Universidade de Évora - Escola de Ciências e Tecnologia

Mestrado em Modelação Estatística e Análise de Dados

Dissertação

Modelação estatística das temperaturas máximas na região do Alentejo

Ana Rita Vaz Ambrosio

Orientador(es) | Lígia Henriques-Rodrigues

Évora 2025

Universidade de Évora - Escola de Ciências e Tecnologia

Mestrado em Modelação Estatística e Análise de Dados

Dissertação

Modelação estatística das temperaturas máximas na região do Alentejo

Ana Rita Vaz Ambrosio

Orientador(es) | Lígia Henriques-Rodrigues

Évora 2025

A dissertação foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor da Escola de Ciências e Tecnologia:

Presidente	Anabela Afonso (Universidade de Évora)
Vogais	Ediclê Duarte (Universidade de Évora) (Arguente) Lígia Henriques-Rodrigues (Universidade de Évora)

Évora 2025

AGRADECIMENTOS

Agradeço primeiramente à minha querida orientadora Professora Lígia, por me ter dado todas as ferramentas necessárias para consolidar todos os conhecimentos indispensáveis à realização da dissertação, esse apoio incondicional e acima de tudo a exigência, foram determinantes para fomentar o meu gosto pela área de extremos.

Agradeço à Professora Maria João Costa (Professora Catedrática do Departamento de Física da Universidade de Évora) por ter disponibilizado todos os dados essenciais para a realização deste trabalho, agradeço também aos Professores do Mestrado em Modelação Estatística e Análise de Dados, nomeadamente a Professora Anabela Afonso, por todo o tempo disponibilizado.

Por fim, agradeço à minha família e amigos, a todos aqueles que me rodeiam com carinho e compreensão, que fazem parte dos meus alicerces, esses que são indispensáveis para todos os desafios e para todas as etapas da minha vida.

“Tell me and I forget. Teach me and I remember.

Involve me and I learn.”

(Benjamin Franklin).

Resumo

Modelação Estatística das Temperaturas Máximas na Região do Alentejo

O clima da Terra tem sofrido mudanças rápidas e sem precedentes desde 1880, com o aumento das temperaturas médias anuais. Em 2022, foi registado um dos anos mais quentes, e a tendência global mostra uma redução nos dias frios e aumento nos dias quentes, principalmente devido às emissões de gases de efeito estufa, intensificadas pelo crescimento económico e populacional. Esses fatores contribuem para eventos climáticos extremos, como ondas de calor e secas, afetando ecossistemas, produção de alimentos e a saúde pública, com maior impacto em populações vulneráveis. A Teoria dos Valores Extremos (TVE), é amplamente usada para prever fenómenos climáticos extremos. Estudos em Portugal aplicaram a TVE para modelar riscos climáticos, incluindo o Alentejo, onde o aumento das temperaturas é evidente. Este estudo utilizou a TVE univariada para modelar as temperaturas máximas, onde se mostra que os verões no Alentejo estão a ficar mais quentes ao longo dos anos.

Palavas chave: Estatística de extremos; estimação paramétrica univariada; quantis extremais, alterações climáticas; temperaturas máximas.

Abstract

Statistical Modelling of Maximum Temperatures in Alentejo

The Earth's climate has experienced rapid, unprecedented changes. Since 1880, the average annual temperature has been rising, with 2022 being one of the hottest years on record, where the cold days and nights are decreasing while hot ones are increasing globally. This rise in temperatures is linked to greenhouse gas emissions, driven by economic and population growth, leading to more extreme weather events like heatwaves and droughts. These events significantly affect ecosystems, food and water production, and public health, especially for vulnerable populations.

The Theory of Extreme Values (EVT) is essential for predicting rare climate phenomena. Several studies in Portugal have used EVT to model climate risks. In the Alentejo region, EVT was applied to model maximum temperatures, given the expected temperature rise. This study used univariate EVT, revealing a gradual increase in maximum temperatures, confirming that summers in Alentejo are becoming progressively hotter over the years.

Keywords: Extreme statistics; univariate parametric estimation; extreme quantiles; climate change; maximum temperatures.

ÍNDICE

ÍNDICE DE FIGURAS

ÍNDICE DE TABELAS

Abreviaturas & Acrónimos

a.a.: Amostra aleatória

AIC: Akaike Information Criterion

BIC: Bayesian Information Criterion

e.o.: Estatística ordinal ou Estatística de ordem

EVI: Índice de Valores Extremos (do inglês, *Extreme Value Index*)

TVE: Teoria de Valores Extremos (do inglês, *Extreme Value Theory*)

f.d.: Função de distribuição

f.d.p.: Função densidade de probabilidade

GEV: Generalizada de Valores Extremos (do inglês, *Generalized Extreme Value*)

g.l.: Graus de Liberdade

GP: Generalizada de Pareto

IC: Intervalo de confiança

i.i.d.: Independente e identicamente distribuída

LAN: *Locally Asymptotically Normal*

log L: log-verosimilhança

ME-plot: Gráfico da função de excesso médio (do inglês, *Mean Excess plot*)

MMA: Método dos Máximos Anuais

MMO: Método das Maiores Observações

MO: Maiores Observações

MV: Máxima Verosimilhança

POT: Excessos acima do *threshold* (do inglês, *Peaks Over Threshold*)

RV: Razão de Verosimilhanças

TRV: Teste de Razão de Verosimilhanças

v.a.: Variável aleatória

1

Introdução

O clima da terra nunca sofreu alterações tão rápidas como atualmente (Fourth National Climate Assessment, 2018). A temperatura média anual tem vindo a aumentar a uma taxa média de 0.08°C por década, e 2022 foi o sexto ano mais quente, desde 1880 (Malakouti, 2023; Blunden et al., 2023). Segundo a *Intergovernmental Panel on Climate Change* (IPCC), o número de dias e noites frios têm vindo a diminuir e o número de dias e noites quentes têm aumentado numa escala global (Intergovernmental Panel on Climate Change, 2015).

Desde 1950, é observado um crescente aumento de incidentes climáticos e meteorológicos extremos, tais como, ondas de calor, secas, inundações, ciclones e incêndios florestais (Intergovernmental Panel on Climate Change, 2015; Costa et al., 2012; Carvalho et al., 2014; Fraga et al., 2017; Pereira et al., 2021;Cardoso et al., 2023; Green, 2024 e Calvin et al., 2023).

Portugal está entre os países europeus mais severamente afetados por eventos climáticos extremos, como ondas de calor, secas e incêndios florestais. Relatórios recentes da Agência Europeia do Ambiente (European Environment Agency - EEA) indicam que está entre os países com mais perdas humanas devido a eventos extremos ao longo das últimas quatro décadas (Gagliardi et al., 2022).

Este tipo de eventos extremos, tem implicações graves, pois conduzem a alterações dos ecossistemas, ao aumento da mortalidade, a impactos ambientais significativos, entre outros (Intergovernmental Panel on Climate Change, 2015; Blunden et al., 2023).

Na tentativa de modelar e prever esses eventos extremos, a Teoria dos Valores Extremos (*Extreme Value Theory* - TVE), tornou-se cada vez mais relevante, sendo uma ferramenta estatística fundamental no estudo de fenómenos climáticos extremos, ajudando a entender e prever eventos raros, mas de grande impacto, como ondas de calor, tempestades intensas e secas prolongadas (Gilleland & Katz, 2006; Cooley, 2009; Katz, 2010; Towler et al., 2010). Desde que foi introduzida por (Fisher & Tippett, 1927) a TVE tem sido amplamente aplicada em diversas áreas, incluindo hidrologia, finanças e climatologia (Coles, 2001; Salvadori et al., 2007; Katz, 2010; Mudelsee, 2020; Schokker et al., 2022).
A TVE não permite apenas uma melhor compreensão dos riscos associados a esses eventos, mas também fornece suporte essencial para a formulação de políticas de adaptação e mitigação (Intergovernmental Panel on Climate Change, 2015; Fourth National Climate Assessment, 2018; Blunden et al., 2023).

Neste estudo, foram aplicadas metodologias de modelação estatística baseadas na Teoria de Valores Extremos univariada, em temperaturas máximas do Alentejo. Isto é particularmente importante nesta região, onde se prevê um aumento significativo das temperaturas nos próximos anos (Santos et al., 2016; Fraga et al., 2017). A capacidade da TVE de lidar com tais complexidades, faz dela uma ferramenta poderosa na análise das consequências futuras das alterações climáticas (Salvadori et al., 2007).
Nos últimos anos, diversos estudos em Portugal têm utilizado a TVE para modelar temperaturas, precipitação e outros fenómenos climáticos, focando-se na avaliação de riscos climáticos, como ondas de calor e secas prolongadas (Costa et al., 2012; Andrade et al., 2014), referindo alguns casos de estudo, sobre caracterização das secas em Portugal e no Alentejo, através do estudo dos índices de seca baseados na TVE para modelar a severidade e frequência das secas, permitindo criar padrões de escassez hídrica na região (Paulo & Pereira, 2006), e estudos sobre como prever o impacto das temperaturas extremas na produção de vinho na Europa, com um foco particular no Alentejo, através da metodologia TVE (Fraga et al., 2012).

O objetivo geral deste trabalho consiste então, em caracterizar a evolução das temperaturas máximas no Alentejo de forma a perceber se, de facto, as temperaturas máximas observadas estão a aumentar e se, os verões alentejanos estão a ficar mais quentes.

Os objetivos específicos deste trabalho consistem na aplicação das metodologias paramétricas de Estatística de Extremos univariada para modelar as temperaturas máximas, utilizando o método dos máximos anuais, das maiores excedências e das k-maiores observações. Para cada um dos métodos e para cada uma das componentes em estudo serão apresentadas uma análise descritiva preliminar, seguida de uma análise inferencial e de uma seleção do melhor modelo. Com base neste modelo serão então obtidas estimativas de quantidades de interesse como a probabilidade de excedência de níveis elevados, níveis de retorno entre outros. De seguida, e com base nos resultados obtidos, construímos gráficos adequados para as quantidades de interesse (Mudelsee, 2020), de forma a ilustrar a evolução da temperatura máxima, e tentando prever o comportamento futuro da mesma. Como os fenómenos ambientais são, em geral, não estacionários, consideramos também a modelação adequada da não estacionariedade entrando em linha de conta com um efeito temporal que poderá afetar o comportamento das temperaturas máximas.

Este trabalho está dividido em 5 capítulos: o capítulo 2 aborda a Teoria dos Valores Extremos univariada, no capítulo 3 são descritos os vários modelos paramétricos utilizados neste trabalho, o capítulo 4 é relativo ao caso de estudo das temperaturas máximas do Alentejo. Por fim, no capítulo 5, são discutidas as principais conclusões obtidas e futuras considerações.

2

Teoria de Valores Extremos

A Teoria de Valores Extremos (TVE) é um ramo da estatística e da teoria das probabilidades que lida com desvios extremos da mediana em distribuições de probabilidades, onde a ordenação da amostra é primordial. O principal objetivo da TVE, é compreender e modelar a probabilidade de ocorrência de eventos raros, os mais extremos num conjunto de dados (Gomes et al., 2013; Haan & Ferreira, 2006).

Seja $X_n \coloneqq (X_1, \dots, X_n)$ uma amostra aleatória (a.a.), de n variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.) com função de distribuição (f.d.) comum $F(X)$, e considere-se a amostra de estatísticas ordinais $X_{1:n} \leq \dots \leq X_{n:n}$, onde $X_{i:n}$ é a i - ésima estatística ordinal (e.o.) ascendente, $1 \leq i \leq n$. Neste contexto, $X_{1:n}$ e $X_{n:n}$ são, respetivamente, o mínimo e o máximo da amostra, também designadas de extremos (Castillo et al., 2005; Coles, 2001) i.e.,

$$X_{n:n} = M_n = \max\{X_1, \dots, X_n\} \tag{2.1}$$

$$X_{1:n} = m_n = \min\{X_1, \dots, X_n\}. \tag{2.2}$$

Uma das primeiras aplicações da análise estatística de valores extremos consistiu em determinar a altura de diques, tendo em conta a possibilidade de inundação, em média uma vez a cada cem anos. Atualmente tem aplicação em variadíssimas situações e domínios, tais como, precipitação, velocidade do vento, na indústria, em crises financeiras, em seguros para eventos catastróficos e em ciências ambientais como concentração de ozono no ar, entre outras aplicações (Charras-Garrido & Lezaud, 2013).

2.1 DISTRIBUIÇÕES LIMITE DO MÁXIMO

A função de distribuição do M_n (desconhecida) pode ser derivada exatamente para todos os valores de n como,

$$\begin{aligned} F_{M_n}(x) &= P[M_n \leq x] = P\{X_1 \leq x, \dots, X_n \leq x\} \\ &= \prod_{i=1}^{n} P[X_i \leq x] = (P[X_i \leq x])^n \\ &= \prod_{i=1}^{n} F(x) = [F(x)]^n \\ &= F^n(x), x \in \mathbb{R}, \quad n \in \mathbb{N}. \end{aligned} \tag{2.3}$$

Se o comportamento distribucional exato de X_i fosse conhecido, então o correspondente comportamento de M_n poderia ser obtido de forma exata. Porém, na prática a distribuição de X_i é desconhecida, dificultando a determinação da distribuição de M_n.

Diz-se que X_n converge em distribuição para X, quando $n \to \infty$, e usa-se a notação $X_n \xrightarrow[n\to\infty]{d} X$, se $\lim_{n \to \infty} F_n(x) = F(x)$, para todo ponto de continuidade, x, de $F(x)$. Para além disso, o limite superior do suporte de F, x^F, é definido por $x^F \coloneqq sup\{x: F(x) < 1\}$.
Deste modo, quando $n \to \infty$, a função de distribuição do máximo converge para uma distribuição degenerada, i.e.,

$$F^n(x) \to \begin{cases} 0, & se\ F(x) < 1, \\ 1, & se\ F(x) = 1. \end{cases} \tag{2.4}$$

Pelo que $M_n \xrightarrow[n\to\infty]{d} x^F$. A convergência em distribuição anterior implica, a convergência em probabilidade para a mesma constante, $M_n \xrightarrow[n\to\infty]{p} x^F$, mesmo que $x^F = \infty$. A distribuição assintótica de M_n é então degenerada, assumindo apenas valores de 0 e 1 (Castillo et al., 2005; Gomes et al., 2013).

Uma vez que, a distribuição limite de M_n é degenerada é necessário recorrer a uma normalização (transformação linear) de modo a obter uma lei limite não-degenerada, de forma semelhante ao que se passa no caso da teoria assintótica para somas ou médias, através do Teorema do Limite Central.

Considerando sequências de constantes normalizadoras a_n e b_n, com $a_n > 0$ e $b_n \in \mathbb{R}$, e definindo ${M_n}^*$ como,

$$ {M_n}^* = \frac{M_n - b_n}{a_n}, \tag{2.5} $$

obtemos uma sequência de máximos linearmente normalizados, onde a_n e b_n estabilizam a escala e a localização, respetivamente (Coles, 2001). As possíveis distribuições limite de ${M_n}^*$, i.e., quando $n \to \infty$, são dadas pelo Teorema de Tipos Extremais (Gnedenko, 1943).

Admitindo que, existem sucessões de constantes $\{a_n\}_{n\geq 1}(a_n > 0)$, $\{b_n\}_{n\geq 1}$e uma f.d. não degenerada G, tal que,

$$ \lim_{n\to\infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x). \tag{2.6} $$

Então, G pertence a um dos três tipos de distribuições de Valores Extremos descritos abaixo, onde $\gamma > 0$ representa o parâmetro de forma da distribuição,

Tipo I $$ Gumbel: G(x) = exp\left\{- exp\left[-\left(\frac{x-b}{a}\right)\right]\right\} \quad -\infty < x < \infty. \tag{2.7} $$

Tipo II $$ Fréchet: G(x|\gamma) = \begin{cases} 0, & x \leq b, \\ exp\left\{-\left(\frac{x-b}{a}\right)^{-\gamma}\right\}, & x > b. \end{cases} \tag{2.8} $$

Tipo III $$ Max - Weibull: G(x|\gamma) = \begin{cases} exp\left\{-\left[-\left(\frac{x-b}{a}\right)^{\gamma}\right]\right\}, & x < b, \\ 1, & x \geq b. \end{cases} \tag{2.9} $$

Cada família possui um parâmetro de localização (b) e escala (a), adicionalmente as distribuições de Fréchet e Weibull possuem o parâmetro γ de forma (Coles, 2001).
Isto implica que, sempre que M_n for estabilizada com sucessões de constantes $\{a_n\}$ e $\{b_n\}$, a variável correspondente normalizada $M_n{}^*$ possui uma distribuição limite que apenas pode ser uma dessas três distribuições de Valores Extremos, independentemente da distribuição F da população, sendo neste sentido a semelhança ao Teorema do Limite Central (Coles, 2001).

Na Tabela 1, são apresentadas as f.d.'s e as correspondentes funções quantil, denotando o parâmetro de localização por λ , o parâmetro forma por γ e o parâmetro de escala por δ.

Tabela 1 - Algumas características das distribuições de Valores Extremos (Castillo et al., 2005).

FUNÇÃO	GUMBEL	FRÉCHET	MAX- WEIBULL
DISTRIBUIÇÃO	$exp\left[-exp\left(\frac{\lambda - x}{\delta}\right)\right]$	$exp\left[-\left(\frac{\delta}{x-\lambda}\right)^{\gamma}\right]$	$exp\left[-\left(\frac{\lambda - x}{\delta}\right)^{\gamma}\right]$
QUANTIL	$\lambda - \delta\ (\log(-\log p))$	$\lambda + \delta\ (-\log p)^{-1/\gamma}$	$\lambda - \delta(-\log p)^{1/\gamma}$

Estas distribuições também são denominadas de max-estáveis, uma vez que, existem constantes reais $A_\mathrm{k} > 0$ e $\mathrm{B_k}$ (para todo o k) tais que,

$$G^k(x) = G(A_k x + B_k). \tag{2.10}$$

2.2 DISTRIBUIÇÃO GENERALIZADA DE VALORES EXTREMOS

As três distribuições limite mencionadas da secção 2.1, têm comportamentos distintos, nomeadamente diferentes comportamentos de cauda para a distribuição $\mathrm{F}(x)$.

Von Mises (1936; 1954) e Jenkinson (1955) combinaram as três famílias, Gumbel, Fréchet e Max-Weibull, dando origem à f.d. geral de valores extremos (GEV, do inglês "*generalized extreme value*"), tendo a seguinte f. d.,

$$G_\gamma(x;\lambda,\delta)=\begin{cases} exp\left\{-\left[1+\gamma\left(\frac{x-\lambda}{\delta}\right)\right]^{-1/\gamma}\right\}, & \gamma\neq 0,\\ exp\left\{-exp\left(\frac{\lambda-x}{\delta}\right)\right\}, & \gamma=0.\end{cases} \tag{2.11}$$

A distribuição GEV possui três parâmetros (λ, δ, γ), localização, escala e forma, respetivamente, que tomam valores em ($\mathbb{R}, \mathbb{R}^+, \mathbb{R}$), (Castillo et al., 2005; Coles, 2001; Gomes et al., 2013).

O parâmetro de forma, γ, é designado de índice de valores extremos ou índice de cauda (EVI, do inglês "*extreme value index*"), e é o parâmetro fundamental em estatística de extremos, estando diretamente relacionado com o peso da cauda da distribuição, $\bar{F}(x) := 1 - F(x)$, representado na Figura 1, onde se pode observar que $\bar{F}(x)$ decai para zero quando $x \to x^F \coloneqq sup\{x: F(x) < 1\}$.

A unificação das três famílias de distribuição de valores extremos, numa distribuição única, veio simplificar a identificação do modelo limite. A determinação de estimativas pontuais, intervalares e a realização de testes de hipóteses para parâmetro EVI permite determinar o tipo apropriado de comportamento de cauda, não sendo necessário realizar avaliações subjetivas.

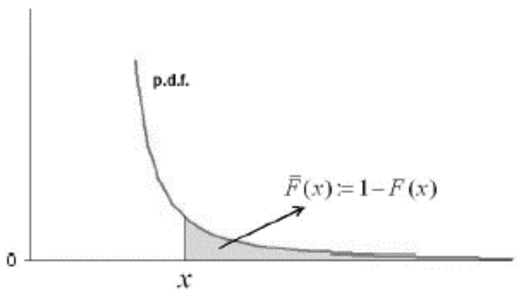


Figura 1 - Cauda direita da distribuição F(x), representada pela área assinalada a cinzento (Gomes et al., 2013).

Na Tabela 2, estão representadas as famílias de distribuições padrão (com parâmetro de localização igual a 0 e parâmetro de escala igual a 1), em função do sinal do EVI.

Tabela 2 - Influência do parâmetro EVI na definição do máx-domínio de atração.

$\boldsymbol{\gamma = 0}$	$\boldsymbol{G_0(x) = exp[-exp(-x)],\ x} \in \mathbb{R}.$	**GUMBEL**
$\boldsymbol{\gamma > 0}$	$G_\gamma(x) = \begin{cases} 0, & x < -1/\gamma, \\ exp\{-[1+\gamma x]^{-1/\gamma}\}, & x \geq -1/\gamma. \end{cases}$	**FRÉCHET**
$\boldsymbol{\gamma < 0}$	$G_\gamma(x) = \begin{cases} exp\{-[1+\gamma x]^{-1/\gamma}\}, & x \leq -1/\gamma, \\ 1, & x > 1/\gamma \end{cases}$	**MAX-WEIBULL**

Consoante o valor do parâmetro forma, é possível ter uma noção da forma da cauda da distribuição, como já foi referido, à medida que nos aproximamos do x^F, $\bar{F}(x)$ decai para zero, havendo assim, três tipos possíveis de cauda (Figura 2),

- Máx-Weibull: se $\gamma < 0$, modelo diz-se de cauda curta com limite superior do suporte finito, $x^F < \infty$. As distribuições Uniforme, Beta, Reversed Burr e Max-Weibull, são alguns exemplos que se incluem nesta família.
- Gumbel: se γ=0, modelo diz-se de cauda leve (a cauda apresenta um comportamento exponencial), $\mathrm{x^F} < \infty \text{ ou } \mathrm{x^F} = \infty$. As distribuições Exponencial, Weibull (de mínimos), Logística, Gumbel, Normal, Log-Normal, Gama e Min-Fréchet, são alguns exemplos que se incluem nesta família.
- Fréchet: se $\gamma > 0$, modelo diz-se de cauda pesada (a cauda apresenta um comportamento polinomial ou de tipo Pareto)~~,~~ $\mathrm{x^F} = \infty$. As distribuições Pareto, Generalizada de Pareto, Burr, Fréchet, t-Student, Cauchy e Log-Gama, são alguns exemplos que se incluem nesta família.

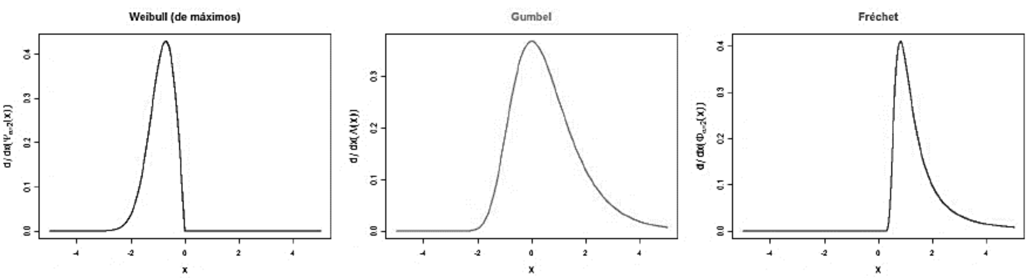


Figura 2 - Funções densidade de probabilidade das distribuições padrão: Max-Weibull ($\alpha = 2$), com cauda curta (esquerda), Gumbel com suporte no eixo real (centro), e Fréchet ($\alpha = 2$), com cauda pesada (direita)(Gomes et al., 2013).

2.3 Distribuição limite das excedências

Temos vindo a abordar a teoria de distribuições limite do máximo, no entanto ao longo dos anos houve a necessidade de analisar v.a.'s que representem os excessos acima de um determinado nível. As observações que excedem um determinado nível extremo, designado de *threshold* (u), são chamadas de excedências acima de um nível $(x_i : x_i > u)$ (Castillo et al., 2005). Seja $X_1, \dots, X_n$ uma sequência de variáveis aleatórias i.i.d. e considerando um nível u elevado (fixo), a f.d. dos excessos F_u, é dada por, $F_u(x) = P(X - u \leq x | X > u) = P(X \leq u + x | X > u), 0 \leq x \leq x^F - u$,

$$F_u(x) = \frac{F(u+x) - F(u)}{1 - F(u)}. \tag{2.12}$$

Pickands (1975) demonstrou que, quando u tende para o limite superior de suporte de uma variável aleatória, x^F, as excedências acima de um *threshold* u fixo, Y:= $X - u$ seguem uma distribuição Generalizada de Pareto (GP) (Castillo et al., 2005),

$$H(Y; \delta, \gamma) = \begin{cases} 1 - \left(1 + \gamma \frac{y}{\delta}\right)^{-1/\gamma}, & y \in (0, \infty), \quad se\ \gamma > 0, \\ 1 - exp\left(-\frac{y}{\delta}\right), & y\ (0, \infty), \quad se\ \gamma = 0, \\ 1 - \left(1 + \gamma \frac{y}{\delta}\right)^{-1/\gamma}, & y \in \left(0, -\frac{\delta}{\gamma}\right), \quad se\ \gamma < 0. \end{cases} \tag{2.13}$$

Pickands (1975) e Balkema and de Haan (1974) estabeleceram a dualidade entre a GEV(γ) e a GP(γ), propondo para níveis elevados u a aproximação: $F_u(x) \approx G_\gamma(x; \delta_u)$, ou seja, se o máximo converge para uma GEV(γ) e se X segue distribuição GP (δ, γ), então $X - u$, sendo $X > u$, também segue uma distribuição GP $(\delta - \gamma_u, \gamma)$, para qualquer valor de u. Esta propriedade permite que, se um conjunto de dados e um *threshold* u_0 são consistentes para esta lei, então qualquer outro *threshold* $u_1 > u_0$, também o será (Castillo et al., 2005). Alguns casos especiais da distribuição GP (Castillo et al., 2005),

- Se $\gamma = 0$, a distribuição GP, reduz-se à distribuição exponencial,
- Se $\gamma = -1$, a distribuição GP, reduz-se à distribuição uniforme $U[0, \delta]$,
- Se $\gamma \geq 1/2$, $Var(X) = \infty$,
- Se $\gamma > 0$, a distribuição GP reduz-se à distribuição de um modelo Pareto,
- Se $\gamma < 0$, a distribuição GP reduz-se à distribuição de um modelo Beta.

2.4 TÉCNICAS GRÁFICAS PRELIMINARES

Existem diversas ferramentas gráficas que ajudam na visualização de uma distribuição e a ter noção na forma da cauda, sendo um complemento aos procedimentos inferenciais para a seleção do modelo max-estável, os mais usados são os histogramas, os *boxplots*, os QQ-plot (do inglês "*Quantile-Quantile plot*"), os ME-plot (do inglês "*Mean Excess plot*") e os TC-plot, do inglês "*Threshold Choice plot*").

Os QQ-plot permitem avaliar a adequação de um modelo de extremos de forma mais precisa e ajudam também na validação empírica desse mesmo modelo, pois mostram diretamente a relação entre os quantis teóricos Q(p) e os quantis teóricos observados do modelo standard de uma determinada família (como a distribuição de Gumbel, Fréchet ou Max-Weibull), Qs(p). Se o padrão de pontos do QQ-plot exibir uma tendência linear, indica que o modelo de ajustado é adequado. Desvios significativos nas caudas do QQ-plot, significam que o modelo pode estar a subestimar ou a superestimar, não permitindo ter previsões fiáveis dos eventos raros.

Para a distribuição Gumbel, tendo em conta a f.d. do modelo Gumbel, apresentada na equação (2.7), é possível verificar que os quantis do modelo Gumbel, $Q_{\lambda,\delta}$, estão relacionados com os quantis do modelo Gumbel padrão $Q_{0,1}$, através da seguinte equação:

$$Q_{\lambda,\delta}(p) = \lambda + \delta\,.\,(-\log(-\log p)) = \lambda + \delta\,.\,Q_{0,1}(p), \qquad p \in (0,1), \lambda \in \mathbb{R}, \delta > 0 \qquad (2.14)$$

Logo as coordenadas do QQ-plot são, $\big((-\log(-\log p_i)), y_{i:n}\big), i = 1, \dots, n$ (em que a localização (λ) é a ordenada na origem e a escala (δ) é o declive da reta).

Para a distribuição GEV, tendo em conta a f.d. do modelo GEV, apresentada na equação (2.11), é possível verificar que os quantis do modelo GEV, $Q_{\gamma,\lambda,\delta}$, estão relacionados com os quantis do modelo GEV padrão, $Q_{\gamma,0,1}$,

$$Q_{\gamma,\lambda,\delta}(p) = \lambda + \delta\,.\frac{(-\log(\mathrm{p}))^{-\gamma} - 1}{\gamma} = \lambda + \delta\,.\,Q_{\gamma,0,1}(p), \qquad p \in (0,1), \lambda, \gamma \in \mathbb{R}, \delta > 0 \qquad (2.15)$$

Logo as coordenadas do QQ-plot são, $\left(\frac{(-\log(\mathrm{p}_i))^{-\gamma}-1}{\gamma}, y_{i:n}\right), i = 1, \dots, n$ (em que a localização (λ) é a ordenada na origem e a escala (δ) é o declive da reta). Para a construção do QQ-plot do modelo GEV é necessário obter uma estimativa preliminar do parâmetro de forma, γ.

Para a distribuição Generalizada de Pareto, tendo em conta a f.d. da GP descrita na equação (2.13), observa-se que os quantis do modelo GP, $Q_{\gamma,\delta}$, estão relacionados com os quantis do modelo GP padrão, $Q_{\gamma,1}$, da seguinte forma,

$$Q_{\gamma,\delta}(p) = \delta\frac{(1-p)^{-\gamma}-1}{\gamma} = \delta.Q_{0,1}(p), \qquad p \in (0,1), \gamma \in \mathbb{R}, \delta > 0 \tag{2.16}$$

Logo as coordenadas do QQ-plot são, $\left(\frac{(1-p)^{-\gamma}-1}{\gamma}, y_{i:n}\right), i = 1, \dots, n$ (com ordenada na origem zero e em que a escala (δ) é o declive da reta). Para a construção do QQ-plot do modelo GP é necessário obter uma estimativa preliminar do parâmetro de forma, γ.

Assim sendo, se a distribuição for bem ajustada, é esperado que a nuvem de pontos se distribua ao longo de uma reta, mostrando que existe uma relação linear entre os quantis teóricos Q(p) e Qs(p).

Os ME-plot, são úteis para escolher um limiar adequado na abordagem POT (do inglês, '*peaks over threshold*') ou no modelo Paretiano de Excessos. Sendo, Y uma v.a. com distribuição GP com parâmetros δ e γ (Davison & Smith, 1990), então a função de excesso médio $\mathrm{e}(u)$ é definida como,

$$\mathrm{E}(Y) = \frac{\delta}{1-\gamma},\ \mathrm{e}(u) \coloneqq E(X-u|X>u) = E(Y|Y>0) = \frac{\delta+\gamma u}{1-\gamma},\ \gamma < 1. \tag{2.17}$$

Portanto $\mathrm{e}(u)$ é uma função linear de u, na prática $\mathrm{e}(u)$ é substituída pela sua contrapartida empírica, $\hat{\mathrm{e}}(u)$, obtida com base na amostra de dados observados $X_1, \dots, X_n$, $\hat{\mathrm{e}}_n(t) = \frac{\sum_{i=1}^{n} x_i I_{(u,\infty)}(x_i)}{\sum_{i=1}^{n} I_{(u,\infty)}(x_i)} - u$, $I_{(u,\infty)}(x_i) = 1, se\ \ x_i > u$ ou 0, caso contrário.

Escolhendo, $\mathrm{u} = \mathrm{x}_{\mathrm{n-k:n}}, \mathrm{k} = 1, 2, \dots, \mathrm{n} - 1, \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i} \mathrm{I}_{(\mathrm{u},\infty)}(\mathrm{x_i}) = \sum_{\mathrm{j=1}}^{\mathrm{k}} \mathrm{x}_{\mathrm{n-j+1:n}}$, com $x_i > u$, e as estimativas dos excessos médios são dadas por, $\hat{e}_{k,n} \coloneqq \hat{e}_n(\mathrm{x}_{\mathrm{n-k:n}}) = \frac{1}{k} \sum_{\mathrm{j=1}}^{\mathrm{k}} \mathrm{x}_{\mathrm{n-j+1:n}} - (\mathrm{x}_{\mathrm{n-k:n}})$.
A função de excesso médio do modelo GP é linear em u, assim, quando se representa o ME-plot empírico, o *threshold* u deverá ser escolhido no ponto a partir do qual se verifica a linearidade à direita de $\mathrm{x}_{\mathrm{n-k:n}}$ (podendo esse ponto ser ou não um valor da amostra), $\{\mathrm{x}_{\mathrm{n-k:n}}, \hat{e}_n(\mathrm{x}_{\mathrm{n-k:n}}), 1 \leq k \leq n - 1\}$. Na parte final do ME-plot, essa linearidade não acontece porque se calculam médias de um número pequeno de excessos elevados. Deste modo, e para determinação do u esses pontos podem ser ignorados. A forma da função $\mathrm{e}(u)$ dá indicação se o modelo tem caudas mais pesadas que a da exponencial (HTE, do inglês "*heavier than exponential*") ou mais leves que a da exponencial (LTE, do inglês "*lighter than exponential*"): se X tiver cauda HTE: a função $\mathrm{e}(u)$ é monótona crescente para valores elevados de u e se X tiver cauda LTE: a função $\mathrm{e}(u)$ é monótona decrescente para valores elevados de u.

Admitindo que o modelo GP (δ, γ), é correto para modelar os excessos acima de u_0, então para qualquer *threshold* $u > u_0$, os novos excessos continuam a ter distribuição GP com o mesmo parâmetro de forma, γ e com a escala alterada para (Coles, 2001), $\delta_u = \delta_{u_0} + \gamma(u - u_0)$, ou seja, o parâmetro de escala varia com u, a menos que $\gamma = 0$. Esta dificuldade pode ser ultrapassada através da reparametrização deste parâmetro, $\delta^* = \delta_u - \gamma_u$. Deste modo, os parâmetros do modelo λ^*e γ são constantes para qualquer *threshold* u acima de u_0. Este comportamento está na base da construção dos gráficos de estabilidade dos parâmetros para seleção do *threshold, os* TC-plot.

De maneira a fazer um estudo preliminar do sinal do parâmetro EVI (γ) e consequentemente do modelo mais apropriado para um bom ajustamento aos dados, é necessário seguir algumas diretrizes: usar os métodos gráficos, como o QQ-plots e ME-plots; se a cauda à direita mostrar uma tendência linear, então $\gamma = 0$, e o modelo pertence à família Gumbel; se a cauda de interesse apresentar uma assimptota vertical, então $\gamma < 0$, e o modelo pertence à família Weibull; se a cauda de interesse apresentar uma assimptota horizontal, então $\gamma > 0$, e o modelo pertence à família Fréchet. Em caso de dúvida, usar sempre o modelo com a cauda mais pesada (o modelo Fréchet tem a cauda mais pesada do que o Gumbel, que tem caudas mais pesadas quando comparada com o modelo Weibull). Por outro lado, o modelo Gumbel é o mais simples, pois possui menos um parâmetro relativamente aos outros dois modelos, sendo assim, o modelo Gumbel, deve ser usado devido à sua simplicidade. Na Figura 3, é exibido o comportamento do QQ-plots e ME-plots, para modelos com caudas HTE, caudas Exponencial e caudas LTE.

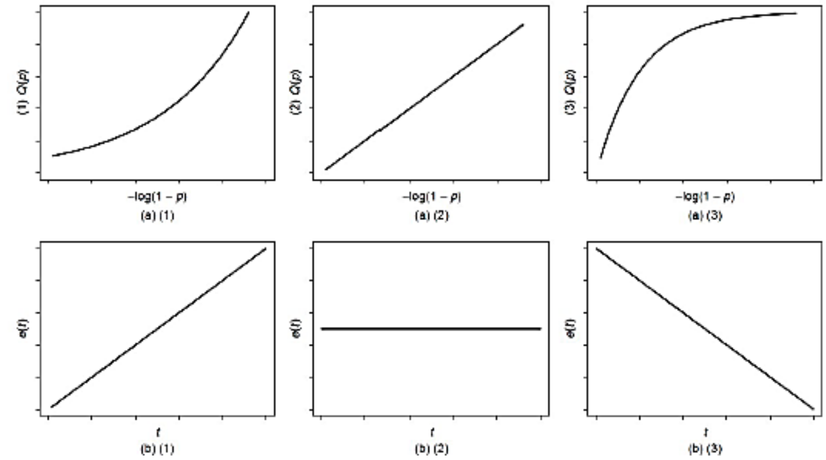


Figura 3 - (a) QQ-plot exponencial; (b) ME-plot, para exemplos de distribuições com caudas tipo: 1 - HTE, 2 - Exponencial, e 3- LTE (Beirlant, 2004).

