

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**ANÁLISE DE VALORES EXTREMOS PARA NÍVEIS
PLUVIOMÉTRICOS EM BARCELOS**

Pedro Alexandre Gonzaga Rosário

Dissertação
Mestrado em Estatística e Investigação Operacional
Especialização em Estatística

2013

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



**ANÁLISE DE VALORES EXTREMOS PARA NÍVEIS
PLUVIOMÉTRICOS EM BARCELOS**

Pedro Alexandre Gonzaga Rosário

Dissertação
Mestrado em Estatística e Investigação Operacional
Especialização em Estatística

Dissertação orientada pela Prof. Doutora Maria Isabel Fraga Alves

2013

Agradecimentos

O meu primeiro agradecimento vai para a minha querida orientadora, a Professora Doutora Maria Isabel Fraga Alves. Quero agradecer todo o apoio que me forneceu, todas as respostas e esclarecimentos que me deu. Obrigado pela paciência, pela compreensão! Obrigado por ser assim! Ainda me recordo do meu nervosismo, no nosso primeiro encontro semanal, e de como este desapareceu instantaneamente. Foram encontros onde expus as minhas dúvidas, apresentei o meu trabalho, onde me transmitiu conhecimento, me deu a sua opinião, me incentivou, mas não só... Foram também momentos de conversa, de confiança, de amizade. Foi um prazer ter sido seu aluno ao longo de todo o mestrado. Começou com ME, depois ECS. Avancámos por MAR a dentro e cá continuaremos, assim espero! Sem a professora, esta dissertação não teria sido possível. Obrigado por acreditar em mim! Obrigado por tudo!

Quero também agradecer, e ao mesmo tempo pedir desculpa, aos meus pais, ao meu irmão, à minha avó e à minha bisavó, pela compreensão da minha ausência em tantas situações do dia-a-dia e em eventos familiares, durante este ano. Sem o vosso apoio teria sido mais difícil! Obrigado por tudo o que fazem por mim e por serem o motivo de eu ser assim, como sou hoje! Vovó, para ti vai o meu mais profundo agradecimento e a ti dedico esta dissertação. Obrigado por seres a melhor avó do mundo!

Não poderia, também, deixar de agradecer a todos os meus amigos pelo forte apoio que sempre transmitiram. Quero e tenho mesmo de destacar alguns. João, obrigado por tudo. Pela tua paciência para me ouvir, pela ajuda, pelos incentivos, pelas sugestões, pelas traduções, pelas correções, por seres um "dicionário de sinónimos"... Obrigado por seres a voz da razão, naqueles momentos mais difíceis, e por me fazeres ver que "a vida, muitas vezes, troca-nos os planos, sem sequer pedir! Há dias perdidos e outros sem fim. Mas, por vezes, a vida devolve aquilo que nos roubou!" És um exemplo de grande coragem! Obrigado, parceiro de mestrado... Mais do que isso, um grande amigo. Tânia "mimi", a minha alma gémea, a minha eterna confidente. Obrigado pelos conselhos, pela ternura que transmites, por teres sempre esse sorriso e por conseguires ter essa visão de que tudo é perfeito. Quase todos os dias me perguntaste como ia a tese! Obrigado por

essa preocupação constante. Obrigado por estares sempre aqui! Tânia Costa, obrigado pela amizade que cresceu entre nós. Obrigado pelas palavras de incentivo constante, pelo interesse que foste demonstrando por esta tese, pela paciência para me ouvir falar sobre o assunto, mesmo que não estivesses a entender nada!

Um agradecimento especial ao Serge pelas suas palavras confortantes, pela sua disponibilidade e pela paciência de me responder, sempre com prontidão, às minhas dúvidas sobre o LaTeX. Muito obrigado Serge, o meu antecessor nesta aventura, que é escrever uma dissertação.

Por último, mas muito importantes na minha vida, os meus lindos cães: Cindy, Luís Miguel e Tomás. Em dias de extremo cansaço ou de profundo desalento, em que não há nada que nos faça sorrir... Estão lá vocês! A vossa dedicação, o vosso ar ternurento, as brincadeiras e as vossas atitudes conseguem trazer a boa disposição e, tudo o que era escuro, desaparece, como num sopro. Cindy, já não estás aqui para poder presenciar este momento. Sinto muito a tua falta. Serás sempre a minha linda! Sempre que preciso de me acalmar, ou por stress ou por receio do desconhecido, continua e continuará a ser sempre em ti que penso. O teu olhar penetrante, o nosso carinho mútuo continua sempre tão perto! Um dia voltaremos a brincar! Obrigado pelos anos maravilhosos que pudemos partilhar.

Um pouco de cada um de vós está nesta tese. Obrigado! Um abraço fechado.

Pedro Rosário

À minha avó

Conteúdo

1	Introdução	1
2	A Teoria Assintótica de Valores Extremos	5
2.1	Introdução	5
2.2	Distribuições Limite do Máximo	7
2.3	Distribuição Generalizada de Valores Extremos	10
2.4	Distribuições Limite do Mínimo	16
2.5	Domínio de Atração para uma Distribuição	18
2.6	Distribuições Limite para os Excessos - POT	26
2.7	Distribuições Limite para Estatísticas de Ordem Elevada - PORT	29
2.7.1	Caso I: Caso Extremal	30
2.7.2	Caso II: Caso Central (Quantil)	33
2.7.3	Caso III: Caso Intermédio	34
3	Abordagem Paramétrica em Valores Extremos	35
3.1	Introdução	35
3.2	Método Gumbel ou Método dos Máximos Anuais	36
3.2.1	Estimadores de Máxima Verosimilhança dos Parâmetros	37
3.2.2	Estimadores dos Momentos Ponderados de Probabilidade dos Parâmetros	38
3.2.3	Estimação de Outros Parâmetros de Interesse em EVT	40
3.2.4	Intervalos de Confiança para os Parâmetros	42

3.2.5	Testes de Hipóteses para o Modelo GEV	43
3.2.6	Não-Estacionariedade: Sazonalidade e Tendência	50
3.3	Método POT	53
3.3.1	Escolha do <i>Threshold</i>	54
3.3.2	Estimadores de Máxima Verosimilhança dos Parâmetros	57
3.3.3	Estimadores dos Momentos Ponderados de Probabilidade dos Pa- râmetros	58
3.3.4	Estimação de Outros Parâmetros de Interesse em EVT	59
3.3.5	Intervalos de Confiança para os Parâmetros	61
3.3.6	Testes de Hipóteses para o Modelo GP	61
3.4	Método das Maiores Observações (MO)	65
3.4.1	Estimação do Modelo para as MO	67
4	Abordagem Semi-Paramétrica em Valores Extremos	71
4.1	Introdução	71
4.2	Condição de Regularidade de Segunda Ordem	72
4.3	Estimação do Índice de Valores Extremos	73
4.3.1	Estimador de Hill	74
4.3.2	Estimador de Pickands	75
4.3.3	Estimador dos Momentos	76
4.3.4	Estimador Generalizado de Hill	78
4.3.5	Estimador Mixed-Moment	78
4.4	Invariância e Não-Invariância dos Estimadores do EVI	80
4.5	Normalidade assintótica dos estimadores do EVI	81
4.6	Estimação semi-paramétrica de outros parâmetros de interesse em EVT . .	82
4.6.1	Testes para os Domínios de Atração	86

5	Estudo de Caso: Aplicação a Níveis Pluviométricos	91
5.1	Introdução	91
5.2	Análise Paramétrica dos Dados	94
5.2.1	Método dos Máximos Anuais	95
5.2.2	Método das Maiores Observações - Níveis Pluviométricos Anuais . .	148
5.2.3	Método POT	153
5.3	Análise Semi-Paramétrica dos Dados	177
5.3.1	Níveis Pluviométricos Máximos Anuais	177
5.3.2	Níveis Pluviométricos Máximos Mensais	187
6	Considerações finais e problemas em aberto	199
A	Referências Bibliográficas	203

Resumo

A Teoria de Valores Extremos (EVT) focou inicialmente a sua atenção no estudo do comportamento do máximo de sucessões de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). A publicação do Teorema de Tipos Extremos por Leonard Tippett, Sir Ronald Fisher e Boris Gnedenko, na primeira metade do século XX, permitiu concluir que esse comportamento pode ser bem modelado por uma das três distribuições max-estáveis - Gumbel, Fréchet e Weibull. Von Mises e Jenkinson definiram que a classe de distribuições max-estáveis pode ser expressa por uma única expressão, a distribuição Generalizada de Valores Extremos (GEV), dependente de apenas um único parâmetro, a menos de localização e escala, designado por Índice de Valores Extremos (EVI) e que dá a forma da distribuição.

A inferência estatística acerca de acontecimentos raros poderá, por exemplo, ser feita por dois caminhos: as abordagens paramétrica e semi-paramétrica. Na primeira metodologia supõe-se que há um modelo paramétrico que se ajusta à amostra de observações i.i.d. Existem alguns métodos neste tipo de abordagem, por exemplo, o Método dos Máximos Anuais, o Método POT (Peaks over Threshold) e o Método das Maiores Observações. Já a abordagem semi-paramétrica, pressupõe apenas que a distribuição subjacente aos dados em estudo pertence a algum domínio max-estável, mas não é proposto nenhum modelo paramétrico. O objetivo principal prende-se com a estimação do EVI, que será feita selecionando as observações da amostra que se encontram acima de um determinado nível aleatório (Peaks-Over-Random-Threshold - PORT).

A EVT é cada vez mais aplicada em diversas áreas, sendo a hidrologia uma delas. Neste campo é fundamental ter o conhecimento sobre precipitação elevada, a sua duração e intensidade. O planeamento de estratégias de impacto-contenção de inundações, principalmente em áreas densamente povoadas, bem como o desenvolvimento de estruturas hidroelétricas, são alvo da análise e modelação deste tipo de acontecimentos extremos.

Nesta tese será aplicada uma análise em valores extremos a uma amostra de níveis pluviométricos de Barcelos. Serão usadas as duas metodologias mencionadas anteriormente, com o objetivo de descrever e prever o comportamento de níveis elevados de precipitação,

nesta cidade. Em particular, serão estimados quantis extremos, níveis de retorno, períodos de retorno, probabilidades de excedência e limite superior do suporte. Na abordagem paramétrica serão usados os métodos de máxima verosimilhança e dos momentos ponderados de probabilidade para a estimação pontual, e o método profile log-likelihood para a estimação intervalar.

Palavras-chave: Teoria de Valores Extremos, Método dos Máximos Anuais, Método POT, Método das Maiores Observações, Abordagem Semi-Paramétrica, Precipitação.

Abstract

Extreme Value Theory (EVT) deals with the stochastic behavior of the extreme values in a process. The behavior of the maxima of independent and identically distributed (i.i.d.) random variables can be described by the three max-stable distributions - Gumbel, Fréchet and Weibull - as suggested by Leonard Tippett, Sir Ronald Fisher and Boris Gnedenko (1928). Von Mises and Jenkinson defined the Generalized Extreme Value (GEV) distribution, which comprises the three extreme value distributions, depending on the value of the shape parameter, called Extreme Value Index - EVI.

There are different ways of mapping observations of a rare event yielding alternative approaches to statistical inference on extreme values: the parametric and the semi-parametric methods. In a parametric approach, the main assumption regards the existence of a suitable class of models for describing the random variable attached to the process that is generating the data under study. There are different methods in this approach as the Block of Annual Maxima method, the Peaks-Over-Threshold (POT) method and the Largest Observations method. Following a semi-parametric approach, the only assumption made is that the underlying distribution function belongs to the max-domain of attraction of some extreme value distribution. The primary goal is to estimate the EVI. In this method, all the inference is based on a portion of the sample above a random level to be determined (Peaks-Over-Random-Threshold - PORT).

EVT can be applied to several areas. One of these is hydrology. Precipitation extremes, their duration and their intensity are of interest in the field of hydrology. The planning of flood impact-curbing strategies, especially in highly populated areas, as well as the development of hydroelectric structures, is closely linked to the analysis and modelling of this type of extreme event.

This thesis provides the application of extreme value analysis to rainfall data collected at Barcelos. The previous two approaches are applied to the data to describe the extremes of rainfall and to predict its future behavior. In particular, estimates are given for extreme quantiles, return levels, return periods, exceedance probability and right endpoint. In the parametric approach, it will be used the maximum likelihood or probability

weighted moments methods for estimation and the profile log-likelihood method for the corresponding confidence intervals.

Keywords: Extreme Value Theory, Block of Annual Maxima method, POT method, Largest Observations method, Semi-parametric approach, Rainfall.

Lista de Figuras

2.1	Cauda direita de uma distribuição F	15
2.2	Representação gráfica das funções densidade de uma distribuição standard Weibull($\alpha = 2$), Gumbel e Fréchet($\alpha = 2$)	16
2.3	Relação entre as funções de distribuição F e F_u	27
3.1	Máximos Anuais	36
3.2	Excedências de um nível elevado u	54
3.3	Maiores Observações	67
4.1	Comparação das variâncias assintóticas para os estimadores do EVI	79
5.1	Rio Cávado em Barcelos: no lado esquerdo, um dia habitual e no lado direito, um dia com inundações	91
5.2	Precipitação média anual em Portugal (1961-1990), in Miranda <i>et al.</i> (2006)	92
5.3	Precipitação média sazonal acumulada em Portugal (1961-1990): (a) inverno, (b) primavera, (c) verão e (d) outono, in Miranda <i>et al.</i> (2006)	93
5.4	Níveis pluviométricos máximos anuais, em Barcelos (1932-2008)	95
5.5	Gráfico de Autocorrelação Parcial	96
5.6	Boxplot e histograma dos níveis pluviométricos máximos anuais, em Barcelos (1932-2008)	96
5.7	QQ-Plot Exponencial e ME-plot dos níveis pluviométricos máximos anuais, em Barcelos (1932-2008)	98
5.8	QQ-Plot Gumbel dos níveis pluviométricos máximos anuais, em Barcelos (1932-2008)	99
5.9	Gráfico de correlação entre $\hat{Q}_{\gamma,\lambda,\delta}(p)$ e $Q_{\gamma,0,1}(p)$	101

5.10	QQ-Plot GEV dos níveis pluviométricos máximos anuais, em Barcelos (1932-2008)	101
5.11	Densidades dos modelos Gumbel e GEV estimados ajustados aos níveis pluviométricos máximos anuais	103
5.12	Intervalos de 95% de confiança para os parâmetros do modelo Gumbel, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais	104
5.13	Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos anuais	105
5.14	Intervalos de 95% de confiança para quantis extremos do modelo Gumbel, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais	106
5.15	Intervalos de 95% de confiança para níveis de retorno do modelo Gumbel, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais	108
5.16	Intervalos de 95% de confiança para os parâmetros do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais	110
5.17	Diagnóstico gráfico do ajustamento GEV aos níveis pluviométricos máximos anuais	111
5.18	Intervalos de 95% de confiança para quantis extremos do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais	112
5.19	Intervalos de 95% de confiança para níveis de retorno do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais	112
5.20	Ajustamento linear na localização, para os níveis pluviométricos máximos anuais	118
5.21	Níveis pluviométricos máximos mensais, em Barcelos (1932-2008)	120
5.22	Gráfico de Autocorrelação Parcial	120
5.23	Boxplot e histograma dos níveis pluviométricos máximos mensais, em Barcelos (1932-2008)	121
5.24	QQ-Plot Exponencial e ME-plot dos níveis pluviométricos máximos mensais, em Barcelos (1932-2008)	122

5.25	QQ-Plot Gumbel dos níveis pluviométricos máximos mensais, em Barcelos (1932-2008)	123
5.26	Gráfico de correlação entre $\hat{Q}_{\gamma,\lambda,\delta}(p)$ e $Q_{\gamma,0,1}(p)$ e QQ-Plot GEV dos níveis pluviométricos máximos mensais, em Barcelos (1932-2008)	124
5.27	Densidades dos modelos Gumbel e GEV estimados ajustados aos níveis pluviométricos máximos mensais	125
5.28	Intervalos de 95% de confiança para os parâmetros do modelo Gumbel, baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais	126
5.29	Intervalos de 95% de confiança para os parâmetros do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais	126
5.30	Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos mensais	127
5.31	Diagnóstico gráfico do ajustamento GEV aos níveis pluviométricos máximos mensais	127
5.32	Quantis elevados para os níveis pluviométricos máximos mensais, em Barcelos (Nascimento <i>et al.</i> , 2012)	128
5.33	Intervalos de 95% de confiança para níveis de retorno dos modelos Gumbel (linha superior) e GEV (linha inferior), baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais	130
5.34	Ajustamento linear na localização, para os níveis pluviométricos máximos mensais	134
5.35	Gráficos de Autocorrelação Parcial	136
5.36	Box-plots e histogramas dos níveis pluviométricos máximos anuais, no outono e inverno, em Barcelos (1932-2008)	137
5.37	QQ-Plot Exponencial e ME-plot dos níveis pluviométricos máximos anuais, no outono e inverno, em Barcelos (1932-2008)	137
5.38	QQ-Plots Gumbel e GEV dos níveis pluviométricos máximos anuais, no outono e inverno, em Barcelos (1932-2008)	139
5.39	Densidades dos modelos Gumbel e GEV estimados ajustados aos níveis pluviométricos máximos anuais, no outono e inverno	139

5.40	Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos anuais, no outono	144
5.41	Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos anuais, no inverno	144
5.42	Figura superior: Série temporal dos níveis pluviométricos máximos anuais no outono e no inverno; figura inferior: zoom dos níveis acima de 60 mm, com demarcação da zona entre 60 mm e 80 mm	147
5.43	Ajustamento linear na localização, para os níveis pluviométricos máximos anuais no outono (figura superior) e no inverno (figura inferior)	147
5.44	Os 5 maiores níveis pluviométricos diários, por ano, em Barcelos	148
5.45	Os 10 maiores níveis pluviométricos diários, por ano, em Barcelos	149
5.46	Diagnóstico gráfico do ajustamento GEV aos níveis pluviométricos máximos anuais, baseado nas 5 MO's	150
5.47	Diagnóstico gráfico do ajustamento GEV aos níveis pluviométricos máximos anuais, baseado nas 10 MO's	151
5.48	Ajustamento linear na localização, para os níveis pluviométricos máximos anuais, baseado nas 5 (figura superior) e 10 (figura inferior) MO's	151
5.49	ME-plot, sem e com retas ajustadas, para os níveis pluviométricos máximos anuais	153
5.50	Estimativas dos parâmetros de forma e escala versus <i>threshold</i>	154
5.51	Excedências de $u=62.5$, na amostra dos níveis pluviométricos máximos anuais	155
5.52	QQ-plot Exponencial dos excessos, referentes à amostra de níveis pluviométricos máximos anuais	156
5.53	Gráfico de correlação entre $\hat{Q}_{\gamma,\sigma}(p)$ e $Q_{\gamma,1}(p)$	157
5.54	QQ-plot GP dos excessos, referentes à amostra de níveis pluviométricos máximos anuais	157
5.55	Intervalos de 95% de confiança para os parâmetros dos modelos Exponencial e GP , baseados no profile log-likelihood	159
5.56	Diagnóstico gráfico do ajustamento Exponencial aos excessos	160
5.57	Diagnóstico gráfico do ajustamento GP($\gamma < 0$) aos excessos	160

5.58	Intervalos de 95% de confiança para quantis extremais baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais, através dos modelos exponencial (parte superior) e GP (parte inferior)	162
5.59	Intervalos de 95% de confiança para níveis de retorno baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais, através dos modelos exponencial (parte superior) e GP (parte inferior)	163
5.60	ME-plot, com retas ajustadas, para os níveis pluviométricos máximos mensais	166
5.61	Estimativas dos parâmetros de forma e escala versus threshold	167
5.62	Excedências de $u=34.5$, na amostra dos níveis pluviométricos máximos mensais	168
5.63	Gráfico de Autocorrelação Parcial	168
5.64	QQ-plots GP (com $\gamma = 0$ e $\gamma \neq 0$) dos excessos, referentes à amostra de níveis pluviométricos máximos mensais	169
5.65	Intervalos de 95% de confiança para os parâmetros dos modelos Exponencial e GP , baseados no profile log-likelihood	170
5.66	Diagnóstico gráfico do ajustamento Exponencial aos excessos	171
5.67	Diagnóstico gráfico do ajustamento GP($\gamma < 0$) aos excessos	171
5.68	Intervalos de 95% de confiança para níveis de retorno baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais, através dos modelos exponencial (parte superior) e GP (parte inferior)	174
5.69	Estimativas do EVI, para a amostra dos níveis pluviométricos máximos anuais	179
5.70	Estimativas do EVI sobre o mesmo gráfico, para a amostra dos níveis pluviométricos máximos anuais	179
5.71	Estimativas para os quantis extremais, para a amostra dos níveis pluviométricos máximos anuais	182
5.72	Estimativas para a probabilidade de excedência e período de retorno do nível 146 mm, para a amostra dos níveis pluviométricos máximos anuais . .	183
5.73	Estimativas para o limite superior do suporte, para a amostra dos níveis pluviométricos máximos anuais	185
5.74	Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão bilateral	186

5.75	Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão unilateral	187
5.76	Estimativas do EVI, para a amostra dos níveis pluviométricos máximos mensais	189
5.77	Estimativas do EVI sobre o mesmo gráfico, para a amostra dos níveis pluviométricos máximos mensais	189
5.78	Estimativas para os quantis extremais, para a amostra dos níveis pluviométricos máximos mensais	191
5.79	Estimativas para os níveis de retorno, para a amostra dos níveis pluviométricos máximos mensais	192
5.80	Estimativas para a probabilidade de excedência e período de retorno do nível 146 mm, para a amostra dos níveis pluviométricos máximos mensais .	193
5.81	Estimativas para o limite superior do suporte, para a amostra dos níveis pluviométricos máximos mensais	195
5.82	Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão bilateral	196
5.83	Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão unilateral	197

Lista de Tabelas

2.1	Relações entre os parâmetros da GEV (assinalados com *) e das distribuições Gumbel, Fréchet e Weibull	13
3.1	Percentis para a estatística de Kolmogorov-Smirnov, adaptada para a distribuição Gumbel	50
3.2	Percentis para as estatísticas de Cramér-von Mises e de Anderson-Darling, adaptadas para a distribuição Gumbel	50
3.3	Pontos críticos, obtidos por simulação, para a estatística de teste G_m^*	64
3.4	Pontos críticos simulados para a estatística de Kolmogorov-Smirnov adaptada para o modelo Exponencial com parâmetros desconhecidos.	65
3.5	Pontos críticos simulados para a estatística de Cramér-von Mises adaptada para o modelo GP com parâmetros desconhecidos.	66
3.6	Pontos críticos simulados para a estatística de Anderson-Darling adaptada para o modelo GP com parâmetros desconhecidos.	66
5.1	Estimativas MV e MPP dos parâmetros de localização e escala do modelo Gumbel, para os níveis pluviométricos máximos anuais	103
5.2	Estimativas pontuais MV e MPP e IC a 95% de quantis extremos do modelo Gumbel, para os níveis pluviométricos máximos anuais	106
5.3	Quantis empíricos para os níveis pluviométricos máximos anuais	107
5.4	Estimativas pontuais MV e MPP e IC a 95% de níveis de retorno do modelo Gumbel, para os níveis pluviométricos máximos anuais	107
5.5	Estimativas MV e MPP do período de retorno e probabilidade de exce-dência do nível 146 mm, no modelo Gumbel, para os níveis pluviométricos máximos anuais	108

5.6	Estimativas MV e MPP dos parâmetros de forma, localização e escala do modelo GEV, para os níveis pluviométricos máximos anuais	109
5.7	Estimativas pontuais MV e MPP e IC a 95% de quantis extremais do modelo GEV, para os níveis pluviométricos máximos anuais	110
5.8	Estimativas pontuais MV e MPP e IC a 95% de níveis de retorno do modelo GEV, para os níveis pluviométricos máximos anuais	111
5.9	Estimativas MV e MPP do período de retorno e probabilidade de excedência do nível 146 mm, no modelo GEV, para os níveis pluviométricos máximos anuais	113
5.10	Estimativas MV e MPP do limite superior do suporte, no modelo GEV, para os níveis pluviométricos máximos anuais	113
5.11	Estimativas dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos anuais	114
5.12	Estimativas MV dos modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos anuais	119
5.13	AIC e BIC para os modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos anuais	119
5.14	Estimativas dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos mensais	125
5.15	Quantis extremais empíricos e estimativas pontuais MV e MPP e IC a 95% para esses quantis, com modelos Gumbel e GEV para os níveis pluviométricos máximos mensais	129
5.16	Estimativas pontuais MV e MPP e IC a 95% para níveis de retorno, nos modelos Gumbel e GEV, para os níveis pluviométricos máximos mensais .	130
5.17	Estimativas MV e MPP do período de retorno e probabilidade de excedência do nível 146 mm nos modelos Gumbel e GEV, para os níveis pluviométricos máximos mensais	131
5.18	Resultados obtidos nos testes às hipóteses definidas em (5.11)	132
5.19	Resultados obtidos nos testes à qualidade do ajustamento Gumbel	133
5.20	Estimativas MV dos modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos mensais	134

5.21	AIC e BIC para os modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos mensais	135
5.22	Estimativas preliminares dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos anuais, no outono e inverno	138
5.23	Estimativas dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos anuais, no outono e inverno	140
5.24	Resultados obtidos nos testes às hipóteses definidas em (5.13) (outono)	141
5.25	Resultados obtidos nos testes à qualidade do ajustamento Gumbel (outono)	142
5.26	Resultados obtidos nos testes às hipóteses definidas em (5.13) (inverno)	142
5.27	Resultados obtidos nos testes à qualidade do ajustamento Gumbel (inverno)	143
5.28	Estimativas MV dos parâmetros para o modelo Gumbel, para os níveis pluviométricos máximos anuais, no outono e inverno	143
5.29	Estimativas pontuais MV e IC a 95% dos quantis extremais, para os níveis pluviométricos máximos anuais no outono e inverno	145
5.30	Estimativas pontuais MV e IC a 95% para os níveis de retorno, no outono e inverno	145
5.31	Estimativas MV do período de retorno e probabilidade de excedência do nível 146 mm, para os níveis pluviométricos máximos anuais, no outono e no inverno	146
5.32	Estimativas MV dos parâmetros para o modelo GEV e respectivos erros padrão, quando considerados os $k=1,5,10$ maiores níveis pluviométricos anuais	149
5.33	Estimativas MV de outros parâmetros para o modelo GEV, quando considerados os $k=1,5,10$ maiores níveis pluviométricos anuais	152
5.34	Estimativas dos parâmetros para os modelos Exponencial e GP ($\gamma \neq 0$), para os excessos na amostra de níveis pluviométricos máximos anuais	159
5.35	Estimativas pontuais e intervalares dos quantis extremais e níveis de retorno, para os níveis pluviométricos máximos anuais, obtidos pelo método POT	161
5.36	Estimativas do período de retorno e probabilidade de excedência do nível 146 mm, para os níveis pluviométricos máximos anuais, obtidos pelo método POT	162

5.37	Resultados obtidos nos testes às hipóteses definidas em (5.15)	164
5.38	Resultados obtidos nos testes às hipóteses definidas em (5.16)	165
5.39	Resultados obtidos nos testes à qualidade do ajustamento do modelo GP com $\gamma = 0$	165
5.40	Estimativas dos parâmetros para os modelos Exponencial e GP ($\gamma \neq 0$), para os excessos na amostra de níveis pluviométricos máximos mensais	170
5.41	Estimativas pontuais e intervalares dos quantis extremais, para os níveis pluviométricos máximos mensais, obtidos pelo método POT	172
5.42	Estimativas pontuais e intervalares dos níveis de retorno, para os níveis pluviométricos máximos mensais, obtidos pelo método POT	173
5.43	Estimativas do período de retorno e probabilidade de excedência do nível 146 mm, para os níveis pluviométricos máximos mensais, obtidos pelo método POT	174
5.44	Resultados obtidos nos testes às hipóteses definidas em (5.17)	176
5.45	Resultados obtidos nos testes às hipóteses definidas em (5.18)	176
5.46	Resultados obtidos nos testes à qualidade do ajustamento do modelo GP com $\gamma = 0$	176
5.47	Estimativas pontuais e intervalares para o EVI, quando $k = 56$, para os níveis pluviométricos máximos anuais	180
5.48	Estimativas pontuais para os quantis extremais ou níveis de retorno, quando $k = 56$, para os níveis pluviométricos máximos anuais	181
5.49	Estimativas pontuais para o período de retorno e probabilidade de excedência do nível 146 mm, quando $k = 56$, para os níveis pluviométricos máximos anuais	184
5.50	Estimativas pontuais e intervalares para o EVI, quando $k = 371$, para os níveis pluviométricos máximos mensais	190
5.51	Estimativas pontuais para os quantis extremais e níveis de retorno, quando $k = 371$, para os níveis pluviométricos máximos mensais	192
5.52	Estimativas pontuais para o período de retorno e probabilidade de excedência do nível 146 mm, quando $k = 371$, para os níveis pluviométricos máximos mensais	194

Lista de acrónimos e abreviaturas

- a.a.*: Amostra aleatória
AIC: Akaike Information Criterion
BIC: Bayesian Information Criterion
e.o.: Estatística ordinal ou Estatística de ordem
 \mathcal{ERV}_γ : Variação regular generalizada de índice γ (do inglês, *Extended Regular Variation*)
EVI: Índice de Valores Extremos (do inglês, *Extreme Value Index*)
EVT: Teoria de Valores Extremos (do inglês, *Extreme Value Theory*)
f.d.: Função de distribuição
f.d.p.: Função de densidade de probabilidade
GEV: Generalizada de Valores Extremos (do inglês, *Generalized Extreme Value*)
g.l.: Graus de Liberdade
GP: Generalizada de Pareto
IC: Intervalo de confiança
i.i.d.: Independente e identicamente distribuído
LAN: *Locally Asymptotically Normal*
ME-plot: Gráfico da função de excesso médio (do inglês, *Mean Excess plot*)
mm: Milímetros
MMA: Método dos Máximos Anuais
MMO: Método das Maiores Observações
MO: Maiores Observações
MPP: Momentos Ponderados de Probabilidade
MV: Máxima Verosimilhança
PORT: *Peaks Over Random Threshold*
POT: *Peaks Over Threshold*
q.c.: Quase certamente
 \mathcal{RV}_γ : Variação regular de índice γ (do inglês, *Regular Variation*)
sse: Se e só se
TRV: Teste de Razão de Verosimilhanças
v.a.: Variável aleatória

Capítulo 1

Introdução

Embora as atividades humanas possam provocar catástrofes muito graves para a vida do planeta, os maiores cataclismos que a ameaçam resultam de processos naturais que não os podemos impedir nem controlar. Essas catástrofes são inevitáveis, mesmo quando previsíveis. Vivemos num planeta vivo, em mudança desde que nasceu, ou seja, há 4.5 bilhões de anos. Muitos dos desastres que o atingem resultam desta instabilidade. Uma dessas grandes manifestações são as tempestades e a precipitação excessiva, que será, precisamente, o foco desta dissertação.

Por cima da crosta terrestre e dos oceanos estende-se uma camada externa móvel, a atmosfera. É composta por uma mistura de gases retida à volta da Terra pela gravidade, mas que se torna mais leve à medida que a altitude aumenta, antes de se perder no vazio sideral. Os raios de sol passam através deste invólucro gasoso para aquecer as superfícies sólidas e líquidas do planeta; uma pequena parte deste calor é refletida para cima, formando assim correntes de ar verticais. O calor do sol, associado a estas correntes atmosféricas, comanda assim o ciclo da água - evaporação e precipitação. Este processo contínuo faz circular a água entre os oceanos, a atmosfera e a terra, transformando-a em vapor e em gelo, e inversamente. Esta atividade dá origem a tempestades, por vezes, muito violentas. Os rios transbordam e arrasam as margens, inundando regiões de terras normalmente secas. Imparáveis, infiltrando-se por todo o lado, as inundações são quase sempre uma verdadeira calamidade para as atividades humanas. Em contrapartida, as condições atmosféricas podem privar de humidade, zonas habitualmente bem irrigadas. A seca destrói então toda a vegetação e tem efeitos devastadores na fauna, além de favorecer incêndios.

As nossas vidas são demasiado curtas, e o Homem apareceu há demasiado pouco tempo para que possamos entender como normais, os fenómenos que afetam a Terra. Este planeta foi formado por 4.5 bilhões de anos de atividade, como já referido. As civilizações humanas

têm apenas uma milionésima parte desse tempo. A duração média da vida de um terrestre é inferior a um século. Por isso, não podemos mais do que entrever essas manifestações naturais. Há muito que os seres humanos se consideram os senhores do Mundo, e ficam simultaneamente surpreendidos e horrorizados quando a Natureza se enfurece e ameaça uma cidade, um porto, plantações inteiras, ou seja, tudo o que esteja relacionado com as suas vidas. Uma vez assente o pó, socorridas as vítimas e limpos os destroços, esquecemos tudo, para conhecer novamente a surpresa e o horror quando a catástrofe se repete.

Não se consegue impedir essas manifestações naturais de acontecerem, mas o ser humano pode tentar prever a sua ocorrência. Muitos pensam que, por exemplo, a quantidade de água que cai em forma de precipitação não segue nenhuma regra, mas talvez as coisas não sejam bem assim. A Teoria de Valores Extremos (EVT, do inglês Extreme Value Theory) pode dar uma ajuda nessa interpretação. A EVT tem-se revelado cada vez mais importante em várias áreas onde é fundamental ter o conhecimento do comportamento de valores demasiado elevados (ou baixos). Compreender o padrão de ocorrências desses valores pode dar um auxílio na prevenção e, conseqüentemente, na redução dos potenciais efeitos catastróficos que poderão causar na Humanidade.

A EVT tornou-se um dos mais importantes ramos da estatística nos últimos 70 anos. A principal característica distintiva de uma análise em valores extremos é o facto de se pretender quantificar o comportamento estocástico de um processo com níveis extremamente elevados ou baixos, como já referido. Em particular, este tipo de análise, geralmente, requer a estimação de probabilidades de acontecimentos que são mais extremos do que qualquer um que já tenha sido observado, ou seja, a EVT permite a extrapolação para além da amostra disponível.

Cada vez mais, outras ciências recorrem a técnicas de valores extremos para obterem respostas fundamentais. O domínio de aplicação da EVT é bastante diversificado: na hidrologia, na meteorologia, na geologia, nos seguros, na economia e finanças, nas telecomunicações, nas engenharias civil, biomédica e aeronáutica, etc.

No campo da hidrologia, são considerados pontos chave a precipitação elevada, a sua duração e intensidade. As inundações, devido a elevados níveis pluviométricos, sempre foram um problema para a Sociedade, devido à grande perda humana, material e financeira. Quantas vezes não vemos, ouvimos ou lemos notícias como:

"Chuva intensa provoca derrocadas e inundações na Madeira."

"Ministro das Finanças diz que as condições meteorológicas nos primeiros três meses do ano prejudicaram a atividade da construção, o que prejudicou o investimento!"

In Agência Lusa, 07-06-2013

"Chuva intensa afecta produção agrícola."

In Jornal de Notícias, 18-05-2011

"Pelo menos 45 mortos e 50 desaparecidos por chuvas na Índia."

In UOL.NOTÍCIAS, 17-06-2013

"Chuvas mataram 1310 portugueses e deixam quase 42 mil desalojados, nos últimos 150 anos."

In O Público, 25-11-2012

Portanto, ter o conhecimento da frequência com que estes eventos extremos ocorrem, é de grande importância para que possa haver um trabalho de prevenção, para se evitar ou diminuir as perdas daí resultantes. A EVT tem auxiliado na predição destes valores. O planeamento de estratégias de impacto-contenção de inundações, principalmente em áreas densamente povoadas, bem como o desenvolvimento de estruturas hidroelétricas, estão intimamente associados à análise e modelação deste tipo de acontecimentos extremos. Com a EVT, por exemplo, poder-se-á estabelecer valores confiáveis para os níveis de precipitação e para os períodos de retorno desses níveis, definir um limite a partir do qual se poderá dizer que se está num estado de calamidade/desastre e quantificar a probabilidade disso acontecer. A EVT vem, assim, mudar a visão errada da Sociedade, de que os danos acontecem sempre devido a uma exceção, que ocorrem ao acaso, provocado por algum agente externo.

Na sequência da Tese de Doutoramento de Nascimento (Nascimento, 2009), foi publicado por Nascimento, Gamerman and Lopes (2012) uma abordagem Bayesiana semi-paramétrica à estimação de valores extremos. Nesse estudo são estimados quantis elevados, de probabilidades 0.95, 0.99, 0.999, 0.9999 e 0.99999, para os níveis pluviométricos na cidade de Barcelos. Nesta dissertação pretende-se aplicar uma análise em valores extremos alternativa a esse mesmo conjunto de observações. Um dos objetivos primordiais será identificar o comportamento da cauda direita da distribuição subjacente a estes níveis de precipitação. Consequentemente, poder-se-ão obter estimativas desses mesmos

quantis, que posteriormente poderão ser comparadas com as obtidas nesse estudo recente, e de outras quantidades importantes que poderão ajudar a perceber, um pouco melhor, o comportamento dos níveis de precipitação, nesta cidade portuguesa, que por se situar no norte de Portugal e numa zona predominantemente montanhosa, costuma ter invernos muito pluviosos.

Com vista à concretização destes objetivos, o capítulo dois desta tese será dedicado à apresentação dos resultados fulcrais da teoria de valores extremos. Os dois capítulos seguintes focarão duas abordagens possíveis em valores extremos, a paramétrica, no capítulo três, e a semi-paramétrica, no capítulo quatro. Em ambos serão evidenciados os tópicos mais importantes para cada uma destas metodologias. No capítulo cinco será efetuada a análise em valores extremos ao conjunto de dados referentes aos níveis pluviométricos em Barcelos. Serão aplicadas as metodologias apresentadas nos capítulos anteriores e, espera-se assim, obter respostas sobre o comportamento dos níveis elevados de precipitação, nesta região portuguesa, e conseqüentemente, ir ao encontro dos objetivos inicialmente propostos e acima apresentados. Por fim, no capítulo seis, serão feitas algumas considerações finais sobre o estudo realizado e indicados possíveis problemas em aberto que poderiam ser desenvolvidos.

Capítulo 2

A Teoria Assintótica de Valores Extremos

2.1 Introdução

A Teoria de Valores Extremos (EVT, do inglês Extreme Value Theory) é um ramo probabilista de suporte à Estatística que lida com situações em que os dados são inexistentes ou se existem são escassos, os chamados acontecimentos raros ou extremos. A EVT vai ajudar a descrever e a quantificar o comportamento desses acontecimentos, procurando estimar uma possível distribuição limite para os máximos (ou mínimos) da amostra composta por variáveis aleatórias (v.a.'s) independentes e identicamente distribuídas (i.i.d.).

Considere-se que os dados em estudo são observações provenientes de uma amostra de v.a.'s (X_1, X_2, \dots, X_n) i.i.d. com função distribuição (f.d.) F e função densidade de probabilidade (f.d.p.) f . Se se ordenar esta amostra por ordem crescente, as correspondentes estatísticas ordinais (e.o.'s) poderão ser denotadas por $(X_{1:n}, X_{2:n}, \dots, X_{n:n})$, com $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$. O i -ésimo elemento desta sequência, $X_{i:n}$, será a i -ésima estatística de ordem da amostra, com $i = 1, \dots, n$. Sendo assim, a primeira e a última estatísticas de ordem, $X_{1:n}$ e $X_{n:n}$, serão, respetivamente, o mínimo e o máximo de (X_1, X_2, \dots, X_n) .

Nesta dissertação, ir-se-á dar destaque ao máximo e por questões de comodidade este passará a ser representado por M_n , ou seja, $M_n = X_{n:n} = \max\{X_1, X_2, \dots, X_n\}$. A f.d. do máximo exprime-se facilmente em termos da f.d. F :

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

Contudo, este resultado não é útil quando não se conhece a f.d. F . Sendo assim, e porque muitas vezes o que se pretende é o máximo de um grande número de variáveis, pode-se modelar este máximo usando argumentos assintóticos. Se se observar o comportamento de F^n quando $n \rightarrow \infty$, constata-se que:

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

Tem-se desta forma que o máximo converge em distribuição para o limite superior do suporte, x^F , de F . Ou seja,

$$M_n \xrightarrow[n \rightarrow \infty]{d} x^F, \quad \text{com } x^F := \sup\{x : F(x) < 1\}.$$

Sendo assim, a distribuição assintótica de M_n é degenerada (pois só toma os valores 0 ou 1). Mas, para fazer qualquer tipo de inferência para o máximo é preciso que este tenha um comportamento assintótico não-degenerado. A solução será proceder a uma normalização linear de M_n , tal como se passa no caso da teoria assintótica para as somas, com o Teorema do Limite Central.

Teorema 2.1 (Teorema do Limite Central). *Considere-se uma sequência de v.a.'s i.i.d $X_1, X_2, \dots, X_n, \dots$, com $E[X_i] = \mu$ e $Var[X_i] = \sigma^2 < +\infty$. Então,*

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow[n \rightarrow \infty]{d} Z \sim \mathcal{N}(0, 1).$$

Efetuada um procedimento semelhante, o máximo normalizado terá a forma

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

para sequências reais $a_n > 0$ e b_n , tal que:

$$\frac{M_n - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} Y \sim G, \tag{2.1}$$

com G uma f.d. não-degenerada.

2.2 Distribuições Limite do Máximo

A f.d. G indicada em (2.1) é não-degenerada, então, para qualquer ponto x pertencente ao domínio de continuidade de G , tem-se:

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) &= \lim_{n \rightarrow \infty} P(M_n \leq a_n x + b_n) \\ &= \lim_{n \rightarrow \infty} F^n(a_n x + b_n) \\ &= G(x). \end{aligned} \tag{2.2}$$

Quando tal é possível (e, portanto, existem as sucessões reais $a_n > 0$ e b_n), diz-se que a f.d. F pertence ao domínio de atração da distribuição limite do máximo (G). Este conceito pode ser visto em detalhe na definição que se segue.

Definição 2.1 (Domínio de Atração). *Uma f.d. F pertence ao max-domínio de atração de G , e escreve-se $F \in \mathcal{D}(G)$, se existem duas sucessões reais $a_n > 0$ e b_n tais que, para todo o ponto x pertencente ao domínio de continuidade de G , se tem (2.2).*

Antes de se determinar que distribuições limite pode assumir a função G , vejamos o que são distribuições do mesmo tipo:

Definição 2.2 (Distribuições do mesmo tipo). *Duas distribuições F_1 e F_2 dizem-se do mesmo tipo se existem constantes reais $a > 0$ e b tais que*

$$F_2(ax + b) = F_1(x).$$

Quer isto dizer que, distribuições do mesmo tipo são a mesma a menos de parâmetros de localização e de escala.

Associado a este conceito, o teorema seguinte sobre a convergência de distribuições do mesmo tipo, vem demonstrar que as sucessões a_n e b_n não são únicas e que se uma f.d. F pertence a um determinado domínio de atração, então não pertence a mais nenhum.

Teorema 2.2 (Convergência de tipos de Khinchin). *Sejam F_1 e F_2 duas distribuições não-degeneradas.*

i. Suponha-se que existem f.d. F_n e sucessões reais $a_n > 0$, $\alpha_n > 0$, $b_n \in \mathbb{R}$ e $\beta_n \in \mathbb{R}$ tais que $F_n(a_n x + b_n)$ e $F_n(\alpha_n x + \beta_n)$ convergem fracamente para $F_1(x)$ e $F_2(x)$, respetivamente, isto é, $F_n(a_n x + b_n) \xrightarrow{w} F_1(x)$ e $F_n(\alpha_n x + \beta_n) \xrightarrow{w} F_2(x)$. Então, quando $n \rightarrow \infty$,

$$\frac{\alpha_n}{a_n} \rightarrow a, \quad \frac{\beta_n - b_n}{a_n} \rightarrow b \quad e \quad F_2(x) = F_1(ax + b).$$

(Isto é, (α_n, β_n) são assintoticamente equivalentes a (a_n, b_n) e F_1 e F_2 são distribuições do mesmo tipo.)

ii. Inversamente, se $\frac{\alpha_n}{a_n} \rightarrow a > 0$, $\frac{\beta_n - b_n}{a_n} \rightarrow b$ e $F_n(a_n x + b_n) \xrightarrow[n \rightarrow \infty]{w} F_1(x)$ então

$$F_n(\alpha_n x + \beta_n) \xrightarrow[n \rightarrow \infty]{w} F_1(ax + b).$$

Então quais são as distribuições limite para o máximo normalizado de uma amostra aleatória (a.a.)? Essa resposta foi inicialmente dada por Fisher and Tippett (1928). Gnedenko (1943) faz a primeira demonstração rigorosa deste resultado. Mais tarde, de Haan (1970) veio desenvolver o conceito de variação regular, o qual passou a ser usado como uma importante ferramenta analítica na teoria de valores extremos.

Teorema 2.3 (Teorema de Fisher-Tippett-Gnedenko). *Se existem sucessões reais $a_n > 0$ e b_n , tais que*

$$\lim_{n \rightarrow \infty} P \left[\frac{M_n - b_n}{a_n} \leq x \right] = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x),$$

para alguma f.d. G não-degenerada, então G é do mesmo tipo de uma das seguintes distribuições:

i. *Tipo I:* $G(x) = \Lambda(x) = \exp(-\exp(-x))$, para $x \in \mathbb{R}$;

ii. *Tipo II:* $G(x|\alpha) = \Phi_\alpha(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & x > 0, \alpha > 0; \end{cases}$

iii. *Tipo III:* $G(x|\alpha) = \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0, \alpha > 0, \\ 1, & x \geq 0. \end{cases}$

O teorema 2.3 afirma, portanto, que o máximo devidamente normalizado converge em distribuição para uma v.a. cuja distribuição será uma das identificadas por Tipo I, II e III. Estas três famílias de distribuição são conhecidas pelas Distribuições de Valores Extremos. Em particular, a distribuição do Tipo I é chamada de **Gumbel**, a de Tipo II de **Fréchet** e a de Tipo III de **Weibull** (de máximos). Cada uma destas famílias de distribuição pode ter os seus parâmetros de localização (λ) e de escala (δ). Adicionalmente, a família de distribuições de Fréchet e Weibull têm um parâmetro de forma (α). Se a localização e a escala forem diferentes de 0 e 1, respetivamente, então, nas expressões indicadas no teorema 2.3, bastará aplicar a transformação habitual à v.a., isto é:

$$\Lambda(x|\lambda, \delta) = \Lambda\left(\frac{x - \lambda}{\delta}\right); \Phi_\alpha(x|\lambda, \delta) = \Phi_\alpha\left(\frac{x - \lambda}{\delta}\right) \text{ e } \Psi_\alpha(x|\lambda, \delta) = \Psi_\alpha\left(\frac{x - \lambda}{\delta}\right),$$

com $\lambda \in \mathbb{R}$ e $\delta > 0$.

Estas distribuições também são denominadas de max-estáveis. Vejamos, em primeiro lugar, o significado desse conceito:

Definição 2.3 (Max-estabilidade). *Uma distribuição G diz-se max-estável se existem constantes reais $A_k > 0$ e B_k tais que*

$$G^k(x) = G(A_k x + B_k), \forall k.$$

Se existe uma distribuição limite para o máximo normalizado, esta terá que ser max-estável. Ora, como as únicas possibilidades são as três famílias referidas anteriormente, vem que estas são as únicas distribuições max-estáveis. A ideia da demonstração deste facto, passa por se considerar que se tem k a.a., cada uma de dimensão n . Então, poder-se-á considerar que M_{nk} é o máximo de k máximos M_n (em que cada um deles é o máximo de uma amostra de dimensão n), ou que M_{nk} é o máximo da amostra conjunta de dimensão $n \times k$. Então, considerando que a distribuição limite de $\frac{M_n - b_n}{a_n}$ é G , com $a_n > 0$ e $b_n \in \mathbb{R}$, para um n suficientemente elevado ter-se-á $P\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx G(x)$ (resultado (2.2)), ou equivalentemente, $P(M_n \leq x) \approx G(a_n x + b_n)$. Logo, para qualquer inteiro k , $n \times k$ também será suficientemente elevado e, como tal,

$$P\left(\frac{M_{nk} - b_{nk}}{a_{nk}} \leq x\right) \approx G(x),$$

ou seja,

$$P(M_{nk} \leq x) \approx G(a_{nk}x + b_{nk}). \quad (2.3)$$

Mas, M_{nk} é o máximo de k variáveis que têm a mesma distribuição de M_n e, assim,

$$P\left(\frac{M_{nk} - b_{nk}}{a_{nk}} \leq x\right) = \left[P\left(\frac{M_n - b_n}{a_n} \leq x\right) \right]^k \approx G^k(x),$$

ou equivalentemente,

$$P(M_{nk} \leq x) = [P(M_n \leq x)]^k \approx G^k(a_n x + b_n). \quad (2.4)$$

Segundo Fisher and Tippett (1928), estas duas perspectivas, (2.3) e (2.4), são equivalentes. As duas distribuições limite G e G^k são do mesmo tipo e as constantes normalizadoras são assintoticamente equivalentes (Teorema 2.2), isto é,

$$\frac{a_{nk}}{a_n} \xrightarrow{n \rightarrow \infty} A_k, \quad \frac{b_{nk} - b_n}{a_n} \xrightarrow{n \rightarrow \infty} B_k$$

e, assim,

$$G(x) = G^k(A_k x + B_k).$$

Portanto, tem-se que G é max-estável. E sendo G a distribuição limite do máximo normalizado, vem que G pertence a um dos três domínios de atração já referidos, Gumbel, Fréchet ou Weibull.

2.3 Distribuição Generalizada de Valores Extremos

Já foi referido que existem três possíveis distribuições limite para o máximo devidamente normalizado. Contudo, é possível trabalhar com uma família que abrange essas três possíveis famílias de distribuições limite. von Mises (1936) e Jenkinson (1955) unificaram as famílias Gumbel, Fréchet e Weibull na Generalizada de Valores Extremos (GEV) com f.d. dada por:

$$G(x|\gamma) = G_\gamma(x) = \begin{cases} \exp\left(- (1 + \gamma x)^{-1/\gamma}\right), & 1 + \gamma x > 0, \gamma \neq 0, \\ \exp\left(- \exp(-x)\right), & x \in \mathbb{R}, \gamma = 0, \end{cases} \quad (2.5)$$

onde γ é o parâmetro de forma, também denominado de Índice de Valores Extremos (EVI, do inglês Extreme Value Index). Pode ser obtida uma versão mais geral para a f.d. GEV, introduzindo dois parâmetros, um de forma ($\lambda \in \mathbb{R}$) e outro de escala ($\delta > 0$):

$$G_\gamma(x|\lambda, \delta) = G_\gamma\left(\frac{x - \lambda}{\delta}\right) = \begin{cases} \exp\left(- (1 + \gamma \left(\frac{x - \lambda}{\delta}\right))^{-1/\gamma}\right), & 1 + \gamma \left(\frac{x - \lambda}{\delta}\right) > 0, \gamma \neq 0, \\ \exp\left(- \exp\left(- \left(\frac{x - \lambda}{\delta}\right)\right)\right), & x \in \mathbb{R}, \gamma = 0. \end{cases} \quad (2.6)$$

As famílias de distribuições de valores extremos Fréchet e Weibull correspondem, respetivamente, aos casos $\gamma > 0$ e $\gamma < 0$, na distribuição GEV. Quando $\gamma = 0$, a família GEV

obtida é interpretada como o limite por continuidade quando $\gamma \rightarrow 0$, e que corresponde ao domínio de atração max-estável Gumbel. Vejamos, então, uma demonstração rápida destes três casos:

1. Caso $\gamma > 0$:

Neste caso, como já foi referido, as f.d.'s $G_\gamma(x)$ e $\Phi_\alpha(x)$ são do mesmo tipo. Considere-se, na expressão em (2.6), que $\lambda = 1$, $\delta = \gamma$ e $\gamma = \frac{1}{\alpha}$, com $\alpha > 0$, então:

$$\begin{aligned}
 G_\gamma(x | 1, \gamma) &= G_\gamma\left(\frac{x-1}{\gamma}\right) \\
 &= \exp\left(-\left(1 + \gamma\left(\frac{x-1}{\gamma}\right)\right)^{-\frac{1}{\gamma}}\right) \\
 &= \exp\left(-\left(1 + x - 1\right)^{-\frac{1}{\gamma}}\right) \\
 &= \exp\left(-x^{-\frac{1}{\gamma}}\right) \\
 &= \exp\left(-x^{-\alpha}\right) \\
 &= \Phi_\alpha(x)
 \end{aligned}$$

2. Caso $\gamma < 0$:

A demonstração desta situação é análoga ao caso anterior, contudo considera-se, na equação (2.6) que $\lambda = -1$, $\delta = -\gamma$ e $\gamma = -\frac{1}{\alpha}$, com $\alpha > 0$:

$$\begin{aligned}
 G_\gamma(x | -1, -\gamma) &= G_\gamma\left(\frac{x+1}{-\gamma}\right) \\
 &= \exp\left(-\left(1 + \gamma\left(\frac{x+1}{-\gamma}\right)\right)^{-\frac{1}{\gamma}}\right) \\
 &= \exp\left(-\left(1 - x - 1\right)^{-\frac{1}{\gamma}}\right) \\
 &= \exp\left(-(-x)^{-\frac{1}{\gamma}}\right) \\
 &= \exp\left(-(-x)^{-\alpha}\right) \\
 &= \Psi_\alpha(x)
 \end{aligned}$$

Logo $G_\gamma(x)$ e $\Psi_\alpha(x)$ são distribuições do mesmo tipo.

3. Caso $\gamma = 0$:

Neste caso, f.d. GEV pode ser obtida como o prolongamento por continuidade

de $G_\gamma(x)$ em $\gamma = 0$, que será o segundo ramo na expressão em (2.5). Calculando, então, os limites laterais, tem-se:

$$\begin{aligned}
\lim_{\gamma \rightarrow 0^+} G_\gamma(x) &= \lim_{\gamma \rightarrow 0^+} \exp\left(- (1 + \gamma x)^{-\frac{1}{\gamma}}\right) \\
&= \exp\left(- \lim_{\gamma \rightarrow 0^+} (1 + \gamma x)^{-\frac{1}{\gamma}}\right) \\
&= \exp\left(- \lim_{\gamma \rightarrow 0^+} \left(1 + \frac{x}{\frac{1}{\gamma}}\right)^{-\frac{1}{\gamma}}\right) \\
&= \exp\left(- \lim_{\gamma \rightarrow 0^+} \left(\left(1 + \frac{x}{\frac{1}{\gamma}}\right)^{\frac{1}{\gamma}}\right)^{-1}\right) \\
&= \exp\left(- (\exp(x))^{-1}\right) \\
&= \exp\left(- \exp(-x)\right)
\end{aligned}$$

$$\begin{aligned}
\lim_{\gamma \rightarrow 0^-} G_\gamma(x) &= \lim_{\gamma \rightarrow 0^-} \exp\left(- (1 + \gamma x)^{-\frac{1}{\gamma}}\right) \\
&= \exp\left(- \lim_{\gamma \rightarrow 0^-} (1 + \gamma x)^{-\frac{1}{\gamma}}\right) \\
&= \exp\left(- \lim_{\gamma \rightarrow 0^-} \left(1 + \frac{x}{\frac{1}{\gamma}}\right)^{-\frac{1}{\gamma}}\right) \\
&= \exp\left(- \lim_{\gamma \rightarrow 0^-} \left(1 + \frac{-x}{-\frac{1}{\gamma}}\right)^{-\frac{1}{\gamma}}\right) \\
&= \exp\left(- \exp(-x)\right)
\end{aligned}$$

Logo,

$$\lim_{\gamma \rightarrow 0^+} G_\gamma(x) = \lim_{\gamma \rightarrow 0^-} G_\gamma(x) = \lim_{\gamma \rightarrow 0} G_\gamma(x) = \exp\left(- \exp(-x)\right) = \Lambda(x)$$

Ou seja, para $\gamma = 0$, a distribuição GEV e a distribuição Gumbel são do mesmo tipo.

Se se considerar os modelos com parâmetros de localização e de escala, facilmente se verifica a existência de relações entre esses parâmetros na distribuição GEV e nas distribuições Fréchet, Weibull e Gumbel (tabela 2.1).

Podemos, então, reformular o teorema 2.3,

Teorema 2.4 (Teorema de Gnedenko). *Se existem sucessões reais $a_n > 0$ e b_n , tais que*

$$\lim_{n \rightarrow \infty} P \left[\frac{M_n - b_n}{a_n} \leq x \right] = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x),$$

para alguma f.d. G não-degenerada, então G é do mesmo tipo da distribuição GEV

$$G(x|\gamma) = G_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}), & 1+\gamma x > 0, \gamma \neq 0, \\ \exp(-\exp(-x)), & x \in \mathbb{R}, \gamma = 0. \end{cases}$$

Tabela 2.1: *Relações entre os parâmetros da GEV (assinalados com *) e das distribuições Gumbel, Fréchet e Weibull*

Distribuição	Relações com GEV
Gumbel	$\lambda = \lambda^* \Leftrightarrow \lambda^* = \lambda$
	$\delta = \delta^* \Leftrightarrow \delta^* = \delta$
Fréchet	$\lambda = \lambda^* - \delta^*/\gamma^* \Leftrightarrow \lambda^* = \lambda - \delta$
	$\delta = \delta^*/\gamma^* \Leftrightarrow \delta^* = \delta/\alpha$
	$\alpha = 1/\gamma^* \Leftrightarrow \gamma^* = 1/\alpha$
Weibull	$\lambda = \lambda^* - \delta^*/\gamma^* \Leftrightarrow \lambda^* = \lambda - \delta$
	$\delta = -\delta^*/\gamma^* \Leftrightarrow \delta^* = \delta/\alpha$
	$\alpha = -1/\gamma^* \Leftrightarrow \gamma^* = -1/\alpha$

Quando se está nas condições do teorema 2.4, diz-se que a f.d. F pertence ao domínio de atração de G_γ , e que pode ser escrito da forma $F \in \mathcal{D}(G_\gamma)$. Repare-se que este teorema não garante a existência de uma lei limite não-degenerada e, caso ela exista, também não indica qual é o tipo dessa distribuição limite, ou seja, não identifica o valor de γ . Este resultado é semelhante ao Teorema do Limite Central, como já tinha sido referido, contudo não são conhecidas as constantes normalizadoras a_n e b_n e existe um conjunto infinito de distribuições limite indexadas em γ . Vejam-se alguns exemplos:

Exemplo 2.1. *Sejam X_1, X_2, \dots v.a.'s independentes com distribuição exponencial standard $Exp(1)$, cuja f.d é dada por $F(x) = 1 - \exp(-x)$, com $x > 0$. Neste caso, considerando*

$a_n = 1$, $b_n = \log n$ e $x > 0$ tem-se:

$$\begin{aligned}
\lim_{n \rightarrow \infty} P \left[\frac{M_n - b_n}{a_n} \leq x \right] &= \lim_{n \rightarrow \infty} F^n(x + \log n) \\
&= \lim_{n \rightarrow \infty} \left[1 - \exp(- (x + \log n)) \right]^n \\
&= \lim_{n \rightarrow \infty} \left[1 - \exp(-x) \cdot n^{-1} \right]^n \\
&= \lim_{n \rightarrow \infty} \left[1 - \frac{\exp(-x)}{n} \right]^n \\
&= \exp(-\exp(-x)) \\
&= \Lambda(x) = G_{\gamma=0}(x).
\end{aligned}$$

Portanto, com estas sucessões a_n e b_n , a distribuição limite do máximo M_n é uma Gumbel, que corresponde ao $\gamma = 0$ na distribuição GEV.

