

Universidade de Évora - Escola de Ciências e Tecnologia

Mestrado em Modelação Estatística e Análise de Dados

Dissertação

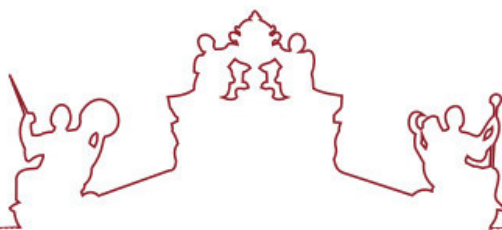
**Uma aplicação da teoria de valores extremos na modelação
da qualidade do ar durante a época de fogos extremos**

Edelise Brito Rodrigues Moreno

Orientador(es) | Lúgia Henriques-Rodrigues
Ediclê Duarte

Évora 2025





Universidade de Évora - Escola de Ciências e Tecnologia

Mestrado em Modelação Estatística e Análise de Dados

Dissertação

**Uma aplicação da teoria de valores extremos na modelação
da qualidade do ar durante a época de fogos extremos**

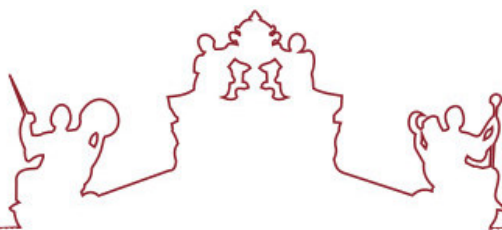
Edelise Brito Rodrigues Moreno

Orientador(es) | Lígia Henriques-Rodrigues

Ediclê Duarte

Évora 2025





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

Presidente | Dulce Gamito Pereira (Universidade de Évora)

Vogais | Anabela Afonso (Universidade de Évora) (Arguente)
Lígia Henriques-Rodrigues (Universidade de Évora)

AGRADECIMENTOS

Agradeço primeiramente, a Deus, todo-poderoso, pela força, proteção e amparo concedidos em todos os momentos desta caminhada.

Expresso a minha profunda gratidão à minha orientadora Professora Lúcia Rodrigues, pela disponibilidade constante, pelo acompanhamento atento e pelos contributos indispensáveis para a realização desta dissertação.

Expresso igualmente a minha sincera gratidão ao meu coorientador Professor Ediclê Duarte, pela orientação rigorosa, pelas sugestões valiosas e pelo incentivo constante ao longo deste percurso académico.

Por último, mas não menos importante, à minha família e aos meus amigos, sou igualmente grata pelo carinho e pelas palavras de incentivo que iluminaram os momentos mais desafiantes.

RESUMO

Uma Aplicação da Teoria de Valores Extremos na Modelação da Qualidade do Ar Durante a Época de Fogos Extremos em Portugal

Este trabalho analisou eventos extremos de poluição por PM_{10} nas regiões Norte e Centro de Portugal (2011–2020), com ênfase nos impactos dos incêndios florestais. Aplicou-se a Teoria dos Valores Extremos (TVE), utilizando modelos GEV e GPD, estacionários e não estacionários.

Os modelos não estacionários, incorporando covariáveis como tempo, temperatura, vento e área ardida, proporcionaram melhor descrição das variações dos extremos. Destaca-se o GPD não estacionário com escala log-linear dependente da área ardida, que evidenciou esta variável como principal fator de severidade dos episódios.

Os resultados confirmam que maiores áreas ardidas estão associadas a aumentos nos níveis e variabilidade de PM_{10} reforçando a relação entre incêndios e degradação da qualidade do ar. A modelação proposta constitui uma ferramenta eficaz para avaliação de risco e apoio à gestão da qualidade do ar em cenários de vulnerabilidade climática crescente.

Palavras chave: Teoria dos Valores Extremos, PM_{10} , poluição atmosférica, não estacionariedade, Portugal.

ABSTRACT

An Application of Extreme Value Theory in Air Quality Modeling During the Extreme Fire Season in Portugal

This study investigates extreme PM_{10} air pollution events in Northern and Central Portugal during 2011–2020, focusing on the impact of wildfires. Extreme Value Theory (EVT) was applied, using both block maxima (GEV) and threshold exceedance (GPD) models in stationary and non-stationary forms.

Non-stationary models, incorporating covariates such as time, temperature, wind speed, and burned area, provided a more realistic description of extreme variability. The log-linear non-stationary GPD, with scale dependent on burned area, performed best, highlighting this variable as the main driver of extreme PM_{10} events.

Results indicate that larger burned areas are associated with higher PM_{10} concentrations and variability, reinforcing the link between wildfires and air quality deterioration. The proposed modeling approach offers a robust tool for risk assessment and air quality management under increasing climate vulnerability.

Keywords: Extreme Value Theory, PM_{10} , air pollution, non-stationarity, Portugal.



ÍNDICE

1	INTRODUÇÃO.....	1
1.1	Descrição do problema	1
1.2	Contextualização.....	3
1.3	Objetivos	5
1.4	Descrição dos dados	6
1.4.1	Variáveis em estudo	7
1.4.2	Área de estudo	9
2	TEORIA DE VALORES EXTREMOS	11
2.1	Introdução.....	11
2.2	Modelos assintóticos	12
2.3	Parâmetros de interesse	14
2.4	Modelo dos máximos anuais.....	15
2.4.1	Modelo GEV estacionário.....	16
2.4.2	Modelo GEV não estacionário	21
2.5	Método de excedências	25
2.5.1	Modelo GPD estacionário	26
2.5.2	Modelo GPD não estacionário.....	29
3	ANÁLISE E DISCUSSÃO DOS RESULTADOS	33
3.1	Método dos máximos anuais	33
3.1.1	Estatística descritiva das variáveis.....	34
3.1.2	Ajuste do modelo GEV estacionário.....	41
3.1.3	Ajuste do modelo GEV não estacionário	46



3.2	Método de Excedências.....	53
3.2.1	Estatística descritiva das variáveis.....	53
3.2.2	Ajuste do modelo GPD estacionário.....	56
3.2.3	Ajuste do modelo GPD não estacionário.....	65
4	CONCLUSÕES	75
	BIBLIOGRAFIA	77

ÍNDICE DE FIGURAS

Figura 1: Regiões de Portugal (NUTII). Estão assinaladas as estações de monitorização da qualidade do ar (n=21) representadas pelos triângulos vermelho e as estações meteorológicas (n=23) representadas pelos triângulos azuis utilizadas neste estudo. Fontes: Agência Portuguesa do Ambiente (APA) e Instituto Português do Mar e da Atmosfera (IPMA).	10
Figura 2: Histograma da concentração PM_{10} ($\mu g/m^3$), temperatura do ar ($^{\circ}C$), velocidade do vento (m/s), e área ardida (ha). Dados recolhidos nas regiões Norte e Centro de Portugal, de 2011 a 2020.	36
Figura 3: Séries Temporais de PM_{10} ($\mu g/m^3$), Temperatura ($^{\circ}C$), Vento (m/s) e Área ardida (ha) nas Regiões Norte e Centro de Portugal (2011-2020).....	37
Figura 4: Distribuição mensal de PM_{10} , temperatura, vento e área ardida (2011-2020). Boxplots mostram a mediana, quartis e <i>outliers</i> . Datas indicam <i>outliers</i>	38
Figura 5: Matriz de correlação mostrando as relações lineares entre os máximos mensais de PM_{10} (Max_PM10), temperatura (Max_TEMP), velocidade do vento (Max_WS) e área total ardida (Max_AreaTotal_ha). As cores indicam a força e direção da correlação: azul para c correlação positiva, vermelho para correlação negativa e branco para correlação fraca ou nula. A intensidade da cor representa a magnitude da correlação, conforme a escala à direita. Dados referentes ao período de junho a outubro, de 2011 a 2020, nas regiões Norte e Centro de Portugal.	39
Figura 6: Gráficos das funções de autocorrelação e de autocorrelação parcial do PM_{10}	40
Figura 7: Gráficos do ajustamento das máximas mensais de PM_{10} da região Norte e Centro à distribuição GEV estacionário.	43
Figura 8: Ajuste do modelo GEV não estacionário para PM_{10} (modelo 1,4), em que a localização depende linearmente do tempo e da área ardida, a escala é constante e o parâmetro de forma é fixo.	50
Figura 9: Distribuição das variáveis PM_{10} , Temperatura, Velocidade do Vento e Área Ardida	55
Figura 10: Distribuição anual da serie dos máximos diários de quatro variáveis: PM_{10} , Temperatura, Velocidade do Vento e Área Ardida ao longo de um período de 10 anos, de 2011 a 2020.	56
Figura 11: Distribuição dos valores máximos diários de PM_{10} (limiar $> 45 \mu g/m$).	58



Figura 12: Distribuição de PM_{10} em relação ao limiar de $45 \mu\text{g}/\text{m}^3$	60
Figura 13: Gráficos do ajustamento de PM_{10} diários da região Norte e Centro à distribuição GPD	61
Figura 14: Gráficos de diagnóstico para o modelo GPD não estacionário ajustado.....	70



ÍNDICE DE TABELAS

Tabela 1: Estatística descritiva das variáveis em estudo	34
Tabela 2: Estimativas dos parâmetros do modelo GEV estacionário para os máximos mensais de PM_{10}	41
Tabela 3: Resultados do teste de razão de verossimilhança para comparação entre os modelos GEV e Gumbel na modelação dos máximos mensais de PM_{10}	44
Tabela 4: Estimativa de níveis de retorno de PM_{10} ($\mu g/m^3$) para diferentes períodos.	45
Tabela 5: Probabilidade de excedência para diferentes concentrações de PM_{10}	45
Tabela 6: Quantis extremais de PM_{10}	46
Tabela 7: Resultados dos modelos GEV ($\mu t, \sigma, \xi$),	48
Tabela 8: Resultados dos modelos GEV ($\mu t, \sigma A = exp\beta 0 + \beta 1A, \xi$)	48
Tabela 9: Estimativas dos parâmetros do modelo GEV não estacionário (modelo 1,4) para PM_{10} , em que o parâmetro de localização depende linearmente do tempo e da área ardida, a escala é constante e o parâmetro de forma é fixo.	50
Tabela 10: Estimativa dos níveis de retorno de PM_{10} ($\mu g/m^3$) para diferentes períodos de retorno.....	51
Tabela 11: Probabilidade de excedência para diferentes limiares de PM_{10}	52
Tabela 12: Quantis extremais estimados de PM_{10} ($\mu g/m^3$) para junho de 2025.....	52
Tabela 13: Estatística descritiva dos máximos diários das variáveis em estudo.....	54
Tabela 14: Estatísticas descritivas das excedências acima de $45 \mu g/m^3$ da concentração diária de PM_{10}	57
Tabela 15: Estimativas dos parâmetros do modelo GPD estacionário para PM_{10}	61
Tabela 16: Resultados do teste de razão de verossimilhança para comparação entre os modelos GPD e modelo exponencial na modelação dos máximos mensais de PM_{10}	62
Tabela 17: Estimativas dos níveis de retorno de PM_{10} ($\mu g/m^3$) para diferentes períodos de retorno.....	63
Tabela 18: Probabilidade de excedência para diferentes limiares de PM_{10}	64
Tabela 19: Quantis extremais estimados de PM_{10} ($\mu g/m^3$)	64
Tabela 20: Resultados dos modelos GPD com parâmetro de escala linear	67



Tabela 21: Resultados dos modelos GPD com parâmetro de escala log-linear	68
Tabela 22: Estimativas dos parâmetros do modelo GPD não estacionário para PM_{10}	68
Tabela 23: Estimativas dos níveis de retorno de PM_{10} ($\mu\text{g}/\text{m}^3$) para diferentes períodos de retorno no modelo GPD não estacionário	72
Tabela 24: Probabilidades de excedência de PM_{10} ($\mu\text{g}/\text{m}^3$) no modelo GPD não estacionário	72
Tabela 25: Quantis extremos estimados de PM_{10} ($\mu\text{g}/\text{m}^3$) no modelo GPD não estacionário	73

ABREVIATURAS & ACRÓNIMOS

AIC: Critério de informação de Akaike (do inglês, *Akaike Information Criterion*)

BIC: Critério de informação Bayesiano (do inglês, *Bayesian Information Criterion*)

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

FDA: Função de distribuição acumulada

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

GPD: Distribuição Generalizada de Pareto (do inglês, *Generalized Pareto distribution*)

IC: Intervalo de confiança

i.i.d.: Independente e identicamente distribuída

QQ: Gráfico Quantil-Quantil

MLE: Estimadores de máxima verosimilhança (do inglês, *Maximum likelihood estimator*)

PP: Gráfico Probabilidade-Probabilidade

POT: Método das excedências (do inglês, *Peaks Over Threshold*)

RV: Razão de Verosimilhanças

TRV: Teste de Razão de Verosimilhanças

v.a.: Variável aleatória

INTRODUÇÃO

Os incêndios florestais são um perigo ambiental recorrente que afeta várias regiões em todo o mundo e representam ameaças significativas para a saúde pública (Monteiro dos Santos *et al.*, 2024).

A complexa relação entre incêndios florestais, variáveis meteorológicas e poluição do ar requer uma análise aprofundada para melhor compreender as suas implicações. Este projeto tem como objetivo aprofundar o conhecimento sobre o impacto multifacetado dos incêndios florestais com ênfase no estudo da qualidade do ar em Portugal durante as épocas de incêndios de 2011 a 2020.

1.1 Descrição do problema

Portugal tem enfrentado nas últimas décadas um aumento na frequência e severidade dos incêndios florestais, particularmente durante os meses de verão, caracterizados por condições meteorológicas propícias à ignição e propagação do fogo (Cruz *et al.*, 2025) Estes eventos, que por vezes atingem proporções extremas em termos de área ardida e duração, representam uma ameaça significativa não apenas para os ecossistemas e a economia, mas também para a qualidade do ar e a saúde pública (Monteiro dos Santos *et al.*, 2024).

Os incêndios florestais extremos em Portugal constituem um desafio ambiental e social de elevada complexidade, cuja análise exige uma abordagem que transcenda as condições meteorológicas locais. A sua ocorrência está estreitamente relacionada com fatores climáticos de maior escala, como as variações de temperatura, humidade e velocidade do vento, que influenciam de forma determinante a intensidade e a propagação do fogo (Carmo, 2022). Esta perspectiva mais abrangente evidencia que os incêndios não devem ser encarados como eventos isolados, mas sim como parte de padrões

climáticos complexos, cuja análise em múltiplas escalas temporais e espaciais é fundamental para compreender a sua dinâmica e sustentar estratégias eficazes de prevenção e gestão de risco (Campos *et al.*, 2023)

Para uma gestão de risco eficaz, é crucial integrar diversas fontes de dados. A combinação de informação meteorológica, climática e de uso do solo permite uma melhor antecipação dos incêndios extremos e a definição de medidas adequadas de proteção civil e planeamento florestal. Este tipo de abordagem multidisciplinar e adaptada às especificidades regionais é essencial para o desenvolvimento de políticas robustas de prevenção e gestão de incêndios florestais em Portugal.

A queima de biomassa florestal liberta para a atmosfera grandes quantidades de poluentes, entre os quais se destaca o material particulado inalável. Este inclui as partículas com diâmetro aerodinâmico inferior a 10 micrómetros (PM_{10}) e as partículas finas com diâmetro inferior a 2,5 micrómetros ($PM_{2,5}$). Ambas conseguem penetrar no sistema respiratório humano, estando associadas a efeitos adversos comprovados na saúde respiratória e cardiovascular (O.M.S, 2021). Em situações de fogos extremos, a magnitude das emissões pode levar a episódios de poluição do ar de intensidade também extrema, expondo vastas áreas e populações a concentrações elevadas de PM_{10} por períodos prolongados.

A modelação da qualidade do ar em cenários de incêndios florestais constitui um desafio particularmente complexo, condicionado por múltiplos fatores, entre os quais se destacam a quantidade e a composição da biomassa queimada, as condições meteorológicas (como temperatura, velocidade e direção do vento) e as características topográficas da região afetada.

A qualidade do ar durante os incêndios florestais em Portugal constitui um tema de estudo relevante, com diversas investigações a demonstrarem que a população fica exposta a partículas poluentes (PM_{10} e $PM_{2,5}$), o que pode ter efeitos significativos na saúde e na mortalidade (Augusto, 2020; Carmo, 2022; Slezakova *et al.*, 2013)

Apesar disso, é necessário aprofundar a investigação sobre a modelação estatística para prever picos de concentração de PM_{10} durante os incêndios. A influência de variáveis como a temperatura, a velocidade do vento e a área queimada na qualidade do ar ainda precisa de ser mais bem compreendidas. É crucial perceber como eventos como as ondas de calor e as secas podem interagir com os incêndios, criando fenómenos em cadeia que afetam não só o país, mas também a Europa (Sutanto *et al.*, 2020;

Vitolo *et al.*, 2019). Estudos no Norte de Portugal já confirmaram que os níveis de partículas no ar aumentam de forma significativa durante os incêndios (Slezkova *et al.*, 2013). Em suma, é necessária uma abordagem multidisciplinar para compreender e reduzir os efeitos dos incêndios na qualidade do ar em Portugal (Carmo, 2022).

Neste contexto, a Teoria de Valores Extremos (TVE) constitui uma abordagem estatística particularmente adequada para o estudo de fenómenos de ocorrência rara, mas de elevada intensidade, permitindo caracterizar e quantificar a probabilidade de eventos extremos, como picos de concentração de poluentes atmosféricos.

A sua aplicação à modelação dos extremos de PM₁₀ em Portugal, durante episódios de incêndios florestais severos, e à análise da sua relação com variáveis como a temperatura, a velocidade do vento e a área ardida, permanece ainda pouco explorada. Este trabalho pretende colmatar esta lacuna, aplicando a Teoria de Valores Extremos para avaliar como estas variáveis influenciam a ocorrência e a magnitude dos episódios de poluição atmosférica, contribuindo assim para o desenvolvimento de estratégias de mitigação e sistemas de alerta mais eficazes.

1.2 Contextualização

O ar que respiramos é fundamental para a saúde humana e para o equilíbrio ambiental. Contudo, a poluição atmosférica, proveniente de atividades humanas e de fontes naturais, constitui um problema global com efeitos negativos amplamente documentados (Associação Portuguesa dos Recursos Hídricos, 2020). Em Portugal, a Agência Portuguesa do Ambiente (APA) é responsável pelo monitoramento da qualidade do ar, avaliando poluentes como material particulado (PM₁₀ e PM_{2,5}), ozono troposférico (O₃), dióxido de azoto (NO₂), dióxido de enxofre (SO₂), monóxido de carbono (CO) e benzeno (C₆H₆). Os padrões de qualidade do ar resultam da conjugação de evidências epidemiológicas com critérios de viabilidade regulatória, seguindo as diretivas europeias e recomendações da Agência Europeia do Ambiente (APA, 2022).

A Organização Mundial da Saúde (O.M.S.) estabelece diretrizes baseadas em evidências robustas, recomendando valores mais restritivos do que a legislação europeia: para PM₁₀, uma média anual de 15 µg/m³ e diária de 45 µg/m³, e para PM_{2,5}, 5 µg/m³ (anual) e 15 µg/m³ (diária) (World Health Organization [WHO], 2021). A legislação europeia e nacional permite limites diários de 50 µg/m³ para PM₁₀ (com até 35 excedências

por ano) e $25 \mu\text{g}/\text{m}^3$ anuais para $\text{PM}_{2,5}$, refletindo uma abordagem que combina saúde pública e viabilidade técnica e económica. Eventos extremos, como incêndios florestais, podem provocar aumentos abruptos de PM_{10} , representando riscos agudos para a saúde, particularmente nas regiões Norte e Centro de Portugal (WHO, 2021; Moreira, 2020).

A qualidade do ar em Portugal é afetada por fontes antropogénicas, como tráfego rodoviário, indústria e agricultura, e por fontes naturais, como poeiras do Saara, aerossóis marinhos e incêndios florestais (APA, 2022). Entre estas, os incêndios têm vindo a aumentar em frequência e intensidade devido a condições meteorológicas favoráveis à ignição e propagação, como altas temperaturas, baixa humidade e ventos fortes (Moreira, 2020). Estes episódios extremos libertam grandes quantidades de poluentes, destacando-se o PM_{10} , cuja inalação está associada a agravamento de doenças crónicas e aumento da mortalidade (WHO, 2021). A relação entre incêndios e aumento de PM_{10} tem sido estudada em Portugal (Augusto, 2020; Chen *et al.*, 2021; de Souza Fernandes Duarte *et al.*, 2021; Faustini *et al.*, 2015; Hänninen *et al.*, 2009; Slezakova *et al.*, 2013), e os dados de monitorização confirmam aumentos significativos durante e após grandes incêndios (APA, 2023). Estudos também analisaram o papel de variáveis meteorológicas, como temperatura e vento, na propagação do fogo e dispersão dos poluentes (Campos *et al.*, 2023; Couto *et al.*, 2022; Turco *et al.*, 2013, 2017, 2019; Viegas, 2018).

A interação entre incêndios, poluentes e variáveis meteorológicas em Portugal continua pouco explorada. Condições de calor extremo e seca favorecem a propagação dos incêndios, libertando poluentes que degradam significativamente a qualidade do ar. A ocorrência simultânea destes perigos, denominada risco composto, bem como a sua manifestação sequencial (eventos em cadeia), evidencia como as condições meteorológicas podem amplificar tanto a frequência como a intensidade dos incêndios e da poluição resultante, aumentando os riscos de morbilidade e mortalidade por doenças cardiorrespiratórias (Sutanto *et al.*, 2020; Tarín-Carrasco *et al.*, 2021).

A modelação estatística de eventos extremos de PM_{10} durante incêndios severos é desafiadora devido à sua natureza rara e à elevada variabilidade. Métodos tradicionais apresentam limitações na caracterização da cauda da distribuição, comprometendo a análise das situações mais críticas. A Teoria de Valores Extremos (TVE) constitui uma

abordagem robusta para estudar eventos raros e de grande magnitude em diferentes domínios, incluindo o ambiental (Coles, 2001). Apesar do seu potencial, a aplicação da TVE a PM_{10} em contextos de incêndios em Portugal, considerando variáveis como temperatura, velocidade do vento e área ardida, é ainda escassa na literatura. A quantificação da probabilidade e magnitude destes eventos é essencial para uma avaliação mais rigorosa dos riscos à saúde pública e para o planeamento de estratégias de mitigação e alerta.