3

Estatística de Extremos: Abordagem Paramétrica

Como foi referido foi referido no capítulo anterior, a unificação das três famílias de distribuição de valores extremos numa única família, veio simplificar a identificação do modelo subjacente. Deste modo, estimando adequadamente o parâmetro EVI (γ), é possível determinar o modelo apropriado para descrever o comportamento de cauda da distribuição.

Para além da estimação do EVI, é necessário estimar conjuntamente os parâmetros de localização λ, e de escala δ. A estimação destes parâmetros é fundamental na estimação de outros parâmetros de acontecimentos raros como, os níveis e períodos de retorno de um nível elevado, os quantis extremais, a probabilidade de excedência de níveis elevados e o limite superior do suporte (somente no caso de $\gamma < 0$) [ver secções 3.1 e 3.3].
Uma das dificuldades da análise de valores extremos, é a limitação da quantidade de dados, dificultando assim o processo de estimação. O objetivo da modelação estatística consiste em usar a informação da amostra para inferir sobre a distribuição de probabilidade da população da qual os dados surgiram. No caso mais simples, é assumido que o conjunto de dados são realizações independentes da distribuição da população.
A inferência consiste então, na estimação da distribuição da população, existindo duas abordagens distintas: paramétrica e não-paramétrica. Nesta dissertação será utilizada a abordagem paramétrica (Castillo et al., 2005; Gomes et al., 2013; Beirlant, 2004).

De seguida são descritas três abordagens paramétricas distintas em estatística de extremos, nomeadamente: o Método dos Máximos Anuais (MMA) [secção 3.1], o Método das Maiores Observações (MMO) [secção 3.2] e o Método das excedências acima de um nível elevado (POT, do inglês "*Peaks Over Threshold*") [secção3.3].

3.1 MODELO GEV UNIVARIADO OU MÉTODO DOS MÁXIMOS ANUAIS

A primeira abordagem à modelação deriva da distribuição GEV, uma vez que o sinal do EVI, especifica o tipo de cauda (e de distribuição limite): cauda pesada (Fréchet), cauda exponencial (Gumbel) e cauda curta ou leve (Max-Weibull), como mencionado anteriormente.
Esta abordagem também é designada na literatura de modelo dos Máximos anuais ou modelo de Gumbel (Gomes et al., 2013).
Este modelo divide uma série de variáveis i.i.d. $(X_1, X_2 \dots, X_n)$, em m sub-amostras de tamanho k, sendo k de grandes dimensões tal que, $n = m \times k$, ou seja, a série original é dividida em m blocos e em cada bloco de dimensão k, obtém-se o máximo,

$$M_i = max\left(X_{1+(\mathrm{i}-1)\mathrm{k}}, \dots, X_{ik}\right),\ i = 1, .., m.$$

A sucessão de máximos $(M_1, M_2 \dots, M_m)$ converge para uma distribuição GEV, pois são distribuições max-estáveis. Muitas vezes esses blocos correspondem a um período de um determinado ano, sendo k o número de observações nesse ano, e os blocos correspondem ao número de anos em análise (Coles, 2001; Beirlant, 2004; Castillo et al., 2005).
De seguida, é então necessário fazer inferência estatística, para estimação dos parâmetros, de escala, localização e forma. Existem diversas técnicas para se obter as estimativas destes parâmetros, as estimativas iniciais podem ser obtidas utilizando técnicas gráficas baseadas nos QQ-plot, como referido na secção 2.4 e de seguida através dos métodos clássicos de estimação propostos em Castillo et al. (2005), nomeadamente: o método de máxima-verosimilhança (MV) e o método dos momentos ponderados de probabilidade (PWM, do inglês '*probability weighted moments*'). Neste trabalho apenas se irá usar o método MV.

3.1.1 ESTIMADORES DE MÁXIMA VEROSIMILHANÇA

O método de máxima-verosimilhança (MV), é um método baseado na maximização da log-verosimilhança da amostra em análise.

Assumindo que, $(Y_1, \dots, Y_m)$ são variáveis i.i.d. com distribuição GEV (Castillo et al., 2005; Coles, 2001), a equação de log-verosimilhança para $\gamma \neq 0$, é dada por

$$\log L\,(\gamma, \lambda, \delta) = -m \log \delta \; - \; \left(\frac{1}{\gamma} + 1\right) \sum_{\mathrm{i}=1}^{\mathrm{m}} \log(1 + \gamma z_i) - \sum_{\mathrm{i}=1}^{\mathrm{m}} (1 + \gamma z_i)^{-\frac{1}{\gamma}} \;, \tag{3.1}$$

sendo

$$z_i = \frac{x_i - \lambda}{\delta} \tag{3.2}$$

e admitindo que $1 + \gamma z_i > 0, i = 1, \dots, m$, para $\gamma \; = \; 0$, a equação de log-verosimilhança é

$$\log L\,(\gamma, \lambda, \delta) = -m\, log\, \delta \; - \sum_{i=1}^{m} \exp(-z_i) - \sum_{i=1}^{m} z_i \,. \tag{3.3}$$

Apesar dos estimadores de MV apresentarem, em geral, boas propriedades assintóticas, quando aplicados ao modelo GEV, Coles (2001) refere as seguintes limitações:

- Quando $\gamma > -0.5$, os estimadores de MV são regulares, ou seja, são consistentes, assintoticamente eficientes e $\sqrt{m}\left(\left(\hat{\gamma}, \hat{\lambda}, \widehat{\delta}\right) - (\gamma, \lambda, \delta)\right)$ é assintoticamente Normal (Beirlant, 2004);
- Quando $-1 < \gamma < -0.5$, os estimadores MV podem são obtidos, mas não é possível estabelecer as suas propriedades assimptóticas;
- Quando $\gamma < -1$, os estimadores MV dificilmente são obtidos.

3.1.2 OUTROS PARÂMETROS DE INTERESSE DE ACONTECIMENTOS RAROS

Um dos parâmetros cruciais em Estatística de Extremos como já foi referido anteriormente é o EVI, γ. Para além deste, existem outros parâmetros de interesse em EVT, nomeadamente, o limite superior de suporte, já abordado no capítulo anterior, o quantil extremal, o período de retorno, o nível de retorno e a probabilidade de excedência.

Seja F uma f.d. contínua com inversa generalizada, $F^{\leftarrow}(u) = inf\,\{x : F(x) \geq u\}$, onde a correspondente função quantil de cauda é definida como, $U(t) = F^{\leftarrow}\left(1-\frac{1}{t}\right)$, para $t \in [1,\infty]$. A função quantil de cauda é monótona não-decrescente e o seu valor em 1 e em ∞ é, respetivamente, o limite inferior e superior do suporte da distribuição F:

- $U(1) = F^{\leftarrow}(0) = inf\,\{x : F(x) \geq 0\} = x_F$;
- $U(\infty) = F^{\leftarrow}(1) = inf\,\{x : F(x) \geq 1\} = sup\,\{x : F(x) < 1\} = x^F$.

Um quantil extremal é definido como $x_p \coloneqq F^{\leftarrow}(1-p) = U\left(\frac{1}{p}\right)$, com p pequeno, normalmente, $p < 1/n$. Enquanto, o nível de retorno do valor t, sendo $u \coloneqq U(t)$, é dado por $u = U\big(T(u)\big)$, então o período de retorno do nível u, ou seja, o tempo médio entre duas ocorrências do nível u, é definido como, $T(u) = \frac{1}{1-F(u)}$, onde $F(U(t)) = 1 - 1/t$.

Podemos reescrever estes parâmetros de interesse associados ao ajustamento GEV (Gomes et al., 2013). O quantil extremal de probabilidade $1-p$, também pode ser obtido através da função quantil de cauda (basta inverter a equação (2.11)), considerando $t = \frac{1}{p}$,

$$q_{1-p} = U\left(\frac{1}{p}\right) = G_{\gamma}{}^{\leftarrow}(1-p;\lambda,\delta) = \begin{cases} \lambda + \dfrac{\delta}{\gamma}[(-\log(1-p))^{-\gamma} - 1], \gamma \neq 0 \\ \lambda - \delta\log(-\log(1-p))\,, \qquad \gamma = 0 \end{cases} \tag{3.4}$$

A probabilidade de excedência de níveis elevados de u para Y, é dada por,

$$P(Y > u) = 1 - G_{\gamma}(u;\lambda,\delta) = \begin{cases} 1 - exp\left\{-\left[1+\gamma\left(\dfrac{u-\lambda}{\delta}\right)\right]^{-1/\gamma}\right\}, & \gamma \neq 0, \\ 1 - exp\left\{-exp\left(-\dfrac{u-\lambda}{\delta}\right)\right\}, & \gamma = 0. \end{cases} \tag{3.5}$$

O período de retorno do nível elevado u para Y, obtém-se a partir da equação,

$$T(u) = \frac{1}{1 - G_{\gamma}(u;\lambda,\delta)}\,. \tag{3.6}$$

Para estimar o nível de retorno a T-anos para Y sendo $p = 1/T$, basta inverter a equação (2.11),

$$U(t) = q_{1-\frac{1}{T}} = G_{\gamma}^{\leftarrow}\left(1 - \frac{1}{T}; \lambda, \delta\right) = \begin{cases} \lambda + \dfrac{\delta}{\gamma}\left[\left(-log(1-p)\right)^{-\gamma} - 1\right], & se \quad \gamma \neq 0, \\ \lambda - \delta log\left(-\,log(1-p)\right), & se \quad \gamma = 0. \end{cases} \tag{3.7}$$

E finalmente, sendo $\gamma < 0$, o limite superior do suporte para Y é dado por,

$$x^F = q_{Y,0} = \lambda + \frac{\delta}{\gamma}\,. \tag{3.8}$$

Sendo $(\hat{\gamma}, \hat{\lambda}, \hat{\delta})$ as estimativas dos parâmetros de forma, localização e de escala, respetivamente, obtidas através do método MV, as estimativas das quantidades de interesse obtêm-se substituindo o vetor de parâmetros desconhecidos $(\gamma, \lambda, \delta)$ por $(\hat{\gamma}, \hat{\lambda}, \hat{\delta})$.

Uma vez que, os quantis permitem que os modelos de probabilidade sejam expressos nas unidades de medida da variável em estudo, a relação do modelo GEV com os seus parâmetros é mais facilmente interpretada em termos das expressões dos quantis extremais. Definindo $y_p = -\log{(1-p)}$,

$$q_{1-p} = \begin{cases} \lambda + \dfrac{\delta}{\gamma}\left[\left(y_p\right)^{-\gamma} - 1\right], & \gamma \neq 0 \\ \lambda - \delta \log\left(y_p\right), & \gamma = 0 \end{cases} \tag{3.9}$$

Caso o parâmetro EVI seja, $\gamma = 0$, ao se representar graficamente q_{1-p} vs y_p, o gráfico será linear, se $\gamma < 0$ o gráfico será convexo com comportamento assintótico, quando $p \to 0$ para $\lambda - \frac{\delta}{\gamma}$ e em caso de $\gamma > 0$, o gráfico será concavo e sem limite finito. O gráfico que se obtém de representar graficamente q_{1-p} vs y_p tem a designação de gráfico de níveis de retorno (do inglês "*return level plot*").

3.1.3 INTERVALOS DE CONFIANÇA PARA OS PARÂMETROS

Tendo em conta a normalidade assintótica dos estimadores de MV, pode-se construir os Intervalos de confiança (IC's) para os parâmetros da população e para os quantis. Como já foi referido os IC's para os parâmetros da GEV $(\gamma, \lambda, \delta)$ decorrem da aproximação à Normal dos estimadores de MV.

O intervalo de aproximadamente $(1-\alpha)100\%$ de confiança para θ, em que θ representa genericamente os parâmetros $(\gamma, \lambda, \delta)$, onde $s.e.$ representa o erro padrão, $z_{1-\frac{\alpha}{2}}$ é o quantil $(1-\alpha/2)$ da distribuição Normal padrão, é definido da seguinte forma,

$$IC_{(1-\alpha)100\%}(\theta) = \left(\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \times s.e.\right). \tag{3.10}$$

Este procedimento não é muito recomendado, uma vez que, admite que o comportamento limite da função de verosimilhança, junto do máximo, seja quadrático. Dado que comportamento da função de verosimilhança costuma ser bastante assimétrico, deverá ser utilizada a função de verosimilhança de perfil, para se obter estimativas intervalares de melhor qualidade, que não são necessariamente centradas na estimativa de MV. A função de log-verosimilhança de perfil (Barndorff-Nielsen & Cox, 1994) de γ, é a log-verosimilhança maximizada relativamente aos outros parâmetros (λ, δ) e é definida como (Beirlant, 2004), $log\, L_p(\gamma) = max_{\lambda,\delta|\gamma} logL(\gamma, \lambda, \delta)$

A construção dos IC's para γ tem por base a estatística de razão de verosimilhança (RV ou LR, do inglês *"likelihood ratio"),*

$$\Lambda = \frac{L_p(\gamma_0)}{L_p(\hat{\gamma})}.$$

À semelhança do Teste de Razão de Verosimilhanças (TRV) para testar as Hipóteses:

$$H_0: \gamma = \gamma_0 = 0 \qquad (Gumbel) \qquad vs \qquad H_1: \gamma \neq \gamma_0 = 0 \qquad (Fréchet\ ou\ Weibull)$$

A estatística de teste, RV, tem distribuição assintótica qui-quadrado com um grau de liberdade,

$$RV = -2\, log\, \Lambda \xrightarrow[m\to\infty]{d} \chi_1^2$$

O teste assintótico de nível α rejeita H_0 se $-2\ log\ \Lambda > \ \chi_1^2(1-\alpha)$, logo o intervalo para o parâmetro γ através do log-verosimilhança de perfil com um grau de confiança$(1-\alpha)100\%$, é dado por,

$$IC_\gamma = \{\gamma: -2log\Lambda \leq \ \chi_1^2(1-\alpha)\},$$

$$IC_\gamma = \left\{\gamma: log\ L_p(\gamma) \geq \ log\ L_p(\hat{\gamma}) - \frac{\chi_1^2(1-\alpha)}{2}\right\} \tag{3.11}$$

Esta metodologia pode também ser aplicada, de forma análoga, aos restantes parâmetros da distribuição GEV.

No entanto, para se determinar um IC para um determinado quantil extremal, q_{1-p} é necessário fazer uma reparametrização do modelo GEV, de forma que q_{1-p} seja um dos parâmetros do modelo. A função log-verosimilhança de perfil é então obtida por maximização da log-verosimilhança em relação aos restantes parâmetros. Considerando, sem perda de generalidade, o caso $\gamma \neq 0$,

$$q_{1-p} = \ \lambda\ + \ \frac{\delta}{\gamma}\left[\left(-log(1-p)\right)^{-\gamma} - 1\right] \Leftrightarrow \lambda\ = q_{1-p} + \frac{\delta}{\gamma}\left[1 - \left(-log(1-p)\right)^{-\gamma}\right] \tag{3.12}$$

Substituindo λ na equação de log-verosimilhança do modelo GEV é obtida a função de log-verosimilhança do modelo GEV em função do vetor de parâmetros $\left(q_{1-p}; \lambda, \delta\right)$, procedendo-se de modo usual.

3.1.4 Testes de hipóteses para o modelo GEV

Na TVE, a validação do modelo e os testes de hipóteses são essenciais para a seleção do melhor modelo.
Como temos visto ao longos dos vários capítulos, o modelo GEV é a única distribuição limite para o máximo para variáveis i.i.d., no entanto a GEV tem três casos especiais: Gumbel, Weibull e Fréchet, que têm características diferentes e que estão associadas a diferentes comportamentos de cauda (ver Figura 3 em Castillo et al., 2005).

De seguida, são apresentados os testes de hipóteses mais comuns para o modelo GEV. Estes testes têm como base de ajustamento o modelo Gumbel, pois é o modelo mais simples assim como a inferência a ele associada. São eles, o teste de razão de verosimilhanças já definido na secção 3.1.3; testes de estatísticas de ajustamento (*Goodness-of-Fit Tests*), como, o teste de Kolmogorov-Smirnov, Cramér-von-Mises e o teste de Anderson Darling (Chandra et al., 1981).

A f.d. Gumbel tem um $\gamma = 0$ e é considerada como um ponto de transição de uma família de distribuições com limite superior do suporte finito ($\gamma < 0$), para uma outra com limite superior do suporte infinito ($\gamma > 0$). Sendo assim, começamos a gerar as hipóteses sempre com o modelo mais simples:

$$H_0: \gamma = 0 \qquad (Gumbel) \qquad vs \qquad H_1: \gamma \neq 0 \qquad (Fréchet\ ou\ Max - Weibull)$$

No caso de rejeição do modelo Gumbel, poderá ser feito um teste unilateral para se determinar qual dos outros modelos máx-estáveis será o adequado, Fréchet ou Max-Weibull (Wang' & Cooke, 1996):

$$H_0: \gamma = 0 \qquad (Gumbel) \qquad vs \qquad H_1: \gamma > 0 \qquad (Fréchet\ )$$

$$H_0: \gamma = 0 \qquad (Gumbel) \qquad vs \qquad H_1: \gamma < 0 \qquad (max - Weibull\ )$$

Existem outras metodologias de testes de hipóteses, igualmente relevantes no contexto da TVE e na avaliação estatística do modelo, como o teste Deviance, o teste Score (Rao), o teste Unilateral de Tiago de Oliveira (de Oliveira & Gomes, 1984), e o teste *Locally Asymptotically Normal* (LAN).

O Teste Deviance é um teste de hipóteses que compara as verosimilhanças entre dois modelos, isto é, é um método alternativo para medir a incerteza dos estimadores MV. Sendo $\mathcal{M}_0$ e $\mathcal{M}_1$ dois modelos encaixados ou aninhados, isto é, $\mathcal{M}_0 \subset \mathcal{M}_1$, a função Deviance (Castillo et al. ,2005) é definida como,

$$\mathcal{D} = 2\{max\ell_1(\mathcal{M}_1) - max\ell_0(\mathcal{M}_0)\}, \tag{3.13}$$

onde, $\ell_0(\mathcal{M}_0)$ e $\ell_1(\mathcal{M}_1)$, são, respetivamente, os valores da função de log-verosimilhança para os modelos ajustados, $\mathcal{M}_0$ e $\mathcal{M}_1$. Tal como definido anteriormente, $\theta \coloneqq (\gamma, \lambda, \delta)$ então $\hat{\theta} \coloneqq (\widehat{\gamma}, \hat{\lambda}, \widehat{\delta})$. Para a distribuição GEV vem,

$$\mathcal{D} \ = \ 2\left\{max\ell\left(\hat{\theta}_{G_\gamma}|Y_1, \dots, Y_m\right) - max\ell\left(\hat{\theta}_{G_0}|Y_1, \dots, Y_m\right)\right\}. \tag{3.14}$$

.

Sob a validade de H_0,

$$\mathcal{D} \xrightarrow[n\to\infty]{d} Q_1 \sim \chi_1^2. \tag{3.15}$$

Em 1984, Hosking (J. R. M. Hosking, 1984) sugeriu uma modificação no parâmetro da razão de verosimilhança que permite uma melhor aproximação à distribuição assintótica de RV (Castillo et al., 2005):

$$\mathcal{D}^* = \frac{\mathcal{D}}{1 + \frac{2.8}{n}} \ \xrightarrow[n\to\infty]{d} \ Q_1 \sim \chi_1^2. \tag{3.16}$$

O teste assintótico de nível α rejeita H_0 (Modelo Gumbel), se $\mathcal{D}^* \geq \chi_1^2(1-\alpha)$.
A decisão também pode ser tomada com base no valor-p,

$$p(\mathcal{D}^*) = \mathbb{P}(Q_1 > \mathcal{D}^*) = 1 - \mathbb{P}(Q_1 \leq \mathcal{D}^*). \tag{3.17}$$

O teste assintótico de nível α rejeita H_0 (Modelo Gumbel), se $p(\mathcal{D}^*) \leq \alpha$.

O Teste de Score ou Teste Rao (*Lagrange Multiplier Test*), é um teste estatístico que permite avaliar se um modelo mais simples com parâmetros fixos é adequado ou se é preferível um modelo mais complexo com parâmetros variáveis. Este teste verifica quando a derivada da função de log-verosimilhança, em ordem a γ, no ponto $\gamma = 0$, é significativamente diferente de zero (Bera & Bilias, 2001; Rao, 1947).

Se a v.a. X segue distribuição GEV, a função de log-verosimilhança para $(X_1, \dots, X_n)$ é dada por, $\ell(\gamma, \lambda, \delta | X_1, \dots, X_n) = \sum_{i=1}^{n} \log GEV_\gamma(\gamma_i | \lambda, \delta)$.

A função score em relação a γ,

$$V(\gamma | \lambda, \delta, X_1, \dots, X_n) = \frac{\partial \ell(\gamma, \lambda, \delta | X_1, \dots, X_n)}{\partial \gamma} = \sum_{i=1}^{n} \frac{\partial \log GEV_\gamma(\gamma_i | \lambda, \delta)}{\partial \gamma}. \tag{3.18}$$

Sendo $(\hat{\lambda}_{G_0}, \hat{\delta}_{G_0})$ os estimadores MV para o modelo Gumbel (G_0), a estatística score para o teste,

$$H_0: \gamma = 0 \qquad (Gumbel) \qquad vs \qquad H_1: \gamma \neq 0 \qquad (Fréchet\ ou\ Weibull\ )$$

$V_n = \sum_{i=1}^{n} \lim_{\gamma \to 0} \frac{\partial \log GEV_\gamma(\gamma_i | \hat{\lambda}_{G_0}, \hat{\delta}_{G_0})}{\partial \gamma} = \sum_{i=1}^{n} \left(\frac{1}{2} {Z_i}^2 - Z_i - \frac{1}{2} {Z_i}^2 exp(-Z_i) \right)$, com $Z_i = \frac{X_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}}, i = 1, \dots, n$.

A estatística de teste pode seguir uma distribuição Gaussiana ou Qui-Quadrado,

$$V_n^* = \frac{V_n}{\sqrt{2.09797n}} \xrightarrow[n \to \infty]{d} Z_1 \sim \mathcal{N}(0,1) \tag{3.19}$$

$$V_n^{*2} = \frac{{V_n}^2}{2.09797n} \xrightarrow[n \to \infty]{d} Q_2 \sim \chi_1^2 \tag{3.20}$$

O teste assintótico bilateral de nível α rejeita H_0 (Modelo Gumbel), se $|V_n^*| \geq z_{(1-\alpha/2)}$ ou $V_n^{*2} \geq \chi_{1,1-\alpha}^2$. Os valores-p podem ser calculados do seguinte modo: $p(V_n^*) = 2 - 2\varphi(|V_n^*|)$ ou $p(V_n^{*2}) = 1 - \mathbb{P}(Q_2 \leq V_n^{*2})$.

O teste unilateral proposto por Tiago de Oliveira (de Oliveira & Gomes, 1984), foi construído para testar hipóteses direcionais especificas, normalmente é usado para verificar se um parâmetro (por exemplo, o parâmetro de forma no modelo GEV ou GP) é maior ou menor que um dado valor. Este teste apresenta uma estatística de teste, de Gumbel, dada por,

$$GS_m = \frac{Y_{m:m} - Y_{(m/2)+1:m}}{Y_{(m/2)+1:m} - Y_{1:m}} \ . \tag{3.21}$$

Sob a validade de H_0,

$$GS_m{}^* = \frac{GS_m - \beta_m}{\alpha_m} \xrightarrow[n\to\infty]{d} W \sim \Lambda. \tag{3.22}$$

Os parâmetros, α_me β_m são dados por, $\alpha_m = \frac{1}{log(\log m)}$ e $\beta_m = \frac{\log m + \log(\log 2)}{\log(\log m) - \log(\log 2)}$.

Sendo q_α o quantil de probabilidade α para uma distribuição Gumbel padrão.

Para $\mathrm{H}_1: \gamma < 0$,

$$H_0: \gamma = 0 \qquad (Gumbel) \qquad vs \qquad H_1: \gamma < 0,$$

o teste assintótico de nível α rejeita-se H_0 , se $q_\alpha \; = GS_m \leq \beta_m - \alpha_m \log(-\log \alpha) \Leftrightarrow \; GS_m{}^* \leq -\log(-\log \alpha)$, sendo $p(GS_m{}^*) = \Lambda(GS_m{}^*)$.

Para $\mathrm{H}_1: \gamma > 0$,

$$H_0: \gamma = 0 \qquad (Gumbel) \qquad vs \qquad H_1: \gamma > 0,$$

o teste assintótico de nível α rejeita-se H_0 , se $q_{1-\alpha} \; = GS_m \geq \beta_m - \alpha_m \log(-\log(1-\alpha)) \Leftrightarrow$ $GS_m{}^* \geq -\log(-\log(1-\alpha))$, sendo $p(GS_m{}^*) = 1 - \Lambda(GS_m{}^*)$.

O teste de LAN (*Locally Asymptotically Normal*) é usado em modelos com parâmetros assintóticos, onde o teste da razão de verosimilhança é aproximado a uma função quadrática, permitindo que o teste seja assintoticamente normal. A hipótese nula considera que os parâmetros seguem uma distribuição assintótica, enquanto na hipótese alternativa, os parâmetros não seguem uma distribuição assintótica (Marohn, 2000).

3.2 MODELO GEV MULTIVARIADO OU DAS K-MAIORES OBSERVAÇÕES

Como já se tem vindo a referir uma das maiores dificuldades na análise de valores extremos é a quantidade limitada de dados para estimação de modelos. Esta limitação permitiu o desenvolvimento de outras técnicas que permitiram modelar dados extremos, para além do modelo de blocos Máximos ou método dos Máximos Anuais, são eles o método das k-Maiores Observações (MMO) disponíveis da amostra através do comportamento conjunto das k maiores observações e o Modelo dos Excessos acima de um nível extremo [secção 3.3] (Coles, 2001).

Tal como nas secções anteriores, seja $M_n \coloneqq (X_1, \dots, X_n)$ uma amostra aleatória de variáveis aleatórias i.i.d., mas neste caso as variáveis correspondem às maiores observações, onde

$$X_1, \dots, X_n, com\ X_j = X_{1j} > X_{2j} > \cdots > X_{kj}, 1 \leq j \leq m.$$

Para comparar com o método MMA, assumimos que efetuamos uma partição da a.a. $X_1, \dots, X_n$ em m blocos (geralmente cada bloco corresponde a um ano - m), cada um de dimensão q. Para cada um desses blocos extraímos as k maiores observações ($k < q$), dando um conjunto de m vetores aleatórios $k -$ dimensionais i.i.d. X_j, ao qual se ajusta o modelo GEV multivariado,

$$P\left\{\left(\frac{{M_n}^k - b_n}{a_n}\right) \leq x\right\} \underset{n\to\infty}{\longrightarrow} G(x) \quad \in \mathcal{D}_{\mathcal{M}}(G)\,, M_n^k = k\ maiores\ observações\ \{X_1, \dots, X_n\}$$

Ficando assim ${M_n}^k$ a pertencer a uma distribuição GEV para um valor fixo de k.
Se considerarmos $k = 1$, este não é nada mais que o Modelo MA. Na Figura 4, está representado o modelo MO, para uma amostra com $m = 4$ blocos, cada um com dimensão $q\ = 7$ de onde se escolhem as $k = 3$ maiores observações de cada bloco.

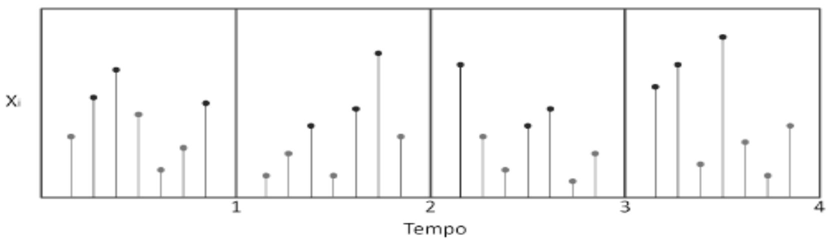


Figura 4 - Representação de uma amostra com partição em 4 blocos e respetivos máximos.

Quando $k\ =\ 1$, a função de log-verosimilhança não é mais que o a log-verosimilhança do modelo GEV, para os máximos anuais. O modelo MMO fornece uma verosimilhança cujos parâmetros correspondem aos da distribuição GEV, obtidos pelo método dos máximos anuais, mas neste caso foram incorporadas mais observações extremais da amostra original (Coles, 2001).

3.2.1 ESTIMAÇÃO DO MODELO PARA AS MMO

Neste modelo, as k maiores observações num bloco são normalizadas tal como um máximo, onde os parâmetros correspondem aos parâmetros de uma distribuição GEV de blocos máximos. Neste modelo é mais fácil aumentar a dimensão da amostra, não sendo necessário um número tão grande de observações (Gomes et al., 2013).

$$(W_1, \dots, W_k) = \left(\frac{X_{n:n}-b_n}{a_n}, \dots, \frac{X_{n-k+1:n}-b_n}{a_n}\right), \quad a_n > 0 \; e \; b_n \in \mathbb{R},$$

$$g_{X_{n:n},\dots,X_{n-k+1:n}}(W_1, \dots, W_k) \coloneqq G_\gamma(W_k) \cdot \prod_{i=1}^{k} \frac{g_\gamma(W_i)}{G_\gamma(W_i)} = G_\gamma\left(\frac{\lambda - x_k}{\delta}\right) \cdot \prod_{i=1}^{k} \frac{g_\gamma\left(\frac{\lambda - x_j}{\delta}\right)}{G_\gamma\left(\frac{\lambda - x_j}{\delta}\right)}. \qquad (3.23)$$

Após normalização, passamos à estimação dos parâmetros de forma, localização e escala, para posterior inferência sobre os parâmetros de interesse em acontecimentos raros e à semelhança da secção anterior, podemos usar o método de MV para obter estimativas dos parâmetros do modelo MMO equação (3.1) (Coles, 2001; Gomes et al., 2013).

Uma vez que, os parâmetros do modelo MMO correspondem aos parâmetros de uma distribuição GEV, o cálculo das outras quantidades de interesse é semelhante ao apresentado na secção 3.1.2.

3.3 MODELO PARETIANO - EXCESSOS DE NÍVEL

Nas secções 3.1 e 3.2, foram abordadas distribuições onde a ordem era importante, no entanto por vezes é mais útil analisar variáveis aleatórias que excedem ou estão abaixo de um certo nível -*threshold*, como já mencionado na secção 2.3 (Castillo et al., 2005; Coles, 2001).

Na Figura 5, é apresentada a f.d. dos excessos F_u, tendo em conta a definição de excesso acima de um determinado u, como $(X - u|X > u)$.

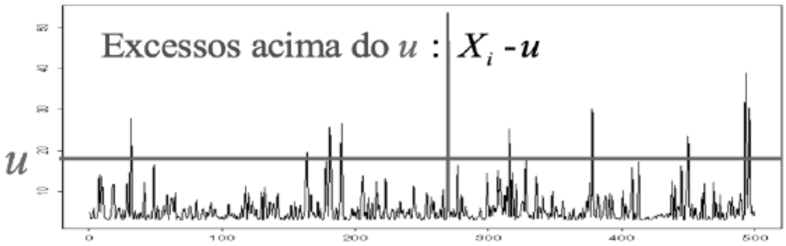


Figura 5 - Excedências de um nível elevado u (Gomes et al., 2013).

Sendo, $x = u + y$, a GP pode também ser expressa em termos de três parâmetros: forma (γ), localização (λ) e escala (δ), estes parâmetros são exclusivamente determinados a partir dos associados à distribuição GEV de blocos máximos, ou seja, existe uma dualidade entre a GP e a GEV, o que significa que o parâmetro de forma γ, é fundamental na determinação do comportamento qualitativo da distribuição GP (Coles, 2001). Esta dualidade é chamada de Teorema de Pickands-Balkema-de Haan, propondo, para níveis elevados u a aproximação, $F_u(x) \approx H(x;\gamma,\delta)$, para $y \in [0; x^F - u], se\ \ \gamma \geq 0\ e\ para\ \ y \in [0; -\delta/y],\ \ se\ \ y < 0.$

Assim, $F \in \mathcal{D}_{\mathcal{M}}(G_\gamma), \gamma \in \mathbb{R} \iff \lim_{u \to x^F} sup|F_u(x) - H(x;\gamma,\delta)| = 0\,,\ \ 0 < y < x^F - u.$

Esta abordagem é designada de Modelo Paretiano de Excessos ou Modelo POT (Gomes et al., 2013),

$$H(x;\gamma,\delta) = \begin{cases} 1 - \left(1 + \gamma \dfrac{x-u}{\delta}\right)_+^{-1/\gamma}, & \gamma \neq 0, \\ 1 - exp\left(-\dfrac{x-u}{\delta}\right), & \gamma = 0. \end{cases} \tag{3.24}$$

$com\ (z)_+ = max(0;z)$

3.3.1 Estimação de parâmetros no modelo GP

Nesta secção vamos usar um dos métodos mais conhecidos para estimar os parâmetros e quantis da distribuição GP, tal como nas secções anteriores, o método MV.

Assumindo que, $(X_1, \ldots, X_{N_u})$ são variáveis i.i.d. com distribuição GP, com f.d $H(X;\gamma,\delta)$, a equação de log-verosimilhança (Gomes et al., 2013), Se $\gamma \neq 0$ e para $1+\frac{\gamma x}{\delta}>0$, $i = 1,\dots,N_u$, $logL(\delta_u,\gamma) = -N_u log\delta_u - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{N_u} log\ \left(1+\frac{\gamma X_i}{\delta_u}\right)$.

Se $\gamma = 0$,

$$logL(\lambda_u,0) = -N_u log\delta_u - \frac{1}{\delta_u}\sum_{i=1}^{N_u} X_i. \tag{3.25}$$

A f.d. $H_0(X;\delta)$ é a f.d. de um modelo exponencial, então o estimador de MV para δ é $\hat{\delta}_u^{\ ML} = \overline{\mathrm{X}}$. Se $\gamma \neq 0$, é necessária uma reparametrização, $(\delta_u,\gamma) \rightsquigarrow (\tau,\gamma), onde\ \tau := \frac{\gamma}{\delta_u}$,

$$logL(\lambda_u,\tau) = -N_u log\gamma + N_u log\tau - \left(\frac{1}{\gamma}+1\right)\sum_{i=1}^{N_u} log(1+\tau X_i). \tag{3.26}$$

Os estimadores de MV de τ e γ, são dados por,

$$\frac{1}{\hat{\tau}^{ML}} - \left(\frac{1}{\hat{\gamma}^{ML}+1}\right)\frac{1}{N_u}\sum_{i=1}^{N_u}\frac{X_i}{1+\hat{\tau}^{ML}.X_i} = 0. \tag{3.27}$$

Com $\hat{\tau}^{ML} = \frac{1}{N_u}\sum_{i=1}^{N_u} log\ (1+\hat{\tau}^{ML}.X_i)$.

A obtenção dos intervalos de confiança para os parâmetros da distribuição GP(δ_u,γ) decorre da normalidade assintótica dos estimadores de MV (Gomes et al., 2013). Tal como foi feito na abordagem MMA, podem ser obtidas estimativas intervalares de melhor qualidade usando a função de log-verosimilhança de perfil (Castillo et al., 2005; Gomes et al., 2013), tal como apresentado na secção 3.1.1.

Os testes de hipóteses, com as devidas adaptações são semelhantes aos apresentados na secção 3.1.4., neste modelo, as hipóteses em estudo são:

$$H_0: \gamma = 0 \qquad (Exponencial) \qquad\qquad vs \qquad\qquad H_1: \gamma \neq 0 \qquad (Pareto\ ou\ Beta\ )$$

Outros testes específicos para esta metodologia são os testes de Marohn (Marohn, 2000) e os de Gomes e Van Monfort (Gomes et al., 1985), descritos abaixo.

Teste de Marohn (Marohn, 2000): sejam $W_1, \dots, W_m$, as m excedências acima do threshold u obtidas da a.a inicial. A estatística de teste é dada por:

$$T_m = \frac{1}{2}\left(\frac{S_w^2}{(\bar{W} - u)^2} - 1\right), \text{ onde } S_w^2 \text{ é a variância amostral.} \tag{3.28}$$

A estatística de teste, sob a validade da hipótese nula, é assintoticamente normal padrão, i.e. $T_m{}^* = \sqrt{m} T_m \xrightarrow[m\to\infty]{d} Z \sim N(0,1)$. A hipótese nula é rejeitada, se $|T_m{}^*| \geq Z_{1-\alpha/2}$ ou se, $T_m{}^* \leq Z_\alpha$, ou se, $T_m{}^* \geq Z_{1-\alpha}$. Os valores-p dos testes bilaterais e unilaterais são: $p(T_m{}^*) = 2 - 2\Phi(|T_m{}^*|)$ ou $p(T_m{}^*) = \Phi(T_m{}^*)$ ou $p(T_m{}^*) = 1 - \Phi(T_m{}^*)$, sendo H_0 rejeitada ao nível assintótico α, se $p(T_m{}^*) \leq \alpha$. A estatística de teste para a versão bilateral é enviesada e com uma baixa potência para amostras de pequena ou moderada dimensão (m < 500).

