na figura abaixo).



**Figura 2.** Evolução dos níveis de precipitação diários em Casal Soeiro entre 1961 e 1989 e entre 1961 e 1963; *boxplots* dos dados mensais entre 1961 e 1989.

Na Tabela 1 são apresentados os valores, em milímetros (*mm*), de algumas medidas descritivas do nível pluviométrico diário em Casal Soeiro, nomeadamente o mínimo (*min*), o primeiro quartil ( $P_{25}$ ), a mediana ( $P_{50}$ ), o terceiro quartil ( $P_{75}$ ), o máximo (*max*), a média ( $\bar{x}$ ) e o desvio padrão (*s*). Note-se que mais de metade das observações são nulas, pois, só em 3778 dos 10592 dias observados é que choveu (rondando 35.67% dias). Na Tabela 2 apresentamos a média e o desvio padrão dos níveis pluviométricos diários em cada mês durante o período em análise. Os dados claramente evidenciam as diferenças existentes entre os meses, com uma quantidade de precipitação bastante superior nos meses de janeiro, fevereiro e dezembro (que correspondem aos meses com maior dispersão em termos de pluviosidade), e mais reduzidos em julho e agosto (que correspondem aos meses com a menor dispersão em termos de pluviosidade). Deste modo, a distribuição subjacente ao nível pluviométrico diário não será a mesma em todos os meses do ano (terá, pelo menos, diferenças em termos de localização e escala).

**Tabela 1.** Medidas descritivas do nível pluviométrico diário em Casal Soeiro.

Medidas	Min.	P25	P50	P75	Max.	$\bar{x}$	<i>s</i>
(mm)	0	0	0	3.6	168.5	5.36	12.9217

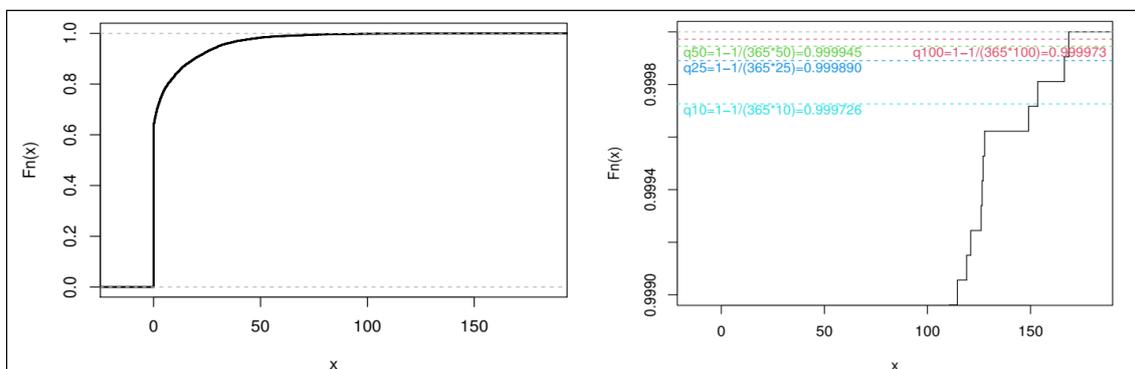
**Tabela 2.** Diferenças dos níveis pluviométricos mensais em Casal Soeiro.

Medidas	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
	9.3	9.6	6.3	4.9	4.1	2.5	1.0	0.9	3.5	6.2	7.1	9.3
Desvio padrão	16.6	16.6	14.6	10.4	8.8	7.2	4.1	4.2	11.5	14.8	14.4	17.6

Foi, igualmente, determinada a função de distribuição empírica dos níveis de precipitação diários em Casal Soeiro (gráfico da esquerda da Figura 3) e apresentamos um *zoom* para os valores mais elevados (superiores a 0.999) da função de distribuição empírica (gráfico da direita da Figura 3) no qual estão ilustradas as linhas associadas aos retornos de 10, 25, 50 e 100 anos (associadas às probabilidades  $p = 1 - 1/(365 \times 10)$ , ou seja, aproximadamente 0.999726, 0.999890, 0.999945 e 0.999973, respetivamente). Como é visível, não é possível extrapolar com base na função de distribuição empírica estes valores (por exemplo, os níveis de retorno a 50 e a 100 anos teriam o mesmo valor) e, como tal, teremos de modelar os valores recorrendo à distribuição GEV de forma a ser possível obter estimativas dos valores pretendidos.

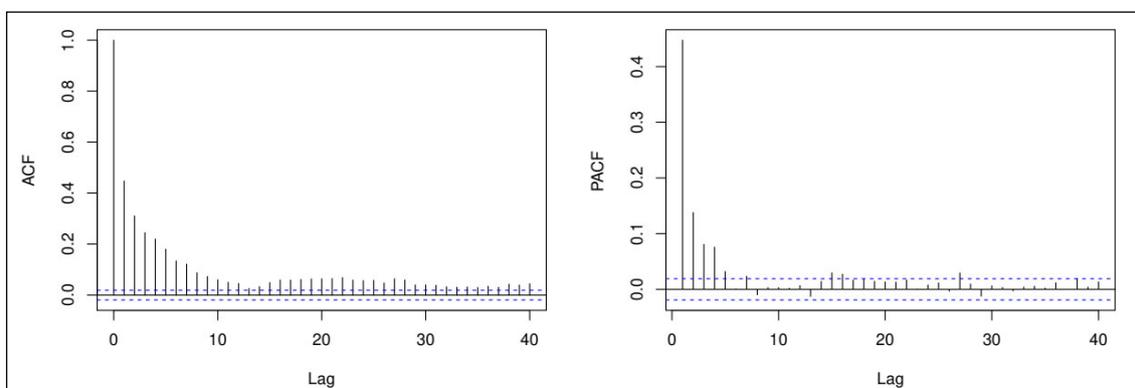
Por outro lado, as metodologias de estimação são baseadas em amostra aleatória (observações independentes) e, ao analisarmos a evolução dos níveis pluviométricos diários, espera-se que haja autocorrelação derivada da sequência

temporal das observações (associações fortes do resultado de um dia com o(s) do(s) dia(s) vizinhos) além da, previamente, referida sazonalidade.



**Figura 3.** Função de distribuição empírica dos níveis de precipitação diários em Casal Soeiro, com todos os valores (esquerda), com os dados relativos à cauda e respetivas níveis de retorno (direita).

Na Figura 4 podem ser observadas as funções de autocorrelação e de autocorrelação parcial que claramente evidenciam a referida correlação (dias próximos com chuva). Além disso, nos testes de existência de autocorrelação de *Durbin-Watson* e de *Ljung-Box* os valores-*p* são ambos inferiores a  $2.210^{16}$ , o que claramente evidência a existência de uma autocorrelação significativa.

