

DM

Um Estudo Sobre Valores Extremos por Blocos

DISSERTAÇÃO DE MESTRADO

Afonso Guilherme Ferro Silva

MESTRADO EM MATEMÁTICA, ESTATÍSTICA E APLICAÇÕES



UNIVERSIDADE da MADEIRA

A Nossa Universidade

www.uma.pt

setembro | 2025

Um Estudo Sobre Valores Extremos por Blocos

DISSERTAÇÃO DE MESTRADO

Afonso Guilherme Ferro Silva

MESTRADO EM MATEMÁTICA, ESTATÍSTICA E APLICAÇÕES

ORIENTAÇÃO

Délia Canha Gouveia Reis

Júri:

Doutora Ana Maria Cortesão Pais Figueira da Silva Abreu

Professora Associada da Universidade da Madeira

Doutora Sandra Maria Freitas Mendonça

Professora Associada da Universidade da Madeira

Doutora Délia Canha Gouveia Reis

Professora Auxiliar da Universidade da Madeira

Agradecimentos

À Professora Délia, expresso a minha mais sincera gratidão pela ajuda e paciência, pela orientação científica e pelo contributo decisivo para que este trabalho pudesse alcançar uma melhor qualidade.

À minha mãe, cuja presença e influência moldaram de forma decisiva a minha vida, expresso a minha profunda gratidão.

A ambas, apresento os meus mais respeitosos agradecimentos.

Resumo

A análise de valores extremos desempenha um papel essencial na compreensão e previsão de fenómenos climáticos e ambientais com grande impacto económico e social. A presente dissertação tem como objetivo analisar séries temporais de valores de temperatura e de precipitação, recorrendo à Teoria dos Valores Extremos. A análise foi conduzida sobre dados da Região Autónoma da Madeira, em particular, para a cidade do Funchal.

Este estudo contempla uma revisão da literatura e a exposição da metodologia seguida na análise das séries consideradas. A metodologia adotada tem por base a modelação de valores extremos por blocos, com particular destaque para a escolha estatística entre a Distribuição Generalizada de Valores Extremos não Gumbel e a distribuição de Gumbel, sendo a seleção entre estes dois modelos apoiada em critérios de informação e procedimentos de validação gráfica, como também no teste de razão de verosimilhanças. Este teste estatístico e os critérios de informação foram também aplicados na seleção entre estes dois modelos e correspondentes modelos com tendência linear.

Os resultados revelam comportamentos contrastantes entre a temperatura e a precipitação: a hipótese do parâmetro de forma nulo é rejeitada exclusivamente na análise da precipitação. Quanto aos máximos anuais das médias mensais dos máximos diários de temperatura, a série longa denota não estacionaridade, contrastando com a predominância de modelos estacionários nos subperíodos analisados. Na análise da série longa dos mínimos anuais das médias mensais dos totais diários de precipitação, como nos correspondentes subperíodos, observa-se a escolha da Distribuição Generalizada de Valores Extremos não Gumbel em oposição aos modelos com tendência linear analisados.

Palavras-Chave: Valores Extremos, Distribuição Generalizada de Valores Extremos, Seleção de Modelos, Análise de Tendência

Abstract

Extreme value analysis plays an essential role in understanding and predicting climate and environmental phenomena with significant economic and social impact. This dissertation aims to analyze time series of temperature and precipitation values using Extreme Value Theory. The analysis was conducted on data from the Autonomous Region of Madeira, specifically for the city of Funchal.

This study includes a literature review and an exposition of the methodology followed in the analysis of the considered series. The adopted methodology is based on block extreme value modeling, with particular emphasis on the statistical choice between the non-Gumbel Generalized Extreme Value Distribution and the Gumbel distribution. The selection between these two models was supported by information criteria and graphical validation procedures, as well as the likelihood ratio test. This statistical test and the information criteria were also applied to select between these stationary models and their corresponding models incorporating linear trend.

The results reveal contrasting behaviors between temperature and precipitation: the null shape parameter hypothesis is rejected exclusively in the precipitation analysis. Regarding the annual maximums of the monthly averages of daily maximum temperatures, the long-term series denotes non-stationarity, contrasting with the predominance of stationary models in the analyzed sub-periods. In the analysis of the long-term series of annual minimums of monthly averages of daily precipitation totals, as well as in the corresponding sub-periods, the selection of the non-Gumbel Generalized Extreme Value Distribution is generally observed, in opposition to the analyzed linear trend models.

Keywords: Extreme Values, Generalized Extreme Value Distribution, Model Selection, Trend Analysis

Índice

Lista de Figuras	xvi
Lista de Tabelas	xviii
1 Introdução	1
2 Valores Extremos por Blocos	3
2.1 Distribuições Limite do Máximo e do Mínimo	4
2.2 Estimação	10
2.3 Aplicabilidade, Validação e Seleção de Modelos	13
2.4 Exemplo de Aplicação	17
3 Máximos Anuais das MMMD de Temperatura	25
3.1 Estatística Descritiva	26
3.2 Estacionaridade e Independência	28
3.3 Seleção e Validação de Modelos	29
3.3.1 Modelo M_0 versus Modelo M_1	29
3.3.2 Dados (1865-2018)	30
3.3.3 Análise de Períodos Temporais I	33
3.3.4 Análise de Períodos Temporais II	40
3.4 Considerações finais	47
4 Mínimos Anuais das MMTD de Precipitação	51
4.1 Estatística Descritiva	52
4.2 Estacionaridade e Independência	54
4.3 Seleção e Validação de Modelos	55
4.3.1 Modelo M_0 versus Modelo M_1	55
4.3.2 Dados (1865 – 2018)	57
4.3.3 Análise de Períodos Temporais I	58
4.3.4 Análise de Períodos Temporais II	61
4.4 Considerações finais	64

5 Conclusão	65
Anexo	68
A Máximos Anuais do Nível do Mar - Port Pirie	71
B Máximos Anuais do Nível do Mar - Fremantle	75
C Gráficos de funções densidade	77
D Máximos Anuais das MMMD de Temperatura	81
E Mínimos Anuais das MMTD de Precipitação	99
Bibliografia	112

Lista de Figuras

2.1	Densidades das distribuições de Fréchet, Gumbel e Max–Weibull.	8
2.2	Densidades das distribuições de extremos mínimos.	9
2.3	Gráficos de diagnóstico do Modelo M_1 - Máximos anuais do nível do mar (<i>Port Pirie</i>).	18
2.4	Gráficos de diagnóstico do Modelo M_0 - Máximos anuais do nível do mar (<i>Port Pirie</i>).	19
2.5	Gráficos de diagnóstico do Modelo M_1 - Máximos anuais do nível do mar (<i>Fremantle</i>).	21
2.6	Gráficos de diagnóstico do Modelo M_0 - Máximos anuais do nível do mar (<i>Fremantle</i>).	22
3.1	Gráficos de diagnóstico para o Modelo M_0 - Dados (1865–2018).	31
3.2	Gráficos de diagnóstico para o Modelo M_1 - Dados (1865–2018).	32
3.3	Gráficos de diagnóstico para o Modelo M_0 - Dados 1 (1865–1897).	34
3.4	Gráficos de diagnóstico para o Modelo M_1 - Dados 1 (1865–1897).	34
3.5	Gráficos de diagnóstico para o Modelo M_0 - Dados 2 (1897–1974).	36
3.6	Gráficos de diagnóstico para o Modelo M_1 - Dados 2 (1897–1974).	37
3.7	Gráficos de diagnóstico para o Modelo M_0 - Dados 3 (1975–2018).	39
3.8	Gráficos de diagnóstico para o Modelo M_1 - Dados 3 (1975–2018).	39
3.9	Gráficos de diagnóstico para o Modelo M_0 - Dados 4 (1865–1950).	41
3.10	Gráficos de diagnóstico para o Modelo M_1 - Dados 4 (1865–1950).	42
3.11	Gráficos de diagnóstico para o Modelo M_0 - Dados 5 (1951–1996).	43
3.12	Gráficos de diagnóstico para o Modelo M_1 - Dados 5 (1951–1996).	44
3.13	Gráficos de diagnóstico para o Modelo M_0 - Dados 6 (1997–2018).	46
3.14	Gráficos de diagnóstico para o Modelo M_1 - Dados 6 (1997–2018).	47
4.1	Gráficos de diagnóstico para o Modelo M_1 - Dados (1865–2018).	57
4.2	Gráficos de diagnóstico para o Modelo M_1 - Dados 1 (1865–1887).	59
4.3	Gráficos de diagnóstico para o Modelo M_1 - Dados 2 (1888–2018).	59
4.4	Gráficos de diagnóstico para o Modelo M_1 - Dados 3 (1865–1950).	61
4.5	Gráficos de diagnóstico para o Modelo M_1 - Dados 4 (1951–1996).	62

4.6 Gráficos de diagnóstico para o Modelo M_1 - Dados 5 (1997-2018). 63

Lista de Tabelas

2.1	Relação entre os parâmetros da distribuição GEV e os parâmetros das distribuições das variáveis aleatórias de Fréchet, de Gumbel e de Weibull.	9
2.2	Especificações dos parâmetros de localização e de escala para os modelos M_1 , $M_{1.A}$, $M_{1.B}$, $M_{1.C}$, M_0 , $M_{0.A}$, $M_{0.B}$ e $M_{0.C}$	12
2.3	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC para escolha entre Modelos M_0 e M_1 - Máximos anuais do nível do mar (<i>Port Pirie</i>).	19
2.4	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC - Máximos anuais do nível do mar (<i>Port Pirie</i>).	20
2.5	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC para escolha entre os Modelos M_0 e M_1 - Máximos anuais do nível do mar (<i>Fremantle</i>).	23
2.6	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC - Máximos anuais do nível do mar (<i>Fremantle</i>).	23
3.1	Dimensão, média, desvio padrão e coeficiente de variação - Máximos Anuais das MMMD de Temperatura.	27
3.2	Mediana (M), primeiro e terceiro quartis, extremos e amplitude amostral (R_n) - Máximos Anuais das MMMD de Temperatura.	27
3.3	Valores- p para os testes de Mann-Kendall e de Wald-Wolfowitz.	29
3.4	Valores de d_{obs} e valor- p para o teste de razão de verossimilhanças.	30
3.5	Valores dos critérios de informação AIC e BIC.	30
3.6	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados (1865–2018).	33
3.7	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 1 (1865–1897).	35
3.8	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 2 (1897–1974).	37

3.9	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 3 (1975–2018).	40
3.10	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 4 (1865–1950).	42
3.11	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 5 (1951–1996).	45
3.12	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 6 (1997–2018).	47
3.13	Escolha de modelos M_0 , $M_{0.A}$, $M_{0.B}$ e $M_{0.C}$.	48
3.14	Escolha de modelos M_1 , $M_{1.A}$, $M_{1.B}$ e $M_{1.C}$.	48
4.1	Dimensão, média, desvio padrão e coeficiente de variação - Mínimos Anuais das MMTD de Precipitação.	53
4.2	Mediana (M), primeiro e terceiro quartis - Mínimos Anuais das MMTD de Precipitação.	54
4.3	Mínimo, máximo e amplitude amostral (R_n) - Mínimos Anuais das MMTD de Precipitação.	54
4.4	Resultados dos testes de tendência e independência para os subconjuntos de dados.	55
4.5	Valores de d_{obs} e valor- p para o teste de razão de verossimilhanças.	56
4.6	Valores dos critérios de informação AIC e BIC.	56
4.7	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados (1865–2018).	58
4.8	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 1 (1865–1887).	60
4.9	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 2 (1888–2018).	60
4.10	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC - Dados 3 (1865–1950).	62
4.11	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC - Dados 4 (1951–1996).	63
4.12	Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 5 (1997-2018).	63
4.13	Escolha de modelos M_1 , $M_{1.A}$, $M_{1.B}$ e $M_{1.C}$.	64

Capítulo 1

Introdução

O estudo dos fenómenos extremos assume um papel central em várias áreas científicas e aplicadas, nomeadamente na climatologia, hidrologia, finanças, engenharia e ciências ambientais. A ocorrência de valores extremos pode ter impactos significativos a nível económico, social e ambiental, sendo por isso fundamental compreender a sua natureza estatística e desenvolver modelos robustos que permitam não apenas descrever o comportamento observado, mas também prever eventos futuros de elevada intensidade.

A Teoria dos Valores Extremos (TVE) constitui o enquadramento teórico que suporta este tipo de análise. Em particular, a distribuição generalizada dos valores extremos, denotada por GEV (do inglês, *Generalized Extreme Value*), surge como modelo de referência na análise por blocos. Contudo, a sua aplicação prática requer uma avaliação cuidada de pressupostos como a independência, a estacionaridade e a escolha adequada de blocos temporais, aspetos que têm implicações diretas na qualidade das estimativas e na validade das conclusões.

Os dados utilizados neste estudo foram obtidos no website oficial <https://www.ipma.pt/pt/oclima/series.longas/?loc=Funchal&type=raw> no âmbito da política de dados abertos promovida pelo Instituto Português do Mar e da Atmosfera (IPMA), que disponibiliza séries longas de dados recolhidos em estações meteorológicas distribuídas pelo território nacional. Estes conjuntos de dados incluem observações diárias e mensais de diversos parâmetros meteorológicos, nomeadamente temperatura do ar (mínima, média e máxima), precipitação acumulada e pressão atmosférica reduzida ao nível médio do mar. Em particular, recorreremos aos registos da estação meteorológica Funchal Observatório (estação número 522), situada na cidade do Funchal, ilha da Madeira, com coordenadas geográficas aproximadas

de $32^{\circ}38'$ N de latitude e $16^{\circ}53'$ W de longitude, a uma altitude de 58 metros.

A estação meteorológica Funchal Observatório apresenta uma das séries meteorológicas mais longas disponíveis para território nacional, com início em 1 de dezembro de 1864, tendo a sua localização e as condições de observação da estação sofrido alterações ao longo do tempo. Entre 1864 e 1950, os dados foram recolhidos na fortaleza de São Lourenço, a uma altitude de 10 metros entre 1864 e 1920 e a 13 metros entre 1921 e 1950. Neste período, as observações meteorológicas eram realizadas às 9h, 15h e 21h. A partir de 1950, a estação foi transferida para novas instalações nos Louros, na orla leste da cidade do Funchal, passando a situar-se a 58 metros de altitude, com observações efetuadas às 8h, 14h e 20h. Esta mudança não só implicou uma nova localização física, como também alterou a exposição dos instrumentos, que passaram a estar instalados ao nível do solo, em vez de numa torre. A partir de 1997, a estação passou a funcionar em modo automático (*AWS Automatic Weather Station*), com registos digitais contínuos. No entanto, a longevidade, consistência e qualidade da série de dados desta estação tornam-na particularmente valiosa para o estudo de valores extremos.

Este trabalho está estruturado da seguinte forma: no Capítulo 2, é apresentado conceitos e metodologias da Teoria dos Valores Extremos. Este capítulo também inclui um exemplo de aplicação, relativo aos máximos anuais de níveis do mar, já utilizado na literatura científica da área. Nos Capítulos 3 e 4 são analisados os máximos anuais das médias mensais dos máximos diários (MMMD) de temperatura e os mínimos anuais das médias mensais dos totais diários (MMTD) de precipitação da estação meteorológica Funchal Observatório. Finalmente, no Capítulo 5, são discutidas as principais conclusões, implicações práticas e possíveis direções para trabalhos futuros.

Capítulo 2

Valores Extremos por Blocos

Este capítulo apresenta alguns fundamentos teóricos e metodológicos da modelação de valores extremos por blocos, incluindo resultados probabilísticos básicos e a revisão de uma aplicação prática apresentada na literatura.

Na Secção 2.1 são introduzidos os conceitos fundamentais associados às estatísticas de ordem, bem como os principais resultados teóricos que conduzem às distribuições limite do máximo e do mínimo, culminando na definição da distribuição Generalizada de Valores Extremos.

A Secção 2.2 é dedicada à estimação dos parâmetros dos modelos GEV através do método da máxima verosimilhança, considerando modelos estacionários e extensões que incorporam tendência temporal.

Na Secção 2.3 são apresentados os procedimentos de verificação das premissas do modelo, bem como os métodos de seleção e validação utilizados na literatura, recorrendo a testes estatísticos e ferramentas gráficas de diagnóstico.

Por fim, na Secção 2.4 é realizada a revisão e discussão de uma aplicação prática apresentada por Coles [5], com o objetivo de ilustrar a implementação da metodologia de valores extremos em dados reais através do pacote `ismev` do R.

2.1 Distribuições Limite do Máximo e do Mínimo

Seja $x = (x_1, \dots, x_n)$ uma realização do vetor aleatório (X_1, \dots, X_n) . Denotamos a ordenação ascendente de x por $x_{1:n}, \dots, x_{n:n}$, onde $x_{1:n} \leq \dots \leq x_{n:n}$. O vetor que para a realização (x_1, \dots, x_n) de (X_1, \dots, X_n) tem como realização $x_{1:n}, \dots, x_{n:n}$ é denominado de vetor das estatísticas ordinais associado ao vetor aleatório (X_1, \dots, X_n) .

Sejam

$$X_{1:n} = \min_{1 \leq i \leq n} X_i \quad \text{e} \quad X_{n:n} = \max_{1 \leq i \leq n} X_i. \quad (2.1)$$

Cada componente univariada $X_{i:n}$ do vetor das estatísticas ordinais $(X_{1:n}, \dots, X_{n:n})$ é denominado i -ésima estatística ordinal ascendente do vetor aleatório (X_1, \dots, X_n) [11]. Consideramos ainda que as marginais do vetor (X_1, \dots, X_n) são variáveis aleatórias independentes e identicamente distribuídas a uma variável aleatória X ($X_i \stackrel{d}{=} X$) com função de distribuição F definida por

$$F(x) := P[X \leq x], \quad x \in \mathbb{R}. \quad (2.2)$$

A função de distribuição empírica é dada por

$$\hat{F}_n(x) = \frac{i}{n}, \quad \text{se } x \in [x_{i:n}, x_{i+1:n}), \quad i = 1, \dots, n-1, \quad (2.3)$$

onde $x_{i:n}$ é o i -ésimo valor amostral [2].

Denotemos $X_{n:n}$ por M_n e $X_{1:n}$ por m_n . A função de distribuição de M_n é dada por:

$$\begin{aligned} H_n(x) = P[M_n \leq x] &= P[X_1 \leq x, \dots, X_n \leq x] \\ &= P[X_1 \leq x] \times \dots \times P[X_n \leq x] \\ &= F^n(x), \end{aligned}$$

e a função de distribuição para m_n é dada por

$$\begin{aligned} L_n(x) = P[m_n \leq x] &= 1 - P[X_1 > x, \dots, X_n > x] \\ &= 1 - P[X_1 > x] \times \dots \times P[X_n > x] \\ &= 1 - [(1 - P[X_1 \leq x]) \times \dots \times (1 - P[X_n \leq x])] \\ &= 1 - (1 - F(x))^n. \end{aligned}$$

Notemos que as funções de distribuição obtidas não têm grande utilidade na prática dado F ser desconhecida. Uma forma de resolução deste problema

seria estimar a função de distribuição F a partir de dados observados, na qual a função de distribuição empírica, definida em (2.3), teria um papel importante. No entanto, não é esta a forma usualmente utilizada porque pequenas discrepâncias na função estimativa para F podem originar grandes discrepâncias para F^n . Outra forma de resolução é partir do facto que F é desconhecida e focalizar o estudo no comportamento da função de distribuição F^n quando $n \rightarrow +\infty$. Ora, quando $n \rightarrow +\infty$ temos

$$\lim_{n \rightarrow \infty} H_n(x) = \lim_{n \rightarrow \infty} [F(x)]^n = \begin{cases} 1, & \text{se } F(x) = 1, \\ 0, & \text{se } F(x) < 1, \end{cases}$$

e

$$\lim_{n \rightarrow \infty} L_n(x) = \lim_{n \rightarrow \infty} 1 - [1 - F(x)]^n = \begin{cases} 0, & \text{se } F(x) = 0, \\ 1, & \text{se } F(x) > 0, \end{cases}$$

o que significa que as distribuições limite de H_n e de L_n são degeneradas, em um e em zero, respetivamente. No entanto, a dificuldade de termos como distribuições limite distribuições degeneradas é evitada se for possível encontrar transformações lineares tais que as distribuições limite

$$\lim_{n \rightarrow \infty} H_n(a_n + b_n x) = \lim_{n \rightarrow \infty} [F(a_n + b_n x)]^n = H(x), \quad \forall x, \quad (2.4)$$

e

$$\lim_{n \rightarrow \infty} L_n(c_n + d_n x) = \lim_{n \rightarrow \infty} 1 - [1 - F(c_n + d_n x)]^n = L(x), \quad \forall x, \quad (2.5)$$

não sejam degeneradas, onde a_n , b_n , c_n e d_n são constantes que dependem de n [3]. Dizemos que uma dada função, F , pertence ao *max*-domínio de atracção de H se for possível encontrar sucessões de constantes reais $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ ($b_n > 0$) que satisfaçam (2.4) e, analogamente, que F pertence ao *min*-domínio de atracção de L se for possível encontrar sucessões de constantes reais $\{c_n\}_{n \geq 1}$ e $\{d_n\}_{n \geq 1}$ que satisfaçam (2.5). A questão que surge naturalmente é quais serão as distribuições limite associadas aos domínios de atracção referidos. A resposta a esta questão é dada por dois teoremas habitualmente designados, respetivamente, por teorema dos tipos extremais para máximos e por teorema dos tipos extremais para mínimos (cf. [3, 5, 6]).

Teorema 2.1 (Teorema dos tipos extremais para máximos) *A única família de distribuições não degenerada que satisfaz (2.4) é:*

$$H_\xi(x; \mu, \sigma) = \exp \left\{ - \left[1 - \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{\frac{1}{\xi}} \right\}, \quad 1 - \xi \left(\frac{x - \mu}{\sigma} \right) \geq 0, \quad \xi \neq 0, \quad (2.6)$$

onde $x \leq \mu + \sigma/\xi$, se $\xi > 0$, ou $x \geq \mu + \sigma/\xi$, se $\xi < 0$.

Tomando o limite de (2.6), quando $\xi \rightarrow 0$, obtemos a família de distribuições para o caso $\xi = 0$

$$H_0(x; \mu, \sigma) = \exp \left[-\exp \left(\frac{\mu - x}{\sigma} \right) \right], \quad -\infty < x < \infty. \quad (2.7)$$

As distribuições em (2.6) e (2.7) constituem a distribuição generalizada de valores extremos máximos, aqui denotada por $GEV_M(\mu, \sigma, \xi)$ [3].

Os quantis de ordem p correspondentes, x_p tal que $H(x_p, \mu, \sigma) = p$, podem ser obtidos com a inversão de H_ξ e de H_0 em (2.6) e (2.7), resultando

$$x_p = \begin{cases} \mu + \sigma [1 - (-\log p)^\xi] / \xi, & \text{se } \xi \neq 0, \\ \mu - \sigma \log(-\log p), & \text{se } \xi = 0. \end{cases} \quad (2.8)$$

Teorema 2.2 (Teorema dos tipos extremais para mínimos) *A única família não-degenerada de distribuições que satisfaz (2.5) é*

$$L_\xi(x; \mu, \sigma) = 1 - \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{1/\xi} \right\}, \quad 1 + \xi \left(\frac{x - \mu}{\sigma} \right) \geq 0, \quad \xi \neq 0, \quad (2.9)$$

onde, $x \geq \mu - \sigma/\xi$, se $\xi > 0$, ou $x \leq \mu - \sigma/\xi$, se $\xi < 0$.

Tomando o limite de (2.5), quando $\xi \rightarrow 0$, obtemos a família de distribuições para o caso $\xi = 0$

$$L_0(x; \mu, \sigma) = 1 - \exp \left[-\exp \left(\frac{x - \mu}{\sigma} \right) \right], \quad \infty < x < \infty. \quad (2.10)$$

As distribuições em (2.9) e (2.10) constituem a distribuição generalizada de valores extremos mínimos, aqui denotada por $GEV_m(\mu, \sigma, \xi)$ [3].

Os quantis de ordem p correspondentes, x_p tal que $H(x_p, \mu, \sigma) = p$, podem ser obtidos com a inversão de L_ξ e de L_0 em (2.9) e em (2.10) obtemos

$$x_p = \begin{cases} \mu - \sigma [1 - (-\log(1 - p))^\xi] / \xi, & \text{se } \xi \neq 0, \\ \mu + \sigma \log(-\log(1 - p)), & \text{se } \xi = 0. \end{cases} \quad (2.11)$$

A transformação de uma variável aleatória que segue uma distribuição de extremos mínimos para a correspondente distribuição de extremos máximos pode ser obtida de forma direta. O Teorema 2.3 estabelece a equivalência

entre a GEV_m e a GEV_M através de uma simples transformação de sinal na variável aleatória. Este teorema constitui a base para diversas propriedades associadas às distribuições de valores extremos e tem implicações diretas na modelação estatística de fenómenos extremos [3].

Teorema 2.3 (Obtenção da GEV_m a partir da GEV_M) Se $X \sim GEV_m(\mu, \sigma, \xi)$, então $Y = -X \sim GEV_M(-\mu, \sigma, \xi)$.

Distribuições de Fréchet, de Gumbel e de Weibull

Consideremos as distribuições de Fréchet, Gumbel e Max-Weibull, definidas por:

- Distribuição da variável aleatória de Fréchet:

$$H(x) = \begin{cases} 0, & x < \mu, \\ \exp\left[-\left(\frac{\sigma}{x-\mu}\right)^\xi\right], & x \geq \mu, \quad \xi > 0. \end{cases}$$

- Distribuição da variável aleatória de Gumbel:

$$H(x) = \exp\left[-\exp\left(\frac{\mu-x}{\sigma}\right)\right], \quad x \in \mathbb{R}.$$

- Distribuição da variável aleatória de Max-Weibull:

$$H(x) = \begin{cases} \exp\left[-\left(\frac{\mu-x}{\sigma}\right)^\xi\right], & x \leq \mu, \quad \xi > 0, \\ 1, & x > \mu. \end{cases}$$

Estas distribuições constituem casos particulares da $GEV_M(\mu, \sigma, \xi)$, introduzida no Teorema 2.1. A distribuição de Gumbel corresponde ao caso $\xi = 0$, a de Fréchet a $\xi > 0$ e a Max-Weibull a $\xi < 0$. Na Figura 2.1, obtida com o código disponibilizado no Anexo C, estão representadas funções de densidade de probabilidade de exemplos de variáveis aleatórias de Fréchet, de Gumbel e de Max-Weibull.

Consideremos agora as distribuições de extremos mínimos, Min-Fréchet, Min-Gumbel e Weibull:

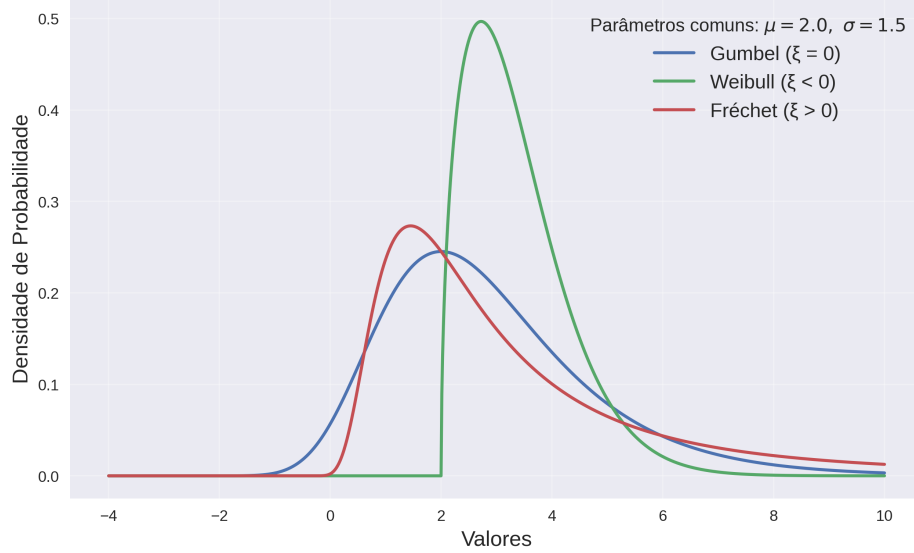


Figura 2.1: Densidades das distribuições de Fréchet, Gumbel e Max–Weibull.