Exemplo 2.2. Sejam X_1, X_2, \dots v.a.'s independentes com distribuição Fréchet standard, com f.d $\Phi_1(x) = \exp(-x^{-1})$, com $x > 0$. Neste caso, seja $a_n = n$ e $b_n = 0$, então,

$$\begin{aligned}
\lim_{n \rightarrow \infty} P \left[\frac{M_n - b_n}{a_n} \leq x \right] &= \lim_{n \rightarrow \infty} \Phi^n(nx) \\
&= \lim_{n \rightarrow \infty} \left[\exp(- (nx)^{-1}) \right]^n \\
&= \lim_{n \rightarrow \infty} \exp(- (nx)^{-1} \cdot n) \\
&= \exp(-x^{-1}) \\
&= \Phi_1(x) = G_{\gamma=1}(x).
\end{aligned}$$

Neste caso, o limite é um resultado exato independente de n . Tal facto deve-se à propriedade da max-estabilidade de F . A distribuição limite do máximo continua a ser uma Fréchet. O parâmetro de forma da GEV, neste exemplo, toma o valor $\gamma = 1$.

Exemplo 2.3. Sejam X_1, X_2, \dots v.a.'s independentes com distribuição Uniforme(0,1), com f.d $F(x) = x$, para $0 \leq x \leq 1$. Fixando $x < 0$ e supondo que $n > -x$, seja $a_n = n^{-1}$ e $b_n = 1$, então,

$$\begin{aligned}
\lim_{n \rightarrow \infty} P \left[\frac{M_n - b_n}{a_n} \leq x \right] &= \lim_{n \rightarrow \infty} F^n(n^{-1}x + 1) \\
&= \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n \\
&= \exp(x) \\
&= \Psi_{-1}(x) = G_{\gamma=-1}(x).
\end{aligned}$$

Obtem-se, assim, uma distribuição do tipo Weibull para o máximo. Na família GEV o índice extremal será $\gamma = -1$.

O parâmetro γ na distribuição GEV, como já foi indicado, é o parâmetro de forma e está diretamente relacionado com o peso da cauda direita da distribuição e, por esse motivo, também é denominado de índice de cauda. A cauda direita de uma f.d. F pode ser definida por

$$\bar{F}(x) := 1 - F(x) = P(X > x),$$

e, de uma forma genérica, a sua representação gráfica traduz-se na figura 2.1.

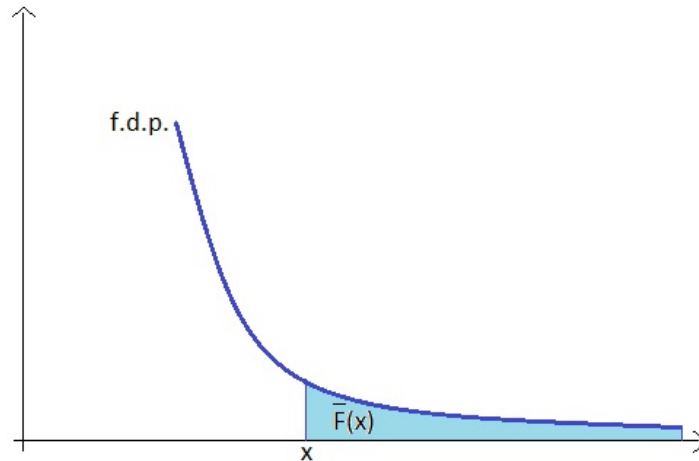


Figura 2.1: *Cauda direita de uma distribuição F*

Consoante o valor do parâmetro de forma, γ , poder-se-á ter uma noção do comportamento da cauda da distribuição, à medida que nos aproximamos do limite superior do suporte, x^F , isto é, da forma como esta decresce para zero.

No caso em que $\gamma < 0$, que corresponde ao max-domínio estável Weibull, as distribuições são caracterizadas por terem caudas curtas ou leves e têm limite superior do suporte finito. As distribuições Uniforme, Beta, Reversed Burr e Weibull de máximos são alguns exemplos que se incluem neste caso.

Quando $\gamma = 0$, correspondente ao domínio de atração Gumbel, a cauda da distribuição é do tipo exponencial e o limite superior do suporte pode ser finito ou infinito. São exemplos as distribuições Exponencial, Normal, Gama, Weibull de mínimos, Gumbel, Logística, entre outras.

Por fim, no caso em que $\gamma > 0$ (max-domínio Fréchet), as distribuições apresentam uma cauda pesada, do tipo polinomial. O limite superior do suporte é infinito. A este domínio max-estável pertencem distribuições como a Pareto, a Generalizada de Pareto, a Fréchet, a T-student, a Cauchy, a Log-Gama... Na figura 2.2 podem ser visualizados os três tipos de caudas referidas, pela mesma ordem que foram apresentadas.

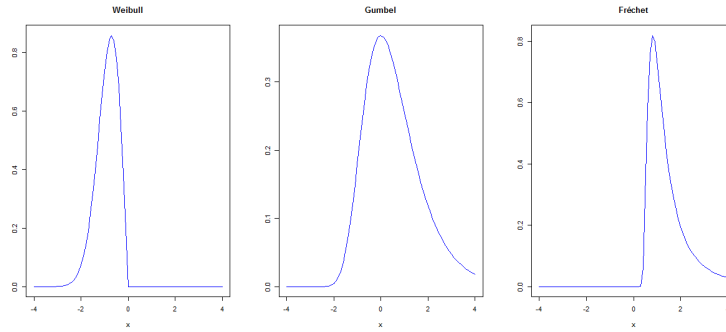


Figura 2.2: Representação gráfica das funções densidade de uma distribuição standard Weibull($\alpha = 2$), Gumbel e Fréchet($\alpha = 2$)

Mais recentemente, Neves and Fraga Alves (2008a) introduziram o conceito de distribuições com cauda super-pesada. Tratam-se de caudas de variação lenta, isto é,

$$\lim_{t \rightarrow \infty} \frac{\bar{F}(tx)}{\bar{F}(t)} = x^0 = 1, \text{ para } x > 0,$$

e escreve-se $\bar{F} \in \mathcal{RV}_0$. Nessas situações não é possível proceder à linearização do máximo de forma a obter uma distribuição limite não-degenerada e, como tal, essas distribuições não pertencem a nenhum dos max-domínios de atração. Distribuições como a Log-Pareto, Log-Weibull, Log-Cauchy e Log-Fréchet têm caudas direitas super-pesadas.

2.4 Distribuições Limite do Mínimo

Como já foi referido, nesta dissertação ir-se-á dar destaque ao máximo e, como tal, todos os resultados teóricos são referentes a este. Contudo, não poderia deixar de ser feito um breve apontamento à teoria assintótica para mínimos. Estes resultados são uma consequência imediata da Teoria de Valores Extremos para máximos, basta ter em consideração a seguinte dualidade:

$$m_n := \min(X_1, X_2, \dots, X_n) = -\max(-X_1, -X_2, \dots, -X_n)$$

Sendo assim,

$$\begin{aligned} P(m_n \leq x) &= P(-M_n \leq x) \\ &= P(M_n \geq -x) \\ &= 1 - P(M_n < -x) \\ &\stackrel{\text{v.a. contínua}}{=} 1 - P(M_n \leq -x) \end{aligned} \tag{2.7}$$

Ora, se estiverem reunidas as condições indicadas no teorema 2.3, a segunda parcela de (2.7) vai convergir para uma das três distribuições max-estáveis, Weibull, Gumbel ou Fréchet.

Pode ser evocado um resultado, tal como para o máximo, onde são definidos três domínios, neste caso, ditos de min-estáveis,

Teorema 2.5 (Teorema dos tipos extremais para mínimos). *Se existem sucessões reais $a_n^* > 0$ e b_n^* , tais que*

$$\lim_{n \rightarrow \infty} P \left[\frac{m_n - b_n^*}{a_n^*} \leq x \right] = G^*(x),$$

para alguma f.d. G^* não-degenerada, então G^* é do mesmo tipo de uma das seguintes distribuições:

i. *Tipo I:* $G^*(x) = \Lambda^*(x) = 1 - \Lambda(-x) = 1 - \exp(-\exp(x))$, para $x \in \mathbb{R}$;

ii. *Tipo II:* $G^*(x|\alpha) = \Phi_\alpha^*(x) = 1 - \Phi_\alpha(-x) = \begin{cases} 1 - \exp(x^{-\alpha}), & x < 0, \alpha > 0, \\ 1, & x \geq 0; \end{cases}$

iii. *Tipo III:* $G^*(x|\alpha) = \Psi_\alpha^*(x) = 1 - \Psi_\alpha(-x) = \begin{cases} 0, & x \leq 0, \\ 1 - \exp(-x^\alpha), & x > 0, \alpha > 0. \end{cases}$

Os três modelos apresentados são, respetivamente, Gumbel de mínimos, Fréchet de mínimos e Weibull de mínimos. Mas, também aqui, estes três modelos podem ser unificados numa só família, a distribuição GEV para mínimos:

$$G^*(x|\gamma) = G_\gamma^*(x) = 1 - G_\gamma(-x) = \begin{cases} 1 - \exp(-(1 - \gamma x)^{-1/\gamma}), & 1 - \gamma x > 0, \gamma \neq 0, \\ 1 - \exp(-\exp(x)), & x \in \mathbb{R}, \gamma = 0. \end{cases} \quad (2.8)$$

Mais uma vez, o parâmetro de forma γ dá a indicação do domínio min-estável com que se está a trabalhar. Assim, se $\gamma = 0$ o modelo é o Gumbel de mínimos. Para $\gamma < 0$, será o domínio Weibull de mínimos. Já se $\gamma > 0$, encontramos no caso da distribuição Fréchet de mínimos. Em todas estas distribuições poderão ser introduzidos parâmetros de localização e escala, $\lambda \in \mathbb{R}$ e $\delta > 0$ respetivamente.

Perante isto, quando o objetivo é fazer inferência sobre mínimos, poder-se-á aplicar diretamente os domínios min-estáveis e a GEV de mínimos. Em alternativa, poder-se-á fazer uso da dualidade existente entre as distribuições para máximos e para mínimos.

2.5 Domínio de Atração para uma Distribuição

Um dos principais problemas em EVT está relacionado com a identificação do domínio de atração de uma dada f.d. F . Ou seja, conhecendo a f.d. F , pretende-se encontrar as sucessões a_n e b_n , tais que se verifique a convergência em distribuição do máximo para uma v.a. com distribuição não-degenerada, e determinar esse limite. Equivalentemente, o problema pode ser encarado como, tendo uma distribuição limite, caracterizar o conjunto de distribuições F para as quais o máximo, convenientemente normalizado, converge para esse limite.

von Mises (1936) forneceu um conjunto de condições que garantem que uma dada distribuição F pertença a um determinado domínio de atração. Veja-se, em primeiro lugar, o conceito de função quantil de cauda que poderá ser incorporado nas referidas condições.

Definição 2.4 (Função quantil de cauda). *Seja F uma f.d. contínua com inversa*

$$F^{\leftarrow}(u) = \inf\{x : F(x) \geq u\}.$$

A correspondente função quantil de cauda é definida por

$$U(t) := F^{\leftarrow}(1 - 1/t), \quad t \in [1, \infty].$$

A função quantil de cauda é monótona não-decrescente e o seu valor em 1 e em ∞ é, respetivamente, o limite inferior e superior do suporte da distribuição F :

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

As condições de von Mises são apresentadas no teorema seguinte.

Teorema 2.6 (Condições suficientes de von Mises). *Para uma distribuição F absolutamente contínua e existindo $f = F'(x)$ e $F''(x)$, defina-se a função de hazard e o seu inverso, respetivamente, por*

$$h(x) := \frac{f(x)}{1 - F(x)} \quad e \quad r(x) := \frac{1 - F(x)}{f(x)}.$$

Se

$$\lim_{x \rightarrow x^F} r'(x) = \gamma$$

então

$$F \in \mathcal{D}(G_\gamma)$$

e tem-se que

$$b_n = F^{\leftarrow}(1 - 1/n) = U(n) \quad e \quad a_n = r(b_n) = \frac{1}{n \cdot f(b_n)} = n \cdot U'(n).$$

Repare-se que este teorema identifica, quase na totalidade, a distribuição limite do máximo de uma amostra com distribuição F , isto é, não só indica como obter as sucessões normalizadoras como também o valor do respetivo parâmetro de forma γ da distribuição GEV. Contudo, não podemos esquecer que estas condições são apenas suficientes, e não necessárias, e só se aplicam unicamente a distribuições absolutamente contínuas. Contudo, são muito fáceis de serem verificadas, requerendo apenas a existência das derivadas de primeira e segunda ordens da f.d. F .

O teorema 2.6 é a versão unificada das condições de von Mises, contudo, estas condições podem ser particularizadas a cada um dos três conjuntos já indicados, e que formam uma partição do espaço-parâmetro de γ .

Teorema 2.7 (Condições suficientes de von Mises). *Sob as condições referidas no teorema 2.6, suponha-se que:*

i. $x^F = \infty$ e que existe $F' = f$. Se para algum $\gamma > 0$,

$$\lim_{t \rightarrow \infty} t \cdot h(t) = \frac{1}{\gamma} = \alpha,$$

então $F \in \mathcal{D}(G_\gamma)$, com $\gamma > 0$, ou seja, $F \in \mathcal{D}(\Phi_\alpha)$;

ii. $x^F < \infty$ e que existe $F' = f$. Se para algum $\gamma < 0$,

$$\lim_{t \rightarrow x^F} (x^F - t) \cdot h(t) = -\frac{1}{\gamma} = \alpha,$$

então $F \in \mathcal{D}(G_\gamma)$, com $\gamma < 0$, ou seja, $F \in \mathcal{D}(\Psi_\alpha)$;

iii. existe $r(t)$ e que é diferenciável numa vizinhança de x^F . Se

$$\lim_{t \rightarrow x^F} r'(t) = 0,$$

então $F \in \mathcal{D}(G_0)$, ou seja, $F \in \mathcal{D}(\Lambda)$.

Como já foi referido, as condições de von Mises são apenas suficientes e não necessárias. Mais tarde, em 1943, Gnedenko vem resolver este problema, apresentando um novo conjunto de condições, necessárias e suficientes, para que uma distribuição $F \in \mathcal{D}(G_\gamma)$, com $\gamma \in \mathbb{R}$.

Teorema 2.8 (Teorema de Gnedenko (1943)).

i. $F \in \mathcal{D}(G_\gamma)$, com $\gamma > 0$ sse

$$x^F = \infty \quad e \quad \lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = \lim_{t \rightarrow \infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-1/\gamma}, \quad \forall x > 0;$$

ii. $F \in \mathcal{D}(G_\gamma)$, com $\gamma < 0$ sse

$$x^F < \infty \quad e \quad \lim_{t \rightarrow 0} \frac{1 - F(x^F - tx)}{1 - F(x^F - t)} = \lim_{t \rightarrow 0} \frac{\overline{F}(x^F - tx)}{\overline{F}(x^F - t)} = x^{-1/\gamma}, \quad \forall x > 0;$$

iii. $F \in \mathcal{D}(G_\gamma)$, com $\gamma = 0$ sse

$$x^F \leq \infty \quad e \quad \lim_{t \rightarrow x^F} \frac{1 - F(t + xg(t))}{1 - F(t)} = \lim_{t \rightarrow x^F} \frac{\overline{F}(t + xg(t))}{\overline{F}(t)} = \exp(-x), \quad \forall x \in \mathbb{R},$$

$$\text{com } \int_t^{x^F} (1 - F(s)) ds < \infty.$$

A função $g(t)$, indicada no item (iii) do teorema anterior, terá de ser monótona e positiva. Uma escolha possível é a função de excesso médio cuja expressão é dada por:

$$g(t) = E[X - t | X > t] := \frac{\int_t^{x^F} (1 - F(s)) ds}{1 - F(t)}, \quad \forall t < x^F. \quad (2.9)$$

Esta função assume um papel importante em EVT, devido ao padrão característico que apresenta para cada um dos três domínios max-estáveis, Gumbel, Fréchet e Weibull. Este tema será abordado mais à frente.

Gnedenko também apresenta um resultado em relação à escolha das constantes normalizadoras para que uma f.d. F pertença a um dos três domínios max-estáveis.

Teorema 2.9 (Constantes Normalizadoras para $F \in \mathcal{D}(\Phi_\alpha)$, $F \in \mathcal{D}(\Psi_\alpha)$ e $F \in \mathcal{D}(\Lambda)$).

Suponhamos que $F \in \mathcal{D}(G_\gamma)$, então:

i. para $\gamma > 0$,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \exp(-x^{-1/\gamma}) = \Phi_{1/\gamma}(x),$$

para $x > 0$, com

$$a_n = U(n) \quad e \quad b_n = 0;$$

ii. para $\gamma < 0$,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \exp(-(-x)^{-1/\gamma}) = \Psi_{-1/\gamma}(x),$$

para $x < 0$, com

$$a_n = x^F - U(n) \quad e \quad b_n = x^F;$$

iii. para $\gamma = 0$,

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \exp(-\exp(-x)) = \Lambda(x),$$

para $x \in \mathbb{R}$, com

$$a_n = g(U(n)) \quad e \quad b_n = U(n)$$

e $g(t)$ é a função de excesso médio definida em (2.9).

Como já foi referido, de Haan (1970) desenvolveu o conceito de funções de variação regular. Esta classe de funções é usada em diversos domínios da Matemática, sendo um deles a Teoria de Valores Extremos. Antes de serem apresentados alguns resultados importantes, vejamos-se primeiro algumas generalidades sobre este tipo de funções.

Definição 2.5 (Variação Regular). Uma função $f : \mathbb{R}^+ \rightarrow \mathbb{R}$, positiva e mensurável, que verifique

$$\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^\gamma, \quad x > 0, \quad \gamma \in \mathbb{R}$$

diz-se de variação regular de índice γ (no infinito) e pode denotar-se por $f \in \mathcal{RV}_\gamma$.

No caso de se verificar $\gamma = 0$, ou seja, $\lim_{t \rightarrow \infty} \frac{f(tx)}{f(t)} = x^0 = 1$, para $x > 0$, f diz-se de variação lenta (no infinito) e escreve-se $f \in \mathcal{RV}_0$, como já tinha sido apresentado no final da secção 2.3.

Se se observar a condição (i), do teorema 2.8, é fácil perceber que \bar{F} é de variação regular de índice $-1/\gamma$ ($\bar{F} \in \mathcal{RV}_{-1/\gamma}$).

Mas o conceito de variação regular poderá ter uma forma mais geral, surgindo assim, um outro conjunto de funções denominadas de Π -variação.

Definição 2.6 (Variação Regular Generalizada e Π -Variação). Seja $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ função mensurável e suponha-se que existe uma função real $a > 0$ tal que

$$\lim_{t \rightarrow \infty} \frac{f(tx) - f(t)}{a(t)} = \tau(x), \quad x > 0,$$

onde τ é uma função definida por:

$$\tau(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma}, & \gamma \neq 0, \\ \log x, & \gamma = 0. \end{cases}$$

Então, diz-se que f é de variação regular generalizada e escreve-se $f \in \mathcal{ERV}_\gamma$. A função a é chamada de função auxiliar de f . No caso particular de $\gamma = 0$, diz-se que f é de Π -variação ($f \in \Pi(a)$).

Usando estes conceitos de variação regular e de variação regular generalizada, juntamente com a função quantil de cauda, de Haan encontrou novas relações para que uma distribuição pertença a um determinado domínio de atração e elaborou novas versões das condições de von Mises e de Gnedenko. Muitas das demonstrações desses resultados poderão ser encontradas em de Haan (1976) e em de Haan and Ferreira (2006). Um teorema que está relacionado com a definição de domínio de atração apresentada na definição 2.1 é o seguinte:

Teorema 2.10 (de Haan and Ferreira, 2006, Theorem 1.1.2). *Sejam as sucessões reais $a_n > 0$ e b_n e considere-se a f.d. não-degenerada G . As seguintes afirmações são equivalentes:*

i.

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x),$$

para todo o ponto de continuidade x de G ;

ii.

$$\lim_{t \rightarrow \infty} t \cdot (1 - F(a(t)x + b(t))) = -\log G(x),$$

para todo o ponto de continuidade x de G , onde $0 < G(x) < 1$ e $a(t) = a_{[t]}$ e $b(t) = b_{[t]}$ (em que $[t]$ representa a parte inteira de t);

iii.

$$\lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} = D(x),$$

para todo o ponto de continuidade x de $D(x) = G^{-1}(\exp(-1/x))$, $a(t) = a_{[t]}$ e $b(t) = b_{[t]}$.

Ora $G(x) = G_\gamma(x)$ é a f.d. GEV, logo é fácil obter a sua inversa, a função quantil:

$$G_\gamma^{-1}(x) = \begin{cases} \frac{1}{\gamma(-\log x)^\gamma} - \frac{1}{\gamma}, & \gamma \neq 0, \\ -\log(-\log x), & \gamma = 0, \end{cases} \quad (2.10)$$

para $0 < x < 1$. Sendo assim, para todo o $x > 0$ a função $D(x)$ mencionada no item (iii) do teorema 2.10 toma a forma explícita

$$\begin{aligned} D(x) = G^{-1}(\exp(-1/x)) = D_\gamma(x) &= \begin{cases} \frac{1}{\gamma(-\log(\exp(-1/x)))^\gamma} - \frac{1}{\gamma}, & \gamma \neq 0 \\ -\log(-\log(\exp(-1/x))), & \gamma = 0 \end{cases} \\ &= \begin{cases} \frac{x^\gamma - 1}{\gamma}, & \gamma \neq 0 \\ \log x, & \gamma = 0. \end{cases} \end{aligned}$$

Então, o teorema anterior (teorema 2.10) poder ser reformulado, obtendo-se assim:

Teorema 2.11 (de Haan and Ferreira, 2006, Theorem 1.1.6). *Sejam $\gamma \in \mathbb{R}$ e G_γ a f.d. GEV apresentada em (2.5). As seguintes afirmações são equivalentes:*

i. *Existem duas constantes reais $a_n > 0$ e b_n , tais que*

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G_\gamma(x) \Leftrightarrow F \in \mathcal{D}(G_\gamma),$$

para todo o ponto de continuidade x de G_γ ;

ii. *Existe uma função positiva a tal que*

$$\lim_{t \rightarrow \infty} t \cdot (1 - F(a(t)x + U(t))) = -\log G_\gamma(x) = (1 + \gamma x)^{-1/\gamma},$$

para todo o ponto de continuidade x de G_γ e com $1 + \gamma x > 0$;

iii. *Existe uma função positiva a tal que, para $x > 0$ se tem*

$$\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma}, & \gamma \neq 0 \\ \log x, & \gamma = 0; \end{cases}$$

iv. *Existe uma função positiva g tal que*

$$\lim_{t \rightarrow x^F} \frac{\bar{F}(t + x \cdot g(t))}{\bar{F}(t)} = (1 + \gamma x)^{-1/\gamma}$$

para todo o x em que $1 + \gamma x > 0$.

Além disso, tem-se que $b_n = U(n)$, $a_n = a(n)$ e por (iv) tem-se que $g(t) = a\left(\frac{1}{\bar{F}(t)}\right)$.

Há uma relação evidente entre as condições (iii) dos teoremas 2.10 e 2.11, basta ter em atenção que:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} &= \lim_{t \rightarrow \infty} \frac{U(tx) - b(t) + b(t) - U(t)}{a(t)} \\ &= \lim_{t \rightarrow \infty} \frac{U(tx) - b(t)}{a(t)} - \frac{U(t) - b(t)}{a(t)} \\ &= D_\gamma(x) - D_\gamma(1) \\ &= D_\gamma(x) - 0 \\ &= D_\gamma(x). \end{aligned}$$

Note-se, também, que a condição (iii) do teorema 2.11 é equivalente a dizer que a função quantil de cauda é de variação regular generalizada. Sendo assim, $U \in \mathcal{ERV}_\gamma$ é uma condição necessária e suficiente para que $F \in \mathcal{D}(G_\gamma)$. de Haan chamou a este resultado condição de primeira ordem.

Teorema 2.12 (Condição de 1ª ordem).

$$F \in \mathcal{D}(G_\gamma), \gamma \in \mathbb{R} \quad \text{sse} \quad \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},$$

para alguma função positiva a , com $x > 0$. Quando $\gamma = 0$, o segundo membro é interpretado como $\log x$, correspondendo ao limite por continuidade.

Na condição (iv) do teorema 2.11 é referida a existência de uma função g positiva. de Haan demonstrou que g depende do sinal do parâmetro γ , podendo ser expressa por

$$g(t) = \begin{cases} \gamma t, & \gamma > 0 \\ -\gamma(x^F - t), & \gamma < 0 \\ \frac{\int_t^{x^F} (1-F(s))ds}{1-F(t)}, & \gamma = 0. \end{cases} \quad (2.11)$$

No caso de $\gamma = 0$, a sugestão para g é a função de excesso médio, já apresentada anteriormente. Usando a condição (iv) do teorema 2.11 e a função g em (2.11), de Haan criou, assim, uma versão alternativa das condições necessárias e suficientes de Gnedenko,

Teorema 2.13 (Condições necessárias e suficientes - versão alternativa de de Haan).

$$F \in \mathcal{D}(G_\gamma) \quad \text{sse} \quad \lim_{t \rightarrow x^F} \frac{\overline{F}(t + x \cdot g(t))}{\overline{F}(t)} = (1 + \gamma x)^{-1/\gamma}$$

para todo o x em que $1 + \gamma x > 0$ e g é uma função positiva. Para além disso, se o limite acima é verificado, então a função g pode ser dada pela expressão (2.11).

Em relação à nova versão unificada das condições de Von Mises, da autoria de Laurens de Haan, é obtida recorrendo ao conceito de função de quantil de cauda, $U(t)$:

Teorema 2.14 (Condições suficientes de von Mises, em termos de U). Sob as condições do teorema 2.6 se

$$\lim_{t \rightarrow x^F} r'(t) = \gamma,$$

o que é equivalente a

$$\lim_{t \rightarrow x^F} r(t) \cdot \frac{f'(t)}{f(t)} = -\gamma - 1$$

e que será o mesmo que

$$\lim_{t \rightarrow \infty} \frac{t \cdot U''(t)}{U'(t)} = \gamma - 1$$

então,

$$F \in \mathcal{D}(G_\gamma).$$

Como não poderia de deixar de ser, também existe uma readaptação do teorema 2.7, a versão específica destas condições para os domínios Fréchet, Weibull e Gumbel.

Teorema 2.15 (Condições suficientes de von Mises, em termos de U). *Sob as condições do teorema 2.6*

i. se

$$\lim_{t \rightarrow \infty} \frac{t \cdot U'(t)}{U(t)} = \gamma = \alpha,$$

então $F \in \mathcal{D}(G_\gamma)$, com $\gamma > 0$, ou seja, $F \in \mathcal{D}(\Phi_\alpha)$;

ii. se

$$\lim_{t \rightarrow \infty} \frac{t \cdot U'(t)}{U(\infty) - U(t)} = -\gamma = \alpha,$$

então $F \in \mathcal{D}(G_\gamma)$, com $\gamma < 0$, ou seja, $F \in \mathcal{D}(\Psi_\alpha)$;

iii. se

$$\lim_{t \rightarrow x^F} r(t) \cdot \frac{f'(t)}{f(t)} = -1, \text{ o que é equivalente a } \lim_{t \rightarrow \infty} \frac{t \cdot U''(t)}{U'(t)} = -1$$

então $F \in \mathcal{D}(G_0)$, ou seja, $F \in \mathcal{D}(\Lambda)$.

Como já foi referido, também as condições necessárias e suficientes de Gnedenko, apresentadas no teorema 2.8, têm uma versão em termos da função quantil de cauda U . Mas antes, veja-se um lema bastante útil que fornece um conjunto de condições necessárias para que uma f.d. $F \in \mathcal{D}(G_\gamma)$, recorrendo também à função U .

Lema 2.1 (Condições necessárias para $F \in \mathcal{D}(G_\gamma)$, em termos de U). *Se $F \in \mathcal{D}(G_\gamma)$, com $\gamma \in \mathbb{R}$, isto é, se $\lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma}$*

i. no caso de $\gamma > 0$, então

$$U(\infty) = \infty \quad e \quad \lim_{t \rightarrow \infty} \frac{U(t)}{a(t)} = \frac{1}{\gamma};$$

ii. no caso de $\gamma < 0$, então

$$U(\infty) < \infty \quad e \quad \lim_{t \rightarrow \infty} \frac{U(\infty) - U(t)}{a(t)} = -\frac{1}{\gamma};$$

iii. no caso de $\gamma = 0$, então

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = 1 \Leftrightarrow U \in \mathcal{RV}_0, \forall x > 0 \quad e \quad \lim_{t \rightarrow \infty} \frac{a(t)}{U(t)} = 0.$$

Além disso, se $U(\infty) < \infty$,

$$\lim_{t \rightarrow \infty} \frac{U(\infty) - U(tx)}{U(\infty) - U(t)} = 1, \forall x > 0 \quad e \quad \lim_{t \rightarrow \infty} \frac{a(t)}{U(\infty) - U(t)} = 0.$$

Tem-se ainda que

$$\lim_{t \rightarrow \infty} \frac{a(tx)}{a(t)} = 1, \forall x > 0,$$

isto é, $a \in \mathcal{RV}_0$.

Usando este Lema, poder-se-á, então, reformular as condições de Gnedenko quando o parâmetro de forma é diferente de zero:

Teorema 2.16 (Condições necessárias e suficientes para $F \in \mathcal{D}(G_\gamma)$, com $\gamma \neq 0$ em termos de U).

$F \in \mathcal{D}(G_\gamma)$, com $\gamma \neq 0$ sse

i. no caso de $\gamma > 0$,

$$\lim_{t \rightarrow \infty} \frac{U(tx)}{U(t)} = x^\gamma \Leftrightarrow U \in \mathcal{RV}_\gamma, \quad \forall x > 0;$$

ii. no caso de $\gamma < 0$,

$$U(\infty) < \infty \quad e \quad \lim_{t \rightarrow \infty} \frac{U(\infty) - U(tx)}{U(\infty) - U(t)} = \lim_{t \rightarrow \infty} \frac{x^F - U(tx)}{x^F - U(t)} = x^\gamma \Leftrightarrow x^F - U \in \mathcal{RV}_\gamma, \quad \forall x > 0.$$

Algumas das demonstrações destes resultados podem ser encontradas em de Haan (1976) e em de Haan and Ferreira (2006), como já referido anteriormente, e também em Vicente (2012). Em Beirlant *et al.* (2004) são apresentados diversos exemplos com aplicação direta destes resultados.

2.6 Distribuições Limite para os Excessos - POT

Nos subcapítulos anteriores foi apresentada a teoria sobre as distribuições limite do máximo. Contudo, por vezes poderá ser mais útil analisar v.a.'s que representem os excessos acima de um determinado nível extremo, chamado de *threshold*. Esta metodologia tem o nome de POT, do inglês Peaks Over Threshold. As observações que excedem um

dado *threshold* u são chamadas de excedências. Estas excedências têm um papel bastante importante em EVT e, como tal, o seu comportamento assintótico é de grande interesse.

Dada uma v.a. X com f.d. F e fixado um *threshold* u , o que se pretende estudar será, portanto, a v.a. dos excessos dada por $(X - u)|X > u$ ou, se se definir que $Y = X - u$, então, será $Y|Y > 0$. Mais em concreto, pretende-se caracterizar a distribuição dos excessos acima do nível u , a qual será denotada por F_u e assume a expressão

$$F_u(y) = P[X - u \leq y | X > u] = \frac{P(u < X \leq y + u)}{P(X > u)} = \frac{F(y + u) - F(u)}{1 - F(u)}, \quad (2.12)$$

com $0 \leq y \leq x^F - u$, sendo x^F o limite superior do suporte da v.a. X . Na figura 2.3 pode ser visualizada a representação da f.d. F_u à custa da f.d. F , embora com escalas diferentes.

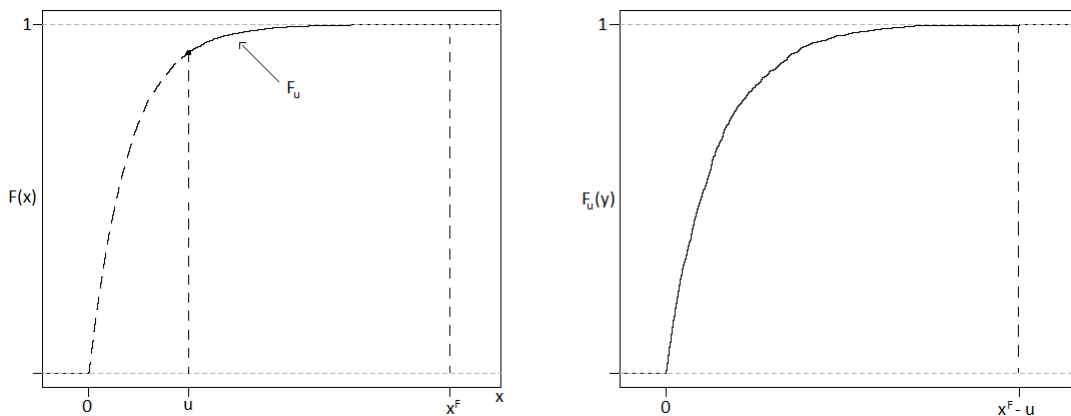


Figura 2.3: Relação entre as funções de distribuição F e F_u

Balkema and de Haan (1974) e Pickands (1975) provam que a distribuição Generalizada Pareto (GP) é uma aproximação para a distribuição dos excessos acima de um *threshold* elevado u .

A f.d. GP é dada por

$$H_\gamma(x|u, \sigma_u) := \begin{cases} 1 - \left(1 + \gamma \frac{x-u}{\sigma_u}\right)^{-1/\gamma}, & x \geq u & \text{se } \gamma > 0, \\ 1 - \exp\left(-\frac{x-u}{\sigma_u}\right), & x \geq u & \text{se } \gamma = 0, \\ 1 - \left(1 + \gamma \frac{x-u}{\sigma_u}\right)^{-1/\gamma}, & u \leq x \leq u - \frac{\sigma_u}{\gamma} & \text{se } \gamma < 0. \end{cases} \quad (2.13)$$

Para $\gamma > 0$, $\gamma = 0$ e $\gamma < 0$ a f.d. GP reduz-se à f.d. de um modelo Pareto, Exponencial e

Beta, respetivamente. O *threshold* u será o parâmetro de localização, σ_u o parâmetro de escala e γ o habitual parâmetro de forma.

A f.d. GP também poderá ser escrita em termos da variável $Y = X - u$, assumindo a forma

$$H_\gamma(y|\sigma_u) := \begin{cases} 1 - \left(1 + \gamma \frac{y}{\sigma_u}\right)^{-1/\gamma}, & y \geq 0 & \text{se } \gamma > 0, \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & y \geq 0 & \text{se } \gamma = 0, \\ 1 - \left(1 + \gamma \frac{y}{\sigma_u}\right)^{-1/\gamma}, & 0 \leq y \leq -\frac{\sigma_u}{\gamma} & \text{se } \gamma < 0. \end{cases} \quad (2.14)$$

Portanto, Pickands e Balkema and de Haan propuseram, para níveis elevados de u , a aproximação

$$F_u(y) \approx H_\gamma(y|\sigma_u), \text{ com } y = x - u > 0.$$

Este resultado é enunciado no teorema seguinte,

Teorema 2.17 (Teorema de Pickands-Balkema-de Haan).

$$F \in \mathcal{D}(G_\gamma), \gamma \in \mathbb{R} \quad \text{sse} \quad \lim_{u \rightarrow x^F} \sup_{0 < y < x^F - u} |F_u(y) - H_\gamma(y|\sigma_u)| = 0.$$

O Teorema 2.17 garante a existência de uma dualidade entre as distribuições GP e GEV, mais propriamente, se o máximo converge para uma distribuição $GEV(\gamma)$ e, como tal, a f.d. F pertence a algum dos max-domínios de atração, será equivalente a ter que a distribuição dos excessos acima de um nível u , dado que u é excedido, converge para uma distribuição $GP(\gamma)$. Este resultado encontra-se mais explícito no teorema que se apresenta de seguida, embora este seja equivalente ao teorema anterior. É de referir, também, que o parâmetro de forma γ é o mesmo nas distribuições GEV e GP e que o parâmetro de escala, σ_u , depende do nível u . Assim, a informação acerca do tipo de cauda direita da distribuição F , que era disponibilizada pelo sinal de γ , continua a ser idêntica à já mencionada.

Teorema 2.18.

$$F \in \mathcal{D}(G_\gamma), \gamma \in \mathbb{R} \quad \text{sse} \quad F_u(y) \approx H_\gamma(y|\sigma_u),$$

com $y = x - u > 0$ e H a f.d. GP definida em (2.13).

Uma demonstração deste teorema poderá ser encontrada em Coles (2001).

2.7 Distribuições Limite para Estatísticas de Ordem Elevada - PORT

Nas secções anteriores, com o objetivo de obter informações sobre o comportamento da cauda direita de uma distribuição, começou-se por estudar o máximo de uma a.a., $M_n = X_{n:n}$, e, de seguida, passou-se às excedências (e excessos) de um nível elevado u . Neste momento, focar-se-á o caso referente às k maiores observações de uma a.a. Quer isto dizer que se irá usar o conjunto de estatísticas

$$\{X_{n-k+1:n}, X_{n-k+2:n}, \dots, X_{n:n}\}$$

de uma a.a. ordenada. De certa forma, isso significa que se vai considerar um determinado nível, $X_{n-k+1:n}$, que será aleatório. A esta metodologia dá-se o nome de metodologia PORT, do inglês Peaks Over Random *Threshold*.

Tendo a k -ésima estatística de ordem, ou estatística ordinal, (e.o.), X_k , é possível conhecer a sua distribuição exata, usando argumentos combinatórios. A f.d. e a f.d.p de $X_{k:n}$ são apresentadas no teorema seguinte,

Teorema 2.19 (f.d. e f.d.p. exatas de $X_{k:n}$). *Seja (X_1, X_2, \dots, X_n) uma amostra de v.a.'s i.i.d. a X com f.d. F e f.d.p. f . As f.d e f.d.p de $X_{k:n}$ são dadas, respetivamente, por*

$$\begin{aligned} F_{X_{k:n}}(x) &= P(X_{k:n} \leq x) \\ &= P(\text{pelo menos } k \text{ das } n \text{ v.a.'s } X_i \leq x) \\ &= \sum_{i=k}^n \binom{n}{i} F^i(x) (1 - F(x))^{n-i} \end{aligned}$$

e

$$f_{X_{k:n}}(x) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x) (1 - F(x))^{n-k} f(x).$$

Nesta subcapítulo, o objetivo principal é o de estudar a distribuição assintótica da e.o. $X_{k:n}$, devidamente normalizada. Existe uma relação entre a ordem k e a dimensão da amostra n , dada por

$$\frac{k}{n} \approx p, \text{ quando } n \rightarrow \infty.$$

Esta relação pode assumir três vertentes diferentes. Uma primeira, a qual é chamada de **caso extremal**, é referente às situações em que k ou $n - k$ são fixos. Um segundo caso, o **caso central** ou **quantil**, é quando $0 < p < 1$ e, como tal, tem-se que k depende de n , $k \equiv k(n)$. Por fim, a terceira situação, o **caso intermédio**, que acontece quando se tem $p = 0$ ou $p = 1$. Também aqui k depende de n e, para além disso, tende para infinito, ou seja, $k = k(n) \rightarrow \infty$.

2.7.1 Caso I: Caso Extremal

Comece-se por constatar que o mínimo e o máximo são e.o.'s. que se incluem neste caso. Se $k = 1$ obtem-se $X_{k:n} = X_{1:n} = m_n$, ou, se $n - k = 0$ vem $X_{n-k+1:n} = X_{1:n} = m_n$, ou seja, o mínimo. Já se, $n - k = 0$ tem-se $X_{k:n} = X_{n:n} = M_n$, alternativamente, com $k = 1$ vem $X_{n-k+1:n} = X_{n:n} = M_n$, que será, portanto, o máximo. É possível usar os resultados anteriormente apresentados para o máximo, devidamente transformados, e com eles obter informações sobre a distribuição limite da e.o. $X_{n-k+1:n}$.

Teorema 2.20 (Distribuição assintótica da k -ésima e.o. superior, $X_{n-k+1:n}$). *Considere-se a e.o. $X_{n-k+1:n}$, com k fixo, então existem $a_n > 0$ e $b_n \in \mathbb{R}$ tais que $F \in \mathcal{D}(G)$, isto é,*

$$\lim_{n \rightarrow \infty} P(M_n \leq a_n x + b_n) = G(x),$$

com $G(x)$ não-degenerada, o que é equivalente a ter

$$\lim_{n \rightarrow \infty} P(X_{n-k+1:n} \leq a_n x + b_n) = G(x) \cdot \sum_{i=0}^{k-1} \frac{(-\log G(x))^i}{i!}.$$

As constantes normalizadoras para $X_{n-k+1:n}$ indicadas no teorema 2.20 são as mesmas que foram obtidas na normalização do máximo. Evidentemente que a f.d. G presente na distribuição assintótica da e.o. $X_{n-k+1:n}$ é a mesma que aparece no max-domínio de atração e será, portanto, ou Gumbel (Λ), ou Fréchet (Φ_α), ou Weibull (Ψ_α). Alternativamente, G também poderá ser a f.d. GEV, G_γ .

O resultado apresentado no teorema 2.20 justifica-se, essencialmente, devido ao número de excedências de um nível ter um comportamento limite Poisson. Vejam-se, sucintamente, alguns resultados que estão por detrás disso.

Lema 2.2. *Seja $\{\tau_n, n \geq 1\}$ uma sucessão de números reais. Então,*

$$\tau_n \xrightarrow[n \rightarrow \infty]{} \tau \Leftrightarrow \left(1 - \frac{\tau_n}{n}\right)^n \xrightarrow[n \rightarrow \infty]{} \exp(-\tau).$$

Considerando no lema anterior a sucessão $\tau_n = n \cdot (1 - F(u_n))$, tem-se

Lema 2.3. *Sejam $u_n \equiv u_n(\tau)$ níveis elevados. Então tem-se a seguinte equivalência:*

$$P[X_{n:n} \leq u_n] \xrightarrow[n \rightarrow \infty]{} \exp(-\tau) \Leftrightarrow n \cdot (1 - F(u_n)) \xrightarrow[n \rightarrow \infty]{} \tau.$$

Sendo assim, se se tiver $\{X_i\}_{i \geq 1}$ v.a.'s i.i.d com f.d. F e $u_n \equiv u_n(\tau)$ níveis elevados tais que $1 - F(u_n) = \frac{\tau}{n} + o\left(\frac{1}{n}\right)$, com $n \rightarrow \infty$, o número de excedências do nível u_n por X_i , com $1 \leq i \leq n$, pode ser expresso por

$$S_n := \#\{i : 1 \leq i \leq n, X_i > u_n\} = \sum_{i=1}^n \mathbf{1}_{\{X_i > u_n\}},$$

com

$$\mathbf{1}_A := \begin{cases} 1 & \text{se } A, \\ 0 & \text{c.c.} \end{cases}.$$

Quer isto dizer que S_n pode ser expressa como uma soma de v.a.'s Bernoulli independentes. E mais, como $P(X_i > u_n) = 1 - F(u_n)$, então tem-se que

$$S_n \sim \text{Binomial}(n, 1 - F(u_n)).$$

Ora, como $n \cdot (1 - F(u_n)) \xrightarrow{n \rightarrow \infty} \tau$, vem que S_n tem distribuição assintótica de Poisson(τ), ou seja,

$$\lim_{n \rightarrow \infty} P(S_n = k) = \frac{e^{-\tau} \tau^k}{k!}, \quad k = 0, 1, 2, \dots$$

Adaptando este resultado para a k -ésima e.o. de topo, $X_{n-k+1:n}$, obtém-se

$$\begin{aligned} P(X_{n-k+1:n} \leq u_n) &= P(\text{no máximo } k-1 \text{ das v.a.'s } X_i > u_n) \\ &= P(S_n \leq k-1) \\ &\xrightarrow{n \rightarrow \infty} e^{-\tau} \sum_{i=0}^{k-1} \frac{\tau^i}{i!} \end{aligned}$$

Substituindo, neste resultado, u_n por $a_n x + b_n$, τ_n por $n \cdot (1 - F(a_n x + b_n))$, $e^{-\tau}$ por $G(x)$ e, conseqüentemente, τ por $-\log G(x)$, ir-se-á obter o resultado apresentado no teorema 2.20:

$$P(X_{n-k+1:n} \leq a_n x + b_n) \xrightarrow{n \rightarrow \infty} G(x) \cdot \sum_{i=0}^{k-1} \frac{(-\log G(x))^i}{i!}, \quad \forall k \geq 1, \quad k \text{ fixo.}$$

Como é evidente, para se obter a f.d.p. limite da e.o. $X_{n-k+1:n}$, bastará derivar a f.d.:

$$\frac{d}{dx} \left\{ G(x) \cdot \sum_{i=0}^{k-1} \frac{(-\log G(x))^i}{i!} \right\} = g(x) \cdot \frac{(-\log G(x))^{k-1}}{(k-1)!}, \quad (2.15)$$

com g a f.d.p. de G .

No caso de se estar interessado na k -ésima e.o. inferior, com k fixo, o resultado assintótico correspondente é indicado no teorema seguinte.

Teorema 2.21 (Distribuição assintótica da k -ésima e.o. inferior, $\mathbf{X}_{k:n}$). *Considere-se a e.o. $X_{k:n}$, com k fixo, então existem $a_n^* > 0$ e $b_n^* \in \mathbb{R}$ tais que $F \in \mathcal{D}(G^*)$, isto é,*

$$\lim_{n \rightarrow \infty} P(m_n \leq a_n^* x + b_n^*) = G^*(x),$$

com $G^*(x)$ não-degenerada, o que é equivalente a ter

$$\lim_{n \rightarrow \infty} P(X_{k:n} \leq a_n^* x + b_n^*) = 1 - (1 - G^*(x)) \cdot \sum_{i=0}^{k-1} \frac{[-\log(1 - G^*(x))]^i}{i!}.$$

A f.d.p. limite é dada por

$$g^*(x) \cdot \frac{[-\log(1 - G^*(x))]^{k-1}}{(k-1)!}.$$

Os resultados apresentados nos teoremas 2.20 e 2.21 dão indicação das distribuições assintóticas para uma e.o. superior e inferior, respetivamente. Mas, por vezes, poder-se-á estar interessado no comportamento assintótico de um conjunto de e.o.'s superiores ou inferiores. Em modelação de acontecimentos raros, uma possível abordagem paramétrica consiste nisso mesmo, ou seja, na modelação conjunta das maiores ou menores observações de uma amostra. Como se constatou, a distribuição limite da k -ésima e.o. superior foi obtida considerando o número de excedências de um determinado nível u_n , por parte das v.a.'s X_1, \dots, X_n . Utilizando argumentos semelhantes, consegue-se chegar facilmente à distribuição conjunta assintótica das k e.o.'s superiores, quando k é fixo.

Teorema 2.22 (Distribuição assintótica das k e.o.'s superiores). $F \in \mathcal{D}(G)$, para constantes $a_n > 0$ e $b_n \in \mathbb{R}$, com $g = G'$ a f.d.p. de G

sse

o k -vetor $\left(\frac{X_{n:n} - b_n}{a_n}, \dots, \frac{X_{n-k+1:n} - b_n}{a_n} \right)$ tem distribuição limite com f.d.p. conjunta dada por

$$g_{X_{n:n}, \dots, X_{n-k+1:n}}(w_1, \dots, w_k) := G(w_k) \cdot \prod_{i=1}^k \frac{g(w_i)}{G(w_i)}, \quad \text{para } w_1 > \dots > w_k.$$

Como não poderia deixar de ser, embora nesta a dissertação o destaque é dado aos grandes valores, também existe um resultado para a distribuição conjunta limite das k menores e.o.'s.

Teorema 2.23 (Distribuição assintótica das k e.o.'s inferiores). $F \in \mathcal{D}(G^*)$, para constantes $a_n^* > 0$ e $b_n^* \in \mathbb{R}$, com $g^* = G^{*'}$ a f.d.p. de G^*

sse

o k -vetor $\left(\frac{X_{1:n} - b_n^*}{a_n^*}, \dots, \frac{X_{k:n} - b_n^*}{a_n^*} \right)$ tem distribuição limite com f.d.p. conjunta dada por

$$g_{X_{1:n}, \dots, X_{k:n}}^*(w_1, \dots, w_k) := (1 - G^*(w_k)) \cdot \prod_{i=1}^k \frac{g^*(w_i)}{1 - G^*(w_i)}, \quad \text{para } w_1 < \dots < w_k.$$

2.7.2 Caso II: Caso Central (Quantil)

Como já foi referido, neste caso a e.o. central $X_{k:n}$ não tem o valor de k fixado. Este depende de n , $k \equiv k(n)$, e é tal que $k(n) \xrightarrow{n \rightarrow \infty} \infty$ e $\frac{k(n)}{n} \xrightarrow{n \rightarrow \infty} p \in (0, 1)$.

A distribuição limite da e.o. central $X_{k:n}$ é apresentada no teorema seguinte.

Teorema 2.24 (Distribuição assintótica da e.o. central (Quantil)).

Seja $0 < p < 1$, $k \equiv k(n) \approx np$, quando $n \rightarrow \infty$, por exemplo, $k := [np] + 1$, e F f.d. contínua com f.d.p. f . Denote-se o quantil- p por $\chi_p := F^{\leftarrow}(p)$, com $0 < f(\chi_p) < \infty$. Então, quando $n \rightarrow \infty$,

$$\sqrt{n}(X_{k:n} - \chi_p) \xrightarrow{d} W \sim \mathcal{N}\left(0, \frac{p(1-p)}{[f(\chi_p)]^2}\right)$$

ou, equivalentemente,

$$\frac{\sqrt{n}f(\chi_p)}{\sqrt{p(1-p)}}(X_{k:n} - \chi_p) \xrightarrow{d} Z \sim \mathcal{N}(0, 1).$$

Portanto, na zona central de uma amostra, o comportamento limite de $X_{k:n}$ é bem aproximado por uma distribuição Gaussiana e depende do quantil de probabilidade p e do valor da densidade f nesse quantil. Este resultado é bastante útil se se pretender estimar quantis centrais. No entanto, não existe qualquer relação entre o comportamento das maiores e.o.'s e dos quantis longe das caudas da distribuição. Porém, em análise de valores extremos, os quantis são, por vezes, utilizados em simultâneo com as e.o.'s do caso extremal. Um exemplo disso, são as estatísticas envolvendo os excessos das observações acima de quantis empíricos (abordagem PORT), unicamente através de ratios do tipo

$$\frac{X_{i:n} - X_{[np]:n}}{X_{j:n} - X_{[np]:n}}, \quad i, j \geq [np].$$

Estas situações são bastante atrativas pois gozam da propriedade de ancilaridade, ou seja, são invariantes perante transformações de localização e escala.

No caso de se considerar um conjunto de e.o.'s centrais, a sua distribuição assintótica continua a ser Normal mas, neste caso, multivariada.

Teorema 2.25 (Distribuição assintótica conjunta dos quantis).

Sejam $0 < p_1 < p_2 < \dots < p_r < 1$ e $k_i := [np_i] + 1$, com $i = 1, 2, \dots, r$. Considere-se, também, F f.d. contínua com f.d.p. f . Denote-se para cada i o quantil- p_i por $\chi_{p_i} := F^{\leftarrow}(p_i)$, com $0 < f(\chi_{p_i}) < \infty$. Então, considerando as r e.o.'s centrais, quando $n \rightarrow \infty$, a distribuição conjunta de

$$\sqrt{n}(X_{k_1:n} - \chi_{p_1}), \dots, \sqrt{n}(X_{k_r:n} - \chi_{p_r})$$

é assintoticamente r -dimensional Normal Multivariada com vetor de valores médios nulo e matriz de covariâncias $[\sigma_{ii'}]$, em que,

$$\sigma_{ii'} := \frac{p_i(1-p_{i'})}{f(x_{p_i}) \cdot f(x_{p_{i'}})}, \quad 1 \leq i \leq i' \leq r.$$

Através deste teorema, constata-se que duas e.o.'s centrais são dependentes, mesmo assintoticamente. Ora, esta situação vem contrastar com o que acontece no caso extremal, em que há independência assintótica entre as e.o.'s inferiores e superiores.

2.7.3 Caso III: Caso Intermédio

Neste caso, tal como no caso central, a ordem k depende de n e é tal que $k \equiv k(n) \rightarrow \infty$, com $n \rightarrow \infty$. Contudo, aqui tem-se que a sucessão intermédia k tende para infinito, mas com uma velocidade inferior à de n , $k = o(n)$, isto é,

$$\frac{k}{n} \rightarrow 0.$$

Nesta situação, $X_{n-k+1:n}$ tem o nome de e.o. intermédia superior e $X_{k:n}$ será a e.o. intermédia inferior. No teorema que se segue é indicada qual a distribuição limite da e.o. intermédia superior, bem como os resultados que permitem encontrar as constantes normalizadoras a_n e b_n .

Teorema 2.26 (Distribuição assintótica da e.o. intermédia).

Suponha-se que as condições de von Mises (teorema 2.6) se verificam para algum domínio de atração. Denote-se $f := F'$. Então, se $k \equiv k(n) \rightarrow \infty$, com $\frac{k}{n} \rightarrow 0$ quando $n \rightarrow \infty$, tem-se que

$$\frac{X_{n-k+1:n} - b_n}{a_n} \xrightarrow[n \rightarrow \infty]{d} Z \sim \mathcal{N}(0, 1).$$

Para além disso, as constantes normalizadoras são dadas por

$$b_n = F^{\leftarrow} \left(1 - \frac{k}{n} \right) = U \left(\frac{n}{k} \right) \quad e \quad a_n = \frac{\sqrt{k}}{n \cdot f(b_n)} = \frac{\frac{n}{k} \cdot U' \left(\frac{n}{k} \right)}{\sqrt{k}}.$$

Em Beirlant *et al.* (2004) e Embrechts *et al.* (1997) encontram-se diversos exemplos de aplicação de cada um dos três casos apresentados, bem como algumas demonstrações de resultados.

Capítulo 3

Abordagem Paramétrica em Valores Extremos

3.1 Introdução

No capítulo anterior apresentou-se o resultado fundamental em EVT onde se estipula que as distribuições max-estáveis são as únicas que modelam o comportamento limite do máximo de uma a.a., devidamente normalizado (teorema 2.3). Estas distribuições podem ser unificadas numa única família, a GEV (teorema 2.4). Foi também evidenciado o papel fundamental do EVI, γ , que dá a forma diferenciada do tipo de cauda direita da distribuição limite. Se o valor deste parâmetro for positivo, a distribuição pertencerá ao domínio Fréchet (cauda pesada), se for negativo, já será ao domínio Weibull (cauda curta) e se for nulo pertencerá ao domínio Gumbel (cauda exponencial). Será, portanto, necessário proceder-se à estimação do EVI, γ , conjuntamente com os outros parâmetros da GEV, nomeadamente, os parâmetros de localização λ e de escala δ . Para além destes três parâmetros associados diretamente à distribuição GEV, existem outras quantidades bastante importantes em Acontecimentos Raros como os níveis e períodos de retorno, os quantis elevados, a probabilidade de excedência de níveis elevados e o limite superior do suporte (somente no caso de $\gamma < 0$). Mais à frente serão apresentados estes conceitos e, facilmente, poder-se-á verificar que estes são função das estimativas obtidas para $(\gamma, \lambda, \delta)$.

Nesta dissertação serão apresentadas duas abordagens para fazer inferência sobre os parâmetros indicados: a abordagem paramétrica e a abordagem semi-paramétrica. Neste capítulo irá ser desenvolvida a metodologia paramétrica, onde se supõe que há um modelo paramétrico G que se ajusta à amostra de observações i.i.d. Tendo em conta que os elementos da amostra são máximos, é sensato considerar para esse modelo a distribuição

GEV, apresentada em (2.6). Logo qualquer inferência que seja feita em relação aos diversos parâmetros indicados anteriormente, partem sempre do pressuposto que os dados em estudo seguem uma distribuição GEV. Irão ser descritos dois métodos para efetuar essa inferência - o Método Gumbel ou Método dos Máximos Anuais (MMA) e o Método POT (Peaks Over Threshold). Será também referenciado, mais superficialmente, um terceiro método chamado Método das Maiores Observações (MMO).

3.2 Método Gumbel ou Método dos Máximos Anuais

No Método Gumbel ou, também chamado de Método dos Máximos Anuais, a sequência de v.a.'s i.i.d., X_1, X_2, \dots, X_n , é partida em m blocos de dimensão k , com k suficientemente grande e, tal que, $n = m \times k$. Para cada um desses blocos extrai-se unicamente a maior observação, obtendo-se assim uma amostra de dimensão m de máximos. Este procedimento encontra-se ilustrado na figura 3.1, onde estão representados os máximos em cada um dos quatro períodos ou blocos de dimensão $n = 7$.

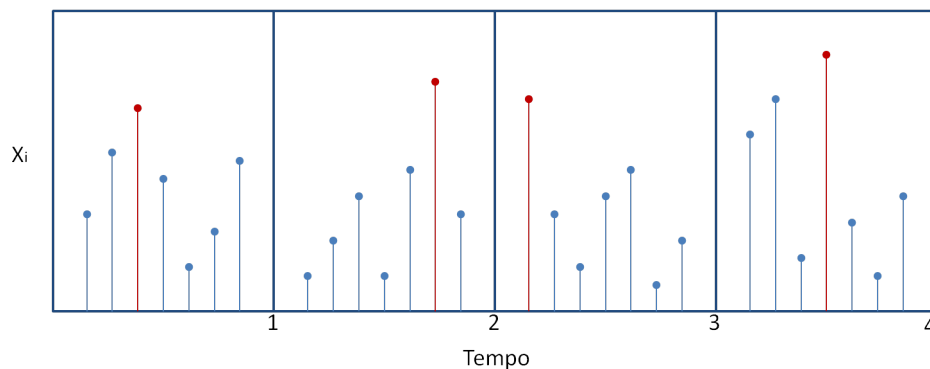


Figura 3.1: *Máximos Anuais*

Esta abordagem é usualmente implementada em situações onde se considera que um bloco equivale a um ano, logo diz-se que se tem uma amostra de dimensão m de máximos anuais. Ou seja, a amostra à qual se irá ajustar o modelo GEV terá a forma:

$$(Y_1, Y_2, \dots, Y_m)$$

em que,

$$Y_i = M_{ik} := \max(X_{i1}, X_{i2}, \dots, X_{ik}), \quad i = 1, 2, \dots, m.$$

Existem diversas técnicas para se obter estimativas dos parâmetros em EVT. Uma gráfica, outras numéricas, todas elas com vantagens e desvantagens. Apenas dois métodos serão utilizados, devido à sua fácil utilização e adaptação a diversos modelos e, que por esse mesmo motivo, se tornam particularmente atrativos - o método da Máxima Verosimilhança (MV) e o método dos Momentos Ponderados de Probabilidade (MPP).

3.2.1 Estimadores de Máxima Verosimilhança dos Parâmetros

Na secção 2.3 foi apresentada a família GEV, cuja f.d. é dada por

$$G_\gamma(x|\lambda, \delta) = \begin{cases} \exp\left(-\left(1 + \gamma\left(\frac{x-\lambda}{\delta}\right)\right)^{-1/\gamma}\right), & 1 + \gamma\left(\frac{x-\lambda}{\delta}\right) > 0, \gamma \neq 0, \\ \exp\left(-\exp\left(-\left(\frac{x-\lambda}{\delta}\right)\right)\right), & x \in \mathbb{R}, \gamma = 0. \end{cases}$$

Existem, portanto, três parâmetros a serem estimados: a localização λ , a escala δ e a forma γ .

A log-verosimilhança para a amostra obtida pelo método Gumbel, (Y_1, Y_2, \dots, Y_m) , i.i.d. à GEV, no caso de $\gamma \neq 0$, é dada por

$$\begin{aligned} l(\gamma, \lambda, \delta|y_1, \dots, y_m) &= \log L(\gamma, \lambda, \delta|y_1, \dots, y_m) = & (3.1) \\ &= -m \log \delta - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^m \log\left(1 + \gamma \frac{y_i - \lambda}{\delta}\right) - \sum_{i=1}^m \left(1 + \gamma \frac{y_i - \lambda}{\delta}\right)^{-\frac{1}{\gamma}}, \end{aligned}$$

com $1 + \gamma \frac{y_i - \lambda}{\delta} > 0$, $i = 1, \dots, m$.