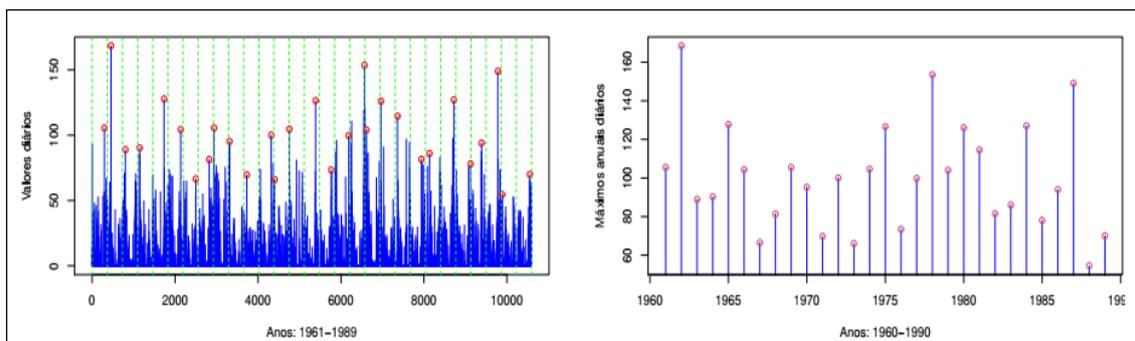


**Figura 4.** Funções de autocorrelação e de autocorrelação parcial dos níveis pluviométricos diários em Casal Soeiro.

Tendo em consideração o descrito, consideraram-se as três metodologias paramétricas que podem ser utilizadas numa análise de valores extremos: o método dos máximos anuais (MMA) ou método de *Gumbel*, o método das maiores observações (MMO) e o método dos valores superiores a um limiar POT. Assim, começar-se-á pela aplicação do MMA às amostras de níveis pluviométricos referenciadas anteriormente. De seguida, será aplicado o MMO às 5 e 10 maiores observações anuais. Por fim, será aplicada a metodologia POT com as amostras dos máximos mensais.

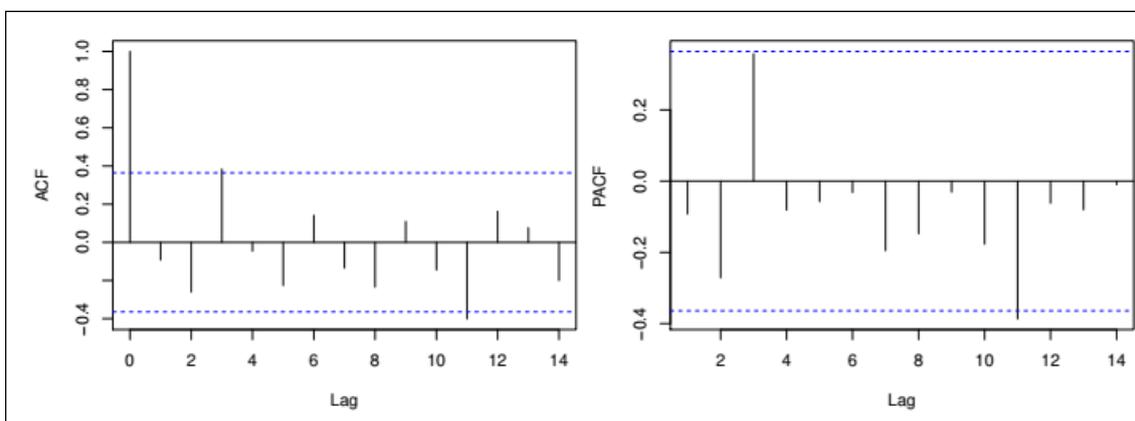
## O máximo anual

É possível observar, na Figura 5, os dados disponíveis para o período em análise (1961-1989) em Casal Soeiro, dividido em  $m = 29$  blocos (anos) e o máximo em cada bloco, formando a sequência dos máximos anuais (gráfico da direita). Assim, cada valor da nova sequência é obtido através do máximo de 365 (ou 366) observações. Pretende-se, deste modo, retirar a autocorrelação e a sazonalidade previamente referida.



**Figura 5.** Evolução do nível pluviométrico máximo anual em Casal Soeiro. Dados diários (azul), blocos anuais (verde) e máximos anuais (vermelho).

Para analisar a autocorrelação desta nova série, foram determinadas as funções de autocorrelação (figura da esquerda da Figura 6) e de autocorrelação parcial (figura da direita da Figura 6) do máximo anual, as quais não evidenciam a existência de autocorrelação. Foram ainda, aplicados os testes de existência de *Durbin-Watson* e de *Ljung-Box* obtendo-se, respetivamente, os valores 0.6049 e 0.6005 para  $p$ . Por conseguinte, parece que o processo de utilização do máximo anual eliminou a autocorrelação da série.

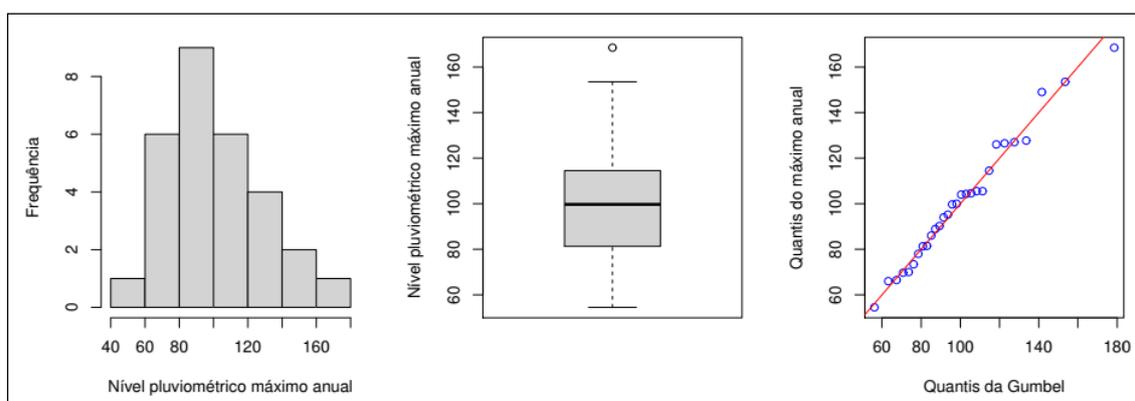


**Figura 6.** Funções de autocorrelação e de autocorrelação parcial do nível pluviométrico máximo anual em Casal Soeiro.

Apresentamos na Tabela 3 os valores, em milímetros (*mm*), de algumas medidas descritivas do máximo anual, nomeadamente o mínimo (min), o primeiro quartil ( $P_{25}$ ), a mediana ( $P_{50}$ ), o terceiro quartil ( $P_{75}$ ), o máximo (max), a média ( $\bar{x}$ ) e o desvio padrão (*s*). Na Figura 7 podemos observar o histograma e o *boxplot*. Na Figura 7 podemos observar que a variável aleatória, máximo anual, tem uma distribuição ligeiramente assimétrica (histograma e *boxplot*). No *QQ-plot* pode-se verificar que não parece existir uma diferença significativa entre a distribuição que caracteriza o máximo anual e a distribuição de *Gumbel*.