- Distribuição da variável aleatória de Min–Fréchet:

$$L(x) = \begin{cases} 1 - \exp\left[-\left(\frac{\sigma}{\mu - x}\right)^\xi\right], & x \leq \mu, \\ 1, & x > \mu. \end{cases}$$

- Distribuição da variável aleatória de Min–Gumbel:

$$L(x) = 1 - \exp\left[-\exp\left(\frac{x - \mu}{\sigma}\right)\right], \quad x \in \mathbb{R}.$$

- Distribuição da variável aleatória de Weibull:

$$L(x) = \begin{cases} 0, & x < \mu, \\ 1 - \exp\left[-\left(\frac{x - \mu}{\sigma}\right)^\xi\right], & x \geq \mu. \end{cases}$$

Estas distribuições constituem casos particulares da $GEV_m(\mu, \sigma, \xi)$, introduzida no Teorema 2.2. A representação gráfica de funções de densidade de probabilidade de exemplos de variáveis aleatórias de Min–Fréchet, de Min–Gumbel e de Weibull encontra-se na Figura 2.2, obtida com o código disponibilizado no Anexo C.

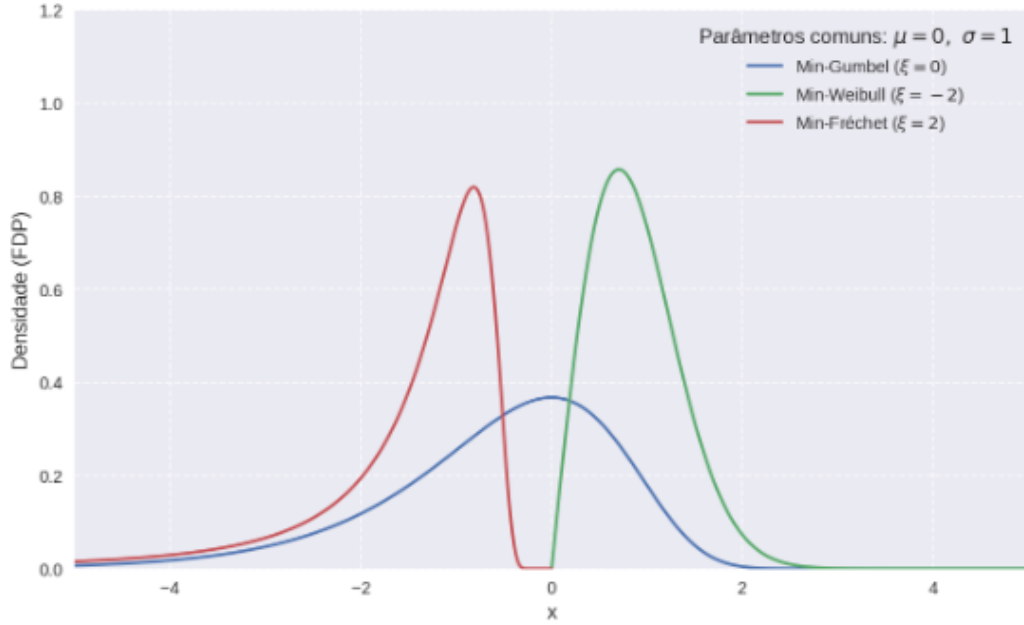


Figura 2.2: Densidades das distribuições de extremos mínimos.

A Tabela 2.1 sintetiza a relação entre os parâmetros da distribuição GEV e os das distribuições clássicas (de Fréchet, de Gumbel e de Weibull), destacando como os parâmetros da GEV (denotados por asteriscos, como μ^* , σ^* e ξ^*) se convertem nos parâmetros tradicionais (μ , σ , ξ) e vice-versa.

Tabela 2.1: Relação entre os parâmetros da distribuição GEV e os parâmetros das distribuições das variáveis aleatórias de Fréchet, de Gumbel e de Weibull.

Distribuição	Máximo		Mínimo	
Fréchet	$\mu = \mu^* + \sigma^*/\xi^*$ $\sigma = -\sigma^*/\xi^*$ $\xi = -1/\xi^*$	$\mu^* = \mu + \sigma$ $\sigma^* = \sigma/\xi$ $\xi^* = -1/\xi$	$\mu = \mu^* - \sigma^*/\xi^*$ $\sigma = -\sigma^*/\xi^*$ $\xi = -1/\xi^*$	$\mu^* = \mu - \sigma$ $\sigma^* = \sigma/\xi$ $\xi^* = -1/\xi$
Gumbel	$\mu = \mu^*$ $\sigma = \sigma^*$	$\mu^* = \mu$ $\sigma^* = \sigma$	$\mu = \mu^*$ $\sigma = \sigma^*$	$\mu^* = \mu$ $\sigma^* = \sigma$
Weibull	$\mu = \mu^* + \sigma/\xi^*$ $\sigma = \sigma^*/\xi^*$ $\xi = 1/\xi^*$	$\mu^* = \mu - \sigma$ $\sigma^* = \sigma/\xi$ $\xi^* = 1/\xi$	$\mu = \mu^* - \sigma^*/\xi^*$ $\sigma = \sigma^*/\xi^*$ $\xi = 1/\xi^*$	$\mu^* = \mu + \sigma$ $\sigma^* = \sigma/\xi$ $\xi^* = 1/\xi$

2.2 Estimação

O método da máxima verosimilhança é, de entre os diferentes métodos de estimação dos parâmetros de uma distribuição generalizada de valores extremos existentes na literatura da área, um dos mais utilizados pela sua versatilidade e adaptabilidade na construção de modelos complexos [5].

No método dos máximos por blocos, a série temporal é dividida em blocos disjuntos de igual dimensão, sendo extraído em cada bloco apenas o valor máximo observado. Um bloco pode corresponder, por exemplo, a um ano, uma estação ou outro intervalo temporal fixo, dependendo da natureza dos dados em análise.

Consideremos agora somente $GEV_M(\mu, \sigma, \xi)$ introduzida no Teorema 2.1, que também pode ser escrita da seguinte forma

$$H(x; \mu, \sigma, \xi) = \begin{cases} e^{-\left[1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}}, & \text{se } \xi \neq 0, \\ e^{-e^{-\left(\frac{x - \mu}{\sigma}\right)}}, & \text{se } \xi = 0, \end{cases} \quad (2.12)$$

válida para todos os valores de x tais que $1 + \xi \left(\frac{x - \mu}{\sigma}\right) > 0$, com $\mu \in \mathbb{R}$, $\sigma > 0$ e $\xi \in \mathbb{R}$ [5].

A log-verosimilhança associada a uma amostra de máximos independentes x_1, \dots, x_n é dada por

$$\ell(\mu, \sigma, \xi) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma}\right)\right]^{-1/\xi},$$

válida quando

$$1 + \xi \left(\frac{x_i - \mu}{\sigma}\right) > 0, \quad i = 1, \dots, n.$$

No caso $\xi = 0$, a log-verosimilhança reduz-se a

$$\ell(\mu, \sigma, 0) = -n \log \sigma - \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^n \exp\{-(x_i - \mu/\sigma)\}.$$

Os estimadores de máxima verosimilhança $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$ são obtidos por maximização numérica da log-verosimilhança. Este procedimento encontra-se implementado no pacote `ismev` do R, através da função `gev.fit()`, que permite ajustar modelos estacionários e não estacionários. No caso não estacionário, a função `gev.fit()` admite a inclusão de covariáveis através dos argumentos `ydat`, `mul`, `sigl` e `shl`, permitindo modelar os parâmetros de localização, escala e forma como funções explícitas do tempo ou de outras variáveis explicativas.

O parâmetro de forma ξ determina o comportamento assintótico da cauda da distribuição. Admitindo a aproximação assintótica Normal dos estimadores de máxima verosimilhança, tem-se

$$\hat{\xi} \sim N\left(\xi, \text{Var}(\hat{\xi})\right),$$

pelo que um intervalo de confiança aproximado a $100(1 - \alpha)\%$ para ξ é dado por

$$\hat{\xi} \pm z_{1-\alpha/2} SE(\hat{\xi}),$$

onde $SE(\hat{\xi})$ representa o erro padrão do estimador $\hat{\xi}$ e $z_{1-\alpha/2}$ é o quantil $1 - \alpha/2$ da distribuição Normal Standard.

Na formulação não estacionária, admite-se que os parâmetros dependem explicitamente do tempo, escrevendo

$$\mu_t = \mu(t), \quad \sigma_t = \sigma(t), \quad e \quad \xi_t = \xi(t).$$

De acordo com Coles [5], neste caso, a log-verosimilhança total assume a forma

$$\ell = -\sum_{t=1}^n \log \sigma_t - \sum_{t=1}^n \left(1 + \frac{1}{\xi_t}\right) \log \left[1 + \xi_t \left(\frac{x_t - \mu_t}{\sigma_t}\right)\right] - \sum_{t=1}^n \left[1 + \xi_t \left(\frac{x_t - \mu_t}{\sigma_t}\right)\right]^{-1/\xi_t},$$

sujeita à condição

$$1 + \xi_t \left(\frac{x_t - \mu_t}{\sigma_t}\right) > 0, \quad t = 1, \dots, n.$$

Quando $\xi_t = 0$, obtém-se

$$\ell = -\sum_{t=1}^n \log \sigma_t - \sum_{t=1}^n \frac{x_t - \mu_t}{\sigma_t} - \sum_{t=1}^n \exp\left\{-\frac{x_t - \mu_t}{\sigma_t}\right\}.$$

Em aplicações práticas, é habitual admitir não estacionariedade apenas no parâmetro de localização e, eventualmente, no parâmetro de escala, mantendo o parâmetro de forma constante. Conforme referido por Coles [5], a estimação de ξ é geralmente mais instável, sobretudo em amostras de dimensão reduzida, podendo a introdução de dependência temporal neste parâmetro aumentar significativamente a variabilidade das estimativas e comprometer a robustez do ajustamento. Assim, salvo evidência empírica clara em contrário, considera-se frequentemente ξ constante ao longo do tempo, reservando a modelação temporal para μ_t e σ_t .

No restante texto desta dissertação, a distribuição generalizada de valores extremos denotada por $GEV_M(\mu, \sigma, \xi)$ será referida como Modelo M_1 quando $\xi \neq 0$ e como Modelo M_0 quando $\xi = 0$. Denotemos por Modelos $M_{1.A}$ e $M_{0.A}$, respectivamente, os modelos com a inclusão de tendência linear apenas no parâmetro de localização, $GEV(\mu(t), \sigma, \xi)$ e $GEV(\mu(t), \sigma, 0)$, com $\mu(t) = \beta_0 + \beta_1 t$, onde β_0 e $\beta_1 \in \mathbb{R}$. A inclusão de tendência linear no parâmetro da dispersão por ser realizada considerando $\sigma(t) = \exp(\beta_2 + \beta_3 t)$, com β_2 e $\beta_3 \in \mathbb{R}$, conduzindo às distribuições $GEV(\mu, \sigma(t), \xi)$ e $GEV(\mu, \sigma(t), 0)$, denotadas por Modelos $M_{1.B}$ e $M_{0.B}$, respectivamente. Os Modelos $M_{1.C}$ e $M_{0.C}$ corresponderão a $GEV(\mu(t), \sigma(t), \xi)$ e $GEV(\mu(t), \sigma(t), 0)$, respectivamente. Na Tabela 2.2 estão sintetizadas as especificações dos parâmetros de localização e de escala para os modelos M_1 , $M_{1.A}$, $M_{1.B}$, $M_{1.C}$, M_0 , $M_{0.A}$, $M_{0.B}$ e $M_{0.C}$.

Tabela 2.2: Especificações dos parâmetros de localização e de escala para os modelos M_1 , $M_{1.A}$, $M_{1.B}$, $M_{1.C}$, M_0 , $M_{0.A}$, $M_{0.B}$ e $M_{0.C}$.

Modelo	Localização (μ)	Escala (σ)
M_1 / M_0	μ	σ
$M_{1.A} / M_{0.A}$	$\mu(t) = \beta_0 + \beta_1 t$	σ
$M_{1.B} / M_{0.B}$	μ	$\sigma(t) = \exp(\beta_2 + \beta_3 t)$
$M_{1.C} / M_{0.C}$	$\mu(t) = \beta_0 + \beta_1 t$	$\sigma(t) = \exp(\beta_2 + \beta_3 t)$

2.3 Aplicabilidade, Validação e Seleção de Modelos

Aplicabilidade de Modelos

Nesta secção, são apresentados os testes não-paramétricos utilizados neste estudo para avaliar a independência e a estacionaridade, implementados nos pacotes `trend` [12] e `RecordTest` [4] do R.

O teste de Mann-Kendall é um teste estatístico não paramétrico usado para identificar a presença de uma tendência monotónica (consistente e numa única direção) em séries temporais, como dados de caudal de rios, temperatura ou precipitação (e.g. [9, 13, 15, 17, 18]). O teste foi originalmente introduzido por Mann [10] e posteriormente desenvolvido por Kendall [7].

Para uma dada série temporal $\{X_i, i = 1, 2, \dots, n\}$, a hipótese nula H_0 indica ausência de tendência monotónica e a hipótese alternativa H_1 estabelece que existe uma tendência monotónica. O valor observado para a estatística S do teste de Mann-Kendall é dado por:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i), \quad (2.13)$$

onde $\text{sgn}(\cdot)$ representa a função sinal, definida por

$$\text{sgn}(x) = \begin{cases} +1, & \text{se } x > 0, \\ 0, & \text{se } x = 0, \\ -1, & \text{se } x < 0. \end{cases}$$

O teste de sequências de Wald-Wolfowitz [16] é um teste não paramétrico utilizado para avaliar se uma sequência X_1, X_2, \dots, X_n apresenta aleatoriedade serial, o que, em aplicações de séries temporais, é habitualmente interpretado como evidência a favor da independência.

Sejam x_1, x_2, \dots, x_n os dados observados. O valor observado para a estatística R do teste de Wald-Wolfowitz é dado por:

$$\sum_{i=1}^{n-1} x_i x_{i+1} + x_1 x_n.$$

Para amostras suficientemente grandes, a estatística de teste dada por $\frac{R - \mu_R}{\sigma_R}$ tem distribuição normal padrão. A implementação destes dois testes

é realizada por meio das funções `mk.test` e `ww.test` do pacote `trend` do R, respetivamente.

Validação dos Modelos M_0 e M_1

Aqui são apresentados os gráficos de diagnóstico utilizados para avaliar o ajustamento dos Modelos M_0 e M_1 [5]. Os gráficos de probabilidades, de quantis, e de níveis de retorno tal como o histograma com a densidade ajustada para os Modelos M_0 e M_1 são obtidos pela aplicação das funções `gum.diag` e `gev.diag` do pacote `isnev` do R, respetivamente.

Com os máximos por blocos ordenados, $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(m)}$, a função de distribuição empírica avaliada em $z_{(i)}$ é dada por:

$$\tilde{G}(z_{(i)}) = i/(m+1) \text{ com } i = 1, \dots, m.$$

O gráfico de probabilidades é formado pelos pontos

$$\left\{ \left(\tilde{G}(z_{(i)}), \hat{G}(z_{(i)}) \right), i = 1, \dots, m \right\}.$$

com

$$\hat{G}(z_{(i)}) = \exp \left\{ - \left[1 + \hat{\xi} \left(\frac{z_{(i)} - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-1/\hat{\xi}} \right\},$$

onde $\hat{\mu}$, $\hat{\sigma}$ e $\hat{\xi}$ são, respetivamente, as estimativas dos parâmetros μ , σ e ξ obtidas pelo método da máxima versomilhança. Se o modelo considerado for adequado então temos:

$$\hat{G}(z_{(i)}) \approx \tilde{G}(z_{(i)}),$$

para cada i , e neste caso, os pontos devem situar-se numa diagonal de 45 graus.

O gráfico de quantis compara os quantis empíricos com os quantis teóricos da distribuição ajustada, consistindo nos seguintes pontos [5]:

$$\left\{ \left(\hat{G}^{-1}(i/(m+1)), z_{(i)} \right), i = 1, \dots, m \right\}$$

com

$$\hat{G}^{-1}\left(\frac{i}{m+1}\right) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left[1 - \left\{-\log\left(\frac{i}{m+1}\right)\right\}^{-\hat{\xi}}\right].$$

Desvios da linearidade no gráfico de quantis também aqui indiciam um menor ajuste do modelo.

O gráfico de níveis de retorno é particularmente relevante na teoria dos valores extremos, pois permite interpretar diretamente a magnitude de eventos raros. O nível de retorno associado a um período de retorno T corresponde ao quantil excedido, em média, uma vez em cada T períodos, sendo definido por

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left(1 - y_p^{-\hat{\xi}}\right), & \text{se } \hat{\xi} \neq 0, \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{se } \hat{\xi} = 0, \end{cases} \quad (2.14)$$

com $y_p = -\log(1 - p)$.

O gráfico de níveis de retorno é a representação dos seguintes pares

$$\{(\log y_p, \hat{z}_p) : 0 < p < 1\}.$$

A utilização de escala logarítmica no eixo horizontal permite comprimir a cauda da distribuição, facilitando a visualização de níveis de retorno associados a períodos longos. No caso particular $\xi = 0$, a relação entre z_p e $\log y_p$ é linear, servindo esta linearidade como referência para avaliar o efeito do parâmetro de forma ξ . Desvios em relação a esse comportamento podem indicar inadequação do modelo ajustado [5].

Os gráficos de probabilidades, quantis e níveis de retorno baseiam-se na comparação entre o modelo teórico e as estimativas empíricas da função distribuição. Adicionalmente, o histograma constitui um diagnóstico complementar simples, permitindo comparar visualmente a distribuição empírica dos dados com a densidade ajustada pelo modelo GEV, embora esta abordagem seja geralmente menos informativa devido à dependência da escolha das classes [5].

Seleção de Modelos

Para a seleção de um modelo de entre os modelos apresentados na Tabela 2.2 da Secção 2.2, aplicaremos neste estudo, o teste da razão de verossimilhança (*likelihood ratio test*, LRT), particularmente adequado quando os modelos

em análise são encaixados, isto é, quando um modelo mais simples pode ser obtido impondo restrições paramétricas a um modelo mais geral [5]. Os modelos estacionários constituem casos particulares dos modelos não estacionários, obtidos pela imposição de nulidade aos coeficientes associados às tendências temporais nos parâmetros de localização e/ou de escala.

Para dois modelos encaixados, designando por modelo reduzido aquele sujeito a restrições paramétricas e por modelo completo o modelo mais geral, a estatística de teste é definida por

$$D = 2 \left[\ell(\hat{\theta}_f) - \ell(\hat{\theta}_r) \right],$$

onde $\ell(\hat{\theta}_f)$ e $\ell(\hat{\theta}_r)$ representam os valores máximos da log-verosimilhança sob o modelo completo e sob o modelo reduzido, respetivamente. Sob a hipótese nula de que o modelo reduzido é suficiente, a estatística D converge assintoticamente para uma distribuição $\chi^2_{(k_f - k_r)}$, em que $k_f - k_r$ corresponde à diferença no número de parâmetros entre os dois modelos [5]. Assim, rejeita-se a hipótese nula quando

$$D > \chi^2_{1-\alpha, k_f - k_r}.$$

Em particular, testar

$$H_0 : \xi = 0 \quad \text{versus} \quad H_1 : \xi \neq 0$$

resulta numa seleção entre os Modelos M_0 (Gumbel) e M_1 (GEV não Gumbel) definidos na Secção 2.2. Assim, o Modelo M_0 será rejeitado se o valor observado para a estatística do teste de razão de verosimilhanças correspondente for maior do que o quantil $1 - \alpha$ da distribuição qui-quadrado com um grau de liberdade.

Complementarmente, a comparação entre modelos pode também ser realizada com base em critérios de informação [8]. O critério de Akaike [1] (*Akaike Information Criterion*, AIC) é definido por

$$AIC = -2\ell(\hat{\theta}) + 2k,$$

onde $\ell(\hat{\theta})$ representa o valor máximo da log-verosimilhança e k o número de parâmetros estimados. Este critério procura um compromisso entre qualidade de ajustamento e complexidade paramétrica, favorecendo o modelo com menor valor de AIC.

Por sua vez, o critério bayesiano de Schwarz [14] (*Bayesian Information Criterion* (BIC)) é dado por

$$BIC = -2\ell(\hat{\theta}) + k \log(n),$$

onde n representa a dimensão amostral.

A escolha do critério de seleção de modelos deve considerar o tamanho da amostra e o tipo de modelo, estacionário ou não estacionário. Com base no estudo de *Kim et al.* [8], que avaliou, através de simulações de Monte Carlo, o desempenho do AIC, BIC e do teste da razão de verossimilhanças na seleção de modelos GEV estacionários e não estacionários, verificou-se que o BIC tende a apresentar melhor desempenho em modelos estacionários, especialmente quando a amostra é relativamente grande, enquanto o AIC demonstra maior capacidade de seleção para modelos não estacionários, em particular em amostras pequenas. Nos modelos com dois parâmetros não estacionários, o AIC apresentou o melhor desempenho de forma consistente. Foram ainda identificados pontos de reversão (*reversed points*) em que o BIC ou o RV podem superar o AIC, dependendo da inclinação da tendência e do tamanho da amostra.

2.4 Exemplo de Aplicação

De modo a ilustrar a aplicação da metodologia dos valores extremos por bloco exposta nas secções anteriores, consideramos os dados relativos aos níveis máximos anuais do mar em duas cidades australianas, *Port Pirie* e *Fremantle*, disponíveis no pacote *ismev* do R.

Iniciámos a análise da série de máximos anuais do nível do mar em *Port Pirie*, abrangendo o período de 1923 a 1987, com a aplicação dos testes de Mann–Kendall e de Wald–Wolfowitz, descritos na Subsecção 2.3. O código utilizado está disponível no Anexo A.

O teste de Mann–Kendall conduziu a um valor- p aproximadamente igual a 0,7944, não permitindo rejeitar a hipótese nula de presença de tendência ao nível de significância de 5%. De igual modo, o teste de Wald–Wolfowitz produziu um valor- p de aproximadamente 0,8838, não havendo evidência estatística contra a aleatoriedade serial. Assim, assumimos que temos estacionaridade e independência como pressupostos.

Procedemos então à modelação dos dados através do ajuste do Modelo M_1 , recorrendo ao procedimento descrito no Anexo A. A maximização do logaritmo da função de verosimilhança conduziu às estimativas

$$(\hat{\mu}, \hat{\sigma}, \hat{\xi}) = (3,8747; 0,1980; -0,0501).$$

Os erros padrão associados são aproximadamente iguais a 0,028, 0,020 e 0,098, respetivamente. As estimativas dos níveis de retorno foram obtidas por substituição direta na expressão (2.14). Para um período de retorno de 10 anos, obtivemos $z_{0.1} = 4,30$, enquanto para 100 anos se obteve $z_{0.01} = 4,69$.

A adequação do Modelo M_1 foi avaliada através dos gráficos de diagnóstico obtidos com a função `gev.diag`, apresentados na Figura 2.3. Os gráficos de probabilidades e quantis evidenciam um bom alinhamento com a reta de referência, e a densidade ajustada é consistente com o histograma empírico, validando o ajustamento do modelo.

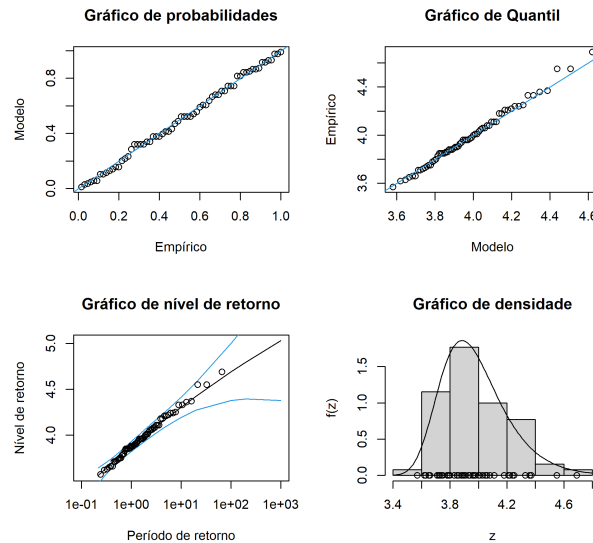


Figura 2.3: Gráficos de diagnóstico do Modelo M_1 - Máximos anuais do nível do mar (*Port Pirie*).

Como alternativa, ajustámos também o Modelo M_0 aos dados de Port Pirie, tendo obtido as estimativas

$$(\hat{\mu}, \hat{\sigma}) = (3,8694; 0,1949),$$

com erros padrão de 0,0255 e 0,0189, respetivamente. Os gráficos de diagnóstico correspondentes estão representados na Figura 2.4.

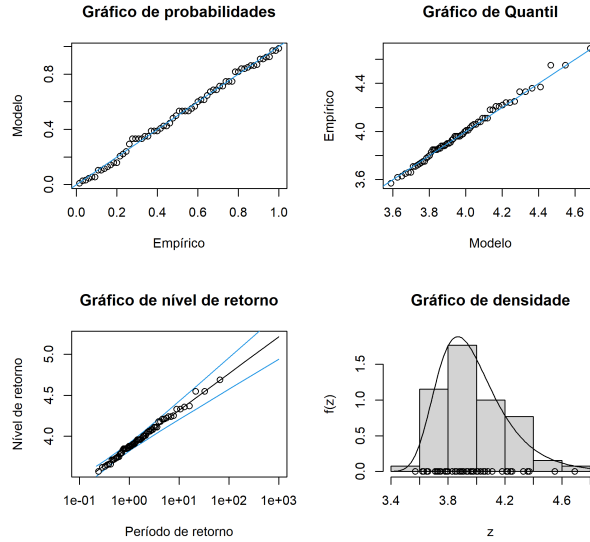


Figura 2.4: Gráficos de diagnóstico do Modelo M_0 - Máximos anuais do nível do mar (*Port Pirie*).

Para avaliar formalmente qual dos modelos fornece um melhor ajustamento aos dados de máximos anuais do nível do mar em *Port Pirie*, procedemos à comparação entre o Modelo M_0 e o Modelo M_1 , com base nos critérios de informação AIC e BIC e no teste da razão de verossimilhanças. O procedimento computacional utilizado está descrito no Anexo A. A Tabela 2.3 apresenta os valores obtidos para os critérios de informação para cada um dos modelos, bem como o valor observado da estatística de teste do teste da razão de verossimilhanças e correspondente valor- p .

Tabela 2.3: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC para escolha entre Modelos M_0 e M_1 - Máximos anuais do nível do mar (*Port Pirie*).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_0	-4,44	-0,09	0,24 ($\approx 0,6222$)
M_1	-2,68	3,85	

Observamos que o Modelo M_0 apresenta valores de AIC e BIC ligeiramente inferiores aos do Modelo M_1 , sugerindo um melhor compromisso entre qualidade de ajustamento e complexidade do modelo segundo estes critérios. Adicionalmente, no teste da razão de verossimilhanças não é rejeitada a hipótese nula de que o parâmetro de forma é nulo, uma vez que o valor- p obtido é elevado ($\approx 0,6222$). Assim, não existe evidência estatística

suficiente para a rejeição do Modelo M_0 em detrimento do Modelo M_1 para os máximos anuais do nível do mar (*Port Pirie*). Estes resultados são consistentes com a estimativa do parâmetro de forma próxima de zero e com os gráficos de diagnóstico anteriormente apresentados, indicando que o Modelo M_0 constitui uma descrição adequada dos extremos anuais do nível do mar em *Port Pirie*.