No caso de $\gamma = 0$, essa log-verosimilhança toma a forma

$$\begin{aligned} l(0, \lambda, \delta|y_1, \dots, y_m) &= \log L(0, \lambda, \delta|y_1, \dots, y_m) = & (3.2) \\ &= -m \log \delta - \sum_{i=1}^m \exp\left(-\frac{y_i - \lambda}{\delta}\right) - \sum_{i=1}^m \frac{y_i - \lambda}{\delta}. \end{aligned}$$

Os estimadores de máxima verosimilhança $(\hat{\gamma}, \hat{\lambda}, \hat{\delta})$ e $(\hat{\lambda}, \hat{\delta})$ são obtidos maximizando as expressões (3.1) e (3.2), respetivamente. Para tal, consideram-se as equações

$$\frac{\partial l(\gamma, \lambda, \delta|y_1, \dots, y_m)}{\partial \gamma} = 0, \quad \frac{\partial l(\gamma, \lambda, \delta|y_1, \dots, y_m)}{\partial \lambda} = 0 \quad \text{e} \quad \frac{\partial l(\gamma, \lambda, \delta|y_1, \dots, y_m)}{\partial \delta} = 0,$$

no caso de $\gamma \neq 0$, e

$$\frac{\partial l(0, \lambda, \delta|y_1, \dots, y_m)}{\partial \lambda} = 0 \quad \text{e} \quad \frac{\partial l(0, \lambda, \delta|y_1, \dots, y_m)}{\partial \delta} = 0,$$

no caso de $\gamma = 0$.

Não existe solução analítica para estas equações e, como tal, terão de ser usados métodos numéricos de otimização. É necessário algum cuidado com estes algoritmos, pois podem dar soluções que violam as condições de regularidade exigidas nas propriedades assintóticas dos estimadores de MV. Tais condições não são satisfeitas no modelo GEV porque o suporte desta distribuição depende dos parâmetros desconhecidos. Logo as estimativas de MV obtidas para os parâmetros da GEV, por vezes, não são aplicáveis, pois poderão violar a condição $1 + \gamma \frac{y_i - \lambda}{\delta} > 0$, $i = 1, \dots, m$. Smith (1985) estudou detalhadamente este problema e obteve os seguintes resultados:

- quando $\gamma > -0.5$, os estimadores de MV são regulares, ou seja, gozam das propriedades de consistência, eficiência e normalidade assintóticas. Mais propriamente, tem-se que $\sqrt{m}((\hat{\gamma}, \hat{\lambda}, \hat{\delta}) - (\gamma, \lambda, \delta))$ é assintoticamente Normal de vetor de valores médios nulo e matriz de covariâncias igual à inversa da Matriz de Informação de Fisher, $I_{(\gamma, \lambda, \delta)}^{-1}$. Logo neste caso, os intervalos de confiança e outras formas de inferência são elaborados tendo como base esta aproximação à Normal;
- quando $-1 < \gamma < -0.5$, consegue-se, na maioria dos casos, encontrar os estimadores de MV, mas estes não gozam das propriedades assintóticas;
- quando $\gamma < -1$, geralmente, os estimadores de MV não existem.

Um outro pormenor que é preciso alguma atenção, prende-se com o facto de que computacionalmente podem surgir casos de falta de convergência dos métodos numéricos para a obtenção das estimativas de MV.

Em Castillo *et al.* (2005) é apresentado, com maior detalhe, um método iterativo para a obtenção das estimativas de MV dos parâmetros $(\gamma, \lambda, \delta)$, bem como a forma de obtenção da matriz de Informação de Fisher $I_{(\gamma, \lambda, \delta)}$. Também em Beirlant *et al.* (2004) poderão ser encontrados mais detalhes sobre o assunto.

3.2.2 Estimadores dos Momentos Ponderados de Probabilidade dos Parâmetros

O Momentos Ponderados de Probabilidade (MPP) para uma v.a. Y com f.d. F foram introduzidos por Greenwood *et al.* (1979) e são dados por:

$$M_{p,r,s} = E\{Y^p [F(Y)]^r [1 - F(Y)]^s\}, \quad p, r, s \in \mathbb{R}. \quad (3.3)$$

O caso específico da estimação MPP dos parâmetros da GEV foi desenvolvido por Hosking *et al.* (1985). Para o caso em que $\gamma \neq 0$, em específico com $\gamma < 1$, tem-se para $p = 1$, $r = 0, 1, 2, \dots$ e $s = 0$ o MPP tem a forma

$$M_{1,r,0} = E\{Y[F(Y)]^r\} = \frac{1}{r+1} \left\{ \lambda - \frac{\delta}{\gamma} [1 - (r+1)^\gamma \Gamma(1-\gamma)] \right\}, \quad (3.4)$$

em que $\Gamma(\cdot)$ é a *função gamma* dada por $\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx$, $t \geq 0$. Os estimadores MPP $(\hat{\gamma}, \hat{\lambda}, \hat{\delta})$ para $(\gamma, \lambda, \delta)$ são a solução do sistema de equações obtido através da expressão em (3.4):

$$\begin{cases} M_{1,0,0} = \lambda - \frac{\delta}{\gamma} [1 - \Gamma(1-\gamma)] \\ 2M_{1,1,0} - M_{1,0,0} = \frac{\delta}{\gamma} \Gamma(1-\gamma)(2^\gamma - 1) \\ \frac{3M_{1,2,0} - M_{1,0,0}}{2M_{1,1,0} - M_{1,0,0}} = \frac{3^\gamma - 1}{2^\gamma - 1}. \end{cases} \quad (3.5)$$

Substituindo de seguida $M_{1,r,0}$, com $r = 0, 1, 2$, pelo estimador centrado dado por Landwerh *et al.* (1979)

$$\hat{M}_{1,r,0} = \frac{1}{m} \sum_{j=1}^m \left(\prod_{l=1}^r \frac{(j-l)}{(m-l)} \right) \cdot Y_{j:m}, \quad (3.6)$$

e reescrevendo as duas primeiras equações de (3.5), em ordem a λ e a δ , respetivamente, obtém-se o estimador MPP, $(\hat{\gamma}, \hat{\lambda}, \hat{\delta})$:

$$\begin{cases} \hat{\lambda} = \hat{M}_{1,0,0} + \frac{\hat{\delta}}{\hat{\gamma}} [1 - \Gamma(1-\hat{\gamma})] \\ \hat{\delta} = \frac{\hat{\gamma}(2\hat{M}_{1,1,0} - \hat{M}_{1,0,0})}{\Gamma(1-\hat{\gamma})(2^{\hat{\gamma}} - 1)} \\ \frac{3\hat{M}_{1,2,0} - \hat{M}_{1,0,0}}{2\hat{M}_{1,1,0} - \hat{M}_{1,0,0}} = \frac{3^{\hat{\gamma}} - 1}{2^{\hat{\gamma}} - 1}. \end{cases} \quad (3.7)$$

Note-se que, para se obter $\hat{\gamma}$, a última equação do sistema em (3.7) terá de ser resolvida por métodos numéricos.

Já na situação em que $\gamma = 0$, ou seja, o caso da distribuição Gumbel, tem-se para $p = 1$, $r = 0, 1, 2, \dots$ e $s = 0$ que o MPP é dado por

$$M_{1,r,0} = E\{Y[F(Y)]^r\} = \frac{1}{r+1} [\lambda + \delta(-\psi(1) + \log(1+r))], \quad (3.8)$$

em que, $\psi(x) = \frac{d}{dx} \log \Gamma(x)$ é a *função digamma* e $-\psi(1) \simeq 0.5772156649$ é a *constante de Euler-Mascheroni*. Ora, neste caso, tem-se apenas os parâmetros de localização e escala (λ, δ) que precisam de ser estimados. Assim, substituindo em (3.8) a variável r por 0 e 1 e, resolvendo as equações obtidas em ordem aos dois parâmetros, vem que

$$\begin{cases} \lambda = M_{1,0,0} + \delta\psi(1) \\ \delta = \frac{2M_{1,1,0} - M_{1,0,0}}{\log 2}. \end{cases} \quad (3.9)$$

Novamente usando o estimador para $M_{1,r,0}$, apresentado em (3.6), e substituindo em (3.9), obtém-se os estimadores MPP para a localização e escala do modelo GEV, com $\gamma = 0$:

$$\begin{cases} \hat{\lambda} = \hat{M}_{1,0,0} + \hat{\delta}\psi(1) \\ \hat{\delta} = \frac{2\hat{M}_{1,1,0} - \hat{M}_{1,0,0}}{\log 2}. \end{cases} \quad (3.10)$$

Para estes estimadores baseados nos MPP, para $\gamma < 1$ e quando $m \rightarrow \infty$, verifica-se que

$$\sqrt{m}((\hat{\gamma}, \hat{\lambda}, \hat{\delta}) - (\gamma, \lambda, \delta))$$

é assintoticamente Normal. Mais detalhes poderão ser vistos em Beirlant *et al.* (2004).

3.2.3 Estimação de Outros Parâmetros de Interesse em EVT

No início deste capítulo, foi referenciado que para além dos três parâmetros diretamente associados à f.d. GEV, existem outras quantidades bastante usuais e importantes em EVT, e que também precisam de ser estimadas, nomeadamente, os quantis extremais, os níveis e períodos de retorno, probabilidade de excedência de um nível elevado e, no caso de $\gamma < 0$, o limite superior do suporte.

Os **quantis extremais** podem ser obtidos diretamente por inversão da f.d. GEV. Sendo assim, o quantil de probabilidade $1 - p$, q_{1-p} , ou também chamado de quantil extremal de probabilidade de excedência p , é dado por:

$$q_{1-p} = G_{\gamma}^{\leftarrow}(1 - p \mid \lambda, \delta) = \begin{cases} \lambda + \frac{\delta}{\gamma} ((-\log(1 - p))^{-\gamma} - 1), & \gamma \neq 0 \\ \lambda - \delta \log(-\log(1 - p)), & \gamma = 0. \end{cases} \quad (3.11)$$

O quantil extremal de probabilidade $1 - p$ também pode ser escrito usando a função quantil de cauda, apresentada na definição 2.4. Considerando $t = \frac{1}{p}$, tem-se

$$q_{1-p} = U\left(\frac{1}{p}\right) = G_\gamma^\leftarrow\left(1 - \frac{1}{1/p} \mid \lambda, \delta\right) = G_\gamma^\leftarrow(1 - p \mid \lambda, \delta) \quad (3.12)$$

Para se obter a estimativa do quantil, bastará substituir $(\gamma, \lambda, \delta)$ pelas estimativas de MV ou dos MPP obtidas para estes parâmetros.

Outro conceito muito usado em EVT é o de valor de retorno. Este engloba duas quantidades: o **nível de retorno** U e o **período de retorno** T . Estes dois conceitos estão interligados, pois, o nível de retorno $U(T)$ é o valor que é excedido, em média, uma vez em cada T anos (período de retorno do nível U). O **nível de retorno T -anos** para a v.a. Y (v.a. obtida pelo método dos máximos anuais), não é mais do que o quantil de probabilidade $1 - \frac{1}{T}$ e pode ser dado por

$$U(T) = q_{1-\frac{1}{T}} = G_\gamma^\leftarrow\left(1 - \frac{1}{T} \mid \lambda, \delta\right). \quad (3.13)$$

O **período de retorno do nível u** é calculado através de

$$T_u = \frac{1}{P(Y > u)} = \frac{1}{1 - G_\gamma(u \mid \lambda, \delta)}. \quad (3.14)$$

Mais uma vez, para se obter estimativas destes dois valores, substitui-se os parâmetros de forma, localização e escala pelas estimativas encontradas pelo método de MV ou pelo método MPP.

Em EVT, também é usual estar-se interessado na estimativa do valor da **probabilidade de excedência do nível elevado u** , bastando, para isso, calcular

$$P(Y > u) = 1 - G_\gamma(u \mid \lambda, \delta), \quad (3.15)$$

e usar as estimativas (MV ou MPP) dos parâmetros da distribuição GEV.

No caso do EVI ser negativo ($\gamma < 0$), ou seja, quando se está no domínio max-estável Weibull, como já tinha sido indicado, o **limite superior do suporte**, x^F , é finito e é dado por

$$x^F = q_1 = U(\infty) = \lambda - \frac{\delta}{\gamma}. \quad (3.16)$$

A estimativa para este valor pode ser obtida substituindo $(\gamma, \lambda, \delta)$ por $(\hat{\gamma}, \hat{\lambda}, \hat{\delta})$, obtidos pelos dois métodos de estimação apresentados.

Caso se pretenda inferir sobre parâmetros análogos, mas para a população referente aos m blocos de observações de X_1, \dots, X_n i.i.d a $X \sim F$, ter-se-á que recorrer à propriedade de max-estabilidade, ou mais simplesmente, ao facto de que

$$Y := \max X_i = M_n \text{ possuir distribuição } F_Y \equiv F_{M_n} = F^n \approx G_\gamma,$$

onde n é a dimensão de cada bloco. Assim, por exemplo, para blocos de tamanho n , os quantis de probabilidade $1 - p$ ajustados para X , q_{1-p}^* , em que $F(q_{1-p}^*) = 1 - p$, são tais que

$$F^n(q_{1-p}^*) = (1 - p)^n \approx G_\gamma(q_{1-p}^* | \lambda, \delta).$$

Logo, explicitamente, estes tomam a forma

$$q_{1-p}^* = G_\gamma^\leftarrow((1 - p)^n | \lambda, \delta) = \begin{cases} \lambda + \frac{\delta}{\gamma} ((-\log(1 - p)^n)^{-\gamma} - 1), & \gamma \neq 0 \\ \lambda - \delta \log(-\log(1 - p)^n), & \gamma = 0. \end{cases}$$

3.2.4 Intervalos de Confiança para os Parâmetros

Nas secções 3.2.1 e 3.2.2 foi referido que os estimadores de MV e MPP dos parâmetros da GEV têm uma distribuição assintoticamente Normal. Assim, para construir intervalos de confiança para γ , λ e δ , e mesmo para os parâmetros apresentados no tópico anterior, em especial, os quantis extremos e os níveis de retorno, pode fazer-se uso dessa normalidade assintótica. Assim, um intervalo de, aproximadamente, $(1 - \alpha)100\%$ de confiança para um destes parâmetros, designe-se genericamente por θ , é dado por

$$\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \times s.e.$$

em que, *s.e.* é o erro padrão (do inglês, *standard error*) e $z_{1-\frac{\alpha}{2}}$ é o quantil de probabilidade $1 - \frac{\alpha}{2}$ de uma Normal standard. Uma vez que a distribuição Normal é simétrica, estes intervalos de confiança são centrados nas respetivas estimativas pontuais dos parâmetros. Todavia, este procedimento não é muito recomendado, uma vez que este admite que o comportamento limite da forma da verosimilhança, junto do máximo, é quadrático. Na verdade, esta curva costuma ser bastante assimétrica. Sendo assim, poderão ser obtidas estimativas intervalares de melhor qualidade usando a função de *profile log-likelihood* (Beirlant *et al.*, 2004). Por este método, os intervalos de confiança não são obrigatoriamente centrados na estimativa pontual de MV.

Definição 3.1 (Função de Profile Log-Likelihood). *A função de profile log-likelihood para γ , $\log L_p(\gamma)$, é dada por*

$$\log L_p(\gamma) = \max_{\lambda, \delta | \gamma} \log L(\lambda, \delta | \gamma, y_1, \dots, y_m),$$

ou seja, para cada valor de γ , a função de profile log-likelihood é a log-verosimilhança maximizada em ordem aos parâmetros λ e δ .

A construção dos intervalos de confiança tem por base a estatística de Razão de Verosimilhanças

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

do teste de Razão de Verosimilhanças, para testar

$$H_0 : \gamma = \gamma_0 \quad vs \quad H_1 : \gamma \neq \gamma_0.$$

Esta estatística tem por distribuição assintótica, quando $m \rightarrow \infty$,

$$-2 \log \Lambda \xrightarrow{d} Q \sim \chi_1^2.$$

O teste, de nível assintótico α , vai rejeitar H_0 se $-2 \log \Lambda > \chi_1^2(1 - \alpha)$. Sendo assim, o intervalo para o parâmetro γ baseado no profile log-likelihood, com um grau de confiança de $(1 - \alpha)100\%$, é dado por

$$\{\gamma : -2 \log \Lambda \leq \chi_1^2(1 - \alpha)\} = \left\{ \gamma : \log L_p(\gamma) \geq \log L_p(\hat{\gamma}) - \frac{\chi_1^2(1 - \alpha)}{2} \right\}. \quad (3.17)$$

Esta metodologia pode também ser aplicada, de forma análoga, aos restantes parâmetros da distribuição GEV. Para além disso, podem também ser obtidos intervalos de confiança para os parâmetros apresentados na secção 3.2.3. Em particular, para obter um intervalo de confiança para um determinado quantil extremal, q_{1-p} , é necessário fazer-se uma reparametrização do modelo GEV, de forma a que q_{1-p} seja um dos parâmetros do modelo. De seguida, a função de profile log-likelihood é obtida por maximização da log-verosimilhança em relação aos restantes parâmetros, como explicado anteriormente. Ora, a partir da expressão em (3.11), e considerando, sem perda de generalidade, apenas o caso de $\gamma \neq 0$, obtém-se

$$q_{1-p} = \lambda + \frac{\delta}{\gamma} ((-\log(1-p))^{-\gamma} - 1) \Leftrightarrow \lambda = q_{1-p} + \frac{\delta}{\gamma} (1 - (-\log(1-p))^{-\gamma}). \quad (3.18)$$

Substituindo λ na equação $l(\gamma, \lambda, \delta | y_1, \dots, y_m)$, apresentada em (3.1), pela expressão em (3.18), obtém-se a função log-verosimilhança da distribuição GEV em função do vetor de parâmetros $(\gamma, q_{1-p}, \delta)$. A partir deste ponto o processo é análogo ao apresentado para a estimação intervalar de γ . Para os restantes parâmetros de interesse em EVT, a metodologia para a obtenção de intervalos de confiança é semelhante à apresentada.

3.2.5 Testes de Hipóteses para o Modelo GEV

Como já foi indicado, o modelo GEV é a única distribuição limite para o máximo. Contudo, este modelo engloba três casos: Weibull, Gumbel e Fréchet. De facto, viu-se

que uma distribuição com cauda direita limitada não pode estar no domínio de atração Fréchet, e uma distribuição com um suporte não limitado superiormente não pertence ao max-domínio Weibull. Para além disso, a família Gumbel tem caudas mais leves do que a família Fréchet, mas tem caudas mais pesadas quando comparada com a família Weibull. Por outro lado, o modelo Gumbel é considerado mais simples pois tem menos um parâmetro que os modelos Fréchet e Weibull. Assim, selecionar um modelo errado para descrever uma determinada amostra de dados, pode ter, na prática, consequências graves, pois leva a seleção errónea dos parâmetros em jogo. Logo, será sensato tomar-se uma decisão sobre qual dos três modelos melhor se ajusta à a.a. em estudo, o famoso trilema estatístico.

Poderão ser usadas técnicas gráficas, como os papéis de probabilidade (P-P plot e Q-Q plot), histogramas com as curvas da densidade sobrepostas, por exemplo, para fazer um diagnóstico à qualidade do ajustamento do modelo selecionado. Mas, também, podem ser efetuados testes de hipóteses para a seleção do modelo max-estável, nomeadamente ao parâmetro de forma γ .

A f.d. Gumbel é uma opção atrativa devido à simplicidade do modelo, bem como da inferência a ele associado. Para além disso, ao modelo Gumbel está associado um EVI nulo, $\gamma = 0$, e, por isso mesmo, é considerado como um ponto de transição de $\gamma < 0$ para $\gamma > 0$, isto é, de uma família de distribuições com limite superior do suporte finito para uma outra com limite superior do suporte infinito. Por estes motivos, a escolha mais plausível será a de testar a hipótese do modelo ser Gumbel contra a hipótese alternativa de não ser, que se traduz por

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma \neq 0. \quad (3.19)$$

No caso de rejeição do modelo Gumbel, poderá ser feito um teste unilateral para se determinar qual dos outros dois modelos max-estáveis será o adequado. Para a distribuição Weibull, as hipóteses serão,

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma < 0, \quad (3.20)$$

e, no caso da distribuição Fréchet, essas hipóteses assumirão a forma

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma > 0. \quad (3.21)$$

Estes testes têm sido largamente focados na literatura. É de destacar alguns deles, nomeadamente, Van Montfort (1970), Bardsley (1977), Otten and Van Montfort (1978), Tiago de Oliveira (1981), Gomes (1982), Tiago de Oliveira (1984), Tiago de Oliveira and Gomes (1984), Hosking (1984), Marohn (1994), Wang *et al.* (1996) e Marohn (1998a).

O primeiro teste apresentado, e que pode ser visto em detalhe em Hosking (1984), é um Teste de Razão de Verosimilhanças (TRV) e é aplicado para as hipóteses definidas em (3.19). A estatística do TRV é dada pela *Deviance*.

Definição 3.2 (Deviance). *Sejam \mathcal{M}_0 e \mathcal{M}_1 dois modelos encaixados ou aninhados, isto é, $\mathcal{M}_0 \subset \mathcal{M}_1$. A estatística deviance é definida por*

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

em que $l(\mathcal{M}_0)$ e $l(\mathcal{M}_1)$ são, respetivamente, os valores da função de log-verosimilhança para os modelos ajustados \mathcal{M}_0 e \mathcal{M}_1 .

Adaptando-se a nomenclatura a este caso específico, vem que

$$\mathcal{D} = 2 \cdot \left\{ l(\hat{\gamma}_{G_\gamma}, \hat{\lambda}_{G_\gamma}, \hat{\delta}_{G_\gamma} | Y_1, \dots, Y_m) - l(0, \hat{\lambda}_{G_0}, \hat{\delta}_{G_0} | Y_1, \dots, Y_m) \right\}, \quad (3.22)$$

em que $l(\gamma, \lambda, \delta | Y_1, \dots, Y_m)$ e $l(0, \lambda, \delta | Y_1, \dots, Y_m)$ são as funções log-verosimilhança para os modelos GEV (G_γ) e Gumbel (G_0), respetivamente, apresentadas em (3.1) e (3.2). $(\hat{\lambda}_{G_0}, \hat{\delta}_{G_0})$ e $(\hat{\gamma}_{G_\gamma}, \hat{\lambda}_{G_\gamma}, \hat{\delta}_{G_\gamma})$ são os estimadores de MV para os parâmetros dos modelos G_0 e G_γ .

Sob a validade de H_0 , tem-se

$$\mathcal{D} \xrightarrow[m \rightarrow \infty]{d} Q_1 \sim \chi_{(1)}^2.$$

Hosking (1984) recomenda a correção de Bartlett para se obter uma maior precisão na aproximação ao qui-quadrado. Assim,

$$\mathcal{D}^* = \frac{\mathcal{D}}{1 + \frac{2.8}{m}} \xrightarrow[m \rightarrow \infty]{d} Q_1 \sim \chi_{(1)}^2. \quad (3.23)$$

Rejeita-se H_0 , ao nível assintótico α , isto é, rejeita-se o modelo Gumbel, quando

$$\mathcal{D}^* \geq \chi_{1;1-\alpha}^2,$$

onde $\chi_{1;1-\alpha}^2$ é o quantil de probabilidade $1-\alpha$ de um qui-quadrado com 1 grau de liberdade (g.l.). A decisão também pode ser tomada tendo como base o p-value, dado por

$$p(\mathcal{D}^*) = P(Q_1 > \mathcal{D}^*) = 1 - P(Q_1 \leq \mathcal{D}^*),$$

e, como é habitual, rejeita-se H_0 ao nível assintótico α se $p(\mathcal{D}^*) \leq \alpha$.

Um outro teste para as mesmas hipóteses do caso anterior, desenvolvido por Tiago de Oliveira (1981) e também por Hosking (1984), é o Teste Score de Rao. Este teste averigua quando é que a derivada da função log-verosimilhança, em ordem a γ , no ponto $\gamma = 0$,

é significativamente diferente de zero. Valores não nulos sugerem que $\gamma \neq 0$ e, como tal, levam à rejeição da hipótese nula.

Se $g_\gamma(y | \lambda, \delta)$ for a f.d.p. da distribuição GEV, $G_\gamma(y | \lambda, \delta)$, a função log-verosimilhança para uma a.a. observada (y_1, \dots, y_m) pode ser dada, genericamente, por

$$l(\gamma, \lambda, \delta | y_1, \dots, y_m) = \sum_{i=1}^m \log g_\gamma(y_i | \lambda, \delta).$$

A função score em relação a γ é:

$$V(\gamma | \lambda, \delta, y_1, \dots, y_m) = \frac{\partial l(\gamma, \lambda, \delta | y_1, \dots, y_m)}{\partial \gamma} = \sum_{i=1}^m \frac{\partial \log g_\gamma(y_i | \lambda, \delta)}{\partial \gamma}.$$

Sendo $(\hat{\lambda}_{G_0}, \hat{\delta}_{G_0})$ os estimadores de MV para o modelo Gumbel, G_0 , a estatística score para o teste (3.19) é definida como

$$V_m = \sum_{i=1}^m \lim_{\gamma \rightarrow 0} \frac{\partial \log g_\gamma(y_i | \hat{\lambda}_{G_0}, \hat{\delta}_{G_0})}{\partial \gamma} = \sum_{i=1}^m \left(\frac{1}{2} Z_i^2 - Z_i - \frac{1}{2} Z_i^2 \exp(-Z_i) \right),$$

com $Z_i = \frac{Y_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}}$, $i = 1, \dots, m$.

A distribuição assintótica desta estatística, sob a validade de H_0 , poderá ser Gaussiana (Tiago de Oliveira, 1981) ou um Qui-quadrado: (Hosking, 1984):

$$V_m^* = \frac{V_m}{\sqrt{2.09797m}} \xrightarrow[m \rightarrow \infty]{d} Z_1 \sim N(0, 1); \quad (3.24)$$

$$V_m^{*2} = \frac{V_m^2}{2.09797m} \xrightarrow[m \rightarrow \infty]{d} Q_2 \sim \chi_{(1)}^2. \quad (3.25)$$

A região de rejeição deste teste para o caso (3.19), para um nível assintótico α , é dada por

$$|V_m^*| \geq z_{1-\frac{\alpha}{2}} \quad \text{ou por} \quad V_m^{*2} \geq \chi_{1;1-\alpha}^2,$$

com $z_{1-\frac{\alpha}{2}}$ o quantil de probabilidade $1 - \frac{\alpha}{2}$ de uma Normal standard e $\chi_{1;1-\alpha}^2$ o quantil de probabilidade $1 - \alpha$ de um Qui-quadrado com 1 g.l. Os p-values podem ser calculados através de

$$p(V_m^*) = 2 - 2 \cdot \Phi(|V_m^*|) \quad \text{ou} \quad p(V_m^{*2}) = 1 - P(Q_2 \leq V_m^{*2}).$$

Esta estatística de teste também poderá ser usada nas situações (3.20) e (3.21). Para cada um destes casos, rejeita-se H_0 , ao nível assintótico α , se

$$V_m^* \leq z_\alpha \quad \text{ou} \quad V_m^* \geq z_{1-\alpha},$$

respetivamente. Os p-values podem ser obtidos por

$$p(V_m^*) = \Phi(V_m^*) \quad \text{ou por} \quad p(V_m^*) = 1 - \Phi(V_m^*),$$

respetivamente.

Em Tiago de Oliveira and Gomes (1984) é apresentado um teste específico para as hipóteses apresentadas em (3.20) e (3.21). A estatística de teste, a estatística Gumbel, é dada por:

$$GS_m = \frac{Y_{m:m} - Y_{[m/2]+1:m}}{Y_{[m/2]+1:m} - Y_{1:m}}$$

Sob a validade de H_0 , tem-se que

$$GS_m^* = \frac{GS_m - \beta_m}{\alpha_m} \xrightarrow[m \rightarrow \infty]{d} W \frown \Lambda, \quad (3.26)$$

onde α_m e β_m são dados por:

$$\alpha_m = \frac{1}{\log(\log m)} \quad \text{e} \quad \beta_m = \frac{\log m + \log(\log 2)}{\log(\log m) - \log(\log 2)}.$$

Considerando que \mathcal{G}_α é o quantil de probabilidade α para uma distribuição Gumbel standard, para o teste de hipóteses (3.20), ao nível assintótico α , rejeita-se H_0 se

$$GS_m \leq \beta_m - \alpha_m \log(-\log \alpha)$$

$$\Updownarrow$$

$$GS_m^* \leq -\log(-\log \alpha) = \mathcal{G}_\alpha.$$

O p-value é dado por

$$p(GS_m^*) = \Lambda(GS_m^*).$$

Já no caso das hipóteses serem as apresentadas em (3.21), ao nível assintótico α , H_0 é rejeitada quando

$$GS_m \geq \beta_m - \alpha_m \log(-\log(1 - \alpha))$$

$$\Updownarrow$$

$$GS_m^* \geq -\log(-\log(1 - \alpha)) = \mathcal{G}_{1-\alpha}.$$

Para se obter o p-value, bastará calcular

$$p(GS_m^*) = 1 - \Lambda(GS_m^*).$$

Em Marohn (2000) é explicado detalhadamente um outro teste para as três possibilidades de hipóteses apresentadas - o Teste LAN (Locally Asymptotically Normal).

A estatística do teste LAN é dada por

$$T_m = \frac{1}{3.451} \cdot \left(\frac{1.6449}{\sqrt{m}} \cdot \mathcal{V}_{1,m} - \hat{\delta}_{G_0} \cdot \frac{0.5066}{\sqrt{m}} \cdot \mathcal{V}_{2,m} - \hat{\delta}_{G_0} \cdot \frac{0.8916}{\sqrt{m}} \cdot \mathcal{V}_{3,m} \right).$$

$\mathcal{V}_{1,m}$, $\mathcal{V}_{2,m}$ e $\mathcal{V}_{3,m}$ são as derivadas de primeira ordem da função log-verosimilhança da distribuição GEV, com respeito aos parâmetros γ , δ e λ , respetivamente, quando $\gamma = 0$:

$$\begin{pmatrix} \mathcal{V}_{1,m} \\ \mathcal{V}_{2,m} \\ \mathcal{V}_{3,m} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^m \lim_{\gamma \rightarrow 0} \frac{\partial \log g_\gamma(Y_i | \lambda, \delta)}{\partial \gamma} \\ \sum_{i=1}^m \frac{\partial \log g_0(Y_i | \lambda, \delta)}{\partial \delta} \\ \sum_{i=1}^m \frac{\partial \log g_0(Y_i | \lambda, \delta)}{\partial \lambda} \end{pmatrix}$$

E, que neste caso, assumem a forma,

$$\begin{pmatrix} \sum_{i=1}^m \left\{ -\frac{Y_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}} + \frac{1}{2} \left(\frac{Y_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}} \right)^2 \left(1 - \exp \left(-\frac{Y_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}} \right) \right) \right\} \\ \sum_{i=1}^m \left\{ -\frac{1}{\hat{\delta}_{G_0}} + \left(\frac{Y_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}^2} \right) \left(1 - \exp \left(-\frac{Y_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}} \right) \right) \right\} \\ \sum_{i=1}^m \left\{ \frac{1}{\hat{\delta}_{G_0}} - \frac{1}{\hat{\delta}_{G_0}} \exp \left(-\frac{Y_i - \hat{\lambda}_{G_0}}{\hat{\delta}_{G_0}} \right) \right\} \end{pmatrix},$$

em que $(\hat{\lambda}_{G_0}, \hat{\delta}_{G_0})$ são os estimadores de MV dos parâmetros do modelo G_0 .

Sob a validade de H_0 , tem-se que

$$T_m^* = \frac{T_m}{0.6904} \xrightarrow[m \rightarrow \infty]{d} Z_2 \sim N(0, 1), \quad (3.27)$$

e ao nível de significância assintótico α , H_0 é rejeitada se

$$|T_m^*| \geq z_{1-\frac{\alpha}{2}} \quad \text{ou} \quad T_m^* \leq z_\alpha \quad \text{ou} \quad T_m^* \geq z_{1-\alpha},$$

para os casos definidos em (3.19), (3.20) e (3.21), respetivamente. Os correspondentes p-values são obtidos através do cálculo de

$$p(T_m^*) = 2 - 2 \cdot \Phi(|T_m^*|) \quad \text{ou} \quad p(T_m^*) = \Phi(T_m^*) \quad \text{ou} \quad p(T_m^*) = 1 - \Phi(T_m^*).$$

Associado a este problema de determinação do modelo extremal, também podem ser usados os testes de ajustamento (do inglês, goodness-of-fit tests) para o modelo Gumbel e que receberam grande atenção por parte de Stephens (1976), Stephens (1977), Stephens (1986) e Kinnison (1989). Estes testes são baseados, essencialmente, nas estatísticas de ajustamento de Kolmogorov-Smirnov, de Cramér-von Mises e de Anderson-Darling.

Assim, será considerado para hipótese nula

$$H_0 : F(x) = \exp \left(- \exp \left(- \frac{x - \lambda}{\delta} \right) \right), \quad (3.28)$$

ou seja, a distribuição Gumbel, contra uma hipótese alternativa não especificada. Os parâmetros λ e δ são desconhecidos e têm de ser estimados. As estimativas $\hat{\lambda}_{G_0}$ e $\hat{\delta}_{G_0}$ são obtidas por MV. A hipótese nula é composta, uma vez que apenas a forma funcional da f.d. é fornecida, e os parâmetros desta não são especificados.

As estatísticas de ajustamento utilizadas, e já mencionadas, serão dadas por:

i. Kolmogorov-Smirnov

$$D_m = \max_{1 \leq i \leq m} \left\{ \left| G_0(Y_{i:m} | \hat{\lambda}_{G_0}, \hat{\delta}_{G_0}) - \frac{i}{m} \right|, \left| G_0(Y_{i:m} | \hat{\lambda}_{G_0}, \hat{\delta}_{G_0}) - \frac{i-1}{m} \right| \right\}; \quad (3.29)$$

ii. Cramér-von Mises

$$W_m^2 = \sum_{i=1}^m \left(G_0(Y_{i:m} | \hat{\lambda}_{G_0}, \hat{\delta}_{G_0}) - \frac{2i-1}{2m} \right)^2 + \frac{1}{12m}; \quad (3.30)$$

iii. Anderson-Darling

$$A_m^2 = -m - \frac{1}{m} \sum_{i=1}^m \left\{ (2i-1) \log(G_0(Y_{i:m} | \hat{\lambda}_{G_0}, \hat{\delta}_{G_0})) + (2m+1-2i) \log(1 - G_0(Y_{i:m} | \hat{\lambda}_{G_0}, \hat{\delta}_{G_0})) \right\}. \quad (3.31)$$

A hipótese nula será rejeitada se o valor da estatística observada for superior a um determinado quantil de nível assintótico α . Estes quantis encontram-se tabelados e podem ser encontrados em Chandra *et al.* (1981), para a estatística de Kolmogorov-Smirnov, e em Stephens (1977), para as estatísticas de Cramér-von Mises e Anderson-Darling. Ambos estão reproduzidos nas tabelas 3.1 e 3.2, que são apresentadas de seguida.

Tabela 3.1: *Percentis para a estatística de Kolmogorov-Smirnov, adaptada para a distribuição Gumbel*

Estatística	m	Nível de significância α			
		.10	.05	.025	.01
$\sqrt{m} \cdot D_m$	10	.760	.819	.880	.944
	20	.779	.843	.907	.973
	50	.790	.856	.922	.988
	∞	.803	.874	.939	1.007

Chandra *et al.* (1981)Tabela 3.2: *Percentis para as estatísticas de Cramér-von Mises e de Anderson-Darling, adaptadas para a distribuição Gumbel*

Estatística	Modificação	Percentis				
		.75	.90	.95	.975	.99
W_m^2	$W_m^2(1 + 0.2/\sqrt{m})$.073	.102	.124	.146	.175
A_m^2	$A_m^2(1 + 0.2/\sqrt{m})$.474	.637	.757	.877	1.038

Stephens (1977)

3.2.6 Não-Estacionariedade: Sazonalidade e Tendência

Processos não-estacionários têm características que mudam sistematicamente ao longo do tempo. Em muitas situações relacionadas com o meio ambiente, a presença de não-estacionariedade nas observações costuma ser comum, ou por causa dos efeitos sazonais, uma vez que há padrões no clima que se manifestam em determinados meses, ou devido à presença de tendência na a.a., já que, para alguns fenômenos naturais, se tem verificado determinadas mudanças climáticas a longo prazo.

Existem resultados teóricos para modelar algumas forma de não-estacionariedade, contudo, são demasiado restritivos para serem usados na descrição de padrões de tendência presentes em situações reais. Logo, em detrimento desses resultados, é mais comum adotar-se uma abordagem pragmática e usar as distribuições limite max-estáveis, considerando-as como os modelos básicos e que podem ser melhorados através de modelação estatística.

Quando a não-estacionariedade se deve, essencialmente, a efeitos sazonais, o procedimento habitual consiste na partição da amostra de dados, por períodos de tempo, dentro dos quais se poderá assumir a homogeneidade das observações, e aplicar a análise de valores extremos separadamente para cada um desses períodos. Exemplos desta abordagem podem ser encontrados em Smith (1989) e Walshaw (1994). Estes períodos de tempo poderão ser, por exemplo, as estações do ano, ou épocas secas e épocas chuvosas, onde as

variações sazonais são facilmente identificadas. Contudo, esta abordagem tem a principal desvantagem de que é preciso estimar os parâmetros do modelo extremal para cada um dos períodos sazonais e, estabelecer relações entre eles, não é tarefa fácil!

Uma abordagem alternativa consiste em permitir que os parâmetros do modelo variem continuamente ao longo do período de sazonalidade. As séries de Fourier podem ser ajustadas aos parâmetros e o modelo que melhor se ajusta, pode ser selecionado através de testes de razão de verosimilhança (TRV). No entanto, Walshaw (1991) mostrou que a diferença, ao nível da inferência, entre esta abordagem e a primeira metodologia indicada, é muito reduzida. Para além disso, a abordagem baseada nas séries de Fourier mostra-se pouco eficiente, já que a estimação dos parâmetros que variam continuamente é computacionalmente morosa. Deste modo, esta metodologia não se revela uma abordagem vantajosa quando comparada com a abordagem em períodos sazonais.

No caso de se verificar não-estacionariedade devido à existência de tendência nas observações, poder-se-á criar uma série temporal Z_t , que corresponderá ao máximo verificado no instante t , e ajustar-lhe o modelo extremal GEV. Essa tendência poderá estar presente na localização ou na escala, e pode ser representada, no caso de ser linear, por exemplo, como

$$\begin{cases} \lambda(t) = \beta_0 + \beta_1 t \\ \delta(t) = \delta_0 + \delta_1 t. \end{cases} \quad (3.32)$$

Quantas vezes não se ouviu já falar no aquecimento global, e de que a temperatura média tem vindo a aumentar gradualmente, ou de que a precipitação em determinadas zonas do planeta é cada vez menor. Isto são algumas das evidências da presença de tendência no meio ambiente. As variações observadas numa determinada categoria ambiental são bem modeladas, em geral, por uma tendência linear no parâmetro de localização, da distribuição GEV. Esta forma de tendência está representada na primeira expressão de (3.32), e será a destacada nesta dissertação. O parâmetro β_1 corresponde à taxa de alteração média anual (já que se está na metodologia dos máximos anuais) no máximo anual do processo em estudo. Outras formas para modelar a tendência poderão ser vistas em Coles (2001).

Para se obter estimativas do vetor de parâmetros $(\gamma, \lambda(t), \delta)$, poderá ser usado o método de MV, de forma semelhante ao apresentado na secção 3.2.1.

Como existe a possibilidade de se ajustar uma vasta combinação de modelos a um conjunto de dados, pois qualquer um dos parâmetros pode ser escrito como função do tempo ou mesmo de outras covariáveis, a questão da seleção do melhor modelo é bastante importante. O princípio básico será sempre o de escolher o modelo parcimonioso, isto é,

o modelo mais simples, com o menor número de parâmetros possíveis, e que explique a maior percentagem de variância presente na amostra. Comparações entre o modelo GEV com e sem tendência poderão ser feitas usando a estatística Deviance, \mathcal{D} , apresentada na definição 3.2.

Se se considerar que \mathcal{M}_1 e \mathcal{M}_0 são os modelos ajustados com e sem tendência, respetivamente, valores elevados de \mathcal{D} indicam que o modelo \mathcal{M}_1 explica, substancialmente, uma maior parte da variabilidade da amostra do que \mathcal{M}_0 ; valores baixos de \mathcal{D} sugerem que aumentar o número de parâmetros não melhora a capacidade preditiva do modelo. Contudo, isto é muito vago porque não se sabe a partir de que valores se poderá considerar que a deviance é ou não elevada. Será, por isso, necessário aplicar-se um teste de hipóteses, neste caso, um TRV.

A distribuição assintótica da deviance é um qui-quadrado com os g.l. dados pela diferença de parâmetros entre os dois modelos, ou seja,

$$\mathcal{D} = 2 \cdot \{\max l(\mathcal{M}_1) - \max l(\mathcal{M}_0)\} \xrightarrow[n \rightarrow \infty]{d} Q \sim \chi_{p_1 - p_0}^2 = \chi_k^2,$$

em que p_0 e p_1 são, respetivamente, o número de parâmetros dos modelos \mathcal{M}_0 e \mathcal{M}_1 e $p_1 = p_0 + k$ para $k = 1, 2, \dots$

As hipóteses do teste serão:

$$H_0 : \text{Não há tendência na amostra} \quad vs \quad H_1 : \text{Há tendência na amostra}$$

$$\Updownarrow$$

$$H_0 : \text{O modelo tem } p_0 \text{ parâmetros} \quad vs \quad H_1 : \text{O modelo tem } p_1 \text{ parâmetros}$$

A estatística de teste é dada por \mathcal{D} e rejeita-se a hipótese nula, ou seja, aceita-se a presença de tendência no conjunto de dados, a um nível de significância α , quando

$$\mathcal{D} \geq \chi_{k;1-\alpha}^2,$$

em que $\chi_{k;1-\alpha}^2$ é o quantil de probabilidade $1 - \alpha$ de um qui-quadrado com k g.l. Alternativamente, também poderá ser calculado o p-value, neste caso dado por

$$p(\mathcal{D}) = P(Q \geq \mathcal{D})$$

e, como habitualmente, rejeitar-se-à a hipótese nula, ao nível de significância α , se $p(\mathcal{D}) \leq \alpha$.

Há que ter em consideração que, a distribuição assintótica da estatística de teste de razão de verosimilhanças só é válida se a hipótese nula não estiver na fronteira. De acordo com Stram and Lee (1994), quando tal acontece, a distribuição assintótica do TRV é

uma mistura de qui-quadrados. Assim, quando se testa a hipótese de existência de p_0 parâmetros *versus* a hipótese alternativa de existirem $p_1 = p_0 + 1$ parâmetros, tem-se uma mistura 50 : 50 de dois qui-quadrados com p_0 e $p_0 + 1$ g.l..

Quando o modelo da hipótese nula não está encaixado no modelo da hipótese alternativa, por exemplo, quando se tem um modelo com tendência na localização e outro com tendência na escala, o TRV não é apropriado. Nesse caso, a comparação entre modelos não encaixados pode ser feita com base nos critérios de informação, como por exemplo, o critério **AIC** (Akaike Information Criterion)

$$AIC = -2 \cdot \max l(\mathcal{M}) + 2 \cdot p, \quad (3.33)$$

ou o critério **BIC** (Bayesian Information Criterion),

$$BIC = -2 \cdot \max l(\mathcal{M}) + 2 \cdot p \cdot \log N, \quad (3.34)$$

em que p é o número de parâmetros do modelo \mathcal{M} e N a dimensão da amostra. Quanto menor o valor do AIC ou do BIC, melhor será o modelo. Estes critérios também podem ser aplicados no caso de modelos encaixados.

3.3 Método POT

No subcapítulo 2.6 já foi apresentado o conceito de excesso acima de um determinado *threshold* u , e que foi denotado por $(X - u)|X > u$ ou por $Y|Y > 0$, com $Y = X - u$. Foi também visto no Teorema 2.17, Teorema de Pickands-Balkema-de Haan, que a f.d. condicional destes excessos pode ser bem aproximada por uma Generalizada Pareto (GP), cuja expressão é novamente apresentada de seguida:

$$H_\gamma(x|u, \sigma_u) := \begin{cases} 1 - \left(1 + \gamma \frac{x-u}{\sigma_u}\right)^{-1/\gamma}, & x \geq u & \text{se } \gamma > 0, \\ 1 - \exp\left(-\frac{x-u}{\sigma_u}\right), & x \geq u & \text{se } \gamma = 0, \\ 1 - \left(1 + \gamma \frac{x-u}{\sigma_u}\right)^{-1/\gamma}, & u \leq x \leq u - \frac{\sigma_u}{\gamma} & \text{se } \gamma < 0. \end{cases} \quad (3.35)$$

Em termos de Y a f.d. GP assume a forma

$$H_\gamma(y|\sigma_u) := \begin{cases} 1 - \left(1 + \gamma \frac{y}{\sigma_u}\right)^{-1/\gamma}, & y \geq 0 & \text{se } \gamma > 0, \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & y \geq 0 & \text{se } \gamma = 0, \\ 1 - \left(1 + \gamma \frac{y}{\sigma_u}\right)^{-1/\gamma}, & 0 \leq y \leq -\frac{\sigma_u}{\gamma} & \text{se } \gamma < 0. \end{cases} \quad (3.36)$$

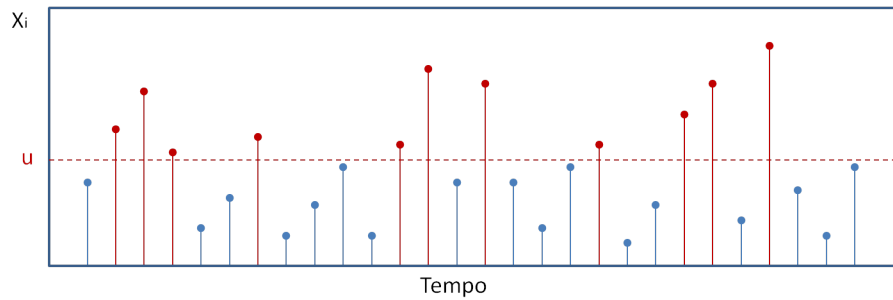


Figura 3.2: *Excedências de um nível elevado u*

No método POT (Peaks Over Threshold) não se irá trabalhar com a a.a. original, X_1, \dots, X_n . Fixando um determinado nível elevado u , essa amostra irá ser censurada e apenas serão retidas as m excedências desse nível, como pode ser visualizado na figura 3.2. De seguida, constrói-se a nova amostra, neste caso, dos excessos, através da transformação

$$Y_j := X_i - u \mid X_i > u, \quad j = 1, 2, \dots, m.$$

Com m suficientemente elevado, poderá ser ajustado o modelo GP a esta nova a.a. e, conseqüentemente, poderá ser feita inferência sobre parâmetros importantes em acontecimentos raros.

3.3.1 Escolha do *Threshold*

O *threshold* u tem um papel bastante importante na metodologia POT, logo é crucial que a sua seleção seja feita de forma razoável. Contudo, a escolha desse nível é uma questão em aberto e com alguma controvérsia. Se forem escolhidos valores pequenos para u , ter-se-á um maior número de observações e, como consequência, os estimadores terão uma menor variância, mas, em compensação, o seu viés será maior. Para além disso, a aproximação à distribuição GP será questionável, uma vez que não se tem $u \rightarrow x^F$. Já se for escolhido um valor elevado para o nível u , esse viés é menor mas os estimadores apresentarão comportamentos com uma grande variabilidade. Logo, será necessário haver um compromisso entre estes dois lados para se conseguir obter estimadores com valores baixos quer para a variância, quer para o viés.

Em Gong (2012), é referido que, em geral, é aplicado um de três critérios de seleção do *threshold*.

No primeiro método, considerado o mais simples, escolhe-se como *threshold*, o valor de um determinado quantil empírico de probabilidade a variar entre 0.9 e 0.97.

O segundo método referido, é baseado em Davison and Smith (1990), onde é proposto para a escolha de um nível u conveniente o estudo da função de excesso médio (mean excess function), já definida em (2.9), que no caso GP assume a expressão

$$e(u) := E[X - u | X > u] = E[Y | Y > 0] = \frac{\sigma_u + \gamma u}{1 - \gamma}, \quad \text{se } \gamma < 1, \quad (3.37)$$

Este resultado é obtido utilizando-se a definição de função de excesso médio para os três casos da distribuição GP, com $\gamma > 0$, $\gamma < 0$ e $\gamma \neq 0$. Assim, para $0 < \gamma < 1$,

$$\begin{aligned} e(u) &= \frac{\int_u^{+\infty} (1 - H_\gamma(s)) ds}{1 - H_\gamma(u)} = \frac{\int_u^{+\infty} \left(1 + \gamma \frac{s}{\sigma_u}\right)^{-\frac{1}{\gamma}} ds}{\left(1 + \gamma \frac{u}{\sigma_u}\right)^{-\frac{1}{\gamma}}} = \\ &= \frac{1}{\left(1 + \gamma \frac{u}{\sigma_u}\right)^{-\frac{1}{\gamma}}} \cdot \lim_{A \rightarrow +\infty} \left[\frac{\sigma_u}{\gamma - 1} \left(1 + \gamma \frac{s}{\sigma_u}\right)^{\frac{\gamma-1}{\gamma}} \right]_u^A = \\ &= \frac{1}{\left(1 + \gamma \frac{u}{\sigma_u}\right)^{-\frac{1}{\gamma}}} \cdot \left(0 - \frac{\sigma_u}{\gamma - 1} \left(1 + \gamma \frac{u}{\sigma_u}\right)^{\frac{\gamma-1}{\gamma}}\right) = \\ &= -\frac{\sigma_u}{\gamma - 1} \left(1 + \gamma \frac{u}{\sigma_u}\right) = \\ &= \frac{\sigma_u + \gamma u}{1 - \gamma}. \end{aligned}$$

Para $\gamma < 0$, a demonstração é muito semelhante, no entanto, neste caso, o limite superior do suporte é finito e igual a $-\frac{\sigma_u}{\gamma}$, logo o extremo superior do integral coincidirá com este valor. Por fim, para $\gamma = 0$ tem-se

$$\begin{aligned} e(u) &= \frac{\int_u^{+\infty} (1 - H_0(s)) ds}{1 - H_0(u)} = \frac{\int_u^{+\infty} \exp\left(-\frac{s}{\sigma_u}\right) ds}{\exp\left(-\frac{u}{\sigma_u}\right)} = \\ &= \frac{1}{\exp\left(-\frac{u}{\sigma_u}\right)} \cdot \lim_{A \rightarrow +\infty} \left[-\sigma_u \exp\left(-\frac{s}{\sigma_u}\right) \right]_u^A = \\ &= \frac{1}{\exp\left(-\frac{u}{\sigma_u}\right)} \cdot \left(0 + \sigma_u \exp\left(-\frac{u}{\sigma_u}\right)\right) = \\ &= \sigma_u = \frac{\sigma_u + \gamma u}{1 - \gamma}. \end{aligned}$$

Na prática, $e(u)$ é substituída pela sua contrapartida empírica, $\hat{e}_n(u)$, que é definida,

com base na amostra de dados observados, x_1, \dots, x_n , por

$$\hat{e}_n(u) := \frac{\sum_{i=1}^n x_i \cdot \mathbf{1}_{(u,+\infty)}(x_i)}{\sum_{i=1}^n \mathbf{1}_{(u,+\infty)}(x_i)} - u, \quad (3.38)$$

$$\text{em que } \mathbf{1}_{(u,+\infty)}(x_i) := \begin{cases} 1 & \text{se } x_i > u, \\ 0 & \text{cc} \end{cases}.$$

Usualmente considera-se que $u = x_{n-k:n}$, com $k = 1, 2, \dots, n-1$, e assim tem-se que $\sum_{i=1}^n x_i \cdot \mathbf{1}_{(u,+\infty)}(x_i) = \sum_{j=1}^k x_{n-j+1:n}$, com $k \equiv \#\{x_i : x_i > u\}$. Sendo assim, as estimativas dos excessos médios poderão ser dadas por

$$\hat{e}_n(x_{n-k:n}) = \frac{1}{k} \sum_{j=1}^k x_{n-j+1:n} - x_{n-k:n}. \quad (3.39)$$

Repare-se que a função de excesso médio (3.37), para a distribuição GP, é linear em u . Assim, quando se representa o ME-plot empírico, isto é, o gráfico obtido através de

$$\{(x_{n-k:n}, \hat{e}_n(x_{n-k:n})), 1 \leq k \leq n-1\},$$

espera-se que haja uma linearidade à direita de $x_{n-k:n}$. O *threshold* u deverá ser escolhido, precisamente no ponto, a partir do qual se verifica essa linearidade, podendo esse ponto estar ou não presente na amostra (mas sempre tendo em atenção que u não pode exceder o limite superior do suporte da variável). Contudo, na parte final do ME-plot, em geral, essa linearidade não acontece, uma vez que estão a ser calculadas médias de um número reduzido de excessos elevados. Sendo assim, quando se pretende determinar o valor de u , esses pontos do ME-plot poderão ser ignorados.

O terceiro procedimento sugerido, e que também pode ser encontrado em Coles (2001) e em Furrer *et al.* (2010), consiste na escolha de alguns candidatos a *threshold* e, de seguida, para cada um deles, estimam-se os parâmetros do modelo GP. Um valor adequado para o *threshold* será aquele, a partir do qual, os estimadores dos parâmetros de forma γ e de escala σ , com este último devidamente transformado, se tornam estáveis. De facto, se o modelo GP com parâmetros de forma γ e de escala σ_{u_0} for o correto para os excessos acima do nível u_0 , então, para qualquer *threshold* $u > u_0$, os novos excessos continuarão a ter distribuição GP com a mesma forma mas, com a escala alterada para

$$\sigma_u = \sigma_{u_0} + \gamma(u - u_0).$$

Se for usada uma nova versão do parâmetro de escala,

$$\sigma^* = \sigma_u - \gamma u$$

então, ambos os parâmetros do modelo, σ^* e γ , serão constantes para qualquer *threshold* acima de u_0 . Este método é considerado uma ferramenta bastante útil para avaliar a escolha do *threshold* u , pois, recorre a representações gráficas, o que torna mais fácil a visualização do comportamento dos dois parâmetros mencionados, em função dos valores de u .

Nesta dissertação serão usados os dois últimos métodos para a seleção de um *threshold* adequado.

3.3.2 Estimadores de Máxima Verosimilhança dos Parâmetros

Tal como foi apresentado para os parâmetros da distribuição GEV, o método de MV poderá ser usado para obter estimadores dos parâmetros da distribuição GP. Seja Y_1, \dots, Y_m , a a.a. de m excessos obtida a partir da a.a. original, X_1, \dots, X_n , para um *threshold* u . Tem-se que, para todo o $i = 1, \dots, m$, Y_i é i.i.d. a $Y \sim GP$, com f.d. $H_\gamma(y|\sigma_u)$, definida em (3.36). A log-verosimilhança para Y_1, \dots, Y_m , quando $\gamma \neq 0$ é dada por

$$\begin{aligned} l(\gamma, \sigma_u | y_1, \dots, y_m) &= \log L(\gamma, \sigma_u | y_1, \dots, y_m) = \\ &= -m \log \sigma_u - \left(\frac{1}{\gamma} + 1 \right) \sum_{i=1}^m \log \left(1 + \gamma \frac{y_i}{\sigma_u} \right), \end{aligned} \quad (3.40)$$

com $1 + \gamma \frac{y_i}{\sigma_u} > 0$, $i = 1, \dots, m$.

Para $\gamma = 0$, a log-verosimilhança é

$$\begin{aligned} l(0, \sigma_u | y_1, \dots, y_m) &= \log L(0, \sigma_u | y_1, \dots, y_m) = \\ &= -m \log \sigma_u - \frac{1}{\sigma_u} \sum_{i=1}^m y_i. \end{aligned} \quad (3.41)$$

Através da maximização da função de log-verosimilhança, obter-se-ão os estimadores de MV, $\hat{\gamma}$ e $\hat{\sigma}_u$, de γ e σ_u , respetivamente. Para tal, resolvem-se as equações

$$\frac{\partial l(\gamma, \sigma_u | y_1, \dots, y_m)}{\partial \gamma} = 0 \quad \text{e} \quad \frac{\partial l(\gamma, \sigma_u | y_1, \dots, y_m)}{\partial \sigma_u} = 0,$$

para o caso $\gamma \neq 0$, e

$$\frac{\partial l(0, \sigma_u | y_1, \dots, y_m)}{\partial \sigma_u} = 0,$$

quando $\gamma = 0$.

É de lembrar que, no caso $\gamma = 0$, a f.d. $H_0(y|\sigma_u)$ não é mais do que a f.d. de um modelo exponencial, logo consegue-se obter explicitamente o estimador de MV para σ_u , sendo dado por $\hat{\sigma}_u = \bar{Y}$.

No caso $\gamma \neq 0$ é vantajoso, em termos computacionais, proceder-se à reparametrização sugerida por Davison (1984),

$$(\gamma, \sigma_u) \rightsquigarrow (\gamma, \tau), \text{ com } \tau := \frac{\gamma}{\sigma_u},$$

e substituindo em (3.40), resulta

$$\begin{aligned} l(\gamma, \tau|y_1, \dots, y_m) &= \log L(\gamma, \tau|y_1, \dots, y_m) = \\ &= -m \log \gamma + m \log \tau - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^m \log(1 + \tau y_i), \end{aligned} \quad (3.42)$$

com $1 + \tau y_i > 0$, $i = 1, \dots, m$.

Os estimadores de MV $(\hat{\gamma}, \hat{\tau})$ dos parâmetros (γ, τ) , são obtidos, usando processos numéricos, a partir de

$$\frac{1}{\hat{\tau}} - \left(\frac{1}{\hat{\gamma}} + 1\right) \cdot \frac{1}{m} \cdot \sum_{i=1}^m \frac{Y_i}{1 + \hat{\tau} Y_i} = 0, \quad (3.43)$$

$$\text{com } \hat{\gamma} = \frac{1}{m} \cdot \sum_{i=1}^m \log(1 + \hat{\tau} Y_i).$$

3.3.3 Estimadores dos Momentos Ponderados de Probabilidade dos Parâmetros

O método MPP para a distribuição GP foi trabalhado por Hosking and Wallis (1987). Os momentos ponderados de probabilidade, $M_{p,r,s}$, são obtidos através da expressão definida em (3.3), e, para a distribuição GP com $\gamma \neq 0$, considerando $p = 1$, $r = 0$ e $s = 0, 1, \dots$, estes tomam a forma

$$M_{1,0,s} = \frac{\sigma_u}{(s+1) \cdot (s+1-\gamma)}, \quad \gamma < 1. \quad (3.44)$$

Note-se que os momentos GP só existem para ordens r tais que $\gamma < \frac{1}{r}$.

Também aqui se procede à substituição de $M_{1,0,s}$ pela sua contrapartida empírica $\hat{M}_{1,0,s}$, dada por

$$\hat{M}_{1,0,s} = \frac{1}{m} \cdot \sum_{i=1}^m \left(\prod_{k=1}^s \frac{m-i-k+1}{m-k} \right) \cdot Y_{i:m}. \quad (3.45)$$

Para se obter os estimadores MPP para os parâmetros de forma e de escala, bastará considerar $s = 0$ e $s = 1$ nas expressões (3.44) e (3.45) e igualá-las de seguida. Obter-se-á, assim,

$$\begin{cases} \hat{\gamma} = 2 - \frac{\hat{M}_{1,0,0}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}} \\ \hat{\sigma}_u = \frac{2\hat{M}_{1,0,0} \cdot \hat{M}_{1,0,1}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}}. \end{cases} \quad (3.46)$$

A aplicação do método MPP não é isenta de problemas, pois, para $\gamma \geq 1$, os estimadores MPP não existem. Para além disso, a sua implementação prática poderá conduzir a valores inadmissíveis. Uma situação típica acontece quando $\gamma < 0$, em que a inferência poderá ir além do limite superior do suporte, que, como se sabe, neste caso é finito.