**Tabela 3.** Medidas descritivas do nível pluviométrico máximo anual em Casal Soeiro.

Medidas	Min.	$P_{25}$	$P_{50}$	$P_{75}$	Max.	$\bar{X}$	<i>s</i>
(mm)	54.5	81.3	99.7	114.5	168.5	100.4	27.82



**Figura 7.** Histograma, *boxplot* e *QQ plot* (*Gumbel*) do nível pluviométrico máximo anual em Casal Soeiro.

Além disso, no teste de *Kolmogorov-Smirnov* (validação da assimetria) o valor-*p* assume o valor 0.9817 quando os dados são comparados com uma distribuição de *Gumbel* com  $\hat{\mu} = 87.584$  e  $\hat{\sigma} = 22.447$  (valores estimados pelo método da máxima verosimilhança recorrendo à função *Gumbel* do *package evir*).

Os parâmetros foram ainda estimados recorrendo à função *fgev* ao *package evd*, primeiro restringindo a  $\gamma = 0$  (através de *forma = 0*) (isto é, restrito à distribuição de *Gumbel*) e depois sem restrições (utilizando a distribuição *GEV*). As estimativas obtidas para os parâmetros de interesse, nomeadamente forma ( $\gamma$ ), localização ( $\mu$ ) e escala ( $\sigma$ ), que apresentamos na Tabela 4, onde o erro padrão de cada estimativa é apresentado entre parêntesis.

Refira-se que as estimativas obtidas para os parâmetros de localização e escala pelos três métodos são quase idênticas. Por outro lado, a estimativa obtida para o parâmetro de forma é próxima de zero e parece não ser significativa (tendo

em consideração o valor da estimativa e do erro padrão associado). Assim, para comparar os dois modelos estimados pela função  $fg_{ev}$ , foi realizado um teste ANOVA ( $H_0$ : a qualidade do ajustamento é igual nos dois modelos) obtendo-se um valor- $p$  igual a 0.7090. Por conseguinte, a  $H_0$  não é rejeitada, evidenciando que não há diferenças significativas entre a estimação com base na distribuição de *Gumbel* ou a distribuição GEV para estes dados. De referir que o intervalo com 95% de confiança para  $\hat{\gamma}$  é  $(-0.3643, 0.2456)$  que, como esperado, inclui o valor 0 associado à distribuição de *Gumbel*.

Utilizando a distribuição de *Gumbel* para os dados de precipitação de Casal Soeiro (com as estimativas dos seus parâmetros) podemos estimar o período de retorno para os níveis pretendidos ou os níveis de retorno a  $T$ -anos. Por exemplo, considerando o valor máximo da amostra (168.5 mm) temos que o seu  $\hat{U}(50) = 168.0 \text{ mm}$  e  $\hat{U}(100) = 180.4 \text{ mm}$  utilizando a GEV.

**Tabela 4.** Estimativas dos parâmetros da distribuição GEV do nível pluviométrico máximo anual em Casal Soeiro.

	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$
Gumbel	-	87.5844	22.4472
	(-)	(4.3966)	(3.2560)
fFgve /shape=0)	-	87.5861	22.4389
	(-)	4.3950	3.2532
fgve	-0.0593	87.5861	22.8857
	(0.1556)	4.3950	3.5437

Apresentamos na Tabela 5 as estimativas obtidas pelo modelo *Gumbel* e pelo modelo GEV, para o período de retorno associado aos níveis de precipitação diários  $u \in \{130, 140, 150, 160, 170, 180, 190, 200\}$  (em  $mm$ ), bem como para o nível de retorno (em  $mm$ ) associado a  $T$ -anos com  $T \in \{5, 15, 25, 50, 75, 100, 150, 250\}$  e às probabilidades  $p \in \{0.1, 0.050, 0.025, 0.01, 0.005, 0.001, 0.0001, 0.00001\}$  de excedência.

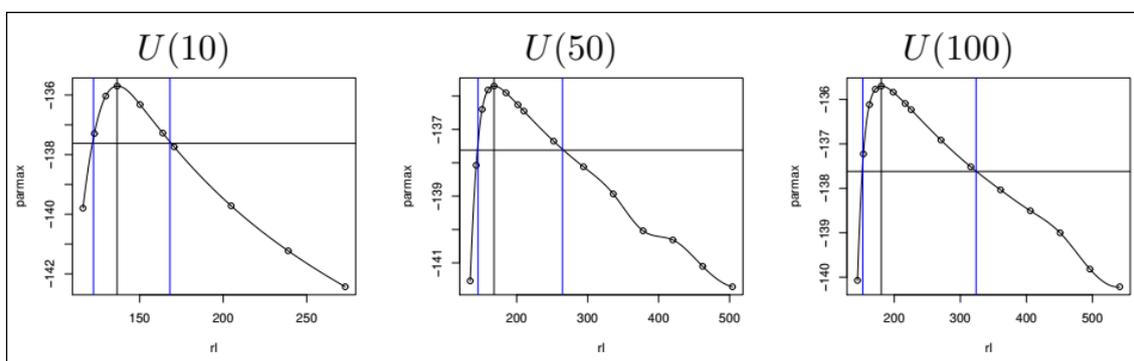
Nas estimativas apresentadas na Tabela 5 note-se que, apesar de serem próximas para os primeiros valores, tendem a distanciar-se à medida que aumentamos o nível e/ou probabilidade associada a essa estimativa (como seria de esperar, pois quando mais extremo for o evento, maior variabilidade existirá associada à sua estimação). Com a aplicação da distribuição de *Gumbel* são obtidos níveis com valores mais elevados para a mesma probabilidade de excedência e, por conseguinte, períodos de retorno mais reduzidos para o mesmo nível (elevado) de precipitação. Note-se, igualmente, que para probabilidades maiores ( $p$  igual a 0.10 e 0.05) as estimativas do nível de retorno encontram-se

abaixo do valor máximo observado (168.50 mm), pois têm associados períodos (10 e 20 anos) inferiores à dimensão da amostra em estudo. Para valores menores da probabilidade ( $p \in [0.01, 0.00001]$ ), o nível estimado ultrapassa o máximo observado na amostra, podendo mesmo chegar a valores muito superiores (como os 346 mm diários). Para as estimativas pontuais apresentadas na Tabela 5 podem ser igualmente obtidas estimativas intervalares.

**Tabela 5.** Estimativas do período de retorno para o nível  $u$  e o nível de retorno a  $T$ -anos em Casal Soeiro com base nos máximos anuais (Gumbel).