Teste de Gomes e van Monfort (Gomes et al., 1985): é utilizado para hipóteses unilaterais, ou seja, $H_1: \gamma < 0 \ e \ H_1: \gamma > 0$. Este teste recorre ao máximo e à mediana da amostra de excedências, e a estatística de teste é,

$$G_m = \frac{W_{m:m}}{W_{\left[\frac{m}{2}\right]+1:m}}. \tag{3.29}$$

Sob a validade de H_0, $G_m{}^* = \log 2 \cdot G_m - \log m \xrightarrow[m\to\infty]{d} U \sim \Lambda$. A hipótese $H_0: \gamma = 0$, será rejeitada ao nível assintótico α, se $\mathrm{G_m}^*$ tiver o seguinte comportamento com os quantis de probabilidade, $\mathrm{G_m}^* \leq q_\alpha$ ou $\mathrm{G_m}^* \geq \mathrm{q}_{1-\alpha}$. Os valores-p são: $p(G_m{}^*) = \Lambda(G_m{}^*)$ e $p(G_m{}^*) = 1 - \Lambda(G_m{}^*)$.

3.3.2 ESCOLHA DO THRESHOLD

O *threshold* u tem um papel bastante importante na metodologia POT, e a escolha deste nível é análoga à escolha do tamanho do bloco no método MA, implicando um equilíbrio entre o viés e a variância, o que gera algumas dificuldades na sua seleção. Valores baixos de *threshold* correspondem à utilização de um número maior de observações, e possivelmente à violação assintótica do modelo, aumentando o viés e diminuindo a variância dos estimadores. Enquanto, valores elevados de *threshold* geram poucos excessos, provocando maior variância e uma redução do viés dos estimadores (Coles, 2001).

A metodologia padrão consiste em adotar o menor valor de u que permita uma razoável aproximação do modelo GP. Segundo (Coles, 2001), existem dois métodos (já descritos na secção 2.4) para determinar o valor correto de u:

- 1º Método - Função de excesso médio de uma distribuição GP através da representação do ME-plot, onde o *threshold* u deverá ser escolhido no ponto a partir do qual se verifica a linearidade.
- 2º Método - Escolha de um conjunto de *threshold*, este método consiste na escolha de alguns candidatos a *threshold*. Primeiro são estimados os parâmetros da GP (δ, γ), e de seguida é escolhido um valor de u a partir do qual os estimadores dos parâmetros de forma γ e de escala δ, se tornam estáveis (Coles, 2001).

3.3.3 OUTROS PARÂMETROS DE INTERESSE DE ACONTECIMENTOS RAROS

Os parâmetros de acontecimentos raros com interesse na EVT, podem ser calculados também na abordagem GP, introduzindo o parâmetro *threshold* u, nas equações definidas acima, são eles, a estimação de quantis extremais e do limite superior do suporte (Gomes et al., 2013).

Invertendo a f.d. da GP, os quantis extremais de probabilidade de excedência p, são representados da seguinte forma,

$$q_{1-p} = U_{H_\gamma}\left(\frac{1}{p}\right) = H_\gamma^{\leftarrow}(1 - p/\delta_u), \tag{3.30}$$

com

$$U_{H_\gamma}\left(\frac{1}{p}\right) = \begin{cases} \dfrac{\delta_u}{\gamma}(p^{-\gamma} - 1), & \gamma \neq 0, \\ -\delta_u \log p\,, & \gamma = 0. \end{cases} \tag{3.31}$$

Se $\gamma < 0$, o limite superior do suporte da GP é finito e dado por,

$$U_{H_\gamma}(\infty) = \mathrm{q}_1 = H_\gamma^{\leftarrow}(1/\delta_u) = -\frac{\delta_u}{\gamma}\ . \tag{3.32}$$

Ao substituir o vetor de parâmetros desconhecidos (δ_u, γ), pelas estimativas $(\hat{\delta}_u, \hat{\gamma})$ de MV teremos estimativas para os quantis extremais e para o limite superior de suporte da GP, H_γ (Gomes et al., 2013), tal como secção 3.1.2.

Considerando que $X \frown F$, vimos que a distribuição condicional dos excessos acima de u, é aproximada pela GP através da equação,

$$F_u(y) = \mathbb{P}[X - u \leq y | X > u] = \frac{F(u+y) - F(u)}{1 - F(u)} \approx H_\gamma(y; \delta_u). \tag{3.33}$$

Deste modo, para estimar a probabilidade de excedência de x elevado, $\bar{F}(x)$, é feita a seguinte aproximação,

$$\bar{F}(x) \approx \bar{F}(u)\left[1 - H_\gamma(x - u; \delta_u)\right] = \bar{F}(u)\left(1 + \frac{\gamma(x-u)}{\delta_u}\right)^{-\frac{1}{\gamma}}. \tag{3.34}$$

Estimando $\bar{F}(u)$, pela frequência relativa Nu/n das observações que excedem u na amostra original de dimensão n, $X_1, \ldots, X_n$, vem que,

$$F(x) = \frac{N_u}{n}\left(1 + \frac{\gamma(x-u)}{\delta_u}\right)^{-\frac{1}{\gamma}}. \tag{3.35}$$

Tendo, $U(1/p) = \mathrm{F}^{\leftarrow}(1-p)$, considerando que $F(x) = 1 - p$ na expressão da probabilidade de excedência, é feita a inversão da equação para estimar os quantis elevados de F,

$$U(1/p) = u + \frac{\delta_u}{\gamma}\left(\left(\frac{np}{N_u}\right)^{-\gamma} - 1\right). \tag{3.36}$$

Para $\gamma < 0$, o limite superior do suporte para F é dado por,

$$x^F = U(\infty) = q_1 = u - \frac{\delta}{\gamma}. \tag{3.37}$$

Para se obter o nível de retorno de T-anos, basta calcular,

$$U(\mathrm{T}) = q_{1-\frac{1}{T}}. \tag{3.38}$$

Para se estimar o período de retorno de um nível v,

$$T_v = \frac{1}{\bar{F}(v)} \,. \tag{3.39}$$

3.4 NÃO-ESTACIONARIDADE: SAZONALIDADE E TENDÊNCIA

Quando se trabalha com sequências de variáveis aleatórias independentes e identicamente distribuídas, é assumido um contexto estacionário. Na prática, contudo, os dados raramente são estacionários, tal como, dados de meteorologia, mudanças climáticas normalmente têm uma componente sazonal muito forte possivelmente devido aos diferentes padrões de temperatura nos diferentes meses; dados de extremos de níveis do mar (Dixon et al., 1998; Kotz & Nadarajah, 2000); dados financeiros ou económicos também apresentam flutuações, e um padrão ascendente e descendente (Beirlant, 2004; Embrechts et al., 2003; Reiss et al., 2007), ou seja, neste tipo de dados parece existir uma dependência temporal (Bocciolone et al., 1993; Coles, 2001).

A não-estacionariedade significa que as propriedades dos dados geram um processo que varia com o tempo (Mudelsee, 2020; Salvadori et al., 2007). Existem resultados teóricos para modelar alguns tipos de não-estacionariedade, mas que não traduzem a diversidade de padrões presentes em situações reais. Uma opção aparentemente simples é a extensão de modelos estacionários através de uma dependência temporal nos parâmetros (Mudelsee, 2020).

No caso dos fenómenos climáticos, uma abordagem frequentemente utilizada é admitir que o parâmetro de localização depende linearmente do tempo, i.e., que $\lambda(t) = \beta_0 + \beta_1 t$ ou que $\delta(t) = \beta_0 + \beta_1 t$, passando as distribuições em contexto não estacionário a serem representadas como,

$$X_t \sim GEV(\lambda(t), \delta, \gamma) \tag{3.40}$$

$$Y_t \sim GPD(\delta(t), \gamma) \tag{3.41}$$

Desta maneira, as variações ao longo tempo que se observam no processo, são modeladas com tendência linear, no parâmetro de localização no caso do modelo GEV e no parâmetro escala no caso do modelo GPD em dados ambientais.

O parâmetro β_1 corresponde à taxa de alteração média anual (se aplicado a uma metodologia dos máximos anuais)(Coles, 2001).

Como podemos ajustar uma vasta combinação de modelos a um conjunto de dados, escrevendo qualquer um dos parâmetros como uma função do tempo ou mesmo de outras covariáveis, torna-se crucial a escolha do melhor modelo. Devemos escolher um modelo parcimonioso, i.e., o modelo mais simples com o menor número de parâmetros, e que explique a maior percentagem de variância presente na amostra (Coles, 2001).

Posteriormente, podemos utilizar o método de MV para obter as estimativas dos parâmetros associados ao modelo não-estacionário e de seguida usar a estatística Deviance ($\mathcal{D}$) já definida em (3.13) para medir a incerteza dos estimadores MV.

Neste caso, $\mathcal{M}_0$ seria o modelo ajustado sem tendência (modelo estacionário) e $\mathcal{M}_1$ o modelo ajustado com tendência (modelo não-estacionário), uma vez que, se trata de modelos encaixados.

Mas, como neste caso adicionámos mais parâmetros ao processo, os g.l. da distribuição qui-quadrado vão ser a diferença do número de parâmetros entre os dois modelos,

$$\mathcal{D} \xrightarrow[n\to\infty]{d} Q_1 \sim \chi^2_{p1-p0} = \chi^2_k \tag{3.42}$$

Onde $p0$ e $p1$ são, respetivamente, o número de parâmetros dos modelos $\mathcal{M}_0$ e $\mathcal{M}_1$, sendo $p1 = p0 + k$.

Valores elevados de $\mathcal{D}$, indicam que o modelo $\mathcal{M}_1$, explica, substancialmente, uma maior parte da variabilidade da amostra do que $\mathcal{M}_0$, enquanto valores baixos de $\mathcal{D}$, sugerem que aumentar o número de parâmetros do modelo não melhora a capacidade preditiva do modelo, indo de encontro com o princípio da parcimónia (Coles, 2001).

À semelhança dos testes de hipóteses apresentados anteriormente,

$$H_0: Não\ há\ tendência\ na\ amostra \qquad vs \qquad H_1: Há\ tendência\ na\ amostra$$

$$H_0: O\ modelo\ tem\ p_0\ parâmetros \qquad vs \qquad H_1: O\ modelo\ tem\ p_1\ parâmetros$$

O teste assintótico de nível α rejeita H_0, se $\mathcal{D} \geq \chi_k^2(1-\alpha)$.
A decisão também pode ser tomada com base no valor-p, $p(\mathcal{D}) = \mathbb{P}\,(Q \geq \mathcal{D})$, e o teste assintótico de nível α rejeita H_0, se $p(\mathcal{D}) \leq \alpha$.
Há que ter em consideração que, a distribuição assintótica da estatística de TRV só é válida se a hipótese nula não estiver na fronteira (Stram & Lee, 1994; Stram & Lee, 1994), quando tal acontece, a distribuição assintótica do TRV é uma mistura de qui-quadrados.
Para $k = 1$, a distribuição assintótica do TRV é uma mistura de qui-quadrados com p_0 e $p_0 + 1$ g.l.
Quando o modelo da hipótese nula não está encaixado no modelo da hipótese alternativa, por exemplo quando se tem um modelo com tendência na localização e outro com tendência na escala, o teste RV não é apropriado. A comparação de modelos não encaixados pode ser feita com base nos critérios de informação AIC (*Akaike Information Criterion*) e BIC (*Bayesian Information Criterion*),

$$AIC = -2\max l(M) + 2p \tag{3.43}$$

$$BIC = -2\max l(M) + plogN \tag{3.44}$$

Sendo p o número de parâmetros do modelo $\mathcal{M}$ e N a dimensão da amostra.
Sabendo que, quanto menor o valor do AIC ou do BIC melhor será o modelo, estes critérios também podem ser usados em modelos encaixados (Gomes et al., 2013).

4

Caso de Estudo: Temperaturas Máximas no Alentejo

O objetivo geral deste trabalho consiste em caracterizar a evolução das temperaturas máximas no Alentejo de forma a perceber se, de facto, as temperaturas máximas observadas estão a aumentar e os verões alentejanos estão a ficar mais quentes, para tal, foram analisadas as temperaturas máximas de quatro concelhos: Portalegre, Évora, Beja e Sines. Os objetivos específicos deste trabalho consistem na aplicação das metodologias paramétricas de Estatística de Extremos univariada para modelar as temperaturas máximas, utilizando o método dos máximos anuais, das k-maiores observações e das maiores excedências/paretiano. A análise estatística será desenvolvida no software RStudio (R Core Team, 2021).

Para cada um dos métodos e para cada uma das componentes em estudo serão apresentadas estimativas de quantidades de interesse como a probabilidade de excedência de níveis elevados, níveis de retorno entre outros. Os dados foram recolhidos na *European Climate Assessment & Dataset* (ECA&D) (https://cds.climate.copernicus.eu/cdsapp - !/dataset/insitu-gridded-observations-europe?tab=overview). O E-OBS é um conjunto de dados observacionais diários espacialmente distribuídos sobre a Europa. A série temporal combinada da rede de estações do projeto ECA&D é utilizada como base para o conjunto de dados espacialmente distribuídos E-OBS. Todos os dados das estações são obtidos diretamente dos Serviços Meteorológicos e Hidrológicos Nacionais (NMHSs) Europeus ou de outras instituições detentoras de dados. Para um número considerável de países, o número de estações usadas é a rede nacional completa e, portanto, muito mais densa do que a rede de estações que é rotineiramente compartilhada entre os NMHSs (que é a base de outros conjuntos de dados em grade).

Os dados recolhidos dizem respeito aos dados diários de temperatura máxima (em °C), desde 1950 a 2022, e correspondendo a 73 anos. Uma vez que, o intuito deste trabalho é analisar as temperaturas máximas anuais e verificar se, estas estão a aumentar ao longo do tempo, de cada série de dados apenas foram usados para o estudo os quatro meses mais quentes: Junho, Julho, Agosto e Setembro.

4.1 MODELO DOS MÁXIMOS ANUAIS

A modelação paramétrica das temperaturas máximas anuais é iniciada com uma análise exploratória dos dados, para averiguar qual o modelo que mais se adapta ao conjunto de dados, e de seguida são realizadas uma análise inferencial e a seleção do melhor modelo.

4.1.1 ESTATÍSTICA DESCRITIVA PRELIMINAR

O conjunto de dados em estudo para o modelo MA é composto pelas 73 temperaturas máximas anuais dos quatro concelhos do Alentejo: Portalegre, Évora, Beja e Sines, em °C, no período de 1950 a 2022, com as séries temporais apresentadas na Figura 6.

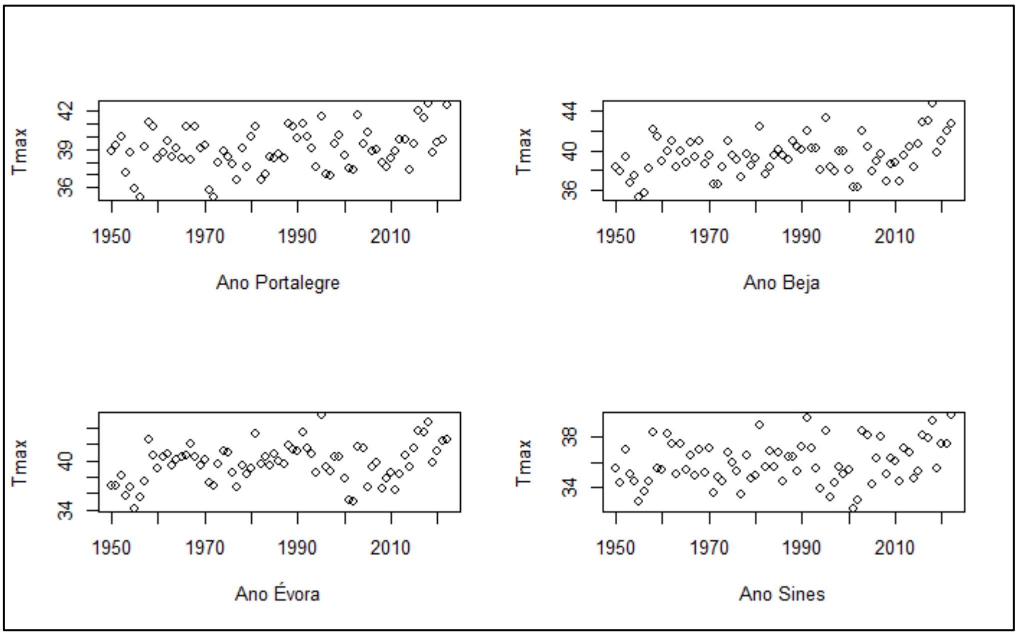


Figura 6 - Temperaturas máximas anuais dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

Através do `runs test` e do `difference.sign.test` foi verificada a aleatoriedade dos dados, a um nível de significância de 5%, e a decisão foi a de não rejeitar a hipótese nula (de aleatoriedade), pois não existem evidências estatística significativas que os dados não são aleatórios, uma vez que, os *p-values* foram todos superiores a 0.05.

Na análise preliminar de dados, começámos por obter o histograma e o *boxplot* associados a este conjunto de dados, apresentados na Figura 7 e na Figura 8.

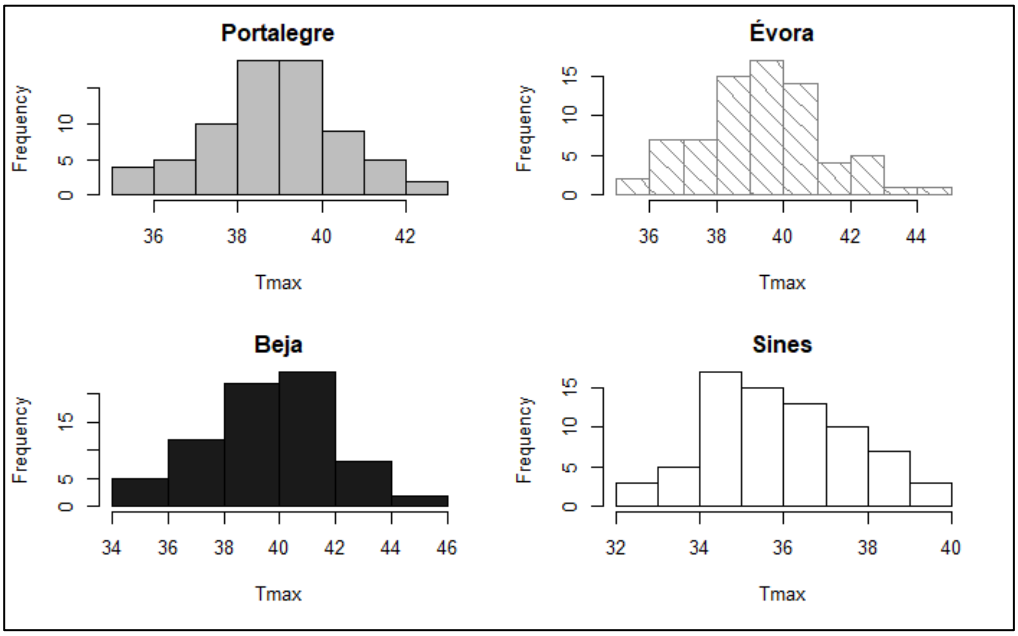


Figura 7 - Histograma das temperaturas máximas anuais dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

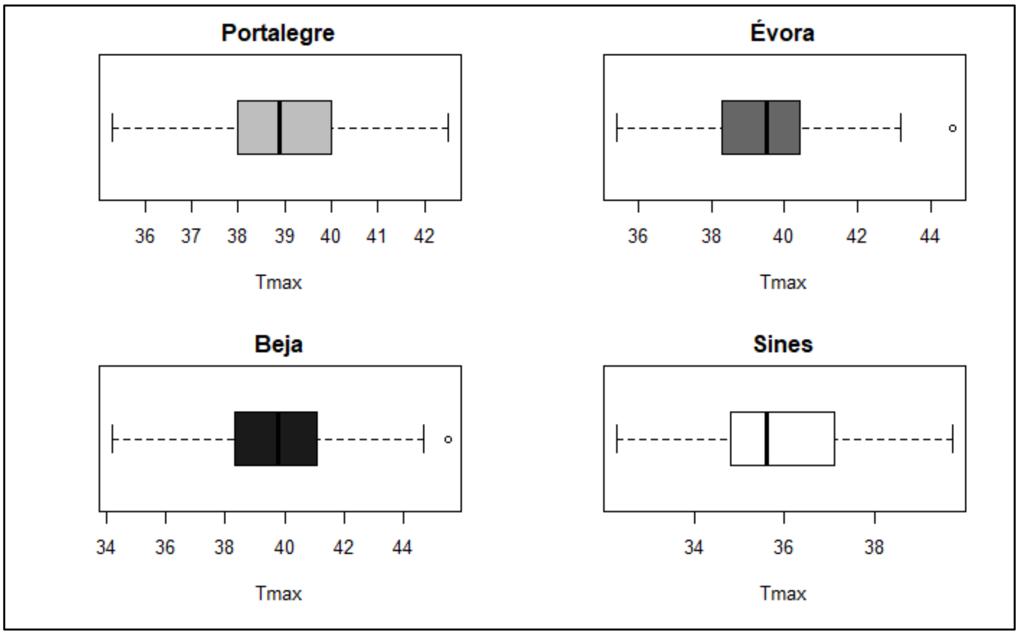


Figura 8 - *Boxplot* das temperaturas máximas anuais dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

Na Tabela 3, estão representadas as estatísticas descritivas de cada concelho, que permitem uma análise mais detalhada dos gráficos acima representados.

Tabela 3 - Estatísticas descritivas dos quatro concelhos do Alentejo.

	Min.	1roQu.	Mediana	Média	3roQu.	Max.	Curtose	Assimetria
Portalegre	35.30	38.00	38.90	38.96	40.00	42.50	-0.27	-0.04
Évora	35.40	38.30	39.50	39.43	40.40	44.60	-0.17	0.25
Beja	34.20	38.30	39.80	39.68	41.10	45.50	-0.20	-0.07
Sines	32.40	34.80	35.60	35.94	37.10	39.70	-0.60	0.18

Como podemos verificar através dos histogramas e das caixas-de-bigodes, Évora e Sines têm uma leve assimetria à direita, enquanto Portalegre e Beja têm uma leve assimetria à esquerda. O coeficiente de curtose em todos os concelhos do Alentejo é negativo podendo indicar que as distribuições das temperaturas máximas anuais nos quatro concelhos do Alentejo apresentam caudas mais leves do que a da distribuição Normal. De forma a ter uma ideia do comportamento de cauda das distribuições de temperaturas máximas anuais, foram construídos os ME-plots apresentados na Figura 9.

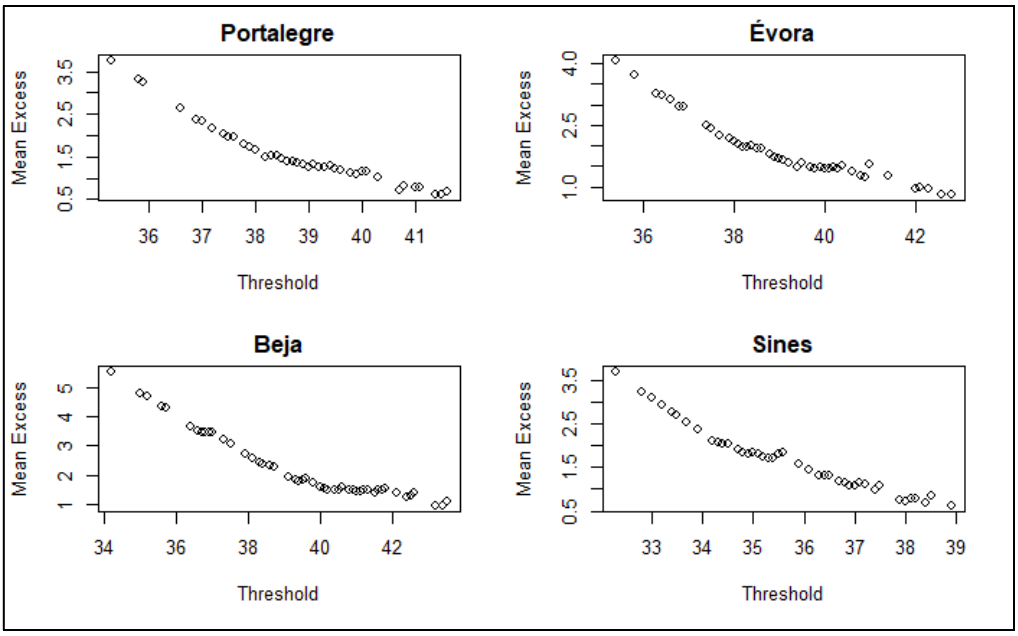


Figura 9 - ME-plot para as temperaturas máximas anuais dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

O comportamento monótono decrescente dos ME-plot indica que as caudas dos modelos subjacente aos dados das temperaturas máximas anuais dos quatro concelhos, Portalegre, Évora, Beja e Sines, não serão mais pesadas que a exponencial.
Dada a natureza dos dados é valida a aplicação de análise estatística de extremos, começando o ajustamento pelo modelo mais simples, o modelo Gumbel.

4.1.2 INFERÊNCIA ESTATÍSTICA

Na estimação dos parâmetros através do método de máxima-verosimilhança (MV), para o modelo Gumbel, usando a biblioteca `ismev`, foram obtidas as estimativas de MV dos parâmetros (λ, δ) e os erros padrão (s.e) correspondentes, apresentados na Tabela 4.

Tabela 4 - Ajustamento dos dados à distribuição Gumbel: estimação dos parâmetros localização, escala e erros padrão correspondentes.

	log L	$\hat{\lambda}$	$\hat{\delta}$	$s.e.(\hat{\lambda})$	$s.e.(\hat{\delta})$
Portalegre	-143.385	38.163	1.579	0.196	0.133
Évora	-151.277	38.516	1.721	0.213	0.149
Beja	-170.331	38.521	2.278	0.283	0.194
Sines	-143.305	35.123	1.536	0.190	0.134

De seguida, é apresentado o ajustamento dos dados ao modelo Gumbel, para o máximo anual das temperaturas, através do QQ-plot exponencial, obtidos através da biblioteca `evd`, onde podemos observar o fraco ajustamento das caudas à distribuição Gumbel (Figura 10).

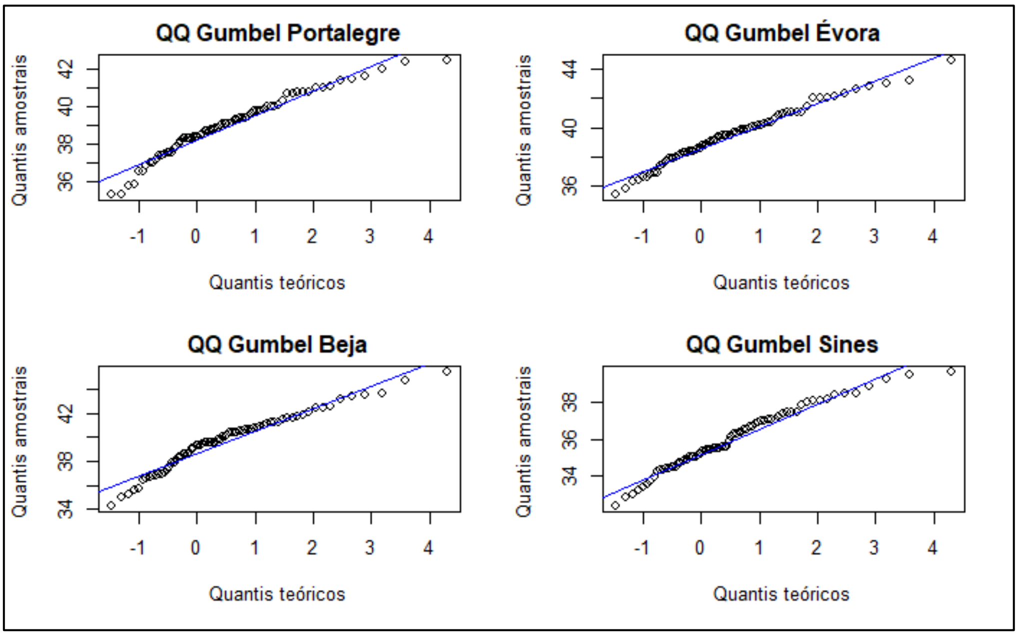


Figura 10 – Gráficos do ajustamento das temperaturas máximas anuais à distribuição Gumbel.

O fraco ajustamento à distribuição Gumbel pode ser visualizado com mais detalhe nos gráficos representados nas Figura 11 à Figura 14, obtidos através da função diagnóstico da biblioteca `ismev`.

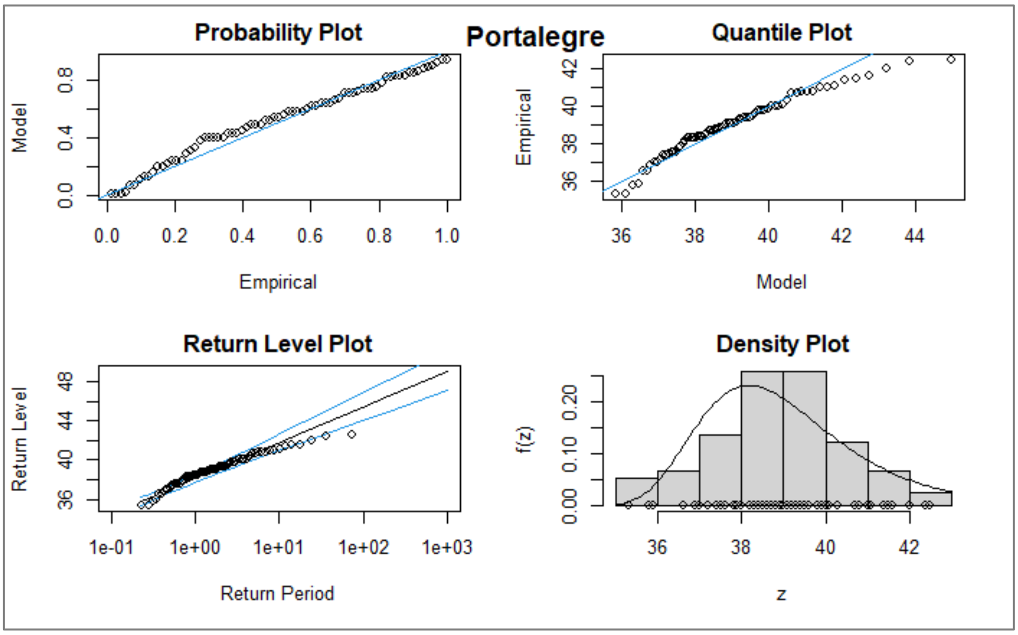


Figura 11 – Gráficos do ajustamento das temperaturas máximas anuais de Portalegre à distribuição Gumbel.

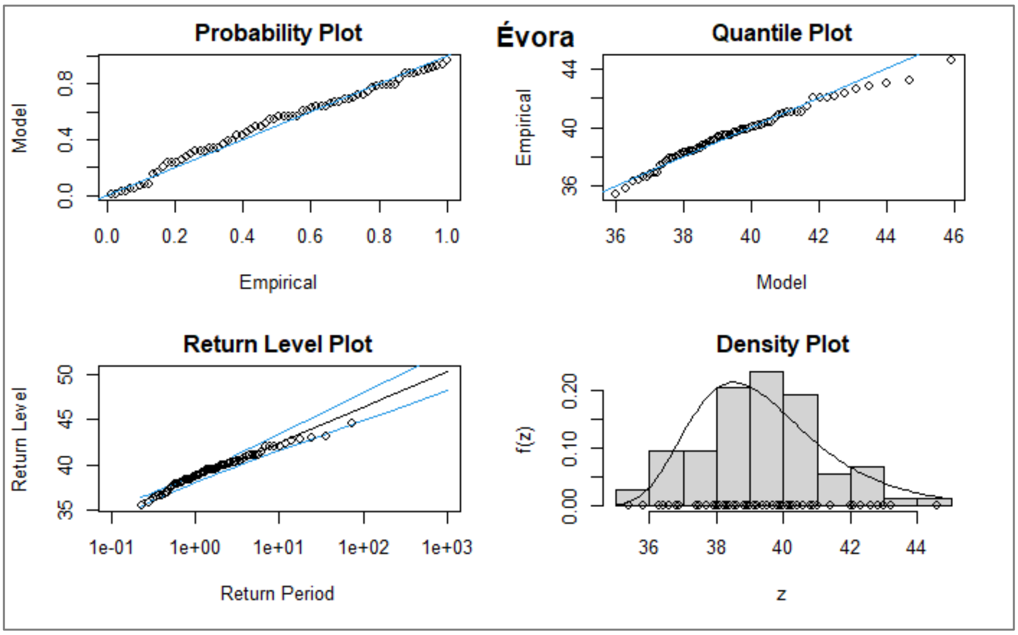


Figura 12- Gráficos do ajustamento das temperaturas máximas anuais de Évora à distribuição Gumbel.

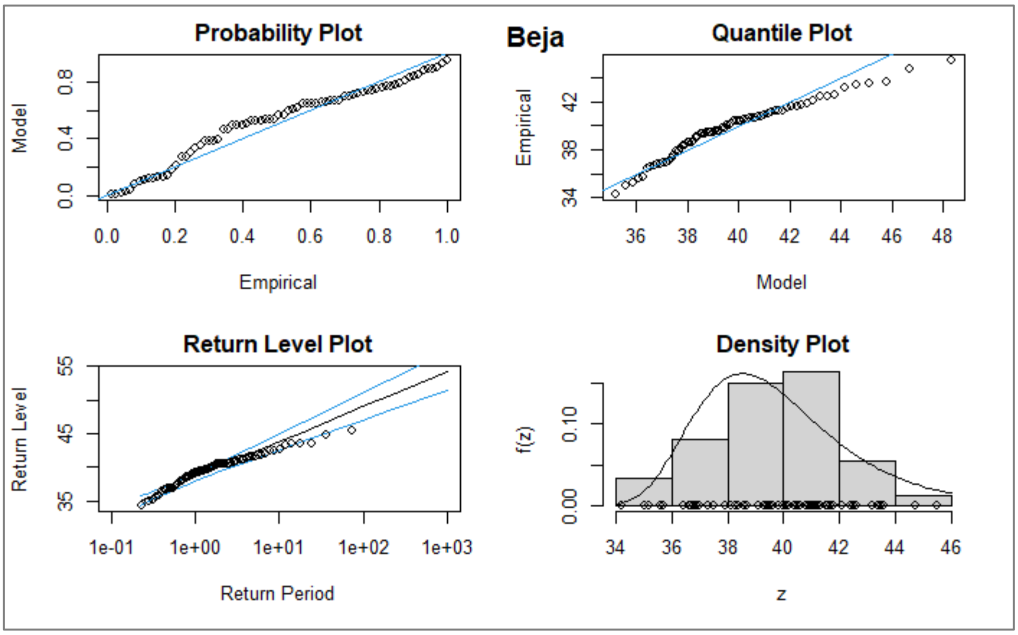


Figura 13 – Gráficos do ajustamento das temperaturas máximas anuais de Beja à distribuição Gumbel.

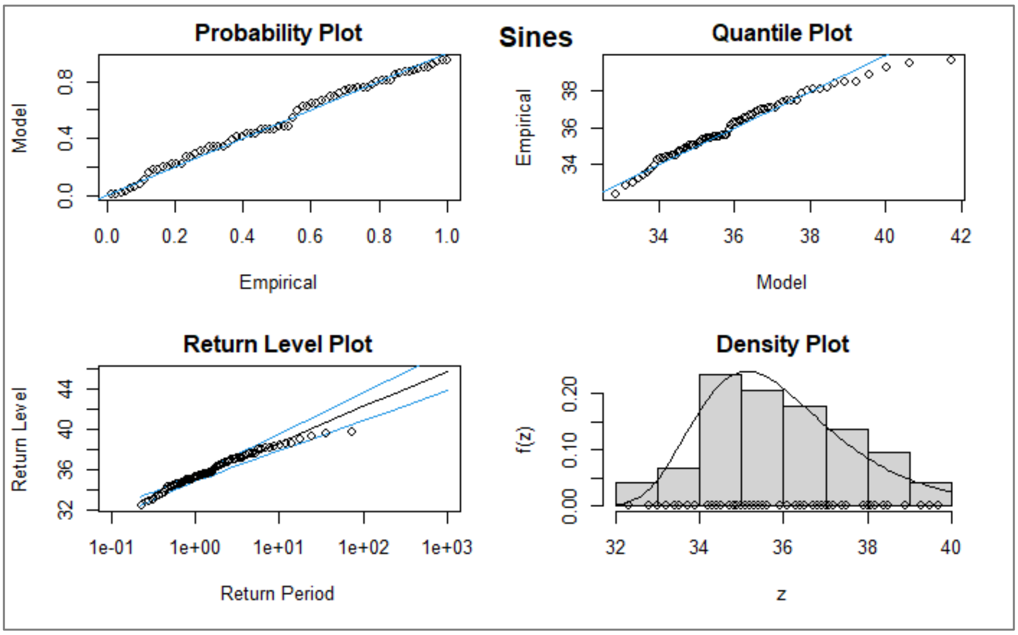


Figura 14 - Gráficos do ajustamento das temperaturas máximas anuais de Sines à distribuição Gumbel.