Esta tese propõe, assim, a aplicação da TVE na modelação da qualidade do ar durante a época de incêndios severos em Portugal, com foco em PM_{10} . O estudo avaliará a influência de variáveis meteorológicas e da área ardida na ocorrência e intensidade dos eventos extremos, gerando conhecimento científico relevante para a gestão da qualidade do ar em cenários de incêndios de grande intensidade.

1.3 Objetivos

O objetivo geral desta tese consiste em aplicar a Teoria de Valores Extremos para modelar os eventos extremos de concentração de PM_{10} durante a época de incêndios severos em Portugal, no período de 2010 a 2020, avaliando a influência de variáveis meteorológicas, como temperatura e velocidade do vento, e da área ardida. Pretende-se assim investigar a relação entre a ocorrência de incêndios e os níveis de PM_{10} , com foco na deteção de eventos extremos de poluição e na avaliação dos riscos à saúde humana decorrentes da exposição aguda e crónica. A caracterização desses episódios é fundamental para fundamentar políticas públicas de mitigação e adaptação, particularmente em contextos de crescente pressão climática.

De forma a conseguir alcançar o objetivo geral desta tese, apresentamos os objetivos específicos que indicam os passos adotados neste estudo:

- Analisar o comportamento individual das variáveis em estudo, identificando padrões e características relevantes;
- Investigar as associações entre as variáveis analisadas, de forma a compreender interações relevantes para a modelação da poluição atmosférica extrema;

- Aplicar metodologias paramétricas de Estatística de Extremos univariada, a fim de caracterizar a qualidade do ar durante a temporada de incêndios florestais, com base em dados de área ardida, concentrações de poluentes e variáveis meteorológicas;
- Ajustar modelos de regressão para eventos extremos, estabelecendo relações entre as variáveis independentes em estudo e uma variável resposta representativa da qualidade do ar;
- Comparar o desempenho dos diferentes modelos de regressão, identificando aqueles que melhor se adequam à descrição dos eventos extremos de PM_{10} .

1.4 Descrição dos dados

Em Portugal, os incêndios florestais representam uma fonte significativa de emissões atmosféricas, com impacto direto no aumento das concentrações de material particulado, especialmente as frações PM_{10} . Para o efeito, foi construída uma base de dados integrada, composta por séries horárias de concentração de PM_{10} , variáveis meteorológicas, área ardida e georreferenciação das estações de monitoramento. As concentrações de PM_{10} , expressas em $\mu\text{g}/\text{m}^3$, foram registadas em 21 estações de fundo da APA localizadas nas regiões Norte e Centro de Portugal continental — áreas estratégicas devido à sua elevada vulnerabilidade a fenômenos climáticos extremos, como secas prolongadas, ondas de calor e incêndios florestais, cuja frequência e severidade tendem a intensificar-se sob o efeito das alterações climáticas (Turco *et al.*, 2019). A partir das séries horárias, foram obtidas estatísticas diárias e mensais (médias e máximos), restritas ao período crítico de incêndios (junho a outubro). Todos os dados utilizados são de domínio público e acessíveis via plataforma QualAR (<https://qualar.apambiente.pt>).

A base de dados QualAR, gerida pela APA, classifica as estações de monitorização da qualidade do ar com base em dois critérios: o contexto geográfico (urbano, suburbano ou rural) e o tipo de influência emissiva predominante (fundo, tráfego ou industrial), conforme estabelecido pela Decisão da Comissão Europeia 2001/752/CE, de 17 de outubro de 2001. As estações classificadas como “de fundo” localizam-se em áreas geográficas distantes de fontes antropogénicas diretas, tais como vias de grande tráfego

ou zonas industriais, o que as torna particularmente adequadas para avaliar o impacto de fenómenos de larga escala, como os incêndios florestais. A rede de monitorização da APA opera em conformidade com a legislação europeia vigente, nomeadamente a Diretiva 2008/50/CE, de 21 de maio de 2008, que define os padrões e procedimentos para a avaliação e gestão da qualidade do ar ambiente.

Os dados meteorológicos utilizados neste estudo foram cedidos pelo Instituto Português do Mar e da Atmosfera (IPMA; www.ipma.pt/). Foram analisadas observações horárias provenientes de 23 estações meteorológicas distribuídas pelas regiões Norte e Centro, incluindo: temperatura do ar a 2 m acima da superfície e velocidade do vento a 10 m de altura em m/s. Essas variáveis foram selecionadas pela sua relevância na modulação da dispersão de poluentes atmosféricos e na dinâmica de propagação de incêndios florestais. A partir das séries horárias, foram calculados valores médios diários e mensais, permitindo a caracterização das condições meteorológicas durante o período crítico de incêndios e sua associação com os níveis de PM₁₀.

Os dados de área ardida (em hectares; ha) utilizados neste estudo foram disponibilizados pelo Instituto da Conservação da Natureza e das Florestas (ICNF; <https://www.icnf.pt/>), referentes ao território continental português no período de 2011 a 2020, com resolução diária. Esses valores são estimados a partir da integração de observações in situ e deteções por sensoriamento remoto, permitindo mapear com precisão a extensão espacial dos incêndios. A delimitação das áreas queimadas considera informações detalhadas sobre o momento de ignição e extinção de cada evento, conforme metodologias validadas por Pereira *et al.* (2011). Adicionalmente, as séries históricas são frequentemente utilizadas para analisar mudanças no regime de incêndios, particularmente em resposta a variações climáticas e à eficácia de políticas de gestão do fogo, como discutido em Parente *et al.* (2016, 2019).

1.4.1 Variáveis em estudo

As variáveis analisadas neste estudo compreendem três conjuntos principais, todos georreferenciados e temporalmente alinhados para o território continental português entre junho de 2011 e outubro de 2020:

1. Variáveis meteorológicas: temperatura do ar a 2 m de altura ($^{\circ}\text{C}$) e velocidade do vento a 10 m (m/s) distribuídas estrategicamente por regiões com elevada recorrência de incêndios, incluindo as zonas Norte e Centro — coincidentes espacialmente com as estações de monitorização de PM_{10} .
2. Área ardida (hectares; ha): estimada com base em dados integrados de observação terrestre e deteção por satélite, fornecidos pelo ICNF, com resolução diária. Esses dados permitem quantificar a extensão espacial dos incêndios e sua evolução temporal ao longo da década estudada.
3. PM_{10} : concentrações horárias de material particulado com diâmetro inferior a $10\ \mu\text{m}$ ($\mu\text{g}/\text{m}^3$), fornecidos pela QualAR (APA), localizadas nas regiões Norte e Centro de Portugal. Essas estações foram selecionadas por estarem distantes de fontes emissoras pontuais (como tráfego ou indústria), tornando-as adequadas para captar a influência de incêndios florestais na qualidade do ar.

A base de dados utilizada neste estudo foi composta por observações diárias das variáveis PM_{10} , temperatura, velocidade do vento e área ardida nas regiões Norte e Centro de Portugal. A partir desses dados, foram calculadas tanto as médias diárias quanto os máximos diários, que, posteriormente, foram agregados em máximos mensais, considerando o período de 2011 a 2020, correspondente à época de maior suscetibilidade a incêndios florestais (junho a outubro). Para cada mês (série dos máximos), selecionou-se o valor máximo entre as duas regiões, originando a série temporal designada NorCentro. Esta série constitui uma medida consolidada das condições mais extremas, permitindo uma avaliação integrada e robusta do risco de incêndios florestais no território.

As variáveis: temperatura, velocidade do vento e área ardida foram submetidas à normalização Z , de forma a padronizar suas magnitudes e permitir comparabilidade entre séries. A normalização consiste em subtrair a média de cada variável e dividir pelo seu desvio padrão, resultando em séries com média zero e desvio padrão unitário. Este procedimento assegura que todas as variáveis contribuam de maneira equilibrada para análises subsequentes, incluindo a avaliação de extremos e a integração das condições meteorológicas e ambientais na série NorCentro.

Todos os dados passaram por um rigoroso protocolo de controlo de qualidade, incluindo preenchimento de falhas, deteção e remoção de *outliers*, e correção de

inconsistências temporais ou espaciais, assegurando a robustez estatística das análises subsequentes.

1.4.2 Área de estudo

Portugal situa-se no sudoeste da Europa, na Península Ibérica, com extensa costa atlântica a oeste e sul (Figura 1), posicionando-se numa zona de transição climática entre as latitudes médias e o domínio subtropical. O território continental apresenta, predominantemente, clima mediterrâneo: quente e seco no verão (Csa, segundo a classificação de Köppen-Geiger) no Sul, e com verões mais amenos (Csb) na maior parte das regiões Norte e Centro, onde se insere a área focal deste estudo. Localmente, observam-se ainda pequenos núcleos com clima semiárido (BSk).

As regiões Norte e Centro foram selecionadas como foco da investigação devido à sua elevada vulnerabilidade a fenómenos climáticos extremos — nomeadamente secas severas, ondas de calor e incêndios florestais — cuja frequência e intensidade se projetam aumentar com as alterações climáticas (Turco *et al.*, 2019). Essa combinação de fatores torna a área particularmente relevante para estudos sobre os impactos atmosféricos do fogo. A Figura 1 apresenta a localização das estações de monitorização utilizadas, estrategicamente distribuídas para captar a variabilidade espacial e temporal das variáveis em análise. Para este estudo, apenas os meses de junho–julho–agosto–setembro–outubro de 2011–2020 foram considerados pois representam o período seco com maior frequência de incêndio florestais.

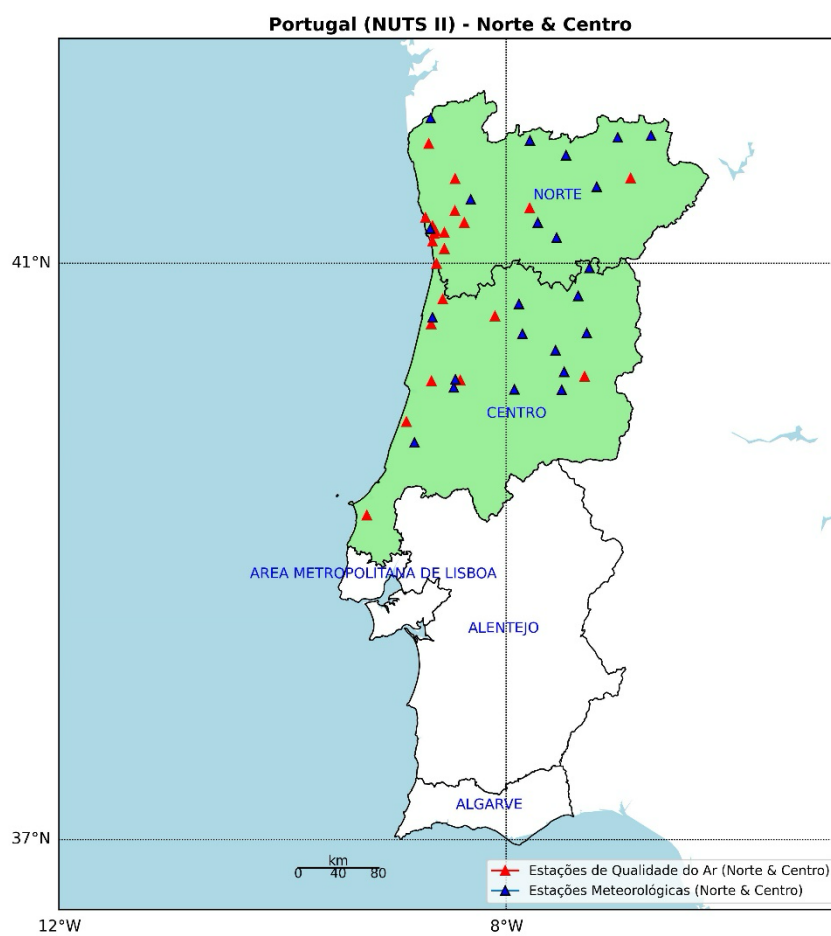


Figura 1: Regiões de Portugal (NUTS II). Estão assinaladas as estações de monitorização da qualidade do ar (n=21) representadas pelos triângulos vermelho e as estações meteorológicas (n=23) representadas pelos triângulos azuis utilizadas neste estudo. Fontes: Agência Portuguesa do Ambiente (APA) e Instituto Português do Mar e da Atmosfera (IPMA).

TEORIA DE VALORES EXTREMOS

O presente capítulo tem como objetivo fornecer a base conceptual e teórica que fundamenta a metodologia e a análise desenvolvida nesta tese. Abordam-se os principais pilares teóricos necessários à modelação estatística de eventos extremos de concentração de PM_{10} durante a época de incêndios florestais em Portugal. Começa-se por apresentar a Teoria de Valores Extremos (TVE), explicitando seus princípios fundamentais e as distribuições estatísticas centrais para a análise de eventos raros: a Distribuição de Valores Extremos Generalizada (GEV) e a Distribuição de Pareto Generalizada (GPD). A articulação destes conceitos fornece o suporte necessário para a compreensão da metodologia adotada e para a interpretação dos resultados apresentados nos capítulos subsequentes.

2.1 Introdução

A TVE representa uma área da estatística dedicada à análise do comportamento de ocorrências pouco comuns, situadas nas extremidades das distribuições de probabilidade. Estes eventos, embora pouco frequentes, podem ter impactos desproporcionais, sendo frequentemente objeto de estudo em áreas como hidrologia, climatologia, finanças, engenharia e, mais recentemente, nas ciências ambientais (Coles, 2001).

Diferentemente da estatística convencional, que se foca na média e na dispersão dos dados, a TVE dedica-se à modelação dos valores mais excepcionais, sejam eles máximos ou mínimos, que se afastam do padrão “típico” da distribuição. A sua relevância é particularmente notória na avaliação de riscos ligados a catástrofes naturais, a níveis de poluição elevados, ou a outros fenómenos agravados pelas mudanças climáticas. O

aumento da ocorrência de eventos extremos, como vagas de calor, períodos de seca intensa e grandes incêndios florestais, sublinha a urgência de dispor de abordagens sólidas para a sua quantificação e previsão (IPCC, 2021).

Os alicerces teóricos da TVE residem em princípios assintóticos da teoria da probabilidade, nomeadamente o teorema de Fisher–Tippett–Gnedenko. Este teorema postula que a distribuição dos valores máximos (devidamente normalizados) de uma amostra tenderá a convergir para uma das três famílias de distribuições de valores extremos: Gumbel, Fréchet ou Weibull. A distribuição generalizada de valores extremos (GEV) permite unificar estas três formas, sendo frequentemente empregada na modelação baseada em blocos (Coles, 2001).

Por outro lado, para análise de valores que ultrapassam um determinado limiar elevado, recorre-se à distribuição generalizada de Pareto (GPD), que descreve o comportamento limite das caudas da distribuição (Pickands, 1975).

Este capítulo apresenta os fundamentos da TVE, com ênfase nos modelos assintóticos, nas abordagens baseadas em máximos anuais (GEV) e nas excedências (GPD), tanto em cenários estacionários como não estacionários. As ferramentas aqui introduzidas fornecerão a base metodológica para a análise estatística dos episódios de poluição extrema associados a fogos florestais em Portugal, que será desenvolvida nos capítulos seguintes.

2.2 Modelos assintóticos

A estrutura teórica da TVE baseia-se em princípios assintóticos que descrevem o comportamento dos valores mais elevados de uma sequência de variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.). O teorema central que sustenta esta teoria é o de Fisher–Tippett–Gnedenko, o qual estabelece que, mediante uma normalização apropriada, a distribuição dos máximos converge para uma das três famílias de distribuições de valores extremos: Gumbel, Fréchet ou Weibull (Embrechts *et al.*, 1997; Coles, 2001).

Seja (X_1, X_2, \dots, X_n) uma sequência de v.a.'s i.i.d. com função de distribuição acumulada (FDA) F . Definimos o máximo da amostra como:

$$M_n = \max(X_1, X_2, \dots, X_n).$$

A FDA do máximo, M_n é dada por:

$$P(M_n \leq z) = F^n(z).$$

No entanto, esta expressão tende para 0 ou 1, quando $n \rightarrow \infty$, não permitindo inferências sobre o comportamento assintótico. Para ultrapassar esta limitação, aplica-se uma normalização linear:

$$Z_n = \frac{M_n - b_n}{a_n},$$

considerando sequências apropriadas de constantes normalizadoras $a_n > 0$, $b_n \in \mathbb{R}$ (Coles, 2001). Sob condições gerais, a distribuição de Z_n converge, quando $n \rightarrow \infty$, para uma distribuição limite não degenerada $G(z)$ (Fisher, 1928) e escreve-se

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq z\right) = G(z).$$

O Teorema de Fisher–Tippett–Gnedenko estabelece que $G(z)$ pertence a uma das três classes de distribuições de valores extremos: Gumbel, Fréchet ou Weibull (Leadbetter, 1983). Estas podem ser unificadas na Distribuição Generalizada de Valores Extremos (GEV), cuja função de distribuição é dada por (Coles, 2001):

$$G(z) = \exp\left\{-\left[1 + \xi\left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \quad \text{para } 1 + \xi\left(\frac{z - \mu}{\sigma}\right) > 0$$

em que:

- $\mu \in \mathbb{R}$ é o parâmetro de localização,
- $\sigma \in \mathbb{R}^+$ é o parâmetro de escala,
- $\xi \in \mathbb{R}$ é o parâmetro de forma.

O parâmetro ξ determina o tipo de cauda da distribuição:

- $\xi > 0$: Fréchet (cauda pesada),
- $\xi = 0$: Gumbel (cauda exponencial),
- $\xi < 0$: Weibull (cauda limitada).

Este enquadramento é apropriado para a modelação de máximos por blocos (ver secção 2.3), como máximos anuais de concentração de poluentes ou de temperaturas extremas (Ambrósio, 2024).

Em alternativa à abordagem por blocos, a teoria das excedências acima de um limiar elevado conduz à Distribuição Generalizada de Pareto (GPD), com FDA:

$$H(y) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-\frac{1}{\xi}}, \quad \text{para } y > 0, \text{ e } 1 + \xi \frac{y}{\sigma} > 0,$$

onde $\sigma > 0$ é o parâmetro de escala, $\xi \in \mathbb{R}$ é o parâmetro de forma da distribuição e Y é a v.a. representa os excessos acima de um limiar u , ou seja, $Y = X - u | X > u$.

Esta abordagem é frequentemente preferida quando os dados contêm múltiplos eventos extremos por bloco, como picos de poluição em épocas de incêndios (Davison, 1990) (ver secção 2.4).

A escolha entre os modelos GEV e GPD depende da estrutura dos dados disponíveis e da natureza do fenómeno em estudo.

2.3 Parâmetros de interesse

Os modelos, GEV e GPD, representam soluções assintóticas que permitem inferir sobre a ocorrência de eventos raros fora do intervalo observado. Para além dos parâmetros associados a cada um dos modelos, a análise de valores extremos permite derivar outras medidas importantes para a avaliação de risco em contextos ambientais e de qualidade do ar (Coles, 2001; Beirlant *et al.*, 2004) como: a probabilidade de excedência, o nível de retorno, o período de retorno e os quantis extremos.

Probabilidade de excedência

A probabilidade de excedência fornece a frequência esperada de eventos que ultrapassam um limiar crítico v :

$$P(X > v) = \begin{cases} 1 - G(v; \mu, \sigma, \xi), & X \sim \text{GEV}, \\ 1 - H(v; \sigma, \xi), & X \sim \text{GPD}. \end{cases}$$

Esta métrica é útil para estimar a ocorrência de episódios de poluição extrema, permitindo decisões de mitigação e prevenção mais fundamentadas.

Níveis de retorno, períodos de retorno e quantis extremos

Um dos principais objetivos da modelação de valores extremos é a estimação de níveis de retorno, isto é, quantis associados a períodos de retorno T . O nível de retorno z_T corresponde ao valor que, em média, é excedido uma vez a cada T períodos (Coles, 2001):

$$P(X > z_T) = \frac{1}{T}.$$

Equivalentemente, z_T é obtido por inversão da FDA da distribuição GEV ou da GPD:

$$z_T := \begin{cases} G^{-1}\left(1 - \frac{1}{T}; \mu, \sigma, \xi\right), & X \sim \text{GEV}, \\ H^{-1}\left(1 - \frac{1}{T}; \sigma, \xi\right), & X \sim \text{GPD}. \end{cases}$$

Esta medida é essencial em estudos ambientais, permitindo estimar concentrações críticas de poluentes atmosféricos ou intensidades de fenómenos climáticos em diferentes horizontes temporais (Katz, Parlange, & Naveau, 2002).

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

$$T(u) = \begin{cases} \frac{1}{1 - G(u; \mu, \sigma, \xi)}, & X \sim \text{GEV} \\ \frac{1}{1 - H(u; \sigma, \xi)}, & X \sim \text{GPD}. \end{cases}$$

Um quantil extremal, x_p , com p pequeno, normalmente, $p < 1/n$ é tal que

$$P(X > x_p) = p$$

e, portanto, um quantil extremal não é mais do que um nível de retorno, i.e., $x_p = z_{1/p}$.

2.4 Modelo dos máximos anuais

Nesta secção iremos apresentar o modelo dos máximos anuais ou modelo GEV em contexto estacionário e não estacionário. Para cada uma das abordagens

apresentaremos os procedimentos inferências associados à estimação dos parâmetros, os métodos de diagnóstico da qualidade do ajuste realizado e de seleção do melhor modelo.

2.4.1 Modelo GEV estacionário

O modelo dos máximos por blocos é uma das metodologias centrais da TVE e é amplamente utilizado para caracterizar a distribuição de valores extremos de uma sequência de v.a.'s i.i.d. A ideia fundamental é dividir uma sequência temporal $\{X_t\}_{t=1}^T$ em blocos de igual duração (por exemplo, ano) e selecionar apenas o máximo observado em cada bloco (Coles, 2001; Beirlant *et al.*, 2004). Assim, define-se:

$$Z_i = \text{máx}\{X_{(i-1)r+1}, X_{(i-1)r+2}, \dots, X_{ir}\}, \quad i = 1, \dots, n$$

onde r é o tamanho do bloco e n o número total de blocos. Na prática, assume-se que os Z_i seguem aproximadamente uma distribuição Generalizada de Valores Extremos (GEV), com FDA, $G(z)$ (ver pág. 15). Quando se assume que o comportamento estatístico dos máximos não se altera ao longo do tempo, ou seja, não há tendência ou mudanças estruturais, diz-se que o modelo é estacionário. Nesse caso, os parâmetros (μ, σ, ξ) são considerados constantes durante todo o período analisado (Fisher & Tippett, 1928; Gnedenko, 1943; Coles, 2001).