Nível $u$ (mm)		130	140	150	160	170	180	190	200
Período de retorno	Gumbel	7.1	10.8	16.6	25.7	39.8	61.9	96.3	150.1
estimado (anos)	GEV	7.4	11.8	19.4	32.5	55.8	97.6	174.5	318.8
Período $T$ (anos)		5	15	25	50	75	100	150	250
Nível de retorno	Gumbel	121.3	147.6	159.4	175.2	184.3	190.8	200.0	211.5
estimado (mm)	GEV	121.1	144.9	155.0	168.0	175.3	180.4	187.4	196.0
Probabilidade $p$		0.1	0.050	0.025	0.01	0.005	0.001	0.0001	0.00001
Nível de retorno	Gumbel	138.1	154.3	170.1	190.8	206.5	242.6	294.3	346.0
estimado (mm)	GEV	136.5	150.6	163.9	180.4	192.3	218.0	250.7	279.2

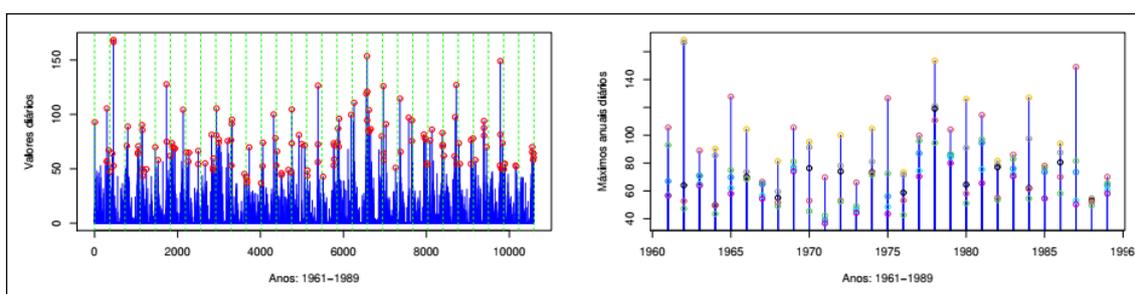
Apresentamos, na Figura 8, as estimativas de máxima verosimilhança e respetivos intervalos com 95% de confiança, utilizando a distribuição GEV. Assim, para 10 anos obtivemos  $\hat{U}(10) = 136.50 \text{ mm}$  e  $\hat{U}(10) \in (122.34, 168.10)$ ; para 50 anos  $\hat{U}(50) = 167.98 \text{ mm}$  e  $\hat{U}(50) \in (145.53, 264.38)$  e para 100 anos  $\hat{U}(100) = 180.38 \text{ mm}$  e  $\hat{U}(100) \in (152.28, 323.78)$ . Estes dados estão dentro do intervalo dos valores mais elevados de precipitação diária, relembrando que o mais alto atingido ronda os 168 mm.



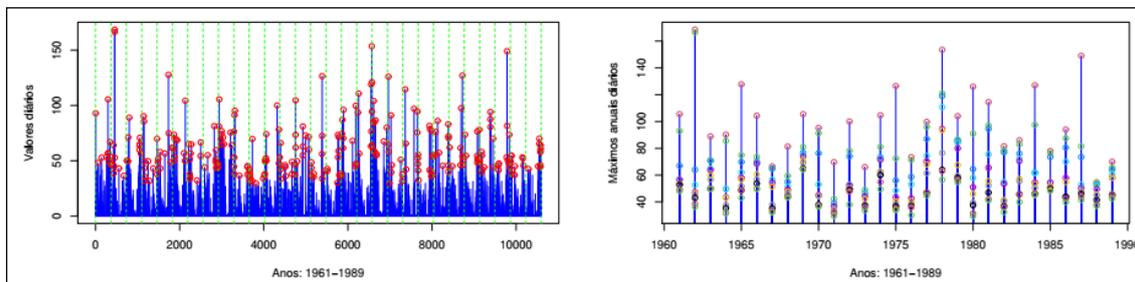
**Figura 8.** Estimativa pontual e intervalar do retorno a 10, 50 e 100 anos utilizando a GEV.

## Os $r$ máximos anuais

Em alternativa ao método ilustrado anteriormente (máximo anual), podemos usar o método do “top” anual, quando por exemplo a amostra é muito pequena. Este é um método paramétrico que se baseia numa amostra com as  $r$  maiores observações anuais, observadas para  $m$  anos. Assim, considerando blocos de um ano ( $m = 29$ ), seleccionamos as 5 e as 10 maiores observações anuais ( $r \in \{5, 10\}$ ), para ilustrar esta abordagem. Com esta alteração, ficamos com uma série de 145 observações e outra com 290 observações. Nas Figuras 9 e 10 é possível observar os dados disponíveis para o período em análise (1951-1989), dividido em  $m = 29$  blocos (anuais) com as 5 e 10 observações máximas, respetivamente.



**Figura 9.** Determinação e evolução dos 5 valores máximos anuais em Casal Soeiro.



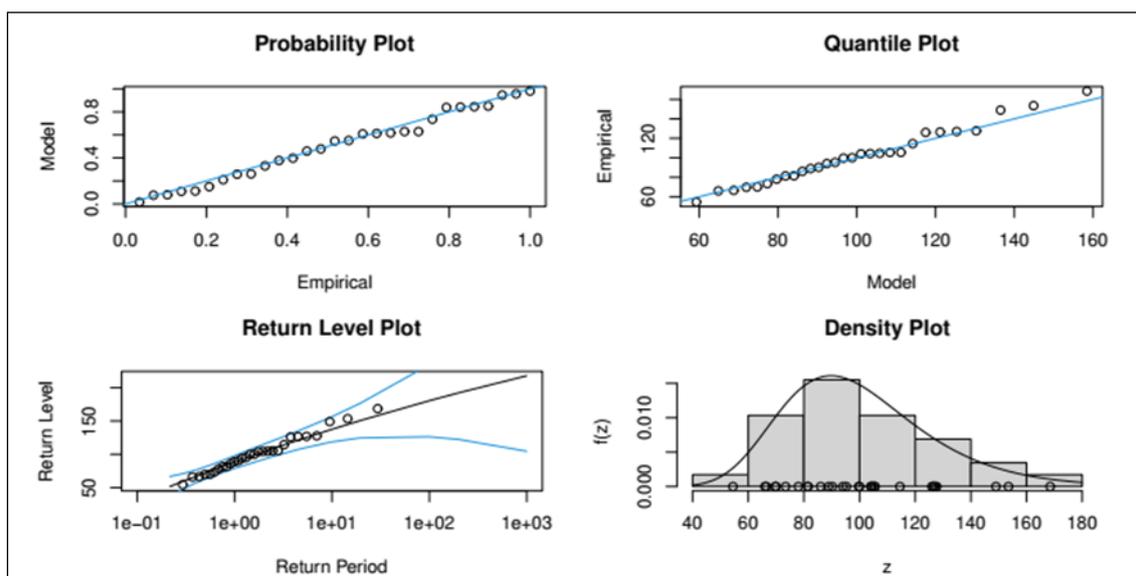
**Figura 10.** Determinação e evolução dos 10 valores máximos anuais em Casal Soeiro.