Apresentam-se de seguida os resultados da comparação entre os Modelos M_0 e M_1 , considerando versões estacionárias e não estacionárias com dependência temporal nos parâmetros de localização (μ) e de escala (σ), ajustados aos máximos anuais do nível do mar observados em *Port Pirie*. A avaliação dos modelos é efetuada com base nos critérios de informação AIC e BIC, bem como nos testes da razão de verosimilhança, com o objetivo de analisar a eventual presença de tendências temporais nos parâmetros dos modelos.

Tabela 2.4: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Máximos anuais do nível do mar (*Port Pirie*).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_0	-4,435	-0,087	-
$M_{0.A}$	-2,501	4,022	0,066 (0,797)
$M_{0.B}$	-2,422	4,101	-0,013 (1,000)
$M_{0.C}$	-0,436	8,262	0,001 (0,999)
M_1	-2,678	3,845	-
$M_{1.A}$	-0,750	7,947	0,072 (0,788)
$M_{1.B}$	-0,671	8,027	-0,007 (1,000)
$M_{1.C}$	1,322	12,194	-0,001 (1,000)

De um modo geral, verifica-se que, tanto no enquadramento do Modelo M_0 como do Modelo M_1 , as versões estacionárias apresentam os valores mais baixos dos critérios AIC e BIC. Adicionalmente, os testes da razão de verosimilhança não revelam evidência estatisticamente significativa que justifique a introdução de não estacionaridade nos parâmetros de localização ou de escala. Estes resultados indicam que a hipótese de estacionaridade é adequada para a modelação dos máximos anuais do nível do mar em *Port Pirie*, não sendo detetada uma tendência temporal relevante nos parâmetros dos modelos considerados.

Passamos agora à análise da série de máximos anuais do nível do

mar em *Fremantle*, abrangendo o período de 1897 a 1989. Os testes estatísticos de Mann–Kendall e de Wald–Wolfowitz foram realizados em R, conforme indicado no Anexo B. O teste de Mann–Kendall produziu um valor da estatística aproximadamente igual a 2,9317 e um valor- p aproximado de 0,003371, indicando a presença de uma tendência monotônica estatisticamente significativa ao nível de significância de 5%. Por sua vez, o teste de Wald–Wolfowitz apresentou um valor- p aproximadamente igual a 0,1649, não fornecendo evidência estatística suficiente para rejeitar a hipótese de aleatoriedade serial. Assim, concluímos que os máximos anuais do nível do mar em *Fremantle* exibem uma tendência crescente significativa, não se detetando, contudo, indícios claros de violação da independência dos dados.

Procedemos então ao ajustamento do Modelo M_1 , recorrendo ao método da máxima verosimilhança, conforme descrito no Anexo B. A adequação do ajustamento foi avaliada através dos gráficos de diagnóstico obtidos com a função `gev.diag`, apresentados na Figura 2.5.

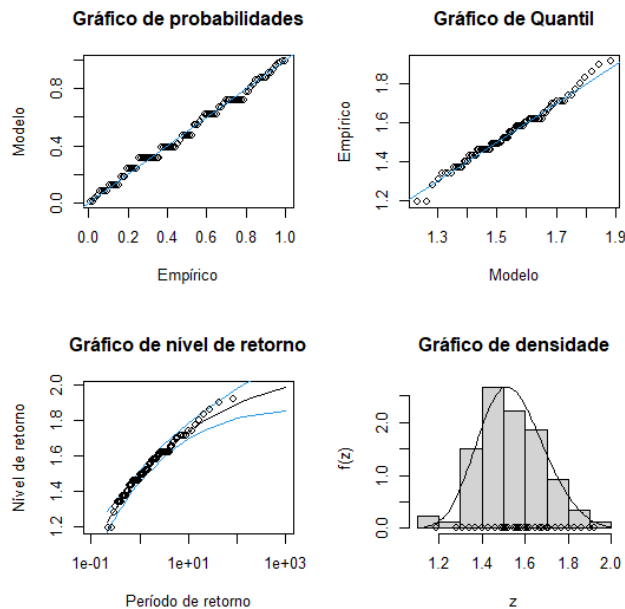


Figura 2.5: Gráficos de diagnóstico do Modelo M_1 - Máximos anuais do nível do mar (*Fremantle*).

Os gráficos de probabilidades e de quantis revelam um alinhamento satisfatório dos pontos com a reta de referência, não evidenciando desvios sistemáticos relevantes. O gráfico dos níveis de retorno apresenta um

comportamento aproximadamente linear, enquanto a densidade ajustada se mostra coerente com o histograma empírico, sugerindo um bom ajustamento global do Modelo M_1 aos dados observados.

De forma análoga ao procedimento adotado para máximos anuais do nível do mar em *Port Pirie*, ajustamos igualmente o Modelo M_0 aos máximos anuais do nível do mar em *Fremantle*, sendo a adequação deste modelo avaliada com base nos gráficos de diagnóstico obtidos através da função `gum.diag`. Estes gráficos estão representados na Figura 2.6.

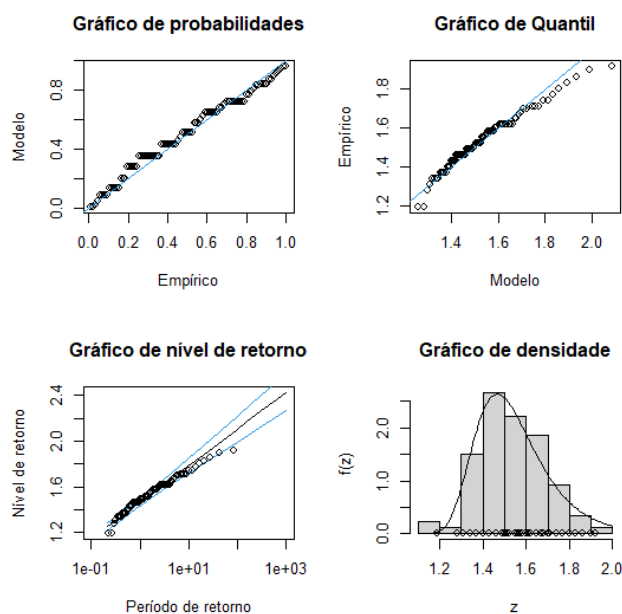


Figura 2.6: Gráficos de diagnóstico do Modelo M_0 - Máximos anuais do nível do mar (*Fremantle*).

Embora os gráficos de diagnóstico do Modelo M_0 apresentem um ajustamento visualmente razoável, observamos uma maior discrepância nos quantis extremos quando comparados com os obtidos para o Modelo M_1 , sugerindo que a restrição $\xi = 0$ poderá não ser adequada para descrever o comportamento da cauda superior da distribuição.

Para avaliar formalmente qual dos dois modelos fornece um melhor ajustamento aos máximos anuais do nível do mar em *Fremantle*, procedemos à comparação entre os Modelos M_0 e M_1 através dos critérios de informação AIC e BIC e do teste da razão de verossimilhanças, conforme descrito no Anexo A. A Tabela 2.5 evidencia que o Modelo M_1 apresenta valores

inferiores de AIC e BIC, indicando um melhor compromisso entre qualidade de ajustamento e complexidade do modelo. O teste da razão de verossimilhanças conduz a um valor- p reduzido, permitindo rejeitar o Modelo M_0 em favor do Modelo M_1 , ao nível de significância de 5%.

Tabela 2.5: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC para escolha entre os Modelos M_0 e M_1 - Máximos anuais do nível do mar (*Fremantle*).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_0	-79,09	-73,73	10,04
M_1	-81,13	-73,77	($\approx 0,00153$)

A estimativa de máxima verossimilhança para o parâmetro de forma do Modelo M_1 é aproximadamente igual $-0,217$, sendo o respetivo intervalo de confiança a 95% dado por $(-0,342; -0,092)$. O facto de este intervalo não incluir o valor zero constitui evidência estatística adicional contra o modelo M_0 , confirmando a necessidade de recorrer ao Modelo M_1 , mais geral, para descrever os extremos anuais do nível do mar em *Fremantle*.

Com o objetivo de avaliar a eventual presença de tendência temporal nos parâmetros do Modelo M_1 , foram consideradas extensões do modelo estacionário que incorporam o ano como covariável nos parâmetros de localização e escala. Em particular, analisámos o modelo com tendência linear em μ (Modelo $M_{1.A}$), o modelo com tendência em σ (Modelo $M_{1.B}$) e o modelo com tendência simultânea em μ e σ (Modelo $M_{1.C}$). Os correspondentes valores dos critérios de informação AIC e BIC estão apresentados na Tabela 2.6, juntamente com o valor observado para a estatística d_{obs} e o respetivo valor- p do teste de razão de verossimilhanças.

Tabela 2.6: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC - Máximos anuais do nível do mar (*Fremantle*).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_1	-81,133	-73,770	-
$M_{1.A}$	-91,826	-82,008	12,692 ($\approx 0,367 \times 10^{-3}$)
$M_{1.B}$	-81,531	-71,713	2,397 ($\approx 0,122$)
$M_{1.C}$	-79,809	-67,537	2,675 ($\approx 0,262$)

A análise da Tabela 2.6 revela que o modelo com tendência linear no parâmetro de localização apresenta uma melhoria estatisticamente

significativa relativamente ao modelo estacionário, evidenciada quer pela redução substancial dos valores de AIC e BIC, quer pelo teste da razão de verossimilhanças, cujo valor- p é inferior a 0,001.

Em contraste, os modelos que introduzem dependência temporal apenas no parâmetro de escala, ou simultaneamente nos parâmetros de localização e escala, não apresentam melhorias estatisticamente significativas face ao modelo base, sendo penalizados pelos critérios de informação.

Assim, ao contrário do observado para os máximos anuais do nível do mar em *Port Pirie*, existe evidência estatística consistente de uma tendência temporal nos máximos anuais do nível do mar em *Fremantle*, a qual se manifesta predominantemente através de uma evolução crescente do parâmetro de localização da distribuição GEV ao longo do tempo.

Capítulo 3

Máximos Anuais das MMMD de Temperatura

O estudo dos extremos climáticos, nomeadamente de valores de temperatura, está revestido de particular importância para a compreensão de fenómenos térmicos extremos, como ondas de calor prolongadas, que têm repercussões relevantes em domínios como a saúde humana, a produção agrícola, a gestão de recursos energéticos e o planeamento ambiental sustentável.

Neste capítulo, foram analisados valores de temperatura registados na estação meteorológica do Funchal. As medições diárias consistiram em três leituras da temperatura ao longo do dia. Para cada dia, registou-se o valor máximo da temperatura; estes valores diários máximos foram posteriormente agregados para obter a média mensal das máximas diárias, calculada como a soma dos máximos diários dividida pelo número de dias do mês. A partir destes dados mensais, foram extraídos os máximos anuais, que constituem a base de dados formada pelos máximos anuais das MMMD de temperatura utilizada neste capítulo.

A análise foi conduzida com recurso a um *script* em R, apresentado no Anexo D, que assegura o pré-processamento dos dados, a definição dos diferentes grupos de análise - incluindo os subconjuntos obtidos por deteção de pontos de mudança através da função `change.point` do pacote `RecordTest`, bem como os períodos associados a alterações geográficas e tecnológicas da estação e o cálculo de medidas descritivas. A divisão dos dados em seis subconjuntos: três resultantes da deteção de pontos de mudança, designados por Períodos Temporais I, e três associados a diferentes sistemas de observação, designados por Períodos Temporais II, e, conseqüente, obtenção de algumas medidas de localização e dispersão estão apresentadas

na Secção 3.1. No *script* estão também os procedimentos realizados para análise dos pressupostos de independência e estacionariedade apresentada na Secção 3.2, tal como o ajustamento dos Modelos M_0 e M_1 , a determinação de níveis de retorno, a seleção e validação de modelos que constituem a análise apresentada na Secção 3.3. O *script* implementa ainda funções auxiliares que garantem a consistência da análise, permitindo aplicar de forma sistemática todos os procedimentos à série completa e a cada subconjunto, bem como organizar automaticamente os resultados, os quais são posteriormente utilizados na construção das tabelas, gráficos e conclusões apresentados ao longo deste capítulo. Uma comparação global dos resultados obtidos para os diferentes intervalos temporais considerados é realizada na Secção 3.4.

3.1 Estatística Descritiva

Na análise da série em estudo, designada por `datatempmax` e denotada a partir de agora por Dados (1865- 2018), foram detetados pontos de mudança com a aplicação da função `change.point` do pacote `RecordTest` do R, os quais resultaram nos seguintes conjuntos de dados: Dados 1 (1865-1896), Dados 2 (1897-1974) e Dados 3 (1975-2018). Tendo em conta as alterações geográficas e tecnológicas na estação do Funchal, os dados foram também divididos segundo os diferentes sistemas de observação, refletindo o impacto da mudança de local e do tipo de equipamento nas medições. Esta abordagem resultou no estudo dos máximos anuais das MMMD de temperatura nos períodos de 1865 e 1950 (Dados 4 (1865- 1950)), de 1951 e 1996 (Dados 5 (1951-1996)) e de 1996 e 2018 (Dados 6 (1997-2018)).

Antes de prosseguir com a modelação dos máximos anuais das MMMD de temperatura, é essencial caracterizar estatisticamente a série de dados e os seus subconjuntos. Os valores de algumas medidas de localização e de dispersão para os Dados (1865- 2018) e para os seis subconjuntos de dados através do *script* em R obtidos estão apresentados nas Tabelas 3.1 e 3.2. Observamos, em primeiro lugar, uma tendência de aumento dos valores centrais ao longo do tempo. A mediana aumenta de $25,0^{\circ}\text{C}$ nos Dados 1 (1865-1896) para $25,2^{\circ}\text{C}$ nos Dados 2 (1897-1974) e atinge $26,5^{\circ}\text{C}$ no período mais recente, refletindo uma elevação gradual dos máximos anuais das MMMD de temperatura ao longo das diferentes fases históricas. De forma consistente, nos Dados 4 a 6 definidos pelas alterações da estação do IPMA, a mediana também aumenta de $25,0^{\circ}\text{C}$ nos Dados 4 (1865-1950) para $25,9^{\circ}\text{C}$ nos Dados 5 (1951-1996) e $26,8^{\circ}\text{C}$ nos Dados 6 (1997-2018), evidenciando que

a tendência de aquecimento é observada independentemente das mudanças nos sistemas de medição.

Tabela 3.1: Dimensão, média, desvio padrão e coeficiente de variação - Máximos Anuais das MMMD de Temperatura.

	n	Média	Desvio-padrão	CV (%)
Dados (1865-2018)	154	25,58	1,00	3,90
Dados 1 (1865-1896)	32	25,07	0,52	2,07
Dados 2 (1897-1974)	78	25,23	0,76	3,02
Dados 3 (1975-2018)	44	26,65	0,91	3,41
Dados 4 (1865-1950)	86	25,08	0,70	2,79
Dados 5 (1951-1996)	46	25,95	0,91	3,49
Dados 6 (1997-2018)	22	26,88	0,73	2,73

Tabela 3.2: Mediana (M), primeiro e terceiro quartis, extremos e amplitude amostral (R_n) - Máximos Anuais das MMMD de Temperatura.

	M	Q_1	Q_3	Mínimo	Máximo	R_n
Dados (1865-2018)	25,6	24,6	26,5	23,7	28,7	5,0
Dados 1 (1865-1896)	25,0	24,6	25,5	24,1	26,2	2,1
Dados 2 (1897-1974)	25,2	24,3	26,1	23,7	27,3	3,6
Dados 3 (1975-2018)	26,5	25,8	27,2	25,1	28,7	3,6
Dados 4 (1865-1950)	25,0	24,4	25,7	23,7	27,3	3,6
Dados 5 (1951-1996)	25,9	25,35	26,6	24,4	28,7	4,3
Dados 6 (1997-2018)	26,8	26,2	27,4	25,6	28,1	2,5

As medidas de dispersão indicam que a variabilidade dos dados permanece relativamente baixa em todos os períodos. O desvio padrão varia entre 0,52 °C e 0,91 °C nos Períodos Temporais I e entre 0,70 °C e 0,91 °C nos Períodos Temporais II, traduzindo a dispersão dos máximos anuais das MMMD de temperatura em torno da respectiva média em cada conjunto de dados. Contudo, tratando-se de uma medida absoluta de variabilidade, a sua utilização deve ser interpretada sobretudo de forma descritiva, não sendo a mais adequada para a comparação direta da variabilidade entre grupos de dados com médias distintas. O coeficiente de variação confirma esta estabilidade relativa, com valores sempre inferiores a 4%, enquanto a amplitude toma os valores 2,1°C e 3,6°C nos Períodos Temporais I e os valores 2,5°C, 3,6°C e 4,3°C nos Períodos Temporais II.

Em termos comparativos, para os Períodos Temporais I e II, observamos que a média, a mediana e os quartis são mais elevados nos períodos mais recentes (1975-2018 e 1997-2018, respetivamente) do que nos períodos iniciais (1865-1896 e 1865-1950, respetivamente), sugerindo um padrão consistente de aquecimento nos máximos anuais das MMMD de temperatura registados na estação do Funchal. Além disso, a relativa estabilidade do coeficiente de variação indica que, apesar do aumento das temperaturas centrais, a variabilidade relativa dos valores ao longo dos anos não apresentou alterações significativas, reforçando a robustez da tendência observada.

A análise dos Dados (1865-2018) evidencia que os máximos anuais das MMMD de temperatura apresentam uma média de $25,58^{\circ}\text{C}$, com valores mínimos e máximos de $23,7^{\circ}\text{C}$ e $28,7^{\circ}\text{C}$, respetivamente. O primeiro quartil ($Q_1 = 24,6^{\circ}\text{C}$), a mediana ($25,6^{\circ}\text{C}$) e o terceiro quartil ($Q_3 = 26,5^{\circ}\text{C}$) indicam uma distribuição relativamente simétrica e moderadamente concentrada em torno da média. A amplitude total de $5,0^{\circ}\text{C}$ e o desvio-padrão de $1,0^{\circ}\text{C}$ revelam uma dispersão reduzida dos valores, enquanto o coeficiente de variação de $3,90\%$ confirma a estabilidade relativa da série. Estes resultados sugerem que, ao longo do período analisado, a variabilidade interanual dos máximos anuais das MMMD de temperatura foi limitada.

3.2 Estacionaridade e Independência

Na Tabela 3.3 estão apresentados os valores- p obtidos na aplicação dos testes não paramétricos de Mann-Kendall e de Wald-Wolfowitz para cada um dos conjuntos de dados descritos na Secção 3.1, aplicados com o objetivo de avaliar os pressupostos de estacionaridade e a independência, respetivamente.

Para um nível de significância de 5% , há evidência estatística suficiente para a rejeição das hipóteses nulas de ausência de tendência e de aleatoriedade nos Dados (1865-2018) uma vez que os correspondentes valores- p ($6,1 \times 10^{-13}$ e $5,8 \times 10^{-10}$) são inferiores a $0,05$. As hipóteses nulas também são rejeitadas para os Dados 4 (1865-1950), onde os valores- p dos testes de Mann-Kendall e de Wald-Wolfowitz são aproximadamente iguais a $0,0001$ e $0,0035$, respetivamente. Na análise dos Dados 5 (1951-1996), a rejeição da hipótese nula ocorre apenas para o teste de Mann-Kendall (valor- $p \approx 0,039$).

Por outro lado, nos restantes subperíodos analisados, os valores- p são maiores do que o nível de significância de $0,05$, não existindo evidência

estatística suficiente para rejeitar as respectivas hipóteses nulas. No entanto, observa-se um valor- p aproximadamente igual a 0,054 para o teste de Mann-Kendall para os Dados 3 (1975-2018). Em resumo, a flutuação da significância estatística entre os diversos intervalos de tempo analisados valida a relevância da análise por subperíodos para uma caracterização do comportamento dos máximos em estudo.

Tabela 3.3: Valores- p para os testes de Mann-Kendall e de Wald-Wolfowitz.

	Teste de Mann-Kendall (valor-p)	Teste de Wald-Wolfowitz (valor-p)
Dados (1865-2018)	$6,1 \times 10^{-13}$	$5,8 \times 10^{-10}$
Dados 1 (1865-1896)	0,721	0,229
Dados 2 (1897-1974)	0,213	0,211
Dados 3 (1975-2018)	0,054	0,461
Dados 4 (1865-1950)	$< 0,0001$	0,0035
Dados 5 (1951-1996)	0,039	0,114
Dados 6 (1997-2018)	0,631	0,789

3.3 Seleção e Validação de Modelos

3.3.1 Modelo M_0 versus Modelo M_1

Na Tabela 3.4 estão apresentados os resultados do teste de razão de verossimilhanças, nomeadamente o valor observado da estatística de teste, d_{obs} , e o correspondente valor- p para todos os conjuntos de dados analisados. Observamos que nos Dados (1865-2018), o valor- p obtido ($\approx 0,663$) é maior que o nível de significância de 0,05. Este valor indica que não existe evidência estatística suficiente para rejeitar $\xi = 0$, sugerindo que o modelo mais simples, M_0 , é suficiente para modelar os máximos anuais das MMMD de temperatura em estudo. No que concerne à análise por subperíodos, o comportamento mantém-se consistente. Em todos os intervalos analisados, desde os Dados 1 (1865-1896) até aos Dados 6 (1997-2018), os valores- p obtidos situam-se acima do nível de significância de 0,05. Importa notar que nos Dados 2 (1897-1974) observa-se o valor- p mais baixo ($\approx 0,094$).

Foram também obtidos para todos os conjuntos de dados analisados, os valores dos critérios de informação AIC e BIC obtidos para os modelos M_0

Tabela 3.4: Valores de d_{obs} e valor- p para o teste de razão de verossimilhanças.

	d_{obs}	valor- p
Dados (1865-2018)	0,19	0,663
Dados 1 (1865-1896)	1,21	0,270
Dados 2 (1897-1974)	2,80	0,094
Dados 3 (1975-2018)	0,90	0,344
Dados 4 (1865-1950)	0,27	0,606
Dados 5 (1951-1996)	0,30	0,586
Dados 6 (1997-2018)	1,39	0,239

e M_1 , os quais estão indicados na Tabela 3.5. Verifica-se que, em todos os casos analisados, o modelo M_0 apresenta valores inferiores de AIC e de BIC relativamente ao modelo M_1 . Atendendo a que ambos os critérios avaliam a qualidade do ajustamento penalizando a complexidade do modelo, a obtenção de valores mais reduzidos constitui um indicador de melhor desempenho. Assim, os resultados evidenciam, de forma sistemática, uma preferência pelo modelo M_0 .

Tabela 3.5: Valores dos critérios de informação AIC e BIC.

	Modelo M_0		Modelo M_1	
	AIC	BIC	AIC	BIC
Dados (1865–2018)	447,36	453,56	449,17	458,47
Dados 1 (1865–1897)	56,30	59,41	57,08	61,75
Dados 2 (1897–1974)	56,30	59,41	57,08	61,75
Dados 3 (1975–2018)	118,08	121,64	119,18	124,53
Dados 4 (1865-1950)	188,22	192,72	189,95	196,70
Dados 5 (1951-1996)	116,92	120,66	118,62	124,23
Dados 6 (1997–2018)	52,52	54,71	53,14	56,41

Analisaremos em seguida mais detalhadamente a escolha do modelo em cada um dos períodos estudados.

3.3.2 Dados (1865-2018)

Para os máximos anuais das MMMD de temperatura, procedeu-se ao ajustamento dos modelos de valores extremos M_0 e M_1 por máxima verossimilhança. No caso do modelo M_0 , obtivemos uma estimativa do parâmetro de localização de $\hat{\mu} = 25,118$ e do parâmetro de escala de $\hat{\sigma} = 0,800$. Relativamente ao modelo M_1 , as estimativas dos parâmetros de

localização e escala foram $\hat{\mu} = 25,130$ e $\hat{\sigma} = 0,807$, respetivamente, valores muito próximos dos obtidos com o modelo de Gumbel, o que sugere uma boa concordância entre os dois ajustamentos.

A estimativa do parâmetro de forma foi $\hat{\xi} = -0,027$, com um intervalo de confiança a 95 % compreendido entre $-0,149$ e $0,094$. Uma vez que o valor zero se encontra contido neste intervalo, não existe evidência estatística suficiente para rejeitar a hipótese de $\xi = 0$, sendo o Modelo M_0 um modelo que poderá ser aplicado na modelação dos máximos anuais das MMMD de temperatura na estação do Funchal.

Nas Figuras 3.1 e 3.2 estão apresentados os gráficos de diagnóstico associados aos ajustamentos dos Modelos M_0 e M_1 , respetivamente, permitindo avaliar visualmente a adequação dos modelos aos dados observados.

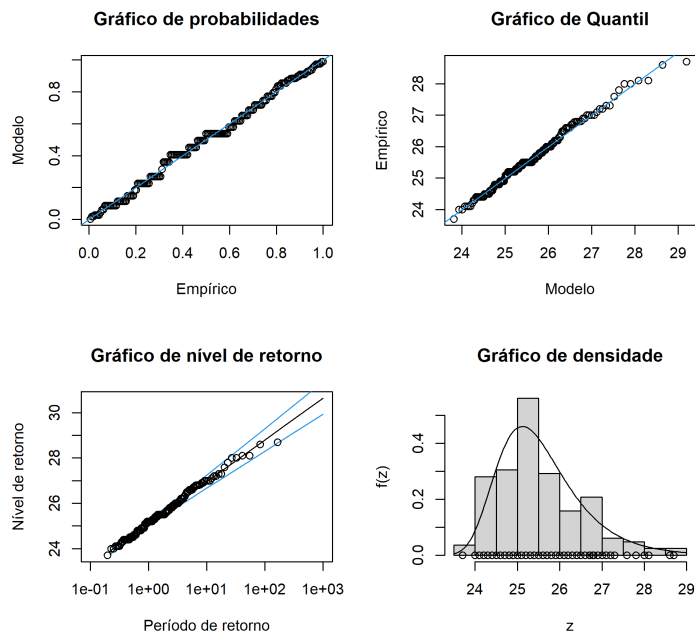


Figura 3.1: Gráficos de diagnóstico para o Modelo M_0 - Dados (1865–2018).

A consideração do Modelo M_1 conduz à estimação de níveis de retorno que, para um período de retorno de 10 anos, resultam num valor extremo de $z_{0,1} = 26,89^\circ\text{C}$ e para um período de retorno de 100 anos, o nível de retorno estimado é $z_{0,01} = 28,62^\circ\text{C}$. Por outro lado, o ajustamento do Modelo M_0 conduz a níveis de retorno de 10 e 100 anos de $26,95^\circ\text{C}$ e $28,84^\circ\text{C}$,

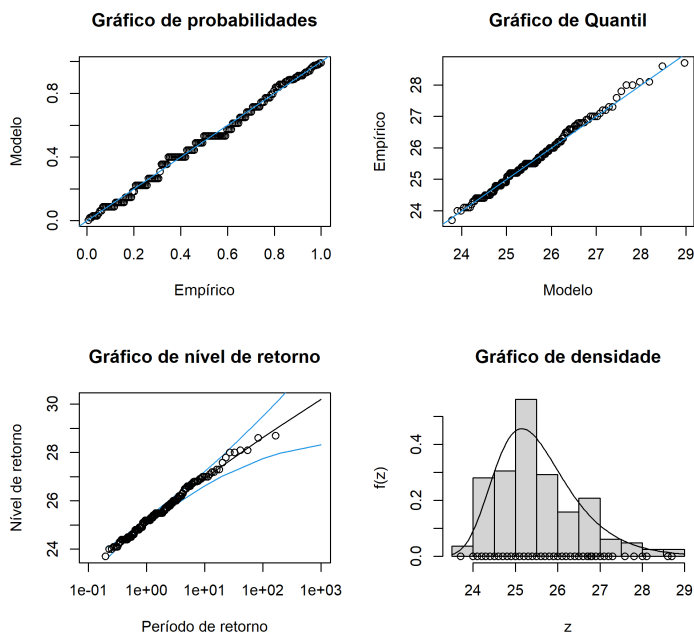


Figura 3.2: Gráficos de diagnóstico para o Modelo M_1 - Dados (1865–2018).

respetivamente, que constituem valores muito próximos dos obtidos com o Modelo M_1 .