Inferência no modelo GEV estacionário

A inferência no modelo GEV estacionário é essencial na aplicação da TVE, uma vez que permite estimar os parâmetros (μ, σ, ξ) e obter indicadores essenciais para quantificar a probabilidade e a intensidade de eventos extremos. A precisão destas estimativas influencia de forma decisiva as conclusões obtidas em estudos de natureza ambiental, hidrológica ou climatológica (Coles, 2001; Beirlant *et al.*, 2004). A estimação dos parâmetros da distribuição GEV pode ser realizada através de diferentes metodologias, cada uma com vantagens e limitações específicas.

O método da Máxima Verosimilhança (MLE, do inglês "Maximum Likelihood Estimation") é o procedimento mais comum e robusto na inferência paramétrica. Ele procura os valores dos parâmetros que maximizam a probabilidade de a amostra observada

ter ocorrido, expressa pela função de verosimilhança (L). Para uma amostra de tamanho n de uma distribuição GEV, a função de verosimilhança é dada por:

$$L(\mu, \sigma, \xi; z_1, \dots, z_n) = \prod_{i=1}^n \frac{1}{\sigma} \exp \left\{ - \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \left[1 + \xi \left(\frac{z_i - \mu}{\sigma} \right) \right]^{-1/\xi - 1}.$$

As estimativas (μ, σ, ξ) são obtidas numericamente ao maximizar esta função. A grande vantagem do MLE reside nas suas propriedades assintóticas desejáveis. Para grandes amostras, as estimativas convergem para os verdadeiros valores (consistência), têm a menor variância possível (eficiência) e seguem uma distribuição normal, o que facilita a construção de intervalos de confiança e a realização de testes de hipóteses. No entanto, a sua principal limitação reside em cenários com amostras reduzidas ou quando o parâmetro de forma (ξ) se aproxima de $-1/2$, o que pode levar a problemas de estabilidade numérica e a estimativas instáveis (Smith, 1985; Davison & Huser, 2015).

Uma alternativa de grande relevância é o método dos momentos-L, introduzido por Hosking (1990). Este procedimento baseia-se em combinações lineares de estatísticas de ordem e apresenta propriedades de robustez acrescida face à presença de *outliers* e a tamanhos amostrais reduzidos. Devido a estas características, é extensivamente aplicado em áreas como a hidrologia e a climatologia, onde a disponibilidade de dados pode ser limitada e a ocorrência de valores extremos pode enviesar outras técnicas de estimação (Hosking & Wallis, 1997).

Mais recentemente, têm ganho destaque as abordagens Bayesianas, que permitem a incorporação explícita de informação prévia sobre os parâmetros do modelo. Este enquadramento possibilita não apenas a obtenção de estimativas pontuais, mas também a caracterização completa da distribuição a posteriori, fornecendo intervalos de credibilidade e quantificação formal da incerteza associada. Estas metodologias revelam-se particularmente úteis em contextos onde a informação é escassa ou onde se pretende integrar conhecimento proveniente de estudos anteriores, sendo aplicadas com sucesso em problemas de modelação ambiental e avaliação de riscos (Martins & Stedinger, 2000; Cooley, 2009).

A quantificação da incerteza é um passo crucial na análise de valores extremos, complementando a estimação pontual dos parâmetros do modelo GEV. Não é suficiente obter as melhores estimativas para a localização, escala e forma; é igualmente vital compreender a sua fiabilidade e a gama de valores plausíveis. Este processo é fundamental para a tomada de decisões em contextos de risco, como o dimensionamento de infraestruturas ou a avaliação de impactos de eventos climatéricos extremos (Coles, 2001; Beirlant *et al.*, 2004).

No enquadramento da estimação, a incerteza dos parâmetros estimados é expressa através de intervalos de confiança. Estes intervalos baseiam-se nas propriedades assintóticas dos estimadores MLE, que tendem a seguir uma distribuição normal com um desvio padrão que pode ser aproximado pela matriz de informação de Fisher. A matriz de covariância assintótica, que é a inversa da matriz de informação de Fisher, fornece as variâncias individuais das estimativas dos parâmetros. Assim, um intervalo de confiança (IC) de 95% para um parâmetro θ é dado por:

$$IC_{95\%}(\theta) = \left[\hat{\theta} \pm 1.96 \times \sqrt{\text{Var}(\hat{\theta})} \right].$$

onde $\hat{\theta}$ é a estimativa pontual do parâmetro e $\text{Var}(\hat{\theta})$ é a estimativa da variância assintótica.

Quando as aproximações assintóticas não são fiáveis, sobretudo em amostras pequenas, recorre-se a métodos de reamostragem, como o bootstrap (Efron & Tibshirani, 1993).

Como alternativa, a abordagem Bayesiana quantifica a incerteza de forma mais direta e intuitiva através dos intervalos de credibilidade. Estes intervalos são construídos a partir da distribuição a posteriori dos parâmetros, que reflete a atualização das crenças iniciais (distribuição a priori) com a informação dos dados observados. Por exemplo, um intervalo de credibilidade de 95% para o parâmetro ξ é o intervalo que contém 95% da área total da sua distribuição a posteriori. Esta metodologia é particularmente útil quando a amostra é reduzida, pois permite incorporar conhecimento prévio, estabilizando a inferência (Martins & Stedinger, 2000).

Diagnóstico de ajuste

A adequação do modelo GEV aos dados observados é uma etapa crítica que deve ser avaliada de forma rigorosa, utilizando tanto ferramentas gráficas como testes estatísticos formais. Esta validação assegura que as inferências e as estimativas de risco são fiáveis (Beirlant *et al.*, 2004).

Os gráficos Probabilidade-Probabilidade (PP) e Quantil-Quantil (QQ) são ferramentas visuais essenciais para comparar a distribuição empírica dos dados com a distribuição teórica do modelo GEV ajustado. Num ajuste adequado, os pontos do gráfico QQ deverão alinhar-se ao longo da linha de 45 graus, indicando que os quantis observados correspondem de perto aos quantis teóricos. Da mesma forma, no gráfico PP, a proximidade dos pontos à diagonal reflete a compatibilidade entre as probabilidades empíricas e teóricas. Desvios significativos desta linha de referência podem indicar um ajuste insatisfatório, sugerindo a presença de *outliers* ou uma inadequação do modelo GEV aos dados.

Para além da inspeção visual, a validação da qualidade do modelo ajustado é reforçada por testes formais de bondade de ajustamento. Testes como o de Kolmogorov–Smirnov e o de Anderson–Darling fornecem uma medida quantitativa da qualidade do ajuste. A sua estatística de teste avalia a distância entre a FDA empírica e a FDA teórica do modelo GEV. Um valor-p elevado (geralmente > 0.05) sugere que não existem evidências estatísticas para rejeitar a hipótese nula de que os dados seguem a distribuição GEV (Stephens, 1986).

Seleção de modelos

Na família GEV, o caso $\xi = 0$ corresponde à distribuição de Gumbel e funciona como fronteira entre as caudas limitadas (Weibull) e as caudas pesadas (Fréchet). Assim, o teste mais simples é: $H_0: \xi = 0$ (Gumbel) versus $H_1: \xi \neq 0$ ($\xi > 0$: Fréchet; $\xi < 0$: Weibull).

O Teste Deviance é um teste de hipóteses utilizado para testar as hipóteses acima descritas ao comparar as verosimilhanças entre dois modelos e desta forma inferir sobre o modelo que origina uma melhor qualidade do ajuste. Se \mathcal{M}_0 e \mathcal{M}_1 são dois modelos encaixados ou aninhados, isto é, se $\mathcal{M}_0 \subset \mathcal{M}_1$, a função Deviance (Castillo *et al.*, 2005) é definida como,

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

onde, $\ell_0(\mathcal{M}_0) = \log L(\mathcal{M}_0)$ e $\ell_1(\mathcal{M}_1) = \log L(\mathcal{M}_1)$, são, respetivamente, os valores da função de log-verosimilhança para os modelos ajustados, \mathcal{M}_0 e \mathcal{M}_1 . Definindo o vetor de parâmetros, $\theta := (\mu, \sigma, \xi)$ então $\hat{\theta} := (\hat{\mu}, \hat{\sigma}, \hat{\xi})$. Para a distribuição GEV vem,

$$\mathcal{D} = 2\left\{\max \ell\left(\hat{\theta}_{G_\xi} | Z_1, \dots, Z_n\right) - \max \ell\left(\hat{\theta}_{G_0} | Z_1, \dots, Z_n\right)\right\},$$

onde $\ell\left(\hat{\theta}_{G_\xi} | Z_1, \dots, Z_n\right)$ e $\ell\left(\hat{\theta}_{G_0} | Z_1, \dots, Z_n\right)$ representam os valores das funções de log-verosimilhança dos modelos GEV e Gumbel, respetivamente. Sob a validade de H_0 ,

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

Em 1984, (Hosking, 1984) sugeriu uma modificação para melhorar a aproximação à distribuição Qui-quadrado (Castillo *et al.*, 2005):

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

O teste assintótico de nível α rejeita H_0 , ou seja, o Modelo Gumbel se $\mathcal{D}^* \geq \chi_1^2(1 - \alpha)$.

Estimação de parâmetros de interesse

A estimação dos parâmetros de interesse apresentadas na Secção 2.3 surge naturalmente substituindo os parâmetros em estudo pelas suas estimativas pontuais.

2.4.2 Modelo GEV não estacionário

A abordagem não estacionária na modelação de valores extremos permite incorporar variações temporais ou dependência de covariáveis nos parâmetros da distribuição GEV(μ, σ, ξ). Esta generalização é essencial quando se observam tendências, sazonalidades ou alterações estruturais nos máximos anuais, como é frequentemente o caso em fenómenos ambientais sujeitos a mudanças climáticas ou impactos de incêndios florestais (Coles, 2001; Katz, Parlange, & Naveau, 2002)

No modelo GEV não estacionário, um ou mais parâmetros podem ser expressos como funções de tempo t ou de covariáveis explicativas $X_{t,j}$ (que estão indexadas no tempo). Na presença de mais do que uma covariável, estes modelos são usualmente designados de modelos de regressão em extremos. Na formulação adotada neste trabalho assumimos que os parâmetros de localização e de escala são funções lineares das covariáveis e que o parâmetro de forma se mantém constante (Henriques-Rodrigues, Duarte, & Rodrigues, 2025). Deste modo, assumimos que

$$Z_t \sim GEV(\mu_t, \sigma_t, \xi)$$

com

$$\begin{aligned} \mu_t &= \mu_0 + \sum_{j=1}^p \beta_j X_{t,j}, \\ \sigma_t &= \exp\left(\sigma_0 + \sum_{j=1}^p \gamma_j X_{t,j}\right), \end{aligned}$$

onde:

- μ_0 é o intercepto (ordenada na origem);
- β_j e γ_j são os coeficientes associados às covariáveis $X_{t,j}$.

Esta parametrização permite modelar, por exemplo, aumentos graduais na concentração de PM₁₀ durante épocas de incêndios ou alterações na variabilidade dos máximos ao longo do tempo.

Inferência no modelo GEV não estacionário

O modelo GEV não estacionário mantém a formulação funcional da distribuição GEV, mas com parâmetros dependentes do tempo e/ou de covariáveis. A FDA de Z_t é,

$$G(z_t; \mu_t, \sigma_t, \xi_t) = \exp \left\{ - \left[1 + \frac{\xi_t(z_t - \mu_t)}{\sigma_t} \right]^{-1/\xi_t} \right\}, \quad 1 + \frac{\xi_t(z_t - \mu_t)}{\sigma_t} > 0.$$

A estimação dos parâmetros é geralmente realizada por MLE, adaptando a função de verosimilhança para refletir a dependência temporal e/ou das covariáveis:

$$\mathcal{L}(\theta) = \prod_{t=1}^n g(z_t; \mu_t, \sigma_t, \xi_t),$$

onde $g(z_t; \mu_t, \sigma_t, \xi_t)$ é a função densidade da GEV no instante t e θ representa o vetor de todos os parâmetros $(\mu_0, \sigma_0, \xi, \beta_1, \dots, \beta_p, \xi_1, \dots, \xi_p)$.

Outras metodologias, como máxima verosimilhança penalizada ou Bayesiana, podem ser utilizadas para garantir estabilidade na estimação, especialmente quando o número de observações extremas é limitado (Coles, 2001; Davison & Huser, 2015).

Na análise da qualidade do ar durante a época de incêndios em Portugal, a modelação não estacionária permite detetar tendências temporais na intensidade dos picos de PM_{10} , avaliar a influência de covariáveis meteorológicas ou de emissão e ainda melhorar a previsão de eventos extremos futuros e o planeamento de medidas de mitigação.

Diagnóstico de ajuste

Após ajustar o modelo GEV não estacionário e obter estimativas pontuais e intervalares para os parâmetros, é necessário verificar o ajustamento do modelo. A heterogeneidade das suposições distribucionais entre observações obriga a pequenas adaptações dos procedimentos gráficos usuais.

O modelo $Z_t \sim \text{GEV}(\mu_t, \sigma_t, \xi)$ equivale a ter uma distribuição GEV diferente para cada t (e para valor das covariáveis em estudo). Para aplicar os diagnósticos padrão, é

necessário padronizar os dados de forma a tratar Y_t como aproximadamente i.i.d. ao longo do tempo (e das covariáveis). Usualmente, transforma-se o conjunto de máximos não estacionários para uma distribuição comum Gumbel, cuja FDA é $F(z) = e^{-e^{-z}}$.

A transformação desejada obtém-se igualando

$$F(z_t) = GEV(x_t; \hat{\mu}_t, \hat{\sigma}_t, \hat{\xi})$$

e resolvendo a igualdade anterior em ordem a y_t obtém-se,

$$z_t = -\ln(-\ln(GEV(x_t; \hat{\mu}_t, \hat{\sigma}_t, \hat{\xi}))).$$

Depois de convertidos para uma distribuição comum, aplicam-se os diagnósticos gráficos habituais. Em particular, comparam-se as probabilidades e os quantis empíricos de $\{z_t\}$ com os respetivos valores teóricos da distribuição de Gumbel (gráficos PP e QQ de quantis residuais).

Seleção de modelos

No processo de modelação estatística, é comum ajustar mais do que um modelo aos dados, com o objetivo de identificar aquele que melhor representa o fenómeno em estudo.

Depois de efetuado o diagnóstico dos modelos ajustados, e no caso de modelos aninhados ou encaixados, podemos também realizar um teste de hipóteses para a seleção do melhor modelo, baseado na estatística de teste *Deviance*.

Considerando dois modelos encaixados, $\mathcal{M}_0 \subset \mathcal{M}_1$, o Teste de Razão de Verossimilhanças (TRV) permite testar as hipóteses:

$$H_0: \theta = \theta_0 \quad (\mathcal{M}_0) \quad \text{vs} \quad H_1: \theta = \theta_1 \quad (\mathcal{M}_1).$$

A estatística de teste é baseada na função Deviance

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

onde, $\ell_0(\mathcal{M}_0) = \log L(\mathcal{M}_0)$ e $\ell_1(\mathcal{M}_1) = \log L(\mathcal{M}_1)$, são, respetivamente, os valores da função de log-verosimilhança para os modelos ajustados, \mathcal{M}_0 e \mathcal{M}_1 . Sob a validade de H_0 ,

$$\mathcal{D} \xrightarrow[n \rightarrow \infty]{d} Q_1 \sim \chi_k^2,$$

em que k é a diferença de dimensionalidade de \mathcal{M}_1 e \mathcal{M}_0 . O teste assintótico de nível α rejeita H_0 se $\mathcal{D} > \chi_k^2(1 - \alpha)$, onde $\chi_k^2(1 - \alpha)$ é o quantil $(1 - \alpha)$ 100% da distribuição χ_k^2 .

Quando os modelos da hipótese nula e da alternativa não são aninhados, o TRV deixa de ser apropriado. Nessa situação, a comparação faz-se com critérios de informação, em particular o AIC (Akaike Information Criterion) e BIC (Bayesian Information Criterion):

$$\text{AIC} = -2 \max l_{\mathcal{M}}(\hat{\theta}) + 2p$$

$$\text{BIC} = -2 \max l_{\mathcal{M}}(\hat{\theta}) + p \ln n,$$

Sendo $l_{\mathcal{M}}(\hat{\theta})$ a log-verosimilhança maximizada do modelo \mathcal{M} , p o número de parâmetros do modelo e n , a dimensão da amostra. Valores mais baixos de AIC ou BIC indicam melhor compromisso entre ajustamento e complexidade. Para além disso, o AIC e o BIC equilibram qualidade do ajuste com parcimónia, evitando a sobreparametrização (Akaike, 1973; Schwarz, 1978; Burnham & Anderson, 2002). Esta abordagem garante não só que o modelo escolhido se adapta bem aos dados, mas também que mantém interpretabilidade estatística e aplicabilidade prática, fatores essenciais em contextos de previsão de fenómenos extremos e avaliação de risco ambiental (Beirlant *et al.*, 2004; Coles, 2001). Estes critérios podem também ser usados quando os modelos são aninhados e em contexto estacionário.

Estimação de parâmetros de interesse

No caso não estacionário a estimação dos outros parâmetros de interesse é mais complexa dada a dependência do tempo e das covariáveis. Desta forma, é necessário

considerar valores apropriados para as covariáveis e para o tempo de forma a obter estimativas dos parâmetros do modelo e com base nestas obter as estimativas das quantidades de interesse.

2.5 Método de excedências

A abordagem dos máximos por blocos, desenvolvida na secção 2.4, constitui uma das principais metodologias da TVE. Contudo, pode ser pouco eficiente quando cada bloco contém muitas observações, porque utiliza apenas o máximo de cada bloco, e não considera informação sobre outros valores, resultando numa perda substancial de informação. Como alternativa, surge a abordagem das excedências acima de um limiar (*peaks-over-threshold*, POT), que aproveita todos os valores que ultrapassam um determinado nível elevado, permitindo uma utilização mais eficiente da amostra (Davison & Smith, 1990; Coles, 2001).

O resultado teórico fundamental que sustenta esta metodologia é o Teorema de Pickands–Balkema–de Haan, o qual estabelece que, sob condições gerais, a distribuição dos excessos acima de um limiar elevado pode ser aproximada pela Distribuição Generalizada de Pareto (GPD) (Pickands, 1975; Balkema & de Haan, 1974). Esta distribuição é caracterizada por dois parâmetros principais: o parâmetro de escala (σ) que controla a dispersão das excedências, e o parâmetro de forma (ξ) que define o comportamento da cauda.

A abordagem POT apresenta diversas vantagens em relação ao modelo dos máximos anuais. Em particular, permite uma maior eficiência estatística ao considerar múltiplos eventos extremos por período, facto especialmente relevante em fenómenos ambientais e climatológicos, como episódios de poluição atmosférica, cheias repentinas ou ondas de calor (Davison & Huser, 2015). Contudo, a sua aplicação requer a escolha criteriosa de um limiar adequado: demasiado baixo, introduz viés por incluir observações não extremas; demasiado alto, reduz drasticamente o tamanho da amostra disponível (Coles, 2001).

Deste modo, o modelo das excedências constitui uma ferramenta fundamental da TVE, complementando a análise baseada nos máximos anuais e oferecendo uma caracterização mais detalhada e flexível da cauda da distribuição de fenômenos extremos.

2.5.1 Modelo GPD estacionário

O modelo das excedências acima um limiar elevado baseia-se na aproximação assintótica proporcionada pela GPD. O Teorema de Pickands–Balkema–de Haan afirma que, para uma larga classe de distribuições, a distribuição condicional dos excessos acima de um limiar suficientemente alto converge para a GPD (Balkema & de Haan, 1974; Pickands, 1975).

Seja $Y = X - u \mid X > u$ a v.a. que representa os excessos (ou as excedências) acima de um limiar u . A FDA da GPD é dada por:

$$H(y; \sigma, \xi) = 1 - \left(1 + \xi \frac{y}{\sigma}\right)^{-1/\xi}, \quad y > 0, \quad 1 + \xi \frac{y}{\sigma} > 0.$$

No caso estacionário, assume-se que os parâmetros (σ, ξ) permanecem constantes ao longo do tempo, ou seja, não há dependência de covariáveis nem variação temporal. Este enquadramento fornece uma representação parsimoniosa e robusta dos valores extremos, sendo particularmente adequado quando se dispõe de dados homogêneos e sem tendências estruturais (Coles, 2001).

O modelo GPD estacionário fornece uma descrição essencial da cauda da distribuição, permitindo inferir sobre probabilidades de excedência e níveis de retorno com base numa formulação assintótica rigorosa.

Inferência no modelo GPD estacionário

A estimação dos parâmetros da GPD em regime estacionário constitui uma etapa central da abordagem POT. Tal como no caso da GEV, diferentes metodologias têm sido desenvolvidas para inferir os parâmetros (σ, ξ) cada uma apresentando vantagens e limitações específicas em termos de eficiência estatística, robustez e aplicabilidade prática.

O método mais utilizado é o MLE, devido às suas propriedades assintóticas desejáveis de consistência e eficiência (Smith, 1987). A função de log-verosimilhança para uma amostra de m excedências (y_1, y_2, \dots, y_m) é dada por:

$$l(\xi, \sigma) = -m \ln(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \ln\left(1 + \xi \frac{y_i}{\sigma}\right), \quad 1 + \xi \frac{y_i}{\sigma} > 0.$$

Embora a MLE seja assintoticamente eficiente, pode apresentar instabilidade numérica em amostras pequenas ou quando ξ assume valores próximos de zero, levando a estimativas pouco fiáveis (Davison & Smith, 1990).

Independentemente do método de estimação escolhido, a qualidade da inferência depende fortemente da escolha do limiar u . Assim, a seleção do limiar é geralmente acompanhada por análises gráficas, como o *mean residual life plot* ou pela avaliação da estabilidade dos parâmetros ao variar o limiar (Davison & Smith, 1990; Coles, 2001). Uma descrição dos métodos de escolha do limiar u pode ser consultada em (Ambrósio, 2024). Valores relativamente constantes dos parâmetros, à medida que o limiar aumenta, indicam que a teoria assintótica é válida e, portanto, que o limiar escolhido é adequado para a modelação.

Em síntese, a inferência no modelo GPD estacionário exige um equilíbrio entre eficiência estatística, robustez e aplicabilidade prática. A escolha do método de estimação deve considerar o tamanho da amostra, a qualidade dos dados e os objetivos do estudo (Coles, 2001; Beirlant *et al.*, 2004).

Diagnóstico de ajuste

A verificação da adequação do modelo GPD estacionário é essencial para garantir que as inferências sobre probabilidades de excedência e níveis de retorno sejam confiáveis. Para isso, recomenda-se uma abordagem multifacetada, combinando ferramentas gráficas, testes formais e análise de estabilidade do limiar (Coles, 2001; Hosking & Wallis, 1997).