Na Tabela 6 encontra-se os valores estimados para os parâmetros da GEV utilizando as  $r$  maiores valores de cada ano recorrendo à função *rlarg.fit* do *package ismev*. O caso  $r = 1$  corresponde à situação analisada na secção anterior, pois temos unicamente como base o máximo anual. Apesar da função utilizada para a estimação ser distinta, as estimativas e respetivos erros padrão são quase iguais (compare-se os valores para  $r = 1$  na Tabela 6 com os valores da função *fgev* na Tabela 4). Com o aumento do  $r$  utilizamos mais informação e, como tal, os erros padrão associados diminuem. Todavia, saliente-se que nos três casos analisados a estimativa de  $\gamma$  é sempre próxima de zero, pouco significativa se tivermos em consideração o erro padrão associado, e nem sempre tem o mesmo sinal. Na Figura 11, podemos observar a qualidade do ajustamento para  $r = 1$  (máximo

anual), quer o *PP-plot* quer o *QQ-plot* -, ambos parecem revelar um ajustamento razoável, o nível de retorno e os intervalos de confiança associados (com o aumento significativo da amplitude dos intervalos com o aumento no período de retorno) e o histograma que parece ajustar-se à densidade da distribuição proposta.

**Tabela 6.** Estimativas dos parâmetros da distribuição GEV do nível pluviométrico máximo anual em Casal Soeiro utilizando os  $r$  maiores valores de cada ano.

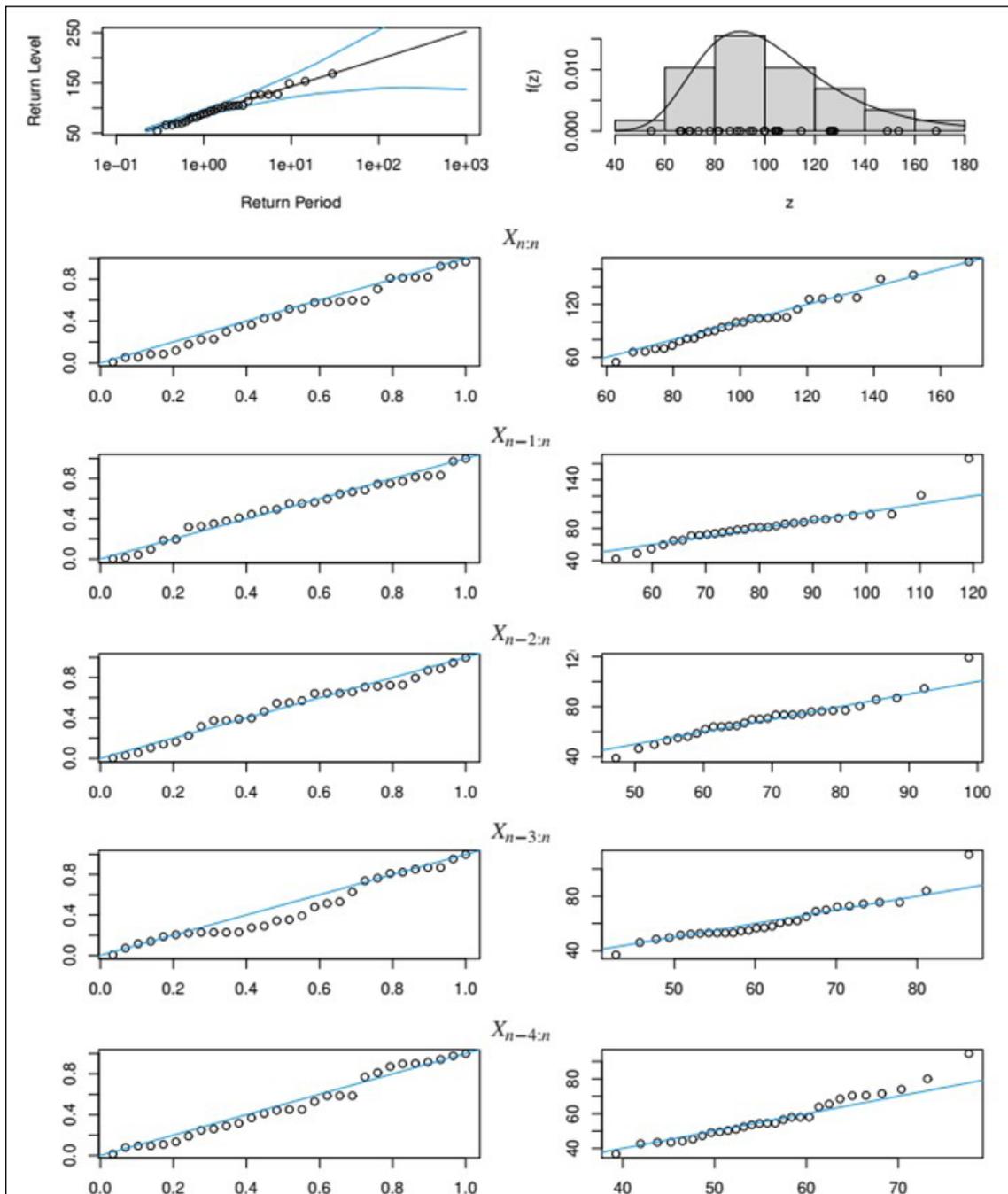
$r=1$	$\hat{\gamma}$	$\hat{\mu}$	$\hat{\sigma}$
	(0.1555)	(4.8679)	(3.5417)
$r=5$	0.0098	90.4071	22.6661
	(0.0806)	(3.5240)	(2.3487)
$r=10$	0.0107	90.5167	22.2564
	(0.0561)	(3.2904)	(2.2194)



**Figura 11.** Ajustamento para o máximo anual baseado no maior valor anual.

A qualidade do ajustamento para as 5 maiores observações anuais ( $r = 5$ ) é apresentado na Figura 12.

Tal como no caso  $r = 1$ , representamos o nível de retorno e os intervalos de confiança associados, com o aumento significativo da amplitude dos intervalos, face ao aumento no período de retorno. O histograma reforça a grande semelhança com a densidade da distribuição GEV. São, ainda, apresentados o *PP-plot* e o *QQ-plot* para a  $i$ -ésima maior observação, com  $i = 1, \dots, 5$  (ou seja,  $X_{n:n}, X_{n-1:n}, X_{n-2:n}, X_{n-3:n}, X_{n-4:n}$ ), estimadas com base nas 5 maiores observações, parecendo revelar um ajustamento razoável em todos os casos.