De seguida são apresentados os intervalos de confiança (IC) para os parâmetros de escala e localização obtidos pelo ajustamento à distribuição Gumbel, baseados na função de log-verosimilhança de perfil, com um grau de confiança de 95%, aproximadamente.

Na Figura 15, obtida pela biblioteca `ismev`, são apresentados os ICs. As duas retas verticais de cada extremo dos gráficos, correspondem ao limite inferior e superior dos intervalos, enquanto as retas verticais centrais correspondem às estimativas de MV dos parâmetros de localização e escala, apresentados na Tabela 5.

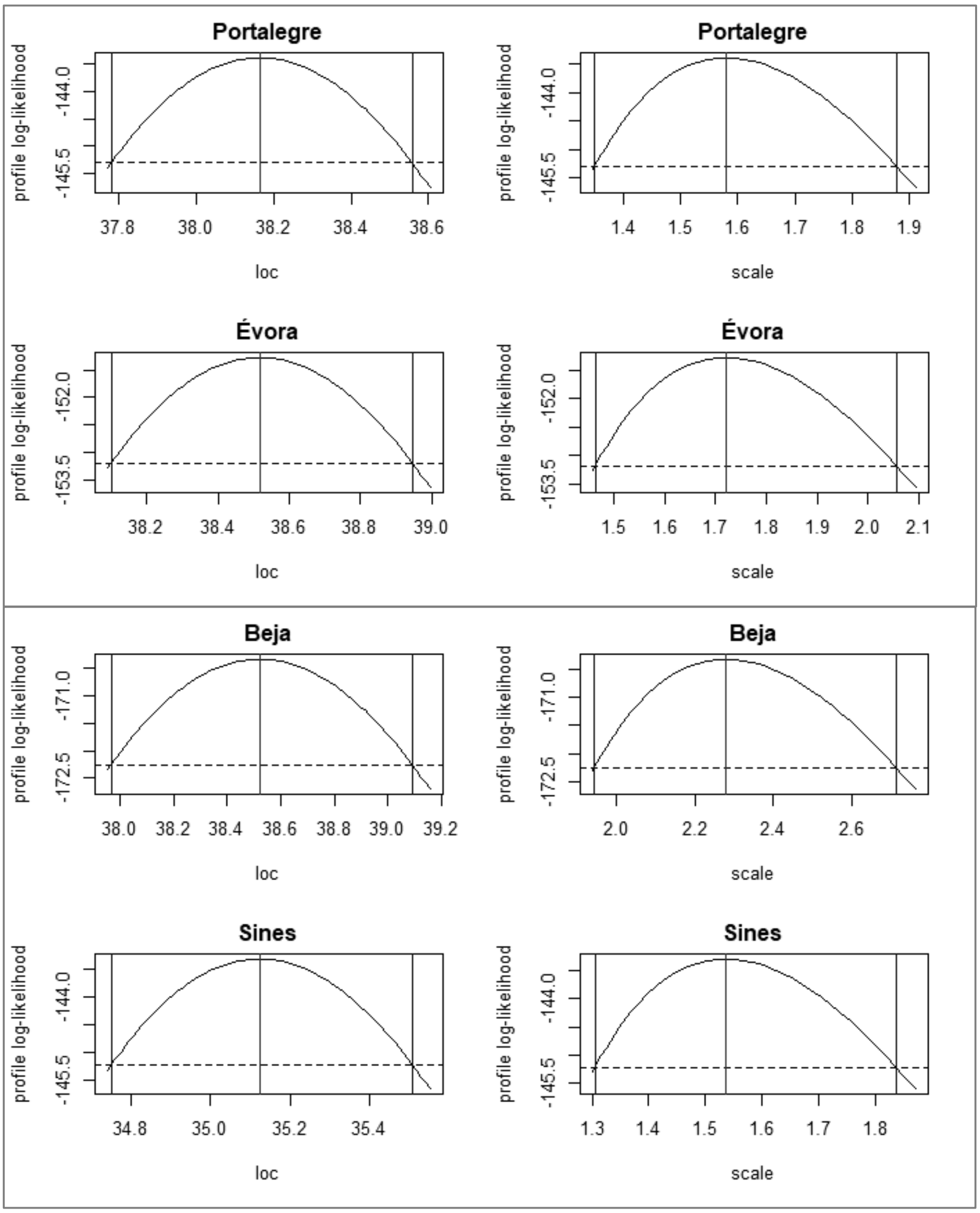


Figura 15 - Intervalos de confiança a 95% para os parâmetros de localização (esquerda) e de escala (direita) para os quatro concelhos.

Na Tabela 5, estão representados os ICs com grau de confiança 95% , para os parâmetros de localização (λ) e escala (δ), para os quatro concelhos.

Tabela 5 - Intervalos de confiança dos parâmetros de localização e de escala para um grau de confiança de 95% aproximadamente, do modelo Gumbel.

	$IC_{95\%}(\lambda)$	$IC_{95\%}(\delta)$
Portalegre	$(37.781; 38.558)$	$(1.349; 1.880)$
Évora	$(38.100; 38.946)$	$(1.463; 2.058)$
Beja	$(37.969; 39.091)$	$(1.944; 2.715)$
Sines	$(34.751; 35.506)$	$(1.305; 1.839)$

Uma vez que, os dados têm fraco ajustamento à distribuição Gumbel, passamos para uma distribuição mais complexa, a distribuição GEV de maneira a verificar se estes têm bom ajustamento. No entanto, primeiramente é necessário determinar uma estimativa preliminar do parâmetro de forma (γ) pelo método dos mínimos quadrados, representado na Figura 16.

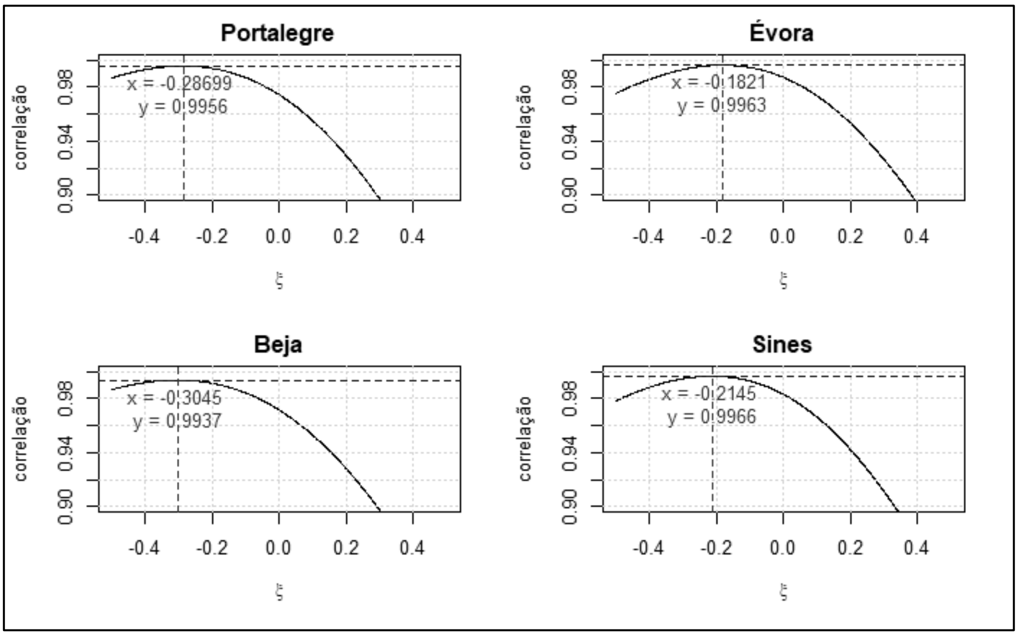


Figura 16 - Estimativas preliminares do parâmetro de forma (representado como ξ), associadas à correlação máxima, para os quatro concelhos.

As estimativas preliminares de γ correspondem aos valores de γ que maximizam o coeficiente de correlação no QQ-plot. Considerando $\gamma \in [-0.5, 0.5]$, as estimativas obtidas para cada concelho foram, $\hat{\gamma}_{Portalegre} = -0.287$, $\hat{\gamma}_{\text{Évora}} = -0.182$, $\hat{\gamma}_{Beja} = -0.305$ e $\hat{\gamma}_{Sines} = -0.215$, a que correspondem coeficientes de correlação de $r_{Portalegre} = 0.996$, $r_{\text{Évora}} = 0.996$, $r_{Beja} = 0.994$ e $r_{Sines} = 0.997$.

De seguida, foi feito o ajustamento à distribuição GEV, através da biblioteca `ismev`, que usa o método de MV na estimação dos parâmetros, onde se obteve os resultados apresentados na Tabela 6.

Tabela 6 - Ajustamento dos dados à distribuição GEV: estimação dos parâmetros de localização, escala, forma e os erros-padrão correspondentes.

	log L	$\hat{\boldsymbol{\lambda}}$	$\hat{\boldsymbol{\delta}}$	$\hat{\boldsymbol{\gamma}}$	$\boldsymbol{s.e.(\hat{\lambda})}$	$\boldsymbol{s.e.(\hat{\delta})}$	$\boldsymbol{s.e.(\hat{\gamma})}$
Portalegre	-137.358	38.419	1.6130	-0.305	0.208	0.147	0.075
Évora	-148.482	38.706	1.775	-0.204	0.229	0.160	0.072
Beja	-164.592	38.866	2.325	-0.281	0.297	0.205	0.063
Sines	-140.351	35.322	1.6102	-0.241	0.212	0.152	0.090

As estimativas dos parâmetros de forma são todas negativas indicando que as caudas das distribuições subjacentes aos dados são leves e com limite superior de suporte finito. Os valores da log-verosimilhança associados ao modelo GEV são inferiores aos obtidos com o ajustamento ao modelo Gumbel, indicando assim uma superioridade do ajuste GEV em relação ao Gumbel.

Os resultados do ajustamento obtido pelo modelo GEV podem também ser observados da Figura 17 à Figura 20, para cada concelho, obtidos através da função diagnóstico da mesma biblioteca.

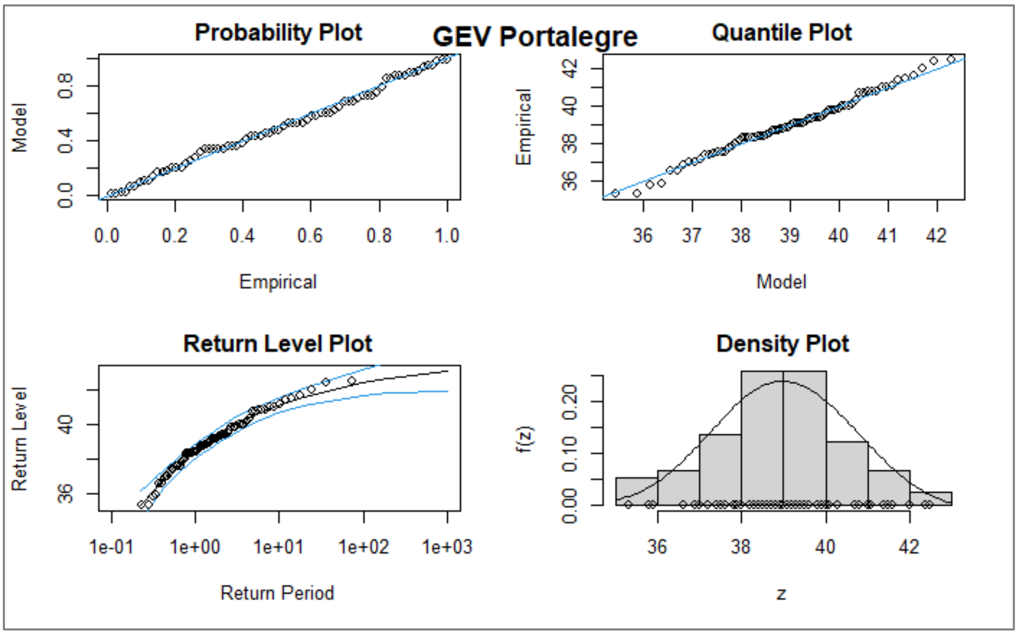


Figura 17 - Gráficos do ajustamento das temperaturas máximas anuais de Portalegre à distribuição GEV.

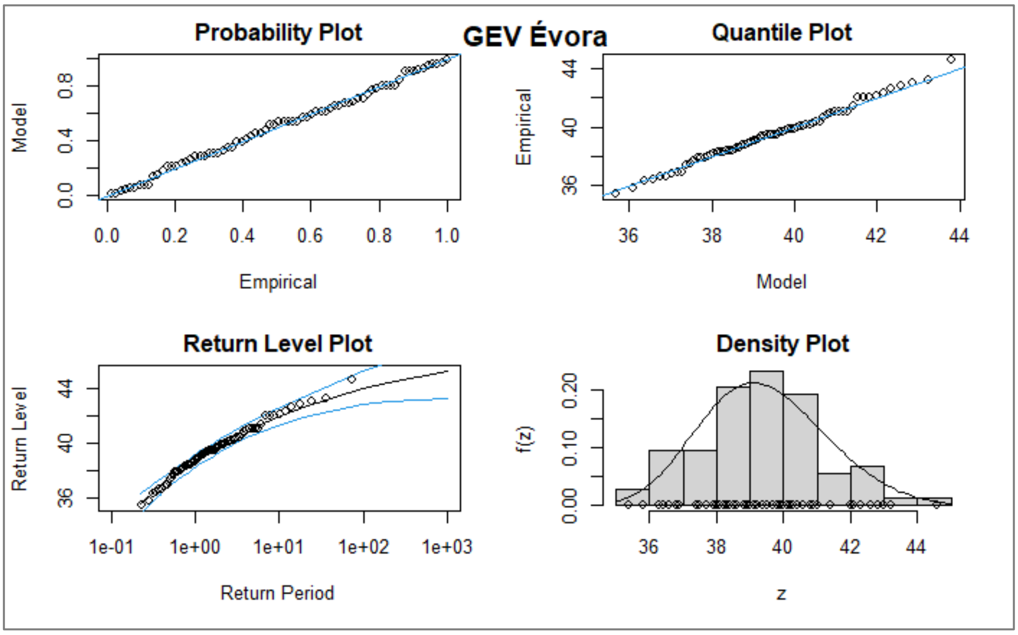


Figura 18 - Gráficos do ajustamento das temperaturas máximas anuais de Évora à distribuição GEV.

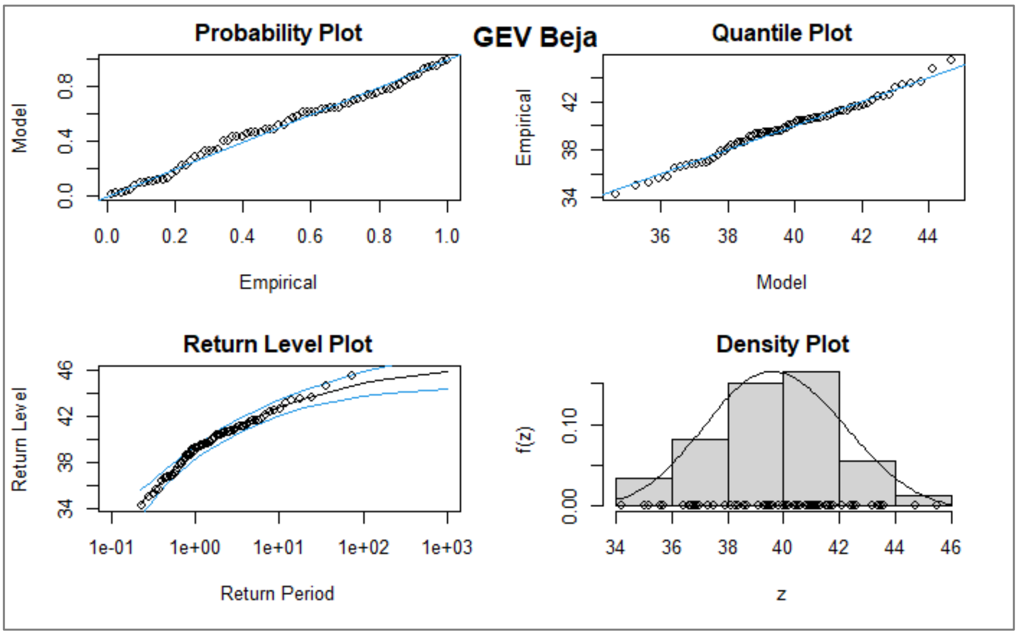


Figura 19 - Gráficos do ajustamento das temperaturas máximas anuais de Beja à distribuição GEV.

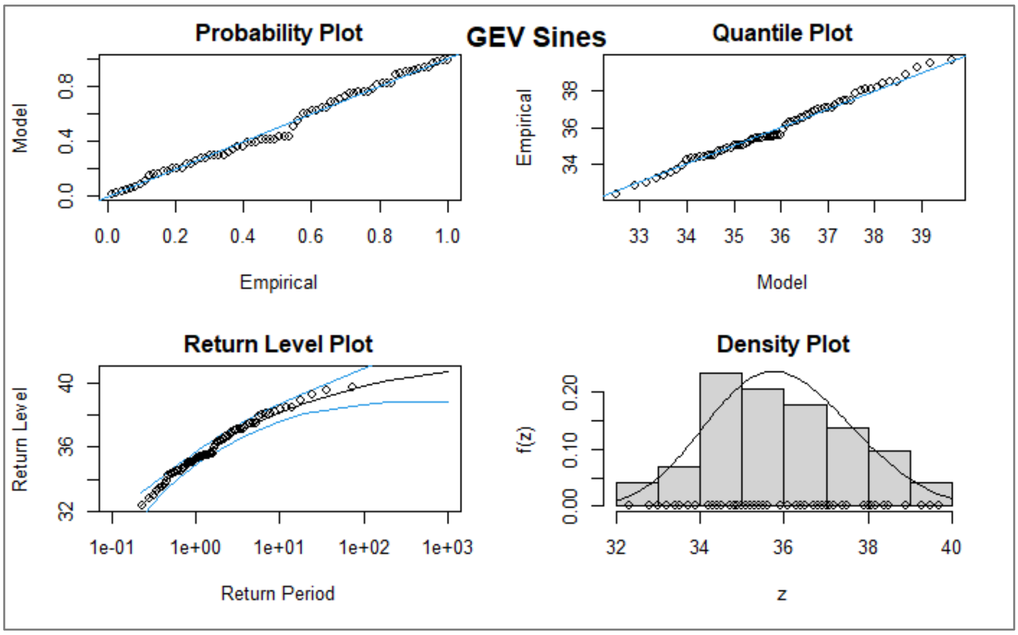


Figura 20 - Gráficos do ajustamento das temperaturas máximas anuais de Sines à distribuição GEV.

Da análise dos PP-plot e dos QQ-plot, parece existir um bom ajustamento dos dados à distribuição GEV, que é ainda validado pelo *Return level plot,* uma vez que em todos os concelhos, os dados estão dentro das bandas dos intervalos de confiança.

4.1.3 Estimação de quantidades de interesse

Usando as estimativas obtidas nos modelos GEV ajustados, para os parâmetros de forma, localização e escala, apresentados na Tabela 6, podem ser obtidas quantidades importantes em modelação de acontecimentos raros.

Na Tabela 7, estão representadas as estimativas de MV para os quantis extremais ($\hat{q}$) de probabilidades 0.90, 0.95, 0.99, para os níveis de retorno ($\hat{U}$) de 10, 50 e 100 anos, para o período de retorno ($\hat{T}$) da temperatura máxima observada em cada concelho, para a probabilidade de excedência desse mesmo máximo ($\hat{P}(Y > Tmax)$) e para o limite superior do suporte ($\hat{x}^F$).

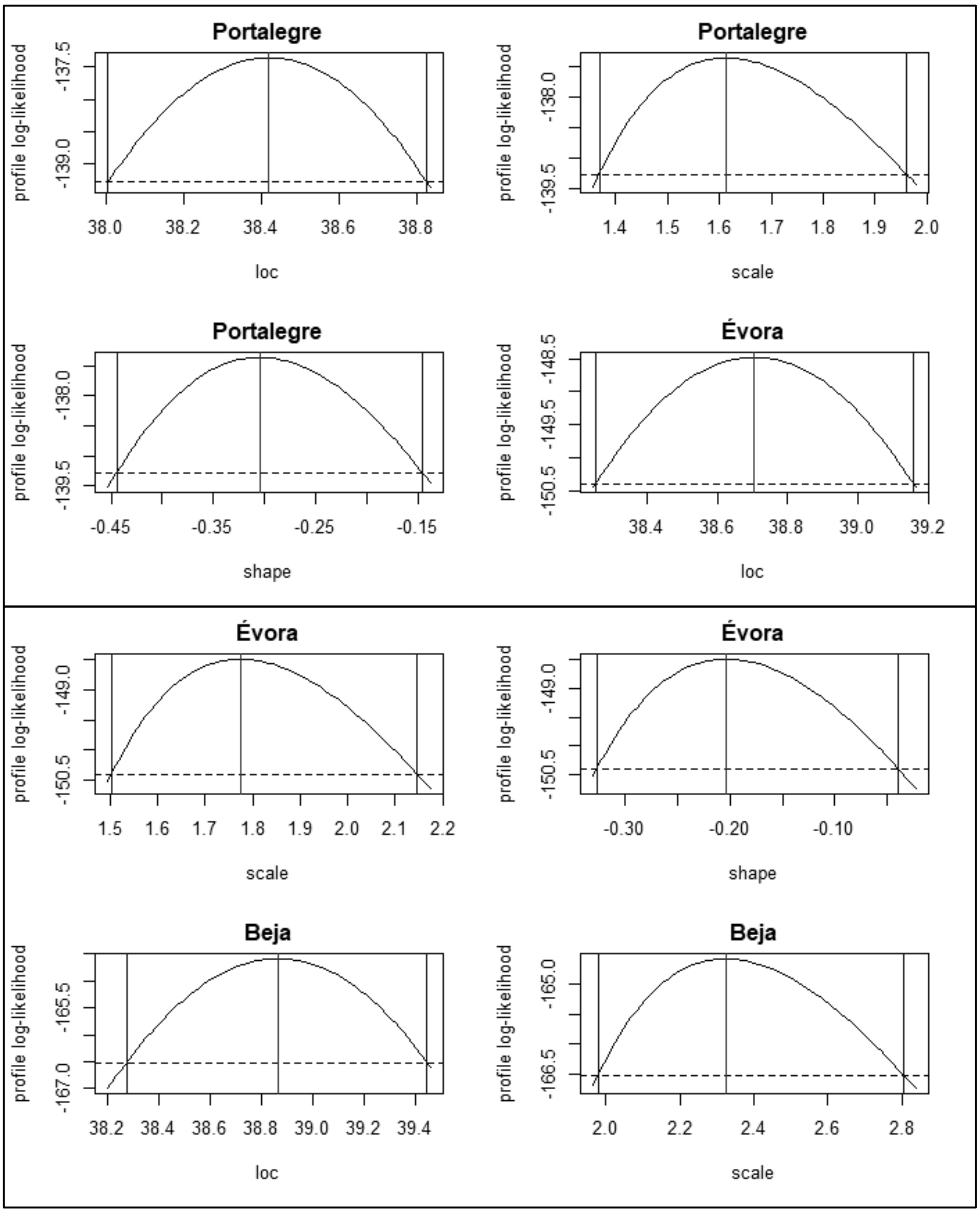
Tabela 7 - Estimativas de MV de outros parâmetros de interesse para o modelo GEV (Tmax=42.50°C*;44.60°C**;45.50°C***;39.70°C****)

	Portalegre	Évora	Beja	Sines
$\hat{q}_{0.90}$	41.05	41.91	42.74	38.12
$\hat{q}_{0.95}$	41.57	42.66	43.55	38.74
$\hat{q}_{0.99}$	42.41	44.00	44.86	39.80
$\hat{U}(10)$	41.05	41.91	42.74	38.12
$\hat{U}(50)$	42.10	43.48	44.37	39.39
$\hat{U}(100)$	42.41	44.00	44.86	39.80
$\hat{T}(Tmax)$	*127	**254	***321	****83
$\hat{P}(Y > Tmax)$	*0.0079	**0.0039	***0.0031	****0.0119
$\hat{x}^F$	43.71	47.42	47.13	42.00

Apenas 10% das temperaturas excedem 41.05ºC em Portalegre, 41.91ºC em Évora, 42.74ºC em Beja e 38.12ºC em Sines, no entanto, 1% das temperaturas excedem 42.41ºC em Portalegre, 44.00ºC em Évora, 44.86ºC em Beja e 39.80ºC em Sines. Relativamente ao período de retorno da temperatura máxima, para Portalegre é de 127 anos, 254 para Évora, 321 anos para Beja e 83 anos para Sines, sendo a probabilidade de excedência dessas temperaturas de 0.0079, 0.0039, 0.0031 e 0.0119, respetivamente. Analisando o limite superior de suporte para cada concelho, significa que, nas condições atuais, as temperaturas máximas podem atingir os valores de 43.71ºC, 47.42ºC, 47.13ºC, e 42.00ºC respetivamente para Portalegre, Évora, Beja e Sines.

4.1.4 INTERVALOS DE CONFIANÇA PARA OS PARÂMETROS

De seguida são apresentados, na Figura 21, os ICs para os parâmetros de localização, escala e forma baseados na função de log-verosimilhança de perfil, com grau de confiança 95% aproximadamente, os quais foram obtidos através da biblioteca `evd`. As duas rectas verticais dos extremos dos gráficos, correspondem ao limite inferior e superior dos intervalos, enquanto as rectas verticais centrais correspondem às estimativas de MV dos parâmetros, apresentados na Tabela 18.



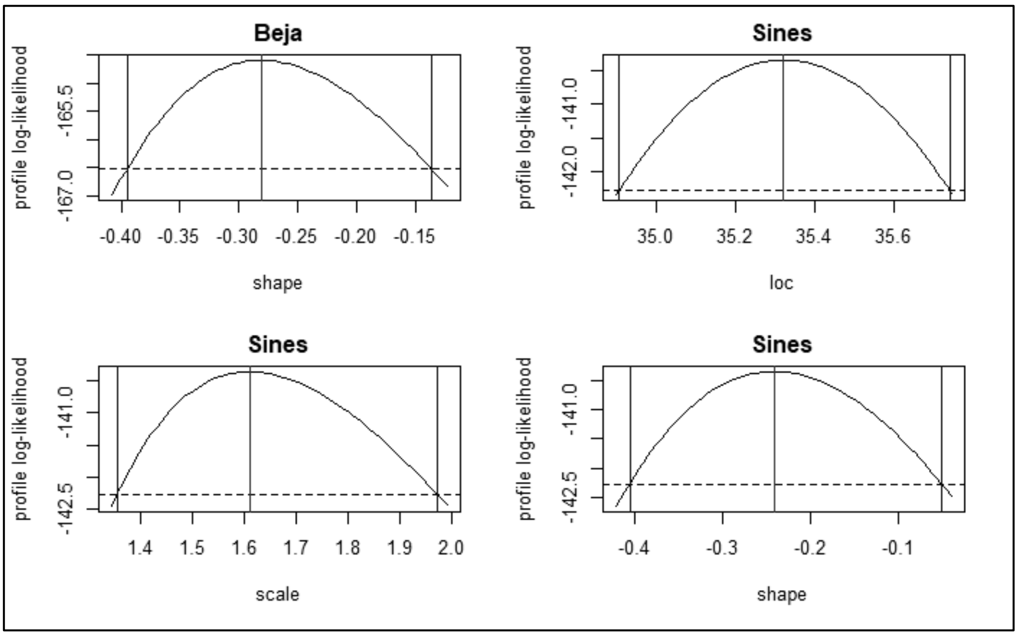


Figura 21 - ICs dos parâmetros de forma ($shape - \gamma$), localização ($loc - \lambda$) e escala ($scale - \delta$), para os quatro concelhos.

Na Tabela 8, estão representados os ICs com grau de confiança 95%, para os parâmetros de forma (γ), localização (λ) e escala(δ), para os quatro concelhos.

Tabela 8 - Intervalos de confiança dos parâmetros localização, escala e forma para um grau de confiança de 95% aproximadamente, do modelo GEV.

	$\boldsymbol{IC_{95\%}(\lambda)}$	$\boldsymbol{IC_{95\%}(\delta)}$	$\boldsymbol{IC_{95\%}(\gamma)}$
Portalegre	(38.004; 38.825)	(1.370; 1.961)	(-0.444; -0.146)
Évora	(38.253; 39.158)	(1.504; 2.146)	(-0.326; -0.040)
Beja	(38.272; 39.445)	(1.980; 2.806)	(-0.395; -0.137)
Sines	(34.906; 35.742)	(1.356; 1.972)	(-0.405; -0.051)

Nos quatro concelhos observamos que $0 \notin IC_{95\%}(\gamma)$, logo existe evidência estatística, ao nível de significância de 5%, que o modelo GEV é o que melhor se ajusta aos dados das temperaturas máximas dos quatro concelhos.

4.1.5 VALIDAÇÃO DO MODELO E ESTATÍSTICAS DE AJUSTAMENTO

Os testes Deviance (D), de Rao (V) e LAN (T), descritos anteriormente, serão aplicados às hipóteses: $H_0: \gamma = 0\ (Gumbel)\ vs\ H_1: \gamma \neq 0\ (GEV)$, estando os resultados sumariados na Tabela 9.

Tabela 9 - Valores observados da estatística deviance (D), deviance corrigida(D^*), teste de Rao (V) e Rao corrigido (V^{2*}) e teste de LAN (T) e LAN corrigido (T^*), para os dados das temperaturas máximas dos quatro concelhos.

	$\mathbf{D}$	$\mathbf{D}^*$	$\mathbf{V}$	$\mathbf{V}^{2*}$	$\mathbf{T}$	$\mathbf{T}^*$
Portalegre (valor-p)	12.05	11.61 (<0.001)	1095.96	7.16 (0.007)	-1.85	-2.67 (0.007)
Évora (valor-p)	5.59	5.38 (0.020)	554.34	3.62 (0.057)	-1.31	-1.90 (0.057)
Beja (valor-p)	11.48	11.06 (<0.001)	1023.41	6.68 (<0.001)	-1.78	-2.59 (0.010)
Sines (valor-p)	5.91	5.69 (0.017)	494.74	3.23 (0.072)	-1.24	-1.79 (0.07)

Os resultados obtidos para os três testes de hipóteses, considerando um nível significância de 10%, permitem-nos rejeitar a hipótese H_0, existindo assim evidências estatísticas de que o modelo que melhor se ajusta aos dados é o modelo GEV.

Neste estudo, os modelos também foram comparados com base nos critérios de informação de ajustamento AIC e BIC, apresentados na Tabela 10.

Tabela 10 - Critérios de Informação de Akaike e Bayesiano dos modelos GEV e Gumbel.

		AIC	BIC
PORTALEGRE	GUMBEL	290.77	295.35
	GEV	280.72	287.59
ÉVORA	GUMBEL	306.55	311.14
	GEV	302.96	309.85
BEJA	GUMBEL	344.66	349.24
	GEV	335.18	342.05
SINES	GUMBEL	290.61	295.19
	GEV	286.70	293.57

É de observar que no caso do ajustamento à GEV, os valores dos critérios de informação assumem valores inferiores relativamente aos valores do ajustamento à Gumbel, mostrando mais uma vez que, o modelo a adotar é o modelo GEV.

4.1.6 MODELAÇÃO DE FENÓMENOS NÃO-ESTACIONÁRIOS

Como temos vindo a referir os fenómenos ambientais raramente são estacionários, e uma análise da série temporal apresentada na Figura 6, deixa transparecer visualmente a presença de uma sazonalidade e tendência.

Por tal motivo, e como já foi referido anteriormente, considera-se que o parâmetro de localização é uma função linear do tempo, t, representada na Figura 22, sendo,

$$\lambda(t) = \beta_0 + \beta_1 t, \text{ com } t = 1, \dots, 73 \text{ anos.}$$

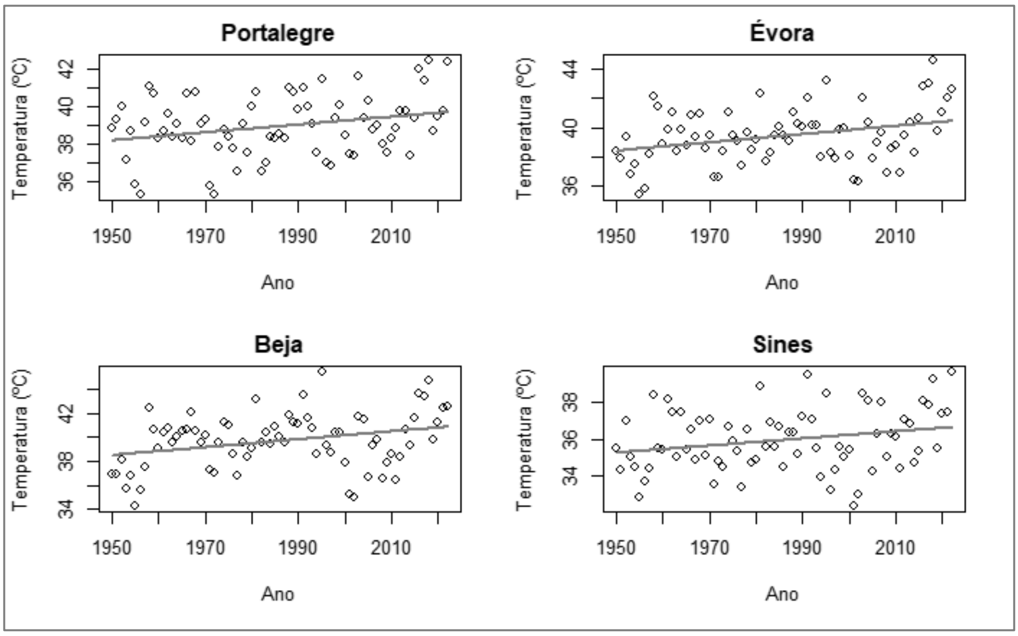


Figura 22 - Ajustamento linear à nuvem das temperaturas máximas anuais do Alentejo.

Como podemos verificar, as retas ajustadas à localização do modelo dos máximos anuais, apresentam um declive positivo pouco acentuado, evidenciando estarmos na presença de fenómenos não estacionários.

Através, da estimação MV da tendência linear para a função $\lambda(t) = \beta_0 + \beta_1 t$, foi possível obter as estimativas para os parâmetros a localização β_0 e β_1, escala δ e forma γ do modelo GEV e os respetivos erros padrão apresentados na Tabela 11.

Tabela 11 - Estimação de MV do modelo GEV($\lambda(t), \delta, \gamma$).

	log L	$\widehat{\boldsymbol{\beta_0}}$ $(\boldsymbol{s.e.}(\widehat{\boldsymbol{\beta_0}}))$	$\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$ $(\boldsymbol{s.e.}(\widehat{\boldsymbol{\beta_1}}))$	$\widehat{\boldsymbol{\delta}}$ $\boldsymbol{s}(.\boldsymbol{e.}(\widehat{\boldsymbol{\delta}}))$	$\widehat{\boldsymbol{\gamma}}$ $(\boldsymbol{s.e.}(\widehat{\boldsymbol{\gamma}}))$
Portalegre (valor-p)	-133.497	37.653 (0.334)	0.023 (0.007)	1.587 (0.162)	-0.387 (0.106)
Évora (valor-p)	-144.759	37.748 (0.417)	0.028 (0.010)	1.741 (0.162)	-0.264 (0.083)
Beja (valor-p)	-161.055	37.664 (0.539)	0.034 (0.013)	2.254 (0.198)	-0.306 (0.059)
Sines (valor-p)	-138.273	34.681 (0.381)	0.019 (0.009)	1.615 (0.153)	-0.295 (0.090)

Podemos verificar que, as estimativas para o declive da tendência linear são positivas, mas próximas de zero. Deste modo, espera-se que, em média haja um aumento de 0.023ºC por ano em Portalegre, de 0.028ºC por ano em Évora, de 0.034ºC por ano em Beja e de 0.019ºC por ano em Sines.

4.1.7 TESTES DE HIPÓTESES E SELEÇÃO DO MELHOR MODELO

Foi realizado também o teste da razão de verosimilhanças (Tabela 12), para se averiguar se a modelação não-estacionária era preferível ao modelo estacionário (são modelos encaixados), e aplicados os Teste Deviance (D) e Deviance corrigida (D^*) às hipóteses:

$$H_0\text{: } GEV\ Estacionário\ vs\ H_1\text{: } GEV\ Não - Estacionário,$$

Tabela 12 - Valores observados da estatística de deviance (D), deviance corrigida(D^*), na verificação do modelo Estacionário vs Não-Estacionário para os dados das temperaturas máximas dos quatro concelhos.