Até este ponto, a análise incidiu sobre os modelos estacionários M_0 e M_1 . Seguidamente, procedemos à análise dos modelos apresentados na Tabela 2.2, que incorporam a variável temporal *ano* como covariável nos parâmetros de localização e dispersão. Os valores dos critérios de informação AIC e BIC, juntamente com o valor observado para a estatística d_{obs} e o respetivo valor- p do teste da razão de verosimilhança para cada teste entre os Modelos M_0 e M_1 e os correspondentes modelos com a covariável *ano*, estão apresentados na Tabela 3.6.

A comparação entre modelos através do teste de razão de verosimilhança revelou que a inclusão da covariável na média relativa ao Modelo M_1 (Modelo $M_{1.A}$) não produziu uma melhoria estatisticamente significativa ($d_{obs} \approx 0,229$, valor- $p \approx 6.32 \times 10^{-1}$), enquanto a inclusão na escala (Modelo $M_{1.B}$) evidenciou uma diferença significativa relativamente ao modelo base ($d_{obs} \approx 5,355$, valor- $p \approx 2.07 \times 10^{-2}$). O Modelo $M_{1.C}$, com a covariável *ano* incluída simultaneamente na média e na escala, apresentou a maior melhoria em termos de log-verosimilhança, sendo a diferença estatisticamente altamente significativa ($d_{obs} \approx 28,757$, valor- $p \approx 5.70 \times 10^{-7}$). Com base

Tabela 3.6: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Dados (1865–2018).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_0	447,166	453,365	–
$M_{0,A}$	448,937	458,236	0,229 ($\approx 6.32 \times 10^{-1}$)
$M_{0,B}$	443,811	453,110	5,355 ($\approx 2.07 \times 10^{-2}$)
$M_{0,C}$	422,409	434,808	28,757 ($\approx 5.70 \times 10^{-7}$)
M_1	449,166	458,465	–
$M_{1,A}$	450,937	463,336	0,229 ($\approx 6.32 \times 10^{-1}$)
$M_{1,B}$	445,811	458,210	5,355 ($\approx 2.07 \times 10^{-2}$)
$M_{1,C}$	424,409	439,908	28,757 ($\approx 5.70 \times 10^{-7}$)

nos critérios de informação AIC e BIC, observa-se também que o modelo preferível é o modelo $M_{1,C}$. Analogamente, é observado que o modelo preferível é o modelo $M_{0,C}$ quando é considerado o Modelo M_0 como modelo base.

3.3.3 Análise de Períodos Temporais I

Dados 1 (1865–1897)

A análise aqui efetuada, correspondente ao período de 1865 a 1897, seguiu o mesmo procedimento adotado para na secção anterior relativa aos Dados (1865–2018). O Modelo M_1 foi ajustado aos máximos anuais das MMMD de temperatura, tendo sido obtido estimativas dos parâmetros de localização, escala e forma de aproximadamente 24, 867, 0, 470 e $-0, 177$, respetivamente.

O parâmetro de forma apresenta um intervalo de confiança a 95% dado por $[-0, 476, 0, 121]$. O intervalo de confiança a 95% do parâmetro de forma inclui assim o valor zero, não existindo evidência estatisticamente significativa de comportamento de cauda distinto do Modelo M_0 . Por sua vez, o Modelo M_0 estimou os parâmetros de localização e escala em cerca de 24, 824 e 0, 446, respetivamente.

De seguida, são apresentados nas Figuras 3.3 e 3.4 os gráficos de diagnóstico associados aos ajustamentos dos Modelos M_0 e M_1 , respetivamente, permitindo avaliar visualmente a adequação dos modelos aos dados observados.

34CAPÍTULO 3. MÁXIMOS ANUAIS DAS MMMD DE TEMPERATURA

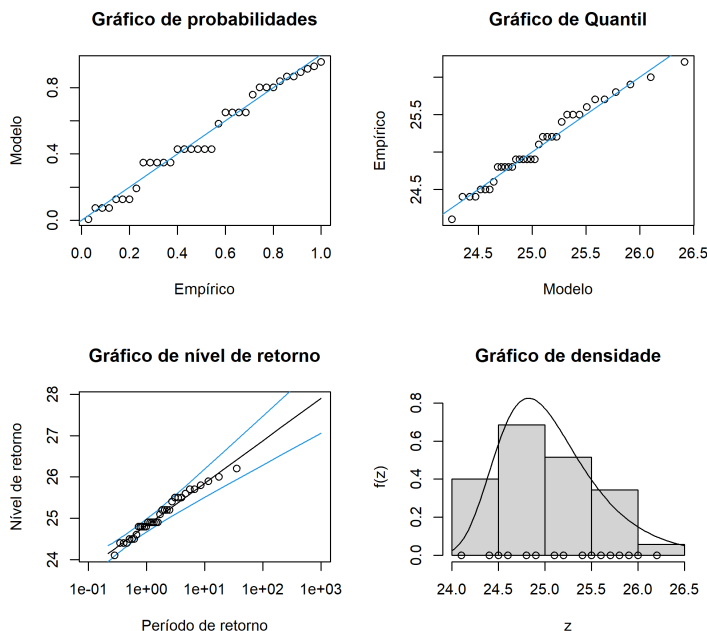


Figura 3.3: Gráficos de diagnóstico para o Modelo M_0 - Dados 1 (1865–1897).

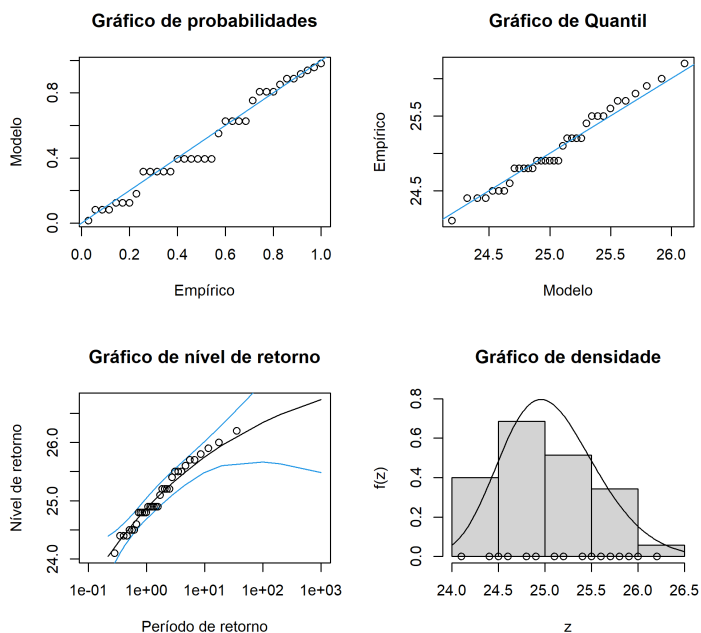


Figura 3.4: Gráficos de diagnóstico para o Modelo M_1 - Dados 1 (1865–1897).

Relativamente aos níveis de retorno, verificamos que, para este período e considerando o Modelo M_1 , o nível de retorno de 10 anos é 25,74 °C e o nível

de retorno de 100 anos é 26,34°C. Se for o Modelo M_0 o modelo considerado, os níveis de retorno calculados para 10 e 100 anos são aproximadamente iguais a 25,92°C e 27,03°C, respectivamente.

Seguidamente, foram considerados modelos GEV não estacionários com a inclusão da variável temporal *ano* como covariável nos parâmetros de localização (Modelo $M_{1.A}$), de escala (Modelo $M_{1.B}$) e simultaneamente em ambos os parâmetros (Modelo $M_{1.C}$), bem como o modelo Gumbel estacionário (Modelo M_0) e variantes não estacionárias (Modelos $M_{0.A}$, $M_{0.B}$, $M_{0.C}$). Os correspondentes valores dos critérios de informação AIC e BIC, juntamente com o valor observado para a estatística d_{obs} e o respetivo valor- p do teste da razão de verosimilhanças, encontram-se apresentados na Tabela 3.7.

Tabela 3.7: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Dados 1 (1865–1897).

Modelo	AIC	BIC	d_{obs}	p-valor
M_0	55,081	58,191	–	–
$M_{0.A}$	57,048	61,714	0,033	0,856
$M_{0.B}$	59,433	64,100	-2,353	1,000
$M_{0.C}$	61,361	67,582	-2,280	1,000
M_1	57,081	61,747	–	–
$M_{1.A}$	59,048	65,269	0,033	0,856
$M_{1.B}$	61,433	67,655	-2,353	1,000
$M_{1.C}$	63,361	71,137	0,073	0,787

Com base nos critérios de informação AIC e BIC, não se observa melhoria significativa do ajuste ao considerar modelos não estacionários qualquer que seja o modelo base considerado (Modelo M_0 ou Modelo M_1). Observou-se também que nenhum dos Modelos M_0 e M_1 é rejeitado face aos correspondentes modelos não estacionários, na aplicação do teste de razão de versomilhanças, uma vez que todos os valores- p observados são maiores do que 0.05.

Dados 2 (1897–1974)

Considerando o período de 1897 a 1974, o Modelo M_1 forneceu estimativas para os parâmetros de localização, escala e forma aproximadamente iguais a 24,924, 0,865 e $-0,146$, respectivamente. O Modelo M_0 apresentou

estimativas para os parâmetros de localização e escala da ordem de 26,871 e 0,661, respetivamente.

O parâmetro de forma apresenta um intervalo de confiança a 95% dado por $[-0,299; 0,008]$, indicando que não difere significativamente de zero.

De seguida, são apresentados nas Figuras 3.5 e 3.6 os gráficos de diagnóstico associados aos ajustamentos dos Modelos M_0 e M_1 , respetivamente, permitindo avaliar visualmente a adequação dos modelos aos dados observados.

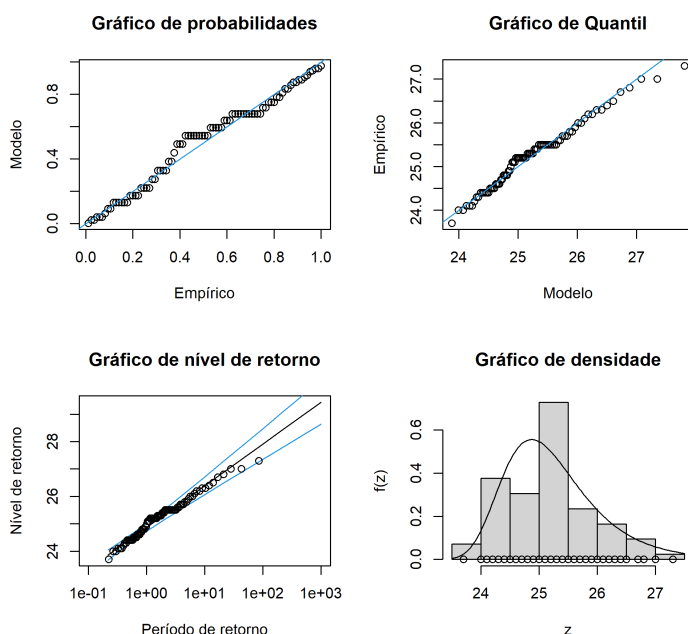


Figura 3.5: Gráficos de diagnóstico para o Modelo M_0 - Dados 2 (1897–1974).

No que respeita aos níveis de retorno, o nível associado a um período de retorno de 10 anos é $26,24^\circ\text{C}$ e para um período de retorno de 100 anos, o valor extremo esperado é $27,22^\circ\text{C}$. Observamos um aumento das estimativas pontuais dos níveis de retorno relativamente ao período anterior, sugerindo uma elevação gradual dos máximos anuais das MMMD de temperatura. Se for o Modelo M_0 o modelo considerado, os níveis de retorno calculados para 10 e 100 anos são aproximadamente iguais a $26,4792^\circ\text{C}$ e $28,08^\circ\text{C}$, respetivamente.

Adicionalmente, foram ajustados modelos com dependência temporal

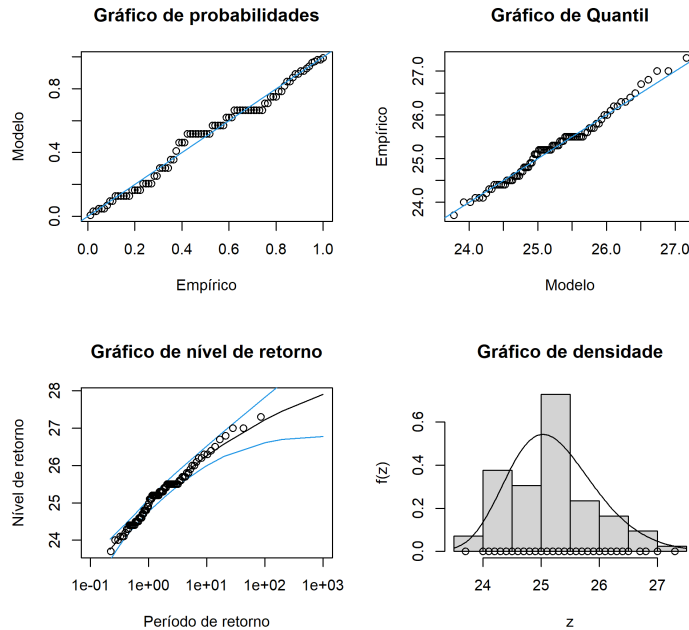


Figura 3.6: Gráficos de diagnóstico para o Modelo M_1 - Dados 2 (1897–1974).

nos parâmetros de localização (Modelo $M_{1.A}$), de escala (Modelo $M_{1.B}$) e em ambos os parâmetros (Modelo $M_{1.C}$), considerando o *ano* como covariável explicativa. A Tabela 3.8 apresenta os correspondentes valores dos critérios de informação AIC e BIC, bem como o valor observado para a estatística d_{obs} e o respectivo valor- p do teste da razão de verossimilhanças, incluindo igualmente os valores correspondentes para os modelos relativos ao Modelo M_0 .

Tabela 3.8: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Dados 2 (1897–1974).

Modelo	AIC	BIC	d_{obs}	p-valor
M_0	194,174	199,060	–	–
$M_{0.A}$	196,121	203,449	0,054	0,817
$M_{0.B}$	203,595	210,923	-7,421	1,000
$M_{0.C}$	204,748	214,519	-6,574	1,000
M_1	196,174	203,502	–	–
$M_{1.A}$	198,121	207,891	0,054	0,817
$M_{1.B}$	205,595	215,366	-7,421	1,000
$M_{1.C}$	206,748	218,962	-6,574	1,000

Analogamente ao observado na secção anterior, com base nos critérios de informação AIC e BIC, não se observa melhoria significativa do ajuste ao considerar modelos não estacionários qualquer que seja o modelo base considerado (Modelo M_0 ou Modelo M_1). Observou-se também que nenhum dos Modelos M_0 e M_1 é rejeitado face aos correspondentes modelos não estacionários, na aplicação do teste de razão de versosmilhanças, uma vez que todos os valores- p observados são maiores do que 0.05.

Dados 3 (1975–2018)

Considerando o período de 1975 a 2018, o ajuste do Modelo M_1 conduziu a estimativas dos parâmetros de localização, escala e forma aproximadamente iguais a $\hat{\mu} \approx 26,271$, $\hat{\sigma} \approx 0,801$ e $\hat{\xi} \approx -0,135$, respetivamente. Por sua vez, o Modelo M_0 apresentou estimativas $\hat{\mu} \approx 26,214$ e $\hat{\sigma} \approx 0,767$, evidenciando valores muito próximos para os parâmetros de localização e dispersão quando comparado com o modelo mais geral.

Relativamente ao parâmetro de forma do Modelo M_1 foi obtido o intervalo de confiança a 95% igual a $[-0,395, 0,125]$. Uma vez que o intervalo inclui o valor zero, não se observa evidência estatisticamente significativa de uma cauda distinta da distribuição de Gumbel, o que reforça a adequação do Modelo M_0 para este conjunto de dados.

Os gráficos de diagnóstico associados aos ajustamentos dos Modelos M_0 e M_1 estão apresentados nas Figuras 3.7 e 3.8, permitindo avaliar visualmente a qualidade do ajustamento aos dados observados.

Quando ajustados os Modelos M_0 e M_1 aos máximos anuais das MMMD de temperatura, são obtidas estimativas para os níveis de retorno mais elevadas do que nos períodos anteriores. Em particular, quando consideramos o Modelo M_1 , o nível de retorno associado a um período médio de 10 anos é $27,83^\circ\text{C}$ e o nível de retorno de 100 anos é $29,02^\circ\text{C}$. Se for o Modelo M_0 o modelo considerado, os níveis de retorno calculados para 10 e 100 anos são aproximadamente iguais a $28,07^\circ\text{C}$ e $29,96^\circ\text{C}$, respetivamente.

Analogamente, foi em seguida realizada uma análise entre os Modelos M_0 e M_1 e os modelos com dependência temporal, considerando o *ano* como covariável explicativa na localização e/ou na escala. A Tabela 3.9 apresenta os valores dos critérios de informação AIC e BIC obtidos, bem como o valor observado da estatística de teste d_{obs} e o respetivo valor- p do teste da razão de verosimilhanças.

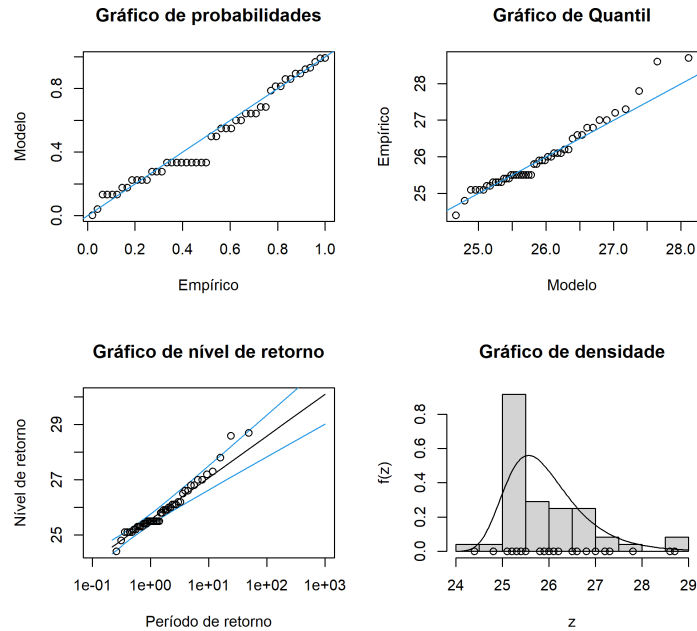


Figura 3.7: Gráficos de diagnóstico para o Modelo M_0 - Dados 3 (1975–2018).

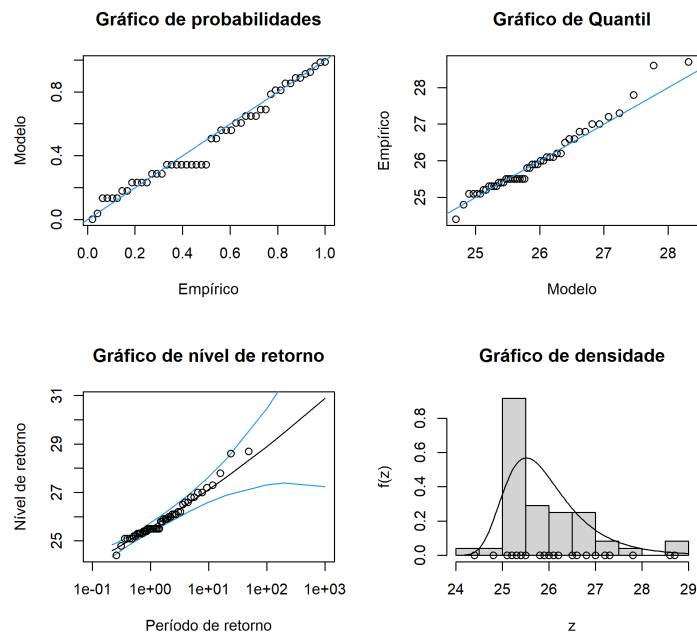


Figura 3.8: Gráficos de diagnóstico para o Modelo M_1 - Dados 3 (1975–2018).

De acordo com os critérios de informação, o Modelo $M_{1.B}$ apresenta o menor valor para os critérios de informação AIC e BIC, sendo, por isso, o

Tabela 3.9: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Dados 3 (1975–2018).

Modelo	AIC	BIC	d_{obs}	valor- p
M_0	117,178	120,746	0,000	–
$M_{0.A}$	119,147	124,499	0,031	0,860
$M_{0.B}$	120,156	125,509	-0,978	1,000
$M_{0.C}$	122,859	129,996	-1,681	1,000
M_1	119,178	124,530	–	–
$M_{1.A}$	121,147	128,283	0,031	0,860
$M_{1.B}$	122,156	129,293	-0,978	1,000
$M_{1.C}$	124,859	133,780	-1,681	1,000

modelo preferível segundo estes critérios quando o Modelo M_1 é o modelo base. Notamos também que no teste do Modelo M_1 *versus* o Modelo $M_{1.B}$, o Modelo M_1 é rejeitado na aplicação do teste de verosimilhanças (valor- $p \approx 0.047$). Quando é considerado como modelo base o Modelo M_0 , não se observa melhoria significativa do ajuste ao considerar modelos não estacionários com base nos critérios de informação AIC e BIC nem o Modelo M_0 é rejeitado face aos correspondentes modelos não estacionários, na aplicação do teste de razão de verosimilhanças, uma vez que todos os valores- p observados são maiores do que 0.05.

3.3.4 Análise de Períodos Temporais II

Dados 4 (1865-1950)

Considerando os dados obtidos na estação do Funchal quando estava localizada na fortaleza de São Lourenço, o ajuste do Modelo M_1 conduziu às seguintes estimativas dos parâmetros de localização, escala e forma $\hat{\mu} \approx 24,778$, $\hat{\sigma} \approx 0,590$ e $\hat{\xi} \approx -0,070$, respetivamente. Por sua vez, o Modelo M_0 apresentou estimativas $\hat{\mu} \approx 24,756$ e $\hat{\sigma} \approx 0,578$, evidenciando valores muito próximos para os parâmetros de localização e escala entre os dois modelos.

Relativamente ao parâmetro de forma do Modelos M_1 foi obtido o intervalo a 95% de confiança $[-0,219, 0,077]$, indicando não existir evidência estatística suficiente para o Modelo M_0 ($\xi = 0$) ser rejeitado, uma vez que $0 \in [-0,219, 0,077]$.

Os gráficos de diagnóstico associados aos ajustamentos dos Modelos M_0 e M_1 são apresentados nas Figuras 3.9 e 3.10, respetivamente, permitindo

avaliar visualmente a adequação dos modelos aos dados observados.

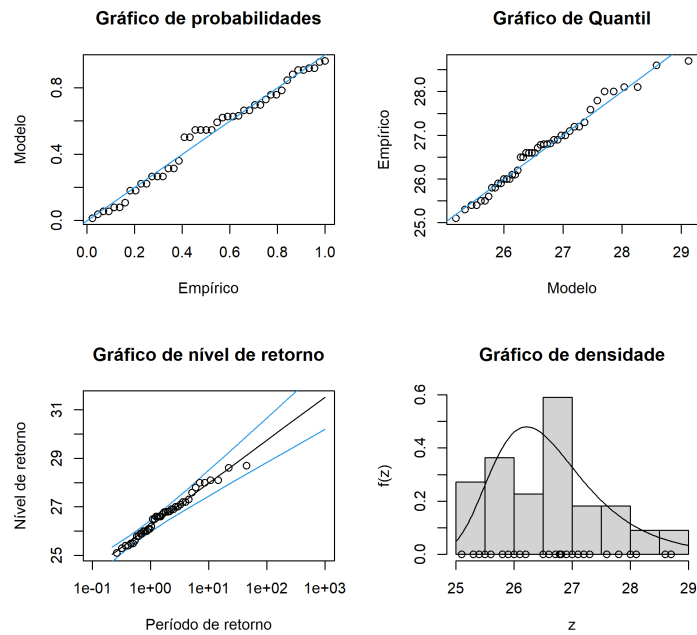


Figura 3.9: Gráficos de diagnóstico para o Modelo M_0 - Dados 4 (1865-1950).

Os níveis de retorno estimados indicam $z_{0,1} \approx 27,49^\circ\text{C}$ para 10 anos e $z_{0,01} \approx 29,04^\circ\text{C}$ para 100 anos. Se for o Modelo M_0 o modelo considerado, os níveis de retorno calculados para 10 e 100 anos são aproximadamente iguais a $26,10^\circ\text{C}$ e $27,49^\circ\text{C}$, respectivamente.

Aqui também foi realizada uma análise entre os Modelos M_0 e M_1 e os modelos com dependência temporal, considerando o *ano* como covariável explicativa na localização e/ou na escala. A Tabela 3.10 apresenta os valores dos critérios de informação AIC e BIC, bem como a estatística d_{obs} e o respectivo valor- p do teste da razão de verossimilhanças, incluindo igualmente os resultados obtidos para os modelos M_0 . Considerando o Modelo M_0 como modelo base, os critérios de informação AIC e BIC sugerem alguma melhoria do ajuste quando se introduz não estacionaridade em certos parâmetros. Em particular, os Modelos $M_{0,B}$ e $M_{0,C}$ apresentam valores de AIC inferiores ao do modelo estacionário M_0 , indicando um melhor compromisso entre qualidade de ajuste e complexidade. Entre estes, o Modelo $M_{0,C}$ apresenta o menor valor de AIC dentro desta classe enquanto que o Modelo $M_{0,B}$ apresenta o menor valor de BIC.

Adicionalmente, os resultados do teste baseado em d_{obs} indicam evidência

42CAPÍTULO 3. MÁXIMOS ANUAIS DAS MMMD DE TEMPERATURA

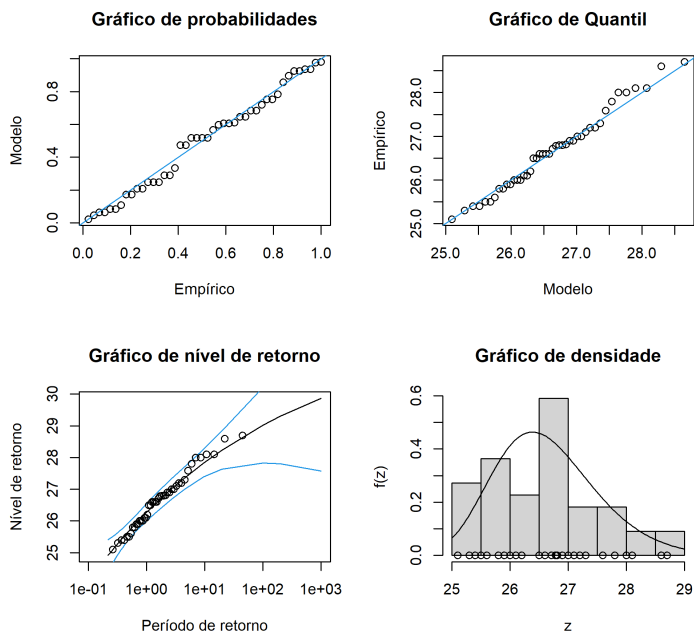


Figura 3.10: Gráficos de diagnóstico para o Modelo M_1 - Dados 4 (1865-1950).

Tabela 3.10: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 4 (1865-1950).

Modelo	AIC	BIC	d_{obs}	valor- p
M_0	194,083	199,169	0,000	–
$M_{0.A}$	200,173	207,803	-4,090	1,000
$M_{0.B}$	191,867	199,497	4,215	0,040
$M_{0.C}$	191,590	201,763	6,493	0,039
M_1	196,083	203,712	–	–
$M_{1.A}$	202,173	212,346	-4,090	1,000
$M_{1.B}$	193,867	204,040	4,215	0,040
$M_{1.C}$	193,590	206,306	6,493	0,039

estatisticamente significativa para os Modelos $M_{0.B}$ e $M_{0.C}$ (valor- $p \approx 0,04$), sugerindo que a introdução de dependência temporal, em particular no parâmetro de escala, pode conduzir a uma melhoria do ajuste relativamente ao modelo estacionário. Se for considerado o Modelo M_1 como modelo base, os resultados evidenciam um comportamento análogo.