O gráfico QQ compara os quantis empíricos dos excedentes $Y_i = X_i - u$ com os quantis teóricos da GPD ajustada. Pontos próximos à linha de 45° indicam bom ajuste; desvios sistemáticos sugerem inadequação do modelo ou a necessidade de ajustar o limiar

u (Coles, 2001). O gráfico PP compara as probabilidades acumuladas empíricas com as teóricas. A proximidade dos pontos à diagonal reflete a compatibilidade entre o modelo e os dados (Brabson & Palutikof, 2000).

Podem também ser aplicados os testes de ajustamento de Kolmogorov-Smirnov ou de Anderson-Darling, referidos na secção 2.4.1.

Seleção de modelos

A seleção do modelo mais adequado constitui uma etapa essencial na análise de valores extremos, assegurando que as inferências e previsões são robustas e fiáveis. No contexto do modelo GPD estacionário as hipóteses de interesse são:

$$H_0: \xi = 0 \quad (\textit{Exponencial}) \quad \textit{vs} \quad H_1: \xi \neq 0 \quad (\textit{Pareto ou Beta})$$

Sendo $\theta := (\sigma, \xi)$ então $\hat{\theta} := (\hat{\sigma}, \hat{\xi})$, a estatística Deviance, descrita na secção 2.4.1, é dada por

$$\mathcal{D} = 2 \left\{ \max_{\hat{\theta}_{H_1}} \ell(\hat{\theta}_{H_1} | Y_1, \dots, Y_m) - \max_{\hat{\theta}_{H_0}} \ell(\hat{\theta}_{H_0} | Y_1, \dots, Y_m) \right\}.$$

Sob a validade de H_0 , $\mathcal{D} \xrightarrow[n \rightarrow \infty]{d} \chi_1^2$. Reiss and Thomas (2007) propuseram a correção abaixo de forma a melhorar a aproximação da estatística de teste à distribuição χ_1^2 .

$$\mathcal{D}^* = \frac{\mathcal{D}}{1 + \frac{4}{m}} \xrightarrow[n \rightarrow \infty]{d} \chi_1^2.$$

Estimação de parâmetros de interesse

Para além da estimação direta dos parâmetros (σ, ξ) , o modelo GPD estacionário permite derivar medidas adicionais de grande relevância para análise e gestão do risco de eventos extremos. Estas medidas constituem ferramentas fundamentais na avaliação da probabilidade de ocorrência e da magnitude de excedências raras, sendo amplamente aplicadas em áreas como hidrologia, climatologia e ciências ambientais (Coles, 2001; Davison & Smith, 1990).

Uma vez ajustado o modelo GPD às excedências, $Y = X - u | X > u$, o interesse reside em estimar os parâmetros de interesse da variável original, X . Desta forma é necessário proceder à estimação da taxa de excedência, λ_u . Dada uma série de n observações, em que N_u representam o número de excedências, a taxa é estimada por:

$$\hat{\lambda}_u = \frac{N_u}{n}.$$

A probabilidade de excedência acima de um nível elevado, u , é então dada por

$$P(X > x) = \hat{\lambda}_u \left(1 + \frac{\xi(x - u)}{\sigma} \right)^{-\frac{1}{\xi}}.$$

Os quantis extremos, x_p , são obtidos a partir de

$$x_p = u + \frac{\sigma}{\xi} \left(\left(\frac{p}{\hat{\lambda}_u} \right)^{-\xi} - 1 \right).$$

Utilizando as relações apresentadas na secção 2.3 obtêm-se facilmente os níveis e o período de retorno.

Em síntese, as quantidades acima descritas constituem extensões diretas da inferência sobre os parâmetros da GPD. A sua correta estimação é essencial para a avaliação quantitativa do risco, permitindo traduzir resultados estatísticos em medidas operacionais úteis para a gestão ambiental e de riscos extremos (Coles, 2001; Davison & Smith, 1990; Embrechts *et al.*, 1997).

2.5.2 Modelo GPD não estacionário

A suposição de estacionariedade na GPD, isto é, de que os parâmetros permanecem constantes ao longo do tempo, pode ser demasiado restritiva em contextos onde há evidência de variabilidade associada a fatores ambientais, climáticos ou sazonais. Neste enquadramento, o modelo GPD não estacionário surge como uma extensão natural, permitindo que os parâmetros da distribuição sejam modelados como funções

de covariáveis explicativas (Davison & Smith, 1990; Chavez-Demoulin & Davison, 2012).

Seja $Y_t = X_t - u \mid X_t > u$ a variável excedência acima de um limiar u . No caso não estacionário, a função de densidade de probabilidade de $Y_t \sim \text{GPD}(\sigma_t, \xi_t)$ é dada por:

$$h(y_t; \sigma_t, \xi_t) = \frac{1}{\sigma_t} \left(1 + \xi_t \frac{y_t}{\sigma_t}\right)^{-1/\xi_t - 1}, \quad y_t > 0, 1 + \xi_t \frac{y_t}{\sigma_t} > 0,$$

onde $\sigma_t > 0$ e ξ_t dependem de covariáveis que variam em função do tempo, t .

A dependência em relação a covariáveis é frequentemente especificada através de funções lineares generalizadas. Na nossa abordagem e à semelhança do efetuado para o modelo GEV não estacionário, assumimos que o parâmetro de forma é constante e que apenas o parâmetro de escala varia em função das covariáveis. Assim, o modelo $\text{GPD}(\sigma_t, \xi)$ caracteriza-se por:

$$\log(\sigma_t) = \beta_0 + \sum_{j=1}^p \beta_j X_{t,j}$$

onde $X_{t,j}$ representam covariáveis (por exemplo, temperatura, humidade, índices de seca ou indicadores sazonais), e $\beta_j, j=0, \dots, p$, são coeficientes a estimar.

Esta formulação assegura a positividade de σ_t e permite incorporar explicitamente a influência de variáveis externas no comportamento das excedências.

Inferência no modelo GPD não estacionário

A estimação dos parâmetros é realizada, em geral, por MLE, com a função de log-verosimilhança definida como:

$$l(\beta_0, \beta_1, \dots, \beta_p, \xi) = - \sum_{t=1}^m \ln(\sigma_t) - \sum_{t=1}^m \left(1 + \frac{1}{\xi}\right) \ln\left(1 + \xi \frac{y_t}{\sigma_t}\right)$$

sujeito à condição $1 + \xi \frac{y_t}{\sigma_t} > 0$. Métodos Bayesianos também têm sido aplicados com sucesso, oferecendo maior flexibilidade para lidar com incerteza e modelos complexos (Eastoe & Tawn, 2009).

O modelo GP não estacionário é especialmente relevante em situações onde os extremos estão sujeitos a variabilidade sazonal ou a mudanças de longo prazo. Exemplos incluem:

- Análise de precipitações extremas condicionadas pela estação do ano;
- modelação de excedências de poluição atmosférica em função de indicadores meteorológicos;
- avaliação de riscos climáticos sob cenários de mudança climática.

Ao permitir que a intensidade e a forma da distribuição se adaptem dinamicamente a covariáveis, este modelo fornece uma representação mais realista do risco extremo, sobretudo em sistemas ambientais complexos e não estacionários (Coles, 2001; Chavez-Demoulin & Davison, 2012).

Diagnóstico de ajuste

O diagnóstico do modelo GP não estacionário é mais exigente do que no caso estacionário, uma vez que a adequação deve ser avaliada não apenas em termos globais, mas também relativamente à influência das covariáveis e da estrutura temporal.

Procedimentos de validação cruzada podem ser usados para avaliar o desempenho preditivo do modelo em subconjuntos dos dados. Esta abordagem é útil para verificar se a inclusão de covariáveis melhora efetivamente a capacidade de previsão de excedências extremas (Cooley, 2009). Esta metodologia pode também ser aplicada no modelo GEV não estacionário.

O modelo $Y_t \sim \text{GPD}(\sigma_t, \xi)$ equivale a ter uma GPD diferente para cada t (e para valor das covariáveis em estudo). Para aplicar os diagnósticos padrão, é necessário padronizar os dados de forma a tratar Y_t como aproximadamente i.i.d. ao longo do tempo (e das covariáveis), de modo análogo ao efetuado no modelo GEV não estacionário.

Neste caso, transforma-se o conjunto de excedências não estacionárias para uma distribuição comum de Exponencial, cuja função de distribuição é $F(y) = 1 - e^{-y}$.

A transformação desejada obtém-se igualando

$$F(y_t) = GPD(x_t; \hat{\sigma}_t, \hat{\xi})$$

e resolvendo a igualdade anterior em ordem a y_t , obtém-se

$$y_t = -\ln(1 - GPD(x_t; \hat{\sigma}_t, \hat{\xi})).$$

Depois de convertidos para uma distribuição comum, aplicam-se os diagnósticos gráficos habituais.

Seleção de modelos

Tal como discutido na secção 2.4.2, quando os modelos são encaixados, isto é, quando um modelo é um caso particular do outro, a comparação pode ser feita através de um TRV, permitindo avaliar se a complexidade adicional resulta numa melhoria significativa do ajuste (Coles, 2001; Castillo *et al.*, 2005). Para modelos não encaixados, recorrem-se a critérios de informação, como o AIC e o BIC.

Estimação de parâmetros de interesse

As expressões para a estimação de quantidades de interesse mantêm as equações apresentadas no modelo GPD estacionário, escolhendo-se valores de referência para as covariáveis.

ANÁLISE E DISCUSSÃO DOS RESULTADOS

Este capítulo destina-se à análise e discussão aprofundada dos resultados obtidos na investigação da qualidade do ar, especificamente nas regiões Norte e Centro, durante a época de incêndios. O foco recai sobre o período de junho a outubro, entre 2011 e 2020. Esta década foi particularmente desafiadora devido à ocorrência de significativos incêndios florestais que impactaram diretamente na qualidade do ar ao contribuir para o aumento de concentrações de poluentes como o PM_{10} . Para uma caracterização robusta dos eventos extremos de poluição, foram aplicados e contrastados múltiplos métodos da teoria de valores extremos: o modelo GEV aos máximos mensais; o modelo GEV não estacionário, uma abordagem que permite identificar e quantificar tendências ou sazonalidades nos extremos, como a variação da intensidade dos incêndios ao longo do tempo; o método POT, em contexto estacionário e não estacionário, que permite analisar todos os eventos que superam um patamar pré-definido, oferecendo uma visão mais completa da cauda da distribuição. A análise dos modelos não estacionários será enriquecida pela consideração de covariáveis como temperatura, velocidade do vento e área ardida, para além do tempo, fornecendo um entendimento multifacetado da dinâmica da qualidade do ar em cenários de extremos.

3.1 Método dos máximos anuais

A primeira abordagem empregada para investigar os eventos extremos de concentração de PM_{10} consiste na análise dos máximos mensais. A escolha dos máximos mensais, ao invés dos máximos anuais, justifica-se pela elevada frequência de eventos

extremos de poluição relacionados a incêndios florestais dentro de um mesmo ano e pela necessidade de capturar variações sazonais intrínsecas ao fenômeno. Para cada mês (junho, julho, agosto, setembro e outubro) e para cada ano do período de estudo (2011-2020), extraiu-se o valor máximo diário de PM_{10} . Da série de máximos diários de PM_{10} obteve-se a série de máximos mensais, para cada um dos meses e para cada ano em estudo.

3.1.1 Estatística descritiva das variáveis

A caracterização inicial dos dados é fundamental para compreender a natureza e a variabilidade das concentrações de PM_{10} nos meses de estudo. Para cada um dos meses (junho a outubro), a série de máximos mensais de PM_{10} (2011-2020) foi submetida a uma análise exploratória detalhada.

Esta amostra é constituída por $n = 50$ observações de máximos mensais de todas as variáveis da região Norte/Centro, nos meses de junho-julho-agosto-setembro-outubro, no período de 2011 a 2020.

Tabela 1: Estatística descritiva das variáveis em estudo

Estatística	PM10 ($\mu\text{g}/\text{m}^3$)	Temperatura ($^{\circ}\text{C}$)	Velocidade do vento (m/s)	Área ardida (ha)
Média	41,94	24,74	4,73	1560,32
Mediana	34,44	24,98	4,58	889,64
Desvio Padrão	24,78	3,00	0,93	1928,65
Mínimo	20,11	17,10	3,18	17,19
Máximo	169,80	30,30	7,89	7910,02
Assimetria	3,07	-0,62	1,05	1,79
Curtose	12,15	-0,20	1,36	2,59

Durante o período analisado, os níveis máximos de PM_{10} apresentaram uma média de $41,94 \mu\text{g}/\text{m}^3$. Em alguns momentos, as concentrações chegaram a valores tão baixos quanto $20,11 \mu\text{g}/\text{m}^3$, enquanto em outros atingiram picos próximos de $169,80 \mu\text{g}/\text{m}^3$. Essa distribuição mostrou-se assimétrica à direita, evidenciando que eventos de alta concentração foram mais frequentes do que eventos muito abaixo da média. Além disso, a presença de uma curtose leptocúrtica indica que não apenas houve picos elevados, mas que eles ocorreram com maior frequência do que se esperaria numa situação estatisticamente “normal”, refletindo dias pontuais de poluição intensa.

A temperatura máxima, por sua vez, oscilou entre 17,01°C e 30,30°C, com valores médios em torno de 24,98°C, sua distribuição apresentou assimetria à direita, sugerindo que houve mais dias relativamente frios do que extremamente quentes. A curtose platicúrtica reforça a ideia de estabilidade térmica, com menos ocorrências de extremos e predominância de dias com temperaturas máximas moderadas, sem grandes desvios.

No caso da velocidade máxima do vento, observou-se um padrão intermediário: a maioria dos dias apresentou ventos moderados, entre 3,18 m/s e 7,89 m/s, com mediana de 4,58 m/s. A assimetria positiva indica que, ocasionalmente, ocorreram dias com ventos máximos mais fortes, mas sem frequência suficiente para gerar uma cauda muito alongada. A curtose mesocúrtica confirma que a distribuição foi relativamente equilibrada, sem excesso nem escassez marcante de eventos extremos.

Por fim, a análise da área ardida revelou que, embora a média tenha sido de 1560,32 hectares, a variação foi expressiva, indo de incêndios muito pequenos, com apenas 17,19 hectares, até ocorrências mais graves, que chegaram a destruir 7910 hectares, a assimetria positiva demonstra que, em alguns casos, as áreas queimadas foram significativamente maiores do que a média. No entanto, a curtose platicúrtica indica que grandes incêndios não foram tão frequentes quanto se poderia esperar em uma distribuição com alta concentração de eventos extremos, revelando uma dispersão mais uniforme no tamanho das áreas afetadas.

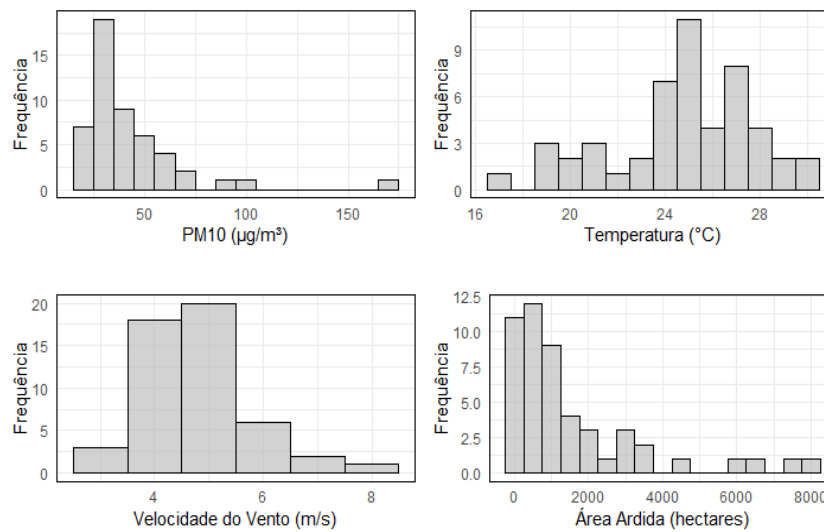


Figura 2: Histograma da concentração PM_{10} ($\mu\text{g}/\text{m}^3$), temperatura do ar ($^{\circ}\text{C}$), velocidade do vento (m/s), e área ardida (ha). Dados recolhidos nas regiões Norte e Centro de Portugal, de 2011 a 2020.

A análise da Figura 2 indica que, na maior parte do tempo, a qualidade do ar, medida pela concentração de PM_{10} , mantém-se dentro de padrões aceitáveis (abaixo de $50 \mu\text{g}/\text{m}^3$). No entanto, observa-se que essa estabilidade é interrompida por episódios de aumento abrupto nas concentrações, possivelmente associados a incêndios de grande dimensão. Esses eventos libertam grandes quantidades de partículas em suspensão, degradando rapidamente a qualidade do ar. As variáveis meteorológicas ajudam a contextualizar esses padrões. As temperaturas elevadas, típicas de um clima quente, associadas a ventos predominantemente fracos, criam condições propícias para a propagação de incêndios e a acumulação de poluentes na atmosfera. A assimetria acentuada nas distribuições de PM_{10} e área ardida sugere que ambas são influenciadas por valores extremos, indicando a presença de poucos eventos de grande magnitude, mas com forte impacto ambiental.

As séries temporais, apresentadas na Figura 3, tornam evidente a relação entre o comportamento das variáveis: picos de PM_{10} tendem a coincidir com dias de grandes áreas queimadas, sugerindo um vínculo direto entre a ocorrência de incêndios florestais e a deterioração da qualidade do ar. Essa associação reforça a importância de integrar informações meteorológicas e ambientais na previsão e mitigação de impactos causados por incêndios.

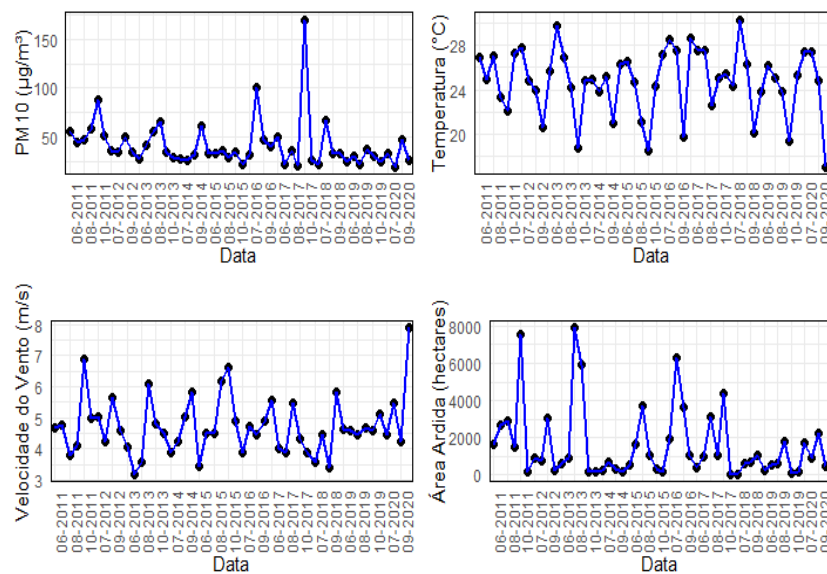


Figura 3: Séries Temporais de PM_{10} ($\mu\text{g}/\text{m}^3$), Temperatura ($^{\circ}\text{C}$), Vento (m/s) e Área ardida (ha) nas Regiões Norte e Centro de Portugal (2011-2020)

A análise da Figura 4, evidencia que as concentrações de PM_{10} variam ao longo do tempo, mantendo-se, na maior parte dos dias, abaixo de $50 \mu\text{g}/\text{m}^3$, mas apresentando picos ocasionais que ultrapassam os $100 \mu\text{g}/\text{m}^3$. Entre esses eventos, destaca-se outubro de 2017, quando foi registado um valor excecionalmente elevado, caracterizando um episódio de poluição extrema. Outros picos menores também são observados em diferentes períodos. A temperatura apresenta uma variação relativamente estreita, entre aproximadamente 18°C e 28°C , seguindo um padrão sazonal bem definido, com máximos durante os meses de verão (junho a setembro) e mínimos no inverno.

A velocidade do vento, por sua vez, mostra flutuações irregulares ao longo do tempo, sem um padrão sazonal claro, permanecendo a maior parte do tempo abaixo de $6 \text{ m}/\text{s}$, com episódios pontuais de ventos mais fortes. No que se refere à área ardida, a maioria dos valores registados é baixa, mas ocorrem picos significativos que chegam a atingir cerca de 8.000 hectares, revelando incêndios de grande escala. Um desses eventos extremos, também em outubro de 2017, coincide com o pico mais alto de PM_{10} , reforçando a associação entre incêndios florestais e deterioração da qualidade do ar. Assim, os dados indicam que os picos de concentração de partículas estão frequentemente ligados a grandes áreas queimadas, enquanto a temperatura mantém o comportamento sazonal esperado e a velocidade do vento apresenta maior irregularidade.

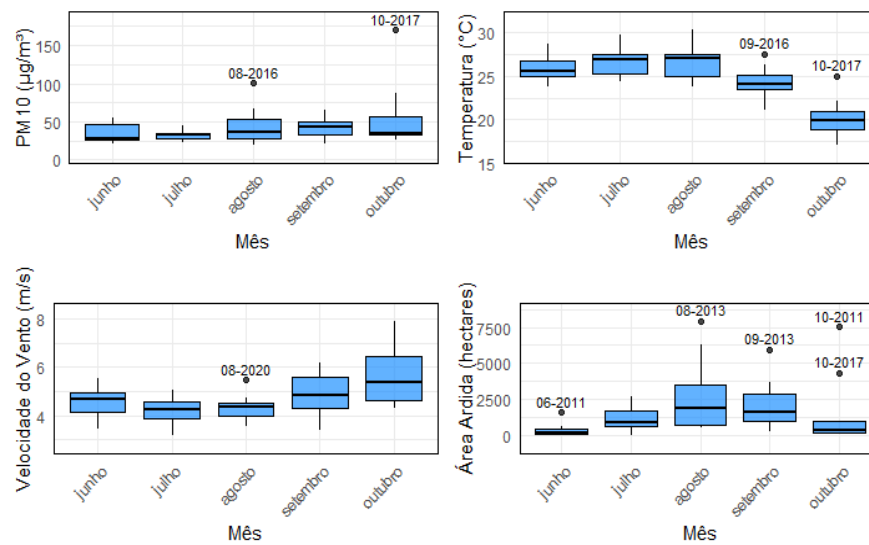


Figura 4: Distribuição mensal de PM₁₀, temperatura, vento e área ardida (2011-2020). Boxplots mostram a mediana, quartis e outliers. Datas indicam outliers.