	D	**D***
Portalegre ((valor-p)	7.722	7.437 (0.006)
Évora (valor-p)	7.446	7.171 (0.007)
Beja (valor-p)	39.665	38.200 (<0.001)
Sines (valor-p)	4.155	4.001 (0.045)

Para um nível significância de 5%, a hipótese H_0 é rejeitada para todos os concelhos, sugerindo assim que, o melhor modelo que se ajusta aos dados é o não estacionário.

O modelo estacionário (GEV_E) e não-estacionário (GEV_{NE}) também foram comparados, com base nos critérios de informação de ajustamento AIC e BIC, apresentados na Tabela 13.

Tabela 13 - Critérios de Informação de Akaike e Bayesiano dos modelos Estacionário e Não-Estacionário.

		AIC	**BIC**
PORTALEGRE	GEV_E	280.72	287.59
	GEV_{NE}	274.99	284.16
ÉVORA	GEV_E	302.96	309.85
	GEV_{NE}	297.52	306.68
BEJA	GEV_E	335.18	342.05
	GEV_{NE}	330.11	339.27
SINES	GEV_E	286.70	293.57
	GEV_{NE}	284.55	293.71

Analisando, os valores de AIC e BIC dos modelos estacionário e não estacionário observamos que os valores associados aos modelos não estacionários são inferiores aos valores dos modelos estacionários, para os concelhos de Portalegre, Évora e Beja, indicando que os modelos não estacionários se ajustam melhor aos dados.

No concelho de Sines, os valores de BIC são semelhantes e o valor do AIC para o modelo não estacionário é ligeiramente inferior ao AIC do modelo estacionário, indicando que possivelmente a adição de uma tendência linear no parâmetro de localização, não seja, estatisticamente tão vantajoso como os restantes casos (note-se que o valor-p do TRV é de 4.5%).

Na Figura 23 mostra-se a série temporal das previsões dos níveis de retorno a 10 anos para os quatro concelhos do Alentejo, para os anos 2023 a 2122.

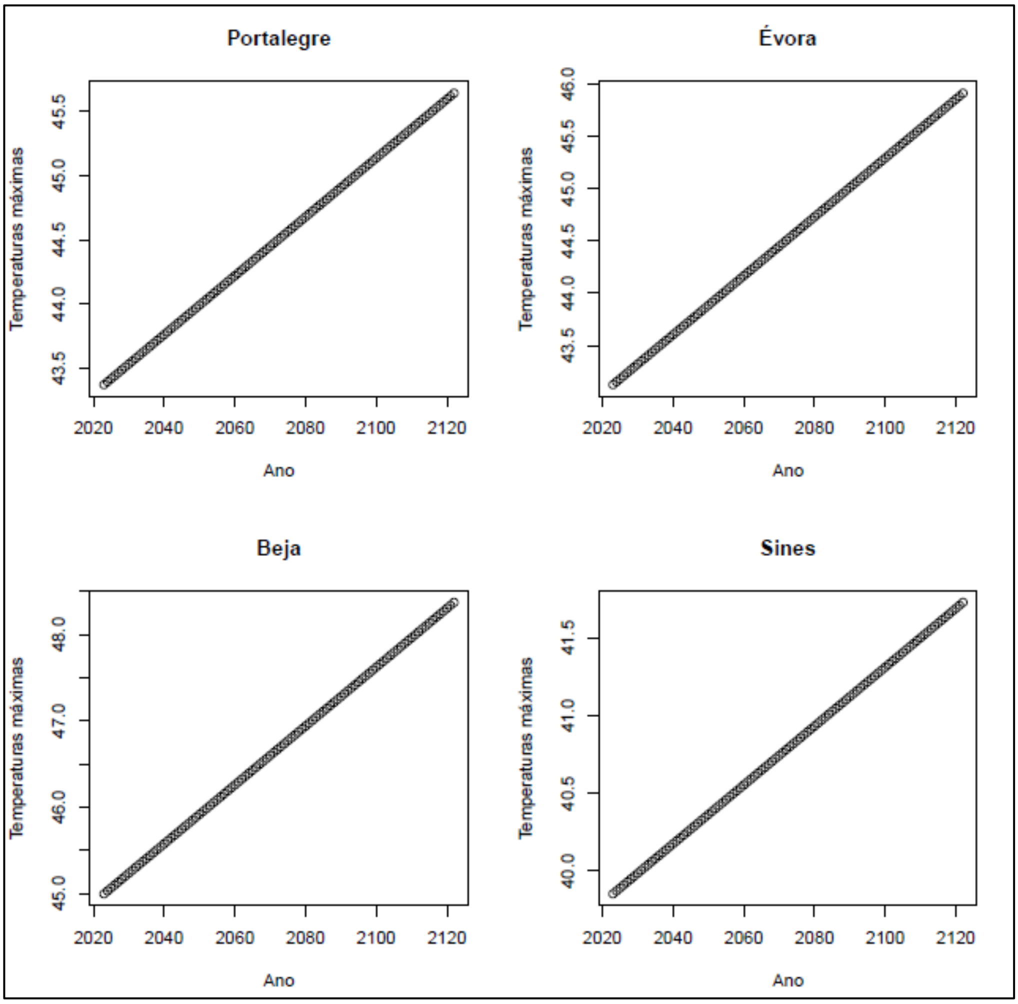


Figura 23 - Séries temporais dos níveis de retorno a 10 anos para os quatro concelhos do Alentejo, para os anos 2023 a 2122, para GEV não-estacionária.

De acordo com modelo GEV não-estacionário, espera-se que, em média, Portalegre atinja em 2040 a temperatura máxima de aproximadamente 43°C e em 2120 de aproximadamente 45°C. Da mesma forma, Évora registará em 2040 uma temperatura máxima de aproximadamente 43°C e em 2120 de aproximadamente 45.5°C, para Beja em 2040 a temperatura máxima prevista será de aproximadamente 45.5°C e em 2120 de aproximadamente 48°C, finalmente para Sines em 2040 a temperatura máxima prevista será aproximadamente 40°C e em 2120 de aproximadamente 41.5°C.

4.2 MODELO DAS K-MAIORES OBSERVAÇÕES

Ao considerar apenas a temperatura máxima anual, o modelo GEV não leva em conta outras informações relevantes como são, por exemplo, as k-maiores temperaturas anuais ($k \geq 1$).

4.2.1 ESTATÍSTICA DESCRITIVA PRELIMINAR

Para cada ano, e para cada concelho, foram selecionadas as 10 maiores ($k = 10$) temperaturas, no período de 1950 a 2022, onde se obteve para cada concelho uma série de temperaturas máximas com 730 observações, apresentadas na Figura 24.

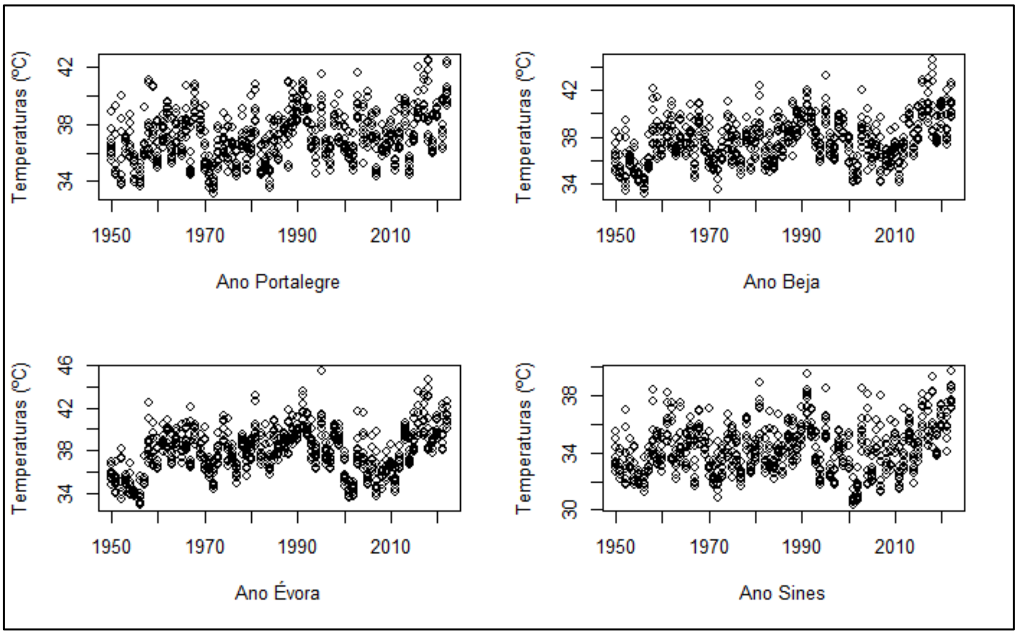


Figura 24 – Séries temporais das 10 maiores temperaturas por ano dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

Através do `runs test` e do *`difference.sign.test`* foi verificada a aleatoriedade dos dados, a um nível de significância de 5%, e a decisão foi a de não rejeitar a hipótese nula (de aleatoriedade), pois não existe evidências estatística significativas que os dados não são aleatórios, uma vez que, os valores-p foram todos superiores a 0.05.

Para uma análise preliminar de dados, foram feitos os gráficos histograma e o *boxplot* associados a este conjunto de dados, apresentados na Figura 25 e na Figura 26.

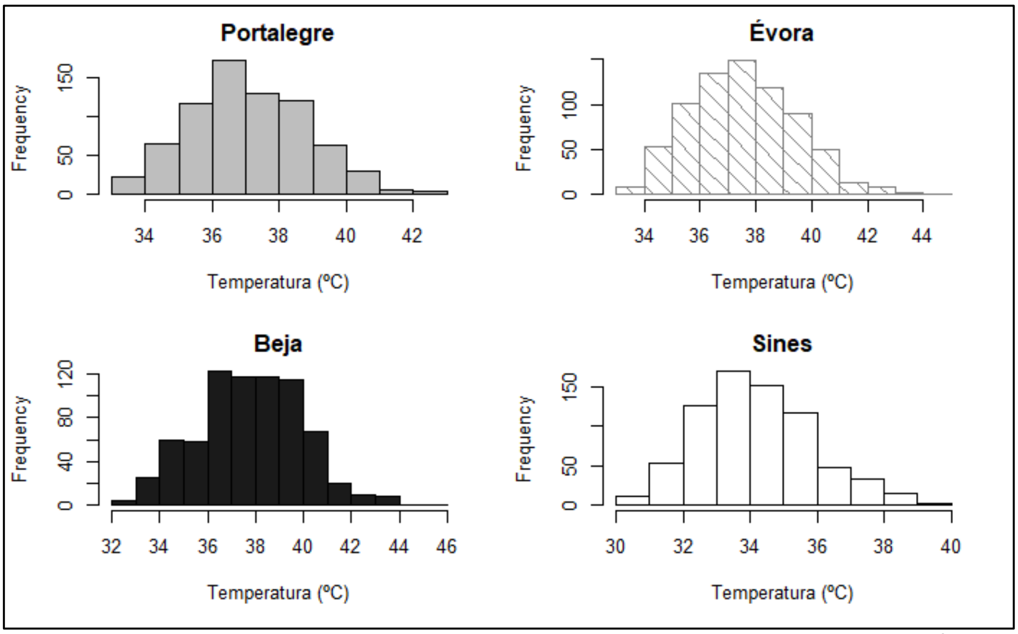


Figura 25 - Histograma das 10 maiores temperaturas por ano, dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

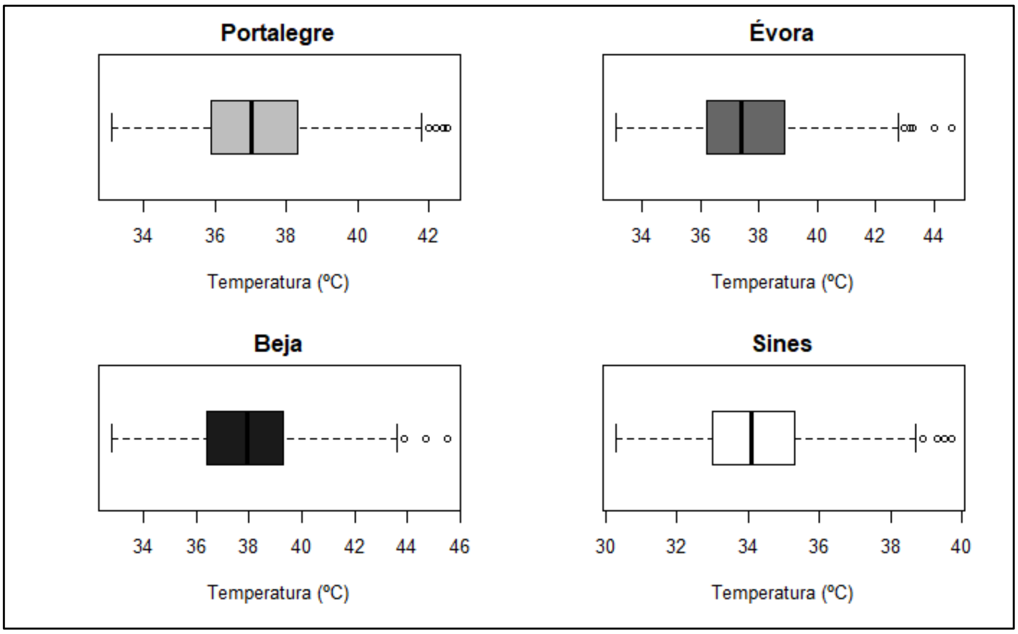


Figura 26 - Boxplot das 10 maiores temperaturas por ano, dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

Na Tabela 14, estão representadas as estatísticas descritivas de cada concelho, que permitem uma análise mais detalhada dos gráficos acima representados.

Tabela 14 - Estatísticas descritivas das 10 maiores temperaturas anuais dos quatro concelhos do Alentejo.

	Min.	1roQu.	Mediana	Média	3roQu.	Max.	Curtose	Assimetria
Portalegre	33.10	35.90	37.00	37.13	38.30	42.50	-0.21	0.29
Évora	33.10	36.20	37.40	37.59	38.90	44.60	-0.05	0.36
Beja	32.80	36.40	37.90	37.84	39.30	45.50	-0.16	0.12
Sines	30.30	33.00	34.10	34.22	35.30	39.70	-0.03	0.42

Ao considerarmos as 10 maiores temperaturas anuais, para cada concelho, através da análise dos histogramas e dos *boxplots* é observada uma assimetria positiva em todas as amostras, comprovada também pelo coeficiente de assimetria da Tabela 14. À semelhança do verificado na análise das temperaturas máximas anuais, o coeficiente de curtose é negativo, indiciando que as distribuições das 10 maiores temperaturas anuais dos quatro concelhos do Alentejo têm uma cauda mais leve do que a Normal. De forma a ter uma ideia do comportamento de cauda das distribuições das 10 maiores temperaturas anuais, foram construídos os ME-plots apresentados na Figura 27.

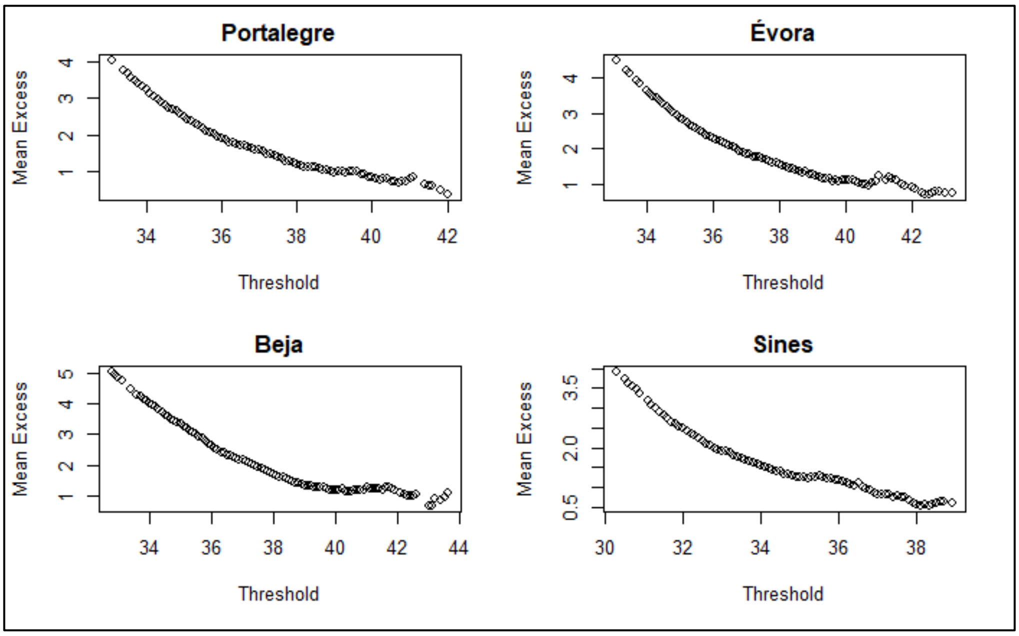


Figura 27 - ME-plot das 10 maiores observações das temperaturas por ano, dos quatro concelhos: Portalegre, Évora, Beja e Sines, no período de 1950 a 2022.

Mais uma vez, e à semelhança do que verificado na análise do tipo de cauda associado à distribuição da temperatura máxima anual, observamos que a cauda das distribuições subjacentes às 10 maiores temperaturas anuais, dos quatro concelhos, Portalegre, Évora, Beja e Sines, não serão mais pesadas que a exponencial. Dada a natureza dos dados e o facto de se verificar uma assimetria à direita, é valido a aplicação de análise estatística de extremos. A inferência será realizada admitindo que a série das 10 maiores temperaturas anuais é estacionária, e de seguida é realizada uma análise inferencial em contexto não estacionário. De modo a averiguar qual o valor de k a usar na modelação, iremos apresentar os resultados para $k = \{2, 5, 10\}$. De notar que $k = 1$ corresponde à análise da temperatura máxima anual, efetuada na secção anterior.

4.2.2 INFERÊNCIA ESTATÍSTICA

Na estimação dos parâmetros através do método de MV, para o modelo GEV, usando a biblioteca `ismev`, foram obtidas as estimativas dos parâmetros do modelo GEV_k (λ, δ, γ) e os respetivos erros padrão (s.e) para os valores de k em estudo, e apresentados na Tabela 15.

Tabela 15 - Estimativas dos parâmetros do modelo $GEV_k(\lambda, \delta, \gamma)$.

	$\log L$	$\hat{\lambda}$	$\hat{\delta}$	$\hat{\gamma}$	$s.e.(\hat{\lambda})$	$s.e.(\hat{\delta})$	$s.e.(\hat{\gamma})$
$k = 2$							
Portalegre	-198.54	38.98	1.45	-0.30	0.16	0.08	0.06
Évora	-221.55	39.44	1.72	-0.25	0.19	0.09	0.05
Beja	-240.71	39.79	1.97	-0.28	0.22	0.11	0.04
Sines	-199.28	35.95	1.46	-0.30	0.16	0.07	0.05
$k = 5$							
Portalegre	-248.68	39.38	1.28	-0.31	0.12	0.05	0.03
Évora	-278.62	40.05	1.48	-0.25	0.14	0.06	0.03
Beja	-314.51	40.56	1.63	-0.28	0.16	0.06	0.03
Sines	-233.87	36.40	1.26	-0.28	0.12	0.05	0.04
$k = 10$							
Portalegre	-97.92	39.57	1.15	-0.31	0.10	0.04	0.02
Évora	-162.60	40.33	1.35	-0.27	0.12	0.04	0.02
Beja	-198.66	40.86	1.44	-0.28	0.13	0.04	0.02
Sines	-65.67	36.55	1.16	-0.27	0.10	0.05	0.02

Para valores crescentes de k observa-se um decréscimo dos erros padrão. As estimativas dos parâmetros de localização e de forma não variam muito em função de k, o mesmo não acontecendo com as estimativas do parâmetro de escala que diminuem à medida que k aumenta. Como seria de esperar valor da log-verosimilhança também aumenta quando k aumenta.
Para todas as situações o valor obtido para o EVI é negativo, o que significa que, a distribuição terá uma cauda leve e com limite superior do suporte finito.

A qualidade do ajustamento para as k maiores observações das temperaturas, dos quatro concelhos, através das 2, 5 e das 10 maiores observações anuais, pode ser investigado através dos gráficos de diagnóstico, obtidos através da biblioteca `evd`. Nas Figura 28 à Figura 31, estão representados os gráficos de diagnósticos associados a k=2, e os gráficos relativos a k=5 e k=10 encontram-se no Apêndice A.

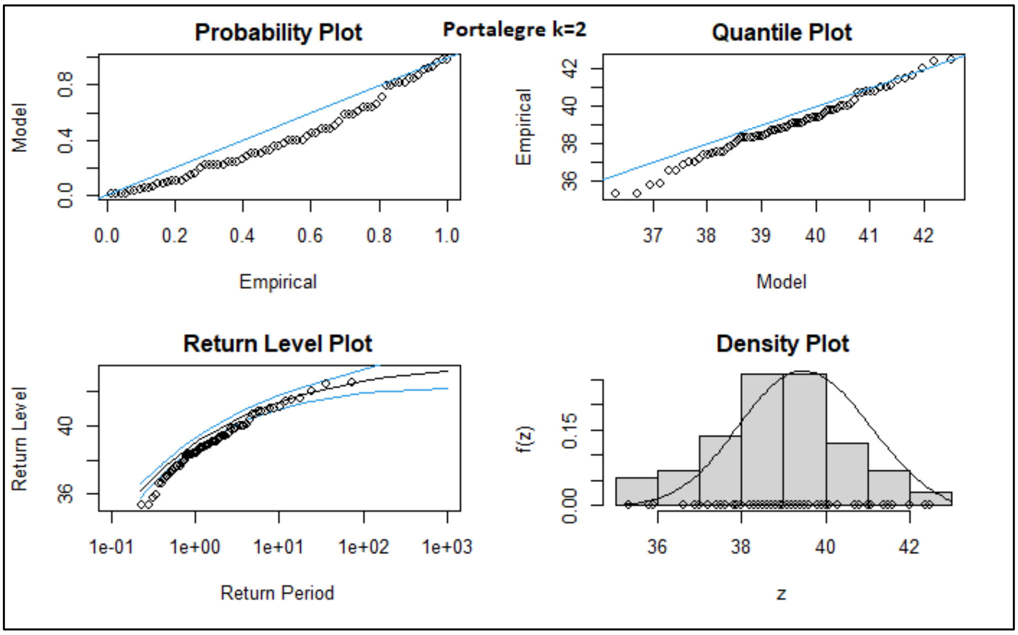


Figura 28 - Gráficos do ajustamento ao modelo GEV das 2 maiores observações, para Portalegre.

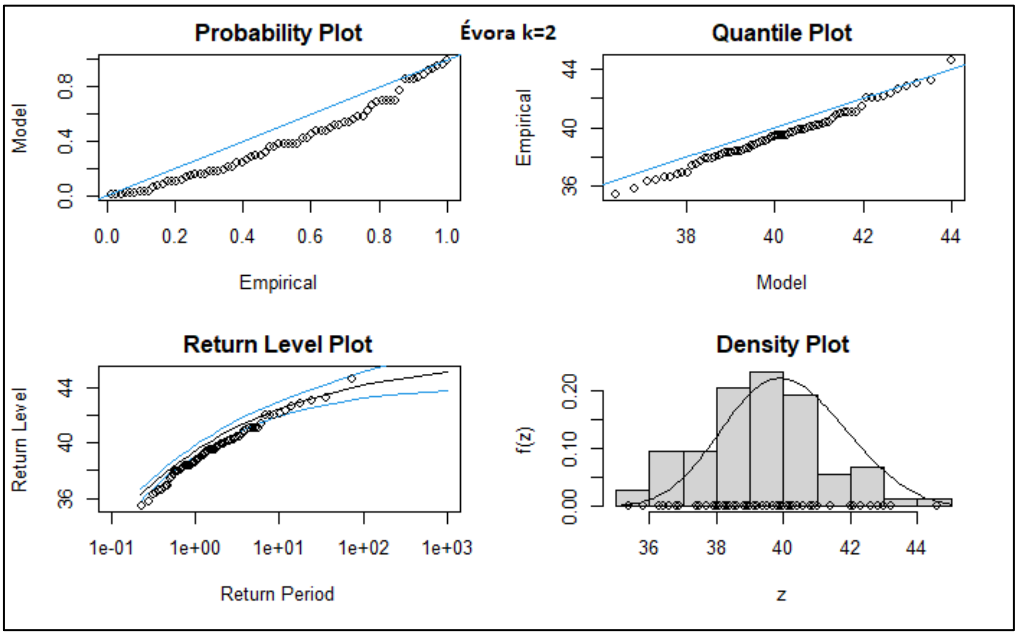


Figura 29 - Gráficos do ajustamento ao modelo GEV das 2 maiores observações, para Évora.

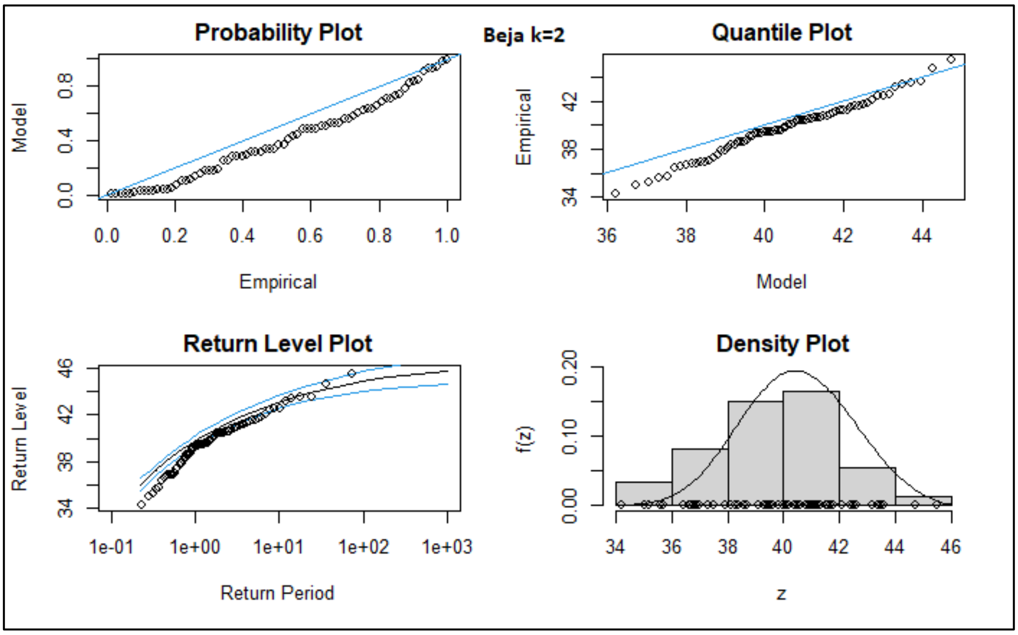


Figura 30 - Gráficos do ajustamento ao modelo GEV das 2 maiores observações, para Beja.

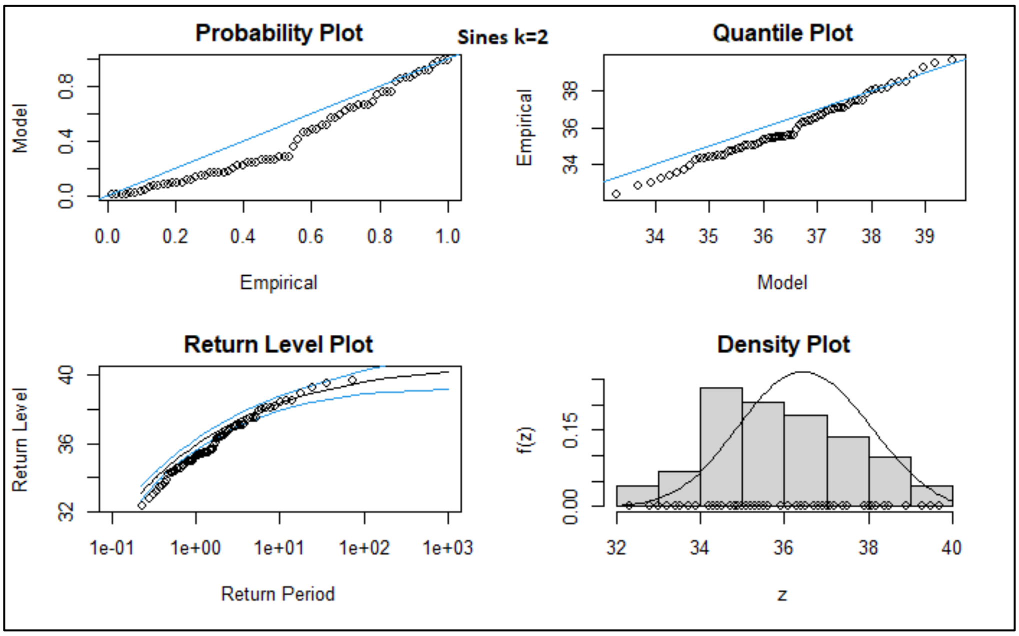


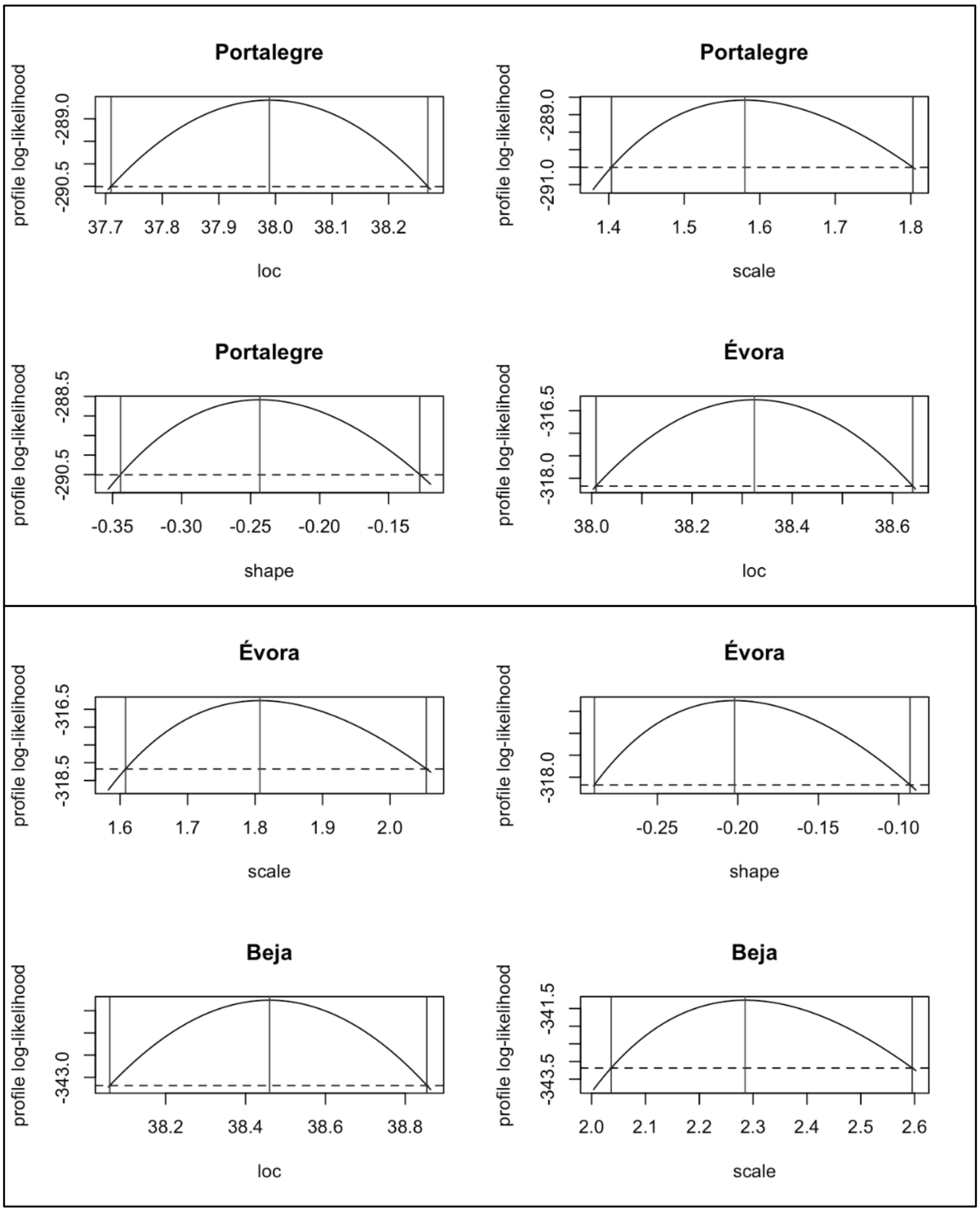
Figura 31 - Gráficos do ajustamento ao modelo GEV das 2 maiores observações, para Sines.

Como se pode observar, o ajustamento ao modelo GEV das 2 maiores observações, para os quatro concelhos é superior ao observado para $k = 5$ e $k = 10$. De qualquer modo, mesmo com $k = 2$, o ajuste obtido, é inferior ao que se obteve para o máximo anual ($k = 1$), como o que foi observado nas Figura 17 à Figura 20, onde parece existir um melhor ajustamento do modelo.

4.2.3 Intervalos de confiança para os parâmetros

De seguida são apresentados, na Figura 32, os ICs para os parâmetros de localização, escala e forma baseados na função de log-verosimilhança de perfil, com grau de confiança de 95% aproximadamente, os quais foram obtidos através da biblioteca `evd`.

As duas retas verticais dos extremos dos gráficos, correspondem ao limite inferior e superior dos intervalos, enquanto as retas verticais centrais correspondem às estimativas de MV dos parâmetros, através da distribuição GEV, baseados no modelo limite das k maiores observações, apresentados na Tabela 16.



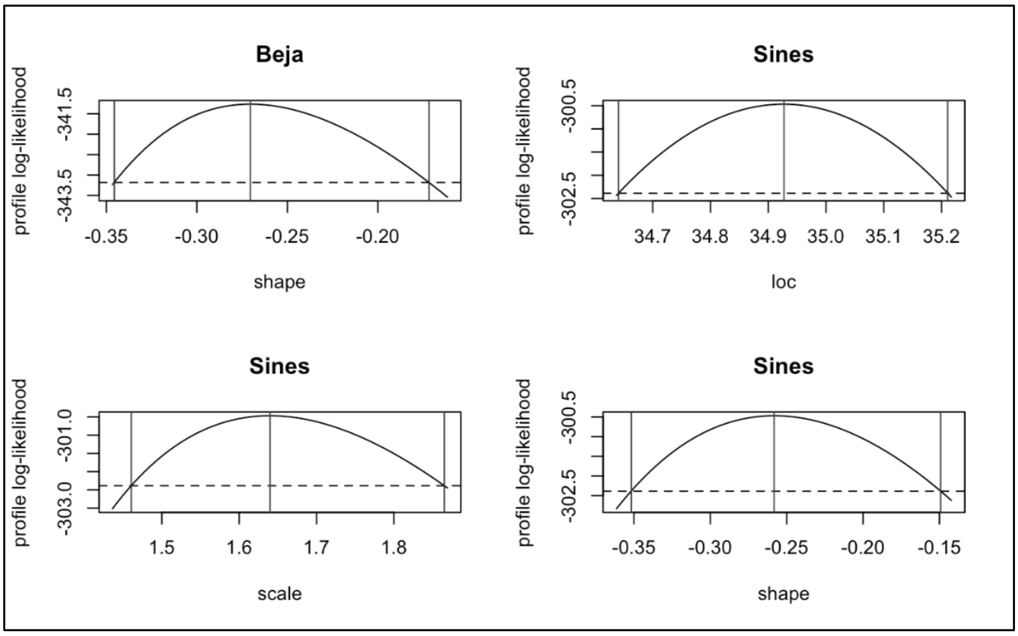


Figura 32 - IC baseados no *profile-likelihood* em função de o parâmetro de forma ($shape - \gamma$), localização($loc - \lambda$) e escala ($scale - \delta$), para os quatro concelhos, baseados no modelo limite das 2 maiores observações.

Na Tabela 16, estão representados os IC com grau de confiança de 95% , para os parâmetros de forma (γ), localização (λ) e escala (δ), para os três concelhos, baseados no modelo limite das 2 maiores observações.

Tabela 16 - Intervalos de confiança dos parâmetros localização, escala e forma para um grau de confiança de 95% aproximadamente, do modelo GEV, baseados no modelo limite das 2 maiores observações.

	$\boldsymbol{IC_{95\%}(\lambda)}$	$\boldsymbol{IC_{95\%}(\delta)}$	$\boldsymbol{IC_{95\%}(\gamma)}$
Portalegre	(37.710; 38.269)	(1.403; 1803)	(-0.344; -0.127)
Évora	(38.008; 3(.640)	(1.608; 2.053)	(-0.289; -0.093)
Beja	(38.060; 38.854)	(2.036; 2.595)	(-0.345; -0.172)
Sines	(34.641; 35.211)	(1.460; 1.865)	(-0.352; -0.149)

Nos quatro concelhos observamos que $0 \notin IC_{95\%}(\gamma)$, logo existe evidência estatística, ao nível de significância de 5%, que o modelo GEV_2 (baseados no modelo limite das 2 maiores observações), é o que melhor se ajusta aos dados das temperaturas máximos dos quatro concelhos, quando comparado com o modelo Gumbel.