Dados 5 (1951-1996)

Considerando os dados obtidos na estação do Funchal quando estava localizada nos Louros, o ajuste do Modelo M_1 conduziu às seguintes estimativas dos parâmetros de localização, escala e forma iguais a $\hat{\mu} \approx 25,545$, $\hat{\sigma} \approx 0,647$ e $\hat{\xi} \approx 0,050$, respetivamente. Por sua vez, o Modelo M_0 apresentou estimativas $\hat{\mu} \approx 25,562$ e $\hat{\sigma} \approx 0,657$, revelando valores muito próximos para os parâmetros de localização e de dispersão entre ambos os modelos.

O intervalo a 95% de confiança para o parâmetro de forma obtido foi o seguinte $[-0,135, 0,236]$, indicando não existir evidência estatística suficiente para o Modelo M_0 ($\xi = 0$) ser rejeitado.

Os gráficos de diagnóstico para os Modelos M_0 e M_1 permitem avaliar visualmente a adequação dos ajustamentos aos dados observados (Figuras 3.11 e 3.12, respetivamente).

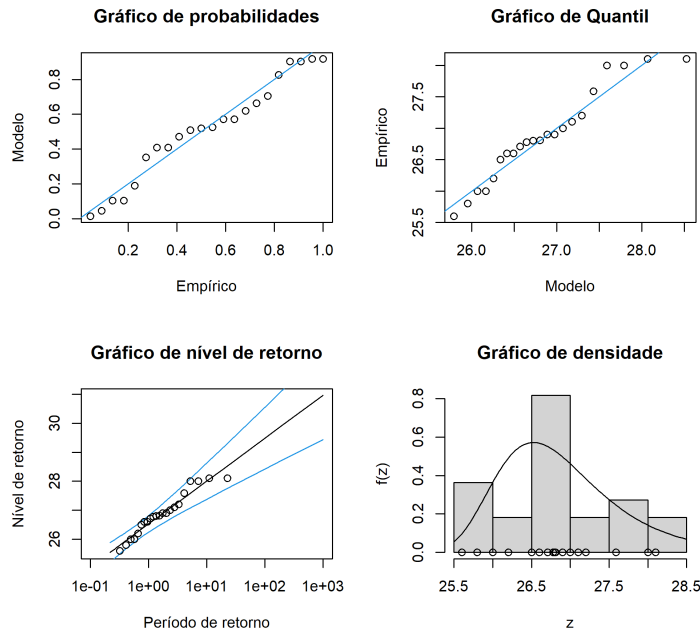


Figura 3.11: Gráficos de diagnóstico para o Modelo M_0 - Dados 5 (1951-1996).

Os níveis de retorno estimados são: para um período médio de 10 anos ($z_{0,1}$) a estimativa pontual é $27,09^\circ\text{C}$ e para um período de 100 anos ($z_{0,01}$)

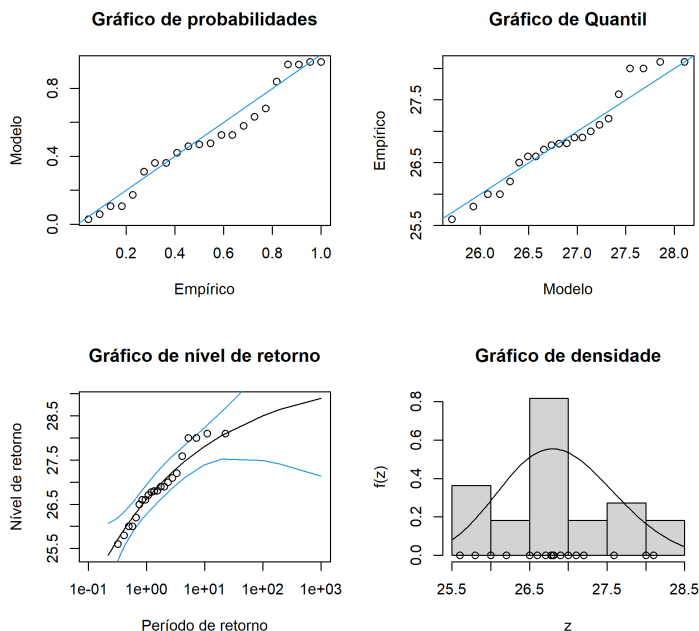


Figura 3.12: Gráficos de diagnóstico para o Modelo M_1 - Dados 5 (1951-1996).

o valor obtido foi de aproximadamente $28,90^\circ\text{C}$. Se for o Modelo M_0 o modelo considerado, os níveis de retorno calculados para 10 e 100 anos são aproximadamente iguais a $27,00^\circ\text{C}$ e $28,52^\circ\text{C}$, respectivamente.

Adicionalmente, foram ajustados modelos com dependência temporal nos parâmetros de localização (Modelo $M_{1,A}$), de escala (Modelo $M_{1,B}$) e em ambos os parâmetros (Modelo $M_{1,C}$), considerando o *ano* como covariável explicativa. A Tabela 3.11 apresenta os correspondentes valores dos critérios de informação AIC e BIC, bem como o valor observado para a estatística d_{obs} e o respectivo valor- p do teste da razão de verossimilhanças, incluindo igualmente os valores correspondentes para os modelos relativos ao Modelo M_0 . Considerando o Modelo M_0 como modelo base, observamos que o Modelo $M_{0,B}$ apresenta o menor valor de AIC, sendo também acompanhado por um valor da estatística d_{obs} estatisticamente significativo (valor- $p \approx 0,047$), o que sugere que a introdução de dependência temporal no parâmetro de escala pode melhorar o ajustamento relativamente ao modelo estacionário M_0 .

O modelo $M_{0,C}$ apresenta igualmente um valor de AIC inferior ao do Modelo M_0 , embora este modelo não seja rejeitado no teste de razão de

Tabela 3.11: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 5 (1951–1996).

Modelo	AIC	BIC	d_{obs}	valor- p
M_0	116,619	120,362	0,000	–
$M_{0.A}$	118,616	124,229	0,004	0,952
$M_{0.B}$	114,657	120,271	3,962	0,047
$M_{0.C}$	115,164	122,649	5,455	0,065
M_1	118,619	124,233	–	–
$M_{1.A}$	120,616	128,101	0,004	0,952
$M_{1.B}$	116,657	124,142	3,962	0,047
$M_{1.C}$	117,164	126,520	5,455	0,065

versomilhanças uma vez que o correspondente valor- $p \approx 0,065$ é maior do que 0.05. Tal como no período anterior, se for considerado o Modelo M_1 como modelo base, os resultados evidenciam um comportamento análogo.

Dados 6 (1997–2018)

Considerando o período no qual a estação do Funchal funcionou em modo automático, o ajuste do Modelo M_1 conduziu a estimativas dos parâmetros de localização, escala e forma aproximadamente iguais a 26,612, 0,685 e $-0,244$, respetivamente. Por sua vez, o Modelo M_0 apresentou estimativas $\hat{\mu} \approx 26,527$ e $\hat{\sigma} \approx 0,643$, evidenciando valores próximos para os parâmetros de localização e de dispersão entre os dois modelos.

Relativamente ao parâmetro de forma do Modelo M_1 foi obtido o intervalo de confiança a 95% igual a $[-0,6421, 0,153]$. Uma vez que o intervalo inclui o valor zero, não se observa evidência estatisticamente significativa de uma cauda distinta da distribuição de Gumbel, o que reforça a adequação do Modelo M_0 para este conjunto de dados.

Os níveis de retorno a 10 e 100 anos foram estimados em $z_{0,1} = 27,80^\circ\text{C}$ e $z_{0,01} = 28,51^\circ\text{C}$, respetivamente, quando considerado o Modelo M_1 . Se for o Modelo M_0 o modelo considerado, os níveis de retorno calculados para 10 e 100 anos são aproximadamente iguais a $28,15^\circ\text{C}$ e $29,76^\circ\text{C}$, respetivamente.

Os gráficos de diagnóstico para os Modelos M_0 e M_1 presentes nas Figuras 3.13 e 3.14, respetivamente, permitem avaliar a adequação dos modelos analisados.

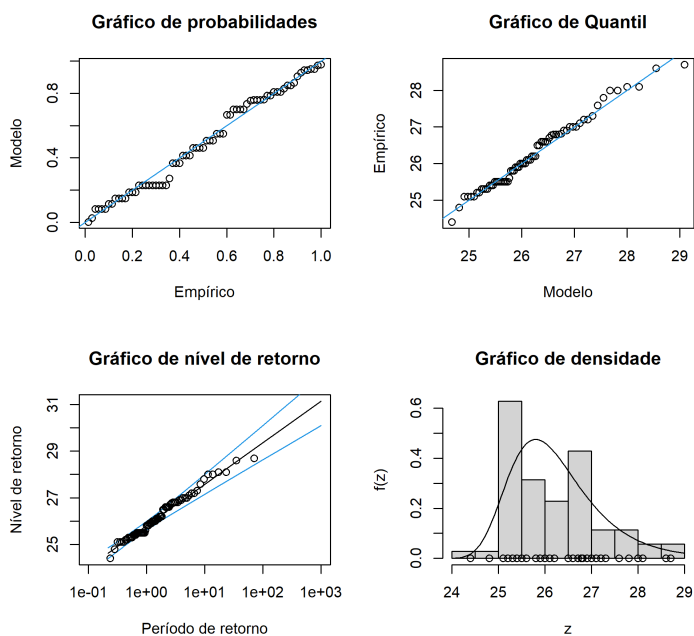


Figura 3.13: Gráficos de diagnóstico para o Modelo M_0 - Dados 6 (1997–2018).

Aqui também foi realizada uma análise entre os Modelos M_0 e M_1 e os modelos com dependência temporal, considerando o *ano* como covariável explicativa na localização e/ou na escala. A Tabela 3.12 apresenta os valores dos critérios de informação AIC e BIC, bem como a estatística d_{obs} e o respectivo valor- p do teste da razão de verossimilhanças, incluindo igualmente os resultados obtidos para os modelos M_0 .

Considerando o Modelo M_0 como modelo base, observa-se que é este o modelo que apresenta os menores valores de AIC e BIC, indicando o melhor compromisso entre qualidade de ajustamento e complexidade. Os Modelos $M_{0,A}$, $M_{0,B}$ e $M_{0,C}$ apresentam valores superiores destes critérios, não se verificando melhorias relevantes com a introdução de dependência temporal nos parâmetros. Adicionalmente, os valores da estatística d_{obs} e os respectivos valores- p não evidenciam ganhos estatisticamente significativos relativamente ao modelo estacionário. Se for considerado o Modelo M_1 como modelo base, os critérios de informação AIC e BIC também não indicam melhorias claras face ao Modelo M_1 , uma vez que todos os modelos não estacionários apresentam valores superiores destes critérios e os valores- p observados são maiores do que 0.05.

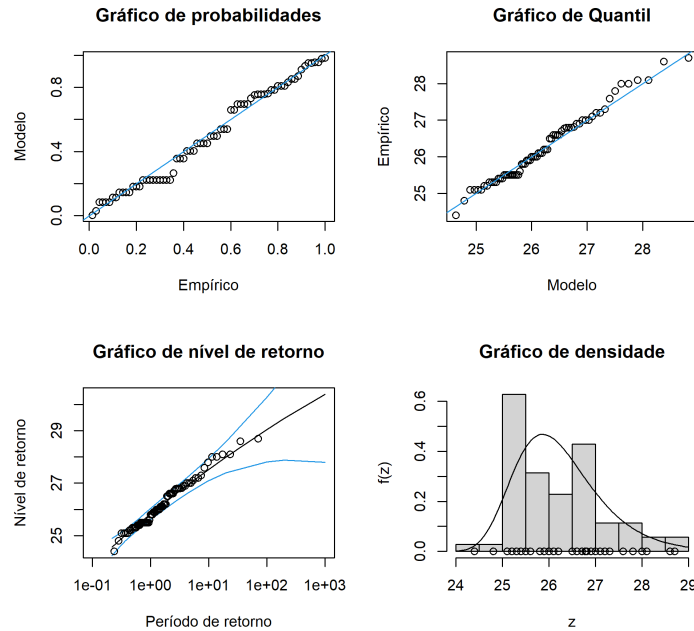


Figura 3.14: Gráficos de diagnóstico para o Modelo M_1 -Dados 6 (1997–2018).
 Tabela 3.12: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 6 (1997–2018).

Modelo	AIC	BIC	d_{obs}	valor- p
M_0	51,136	53,318	0,000	–
$M_{0,A}$	55,742	59,015	-2,605	1,000
$M_{0,B}$	53,338	56,611	-0,202	1,000
$M_{0,C}$	53,486	57,850	1,650	0,438
M_1	53,136	56,409	–	–
$M_{1,A}$	57,742	62,106	-2,605	1,000
$M_{1,B}$	55,338	59,702	-0,202	1,000
$M_{1,C}$	55,486	60,941	1,650	0,438

3.4 Considerações finais

Numa perspectiva de síntese e comparação global dos resultados obtidos para os diferentes intervalos temporais considerados, é apresentada na Tabela 3.13 a escolha do modelo selecionado em cada caso, com base nos critérios de informação e nos testes de razão de verossimilhança.

Ainda assim, é prática comum na literatura de valores extremos

Tabela 3.13: Escolha de modelos M_0 , $M_{0.A}$, $M_{0.B}$ e $M_{0.C}$.

	Modelo selecionado
Dados (1865-2018)	$M_{0.C}$
Dados 1 (1865-1896)	M_0
Dados 2 (1897-1974)	M_0
Dados 3 (1975-2018)	M_0
Dados 4 (1865-1950)	$M_{0.B}$ (BIC) $M_{0.C}$ (AIC)
Dados 5 (1951-1996)	$M_{0.B}$
Dados 6 (1997-2018)	M_0

considerar o Modelo M_1 mesmo quando a hipótese $\xi = 0$ não é rejeitada, uma vez que este modelo fornece um enquadramento mais geral e flexível, permitindo acomodar eventuais desvios do Modelo M_0 e facilitando comparações com estudos realizados noutras regiões ou contextos climáticos. Logo, considerando o Modelo M_1 , é apresentada na Tabela 3.14 a escolha do modelo selecionado em cada caso, com base nos critérios de informação e nos testes de razão de verossimilhança.

Tabela 3.14: Escolha de modelos M_1 , $M_{1.A}$, $M_{1.B}$ e $M_{1.C}$.

	Modelo selecionado
Dados (1865-2018)	$M_{1.C}$
Dados 1 (1865-1896)	M_1
Dados 2 (1897-1974)	M_1
Dados 3 (1975-2018)	M_1
Dados 4 (1865-1950)	$M_{1.B}$ (BIC) $M_{1.C}$ (AIC)
Dados 5 (1951-1996)	$M_{1.B}$
Dados 6 (1997-2018)	M_1

Para Dados (1865–2018), a seleção dos modelos $M_{0.C}$ e $M_{1.C}$ indica que a inclusão simultânea de tendência linear nos parâmetros de localização e de escala conduz a um ganho de ajustamento relativamente às parametrizações estacionárias. Em particular, os modelos $GEV(\mu(t), \sigma(t), 0)$ e $GEV(\mu(t), \sigma(t), \xi)$, com $\mu(t) = \beta_0 + \beta_1 t$ e $\sigma(t) = \exp(\beta_2 + \beta_3 t)$, acomodam de forma mais adequada a evolução temporal da série, evidenciando não estacionaridade persistente em ambos os parâmetros.

Os subconjuntos Dados 1 (1865-1896), Dados 2 (1897-1974) e Dados 3 (1975-2018) correspondem aos intervalos delimitados pelos pontos de mudança estimados através da função *change.point* do pacote `RecordTest`. Em todos estes casos, os critérios de seleção favorecem os modelos estacionários M_0 e M_1 , consoante se considere nulo ou não nulo o parâmetro de forma, não se justificando a introdução de dependência temporal em μ ou σ .

Os subconjuntos Dados 4 (1865-1950), Dados 5 (1951-1996) e Dados 6 (1997-2018) correspondem a períodos definidos com base em alterações na localização e no sistema de observação da estação meteorológica. Para os Dados 4 (1865-1950), observa-se divergência entre os critérios de informação: o AIC favorece os modelos com tendência simultânea em μ e σ , enquanto o BIC seleciona modelos com dependência temporal apenas em σ , traduzindo a penalização acrescida da complexidade paramétrica imposta por este último critério. Com base no estudo de *Kim et al.* [8], a escolha deverá ser baseada no critério AIC e, portanto, serão os Modelos $M_{0.C}$ ou $M_{1.C}$ os modelos selecionados. Nos Dados 5 (1951-1996), tanto considerando $\xi = 0$ como $\xi \neq 0$, os modelos selecionados são $M_{0.B}$ e $M_{1.B}$, respetivamente, indicando evidência consistente de dependência temporal exclusivamente no parâmetro de escala. Para os Dados 6 (1997-2018), a seleção dos modelos estacionários M_0 e M_1 não revela dependência temporal estatisticamente relevante.

Capítulo 4

Mínimos Anuais das MMTD de Precipitação

A análise de extremos climáticos, em particular dos mínimos anuais das MMTD de precipitação, é fundamental para compreender a ocorrência de secas severas e eventos hidrológicos críticos, que impactam significativamente setores como a agricultura, a gestão hídrica e o planeamento ambiental.

A base de dados original do IPMA, utilizada neste capítulo, fornece os totais mensais de precipitação, calculados a partir de três medições diárias realizadas ao longo de cada dia. Para cada mês, é calculada a soma total da precipitação registada nos dias correspondentes, resultando numa série mensal de precipitação acumulada disponível em formato CSV. A partir da série mensal, procedeu-se à extração dos mínimos anuais das MMTD de precipitação: para cada ano, selecionou-se o valor mensal mais baixo, correspondente ao mês com menor precipitação acumulada.

O estudo dos mínimos anuais das MMTD de precipitação foi realizado com base num *script*, cujo conteúdo integral se encontra no Anexo E. Neste *script* está o pré-processamento dos dados, incluindo a construção da variável `precip_inv`, correspondente a multiplicação dos valores em estudo por menos um para permitir modelar mínimos como máximos até à definição dos grupos de análise, resultantes da aplicação da função `change.point` do pacote `RecordTest`, bem como dos períodos associados a alterações geográficas e tecnológicas da estação. A divisão dos dados em cinco subconjuntos: dois resultantes da deteção de pontos de mudança, designados por Períodos Temporais I, e três associados a diferentes sistemas de observação, designados por Períodos Temporais II, e, conseqüente, obtenção de algumas medidas de localização e dispersão estão apresentadas

na Secção 4.1. No *script* estão também os procedimentos realizados para análise dos pressupostos de independência e estacionariedade apresentada na Secção 4.2, tal como o ajustamento dos Modelos M_0 e M_1 , a determinação de níveis de retorno, a seleção e validação de modelos que constituem a análise apresentada na Secção 4.3. O *script* inclui ainda funções auxiliares que garantem a consistência da análise, permitindo aplicar de forma sistemática todos os procedimentos à série completa e a cada subconjunto, bem como organizar automaticamente os resultados, os quais são posteriormente utilizados na construção das tabelas, gráficos e conclusões apresentados ao longo deste capítulo. Uma comparação global dos resultados obtidos para os diferentes intervalos temporais considerados é realizada na Secção 4.4.

4.1 Estatística Descritiva

Na análise, foi detetado um pontos de mudança com a aplicação da função `change.point` do pacote `RecordTest` do R, os quais resultaram nos seguintes conjuntos de dados: Dados 1 (1865-1887) e Dados 2 (1888-2018). Tendo em conta as alterações geográficas e tecnológicas na estação do Funchal, os dados foram também divididos segundo os diferentes sistemas de observação, refletindo o impacto da mudança de local e do tipo de equipamento nas medições. Esta abordagem resultou no estudo dos mínimos anuais das MMTD de precipitação nos períodos de 1865 e 1950 (Dados 3 (1865-1950)), de 1951 e 1996 (Dados 4 (1951-1996)) e de 1996 e 2018 (Dados 5 (1997-2018)).

As Tabelas 4.1, 4.2 e 4.3 apresentam valores de medidas de localização e dispersão dos mínimos anuais das MMTD de precipitação para a série completa (Dados (1865-2018)) e para todos os cinco subconjuntos.

A Tabela 4.1 apresenta as medidas de dispersão dos mínimos anuais das MMTD de precipitação. Relativamente ao conjunto global dos Dados (1865-2018), observa-se que os mínimos anuais das MMTD de precipitação apresentam uma média de 619,84 mm, um desvio-padrão de 214,33 mm e um coeficiente de variação de 34,58%, o que indica uma variabilidade considerável em torno do valor médio. Nos Períodos Temporais I, verifica-se que Dados 1 (1865-1887) apresentam uma média superior (697,60 mm) face a Dados 2 (1888-2018), cuja média é de 607,94 mm. O desvio-padrão é também ligeiramente superior no primeiro período (232,19 mm contra 209,73 mm), enquanto os coeficientes de variação são semelhantes (33,77% e 34,50%), sugerindo que, apesar da diminuição dos valores médios, a variabilidade

relativa se mantém praticamente constante. Nos Períodos Temporais II, observa-se uma diminuição progressiva da média ao longo dos subperíodos: Dados 3 (1865-1950) registam 644,05 mm, Dados 4 (1951-1996) 591,22 mm e Dados 5 (1997-2018) 585,02 mm. O coeficiente de variação é menor em Dados 4 (1951-1996) (30,58%), indicando maior estabilidade relativa nesse intervalo, enquanto Dados 5 (1997-2018) apresentam o valor mais elevado (36,25%), refletindo maior dispersão relativa recente.

Tabela 4.1: Dimensão, média, desvio padrão e coeficiente de variação - Mínimos Anuais das MMTD de Precipitação.

	n	Média	Desvio-padrão	CV (%)
Dados (1865-2018)	154	619,84	214,33	34,58
Dados 1 (1865-1887)	23	697,60	232,19	33,77
Dados 2 (1888-2018)	131	607,94	209,73	34,50
Dados 3 (1865-1950)	86	644,05	230,05	35,72
Dados 4 (1951-1996)	46	591,22	180,79	30,58
Dados 5 (1997-2018)	22	585,02	212,06	36,25

Na Tabela 4.2 estão apresentados os valores da mediana e dos quartis das amostras analisadas. Nos Dados (1865-2018), a mediana dos mínimos anuais das MMTD de precipitação é de 603,15 mm, com primeiro quartil de 463,90 mm e terceiro quartil de 725,25 mm, indicando que 50% dos valores se concentram nesse intervalo. Nos Períodos Temporais I, Dados 1 (1865-1887) apresentam uma mediana de 712,30 mm, claramente superior à observada em Dados 2 (1888-2018), onde a mediana é de 588,30 mm. Também os quartis mostram valores mais elevados no primeiro período, evidenciando que os mínimos anuais eram, em geral, superiores no intervalo inicial. Nos Períodos Temporais II, Dados 3 (1865-1950) apresentam a mediana mais elevada (671,70 mm), seguida de Dados 4 (1951-1996) com 562,75 mm e Dados 5 (1997-2018) com 537,75 mm. Esta evolução sugere uma redução gradual dos valores centrais da distribuição ao longo do tempo. O intervalo interquartil é mais reduzido em Dados 5 (1997-2018), indicando menor dispersão dos valores centrais nesse período recente.

A Tabela 4.3 apresenta os valores extremos e a amplitude amostral das amostras estudadas. Para os Dados (1865-2018), os mínimos anuais variam entre 200,4 mm e 1418,7 mm, correspondendo a uma amplitude amostral de 1218,3 mm, o que evidencia elevada variabilidade entre anos extremos. Nos Períodos Temporais I, Dados 1 (1865-1887) apresentam um mínimo de 332,0 mm e máximo de 1206,7 mm, enquanto Dados 2 (1888-2018)

Tabela 4.2: Mediana (M), primeiro e terceiro quartis - Mínimos Anuais das MMTD de Precipitação.

	M	Q_1	Q_3
Dados (1865-2018)	603,15	463,90	725,25
Dados 1 (1865-1887)	712,30	484,35	841,35
Dados 2 (1888-2018)	588,30	464,20	715,10
Dados 3 (1865-1950)	671,70	456,4	770,48
Dados 4 (1951-1996)	562,75	425,10	721,05
Dados 5 (1997-2018)	537,75	475,90	651,45

registam os valores extremos globais (200,4 mm e 1418,7 mm). Isto mostra que os eventos mais extremos ocorreram sobretudo no segundo subperíodo. Nos Períodos Temporais II, Dados 3 (1865-1950) mantêm os extremos globais e a amplitude de 1218,3 mm. Já Dados 4 (1951-1996) apresentam menor amplitude (695,4 mm), com valores entre 310,8 mm e 1006,2 mm, sugerindo menor variabilidade extrema. Em Dados 5 (1997-2018) observa-se novamente um aumento da amplitude (1103,7 mm), devido ao máximo de 1403,2 mm, refletindo o reaparecimento de valores extremos recentes.

Tabela 4.3: Mínimo, máximo e amplitude amostral (R_n) - Mínimos Anuais das MMTD de Precipitação.

	Mínimo	Máximo	R_n
Dados (1865-2018)	200,4	1418,7	1218,3
Dados 1 (1865-1887)	332,0	1206,7	1218,3
Dados 2 (1888-2018)	200,4	1418,7	1218,3
Dados 3 (1865-1950)	200,4	1418,7	1218,3
Dados 4 (1951-1996)	310,8	1006,2	695,4
Dados 5 (1997-2018)	299,5	1403,2	1103,7

4.2 Estacionaridade e Independência

Na Tabela 4.4 estão apresentados os valores- p obtidos na aplicação dos testes não paramétricos de Mann-Kendall e de Wald-Wolfowitz para cada um dos conjuntos de dados descritos na Secção 4.1, aplicados com o objetivo de avaliar os pressupostos de estacionaridade e a independência, respetivamente.

Relativamente aos Dados (1865–2018), o teste de Mann–Kendall indica evidência de tendência estatisticamente significativa ($p \approx 0,021$), enquanto o teste de Wald–Wolfowitz não rejeita a hipótese de independência ($p \approx 0,518$). Nos Períodos Temporais I, observa-se que Dados 1 (1865–1887) apresentam igualmente tendência significativa ($p \approx 0,020$), enquanto em Dados 2 (1888–2018) o valor- p do teste de Mann–Kendall (0,087) não evidencia significância ao nível de 5%. Em ambos os subconjuntos, não há evidência estatística para rejeitar a aleatoriedade das séries em análise de acordo com o teste de Wald–Wolfowitz, sugerindo assim a existência de independência. Nos Períodos Temporais II, Dados 3 (1865–1950) mantêm evidência de tendência significativa ($p \approx 0,020$), ao contrário de Dados 4 (1951–1996) e Dados 5 (1997–2018), onde não se observam tendências estatisticamente significativas. Em todos estes subperíodos, os valores- p do teste de Wald–Wolfowitz são maiores do que 0.05, não existindo portanto evidência estatística para rejeitar a aleatoriedade das séries em estudo.

Tabela 4.4: Resultados dos testes de tendência e independência para os subconjuntos de dados.

	Teste de Mann–Kendall (valor-p)	Teste de Wald–Wolfowitz (valor-p)
Dados (1865-2018)	0,021	0,518
Dados 1 (1865-1887)	0,020	0,193
Dados 2 (1888-2018)	0,087	0,159
Dados 3 (1865-1950)	0,020	> 0,656
Dados 4 (1951-1996)	0,121	0,233
Dados 5 (1997-2018)	0,310	0,662

4.3 Seleção e Validação de Modelos

4.3.1 Modelo M_0 versus Modelo M_1

A comparação entre o Modelo M_0 e o Modelo M_1 foi realizada utilizando o teste de razão de verossimilhanças, cujos resultados se apresentam na Tabela 4.5. Para os Dados (1865–2018), o valor observado $d_{obs} = 80,072$ acompanhado de um valor- p inferior a 0,001 indica que o modelo M_1 fornece um ajuste significativamente melhor do que M_0 . Esta tendência mantém-se ao longo dos subperíodos analisados, com todos os valores- p bastante baixos e inferiores a 0.05, demonstrando de forma consistente a superioridade do

Modelo M_1 em capturar a variabilidade da precipitação.