A análise mensal mostra que a mediana das concentrações de PM₁₀ apresenta uma ligeira tendência de aumento ao longo do ano, atingindo os valores mais elevados nos meses de setembro e outubro, período em que também se observa maior variabilidade, sugerindo uma instabilidade crescente na qualidade do ar. Eventos extremos são evidenciados por outliers significativos, como os registados em agosto de 2016 e outubro de 2017, que correspondem a episódios de poluição intensa. Esse comportamento indica que a qualidade do ar tende a deteriorar-se nos meses finais do verão e início do outono. A temperatura, por sua vez, exibe um padrão sazonal claro, com valores mais altos entre junho e agosto e mais baixos em outubro, mantendo variabilidade relativamente baixa ao longo do ano, exceto por um outlier em outubro de 2017, que aponta para uma temperatura atipicamente baixa nesse mês. Já a velocidade do vento mostra uma tendência de aumento gradual ao longo dos meses, alcançando valores mais altos em setembro e outubro, acompanhada por maior variabilidade nessa época; destaca-se ainda um outlier em agosto de 2020, associado a um episódio de vento excepcionalmente forte. Quanto à área ardida, observa-se grande variabilidade durante todo o ano, embora a mediana permaneça baixa, com picos ocasionais que revelam incêndios de grande escala, como os registados em junho de 2011, agosto e setembro de 2013 e em outubro de 2011 e 2017. Esses eventos demonstram que, embora a maioria dos

incêndios tenha pequena extensão, ocorrem, ocasionalmente, episódios capazes de destruir áreas significativamente amplas.

A Figura 5, representa o gráfico da correlação entre quatro variáveis: o valor máximo de PM_{10} , a temperatura máxima, a velocidade máxima do vento e a área total máxima ardida. Observa-se uma correlação positiva moderada entre o máximo de PM_{10} e a área ardida máxima, o que sugere que, quanto maior a extensão queimada, maiores tendem a ser as concentrações de partículas em suspensão na atmosfera. Essa relação é coerente com o impacto dos incêndios florestais na qualidade do ar, já que a queima de biomassa libera grandes quantidades de poluentes. Por outro lado, há uma correlação negativa moderada entre a temperatura máxima e a velocidade máxima do vento, indicando que períodos mais quentes costumam estar associados a ventos mais fracos, enquanto ventos mais intensos tendem a ocorrer em dias menos quentes. As demais correlações observadas são fracas ou próximas de zero, revelando ausência de uma relação linear marcante entre essas variáveis. Em síntese, os dados reforçam que incêndios de maior dimensão estão ligados a níveis mais elevados de poluição atmosférica e que, de forma geral, o calor intenso costuma vir acompanhado de condições de vento mais calmo.

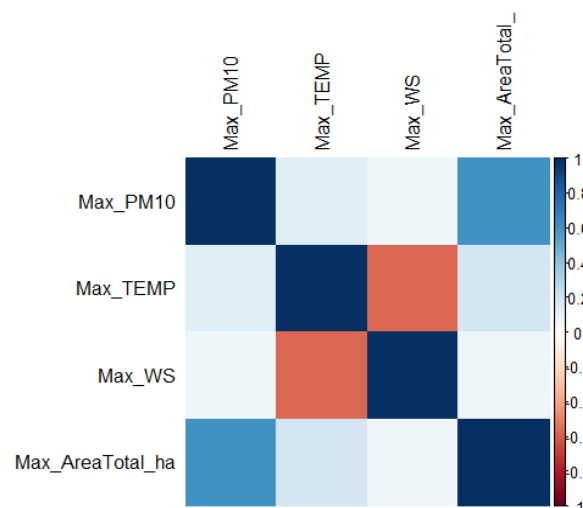


Figura 5: Matriz de correlação mostrando as relações lineares entre os máximos mensais de PM_{10} (Max_PM10), temperatura (Max_TEMP), velocidade do vento (Max_WS) e área total ardida (Max_AreaTotal_ha). As cores indicam a força e direção da correlação: azul para correlação positiva, vermelho para correlação negativa e branco para correlação fraca ou nula. A intensidade da cor representa a magnitude da correlação, conforme a escala à direita. Dados referentes ao período de junho a outubro, de 2011 a 2020, nas regiões Norte e Centro de Portugal.

Estacionariedade

A análise dos gráficos de autocorrelação e de autocorrelação parcial (Figura 6) indicam que a série de máximos mensais de PM_{10} exibe independência. Para verificar a estacionariedade foi feito o teste de *Dickey-Fuller* tendo-se obtido um valor-p de 0.05. Para um nível de significância de 5%, não se rejeita a hipótese nula de presença de raiz unitária, dando indícios de que a série é não estacionária.

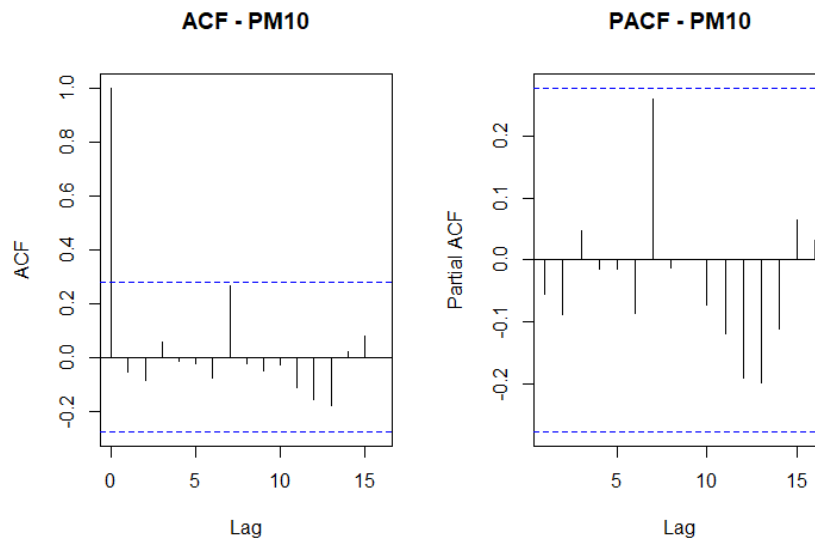


Figura 6: Gráficos das funções de autocorrelação e de autocorrelação parcial do PM_{10} .

Como a análise exploratória inicial levantou dúvidas relativamente à estacionariedade da série de máximos anuais de PM_{10} , iremos proceder à sua análise em contexto estacionário e não estacionário. Embora o ajuste de um modelo GEV estacionário (conforme apresentado na Tabela 2) constitua um ponto de partida válido para uma caracterização inicial, os resultados da análise de dependência temporal justificam a exploração de modelos GEV não estacionários. Tais modelos permitirão que os parâmetros da distribuição GEV (localização, escala e/ou forma) variem em função de covariáveis temporais ou sazonais, abordagem que será desenvolvida na Secção seguinte, com o intuito de capturar a dinâmica temporal subjacente e proporcionar uma representação mais precisa do comportamento dos eventos extremos de PM_{10} .

3.1.2 Ajuste do modelo GEV estacionário

Após a análise exploratória de máximos mensais de PM_{10} , procedeu-se ao ajuste do modelo de GEV a esta série. Este ajuste foi realizado sob a suposição de estacionariedade, o que implica que os três parâmetros da distribuição GEV — localização (μ), escala (σ) e forma (ξ) — são considerados constantes para todo o período de estudo e para todos os meses da época de fogos incluídos na série.

O ajuste do modelo GEV foi realizado utilizando o método MLE, que fornece os valores dos parâmetros que maximizam a função de verossimilhança, ou seja, os valores que tornam a probabilidade de observar os dados amostrais a mais alta possível. Para a implementação computacional, utilizou-se o ambiente de programação estatística R (*R Project for Statistical Computing*), no ambiente *RStudio*, e com recurso à biblioteca *ismev*, que oferece funções para a estimação de modelos de valores extremos.

Tabela 2: Estimativas dos parâmetros do modelo GEV estacionário para os máximos mensais de PM_{10} .

Parâmetro	Estimativa	Erro Padrão	IC 95%
μ (Localização)	30,90	1,62	[27,72; 34,08]
σ (Escala)	9,88	1,44	[7,06; 12,70]
ξ (Forma)	0,38	0,14	[0,10; 0,66]

As estimativas dos parâmetros GEV obtidas para a série agregada de máximos mensais fornecem uma caracterização geral do comportamento dos eventos extremos de PM_{10} durante a época de fogos no período de estudo, sob a premissa de estacionariedade. Analisando a Tabela 2, podemos constatar que as estimativas do modelo GEV para os máximos mensais de PM_{10} durante a época de fogos fornecem evidências relevantes sobre o comportamento extremo da poluição atmosférica no período analisado. A estimativa do parâmetro de localização ($\hat{\mu} = 30,90 \mu g/m^3$) indica o nível de referência em torno do qual se concentram os valores extremos, enquanto a dispersão moderada associada à estimativa do parâmetro de escala ($\hat{\sigma} = 9,88 \mu g/m^3$) sugere variações significativas entre diferentes episódios. A estimativa do parâmetro de forma ($\hat{\xi}=0,38$) é positiva e seu intervalo de confiança a 95% ([0,10; 0,66]) não inclui zero. Este

resultado é crucial, pois indica que a distribuição dos máximos mensais de PM_{10} , quando considerados em conjunto, pertence ao domínio de atração da distribuição Fréchet. Isso implica que cauda superior da distribuição é pesada, sugerindo uma probabilidade não desprezível de ocorrência de eventos de poluição por PM_{10} de magnitude extremamente elevada. Essa característica é particularmente relevante no contexto de incêndios florestais, que podem gerar picos de concentração muito altos e potencialmente nocivos para a saúde pública.

Diagnóstico de ajuste

A validação da adequação do modelo GEV estacionário ajustado aos máximos mensais de PM_{10} é um passo fundamental para assegurar a confiabilidade das inferências e previsões. Para este fim, foram utilizadas diversas ferramentas de diagnóstico gráfico, as quais permitem uma avaliação visual da aderência do modelo aos dados observados.

Os resultados desses diagnósticos são apresentados na Figura 7, composta por quatro gráficos distintos: o gráfico PP, o gráfico QQ, o gráfico dos níveis de retorno (do inglês, *Return Level Plot*) e o gráfico da densidade ajustada (do inglês, *Density Plot*). A interpretação conjunta destes gráficos fornece uma compreensão abrangente da qualidade do ajuste do modelo, bem como identifica possíveis áreas onde o modelo pode apresentar limitações.

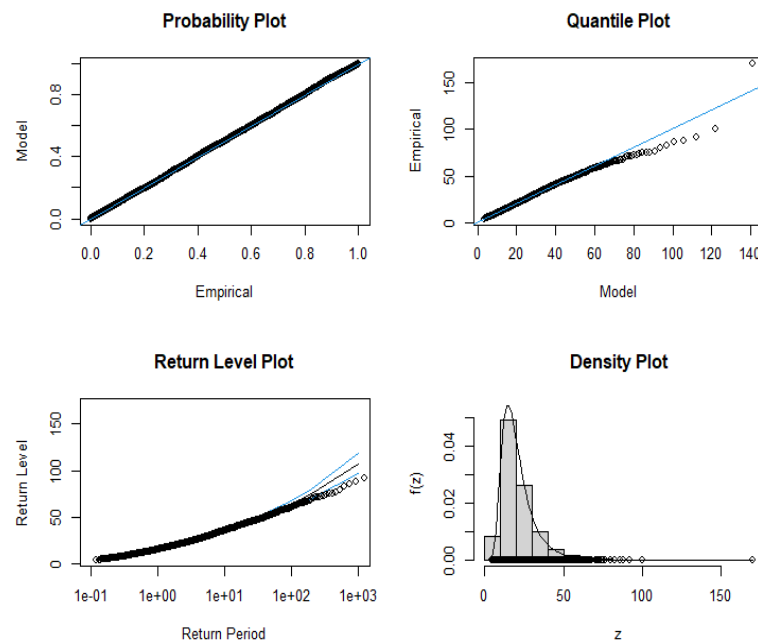


Figura 7: Gráficos do ajustamento das máximas mensais de PM_{10} da região Norte e Centro à distribuição GEV estacionário.

A análise gráfica sugere que o modelo GEV estacionário, ajustado aos máximos mensais de PM_{10} , proporciona um bom ajuste global aos dados. Os gráficos PP e a densidade ajustada, em particular, reforçam a adequação do modelo. Contudo, um ligeiro indicativo de subestimação nos quantis e níveis de retorno mais elevados (observado no gráfico QQ e no *Return Level Plot*) constitui um ponto relevante a considerar. Os dados mais extremos não apresentam um ajuste tão bom e a estimação de níveis de retorno para período de retorno elevados também sai dos limites de confiança do modelo. Esta observação permite levantar a hipótese de que a premissa de estacionariedade, ao considerar os parâmetros constantes ao longo do tempo, possa representar uma simplificação para os eventos de PM_{10} de maior magnitude. Esta percepção servirá de justificação para a exploração de abordagens não estacionárias na secção seguinte, visando aprimorar a capacidade de modelar e prever os eventos de poluição mais severos.

Seleção do modelo

Para avaliar a significância do parâmetro de forma ξ na caracterização dos máximos mensais de PM_{10} , foi realizado um TRV comparando o modelo GEV completo com o modelo Gumbel, que assume $\xi=0$. Os resultados estão apresentados na Tabela 3.

Tabela 3: Resultados do teste de razão de verossimilhança para comparação entre os modelos GEV e Gumbel na modelação dos máximos mensais de PM_{10} .

Estatística	Valor
LRstat	14,05
p-value	0,00017
Log-verossimilhança GEV	-204,23
Log-verossimilhança Gumbel	-211,25

A estatística de razão de verossimilhança ($LR = 14,05$) e o valor-p ($p = 0,0001$) indicam que a hipótese nula $H_0: \xi = 0$ deve ser rejeitada ao nível de significância de 5%. Assim, o parâmetro de forma é significativamente diferente de zero, justificando a escolha da distribuição GEV em detrimento da distribuição Gumbel. Este resultado evidencia a presença de cauda superior pesada na distribuição dos máximos mensais de PM_{10} , confirmando que a ocorrência de concentrações extremamente elevadas é mais provável do que se assumisse uma distribuição Gumbel.

Em conjunto, estas análises permitem concluir que os eventos extremos de PM_{10} , particularmente durante períodos de incêndios florestais, apresentam variabilidade significativa e uma probabilidade não desprezível de ocorrência de concentrações elevadas, sendo a utilização da distribuição GEV adequada para modelar e prever estes extremos.

Estimação de parâmetros de interesse

Períodos e níveis de retorno

Os níveis de retorno calculados para as concentrações de PM_{10} , com base no modelo GEV ajustado aos máximos mensais durante a época de incêndios (junho–outubro), encontram-se apresentados na Tabela 4. Estes valores representam as concentrações que se esperam ser ultrapassadas, em média, uma vez a cada T anos. A interpretação

destes resultados evidencia que, em média, concentrações de PM_{10} superiores a cerca de $66 \mu\text{g}/\text{m}^3$ são esperadas uma vez a cada 10 anos. Concentrações mais elevadas, como $119 \mu\text{g}/\text{m}^3$, são eventos mais raros, previstos para ocorrerem, em média, uma vez a cada 50 anos. Estes resultados permitem quantificar de forma robusta a ocorrência de episódios extremos de poluição atmosférica, fornecendo uma base sólida para a avaliação de risco e planejamento de medidas de mitigação.

Tabela 4: Estimativa de níveis de retorno de PM_{10} ($\mu\text{g}/\text{m}^3$) para diferentes períodos.

Período de retorno (anos)	Nível de retorno ($\mu\text{g}/\text{m}^3$)
10	66,04
20	85,27
50	119,40

Probabilidade de excedência

A Tabela 5 apresenta as probabilidades de ocorrência de concentrações de PM_{10} superiores a determinados valores críticos, calculadas com base no modelo GEV ajustado aos máximos mensais durante a época de incêndios (junho–outubro). Observa-se que a probabilidade de ultrapassar $50 \mu\text{g}/\text{m}^3$ é moderada ($\approx 21\%$), diminuindo progressivamente para valores mais elevados de PM_{10} , como $100 \mu\text{g}/\text{m}^3$ ($\approx 3\%$) e $120 \mu\text{g}/\text{m}^3$ ($\approx 2\%$). Estes resultados permitem quantificar a frequência esperada de episódios de poluição extrema, fornecendo uma base útil para planejamento e medidas de mitigação da qualidade do ar.

Tabela 5: Probabilidade de excedência para diferentes concentrações de PM_{10}

Concentração PM_{10} ($\mu\text{g}/\text{m}^3$)	Probabilidade de excedência $P(X > v)$
50	0,2091
75	0,0708
100	0,0324
120	0,0197

Quantis extremais

A Tabela 6 apresenta os quantis extremos, correspondentes às probabilidades estimadas de 90%, 95% e 99%, de concentração de PM_{10} estimados a partir do modelo GEV ajustado aos máximos mensais durante a época de incêndios (junho–outubro).

Tabela 6: Quantis extremais de PM_{10} .

Quantil	Concentração estimada ($\mu\text{g}/\text{m}^3$)
90%	66,04
95%	85,27
99%	154,18

A interpretação destes resultados evidencia que, o quantil de 90% corresponde a uma concentração de $66,04 \mu\text{g}/\text{m}^3$, indicando que 90% das concentrações estimadas são inferiores ou iguais a este valor, enquanto apenas 10% dos registos o excedem. No quantil de 95%, a concentração estimada é de $85,27 \mu\text{g}/\text{m}^3$, o que significa que apenas 5% das observações apresentam valores superiores a este nível. Por sua vez, o quantil de 99% apresenta uma concentração de $154,18 \mu\text{g}/\text{m}^3$, evidenciando que apenas 1% dos eventos ultrapassa este valor, caracterizando a ocorrência de episódios extremos de elevada concentração.

3.1.3 Ajuste do modelo GEV não estacionário

Tradicionalmente, o modelo GEV assume que os dados seguem uma distribuição estacionária, ou seja, que as características dos eventos extremos permanecem constantes ao longo do tempo. No entanto, evidências sugerem que a qualidade do ar e os padrões de incêndios florestais estão sujeitos a mudanças ao longo do tempo, influenciados por fatores como as mudanças climáticas, a variabilidade sazonal e a intensidade dos incêndios. Portanto, a suposição de estacionariedade pode não ser apropriada para os nossos dados.

Vamos aplicar o modelo GEV não estacionário, que permite que os parâmetros da distribuição GEV variem em função de covariáveis relevantes, como a área ardida, a temperatura e a velocidade do vento. Ao ajustar um modelo GEV não estacionário, podemos capturar tendências e mudanças nas características dos eventos extremos de PM_{10}

ao longo do tempo, fornecendo estimativas mais precisas dos períodos de retorno e da probabilidade de eventos de poluição extrema. Além disso, podemos avaliar a influência das covariáveis nos extremos de PM_{10} , obtendo informação sobre os fatores que contribuem para a degradação da qualidade do ar durante a época de incêndios florestais.

O conjunto de dados utilizado neste estudo é composto por cinco variáveis principais. A variável de interesse central é a concentração de PM_{10} expressa em microgramas por metro cúbico ($\mu\text{g}/\text{m}^3$). As restantes variáveis, tempo (1), temperatura do ar (2), velocidade do vento (3) e área ardida (4), são consideradas covariáveis explicativas, utilizadas para avaliar a influência de fatores meteorológicos e da severidade dos incêndios florestais na variabilidade das concentrações de PM_{10} . Antes do ajuste dos modelos, procedeu-se à normalização das covariáveis, pelos métodos de normalização Z e $Min-Max$. Entre as alternativas avaliadas, a normalização, subtração da média e divisão pelo desvio padrão, apresentou o melhor desempenho nos testes preliminares, sendo, portanto, adotada em todas as análises subsequentes. Na modelação não estacionária da distribuição GEV, foram assumidos inicialmente os seguintes pressupostos:

- **Parâmetro de forma (ξ):** constante;
- **Parâmetro de escala (σ):** constante, função linear do tempo ou função linear da área ardida;
- **Parâmetro de localização (μ):** função linear das covariáveis.

Ajuste dos modelos GEV não estacionários

O processo de modelação iniciou-se com a avaliação de modelos contendo apenas uma covariável. Nessa fase, verificou-se que o modelo em que o parâmetro de localização depende linearmente da área ardida, com escala constante e o modelo com parâmetro de localização dependente linearmente da área ardida e com a log-escala dependente linearmente da área ardida, apresentaram os melhores desempenhos. Este resultado destacou a relevância da área ardida como variável explicativa central, estabelecendo a expectativa de que esta permanecesse como variável determinante na definição dos modelos seguintes.

Na etapa seguinte, foram testadas todas as combinações de duas covariáveis, mantendo os pressupostos previamente definidos para os parâmetros, prosseguindo-se depois para modelos com três e, finalmente, quatro covariáveis. Embora todos os modelos tenham sido ajustados, no presente trabalho apresentam-se apenas os que evidenciaram os melhores resultados em cada grupo.

As Tabelas 7 e 8 apresentam os resultados dos modelos GEV não estacionários, diferenciando-se pela formulação adotada para a escala e onde p representa o número de parâmetros do modelo. Na primeira coluna encontra-se a parametrização adotada para o parâmetro de localização, μ_t . Assim, na primeira linha o parâmetro de localização é função linear da área ardida (4), na segunda linha é função linear o tempo (1) e da área ardida (4). A interpretação das linhas três e quatro segue o mesmo raciocínio. Os critérios utilizados para a seleção do modelo foram o simétrico da log-verosimilhança (nlh), o critério de informação de Akaike (AIC) e o critério de informação Bayesiano (BIC), que permitem avaliar o equilíbrio entre qualidade do ajustamento e parcimónia do modelo.