**Figura 12.** Ajustamento para o máximo anual baseado nos 5 valores anuais mais elevados.

Apresentamos, na Tabela 7, as estimativas, baseadas nos modelos ajustados com base nas  $r$  maiores observações anuais, com  $r \in \{1, 5, 10\}$ , para o período de retorno associado ao nível de precipitação diário  $u \in \{130, 140, 150, 160, 170, 180, 190, 200\}$  (em  $mm$ ), ao nível de retorno (em  $mm$ ) associado a  $T \in \{5, 15, 25, 50, 75, 100, 150, 250\}$  anos e ao nível de retorno (em  $mm$ ) associado à probabilidade  $p \in \{0.1, 0.050, 0.025, 0.01, 0.005, 0.001, 0.0001, 0.00001\}$  de excedência. Os valores observados para  $r = 1$  são, como seria esperado, praticamente iguais aos apresentados na Tabela 5, para a distribuição GEV. Por

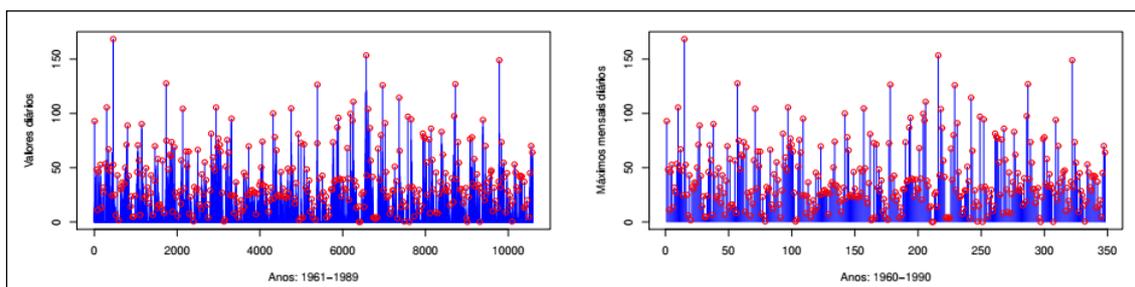
outro lado, os valores obtidos com  $r \in \{5, 10\}$  são mais próximos dos obtidos com a distribuição de *Gumbel*, aquando da utilização do máximo anual.

**Tabela 7 .** Estimativas do período de retorno para o nível  $u$  e o nível de retorno a  $T$ -anos em Casal Soeiro com base nos  $r \in \{1, 5, 10\}$  máximos anuais

Nível $u$ mm:		130	140	150	160	170	180	190	200
Período de retorno	$r=1$	7.4	11.8	19.4	32.6	55.8	97.8	175.1	320.4
estimado (anos)	$r=5$	6.2	9.2	13.9	21.1	32.1	48.8	74.3	113.1
	$r=10$	6.3	9.5	14.5	22.1	33.8	51.8	79.3	121.4
Período $T$ (anos):		5	15	25	50	75	100	150	250
Nível de retorno	$r=1$	121.2	144.9	155.0	168.0	175.3	180.4	187.4	195.9
estimado (mm)	$r=5$	124.7	151.8	164.1	180.6	190.2	197.1	206.7	219.0
	$r=10$	124.2	150.9	162.9	179.2	188.7	195.5	205.0	217.0
Probabilidade $p$ :		0.1	0.050	0.025	0.01	0.005	0.001	0.0001	0.00001
Nível de retorno	$r=1$	136.5	150.6	163.9	180.4	192.2	217.9	250.5	278.9
estimado (mm)	$r=5$	142.0	158.7	175.3	197.1	213.6	252.4	308.9	366.7
	$r=10$	141.2	157.7	174.0	195.5	211.8	250.0	305.9	363.1

### Máximos mensais superiores a um limiar

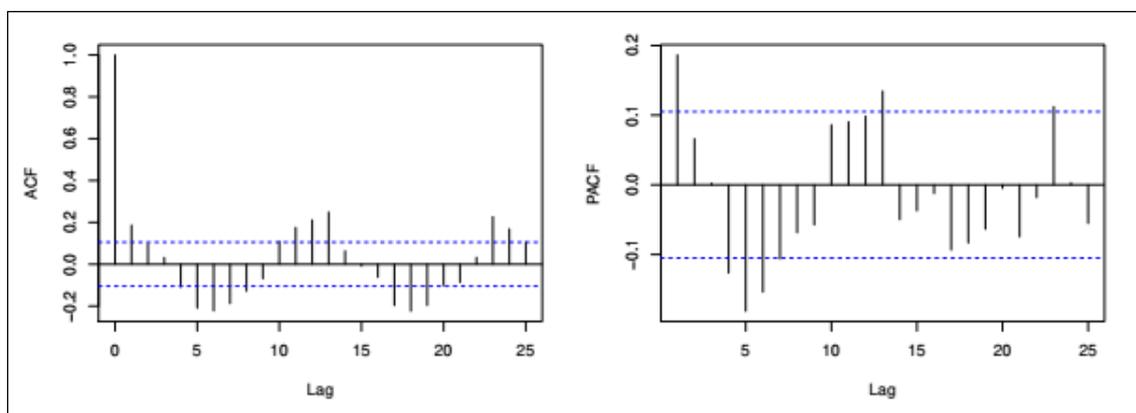
Em primeiro lugar, começamos por criar a série dos máximos mensais durante os 29 anos para análise (1961–1989). Na Figura 13 é apresentada a evolução do nível pluviométrico diário com a indicação dos valores máximos de cada mês (gráfico da esquerda) e na direita unicamente o valor máximo observado em cada um dos 348 meses em análise.



**Figura 13.** Determinação e evolução do nível máximo pluviométrico mensal em Casal Soeiro.

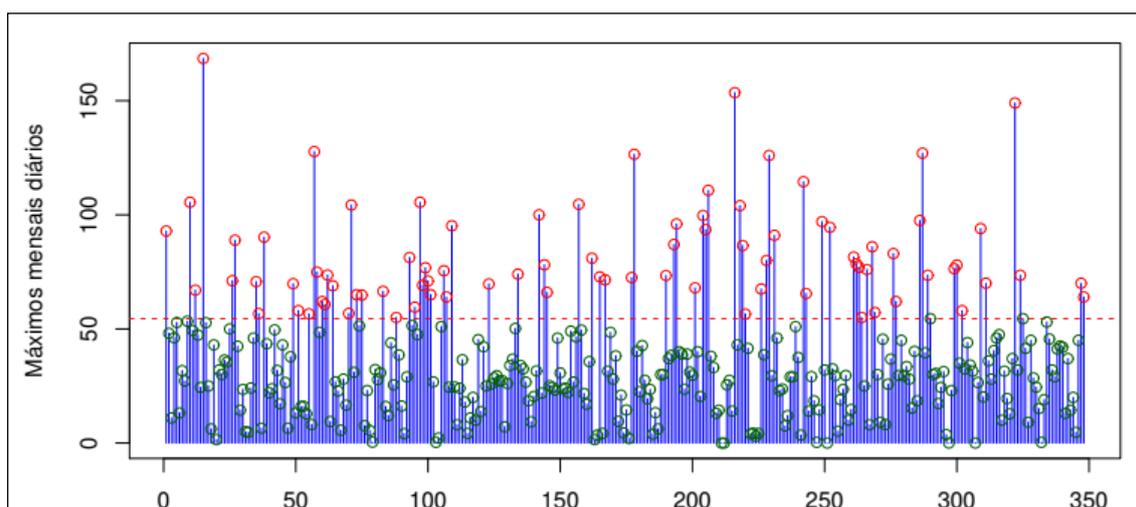
Na Figura 14 podem ser observadas as funções de autocorrelação e de autocorrelação parcial do máximo mensal, mostrando a autocorrelação que está presente. Naturalmente, como previamente referido, os níveis de precipitação são distintos em cada mês, havendo, portanto, sazonalidade. Por outro lado, nos

testes de *Durbin-Watson* o valor- $p$  é igual a 0.0001 e no teste de *Ljung-Box* o valor  $p$  é igual a 0.0005. Assim, ambos os testes indicam que a autocorrelação existe.