Quer os estimadores de MV, quer os estimadores dos MPP, são assintoticamente normais. Em Hosking and Wallis (1987) poderão ser vistos mais detalhes e, em particular, demonstram que quando $0 \leq \gamma \leq 0.4$ e a dimensão da amostra é pequena, os estimadores MPP dão resultados melhores que os estimadores de MV, uma vez que apresentam uma menor dispersão.

3.3.4 Estimação de Outros Parâmetros de Interesse em EVT

Na metodologia POT, também é possível estimar qualquer uma das quantidades apresentadas na secção 3.2.3, para o método dos máximos anuais.

Invertendo a f.d. da GP, $H_\gamma(y|\sigma_u)$, obtêm-se os quantis extremos de probabilidade $1 - p$ (também chamados de quantis extremos de probabilidade de excedência p). Estes também podem ser escritos à custa da função quantil de cauda, definida em 2.4. Assim,

$$q_{1-p} = U_{H_\gamma} \left(\frac{1}{p} \right) = H_\gamma^\leftarrow(1-p|\sigma_u) = \begin{cases} \frac{\sigma_u}{\gamma}(p^{-\gamma} - 1), & \gamma \neq 0, \\ -\sigma_u \log p, & \gamma = 0. \end{cases} \quad (3.47)$$

No caso de $\gamma < 0$, o limite superior do suporte da GP $H_\gamma(y|\sigma_u)$ é finito e dado por

$$U_{H_\gamma}(\infty) = q_1 = H_\gamma^\leftarrow(1|\sigma_u) = -\frac{\sigma_u}{\gamma}. \quad (3.48)$$

Substituindo, nas expressões (3.47) e (3.48), os parâmetros γ e σ_u pelas suas estimativas de MV ou MPP, obtêm-se, respetivamente, as estimativas dos quantis e do limite superior do suporte do modelo GP.

Contudo, os resultados apresentados em (3.47) e (3.48) são para a distribuição de $Y|Y > 0$, que como já foi referido, é aproximadamente GP, com f.d. $H_\gamma(y|\sigma_u)$. Se se trabalhar com a v.a. original X com f.d. desconhecida F , é preciso recorrer-se à relação (2.12), onde se representa a distribuição condicional dos excessos acima de u , e ao facto dessa distribuição poder ser aproximada a um modelo GP, ou seja,

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(y + u) - F(u)}{1 - F(u)} \approx H_\gamma(y|\sigma_u).$$

Considerando que $x = y + u$, vem que

$$\begin{aligned} F_u(x - u) &= \frac{F(x) - F(u)}{1 - F(u)} \Leftrightarrow \\ &\Leftrightarrow F(x) = F_u(x - u) \cdot (1 - F(u)) + F(u) \Leftrightarrow \\ \Leftrightarrow 1 - F(x) &= 1 - F(u) - F_u(x - u) \cdot (1 - F(u)) \Leftrightarrow \\ \Leftrightarrow 1 - F(x) &= (1 - F(u)) \cdot (1 - F_u(x - u)) \Leftrightarrow \\ \Leftrightarrow \bar{F}(x) &= \bar{F}(u) \cdot (1 - F_u(x - u)), \end{aligned}$$

e assim tem-se que

$$\bar{F}(x) = \bar{F}(u) \cdot (1 - H_\gamma(x - u|\sigma_u)).$$

Estimando $\bar{F}(u)$ pela frequência relativa, $\frac{m}{n}$, das observações que excedem o nível u na amostra original de dimensão n , X_1, \dots, X_n , e substituindo os parâmetros da distribuição GP pelos seus estimadores de MV ou MPP, vem que

$$\hat{\bar{F}}(x) = \frac{m}{n} \cdot (1 - H_\gamma(x - u|\hat{\sigma}_u)). \quad (3.49)$$

A expressão (3.49) não é mais do que o estimador da Probabilidade de excedência de um nível x elevado.

O estimador para um quantil elevado, q_{1-p} , de F , poderá ser obtido considerando $\hat{\bar{F}}(x) = p$ e resolvendo em ordem a x . Obtém-se assim,

$$\hat{q}_{1-p} = \hat{U}\left(\frac{1}{p}\right) = \hat{F}^{\leftarrow}(1 - p) = \begin{cases} u + \frac{\hat{\sigma}_u}{\hat{\gamma}} \cdot \left(\left(\frac{np}{m}\right)^{-\hat{\gamma}} - 1 \right) & \text{se } \gamma \neq 0, \\ u - \hat{\sigma}_u \cdot \log\left(\frac{np}{m}\right) & \text{se } \gamma = 0. \end{cases} \quad (3.50)$$

Para se obter o nível de retorno de T -anos basta calcular o quantil de probabilidade $1 - \frac{1}{T}$, como já indicado. Assim, para se obter um seu estimador basta fazer

$$\hat{U}(T) = \hat{q}_{1-\frac{1}{T}}. \quad (3.51)$$

Já no caso de se pretender estimar o período de retorno de um nível v , o seu estimador pode ser dado por

$$\hat{T}_v = \frac{1}{\hat{F}(v)}. \quad (3.52)$$

No caso do parâmetro de forma ser negativo, $\gamma < 0$, o limite superior do suporte é finito e pode ser obtida uma sua estimativa, através de

$$\hat{x}^F = \hat{U}(\infty) = \hat{q}_1 = u - \frac{\hat{\sigma}_u}{\hat{\gamma}}. \quad (3.53)$$

3.3.5 Intervalos de Confiança para os Parâmetros

Os intervalos de confiança para os parâmetros da distribuição GP, (γ, σ_u) , e para outros parâmetros referido na secção anterior, na abordagem POT, decorrem da aproximação à distribuição Normal dos estimadores de MV e MPP, apoiada pela normalidade assintótica de valor médio nulo para $\sqrt{m} \cdot (\hat{\gamma} - \gamma)$, no caso de se pretender intervalos de confiança para o parâmetro de forma γ , e para $\sqrt{m} \cdot \left(\hat{U}\left(\frac{1}{p}\right) - U\left(\frac{1}{p}\right) \right)$, se o objetivo forem os quantis elevados. Em Beirlant *et al.* (2004) é demonstrada a construção deste tipo de intervalos, que, como é sabido, são centrados nas respetivas estimativas pontuais dos parâmetros, pela simetria da Normal.

Tal como foi feito no método dos máximos anuais, poderão ser obtidas estimativas intervalares de melhor qualidade para a abordagem POT, usando a função de profile log-likelihood. Para o parâmetro γ , esta assume a forma

$$\log L_p(\gamma) = \max_{\sigma_u | \gamma} \log L(\sigma_u | \gamma, y_1, \dots, y_m).$$

Usando argumentos semelhantes para o caso da distribuição GEV, o intervalo com $(1 - \alpha)100\%$ de confiança para γ , baseado na função de profile log-likelihood é dado por

$$\left\{ \gamma : \log L_p(\gamma) \geq \log L_p(\hat{\gamma}) - \frac{\chi_1^2(1 - \alpha)}{2} \right\}. \quad (3.54)$$

Para se obter intervalos de confiança, baseados na função de profile log-likelihood, para outros parâmetros, o caminho a seguir será semelhante ao apresentado na secção 3.2.4.

3.3.6 Testes de Hipóteses para o Modelo GP

No método POT, e pelas mesmas razões apresentadas no método dos máximos anuais, tem-se, mais uma vez, o famoso trilema estatístico. Assim, as hipóteses em teste serão:

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma \neq 0, \quad (3.55)$$

ou seja, pretende-se testar se a distribuição Exponencial, que corresponde ao caso em que $\gamma = 0$ na distribuição GP, é ou não um bom modelo para se ajustar aos excessos acima de um nível elevado. Porém, também se pode ter hipóteses alternativas mais específicas, ou seja, uma hipótese alternativa unilateral Beta (modelo GP com $\gamma < 0$),

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma < 0, \quad (3.56)$$

ou uma hipótese alternativa unilateral Pareto (modelo GP com $\gamma > 0$),

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma > 0. \quad (3.57)$$

Existe diversa bibliografia para este tipo de testes, em especial, podem ser destacados Gomes and van Monfort (1986), Marohn (2000), Reiss and Thomas (2007) e Kozubowski *et al.* (2009).

Também poderão se aplicados os testes de ajustamento ao modelo GP. Estes foram estudados, entre outros, por Choulakian and Stephens (2001) e Lilliefors (1969), sendo que este último apresentou o caso especial do teste de Kolmogorov-Smirnov, aplicado à distribuição Exponencial com parâmetros desconhecidos.

O primeiro teste apresentado, e que pode ser aplicado aos três conjuntos de hipóteses definidos, é apresentado em Marohn (2000). Sejam (W_1, \dots, W_m) as m excedências acima do *threshold* u , obtidas da a.a. inicial, a estatística de teste é dada por

$$T_m = \frac{1}{2} \cdot \left(\frac{S_W^2}{(\bar{W} - u)^2} - 1 \right),$$

onde S_W^2 é a variância amostral dada por $S_W^2 = \frac{1}{m} \cdot \sum_{i=1}^m (W_i - \bar{W})^2$.

Sob a validade de H_0 , tem-se que

$$T_m^* = \sqrt{m} \cdot T_m \xrightarrow[m \rightarrow \infty]{d} Z \sim N(0, 1), \quad (3.58)$$

e, ao nível assintótico α , e sendo z_ϵ o quantil de probabilidade ϵ de uma Normal standard, H_0 é rejeitada se

$$|T_m^*| \geq z_{1-\frac{\alpha}{2}} \quad \text{ou se} \quad T_m^* \leq z_\alpha \quad \text{ou se} \quad T_m^* \geq z_{1-\alpha},$$

para os testes (3.55), (3.56) e (3.57), respetivamente. Os p-values correspondentes poderão ser obtidos através de

$$p(T_m^*) = 2 - 2\Phi(|T_m^*|) \quad \text{ou} \quad p(T_m^*) = \Phi(T_m^*) \quad \text{ou} \quad p(T_m^*) = 1 - \Phi(T_m^*),$$

e, neste caso, a hipótese nula será rejeitada, ao nível α , se $p(T_m^*) \leq \alpha$.

Marohn (2000) mostrou, através de simulação, que esta estatística de teste para a versão bilateral é enviesada e com uma potência baixa para amostras de pequena ou moderada dimensão ($m < 500$).

Um segundo teste, para as hipóteses definidas em (3.55), é o já conhecido TRV. Sendo (W_1, \dots, W_m) a a.a. das excedências, onde $W_i \sim H_\gamma$, para $i = 1, \dots, m$, com H_γ definida em (3.35), e $l(\gamma, u, \sigma_u | w_1, \dots, w_m)$ e $l(0, u, \sigma_u | w_1, \dots, w_m)$ as funções de log-verosimilhança para H_γ e H_0 , respetivamente, a estatística TRV (a já mencionada *Deviance*) é dada por

$$\mathcal{D} = 2 \cdot \left\{ l(\hat{\gamma}_{H_\gamma}, u, \hat{\sigma}_{u_{H_\gamma}} | w_1, \dots, w_m) - l(0, u, \hat{\sigma}_{u_{H_0}} | w_1, \dots, w_m) \right\}, \quad (3.59)$$

onde $\hat{\sigma}_{u_{H_0}}$ e $(\hat{\gamma}_{H_\gamma}, \hat{\sigma}_{u_{H_\gamma}})$ são os estimadores de MV para os parâmetros dos modelos H_0 e H_γ , respetivamente.

Sob a validade da hipótese nula,

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

Reiss and Thomas (2007) recomendam a correção de Bartlett, ficando assim

$$\mathcal{D}^* = \frac{\mathcal{D}}{1 + \frac{4}{m}} \xrightarrow[m \rightarrow \infty]{d} Q \sim \chi_1^2. \quad (3.60)$$

Para o teste (3.55), ao nível assintótico α , a hipótese nula é rejeitada se

$$\mathcal{D}^* \geq \chi_{1;1-\alpha}^2,$$

em que $\chi_{1;1-\alpha}^2$ é o quantil de probabilidade $1 - \alpha$ de um qui-quadrado com 1 g.l. O p-value pode ser obtido por

$$p(\mathcal{D}^*) = 1 - P(Q \leq \mathcal{D}^*).$$

Um outro teste, proposto por Gomes and van Monfort (1986), e que será usado nas situações (3.56) e (3.57), recorre ao máximo e à mediana da amostra de excedências. A sua estatística de teste é

$$G_m = \frac{W_{m:m}}{W_{[\frac{m}{2}]+1:m}},$$

e sob a validade de H_0 , tem-se que

$$G_m^* = \log 2 \cdot G_m - \log m \xrightarrow[m \rightarrow \infty]{d} U \sim \Lambda. \quad (3.61)$$

Para o nível assintótico α , H_0 é rejeitada se

$$G_m^* \leq \mathcal{G}_\alpha \quad \text{ou se} \quad G_m^* \geq \mathcal{G}_{1-\alpha},$$

para os testes (3.56) e (3.57), respetivamente, sendo \mathcal{G}_α e $\mathcal{G}_{1-\alpha}$ os quantis de probabilidade α e $1 - \alpha$, para uma distribuição Gumbel standard. Alternativamente poder-se-á calcular os correpondentes p-values:

$$p(G_m^*) = \Lambda(G_m^*) \quad \text{ou} \quad p(G_m^*) = 1 - \Lambda(G_m^*).$$

No caso da dimensão da amostra ser pequena ou moderada, Gomes and van Monfort (1986) apresentam uma tabela com os pontos críticos para esta estatística de teste, para os níveis de 0.05 e 0.1, obtidos por simulação. Parte dessa tabela encontra-se representada na Tabela 3.3 e onde $a \downarrow$ e $a \uparrow$ significam, respetivamente, valores menores e maiores do que a .

Tabela 3.3: *Pontos críticos, obtidos por simulação, para a estatística de teste G_m^**

Estatística	m	Região crítica para		Região crítica para	
		$H_1 : \gamma < 0$		$H_1 : \gamma > 0$	
		.10	.05	.10	.05
G_m^*	20	-0.89↓	-1.19↓	3.15↑	4.04↑
	100	-0.94↓	-1.21↓	2.55↑	3.40↑
	250	-0.86↓	-1.13↓	2.57↑	3.37↑
	∞	-0.83↓	-1.09↓	2.25↑	2.97↑

Gomes and van Monfort (1986)

Como já foi referido, também aqui poderão ser aplicados testes à qualidade do ajustamento, nomeadamente os testes de Kolmogorov-Smirnov, Cramér-von Mises e Anderson-Darling.

O teste de Kolmogorov-Smirnov para a distribuição Exponencial, com parâmetros desconhecidos, recebeu especial atenção por parte de Lilliefors (1969). A estatística de Kolmogorov-Smirnov adaptada para esta hipótese nula, correspondente ao modelo Exponencial, é dada por

$$D_m = \max_{1 \leq i \leq m} \left\{ \left| 1 - \exp\left(-\frac{Y_{i:m}}{\hat{\sigma}_u}\right) - \frac{i}{m} \right|, \left| 1 - \exp\left(-\frac{Y_{i:m}}{\hat{\sigma}_u}\right) - \frac{i-1}{m} \right| \right\}, \quad (3.62)$$

onde $\hat{\sigma}_u$ é o estimador de MV de σ_u , do modelo Exponencial.

A hipótese nula será rejeitada, a um determinado nível, se o valor observado da estatística de teste for superior ao ponto crítico associado a esse nível. Na tabela 3.4 estão representados alguns desses pontos críticos.

Em relação às estatísticas de Cramér-von Mises e Anderson-Darling (Choulakian and Stephens, 2001), estas são dadas por:

Tabela 3.4: Pontos críticos simulados para a estatística de Kolmogorov-Smirnov adaptada para o modelo Exponencial com parâmetros desconhecidos.

Estatística	Níveis de significância para D_m			
	m	.10	.05	.01
D_m	5	.406	.442	.504
	10	.295	.325	.380
	15	.244	.269	.315
	20	.212	.234	.278
	30	.174	.192	.226
	>30	$\frac{.96}{\sqrt{m}}$	$\frac{1.06}{\sqrt{m}}$	$\frac{1.25}{\sqrt{m}}$
	Lilliefors (1969)			

i. Cramér-von Mises

$$W_m^2 = \sum_{i=1}^m \left(H_{\hat{\gamma}}(Y_{i:m} | \hat{\sigma}_{u_{H_\gamma}}) - \frac{2i-1}{2m} \right)^2 + \frac{1}{12m}; \quad (3.63)$$

ii. Anderson-Darling

$$A_m^2 = -m - \frac{1}{m} \sum_{i=1}^m \left\{ (2i-1) \log(H_{\hat{\gamma}}(Y_{i:m} | \hat{\sigma}_{u_{H_\gamma}})) + (2m+1-2i) \log(1 - H_{\hat{\gamma}}(Y_{i:m} | \hat{\sigma}_{u_{H_\gamma}})) \right\}, \quad (3.64)$$

com $\hat{\gamma}$ e $\hat{\sigma}_{u_{H_\gamma}}$ os estimadores de MV para a f.d. GP, H_γ .

Nestes dois casos, a distribuição em H_0 , é uma GP com parâmetros desconhecidos. Essa hipótese será rejeitada se o valor observado das estatísticas excederem os respectivos pontos críticos associados ao nível de significância considerado. Estes valores encontram-se tabelados em Choulakian and Stephens (2001), e estão parcialmente reproduzidos nas tabelas 3.5 e 3.6.

3.4 Método das Maiores Observações (MO)

No capítulo anterior, mais propriamente no teorema 2.22, foi visto que as k maiores observações da a.a., X_1, \dots, X_n , depois de convenientemente linearizadas por constantes $a_n > 0$ e $b_n \in \mathbb{R}$, isto é,

$$(W_1, \dots, W_k) = \left(\frac{X_{n:n} - b_n}{a_n}, \dots, \frac{X_{n-k+1:n} - b_n}{a_n} \right),$$

Tabela 3.5: Pontos críticos simulados para a estatística de Cramér-von Mises adaptada para o modelo GP com parâmetros desconhecidos.

Estatística	γ	Níveis de significância		
		.10	.05	.01
W_m^2	.9	.094	.115	.165
	.5	.101	.124	.179
	.1	.116	.144	.210
	0	.124	.153	.224
	-.1	.129	.160	.236
	-.5	.174	.222	.338

Choulakian and Stephens (2001)

Tabela 3.6: Pontos críticos simulados para a estatística de Anderson-Darling adaptada para o modelo GP com parâmetros desconhecidos.

Estatística	γ	Níveis de significância		
		.10	.05	.01
A_m^2	.9	.641	.771	1.086
	.5	.685	.830	1.180
	.1	.766	.935	1.348
	0	.796	.974	1.409
	-.1	.831	1.020	1.481
	-.5	1.061	1.321	1.958

Choulakian and Stephens (2001)

são bem modeladas pela f.d.p. conjunta dada por

$$g_{X_{n:n}, \dots, X_{n-k+1:n}}(w_1, \dots, w_k) := G_\gamma(w_k) \cdot \prod_{i=1}^k \frac{g_\gamma(w_i)}{G_\gamma(w_i)}, \quad \text{para } w_1 > \dots > w_k, \quad (3.65)$$

onde $g_\gamma(w) := \frac{\partial G_\gamma(w)}{\partial w}$ é a f.d.p. do modelo GEV com f.d. G_γ . Este modelo tem a designação de GEV-multivariado ou GEV-processo extremal e foi introduzido por Weissman (1978). Gomes and Alpuim (1986) e Gomes (1989), por exemplo, utilizaram este modelo para a escolha estatística de G_0 contra G_γ , com $\gamma \neq 0$, ou contra alternativas unilaterais $\gamma < 0$ ou $\gamma > 0$.

Em inferência estatística para acontecimentos raros, uma outra abordagem paramétrica possível é, precisamente, modelar estas k maiores observações disponíveis da amostra, com o comportamento conjunto indicado em (3.65).

3.4.1 Estimação do Modelo para as MO

Nesta abordagem procede-se a uma partição da a.a., X_1, \dots, X_n , em m blocos (geralmente um bloco corresponde a um ano), cada um com dimensão q , tal como para o método dos Máximos Anuais. Para cada um desses blocos, extrai-se as maiores k observações ($k < q$), obtendo-se, assim, um conjunto de m vetores aleatórios k -dimensionais i.i.d., $\{\underline{x}^{(j)}\}_{j=1, \dots, m}$, ao qual se pretende ajustar o modelo descrito anteriormente. Na figura 3.3, é exemplificada uma situação com $m = 4$ blocos, cada um com dimensão $q = 7$, de onde se escolhem as $k = 3$ maiores observações de cada bloco.

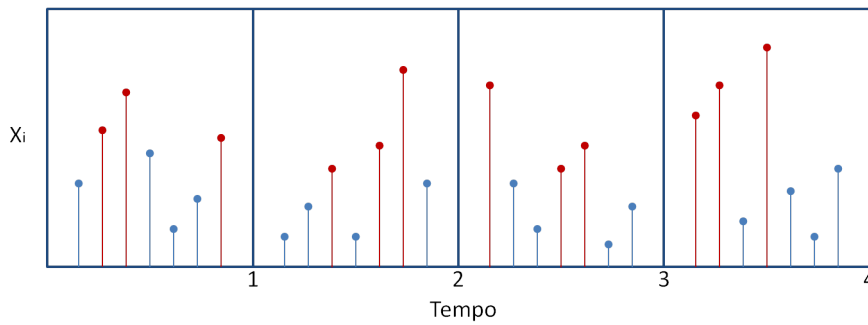


Figura 3.3: *Maiores Observações*

Surge assim a necessidade de estimação dos parâmetros de forma, localização e escala, para posterior inferência sobre os parâmetros de interesse em acontecimentos raros. Irá ser usado o método de MV para se obter estimativas dos parâmetros do modelo.

A função de verosimilhança para este modelo é obtida através de (3.65) e efetuando os produtos por bloco, ou seja,

$$L(\gamma, \lambda, \delta | \underline{x}^{(1)}, \dots, \underline{x}^{(m)}) = \prod_{j=1}^m \left(G_\gamma(x_{q-k+1:q}^{(j)} | \lambda, \delta) \times \prod_{i=1}^k \frac{g_\gamma(x_{q-i+1:q}^{(j)} | \lambda, \delta)}{G_\gamma(x_{q-i+1:q}^{(j)} | \lambda, \delta)} \right).$$

Substituindo g_γ e G_γ pelas expressões da f.d.p. e da f.d., respetivamente, do modelo GEV, obtém-se:

- para $\gamma \neq 0$

$$L(\gamma, \lambda, \delta | \underline{x}^{(1)}, \dots, \underline{x}^{(m)}) = \prod_{j=1}^m \left\{ \exp \left(- \left(1 + \gamma \cdot \frac{x_{q-k+1:q}^{(j)} - \lambda}{\delta} \right)^{-\frac{1}{\gamma}} \right) \times \prod_{i=1}^k \frac{1}{\delta} \left(1 + \gamma \cdot \frac{x_{q-i+1:q}^{(j)} - \lambda}{\delta} \right)^{-\frac{1}{\gamma} - 1} \right\}, \quad (3.66)$$

com $1 + \gamma \cdot \frac{x_{q^{-i+1}:q}^{(j)} - \lambda}{\delta} > 0$, $i = 1, \dots, k$, $j = 1, \dots, m$;

- para $\gamma = 0$

$$L(0, \lambda, \delta | \underline{x}^{(1)}, \dots, \underline{x}^{(m)}) = \prod_{j=1}^m \left\{ \exp \left(- \exp \left(- \frac{x_{q^{-k+1}:q}^{(j)} - \lambda}{\delta} \right) \right) \times \right. \\ \left. \times \prod_{i=1}^k \frac{1}{\delta} \cdot \exp \left(- \frac{x_{q^{-i+1}:q}^{(j)} - \lambda}{\delta} \right) \right\}. \quad (3.67)$$

A função log-verosimilhança será dada por

- para $\gamma \neq 0$

$$l(\gamma, \lambda, \delta | \underline{x}^{(1)}, \dots, \underline{x}^{(m)}) = \sum_{j=1}^m \left\{ - \left(1 + \gamma \cdot \frac{x_{q^{-k+1}:q}^{(j)} - \lambda}{\delta} \right)^{-\frac{1}{\gamma}} - \right. \\ \left. - \sum_{i=1}^k \left(\log \delta + \left(\frac{1}{\gamma} + 1 \right) \cdot \log \left(1 + \gamma \cdot \frac{x_{q^{-i+1}:q}^{(j)} - \lambda}{\delta} \right) \right) \right\}, \quad (3.68)$$

com $1 + \gamma \cdot \frac{x_{q^{-i+1}:q}^{(j)} - \lambda}{\delta} > 0$, $i = 1, \dots, k$, $j = 1, \dots, m$;

- para $\gamma = 0$

$$l(0, \lambda, \delta | \underline{x}^{(1)}, \dots, \underline{x}^{(m)}) = \sum_{j=1}^m \left\{ - \exp \left(- \frac{x_{q^{-k+1}:q}^{(j)} - \lambda}{\delta} \right) - \right. \\ \left. - \sum_{i=1}^k \left(\log \delta + \frac{x_{q^{-i+1}:q}^{(j)} - \lambda}{\delta} \right) \right\}. \quad (3.69)$$

A log-verosimilhança terá que ser maximizada numericamente para se obter as estimativas de MV dos parâmetros do modelo $(\gamma, \lambda, \delta)$. A teoria assintótica associada aos estimadores de MV poderá ser aplicada, por exemplo, para se obter intervalos de confiança para os parâmetros.

É de constatar que quando $k = 1$, a função de log-verosimilhança reduz-se ao caso apresentado para o modelo GEV, no método dos máximos anuais. Aliás, o modelo estatístico para as k MO fornece uma verosimilhança cujos parâmetros correspondem aos da distribuição GEV, obtidos pelo método dos máximos anuais, mas neste caso foram incorporadas mais observações extremas, da amostra original. Assim, a forma de obtenção dos parâmetros de interesse em EVT, bem como a sua interpretação será a mesma, contudo, obtém-se uma melhor precisão nas estimativas destes, devido à inclusão de mais informação.

Um outro pormenor a ter em consideração será o número de observações de topo que deverá ser retido. Este problema é semelhante ao da escolha do *threshold* na metodologia POT. Pequenos valores de k originam amostras mais pequenas, o que irá implicar uma maior variância; por sua vez, valores elevados de k conduzem a um viés maior. É preciso, portanto, ter em atenção estas duas situações e haver uma ponderação na escolha do valor para k . Na prática, é habitual seleccionar um valor tão grande quanto possível e aplicar-se, de seguida, diagnósticos adequados ao modelo.

Quando se consideram as MO de uma a.a., tal como foi apresentado para o máximo anual, poderá haver a presença de tendência nos dados. Nessa situação o procedimento a tomar será análogo ao apresentado na secção 3.2.6.

Capítulo 4

Abordagem Semi-Paramétrica em Valores Extremos

4.1 Introdução

Como já tinha sido referido no início do capítulo anterior, nesta dissertação serão apresentadas duas abordagens possíveis para fazer inferência em valores extremos. Nesse mesmo capítulo foram expostos os pontos fulcrais da abordagem paramétrica que, recorde-se, tem como principal pressuposto a existência de uma classe de modelos paramétricos que descrevam adequadamente a v.a. em estudo. Como se está a trabalhar em valores extremos e, mais propriamente, para valores elevados, esses modelos resumem-se a três casos possíveis, Gumbel, Fréchet e Weibull, os quais dependem fundamentalmente do índice de cauda, γ , e, eventualmente, de parâmetros de localização, λ , e de escala, δ . Estes três modelos poderão, também, ser reunidos numa única família, a distribuição GEV. Falta, portanto, apresentar a abordagem semi-paramétrica.

Na abordagem semi-paramétrica não se supõe nenhum modelo paramétrico, F , subjacente à amostra de observações i.i.d., X_1, \dots, X_n . Em contrapartida, a única suposição que é feita é a de que a função F pertença ao domínio de atração da distribuição GEV, isto é, $F \in \mathcal{D}(G_\gamma)$, para algum $\gamma \in \mathbb{R}$. Neste contexto, qualquer inferência respeitante à cauda da distribuição de F , pode ser baseada nas $k + 1$ observações de topo,

$$X_{n:n} \geq X_{n-1:n} \geq \dots \geq X_{n-k:n},$$

que, por sua vez, são obtidas a partir de um *threshold* aleatório, $X_{n-k:n}$. Este nível aleatório é uma e.o. intermédia superior (ver secção 2.7.3), em que

$$k \equiv k_n \rightarrow \infty, \text{ com } \frac{k_n}{n} \rightarrow 0, \text{ quando } n \rightarrow \infty.$$

A escolha de k é um problema semelhante ao da determinação do *threshold* u , na metodologia POT, apresentada na secção 3.3. A determinação deste valor é um problema em aberto, e muitos autores sugerem diversas soluções.

4.2 Condição de Regularidade de Segunda Ordem

Seguindo uma metodologia semi-paramétrica, para além da condição de primeira ordem apresentada no teorema 2.12, muitas vezes é necessário introduzir uma condição de segunda ordem que especifique a velocidade de convergência de

$$F \in \mathcal{D}(G_\gamma) \Leftrightarrow \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{se } \gamma \neq 0, \\ \log x & \text{se } \gamma = 0; \end{cases}$$

para alguma função positiva a , com $x > 0$.

É comum assumir-se a existência de uma função $A(t)$, que não mude de sinal e tenda para zero quando $t \rightarrow \infty$, tal que

$$\lim_{t \rightarrow \infty} \frac{\frac{U(tx) - U(t)}{a(t)} - \frac{x^\gamma - 1}{\gamma}}{A(t)} = H_{\gamma, \rho}(x) = \frac{1}{\rho} \left(\frac{x^{\gamma + \rho} - 1}{\gamma + \rho} - \frac{x^\gamma - 1}{\gamma} \right), \quad (4.1)$$

para todo o $x > 0$ e em que $\rho \leq 0$ é chamado de parâmetro de segunda ordem. Este parâmetro serve para controlar a velocidade de convergência do máximo linearmente normalizado para a lei limite G_γ . Para além disso, a função A é tal que

$$\lim_{t \rightarrow \infty} \frac{A(tx)}{A(t)} = x^\rho,$$

para todo o $x > 0$, isto é, $|A| \in \mathcal{RV}_\rho$ (de Haan and Stadtmüller, 1996).

A condição de regularidade de segunda ordem pode, então, ser formalizada na definição seguinte.

Definição 4.1 (Condição de Segunda Ordem). *A função U (ou a f.d. F associada) satisfaz a condição de regularidade de segunda ordem se, para uma função positiva a e para alguma função positiva ou negativa A , com $\lim_{t \rightarrow \infty} A(t) = 0$, verifica a condição (4.1).*

A função A , chamada de função auxiliar de segunda ordem, descreve a velocidade de convergência da condição de primeira ordem e é, como já referido, de variação regular de índice ρ . Se $\rho < 0$, essa velocidade é rápida. Já se $\rho = 0$, a velocidade de convergência é lenta. Para além disso, a velocidade com que a sequência intermédia k_n tende para o infinito, tem de estar de acordo com a velocidade de convergência da condição de primeira ordem que, por sua vez, é quantificada pela função A . Logo a escolha de k_n é decidida pela condição de segunda ordem.

4.3 Estimação do Índice de Valores Extremos

Recordando o item (iv) do Teorema 2.11, existe um função positiva não decrescente g , tal que

$$F \in \mathcal{D}(G_\gamma) \Leftrightarrow \lim_{t \rightarrow x^F} \frac{\overline{F}(t + x \cdot g(t))}{\overline{F}(t)} = (1 + \gamma x)^{-1/\gamma}, \quad (4.2)$$

para todo o x em que $1 + \gamma x > 0$ (de Haan and Ferreira, 2006) e γ é o EVI que, como já foi referido diversas vezes, descreve a forma da cauda da distribuição F .

Se se tomar em atenção a secção 2.6, facilmente se constata que

$$\frac{\overline{F}(t + x \cdot g(t))}{\overline{F}(t)} = P(X > t + x \cdot g(t) | X > t) = \overline{F}_{X|X>t}(t + x \cdot g(t)),$$

e sabe-se, também, que

$$\lim_{t \rightarrow x^F} \overline{F}_{X|X>t}(t + x \cdot g(t)) = \overline{H}_\gamma(x),$$

ou, equivalentemente, fazendo uma mudança de variável e usando a definição de cauda direita, obtém-se

$$\lim_{t \rightarrow x^F} F_{X|X>t}(x) = H_\gamma\left(\frac{x-t}{g(t)}\right),$$

onde H_γ é a f.d. GP, apresentada em (2.14). Assim, usando estas relações e tendo em atenção a secção 3.3.4, a condição em (4.2) é equivalente a dizer que, a partir de um determinado nível elevado, por exemplo, t , a cauda da f.d. F pode ser escrita como

$$\overline{F}(x) \simeq \overline{F}(t) \cdot \left\{ 1 - H_\gamma\left(\frac{x-t}{g(t)}\right) \right\}, \quad x > t. \quad (4.3)$$

Espera-se que esta aproximação se verifique em e.o.'s intermédias ou de extremos.

Sejam X_1, \dots, X_n v.a.'s i.i.d. com f.d. F e seja F_n a correspondente f.d. empírica dada por

$$F_n(x) = \frac{1}{n} \cdot \sum_{i=1}^n \mathbf{1}_{]-\infty, x]}(x_i). \quad (4.4)$$

Aplicando a aproximação (4.3) para o *threshold* aleatório $t = X_{n-k:n}$ e com $k = k_n \rightarrow \infty$ e $\frac{k}{n} \rightarrow 0$, quando $n \rightarrow \infty$, obtém-se

$$\overline{F}(x) \simeq \overline{F}(X_{n-k:n}) \cdot \left\{ 1 - H_\gamma\left(\frac{x - X_{n-k:n}}{g(X_{n-k:n})}\right) \right\}, \quad x > X_{n-k:n}.$$

Como $\overline{F}(X_{n-k:n}) \simeq \overline{F}_n(X_{n-k:n}) = \frac{k}{n}$ e, sendo $g(t) = a\left(\frac{1}{\overline{F}(t)}\right)$, então $g(X_{n-k:n}) = a\left(\frac{1}{\overline{F}(X_{n-k:n})}\right) = a\left(\frac{n}{k}\right)$ (ver item (iv) do teorema 2.11), vem que a relação anterior

pode ser escrita como

$$\bar{F}(x) \simeq \frac{k}{n} \cdot \left\{ 1 - H_\gamma \left(\frac{x - X_{n-k:n}}{a\left(\frac{n}{k}\right)} \right) \right\}, \quad x > X_{n-k:n}. \quad (4.5)$$

Para se conseguir concretizar esta aproximação, é, então, preciso estimar o valor do parâmetro γ e da função a no ponto $\frac{n}{k}$. Repare-se que esta aproximação é válida para qualquer $x > X_{n-k:n}$, e mesmo para $x > X_{n:n}$, ou seja, é possível de ser utilizada fora do alcance das observações, o que é, precisamente, o ponto nevrálgico da aplicação da EVT.

Pelo que se acabou de observar, na abordagem semi-paramétrica, é crucial a estimação do EVI. Para tal, vejam-se alguns estimadores clássicos para este parâmetro.

4.3.1 Estimador de Hill

Para caudas pesadas, isto é, para $\gamma > 0$, Hill (1975) propôs um estimador simples para o parâmetro γ . As suas propriedades foram estudadas por diversos autores entre os quais se pode destacar Hall (1982), Mason (1982), Marohn (1997), Embrechts *et al.* (1997) e de Haan and Ferreira (2006). A expressão do estimador de Hill, $\hat{\gamma}_{k,n}^H$, é dada por

$$\hat{\gamma}_{k,n}^H = \frac{1}{k} \cdot \sum_{i=1}^k (\log X_{n-i+1:n} - \log X_{n-k:n}). \quad (4.6)$$

Este estimador goza da propriedade de consistência, mas acrescentando uma condição extra, essa consistência poderá tornar-se forte. Estes resultados são apresentados nos dois teoremas seguintes.

Teorema 4.1 (Consistência do Estimador de Hill). *Sejam X_1, \dots, X_n, \dots , v.a.'s i.i.d. com f.d. $F \in \mathcal{D}(G_\gamma)$ e $\gamma > 0$. Então, se $k = k_n \rightarrow \infty$ e $\frac{k_n}{n} \rightarrow 0$, quando $n \rightarrow \infty$, tem-se que*

$$\hat{\gamma}_{k,n}^H \xrightarrow{P} \gamma,$$

ou seja, o estimador de Hill de γ tende em probabilidade para γ .

Teorema 4.2 (Consistência Forte do Estimador de Hill). *Sejam X_1, \dots, X_n, \dots , v.a.'s i.i.d. com f.d. $F \in \mathcal{D}(G_\gamma)$ e $\gamma > 0$. Se $k = k_n \rightarrow \infty$, $\frac{k_n}{n} \rightarrow 0$ e $\frac{k_n}{\log(\log n)} \rightarrow \infty$, quando $n \rightarrow \infty$, então*

$$\hat{\gamma}_{k,n}^H \xrightarrow{q.c.} \gamma,$$

ou seja, o estimador de Hill de γ converge quase certamente para γ .

Considerando condições suplementares para a cauda da distribuição F e para a velocidade de convergência da sucessão intermédia $k = k_n \rightarrow \infty$, mais propriamente, sob a validade da condição de segunda ordem apresentada na definição 4.1, o estimador de Hill é assintoticamente Normal:

$$\sqrt{k} (\hat{\gamma}_{k,n}^H - \gamma) \xrightarrow[n \rightarrow \infty]{d} W \sim \mathcal{N}(0, \sigma_H^2 = \gamma^2).$$

O estimador de Hill é invariante perante modificações de escala. Tal não se verifica quando as alterações ocorrem na localização. Outro ponto negativo deste estimador, tem a ver com o facto de ser sensível à velocidade de crescimento de k em relação a n , podendo apresentar um grande viés para determinadas taxas de crescimento.

Na tentativa de ultrapassar estas desvantagens, alguns autores sugeriram outras versões deste estimador. Por exemplo, a não invariância perante modificações na localização do estimador de Hill tradicional, levou Fraga Alves (2001) a considerar uma nova versão deste estimador que verificasse essa propriedade:

$$\hat{\gamma}_{k,k_0,n}^H = \frac{1}{k_0} \cdot \sum_{i=1}^k \log \left(\frac{X_{n-i+1:n} - X_{n-k:n}}{X_{n-k_0+1:n} - X_{n-k:n}} \right), \quad (4.7)$$

com um adequado $k_0 < k$.

4.3.2 Estimador de Pickands

O estimador mais simples e antigo para o EVI, é o estimador de Pickands (Pickands, 1975). Para um EVI real, ou seja, $\gamma \in \mathbb{R}$, o estimador de Pickand tem a forma funcional

$$\hat{\gamma}_{k,n}^P = \frac{1}{\log 2} \cdot \log \left(\frac{X_{n-\lfloor \frac{k}{4} \rfloor + 1:n} - X_{n-\lfloor \frac{k}{2} \rfloor + 1:n}}{X_{n-\lfloor \frac{k}{2} \rfloor + 1:n} - X_{n-k+1:n}} \right), \quad \frac{k}{4} = 1, 2, \dots, \left[\frac{n}{4} \right], \quad (4.8)$$

em que $[x]$ denota, como habitualmente, a parte inteira de x .

A consistência deste estimador do EVI foi provada por Pickands (1975) para um $\gamma \in \mathbb{R}$. O estudo sobre a consistência forte e as propriedades assintóticas deste estimador foi feito, posteriormente, por Dekkers and de Haan (1989).

Teorema 4.3 (Consistência do Estimador de Pickands, Pickands (1975)). *Sejam X_1, \dots, X_n, \dots , v.a.'s i.i.d. com f.d. $F \in \mathcal{D}(G_\gamma)$ e $\gamma \in \mathbb{R}$. Então, se $k = k_n \rightarrow \infty$ e $\frac{k_n}{n} \rightarrow 0$, quando $n \rightarrow \infty$, tem-se que*

$$\hat{\gamma}_{k,n}^P \xrightarrow{P} \gamma.$$

Teorema 4.4 (Consistência Forte do Estimador de Pickands, Dekkers and de Haan (1989)). *Sejam X_1, \dots, X_n, \dots , v.a.'s i.i.d. com f.d. $F \in \mathcal{D}(G_\gamma)$ e $\gamma \in \mathbb{R}$. Se $k = k_n \rightarrow \infty$, $\frac{k_n}{n} \rightarrow 0$ e $\frac{k_n}{\log(\log n)} \rightarrow \infty$, quando $n \rightarrow \infty$, então*

$$\hat{\gamma}_{k,n}^P \xrightarrow{q.c.} \gamma.$$

Tal como o estimador de Hill, o estimador de Pickands é assintoticamente Normal. Mais detalhadamente, tem-se

$$\sqrt{k} (\hat{\gamma}_{k,n}^P - \gamma) \xrightarrow[n \rightarrow \infty]{d} W \sim \mathcal{N}(0, \sigma_P^2),$$

com

$$\sigma_P^2 = \begin{cases} \frac{\gamma^2(2^{2\gamma+1}+1)}{(\log 2)^2(2^\gamma-1)^2} & \text{se } \gamma \neq 0, \\ \frac{3}{(\log 2)^4} & \text{se } \gamma = 0. \end{cases}$$

Como este estimador é relativamente simples e é invariante a transformações na localização e na escala, torna-se bastante apelativo. A juntar a essas vantagens, há ainda o facto de ser válido para qualquer valor de γ . Contudo, o estimador de Pickands tem uma variância bastante elevada e depende fortemente do valor de k , sendo, por isso, muito sensível à escolha das estatísticas de ordem intermédias, que serão usadas na estimação do parâmetro γ .

Este conjunto de vantagens e desvantagens motivaram diversas versões do estimador de Pickands. Por exemplo, Fraga Alves (1992), Themido Pereira (1993) e Yun (2002), que, com ligeiras diferenças, introduziram um parâmetro de controlo ou ajuste, ν , obtendo-se, assim, a versão generalizada deste estimador:

$$\hat{\gamma}_{k,n,\nu}^P = -\frac{1}{\log \nu} \cdot \log \left(\frac{X_{n-[\nu^2 k]+1:n} - X_{n-[\nu k]+1:n}}{X_{n-[\nu k]+1:n} - X_{n-k+1:n}} \right), \quad 0 < \nu < 1. \quad (4.9)$$

Note-se que para $\nu = \frac{1}{2}$, obtém-se o estimador de Pickands tradicional.

4.3.3 Estimador dos Momentos

Este estimador é semelhante ao de Hill, apresentado na secção 4.3.1, mas tem a grande vantagem de ser aplicável para qualquer $\gamma \in \mathbb{R}$ e não apenas para o caso de $\gamma > 0$.

Considerando que

$$M_{k,n}^{(r)} = \frac{1}{k} \cdot \sum_{i=1}^k (\log X_{n-i+1:n} - \log X_{n-k:n})^r, \quad r = 1, 2 \quad (4.10)$$

é o r -ésimo momento dos log-excessos e que

$$\hat{\gamma}_{k,n}^+ = M_{k,n}^{(1)} = \hat{\gamma}_{k,n}^H \quad \text{e} \quad \hat{\gamma}_{k,n}^- = 1 - \frac{1}{2} \cdot \left(1 - \frac{\left(M_{k,n}^{(1)} \right)^2}{M_{k,n}^{(2)}} \right)^{-1},$$

o estimador dos Momentos para $\gamma = \gamma_- + \gamma_+$, $\hat{\gamma}_{k,n}^M$, de autoria de Dekkers, Einmahl and de Haan (1989), é dado pela expressão

$$\hat{\gamma}_{k,n}^M = \hat{\gamma}_{k,n}^- + \hat{\gamma}_{k,n}^+, \quad (4.11)$$

em que $\gamma_- := \min(0, \gamma)$ e $\gamma_+ := \max(0, \gamma)$.

Quer isto dizer que o estimador dos Momentos é constituído por duas partes, $\hat{\gamma}_{k,n}^+$, para $\gamma > 0$, que é precisamente o estimador de Hill definido em (4.6), e por $\hat{\gamma}_{k,n}^-$, para $\gamma < 0$, que é chamado de estimador dos Momentos Negativo.

O estimador dos Momentos, tal como os estimadores apresentados anteriormente, goza das propriedades de consistência e, sobre determinada condição adicional, de consistência forte.

Teorema 4.5 (Consistência do Estimador dos Momentos). *Sejam X_1, \dots, X_n, \dots , v.a.'s i.i.d. com f.d. $F \in \mathcal{D}(G_\gamma)$ e $\gamma \in \mathbb{R}$. Então, se $k = k_n \rightarrow \infty$ e $\frac{k_n}{n} \rightarrow 0$, quando $n \rightarrow \infty$, tem-se que*

$$\hat{\gamma}_{k,n}^M \xrightarrow{P} \gamma.$$

Teorema 4.6 (Consistência Forte do Estimador dos Momentos). *Sejam X_1, \dots, X_n, \dots , v.a.'s i.i.d. com f.d. $F \in \mathcal{D}(G_\gamma)$ e $\gamma \in \mathbb{R}$. Se $k = k_n \rightarrow \infty$, $\frac{k_n}{n} \rightarrow 0$ e $\frac{k_n}{(\log n)^\eta} \rightarrow \infty$, quando $n \rightarrow \infty$ e para algum $\eta > 0$, então*

$$\hat{\gamma}_{k,n}^M \xrightarrow{q.c.} \gamma.$$

O estimador dos Momentos é invariante perante modificações de escala, mas o mesmo já não acontece para alterações na localização. Assim como os casos anteriores, sob a validade da condição de segunda ordem, é possível garantir a normalidade assintótica deste estimador,

$$\sqrt{k} \left(\hat{\gamma}_{k,n}^M - \gamma \right) \xrightarrow[n \rightarrow \infty]{d} W \sim \mathcal{N}(0, \sigma_M^2),$$

com

$$\sigma_M^2 = \begin{cases} \gamma^2 + 1 & \text{se } \gamma \geq 0, \\ \frac{(1 - \gamma)^2 \cdot (1 - 2\gamma) \cdot (1 - \gamma + 6\gamma^2)}{(1 - 3\gamma) \cdot (1 - 4\gamma)} & \text{se } \gamma < 0. \end{cases}$$

4.3.4 Estimador Generalizado de Hill

Este estimador foi introduzido por Beirlant, Vynckier and Tengels (1996), como uma tentativa de generalização do estimador de Hill para $\gamma \in \mathbb{R}$, e tem a forma

$$\hat{\gamma}_{k,n}^{GH} = \hat{\gamma}_{k,n}^H + \frac{1}{k} \cdot \sum_{i=1}^k (\log \hat{\gamma}_{i,n}^H - \log \hat{\gamma}_{k,n}^H). \quad (4.12)$$

Este estimador usa diretamente o estimador de Hill, como se pode observar pela expressão. É invariante a modificações na escala e é consistente para $\gamma \in \mathbb{R}$. Mediante condições suplementares, este estimador é assintoticamente Normal, ou seja,

$$\sqrt{k} (\hat{\gamma}_{k,n}^{GH} - \gamma) \xrightarrow[n \rightarrow \infty]{d} W \sim \mathcal{N}(0, \sigma_{GH}^2),$$

com

$$\sigma_{GH}^2 = \begin{cases} \gamma^2 + 1 & \text{se } \gamma \geq 0, \\ \frac{(1 - \gamma) \cdot (1 + \gamma + 2\gamma^2)}{(1 - 2\gamma)} & \text{se } \gamma < 0. \end{cases}$$

Um estudo mais aprofundado sobre este estimador foi realizado por Beirlant, Dierckx and Guillou (2005).

4.3.5 Estimador Mixed-Moment

Este estimador foi desenvolvido por Fraga Alves, Gomes, de Haan and Neves (2007) como resultado de uma combinação dos Teoremas 2.6.1 e 2.6.2 de de Haan (1970): uma f.d. F , com limite superior do suporte $x^F \in]0, +\infty]$, pertence ao domínio de atração de G_γ , isto é, $F \in \mathcal{D}(G_\gamma)$, com G_γ a distribuição GEV para algum $\gamma \in \mathbb{R}$, se e só se

$$\lim_{t \rightarrow x^F} \frac{(1 - F(t)) \cdot \int_t^{x^F} \int_y^{x^F} (1 - F(x)) \cdot x^{-2} dx dy}{t^2 \cdot \left(\int_t^{x^F} (1 - F(x)) \cdot x^{-2} dx \right)^2} = \varphi(\gamma) := \begin{cases} 1 + \gamma & \text{se } \gamma > 0, \\ \frac{1 - \gamma}{1 - 2\gamma} & \text{se } \gamma \leq 0. \end{cases} \quad (4.13)$$

Usando a expressão no primeiro membro de (4.13) e substituindo F e t pela f.d. empírica F_n e pelo *threshold* aleatório $X_{n-k:n}$, com $k < n$, respetivamente, obtém-se

$$\hat{\varphi}_n(k) := \frac{M_{k,n}^{(1)} - L_{k,n}^{(1)}}{\left(L_{k,n}^{(1)}\right)^2}, \quad \text{com } L_{k,n}^{(1)} := \frac{1}{k} \cdot \sum_{i=1}^k \left(1 - \frac{X_{n-k:n}}{X_{n-i+1:n}}\right), \quad (4.14)$$

em que $M_{k,n}^{(1)}$ foi definido em (4.10), para $r = 1$. Sendo assim, o estimador Mixed-Moment, para $\gamma \in \mathbb{R}$, é dado pela expressão

$$\hat{\gamma}_{k,n}^{MM} \equiv \hat{\gamma}_{k,n}^{MM}(X_{n-j+1:n}, 1 \leq j \leq k + 1) = \frac{\hat{\varphi}_n(k) - 1}{1 + 2 \cdot \min(\hat{\varphi}_n(k) - 1, 0)}. \quad (4.15)$$

Este estimador surgiu como uma alternativa atrativa aos estimadores mais usuais para o EVI, sendo válido para qualquer $\gamma \in \mathbb{R}$. A sua variância assintótica, para $\gamma < 0$, é igual à do estimador dos Momentos e, para $\gamma \geq 0$, é igual à do estimador de MV do EVI. Para além disso, fornece resultados muito semelhantes aos do estimador de MV para uma grande classe de modelos com $\gamma \geq 0$.

Tal como os estimadores já apresentados, o estimador Mixed-Moment é consistente para qualquer $\gamma \in \mathbb{R}$ e, sobre as mesma condições referidas nos casos anteriores, converge para uma distribuição assintótica Normal,

$$\sqrt{k} (\hat{\gamma}_{k,n}^{MM} - \gamma) \xrightarrow[n \rightarrow \infty]{d} W \sim \mathcal{N}(0, \sigma_{MM}^2),$$

com

$$\sigma_{MM}^2 = \begin{cases} (1 + \gamma)^2 & \text{se } \gamma \geq 0, \\ \frac{(1 - \gamma)^2 \cdot (1 - 2\gamma) \cdot (1 - \gamma + 6\gamma^2)}{(1 - 3\gamma) \cdot (1 - 4\gamma)} & \text{se } \gamma < 0. \end{cases}$$

Na figura 4.1 encontram-se representadas as variâncias de cada um dos estimadores do EVI apresentados. O estimador de Pickands, em geral, tem uma maior variância do que os outros casos. Para um EVI negativo, a variância dos estimadores dos Momentos e Mixed-Moment coincidem e, para um EVI positivo, o mesmo acontece para os estimadores dos Momentos e Generalizado de Hill. Tal situação era facilmente verificada pelas expressões da variância para cada um destes estimadores.

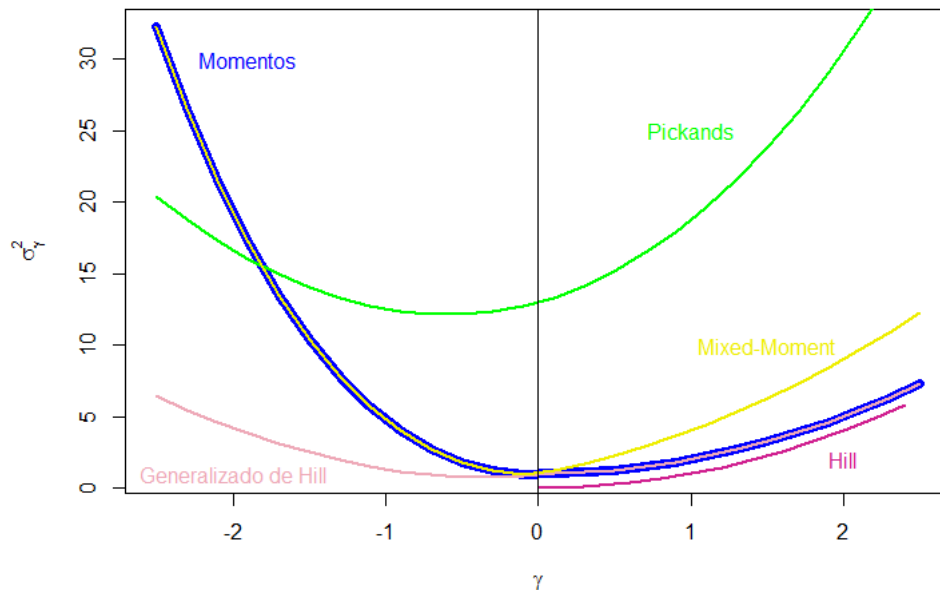


Figura 4.1: Comparação das variâncias assintóticas para os estimadores do EVI

É de realçar que todos os estimadores apresentados nesta secção são assintoticamente centrados apenas para valores de níveis $k = k_n$ convenientes, e que verifiquem a condição de regularidade de segunda ordem. A escolha adequada de k para a estimativa a considerar de γ deve ir ao encontro de um equilíbrio entre o viés e a variância, pois, se a escolha recair num valor pequeno de k , ou equivalentemente num nível elevado, estes estimadores apresentarão uma grande variabilidade e se, pelo contrário, o valor de k selecionado for elevado, que será o mesmo que ter um nível baixo, os estimadores terão um grande viés.

4.4 Invariância e Não-Invariância dos Estimadores do EVI

A maioria dos estimadores apresentados na secção anterior dependem dos log-excessos e, conseqüentemente, são não-invariantes a modificações na localização.

A invariância, perante alterações na localização e escala, é uma propriedade estatística bastante atrativa para um estimador do EVI. Como muitos autores sugerem, por exemplo Smith (1987), o uso inadequado de um destes estimadores não-invariante na localização, numa situação em que haja modificações deste parâmetro, poderá causar erros graves em toda a inferência estatística que seja realizada.

Em Araújo Santos, Fraga Alves and Gomes (2006) é discutido este tema e é apresentada uma transformação dos estimadores de Hill e dos Momentos em estimadores invariantes na localização e escala, mediante a simples transformação dos dados, dada por:

$$X_i^* := X_i - X_{[np]+1}, \quad 0 \leq p < 1, \quad 1 \leq i \leq n. \quad (4.16)$$

Este método tem o nome de metodologia PORT (Peaks Over Random Threshold). Um procedimento semelhante é usado em Fraga Alves, Gomes, de Haan and Neves (2007), onde é, também, proposto uma classe de estimadores do EVI alternativa à do estimador Mixed-Moment, apresentada na secção 4.3.5, invariante a modificações na localização, e dependentes do parâmetro de afinação p , $0 \leq p < 1$.

Estes estimadores têm a mesma expressão funcional dos seus correspondentes originais, denominem-se por T , mas as observações originais X_i são substituídas por X_i^* , com $1 \leq i \leq n$. Podem ser representados, por exemplo, por:

$$\hat{\gamma}_{k,n}^T(p) := \hat{\gamma}_{k,n}^T(X_{n-j+1:n} - X_{[np]+1:n}, \quad 1 \leq j \leq k+1), \quad p \in [0, 1[. \quad (4.17)$$

Esta versão de estimadores invariantes em (4.17) gozam das mesmas propriedades

que os estimadores originais $\hat{\gamma}_{k,n}^T (X_{n-j+1:n}, 1 \leq j \leq k+1)$, desde que se faça a escolha adequada quer para o valor de k , quer para o valor do parâmetro p .

4.5 Normalidade assintótica dos estimadores do EVI

Sob a validade da condição de regularidade de segunda ordem, apresentada na secção 4.2, é possível garantir a normalidade assintótica dos estimadores mencionados no ponto 4.3, como já foi mencionado nessa mesma secção. Segundo de Haan and Ferreira (2006), denotando por T qualquer uma das estatísticas H (Hill), P (Pickands), M (Momentos), GH (Hill Generalizado) ou MM (Mixed-Moment), e por $B(t)$ a função que representa o viés desse estimador e que converge para zero, quando $t \rightarrow \infty$, e que está relacionada com a função $A(t)$ de (4.1), é possível garantir para $\gamma \in D_T$, com D_T o conjunto para o qual γ está definido usando o estimador T ($D_H = \mathbb{R}^+$ e $D_P = D_M = D_{GH} = D_{MM} = \mathbb{R}$), a existência de constantes reais (b_T, σ_T) , com $\sigma_T > 0$, tais que

$$\hat{\gamma}_{k,n}^T \stackrel{d}{=} \gamma + \sigma_T \cdot \frac{P_k^T}{\sqrt{k}} + b_T \cdot B\left(\frac{n}{k}\right) + o\left(B\left(\frac{n}{k}\right)\right), \quad (4.18)$$

em que P_k^T é uma v.a. com distribuição assintótica $\mathcal{N}(0,1)$. Consequentemente, para valores de k tais que $\sqrt{k} \cdot B\left(\frac{n}{k}\right) \xrightarrow[n \rightarrow \infty]{} \lambda$, finito, tem-se

$$\sqrt{k} (\hat{\gamma}_{k,n}^T - \gamma) \xrightarrow[n \rightarrow \infty]{d} W \sim \mathcal{N}(\lambda \cdot b_T, \sigma_T^2). \quad (4.19)$$

Os valores b_T e σ_T^2 são, geralmente, chamados de viés e variância assintóticos para $\hat{\gamma}_{k,n}^T$, respetivamente, e necessitam de ser estimados, o que não é tarefa fácil. Em de Haan and Ferreira (2006) ou Beirlant *et al.* (2004) são apresentadas expressões para estas duas quantidades.

Tendo estimativas para b_T e σ_T^2 , é possível construir intervalos de confiança para γ , baseados no estimador $\hat{\gamma}_{k,n}^T$. Mas, como já referido, a estimação destes parâmetros é complexa e contém, ainda, algumas questões em aberto. de Haan and Ferreira (2006) recomendam que se deva escolher $\lambda = 0$ em (4.19), isto é, $\sqrt{k} \cdot B\left(\frac{n}{k}\right) \xrightarrow[n \rightarrow \infty]{} 0$ e, como tal, obtém-se

$$\sqrt{k} (\hat{\gamma}_{k,n}^T - \gamma) \xrightarrow[n \rightarrow \infty]{d} W \sim \mathcal{N}(0, \sigma_T^2). \quad (4.20)$$

Assim, o intervalo de $(1 - \alpha) \cdot 100\%$ de confiança para γ é dado por

$$\hat{\gamma}_{k,n}^T \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_T^2}{k}}, \quad (4.21)$$

com $z_{1-\frac{\alpha}{2}}$ o quantil de probabilidade $1 - \frac{\alpha}{2}$, de uma distribuição Normal standard. É de salientar que na secção 4.3, é apresentada a expressão de σ_T^2 para cada um dos estimadores do EVI.

4.6 Estimação semi-paramétrica de outros parâmetros de interesse em EVT

A estimação do EVI é determinante, como se sabe, para conhecer o tipo de cauda da distribuição subjacente à amostra em estudo. Mas existem outros parâmetros de grande interesse e com uma grande utilidade em Valores Extremos, como se constatou nas secções 3.2.3 e 3.3.4. Aí foram apresentados estimadores paramétricos para os quantis extremos, os níveis e períodos de retorno, a probabilidade de excedência de um nível elevado e o limite superior do suporte (no caso de $\gamma < 0$). Neste ponto desta dissertação, os mesmos parâmetros serão abordados, mas numa perspectiva semi-paramétrica.