Tabela 7: Resultados dos modelos GEV (μ_t, σ, ξ),

μ_t (Covariáveis)	p	nlh	AIC	BIC
Área ardida (4)	4	193,6147	395,23	402,88
Tempo + Área ardida (1,4)	5	187,9678	385,94	395,50
Tempo + Temperatura + Área ardida (1,2,4)	6	197,9799	407,96	419,43
Tempo + Vento + Área ardida (1,3,4)	6	197,6118	407,22	418,70
Tempo + Temperatura + Vento + Área ardida (1,2,3,4)	7	187,5836	389,17	402,55

Tabela 8: Resultados dos modelos GEV ($\mu_t, \sigma_A = \exp(\beta_0 + \beta_1 A, \xi)$)

μ_t (Covariáveis)	p	nlh	AIC	BIC
Área ardida (4)	5	191,759	393,52	403,08
Tempo + Área ardida (1,4)	6	186,3308	384,66	396,13
Tempo + Temperatura + Área ardida (1,2,4)	7	186,0223	386,04	399,43
Tempo + Vento + Área ardida (1,3,4)	7	186,1549	386,31	399,69
Tempo + Temperatura + Vento + Área ardida (1,2,3,4)	8	186,0227	388,05	403,34

Na comparação direta dos modelos, verificou-se que tanto a formulação com escala constante como a formulação em que a escala depende exponencialmente da área ardida apresentaram os melhores resultados globais. Para comparar o modelo com escala constante com aquele em que a escala depende exponencialmente da área ardida, foi aplicado o TRV. A estatística do teste resultou em $D = 3,274$ com um valor- p aproximado de 0,07, indicando que não há evidência estatisticamente significativa para a adoção do modelo mais complexo. Dessa forma, o modelo com escala constante é considerado mais parcimonioso e robusto, sendo selecionado como modelo final para a análise dos extremos de PM_{10} . Portanto, o modelo GEV não estacionário mais adequado é aquele em que o parâmetro de localização depende linearmente do tempo e da área ardida, a escala é mantida constante e o parâmetro de forma permanece fixo. Este modelo representa o melhor compromisso entre simplicidade e qualidade de ajustamento, sendo adotado para a interpretação e estimação das quantidades de interesse associadas aos extremos de PM_{10} .

Diagnóstico de ajuste

A avaliação gráfica do modelo GEV não estacionário (modelo 1,4), exibida na Figura 8, indica bom ajuste aos dados de PM_{10} . Nos gráficos PP e QQ residuais, a maioria dos pontos alinha-se à linha de referência, mostrando que o modelo descreve adequadamente as probabilidades e quantis observados, com pequenos desvios nos extremos superiores, esperados em séries de máximos anuais. A ausência de padrões sistemáticos nos resíduos, aliada aos menores valores de AIC e BIC e aos resultados do teste de razão de verossimilhança, confirma que o modelo (1,4) é a melhor representação da série, equilibrando parcimônia e qualidade de ajustamento.

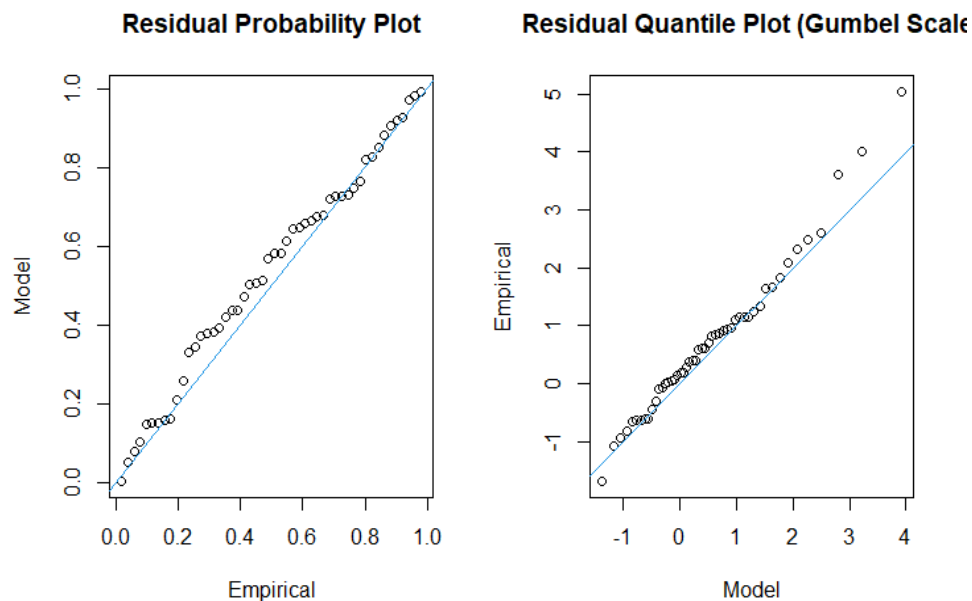


Figura 8: Ajuste do modelo GEV não estacionário para PM_{10} (modelo 1,4), em que a localização depende linearmente do tempo e da área ardida, a escala é constante e o parâmetro de forma é fixo.

Tabela 9: Estimativas dos parâmetros do modelo GEV não estacionário (modelo 1,4) para PM_{10} , em que o parâmetro de localização depende linearmente do tempo e da área ardida, a escala é constante e o parâmetro de forma é fixo.

Parâmetro	Estimativa	Erro Padrão	IC 95%
Localização (intercepto)	40,89	2,14	[36,73; 45,06]
Localização (tempo)	-0,29	0,08	[-0,44; -0,14]
Localização (área ardida)	7,45	0,81	[5,86; 9,04]
Escala (constante)	7,11	1,02	[5,11; 9,11]
Forma (constante)	0,38	0,13	[0,12; 0,65]

Os resultados obtidos na Tabela 9, mostram as estimativas dos parâmetros do modelo e indicam que os extremos de PM_{10} aumentam com a área ardida e diminuem levemente ao longo do tempo. O parâmetro de escala constante sugere que a variabilidade dos extremos se mantém estável, enquanto o parâmetro de forma positivo ($\xi = 0,38$) indica a presença de eventos extremos significativos, com cauda moderadamente pesada.

Os intervalos de confiança a 95% mostram que todos os parâmetros da localização são estatisticamente significativos, confirmando a influência relevante de tempo e área

ardida sobre os valores máximos de PM_{10} . O modelo apresenta bom ajuste aos dados, refletido pelo valor do simétrico da log-verosimilhança ($nllh = 187,9678$) e pelo desempenho gráfico, sendo considerado a melhor representação para a série analisada, equilibrando precisão e parcimónia.

Estimação de parâmetros de interesse

A partir do modelo GEV não estacionário ajustado aos máximos mensais de PM_{10} para o período de 2011–2020, em que o parâmetro de localização (μ) varia em função do tempo e da área ardida padronizada, procedeu-se à estimativa das concentrações extremas esperadas para o mês de junho de 2025, considerando uma área ardida de 5,500 ha. Para este cenário, a área ardida foi padronizada ($z=2,02$), resultando nos seguintes parâmetros estimados do modelo:

$$\hat{\mu}_{2025} = 40,89 - 0,29 * (2025 - 2020) + 7,45 * 2,02 = 54,49, \hat{\sigma} = 7,11, \hat{\xi} = 0,384$$

O valor médio esperado de PM_{10} para junho de 2025 é de $54,24 \mu\text{g}/\text{m}^3$. Este valor representa a concentração de PM_{10} mais provável naquele mês, considerando as condições de tempo e a área ardida. É o nível típico esperado, sem considerar eventos extremos.

Níveis de Retorno

Os níveis de retorno correspondem às concentrações de PM_{10} que se esperam ser ultrapassadas, em média, uma vez a cada T anos. Para junho de 2025, com área ardida de 5,500 ha, os níveis de retorno estimados são apresentados na Tabela 10:

Tabela 10: Estimativa dos níveis de retorno de PM_{10} ($\mu\text{g}/\text{m}^3$) para diferentes períodos de retorno.

Período de Retorno (anos)	Nível de Retorno ($\mu\text{g}/\text{m}^3$)
10	79,66
20	93,66
50	118,60

A análise mostra que concentrações superiores a $\approx 79,7 \mu\text{g}/\text{m}^3$ poderão ocorrer, em média, uma vez a cada 10 anos, enquanto valores na ordem dos $118,6 \mu\text{g}/\text{m}^3$ são extremamente raros, associados a um período de retorno de cerca de 50 anos. Estes valores são significativamente superiores aos observados no modelo estacionário, refletindo o impacto da área ardida na intensificação das concentrações máximas de PM_{10} .

Probabilidade de excedência

A Tabela 11 apresenta as probabilidades de ocorrência de concentrações de PM_{10} superiores a valores críticos sob o mesmo cenário (junho de 2025, área ardida = 5,500 ha).

Tabela 11: Probabilidade de excedência para diferentes limiares de PM_{10} .

Concentração PM_{10} ($\mu\text{g}/\text{m}^3$)	Probabilidade de excedência $P(X > v)$
60	0,3896
80	0,0981
100	0,0384
120	0,0192
150	0,0087

Observa-se que a probabilidade de ultrapassar $80 \mu\text{g}/\text{m}^3$ é de cerca de 9,8%, e a de ultrapassar $100 \mu\text{g}/\text{m}^3$ é de 3,8%. Estes valores indicam uma probabilidade não desprezável de ocorrência de episódios severos de poluição atmosférica durante períodos com elevada área ardida.

Quantis extremos

A Tabela 12 resume os quantis extremos correspondentes às probabilidades de 90%, 95% e 99%, obtidos a partir da distribuição ajustada.

Tabela 12: Quantis extremos estimados de PM_{10} ($\mu\text{g}/\text{m}^3$) para junho de 2025.

Quantil	Concentração estimada ($\mu\text{g}/\text{m}^3$)
90%	79,66
95%	93,66
99%	144,11

Estes quantis mostram que 90% dos valores esperados estarão abaixo de aproximadamente $80 \mu\text{g}/\text{m}^3$, mas que há uma probabilidade de 1% de ocorrerem concentrações

superiores a $144 \mu\text{g}/\text{m}^3$ em junho de 2025, sob condições meteorológicas e de área arida extremas.

3.2 Método de Excedências

Após a análise baseada no método dos máximos anuais, utilizando agregados mensais da concentração de PM_{10} , optou-se por avançar para o método das excedências, aplicando agora a GPD aos dados diários. Esta mudança justifica-se pela maior resolução temporal que a série diária proporciona, permitindo identificar e caracterizar episódios de poluição extrema de forma mais rigorosa. Com efeito, fenômenos críticos podem ser atenuados ou mesmo não detetados quando se utilizam séries agregadas mensalmente, ao passo que a abordagem diária preserva a variabilidade intrínseca dos eventos extremos.

A abordagem POT permite analisar o comportamento da cauda da distribuição dos dados, focando especificamente na frequência e magnitude dos eventos que excedem um determinado limiar. Esta metodologia é particularmente útil quando se dispõe de dados com uma frequência temporal mais elevada, como os dados diários utilizados neste estudo, complementando a análise dos máximos mensais abordada anteriormente.

Assim, esta secção detalha o processo de aplicação do método POT, incluindo a crucial etapa de seleção de um limiar apropriado para a concentração de PM_{10} , o ajuste da GPD aos excessos identificados, à avaliação da adequação do modelo ajustado através de diversas ferramentas de diagnóstico, e a subsequente estimação de quantis extremos e dos seus respetivos períodos de retorno.

3.2.1 Estatística descritiva das variáveis

A caracterização inicial dos dados é essencial para compreender a estrutura e a variabilidade das concentrações diárias de PM_{10} no período em estudo. Para cada mês (junho a outubro), a série de máximos diários de PM_{10} entre 2011 e 2020 foi submetida a uma análise exploratória detalhada. O conjunto de dados reúne observações diárias na região Norte/Centro, restritas aos meses indicados ao longo de 2011–2020, perfazendo por estação até 153 dias por ano ($30+31+31+30+31$) e, no máximo, 1 530

observações. Na Tabela 13 encontram-se sumariadas as principais medidas descritivas das variáveis em estudo.

Tabela 13: Estatística descritiva dos máximos diários das variáveis em estudo

Estatística	PM10 ($\mu\text{g}/\text{m}^3$)	Temperatura ($^{\circ}\text{C}$)	Velocidade do vento (m/s)	Área adida (m^2)
Média	21,11	14,40	3,04	35,17
Mediana	18,36	13,75	2,81	2,08
Desvio Padrão	11,29	5,78	1,11	87,71
Mínimo	4,33	1,14	0,99	0
Máximo	169,80	30,30	9,77	960,63
Assimetria	2,27	—	1,25	4,50
Curtose	12,94	-0,91	2,25	25,90

A análise descritiva evidencia padrões distintos entre as variáveis em estudo. As concentrações de PM_{10} e a Área Ardida demonstram uma acentuada assimetria positiva e uma curtose elevada, características que apontam para a presença de valores extremos e distribuições com caudas pesadas. Este perfil é, de facto, coerente com a natureza dos fenómenos de poluição atmosférica e dos incêndios florestais. Em contrapartida, a Temperatura e a Velocidade do Vento tendem a apresentar distribuições mais simétricas ou com assimetria ligeira, e uma curtose mais baixa, o que sugere uma menor propensão para a ocorrência de valores extremos de magnitude excepcional. Estes conhecimentos iniciais reforçam a pertinência da aplicação de metodologias específicas, como a TVE, para uma compreensão mais aprofundada dos eventos mais severos de PM_{10} .

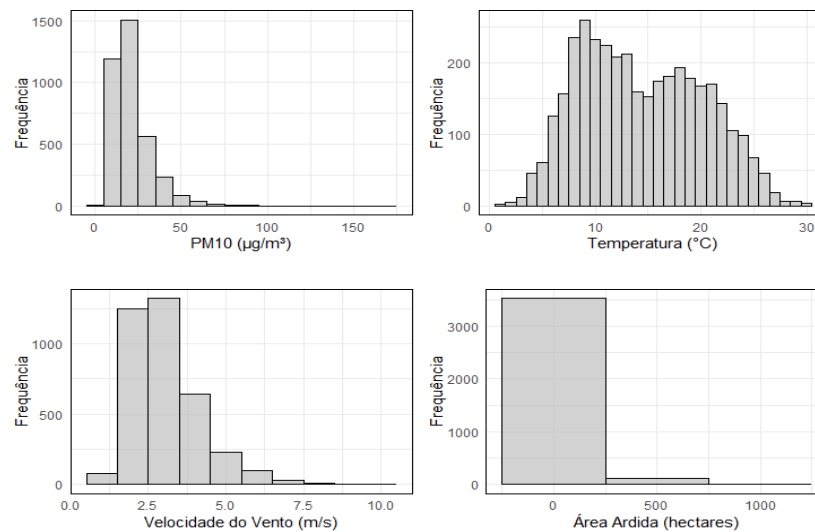


Figura 9: Distribuição das variáveis PM_{10} , Temperatura, Velocidade do Vento e Área Ardida

Através da Figura 9, podemos observar que as variáveis PM_{10} , Velocidade do Vento e Área Ardida apresentam uma, assimetria à direita, dando indícios que os dados são dominados por medições de baixo valor (baixa poluição, ventos leves e incêndios pequenos ou inexistentes). A presença de uma cauda longa indica a existência de eventos extremos que, embora pouco frequentes, podem ter um grande impacto. Por exemplo, um grande incêndio (na cauda da distribuição de Área Ardida) ou um dia de poluição intensa (na cauda de PM_{10}) são eventos de baixa probabilidade, mas de alta relevância para a gestão ambiental.

A distribuição de temperatura, que se assemelha a uma distribuição normal, sugere um comportamento mais previsível. A maioria das medições se agrupa em torno da média de 10°C , e a frequência de temperaturas muito altas ou muito baixas diminui de forma simétrica. Essa é a distribuição mais "bem-comportada" entre as quatro, indicando que a temperatura do local segue um padrão sazonal relativamente estável, sem a mesma propensão a eventos extremos raros que as outras variáveis. No gráfico de Área Ardida observa-se que a grande maioria dos incêndios observados conduz a uma área ardida inferior a 250 hectares. A dispersão subsequente, embora pareça insignificante, mostra que eventos de incêndios com áreas ardidas acima de 500 hectares, embora extremamente raros, ocorreram.

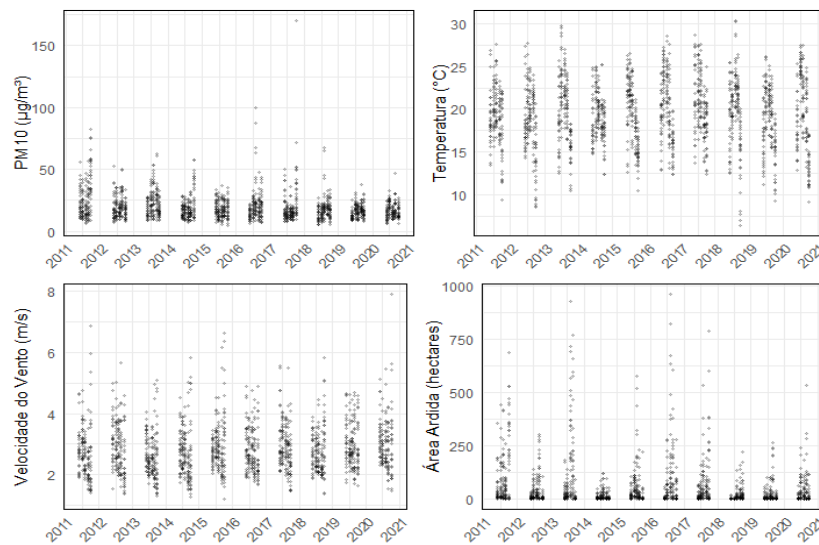


Figura 10: Distribuição anual da serie dos máximos diários de quatro variáveis: PM₁₀, Temperatura, Velocidade do Vento e Área Ardida ao longo de um período de 10 anos, de 2011 a 2020.

Em primeiro lugar, a análise da Figura 10, não mostra uma tendência clara de aumento ou diminuição para nenhuma das variáveis ao longo dos 10 anos. Os dados de PM₁₀, Temperatura e Velocidade do Vento mantiveram-se em faixas de valores consistentes, sem um padrão de mudança significativa de ano para ano. Isso sugere que as condições ambientais gerais na região permaneceram relativamente estáveis durante o período. A Área Ardida, por sua vez, também não apresenta uma tendência de aumento no número ou na escala dos incêndios ao longo do tempo.

Em resumo, os gráficos mostram um sistema ambiental que, na maioria das vezes, opera em condições normais, mas é propenso a eventos raros e de alto impacto que merecem uma investigação mais aprofundada para entender as causas e as inter-relações entre eles.

3.2.2 Ajuste do modelo GPD estacionário

Para a definição do limiar de excedência, foi adotado o valor de $45 \mu\text{g}/\text{m}^3$, considerando a sua relevância de acordo com padrões de qualidade do ar recomendados pela Organização Mundial da Saúde (OMS) para a concentração de partículas inaláveis com diâmetro inferior a 10 micrômetros (PM₁₀). Assim, a escolha deste limiar visa

identificar excedências associadas a condições de risco à saúde humana, permitindo caracterizar adequadamente a cauda superior da distribuição — componente central da modelagem por valores extremos. Portanto, o uso do limiar de $45 \mu\text{g}/\text{m}^3$ está em conformidade com critérios normativos internacionais e reforça o foco da análise na ocorrência de eventos de poluição atmosférica potencialmente prejudiciais à saúde pública.

Após a definição do limiar em $45 \mu\text{g}/\text{m}^3$ para a concentração de PM_{10} , foram identificados todos os episódios diários que ultrapassaram este valor. Estes casos constituem a base da modelação da GPD. Antes do ajuste formal do modelo, procedeu-se a uma análise descritiva destes excedentes, com o objetivo de caracterizar a sua distribuição e variabilidade. A Tabela 14 apresenta a estatística descritiva dos valores de PM_{10} acima do limiar de $45 \mu\text{g}/\text{m}^3$. Foram identificados 148 episódios que ultrapassaram o valor de referência, o que corresponde a uma fração relevante da série temporal e confirma a adequação do limiar escolhido para a análise com base na GPD.

Tabela 14: Estatísticas descritivas das excedências acima de $45 \mu\text{g}/\text{m}^3$ da concentração diária de PM_{10} .

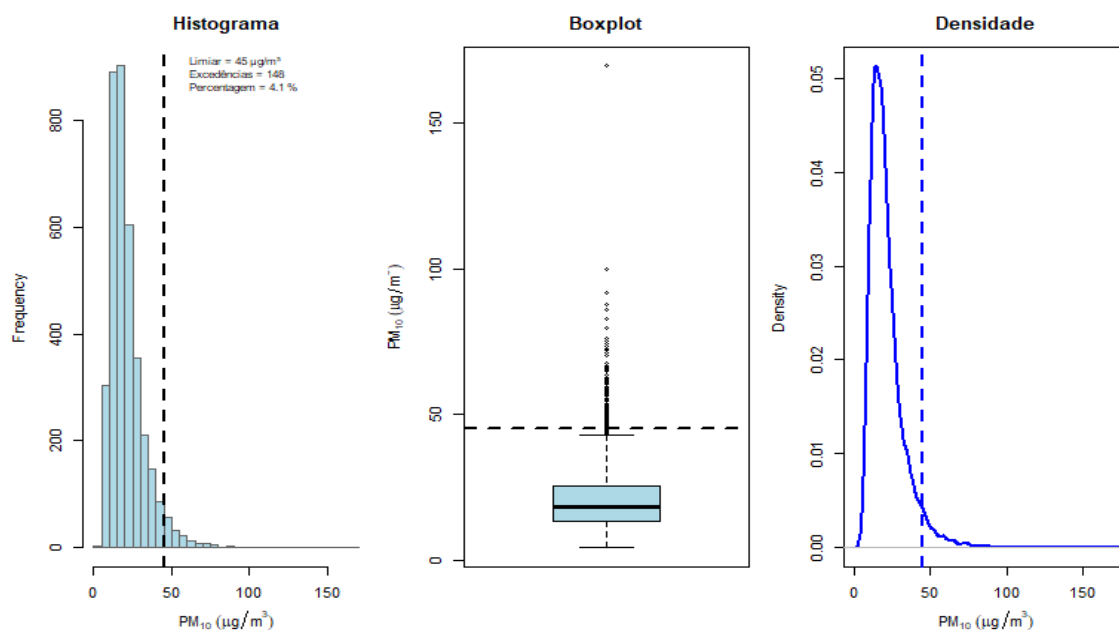
Estatística	Valor
Número de excedências (N)	148
Média	56,51
Desvio padrão	14,04
Mínimo	45,02
Mediana	52,67
Máximo	169,80
Assimetria	4,10
Curtose	27,27

O valor médio das excedências foi de $56,51 \mu\text{g}/\text{m}^3$, com um desvio padrão de $14,04 \mu\text{g}/\text{m}^3$, refletindo uma variabilidade moderada em torno da média. O mínimo observado ($45,02 \mu\text{g}/\text{m}^3$) corresponde ao ponto de corte definido, enquanto a mediana ($52,67 \mu\text{g}/\text{m}^3$) indica que metade dos episódios extremos se situa relativamente próximo do limiar.