4.2.4 SELEÇÃO DO MELHOR MODELO

De modo a selecionar o modelo que melhor se ajusta aos dados, na Tabela 17 é apresentada uma comparação dos critérios de informação de ajustamento AIC e BIC entre os modelos GEV (k=1) e GEV_2 (k=2).

Tabela 17 - Critérios de Informação de Akaike e Bayesiano dos modelos GEV e GEV2.

		AIC	**BIC**
PORTALEGRE	GEV	280.72	287.59
	GEV_2	403.08	412.03
ÉVORA	GEV	302.96	309.85
	GEV_2	449.10	458.05
BEJA	GEV	335.18	342.05
	GEV_2	487.42	496.37
SINES	GEV	286.70	293.57
	GEV_2	404.56	413.501

Como seria de esperar, os valores do AIC e do BIC são maiores para o modelo GEV_2 dado que a dimensão da amostra duplicou e o número de parâmetros manteve-se igual. Vamos de seguida prosseguir o estudo considerando o modelo GEV_2 de modo a poder comparar os resultados com os obtidos no modelo GEV.

4.2.5 ESTIMAÇÃO DE QUANTIDADES DE INTERESSE

Usando as estimativas obtidas nos modelos GEV_2 ajustados, para os parâmetros de forma, localização e escala, representados na Tabela 15, podem ser obtidas quantidades importantes em modelação de acontecimentos raros. Na Tabela 18 estão representadas as estimativas MV para os quantis extremais ($\hat{q}$) de probabilidades 0.90, 0.95, 0.99, para os níveis de retorno ($\widehat{U}$) de 10, 50 e 100 anos, para o período de retorno ($\widehat{T}$) da temperatura máxima observada em casa concelho, para a probabilidade de excedência desse mesmo máximo ($\hat{P}(Y > Tmax)$) e para o limite superior do suporte ($\hat{x}^F$), baseados no modelo limite das 2 maiores observações.

Tabela 18 - Estimativas MV de outros parâmetros para o máximo anual, baseado nas 2 maiores observações do alentejo, no período de 1950 a 2022 (Tmax=42.50°C*;44.60°C**;45.50°C***;39.70°C****).

	Portalegre	Évora	Beja	Sines
$\hat{q}_{0.90}$	41.36	42.40	43.09	38.35
$\hat{q}_{0.95}$	41.84	43.05	43.7769	38.84
$\hat{q}_{0.99}$	42.61	44.15	44.89	39.62
$\widehat{U}(10)$	41.36	42.40	43.09	38.35
$\widehat{U}(50)$	42.32	43.73	44.47	39.33
$\widehat{U}(100)$	42.61	44.15	44.89	39.62
$\widehat{T}(Tmax)$	*75.27	**248.02	***384.69	****126.53
$\widehat{P}(Y > Tmax)$	*0.013	**0.004	***0.003	****0.008
$\hat{x}^F$	43.85	46.38	46.82	40.86

Apenas 10% das temperaturas excedem 41.36ºC em Portalegre, 42.40ºC em Évora, 43.09ºC em Beja e 38.35ºC em Sines, no entanto, 1% das temperaturas excedem 42.61ºC em Portalegre, 44.15ºC em Évora, 44.89ºC em Beja e 39.62ºC em Sines. Relativamente ao período de retorno da temperatura máxima, para Portalegre é de 75 anos, 248 para Évora, 384 anos para Beja e 126 anos para Sines, sendo a probabilidade de excedência dessas temperaturas de 0.013, 0.004, 0.003, 0.008 respetivamente. Analisando o limite superior de suporte para cada concelho, significa que é extremamente improvável que as temperaturas venham a ser superiores a 43.85ºC, 46.38ºC, 46.82ºC, e 40.86ºC respetivamente para Portalegre, Évora, Beja e Sines.

4.2.6 MODELAÇÃO DE FENÓMENOS NÃO-ESTACIONÁRIOS

Como temos vindo a referir os fenómenos ambientais raramente são estacionários, e uma análise da série temporal apresentada na Figura 24, deixa transparecer visualmente a presença de uma sazonalidade e tendência.

Por tal motivo, e como já foi referido anteriormente, considera-se que o parâmetro de localização é uma função linear do tempo, t, representada na Figura 33, para o modelo limite das 2 maiores observações, sendo,

$$\lambda(t) = \beta_0 + \beta_1 t, \text{ com } t = 1, \dots, 73 \text{ anos}$$

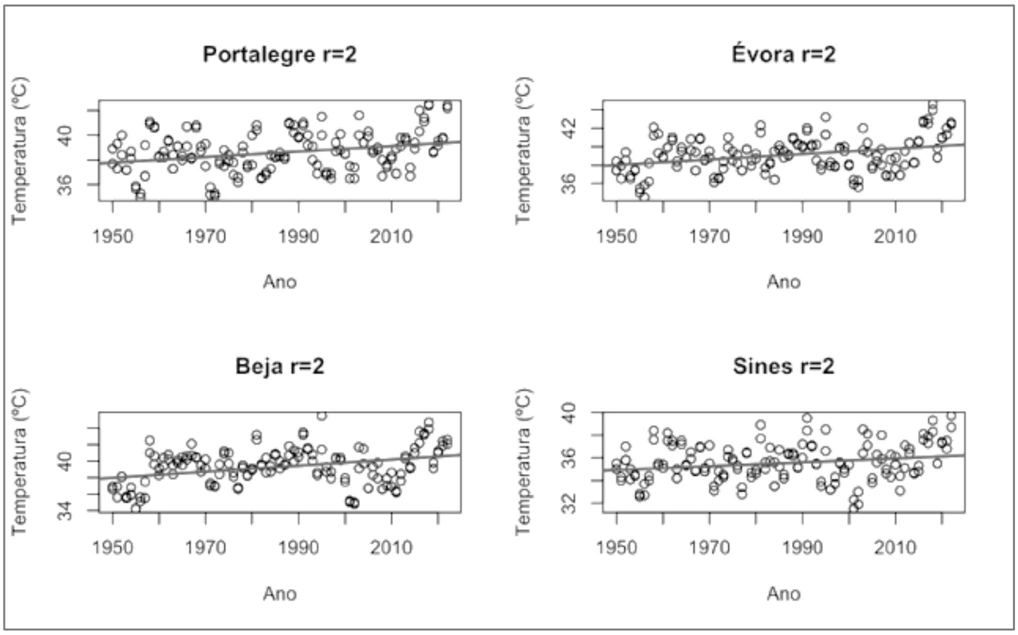


Figura 33 - Ajustamento linear à nuvem das 2 maiores observações anuais, dos quatro concelhos do Alentejo.

Como podemos verificar, as retas ajustadas à localização do modelo limite das 2 maiores observações, apresentam um declive positivo pouco acentuado.

Através, da estimação de MV da tendência linear para a função $\lambda(t) = \beta_0 + \beta_1 t$, foi possível obter as estimativas para os parâmetros a localização β_0 *e* β_1, escala δ e forma γ do modelo GEV baseado nas das 2 maiores observações anuais, apresentados na Tabela 19.

Tabela 19 - Estimação MV da tendência linear, dos parâmetros, no modelo ajustado às 2 maiores observações anuais, com $\lambda(t) = \beta_0 + \beta_1 t$.

$k = 2$	log L	$\widehat{\beta_0}$ $(s.e.(\widehat{\beta_0}))$	$\widehat{\beta}_1$ $(s.e.(\widehat{\beta_1}))$	$\widehat{\delta}$ $(s.e.(\widehat{\delta}))$	$\widehat{\gamma}$ $s(.e.(\widehat{\gamma}))$
Portalegre (valor-p)	-191.325	38.141 (0.259)	0.024 (0.006)	1.365 (0.074)	-0.365 (0.073)
Évora (valor-p)	-213.338	38.307 (0.310)	0.033 (0.008)	1.610 (0.087)	-0.318 (0.059)
Beja (valor-p)	-233.848	38.565 (0.377)	0.034 (0.009)	1.883 0.099	-0.300 0.040
Sines (valor-p)	-196.21	35.380 (0.2779)	0.017 (0.007)	1.429 (0.076)	-0.338 (0.054)

À semelhança do verificado na modelação da temperatura máxima anual em contexto de não estacionariedade, o declive da tendência linear ajustada do parâmetro de localização apresenta a mesma ordem de valores. Deste modo, espera-se que, em média haja um aumento de 0.024ºC por ano em Portalegre, de 0.033ºC por ano em Évora, de 0.034ºC por ano em Beja e de 0.017ºC por ano em Sines.

4.2.7 TESTES DE HIPÓTESES E SELEÇÃO DO MELHOR MODELO

Comparando agora os modelos estacionário e não-estacionário, resultados descritos na Tabela 20. Como o modelo estacionário $GEV_2(\lambda, \delta, \gamma)$ está encaixado no modelo não estacionário $GEV_2(\lambda(t), \delta, \gamma)$, aplica-se o Teste deviance descrito na secção 3.1.5, às hipóteses:

$$H_0: GEV_2\ Estacionário \quad vs \qquad H_1: GEV_2\ Não - Estacionário,$$

Tabela 20 - Valores observados de estatística de deviance (D), deviance corrigida (D^*) na verificação do modelo Estacionário para os dados baseados no modelo limite das 2 maiores observações dos quatro concelhos.

$\boldsymbol{k = 2}$	$\mathbf{D}$	$\mathbf{D^*(valor - p)}$
Portalegre (valor-p)	779.727	765.055 (<0.001)
Évora (valor-p)	869.776	853.409 (<0.001)
Beja (valor-p)	949.114	931.254 (<0.001)
Sines (valor-p)	873.959	857.514 (<0.001)

Para todos concelhos, para um nível significância de 5%, a hipótese H_0 é rejeitada, existindo assim evidências estatísticas que o modelo que melhor se ajusta aos dados é o modelo não estacionário, pois todos os valores de D são bastante elevados, sugerindo uma melhoria substancial relativamente ao modelo estacionário. Permitindo, assim, a dependência com a variável tempo e evidenciando que a inclusão desta variável melhora substancialmente a qualidade de ajuste dos dados.

Os modelos estacionário e não-estacionário, também foram comparados com base nos critérios de informação de ajustamento AIC e BIC, valores apresentados na Tabela 21.

Tabela 21 - Critérios de Informação de Akaike e Bayesiano dos modelos Estacionário e Não-Estacionário.

$\boldsymbol{k = 2}$		$\mathbf{AIC}$	$\mathbf{BIC}$
Portalegre	Estacionário	403.08	412.03
	Não-Estacionário	390.64	402.58
Évora	Estacionário	449.10	458.05
	Não-Estacionário	434.68	446.61
Beja	Estacionário	487.42	496.37
	Não-Estacionário	475.70	487.63
Sines	Estacionário	404.56	413.50
	Não-Estacionário	400.54	412.48

Tal como verificado no modelo GEV, os valores de AIC e BIC do modelo GEV_2 não estacionário são inferiores aos do modelo GEV_2 estacionário, indicando assim que a introdução de uma tendência linear no parâmetro de localização permite obter um melhor ajustamento aos dados.

Na Figura 34, é apresentada a série temporal das previsões dos níveis de retorno a 10 anos para os quatro concelhos do Alentejo, para os anos 2023 a 2122.

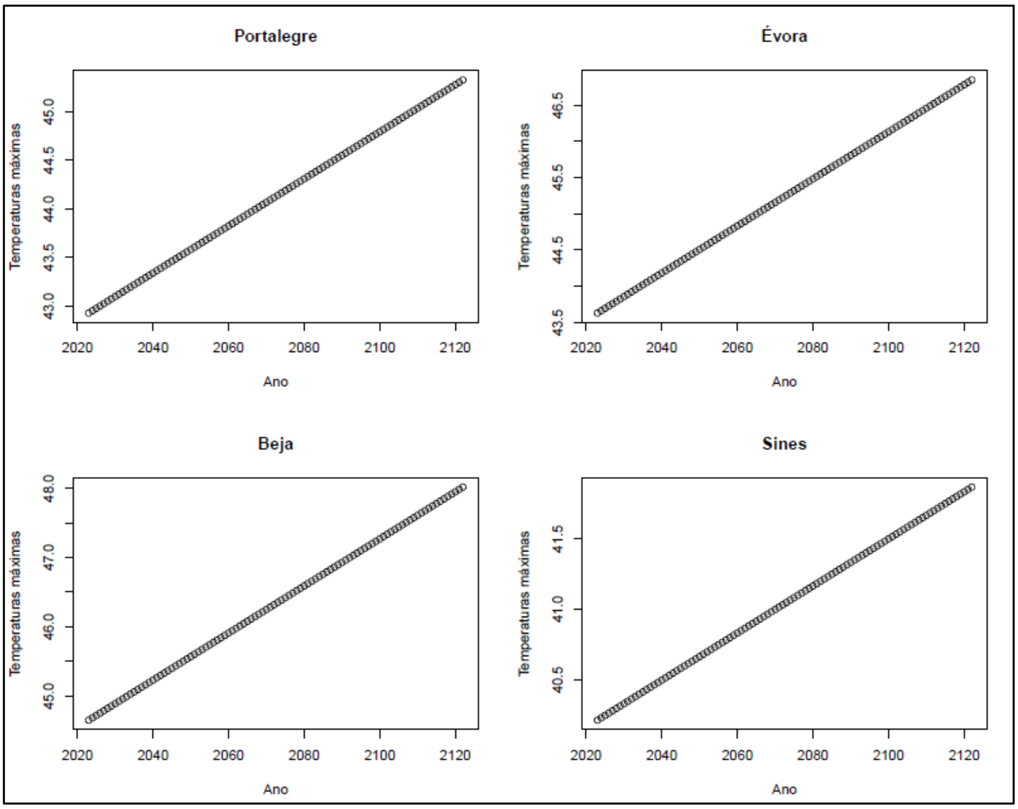


Figura 34 - Séries temporais dos níveis de retorno a 10 anos para os quatro concelhos do Alentejo, para os anos 2023 a 2122, através da GEV para k=2 não-estacionária.

De acordo com o modelo GEV_2 não-estacionário, espera-se que, em média, Portalegre atinja em 2040 uma temperatura máxima de aproximadamente 43.5°C e em 2120 de aproximadamente 45°C. Para Évora, espera-se em 2040 uma temperatura máxima de aproximadamente 44°C e em 2120 de aproximadamente 46.5°C.

Em Beja, espera-se em 2040 uma temperatura máxima de aproximadamente 45.5°C e em 2120 de aproximadamente 48°C. Finalmente para Sines em 2040 a temperatura máxima prevista é de aproximadamente 40.5°C e em 2120 de aproximadamente 42°C, ligeiramente superiores às previsões obtidas no modelo GEV não estacionário [secção 4.1.6].

4.3 MODELO PARETIANO - EXCESSOS DE NÍVEL

Nesta secção vamos analisar as temperaturas diárias que excedem um certo *threshold,* u. Pelo teorema de *Pickands-Balkema-de Haan*, apresentado na secção 3.3, se as temperaturas máximas diárias seguem uma distribuição GEV(δ, λ, γ), então as excedências do nível *u,* seguem uma distribuição GP (δ, γ), e onde o parâmetro de escala δ é função do $threshold, u$.
Inicialmente, procede-se à escolha deste nível, passando pela análise exploratória dos dados. Em seguida, realiza-se uma análise inferencial, seguida da seleção do melhor modelo e, por fim, realiza-se uma modelação não estacionária, admitindo que o parâmetro de escala é uma função linear do tempo.

4.3.1 ESCOLHA DO *THRESHOLD*

A escolha do nível u é o passo mais importante na modelação POT, pois determina que pontos dos dados são considerados extremos e que serão usados para ajustar a Distribuição Generalizada de Pareto (GPD). Para tal escolha, foram realizados os seguintes passos (Beirlant, 2004; Coles, 2001; Davison et al., 1990; Embrechts et al., 2003; McNeil et al., 2005):

1) Primeiramente foi testado um conjunto de *thresholds* no intervalo de quantis extremais [90%; 98%] através da biblioteca `ercv` e da função `thrselect,` que foram também validados pelo gráfico de excesso médio (`mrl.plot)` da biblioteca `ismev` onde é possível observar a média dos excessos acima de u e as bandas de confiança. Nesta avaliação procuram-se valores de u na parte mais linear do gráfico (valores de u mais baixos correspondem a um tamanho da amostra maior), evitando as partes mais instáveis, onde normalmente estão situados os u mais elevados e onde o tamanho da amostra é menor, aumentando assim, a variância devido ao limite de dados disponíveis.

2) Numa segunda fase, foi feito uma nova validação do conjunto de u possíveis com a ajuda da função `tcplot` da biblioteca `POT`. Esta função permite avaliar como os parâmetros de forma e de escala da GPD variam em função do *threshold*. A escolha do *threshold* deve

basear-se na identificação de regiões de estabilidade das estimativas dos parâmetros. À medida que u aumenta, deve-se procurar uma região estável, evitando áreas onde o alargamento do intervalo de confiança indique instabilidade. Além disso, é essencial evitar valores de u demasiado baixos, de modo a minimizar erros sistemáticos (bias) resultantes da inclusão de valores não extremos, que poderiam distorcer a estimação dos parâmetros da GPD.

3) Na terceira fase, foi novamente verificada a estabilidade das estimativas dos parâmetros em função do nível u escolhido através do comando `gpd.fitrange` da biblioteca `ismev`.

Segundo diversos estudos, os valores de quantil mais utilizados e recomendados em dados ambientais situam-se no intervalo [90%; 96%] (Beirlant, 2004; Coles, 2001). A aplicação das três etapas descritas acima levou à consideração dos seguintes *thresholds:* $u_{Portalegre}$=37.4ºC (quantil 96%); u_{Beja}=38.5 ºC (quantil 96%); $u_{Évora}$=37 ºC (quantil 93%) e u_{Sines}=33.5 ºC (quantil 93%).

De seguida, são apresentadas as várias análises para os quatro concelhos (Figura 35 à Figura 38). As linhas verticais das figuras correspondem aos *thresholds* escolhidos para cada concelho.

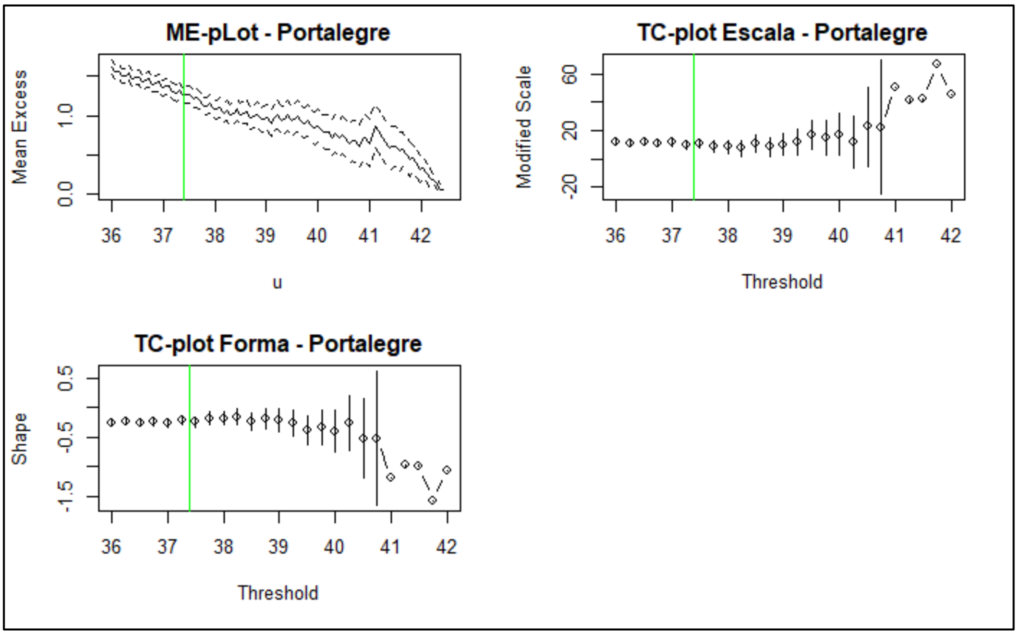


Figura 35 - ME-plot e bandas de confiança a 95% e gráficos de estabilidade do *threshold* para os parâmetros de escala e forma para a Distribuição Generalizada de Pareto das temperaturas máximas anuais de Portalegre para um *threshold* de 37.4°C.

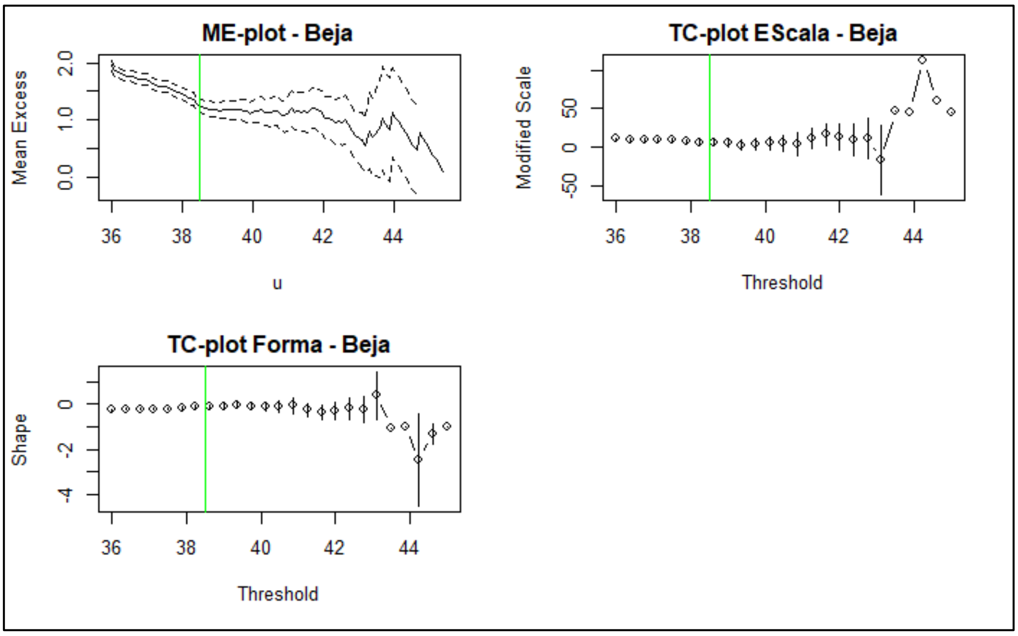


Figura 36 - ME-plot e bandas de confiança a 95% e gráficos de estabilidade para os parâmetros de escala e forma para a Distribuição Generalizada de Pareto das temperaturas máximas anuais de Beja, para um *threshold* de 38.5°C.

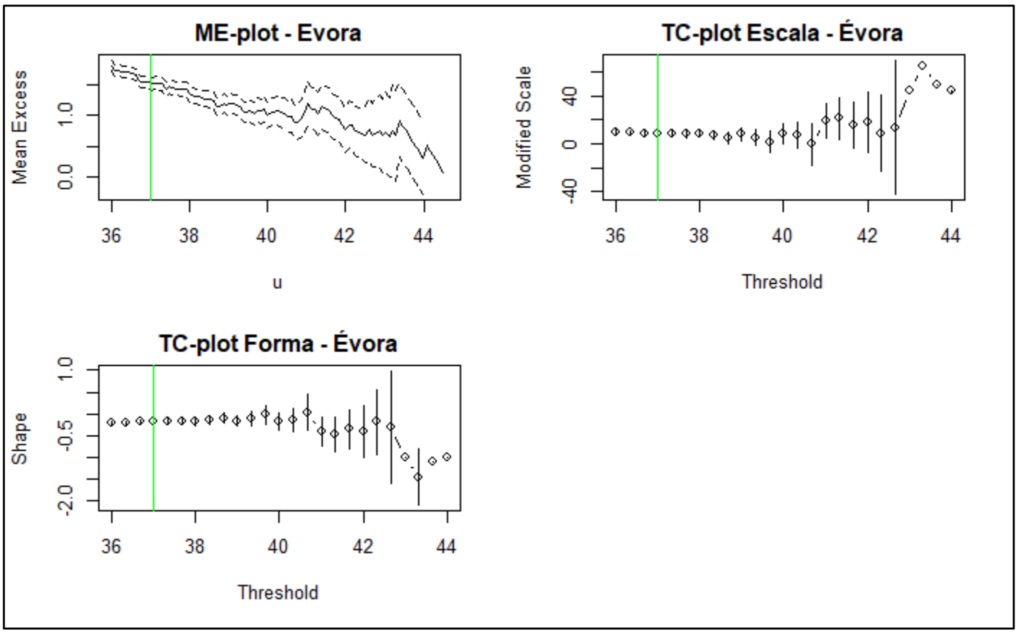


Figura 37 - ME-plot e bandas de confiança a 95% e gráficos de estabilidade para os parâmetros de escala e forma para a Distribuição Generalizada de Pareto das temperaturas máximas anuais de Évora, para um *threshold* de 37.0°C.

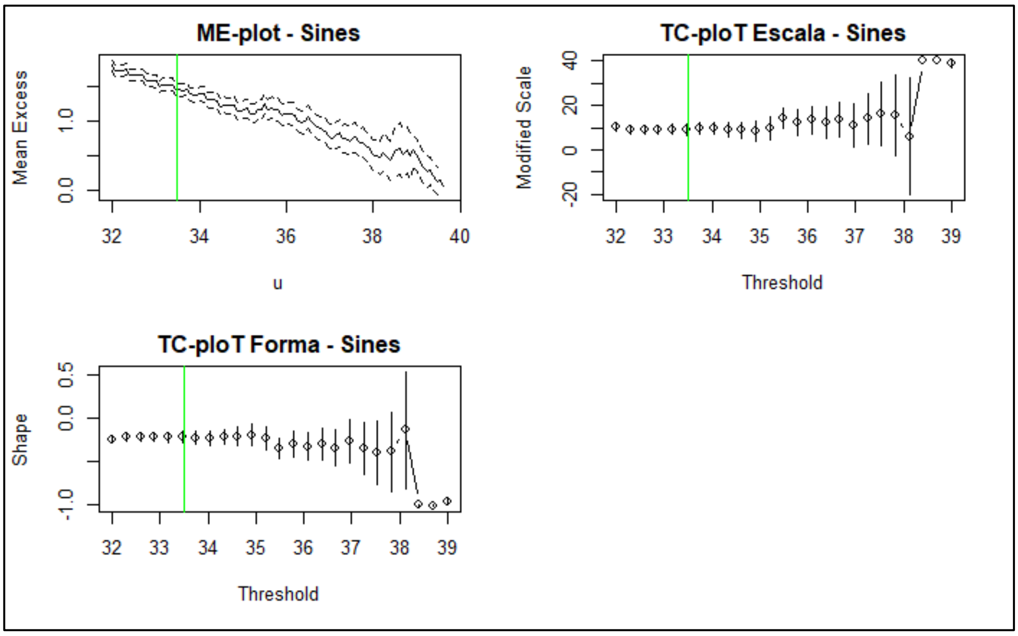


Figura 38 - ME-plot e bandas de confiança a 95% e gráficos de estabilidade para os parâmetros de escala e forma transformados para a Distribuição Generalizada de Pareto das temperaturas máximas anuais de Sines, para um *threshold* de 33.5°C.

4.3.2 ESTATÍSTICA DESCRITIVA PRELIMINAR

Considerando, os níveis 37.4°C para Portalegre (353 excedências), 38.5°C (349 excedências), para Beja, 37.0°C (623 excedências) para Évora e 33.5°C (610 excedências) para Sines, Figura 39 são apresentadas as séries das temperaturas diárias máximas dos meses de Junho, Julho, Agosto e Setembro entre 1950 e 2022 que excedem os respetivos *thresholds*.

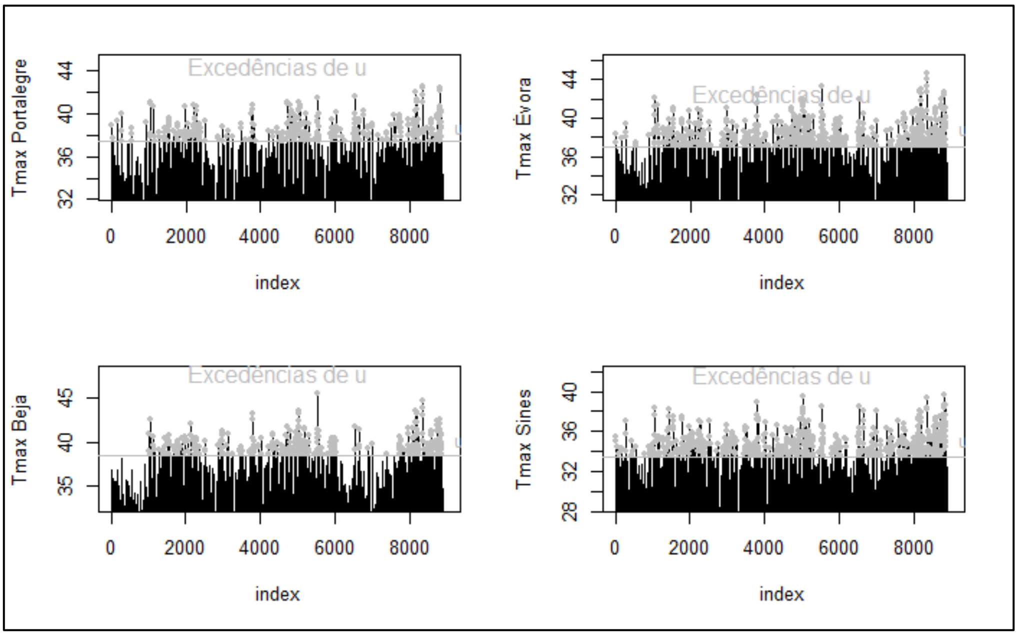


Figura 39 - Série temporal temperaturas máximas diárias com as excedências de u, representadas a cinzento, para os concelhos de Portalegre, Évora, Beja e Sines de 1950 a 2022.

Numa análise preliminar de dados, foram obtidos os histogramas e os *boxplots* associados a este conjunto de dados, apresentados na Figura 40 e Figura 41,

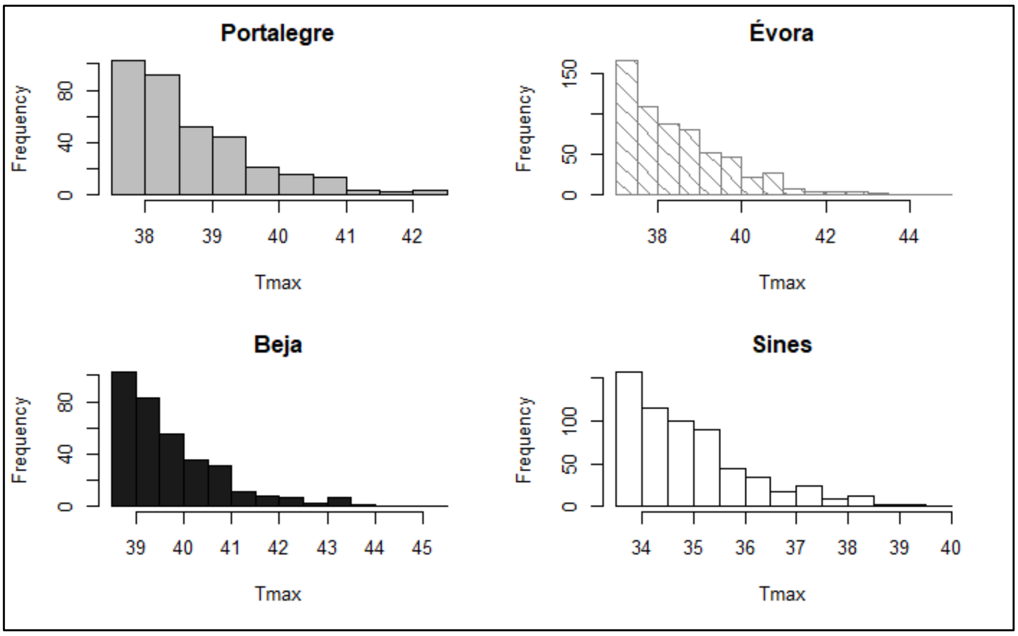


Figura 40 - Histograma das temperaturas máximas diárias que excedem *u*, dos quatro concelhos: Portalegre, Évora, Beja e Sines, de 1950 a 2022.

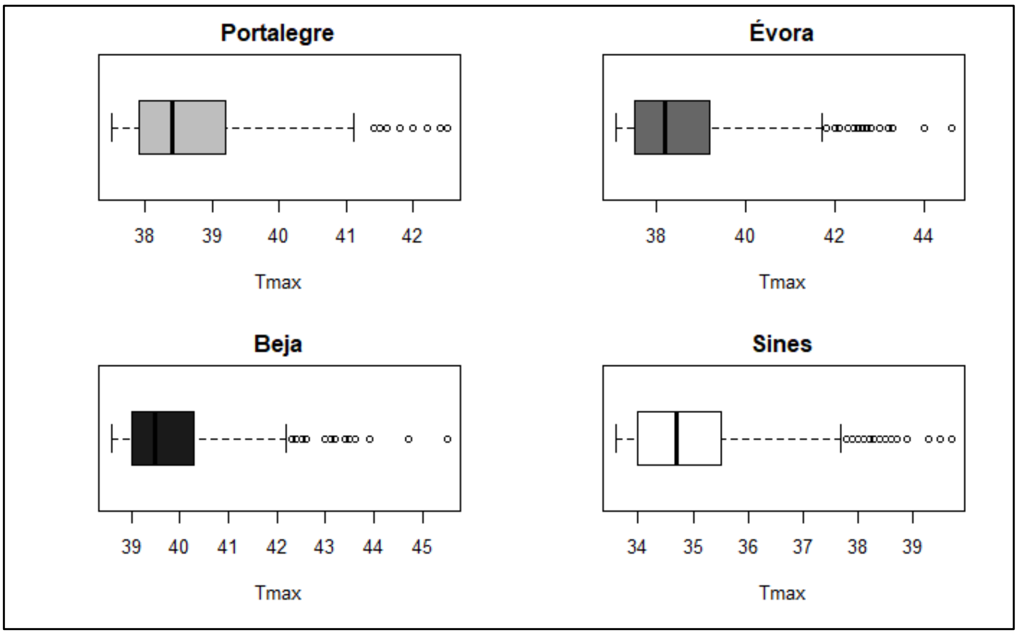


Figura 41 - Boxplot das temperaturas máximas diárias que excedem *u*, dos quatro concelhos: Portalegre, Évora, Beja e Sines, de 1950 a 2022.

Na Tabela 22, são apresentadas as estatísticas descritivas das excedências para cada concelho, que permitem uma análise mais detalhada dos gráficos acima representados.

Tabela 22 - Estatísticas descritivas das temperaturas máximas diárias que excedem u, dos quatro concelhos do Alentejo.

	Min.	1^{ro}Qu.	Mediana	Média	3^{ro}Qu.	Max.	Curtose	Assimetria
Portalegre	37.50	37.90	38.40	38.71	39.20	42.50	1.40	1.25
Évora	37.10	37.50	38.20	38.55	39.20	44.60	1.78	1.27
Beja	38.60	39.00	39.50	39.81	40.30	45.50	3.16	1.63
Sines	33.60	34.00	34.70	35.00	35.50	39.70	1.05	1.17

Como podemos verificar, através da análise dos histogramas e dos *boxplots*, observa-se uma assimetria positiva em todos os concelhos, comprovada também pelo coeficiente de assimetria da Tabela 22. Dada a natureza dos dados e o facto de se verificar uma assimetria à direita, é valido a aplicação de análise estatística de extremos.

Relativamente ao comportamento de cauda das distribuições das temperaturas máximas diárias que excedem u, o coeficiente de curtose é positivo pelo que a cauda direita da das distribuições poderá ser mais pesada do que a da distribuição normal. No entanto, o comportamento decrescente, evidenciado nos ME-plots, indicia que a cauda das distribuições subjacentes às temperaturas máximas diárias que excedem u, dos quatro concelhos, Portalegre, Évora, Beja e Sines, não serão mais pesadas que a exponencial, sendo assim, de cauda leve.

Nesta análise preliminar foi ainda verificado o ajuste ao modelo exponencial para a amostra de excessos (ver Figura 42). Esta análise é de extrema importância uma vez que, caso $\gamma = 0$, a distribuição GP, reduz-se à distribuição exponencial, tal como mencionado na secção 2.3.