Os quantis extremos são, talvez, dos parâmetros mais importantes em eventos extremos, funções do EVI, bem como dos parâmetros de localização e escala. O trabalho de Laurens de Haan é, também aqui, proeminente. Recorde-se a função quantil de cauda, apresentada na definição 2.4,

$$U(t) := F^{\leftarrow} \left(1 - \frac{1}{t} \right),$$

e do facto de que os quantis extremos podem ser expressos em termos desta função por

$$q_{1-p} = U \left(\frac{1}{p} \right) := F^{\leftarrow}(1 - p),$$

sendo p próximo de zero, ou teoricamente, $p \equiv p_n \rightarrow 0$, com $p_n \ll \frac{k}{n}$, quando $n \rightarrow \infty$. No contexto semi-paramétrico, os estimadores dos quantis extremos podem ser facilmente obtidos a partir da condição de primeira ordem, teorema 2.12, em que

$$F \in \mathcal{D}(G_\gamma), \gamma \in \mathbb{R} \Leftrightarrow \lim_{t \rightarrow \infty} \frac{U(tx) - U(t)}{a(t)} = D_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{se } \gamma \neq 0, \\ \log x & \text{se } \gamma = 0. \end{cases}$$

para alguma função positiva $a(\cdot)$ e com $x > 0$.

Assim,

$$\begin{aligned} \frac{U(tx) - U(t)}{a(t)} &\simeq D_\gamma(x) \Leftrightarrow \\ \Leftrightarrow U(tx) &\simeq U(t) + a(t) \cdot D_\gamma(x) \Leftrightarrow \\ \Leftrightarrow U(x) &\simeq U(t) + a(t) \cdot D_\gamma \left(\frac{x}{t} \right). \end{aligned}$$

Considerando, $t := \frac{n}{k}$, que tende para ∞ , quando $n \rightarrow \infty$, uma vez que $k \equiv k_n$ é sucessão

intermédia, e $x := \frac{1}{p}$, com p próximo de zero ($p \equiv p_n \xrightarrow{n \rightarrow \infty} 0$), a expressão anterior fica, então,

$$U\left(\frac{1}{p}\right) \simeq U\left(\frac{n}{k}\right) + a\left(\frac{n}{k}\right) \cdot D_\gamma\left(\frac{k}{np}\right). \quad (4.22)$$

A quantidade $U\left(\frac{n}{k}\right)$ representa o quantil extremal de ordem $1 - \frac{k}{n}$ da f.d. F . Logo um estimador para este valor poderá ser obtido recorrendo à f.d. empírica, já apresentada anteriormente. Sendo assim,

$$\widehat{U}\left(\frac{n}{k}\right) = \widehat{F}^{\leftarrow}\left(1 - \frac{k}{n}\right) = \widehat{F}^{\leftarrow}\left(\frac{n-k}{n}\right) = F_n^{\leftarrow}\left(\frac{n-k}{n}\right) = X_{n-k:n}, \quad (4.23)$$

ou seja, o estimador é exatamente o nível aleatório dado pela e.o. intermédia $X_{n-k:n}$.

Relativamente à quantidade $a\left(\frac{n}{k}\right)$, em de Haan and Ferreira (2006) é demonstrado que um estimador consistente é dado por

$$\hat{a}\left(\frac{n}{k}\right) = X_{n-k:n} \cdot M_{k,n}^{(1)} \cdot (1 - \hat{\gamma}_{k,n}^-), \quad (4.24)$$

em que, $M_{k,n}^{(1)}$ é dado por (4.10), para $r = 1$, e $\hat{\gamma}_{k,n}^-$ é a parte negativa do estimador dos Momentos, apresentado em (4.11), e cuja expressão, recorde-se, é dada por

$$\hat{\gamma}_{k,n}^- = 1 - \frac{1}{2} \cdot \left(1 - \frac{\left(M_{k,n}^{(1)}\right)^2}{M_{k,n}^{(2)}}\right)^{-1}.$$

Sendo assim, o estimador do quantil extremal pode ser escrito como

$$\hat{q}_{1-p} = \widehat{U}\left(\frac{1}{p}\right) = X_{n-k:n} + \hat{a}\left(\frac{n}{k}\right) \cdot D_{\hat{\gamma}}\left(\frac{k}{np}\right).$$

Como,

$$D_\gamma(x) = \begin{cases} \frac{x^\gamma - 1}{\gamma} & \text{se } \gamma \neq 0, \\ \log x & \text{se } \gamma = 0, \end{cases}$$

então, para $\gamma \neq 0$, o quantil extremal de probabilidade $1 - p$ é estimado por

$$\hat{q}_{1-p} = \widehat{U}\left(\frac{1}{p}\right) = X_{n-k:n} + \hat{a}\left(\frac{n}{k}\right) \cdot \frac{\left(\frac{k}{np}\right)^{\hat{\gamma}} - 1}{\hat{\gamma}}, \quad (4.25)$$

com $\hat{\gamma} = \hat{\gamma}_{k,n}^T$, em que $T = H, P, M, GH, MM$.

Para $\gamma = 0$, o estimador do quantil extremal de probabilidade $1 - p$ é dado por

$$\hat{q}_{1-p} = \hat{U}\left(\frac{1}{p}\right) = X_{n-k:n} + \hat{a}\left(\frac{n}{k}\right) \cdot \log\left(\frac{k}{np}\right). \quad (4.26)$$

No entanto, quando o valor de γ é positivo, e se está, por isso, num domínio de caudas pesadas, o estimador do quantil extremal poderá ter uma expressão mais simples, dada por

$$\hat{q}_{1-p} = \hat{U}\left(\frac{1}{p}\right) = X_{n-k:n} \cdot \left(\frac{k}{np}\right)^{\hat{\gamma}}. \quad (4.27)$$

Este estimador tem a designação de estimador de Weissman (Weissman, 1978).

O limite superior do suporte, x^F , pode ser expresso como função de U , com $p = 0$, como já apresentado anteriormente,

$$x^F := F^{\leftarrow}(1) = U(\infty).$$

Para o caso de $\gamma < 0$, isto é, quando se está no max-domínio Weibull caracterizado por f.d.'s com caudas curtas, o limite superior do suporte é finito. Um seu estimador poderá ser encontrado a partir do estimador do Quantil Extremal, em (4.25), considerando-se $p = 0$. Obtém-se assim,

$$\hat{x}^F = \hat{U}(\infty) = X_{n-k:n} - \frac{\hat{a}\left(\frac{n}{k}\right)}{\hat{\gamma}}, \quad (4.28)$$

com $\hat{a}(\cdot)$ definida em (4.24) e $\hat{\gamma} = \hat{\gamma}_{k,n}^T$, em que $T = P, M, GH, MM$, ou ainda, $\hat{\gamma} = \hat{\gamma}_{k,n}^-$.

Na prática, a implementação deste estimador deverá ter em consideração ao facto de que $\hat{x}^F \geq X_{n:n}$, logo, a expressão do estimador do limite superior do suporte, apresentada em (4.28) poderá ser redefinida, obtendo-se

$$\hat{x}^F = \max\left(X_{n:n}, X_{n-k:n} - \frac{\hat{a}\left(\frac{n}{k}\right)}{\hat{\gamma}}\right). \quad (4.29)$$

Um outro parâmetro de especial interesse é a probabilidade de excedência de um nível elevado x , isto é, $p = P(X > x) = 1 - F(x) = \bar{F}(x)$.

Para se obter um estimador desta quantidade dever-se-á ter em atenção ao que foi apresentado no início da secção 4.3. Esses resultados conduzem à expressão para \bar{F} , apresentada em (4.5), e que aqui é novamente reproduzida

$$\bar{F}(x) \simeq \frac{k}{n} \cdot \left\{ 1 - H_{\gamma}\left(\frac{x - X_{n-k:n}}{a\left(\frac{n}{k}\right)}\right) \right\}, \quad x > X_{n-k:n}.$$

Logo o estimador da probabilidade de excedência $p = \bar{F}(x)$ é

$$\hat{p} = \frac{k}{n} \cdot \left\{ 1 - H_{\hat{\gamma}} \left(\frac{x - X_{n-k:n}}{\hat{a} \left(\frac{n}{k} \right)} \right) \right\}, \quad x > X_{n-k:n}. \quad (4.30)$$

Tendo em atenção a expressão da f.d. GP, H_{γ} , então, o estimador para a probabilidade de excedência do nível x , quando $\gamma \neq 0$, é dado por

$$\hat{p} = \frac{k}{n} \cdot \left\{ \max \left(0, 1 + \hat{\gamma} \cdot \frac{x - X_{n-k:n}}{\hat{a} \left(\frac{n}{k} \right)} \right) \right\}^{-\frac{1}{\hat{\gamma}}}, \quad (4.31)$$

com $\hat{\gamma} = \hat{\gamma}_{k,n}^T$, em que $T = H, P, M, GH, MM$.

Quando $\gamma = 0$, esse estimador assume a forma

$$\hat{p} = \frac{k}{n} \cdot \exp \left(- \frac{x - X_{n-k:n}}{\hat{a} \left(\frac{n}{k} \right)} \right). \quad (4.32)$$

Em ambos os casos, $\hat{a}(\cdot)$ é dado por (4.24).

No caso do EVI ser positivo, $\gamma > 0$, o estimador da probabilidade de excedência poderá ser simplificado. Neste caso, tem-se que

$$\hat{a} \left(\frac{n}{k} \right) = X_{n-k:n} \cdot M_{k,n}^{(1)},$$

já que $\hat{\gamma}_{k,n}^- \xrightarrow[n \rightarrow \infty]{p} 0$, e como $\frac{\hat{\gamma}}{M_{k,n}^{(1)}} = \frac{\hat{\gamma}}{\hat{\gamma}_{k,n}^H} \xrightarrow[n \rightarrow \infty]{p} 1$, então,

$$1 + \hat{\gamma} \cdot \frac{x - X_{n-k:n}}{\hat{a} \left(\frac{n}{k} \right)} = 1 + \hat{\gamma} \cdot \frac{x - X_{n-k:n}}{X_{n-k:n} \cdot M_{k,n}^{(1)}} \simeq 1 + \frac{x - X_{n-k:n}}{X_{n-k:n}} = \frac{x}{X_{n-k:n}}.$$

Logo, para $\gamma > 0$, o estimador da probabilidade de excedência do nível x é dado por

$$\hat{p} = \frac{k}{n} \cdot \left(\frac{x}{X_{n-k:n}} \right)^{-\frac{1}{\hat{\gamma}}}, \quad (4.33)$$

com $\hat{\gamma} = \hat{\gamma}_{k,n}^T$, em que $T = H, P, M, GH, MM$.

Relativamente ao nível de retorno T -anos, $U(T)$, como já referido, esta quantidade é um quantil extremal de probabilidade $1 - \frac{1}{T}$, logo um seu estimador pode ser dado recorrendo aos estimadores apresentados em (4.25), (4.26) ou (4.27), quando $\gamma \neq 0$, $\gamma = 0$ ou $\gamma > 0$, respetivamente. Tem-se então,

$$\hat{U}(T) = \hat{q}_{1-\frac{1}{T}}. \quad (4.34)$$

Para o período de retorno de um nível u , T_u , poder-se-á recorrer aos estimadores apresentados para a probabilidade de excedência, em (4.31), (4.32) ou (4.33), se $\gamma \neq 0$, $\gamma = 0$ ou $\gamma > 0$, respetivamente. Recorde-se que o período de retorno de um determinado nível u de uma distribuição F , é dado por

$$T_u = \frac{1}{1 - F(u)} = \frac{1}{\overline{F}(u)},$$

logo um seu estimador é dado por

$$\widehat{T}_u = \frac{1}{\widehat{p}}, \quad (4.35)$$

com \widehat{p} o estimador da probabilidade de excedência do nível u , apresentado em (4.31), (4.32) ou (4.33).

4.6.1 Testes para os Domínios de Atração

Como já referido, ao seguir uma abordagem semi-paramétrica, a única suposição é que a função F pertença ao domínio de atracção $\mathcal{D}(G_\gamma)$, para algum γ real. Assim, e como evidenciado diversas vezes ao longo desta dissertação, o EVI é o parâmetro de destaque, sendo ele quem determina a forma da cauda da f.d., F , subjacente aos dados em estudo. Recorde-se que, o domínio de atracção Fréchet, em que $\gamma > 0$, contém distribuições com cauda pesada, do tipo polinomial, e com limite superior do suporte não finito; no lado oposto, o domínio Weibull, com $\gamma < 0$, as distribuições são caracterizadas por terem caudas curtas e limite superior do suporte finito. Na fronteira entre estes dois casos, encontra-se o Domínio Gumbel, com $\gamma = 0$, que abrange distribuições com caudas do tipo exponencial e limites superiores do suporte finitos e infinitos. Este último caso, por ser mais simples, é considerado, muitas vezes, como a referência, assumindo-se que a distribuição dos dados em estudo pertence a este domínio de atracção, para depois se aplicar metodologias, onde o domínio Gumbel é testado contra os outros dois domínios max-estáveis.

O problema da escolha estatística neste contexto semi-paramétrico pode ser explicitado através do teste com hipótese nula $F \in \mathcal{D}(G_0)$, isto é, da f.d. F pertencer ao domínio de atracção Gumbel, contra uma hipótese alternativa bilateral,

$$H_0 : F \in \mathcal{D}(G_0) \quad vs. \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0} \quad (4.36)$$

ou contra uma de duas hipóteses alternativas unilaterais, ou de F pertencer ao domínio de atracção Weibull,

$$H_0 : F \in \mathcal{D}(G_0) \quad vs. \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma < 0} \quad (4.37)$$

ou de F pertencer ao domínio de atração Fréchet,

$$H_0 : F \in \mathcal{D}(G_0) \quad \text{vs.} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma>0}. \quad (4.38)$$

Este tipo de testes podem ser encontrados em diversas publicações, de onde se destaca Galambos (1982), Castillo *et al.* (1989), Fraga Alves and Gomes (1996), Wang *et al.* (1996), Marohn (1998a) e Marohn (1998b). Mais recentemente, em Neves *et al.* (2006), Neves and Fraga Alves (2007) e Neves and Fraga Alves (2008b) são apresentadas três estatísticas para os testes (4.36), (4.37) e (4.38), que são baseadas nas observações da amostra acima de um nível aleatório. Mais especificamente, vão ser considerados os k excessos acima da $(n - k)$ -ésima e.o. intermédia superior, $X_{n-k:n}$, onde $k \equiv k_n \rightarrow \infty$ e $\frac{k}{n} \rightarrow 0$, quando $n \rightarrow \infty$.

Considerando que $Z_i = X_{n-i+1:n} - X_{n-k:n}$, com $i = 1, \dots, k$, são os excessos acima de $X_{n-k:n}$, as estatísticas referidas são, então:

- i. a estatística de teste da razão, $T_n(k)$, entre o máximo e o valor médio da amostra dos excessos acima de um nível aleatório, $X_{n-k:n}$:

$$T_n(k) = \frac{Z_1}{k^{-1} \cdot \sum_{i=1}^k Z_i}; \quad (4.39)$$

- ii. a estatística de teste de Greenwood (1946), $R_n(k)$, que permite identificar caudas pesadas:

$$R_n(k) = \frac{k^{-1} \cdot \sum_{i=1}^k Z_i^2}{\left(k^{-1} \cdot \sum_{i=1}^k Z_i\right)^2}; \quad (4.40)$$

- iii. a estatística de teste de Hasofer-Wang (1992), $W_n(k)$, que se baseia nos k extremos superiores de uma amostra de dimensão n :

$$W_n(k) = k^{-1} \frac{\left(k^{-1} \cdot \sum_{i=1}^k Z_i\right)^2}{k^{-1} \cdot \sum_{i=1}^k Z_i^2 - \left(k^{-1} \cdot \sum_{i=1}^k Z_i\right)^2} = \frac{1}{k} \cdot \frac{1}{R_n(k) - 1}, \quad (4.41)$$

Estas estatísticas são invariantes a modificações na localização e na escala e, por esse motivo, dizem-se ancilárias.

Com base em conceitos da teoria de variação regular, Neves and Fraga Alves (2007) reformularam as propriedades assintóticas da estatística de teste de Hasofer-Wang, em

(4.41), e de Greenwood, em (4.40), no caso em que $k = k_n$ se comporta como uma sequência intermédia em vez de permanecer fixo (que era o caso original), enquanto a dimensão da amostra aumenta. A estatística de teste da razão, em (4.39), introduzida por Neves *et al.* (2006), surgiu devido às diferenças que existem na contribuição relativa do máximo para a soma dos k excessos, acima do nível aleatório, em diferentes caudas pesadas.

As versões normalizadas de (4.39), (4.40) e (4.41) são, respectivamente:

$$T_n^*(k) = T_n(k) - \log k; \quad (4.42)$$

$$R_n^*(k) = \sqrt{\frac{k}{4}} \cdot (R_n(k) - 2); \quad (4.43)$$

$$W_n^*(k) = \sqrt{\frac{k}{4}} \cdot (k \cdot W_n(k) - 1). \quad (4.44)$$

Sob a hipótese nula de que a f.d. subjacente à amostra, pertence ao domínio de atração de uma distribuição Gumbel, $F \in \mathcal{D}(G_0)$, e sob a condição de regularidade de segunda ordem, apresentada na definição 4.1, para a cauda superior de F e para a taxa de convergência de k_n para ∞ , as estatísticas normalizadas em (4.42), (4.43) e (4.44) convergem em distribuição de acordo com:

$$T_n^* \xrightarrow[n \rightarrow \infty]{d} T \frown \Lambda; \quad (4.45)$$

$$R_n^* \xrightarrow[n \rightarrow \infty]{d} Z_1 \frown \mathcal{N}(0, 1); \quad (4.46)$$

$$W_n^* \xrightarrow[n \rightarrow \infty]{d} Z_2 \frown \mathcal{N}(0, 1). \quad (4.47)$$

Para cada uma das estatísticas de teste apresentadas, a região crítica para o teste bilateral (4.36), com nível de significância α , é dada por

$$T_n^*(k) < \mathcal{G}_{\frac{\alpha}{2}} \quad \text{ou} \quad T_n^*(k) > \mathcal{G}_{1-\frac{\alpha}{2}};$$

$$|R_n^*(k)| > z_{1-\frac{\alpha}{2}}; \quad (4.48)$$

$$|W_n^*(k)| > z_{1-\frac{\alpha}{2}},$$

em que $\mathcal{G}_\varepsilon = -\log(-\log(\varepsilon))$ e $z_\varepsilon = \Phi^{-1}(\varepsilon)$ são os quantis de probabilidade ε das distribuições Gumbel e Normal standards, respetivamente.

Seguindo a abordagem unilateral, as regiões de rejeição, com um nível de significância α , são:

- i. para o teste definido em (4.37), ou seja, para o domínio de atração Weibull ($H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma < 0}$):

$$T_n^*(k) < \mathcal{G}_\alpha; \quad R_n^*(k) < z_\alpha; \quad W_n^*(k) > z_{1-\alpha}; \quad (4.49)$$

- ii. para o teste definido em (4.38), ou seja, para o domínio de atração Fréchet ($H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma > 0}$):

$$T_n^*(k) > \mathcal{G}_{1-\alpha}; \quad R_n^*(k) > z_{1-\alpha}; \quad W_n^*(k) < z_\alpha, \quad (4.50)$$

em que \mathcal{G}_ε e z_ε são os quantis de probabilidade ε das distribuições Gumbel e Normal standards, respetivamente, como já indicado.

Capítulo 5

Estudo de Caso: Aplicação a Níveis Pluviométricos

5.1 Introdução

Como já tinha sido indicado na Introdução desta dissertação, com vista à aplicação prática das metodologias abordadas nos capítulos anteriores, a amostra em estudo compreende níveis pluviométricos diários (em mm) da cidade de Barcelos, em Portugal, referentes ao período de 01/01/1932 a 01/05/2008. Estes dados, bem como outros referentes a questões ambientais, são disponibilizados pelo Sistema Nacional de Informação de Recursos Hídricos, e encontram-se disponíveis em <http://snirh.pt>. Esta amostra tem uma dimensão de 27568 níveis pluviométricos, havendo valores omissos, essencialmente nos últimos anos (por incrível que pareça, já que nos encontramos numa fase de grande desenvolvimento da tecnologia).



Figura 5.1: *Rio Cávado em Barcelos: no lado esquerdo, um dia habitual e no lado direito, um dia com inundações, in <http://www.panoramio.com>*

A estação de monitorização de Barcelos, segundo o histórico de dados, é a estação com os maiores registos de níveis pluviométricos em Portugal. Por este motivo, esta região é uma boa candidata à aplicação de uma análise em valores extremos.

Portugal continental, localizado aproximadamente entre as latitudes de 37°N e 42°N e as longitudes de 9.5°W e 6.5°W, no extremo Sudoeste da Europa, situa-se na zona de transição entre o anticiclone subtropical (anticiclone dos Açores) e a zona das depressões subpolares, sendo o clima fortemente influenciado pela proximidade ao oceano Atlântico. Outro dos fatores determinantes do clima é a orografia da região, com áreas significativas, nas zonas Norte e Centro, a ultrapassarem os 1000 m de altitude.

A variação dos fatores climáticos (latitude, proximidade ao oceano e orografia), embora pequena, é suficiente para induzir variações significativas na temperatura e, principalmente, na precipitação observadas em Portugal continental. Com efeito, a região Noroeste (Minho), e onde se localiza a cidade de Barcelos, é uma das zonas da Europa que regista valores mais elevados de precipitação, atingindo a média de precipitação anual acumulada, nalguns locais, valores superiores a 3000 mm, como é visível na figura 5.2, o que, por vezes, leva à ocorrência de fenómenos extremos associados aos excessos de precipitação (cheias).

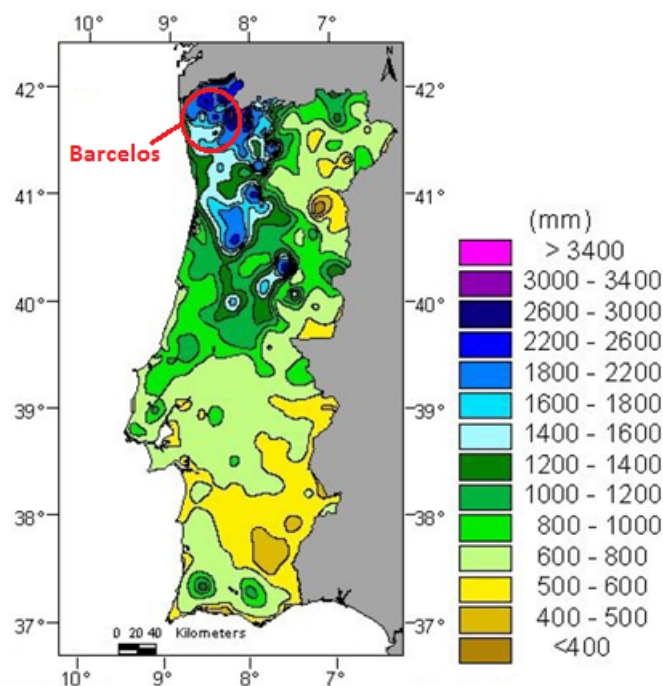


Figura 5.2: Precipitação média anual em Portugal (1961-1990), in Miranda et al. (2006)

O clima de Barcelos, pelo facto de se situar entre serras e o oceano Atlântico, é tipicamente atlântico temperado, ou seja, com quatro estações do ano bem definidas. As primaveras são frescas, com grandes aberturas e ventos suaves. As brisas matinais ocorrem com grande frequência, principalmente, nas maiores altitudes. É de salientar o mês de maio, que é bastante propício a trovoadas, devido ao aquecimento do ar húmido com a chegada do verão. Os verões são quentes, com muito sol e com ventos suaves de Este. Nos dias mais frescos, podem ocorrer espontaneamente chuvas de curta duração. Estas chuvas são bastante importantes para a vegetação da região, pois reabastece os lençóis de água. Isto possibilita a existência de vegetação durante o ano inteiro, daí o termo habitualmente usado de "Verde Minho" para caracterizar esta zona de Portugal. Quanto aos outonos, estes são amenos mas pluviosos e com ventos moderados. Enquanto a temperatura desce, aumenta a pluviosidade. Existe, também, uma maior frequência de nevoeiros. Por fim, os invernos são caracterizados por serem bastante pluviosos e frios, e geralmente com ventos moderados. Em anos muito frios, pode ocorrer a queda de neve, no entanto, devido ao aquecimento global, a queda de neve é um acontecimento muito raro. Na figura 5.3 encontra-se representada a distribuição média de precipitação por estações do ano, em Portugal continental e, com destaque da região de Barcelos. É visível que, nas quatro estações do ano, esta região apresenta níveis de pluviosidade mais elevados do que qualquer outra região do país.

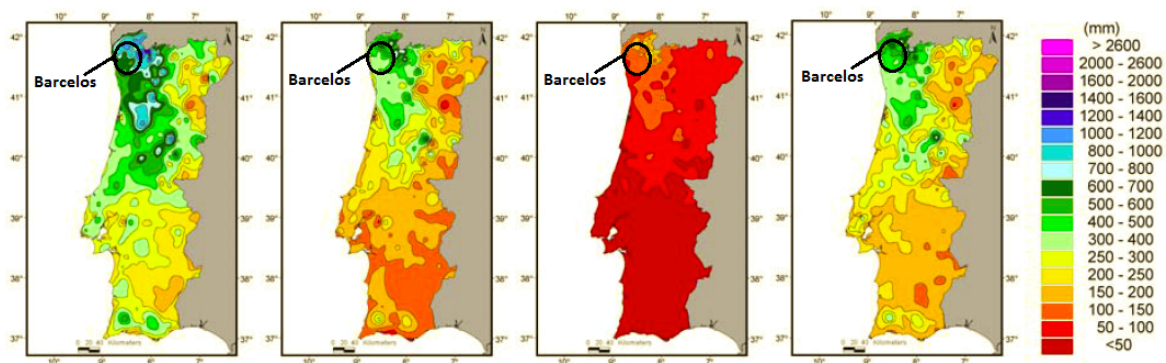


Figura 5.3: *Precipitação média sazonal acumulada em Portugal (1961-1990): (a) inverno, (b) primavera, (c) verão e (d) outono, in Miranda et al. (2006)*

Como foi referido, a base de dados original é constituída por 27568 registos diários de níveis pluviométricos, em Barcelos. Contudo, a análise em valores extremos não será aplicada a todas estas observações, pois, dados ambientais diários não são, geralmente, independentes. De facto, por exemplo, um dia de chuva tem uma maior probabilidade

de ser seguido por outro chuvoso. Assim, foram considerados subconjuntos da amostra original. Criou-se uma sub-amostra de dimensão 77, constituída pela observação máxima em cada um dos anos mencionados. Inicialmente, os anos que tinham dados em falta não foram considerados, pois nada nos garante que o máximo anual não fosse atingido num desses dias em falta. Contudo, os resultados obtidos não se revelaram muito diferentes dos do estudo usando esses mesmos anos e, como tal, resolveu-se abandonar essa amostra mais pequena.

Com a amostra original e, devido à sua dimensão, foi possível criar outras duas sub-amostras. Numa são considerados os 5 maiores níveis pluviométricos por ano e na outra, as 10 maiores observações anuais. Estas amostras permitem a aplicação do método das maiores observações.

Foi, também criada uma outra amostra, mas desta vez, selecionando os máximos mensais. Assim, cada mês foi considerado como um bloco, mas, tendo em vista uma análise sazonal, ou seja, por estações do ano, os meses considerados foram os do calendário (repare-se que, desta forma, cada bloco poderá ter uma dimensão a variar entre 28 a 31 elementos). Obteve-se, assim, uma amostra de 917 observações. Em Fraga Alves (2013) foi efetuada esta mesma análise, mas considerando que um mês era um período de 30 dias. Os resultados obtidos em cada uma destas duas abordagens mensais não são muito diferentes.

Um outro objetivo será o de aplicar a análise em valores extremos separada por estação do ano, mais especificamente no outono e no inverno. Seguindo a proposta de Miranda *et al.* (2006), a primavera compreende os meses de março, abril e maio, o verão os meses de junho, julho e agosto, o outono os meses de setembro, outubro e novembro e o inverno os meses de dezembro, janeiro e fevereiro. Assim, e escolhendo o máximo diário de cada um destes períodos, em cada um dos anos, obtiveram-se quatro amostras, primavera, verão, outono e inverno, cada uma com 77, 76, 75 e 76 observações, respetivamente, contudo a análise será apenas aplicada às amostras referentes ao outono e inverno.

Para cada uma destas amostras, assume-se que os elementos que as constituem são identicamente distribuídos, pois considera-se que a precipitação é gerada pelo mesmo processo físico.

5.2 Análise Paramétrica dos Dados

No capítulo 3 foram apresentadas três metodologias paramétricas, que podem ser utilizadas numa análise em valores extremos: o método dos máximos anuais (MMA) ou

método Gumbel, o método POT e o método das maiores observações (MMO). Assim, começar-se-á pela aplicação do MMA às amostras de níveis pluviométricos referenciadas anteriormente, mais propriamente, aos máximos anuais, aos máximos mensais e aos máximos anuais no outono e inverno. De seguida, será aplicado o MMO às 5 e 10 maiores observações anuais. Por fim, será usada a metodologia POT com as amostras dos máximos anuais e mensais. Todas estas análises serão efetuadas recorrendo ao software R, versão 3.0.1.

5.2.1 Método dos Máximos Anuais

5.2.1.1 Níveis Pluviométricos Máximos Anuais

I) Análise Preliminar dos Dados

Antes de começar a aplicar os procedimentos do MMA a esta amostra, começa-se por fazer uma análise preliminar dos dados. Esta análise será essencialmente gráfica e vai permitir ter uma ideia do comportamento da cauda direita da distribuição associada aos dados referentes aos níveis máximos anuais de pluviosidade, em Barcelos.

Esta amostra, recorde-se, é constituída pelos níveis pluviométricos máximos, por ano, no período de 01/01/1932 a 01/05/2008, tendo por isso uma dimensão igual a $m = 77$ observações. Esta amostra encontra-se representada na figura 5.4.

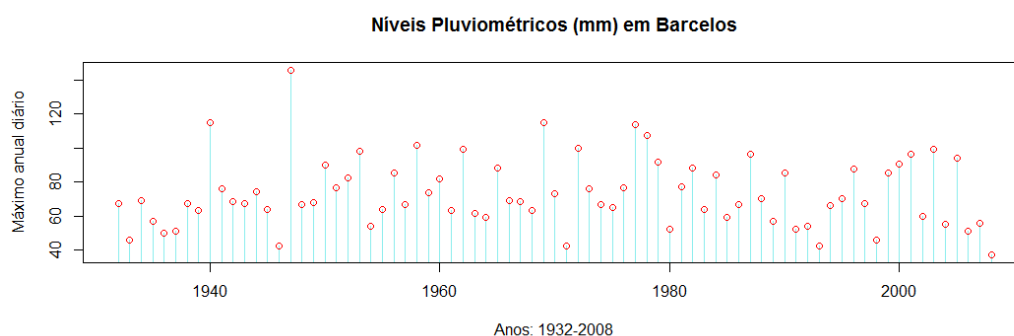


Figura 5.4: *Níveis pluviométricos máximos anuais, em Barcelos, (1932-2008)*

Com a representação de um gráfico de autocorrelação parcial (figura 5.5), observa-se que as observações são fracamente correlacionadas entre si, podendo-se, assim, admitir a independência dos dados.

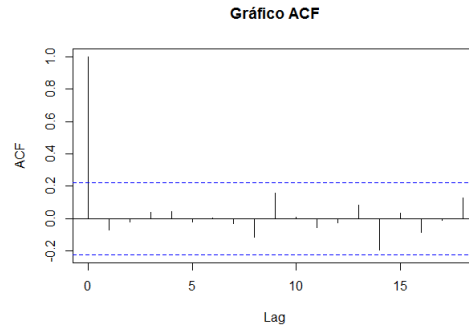


Figura 5.5: *Gráfico de Autocorrelação Parcial*

As suas características amostrais como a média, a mediana, os quartis e os extremos são:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
37.00	59.60	68.40	73.24	85.50	146.00

Observando estes valores, os níveis de pluviosidade variam entre 37 mm a 146 mm, sendo perceptível uma assimetria positiva e que pode ser confirmada pelo boxplot e pelo histograma representados na figura 5.6. Esta assimetria acontece quer na parte central, visível

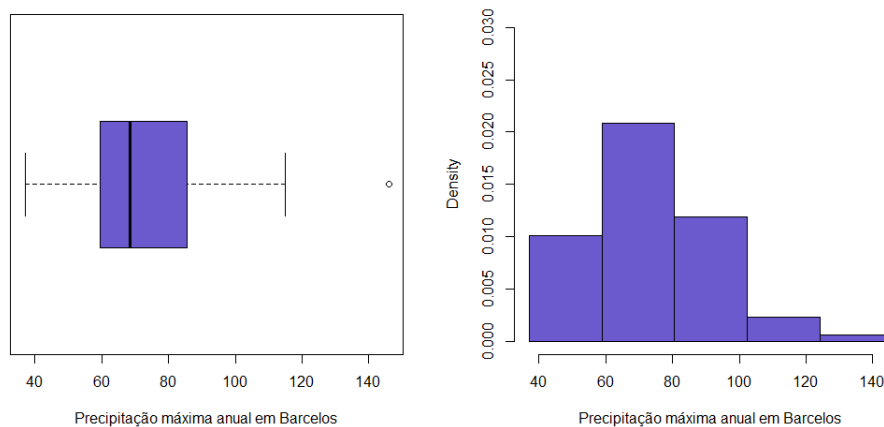


Figura 5.6: *Boxplot e histograma dos níveis pluviométricos máximos anuais, em Barcelos, (1932-2008)*

pela forma da caixa do boxplot, quer no extremo superior, onde também se constata um forte candidato a outlier. Observando o histograma, a cauda direita aparenta ser leve ou do tipo exponencial, pois não é demasiado extensa e as barras não decrescem lentamente.

Será que isto significa que a distribuição de probabilidade subjacente aos níveis máximos anuais de pluviosidade, em Barcelos, é Exponencial? Ou a cauda é, de facto, mais leve e a distribuição será Gumbel, ou até mesmo Normal, ou quem sabe Weibull (de máximos)? Para se obter uma resposta rápida a estas questões, poder-se-á recorrer aos qq-plot. Estes são um bom instrumento para uma confirmação visual e rápida do ajustamento de um modelo probabilístico a um conjunto de dados, permitindo, também, a estimação preliminar de parâmetros desse modelo. Para além disso, poderão também dar indicação do tipo de cauda do modelo de onde os dados são provenientes, quando comparada com a cauda do modelo associado ao qq-plot em questão.

Em EVT, o qq-plot exponencial desempenha um papel bastante importante. Segundo Beirlant *et al.* (2004), se o conjunto de dados estiver sobre uma reta diagonal, significa que este poderá ter uma distribuição aproximadamente exponencial. Também a cauda associada ao modelo de onde os dados são provenientes será exponencial. Por sua vez, se o conjunto de pontos apresentar um padrão convexo, a cauda do modelo subjacente a esses dados será mais pesada que a da exponencial. Já se o padrão for côncavo, a cauda será mais leve. Desta forma, é possível ter suspeitas sobre qual dos domínios max-estáveis está associado ao conjunto de dados.

Se se considerar o modelo $EXP(\lambda)$, com f.d. dada por

$$F(x) = 1 - \exp(-\lambda x), \quad x, \lambda > 0,$$

e com função quantil $Q_\lambda(p)$, com $p \in (0, 1)$, existe a seguinte relação linear entre $Q_\lambda(p)$ e $Q_1(p)$:

$$Q_\lambda(p) = \frac{1}{\lambda} \cdot (-\log(1 - p)) = \frac{1}{\lambda} \cdot Q_1(p), \quad p \in (0, 1).$$

Substituindo os quantis teóricos pelos empíricos, $\hat{Q}_\lambda(p)$, e se $x_{i:m}$ for o i -ésimo valor da amostra ordenada, tem-se que $\hat{Q}_\lambda(p_i) = x_{i:m}$, em que a escolha dos p_i , os *plotting positions*, é dada por Weibull,

$$p_i = \frac{i}{m+1}, \quad i = 1, \dots, m.$$

Sendo assim, a nuvem de pontos num qq-plot exponencial pode ser dada por

$$(-\log(1 - p_i), x_{i:m}). \tag{5.1}$$

Um outro meio gráfico, que fornece indicações acerca do tipo de cauda da distribuição subjacente aos dados em estudo, quando comparada com a da exponencial, é o ME-plot. Recorde-se que este é a representação gráfica da função de excesso médio, e_m , definida em (2.9). Em geral, a distribuição subjacente aos dados não é conhecida e, como tal, terá

que se recorrer à contrapartida empírica da função de excesso médio, \hat{e}_m apresentada em (3.38).

Segundo Beirlant *et al.* (2004), o padrão do ME-plot obtido através de

$$\{(x_{m-k:m}, \hat{e}_m(x_{m-k:m})), 1 \leq k \leq m - 1\}, \quad (5.2)$$

poderá levar a identificação do tipo de cauda da distribuição. Se o ME-plot apresentar um padrão crescente, está-se perante um caso de cauda mais pesada que a da exponencial. Se, pelo contrário for decrescente, a cauda será, então, mais leve que a da exponencial. No caso de se obter um padrão constante, a cauda será do tipo exponencial. Tal como foi referido aquando da explicação do uso do ME-plot para a identificação de um possível *threshold*, na metodologia POT, a parte final desta representação gráfica, em geral, é inconstante, uma vez que estão a ser calculadas médias de um número reduzido de excessos elevados, podendo, por isso, ser ignorada.

Na figura 5.7 estão representados o qq-plot exponencial e o ME-plot para a amostra em estudo. Estes vêm corroborar as suspeitas já evidenciadas atrás. O qq-plot apresenta uma curva com um padrão côncavo o que sugere que a distribuição subjacente a estes dados tem uma cauda mais leve do que a exponencial. O mesmo é indicado no ME-plot que apresenta um padrão decrescente.

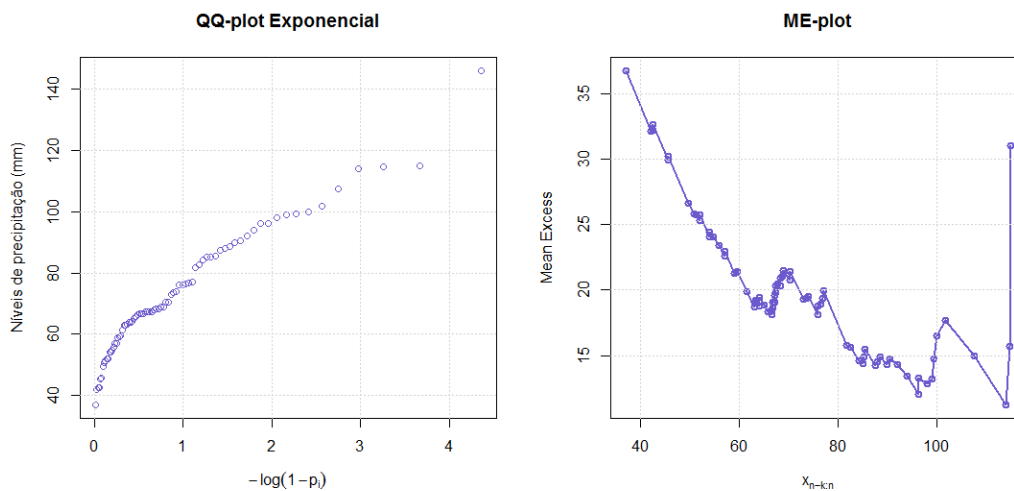


Figura 5.7: *QQ-Plot Exponencial e ME-plot dos níveis pluviométricos máximos anuais, em Barcelos, (1932-2008)*

Tem-se, portanto, a indicação de que a distribuição subjacente à amostra dos níveis máximos anuais de pluviosidade, em Barcelos, terá uma cauda mais leve que a da exponencial. Sendo assim, são candidatas algumas distribuições do domínio max-estável

Gumbel, como a distribuição Normal e a própria Gumbel, e todas as distribuições do domínio max-estável Weibull. Devido à evidente assimetria à direita dos dados, o ajustamento do modelo normal poderá ser posto de parte. Faz, portanto, mais sentido fazer um ajustamento preliminar ao modelo Gumbel, um dos modelos extremais.

Tendo em conta a f.d. do modelo Gumbel, apresentada no teorema 2.3, com parâmetros de localização, $\lambda \in \mathbb{R}$, e de escala, $\delta > 0$, ou seja, $\Lambda(x|\lambda, \delta)$, e se $Q_{\lambda, \delta}(p)$, com $p \in (0, 1)$, for a respetiva função quantil, facilmente se verifica que,

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

Há, portanto, uma relação linear entre $Q_{\lambda, \delta}(p)$ e $Q_{0,1}(p)$ em que a localização e a escala do modelo são dados pela ordenada na origem e pelo declive, respetivamente. Seguindo um raciocínio análogo ao apresentado para a obtenção de um qq-plot exponencial, o qq-plot Gumbel será obtido através de

$$(-\log(-\log(p_i)), x_{i:m}), \quad i = 1, \dots, m. \quad (5.3)$$

Ajustando uma reta de mínimos quadrados a essa nuvem de pontos, obtida no qq-plot, poder-se-á usar a ordenada na origem e o declive como estimativas preliminares dos parâmetros λ e δ , respetivamente.

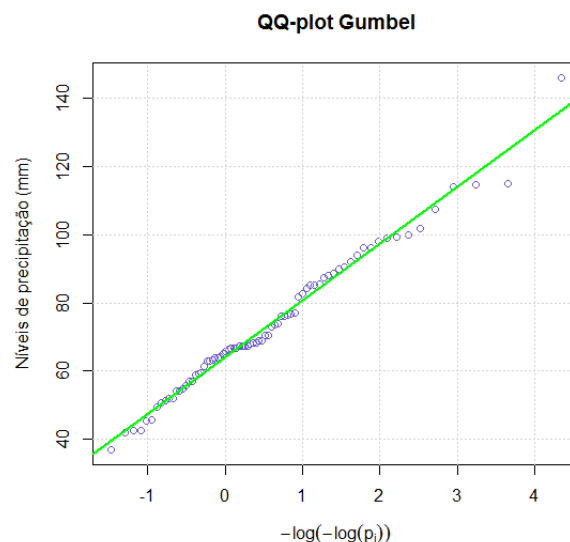


Figura 5.8: *QQ-Plot Gumbel dos níveis pluviométricos máximos anuais, em Barcelos, (1932-2008)*

O qq-plot Gumbel para a amostra em estudo poderá ser visto na figura 5.8. Neste, a linearidade é evidente o que poderá indicar que o ajustamento à Gumbel parece ser

aceitável. De facto, obtém-se uma correlação muito forte a rondar o valor de 0.993210. As estimativas preliminares para os parâmetros, obtidas no R, são:

```
Call:
lm(formula = x1_ord ~ pi)
Coefficients:
(Intercept)          pi
   63.95297    16.69477
```

ou seja,

$$(\hat{\lambda}, \hat{\delta}) = (63.95297, 16.69477). \quad (5.4)$$

Mas todas as distribuições pertencentes ao domínio max-estável Weibull são também candidatas ao ajustamento destes dados, pois têm caudas direitas mais leves que a da exponencial. Neste caso, será razoável utilizar o modelo GEV. Se, de facto, a distribuição que melhor se ajusta a este conjunto de dados for do domínio Weibull, é de esperar que o parâmetro de forma, γ , seja negativo.

Para se obter o qq-plot GEV, os procedimentos a adotar serão semelhantes aos anteriores. Tendo em conta a f.d. $G_\gamma(x|\lambda, \delta)$ e a função quantil da GEV, $Q_{\gamma,\lambda,\delta}(p)$, apresentadas em (2.6) e (3.11), respetivamente, tem-se a seguinte relação

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

Substituindo os quantis teóricos pelos empíricos $\hat{Q}_{\gamma,\lambda,\delta}(p)$ e se $x_{i:m}$ for o i -ésimo valor da amostra ordenada, ter-se-á, mais uma vez, $\hat{Q}_{\gamma,\lambda,\delta}(p_i) = x_{i:m}$, com $p_i = \frac{i}{m+1}$, com $i = 1, \dots, m$. O qq-plot GEV será dado por

$$\left(\frac{(-\log(p_i))^{-\gamma} - 1}{\gamma}, x_{i:m} \right), \quad i = 1, \dots, m. \quad (5.5)$$

Tem-se, assim, definida a relação linear que permite a construção do qq-plot GEV. Mais uma vez, a localização do modelo será dada pela ordenada na origem e a escala pelo declive. Falta, portanto, encontrar uma quantidade que forneça uma estimativa preliminar do parâmetro de forma, γ . Seguindo a sugestão de Beirlant *et al.* (2004), a escolha preliminar para o EVI poderá ser encontrada através da correlação entre os quantis $\hat{Q}_{\gamma,\lambda,\delta}(p) = x_{i:m}$ e $Q_{\gamma,0,1}(p)$. Assim, dever-se-á escolher o γ que maximize essa correlação. Usando o software R, obtém-se os valores da estimativa do parâmetro e da correlação máxima, respetivamente,

```

$maximum
[1] -0.009004
$objective
[1] 0.993235

```

e podem ser visualizados no gráfico reproduzido na figura 5.9. Fixando $\hat{\gamma} = -0.009004$,

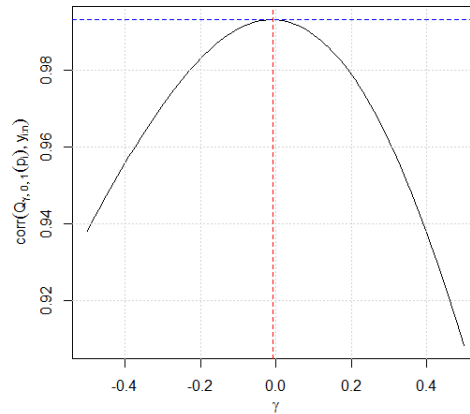


Figura 5.9: Gráfico de correlação entre $\hat{Q}_{\gamma, \lambda, \delta}(p)$ e $Q_{\gamma, 0, 1}(p)$

obtém-se o qq-plot GEV para a amostra de dados, representado na figura 5.10.

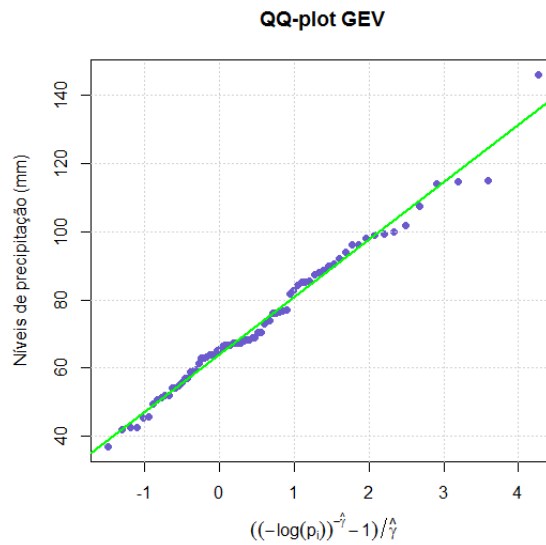


Figura 5.10: *QQ-Plot GEV dos níveis pluviométricos máximos anuais, em Barcelos, (1932-2008)*

O ajustamento da reta aos pontos é bastante bom, o que evidencia um bom ajustamento deste modelo à amostra. À semelhança do qq-plot Gumbel, os pontos estão quase como que "perfeitamente" sobre a reta dos mínimos quadrados e a correlação é praticamente igual sendo, neste caso, de 0.993235 (contra 0.993210). O ajustamento dos mínimos quadrados sugere como estimativas dos parâmetros de localização e escala, respetivamente, os valores:

Call:

```
lm(formula = sort(x) ~ Qgev)
```

Coefficients:

(Intercept)	Qgev
63.99625	16.85117

Portanto, para a amostra em estudo, as estimativas preliminares dos parâmetros do modelo GEV são

$$(\hat{\gamma}, \hat{\lambda}, \hat{\delta}) = (-0.009004, 63.99625, 16.85117). \quad (5.6)$$

Sobrepondo ao histograma dos dados as densidades Gumbel e GEV estimadas (figura 5.11), é fácil constatar que as diferenças entres estas são mínimas, pois as duas curvas parecem quase coincidir ponto por ponto. De facto, as estimativas para os parâmetros de localização e escala são, praticamente, iguais nos dois casos. Repare-se no facto da estimativa preliminar do EVI ser negativa, o que remete a distribuição subjacente a estes dados, para o domínio max-estável Weibull. Contudo, o valor desta estimativa é praticamente igual a zero e, como tal, o domínio max-estável Gumbel é outra forte possibilidade.

A escolha do modelo que melhor se ajusta a estas observações é difícil, mas pela semelhança das estimativas dos parâmetros de localização e escala e pela proximidade de zero do EVI, o modelo Gumbel parece ser o candidato com mais possibilidades de chegar ao pódio. De qualquer forma, ambos os casos vão ao encontro do que tinha sido referido anteriormente, de que a cauda da distribuição subjacente a estes níveis pluviométricos é mais leve que a da exponencial. Não devemos esquecer que esta é apenas uma abordagem preliminar, e que no tópico seguinte obter-se-ão respostas mais consistentes. Contudo, com esta primeira abordagem fica-se com uma ideia sobre o comportamento dos níveis de pluviosidade em estudo, e as estimativas obtidas poderão ser utilizadas como inicializações em processos numéricos que serão posteriormente utilizados.

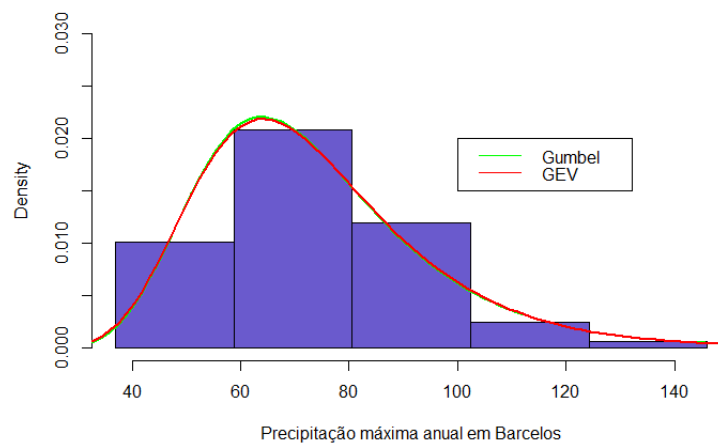


Figura 5.11: *Densidades dos modelos Gumbel e GEV estimados ajustados aos níveis pluviométricos máximos anuais*

II) Estimação dos Parâmetros

Na análise preliminar efetuada, constatou-se que a distribuição Gumbel era uma forte candidata para modelar esta amostra de níveis pluviométricos máximos anuais, em Barcelos. Sendo assim, começar-se-á por fazer o ajustamento deste modelo aos dados, usando os dois métodos de estimação apresentados nas secções 3.2.1 e 3.2.2, o método MV e o método MPP. Recorde-se que o modelo Gumbel é constituído pelos parâmetros de localização, λ , e de escala, δ , e, usando os dois métodos referidos, as suas estimativas são

Tabela 5.1: *Estimativas MV e MPP dos parâmetros de localização e escala do modelo Gumbel, para os níveis pluviométricos máximos anuais*

Método de Estimação	$\hat{\lambda}$ (localização)	$\hat{\delta}$ (escala)
MV	63.97622	16.37472
MPP	63.96075	16.07648

Como é visível, as estimativas obtidas são muito semelhantes para os dois métodos usados, e o mesmo acontece quando comparadas com as estimativas preliminares.

Usando o método profile log-likelihood, apresentado na secção 3.2.4, podem ser construídos intervalos de confiança para os dois parâmetros do modelo. Com um nível assintótico de 95%, os intervalos de confiança para a localização e escala, do modelo Gumbel, são, respetivamente,

```
[1] "profiling loc"
[1] "profiling scale"
      lower  upper
loc    60.13723 67.94500
scale  13.88045 19.61359
```

Estes intervalos também poderão ser visualizados na figura 5.12

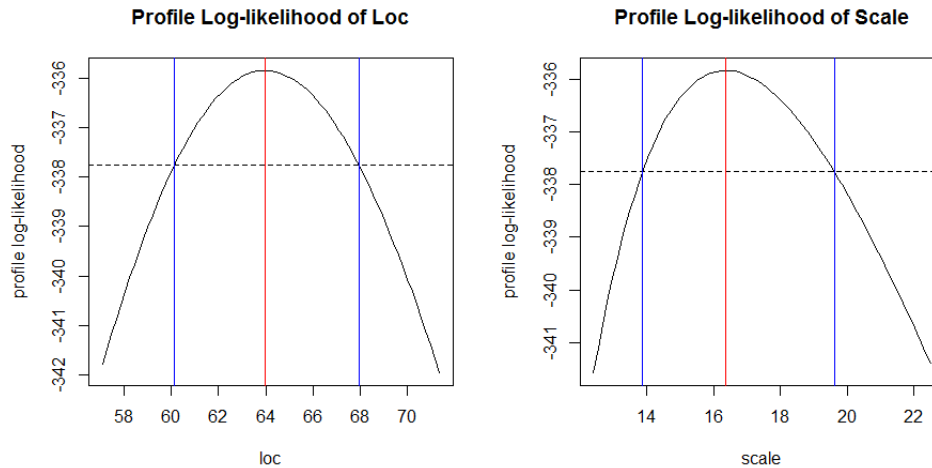


Figura 5.12: Intervalos de 95% de confiança para os parâmetros do modelo Gumbel, baseados no *profile log-likelihood*, para os níveis pluviométricos máximos anuais

A qualidade do ajustamento Gumbel, obtido pelo método MV, pode ser verificada, facilmente, com um diagnóstico gráfico obtido através do software R, e que se encontra reproduzido na figura 5.13. Este ajustamento aparenta ser bastante satisfatório para a amostra em estudo. No qq-plot, apenas as três observações finais apresentam uma pequena divergência em relação à reta, mas essa situação é muito comum. Na zona central do pp-plot, os pontos também se afastam ligeiramente da reta, o que não é grave pois, em acontecimentos raros, a zona de destaque está localizada na parte final. É aí que estão os denominados eventos raros.

Mas não se pretende apenas obter estimativas dos parâmetros diretamente relacionados com o modelo Gumbel. Em modelação de acontecimentos raros, existem outras quantidades de grande interesse e importância. Na secção 3.2.3 foram referidos os quantis extremos, os níveis e períodos de retorno, a probabilidade de excedência de um nível elevado e, caso o EVI seja negativo, o limite superior do suporte, e foram apresentados os seus estimadores, baseados nos estimadores de MV ou dos MPP dos parâmetros da distribuição subjacente aos dados em estudo.

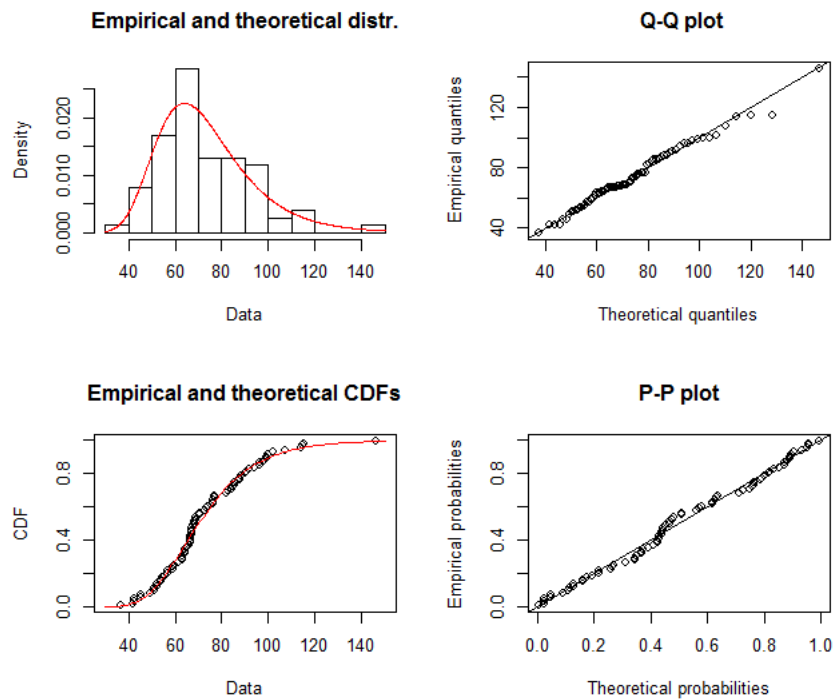


Figura 5.13: *Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos anuais*

No capítulo da Introdução desta dissertação, foi mencionado que em Nascimento (2009) e Nascimento *et al.* (2012) foi feita uma abordagem Bayesiana semi-paramétrica à estimação de valores extremos e, em particular, são apresentadas estimativas de quantis elevados, de probabilidades 0.95, 0.99, 0.999, 0.9999 e 0.99999, para os níveis pluviométricos máximos mensais na cidade de Barcelos. Um dos objetivos será fazer uma comparação entre esses quantis com os obtidos através da metodologia adotada nesta dissertação, para a amostra de máximos mensais. Assim, parece justificável que se estimem os mesmos quantis, para qualquer uma das amostras que serão aqui trabalhadas.

Usando as estimativas MV e MPP para os parâmetros da distribuição Gumbel, apresentadas na tabela 5.1, e a expressão (3.11), para o caso $\gamma = 0$, obtém-se as estimativas pontuais para os quantis extremos $q_{0.95}$, $q_{0.99}$, $q_{0.999}$, $q_{0.9999}$ e $q_{0.99999}$. Também se podem calcular estimativas intervalares para estes mesmos quantis, recorrendo à função de profile log-likelihood e à reparametrização da f.d. Gumbel, de forma a que q_{1-p} , com $p \in (0, 1)$ seja um dos parâmetros. Por exemplo, basta considerar a transformação

$$q_{1-p} = \lambda - \delta \cdot \log(-\log(1-p)) \Leftrightarrow \lambda = q_{1-p} + \delta \cdot \log(-\log(1-p)).$$

Na tabela 5.2 estão representadas as estimativas pontuais e intervalares (com 95% de

confiança) dos quantis extremos mencionados e na figura 5.14 estão reproduzidos esses IC's.