Em contraste, o máximo registado ($169,80 \mu\text{g}/\text{m}^3$) revela a ocorrência de episódios de poluição atmosférica particularmente severos, muito acima do valor limite adotado. A assimetria positiva elevada (4,10) evidencia que a distribuição dos excedentes é fortemente enviesada à direita, traduzindo a presença de valores muito altos em comparação com a maioria das observações. A curtose acentuada (27,27) reforça esta leitura, demonstrando que se trata de uma distribuição com caudas pesadas e elevada propensão para valores extremos.

Estes resultados são consistentes com a expectativa de que a poluição por PM_{10} durante a época de incêndios florestais se caracteriza por episódios raros, mas de grande intensidade, o que justifica a aplicação de metodologias assintóticas da TVE na sua modelação e avaliação de risco.

Figura 11: Distribuição dos valores máximos diários de PM_{10} (limiar $> 45 \mu\text{g}/\text{m}^3$).



A Figura 11 apresenta a distribuição dos valores máximos diários de PM_{10} para o conjunto total de observações, sendo o limiar de $45 \mu\text{g}/\text{m}^3$ assinalado nos gráficos por uma linha tracejada.

No histograma observa-se que a maioria dos valores de PM_{10} se encontra concentrada em níveis baixos, verificando-se uma redução acentuada da frequência à medida que as concentrações aumentam. O limiar considerado localiza-se na extremidade direita da distribuição, evidenciando que apenas uma pequena proporção das observações ultrapassa os $45 \mu\text{g}/\text{m}^3$. No período analisado, as excedências correspondem a 148 registros, o que representa cerca de 4,1% do total, indicando que episódios de poluição mais elevada são pouco frequentes.

O boxplot mostra que tanto a mediana como o intervalo interquartil se situam claramente abaixo do limiar definido. Os valores representados acima do limite superior do diagrama correspondem a ocorrências excepcionais, associadas a concentrações de PM_{10} consideravelmente mais elevadas. Estes valores extremos, embora raros, evidenciam a existência de episódios pontuais de poluição intensa.

O gráfico de densidade confirma o padrão identificado anteriormente, apresentando uma distribuição fortemente assimétrica à direita. O máximo da densidade ocorre para valores baixos de PM_{10} , enquanto a cauda da distribuição se prolonga para concentrações mais elevadas. A posição do limiar numa zona de baixa densidade reforça a ideia de que a maioria dos valores observados se encontra abaixo dos $45 \mu\text{g}/\text{m}^3$.

Em termos globais, a análise evidencia que os níveis de PM_{10} são, na maioria do tempo, relativamente reduzidos, ocorrendo apenas de forma ocasional episódios de concentrações elevadas, os quais justificam uma análise específica no âmbito do estudo de valores extremos e da avaliação do risco associado à qualidade do ar.

Na Figura 12, os pontos azuis representam os dias em que as concentrações de PM_{10} ficaram abaixo ou iguais a $45 \mu\text{g}/\text{m}^3$, indicando condições de poluição consideradas normais. Os pontos pretos correspondem aos dias em que os níveis de PM_{10} superaram $45 \mu\text{g}/\text{m}^3$, evidenciando eventos extremos de poluição. A codificação por cores permite identificar visualmente períodos de normalidade e anormalidade, facilitando a análise de valores atípicos.

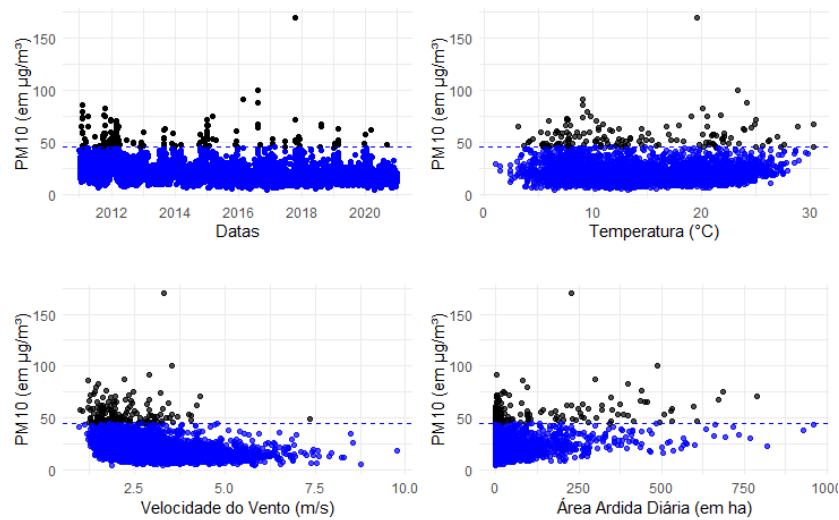


Figura 12: Distribuição de PM_{10} em relação ao limiar de $45 \mu\text{g}/\text{m}^3$

Para modelar os valores extremos de PM_{10} , foi necessário ajustar os parâmetros do modelo estatístico selecionado. O ajuste foi realizado utilizando MLE, que permite estimar os parâmetros de forma a maximizar a probabilidade de observar os dados disponíveis. A Tabela 15 apresenta as estimativas dos parâmetros do modelo GPD estacionário ajustado aos valores extremos de PM_{10} .

O parâmetro de escala (σ) foi estimado em $10,048 \mu\text{g}/\text{m}^3$ e reflete a variabilidade dos excedentes de PM_{10} acima do limiar considerado. Valores mais elevados de σ indicam maior dispersão dos extremos em torno do limiar.

O parâmetro de forma (ξ) foi estimado em $0,124$. Um valor positivo sugere uma cauda ligeiramente pesada, indicando que a ocorrência de eventos extremos elevados de PM_{10} é possível. Contudo, o intervalo de confiança a 95% de $[-0,04; 0,29]$ inclui zero, o que significa que não há evidência estatística suficiente para afirmar que a cauda seja significativamente pesada. Assim, a cauda da distribuição pode comportar-se de forma próxima de uma distribuição exponencial, com ocorrência de extremos menos frequente, embora valores positivos de ξ indiquem que eventos extremos ainda são possíveis.

Em suma, o modelo GPD estacionário permite caracterizar os extremos de PM_{10} , mostrando variabilidade moderada nas excedências e fornecendo uma base sólida para avaliação de risco e planejamento de medidas de mitigação da poluição atmosférica.

Tabela 15: Estimativas dos parâmetros do modelo GPD estacionário para PM_{10}

Parâmetro	Estimativa	Erro Padrão	IC 95%
Escala (σ)	10,048	1,182	[7,73; 12,37]
Forma (ξ)	0,124	0,085	[-0,04; 0,29]

Diagnóstico de ajuste

A adequação do modelo GPD estacionário foi avaliada através da análise de gráficos de diagnóstico, conforme ilustrado na Figura 13. Estes gráficos proporcionam uma validação visual do ajuste.

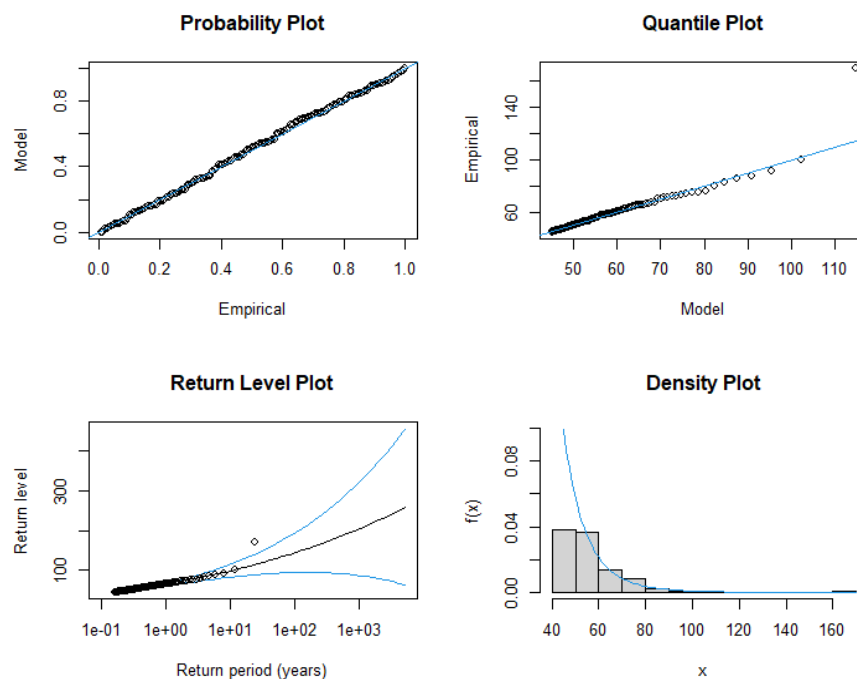


Figura 13: Gráficos do ajustamento de PM_{10} diários da região Norte e Centro à distribuição GPD

Em resumo, o ajustamento do modelo GPD estacionário aos dados diários das excedências de PM_{10} revelou-se eficaz para a sua modelagem, apesar de o parâmetro de forma não ser estatisticamente distinto de zero. As estimativas dos parâmetros são consistentes com a natureza dos fenómenos de poluição extrema e os diagnósticos gráficos confirmam a boa qualidade do ajustamento. Este modelo estabelece uma base sólida para a compreensão dos eventos extremos de PM_{10} , abrindo caminho para o cálculo dos níveis de retorno e para a investigação de modelos GPD não estacionários que possam incorporar a influência de covariáveis.

Seleção do modelo

Para avaliar a significância do parâmetro de forma ξ na caracterização das excedências de PM_{10} acima de $45 \mu\text{g}/\text{m}^3$, foi realizado um TRV comparando o modelo GPD com o modelo exponencial, que assume $\xi=0$. Os resultados estão apresentados na Tabela 16.

Tabela 16: Resultados do teste de razão de verossimilhança para comparação entre os modelos GPD e modelo exponencial na modelação dos máximos mensais de PM_{10} .

Estatística	Valor
LRstat	3,3683
p-value	0,0665
Log-verossimilhança GPD	-507,8745
Log-verossimilhança Exponencial	-509,5586

A estatística de razão de verossimilhança ($LR = 3,3683$) e o valor- $p = 0,0665$ indicam que a hipótese nula $H_0: \xi = 0$ não deve ser rejeitada ao nível de significância de 5%, e que deve ser rejeitada para um nível de significância de 10%. Assim, consoante o nível de significância adotado, não existe (existe) evidência suficiente de que o parâmetro de forma seja diferente de zero, justificando que o modelo exponencial (GPD) seja adequado para descrever as excedências de PM_{10} acima de $45 \mu\text{g}/\text{m}^3$. Em dados ambientais, as caudas pesadas ajustam melhor a probabilidade de eventos extremos, evitando a subestimação de quantis elevados e períodos de retorno que ocorre com caudas exponenciais, e por este motivo o modelo escolhido é o modelo GPD.

Estimação de parâmetros de interesse

Para caracterizar os eventos extremos de PM_{10} acima do limiar de $45 \mu\text{g}/\text{m}^3$, foram calculadas as principais quantidades de interesse utilizando o modelo GPD ajustado às excedências. As análises incluíram:

1. **Níveis de retorno:** indicam a concentração de PM_{10} esperada para diferentes períodos de retorno (2, 5, 10, 20 e 50 anos), fornecendo uma medida direta da magnitude de eventos raros e permitindo avaliar o risco associado a episódios críticos de poluição.
2. **Probabilidades de excedência:** estimadas em base diária e anual, fornecem a chance de ultrapassar determinados limiares em um único dia ou ao longo de um ano, permitindo avaliar a frequência e a recorrência de concentrações elevadas.
3. **Quantis extremos:** correspondem aos percentis superiores da distribuição das excedências (90%, 95% e 99%), representando os episódios mais severos e fornecendo informações essenciais para a gestão ambiental e avaliação de impactos à saúde pública.

A combinação dessas estimativas oferece uma visão abrangente da distribuição das excedências de PM_{10} , evidenciando a utilidade do modelo GPD e/ou Exponencial para capturar a cauda superior da distribuição e fornecer referências robustas para avaliação de risco, planejamento e tomada de decisão em políticas de qualidade do ar.

Tabela 17: Estimativas dos níveis de retorno de PM_{10} ($\mu\text{g}/\text{m}^3$) para diferentes períodos de retorno

Período de retorno (anos)	Nível de retorno ($\mu\text{g}/\text{m}^3$)
2	74,53
5	87,85
10	98,98
20	11,12
50	128,86

A Tabela 17 apresenta os níveis de retorno estimados de PM_{10} para diferentes períodos de retorno, calculados com o modelo GPD. Observa-se que os níveis de retorno do modelo GPD aumentam de forma consistente com o período, variando de aproximadamente $74,5 \mu\text{g}/\text{m}^3$ para um período de 2 anos até $128,9 \mu\text{g}/\text{m}^3$ para 50 anos. Esse

comportamento reflete a presença de uma cauda “ligeiramente” pesada na distribuição das excedências, indicando que eventos de maior intensidade são menos frequentes, mas possíveis em horizontes temporais mais longos.

A Tabela 18 apresenta as probabilidades estimadas de excedência de diferentes concentrações de PM_{10} , calculadas com base no modelo GPD estacionário, ajustado às excedências acima do limiar de $45 \mu\text{g}/\text{m}^3$. As probabilidades foram calculadas considerando a taxa de excedências observada na série. A probabilidade de se observarem concentrações acima de $60 \mu\text{g}/\text{m}^3$, é de aproximadamente 1%. À medida que o limiar aumenta, as probabilidades diminuem rapidamente. Por exemplo, a probabilidade de a concentração ultrapassar $120 \mu\text{g}/\text{m}^3$ é de cerca de 0,02%.

Tabela 18: Probabilidade de excedência para diferentes limiares de PM_{10} .

Concentração de PM_{10} ($\mu\text{g}/\text{m}^3$)	Probabilidade de excedência $P(X>v)$
60	0,01030
80	0,00224
100	0,00062
120	0,00021
150	0,00005

A Tabela 19 apresenta os quantis estimados de PM_{10} para os percentis 90%, 95% e 99%, considerando o modelo GPD. Estes quantis fornecem uma medida clara da frequência e intensidade dos episódios críticos de poluição, complementando a análise baseada em médias e medianas e permitindo avaliar os riscos ambientais e impactos potenciais na saúde pública.

Tabela 19: Quantis extremos estimados de PM_{10} ($\mu\text{g}/\text{m}^3$)

Quantil	Concentração estimada ($\mu\text{g}/\text{m}^3$)
90%	98,98
95%	111,12
99%	143,69

As análises de níveis de retorno, probabilidades de excedência e quantis extremos indicam que o modelo GPD fornece estimativas mais realistas para eventos raros e intensos de PM_{10} , enquanto o modelo exponencial tende a subestimá-los. Essas medidas permitem avaliar a frequência e magnitude de episódios críticos, oferecendo referências robustas para gestão ambiental e planejamento de mitigação da poluição atmosférica. Com base nesses resultados, torna-se pertinente avançar para o ajuste do modelo GPD não estacionário, que permite considerar a influência de variáveis explicativas ao longo do tempo e capturar a dinâmica temporal das concentrações extremas de PM_{10} .

3.2.3 Ajuste do modelo GPD não estacionário

Dada a incerteza observada na estimativa do parâmetro de forma (ξ) do modelo GPD estacionário, optou-se por ajustar um modelo GPD não estacionário, incorporando covariáveis explicativas relevantes, tempo (1), temperatura (2), velocidade do vento (3) e área ardida (4). A inclusão dessas covariáveis permite modelar a variabilidade temporal e as influências ambientais sobre os valores extremos de PM_{10} , oferecendo uma descrição mais realista do comportamento da cauda da distribuição.

O ajuste do modelo foi realizado utilizando o método MLE, estimando como os parâmetros de escala (σ) e forma (ξ) da distribuição GPD variam em função das covariáveis. Esta abordagem possibilita avaliar se fatores ambientais específicos estão associados a aumentos na magnitude ou frequência de eventos extremos de PM_{10} , fornecendo uma interpretação mais detalhada do risco de poluição elevada em diferentes condições.

Modelos com parâmetro de escala linear

Na primeira formulação, o parâmetro de escala foi modelado como uma função linear das covariáveis consideradas, ou seja:

$$\sigma_t = \beta_0 + \beta_1 t^* + \beta_2 V^* + \beta_3 T^* + \beta_4 A^*,$$

onde t^* , V^* , T^* e A^* representam, respectivamente, as covariáveis tempo, velocidade do vento, temperatura e área ardida, normalizadas.

Esta abordagem pressupõe que o efeito de cada covariável sobre o parâmetro de escala é aditivo e independente. Apesar de ser simples e de interpretação direta, esta formulação apresenta uma limitação teórica importante: o modelo linear não garante que σ seja sempre positivo, o que é uma condição necessária para a validade da distribuição GPD.

Modelos com parâmetro de escala log-linear:

Para ultrapassar essa limitação, considerou-se uma segunda parametrização, de natureza log-linear, na qual o logaritmo natural de σ é modelado como uma combinação linear das covariáveis:

$$\sigma_t = \exp(\beta_0 + \beta_1 t^* + \beta_2 V^* + \beta_3 T^* + \beta_4 A^*).$$

Esta formulação assegura que o parâmetro de escala é sempre positivo e permite uma interpretação multiplicativa dos coeficientes. Assim, os parâmetros β podem ser interpretados como o efeito percentual (em termos relativos) das covariáveis sobre σ . Por exemplo, um coeficiente β_4 positivo indica que um aumento na variável área ardida conduz a um aumento exponencial do parâmetro de escala, refletindo uma maior variabilidade e severidade nos eventos extremos.

A formulação linear é conceptualmente mais simples e adequada quando se pretende avaliar relações diretas e aditivas entre as covariáveis e o parâmetro de escala. No entanto, a formulação log-linear é estatisticamente mais robusta, garantindo a positividade de σ e uma interpretação mais realista em contextos de modelação de extremos. Por este motivo, ambas as formulações foram testadas e comparadas, de modo a determinar qual delas apresentava o melhor ajustamento aos dados observados e a descrição mais adequada do comportamento das caudas da distribuição.

Com base nas duas formulações descritas anteriormente, foram ajustados diferentes modelos da GPD, considerando diversas combinações de covariáveis nos parâmetros de escala (σ) e de forma (ξ). O objetivo desta etapa foi identificar a especificação que melhor representa o comportamento das caudas da distribuição, avaliando simultaneamente a influência das variáveis explicativas na severidade dos eventos extremos.

Para a comparação entre modelos, recorreram-se aos critérios simétrico da log-verosimilhança, AIC e BIC, os quais permitem avaliar o equilíbrio entre a qualidade do ajuste e a complexidade do modelo.

Na formulação linear, os modelos que incluíam simultaneamente o tempo e a área ardida apresentaram os menores valores de $nllh$, indicando bom ajustamento aos dados. Contudo, entre estes, o modelo que incluiu apenas as variáveis Tempo + Área ardida apresentou o menor valor de AIC (1014,27), revelando ser o mais parcimonioso dentro desta categoria, como se pode observar na Tabela 20 a seguir:

Tabela 20: Resultados dos modelos GPD com parâmetro de escala linear

Cova- riáveis	Parâmetro de Escala(σ)	Parâmetro de forma(ξ)	$nllh$	AIC	BIC
(4)	0,10	8,5853e-16	508,16	1020,33	1026,33
(1,4)	8,54 + 0,0004t*+ 1,61A*	0,0949	503,12	1014,27	1026,26
(1,3,4)	8,41+0,0004t*-0,26V*+ 1,61A*	0,0898	503,11	1016,23	1031,22
(1,2,3,4)	8,09+0,004t*-0,16T*+ 0,18V*+ 1,69A*	0,0871	503,11	1018,38	1036,36

(1) Tempo; (2) Temperatura; (3) Velocidade do Vento; (4) Área Ardida

Por outro lado, na formulação log-linear, apresentado na Tabela 21, o modelo em que o parâmetro de escala é dado por

$$\sigma_t = \exp(2,16 + 0,1256A^*)$$

$$\xi = 0,1003$$

isto é, o modelo que considera apenas a variável “Área ardida” como covariável, apresentou os valores mais baixos de AIC (1012,62) e BIC (1021,62).

Tabela 21: Resultados dos modelos GPD com parâmetro de escala log-linear

Cova- riáveis	Parâmetro de Escala (σ)	Parâmetro de Forma (ξ)	nllh	AIC	BIC
(4)	$\exp(2,16 + 0,1256A^*)$	0,1003	503,31	1012,62	1021,62
(1,4)	$\exp(2,2927 + 0,000143t^* + 0,386A^*)$	-0,0653	515,72	1039,43	1051,43
(1,3,4)	$\exp(2,07 + 0,004t^* + 0,06V^* + 0,199A^*)$	0,112	505,70	1021,40	1036,39
(1,2,3,4)	$\exp(1,893 + 0,00016t^* - 0,0968V - 0,2289T^* + 0,0380A^*)$	0,0913	508,13	1028,26	1046,26

(1) Tempo; (2) Temperatura; (3) Velocidade do Vento; (4) Área Ardida

A análise comparativa das duas formulações permitiu concluir que, embora a abordagem linear apresente resultados satisfatórios, a formulação log-linear revelou-se mais adequada à modelação dos extremos. Para além de garantir a positividade do parâmetro de escala (σ), esta parametrização proporcionou um melhor ajustamento global aos dados, refletido nos menores valores de AIC e BIC obtidos. Entre os modelos testados, aquele que incluiu apenas a variável “Área ardida” como covariável no parâmetro de escala apresentou o melhor desempenho, evidenciando que esta variável é o principal fator explicativo da variabilidade e severidade dos eventos extremos. Assim, o modelo GPD com parâmetro de escala log-linear e dependente exclusivamente da área ardida foi selecionado como o mais parcimonioso e representativo para descrever o comportamento das caudas da distribuição.

Na Tabela 22 são apresentadas as estimativas dos parâmetros do modelo e os respetivos intervalos de confiança a 95% admitindo que a formulação log-linear para o parâmetro de escala.

Tabela 22: Estimativas dos parâmetros do modelo GPD não estacionário para PM₁₀

Parâmetro	Estimativa	Erro Padrão	IC 95%
Escala (intercepto)	2,160	0,127	[1,910; 2,411]
Escala (Area Ardida)	0,126	0,043	[0,041; 0,210]
Forma (ξ)	0,100	0,082	[-0,06; 0,260]

O parâmetro de forma, ξ , foi estimado em 0,100. O valor positivo indica que a distribuição dos excessos é do tipo cauda pesada, um comportamento característico de fenômenos ambientais extremos, onde a probabilidade de eventos muito grandes decai lentamente. No entanto, o Intervalo de Confiança a 95% para ξ ($[-0,060; 0,260]$) inclui o zero. Este facto implica que o parâmetro de forma não é estatisticamente significativo ao nível de 5% em relação a um modelo exponencial ($\xi=0$). Apesar da evidência empírica apontar para uma cauda pesada, a ausência de significância estatística exige a interpretação cuidadosa das estimativas de risco associadas.