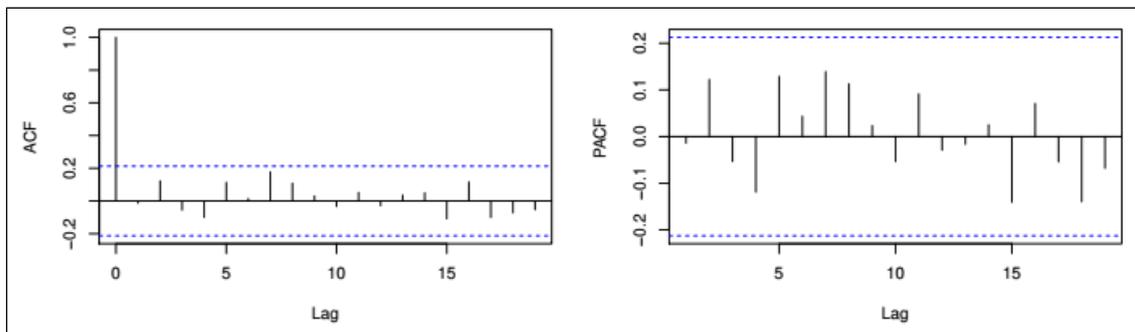
**Figura 14.** Funções de autocorrelação e de autocorrelação parcial do nível pluviométrico máximo mensal em Casal Soeiro.

Logo e de forma a tentar eliminar a autocorrelação, vamos considerar unicamente os valores superiores ao valor mínimo, entre os máximos anuais, que é igual a 54.5 mm, valor que se observou em 1988. Deste modo, vamos considerar unicamente 85 observações, entre os 348 máximos mensais (24.43% das observações). Esta abordagem baseia a inferência estatística nos extremos que excedem um determinado limiar (POT) e, como tal, na modelação com recurso à distribuição de GP. Na Figura 15 assinalamos a evolução do nível máximo de precipitação mensal, com valores superiores a 54.5 mm.



**Figura 15.** Determinação e evolução do nível máximo pluviométrico mensal em Casal Soeiro com valores superiores a 54.5 mm.

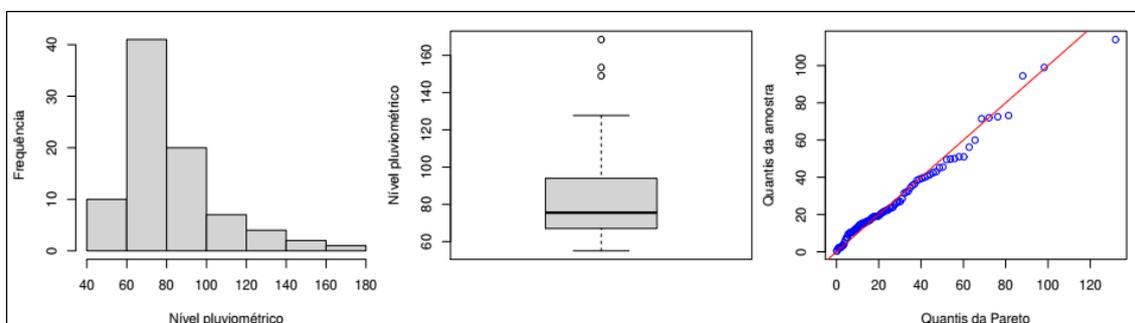
Observamos, na Figura 16, as funções de autocorrelação e de autocorrelação parcial dos máximos mensais superiores a 54.5 mm, que já não parecem evidenciar autocorrelação na série dos níveis máximos pluviométricos mensais, isto é, com valores superiores a  $u = 54.5$  mm. Nos testes de existência de autocorrelação de *Durbin-Watson* o valor- $p$  é igual a 0.5139 e no teste de *Ljung-Box* o valor  $p$  é igual a 0.894, o que significa inexistência de autocorrelação.



**Figura 16.** Funções de autocorrelação e de autocorrelação parcial do nível pluviométrico máximo mensal em Casal Soeiro com valores superiores a 54.5 mm.

Recorrendo à função *gpd.fit* do *package ismev* foram obtidas as estimativas (com os erros padrão das estimativas, escritas entre parênteses) de máxima verosimilhança para os parâmetros da distribuição de GP, assim como o erro padrão das estimativas,  $\hat{\sigma} = 33.3194$  (4.6178391) e  $\hat{\gamma} = -0.1954$  (0.0887917).

Na Figura 17 podemos observar, quer no histograma quer no *boxplot*, que o nível de precipitação máximo mensal com valores superiores a 54.5 mm tem uma distribuição assimétrica. No *QQ-plot* da GP (gráfico da direita) parece não existir uma diferença significativa entre a distribuição que caracteriza o nível de precipitação máximo mensal com valores superiores a 54.5 mm e a distribuição GP, utilizando os valores estimados para os parâmetros.



**Figura 17.** Histograma, *boxplot* e *QQ plot* (GP) do nível pluviométrico máximo mensal em Casal Soeiro com valores superiores a 54.5 mm.

Além disso, no teste de *Kolmogorov-Smirnov* o valor-  $p$  assume o valor 0.4004. Deste modo, parece que os dados dos excessos são bem modelados por uma distribuição GP, utilizando parâmetros iguais às referidas estimativas. Com base nas referidas estimativas para os parâmetros da GP, podemos estimar o período de retorno associado ao nível de precipitação diário  $u \in \{130, 140, 150, 160, 170, 180, 190, 200\}$  (mm), o nível de retorno associado a  $T \in \{5, 15, 25, 50, 75, 100, 150, 250\}$  anos e o nível de retorno associado à probabilidade  $p \in \{0.1, 0.050, 0.025, 0.01, 0.005, 0.001, 0.0001, 0.00001\}$  de excedência (Tabela 8).

**Tabela 8.** Estimativas do período de retorno para o nível  $u$  e o nível de retorno a  $T$ -anos em Casal Soeiro com base nos valores máximos mensais superiores a 54.5 mm.