Tabela 5.2: *Estimativas pontuais MV e MPP e IC a 95% de quantis extremos do modelo Gumbel, para os níveis pluviométricos máximos anuais*

Método de Estimação	$\hat{q}_{0.95}$	$\hat{q}_{0.99}$	$\hat{q}_{0.999}$	$\hat{q}_{0.9999}$	$\hat{q}_{0.99999}$
MV	112.6123	139.3024	177.0806	214.7921	252.4971
MPP	111.7110	137.9150	175.0051	212.0298	249.0480
IC(95%)	(103.4,124.1)	(126.14,156)	(158.36,200.95)	(190.4,246.2)	(222.2,291.3)

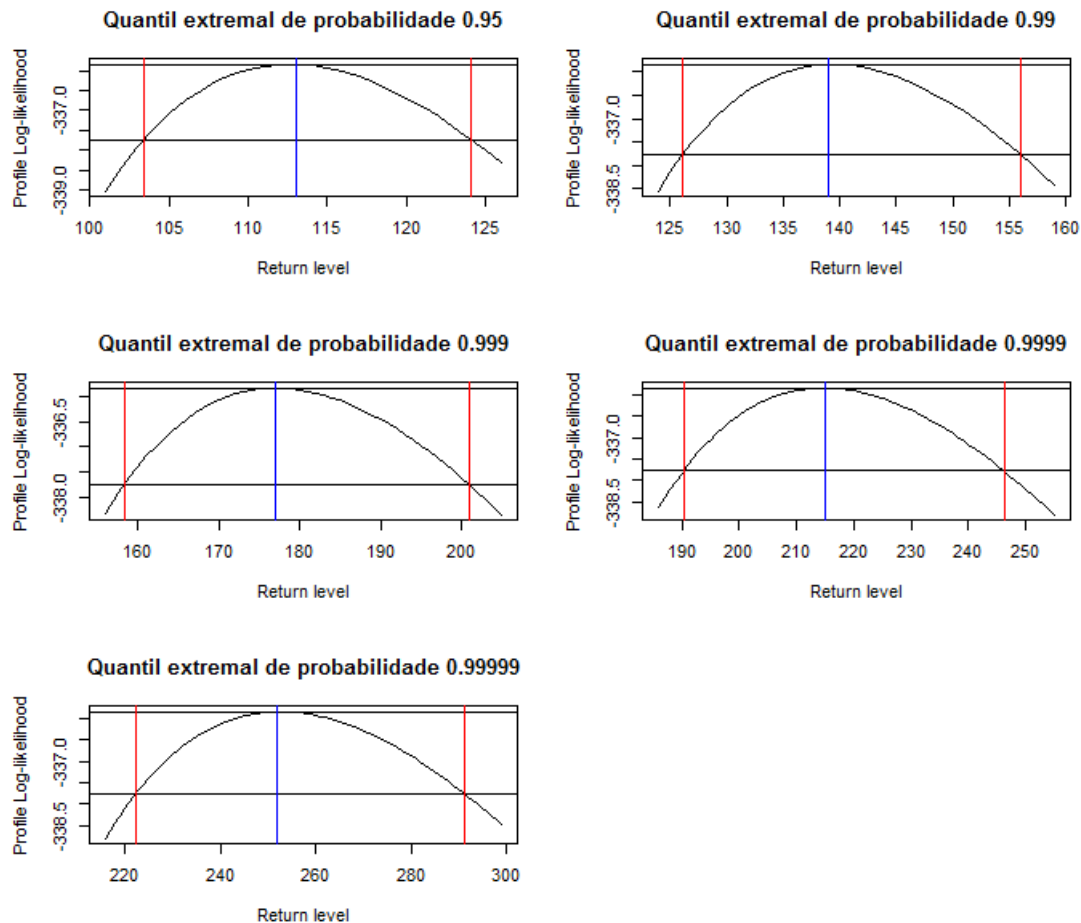


Figura 5.14: *Intervalos de 95% de confiança para quantis extremos do modelo Gumbel, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais*

É de constatar que as estimativas dos quantis extremos não são muito diferentes, quando usados os métodos de MV ou dos MPP. Contudo, os obtidos por este último método são ligeiramente mais baixos. Um pormenor interessante e que demonstra a utilidade

da EVT, tem a ver precisamente com estes quantis. Tomando em consideração, apenas, a amostra disponível de máximos anuais, os quantis empíricos para as probabilidades mencionadas encontram-se na tabela 5.3. Ora os quantis de probabilidade 0.99, 0.999, 0.9999 e 0.99999 coincidem com o máximo registado, pois a amostra é constituída por 77 observações. Ou seja, a dimensão desta amostra não permite uma maior especificidade nos valores destes quantis elevados. Será preciso, por isso, ir além da amostra, ou seja, fazer extrapolação.

Tabela 5.3: *Quantis empíricos para os níveis pluviométricos máximos anuais*

$q_{0.95}$	$q_{0.99}$	$q_{0.999}$	$q_{0.9999}$	$q_{0.99999}$
114	146	146	146	146

O nível de retorno T -anos ($U(T)$) é outra medida que pode ser estimada. Em particular, vão ser calculadas estimativas pontuais e intervalares (baseadas na função de profile log-likelihood, com 95% confiança) para os níveis de retorno de 100, 400 e 1000 anos. Ou seja, vão ser indicados os níveis pluviométricos (em mm) que são excedidos, em média, uma vez em cada 100, 400 e 1000 anos, na cidade de Barcelos. Recorde-se que um nível de retorno T -anos não é mais do que um quantil extremal de probabilidade $1 - \frac{1}{T}$. Assim, as estimativas para os níveis de retorno de 100 e 1000 anos já são conhecidas, pois

$$U(100) = q_{1 - \frac{1}{100}} = q_{0.99} \quad \text{e} \quad U(1000) = q_{1 - \frac{1}{1000}} = q_{0.999}.$$

Quer isto dizer que se está à espera que, num século, o nível 139.3024 mm de precipitação seja, em média, ultrapassado uma vez. Se se alargar o período para 1000 anos, esse nível aumenta para 177.0806 mm. Na tabela 5.4 estão representadas as estimativas dos níveis de retorno pluviométricos indicados, bem como os seus intervalos de 95% confiança, baseados na função de profile log-likelihood.

Tabela 5.4: *Estimativas pontuais MV e MPP e IC a 95% de níveis de retorno do modelo Gumbel, para os níveis pluviométricos máximos anuais*

Método de Estimação	$\hat{U}(100)$	$\hat{U}(400)$	$\hat{U}(1000)$
MV	139.3024	162.0643	177.0806
MPP	137.9150	160.2623	175.0051
IC(95%)	(126.14,156)	(145.6,183.2)	(158.36,200.95)

Na figura 5.15 poderão ser vistos os limites dos IC's para os níveis de retorno de 100, 400 e 1000 anos.

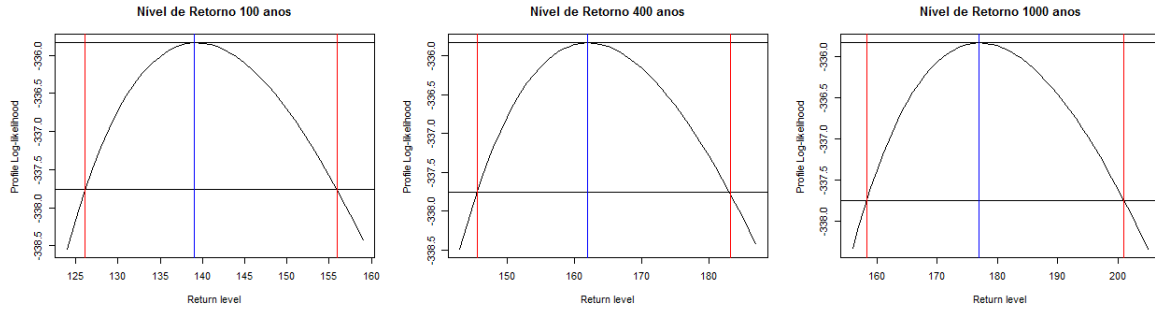


Figura 5.15: Intervalos de 95% de confiança para níveis de retorno do modelo Gumbel, baseados no *profile log-likelihood*, para os níveis pluviométricos máximos anuais

Podem ser, também, calculadas estimativas para o período de retorno e probabilidade de excedência de um determinado nível pluviométrico. Tendo em conta que a maior observação da amostra dos máximos anuais corresponde ao nível de 146 mm de pluviosidade, será este o nível utilizado. Assim, o período de retorno do nível 146 mm, T_{146} , será o número médio de anos de espera até que seja ultrapassada essa quantidade de precipitação. Já a probabilidade disso acontecer, $P(Y > 146)$, será dada pela probabilidade de excedência do nível 146 mm. Na tabela 5.5 encontram-se estimativas destas quantidades.

Tabela 5.5: Estimativas MV e MPP do período de retorno e probabilidade de excedência do nível 146 mm, no modelo Gumbel, para os níveis pluviométricos máximos anuais

Método de Estimação	$\hat{T}(146)$	$\hat{P}(Y > 146)$
MV	150.2811	0.006654195
MPP	165.0252	0.006059682

De acordo com os resultados obtidos, em média, são precisos cerca de 150 anos ou 165 anos, consoante se opte pela estimativa MV ou MPP, para que o nível máximo anual de pluviosidade ultrapasse o valor mais elevado, observado no período de 01/01/1932 a 01/05/2008, na cidade de Barcelos. A probabilidade de se exceder esse nível é de aproximadamente 0.6-0.7%. Repare-se que as estimativas obtidas pelo método MPP indicam que esse acontecimento é, ligeiramente, mais raro do que as obtidas pelo método MV.

Como se está a usar o modelo Gumbel, o que corresponde à distribuição GEV com $\gamma = 0$, a expressão apresentada em (3.16), para o limite superior do suporte, não poderá ser utilizada, pois só é válida para $\gamma < 0$.

Quando foi efetuada a análise preliminar desta amostra, concluiu-se que a distribuição subjacente a estes níveis pluviométricos teria uma cauda direita mais leve que a da exponencial. A distribuição Gumbel seria uma boa candidata, assim como qualquer uma das distribuições pertencentes ao domínio max-estável Weibull. Falta, portanto, considerar este último caso, utilizando para isso o modelo GEV. Se a análise preliminar estiver correta, espera-se encontrar uma estimativa negativa para o parâmetro de forma (γ).

Utilizando os dois métodos de estimação, apresentados nas secções 3.2.1 e 3.2.2, obtiveram-se estimativas para os parâmetros de forma, localização e escala do modelo GEV, que se encontram registadas na tabela 5.6.

Tabela 5.6: *Estimativas MV e MPP dos parâmetros de forma, localização e escala do modelo GEV, para os níveis pluviométricos máximos anuais*

Método de Estimação	$\hat{\gamma}$ (forma)	$\hat{\lambda}$ (localização)	$\hat{\delta}$ (escala)
MV	-0.04967	64.4188	16.6027
MPP	-0.04396	64.2903	16.7161

Tal como foi referido no caso anterior do ajustamento Gumbel, poderão ser obtidos intervalos de confiança para os mesmos parâmetros, recorrendo à função de profile log-likelihood. Assim, as estimativas intervalares com uma confiança de 95% são, respetivamente,

```
[1] "profiling loc"
[1] "profiling scale"
[1] "profiling shape"
      lower      upper
loc    60.3241981 68.6743849
scale  13.9850105 20.0335666
shape  -0.1793777  0.1297275
```

A representação gráfica destes intervalos de confiança encontram-se na figura 5.16.

As estimativas dos parâmetros obtidas pelos métodos MV e MPP, mais uma vez, são muito semelhantes. A estimativa para o EVI é negativa, como já se suspeitava, logo tem-se a indicação de que, a distribuição subjacente aos níveis pluviométricos máximos anuais poderá pertencer ao domínio max-estável Weibull, ao qual fazem parte as distribuições com caudas leves e com limite superior do suporte finito. Mas não pode ser ignorada a proximidade a zero, por parte desta estimativa, bem como o facto do intervalo de confiança

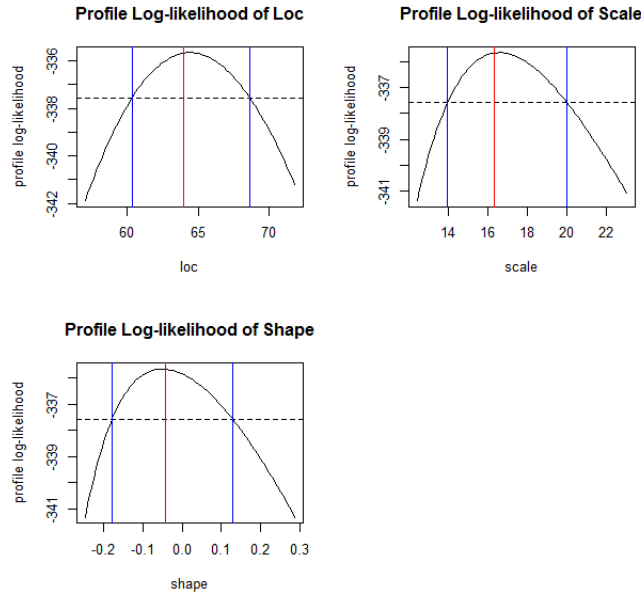


Figura 5.16: Intervalos de 95% de confiança para os parâmetros do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais

obtido para este parâmetro conter o valor zero. Assim, a hipótese da distribuição destes dados pertencer ao max-domínio Gumbel não pode ser descartada. No tópico seguinte aplicar-se-ão testes adequados para se chegar a uma conclusão.

Na figura 5.17 podem ser vistos os gráficos que permitem fazer um diagnóstico rápido à qualidade do ajustamento GEV, obtido pelo método de MV.

O ajustamento GEV parece bastante bom. As semelhanças entre este gráficos e os obtidos no ajustamento Gumbel são imensas, não havendo grandes diferenças que mereçam ser destacadas.

Usando as estimativas obtidas para os parâmetros de forma, localização e escala, poder-se-á proceder à estimação dos outros parâmetros de grande interesse em EVT. Mediante a substituição dessas estimativas na expressão (3.11), para o caso $\gamma \neq 0$, obtêm-se as estimativas dos quantis extremos $q_{0.95}$, $q_{0.99}$, $q_{0.999}$, $q_{0.9999}$ e $q_{0.99999}$, as quais se encontram representadas na tabela 5.7.

Tabela 5.7: Estimativas pontuais MV e MPP e IC a 95% de quantis extremos do modelo GEV, para os níveis pluviométricos máximos anuais

Método de Estimação	$\hat{q}_{0.95}$	$\hat{q}_{0.99}$	$\hat{q}_{0.999}$	$\hat{q}_{0.9999}$	$\hat{q}_{0.99999}$
MV	110.2670	132.6966	161.4951	187.1327	209.9953
MPP	110.8355	133.9102	163.8698	190.8945	215.3128
IC(95%)	(101.41,126.7)	(117.9,169.95)	(134.8,250.4)	(145.8,358.1)	(153,504.1)

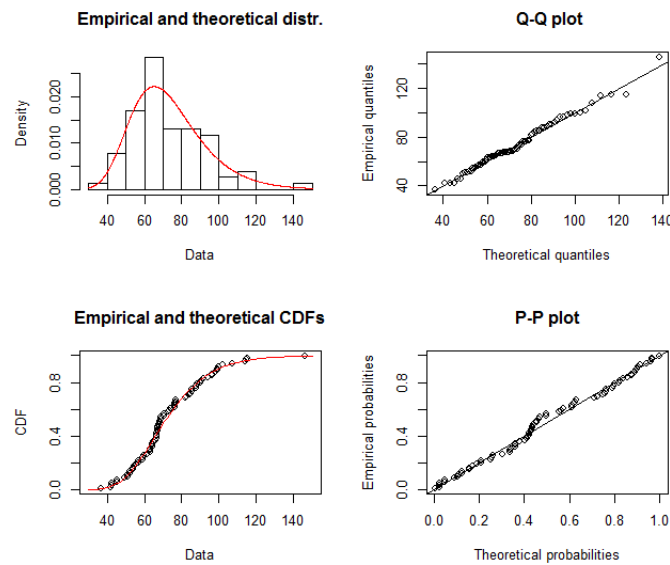


Figura 5.17: *Diagnóstico gráfico do ajustamento GEV aos níveis pluviométricos máximos anuais*

Apesar das estimativas MPP serem ligeiramente maiores, os valores obtidos pelos dois métodos de estimação são bastante semelhantes. Contudo, parece haver uma tendência de diminuição dessa semelhança, à medida que a probabilidade do quantil aumenta. Na mesma tabela podem, também, ser encontrados intervalos com 95% de confiança, obtidos pelo método profile log-likelihood, para esses mesmos quantis, e cuja representação gráfica encontra-se na figura 5.18

Já na tabela 5.8, poderão ser visualizados os níveis de retorno de 100, 400 e 1000 anos, bem como as suas estimativas intervalares. Os valores obtidos são muito semelhantes para os métodos MV e MPP. Assim, usando por exemplo as estimativas MV, espera-se que, em média, um ano em cada 100, 400 ou 1000, terá um dia em que o nível de precipitação é superior a 132.7 mm, 150.4 mm e 161.5 mm, respetivamente.

Tabela 5.8: *Estimativas pontuais MV e MPP e IC a 95% de níveis de retorno do modelo GEV, para os níveis pluviométricos máximos anuais*

Método de Estimação	$\hat{U}(100)$	$\hat{U}(400)$	$\hat{U}(1000)$
MV	132.6966	150.4416	161.4951
MPP	133.9102	152.3238	163.8698
IC(95%)	(117.9,169.95)	(128.8,215.92)	(134.8,250.4)

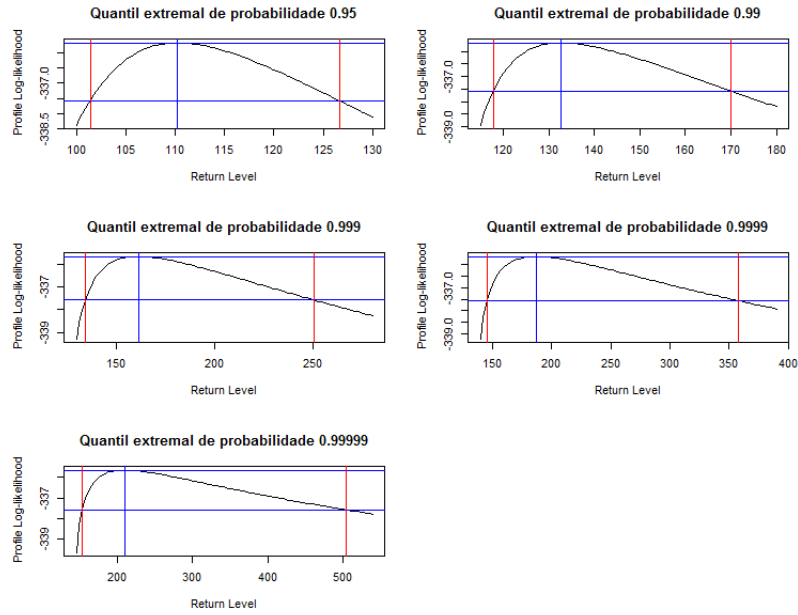


Figura 5.18: *Intervalos de 95% de confiança para quantis extremos do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais*

Na figura 5.19 estão representados os intervalos com uma confiança de 95% para esse níveis de retorno. Repare-se como os intervalos de confiança obtidos via profile log-likelihood não são centrados nas estimativas intervalares, como tinha sido referido aquando da apresentação deste método, na secção 3.2.4.

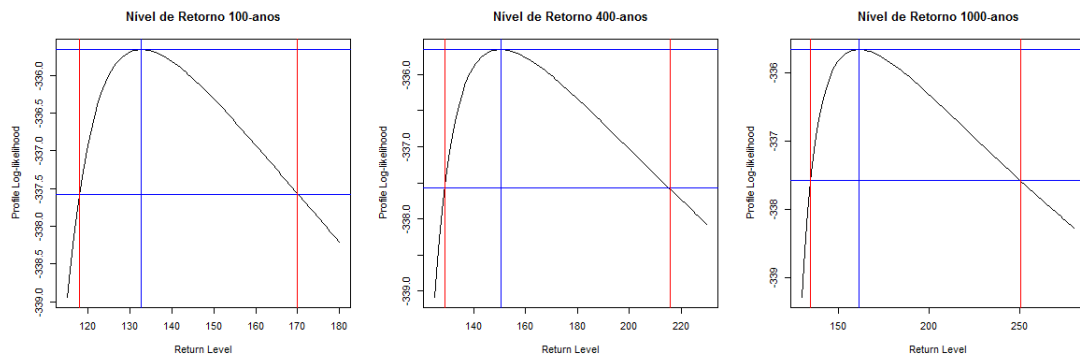


Figura 5.19: *Intervalos de 95% de confiança para níveis de retorno do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais*

Quanto ao período de retorno e à probabilidade de excedência do nível pluviométrico de 146 mm (o máximo da amostra em estudo), usando o modelo ajustado GEV, obtém-se que, em média, uma vez em cada 280 anos esse nível é excedido, sendo a probabilidade de ocorrência dessa excedência de, aproximadamente, 0.00357, usando o método MV. Se se

usar o método MPP, esse nível pluviométrico parece mais fácil de ser atingido. De facto a probabilidade de excedência desse nível aumenta para 0.00407, o que é equivalente a dizer que o nível 146 mm de pluviosidade é ultrapassado uma vez em cada 246 anos. Estes valores estão registados na tabela 5.9.

Tabela 5.9: *Estimativas MV e MPP do período de retorno e probabilidade de excedência do nível 146 mm, no modelo GEV, para os níveis pluviométricos máximos anuais*

Método de Estimação	$\hat{T}(146)$	$\hat{P}(Y > 146)$
MV	280.0469	0.003570831
MPP	245.9725	0.004065494

Como $\hat{\gamma} < 0$, então a distribuição subjacente à amostra poderá pertencer ao domínio Weibull, onde o limite superior do suporte é finito. Assim, poder-se-á obter, também, estimativas para este parâmetro. Usando o estimador indicado em (3.16), obtêm-se os valores registados na tabela 5.10. A estimativa de MV do limite superior do suporte, 398.68 mm, é inferior à obtida usando o método MPP, 444.5298 mm, o que está de acordo com as estimativas dos outros parâmetros já apresentadas, usando o modelo GEV, ou seja, considerando um determinado nível, inferior ao limite superior do suporte, este será mais facilmente atingido se se optar pelas estimativas MPP. Aliás, existem níveis pluviométricos (>398.68 mm) que nunca serão atingidos, quando se usa o método MV, pois excedem o valor da estimativa obtida para o limite superior do suporte.

Tabela 5.10: *Estimativas MV e MPP do limite superior do suporte, no modelo GEV, para os níveis pluviométricos máximos anuais*

Método de Estimação	\hat{x}^F
MV	398.680
MPP	444.5298

Na tabela 5.11 encontram-se todas as estimativas encontradas para os diversos parâmetros, usando os modelos Gumbel e GEV. Os parâmetros diretamente relacionados com as distribuições, nomeadamente, a localização e a escala têm estimativas semelhantes para os dois modelos. Contudo, observando as estimativas dos quantis, dos níveis e períodos de retorno, bem como da probabilidade de excedência, constata-se que a ocorrência de níveis pluviométricos mais elevados é mais rara, se a opção escolhida para modelar esta amostra for a distribuição GEV com forma $\gamma < 0$. Por exemplo, para o modelo Gumbel e usando as estimativas MV, o nível 146 mm é, em média, excedido uma vez em cada

150 anos; já usando o modelo GEV, serão precisos 280 anos para que, num determinado dia, o nível de precipitação em Barcelos ultrapasse esse valor. No entanto este fenómeno é previsível, uma vez que quando $\gamma < 0$, o limite superior do suporte é finito e, como tal, níveis mais elevados (e abaixo do limite superior do suporte) serão mais raros. Nesta mesma tabela, são apresentados os valores da log-verosimilhança obtidos para a estimação dos parâmetros nos dois modelos, que serão utilizados na secção seguinte, nos testes aos modelos extremais.

Tabela 5.11: *Estimativas dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos anuais*

Parâmetros	Modelo Gumbel $\gamma = 0$		Modelo GEV $\gamma < 0$	
	MV	MPP	MV	MPP
$\log L$	-335.8354	-	- 335.6562	-
$\hat{\gamma}$	-	-	-0.04967 (-0.1793777,0.1297275)	-0.04396
$\hat{\lambda}$	63.97622 (60.13723,67.94500)	63.96075	64.4188 (60.3241981,68.6743849)	64.2903
$\hat{\delta}$	16.37472 (13.88045,19.61359)	16.07648	16.6027 (13.9850105,20.0335666)	16.7161
$\hat{q}_{0.95}$	112.6123 (103.4,124.1)	111.7110	110.2670 (101.41,126.7)	110.8355
$\hat{q}_{0.99}$	139.3024 (126.14,156)	137.9150	132.6966 (117.9,169.95)	133.9102
$\hat{q}_{0.999}$	177.0806 (158.36,200.95)	175.0051	161.4951 (134.8,250.4)	163.8698
$\hat{q}_{0.9999}$	214.7921 (190.4,246.2)	212.0298	187.1327 (145.8,358.1)	190.8945
$\hat{q}_{0.99999}$	252.4971 (222.2,291.3)	249.0480	209.9953 (153,504.1)	215.3128
$\hat{U}(100)$	139.3024 (126.14,156)	137.9150	132.6966 (117.9,169.95)	133.9102
$\hat{U}(400)$	162.0643 (145.6,183.2)	160.2623	150.4416 (128.8,215.92)	152.3238
$\hat{U}(1000)$	177.0806 (158.36,200.95)	175.0051	161.4951 (134.8,250.4)	163.8698
$\hat{T}(146)$	150.2811	165.0252	280.0469	245.9725
$\hat{P}(Y > 146)$	0.006654195	0.006059682	0.003570831	0.004065494
\hat{x}^F	-	-	398.6808	444.5298

III) Testes de Hipóteses para o Modelo GEV

Pelas secções anteriores, a distribuição subjacente a esta amostra de níveis pluviométricos máximos anuais poderá pertencer aos domínios max-estáveis Gumbel ou Weibull. Recorde-se que no domínio Gumbel, o EVI é igual a zero e no domínio Weibull este mesmo parâmetro é inferior a zero. Para se tomar a decisão sobre qual será o modelo extremal

mais adequado, poderão ser aplicados os testes apresentados na secção 3.2.5. Neste caso, as hipóteses em teste serão

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma \neq 0, \quad (5.7)$$

ou

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma < 0. \quad (5.8)$$

O primeiro teste a ser aplicado será o Teste de Razão de Verossimilhança (TRV). Este é baseado na estatística Deviance, com distribuição qui-quadrado, e aplica-se para as hipóteses (5.7). Relembre-se que esta estatística é obtida com as log-verossimilhanças dos dois modelos ajustados na secção anterior. Com o auxílio do software R, obtém-se o valor observado da estatística de teste e o respetivo p-value:

d= 0.3582874 d*= 0.3457159 p-value= 0.5565484.

Ora, o valor do p-value obtido é maior do que qualquer nível de significância usual, logo não se rejeita a hipótese nula, isto é, não se rejeita a hipótese de $\gamma = 0$, o que significa que o modelo Gumbel ajustado será o mais adequado para os dados em estudo.

Outro teste, também, para as hipóteses (5.7), é o Teste Score de Rao. A estatística de teste tem duas versões, V_m^* e V_m^{*2} , apresentadas em (3.24) e (3.25). Aplicando os procedimentos do teste, obtêm-se os seguintes valores para as estatísticas e p-values:

Versão Normal: v_m= -6.887117 v_m*= -0.5418667 p-value= 0.5879103;

Versão Qui-quadrado: v^2_m= 47.43238 v^2_m*= 0.2936195 p-value= 0.5879103.

Observando os p-values, que são iguais entre si, uma vez que os testes são equivalentes, ambos são superiores aos níveis de significância habituais e, por isso, mais uma vez não há razões para não se aceitar a nulidade do parâmetro de forma, da distribuição subjacente aos dados. É de lembrar que este teste também poderá ser aplicado às hipóteses definidas em (5.8). Para essa situação tem-se,

Versão Normal: v_m= -6.887117 v_m*= -0.5418667 p-value= 0.2939552

o que leva à decisão de não rejeição da hipótese nula ($\gamma = 0$).

O teste LAN, também pode ser aplicado às duas situações em (5.7) e 5.8. Através do cálculo do valor observado da estatística de teste e da relação (3.27), vem

Teste bilateral: $t_n = -0.3736679$ $t_{n^*} = -0.5412339$ $p\text{-value} = 0.5883464$;

Teste unilateral: $t_n = -0.3736679$ $t_{n^*} = -0.5412339$ $p\text{-value} = 0.2941732$.

Mais uma vez, a hipótese de que a distribuição dos níveis máximos anuais de precipitação é Gumbel, não é rejeitada. Os p-values obtidos são ambos bastante superiores aos níveis de significância habituais de 0.01, 0.05 e 0.1.

Em Tiago de Oliveira and Gomes (1984), recorde-se, é apresentado um outro teste, neste caso, para as hipóteses (5.8). Os valores obtidos para a estatística de teste, a estatística Gumbel, e para o p-value são,

$gs_m = 2.027778$ $gs*_m = -0.2047037$ $p\text{-value} = 0.2931235$.

A conclusão não é diferente das anteriores, não há evidência estatística para não se aceitar que $\gamma = 0$. Um pormenor a salientar é a semelhança entre os p-values para os diferentes testes aplicados (bilaterais com bilaterais e unilaterais com unilaterais) e o facto de nenhum deles se encontrar próximo dos níveis de significância usuais, o que vem reforçar a decisão de não rejeição da hipótese nula.

Na secção 3.2.5 também foram apresentados os testes à qualidade do ajustamento Gumbel, baseados nas estatísticas de ajustamento de Kolmogorov-Smirnov, de Cramér-von Mises e de Anderson-Darling. Estes testes irão permitir concluir se o modelo ajustado Gumbel $G_0(Y|\hat{\lambda}_{G_0}, \hat{\delta}_{G_0})$, com $(\hat{\lambda}_{G_0}, \hat{\delta}_{G_0}) = (63.97622, 16.37472)$, obtido via MV, é ou não adequado para modelar o comportamento dos níveis pluviométricos máximos anuais, em Barcelos. Aceitar esse ajustamento será equivalente a aceitar que $\gamma = 0$, isto é, o modelo Gumbel.

Usando, portanto, $(\hat{\lambda}_{G_0}, \hat{\delta}_{G_0}) = (63.97622, 16.37472)$, os valores observados das referidas estatísticas são:

Goodness-of-fit statistics	
	1-mle-Gumbel
Kolmogorov-Smirnov statistic	0.07328225
Cramer-von Mises statistic	0.05803319
Anderson-Darling statistic	0.32826543

A hipótese nula será rejeitada se o valor observado destas estatísticas, devidamente adaptadas, for superior a um determinado quantil de nível assintótico α .

Para a estatística Kolmogorov-Smirnov, tem-se

$$\sqrt{m} \cdot d_m = \sqrt{77} \cdot 0.07328225 = 0.6430491,$$

e comparando com os quantis registados na tabela 3.1, para o caso $m = 77$, observa-se que, para qualquer um dos níveis de significância α , o valor da estatística é inferior ao quantil de nível assintótico α . Quer isto dizer que não se rejeita a hipótese nula.

Para a estatística Cramér-von Mises, tem-se

$$w_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 0.05803319 \cdot \left(1 + \frac{0.2}{\sqrt{77}}\right) = 0.05935589$$

e observando os quantis da tabela 3.2, na linha correspondente, constata-se que, também neste caso, o valor observado da estatística não excede nenhum dos quantis. Por esse motivo, a conclusão é a de não rejeição de H_0 .

Por fim, para a estatística Anderson-Darling,

$$a_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 0.32826543 \cdot \left(1 + \frac{0.2}{\sqrt{77}}\right) = 0.3357473.$$

Também na tabela 3.2 encontram-se os quantis para esta estatística. Mais uma vez o valor observado da estatística de teste é inferior a qualquer um dos quantis tabelados, não se rejeitando, por isso, a hipótese nula.

Portanto, os três testes à qualidade do ajustamento Gumbel, com os parâmetros estimados, conduzem à não rejeição deste modelo, sendo, por isso adequado para modelar os dados em estudo.

Observando as conclusões dos diversos testes aplicados, estes foram unânimes na decisão de não rejeição de $\gamma = 0$, ou seja, de que o modelo Gumbel é o que melhor se ajusta aos níveis pluviométricos máximos anuais, em Barcelos.

IV) Avaliação da Não-Estacionariedade

Cada vez mais se ouve falar nas mudanças climáticas e no fenómeno aquecimento global. Vários estudos mostram que a partir da última década do século XX, tornou-se clara uma tendência de aquecimento. Em Portugal, desde a década de 1970, a temperatura média subiu a uma taxa de cerca de $0.5^{\circ}\text{C}/\text{década}$, mais do dobro da taxa de aquecimento observada para a temperatura média mundial (Miranda *et al.*, 2006). Esse aquecimento tem provocado alterações nos níveis de precipitação, registando-se, neste caso, uma tendência significativa de diminuição no seu valor médio anual.

Pelos motivos apresentados no parágrafo anterior, faz sentido analisar a existência de tendência nos níveis pluviométricos que constituem a amostra em estudo. Como foi

referido na secção 3.2.6, quase sempre as variações observadas numa determinada categoria ambiental são bem modeladas por uma tendência linear no parâmetro de localização, da distribuição GEV. Assim, pretende-se obter as estimativas dos parâmetros $(\gamma, \lambda(t), \delta)$, com $\lambda(t) = \beta_0 + \beta_1 t$ e $t = 1, \dots, 77$. Assim, aplicando o método MV, obtêm-se as seguintes estimativas para a forma, localização e escala do modelo GEV:

$$(\hat{\gamma}, \hat{\lambda}(t), \hat{\delta}) = (-0.04227, 66.1861 - 0.0468 \cdot t, 16.4855), \quad t = 1, \dots, 77.$$

Na figura 5.20 pode ser visualizado o ajustamento linear na localização, à nuvem dos níveis pluviométricos máximos anuais.

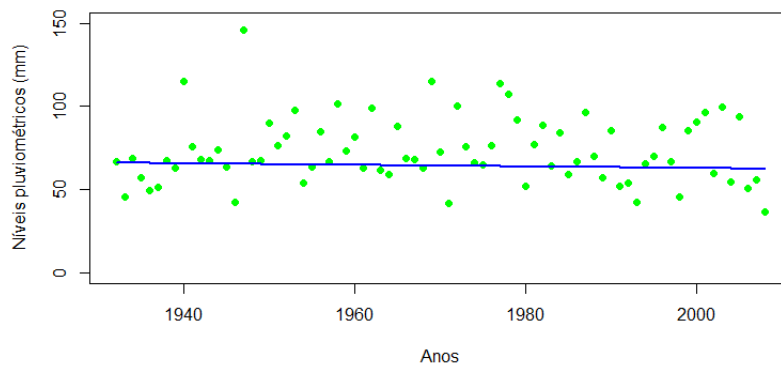


Figura 5.20: *Ajustamento linear na localização, para os níveis pluviométricos máximos anuais*

O declive da reta ajustada é negativo, o que significa que o nível de precipitação médio anual, em Barcelos, está a diminuir. Mas será que essa tendência é estatisticamente significativa? O facto do declive estar tão próximo de zero, e que é visível pela quase horizontalidade da reta, poderá ser um indício de não existência de tendência na localização. Será necessário recorrer a um teste de hipóteses para se obter uma resposta. Esse teste é baseado nas razões de verosimilhança dos modelos com e sem tendência. Na tabela 5.12 é apresentado um resumo das estimativas dos parâmetros dos modelos GEV, com e sem tendência, os quais vão ter os nomes genéricos M_1 e M_0 , respetivamente, bem como os valores da log-verosimilhança que permitem a realização do teste estatístico.

As hipóteses em teste são,

$$H_0 : \text{Não há tendência na amostra} \quad vs \quad H_1 : \text{Há tendência na amostra.}$$

A estatística e a sua distribuição assintótica e a forma da região de rejeição do teste para avaliar a presença de tendência foram apresentadas na secção 3.2.6. Neste caso,

Tabela 5.12: *Estimativas MV dos modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos anuais*

Modelo	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\delta}$	$\log L$
M_0 (sem tendência)	-0.04967	64.4188	16.6027	-335.6562
M_1 (com tendência)	-0.04227	66.1861-0.0468 · t	16.4855	-335.478

para o modelo sem tendência, M_0 , foram estimados três parâmetros e para o modelo com tendência, M_1 , foram estimados quatro parâmetros. Logo, trata-se de um caso de fronteira e, como foi indicado na secção 3.2.6, a distribuição assintótica da estatística de teste, neste caso, será uma mistura 50 : 50 de dois Qui-quadrados com 3 e 4 g.l.

O valor observado da estatística de teste e o p-value obtidos foram, respetivamente:

d_obs=0.3564

p-value=0.9674914

Como o p-value é bastante superior a qualquer nível de significância usual, então, não se rejeita a hipótese nula aos níveis de significância usuais. Assim, conclui-se que a diferença entre os modelos com e sem tendência não é estatisticamente significativa, logo aumentar o número de parâmetros não vai melhorar a capacidade preditiva do modelo. Assim, não há evidência estatística da presença de tendência na localização da distribuição subjacente aos níveis pluviométricos máximos anuais. Através dos critérios de informação AIC e BIC (tabela 5.13) chega-se à mesma conclusão. O modelo sem tendência tem os menores valores de AIC e BIC, logo é considerado o melhor modelo.

Tabela 5.13: *AIC e BIC para os modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos anuais*

Modelo	AIC	BIC
M_0 (sem tendência)	675.3124	688.6876
M_1 (com tendência)	678.956	705.7064

5.2.1.2 Níveis Pluviométricos Máximos Mensais

I) Análise Preliminar dos Dados

Esta amostra é constituída pelos níveis pluviométricos máximos, para cada mês, no período de 01/01/1932 a 01/05/2008. Tem-se, portanto, um total de 917 meses, contudo

não existem registos para 7 meses, logo esta amostra tem uma dimensão igual a $m = 910$ observações. Estes níveis pluviométricos encontram-se representados na figura 5.21.

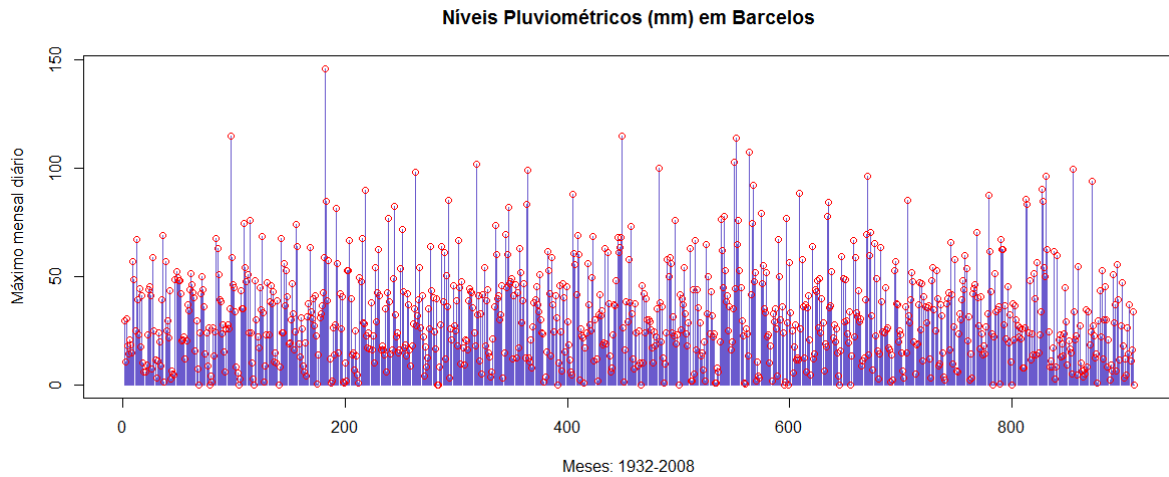


Figura 5.21: *Níveis pluviométricos máximos mensais, em Barcelos, (1932-2008)*

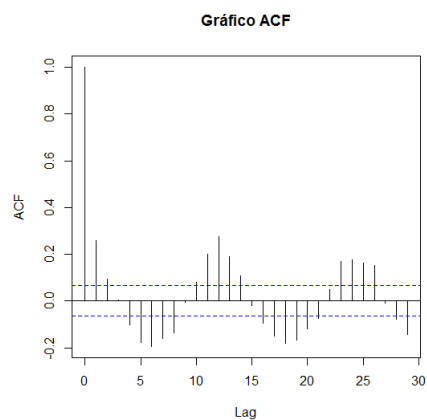


Figura 5.22: *Gráfico de Autocorrelação Parcial*

Através do gráfico de autocorrelação parcial, na figura 5.22, poder-se-á avaliar a independência das observações. Como se pode constatar, existe uma correlação entre as observações, pois as linhas ultrapassam as bandas de confiança limite para se admitir independência entre os dados. Esta correlação não é de estranhar, pois o nível de precipitação varia consoante o período do ano em que nos encontramos. Ou seja, espera-se que os níveis de precipitação nos meses de outono e inverno sejam muito maiores do que nos meses da primavera e verão. Assim, os níveis pluviométricos começam por ser elevados nos dois primeiros meses do ano, para depois começarem a decrescer de março até agosto,

voltando a subir a partir de setembro. Esta relação repete-se de ano para ano, e daí a não independência entre observações mensais. Por outras palavras, suponha-se, por exemplo, que se está num mês de inverno, então o nível pluviométrico do mês seguinte, se ainda pertencer a esta estação do ano, é provável que seja elevado. Já se nos encontrarmos no verão, o mês seguinte tem uma maior probabilidade de ser seco do que chuvoso. Contudo, com o objetivo de fazer uma comparação com os valores obtidos por Nascimento (2009), que considerou uma abordagem Bayesiana aos níveis pluviométricos máximos mensais em Barcelos no mesmo período, como já referido anteriormente, ignorar-se-á esta correlação entre os dados, procedendo-se à aplicação da análise em valores extremos e admitindo a existência de independência. Quando se aplicar a metodologia POT a esta mesma amostra, ter-se-á a oportunidade de ver que esta correlação diminuirá. Também, dividindo esta amostra por estação do ano, obter-se-á uma redução da correlação entre as observações, pois nesse caso estar-se-á a eliminar o efeito da sazonalidade, referido anteriormente.

Para esta amostra as características amostrais habituais são:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	15.00	28.00	31.64	44.00	146.00

Nesta amostra os níveis pluviométricos variam entre os 0 mm e os 146 mm, com o nível mediano de pluviosidade a rondar os 28 mm. Tem-se, portanto, uma amostra com uma acentuada assimetria à direita que, também, pode ser visível no histograma e no box-plot representados na figura 5.23. No box-plot essa assimetria é evidenciada pela quantidade elevada de candidatos a outliers.

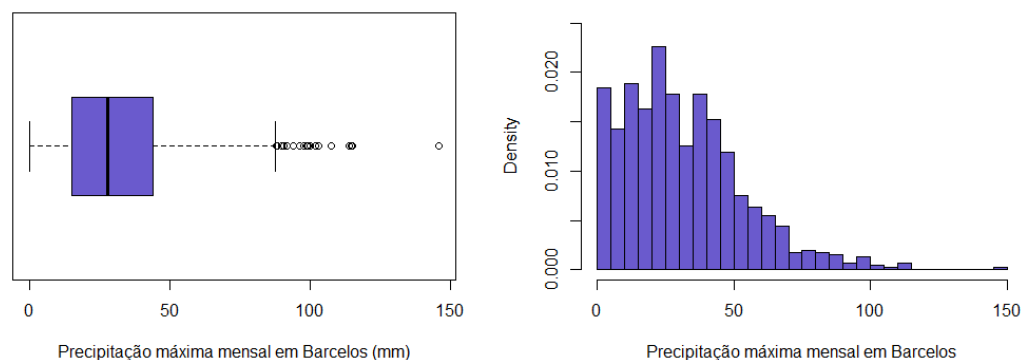


Figura 5.23: *Boxplot e histograma dos níveis pluviométricos máximos mensais, em Barcelos, (1932-2008)*

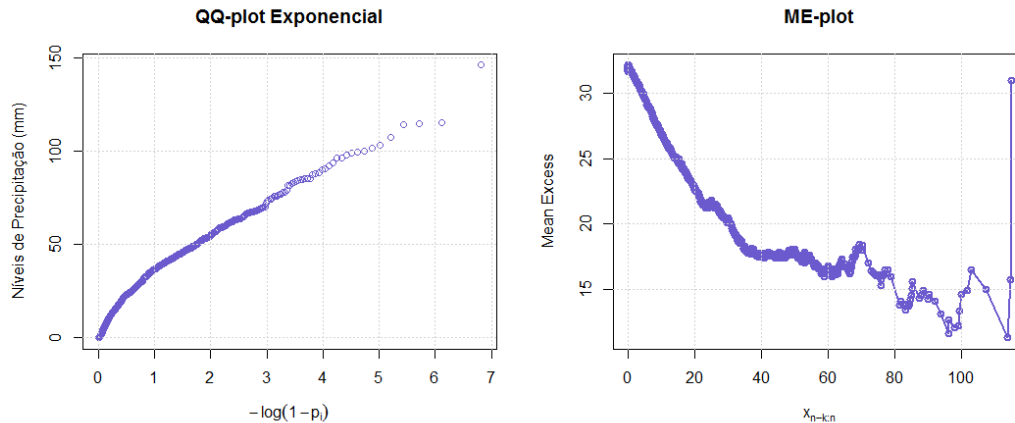


Figura 5.24: *QQ-Plot Exponencial e ME-plot dos níveis pluviométricos máximos mensais, em Barcelos, (1932-2008)*

Com o intuito de ter uma ideia sobre o tipo de cauda da distribuição subjacente a estes dados, poder-se-á recorrer aos, já apresentados, qq-plot exponencial e ME-plot. No qq-plot exponencial representado na figura 5.24, nota-se um ligeiro padrão côncavo na nuvem de pontos, o que sugere que a cauda da distribuição será sensivelmente mais leve que a da exponencial ou mesmo do tipo exponencial. São, então, fortes candidatos os domínios Gumbel ou Weibull. Quanto ao ME-plot, apresentado na mesma figura, inicialmente existe um acentuado decrescimento que depois se torna mais moderado, a partir do nível 35 mm, aproximadamente. Assim, tendo em conta esse pormenor, poder-se-á concluir que também o ME-plot sugere que a distribuição subjacente aos níveis pluviométricos máximos mensais, apresenta uma cauda direita do tipo exponencial ou ligeiramente mais leve.

Seguindo as sugestões dadas pelos dois gráficos anteriores, comece-se por realizar um ajustamento preliminar ao modelo Gumbel. Na figura 5.25 encontra-se representado o qq-plot Gumbel, juntamente com a reta ajustada à nuvem de pontos. Os pontos, na sua maioria, estão sobre a reta e como tal o ajustamento Gumbel parece bastante bom. No início existe um desvio dos pontos em relação à reta correspondendo, precisamente, aos níveis pluviométricos iguais a zero. Também na parte final do qq-plot, como habitual, existem alguns pontos ligeiramente afastados da reta ajustada. Apesar disso, a correlação é muito alta, cerca de 0.9980137, o que vem mostrar a forte associação linear entre os pontos. Através da ordenada na origem e do declive da reta, obtêm-se as seguintes estimativas dos mínimos quadrados para a localização e escala do modelo Gumbel, respetivamente,

Call:

```
lm(formula = x1_ord ~ pi)
```

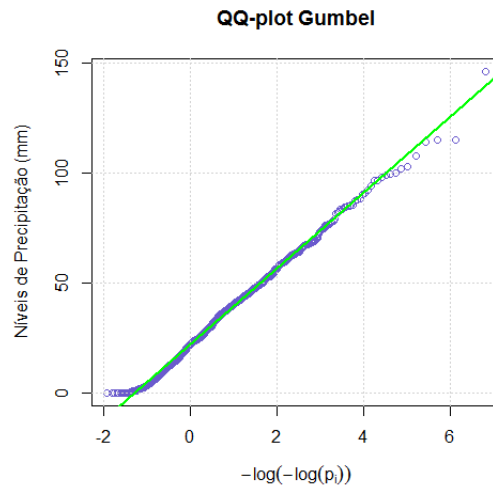


Figura 5.25: *QQ-Plot Gumbel dos níveis pluviométricos máximos mensais, em Barcelos, (1932-2008)*

Coefficients:

(Intercept)	pi
21.72211	17.27529

ou seja,

$$(\hat{\lambda}, \hat{\delta}) = (21.72211, 17.27529). \quad (5.9)$$

Pretendendo-se de seguida fazer um ajustamento preliminar da distribuição GEV aos dados, é preciso, em primeiro lugar, encontrar uma estimativa para o parâmetro de forma. Usando o mesmo método aplicado na amostra dos níveis pluviométricos máximos anuais, obtêm-se para essa estimativa e respetiva correlação máxima (figura 5.26), os valores

`$maximum`

[1] -0.01184;

`$objective`

[1] 0.9980871.

Ou seja, um parâmetro de forma $\gamma < 0$, mas próximo de zero. Fixando $\hat{\gamma} = -0.01184$, obtêm-se o qq-plot GEV representado na figura 5.26 e a reta ajustada aos dados.

Também, neste caso, é evidente a forte associação linear. Apenas existem alguns pontos no início e no fim do qq-plot que se desviam ligeiramente da reta ajustada, tal como acontecia no qq-plot Gumbel. Obtêm-se como estimativas dos parâmetros de localização e escala, os valores

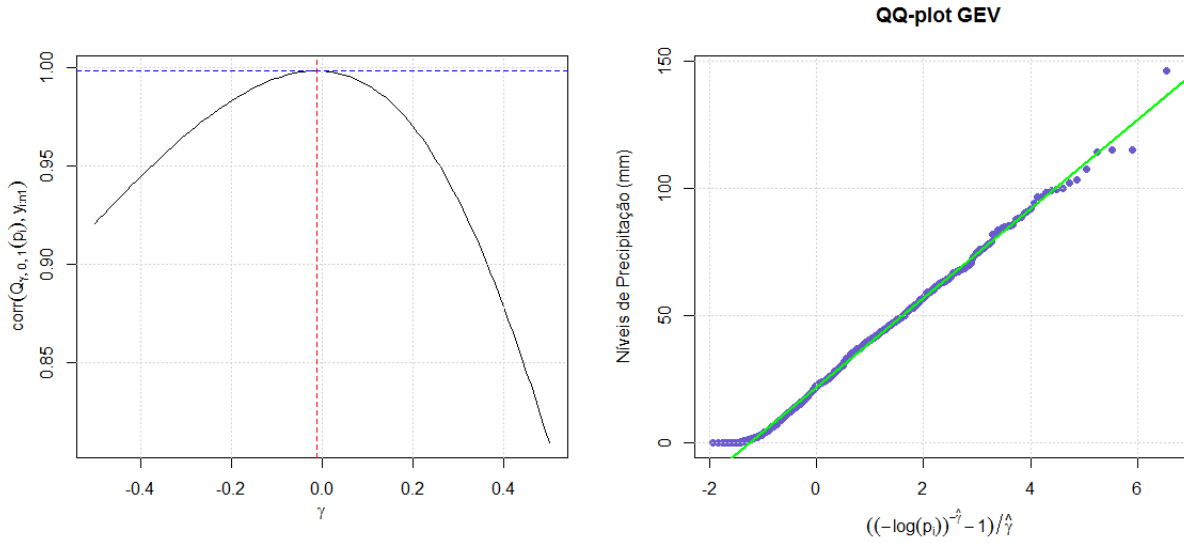


Figura 5.26: Gráfico de correlação entre $\hat{Q}_{\gamma, \lambda, \delta}(p)$ e $Q_{\gamma, 0, 1}(p)$ e QQ-Plot GEV dos níveis pluviométricos máximos mensais, em Barcelos, (1932-2008)

Call:

```
lm(formula = sort(x) ~ Qgev)
```

Coefficients:

(Intercept)	Qgev
21.7759	17.5279

Portanto, para a amostra em estudo, as estimativas preliminares dos parâmetros do modelo GEV são

$$(\hat{\gamma}, \hat{\lambda}, \hat{\delta}) = (-0.01184, 21.7759, 17.5279). \quad (5.10)$$

Na figura 5.27, encontram-se representados o histograma dos níveis pluviométricos máximos mensais e as densidades estimadas dos modelos Gumbel e GEV. As duas curvas estão, praticamente sobrepostas e parecem modelar bastante bem o comportamento da cauda direita da amostra.

Através desta abordagem preliminar, conclui-se que a distribuição dos níveis pluviométricos máximos mensais, em Barcelos, terá uma cauda direita mais leve que a da exponencial, podendo ser ajustada pelos modelos extremos Gumbel ou Weibull. Nas seções seguintes obter-se-ão resultados mais potentes para as estimativas dos parâmetros e sobre a distribuição destas observações.

II) Estimação dos Parâmetros

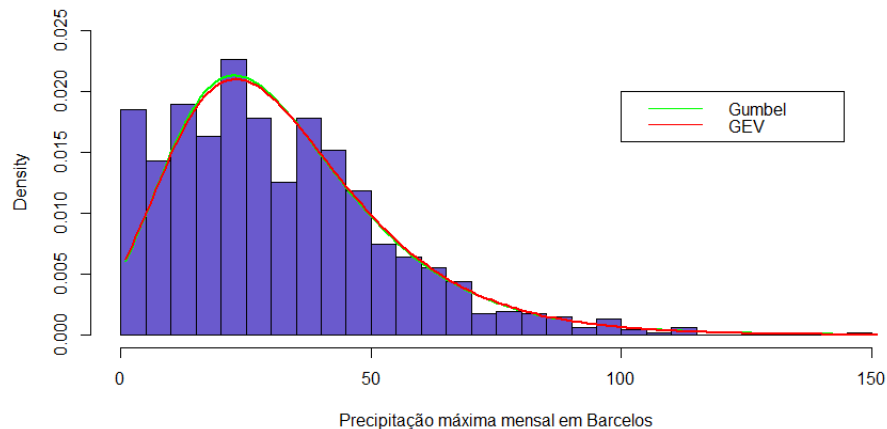


Figura 5.27: Densidades dos modelos Gumbel e GEV estimados ajustados aos níveis pluviométricos máximos mensais

Tal como feito para a amostra de máximos anuais e, seguindo as sugestões da análise preliminar efetuada na secção anterior, irá proceder-se ao ajustamento MV e MPP dos modelos Gumbel (que corresponde à distribuição GEV com parâmetros de forma nulo) e GEV (com $\gamma \neq 0$) para a amostra dos níveis pluviométricos máximos mensais. Na tabela 5.14 estão representadas as estimativas pontuais para os parâmetros de forma (apenas no modelo GEV), localização e escala das duas distribuições. Encontram-se também registados os intervalos a 95% de confiança para esses parâmetros, baseados na função de profile log-likelihood.

Tabela 5.14: Estimativas dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos mensais

Parâmetros	Modelo Gumbel ($\gamma = 0$)		Modelo GEV ($\gamma \neq 0$)	
	MV	MPP	MV	MPP
$\log L$	-4023.362	-	- 4022.886	-
$\hat{\gamma}$	-	-	0.02743 (-0.02618, 0.08713)	-0.01634
$\hat{\lambda}$	21.65848 (20.50152, 22.8323)	21.61636	21.40587 (20.15364, 22.68234)	21.74679
$\hat{\delta}$	17.03685 (16.18907, 17.9492)	17.37081	16.849996 (15.93688, 17.83249)	17.6311

As estimativas dos parâmetros de localização e de escala, para os dois modelos, quer usando o método MV quer usando o método MPP, são muito semelhantes. Quanto ao parâmetro de forma, a estimativa MV, ao contrário do que se estava à espera, é positiva, mas, concluir daqui, que a distribuição subjacente à amostra tem cauda pesada

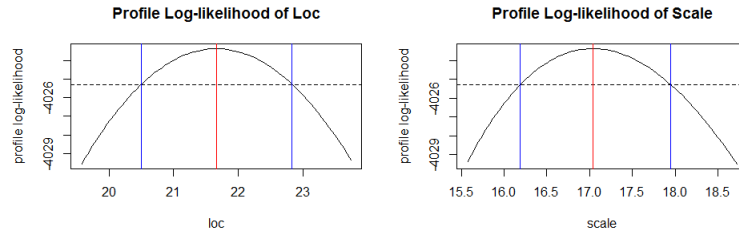


Figura 5.28: Intervalos de 95% de confiança para os parâmetros do modelo Gumbel, baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais

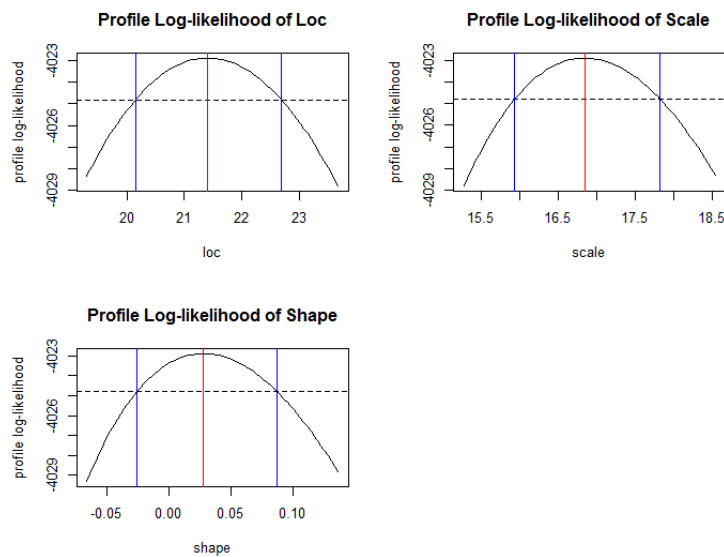


Figura 5.29: Intervalos de 95% de confiança para os parâmetros do modelo GEV, baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais

é questionável, pois o EVI estimado está muito próximo de zero. Já a estimativa MPP é negativa. No entanto, apesar da diferença de sinais, ambas as estimativas estão muito próximas de zero e o intervalo de confiança para este parâmetro também inclui o zero. Logo, é de suspeitar que o modelo Gumbel deverá ser o mais adequado para modelar o comportamento destes dados, mas esse tópico será desenvolvido na próxima seção. Nas figuras 5.28 e 5.29 estão representados os intervalos de confiança para os parâmetros das distribuições Gumbel e GEV.

A qualidade dos dois ajustamentos, obtidos pelo método MV, pode ser avaliada através das figuras 5.30 e 5.31. Através da análise dos gráficos aí representados, poder-se-á dizer que ambos os ajustamentos são bons. Contudo, observando com mais detalhe os qq-plot, nas duas figuras, e em particular a zona final destes, as observações mais elevadas

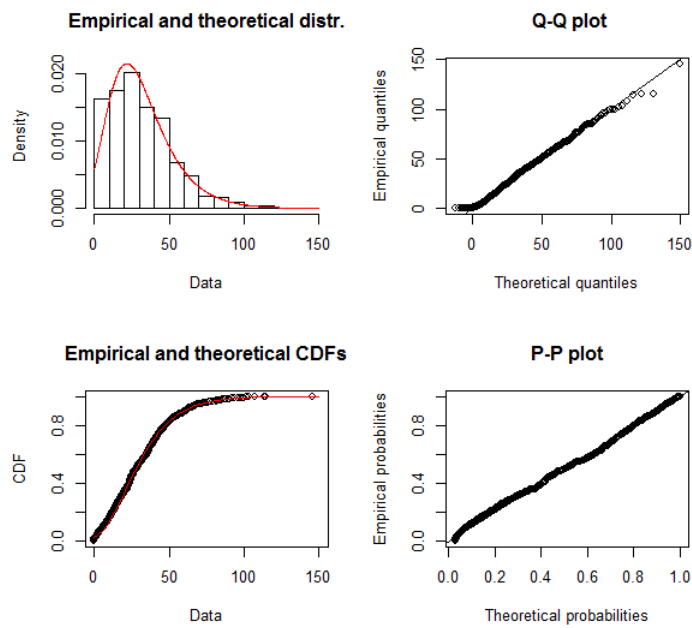


Figura 5.30: *Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos mensais*

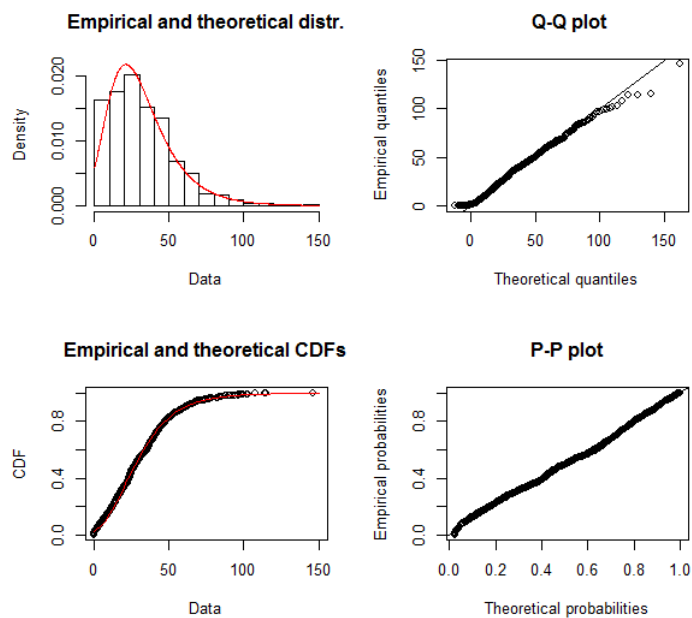


Figura 5.31: *Diagnóstico gráfico do ajustamento GEV aos níveis pluviométricos máximos mensais*

têm uma maior proximidade à reta, no caso do modelo Gumbel, o que poderá ser mais um indício de que este será o melhor modelo para explicar o comportamento dos níveis pluviométricos máximos mensais, em Barcelos.