A análise dos critérios de informação AIC e BIC, apresentados na Tabela 4.6, confirma esta conclusão. Em todos os períodos considerados, o Modelo M_1 apresenta valores de AIC e BIC inferiores aos do Modelo M_0 , o que evidencia um ajuste mais adequado e mais parcimonioso. Esta diferença é particularmente notória nos Dados (1865-2018) e nos subperíodos mais longos, refletindo a capacidade do Modelo M_1 em acomodar melhor a complexidade temporal dos dados.

Em síntese, a combinação do teste de razão de verossimilhanças com a análise dos critérios de informação indicam que o Modelo M_1 é estatisticamente preferível ao Modelo M_0 para a série de precipitação considerada.

Tabela 4.5: Valores de d_{obs} e valor- p para o teste de razão de verossimilhanças.

	d_{obs}	valor- p
Dados (1865-2018)	80,072	$< 0,001$
Dados 1 (1865-1887)	6,456	0,011
Dados 2 (1888-2018)	74,91	$< 0,001$
Dados 3 (1865-1950)	33,681	$6,49 \times 10^{-9}$
Dados 4 (1951-1996)	17,245	$3,29 \times 10^{-5}$
Dados 5 (1997-2018)	25,287	$4,94 \times 10^{-7}$

Tabela 4.6: Valores dos critérios de informação AIC e BIC.

	Modelo M_0		Modelo M_1	
	AIC	BIC	AIC	BIC
Dados (1865-2018)	2161,34	2167,42	2083,27	2092,38
Dados 1 (1865-1887)	323,15	325,42	318,69	322,10
Dados 2 (1888-2018)	1840,07	1845,82	1767,16	1775,78
Dados 3 (1865-1950)	1211,46	1216,37	1179,78	1187,14
Dados 4 (1951-1996)	622,39	626,04	607,14	612,63
Dados 5 (1997-2018)	319,02	321,20	295,73	299,00

4.3.2 Dados (1865 – 2018)

Foi ajustado o Modelo M_1 aos mínimos anuais das MMTD de precipitação. A estimação dos parâmetros do modelo resultou nos seguintes valores: $\hat{\mu} = 690,94$, $\hat{\sigma} = 221,02$ e $\hat{\xi} = -0,425$ (IC 95%: $[-0,493; -0,356]$), indicando uma cauda limitada superiormente, consistente com o comportamento físico expectável da precipitação mínima. Os níveis de retorno estimados refletem a redução mínima esperada de precipitação. Para um período médio de 10 anos, $z_{0,1} = 357,11$ mm e para 100 anos, $z_{0,01} = 222,52$ mm, evidenciando limites inferiores da precipitação.

A inspeção visual dos gráficos de diagnóstico (ver Figura 4.1) indicia a boa adequação do Modelo 1 aos dados observados.

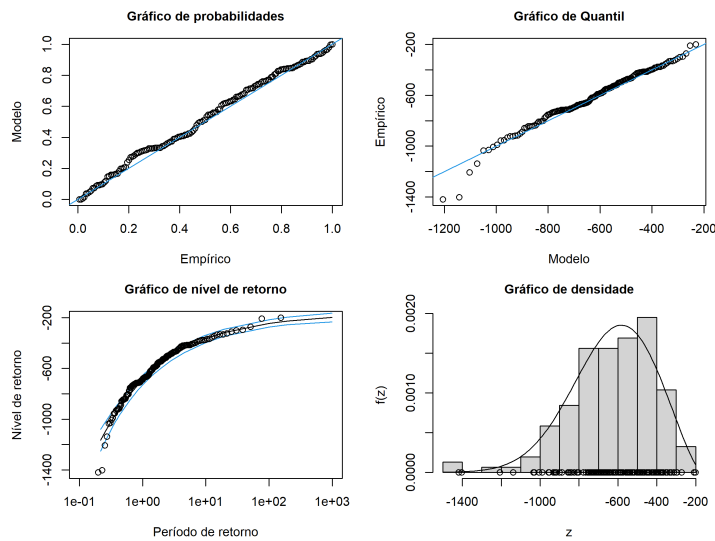


Figura 4.1: Gráficos de diagnóstico para o Modelo M_1 - Dados (1865–2018).

Seguidamente, procedemos à análise dos modelos apresentados na Tabela 2.2, que incorporam a variável temporal *ano* como covariável nos parâmetros de localização e dispersão. Os valores dos critérios de informação AIC e BIC, juntamente com o valor observado para a estatística d_{obs} e o respetivo valor- p do teste da razão de verosimilhança para cada teste entre o Modelo M_1 e os correspondentes modelos com a covariável *ano*, estão apresentados na Tabela 4.7. A análise dos resultados apresentados mostra que o modelo $M_{1,A}$ apresenta o menor valor de AIC (2082, 24), sugerindo um ligeiro melhor ajustamento relativamente ao modelo de referência M_1 . No entanto, o valor de BIC para $M_{1,A}$ (2094, 39) é superior ao de M_1 (2092, 38),

refletindo a penalização associada ao aumento do número de parâmetros. Relativamente ao teste da razão de verosimilhança, a comparação entre M_1 e $M_{1.A}$ conduz a $d_{obs} = 3,027$, com valor- $p \approx 0,0819$, não permitindo rejeitar a hipótese nula ao nível de significância de 5%, embora revele evidência marginal de melhoria do ajustamento.

Tabela 4.7: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Dados (1865–2018).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_1	2083,27	2092,38	–
$M_{1.A}$	2082,24	2094,39	3,027 ($\approx 0,0819$)
$M_{1.B}$	2088,38	2100,53	(≈ 1)
$M_{1.C}$	2083,51	2098,69	3,765 ($\approx 0,152$)

Para os modelos $M_{1.B}$ e $M_{1.C}$, os valores de AIC e BIC não evidenciam vantagens claras face ao modelo base, sendo que os respetivos valores- p indicam ausência de melhoria estatisticamente significativa. De forma geral, os resultados sugerem que a inclusão da covariável temporal *ano* não produz uma melhoria estatisticamente significativa no ajustamento do Modelo M_1 .

4.3.3 Análise de Períodos Temporais I

Na análise correspondente ao período de 1865 a 1887 (Dados 1 (1865-1887)) o Modelo M_1 apresentou os seguintes parâmetros estimados: $\hat{\mu} = 734,60$ mm, $\hat{\sigma} = 253,30$ mm e $\hat{\xi} = -0,574$ (IC 95%: [-0,983;-0,164]), indicando uma cauda limitada superiormente, consistente com o comportamento físico expectável da precipitação mínima. Os níveis de retorno estimados são elevados, refletindo a ocorrência de anos particularmente secos. Para um período médio de 10 anos, observamos $z_{0,1} \approx 414,45$ mm e para 100 anos, $z_{0,01} \approx 324,53$ mm. Considerando Dados 2 (1888-2018), o Modelo M_1 apresentou os seguintes parâmetros estimados: $\hat{\mu} = 677,997$ mm, $\hat{\sigma} = 216,738$ mm e $\hat{\xi} = -0,426$ (IC 95%: [-0,822;-0,380]), indicando uma cauda limitada superiormente. Os níveis de retorno são $z_{0,1} = 361,10$ mm e $z_{0,01} = 264,21$ mm para períodos médios de retorno de 10 e 100 anos, respetivamente.

A avaliação gráfica do ajustamento do Modelo M_1 é apresentada nas Figuras 4.2 e 4.3.

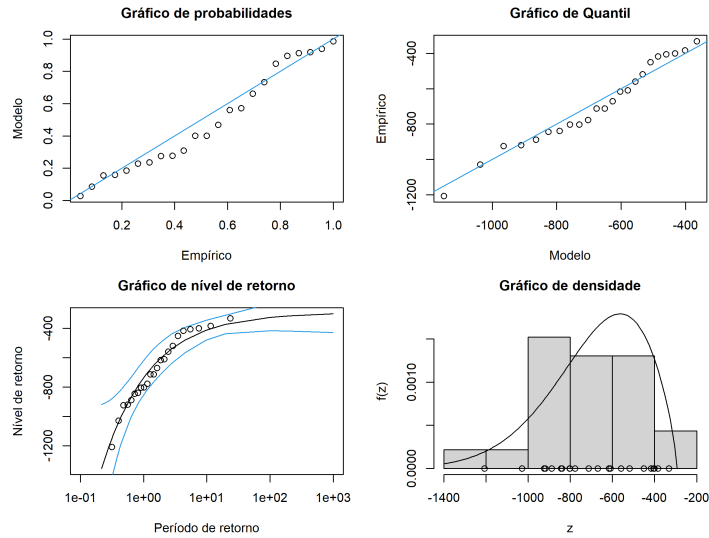


Figura 4.2: Gráficos de diagnóstico para o Modelo M_1 - Dados 1 (1865-1887).

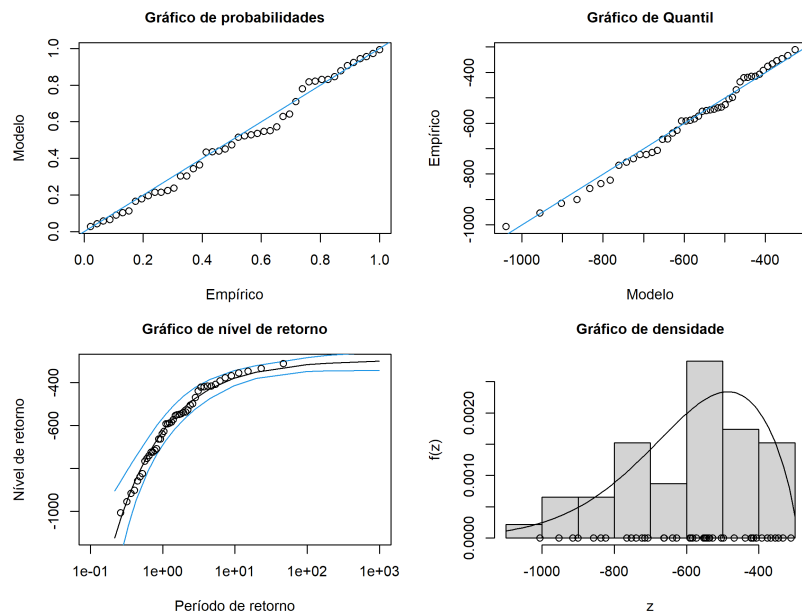


Figura 4.3: Gráficos de diagnóstico para o Modelo M_1 - Dados 2 (1888-2018).

Tal como na secção anterior, foi realizada uma análise dos modelos apresentados na Tabela 2.2, que incorporam a variável temporal *ano* como covariável nos parâmetros de localização e dispersão. Os valores dos critérios de informação AIC e BIC, juntamente com o valor observado para a estatística d_{obs} e o respetivo valor- p do teste da razão de verosimilhança para cada teste entre o Modelo M_1 e os correspondentes modelos com a covariável

ano, estão apresentados na Tabelas 4.8 e 4.9.

Tabela 4.8: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Dados 1 (1865–1887).

Modelo	AIC	BIC	d_{obs}(valor-p)
M_1	318,692	322,098	–
$M_{1.A}$	320,698	325,240	-0,007 (≈ 1)
$M_{1.B}$	321,150	325,692	-0,458 (≈ 1)
$M_{1.C}$	322,653	328,331	0,038 ($\approx 0,981$)

Tabela 4.9: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC-Dados 2 (1888–2018).

Modelo	AIC	BIC	d_{obs}(valor-p)
M_1	1767,157	1775,783	–
$M_{1.A}$	1767,835	1779,336	1,322 ($\approx 0,25$)
$M_{1.B}$	1770,980	1782,481	(≈ 1)
$M_{1.C}$	1769,506	1783,882	1,651 ($\approx 0,438$)

A análise dos resultados apresentados nas Tabelas 4.8 e 4.9 evidencia que, em ambos os períodos considerados, o modelo M_1 apresenta os menores valores dos critérios de informação AIC e BIC, sendo, por isso, o modelo preferido segundo estes critérios. No período 1865–1887, os modelos alternativos $M_{1.A}$, $M_{1.B}$ e $M_{1.C}$ registam valores de AIC e BIC superiores, indicando que o acréscimo de complexidade não conduz a uma melhoria relevante do ajustamento. Para o período 1888–2018, observa-se igualmente a predominância do modelo M_1 , embora o modelo $M_{1.A}$ apresente um valor de AIC próximo. Os valores da estatística d_{obs} e os respetivos valores- p , maiores do que 0.05 nos dois períodos, não fornecem evidência estatística suficiente para justificar a preferência pelos modelos alternativos. Os resultados mostram assim que o Modelo M_1 é o mais adequado seja qual for o período considerado de entre os Períodos Temporais I, não havendo evidência estatística para a inclusão de tendência nos parâmetros de localização e de escala do Modelo M_1 .

4.3.4 Análise de Períodos Temporais II

Na análise correspondente ao período de 1865 a 1950 (Dados 3 (1865-1950)) o Modelo M_1 apresentou os seguintes parâmetros estimados: $\hat{\mu} = 707,51$ mm, $\hat{\sigma} = 244,47$ mm e $\hat{\xi} = -0,436$ (IC 95%: $[-0,822; -0,380]$), indicando uma cauda limitada superiormente, consistente com o comportamento físico expectável da precipitação mínima. Para um período médio de 10 anos, observamos $z_{0,1} \approx 358,1$ mm e para 100 anos, $z_{0,01} \approx 222,7$ mm.

Considerando Dados 4 (1951-1996)), o Modelo M_1 apresentou os seguintes parâmetros estimados: $\hat{\mu} = 625,70$ mm, $\hat{\sigma} = 198,94$ mm e $\hat{\xi} = -0,601$ (IC 95%: $[-1,064; -0,238]$), indicando uma cauda limitada superiormente. Os níveis de retorno são $z_{0,1} = 380,1$ mm e $z_{0,01} = 315,5$ mm para períodos médios de retorno de 10 e 100 anos, respetivamente.

Na análise correspondente ao último período dos Períodos Temporais II, o ajuste do Modelo M_1 forneceu as seguintes estimativas para os parâmetros $\hat{\mu} = 680,03$ mm, $\hat{\sigma} = 233,70$ mm e $\hat{\xi} = -0,587$ (IC 95%: $[-1,002; -0,172]$). Os níveis de retorno são $z_{0,1} = 403,8$ mm e $z_{0,01} = 377,2$ mm para períodos médios de retorno de 10 e 100 anos, respetivamente.

A avaliação gráfica do ajustamento do Modelo M_1 está presente nas Figuras 4.4, 4.5 e 4.6.

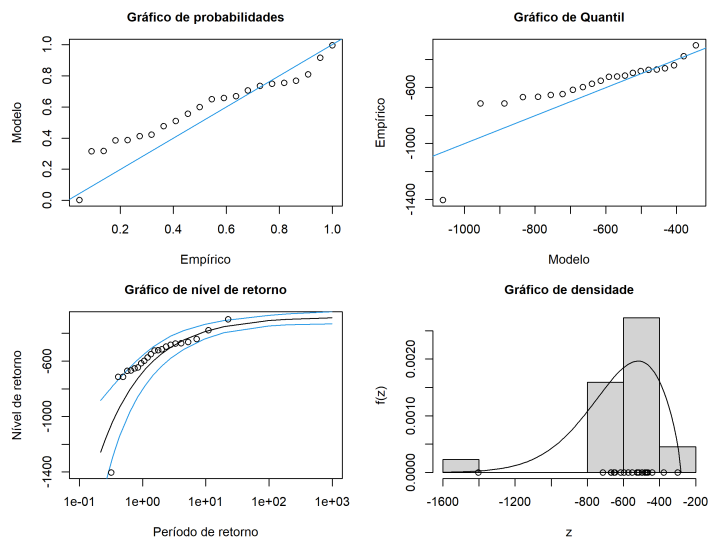


Figura 4.4: Gráficos de diagnóstico para o Modelo M_1 - Dados 3 (1865-1950).

Tal como na secção anterior, foi realizada uma análise dos modelos

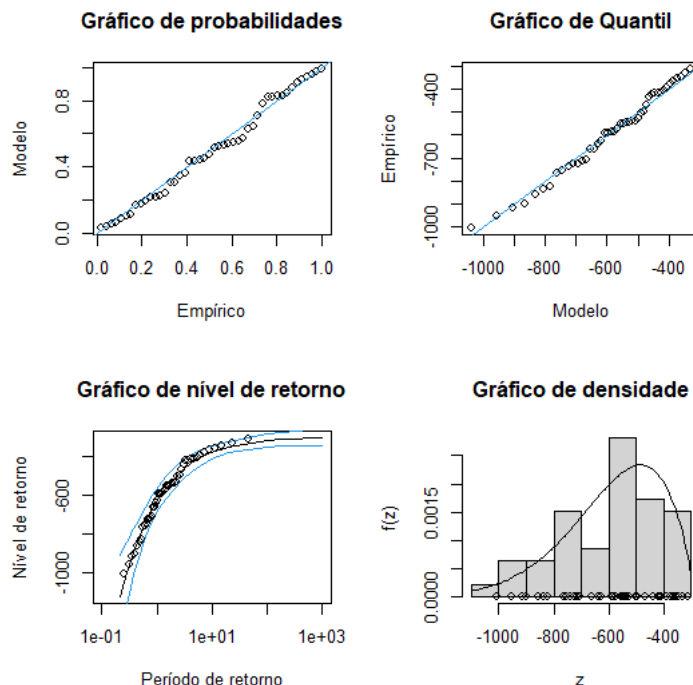


Figura 4.5: Gráficos de diagnóstico para o Modelo M_1 - Dados 4 (1951-1996).

apresentados na Tabela 2.2, que incorporam a variável temporal *ano* como covariável nos parâmetros de localização e dispersão. Os valores dos critérios de informação AIC e BIC, juntamente com o valor observado para a estatística d_{obs} e o respectivo valor- p do teste da razão de verossimilhança para cada teste entre o Modelo M_1 e os correspondentes modelos com a covariável *ano*, estão apresentados na Tabelas 4.10, 4.11 e 4.12.

Tabela 4.10: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC - Dados 3 (1865–1950).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_1	1179,780	1187,143	–
$M_{1.A}$ (μ)	1181,656	1191,473	0,124 ($\approx 0,725$)
$M_{1.B}$ (σ)	1181,192	1191,010	0,588 ($\approx 0,443$)
$M_{1.C}$ (μ, σ)	1185,081	1197,352	–1,301 (≈ 1)

Considerando conjuntamente os três períodos analisados — 1865–1950, 1951–1996 e 1997–2018 — verifica-se que o Modelo M_1 apresenta sistematicamente os menores valores dos critérios de informação AIC e

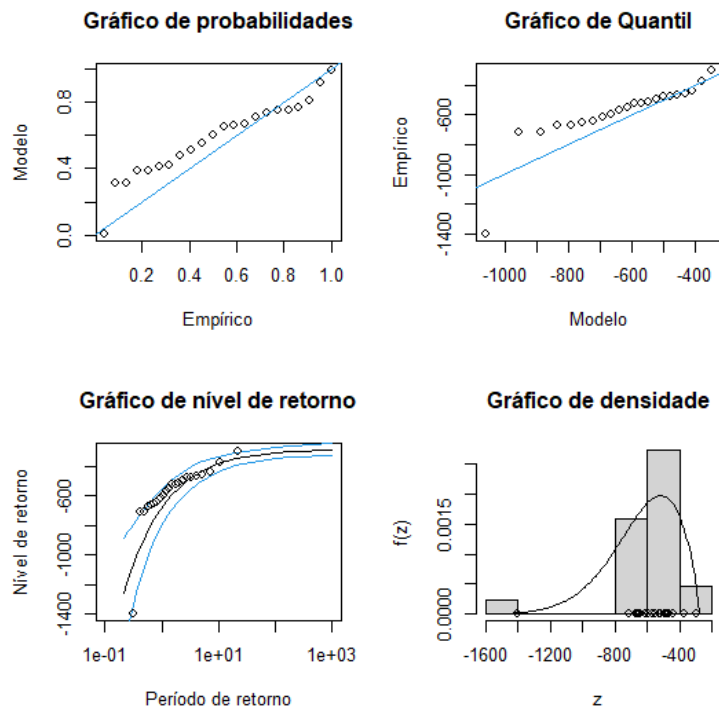


Figura 4.6: Gráficos de diagnóstico para o Modelo M_1 - Dados 5 (1997-2018).
 Tabela 4.11: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC - Dados 4 (1951–1996).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_1	607,141	612,627	–
$M_{1.A}$ (μ)	609,162	616,476	–0,021 (≈ 1)
$M_{1.B}$ (σ)	609,124	616,438	0,017 ($\approx 0,897$)
$M_{1.C}$ (μ, σ)	611,318	620,461	–0,177 (≈ 1)

Tabela 4.12: Valores de d_{obs} (valor- p) e dos critérios de informação AIC e BIC- Dados 5 (1997-2018).

Modelo	AIC	BIC	d_{obs} (valor- p)
M_1	295,730	299,003	–
$M_{1.A}$	296,260	300,624	1,470 ($\approx 0,225$)
$M_{1.B}$	296,371	300,735	1,359 ($\approx 0,244$)
$M_{1.C}$	298,274	303,730	1,456 ($\approx 0,483$)

BIC, evidenciando melhor equilíbrio entre qualidade de ajustamento e

parcimônia em comparação com os modelos não estacionários. Nos três intervalos temporais, os modelos alternativos $M_{1.A}$, $M_{1.B}$ e $M_{1.C}$, que incorporam dependência temporal nos parâmetros μ e/ou σ , não conduzem a melhorias estatisticamente significativas, uma vez que os valores de d_{obs} permanecem reduzidos e os respectivos valores- p são sempre superiores ao nível de significância de 0.05. Assim, os resultados obtidos indicam de forma consistente que, para os três subconjuntos de dados considerados, não há evidência estatística suficiente para rejeitar o Modelo M_1 em detrimento de algum dos modelos $M_{1.A}$, $M_{1.B}$ e $M_{1.C}$.

4.4 Considerações finais

Com o objetivo de explicitar o contributo individual de cada critério estatístico no processo de seleção e de evidenciar eventuais divergências entre abordagens, apresentamos de seguida uma segunda tabela de síntese, na qual se identifica, para cada intervalo temporal, o modelo selecionado segundo os critérios AIC e BIC e o teste do razão de verossimilhanças, considerados separadamente.

Tabela 4.13: Escolha de modelos M_1 , $M_{1.A}$, $M_{1.B}$ e $M_{1.C}$.

	Modelo selecionado
Dados (1865-2018)	$M_{1.A}$ (AIC) e M_1 (BIC e d_{obs})
Dados 1 (1865-1888)	M_1
Dados 2 (1888-2018)	M_1
Dados 4 (1865-1950)	M_1
Dados 5 (1951-1996)	M_1
Dados 6 (1997-2018)	M_1

A Tabela 4.13 mostra que, embora o critério AIC selecione um modelo com tendência no parâmetro de localização considerando os Dados (1865-2018), esta preferência não é confirmada nem pelo critério BIC nem pelo teste de razão de verossimilhanças, ambos favoráveis ao Modelo M_1 . Para todos os restantes subperíodos, observamos uma concordância total entre os três critérios, reforçando a escolha do Modelo M_1 para a descrição dos mínimos anuais das MMTD de precipitação.

Capítulo 5

Conclusão

A análise desenvolvida ao longo desta dissertação permitiu evidenciar a relevância da Teoria dos Valores Extremos como ferramenta de modelação estatística aplicada a fenómenos climáticos e ambientais.

Relativamente aos máximos anuais das MMMD de temperatura, os resultados mostram um contraste claro entre a série completa e os subconjuntos temporais analisados. Para o período total de 1865 a 2018, tanto no enquadramento com $\xi = 0$ como no enquadramento mais geral com $\xi \neq 0$, os modelos selecionados são $M_{0,C}$ e $M_{1,C}$, respetivamente. Isto sugere que, quando considerada a totalidade da série, existe evidência de não estacionaridade persistente, refletindo alterações graduais na estrutura dos máximos observados ao longo do tempo. Em contraste, nos subconjuntos definidos pelos pontos de mudança estatisticamente estimados (Dados 1 (1865-1896), Dados 2 (1897-1974) e Dados 3 (1975-2018)) a seleção recai sistematicamente sobre os modelos estacionários M_0 e M_1 , o que indica que, dentro de cada segmento homogêneo, a variabilidade temporal deixa de justificar parametrizações não estacionárias.

Nos períodos definidos segundo alterações históricas na localização e no sistema de observação da estação meteorológica, os resultados revelam um comportamento intermédio e mais heterogêneo. Para os Dados 4 (1865–1950), verifica-se divergência entre os critérios AIC e BIC: o critério AIC favorece modelos com tendência simultânea em μ e σ , enquanto o critério BIC privilegia modelos com dependência apenas em σ . Seguindo a orientação metodológica proposta por *Kim et al.* [8], a escolha tende a recair nos modelos $M_{0,C}$ ou $M_{1,C}$. Nos Dados 5 (1951–1996), ambos os enquadramentos apontam para $M_{0,B}$ e $M_{1,B}$, evidenciando dependência temporal exclusivamente no parâmetro de escala, enquanto nos Dados 6

(1997–2018) volta a observar-se preferência pelos modelos estacionários. Em síntese, a comparação global sugere que a não estacionaridade identificada na série longa resulta sobretudo da agregação de períodos com comportamentos distintos.

Relativamente aos mínimos anuais das MMTD de precipitação, os resultados confirmam uma predominância clara de modelos estacionários na descrição destes mínimos, verificando-se elevada concordância entre os diferentes critérios de seleção considerados. A única divergência surge na série completa (1865–2018), em que o critério AIC seleciona o modelo $M_{1,A}$, sugerindo dependência temporal no parâmetro de localização, enquanto o critério BIC e o teste da razão de verosimilhanças mantêm a preferência pelo modelo estacionário M_1 , evidenciando uma avaliação mais conservadora face ao acréscimo de complexidade paramétrica. Nos restantes subperíodos, quer os definidos pelos pontos de mudança estimados, quer os estabelecidos com base em alterações históricas na estação meteorológica, observa-se concordância total entre AIC, BIC e d_{obs} , com seleção sistemática do Modelo M_1 , o que reforça a ausência de evidência estatística de não estacionaridade e sugere que a estrutura temporal dos mínimos estudados permanece globalmente estável nos diferentes intervalos analisados.

Assim, a comparação entre os resultados obtidos para os valores de temperatura e precipitação estudados evidencia comportamentos distintos quanto à presença de não estacionaridade nas séries analisadas. No caso da temperatura, a série completa revela preferência por modelos não estacionários, com inclusão simultânea de tendência temporal nos parâmetros de localização e de escala, enquanto a análise por subperíodos mostra que essa não estacionaridade se atenua ou desaparece em segmentos temporalmente mais homogêneos, sugerindo que o efeito observado à escala global resulta da agregação de períodos com características distintas. Em contraste, para a precipitação, a evidência de dependência temporal é globalmente mais fraca: apesar de o critério AIC apontar, na série completa, para um modelo com tendência no parâmetro de localização, essa indicação não é confirmada pelos restantes critérios, que favorecem consistentemente modelos estacionários.