A análise do parâmetro de escala é crucial para quantificar a influência da covariável. O coeficiente de regressão para a Área Ardida (β_1) foi estimado em 0,126. O intervalo de confiança a 95% para este coeficiente ($[0,041; 0,210]$) não inclui o zero, o que confirma a significância estatística da covariável no modelo. Em resumo, o aumento da Área Ardida está diretamente associado a um aumento da variabilidade e da magnitude esperada dos picos de poluição de PM_{10} . Esta conclusão valida o papel da Área Ardida como um fator determinante e significativo na intensidade dos eventos extremos de poluição, justificando a sua inclusão no modelo de risco extremo.

Diagnóstico de ajuste

A avaliação da qualidade do ajustamento do modelo GPD foi realizada através da análise dos gráficos de diagnóstico de resíduos, representado na Figura 14.

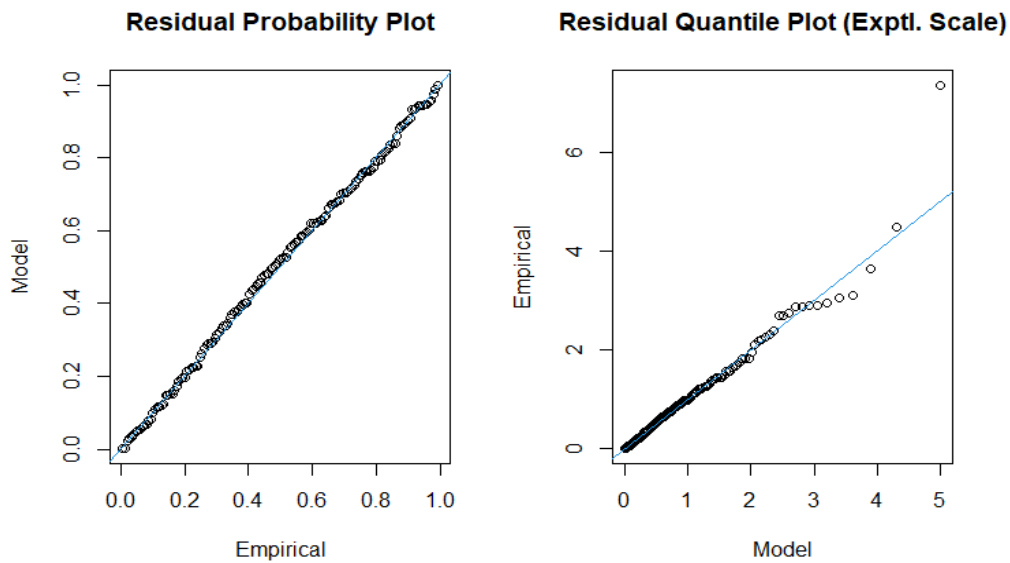


Figura 14: Gráficos de diagnóstico para o modelo GPD não estacionário ajustado

O Gráfico de PP dos resíduos demonstrou um ajuste linear à diagonal de 45 graus, o que fornece uma forte evidência de que a FDA do modelo é altamente adequada para descrever a distribuição dos excessos observados sobre o limiar. Contudo, o bom ajustamento global, o gráfico QQ dos resíduos revelou uma ligeira limitação na modelação dos valores mais extremos. Embora a maior parte dos dados se alinhe estreitamente à linha de referência, o desvio para cima no quantil superior indica que o modelo GPD tende a subestimar a magnitude dos valores residuais máximos.

Consequentemente, apesar de o modelo ser considerado robusto para a maior parte dos dados da cauda, os resultados decorrentes da estimação de quantis muito elevados (como níveis de retorno para grandes períodos) devem ser interpretados com a devida cautela. O ponto de maior desvio deverá ser investigado em detalhe, pois constitui um potencial *outlier* ou um indício de que a cauda da distribuição empírica é marginalmente mais pesada do que a modelada pela GPD.

Seleção do modelo

O teste da razão de verossimilhança entre o modelo GPD não estacionário e o modelo exponencial poderia ter sido utilizado para determinar qual o melhor modelo não estacionário para descrever adequadamente o comportamento da cauda das excedências

de PM_{10} acima de $45 \mu\text{g}/\text{m}^3$. No entanto, e tendo em conta que o modelo exponencial tende a subestimar as quantidades de interesse optou-se pelo modelo GPD não estacionário descrito.

Estimação de parâmetros de interesse

A partir do modelo GPD não estacionário ajustado às excedências de PM_{10} acima de $45 \mu\text{g}/\text{m}^3$, para o período de 2011–2020, em que o parâmetro de escala ($\log(\sigma)$) varia em função da área ardida padronizada, procedeu-se à estimativa das concentrações extremas esperadas, considerando uma área ardida de 550 ha. Embora a área ardida de 5500 ha tenha sido usada como cenário extremo na análise GEV (máximos mensais), no caso do modelo GPD — ajustado às excedências diárias — tal valor encontra-se muito acima da gama de calibração (≈ 62 desvios padrão acima da média histórica).

Para garantir a validade estatística e física das previsões, optou-se por utilizar uma área ardida de 550 ha, correspondente a cerca de 5,9 desvios padrão acima da média. Este valor representa um evento extremo plausível, compatível com episódios históricos de grande severidade, sem gerar extrapolações irrealistas nos parâmetros de escala e forma do modelo GPD.

Para este cenário, a área ardida foi padronizada ($z=5,87$), resultando nos seguintes parâmetros estimados do modelo:

$$\hat{\sigma} = \exp(2,160 + 0,127 * 5,87) = 18,27; \quad \hat{\xi} = 0,100$$

Período e níveis de retorno

A Tabela 23 apresenta os níveis de retorno estimados pelo modelo GPD não estacionário, correspondentes aos valores extremos de PM_{10} para diferentes períodos, considerando uma área ardida de 550 ha e os dias efetivamente observados entre junho e outubro.

Observa-se que os níveis de PM_{10} aumentam com o período de retorno, refletindo eventos mais raros e severos: por exemplo, espera-se que, em média, uma

concentração diária de $138 \mu\text{g}/\text{m}^3$ ocorra uma vez a cada de 10 anos e a concentração de $186 \mu\text{g}/\text{m}^3$, uma vez a cada 50 anos. Estes resultados, apesar de sobrestimarem as concentrações de PM_{10} , demonstram que podem ocorrer concentrações elevadas de PM_{10} , o que é relevante para a gestão e prevenção de episódios extremos de poluição do ar.

Tabela 23: Estimativas dos níveis de retorno de PM_{10} ($\mu\text{g}/\text{m}^3$) para diferentes períodos de retorno no modelo GPD não estacionário

Período de retorno (anos)	Nível de retorno ($\mu\text{g}/\text{m}^3$)
2	96,99
5	119,51
10	137,97
20	157,76
50	186,11

A Tabela 24 apresenta a probabilidade de excedência de PM_{10} para diferentes concentrações, considerando uma área ardida de 550 ha. Observa-se que a probabilidade diminui à medida que a concentração aumenta: por exemplo, 1,84% para $60 \mu\text{g}/\text{m}^3$ e 0,04% para $150 \mu\text{g}/\text{m}^3$. Estes resultados permitem quantificar a frequência esperada de dias com valores elevados de PM_{10} , sendo relevantes para a gestão de episódios extremos de poluição do ar.

Tabela 24: Probabilidades de excedência de PM_{10} ($\mu\text{g}/\text{m}^3$) no modelo GPD não estacionário

Concentração PM_{10} ($\mu\text{g}/\text{m}^3$)	Probabilidade de excedência $P(X > v)$
60	0,01840
80	0,00702
100	0,00291
120	0,00130
150	0,00043

A Tabela 25 apresenta os quantis extremos de PM_{10} estimados pelo modelo GPD não estacionário, considerando uma área ardida de 550 ha. Estes valores permitem identificar as concentrações raras e severas esperadas. Valores como o quantil 99%, próximo de $209,4 \mu\text{g}/\text{m}^3$, indicam a magnitude das concentrações mais elevadas que podem ocorrer em dias críticos. À semelhança do ocorrido nos níveis de retorno também se observa uma sobrestimação dos quantis extremos. No entanto, constituem elementos

indicadores para avaliar a exposição a episódios extremos de poluição e para apoiar decisões de gestão da qualidade do ar, permitindo priorizar ações preventivas e de mitigação.

Tabela 25: Quantis extremos estimados de PM_{10} ($\mu\text{g}/\text{m}^3$) no modelo GPD não estacionário

Quantil	Concentração estimada ($\mu\text{g}/\text{m}^3$)
90%	137,97
95%	157,76
99%	209,35

CONCLUSÕES

O presente estudo permitiu caracterizar de forma detalhada os eventos extremos de poluição por PM_{10} na região Norte e Centro de Portugal, com especial foco na época de incêndios florestais, utilizando abordagens estatísticas robustas da Teoria dos Valores Extremos. A análise compreendeu tanto modelos de máximos mensais (GEV) como de excedências acima de um limiar crítico (GPD), considerando versões estacionárias e não estacionárias, com incorporação de covariáveis ambientais relevantes (tempo, área ardida, temperatura e velocidade do vento).

Os resultados demonstram que os extremos de PM_{10} são eventos raros, mas de elevada intensidade, caracterizados por distribuições assimétricas à direita e caudas pesadas. O modelo GEV estacionário, concentrações máximas anuais de PM_{10} de cerca de $66 \mu\text{g}/\text{m}^3$ poderão ser ultrapassadas, em média, uma vez a cada 10 anos, enquanto eventos mais raros, como $119 \mu\text{g}/\text{m}^3$, correspondem ocorrem em média uma vez a cada 50 anos. No caso do modelo GEV, a modelação não estacionária revelou que a área ardida é o principal determinante dos máximos mensais de PM_{10} , enquanto a tendência temporal apresenta ligeira diminuição ao longo do período analisado. O modelo final selecionado — com parâmetro de localização dependente linearmente de tempo e área ardida, escala constante e parâmetro de forma fixo — apresentou o melhor compromisso entre parcimónia e qualidade de ajustamento, permitindo estimar níveis de retorno e quantis extremos de forma robusta.

Para a modelação por excedências, o modelo GPD estacionário demonstrou que concentrações de PM_{10} de cerca de $99 \mu\text{g}/\text{m}^3$ poderão ser ultrapassadas, em média, uma vez a cada 10 anos, enquanto eventos mais raros, como $129 \mu\text{g}/\text{m}^3$, correspondem a períodos de retorno de 50 anos. A análise não estacionária do GPD evidenciou que a área ardida exerce um efeito estatisticamente significativo sobre o parâmetro de escala, aumentando a variabilidade e a magnitude esperada dos picos de poluição. O parâmetro de forma positivo sugere a presença de cauda ligeiramente pesada, indicando que eventos extremos raros, mas severos não podem ser negligenciados.

A comparação entre modelos lineares e log-lineares para o parâmetro de escala do GPD mostrou que a parametrização log-linear é mais adequada, garantindo positividade, maior robustez estatística e melhor ajustamento global. Entre as covariáveis avaliadas, a área ardida destacou-se como o fator mais relevante, consolidando a relação direta entre a intensidade dos incêndios florestais e os níveis extremos de PM_{10} .

Os níveis de retorno, quantis extremos e probabilidades de excedência obtidos permitem uma avaliação detalhada do risco ambiental e da exposição humana. Estes resultados evidenciam que a gestão da qualidade do ar durante a época de incêndios deve considerar não apenas médias históricas, mas também os eventos extremos que, apesar de raros, têm elevado potencial de impacto sobre a saúde pública e os ecossistemas.

Em síntese, a integração de modelos GEV e GPD, estacionários e não estacionários, fornece uma ferramenta poderosa para compreender a dinâmica dos extremos de PM_{10} , identificando os principais fatores de risco e permitindo estimativas mais precisas de níveis críticos e períodos de retorno. Este estudo contribui significativamente para o planeamento ambiental, políticas de saúde pública e gestão de incêndios florestais, oferecendo uma base científica sólida para a antecipação e mitigação de episódios severos de poluição atmosférica.

BIBLIOGRAFIA

Agência Portuguesa do Ambiente. (2022, 3 de fevereiro). Poluição do ar. <https://apambiente.pt/ar-e-ruído/poluicao-do-ar>

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Akademiai Kiado.

Ambrósio, R. (2024). *Modelação estatística das temperaturas máximas na região do Alentejo* (Dissertação de mestrado, Universidade de Évora).

Augusto, S. (2020). Population exposure to particulate matter and related mortality due to the Portuguese wildfires in October 2017 driven by storm Ophelia. *Environment International*, 144, 106057. <https://doi.org/10.1016/j.envint.2020.106057>

Balkema, A. A., & de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5), 792–804. <https://doi.org/10.1214/aop/1176996548>

Beirlant, J., Goegebeur, Y., Segers, J., & Teugels, J. (2004). *Statistics of extremes: Theory and applications*. Wiley.

Brabson, B. B., & Palutikof, J. P. (2000). Extreme value analysis of daily maximum wind speeds over the British Isles. *International Journal of Climatology*, 20(2), 137–153. [https://doi.org/10.1002/\(SICI\)1097-0088\(200002\)20:2<137::AID-JOC466>3.0.CO;2-7](https://doi.org/10.1002/(SICI)1097-0088(200002)20:2<137::AID-JOC466>3.0.CO;2-7)

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.

Campos, C., Couto, F. T., Filippi, J.-B., Baggio, R., & Salgado, R. (2023). Modelling pyroconvection phenomenon during a mega-fire event in Portugal. *Atmospheric Research*, 290, 106776. <https://doi.org/10.1016/j.atmosres.2023.106776>

Carmo, M. F. (2022). The climatology of extreme wildfires in Portugal, 1980–2018: Contributions to forecasting and preparedness. *International Journal of Climatology*, 42(5), 3123–3146. <https://doi.org/10.1002/joc.7411>

Castillo, E., Hadi, A. S., Balakrishnan, N., & Sarabia, J. M. (2005). *Extreme value and related models with applications in engineering and science*. John Wiley & Sons.

Chavez-Demoulin, V., & Davison, A. C. (2012). Statistical modeling of extreme values. *Stochastic Environmental Research and Risk Assessment*, 26(7), 937–948. <https://doi.org/10.1007/s00477-011-0513-5>

- Chen, H., Samet, J. M., Bromberg, P. A., & Tong, H. (2021). Cardiovascular health impacts of wildfire smoke exposure. *Particle and Fibre Toxicology*, 18(1), Article 2. <https://doi.org/10.1186/s12989-020-00394-8>
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.
- Cooley, D. (2009). Extreme value analysis and the study of climate change. *Climatic Change*, 97(1–2), 77–83. <https://doi.org/10.1007/s10584-009-9627-x>
- Couto, F. T., Trigo, R. M., Turco, M., Rodrigues, M., & Pereira, M. G. (2022). Is Portugal starting to burn all year long? The transboundary fire in January 2022. *Atmosphere*, 13(9), Article 1677. <https://doi.org/10.3390/atmos13091677>
- Cruz, J., Belo-Pereira, M., Fonseca, A., & Santos, J. A. (2025). Spatial–temporal variability of hourly precipitation extremes in Portugal: Two case studies in major wine-growing regions. *International Journal of Climatology*.
- Davison, A. C., & Huser, R. (2015). Statistics of extremes. *Annual Review of Statistics and Its Application*, 2, 203–235. <https://doi.org/10.1146/annurev-statistics-010814-020137>
- Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B*, 52(3), 393–442.
- De Souza Fernandes Duarte, E. F. (2021). Evaluation of atmospheric aerosols in the metropolitan area of São Paulo simulated by the regional EURAD-IM model on high-resolution. *Atmospheric Pollution Research*, 12(2), 451–469. <https://doi.org/10.1016/j.apr.2020.12.006>
- Eastoe, E. F., & Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C*, 58, 25–45. <https://doi.org/10.1111/j.1467-9876.2008.00639.x>
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC.
- Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer.
- Faustini, A., Stafoggia, M., Forastiere, F., Caranci, N., Mallone, S., Chiusolo, M., ... Perrucci, C. A. (2015). Short-term effects of particulate matter on mortality during forest fires in Southern Europe: Results of the MED-PARTICLES Project. *Occupational and Environmental Medicine*, 72(5), 323–329. <https://doi.org/10.1136/oemed-2014-102198>
- Fisher, R. A., & Tippett, L. H. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180–190. <https://doi.org/10.1017/S0305004100015681>
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3), 423–453.
- Hänninen, O., Lanki, T., Kulmala, M., Jantunen, M., Koskentalo, T., Kalliokoski, P., ... Härkönen, J. (2009). Population exposure to fine particles and estimated excess mortality in Finland from an East European wildfire episode. *Journal of Exposure Science & Environmental Epidemiology*. <https://doi.org/10.1038/jes.2008.57>

Henriques-Rodrigues, L., Duarte, E., & Rodrigues, E. (2025). Modelling air quality during Portugal's fire season using extreme value theory. In *Advances in Mathematical Modelling in Science, Engineering and Social Sciences*. Springer-Birkhäuser.

Hosking, J. R. M. (1984). Moments or probability weighted moments? An example comparing two methods for estimating the parameters of the generalized extreme-value distribution. *Biometrika*, 71(2), 543–548. <https://doi.org/10.1093/biomet/71.2.543>

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society: Series B*, 52(1), 105–124.

Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press.

Intergovernmental Panel on Climate Change. (2021). *Climate change 2021: The physical science basis*. Cambridge University Press. <https://doi.org/10.1017/9781009157896>

Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8–12), pp. 1287–1304. Obtido de [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8)

Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737–744. <https://doi.org/10.1029/1999WR900294>

Monteiro dos Santos, D., de Oliveira, A. M., & Duarte, E. S. (2024). Compound dry-hot-fire events connecting Central and Southeastern South America: An unapparent and deadly ripple effect. *npj Natural Hazards*, 1, Article 32. <https://doi.org/10.1038/s44304-024-00031-w>

Moreira, A. O. (2020). *Mitigação de incêndios florestais através da gestão de ignições de origem humana* [Dissertação de mestrado, Universidade do Porto].

Naveau, P., Guillou, A., Cooley, D., & Diebolt, J. (2002). A nonparametric model for extreme events. *Extremes*, 5(3), 235–253.

Parente, J., Pereira, M. G., & Tonini, M. (2016). Space-time clustering analysis of wildfires: The influence of dataset characteristics, fire prevention policy decisions, weather and climate. [doi:10.1016/j.scitotenv.2016.03.055](https://doi.org/10.1016/j.scitotenv.2016.03.055)

Parente, J., Amraoui, M., Menezes, I., & Pereira, M. G. (2019). Drought in Portugal: Current regime, comparison of indices and impacts on extreme wildfires. *Science of the Total Environment*, 685, 150–173. <https://doi.org/10.1016/j.scitotenv.2019.05.163>

Pereira, M. G., Malamud, B. D., Trigo, R. M., & Alves, P. I. (2011). The history and characteristics of the 1980–2005 Portuguese rural fire database. *Natural Hazards and Earth System Sciences*, 11, 3343–3358. <https://doi.org/10.5194/nhess-11-3343-2011>

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, 3(1), pp. 119–131. [doi:10.1214/aos/1176343003](https://doi.org/10.1214/aos/1176343003)

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464. <https://doi.org/10.1214/aos/1176344136>

Slezakova, K., Morais, S., & Pereira, M. C. (2013). Forest fires in Northern region of Portugal: Impact on PM levels. *Atmospheric Research*, 127, 148–153. <https://doi.org/10.1016/j.atmosres.2012.07.012>

Sutanto, S. J., Vitolo, C., Di Napoli, C., D'Andrea, M., & Van Lanen, H. A. J. (2020). Heatwaves, droughts, and fires: Exploring compound and cascading dry hazards at the pan-European scale. *Environment International*, 134, 105276

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72(1), 67–90. <https://doi.org/10.1093/biomet/72.1.67>

Stephens, M. A. (1986). Goodness-of-fit techniques. (D. N. Mason, & M. P. Cox, Edits.) *Tests based on EDF statistics*, pp. 97–193.

Tarín-Carrasco, P., Augusto, S., Palacios-Peña, L., Ratola, N., & Jiménez-Guerrero, P. (2021). Impact of large wildfires on PM₁₀ levels and human mortality in Portugal. *Natural Hazards and Earth System Sciences*, 21(9), 2867–2880. <https://doi.org/10.5194/nhess-21-2867-2021>

Trigo, R. M., Sousa, P. M., Pereira, M. G., Rasilla, D., & Gouveia, C. M. (2016). Modelling wildfire activity in Iberia with different atmospheric circulation weather types: 94 MODELLING WILDFIRE ACTIVITY IN IBERIA. *International Journal of Climatology*, 36(7), pp. 2761–2778. doi:10.1002/joc.4511

Turco, M., San-Miguel-Ayanz, J., Bedia, J., Pereira, M. G., & Trigo, R. M. (2019). Climate drivers of the 2017 devastating fires in Portugal. *Scientific Reports*, 9, Article 50318. <https://doi.org/10.1038/s41598-019-50318-3>

Turco, M., Llasat, M. C., von Hardenberg, J., & Provenzale, A. (2013). Impact of climate variability on summer fires in a Mediterranean environment (northeastern Iberian Peninsula). *Climatic Change*, 116(3-4), pp. 665–678. doi:10.1007/s10584-012 0590-0

Turco, M., von Hardenberg, J., Libertá, G., Provenzale, A., & Trigo, R. M. (2017). On the key role of droughts in the dynamics of summer fires in Mediterranean Europe. *Scientific Reports*, 7, p. 81. doi:10.1038/s41598-017-00116-9

Vitolo, C., Di Napoli, C., Di Giuseppe, F., Cloke, H. L., & Pappenberger, F. (2019). Mapping combined wildfire and heat stress hazards to improve evidence-based decision making. *Environment International*, 127, 21–34. <https://doi.org/10.1016/j.envint.2019.03.008>

Viegas, D. X. (2018). Wildfires in Portugal. *Fire Research*, 2(1), 52. <https://doi.org/10.4081/fire.2018.52> ResearchGate

World Health Organization. (2021). WHO global air quality guidelines: Particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization. <https://www.who.int/publications/i/item/9789240034228>