Como já foi referido, por diversas vezes ao longo desta dissertação, na sequência de Nascimento (2009), em Nascimento *et al.* (2012) foi publicada uma abordagem Bayesiana semi-paramétrica à estimação de valores extremos e, em particular, são apresentadas estimativas de quantis elevados, para os níveis pluviométricos máximos mensais, na cidade de Barcelos, tendo como base a amostra de observações máximas mensais no período de 01/01/1932 a 01/05/2008. Esses quantis estão reproduzidos na figura 5.32. Nesse traba-

<i>Prob</i>	Barcelos (in mm)				
	E	1	2	3	<i>MG_k</i>
0.95	73.5	74.54	77.54	73.29	74.71
0.99	99.4	101.73	105.38	102.24	104.09
0.999	N/A	137.84	139.91	137.77	151.50
0.9999	N/A	171.41	176.12	184.54	233.00
0.99999	N/A	203.13	236.24	454.26	333.00

Prob = $P(X \leq q)$, E = Empirical, 1 = *MGPD*₁, 2 = *MGPD*₂, 3 = *MGPD*₃, *MG_k* refers to the best model in this class

Figura 5.32: *Quantis elevados para os níveis pluviométricos máximos mensais, em Barcelos (Nascimento et al., 2012)*

lho foram utilizadas aproximações não paramétricas através da mistura de distribuições Gama, para a distribuição dos dados abaixo de um determinado limiar, com a distribuição GP, para as observações acima desse limiar. Foram estimados vários modelos, uns só com mistura de k Gammas, os quais foram denominados MG_k , com $k = 1, \dots, 5$, e outros com a mistura de k Gammas com a distribuição GP, $MGPD_k$, com $k = 1, \dots, 3$. Através de critérios de informação, retiveram-se apenas os modelos $MGPD_k$, com $k = 1, \dots, 3$, e o modelo MG_3 . De todos esses modelos, o $MGPD_3$ revelou-se o melhor modelo, por apresentar os menores valores dos critérios de informação utilizados.

Pretende-se, portanto, fazer uma pequena comparação entre os quantis calculados com esses modelos, com os obtidos usando os modelos Gumbel e GEV, anteriormente ajustados, e que se apresentam na figura 5.15.

A primeira constatação a ser feita, tem a ver com os quantis empíricos, que para probabilidades muito elevadas e devido à dimensão da amostra, estes assumem o maior valor observado, não se conseguindo, assim, obter valores mais pormenorizados. Repare-se

Tabela 5.15: *Quantis extremais empíricos e estimativas pontuais MV e MPP e IC a 95% para esses quantis, com modelos Gumbel e GEV para os níveis pluviométricos máximos mensais*

Quantis	Empíricos	Modelo Gumbel $\gamma = 0$		Modelo GEV $\gamma \neq 0$	
		MV	MPP	MV	MPP
$\hat{q}_{0.95}$	72.0	72.26125 (69.2,75.53)	73.21106	73.54891 (69.59,78.54)	72.86399
$\hat{q}_{0.99}$	99.1	100.03053 (95.64,104.73)	101.52468	104.02104 (95.41,115.95)	99.87891
$\hat{q}_{0.999}$	146.0	139.33635 (133,146.14)	141.60098	149.54935 (129.85,179.1)	136.90754
$\hat{q}_{0.9999}$	146.0	178.57282 (170.3,187.42)	181.60656	197.96367 (161.9,256.5)	172.50436
$\hat{q}_{0.99999}$	146.0	217.80238 (207.62,228.67)	221.60511	249.52573 (192.6,350.9)	206.78045

que estes valores têm pequenas diferenças quando comparados com os quantis empíricos obtidos em Nascimento *et al.* (2012). Quando se usa o modelo Gumbel, os quantis obtidos não são muito diferentes entre os métodos MV e MPP. Já quando se usa o modelo GEV, à medida que o nível do quantil aumenta, o mesmo acontece com as diferenças entre as estimativas MV e MPP, o que não é de estranhar devido à diferença de sinal para a forma da distribuição, obtida por cada um desses métodos. Usando a distribuição GEV, o método MPP torna níveis pluviométricos elevados mais raros do que o método MV. De facto, a estimativa MV de γ é um valor positivo, logo o limite superior do suporte não é finito, sendo, por isso, mais fácil obter-se níveis mais elevados de precipitação. Este pormenor também é visível nas amplitudes dos intervalos de confiança obtidos para os modelos Gumbel e GEV, com $\gamma > 0$, usando a função de profile log-likelihood. Por exemplo, a amplitude do IC a 95% para o quantil de probabilidade 0.99999, usando o modelo Gumbel, é cerca de 21 mm, mas quando usada a distribuição GEV estimada, esta amplitude aumenta para 158 mm.

Quando comparados os quantis extremais obtidos com os apresentados em Nascimento *et al.* (2012), verifica-se que a situação de maior semelhança acontece entre o modelo $MGPD_1$ com o modelo GEV ajustado através do método MPP. Em relação aos quantis obtidos no modelo $MGPD_3$, considerado o melhor modelo de entre os ajustados nesse estudo, constata-se que os quantis de probabilidade 0.95, 0.99, 0.999 e 0.9999 são semelhantes aos obtidos usando os modelos Gumbel (com estimativas MV ou MPP) ou Gev (com estimativas MPP). Em relação ao quantil de probabilidade 0.99999, o valor obtido com o modelo $MGPD_3$ é de longe mais elevado do que qualquer um dos obtidos nos modelos Gumbel e GEV. Note-se que esta estimativa também não pertence ao intervalo de confiança para este quantil, no modelo GEV, que tem uma elevada amplitude, como

já referido.

Quanto aos níveis de retorno, pretende-se obter estimativas destes para 100, 400 e 1000 anos. Como se está a usar uma amostra de máximos mensais, então é preciso converter os anos em meses. Assim, pretende-se estimar os níveis pluviométricos (em mm) que são excedidos, em média, uma vez em cada 1200, 4800 e 12000 meses. Na tabela 5.16 apresentam-se as estimativas pontuais e intervalares para esses níveis.

Tabela 5.16: *Estimativas pontuais MV e MPP e IC a 95% para níveis de retorno, nos modelos Gumbel e GEV, para os níveis pluviométricos máximos mensais*

Níveis de Retorno	Modelo Gumbel $\gamma = 0$		Modelo GEV $\gamma \neq 0$	
	MV	MPP	MV	MPP
$\hat{U}(1200)$	142.4440 (136,149.4)	144.7695	153.2733 (132.5,184.6)	139.7760
$\hat{U}(4800)$	166.0674 (158.4,174.24)	168.8560	182.1998 (152,230)	161.3041
$\hat{U}(12000)$	181.6791 (173.27,190.74)	184.7738	201.9289 (164.4,263.2)	175.2657

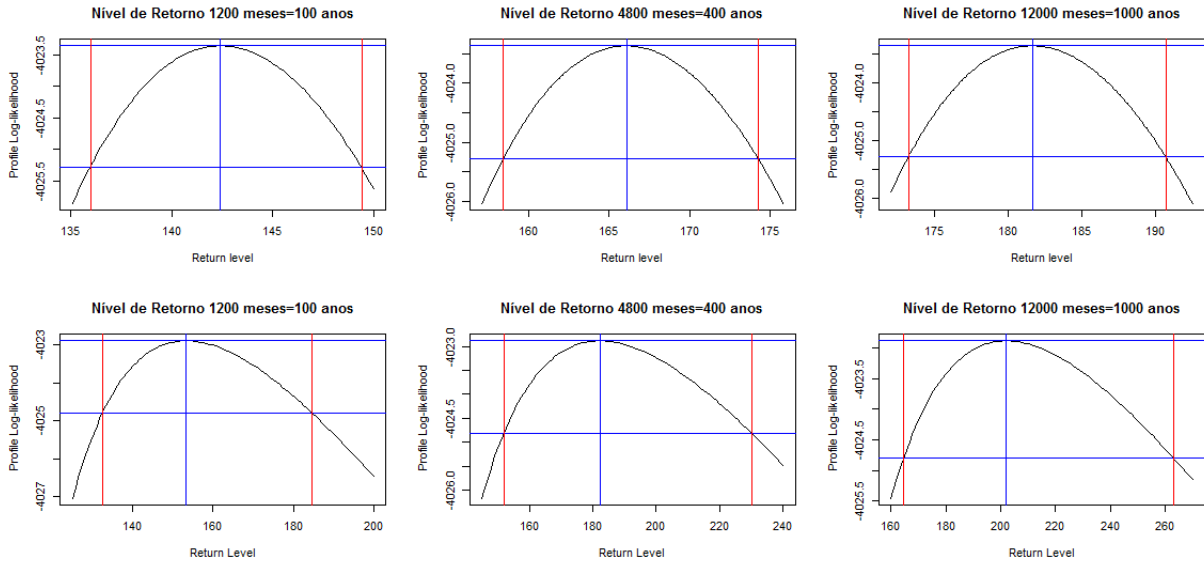


Figura 5.33: *Intervalos de 95% de confiança para níveis de retorno dos modelos Gumbel (linha superior) e GEV (linha inferior), baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais*

Quando usados os modelos Gumbel, com estimativas MV e MPP, e GEV, com estimativas MPP, os níveis de retorno são semelhantes, embora ligeiramente inferiores no

último caso. No caso do modelo GEV, com estimação obtida pelo método MV, os níveis de retorno são mais elevados e, para além disso, os intervalos de 95% de confiança, neste caso, abrangem uma maior quantidade de níveis possíveis para cada um dos horizontes temporais considerados. Essa diferença de amplitudes nas estimativas intervalares, entre os modelos Gumbel e GEV, pode ser visualizada na figura 5.33.

Na tabela 5.17 apresentam-se as estimativas do período de retorno do nível pluviométrico máximo observado, bem como da probabilidade desse nível ser excedido.

Tabela 5.17: *Estimativas MV e MPP do período de retorno e probabilidade de excedência do nível 146 mm nos modelos Gumbel e GEV, para os níveis pluviométricos máximos mensais*

Parâmetros	Modelo Gumbel $\gamma = 0$		Modelo GEV $\gamma \neq 0$	
	MV	MPP	MV	MPP
$\hat{T}(146)$	1478.414 meses (123.2012 anos)	1288.051 meses (107.3376 anos)	839.7848 meses (69.98207 anos)	1785.771 meses (148.8142 anos)
$\hat{P}(Y > 146)$	0.0006764	0.0007764	0.0011908	0.00055998

É notória a diferença entre o valor obtido para o período de retorno, usando o modelo GEV ajustado pelo método MV, e os restantes. No entanto, essa diferença não é de estranhar pelo que já tinha sido referido. Neste caso, através do ajustamento MV, valores elevados de precipitação não são considerados tão raros como pelos outros ajustamentos e, por isso, estima-se que são precisos, em média, cerca de 70 anos para que o nível pluviométrico de 146 mm seja ultrapassado anualmente, sendo a probabilidade desse acontecimento de 0.00119. Note-se que se se considerar o ajustamento GEV pelo método MPP, esse período é de 148.8142 anos, mais do dobro do anterior, e, em compensação, a probabilidade de excedência do nível 146 mm reduz-se para metade, 0.00055998. Usando o modelo Gumbel, há uma maior estabilidade entre os valores obtidos pelos dois métodos de estimação, embora com uma diferença de 16 anos para o período de retorno.

Relembre-se que apenas o método MPP forneceu uma estimativa negativa para o parâmetro de forma da distribuição GEV, logo só neste caso poder-se-á calcular uma estimativa do limite superior do suporte, recorrendo à expressão em (3.16). Assim, obtém-se

```
lss<-21.74679-17.6311/-0.01634182
[1] 1100.641
```

ou seja, o nível de precipitação máximo mensal que alguma vez poderia ser atingido, em Barcelos, seria de 1100.641 mm, uma autêntica catástrofe.

III) Testes ao Modelo GEV

Nesta secção proceder-se-á aplicação dos testes estatísticos apresentados no capítulo 3.2.5, para se determinar qual dos modelos extremais melhor se ajusta à distribuição subjacente a esta amostra de máximos pluviométricos mensais.

A análise preliminar destes dados apontaram para duas situações possíveis, os modelos Gumbel ou Weibull. Contudo, a estimativa de MV obtida para o parâmetro de forma é positiva, o que conduz a distribuição dos dados para o modelo Fréchet. Já tinha sido referido que, devido à proximidade da origem, por parte das estimativas obtidas para esse parâmetro, o modelo Gumbel seria o mais provável, contudo isso foi uma mera suposição. Assim, comece-se por aplicar os testes para as hipóteses:

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma \neq 0, \quad (5.11)$$

Relembre-se que, para estas hipóteses, podem ser utilizados os testes de Razão de Verossimilhanças, Score de RAO (versão Normal e versão Qui-Quadrado) e o teste LAN. Na tabela 5.18 apresentam-se, para cada um dos testes, o valor observado da estatística de teste, o p-value obtido e a respetiva decisão com uma significância de 5%.

Tabela 5.18: Resultados obtidos nos testes às hipóteses definidas em (5.11)

Teste	Valor Observado da Estatística de Teste	p-value	Decisão $\alpha = 0.05$
TRV	$d^* = 0.9506424$	0.3295558	Não rejeitar H_0
Teste Score de RAO (versão Normal)	$v_m^* = 0.8121647$	0.4166971	Não rejeitar H_0
Teste Score de RAO (versão Qui-Quadrado)	$v_m^{2*} = 0.6596115$	0.4166971	Não rejeitar H_0
Teste LAN	$t_n^* = 0.811849$	0.4168783	Não rejeitar H_0

Portanto, a decisão de não rejeitar a hipótese nula é consensual. Na tabela apresentada usou-se um nível de significância de 5%, mas os p-values são superiores aos níveis de significância usuais, logo não se rejeita a hipótese de que a distribuição subjacente aos níveis pluviométricos máximos mensais seja Gumbel.

O teste baseado na estatística Gumbel poderá, também, ser utilizado. Como esta estatística se baseia em valores da amostra ordenada, fará sentido testar a sugestão fornecida pela análise preliminar dos dados, ou seja,

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma < 0. \quad (5.12)$$

O valor observado para a estatística de teste e o p-value são,

gs_m= 2.842105 gs*_m= 0.04067683 p-value= 0.3828395.

Logo, aos níveis de significância usuais, não se rejeita a hipótese de $\gamma = 0$.

Também poderão ser aplicados os testes à qualidade do ajustamento Gumbel, usando as estatísticas de ajustamento de Kolmogorov-Smirnov, de Cramér-von Mises e de Anderson-Darling. Na tabela 5.19 encontram-se representados o valor observado e o respetivo valor transformado, para cada uma das estatísticas mencionadas. Também se apresenta a decisão a ser tomada, com um nível de significância de 5%.

Tabela 5.19: *Resultados obtidos nos testes à qualidade do ajustamento Gumbel*

Teste	Valor Observado da Estatística de Teste	Valor modificado da Estatística de Teste	Decisão $\alpha = 0.05$
Kolmogorov-Smirnov	$d_m = 0.03666204$	$\sqrt{m} \cdot d_m = 1.105955$	Rejeitar H_0
Cramér-von Mises	$w_m^2 = 0.24716236$	$w_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 0.248801$	Rejeitar H_0
Anderson-Darling	$a_m^2 = 1.88996293$	$a_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 1.902493$	Rejeitar H_0

Para estes três testes, a decisão foi a de rejeição da hipótese nula, ou seja, de que este modelo Gumbel ajustado não é adequado para modelar o comportamento destes níveis pluviométricos máximos mensais, quando comparada com qualquer outra distribuição em H_1 . Relembre-se que, para se chegar a essas decisões, foi necessário fazer a comparação do valor transformado, para cada uma das estatísticas de teste, com os quantis representados nas tabelas 3.1 e 3.2. Como o valor transformado é superior aos quantis críticos, rejeitou-se H_0 .

Portanto, os testes à qualidade do ajustamento Gumbel, com os parâmetros estimados, conduzem à rejeição deste modelo. Contudo, tem que se ter em conta que estes testes de ajustamento são baseados em estatísticas que medem a diferença entre as funções de distribuição teórica e empírica, motivo pelo qual, à medida que se aumenta a dimensão da amostra, a proporção das diferenças aumenta, conduzindo, mais frequentemente, à rejeição da hipótese nula. Por este motivo e devido ao facto dos testes aplicados anteriormente não rejeitarem $\gamma = 0$, poder-se-á considerar que a distribuição destes níveis pluviométricos é Gumbel.

IV) Avaliação da Não-Estacionariedade

Tal como foi feito com a amostra de máximos anuais, proceder-se-á nesta secção à avaliação de existência de tendência linear na localização do modelo. Assim, aplicando o método MV para obter as estimativas dos parâmetros $(\gamma, \lambda(t), \delta)$, com $\lambda(t) = \beta_0 + \beta_1 t$ e $t = 1, \dots, 910$, vem que,

$$(\hat{\gamma}, \hat{\lambda}(t), \hat{\delta}) = (0.02731, 21.99323 - 0.00127 \cdot t, 16.85588), \quad t = 1, \dots, 910.$$

As estimativas obtidas para os parâmetros de forma e escala são praticamente iguais, às obtidas para o modelo sem presença de tendência. Quanto ao ajustamento linear na localização, e que pode ser visualizado na figura 5.34, o declive da reta é negativo mas muito próximo de zero. Há, portanto, uma tendência, aparentemente insignificante, de diminuição do nível de precipitação médio mensal, em Barcelos.

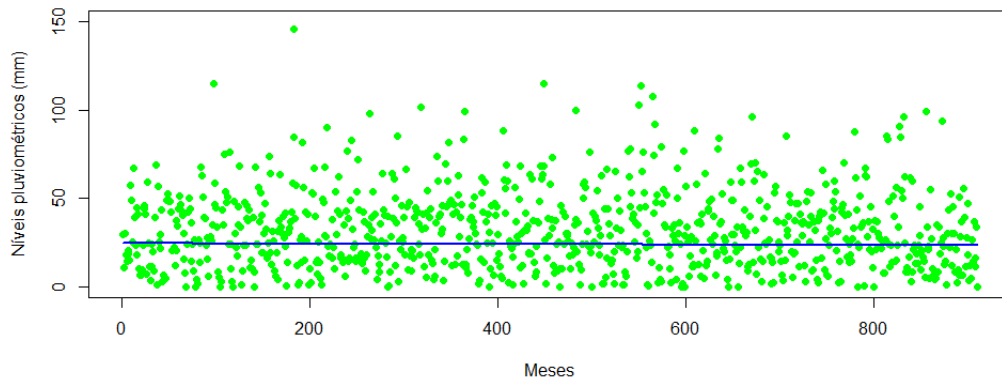


Figura 5.34: Ajustamento linear na localização, para os níveis pluviométricos máximos mensais

Para se averiguar a presença de tendência, será aplicado um TRV aos modelos GEV ajustados com tendência e sem tendência.

Tabela 5.20: Estimativas MV dos modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos mensais

Modelo	$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\delta}$	$\log L$
M_0 (sem tendência)	0.02743	21.40587	16.849996	-4022.886
M_1 (com tendência)	0.02731	21.99323-0.00127 · t	16.85588	-4022.686

Na tabela 5.20 relembram-se as estimativas dos parâmetros dos modelos com e sem tendência, os quais vão ter os nomes genéricos M_1 e M_0 , respetivamente, bem como dos valores da log-verosimilhança que permitem a realização do teste estatístico.

Para os modelos M_0 e M_1 foram estimados, respetivamente, três e quatro parâmetros. Logo, novamente, trata-se de um caso de fronteira e, como tal, a distribuição assintótica da estatística de teste é uma mistura 50:50 de dois Qui-quadrados com 3 e 4 g.l.. Assim, para as hipóteses,

$$H_0 : \text{Não há tendência na amostra} \quad \text{vs} \quad H_1 : \text{Há tendência na amostra,}$$

obtém-se,

d_obs=0.4

p-value=0.9613597

Como o p-value é superior aos níveis de significância usuais, não se rejeita a hipótese nula. Sendo assim, não há evidência estatística da presença de tendência na localização da distribuição subjacente aos níveis pluviométricos máximos mensais. Os critérios AIC e BIC, apresentados na tabela 5.21, também estão de acordo e revelam que o modelo GEV ajustado sem tendência na localização será o melhor modelo.

Tabela 5.21: *AIC e BIC para os modelos GEV com e sem tendência na localização, para os níveis pluviométricos máximos mensais*

Modelo	AIC	BIC
M_0 (sem tendência)	8051.772	8086.653
M_1 (com tendência)	8053.372	8099.88

5.2.1.3 Níveis Pluviométricos Máximos Anuais no outono e no inverno

I) Análise Preliminar dos Dados

Para realizar este estudo, consideraram-se duas amostras, uma para o outono e outra para o inverno, cada uma contendo os níveis pluviométricos diários máximos, por ano, no período de 01/01/1932 a 01/05/2008. Como já foi referido, e seguindo a sugestão de Miranda *et al.* (2006), para a presente análise, considerou-se que o outono abrange os meses de setembro, outubro e novembro e o inverno os meses de dezembro, janeiro e fevereiro. Assim, as amostras referentes ao outono e inverno têm uma dimensão de 75 e 76 observações, respetivamente. De referir que, inicialmente, procedeu-se à análise em valores extremos dos níveis de precipitação nas estações da primavera (março, abril e maio) e do verão (junho, julho e agosto). Todavia, tendo em conta o reduzido nível de pluviosidade registado nesses meses, ignoraram-se estas duas estações, dando-se destaque, apenas, às duas estações do ano com maiores índices de pluviosidade, na cidade de Barcelos.

Na figura 5.35, estão representados os gráficos ACF, para cada uma das amostras. A independência poderá ser assumida para as duas amostras, embora se registre uma ligeira correlação entre as observações correspondentes ao período de inverno.

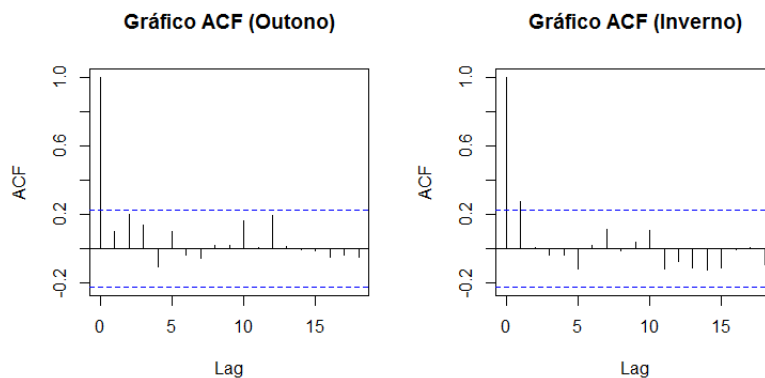


Figura 5.35: Gráficos de Autocorrelação Parcial

Para a amostra dos níveis pluviométricos no outono, o mínimo e o máximo para estes níveis são, respetivamente, 24.6 mm e 103 mm, com um nível mediano a rondar os 53 mm. Já na amostra referente ao inverno, os valores mínimo e máximo são 22.6 mm e 146 mm, respetivamente, com uma mediana nos 55.4 mm.

> OUTONO

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
24.60	39.35	53.00	56.18	67.60	103.00

> INVERNO

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
22.60	45.45	55.40	58.76	67.05	146.00

Os box-plots bem como os histogramas destas duas amostras podem ser visualizados na figura 5.36. Para a amostra referente ao outono, é de salientar a simetria na zona central, evidenciada pela estrutura da caixa, no box-plot. Contudo há uma assimetria à direita, a partir do terceiro quartil, embora não haja candidatos a outliers. Já para a amostra dos níveis pluviométricos, no inverno, a assimetria à direita é relativamente maior, apoiada pela presença de alguns candidatos a outliers. Em relação aos histogramas, a cauda da distribuição subjacente às observações do inverno é mais pesada que a referente ao outono.

Para fazer uma avaliação preliminar sobre o tipo de cauda direita das distribuições subjacentes a cada uma das amostras de níveis de precipitação, poder-se-á recorrer aos meios gráficos, QQ-plot exponencial e ME-plot, que estão representados na figura 5.37.

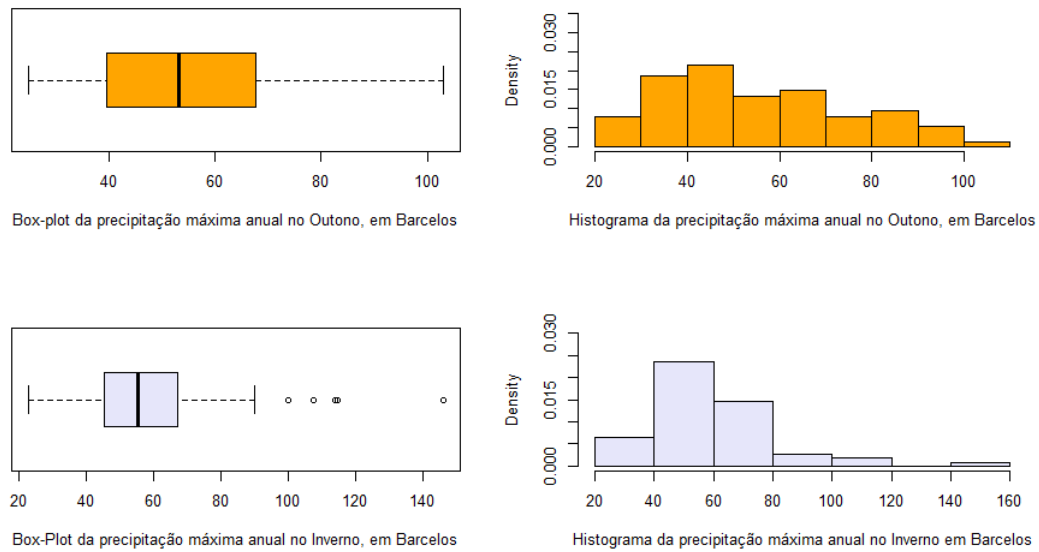


Figura 5.36: *Box-plots e histogramas dos níveis pluviométricos máximos anuais, no outono e inverno, em Barcelos (1932-2008)*

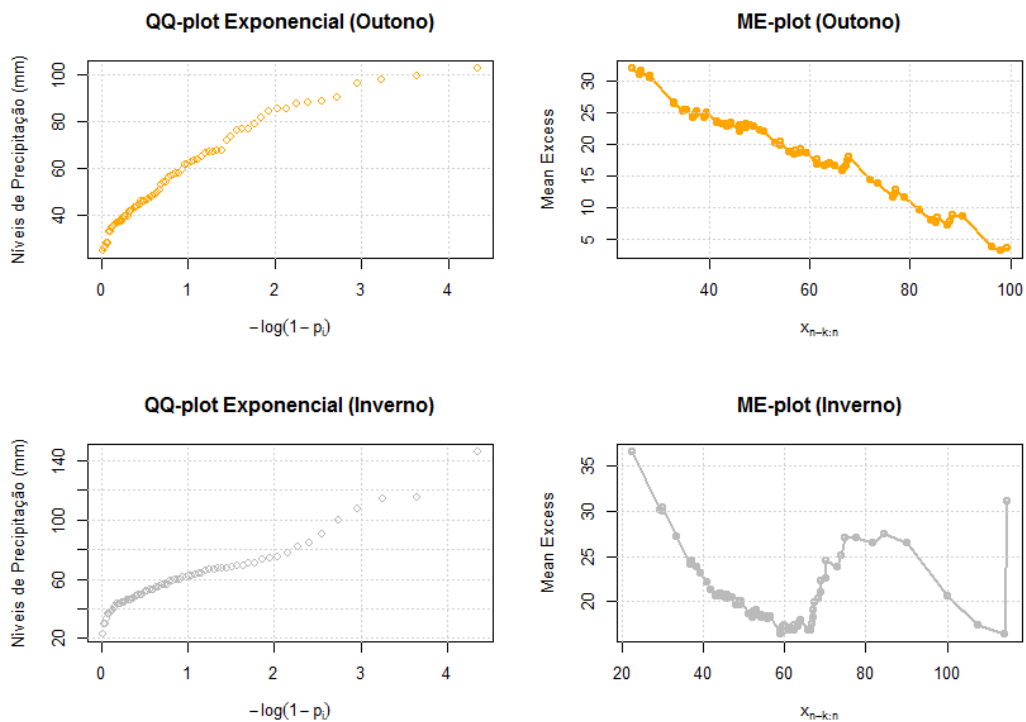


Figura 5.37: *QQ-Plot Exponencial e ME-plot dos níveis pluviométricos máximos anuais, no outono e inverno, em Barcelos (1932-2008)*

Uma vez que, para a amostra referente às observações no período de outono, o QQ-plot exponencial apresenta um padrão côncavo e o ME-plot é decrescente, poder-se-á deduzir que a cauda da distribuição subjacente a estas observações é mais leve que a da exponencial. Já para os níveis pluviométricos no inverno, as conclusões não são tão óbvias. O QQ-plot apresenta inicialmente um padrão ligeiramente côncavo, o que sugere uma cauda leve, mas que depois se torna convexo, o que nos conduz para o domínio das caudas pesadas. Também o ME-plot está de acordo com este padrão, pois começa por decrescer, para depois se tornar crescente (embora o número de observações nesta última situação seja mais reduzido).

Na tabela 5.22 encontram-se as estimativas preliminares dos parâmetros de forma, localização e escala, dos modelos Gumbel e GEV, para as duas amostras de níveis pluviométricos. Estas estimativas foram obtidas pelo processo já explicado anteriormente, com recurso aos qq-plots Gumbel e GEV (figura 5.38). Na mesma tabela indicam-se as correlações obtidas em cada um dos casos.

Tabela 5.22: *Estimativas preliminares dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos anuais, no outono e inverno*

Parâmetros	outono		inverno	
	Gumbel	GEV	Gumbel	GEV
$\hat{\gamma}$	-	-0.1099243	-	0.1943346
$\hat{\lambda}$	48.1022	47.43013	49.29429	48.58989
$\hat{\delta}$	20.41049	18.69038	17.02978	13.51197
Corr	0.9953211	0.9923048	0.9791181	0.9906954

Observando os qq-plots representados na figura 5.38, para a amostra referente aos níveis pluviométricos no outono, parece haver um melhor ajustamento dos pontos à reta, no caso Gumbel. Já para os níveis no inverno, parece haver uma melhor linearidade no qq-plot GEV. Tais situações podem, também, ser verificadas através da identificação das correlações mais elevadas, para o outono e inverno, representadas na tabela 5.22. É de salientar, também, que para os níveis pluviométricos máximos anuais no inverno, a estimativa para o parâmetro de forma é positiva. Assim, não será de descartar o modelo Fréchet para a distribuição das observações no inverno.

Sobrepondo as curvas das densidades estimadas a cada um dos histogramas, na figura 5.39, as conclusões não são muito diferentes ao que já foi referido. Na amostra do outono, as duas curvas apesar de apresentarem ligeiras diferenças, parecem estar bem ajustadas. Quanto à amostra do inverno, a curva da densidade GEV parece estar mais de acordo com as barras do histograma. Contudo, todas estas constatações foram obtidas através

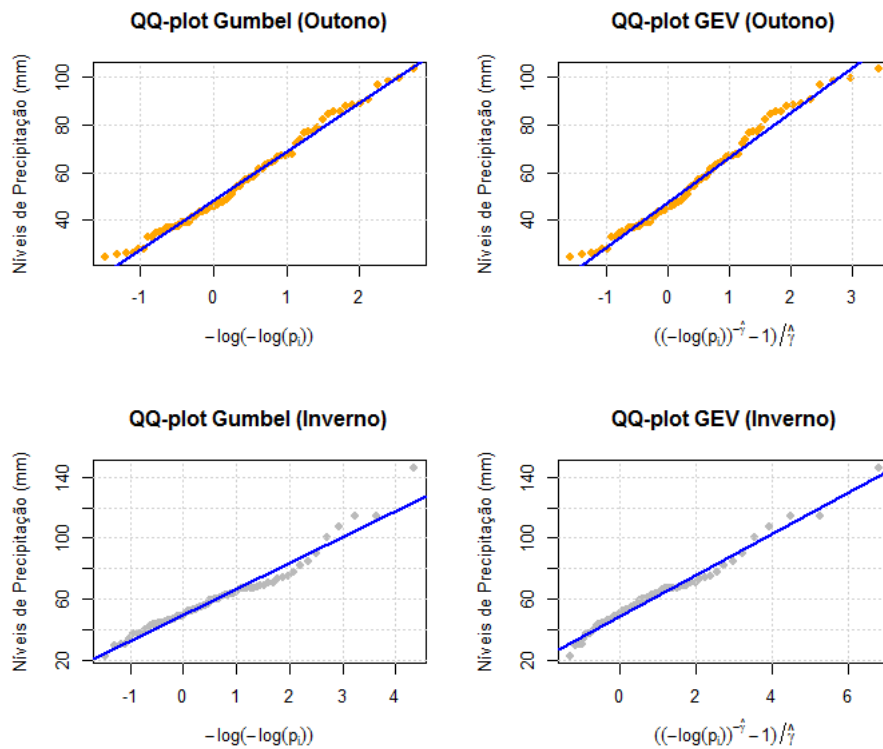


Figura 5.38: *QQ-Plots Gumbel e GEV dos níveis pluviométricos máximos anuais, no outono e inverno, em Barcelos (1932-2008)*

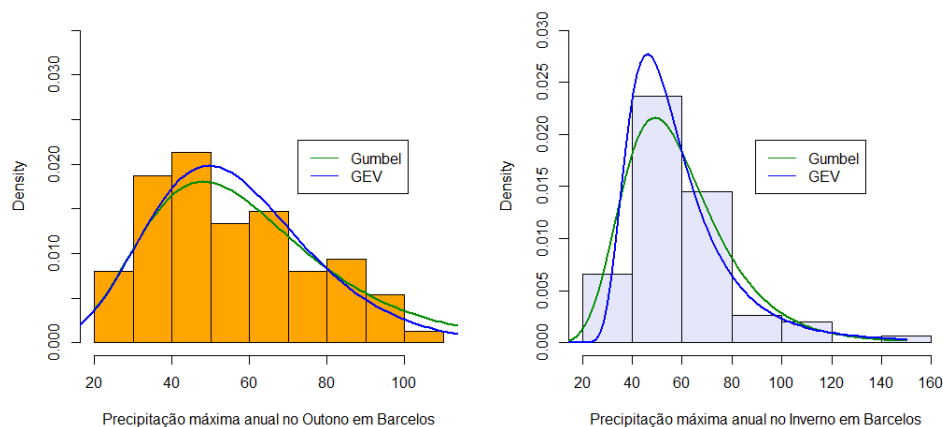


Figura 5.39: *Densidades dos modelos Gumbel e GEV estimados ajustados aos níveis pluviométricos máximos anuais, no outono e inverno*

de uma simples análise preliminar. Na secção seguinte serão apresentados resultados mais consistentes.

II) Estimação dos Parâmetros e Seleção do Modelo GEV

Seguindo as sugestões da análise preliminar e usando os métodos MV e MPP, obtiveram-se estimativas para os parâmetros das distribuições Gumbel (GEV com $\gamma = 0$) e GEV (com $\gamma \neq 0$), para os níveis pluviométricos máximos anuais nas estações do outono e inverno. Essas estimativas estão representadas na tabela 5.23, bem como, intervalos a 95% de confiança para os mesmos parâmetros, baseados na função de profile log-likelihood.

Tabela 5.23: *Estimativas dos parâmetros para os modelos Gumbel e GEV, para os níveis pluviométricos máximos anuais, no outono e inverno*

Estação do Ano	Parâmetros	Modelo Gumbel ($\gamma = 0$)		Modelo GEV ($\gamma \neq 0$)	
		MV	MPP	MV	MPP
outono	$\log L$	-328.6534	-	-328.5963	-
	$\hat{\gamma}$	-	-	-0.04158 (-0.2634389, 0.2125696)	-0.05564
	$\hat{\lambda}$	46.62328 (42.71486, 50.68408)	46.51381	46.99298 (42.64850, 51.7182)	46.95095
	$\hat{\delta}$	16.47903 (13.86418, 19.88309)	16.75102	16.74689 (13.75537, 20.7757)	17.58973
inverno	$\log L$	-327.6886	-	-327.6117	-
	$\hat{\gamma}$	-	-	0.02679 (-0.09255, 0.18449)	0.06348
	$\hat{\lambda}$	49.84256 (46.23258, 53.56803)	49.76896	49.61697 (45.90228, 53.50762)	49.3339
	$\hat{\delta}$	15.32556 (12.98514, 18.39799)	15.58447	15.21137 (12.83706, 18.32706)	14.64553

Ao nível da localização e escala, é notória a estabilidade nas estimativas obtidas pelos dois métodos, quer usando o modelo Gumbel, quer usando a distribuição GEV ($\gamma \neq 0$), para cada uma das estações do ano. Quanto às estimativas do parâmetro de forma, para o outono, obteve-se valores negativos, o que coloca a hipótese da distribuição destas observações ser Weibull. Já no caso do inverno, as estimativas deste parâmetro são ambas positivas, sendo a obtida pelo método MPP ligeiramente superior, e como tal, somos encaminhados para a hipótese do modelo Fréchet, embora as estimativas estejam muito próximas de zero. Repare-se, também, que os intervalos de confiança para este parâmetro, incluem o zero, logo o modelo Gumbel não é descartado em nenhuma das duas situações. Para o caso do outono, é de salientar o facto de como o intervalo está quase centrado no valor zero, o que já não acontece para o inverno, em que este se encontra mais deslocado para a direita.

Tendo em conta o que foi dito no parágrafo anterior, em relação à forma da distribuição, determine-se, em primeiro lugar, qual dos modelos extremos melhor se adequa a cada uma das distribuições subjacentes às amostras de outono e inverno. Comece-se por testar as

hipóteses

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma \neq 0, \quad (5.13)$$

para os níveis pluviométricos máximos anuais, no outono. Aplicando o conjunto de testes apresentados na secção 3.2.5, para as hipóteses definidas em cima, obteve-se os resultados apresentados na tabela 5.24.

Tabela 5.24: Resultados obtidos nos testes às hipóteses definidas em (5.13) (outono)

Teste	Valor Observado da Estatística de Teste	p-value	Decisão $\alpha = 0.05$
TRV	$d^* = 0.1100642$	0.740071	Não rejeitar H_0
Teste Score de RAO (versão Normal)	$v_m^* = -0.2176321$	0.8277158	Não rejeitar H_0
Teste Score de RAO (versão Qui-Quadrado)	$v_m^{2*} = 0.04736371$	0.8277158	Não rejeitar H_0
Teste LAN	$t_n^* = -0.2178095$	0.8275775	Não rejeitar H_0

A hipótese de $\gamma = 0$ não é rejeitada para qualquer nível de significância usual.

Já se se quiser especificar, na hipótese alternativa, o modelo Weibull, sugerido pela abordagem preliminar e pelas estimativas MV e MPP de γ , isto é, $H_1 : \gamma < 0$, poderá ser usado o teste baseado na estatística Gumbel. Assim, obteve-se

gs_m= 1.347305 gs*_m= -1.188653 p-value= 0.03752847.

Neste caso, ao nível de significância de 5%, rejeita-se a hipótese nula, aceitando-se assim que a distribuição subjacente a estes dados é Weibull. Contudo, já ao nível de 1% não se rejeita a hipótese de $\gamma = 0$.

Seguindo os passos aplicados nas outras análises efetuadas nesta dissertação, poderão, também, ser usadas as estatísticas de ajustamento de Kolmogorov-Smirnov, de Cramér-von Mises e de Anderson-Darling, para avaliar a qualidade do ajustamento Gumbel. Na tabela 5.25 encontra-se um resumo dos resultados obtidos, e onde se poderá constatar que não se rejeita a hipótese nula a qualquer nível de significância usual (valores observados inferiores aos quantis representados nas tabelas 3.1 e 3.2).

Portanto, como a maioria dos testes aplicados não rejeitam a hipótese de o parâmetro de forma ser identicamente nulo, a partir deste ponto considerar-se-á que a distribuição subjacente aos níveis de precipitação máximos anuais, no outono, em Barcelos, será Gumbel.

Tabela 5.25: Resultados obtidos nos testes à qualidade do ajustamento Gumbel (outono)

Teste	Valor Observado da Estatística de Teste	Valor modificado da Estatística de Teste	Decisão $\alpha = 0.05$
Kolmogorov-Smirnov	$d_m = 0.06269974$	$\sqrt{m} \cdot d_m = 0.5429957$	Não Rejeitar H_0
Cramér-von Mises	$w_m^2 = 0.06245377$	$w_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 0.06389608$	Não Rejeitar H_0
Anderson-Darling	$a_m^2 = 0.44204409$	$a_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 0.4522527$	Não Rejeitar H_0

Para o caso dos níveis pluviométricos máximos anuais, no inverno, e para as hipóteses definidas em (5.13) os resultados obtidos, para os diversos testes, estão reproduzidos na tabela 5.26

Tabela 5.26: Resultados obtidos nos testes às hipóteses definidas em (5.13) (inverno)

Teste	Valor Observado da Estatística de Teste	p-value	Decisão $\alpha = 0.05$
TRV	$d^* = 0.1481848$	0.7002761	Não rejeitar H_0
Teste Score de RAO (versão Normal)	$v_m^* = 0.4669295$	0.6405503	Não rejeitar H_0
Teste Score de RAO (versão Qui-Quadrado)	$v_m^{2*} = 0.2180232$	0.6405503	Não rejeitar H_0
Teste LAN	$t_n^* = 0.466514$	0.6408476	Não rejeitar H_0

Em nenhum dos testes a hipótese nula é rejeitada, o que quer dizer que não há evidência estatística para não se aceitar que estes níveis pluviométricos sejam modelados pela distribuição Gumbel.

A análise preliminar e a estimação MV e MPP sugeriram, para este caso, um parâmetro de forma positivo, logo poderá ser aplicado o teste específico a esta situação, bastando para isso substituir a hipótese alternativa de (5.13) por $H_1 : \gamma > 0$. Usando o teste baseado na estatística Gumbel, obtém-se,

gs_m= 1.448413 gs*_m= -1.048248 p-value= 0.9423087.

Ora o p-value é maior do que qualquer nível de significância usual, e como tal não se rejeita a hipótese de $\gamma = 0$.

Recorrendo aos testes à qualidade do ajustamento Gumbel, cujos resultados se apresentam na tabela 5.27, todos indicam que este modelo é adequado para modelar os dados em estudo.

Tabela 5.27: Resultados obtidos nos testes à qualidade do ajustamento Gumbel (inverno)

Teste	Valor Observado da Estatística de Teste	Valor modificado da Estatística de Teste	Decisão $\alpha = 0.05$
Kolmogorov-Smirnov	$d_m = 0.07355273$	$\sqrt{m} \cdot d_m = 0.6412178$	Não Rejeitar H_0
Cramér-von Mises	$w_m^2 = 0.04367059$	$w_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 0.04467246$	Não Rejeitar H_0
Anderson-Darling	$a_m^2 = 0.34376686$	$a_m^2 \cdot \left(1 + \frac{0.2}{\sqrt{m}}\right) = 0.3516534$	Não Rejeitar H_0

Assim, os testes aplicados indicam, na sua totalidade, que a distribuição dos níveis pluviométricos máximos anuais, no inverno, é Gumbel. Logo, com vista à obtenção das estimativas de quantidades importantes em modelação de acontecimentos raros, considerar-se-á, a partir daqui, que os modelos ajustados para cada uma das amostras é Gumbel. Apenas se fará uso das estimativas obtidas pelo método MV, e que podem ser revistas na tabela 5.28.

Tabela 5.28: Estimativas MV dos parâmetros para o modelo Gumbel, para os níveis pluviométricos máximos anuais, no outono e inverno

Estação	Estimativas dos Parâmetros	
	$\hat{\lambda}$	$\hat{\delta}$
Outono	46.62328 (42.71486,50.68408)	16.47903 (13.86418,19.88309)
Inverno	49.84256 (46.23258,53.56803)	15.32556 (12.98514,18.39799)

Nas figuras 5.40 e 5.41 são apresentados gráficos que permitem uma avaliação rápida da qualidade do ajustamento dos dois modelos mencionados. O ajustamento feito às observações no outono e no inverno, parece ser bastante aceitável. Apenas os qq-plots apresentam, essencialmente na parte final, um ligeiro afastamento dos pontos em relação à reta ajustada à nuvem de pontos.

Usando os modelos ajustados para cada uma das amostras, poderão ser estimados os quantis extremos, os níveis e períodos de retorno e a probabilidade de excedência de um nível elevado. Saliente-se que os resultados obtidos nesta secção serão sempre valores condicionados à estação do ano, neste caso ao outono e ao inverno. Todavia, tem-se a indicação de que as distribuições subjacentes aos níveis pluviométricos máximos anuais e aos níveis pluviométricos máximos anuais no outono e inverno pertencem à mesma família de distribuições, neste caso, a Gumbel.

Na tabela 5.29 representam-se estimativas pontuais dos quantis extremos, através do método MV, e os seus intervalos de 95% de confiança, baseados na função de profile log-likelihood, para as duas estações do ano referidas.

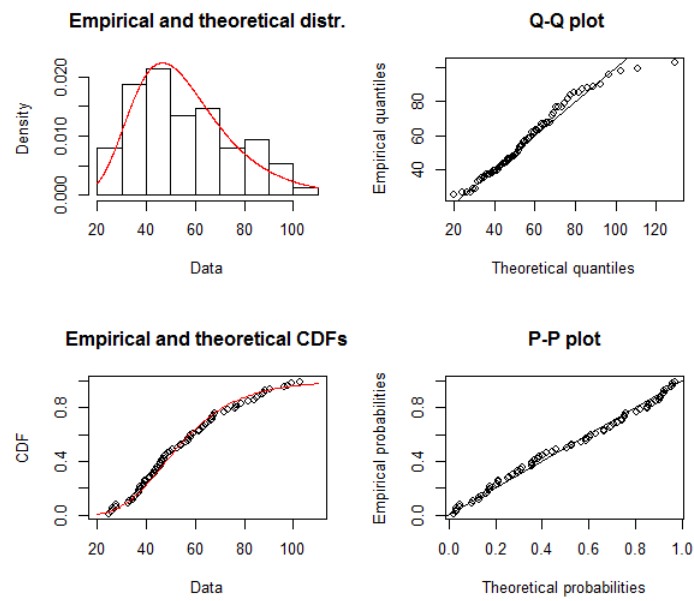


Figura 5.40: *Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos anuais, no outono*

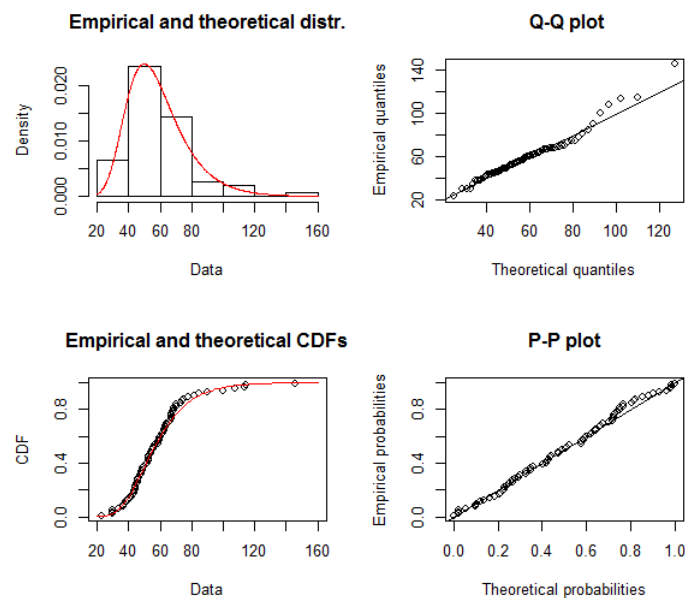


Figura 5.41: *Diagnóstico gráfico do ajustamento Gumbel aos níveis pluviométricos máximos anuais, no inverno*

A semelhança entre os quantis estimados, para as duas estações anuais, é notória, embora, os valores referentes ao outono sejam ligeiramente superiores. Assim, por exemplo,

Tabela 5.29: *Estimativas pontuais MV e IC a 95% dos quantis extremos, para os níveis pluviométricos máximos anuais no outono e no inverno*

Quantis	Outono	Inverno
$\hat{q}_{0.95}$	95.56922 (85.94,107.65)	95.36247 (86.81,106.25)
$\hat{q}_{0.99}$	122.42928 (108.74,139.86)	120.34242 (108.1,136)
$\hat{q}_{0.999}$	160.44814 (140.9,185.5)	155.70011 (138.1,178.22)
$\hat{q}_{0.9999}$	198.39993 (172.98,231.2)	190.99542 (168.18,220.7)
$\hat{q}_{0.99999}$	236.34504 (204.79,277.07)	226.28451 (198.1,262.9)

espera-se que, em média, um em cada 100 outonos, seja excedido o nível 122.43 mm. No caso do inverno esse nível de retorno é de 120.34 mm. Repare-se que estes valores são correspondentes aos quantis de probabilidade 0.99. Relembre-se que níveis de retorno T -anos são quantis de probabilidade $1 - \frac{1}{T}$. Assim, se se pretender usar em específico esse conceito e, para 100, 400 e 1000 anos, as estimativas obtidas para os níveis de retorno no outono e no inverno, poderão ser observadas na tabela 5.30.

Tabela 5.30: *Estimativas pontuais MV e IC a 95% para os níveis de retorno, no outono e inverno*

Estação do Ano	$\hat{U}(100)$	$\hat{U}(400)$	$\hat{U}(1000)$
Outono	122.4293 (108.74,139.86)	145.3362 (128.15,167.27)	160.4481 (140.9,185.5)
Inverno	120.3424 (108.1,136)	141.6459 (126.19,161.4)	155.7001 (138.1,178.22)

Logo espera-se que, em média, uma vez em cada 100, 400 ou 1000 anos, no outono, os níveis de precipitação em Barcelos ultrapassem os valores 122.4293 mm, 145.3362 mm ou 160.4481 mm, respetivamente. Em relação ao inverno, essas quantidades não são muito diferentes. Espera-se que, em média, uma vez em cada 100, 400 e 1000 anos, sejam excedidos os níveis de 120.3424 mm, 141.6459 mm e 155.7001 mm de pluviosidade, respetivamente, durante os meses de dezembro, janeiro ou fevereiro.

Na tabela 5.31 encontram-se as estimativas dos períodos de retorno do nível pluviométrico de 146 mm, no outono e no inverno, bem como os valores obtidos para a probabilidade

de excedência desse nível, para cada uma das duas estações do ano.

Tabela 5.31: *Estimativas MV do período de retorno e probabilidade de excedência do nível 146 mm, para os níveis pluviométricos máximos anuais, no outono e no inverno*

Estação do Ano	$\hat{T}(146)$	$\hat{P}(Y > 146)$
Outono	416.4214	0.002401
Inverno	531.2645	0.001882

Os resultados obtidos estão de acordo com o que já tinha sido constatado com as estimativas dos quantis extremos, para cada uma das épocas do ano. De facto, usando os modelos ajustados para cada uma das situações, o nível de pluviosidade 146 mm, em média será ultrapassado ao fim de 416.42 anos, se se considerar apenas o período de outono, ou ao fim de 531.26 anos se esse período for antes o inverno. A probabilidade disso acontecer, para cada um destes períodos anuais, é, respetivamente, 0.0024 e 0.00188. Não deixa de ser curioso o facto do modelo ajustado ao outono considerar que níveis mais elevados são menos raros que no inverno, quando na amostra referente à última estação do ano existem valores que não são alcançados na amostra do outono. Contudo, tal situação não é assim tão estranha, recorde-se, por exemplo, os histogramas e os box-plots das duas amostras na figura 5.36. Com efeito, existem níveis pluviométricos mais elevados no inverno, bastará olhar para a amplitude entre o terceiro quartil e o máximo, que para o outono é de 35.4 mm e para o inverno é de 78.95 mm. Os valores obtidos para o terceiro quartil em cada uma das amostras são muito semelhantes, 67.6 mm no outono e 67.05 mm no inverno, e é fácil perceber, embora as classes para cada um dos histogramas tenham limites diferentes, que no inverno a maior parte das observações após o terceiro quartil, se situa abaixo dos 80 mm, daí a queda abrupta das barras do histograma, após este nível. O mesmo já não acontece para o outono, em que as barras oscilam ligeiramente entre si, não havendo descidas tão acentuadas, ou seja, há uma distribuição mais equilibrada das observações. Logo os ajustamentos realizados irão ser influenciados por esta situação. Na figura 5.42 estão representadas as séries temporais dos níveis pluviométricos máximos anuais para as duas estações do ano e um zoom dos níveis acima dos 60 mm, onde se pode observar o que foi referido anteriormente.

Uma vez que se obteve um EVI nulo, para as duas situações, não se poderá aplicar o estimador apresentado em (3.16), para o limite superior do suporte.

Foi analisada a presença de tendência em cada uma das amostras, ao nível da localização, tendo esta se revelado estatisticamente não significativa (p-values iguais a 0.2515031 e 0.4077097, para o outono e inverno, respetivamente). Na figura 5.43 estão representadas

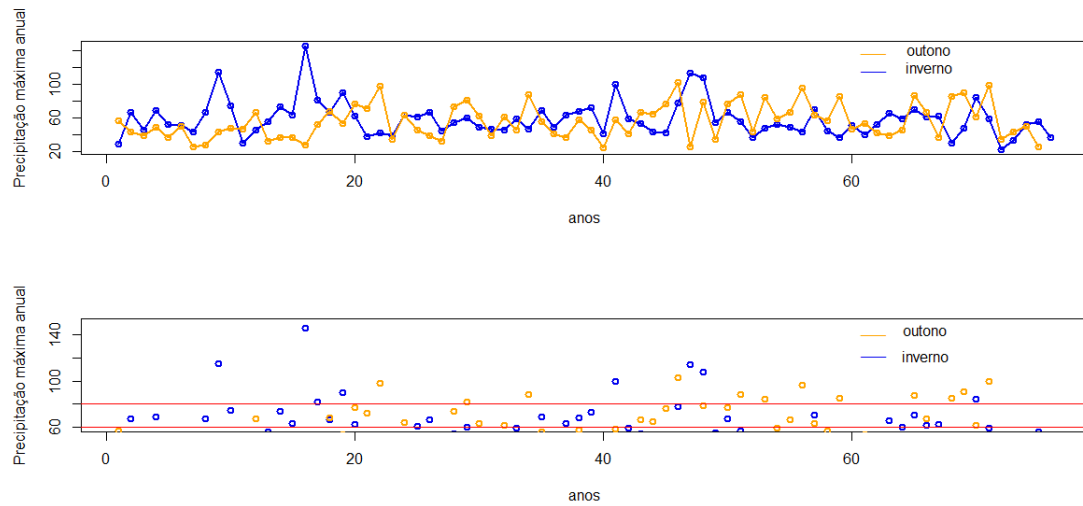


Figura 5.42: *Figura superior: Série temporal dos níveis pluviométricos máximos anuais no outono e no inverno; figura inferior: zoom dos níveis acima de 60 mm, com demarcação da zona entre 60 mm e 80 mm*

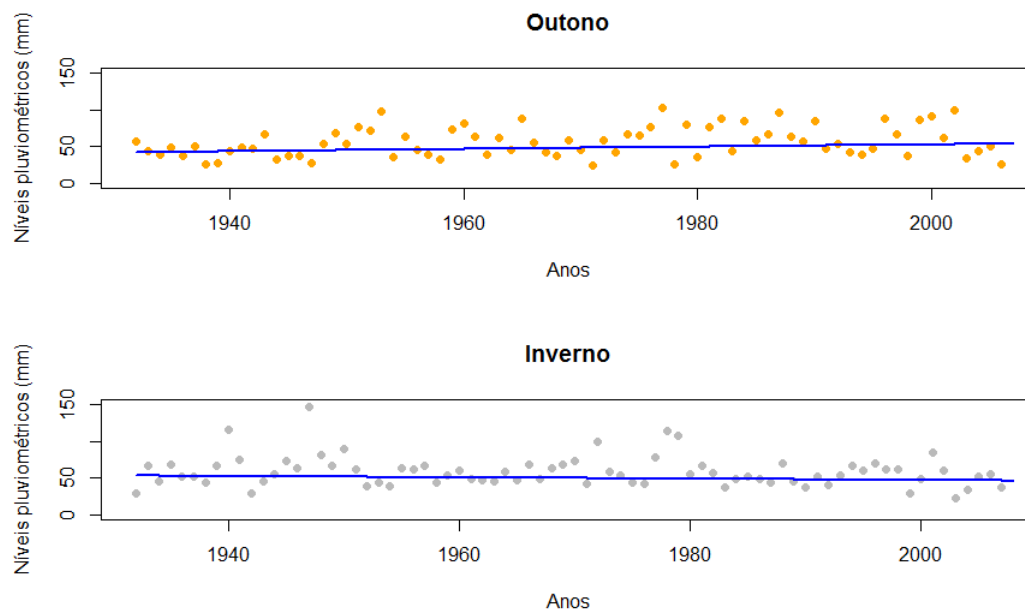


Figura 5.43: *Ajustamento linear na localização, para os níveis pluviométricos máximos anuais no outono (figura superior) e no inverno (figura inferior)*

as retas ajustadas à localização, para as duas estações do ano. Como se pode observar, o declive da reta referente ao outono é positivo, o que sugere que o nível médio máximo anual de precipitação, neste período, tem aumentado ligeiramente. Já para o inverno a reta é decrescente e, como tal, nesta estação do ano, o nível pluviométrico médio máximo

anual tem diminuído ao longo dos anos. Contudo este crescimento e diminuição, não se revelaram significativos, como já indicado.

5.2.2 Método das Maiores Observações - Níveis Pluviométricos Anuais

Na secção 3.4, foi apresentado um outro método para fazer inferência estatística para eventos raros, que consiste em modelar as maiores observações de uma amostra com uma estrutura conjunta (resultado (3.65)). A amostra original, relembre-se era constituída por registos diários de níveis pluviométricos, em Barcelos, desde 1932 a 2008. Assim, com vista a aplicar o Método das Maiores Observações, para cada bloco de um ano, extraíram-se os $k = 5$ e $k = 10$ maiores níveis de precipitação, obtendo-se um conjunto de 77 vetores aleatórios 5-dimensionais e 10-dimensionais. Essas observações encontram-se representadas nas figuras 5.44 e 5.45.

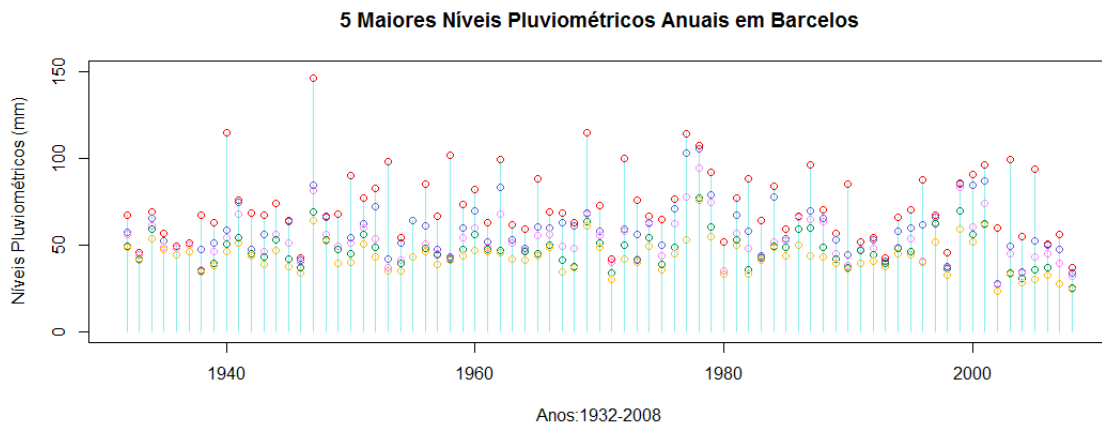


Figura 5.44: Os 5 maiores níveis pluviométricos diários, por ano, em Barcelos

Nesta secção será feita apenas uma rápida análise a estas duas situações e, em paralelo, serão re-apresentados os valores que se obtiveram aquando da aplicação do método Gumbel aos máximos anuais.