O presente estudo evidencia a relevância da modelação de extremos por blocos como instrumento fundamental para caracterizar a variabilidade e a evolução temporal de fenómenos climáticos extremos, permitindo identificar diferenças substanciais entre variáveis meteorológicas e entre escalas temporais de análise. A aplicação sistemática de modelos GEV,

em versões estacionárias e não estacionárias, mostrou-se particularmente útil para avaliar a presença de tendências nos extremos anuais estudados. Os resultados obtidos reforçam a importância de considerar cuidadosamente a segmentação temporal e o contexto histórico das observações, sobretudo quando existem alterações instrumentais ou mudanças de localização das estações meteorológicas. Como perspectivas de investigação futura, destaca-se a realização de análises considerando blocos com diferentes escalas temporais, bem como a exploração de modelos que integrem outras covariáveis explicativas e o desenvolvimento de abordagens espaciais suportadas em séries provenientes de diferentes estações meteorológicas.

Anexos

Apêndice A

Máximos Anuais do Nível do Mar - Port Pirie

Carregamento dos dados

```
# Pacotes
library(ismev)
library(ggplot2)
library(trend)

# Dados
data(portpirie)
dados_pp <- portpirie$SeaLevel
anos_pp <- portpirie$Year
n_pp <- length(dados_pp)

# Pasta para imagens
caminho <- "C:/Users/Utilizador/Documents/dissertação
/dissertação/dissertação/imagens/"
if(!dir.exists(caminho)) dir.create(caminho,
recursive = TRUE)
```

Aplicabilidade dos Modelos

```
# Teste de Mann--Kendall
mk.test(dados_pp)

# Teste de Wald--Wolfowitz
ww.test(dados_pp)
```

Modelação com o Modelo M_1

```
# Ajuste do modelo GEV estacionário
ppfit_gev <- gev.fit(dados_pp, show = TRUE)

# Log-verosimilhança
loglik_gev_pp <- -ppfit_gev$nlh
loglik_gev_pp
```

Níveis de retorno

```
# Níveis de retorno
rl_10_pp <- gev.rl(ppfit_gev, m = 10)
rl_100_pp <- gev.rl(ppfit_gev, m = 100)

rl_10_pp
rl_100_pp
```

Gráficos de diagnóstico do Modelo M_1

```
png(paste0(caminho, "gevpirie.png"),
    width = 2000, height = 2000, res = 300)

par(mfrow = c(2,2), oma = c(0,0,3,0))
gev.diag(ppfit_gev)
mtext("Diagnóstico GEV - Port Pirie", outer = TRUE,
      cex = 1.5)

dev.off()
```

Ajuste e diagnóstico do Modelo M_0

```
# Ajuste do modelo de Gumbel
ppfit_gum <- gum.fit(dados_pp, show = TRUE)

# Log-verosimilhança
loglik_gum_pp <- -ppfit_gum$nlh
loglik_gum_pp

# Diagnóstico gráfico
png(paste0(caminho, "gumbelpirie.png"),
```

```

width = 2000, height = 2000, res = 300)

par(mfrow = c(2,2), oma = c(0,0,3,0))
gum.diag(ppfit_gum)
mtext("Diagnóstico Gumbel-Port-Pirie", outer = TRUE
      , cex = 1.5)

dev.off()

```

Critérios de informação e teste da razão de verossimilhanças

```

# Critérios de informação
aic_gev_pp <- 2*3 - 2*loglik_gev_pp
bic_gev_pp <- log(n_pp)*3 - 2*loglik_gev_pp

aic_gum_pp <- 2*2 - 2*loglik_gum_pp
bic_gum_pp <- log(n_pp)*2 - 2*loglik_gum_pp

# Teste da razão de verossimilhanças
lrt_pp <- -2 * (loglik_gum_pp - loglik_gev_pp)
p_lrt_pp <- 1 - pchisq(lrt_pp, df = 1)

data.frame(
  Modelo = c("Gumbel", "GEV"),
  AIC = c(aic_gum_pp, aic_gev_pp),
  BIC = c(bic_gum_pp, bic_gev_pp)
)

lrt_pp
p_lrt_pp

```

74 APÊNDICE A. MÁXIMOS ANUAIS DO NÍVEL DO MAR - PORT PIRIE

Apêndice B

Máximos Anuais do Nível do Mar - Fremantle

Aplicabilidade dos Modelos

```
mk.test(fremantle$SeaLevel)
ww.test(fremantle$SeaLevel)
```

Ajuste dos Modelos M_0 e M_1

```
# GEV
modelo1dados2 <- gev.fit(fremantle$SeaLevel)

# Gumbel
modelo0dados2 <- gum.fit(fremantle$SeaLevel)
```

Gráficos de diagnóstico

```
png(paste0(caminho, "gevdiafremantle.png"),
    width = 2000, height = 2000, res = 300)

par(mfrow = c(2,2), oma = c(0,0,3,0))
gev.diag(modelo1dados2)
mtext("Diagnóstico GEV - Fremantle", outer = TRUE,
      cex = 1.5)
dev.off()

png(paste0(caminho, "gumbeldiafremantle.png"),
```

```

width = 2000, height = 2000, res = 300)

par(mfrow = c(2,2), oma = c(0,0,3,0))
gum.diag(modelo0dados2)
mtext("Diagnóstico Gumbel - Fremantle", outer = TRUE,
      cex = 1.5)
dev.off()

```

Modelo M_0 versus Modelo M_1

```

loglik_gev_f <- -modelo1dados2$nllh
loglik_gum_f <- -modelo0dados2$nllh
n_f <- length(fremantle$SeaLevel)

aic_gev_f <- 2*3 - 2*loglik_gev_f
bic_gev_f <- log(n_f)*3 - 2*loglik_gev_f

aic_gum_f <- 2*2 - 2*loglik_gum_f
bic_gum_f <- log(n_f)*2 - 2*loglik_gum_f

lrt_f <- -2*(loglik_gum_f - loglik_gev_f)
p_lrt_f <- 1 - pchisq(lrt_f, df = 1)

```

Modelos GEV com tendência

```

cov_f <- matrix(fremantle$Year, ncol = 1)

modelo1Adados2 <- gev.fit(fremantle$SeaLevel, ydat =
  cov_f, mul = 1)
modelo1Bdados2 <- gev.fit(fremantle$SeaLevel, ydat =
  cov_f, sig1 = 1)
modelo1Cdados2 <- gev.fit(fremantle$SeaLevel, ydat =
  cov_f, mul = 1, sig1 = 1)

rv_1A <- 2 * (modelo1Adados2$nllh - modelo1dados2$
  nllh)
rv_1B <- 2 * (modelo1Bdados2$nllh - modelo1dados2$
  nllh)
rv_1C <- 2 * (modelo1Cdados2$nllh - modelo1dados2$
  nllh)

p_1A <- 1 - pchisq(rv_1A, df = 1)
p_1B <- 1 - pchisq(rv_1B, df = 1)
p_1C <- 1 - pchisq(rv_1C, df = 2)

```

Apêndice C

Gráficos de funções densidade

```
1
2
3     #####Funções de densidade probabilidade para as
4         distribuições de Gumbel, Weibull e Fréchet#####
5
6     import numpy as np
7     import matplotlib.pyplot as plt
8     import scipy.stats as stats
9
10    # Parâmetros comuns
11    mu = 2.0
12    sigma = 1.5
13
14    # Geração de valores
15    x = np.linspace(-4, 10, 1000)
16
17    # Distribuições de valores extremos
18    gumbel = stats.gumbel_r.pdf(x, loc=mu, scale=sigma)
19    weibull = stats.weibull_min.pdf(x, 1.5, loc=mu, scale
20        =sigma)
21    frechet = stats.genextreme.pdf(x, -0.5, loc=mu, scale
22        =sigma)
23
24    # Criar figura
25    plt.figure(figsize=(10, 6))
26
27    plt.plot(x, gumbel, label='Gumbel (xi = 0)',
28        linewidth=2)
```

```
26 plt.plot(x, weibull, label='Weibull (xi < 0)',
27         linewidth=2)
28
29 plt.plot(x, frechet, label='Fréchet (xi > 0)',
30         linewidth=2)
31
32 plt.legend(
33     title=r'Parâmetros comuns:  $\mu = 2.0, \sigma = 1.5$ '
34     ,
35     fontsize=13,
36     title_fontsize=12
37 )
38
39 plt.xlabel('Valores', fontsize=14)
40 plt.ylabel('Densidade de Probabilidade', fontsize=14)
41 plt.grid(alpha=0.3)
42
43 # Guardar gráfico
44 plt.savefig(
45     'grafico_valores_extremos.png', # ou .jpg
46     dpi=300,
47     bbox_inches='tight'
48 )
49
50 plt.show()
51
52 #####Funções de densidade de probabilidade para mínimos
53 extremais###
54
55 import numpy as np
56 import matplotlib.pyplot as plt
57 import scipy.stats as stats
58
59 # Parâmetros comuns
60 mu = 2.0
61 sigma = 1.5
62
63 # Geração de valores
64 x = np.linspace(-4, 10, 1000)
65
66 # Distribuições de valores extremos
67 gumbel = stats.gumbel_r.pdf(x, loc=mu, scale=sigma)
```

```
64 weibull = stats.weibull_min.pdf(x, 1.5, loc=mu, scale
    =sigma)
65 frechet = stats.genextreme.pdf(x, -0.5, loc=mu, scale
    =sigma)
66
67 # Criar figura
68 plt.figure(figsize=(10, 6))
69
70 plt.plot(x, gumbel, label='Gumbel (xi = 0)',
    linewidth=2)
71 plt.plot(x, weibull, label='Weibull (xi < 0)',
    linewidth=2)
72 plt.plot(x, frechet, label='Fréchet (xi > 0)',
    linewidth=2)
73
74 plt.legend(
75     title=r'Parâmetros comuns:  $\mu = 2.0, \sigma = 1.5$ 
    ',
76     fontsize=13,
77     title_fontsize=12
78 )
79
80 plt.xlabel('Valores', fontsize=14)
81 plt.ylabel('Densidade de Probabilidade', fontsize=14)
82 plt.grid(alpha=0.3)
83
84 # Guardar gráfico
85 plt.savefig(
86     'grafico_valores_extremos.png', # ou .jpg
87     dpi=300,
88     bbox_inches='tight'
89 )
90
91 plt.show()
```


Apêndice D

Máximos Anuais das MMMD de Temperatura

```
# -----  
# 1 - Pacotes necessários  
# -----  
library(dplyr)  
library(ismev)  
library(trend)  
library(extRemes)  
  
# -----  
# 2 - Carregar dataset principal  
# -----  
load("C:/Users/Utilizador/Documents/dissertação/dissertação/  
dissertação/datatempmax.RData")  
  
# -----  
# 3 - Criar série anual limpa (máximo por ano)  
# -----  
datatempmax_anual <- datatempmax %>%  
group_by(year) %>%  
summarise(Value = max(Value, na.rm = TRUE))  
  
# Verificação  
cat("Número de anos:", nrow(datatempmax_anual), "\n")  
cat("Duplicados anos:", sum(duplicated(datatempmax_anual$  
year)), "\n")  
print(head(datatempmax_anual))  
  
cat("Resumo da série anual:\n")  
print(summary(datatempmax_anual$Value))  
cat("Desvio padrão:", sd(datatempmax_anual$Value), "\n")
```

82APÊNDICE D. MÁXIMOS ANUAIS DAS MMMD DE TEMPERATURA

```
# -----  
# 4 - Criar grupos anuais  
# -----  
grupos_anuais <- list(  
  
# Série completa  
"Serie_Completa" = datatempmax_annual$Value,  
  
# -----  
# Changepoints  
# -----  
"1865-1896" = datatempmax_annual$Value [  
datatempmax_annual$year <= 1896  
],  
  
"1897-1974" = datatempmax_annual$Value [  
datatempmax_annual$year >= 1897 & datatempmax_annual$year <=  
1974  
],  
  
"1975-2018" = datatempmax_annual$Value [  
datatempmax_annual$year >= 1975 & datatempmax_annual$year <=  
2018  
],  
  
# -----  
# Sistemas de observação  
# -----  
"1865-1950" = datatempmax_annual$Value [  
datatempmax_annual$year >= 1865 & datatempmax_annual$year <=  
1950  
],  
  
"1951-1996" = datatempmax_annual$Value [  
datatempmax_annual$year >= 1951 & datatempmax_annual$year <=  
1996  
],  
  
"1997-2018" = datatempmax_annual$Value [  
datatempmax_annual$year >= 1997 & datatempmax_annual$year <=  
2018  
]  
)  
  
# -----  
# 5 - Verificação dos grupos  
# -----  
cat("Observações por grupo:\n")
```

```

print(sapply(grupos_anuais, length))

# -----
# 6 - Validação
# -----

# Sem sobreposição (changepoints)
cat("Sobreposição□changepoints:",
length(intersect(
datatempmax_annual$year[datatempmax_annual$year <= 1896],
datatempmax_annual$year[datatempmax_annual$year >= 1897]
)) == 0, "\n")

# Cobertura total
anos_total <- datatempmax_annual$year

anos_cp <- c(
datatempmax_annual$year[datatempmax_annual$year <= 1896],
datatempmax_annual$year[datatempmax_annual$year >= 1897 &
datatempmax_annual$year <= 1974],
datatempmax_annual$year[datatempmax_annual$year >= 1975]
)

cat("Cobertura□total:", setequal(anos_total, anos_cp), "\n")

# Sem duplicados
cat("Sem□duplicados:", !any(duplicated(anos_cp)), "\n")

# -----
# Função para estatísticas descritivas completas
# -----
estatisticas <- function(x) {
  x <- na.omit(x)
  n <- length(x)
  media <- mean(x)
  mediana <- median(x)
  q1 <- quantile(x, 0.25)
  q3 <- quantile(x, 0.75)
  minimo <- min(x)
  maximo <- max(x)
  amplitude <- maximo - minimo # Rn
  desvio <- sd(x)
  cv <- (desvio / media) * 100
  return(c(
    N = n,
    Media = media,
    Mediana = mediana,
    Q1 = q1,
    Q3 = q3,

```

84 APÊNDICE D. MÁXIMOS ANUAIS DAS MMMD DE TEMPERATURA

```
        Min = minimo ,
        Max = maximo ,
        Rn = amplitude ,
        DesvioPadrao = desvio ,
        CoefVariacao = cv
    ))
}

# -----
# Calcular estatísticas para todos os grupos anuais
# limpos
# -----
tabela_estatisticas <- t(sapply(grupos_anuais,
    estatisticas))

# -----
# Mostrar resultados arredondados
# -----
print(round(tabela_estatisticas, 2))

# =====
# A.3.3 Gráficos de diagnóstico (GEV e Gumbel)
# =====

library(ismev)

diagnosticos_extremos <- function(x){

    x <- na.omit(x)

    # -----
    # Ajuste GEV
    # -----
    ajuste_gev <- gev.fit(x, show = FALSE)

    windows()
    gev.diag(ajuste_gev)

    # -----
    # Ajuste Gumbel
    # -----
    ajuste_gumbel <- gev.fit(x, show = FALSE, fix
        .arg = list(shape = 0))

    windows()
    gev.diag(ajuste_gumbel)
}

# =====
```

```

# Aplicar a todos os grupos
# =====

for(i in seq_along(grupos_lista)){
  diagnosticos_extremos(grupos_lista[[i]])
}

# =====
# A.3 - Função para analisar modelos
# e execução automática para todos os grupos
# =====

library(ismev)
library(trend)
library(extRemes)

cat("=== Análise automática dos grupos de dados ===\n
\n")

analisar_modelos <- function(dados, nome_grupo) {
  cat("-----\n")
  cat("Grupo:", nome_grupo, "\n")
  cat("-----\n")

  # Ajuste dos modelos
  modelo_gev <- gev.fit(dados, show = FALSE)
  modelo_gum <- gum.fit(dados, show = FALSE)

  # Teste de razão de verossimilhança
  lratio <- 2 * ((-modelo_gev$nullh) - (-modelo_gum$nullh))
  pvalor_lrt <- 1 - pchisq(lratio, df = 1)
  cat("LRT=", round(lratio,3), "| p-valor=",
      round(pvalor_lrt,4), "\n")

  # Critérios de informação
  n <- length(dados)
  aic_gev <- -2 * (-modelo_gev$nullh) + 2*3
  bic_gev <- -2 * (-modelo_gev$nullh) + 3*log(n)
  aic_gum <- -2 * (-modelo_gum$nullh) + 2*2
  bic_gum <- -2 * (-modelo_gum$nullh) + 2*log(n)
  cat("AIC_GEV:", round(aic_gev,3), "| BIC_GEV
:", round(bic_gev,3), "\n")
  cat("AIC_Gumbel:", round(aic_gum,3), "| BIC
Gumbel:", round(bic_gum,3), "\n")

  # Testes de tendência e independência
  cat("Teste Mann-Kendall:\n")

```

```

        print(mk.test(dados))

        cat("Teste_Wald-Wolfowitz:\n")
        print(ww.test(dados))

        cat("\n")
    }

# =====
# Aplicar a função a todos os grupos
# =====
cat("=== Análise automática dos grupos de dados ===\n
    \n")

# Executar para todos os grupos
for(nome in names(grupos)) {
    analisar_modelos(grupos[[nome]], nome)
}

cat("=== Fim da análise automática ===\n")

# =====
# A.3.1 Estimativas GEV e Gumbel + IC 95% para xi
# =====

library(ismev)

# Função para estimativas e IC de xi
estimativas_gev_gumbel <- function(x, nome_grupo) {

    x <- as.numeric(na.omit(x))

    # Ajuste GEV
    ajuste_gev <- gev.fit(x, show = FALSE)

    mu_gev      <- ajuste_gev$mle[1]
    sigma_gev   <- ajuste_gev$mle[2]
    xi_gev      <- ajuste_gev$mle[3]
    se_xi       <- sqrt(ajuste_gev$cov[3,3])

    IC_inf_xi   <- xi_gev - 1.96 * se_xi
    IC_sup_xi   <- xi_gev + 1.96 * se_xi

    # Ajuste Gumbel (xi = 0)
    ajuste_gumbel <- gev.fit(x, show = FALSE,
        ydat = x, fix.arg = list(shape = 0))

```

```

mu_gumbel      <- ajuste_gumbel$mle[1]
sigma_gumbel   <- ajuste_gumbel$mle[2]

data.frame(
  Grupo          = nome_grupo ,
  mu_gev         = mu_gev ,
  sigma_gev      = sigma_gev ,
  xi_gev         = xi_gev ,
  xi_IC_inf      = IC_inf_xi ,
  xi_IC_sup      = IC_sup_xi ,
  mu_gumbel      = mu_gumbel ,
  sigma_gumbel   = sigma_gumbel
)
}

# =====
# Aplicação a todos os grupos de temperatura
# =====

grupos_lista <- list(
  "Série_completa" = datatempmax$Value ,
  "Antes_de_1897" = datatempmaxbefore1897$Value ,
  "1897--1974"     = datatempmax1897a1974$Value ,
  "1951--1996"     = datatempmax1951a1996$Value ,
  "1975--2018"     = datatempmax1975a2018$Value ,
  "1997--2018"     = datatempmax1997a2018$Value ,
  ">=1951"         = datatempmaxafterorequal1951$Value
)

tabela_parametros <- do.call(rbind, lapply(seq_along(
  grupos_lista), function(i) {
  estimativas_gev_gumbel(grupos_lista[[i]],
    names(grupos_lista)[i])
}))

# Arredondar colunas numéricas
tabela_parametros[, -1] <- round(tabela_parametros[,
  -1], 3)

# Resultados finais
print(tabela_parametros)

library(ismev)

return_levels_manual <- function(x, nome_grupo ,
  return_periods = c(10, 100)) {
  x <- na.omit(x)

```

```

# Ajuste GEV
ajuste_gev <- gev.fit(x, show = FALSE)
mu <- ajuste_gev$mle[1]
sigma <- ajuste_gev$mle[2]
xi <- ajuste_gev$mle[3]

# Função nível de retorno GEV
gev_rl <- function(T, mu, sigma, xi) {
  mu + (sigma/xi) * ((-log(1 - 1/T))^(xi) - 1)
}

# Calcular níveis de retorno GEV
rl_gev <- sapply(return_periods, function(T)
  gev_rl(T, mu, sigma, xi))

# Ajuste Gumbel
ajuste_gumbel <- gev.fit(x, show = FALSE, fix
  .arg = list(shape = 0))
mu_g <- ajuste_gumbel$mle[1]
sigma_g <- ajuste_gumbel$mle[2]

# Função nível de retorno Gumbel
gumbel_rl <- function(T, mu, sigma) {
  mu - sigma * log(-log(1 - 1/T))
}

# Calcular níveis de retorno Gumbel
rl_gumbel <- sapply(return_periods, function(
  T) gumbel_rl(T, mu_g, sigma_g))

# Data.frame final (sem IC)
df <- data.frame(
  Grupo = nome_grupo,

  RL10GEV = rl_gev[1],
  RL100GEV = rl_gev[2],

  RL10Gumbel = rl_gumbel[1],
  RL100Gumbel = rl_gumbel[2]
)

return(df)
}

# Aplicar a todos os grupos
tabela_return_levels <- do.call(rbind, lapply(seq_
  along(grupos_lista), function(i) {

```

```

        return_levels_manual(grupos_lista[[i]], names
                             (grupos_lista)[i])
    )))

# Arredondar
tabela_return_levels[, -1] <- round(tabela_return_
    levels[, -1], 2)

print(tabela_return_levels)

library(ismev)
library(dplyr)

# =====
# 1) PREPARAR DADOS
# =====

dados_temp <- datatempmax_annual %>%
mutate(
temp = as.numeric(Value)
)

# =====
# 2) DEFINIR GRUPOS DE DADOS
# =====

grupos <- list(
serie_completa = dados_temp,
grupo_1865_1950 = filter(dados_temp, year >= 1865 &
    year <= 1950),
grupo_1951_1996 = filter(dados_temp, year >= 1951 &
    year <= 1996),
grupo_1997_2018 = filter(dados_temp, year >= 1997 &
    year <= 2018),
grupo_1865_1887 = filter(dados_temp, year >= 1865 &
    year <= 1887),
grupo_1888_2018 = filter(dados_temp, year >= 1888 &
    year <= 2018)
)

# =====
# 3) FUNÇÃO PARA AJUSTAR MODELOS E CALCULAR CRITÉRIOS
# =====

compara_modelos_gev <- function(valores, anos) {

    n <- length(valores)
    cov <- matrix(anos, ncol = 1)

```

90 APÊNDICE D. MÁXIMOS ANUAIS DAS MMMD DE TEMPERATURA

```

# Ajuste dos modelos
m1 <- gev.fit(valores, show = FALSE)
m1A <- gev.fit(valores, ydat = cov, mul = 1,
  show = FALSE)
m1B <- gev.fit(valores, ydat = cov, sigl = 1,
  show = FALSE)
m1C <- gev.fit(valores, ydat = cov, mul = 1,
  sigl = 1, show = FALSE)

# Log-verosimilhanças
loglik <- c(-m1$nullh, -m1A$nullh, -m1B$nullh, -
  m1C$nullh)

# Número de parâmetros
k <- c(3, 4, 4, 5)

# AIC e BIC
AIC <- -2 * loglik + 2 * k
BIC <- -2 * loglik + k * log(n)

# Testes de razão de verosimilhança
RV_1_1A <- 2 * (loglik[2] - loglik[1])
RV_1_1B <- 2 * (loglik[3] - loglik[1])
RV_1_1C <- 2 * (loglik[4] - loglik[1])

p_1_1A <- 1 - pchisq(RV_1_1A, df = 1)
p_1_1B <- 1 - pchisq(RV_1_1B, df = 1)
p_1_1C <- 1 - pchisq(RV_1_1C, df = 2)

# Tabela final
tabela <- data.frame(
  tabela <- data.frame(
    Modelo = c("Base", "1A_μ(mu)", "1B_σ(sigma)", "
      1C_μ(mu, σsigma)"),
    AIC = round(AIC, 3),
    BIC = round(BIC, 3),
    RV_pvalor = c(
      "-",
      paste0(round(RV_1_1A, 3), "_(p_μ=μ)", signif(p_
        1_1A, 3), ")"),
      paste0(round(RV_1_1B, 3), "_(p_σ=σ)", signif(p_
        1_1B, 3), ")"),
      paste0(round(RV_1_1C, 3), "_(p_μ=μ)", signif(p_
        1_1C, 3), ")")
    ),
    stringsAsFactors = FALSE
  )
# =====

```

```

# 4) APLICAR A TODOS OS GRUPOS
# =====

resultados <- lapply(grupos, function(d) {
  compara_modelos_gev(
    valores = d$temp,
    anos = d$year
  )
})

# =====
# 5) PRINT
# =====

for (nome in names(resultados)) {
  cat("\n
      =====\n
      ")
  cat("Resultados para:", nome, "\n")
  cat("
      =====\n
      n")
  print(resultados[[nome]])
}

#GUMBEL

library(ismev)
library(dplyr)

# =====
# 1) PREPARAR DADOS
# =====

dados_temp <- datatempmax_annual %>%
mutate(
temp = as.numeric(Value)
)

# =====
# 2) DEFINIR GRUPOS DE DADOS
# =====

grupos <- list(
serie_completa = dados_temp,

```

92 APÊNDICE D. MÁXIMOS ANUAIS DAS MMMD DE TEMPERATURA

```
grupo_1865_1950 = filter(dados_temp, year >=
  1865 & year <= 1950),
grupo_1951_1996 = filter(dados_temp, year >=
  1951 & year <= 1996),
grupo_1997_2018 = filter(dados_temp, year >=
  1997 & year <= 2018),
grupo_1865_1887 = filter(dados_temp, year >=
  1865 & year <= 1887),
grupo_1888_2018 = filter(dados_temp, year >=
  1888 & year <= 2018)
)

# =====
# 3) FUNÇÃO PARA AJUSTAR MODELOS E CALCULAR
# CRITÉRIOS
# =====

compara_modelos_gev <- function(valores, anos
) {

  n <- length(valores)
  cov <- matrix(anos, ncol = 1)

  # Ajuste dos modelos
  m1 <- gev.fit(valores, show = FALSE)
  m1A <- gev.fit(valores, ydat = cov,
    mul = 1, show = FALSE)
  m1B <- gev.fit(valores, ydat = cov,
    sig1 = 1, show = FALSE)
  m1C <- gev.fit(valores, ydat = cov,
    mul = 1, sig1 = 1, show = FALSE)

  # Log-verosimilhanças
  loglik <- c(-m1$nullh, -m1A$nullh, -m1B
    $nullh, -m1C$nullh)

  # Número de parâmetros
  k <- c(3, 4, 4, 5)

  # AIC e BIC
  AIC <- -2 * loglik + 2 * k
  BIC <- -2 * loglik + k * log(n)

  # Testes de razão de verosimilhança
  RV_1_1A <- 2 * (loglik[2] - loglik
    [1])
  RV_1_1B <- 2 * (loglik[3] - loglik
    [1])
```

```

RV_1_1C <- 2 * (loglik[4] - loglik
[1])

p_1_1A <- 1 - pchisq(RV_1_1A, df = 1)
p_1_1B <- 1 - pchisq(RV_1_1B, df = 1)
p_1_1C <- 1 - pchisq(RV_1_1C, df = 2)

# Tabela final
tabela <- data.frame(
tabela <- data.frame(
Modelo = c("Base", "1A_μ(mu)", "1B_μ(
sigma)", "1C_μ(mu, σsigma)"),
AIC = round(AIC, 3),
BIC = round(BIC, 3),
RV_pvalor = c(
"-",
paste0(round(RV_1_1A, 3), "_(p_=" ,
signif(p_1_1A, 3), ")"),
paste0(round(RV_1_1B, 3), "_(p_=" ,
signif(p_1_1B, 3), ")"),
paste0(round(RV_1_1C, 3), "_(p_=" ,
signif(p_1_1C, 3), ")")
),
stringsAsFactors = FALSE
)

# =====
# 4) APLICAR A TODOS OS GRUPOS
# =====

resultados <- lapply(grupos, function
(d) {
  compara_modelos_gev(
valores = d$temp,
anos = d$year
)
})

# =====
# 5) PRINT
# =====

for (nome in names(resultados)) {
  cat("\n=====\n")
  cat("Resultados_μpara:", nome,
"\n")
  cat("=====\n")
  print(resultados[[nome]])
}