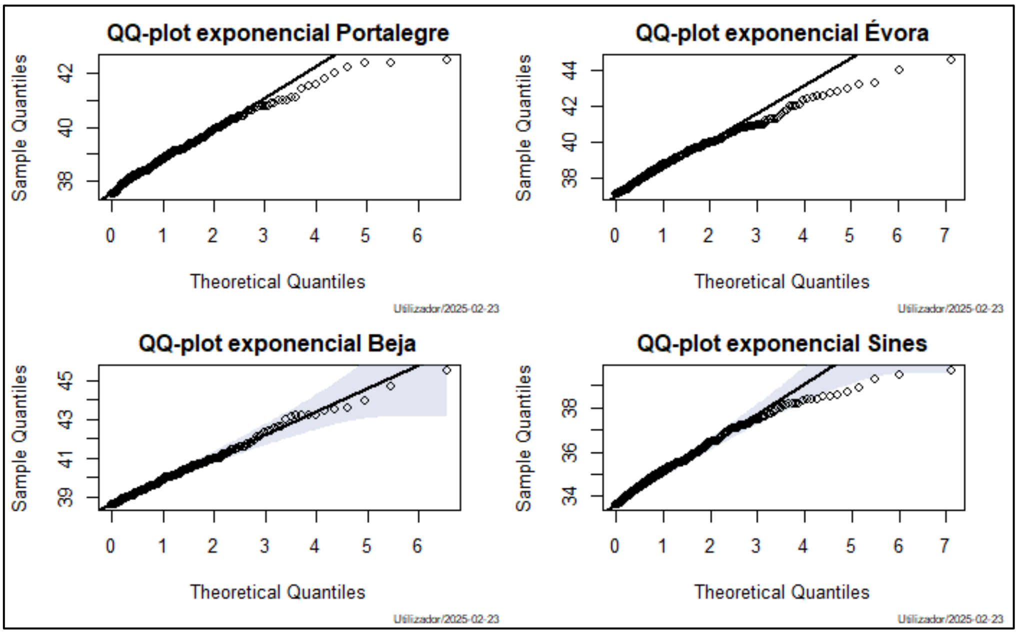


Figura 42 - QQ-plot exponencial das temperaturas máximas diárias que excedem u, dos quatro concelhos: Portalegre, Évora, Beja e Sines, de 1950 a 2022.

Como podemos observar, os dados têm fraco ajustamento à distribuição Exponencial, pois a nuvem de pontos está afastada da reta e fora das bandas de confiança, o que nos leva a admitir que $\gamma \neq 0$.

Para se contruir o QQ-plot da distribuição GP primeiramente é necessário determinar a estimativa preliminar do parâmetro de forma (γ), representada na Figura 43, e obtida de forma a maximizar o coeficiente de correlação entre os quantis teóricos e os quantis empíricos.

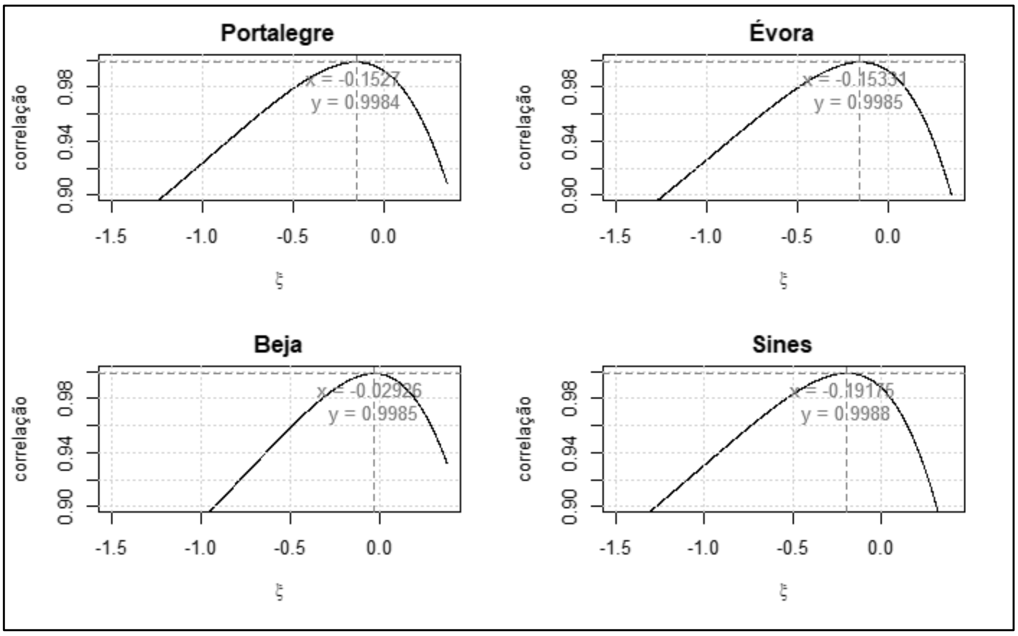


Figura 43 - Estimativas preliminares do parâmetro de forma (representado como ξ), das temperaturas máximas anuais que excedem u, dos quatro concelhos: Portalegre, Évora, Beja e Sines, de 1950 a 2022.

Considerando $\gamma \in [-0.5, 0.5]$, as estimativas obtidas para cada concelho foram, $\hat{\gamma}_{Portalegre} = -0.153$, $\hat{\gamma}_{Évora} = -0.153$, $\hat{\gamma}_{Beja} = -0.029$ e $\hat{\gamma}_{Sines} = -0.192$, a que correspondem coeficientes de correlação de $r_{Portalegre} = 0.998$, $r_{Évora} = 0.998$, $r_{Beja} = 0.999$ e $r_{Sines} = 0.999$.

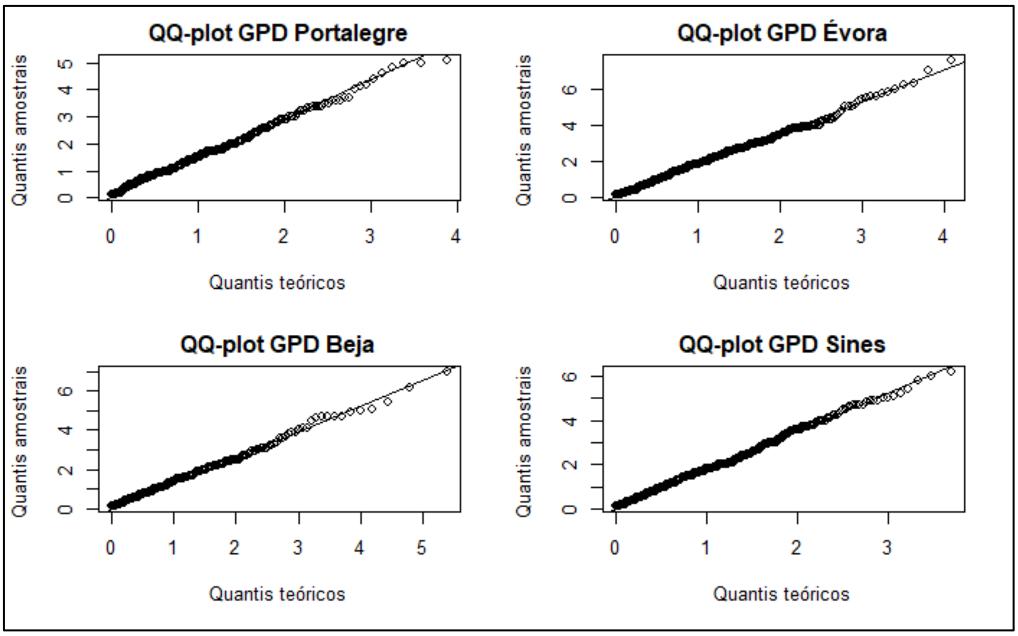


Figura 44 - QQ-Plot da distribuição GP das excedências, referentes às temperaturas máximas diárias, dos quatro concelhos: Portalegre, Évora, Beja e Sines, de 1950 a 2022.

Através do ajuste da reta de mínimos quadrados, obtivemos os seguintes parâmetros de escala para os quatro concelhos: $\hat{\delta}_{Portalegre} = 1.453$, $\hat{\delta}_{Évora} = 1.759$, $\hat{\delta}_{Beja} = 1.313$ e $\hat{\delta}_{Sines} = 1.754$.

Como podemos verificar, a estimativa obtida para o parâmetro de forma para os quatro concelhos, é negativa, logo o modelo GP reduz-se a um modelo Beta. Contudo, o modelo exponencial também pode ser forte candidato ao ajuste dos dados.

4.3.3 INFERÊNCIA ESTATÍSTICA

De seguida, foi feito o ajustamento à distribuição GP, como mencionado acima, através da aplicação da biblioteca `ismev`, que usa o método de MV na estimação dos parâmetros, onde se obteve os resultados apresentados na Tabela 23.

Tabela 23 - Ajustamento dos dados à distribuição GP: estimação dos parâmetros de escala $(\hat{\delta})$, forma $(\hat{\gamma})$ e erros-padrão correspondentes e a taxa (*rate*) de dados que excedem o *threshold*, para 122 observações por ano (*npy*).

	u	**nexc**	**log L**	$\hat{\delta}$	$\hat{\gamma}$	$s.e.(\hat{\delta})$	$s.e.(\hat{\gamma})$	**rate (*npy*=122)**
Portalegre	37.4	353	-437.461	1.642	-0.256	0.108	0.041	0.039
Évora	37	623	-884.259	1.837	-0.189	0.091	0.030	0.070
Beja	38.5	349	-440.416	1.460	-0.116	0.102	0.046	0.039
Sines	33.5	610	-840.445	1.875	-0.251	0.095	0.032	0.068

As estimativas dos parâmetros de forma são todas negativas indicando que as caudas das distribuições subjacentes aos dados são leves e com limite superior de suporte finito.

A qualidade do ajustamento das temperaturas máximas diárias que excedem os respetivos *thresholds*, dos quatro concelhos, pode ser investigada através dos gráficos de diagnóstico, obtidos através da biblioteca `evir` (Figura 45 à Figura 48). Para os quatro concelhos observa-se um bom ajuste do modelo GP.

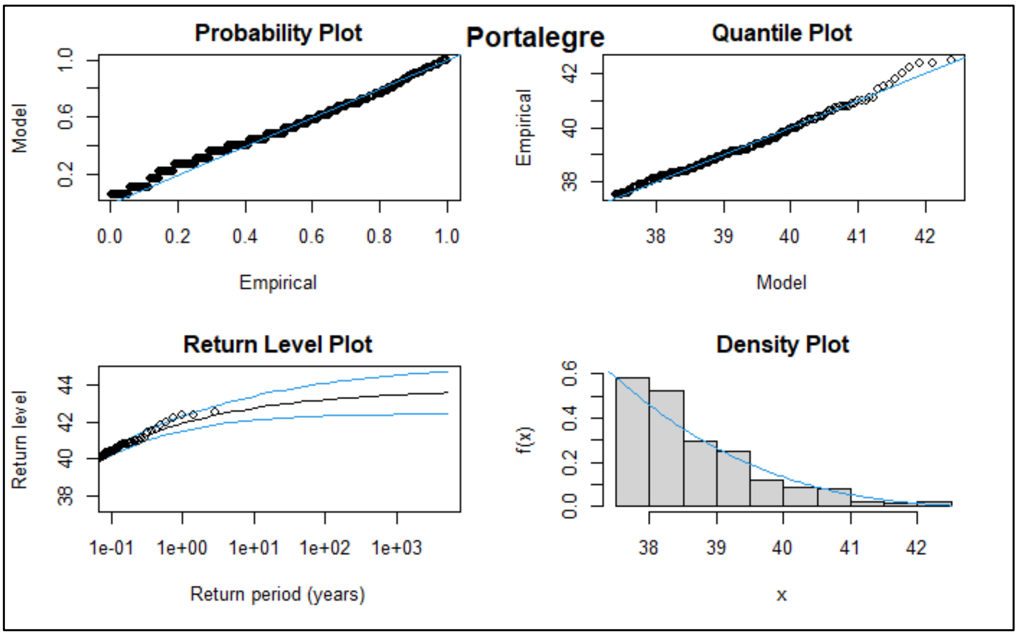


Figura 45 - Ajustamento dos excessos ao modelo GP, referentes às temperaturas máximas diárias em Portalegre de1950 a 2022.

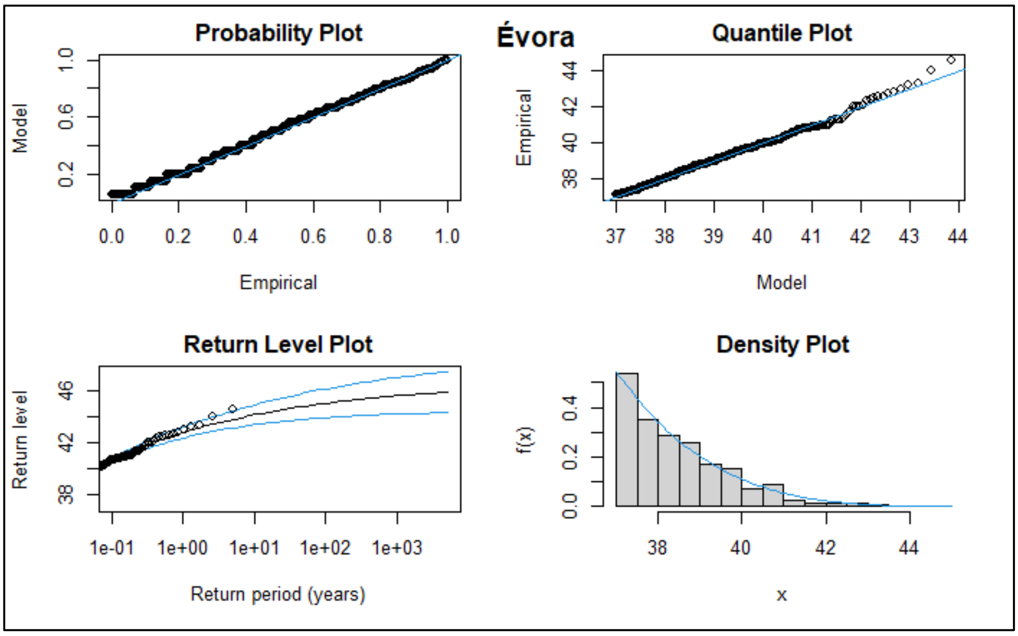


Figura 46 - Ajustamento dos excessos ao modelo GP, referentes às temperaturas máximas diárias em Évora de1950 a 2022.

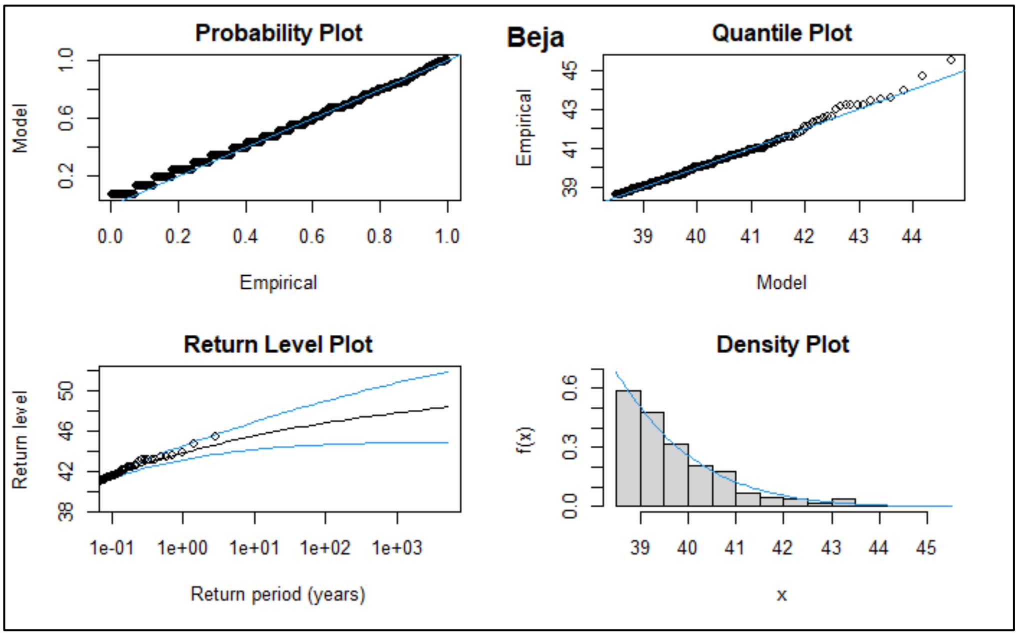


Figura 47 - Ajustamento dos excessos ao modelo GP, referentes às temperaturas máximas diárias em Beja, de1950 a 2022.

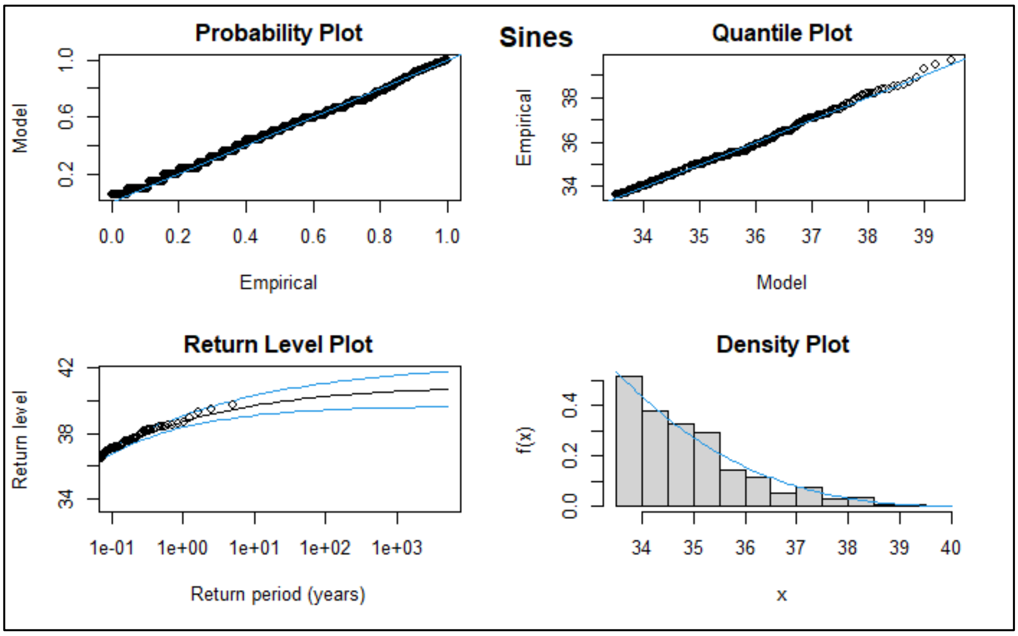


Figura 48 - Ajustamento dos excessos ao modelo GP, referentes às temperaturas máximas diárias em Sines, de1950 a 2022.

4.3.4 INTERVALOS DE CONFIANÇA PARA OS PARÂMETROS

De seguida são apresentados os ICs para os parâmetros de forma na Figura 49 e de escala na Figura 50, baseados na função de log-verosimilhança de perfil, com grau de confiança de 95% aproximadamente, os quais foram obtidos através da biblioteca `evd`.

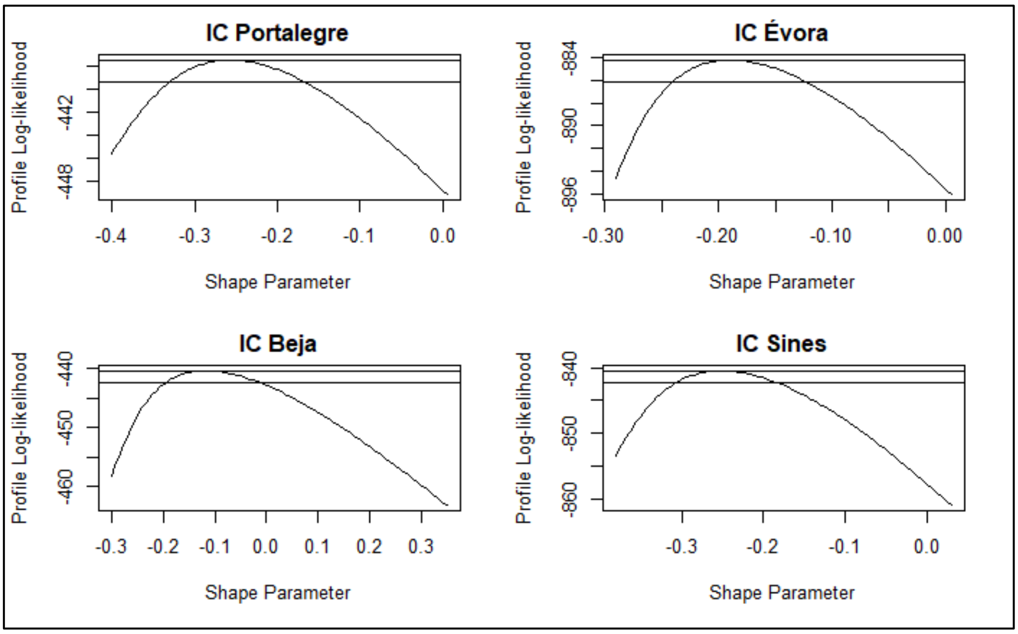


Figura 49 -IC baseados na *profile-likelihood* do parâmetro de forma (γ), para os quatro concelhos, baseados no modelo GP.

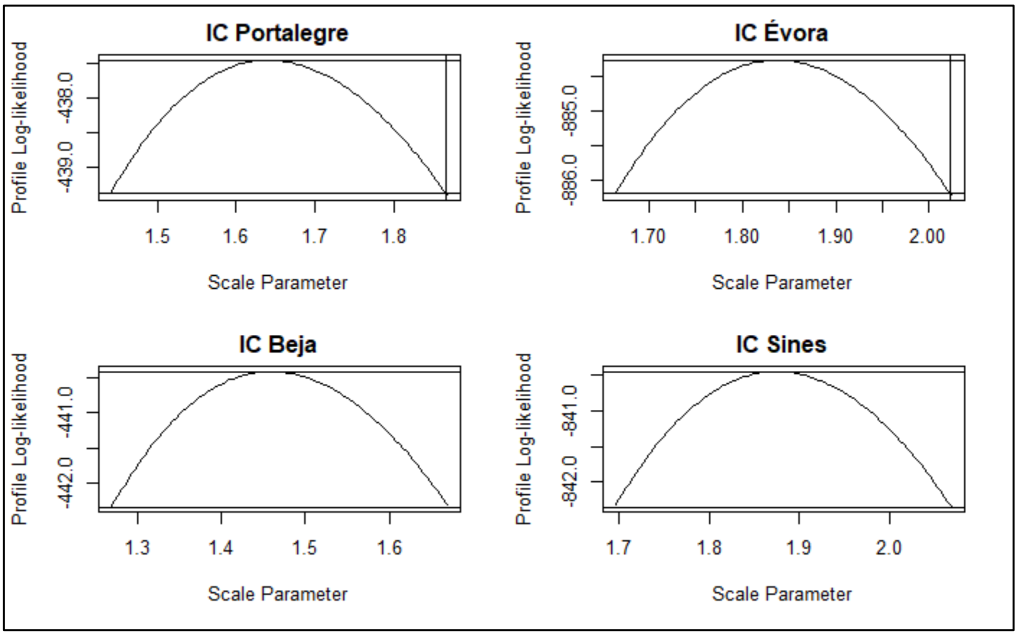


Figura 50 - IC baseados na *profile-likelihood* do parâmetro de escala (δ), para os quatro concelhos, baseados no modelo GP.

Na Tabela 24 estão representados os ICs com grau de confiança de 95% , para os parâmetros de escala (δ) e forma (γ), para os quatro concelhos, baseados no modelo GP.

Tabela 24 - Intervalos de confiança dos parâmetros de escala e de forma, para um grau de confiança de 95% aproximadamente, do modelo GP.

	$\boldsymbol{IC_{95\%}(\delta)}$	$\boldsymbol{IC_{95\%}(\gamma)}$
Portalegre	(1.442; 1.868)	(-0.400; 0.005)
Évora	(1.665; 2.024)	(-0.290; 0.005)
Beja	(1.270; 1.669)	(-0.300; 0.350)
Sines	(1.697; 2.068)	(-0.380; 0.030)

Nos quatro concelhos observamos que $0 \in IC_{95\%}(\gamma)$, logo existe evidência estatística, ao nível de significância de 5%, que o modelo exponencial pode ser adequado para modelar as temperaturas máximas diárias que excedem um nível elevado.

4.3.5 SELEÇÃO DO MELHOR MODELO

De modo a selecionar o modelo que melhor se ajusta aos dados, na Tabela 25 é apresentada uma comparação dos valores da log-verosimilhança e dos critérios de informação de ajustamento AIC e BIC entre os modelos exponencial e GP.

Tabela 25 - Valores da log-verosimilhança e Critérios de Informação de Akaike e Bayesiano dos modelos exponencial e GP.

		Log L	**AIC**	**BIC**
PORTALEGRE	Exponencial	-448.83	899.66	903.53
	GP	-437.46	880.92	892.52
ÉVORA	Exponencial	-895.54	1793.09	1797.53
	GP	-884.26	1774.52	1787.82
BEJA	Exponencial	-442.86	887.73	891.59
	GP	-440.42	886.83	898.40
SINES	Exponencial	-857.73	1717.47	1721.88
	GP	-840.45	1686.90	1700.13

Os valores do AIC e do BIC são maiores para o modelo exponencial, e os valores da log-verosimilhança associados ao modelo GP são inferiores aos obtidos com o ajustamento ao modelo Exponencial, indicando assim que apesar de o intervalo de confiança a 95% para o parâmetro de forma conter o 0, o modelo GP fornece um melhor ajustamento aos dados.

Interessa então saber se a distribuição exponencial é ou não um bom modelo para ajustar aos excessos acima de um nível elevado. Foi então calculado o Teste Deviance com a correção de Bartlett (Reiss et al., 2007):

$$\mathcal{D}^* = \frac{\mathcal{D}}{1+\frac{4}{m}} \xrightarrow[m\to\infty]{d} Q_1 \sim \chi_1^2 \qquad (4.3.5.1.1)$$

As hipóteses testadas foram: $H_0: \gamma = 0$ $(Exponencial)$ $\quad vs \;\; H_1: \gamma \neq 0$ (GP), estando os resultados descritos na Tabela 26.

Tabela 26 -Valores observados da Estatística de Deviance (D), Deviance corrigida (D^*) para verificação do modelo GP para dos quatro concelhos.

	$\mathbf{D}$	$\mathbf{D}^*$ (valor-p)
Portalegre (valor-p)	22.739	22.485 (<0.001)
Évora (valor-p)	22.579	22.435 (<0.001)
Beja (valor-p)	4.899	4.844 (<0.03)
Sines (valor-p)	34.577	34.351 (<0.001)

Para todos concelhos, para um nível significância de 5%, a hipótese H_0 é rejeitada, existindo assim evidências estatísticas que o modelo que melhor se ajusta aos dados é o modelo GP. Dado isto, prosseguimos o estudo considerando o modelo GP.

4.3.6 ESTIMAÇÃO DE QUANTIDADES DE INTERESSE

Usando as estimativas obtidas no modelo GP ajustado, para os parâmetros de forma e escala, representadas na Tabela 23, podem ser obtidas quantidades importantes em modelação de acontecimentos raros, tais como, a probabilidade de excedência, os quantis elevados, o limite superior do suporte, os níveis de retorno e o período de retorno [secção 3.3.3].

Na Tabela 27 estão representadas as estimativas MV para os quantis extremais ($\hat{q}$) de probabilidades 0.90, 0.95, 0.99, para os níveis de retorno ($\widehat{U}$) de 10, 50 e 100 anos, para o período de retorno ($\hat{T}$) da temperatura máxima diária observada em casa concelho, para a probabilidade de excedência desse mesmo máximo ($\hat{P}(Y > Tmax)$) e para o limite superior do suporte ($\hat{x}^F$), baseadas no modelo GP.

Tabela 27 - Estimativas MV de outros parâmetros para referentes às temperaturas máximas diárias: quantis extremais ($\hat{q}$) de probabilidades 0.90, 0.95, 0.99, níveis de retorno ($\widehat{U}$) de 10, 50 e 100 anos, para o período de retorno ($\widehat{T}$) da temperatura máxima observada em casa concelho, para a probabilidade de excedência desse mesmo máximo ($\widehat{P}(Y > Tmax)$) e para o limite superior do suporte ($\hat{x}^F$), baseado no modelo GP para os concelhos do Alentejo, no período de 1950 a 2022 (Tmax=42.50°C*;44.60°C**;45.50°C***;39.70°C****).

	Portalegre	**Évora**	**Beja**	**Sines**
$\hat{q}_{0.90}$	42.25	42.92	44.46	38.83
$\hat{q}_{0.95}$	42.51	43.38	44.97	39.17
$\hat{q}_{0.99}$	42.94	44.26	46.01	39.77
$\widehat{U}(10)$	42.25	42.92	44.46	38.83
$\widehat{U}(50)$	42.78	43.91	45.59	39.54
$\widehat{U}(100)$	42.94	44.26	46.01	39.77
$\widehat{T}(Tmax)$	102.91	369.57	232.74	138.62
$\widehat{P}(Y > Tmax)$	*7.96e-05	**2.22e-05	***3.52e-05	****5.91e-05
$\hat{x}^F$	43.80	46.72	51.05	40.97

Apenas 10% das temperaturas excedem 42.25ºC em Portalegre, 42.92ºC em Évora, 44.46ºC em Beja e 38.83ºC em Sines, no entanto, 1% das temperaturas excedem 42.94ºC em Portalegre, 44.25ºC em Évora, 46.01ºC em Beja e 39.77ºC em Sines.

Relativamente ao período de retorno da temperatura máxima, para Portalegre é, em média, de 103 anos, de 370 anos para Évora, de 233 anos para Beja e de 139 anos para Sines, sendo a probabilidade de excedência dessas temperaturas de 7.96e-05, 2.22e-05, 3.52e-05, 5.91e-05, respetivamente, ou seja, probabilidades de ocorrência muito baixas.

Analisando o limite superior de suporte para cada concelho, significa que é extremamente improvável que as temperaturas venham a ser superiores a 43.80ºC, 46.72ºC, 51.05ºC, e 40.97ºC, respetivamente.

4.3.7 MODELAÇÃO DE FENÓMENOS NÃO-ESTACIONÁRIOS

Como temos vindo a referir os fenómenos ambientais raramente são estacionários, e uma análise da série temporal apresentada na Figura 51, para o modelo GP, deixa transparecer visualmente a presença de uma sazonalidade e tendência. Sendo assim, os dados foram tratados através da modelação linear do parâmetro de escala. Uma vez que, o parâmetro de escala tem de ser positivo, foram escolhidos dois modelos com tendência ao longo do tempo:

$$Modelo\ 1: \delta(t) = \beta_0 + \beta_1 t \tag{4.3.7.1}$$

$$Modelo\ 2: \delta(t) = exp\,\{\beta_0 + \beta_1 t\} \tag{4.3.7.2}$$

Onde, t é o indicador tempo.

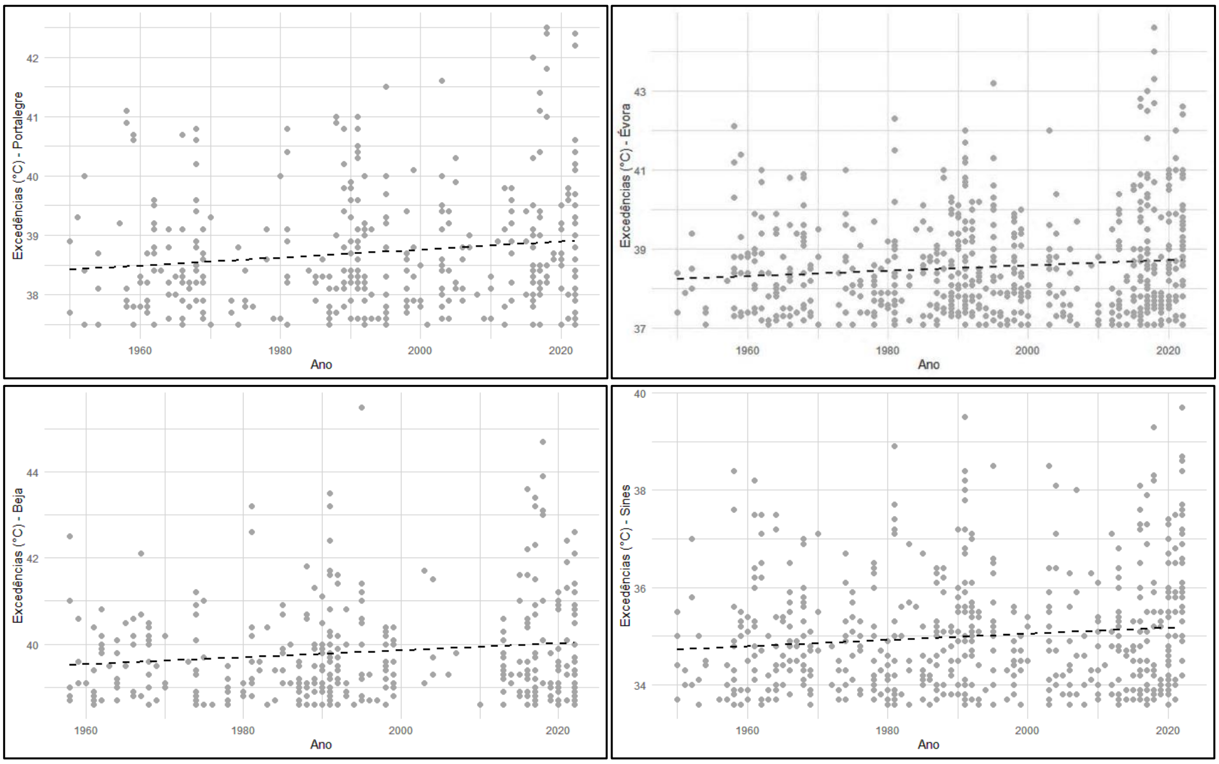


Figura 51 - Ajustamento linear à nuvem das excedências para os quatro concelhos do Alentejo.

Como podemos verificar, as retas ajustadas às excedências, i.e., às temperaturas máximas diárias que excedem um determinado nível elevado, apresentam um declive positivo pouco acentuado, evidenciando a não estacionariedade dos dados.

Na Tabela 28, são apresentadas as estimativas dos parâmetros de escala (β_0 e β_1) e de forma (γ) obtidas para o modelo GP com tendência linear e para o modelo GP com tendência exponencial.

Tabela 28 - Estimação dos parâmetros de escala (β_0 e β_1), forma (γ) para o modelo GP tendência linear e para o modelo GP com tendência exponencial.

Modelo 1	**log L**	$\widehat{\boldsymbol{\beta_0}}$ $(\boldsymbol{s.e.}(\widehat{\boldsymbol{\beta_0}}))$	$\widehat{\boldsymbol{\beta}}_1$ $(\boldsymbol{s.e.}(\widehat{\boldsymbol{\beta_1}}))$	$\widehat{\boldsymbol{\gamma}}$ $(\boldsymbol{s.e.}(\widehat{\boldsymbol{\gamma}}))$
Portalegre	-432.9252	1.4350 (1.1406e-01)	0.0016 (2.0003e-06)	-0.3124 (4.5631e-02)
Évora	-878.3836	1.5669 (9.1485e-02)	0.0011 (2.0011e-06)	-0.2257 (3.0841e-02)
Beja	-436.4853	1.1519 (0.1233)	0.0019 (0.0006)	-0.1408 (0.0447)
Sines	-835.5454	1.6186 (9.7817e-02)	0.0009 (2.0817e-06)	-0.2640 (3.3374e-02)
Modelo 2				
Portalegre	- 432.6649	0.3574 (7.3749e-02)	0.0010 (2.1460e-06)	-0.3141 (5.1595e-02)
Évora	- 878.0874	0.4527 (5.2112e-02)	0.0006 (2.0879e-06)	-0.2276 (3.3343e-02)
Beja	- 436.5147	0.1591 (0.0894)	0.0013 (0.0004)	-0.1395 (0.0470)
Sines	- 835.3487	0.4831 (5.0326e-02)	0.0005 (2.0059e-06)	-0.2636 (3.1776e-02)

Como podemos analisar, os valores da log-verosimilhança são semelhantes nos dois modelos, para os quatro concelhos, assim como, todos os outros parâmetros. Sendo assim, pelo princípio da parcimónia é escolhido o modelo mais simples - Modelo 1.

4.3.8 TESTES DE HIPÓTESES E SELEÇÃO DO MELHOR MODELO

Comparando agora o modelo estacionário com o não-estacionário, estando os resultados descritos na Tabela 29, sabendo que o modelo estacionário GP(δ, γ) está encaixado no modelo não estacionário $GP_{Modelo\ 1}$(δ(t), γ), é possível aplicar o Teste Deviance descrito na secção 3.1.4, às hipóteses:

$$H_0: GP\ Estacionário \quad vs \quad H_1: GP\ Não - Estacionário.$$

Tabela 29 - Valores observados de estatística de Deviance (D), Deviance corrigida (D^*)na verificação do modelo Estacionário para os dados baseados no modelo GP dos quatro concelhos.

	$\mathbf{D}$	$\mathbf{D^*(valor - p)}$
Portalegre (valor-p)	1740.77	1727.07 (<0.001)
Évora (valor-p)	3525.29	3509.51 (<0.001)
Beja (valor-p)	1753.80	1739.84 (<0.001)
Sines (valor-p)	3351.98	3336.67 (<0.001)

Para todos concelhos, para um nível significância de 5%, a hipótese H_0 é rejeitada, existindo assim evidências estatísticas que o modelo que melhor se ajusta aos dados é o modelo não estacionário, pois todos os valores de D são bastante elevados, sugerindo uma melhoria substancial relativamente ao modelo estacionário. Deste modo, a inclusão da dependência temporal evidencia uma melhora substancial da qualidade de ajuste dos dados.