Nível $u$ (mm)	130	140	150	160	170	180	190	200
Período de retorno estimado (anos)	6.8	12.0	22.8	47.4	111.5	311.2	1125.9	6295.5
Período $T$ (anos)	5	15	25	50	75	100	150	250
Nível de retorno estimado (mm)	124.1	143.6	151.3	160.7	165.6	168.8	173.1	178.0
Probabilidade $p$	0.1	0.050	0.025	0.01	0.005	0.001	0.0001	0.00001
Nível de retorno estimado (mm)	136.1	147.7	157.6	168.8	175.9	189.2	202.2	210.4

Os valores obtidos, quando comparados com os valores previamente analisados nas Tabelas 5 e 7, apresentam níveis de retorno significativamente mais baixos e, como tal, períodos de retorno muito superiores.

### Análise dos resultados

As três abordagens usadas são aproximações que requerem que a dimensão da amostra tenda para infinito (baseadas em convergências assintóticas). Na prática, a amostra influencia qualquer uma das três abordagens, pois, enquanto uma tem uma amostra fixa que depende da disponibilidade de dados ( $n$  ou número de anos em análise), a segunda depende da escolha do número ( $r$ ) dos dias máximos escolhidos por ano e a última do valor  $u$  escolhido como limiar.

À semelhança do defendido por outros autores, como Alves e Rosário (2015) e Beirlant et al. (2004) o nosso caso de estudo, da precipitação mais elevada em Casal Soeiro, ilustra bem como as diferentes abordagens podem contribuir para a escolha da subamostra mais adequada para a realização de inferência na TVE. Além disso, a precipitação em Casal Soeiro com valores superiores a 54.5 mm (POT) é propícia a desencadear eventos de cheia, daí a relevância desta investigação.

## Conclusões

Este trabalho foca-se apenas em abordagens paramétricas para estimar os parâmetros de um evento raro. Aplicamos metodologias de estimação de quantis elevados com três abordagens distintas, método de máximo anual ou de *Gumbel* (amostra pequena para alguns anos de estudo),  $r$  máximos anuais (depende da escolha do  $r$ , 1, 5 ou 10 dias) e método POT (depende da escolha do limiar  $\mu$ ). A partir daqui, foi-nos possível estimar o nível de retorno para várias probabilidades e períodos (anos), bem como o tempo de retorno. De uma forma geral, os máximos anuais e  $r$  máximos seguem uma distribuição de *Gumbel*, sendo que o parâmetro da forma se aproxima de 0, em todas as estimativas calculadas. Quando usamos o método de POT é possível verificar que a amostra segue (pelo menos de forma aproximada) a distribuição limite dos excessos ou distribuição de GP, tendo uma distribuição *Beta*, segundo o parâmetro da forma. O valor das estimativas dos parâmetros de interesse, não variam muito entre os métodos dos máximos anuais e dos  $r$  máximos. Isto pode indicar que, na presença de uma amostra mais reduzida, há uma certa confiança em usar o método do  $r$  máximos. Embora próximos, o método de POT é onde se observam resultados mais distintos, que dependem da escolha do limiar  $u$ , embora neste trabalho apenas utilizado um limiar como referência, o que não significa que não pudéssemos procurar outros valores POT. A redução dos riscos e desastres em espaços urbanos e rurais merece cada vez mais atenção em consonância com a densidade de habitantes numa determinada bacia hidrográfica.

## Nota

2 Ao longo deste trabalho vamos restringir a apresentação à análise do máximo, extremo de interesse na aplicação pretendida. Contudo, todos os resultados apresentados para o máximo de uma sequência de v.a. contínuas iid podem ser adaptados para o mínimo uma vez que  $Y_{1:n} = -X_{n:n}$  (e  $Y_{n:n} = -X_{1:n}$ ) se  $Y = -X$ . Se o objetivo fosse a análise do mínimo, em vez de ser analisada a cauda-direita de  $F$  seria investigado o peso da cauda esquerda de  $F$ .

## Referências

ALVES, I. F. & ROSÁRIO, P. Parametric and semi-parametric approaches to extreme rainfall modelling. C.P. Kitsos et al. (eds.): **Theory and Practice of Risk Assessment, Springer Proceedings in Mathematics & Statistics**, 136, Springer, pp. 279–291. 2015.

- BEIRLANT, J., GOEGEBEUR, Y., SEGERS, J., TEUGELS, J. **Statistics of Extremes: Theory and Applications**, Wiley, England. 2004.
- CASTILLO, E., HADI, A. S., BALAKRISHNAN, N., SARABIA, J. M. **Extreme Value and Related Models with Applications in Engineering and Science**, John Wiley & Sons, Hoboken, New Jersey. 2005.
- COLES, S. G. **An Introduction to Statistical Modeling of Extreme Values**. Springer-Verlag. London. 2001.
- Instituto Geográfico do Exército [IGE]. **Carta Militar de Portugal**. Folhas: 3, 4, 8, 9, 16, 16, 28 e 29. [www.igeoe.pt](http://www.igeoe.pt). Escala 1:25 000. 1996, 1997.
- GILLELAND, E. **Package 'ismev'**. Disponível em: <<https://cran.r-project.org/web/packages/ismev/ismev.pdf>>. 2018.
- GOMES, M. I., ALVES, I. F. E NEVES, C. **Análise de Valores Extremos: uma introdução**, sociedade Portuguesa de Estatística, Instituto Nacional de estatística. 2013.
- GOMES M. I. **Desafios em Estatística de Extremos**. Associado à comunicação “Sessão da Classe de Ciências” na Academia das Ciências de Lisboa em 16 de maio de 2019 (Preprint). 2019.
- NEVES, P. **A Teoria de Valores Extremos na Quantificação de Precipitação Elevada**, dissertação de mestrado, Universidade de Aveiro. 2010.
- PFUFF, B., ZIVOT, E., MCNEIL, A. AND STEPHENSON, A. **Package 'evir'**. Disponível em <<https://cran.r-project.org/web/packages/evir/evir.pdf>>. Acesso em 10 de agosto de 2022.. 2018.
- REIS, D. **Statistical Modelling of Extreme Rainfall in Madeira Island**, Tese de doutoramento, Universidade da Madeira. 2014.
- REISS, R. D., THOMAS, M. **Statistical Analysis of Extreme Values, with Application to Insurance, Finance, Hydrology and Other Fields**. 3rd edition, Birkhäuser Verlag. 2007.
- RIZZO, M. L. M. **Statistical computing with R**. 2nd edition. Chapman and Hall/CRC. 2019.
- ROSÁRIO, P. **Análise de valores extremos para níveis pluviométricos em Barcelos**. Tese de doutoramento. Universidade de Lisboa. 2013.
- Sistema Nacional de Informação de Recursos Hídricos (1960/1990). **Bases de dados**. Disponível em <<http://snirh.pt>>. Acesso em 10 de maio de 2021.
- WUERTZ, D., SETZ, T. AND CHALABI, Y. **Package 'fExtremes'**. Disponível em <https://cran.r-project.org/web/packages/fExtremes/fExtremes.pdf>>. Acesso em 3 de agosto de 2022. 2017.
- STEPHENSON, A. **Package 'evd'**, Disponível em <<https://cran.r-project.org/web/packages/evd/evd.pdf>>. Acesso em 10 de julho de 2022. 2018.