```

```
library(ismev)
library(trend)

# Lista dos grupos de dados
grupos <- list(
  geral = datatempmax,
  antes_1897 = datatempmaxbefore1897,
  entre_1897_1974 =
    datatempmax1897a1974,
  entre_1951_1996 =
    datatempmax1951a1996,
  entre_1975_2018 =
    datatempmax1975a2018,
  entre_1997_2018 =
    datatempmax1997a2018,
  apos_1951 =
    datatempmaxafterorequal1951
)

# Função para ajustar os modelos e
# calcular critérios e testes
ajusta_modelos <- function(dados) {
  n <- length(dados$Value)
  cov <- matrix(dados$year,
    ncol = 1)

  # Ajuste dos modelos
  modelo1 <- gev.fit(dados$
    Value, show = FALSE)
  #
  # Estacionário
  modelo1A <- gev.fit(dados$
    Value, ydat = cov, mul =
    1, show = FALSE) #
  # Tendência mu
  modelo1B <- gev.fit(dados$
    Value, ydat = cov, sig1 =
    1, show = FALSE) #
  # Tendência sigma
```

```

modelo1C <- gev.fit(dados$
  Value, ydat = cov, mul =
  1, sigl = 1, show = FALSE)
  # Tendência mu+sigma

# AIC e BIC
loglik <- c(-modelo1$nllh, -
  modelo1A$nllh, -modelo1B$
  nllh, -modelo1C$nllh)
k <- c(3, 4, 4, 5)
aic <- -2*loglik + 2*k
bic <- -2*loglik + k*log(n)

# Testes de razão de
verossimilhança
lr_1_1A <- 2 * ((-modelo1A$
  nllh) - (-modelo1$nllh))
lr_1_1B <- 2 * ((-modelo1B$
  nllh) - (-modelo1$nllh))
lr_1_1C <- 2 * ((-modelo1C$
  nllh) - (-modelo1$nllh))
lr_1B_1C <- 2 * ((-modelo1C$
  nllh) - (-modelo1B$nllh))

p_1_1A <- 1 - pchisq(lr_1_1A,
  df = k[2]-k[1])
p_1_1B <- 1 - pchisq(lr_1_1B,
  df = k[3]-k[1])
p_1_1C <- 1 - pchisq(lr_1_1C,
  df = k[4]-k[1])
p_1B_1C <- 1 - pchisq(lr_1B_1
  C, df = k[4]-k[3])

return(list(
  AIC = aic,
  BIC = bic,
  RV = c(lr_1_1A, lr_1_1B, lr_1
    _1C, lr_1B_1C),
  pvalor = c(p_1_1A, p_1_1B, p_
    1_1C, p_1B_1C)
))
}

# Aplicar a todos os grupos
resultados_completos <- lapply(grupos
  , ajusta_modelos)

# Criar tabela de AIC e BIC
tabela_final <- data.frame(

```

```

Grupo = character(),
Modelo = character(),
AIC = numeric(),
BIC = numeric(),
stringsAsFactors = FALSE
)

for(nome in names(resultados_
completos)) {
  res <- resultados_completos[[
  nome]]
  tabela_final <- rbind(
  tabela_final,
  data.frame(Grupo = nome,
  Modelo = c("1", "1A", "1B"
  , "1C"), AIC = res$AIC,
  BIC = res$BIC)
  )
}

# Mostrar tabela de AIC e BIC
print(tabela_final)

# Criar tabela de RV e p-valor
tabela_RV <- data.frame(
Grupo = character(),
Comparacao = character(),
RV = numeric(),
pvalor = numeric(),
stringsAsFactors = FALSE
)

comparacoes <- c("1 vs 1A", "1 vs 1B"
, "1 vs 1C", "1B vs 1C")

for(nome in names(resultados_
completos)) {
  res <- resultados_completos[[
  nome]]
  tabela_RV <- rbind(
  tabela_RV,
  data.frame(
  Grupo = nome,
  Comparacao = comparacoes,
  RV = res$RV,
  pvalor = res$pvalor
  )
  )
}

```

```
# Mostrar tabela de RV e p-valor
print(tabela_RV)

library(ismev)

# Lista dos grupos de dados
grupos <- list(
  geral = datatempmax,
  antes_1897 = datatempmaxbefore1897,
  entre_1897_1974 =
    datatempmax1897a1974,
  entre_1951_1996 =
    datatempmax1951a1996,
  entre_1975_2018 =
    datatempmax1975a2018,
  entre_1997_2018 =
    datatempmax1997a2018,
  apos_1951 =
    datatempmaxafterorequal1951
)

# Função para calcular níveis de
  retorno Gumbel M0 (sem IC)
niveis_retorno_gumbel <- function(
  dados, nome_grupo, periodos = c
  (10,100)) {
  # Ajuste Gumbel M0
  fit <- gum.fit(dados$Value,
    show = FALSE)
  mu <- fit$mle[1]
  sigma <- fit$mle[2]

  resultados <- data.frame(
    Grupo = nome_grupo,
    Modelo = "M0",
    PerodoRetorno = periodos,
    NivelRetorno = NA
  )

  for(i in seq_along(periodos))
  {
    T <- periodos[i]
    q <- mu - sigma * log
      (-log(1 - 1/T))
  }
}
```

```
                resultados$
                  NivelRetorno[i] <-
                    q
            }

            return(resultados)
    }

    # Aplicar a todos os grupos
    resultados_gumbel_temporal <- do.call
    (
    rbind,
    lapply(names(grupos), function(nome)
    {
        niveis_retorno_gumbel(grupos
        [[nome]], nome)
    })
    )

    # Arredondar valores
    resultados_gumbel_temporal[, 4] <-
        round(resultados_gumbel_temporal[,
        4], 2)

    # Mostrar tabela final
    print(resultados_gumbel_temporal)
```

Apêndice E

Mínimos Anuais das MMTD de Precipitação

```
# =====  
#           PACOTES  
# =====  
library(RecordTest)  
library(trend)  
library(ismev)  
library(ggplot2)  
library(dplyr)  
library(randtests)  
  
# =====  
#           PASTA PARA GUARDAR GRÁFICOS  
# =====  
pasta_graficos <- "C:/Users/Utilizador/Documents/dissertação/  
dissertação/dissertação/imagens"  
  
tema_dissertacao <- theme_minimal(base_size = 12) +  
theme(plot.title = element_blank())  
  
# =====  
#           CARREGAR DADOS  
# =====  
load("C:/Users/Utilizador/Documents/dissertação/dissertação/  
material_aula/Original_não_mexer/datapreciptotalanual.  
RData")  
  
datapreciptotalanual$precip_inv <- datapreciptotalanual$  
preciptotal * -1  
  
# =====
```

100 APÊNDICE E. MÍNIMOS ANUAIS DAS MMTD DE PRECIPITAÇÃO

```

#           DIVISÃO EM GRUPOS
# =====
divide_dataframe_by_year <- function(df, year_column, target_
year) {
  if (!year_column %in% names(df)) stop(paste("Coluna",
year_column, "não existe."))
  if (!is.numeric(df[[year_column]])) {
    if (all(grepl("^\\d{4}$", df[[year_column]]))
) {
      df[[year_column]] <- as.numeric(df[[
year_column]])
    } else {
      stop("Coluna de ano tem de ser
numérica ou convertível.")
    }
  }
  df_before <- df[df[[year_column]] < target_year, ]
  df_after_or_equal <- df[df[[year_column]] >= target_
year, ]
  return(list(before = df_before, after_or_equal = df_
after_or_equal))
}

div1 <- divide_dataframe_by_year(datapreciptotalanual, "year"
, 1931)
grupoA <- div1$before
resto1 <- div1$after_or_equal

div2 <- divide_dataframe_by_year(resto1, "year", 1970)
grupoB <- div2$before
grupoC <- div2$after_or_equal

# Inverter valores para estudar mínimos como máximos
grupoA$precip_inv <- grupoA$preciptotal * -1
grupoB$precip_inv <- grupoB$preciptotal * -1
grupoC$precip_inv <- grupoC$preciptotal * -1

# =====
#           FUNÇÃO DE ANÁLISE DESCRITIVA E GRÁFICOS
# =====
analisar_grupo <- function(grupo, nome_grupo) {

  precip_invertida <- grupo$precip_inv
  df_precip <- data.frame(precip_invertida = precip_
invertida)

  # Estatísticas descritivas
  cat("\n\n=====\n")

```

```

cat("  Análise Descritiva -", nome_grupo, "\n")
cat("=====\n")
print(summary(precip_invertida))
cat("Desvio padrão:", sd(precip_invertida, na.rm =
  TRUE), "\n")
cat("Variância:", var(precip_invertida, na.rm = TRUE)
  , "\n")

# Testes
cat("\n--- Mann-Kendall ---\n")
print(mk.test(precip_invertida))

cat("\n--- Runs Test ---\n")
print(runs.test(precip_invertida))
}

# Aplicar aos três grupos
analisar_grupo(grupoA, "GrupoA")
analisar_grupo(grupoB, "GrupoB")
analisar_grupo(grupoC, "GrupoC")
analisar_grupo(datapreciptotalanual, "SerieCompleta")

# =====
#           AJUSTE DE MODELOS GEV/GUMBEL
# =====
ajustar_modelos <- function(dados_inv, nome_grupo) {
  cat("\n---", nome_grupo, "---\n")

  # GEV
  fit_gev <- gev.fit(dados_inv, show = FALSE)
  loglik_gev <- -fit_gev$nullh
  k_gev <- 3
  n <- length(dados_inv)
  aic_gev <- 2 * k_gev - 2 * loglik_gev
  bic_gev <- log(n) * k_gev - 2 * loglik_gev

  # Gumbel
  fit_gum <- gum.fit(dados_inv, show = FALSE)
  loglik_gum <- -fit_gum$nullh
  k_gum <- 2
  aic_gum <- 2 * k_gum - 2 * loglik_gum
  bic_gum <- log(n) * k_gum - 2 * loglik_gum

```

102 APÊNDICE E. MÍNIMOS ANUAIS DAS MMTD DE PRECIPITAÇÃO

```
# Teste de razão de verossimilhanças
lr_stat <- -2 * (loglik_gum - loglik_gev)
p_valor <- 1 - pchisq(lr_stat, df = 1)

# Mostrar resultados
cat("GEV(mínimos): location=", -fit_gev$mle[1],
    "| scale=", fit_gev$mle[2],
    "| shape=", fit_gev$mle[3], "\n")
cat("AIC_GEV:", round(aic_gev, 3), "| AIC_Gumbel:",
    round(aic_gum, 3), "\n")
cat("BIC_GEV:", round(bic_gev, 3), "| BIC_Gumbel:",
    round(bic_gum, 3), "\n")
cat("LR_Statistic:", round(lr_stat, 3), "| p-valor:",
    signif(p_valor, 3), "\n")

# Gráficos de diagnóstico GEV
op <- par(mfrow = c(2, 2))
gev.diag(fit_gev)
par(op)
}

# Aplicar modelos
ajustar_modelos(grupoA$precip_inv, "GrupoA")
ajustar_modelos(grupoB$precip_inv, "GrupoB")
ajustar_modelos(grupoC$precip_inv, "GrupoC")
ajustar_modelos(datapreciptotalanual$precip_inv, "
  SerieCompleta")

# Testes de tendência e independência
mk.test(datapreciptotalanual$precip_inv)
library(randtests)
runs.test(datapreciptotalanual$precip_inv)

# =====
# PACOTES
# =====
library(dplyr)
library(ggplot2)
library(extRemes)
library(evd)
library(tibble)
library(ismev)

# =====
# DIRETORIO DE IMAGENS
# =====
```

```

dir_imgs <- "C:/Users/Utilizador/Documents/dissertação/
dissertação/dissertação/imagens"
if(!dir.exists(dir_imgs)){
  dir.create(dir_imgs, recursive = TRUE)
  cat("Pasta_criada:", dir_imgs, "\n")
}

# =====
# 1) CARREGAR DADOS
# =====
load("C:/Users/Utilizador/Documents/dissertação/dissertação/
material_aula/Original_não_mexer/datapreciptotalanual.
RData")
datapreciptotalanual <- datapreciptotalanual %>%
mutate(
precip_inv = -as.numeric(preciptotal) # para modelar minimos
      como maximos
)

# =====
# 2) CRIAR GRUPOS
# =====
grupo1 <- filter(datapreciptotalanual, year >= 1865 & year <=
1950)
grupo2 <- filter(datapreciptotalanual, year >= 1951 & year <=
1996)
grupo3 <- filter(datapreciptotalanual, year >= 1997 & year <=
2018)

grupo1_dados <- filter(datapreciptotalanual, year >= 1865 &
year <= 1887)
grupo2_dados <- filter(datapreciptotalanual, year >= 1888 &
year <= 2018)

# =====
# 3) FUNCOES AUXILIARES
# =====
verifica_tamanho <- function(x, min_n = 10) {
  if(length(na.omit(x)) < min_n) {
    warning(paste("Amostra_muito_pequena_(n=",
length(na.omit(x)), ")._Alguns_ajustes_
podem_falhar."))
    return(FALSE)
  }
  TRUE
}

# funcao para imprimir parametros

```

```

print_model_params <- function(fit_gev = NULL, fit_gum = NULL
, nome_grupo = "") {
  cat("\n--- Parametros estimados (" , nome_grupo, ")
  ---\n", sep = "")
  if(!is.null(fit_gev)) {
    mle <- fit_gev$mle
    se <- if(!is.null(fit_gev$se)) fit_gev$se
    else rep(NA, length(mle))
    cat("GEV (minimos como maximos):\n")
    cat("  location (mu)=", sprintf("%.4f", -mle
[1]),
"| se=", if(!is.na(se[1])) sprintf("%.4f",
se[1]) else "NA", "\n")
    cat("  scale (sigma)=", sprintf("%.4f", mle
[2]),
"| se=", if(!is.na(se[2])) sprintf("%.4f",
se[2]) else "NA", "\n")
    cat("  shape (xi)=", sprintf("%.4f", mle
[3]),
"| se=", if(!is.na(se[3])) sprintf("%.4f",
se[3]) else "NA", "\n")
  } else {
    cat("GEV: ajuste nao disponivel.\n")
  }
  if(!is.null(fit_gum)) {
    mle_g <- fit_gum$mle
    se_g <- if(!is.null(fit_gum$se)) fit_gum$se
    else rep(NA, length(mle_g))
    cat("Gumbel (minimos como maximos):\n")
    cat("  location (mu)=", sprintf("%.4f", -mle
_g[1]),
"| se=", if(length(se_g)>=1 && !is.na(se_g
[1])) sprintf("%.4f", se_g[1]) else "NA",
"\n")
    cat("  scale (sigma)=", sprintf("%.4f", mle
_g[2]),
"| se=", if(length(se_g)>=2 && !is.na(se_g
[2])) sprintf("%.4f", se_g[2]) else "NA",
"\n")
  } else {
    cat("Gumbel: ajuste nao disponivel.\n")
  }
  cat("-----\n")
}

ajustar_modelos <- function(dados_inv, nome_grupo) {
  cat("\n---", nome_grupo, "---\n")

```

```

dados_inv <- na.omit(dados_inv)
n <- length(dados_inv)

if(!verifica_tamanho(dados_inv)) {
  return(list(resumo = tibble(
    grupo = nome_grupo, n = n, aic_gev = NA, aic_gum = NA,
    bic_gev = NA, bic_gum = NA, lr_stat = NA, p_valor = NA
  )), fit_gev = NULL, fit_gum = NULL))
}

# fit GEV e Gumbel (ismev)
fit_gev <- gev.fit(dados_inv, show = FALSE)
fit_gum <- gum.fit(dados_inv, show = FALSE)

print_model_params(fit_gev = fit_gev, fit_gum = fit_gum,
  nome_grupo = nome_grupo)

loglik_gev <- -fit_gev$nullh
loglik_gum <- -fit_gum$nullh

aic_gev <- 2*3 - 2*loglik_gev
bic_gev <- log(n)*3 - 2*loglik_gev
aic_gum <- 2*2 - 2*loglik_gum
bic_gum <- log(n)*2 - 2*loglik_gum

lr_stat <- -2 * (loglik_gum - loglik_gev)
p_valor <- 1 - pchisq(lr_stat, df = 1)

cat("AIC_GEV:", round(aic_gev,3), "| AIC_Gumbel:",
  round(aic_gum,3), "\n")
cat("BIC_GEV:", round(bic_gev,3), "| BIC_Gumbel:",
  round(bic_gum,3), "\n")
cat("LR_statistic:", round(lr_stat,3), "| p-valor:",
  signif(p_valor,4), "\n")

list(
  resumo = tibble(
    grupo = nome_grupo,
    n = n,
    aic_gev = aic_gev,
    aic_gum = aic_gum,
    bic_gev = bic_gev,
    bic_gum = bic_gum,
    lr_stat = lr_stat,
    p_valor_lr = p_valor,
    gev_location = -fit_gev$mle[1],
    gev_scale = fit_gev$mle[2],

```

```

    gev_shape = fit_gev$mle[3],
    gum_location = -fit_gum$mle[1],
    gum_scale = fit_gum$mle[2]
  ),
  fit_gev = fit_gev,
  fit_gum = fit_gum
)
}

calcular_return_levels <- function(dados_inv, nome_grupo,
  return.periods = c(10,100)) {
  cat("\n---□Return□levels:", nome_grupo, "---\n")
  dados_inv <- na.omit(dados_inv)

  if(!verifica_tamanho(dados_inv)) return(NULL)

  modelo <- tryCatch(
    fevd(dados_inv, type="GEV", method="MLE"),
    error = function(e) fevd(dados_inv, type="GEV",
      method="Lmoments")
  )

  r1 <- tryCatch(
    return.level(modelo, return.period=return.periods, do
      .ci=TRUE),
    error = function(e) return.level(modelo, return.
      period=return.periods, do.ci=FALSE)
  )

  print(r1)
  invisible(r1)
}

gerar_diagnostico <- function(fit_gev, fit_gum, nome_grupo,
  base_font_pt = 12) {
  rel_cex <- base_font_pt / 12
  safe_name <- iconv(gsub("□", "_", nome_grupo), from="
    UTF-8", to="ASCII//TRANSLIT")

  if(!is.null(fit_gev)) {
    file_out_gev <- file.path(dir_imgs, paste0("
      diagnostico_GEV_", safe_name, ".png"))
    png(file_out_gev, width = 8, height = 6,
      units = "in", res = 300)
    op <- par(no.readonly = TRUE)
    par(cex.lab = rel_cex, cex.axis = rel_cex,
      cex.main = rel_cex*1.1, mar = c(4,4,3,2))
    gev.diag(fit_gev)
  }
}

```

```

        # title(paste("Diagnostico GEV -", nome_grupo
        ), cex.main = rel_cex*1.1) # REMOVIDO
        par(op)
        dev.off()
        cat("Guardado:", file_out_gev, "\n")
    }

    if(!is.null(fit_gum)) {
        file_out_gum <- file.path(dir_imgs, paste0("
            diagnostico_Gumbel_", safe_name, ".png"))
        png(file_out_gum, width = 8, height = 6,
            units = "in", res = 300)
        op <- par(no.readonly = TRUE)
        par(cex.lab = rel_cex, cex.axis = rel_cex,
            cex.main = rel_cex*1.1, mar = c(4,4,3,2))
        gum.diag(fit_gum)
        # title(paste("Diagnostico Gumbel -", nome_
        grupo), cex.main = rel_cex*1.1) #
        REMOVIDO
        par(op)
        dev.off()
        cat("Guardado:", file_out_gum, "\n")
    }
}

analise_grupo <- function(dados, nome_grupo, base_font_pt =
12) {
    cat("\n===== \n")
    cat("Analise_grupo", nome_grupo, "\n")
    cat("===== \n")

    resultado <- ajustar_modelos(dados$precip_inv, nome_
grupo)
    rl <- calcular_return_levels(dados$precip_inv, nome_
grupo)

    gerar_diagnostico(resultado$fit_gev, resultado$fit_
gum, nome_grupo, base_font_pt = base_font_pt)

    list(resumo = resultado$resumo, return_levels = rl)
}

# =====
# 4) APLICAR FUNCOES A SERIE COMPLETA E GRUPOS
# =====
resumo_full <- analise_grupo(datapreciptotalannual, "Serie_
completa", base_font_pt = 12)
resumo_g1 <- analise_grupo(grupo1, "Grupo_1_(1865-1950)",
base_font_pt = 12)

```

108 APÊNDICE E. MÍNIMOS ANUAIS DAS MMTD DE PRECIPITAÇÃO

```
resumo_g2 <- analise_grupo(grupo2, "Grupo_2_(1951-1996)",
  base_font_pt = 12)
resumo_g3 <- analise_grupo(grupo3, "Grupo_3_(1997-2018)",
  base_font_pt = 12)

resultados_modelos <- bind_rows(
  resumo_full$resumo,
  resumo_g1$resumo,
  resumo_g2$resumo,
  resumo_g3$resumo
)

print(resultados_modelos)

# =====
# 5) CHANGEPOINT
# =====
resultado_cp1 <- analise_grupo(grupo1_dados, "Grupo_1_(1865-1887)", base_font_pt = 12)
resultado_cp2 <- analise_grupo(grupo2_dados, "Grupo_2_(1888-2018)", base_font_pt = 12)

resumos_changepoint <- bind_rows(
  resultado_cp1$resumo,
  resultado_cp2$resumo
)

print(resumos_changepoint)

##### Avaliação dos modelos com tendência temporal
#####

library(ismev)
library(dplyr)

# =====
# 1) CARREGAR DADOS
# =====
load("C:/Users/Utilizador/Documents/dissertação/dissertação/
  material_aula/Original_não_mexer/datapreciptotalanual.
  RData")

datapreciptotalanual <- datapreciptotalanual %>%
mutate(
precip_inv = -as.numeric(preciptotal) # mínimos ? máximos
```

```

)

# =====
# 2) DEFINIR GRUPOS DE DADOS
# =====
grupos <- list(
  serie_completa = datapreciptotalanual,
  grupo_1865_1950 = filter(datapreciptotalanual, year >= 1865 &
    year <= 1950),
  grupo_1951_1996 = filter(datapreciptotalanual, year >= 1951 &
    year <= 1996),
  grupo_1997_2018 = filter(datapreciptotalanual, year >= 1997 &
    year <= 2018),
  grupo_1865_1887 = filter(datapreciptotalanual, year >= 1865 &
    year <= 1887),
  grupo_1888_2018 = filter(datapreciptotalanual, year >= 1888 &
    year <= 2018)
)

# =====
# 3) FUNÇÃO PARA AJUSTAR MODELOS E CALCULAR CRITÉRIOS
# =====
compara_modelos_gev <- function(valores, anos) {

  n <- length(valores)
  cov <- matrix(anos, ncol = 1)

  # Ajuste dos modelos
  m1 <- gev.fit(valores, show = FALSE)
  m1A <- gev.fit(valores, ydat = cov, mul = 1, show =
    FALSE)
  m1B <- gev.fit(valores, ydat = cov, sigl = 1, show =
    FALSE)
  m1C <- gev.fit(valores, ydat = cov, mul = 1, sigl =
    1, show = FALSE)

  # Log-verossimilhanças
  loglik <- c(-m1$nullh, -m1A$nullh, -m1B$nullh, -m1C$nullh
    )

  # Número de parâmetros
  k <- c(3, 4, 4, 5)

  # AIC e BIC
  AIC <- -2 * loglik + 2 * k
  BIC <- -2 * loglik + k * log(n)

  # Testes de razão de verossimilhança
  RV_1_1A <- 2 * (loglik[2] - loglik[1])

```

```

RV_1_1B <- 2 * (loglik[3] - loglik[1])
RV_1_1C <- 2 * (loglik[4] - loglik[1])

p_1_1A <- 1 - pchisq(RV_1_1A, df = 1)
p_1_1B <- 1 - pchisq(RV_1_1B, df = 1)
p_1_1C <- 1 - pchisq(RV_1_1C, df = 2)

# Tabela final
tabela <- data.frame(
  Modelo = c("Base", "1A_μ", "1B_σ", "1C_μ,σ"),
  ,
  AIC = round(AIC, 3),
  BIC = round(BIC, 3),
  RV_pvalor = c(
    "-",
    paste0(round(RV_1_1A, 3), "_(p_="), signif(p_1_1A, 3)
    , ")"),
    paste0(round(RV_1_1B, 3), "_(p_="), signif(p_1_1B, 3)
    , ")"),
    paste0(round(RV_1_1C, 3), "_(p_="), signif(p_1_1C, 3)
    , ")")
  ),
  stringsAsFactors = FALSE
)

return(tabela)
}

# =====
# 4) APLICAR A TODOS OS GRUPOS E IMPRIMIR
# =====
resultados <- lapply(grupos, function(d) {
  compara_modelos_gev(
    valores = d$precip_inv,
    anos = d$year
  )
})

# =====
# 5) PRINT DE TODAS AS TABELAS
# =====
for (nome in names(resultados)) {
  cat("\n=====\\n")
  cat("Resultados para:", nome, "\\n")
  cat("=====\\n")
  print(resultados[[nome]])
}

```

Bibliografia

- [1] H. Akaike. Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory*, pages 267–281, 1973. [16](#)
- [2] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Ltd, 2004. [4](#)
- [3] E. Castillo. *Extreme Value and Related Models with Applications in Engineering and Science*. Wiley, 2005. [5](#), [6](#), [7](#)
- [4] J. Castillo-Mateo, Ana C. Cebrián, and Jesús Asín. RecordTest: An R package to analyze non-stationarity in the extremes based on record-breaking events. *Journal of Statistical Software*, 106(5):1–28, 2023. [13](#)
- [5] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001. [3](#), [5](#), [10](#), [11](#), [12](#), [14](#), [15](#), [16](#)
- [6] M.I. Gomes, M. I. Fraga Alves, and C. Neves. *Análise de Valores Extremos: Uma introdução*. Sociedade Portuguesa de Estatística, 2013. [5](#)
- [7] M. G. Kendall. *Rank Correlation Methods*. Charles Griffin, London, 4th edition, 1975. [13](#)
- [8] Hanbeen Kim, Sooyoung Kim, Hongjoon Shin, and Jun-Haeng Heo. Appropriate model selection methods for nonstationary generalized extreme value models. *Journal of Hydrology*, 547:557–574, 2017. [16](#), [17](#), [49](#), [65](#)
- [9] D. Lettenmaier, E. Wood, and J. Wallis. Hydro-climatological trends in the continental united states. *Journal of Climate*, 547:586–607, 1994. [13](#)

- [10] H. B. Mann. Nonparametric tests against trend. *Econometrica*, 13(3):245–259, 1945. [13](#)
- [11] D. Pestana and S. Velosa. *Introdução a Probabilidade e a Estatística*. Fundação Calouste Gulbenkian, 2010. [4](#)
- [12] T. Pohlert. *trend: Non-Parametric Trend Tests and Change-Point Detection*, 2023. R package version 1.1.6. [13](#)
- [13] Y. F. Sang, Z. Wang, and C. Liu. Comparison of the mk test and emd method for trend identification in hydrological time series. *Journal of Hydrology*, 510:293–298, 2014. [13](#)
- [14] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. [17](#)
- [15] F. Serinaldi, C. G. Kilsby, and F. Lombardo. Untenable nonstationarity: an assessment of the fitness for purpose of trend tests in hydrology. *Advances in Water Resources*, 111:132–155, 2018. [13](#)
- [16] A. Wald and J. Wolfowitz. An exact test for randomness in the non-parametric case based on serial correlation. *Annals of Mathematical Statistics*, 14(4):378–388, 1943. [13](#)
- [17] F. Wang, W. Shao, H. Yu, G. Han, X. He, D. Zhang, M. Ren, and G. Wang. Re-evaluation of the power of the mann-kendall test for detecting monotonic trends in hydrometeorological time series. *frontiers in Earth Science*, 8(14), 2020. [13](#)
- [18] S. Yue, P. Pilon, B. Phinney, and G. Cavadias. The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Process*, 16:1807–1829, 2002. [13](#)