Os modelos estacionário e não-estacionário, também foram comparados com base nos critérios de informação de ajustamento AIC e BIC, valores apresentados na Tabela 30.

Tabela 30 - Critérios de Informação de Akaike e Bayesiano dos modelos Estacionário e Não-Estacionário.

Modelo 1		AIC	BIC
Portalegre	Estacionário	880.92	892.52
	Não-Estacionário	871.85	883.45
Évora	Estacionário	1774.52	1787.82
	Não-Estacionário	1762.77	1776.07
Beja	Estacionário	886.83	898.39
	Não-Estacionário	878.97	890.53
Sines	Estacionário	1686.89	1700.13
	Não-Estacionário	1677.09	1690.33

Tal como verificado no modelo GP, os valores de AIC e BIC do modelo GP não estacionário são inferiores aos do modelo GP estacionário, indicando assim que a introdução de uma tendência linear no parâmetro de escala permite obter um melhor ajustamento aos dados.

Na Figura 52, é apresentada a série temporal das previsões dos níveis de retorno a 10 anos para os quatro concelhos do Alentejo, para os anos 2023 a 2130.

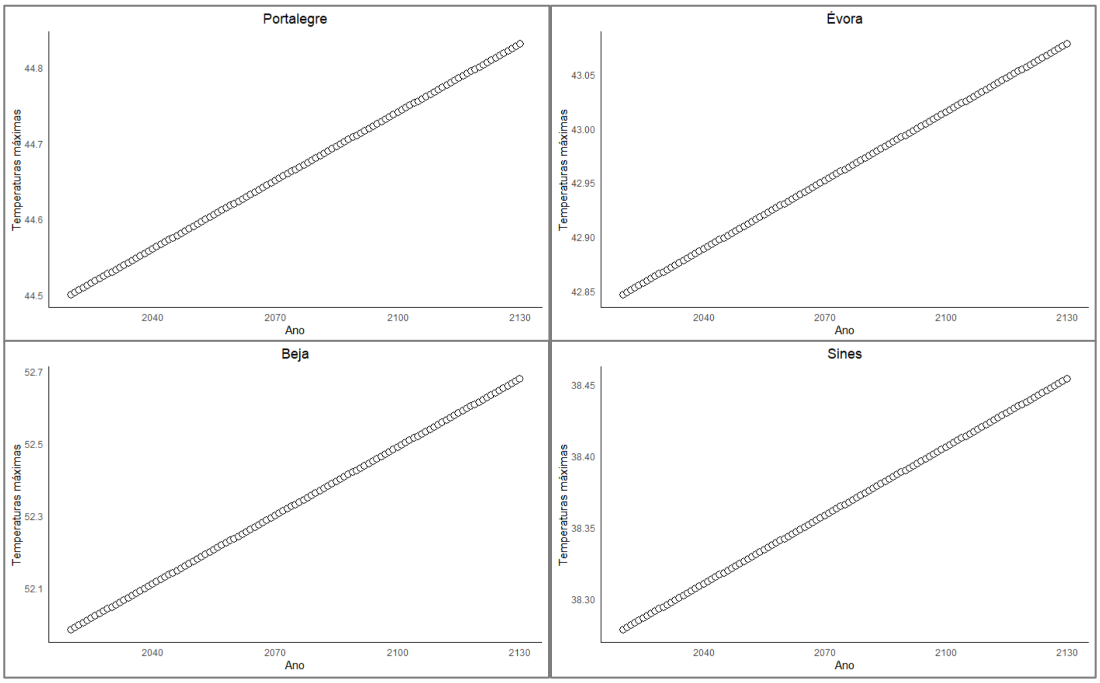


Figura 52 - Séries temporais dos níveis de retorno a 10 anos para os quatro concelhos do Alentejo, para os anos 2023 a 2130, para o modelo GP não-estacionário.

Para o modelo GP não-estacionário obtêm-se as seguintes temperaturas máximas diárias previstas para 2025: $Tmax(2025)_{Portalegre} = 44.51°C$; $Tmax(2025)_{Évora} = 42.87°C$; $Tmax(2025)_{Beja} = 52.02°C$ e $Tmax(2025)_{Sines} = 38.28°C$.

Da observação da Figura 52, espera-se que, em média, Portalegre atinja em 2040 uma temperatura máxima de aproximadamente 44.56°C e em 2120 de aproximadamente 44.80°C, ligeiramente superiores às obtidas no modelo GEV_2 não estacionário [secção 4.2.7]; relativamente a Évora, espera-se em 2040 uma temperatura máxima de aproximadamente 42.89°C e em 2120 de aproximadamente 42.84°C, ligeiramente inferiores às obtidas no modelo GEV_2 não estacionário [secção 4.2.7]; para Beja em 2040 a temperatura máxima prevista é de aproximadamente 52.11°C e em 2120 de aproximadamente 52.61°C, substancialmente superiores às obtidas no modelo GEV_2 não estacionário [secção 4.2.7]. Finalmente para Sines em 2040 a temperatura máxima prevista é de aproximadamente 38.31°C e em 2120 de aproximadamente 38.43°C, ligeiramente inferiores às previsões obtidas no modelo GEV_2 não estacionário [secção 4.2.7].

5
Conclusões

A teoria de valores extremos mostra conclusões interessantes quando aplicada às temperaturas máximas na Região do Alentejo.

No que se refere à análise dos máximos anuais, verificou-se que o modelo GEV com parâmetro de forma negativo se ajusta melhor aos dados em contexto estacionário, quando comparado com o modelo Gumbel. Considerando as quantidades de interesse aplicadas às temperaturas máximas anuais entre 1950 e 2022, observa-se que apenas 10% das temperaturas excedem 41.05ºC em Portalegre, 41.91ºC em Évora, 42.74ºC em Beja e 38.12ºC em Sines, no entanto, 1% das temperaturas excedem 42.41ºC em Portalegre, 44.00ºC em Évora, 44.86ºC em Beja e 39.80ºC em Sines. Analisando as previsões obtidas com base nos níveis de retorno a 10 anos, verifica-se um aumento gradual das temperaturas ao longo dos anos, como esperado neste estudo. Comparando as log-verosimilhanças entre o modelo GEV estacionário e o não-estacionário, observa-se que a introdução de uma tendência linear no parâmetro de localização traz melhorias significativas, sendo preferível o modelo não estacionário, que apresenta melhores resultados, suportando a tese de que as temperaturas máximas estão a aumentar e que os verões alentejanos estão a ficar mais quentes.

Quanto à análise das maiores observações, verificou-se que o modelo GEV baseado no modelo limite das 2 maiores observações anuais, com parâmetro de forma negativo se ajusta melhor aos dados em contexto não estacionário, quando comparado com o modelo GEV baseado no modelo limite das 2 maiores observações anuais, em contexto estacionário. Considerando as quantidades de interesse aplicadas ao modelo das 2 maiores observações anuais entre 1950 e 2022, observa-se que apenas 10% das temperaturas excedem 41.36ºC em Portalegre, 42.40ºC em Évora, 43.09ºC em Beja e 38.35ºC em Sines, no entanto, 1% das temperaturas excedem 42.61ºC em Portalegre, 44.15ºC em Évora, 44.89ºC em Beja e 39.62ºC em Sines. Semelhante à análise dos máximos anuais, a análise das 2 maiores observações mostra que as previsões baseadas nos níveis de retorno a 10 anos indicam um aumento gradual das temperaturas ao longo do tempo, conforme esperado neste estudo. Comparando as log-verosimilhanças entre o modelo GEV_2 estacionário e o não-estacionário, observa-se também que a introdução de uma tendência linear no parâmetro de localização traz melhorias significativas, sendo mais uma vez, preferível o modelo não estacionário, que apresenta melhores resultados.

Quanto à análise dos excessos de nível, verificou-se que o modelo GP não estacionário com tendência linear na escala e com parâmetro de forma negativo se ajusta melhor às temperaturas máximas diárias que excedem um limiar elevado, quando comparado com o modelo GP em contexto estacionário. Considerando as quantidades de interesse aplicadas às temperaturas máximas diárias entre 1950 e 2022, observa-se que apenas 10% das temperaturas excedem 42.25ºC em Portalegre, 42.92ºC em Évora, 44.46ºC em Beja e 38.83ºC em Sines. Assim como observado anteriormente, a análise dos excessos acima de um nível elevado indica que as previsões baseadas nos níveis de retorno mostram um aumento gradual das temperaturas ao longo dos anos, em concordância com as expetativas deste estudo. Comparando as log-verosimilhanças entre o modelo GP estacionário e o não-estacionário, observa-se também que a introdução de uma tendência linear no parâmetro de escala traz melhorias significativas, sendo mais uma vez, preferível o modelo não estacionário, que apresenta melhores resultados. Para o modelo GP não-estacionário obtiveram-se as seguintes temperaturas máximas diárias previstas para 2025: $Tmax(2025)_{Portalegre} = 44.51°C$; $Tmax(2025)_{Évora} = 42.87°C$; $Tmax(2025)_{Beja} = 52.02°C$ e $Tmax(2025)_{Sines} = 38.28°C$.

Podemos assim concluir através da análise destes três modelos que, as temperaturas máximas estão a aumentar e que os verões alentejanos estão a ficar mais quentes.

Dado os resultados obtidos das previsões das temperaturas do Alentejo e a crescente preocupação com eventos extremos devido às alterações climáticas, há várias abordagens futuras que podem ser exploradas na análise e que passamos a descrever de forma sucinta:

- Incorporar covariáveis como, humidade, precipitação, índices de seca[1] (SPI, SPEI), concentração de gases de efeito estufa[2] e variabilidade oceânica[3] (NAO, AMO).
- Aplicação de métodos bayesianos para prever a ocorrência futura de eventos extremos.
- Uso de EVT multivariado para estudar a relação entre temperaturas extremas e eventos de seca prolongada.

Os resultados deste estudo evidenciam a importância da análise de eventos extremos como uma ferramenta essencial para a compreensão e previsão de temperaturas máximas. A utilização de métodos estatísticos apropriados permite não apenas medir e controlar esses eventos, mas também fornecer suporte para a definição de estratégias de mitigação das alterações climáticas. A capacidade de antecipar eventos extremos é fundamental para a adaptação de setores críticos, como a agricultura, os recursos hídricos e infraestruturas urbanas.

Portanto, a modelagem estatística de eventos extremos aplicada a temperaturas máximas é um instrumento essencial para melhorar a adaptação às mudanças climáticas.

[1] Índices de Seca: São métricas utilizadas para quantificar e monitorizar a severidade e a extensão das condições de seca. Estes índices consideram variáveis como precipitação, temperatura, humidade do solo e outros fatores climáticos. Exemplos comuns incluem o Índice de Precipitação Padronizado (SPI) e o Índice de Severidade de Seca de Palmer (PDSI).

[2] Gases de Efeito Estufa: São gases presentes na atmosfera que absorvem e emitem radiação infravermelha, contribuindo para o efeito estufa. Este fenómeno é essencial para manter a temperatura da Terra, mas o aumento na concentração destes gases, como dióxido de carbono (CO_2), metano (CH_4) e óxido nitroso (N_2O), tem levado ao aquecimento global.

[3] Variabilidade Oceânica: Refere-se às flutuações naturais nas condições dos oceanos, incluindo temperatura da superfície do mar, correntes e padrões de salinidade, que ocorrem em diferentes escalas temporais e espaciais. Fenómenos como El Niño e La Niña são exemplos de variabilidade oceânica que influenciam o clima global.

BIBLIOGRAFIA

Andrade, C., Fraga, H., & Santos, J. A. (2014). Climate change multi-model projections for temperature extremes in Portugal. *Atmospheric Science Letters*, *15*(2), 149–156. https://doi.org/10.1002/asl2.485

Barndorff-Nielsen, O. E., & Cox, D. R. (1994). Inference and Asymptotics. In *Compositional Data J. Aitchison* (Vol. 17, Issue 2). Springer US. https://doi.org/10.1007/978-1-4899-3210-5

Beirlant, Jan. (2004). *Statistics of extremes : theory and applications*. Wiley.

Bera, A. K., & Bilias, Y. (2001). Rao's score, Neyman's C(α) and Silvey's LM tests: an essay on historical developments and some new results. In *Journal of Statistical Planning and Inference* (Vol. 97). www.elsevier.com/locate/jspi

Blunden, J., Boyer, T., Bartow-Gillies, E., Beck, H. E., Becker, A., Bedka, K. M., Behe, C., Bell, G. D., Bellouin, N., Belmont, M., Benedetti, A., Bernhard, G. H., Berrisford, P., Berry, D. I., Bhatt, U. S., Bissolli, P., Bjerke, J., Blake, E. S., Blenkinsop, S., … McVicar, T. R. (2023). State of the climate in 2017 -. In *Bulletin of the American Meteorological Society* (Vol. 99, Issue 8, p. Si-S310). American Meteorological Society. https://doi.org/10.1175/2023bamsstateoftheclimate.1

Bocciolone, M., Gasparetto, M., Lagomarsino, S., Piccardo, G., Ratto, C. F., & Solari, G. (1993). Statistical analysis of extreme wind speeds in the Straits of Messina. *Journal of Wind Engineering and Industrial Aerodynamics*, *48*(2–3), 359–377. https://doi.org/10.1016/0167-6105(93)90146-F

Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P. W., Trisos, C., Romero, J., Aldunce, P., Barret, K., Blanco, G., Cheung, W. W. L., Connors, S. L., Denton, F., Diongue-Niang, A., Dodman, D., Garschagen, M., Geden, O., Hayward, B., Jones, C., … Ha, M. (2023). *IPCC, 2023: Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* (P. Arias, M. Bustamante, I. Elgizouli, G. Flato, M. Howden, C. Méndez-Vallejo, J. J. Pereira, R. Pichs-Madruga, S. K. Rose, Y. Saheb, R. Sánchez Rodríguez, D. Ürge-Vorsatz, C.

Xiao, N. Yassaa, J. Romero, J. Kim, E. F. Haites, Y. Jung, R. Stavins, … Y. Park, Eds.). https://doi.org/10.59327/IPCC/AR6-9789291691647.001

Cardoso, R. M., Lima, D. C. A., & Soares, P. M. M. (2023). How persistent and hazardous will extreme temperature events become in a warming Portugal? *Weather and Climate Extremes*, *41*. https://doi.org/10.1016/j.wace.2023.100600

Castillo, E., Hadi, A. S., Balakrishnan, N., & Sarabia, J. M. (2005). *Extreme Value and Related Models with Applications in Engineering and Science*. WILEY-INTERSCIENCE.

Chandra, M., Singpurwalla, N. D., & Stephens, M. A. (1981). Kolmogorov Statistics for Tests of Fit for the Extreme Value and Weibull Distributions. *Journal of the American Statistical Association*, *76*(375), 729. https://doi.org/10.2307/2287539

Charras-Garrido, M., & Lezaud, P. (2013). *Extreme Value Analysis: an Introduction*. http://www.sfds.asso.fr/journal

Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Cooley, D. (2009). Extreme value analysis and the study of climate change: A commentary on Wigley 1988. *Climatic Change*, *97*(1), 77–83. https://doi.org/10.1007/s10584-009-9627-x

Costa, A. C., Santos, J. A., & Pinto, J. G. (2012). Climate change scenarios for precipitation extremes in Portugal. *Theoretical and Applied Climatology*, *108*(1–2), 217–234. https://doi.org/10.1007/s00704-011-0528-3

Davison, A. C., & Smith, R. L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, *52*(3), 393–442. https://www.jstor.org/stable/2345667

Davison, A. C., Smith, R. L., & Davisont, A. C. (1990). Models for Exceedances over High Thresholds. In *Journal of the Royal Statistical Society. Series B (Methodological). J. R. Statist. Soc. B* (Vol. 52, Issue 3).

de Oliveira, J. T., & Gomes, M. I. (1984). Two Test Statistics for Choice of Univariate Extreme Models. In J. T. de Oliveira (Ed.), *Statistical Extremes and Applications* (pp. 651–668). Springer Netherlands. https://doi.org/10.1007/978-94-017-3069-3_50

Dixon, M. J., Tawn, J. A., & Vassie, J. M. (1998). SPATIAL MODELLING OF EXTREME SEA-LEVELS. *ENVIRONMETRICS*. https://doi.org/10.1002/(SICI)1099-095X(199805/06)9:3<283::AID-ENV304>3.0.CO;2

Embrechts, P., Kliippelberg, C., & Mikosch, T. (2003). *Modelling Extremal Events for Insurance and Finance Probability: Vol. Fourth*.

Fisher, R., & Tippett, L. (1927). *Limiting forms of the frequency distribution of the largest or smallest member of a sample*.

Fourth National Climate Assessment. (2018). *Impacts, Risks, and Adaptation in the United States: The Fourth National Climate Assessment, Volume II* (D. R. Reidmiller, C. W. Avery, D. R. Easterling, K. E. Kunkel, K. L. M. Lewis, T. K. Maycock, & B. C. Stewart, Eds.). https://doi.org/10.7930/NCA4.2018

Fraga, H., De Cortázar Atauri, I. G., Malheiro, A. C., Moutinho-Pereira, J., & Santos, J. A. (2017). Viticulture in Portugal: A review of recent trends and climate change projections. *Oeno One*, *51*(2), 61–69. https://doi.org/10.20870/oeno-one.2016.0.0.1621

Fraga, H., Malheiro, A. C., Moutinho-Pereira, J., & Santos, J. A. (2012). An overview of climate change impacts on European viticulture. *Food and Energy Security*, *1*(2), 94–110. https://doi.org/10.1002/fes3.14

Gagliardi, N., Arévalo, P., & Pamies, S. (2022). *Economic and Financial Affairs The Fiscal Impact of Extreme Weather and Climate Events: Evidence for EU Countries*. https://doi.org/10.2765/867213

Gilleland, E., & Katz, R. W. (2006). *ANALYZING SEASONAL TO INTERANNUAL EXTREME WEATHER AND CLIMATE VARIABILITY WITH THE EXTREMES TOOLKIT*.

Gomes, I., Alves, I. F., & Neves, C. (2013). Analise de Valores Extremos: Uma Introdução. *Instituto Nacional de Estatística*. https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=209134994&att_display=n&att_download=y

Gomes, M. I., Gomes, M. I., De Lisboa, U., & Van Montfort, M. A. J. (1985). *Exponentiality versus Generalized Pareto-Quick Tests*. https://www.researchgate.net/publication/356839843

Green, K. P. (2024). *Extreme Weather and Climate Change*.

Haan, L., & Ferreira, A. (2006). *Extreme Value Theory An Introduction*. Springer.

Hosking, J. R. M. (1984). Testing Whether the Shape parameter is Zero in the Generalized Extreme- Value Distribution. *Biometrika*, *71*(2), 367–374. https://about.jstor.org/terms

Intergovernmental Panel on Climate Change. (2015). *Climate change 2014 : synthesis report : longer report.*

Katz, R. W. (2010). Statistics of extremes in climate change. *Climatic Change*, *100*(1), 71–76. https://doi.org/10.1007/s10584-010-9834-5

Kotz, S., & Nadarajah, S. (2000). *Extreme Value Distributions - Theory and Applications*.

Marohn, F. (2000). Testing Extreme Value Models. *Extremes*, *3*(4).

McNeil, A., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: concepts, techniques, and tools*. Princeton University Press.

Mudelsee, M. (2020a). *Statistical Analysis of Climate Extremes*. Cambridge University Press. https://doi.org/10.1017/9781139519441

Mudelsee, M. (2020b). *Statistical Analysis of Climate Extremes*. https://doi.org/10.1017/9781139519441

Paulo, A. A., & Pereira, L. S. (2006). Drought concepts and characterization: Comparing drought indices applied at local and regional scales. *Water International*, *31*(1), 37–49. https://doi.org/10.1080/02508060608691913

Pereira, S. C., Carvalho, D., & Rocha, A. (2021). Temperature and precipitation extremes over the iberian peninsula under climate change scenarios: A review. In *Climate* (Vol. 9, Issue 9). MDPI. https://doi.org/10.3390/cli9090139

R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (4.1.1). R Foundation for Statistical Computing.

Rao, C. R. (1947). Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimatio. *Proceedings of the Cambridge Philosophical Society*, *44*, 50–57.

Reiss, R.-D., Thomas, M., Basel, B., Boston, ·, & Berlin, ·. (2007). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields: Vol. Third Edition*.

Salvadori, G., Kottegoda, N., & Rosso, R. (2007). *Extremes In Nature: An Approach Using Copulas* (Vol. 56). Springer.

SALVADORI, G., MICHELE, C., & KOTTEGODA, N. (2007). *EXTREMES IN NATURE*.

Santos, J. A., Belo-Pereira, M., Fraga, H., & Pinto, J. G. (2016). Understanding climate change projections for precipitation over western europe with a weather typing approach. *Journal of Geophysical Research*, *121*(3), 1170–1189. https://doi.org/10.1002/2015JD024399

Stram, D. O., & Lee, J. W. (1994). Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, *50*(4), 1171. https://doi.org/10.2307/2533455

Towler, E., Rajagopalan, B., Gilleland, E., Summers, R. S., Yates, D., & Katz, R. W. (2010). Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research*, *46*(11). https://doi.org/10.1029/2009WR008876

Wang', J. Z., & Cooke, P. (1996). Determination of Domains of Attraction Based on a Sequence of Maxima. *Australian Journal of Statistics*, *38*(2), 173–181.

APÊNDICE A

Nas Figura 53 à Figura 56, estão representados os gráficos de diagnóstico associados a k=5 e k=10, para o modelo das máximas observações.

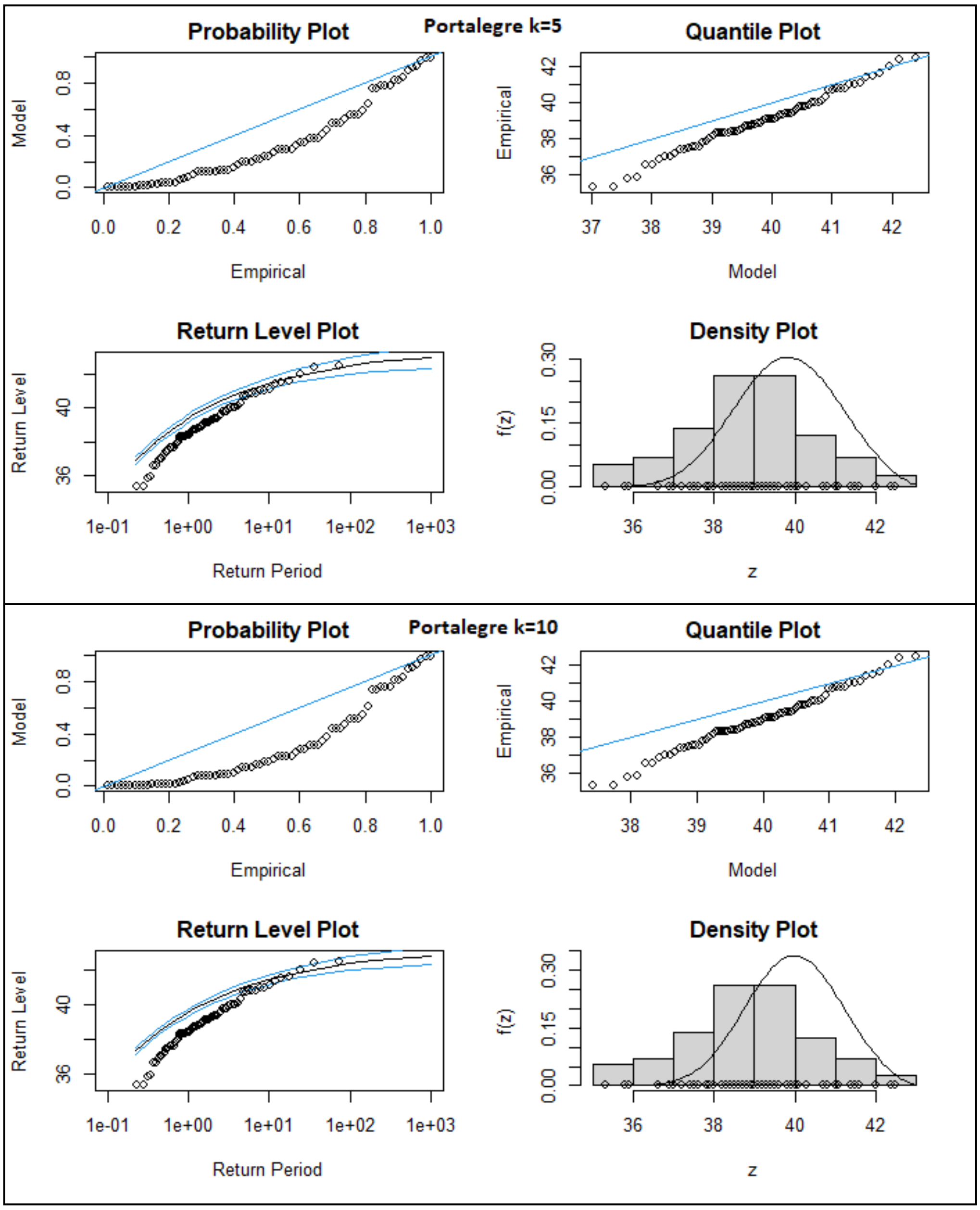


Figura 53 – Gráficos do ajustamento ao modelo GEV das 5 e 10 maiores observações, para Portalegre.

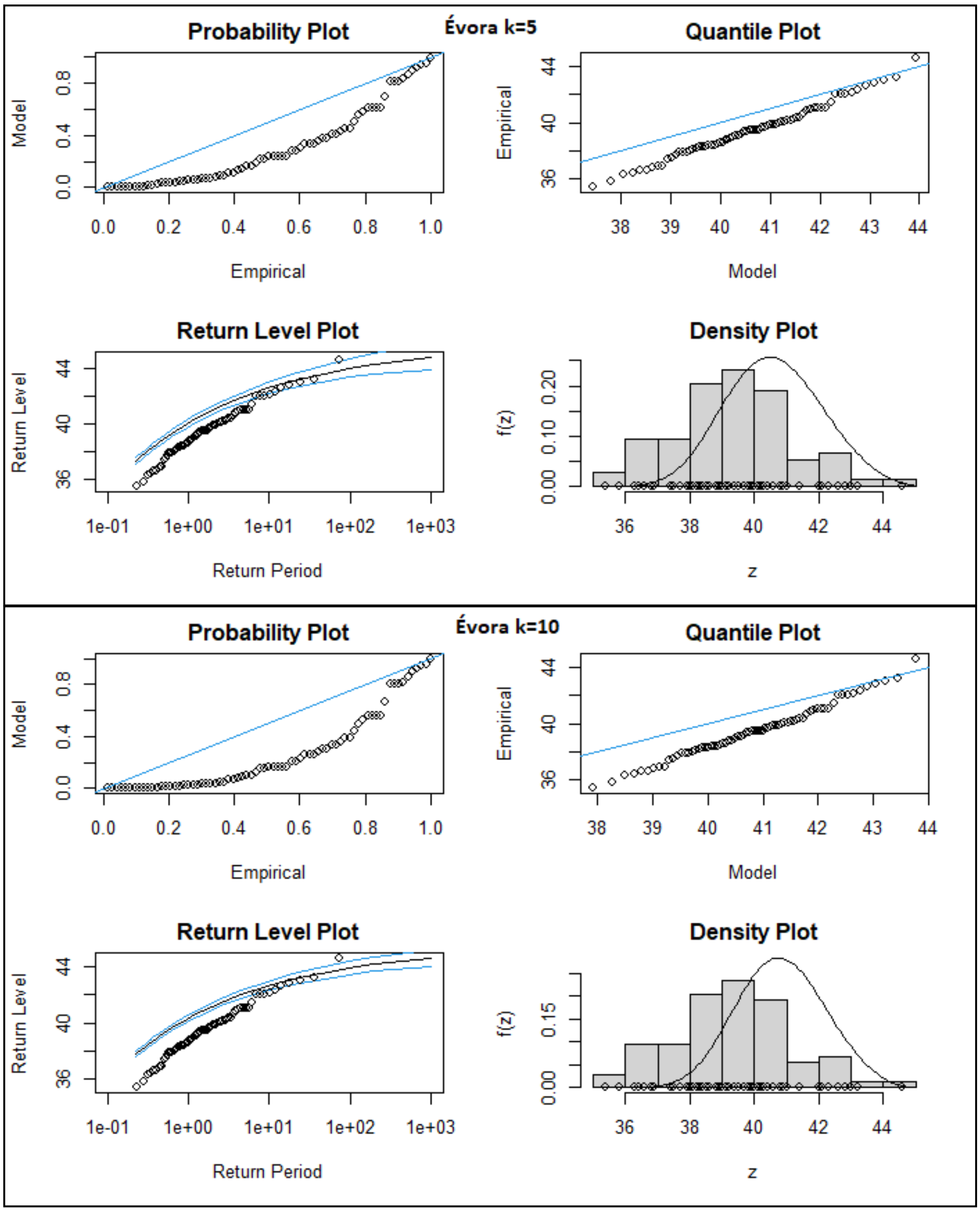


Figura 54 - Gráficos do ajustamento ao modelo GEV das 5 e 10 maiores observações, para Évora.

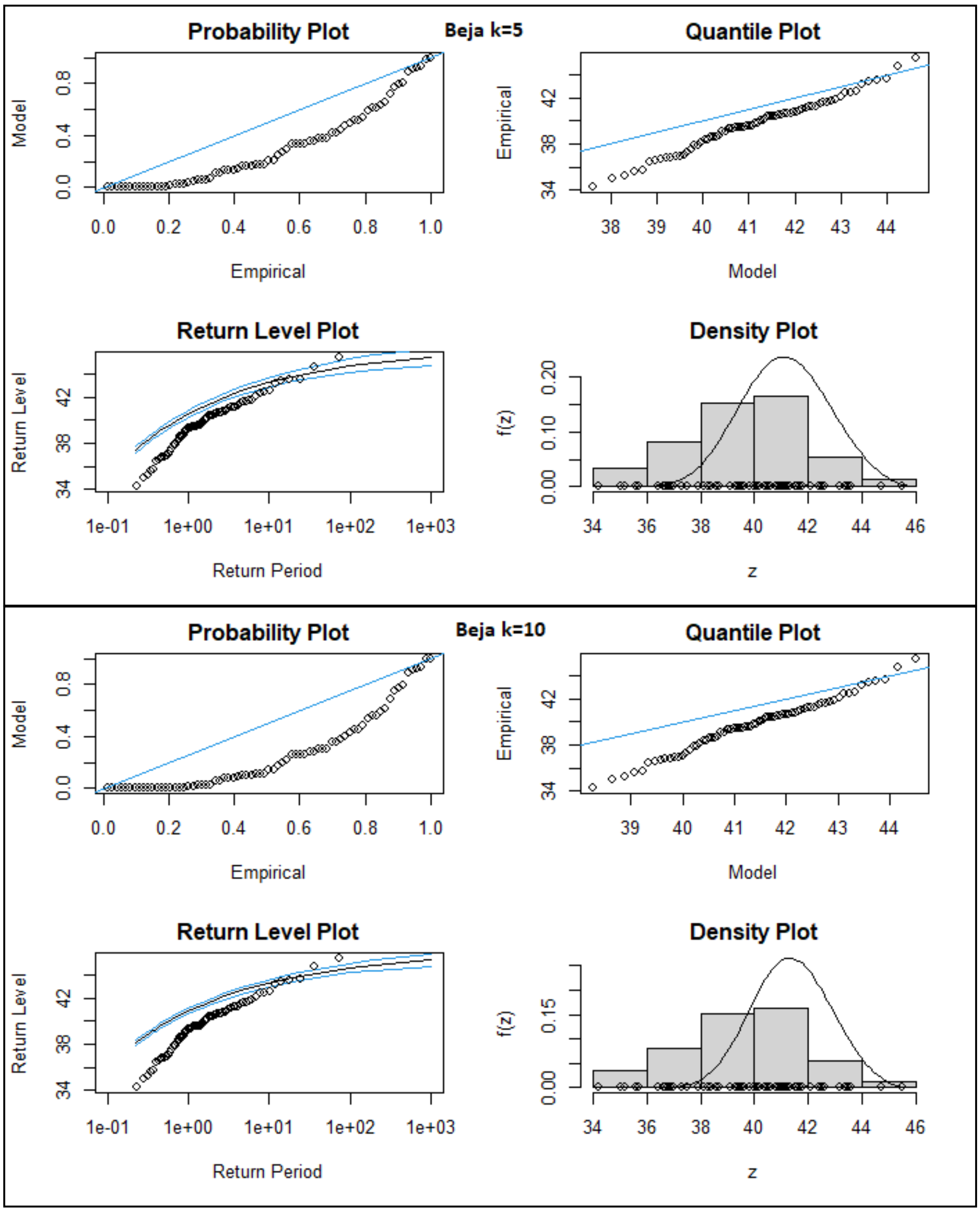


Figura 55 - Gráficos do ajustamento ao modelo GEV das 5 e 10 maiores observações, para Beja.

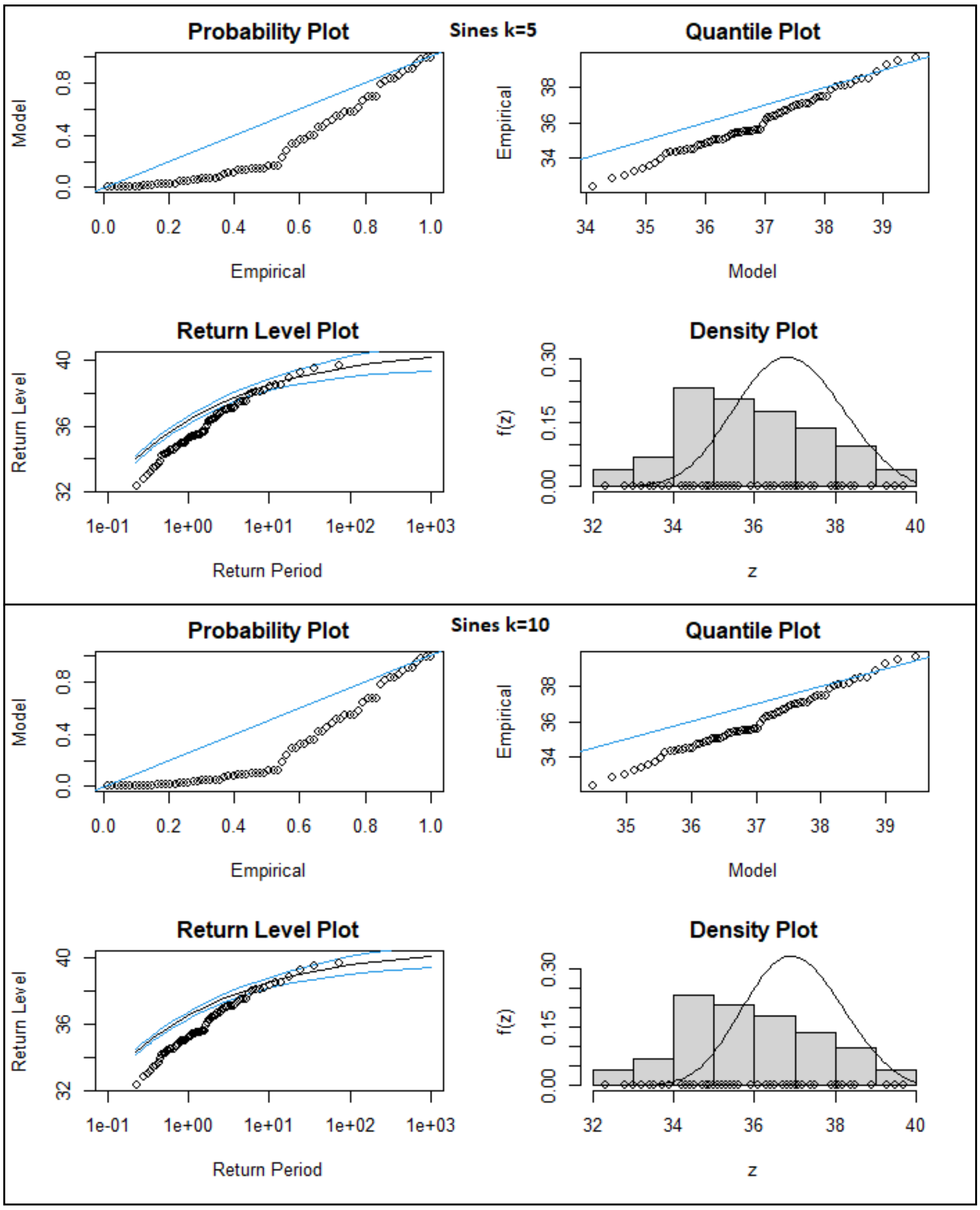


Figura 56 - Gráficos do ajustamento ao modelo GEV das 5 e 10 maiores observações, para Sines.

2024 ANA AMBRÓSIO

MODELAÇÃO ESTATÍSTICA DAS TEMPERATURAS MÁXIMAS
NA REGIÃO DO ALENTEJO

UNIVERSIDADE DE ÉVORA