Utilizando o método MV, cujas expressões da verosimilhança e da log-verosimilhança se encontram na secção 3.4, foram obtidas estimativas para os parâmetros de forma, localização e escala para o modelo GEV. Esses valores são apresentados na tabela 5.32. Para além das estimativas aí apresentadas, também são indicados, entre parênteses, os respetivos erros padrão. Ora é evidente que há medida que aumentamos o número de observações retidas, existe um decréscimo dos erros padrão das estimativas. Os valores

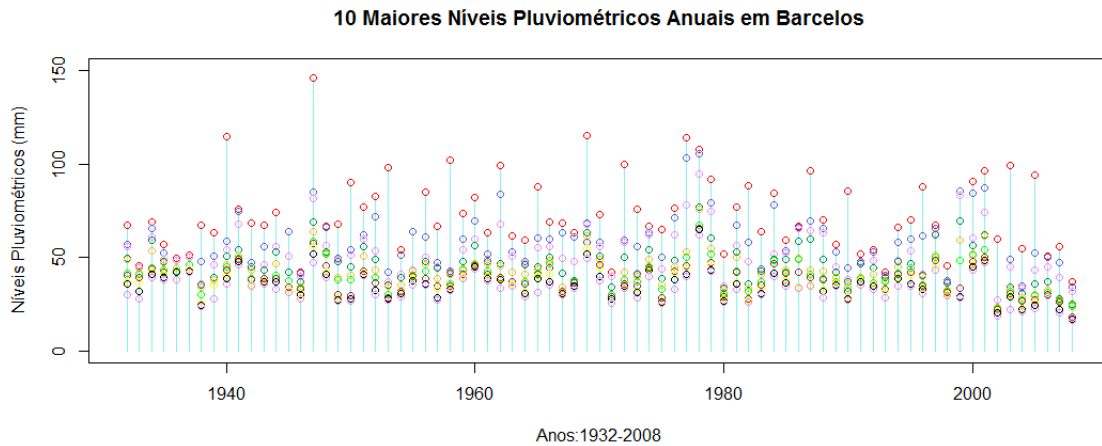


Figura 5.45: Os 10 maiores níveis pluviométricos diários, por ano, em Barcelos

Tabela 5.32: Estimativas MV dos parâmetros para o modelo GEV e respectivos erros padrão, quando considerados os $k=1,5,10$ maiores níveis pluviométricos anuais

Caso	$\log L$	Estimativas dos Parâmetros		
		$\hat{\gamma}$	$\hat{\lambda}$	$\hat{\delta}$
$k = 1$	-335.6562	-0.04967 (0.07946)	64.41883 (2.11780)	16.6027 (1.51555)
$k = 5$	-1166.361	-0.02546 (0.042001)	66.49658 (1.47767)	15.64649 (0.91135)
$k = 10$	-1799.979	-0.02918 (0.02908)	67.02619 (1.33039)	14.81987 (0.80455)

obtidos para as estimativas dos parâmetros não são muito diferentes para cada um dos três casos, contudo nota-se uma maior estabilidade para os casos referentes às 5 e 10 maiores observações. Parece, assim, existir um melhor ajustamento do modelo quando se tomam mais observações por ano, para além do máximo anual.

Para todas as situações o valor obtido para o EVI é negativo, o que significa que a distribuição subjacente a estes níveis pluviométricos poderá ser Weibull. Nesse caso, a distribuição terá uma cauda leve e com limite superior do suporte finito. No entanto, mais uma vez, as estimativas do EVI estão bastante próximas de zero e, por isso, a hipótese da distribuição Gumbel não pode ser posta de lado.

A qualidade do ajustamento para os níveis pluviométricos máximos anuais, através das 5 e das 10 maiores observações anuais, pode ser investigado através dos plots diagnóstico representados nas figuras 5.46 e 5.47. Como é visível, as diferenças não são grandes, contudo, parece haver um melhor ajustamento para o nível pluviométrico máximo anual

quando se retêm as 5 maiores observações em cada ano. Esta conclusão é baseada meramente na análise visual dos PP-plot e QQ-plot em que, na figura 5.46, as nuvens de pontos estão ligeiramente mais próximas da reta. Também no gráfico referente aos níveis de retorno, na figura referida, existem mais pontos entre as bandas de confiança para este parâmetro, do que na figura 5.47. Já se a comparação for feita com os gráficos obtidos no diagnóstico ao ajustamento do máximo anual, baseado apenas na maior observação, e que estão representados na figura 5.17, não existem grandes diferenças entre os PP-plots e os QQ-plots obtidos para esse caso quando comparados com os obtidos, retendo as 5 maiores observações. Note-se que no PP-plot, quando se considera apenas o máximo anual, os eixos das abcissas e das ordenadas são referentes aos valores teóricos e empíricos, respectivamente. Já para o caso das 5 MO's retidas, no PP-plot apresentado, o eixo das abcissas corresponde aos valores empíricos e o das ordenadas aos valores teóricos.

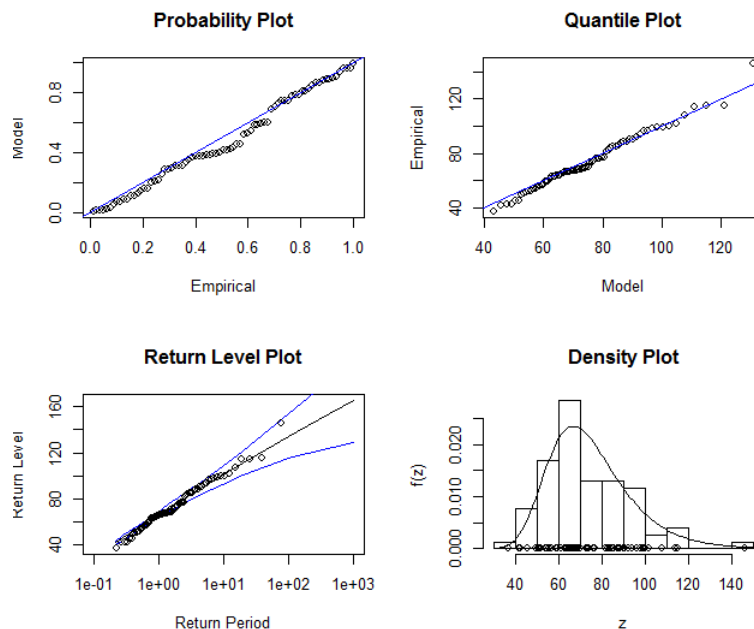


Figura 5.46: *Diagnóstico gráfico do ajustamento GEV aos níveis pluviométricos máximos anuais, baseado nas 5 MO's*

Mais uma vez, a tendência linear na localização não se revelou estatisticamente significativa, obtendo-se p-values de 0.286, para o caso referente às 5 MO's, e de 0.189, para o caso das 10 MO's. Na figura 5.48 estão representadas as retas ajustadas à localização do modelo para o máximo anual, baseado nas 5 e 10 maiores observações, e é notória a quase horizontalidade por parte destas.

Usando as estimativas obtidas para os parâmetros de forma, localização e escala, representados na tabela 5.32, e os estimadores introduzidos na secção 3.2.3, poderão ser

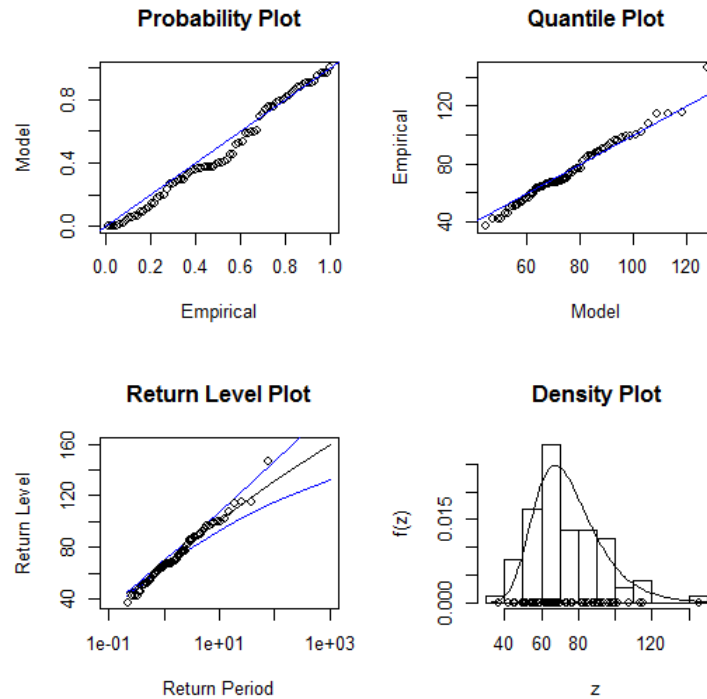


Figura 5.47: Diagnóstico gráfico do ajustamento *GEV* aos níveis pluviométricos máximos anuais, baseado nas 10 *MO*'s

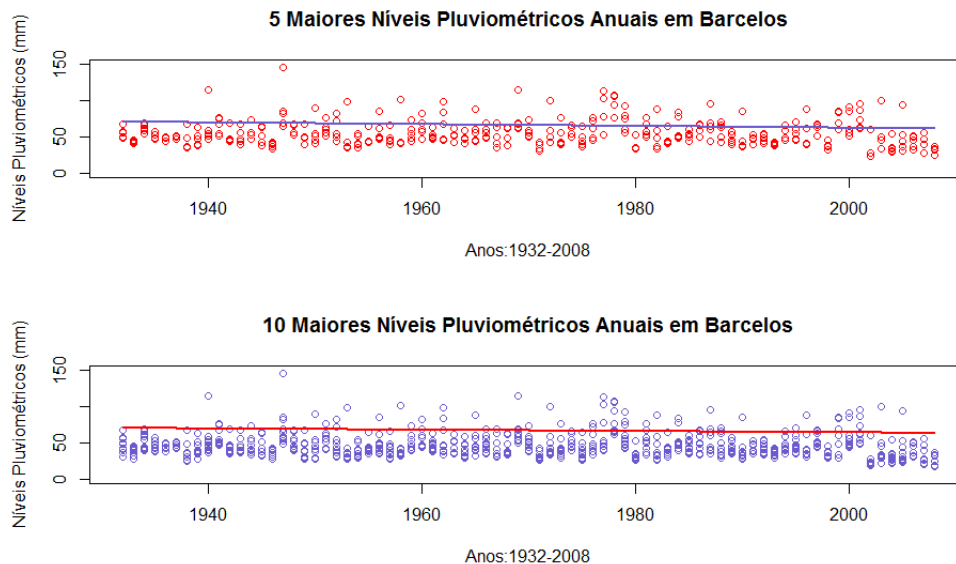


Figura 5.48: Ajustamento linear na localização, para os níveis pluviométricos máximos anuais, baseado nas 5 (figura superior) e 10 (figura inferior) *MO*'s

obtidas quantidades importantes em modelação de acontecimentos raros, e que já foram referidas, por diversas vezes, nesta dissertação. Na tabela 5.33 encontram-se as estimativa

Tabela 5.33: *Estimativas MV de outros parâmetros para o modelo GEV, quando considerados os $k=1,5,10$ maiores níveis pluviométricos anuais*

Parâmetros	Número de MO's retidas		
	$k = 1$	$k = 5$	$k = 10$
$\hat{q}_{0.95}$	110.2670	111.2557	109.1905
$\hat{q}_{0.99}$	132.6966	134.4170	130.8222
$\hat{q}_{0.999}$	161.4951	165.5998	159.7342
$\hat{q}_{0.9999}$	187.1327	194.9531	186.7177
$\hat{q}_{0.99999}$	209.9953	222.6297	211.9431
$\hat{U}(100)$	132.6966	134.4170	130.8222
$\hat{U}(400)$	150.4416	153.4240	148.4746
$\hat{U}(1000)$	161.4951	165.5998	159.7342
$\hat{T}(146)$	280.0469	231.2638	328.14
$\hat{P}(Y > 146)$	0.00357	0.00432	0.00305
\hat{x}^F	398.6808	680.9387	574.8945

MV para os quantis extremais de probabilidades 0.95, 0.99, 0.999, 0.9999 e 0.99999, para os níveis de retorno de 100, 400 e 1000 anos, para o período de retorno do nível pluviométrico máximo observado de 146 mm, para a probabilidade de excedência desse mesmo nível e para o limite superior do suporte. São apresentados os valores estimados considerando os modelos GEV ajustados, baseados na retenção das 5 e 10 MO's anuais, bem como os obtidos quando apenas se considerou a maior observação anual.

Por observação da referida tabela, verifica-se que existe uma grande semelhança entre os valores obtidos para os quantis e níveis de retorno, para os três casos. Contudo, para o caso $k = 5$, o modelo ajustado considera níveis pluviométricos elevados menos raros que nos outros dois casos. De facto, os valores estimados para os quantis, níveis de retorno, probabilidade de excedência e limite superior do suporte são sempre os mais elevados e o período de retorno do nível 146 mm é o mais baixo que nas outras duas situações. Em relação ao modelo ajustado para o caso $k = 10$, aparentemente, dá a ideia de ser o modelo que mais considera invulgar a ocorrência de níveis elevados de precipitação. Por exemplo, observando a estimativa do período de retorno do nível 146 mm, espera-se que, em média, uma vez em cada 328 anos este nível seja excedido, descendo esse período para 280 anos e 231 anos para os casos $k = 1$ e $k = 5$, respetivamente. Todavia, quando se observa os valores obtidos para o quantil de probabilidade 0.99999, este já é relativamente maior que o obtido, no caso de apenas se considerar uma observação máxima por ano, 211.9431 mm contra 209.9953 mm. Pelas estimativas obtidas para o limite superior do suporte,

confirma-se que, afinal, é no caso $k = 1$ que se atingirá valores menos elevados, onde há a indicação de que o nível pluviométrico máximo anual em Barcelos nunca ultrapassará o nível de 398.6808 mm. De facto, este valor é bastante inferior aos 680.9387 mm e 574.8945 mm obtidos através dos modelos ajustados ao máximo anual, baseados nas 5 e 10 maiores observações anuais, respetivamente.

5.2.3 Método POT

5.2.3.1 Níveis Pluviométricos Máximos Anuais

I) Escolha do *Threshold* e Análise Preliminar dos Dados

Pretende-se no presente ponto desta dissertação aplicar a metodologia POT, apresentada na secção 3.3, à amostra de níveis pluviométricos diários máximos anuais, em Barcelos, registados no período de 01/01/1932 a 01/05/2008. Relembre-se, muito resumidamente, que nesta metodologia começa-se por fixar um nível elevado, u , o chamado *threshold*, e com as m excedências desse nível constroem-se os excessos, $(X - u)|X > u$ ou, equivalentemente, se $Y = X - u, Y|Y > 0$. De seguida, pelo teorema 2.17, poder-se-á proceder ao ajustamento de um modelo GP para modelar estes excessos.

Posto isto, o primeiro passo consiste na escolha do *threshold*. Na secção 3.3.1 foram apresentados alguns métodos para efetuar essa escolha. Um deles consistia no estudo da função de excesso médio apresentada em (3.37).

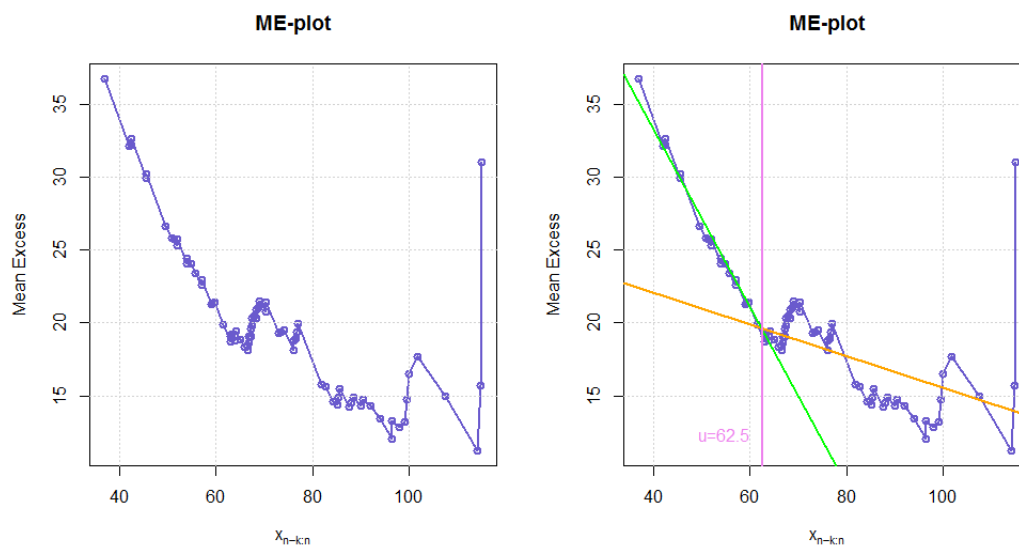


Figura 5.49: *ME-plot*, sem e com retas ajustadas, para os níveis pluviométricos máximos anuais

Na figura 5.49 é novamente apresentado o ME-plot para os níveis pluviométricos máximos anuais. Este apresenta um padrão decrescente, como já tinha sido referido. Contudo, esse decrescimento não acontece sempre ao mesmo ritmo, começando por ser bastante acentuado até, sensivelmente, ao nível pluviométrico de 62.5 mm. A partir desse nível o comportamento do ME-plot é um pouco mais irregular com momentos de crescimento e outros de decrescimento, mas a nuvem de pontos parece rondar uma reta de declive negativo. Este comportamento poderá ser visualizado na figura 5.49, na imagem da direita, com o ajustamento de duas retas aos pontos que formam o ME-plot. Ambas as retas têm declive negativo, mas diferente. Sendo assim, o valor escolhido para o *threshold* será o nível de 62.5 mm.

Na secção 3.3.1 é sugerido um outro processo para a escolha do *threshold*, baseado na estabilidade dos estimadores dos parâmetros de forma e de escala do modelo GP. Um *threshold* adequado será aquele a partir do qual se verifica essa estabilidade. Na figura 5.50 estão representados os gráficos com as estimativas, pontuais e intervalares, dos parâmetros de forma e escala segundo o *threshold* escolhido. Ora, a partir do valor 62.5, parece haver uma forte estabilidade para as estimativas e, como tal, a escolha desse nível para *threshold* revela-se, também aqui, adequada. Na parte final dos gráficos há alguma irregularidade, o que é habitual, pois a escolha de um *threshold* muito elevado implica que os estimadores tenham uma maior variância.

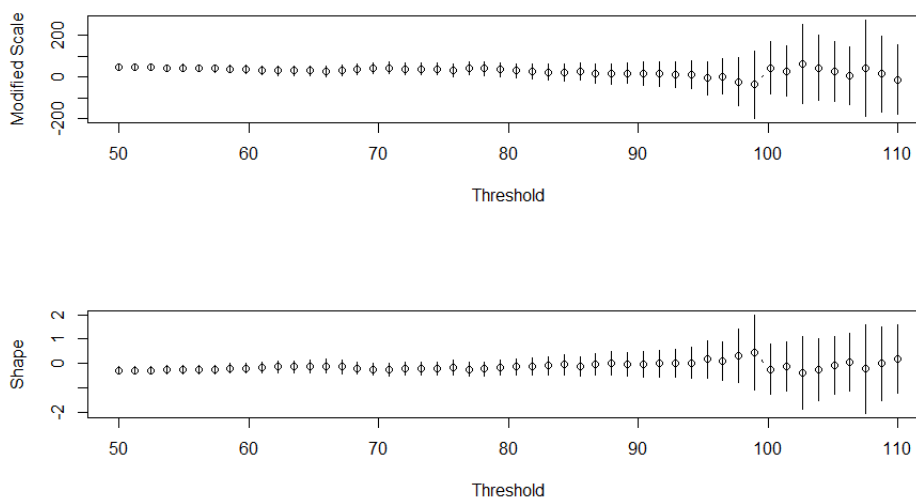


Figura 5.50: *Estimativas dos parâmetros de forma e escala versus threshold*

Considerando, portanto, o nível 62.5, para a amostra referente aos níveis pluviométricos máximos anuais, obtêm-se 56 excedências em 77 valores, o que corresponde a cerca

de 73% da amostra. Estas excedências poderão ser visualizadas na figura 5.51. Tal como

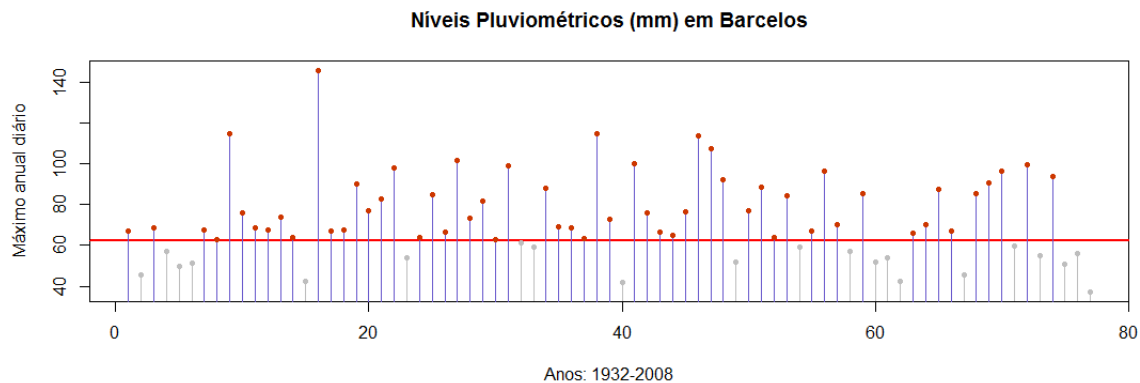


Figura 5.51: *Excedências de $u=62.5$, na amostra dos níveis pluviométricos máximos anuais*

foi feito para a metodologia do Máximo Anual, também aqui poderá ser efetuada uma análise preliminar, essencialmente gráfica, dos excessos obtidos com o *threshold* escolhido. Com esta análise ter-se-á uma ideia de como se comporta a cauda direita da distribuição subjacente aos dados e, também, poderão ser obtidas estimativas preliminares dos parâmetros, que por sua vez poderão ser usadas em inicializações de processos numéricos para obter as estimativas MV ou MPP. Assim, como anteriormente, poder-se-á começar por construir um QQ-plot exponencial para esta amostra de excessos, uma vez que a distribuição exponencial é, precisamente, o caso particular do modelo GP, quando o parâmetro de forma é nulo, ou seja, $\gamma = 0$. Se a nuvem de pontos estiver muito próxima da reta ajustada, será razoável admitir que os dados são do tipo exponencial na cauda ou provêm de uma distribuição pertencente ao domínio de atração Gumbel. O qq-plot encontra-se na figura 5.52.

A nuvem de pontos parece ajustar-se à reta, ocorrendo, apenas, uns desvios maiores nos últimos pontos. A correlação obtida é de 0.9872271, o que é bastante elevada, vindo reforçar a ideia de associação linear entre os pontos. Logo, estes excessos parecem ser bem modelados pelo modelo exponencial.

No entanto, vamos considerar a situação $\gamma \neq 0$ e construir o QQ-plot GP, com o intuito de verificar se existe uma maior linearidade dos pontos. A função quantil do modelo GP, para $\gamma \neq 0$, é dada por

$$Q_{\gamma,\sigma}(p) = \sigma \cdot \frac{(1-p)^{-\gamma} - 1}{\gamma} = \sigma \cdot Q_{\gamma,1}(p), \quad p \in (0, 1), \sigma > 0.$$

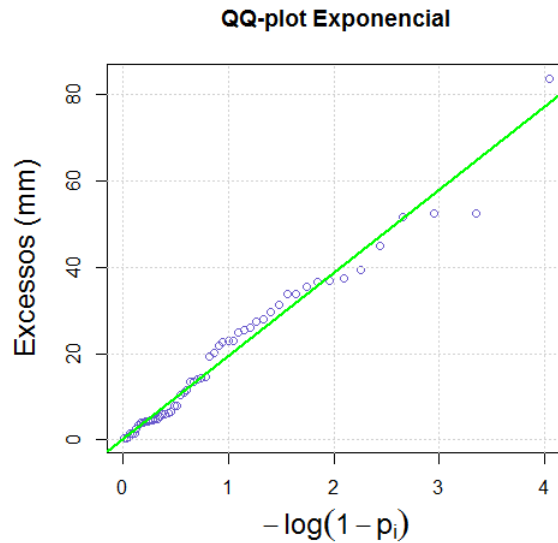


Figura 5.52: *QQ-plot Exponencial dos excessos, referentes à amostra de níveis pluviométricos máximos anuais*

Através dos procedimentos habituais, o QQ-plot GP pode ser dado por,

$$\left(\frac{(1-p_i)^{-\gamma} - 1}{\gamma}, y_{i:m} \right), \quad p_i = \frac{i}{m+1}, \quad i = 1, \dots, m. \quad (5.14)$$

Como se pode verificar, a relação linear tem ordenada na origem nula e declive dado pelo parâmetro de escala, σ . Contudo, antes de construir este qq-plot, é necessário, primeiramente, fixar o valor do parâmetro de forma, γ . Mais uma vez, seguindo a sugestão de Beirlant *et al.* (2004), uma estimativa preliminar para γ será o valor que maximiza a correlação entre os quantis $\hat{Q}_{\gamma,\sigma}(p) = y_{i:m}$ e $Q_{\gamma,1}(p)$. O gráfico com a representação dessa função correlação encontra-se na figura 5.53, onde se poderá constatar que,

```
> gamma_hat
[1] -0.08436
> correl_max
[1] 0.9882729
```

ou seja, a correlação máxima de 0.9882729 é atingida para um $\hat{\gamma}$ igual a -0.08436. Usando esse valor para o EVI, obtém-se o QQ-plot GP para este conjunto de excessos e que se encontra representado na figura 5.54.

O qq-plot revela uma linearidade, por parte da nuvem de pontos, bastante satisfatória e muito semelhante à obtida no qq-plot exponencial. Se se ajustar uma reta dos mínimos

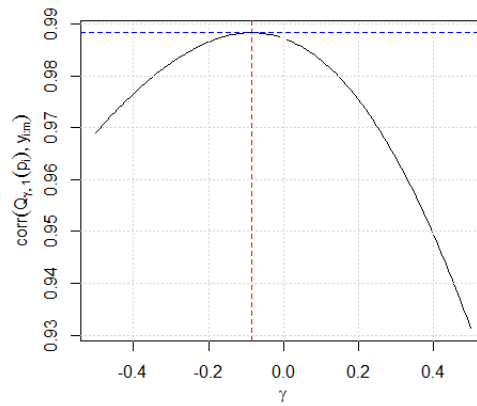


Figura 5.53: Gráfico de correlação entre $\hat{Q}_{\gamma, \sigma}(p)$ e $Q_{\gamma, 1}(p)$

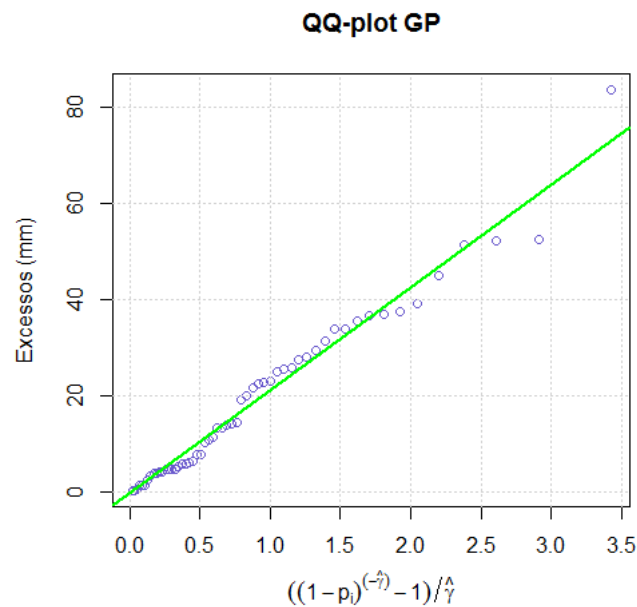


Figura 5.54: *QQ-plot GP dos excessos, referentes à amostra de níveis pluviométricos máximos anuais*

quadrados, obter-se-á para estimativa do parâmetro de escala do modelo, o respectivo valor do declive da reta que, neste caso, é:

Call:

```
lm(formula = yy ~ xx - 1)
```

Coefficients:

xx

21.30119

Ou seja, obtêm-se como estimativas preliminares para os parâmetros do modelo Generalizada de Pareto, os valores,

$$(\hat{\gamma}, \hat{\sigma}) = (-0.08436, 21.30119).$$

A estimativa obtida para o parâmetro de forma é negativa, logo, nesse caso, o modelo GP reduz-se a um modelo Beta. Contudo, não devemos esquecer que o modelo exponencial (distribuição GP com $\gamma = 0$) é, também, uma forte possibilidade. Assim, parece que a cauda da distribuição dos níveis pluviométricos máximos anuais, em Barcelos, balanceia entre a exponencial e a leve. Logo, recorrendo à dualidade existente entre o modelo GP e o modelo GEV (Teorema 2.17), a distribuição limite associada a estes níveis pluviométricos, poderá ser Gumbel ou Weibull. Relembre-se que, no ajustamento preliminar dos níveis pluviométricos máximos anuais, na metodologia Gumbel, as sugestões obtidas foram as mesmas .

II) Estimação dos Parâmetros

Pela análise preliminar efetuada na secção anterior, existem duas fortes possibilidades para o modelo subjacente aos excessos em estudo, o Exponencial (GP com $\gamma = 0$) ou o Beta (GP com $\gamma < 0$). Nesta secção, irão ser usados os métodos de estimação MV e MPP (apenas para o caso $\gamma \neq 0$) para obter estimativas dos parâmetros do modelo, para essas duas situações, bem como das outras quantidades importantes em acontecimentos raros, já apresentadas anteriormente. Assim, as estimativas dos parâmetros de forma e escala do modelo GP, serão obtidas recorrendo às expressões apresentadas nas secções 3.3.2 e 3.3.3, podendo, também, ser construídos intervalos de confiança para esses parâmetros, recorrendo à função de profile log-likelihood tal como foi apresentado na secção 3.3.5. Na tabela 5.34 encontram-se as estimativas pontuais e intervalares para esses parâmetros.

As estimativas para o parâmetro de forma são ambas negativas, no entanto, pelo método MV a cauda subjacente a estes níveis pluviométricos é mais leve do que pelo método MPP, uma vez que a estimativa MV é inferior à estimativa MPP. No entanto, repare-se que o intervalo de confiança para este parâmetro contém o valor zero, logo, mais uma vez, a cauda do tipo exponencial não pode ser posta de parte. Mais a frente serão aplicados testes que permitirão chegar a uma conclusão sobre qual das hipóteses se deve aceitar, $\gamma = 0$ ou $\gamma < 0$. Em relação às estimativas do parâmetro de escala, observa-se que o modelo exponencial é o que apresenta menor dispersão. Os intervalos de confiança

Tabela 5.34: Estimativas dos parâmetros para os modelos Exponencial e GP ($\gamma \neq 0$), para os excessos na amostra de níveis pluviométricos máximos anuais

Parâmetros	Modelo Exponencial ($\gamma = 0$)	Modelo GP ($\gamma \neq 0$)	
	MV	MV	MPP
$\log L$	-220.5189	-220.1558	-
$\hat{\gamma}$	-	-0.12593 (-0.315, 0.224)	-0.05523
$\hat{\sigma}$	18.87498 (14.63, 24.98)	21.27386 (14.30382, 30.08875)	19.91743

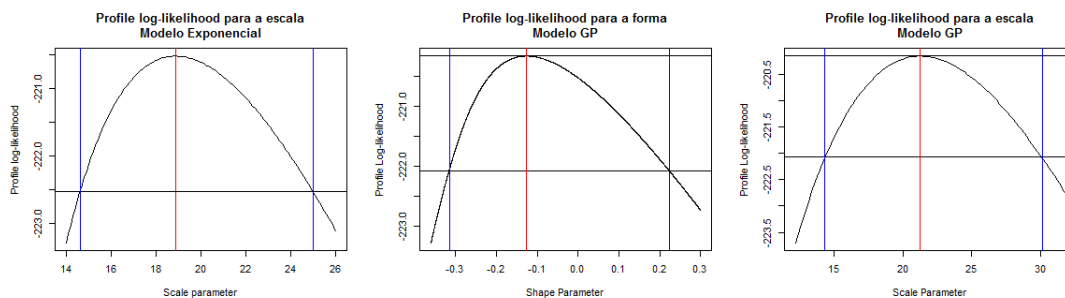


Figura 5.55: Intervalos de 95% de confiança para os parâmetros dos modelos Exponencial e GP, baseados no profile log-likelihood

têm um limite inferior muito próximo, para os dois modelos, no entanto a amplitude de valores possíveis para este parâmetro é maior no caso GP pelo método MV.

É de lembrar que quando foi aplicado o MMA, também se obtiveram estimativas negativas para o EVI, ou seja, houve a indicação de que a distribuição dos dados tinha uma cauda leve, no entanto mais pesada do que as obtidas pelo método POT. Outro pormenor a referir, está relacionado com o parâmetro de localização obtido para o modelo GEV (tabela 5.11), aproximadamente 64 mm, e que está muito próximo do valor escolhido para o *threshold* de 62.5 mm.

Tal como foi feito aquando da aplicação do MMA, também aqui a qualidade dos ajustamentos MV para os modelos Exponencial e GP poderá ser avaliada, rapidamente, através de representações gráficas, que se encontram nas figuras 5.56 e 5.57, respetivamente. Parece haver um melhor ajustamento dos pontos às retas nos pp-plot e qq-plot do modelo ajustado GP, do que no caso exponencial, embora as diferenças não sejam muito acentuadas. Contudo, no qq-plot exponencial há uma maior proximidade do ponto correspondente à maior observação à reta. Nos histogramas com as curvas da densidade, não se verificam grandes diferenças que mereçam ser evidenciadas.

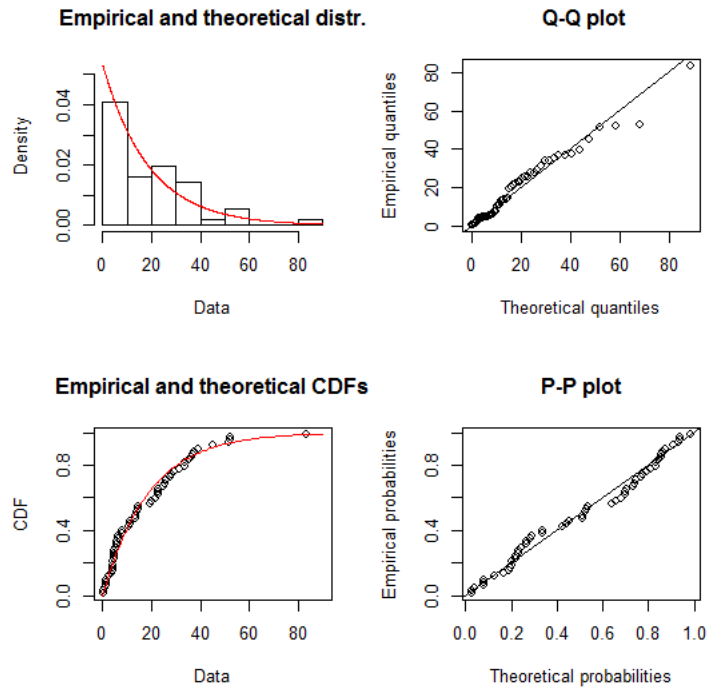


Figura 5.56: *Diagnóstico gráfico do ajustamento Exponencial aos excessos*

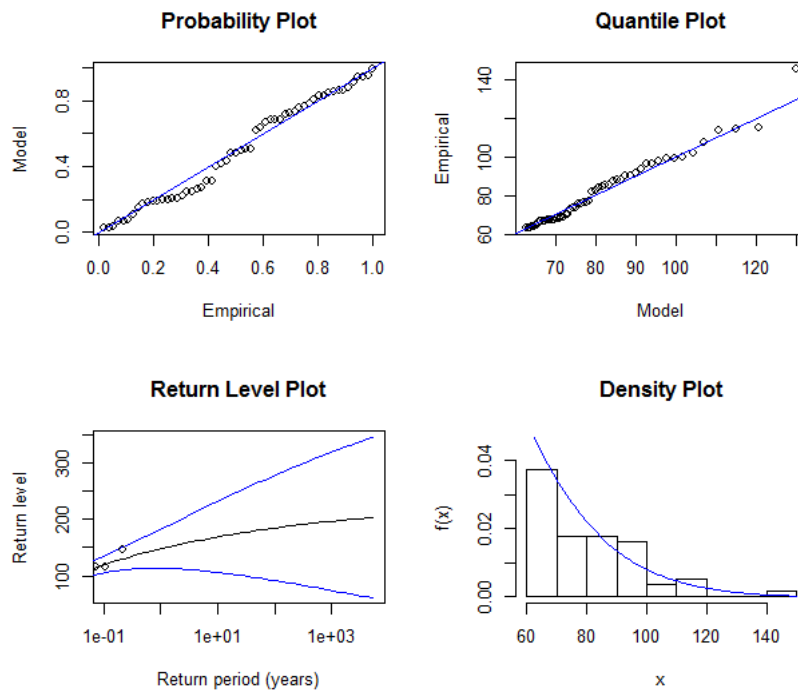


Figura 5.57: *Diagnóstico gráfico do ajustamento $GP(\gamma < 0)$ aos excessos*

Tendo os modelos ajustados, poder-se-á passar à estimação de valores importantes em modelação de eventos raros. Consultando a secção 3.3.4, poderão ser encontradas expressões para os estimadores dos quantis extremais, níveis e períodos de retorno, probabilidade de excedência e limite superior do suporte, quando utilizada a metodologia POT. Na tabela 5.35 estão registadas as estimativas obtidas para os diversos quantis de probabilidade elevada, bem como para os níveis de retorno de 100, 400 e 1000 anos. Como habitualmente, são também apresentados os intervalos de 95% de confiança para essas quantidades.

Tabela 5.35: *Estimativas pontuais e intervalares dos quantis extremais e níveis de retorno, para os níveis pluviométricos máximos anuais, obtidos pelo método POT*

Parâmetros	Modelo Exponencial $\gamma = 0$	Modelo GP $\gamma \neq 0$	
	MV	MV	MPP
$\hat{q}_{0.95}$	113.0336 (101.6,129.35)	110.8486 (101.47,128.5)	112.0695
$\hat{q}_{0.99}$	143.4117 (125.15,169.62)	132.9709 (119,185.25)	138.5261
$\hat{q}_{0.999}$	186.8729 (158.89,227.21)	157.7551 (135.3,312.4)	172.5129
$\hat{q}_{0.9999}$	230.3342 (192.3,284.2)	176.3009 (144,523.9)	202.4412
$\hat{q}_{0.99999}$	273.7954 (225.91,341.84)	190.1785 (149,885)	228.7957
$\hat{U}(100)$	143.4117 (125.15,169.62)	132.9709 (119,185.25)	138.5261
$\hat{U}(400)$	169.5780 (145.3,204.1)	148.7435 (129.9,254)	159.5036
$\hat{U}(1000)$	186.8729 (158.89,227.21)	157.7551 (135.3,312.4)	172.5129

Os valores das estimativas têm ligeiras diferenças, entre os diversos modelos e métodos, que vão aumentando à medida que o quantil se torna mais extremal. Os valores mais elevados obtêm-se para o modelo exponencial, contudo, se se observar os valores dos limites dos intervalos de confiança e as suas representações gráficas nas figuras 5.58 e 5.59, facilmente se verifica que, no modelo GP, os quantis e níveis de retorno poderão atingir valores superiores aos do modelo exponencial. Quando comparadas as estimativas dos quantis e níveis de retorno, obtidas pelos métodos POT e MMA (tabela 5.11), são notórias as semelhanças entre os valores obtidos para os modelos Gumbel e Exponencial, que correspondem ao caso $\gamma = 0$, e entre os modelos GEV e GP, com $\gamma < 0$. O mesmo acontece com os intervalos de confiança mas, há medida que se avança pela cauda direita

das distribuições, facilmente se percebe que pelo método POT, é possível ir-se mais longe, atingindo-se níveis pluviométricos anuais maiores. Por exemplo, recorrendo às estimativas MV para o modelo GEV, espera-se que, em média, uma vez em cada 100000 anos o nível pluviométrico em Barcelos exceda um determinado nível, compreendido entre os 153 mm e os 504.1 mm. Utilizando as estimativas MV para o modelo GP, o intervalo de possibilidades para esse nível de retorno aumenta, passando a ser de 149 mm a 885 mm.

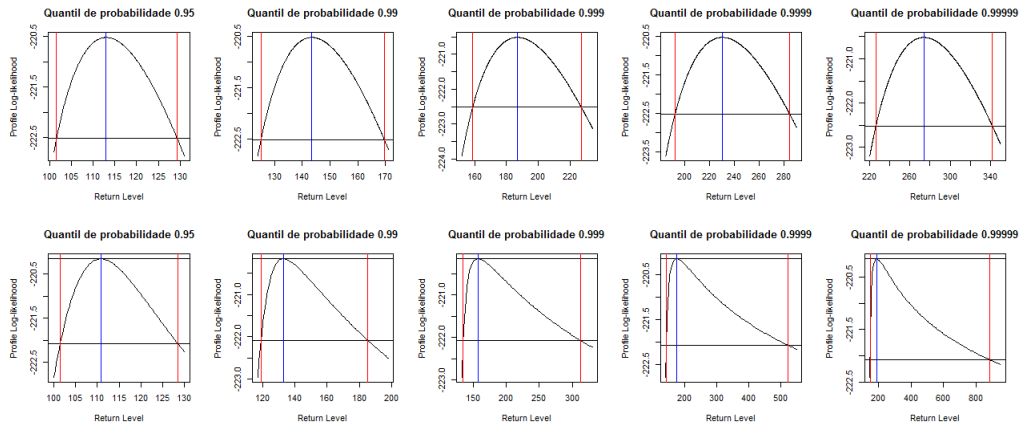


Figura 5.58: Intervalos de 95% de confiança para quantis extremos baseados no *profile log-likelihood*, para os níveis pluviométricos máximos anuais, através dos modelos exponencial (parte superior) e GP (parte inferior)

As estimativas para o período de retorno do nível pluviométrico máximo atingido na amostra, de 146 mm, e para a probabilidade de excedência desse nível podem ser visualizadas na tabela 5.36 e refletem, também, o que já foi dito. Através do modelo

Tabela 5.36: Estimativas do período de retorno e probabilidade de excedência do nível 146 mm, para os níveis pluviométricos máximos anuais, obtidos pelo método POT

Parâmetros	Modelo Exponencial $\gamma = 0$	Modelo GP $\gamma \neq 0$	
	MV	MV	MPP
$\hat{T}(146)$	114.6976	308.6725	161.9036
$\hat{P}(Y > 146)$	0.00872	0.00324	0.00618

exponencial, o nível de 146 mm não é considerado assim tão raro, como no modelo GP, com $\gamma < 0$, o que era esperado, uma vez que, neste último caso, o suporte da variável é finito. Assim, se se considerar que a distribuição destes níveis pluviométricos se encontra no domínio de atração Gumbel, e usando a metodologia POT, em média são precisos cerca de 115 anos para que o nível de precipitação exceda o valor de 146 mm, sendo a probabilidade desse acontecimento, aproximadamente, 0.00872. Já se se considerar o

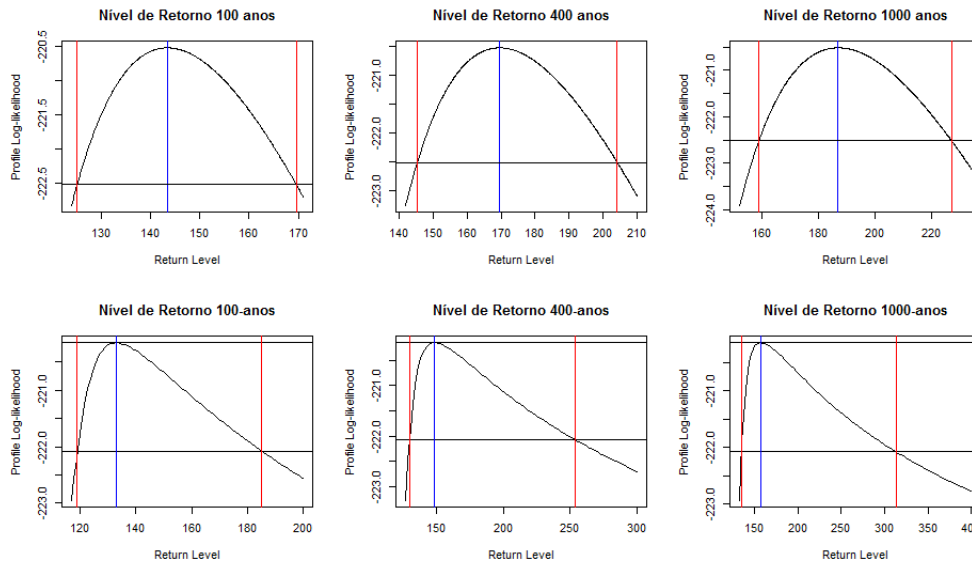


Figura 5.59: Intervalos de 95% de confiança para níveis de retorno baseados no profile log-likelihood, para os níveis pluviométricos máximos anuais, através dos modelos exponencial (parte superior) e GP (parte inferior)

domínio de atração Weibull, e usando a mesma abordagem, o período de retorno do nível 146 mm é de 309 anos, pelo método MV, ou de 162 anos, pelo método MPP, com probabilidades de excedência de 0.00324 e 0.00618, respectivamente. É de referir que a estimativa obtida para o período de retorno, quando o modelo selecionado para os excessos for o exponencial, parece mais adequada tendo em conta o historial da amostra disponível, ou seja, o nível 146 mm foi atingido uma vez em, pelo menos, 77 anos, logo o valor de 115 anos não é muito desfasado desta situação.

Antes de se terminar esta secção, poderão ser obtidas estimativas para o limite superior do suporte, já que se obteve um valor negativo para a forma da distribuição GP. Assim, usando a expressão (3.48), obtém-se,

Estimativa Limite superior do suporte (método MV): 231.4343;

Estimativa Limite superior do suporte (método MPP): 423.1412.

A estimativa MV obtida, neste caso, corresponde ao menor valor obtido para este limite quando usados os métodos POT e dos máximos anuais e, por esse motivo, a probabilidade de excedência do nível 146 é também a mais pequena das obtidas, pois esse nível estará mais perto do valor máximo admitido para um nível pluviométrico anual, em Barcelos.

III) Testes ao modelo GP

Pela análise preliminar efetuada aos excessos, na amostra dos níveis pluviométricos máximos anuais, os modelos GP com $\gamma = 0$ e $\gamma < 0$ pareceram adequados para os modelar. Sendo assim, tem que ser tomada uma decisão sobre qual das duas situações será a mais adequada. Para tal irão ser usados os testes de hipóteses apresentados na secção 3.3.6. As hipóteses em estudo serão, nomeadamente,

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma \neq 0, \quad (5.15)$$

ou

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma < 0. \quad (5.16)$$

Para as hipóteses (5.15), poderão ser aplicados o teste sugerido por Marohn (2000) e o TRV. Na tabela 5.37 encontram-se os valores da estatística de teste, o p-value e a decisão a 5%, para cada um desses testes.

Tabela 5.37: Resultados obtidos nos testes às hipóteses definidas em (5.15)

Teste	Valor Observado da Estatística de Teste	p-value	Decisão $\alpha = 0.05$
Teste POT (Mahron)	$t_m^* = -0.6716027$	0.5018366	Não rejeitar H_0
TRV	$d^* = 0.6777867$	0.4103499	Não rejeitar H_0

Em ambos os testes, não se rejeita a hipótese nula, ou seja, não existe evidências estatísticas para não aceitar que o parâmetro de forma é igual a zero. No entanto, para o primeiro teste, recorde-se que Marohn (2000) mostrou que, para a versão bilateral, a estatística de teste é enviesada e que a sua potência é baixa para amostras de dimensão inferior a 500. Ora, a presente amostra tem dimensão 56, logo encontra-se nessa situação. Sendo assim, poderá ser usada a versão unilateral deste teste, neste caso, para as hipóteses definidas em (5.16). Também para essas hipóteses será usado o teste sugerido por Gomes and van Monfort (1986).

Na tabela 5.38 poderão ser visualizados os resultados obtidos. Usando a versão unilateral do teste apresentado por Marohn (2000), também se obtém como decisão a não rejeição de $\gamma = 0$. Já para o outro teste aí apresentado, a qualquer nível de significância, rejeita-se H_0 , aceitando-se a hipótese de que $\gamma < 0$. Como a dimensão da amostra é relativamente pequena, para este último caso, também poderia ser feita a comparação do valor observado da estatística de teste com os quantis apresentados na tabela 3.3. Escolhendo a linha do $m=100$, a região crítica, para um nível de significância de 5%, é dada

Tabela 5.38: Resultados obtidos nos testes às hipóteses definidas em (5.16)

Teste	Valor Observado da Estatística de Teste	p-value	Decisão $\alpha = 0.05$
Teste POT unilateral (Mahron)	$t_m^* = -0.6716027$	0.2509183	Não rejeitar H_0
Teste de Gomes e van Monfort	$g_m^* = -2.702483$	$3.324696e - 07$	Rejeitar H_0

por $G_m^* < -1.21$. Como o valor observado da estatística é de -2.702484, então rejeita-se a hipótese nula.

Na secção 3.3.6 foram também apresentadas as estatísticas de Kolmogorov-Smirnov, de Cramér-von Mises e de Anderson-Darling, que possibilitam realizar testes à qualidade do ajustamento exponencial, no caso da primeira, e GP com parâmetros desconhecidos, para as duas últimas. Na tabela 5.39 apresentam-se, resumidamente, os valores obtidos para cada uma dessas estatísticas e a respetiva decisão, ao nível de significância de 5%.

Tabela 5.39: Resultados obtidos nos testes à qualidade do ajustamento do modelo GP com $\gamma = 0$

Teste	Valor Observado da Estatística de Teste	Decisão $\alpha = 0.05$
Kolmogorov-Smirnov	$d_m = 0.09397304$	Não Rejeitar H_0
Cramér-von Mises	$w_m^2 = 0.08772998$	Não Rejeitar H_0
Anderson-Darling	$a_m^2 = 0.490572$	Não Rejeitar H_0

Usando a estatística de Kolmogorov-Smirnov, haverá a rejeição da hipótese nula, que neste caso será correspondente a rejeitar a distribuição exponencial, obtida por ajustamento MV, como o modelo adequado para explicar o comportamento dos excessos dos níveis pluviométricos máximos anuais, em relação ao *threshold* de 62.5 mm, se o valor observado da estatística for superior aos quantis registados na tabela 3.4. Para encontrar o quantil crítico, deverá ser utilizada a última linha da tabela, uma vez que a amostra tem dimensão superior a 30. Assim, usando a expressão indicada, obtêm-se os quantis,

$$\begin{aligned} Q_{.10} &= 0.96/\text{sqrt}(56) & Q_{.5} &= 1.06/\text{sqrt}(56) & Q_{.1} &= 1.25/\text{sqrt}(56) \\ Q_{.10} &= 0.1282854 & Q_{.5} &= 0.1416485 & Q_{.1} &= 0.1670383. \end{aligned}$$

Ora o valor 0.09397304 é inferior a qualquer um dos três quantis, e como tal não se rejeita a hipótese nula.

Para os testes de Cramer-von Mises e Anderson Darling, os quantis críticos poderão ser encontrados nas tabelas 3.5 e 3.6, respetivamente. Tendo em conta que se pretende testar

a bondade do ajustamento GP, com forma nula, dever-se-á utilizar a linha correspondente a $\gamma = 0$. Para todas as situações, o valor observado das estatísticas de teste são inferiores aos valores das tabelas, não se rejeitando, portanto, o ajustamento GP com $\gamma = 0$.

Portanto, os testes à qualidade do ajustamento exponencial (GP com $\gamma = 0$), com os parâmetros estimados, indicam que este modelo é adequado para modelar os excessos em estudo.

Tendo em conta todos os testes aplicados, à exceção do teste de Gomes e van Monfort, o modelo exponencial parece ser o mais adequado para modelar esta amostra de excessos. Ora, como o modelo exponencial corresponde ao caso particular do modelo GP com parâmetro de forma $\gamma = 0$, vem que a distribuição dos níveis pluviométricos máximos anuais pertence ao domínio de atração Gumbel, o que está de acordo com a conclusão obtida aquando da aplicação do MMA.

5.2.3.2 Níveis Pluviométricos Máximos Mensais

I) Escolha do *Threshold* e Análise Preliminar dos Dados

Nesta secção proceder-se-á a uma análise POT à amostra dos níveis pluviométricos máximos mensais, em Barcelos. Esta amostra é constituída por 910 observações registadas no período compreendido entre 01/01/1932 a 01/05/2008.

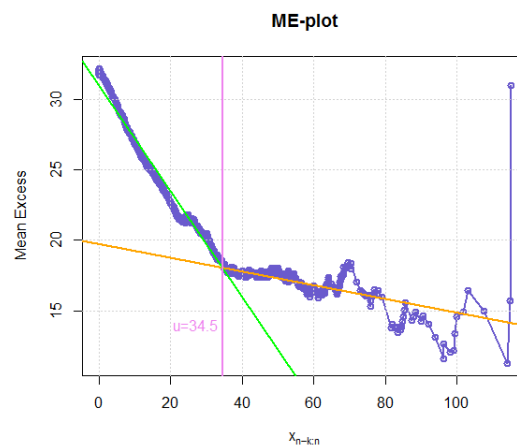


Figura 5.60: *ME-plot*, com retas ajustadas, para os níveis pluviométricos máximos mensais

Com recurso ao ME-plot, escolheu-se para *threshold* o nível de 34.5 mm. Na figura 5.60 poderá ser revisto o gráfico da função de excesso médio, com duas retas ajustadas. Inicialmente, os pontos apresentam um padrão de decrescimento acentuado, para depois ficar mais moderado, quase constante. Essa mudança na velocidade de decrescimento

poderá ser comprovada com as duas retas ajustadas aos pontos. Ambas apresentam um declive negativo, no entanto, a reta ajustada aos pontos finais tem um declive maior e perto de zero. Essa transição de declives dá-se, sensivelmente, no valor 34.5, logo o porquê da escolha desse nível para *threshold*.

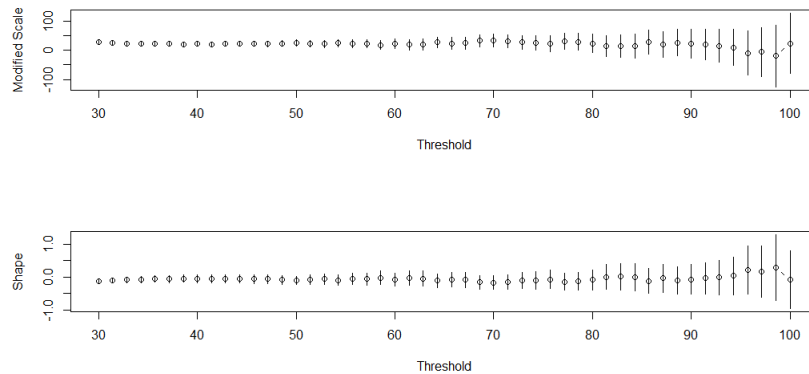


Figura 5.61: *Estimativas dos parâmetros de forma e escala versus threshold*

Na figura 5.61 estão representados os gráficos com as estimativas, pontuais e intervalares, dos parâmetros de forma e escala segundo o *threshold* escolhido. Um outro indicador da adequabilidade deste *threshold* é dada pela estabilidade das estimativas obtidas para a forma e escala do modelo GP, verificada nos dois gráficos. Como habitual, há alguma irregularidade na parte final dos gráficos, que não deverá ser tomada em consideração.

Aplicando o *threshold* de 34.5, à amostra de níveis pluviométricos máximos mensais, obtêm-se 371 excedências, o que corresponde a 41% da amostra. Na figura 5.62 poderão ser visualizadas essas excedências.

Quando foi aplicada a metodologia Gumbel à amostra completa dos máximos mensais, foi referido, por observação do gráfico da autocorrelação parcial da figura 5.22, que as observações não eram independentes devido, essencialmente, à sazonalidade. Quando aplicada a metodologia POT, esse problema é ultrapassado, não se verificando correlação significativa entre os elementos da amostra daí resultante, o que poderá ser comprovado pela figura 5.63.

Como se está perante uma amostra de excessos, sabe-se que a distribuição GP poderá modelar o seu comportamento. Para se obterem estimativas preliminares dos parâmetros do modelo, poderão ser construídos os qq-plots GP para os casos $\gamma = 0$ e $\gamma \neq 0$. Para além disso, com os qq-plots poder-se-á compreender qual a distribuição que melhor se adequa a modelar esses excessos e, com base nisso, saber a qual dos domínio de atração pertencerá a distribuição subjacente aos níveis pluviométricos máximos mensais. Recorde-se que $\gamma > 0$

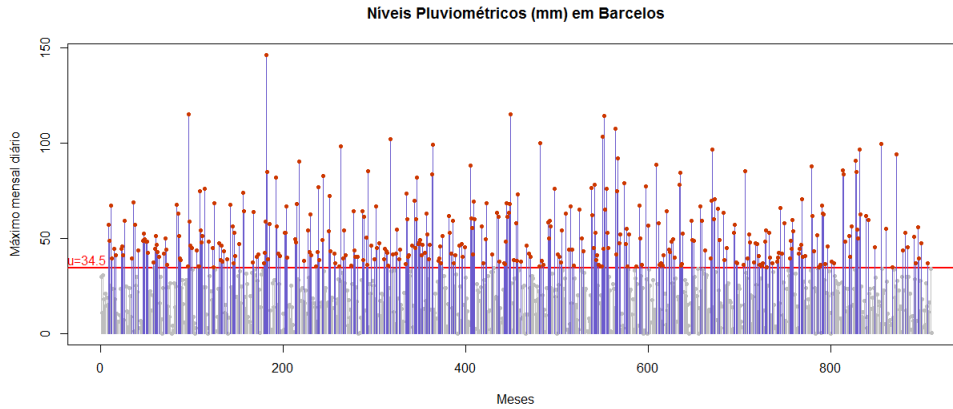


Figura 5.62: *Excedências de $u=34.5$, na amostra dos níveis pluviométricos máximos mensais*

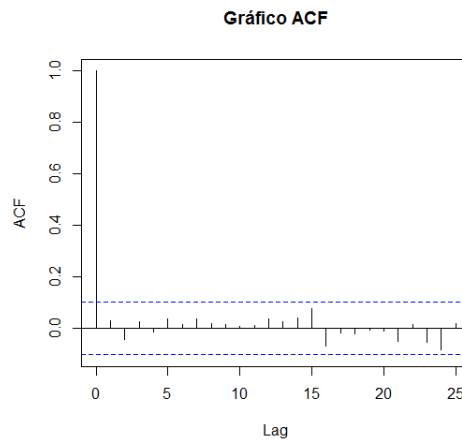


Figura 5.63: *Gráfico de Autocorrelação Parcial*

corresponde ao domínio max-estável Fréchet, com distribuições com cauda direita pesada e limite superior do suporte não finito, $\gamma = 0$ refere-se ao domínio Gumbel, onde as caudas das distribuições são do tipo exponencial e o limite superior do suporte poderá ou não ser finito, e $\gamma < 0$ que será referente ao domínio de atração Weibull, com distribuições com caudas leves e limite superior do suporte finito.

Na figura 5.64 encontram-se os qq-plots referidos no parágrafo anterior, o primeiro correspondente ao modelo GP com forma nula, equivalente ao modelo exponencial, e o segundo ao modelo GP com forma $\hat{\gamma} = -0.03137$, obtida usando o método já explicado anteriormente. A linearidade dos pontos em ambos os gráficos é bastante acentuada, não se verificando grandes diferenças entre as duas situações. A estimativa obtida para o parâmetro de forma, quando considerado o caso de $\gamma \neq 0$ é negativa mas muito próxima de

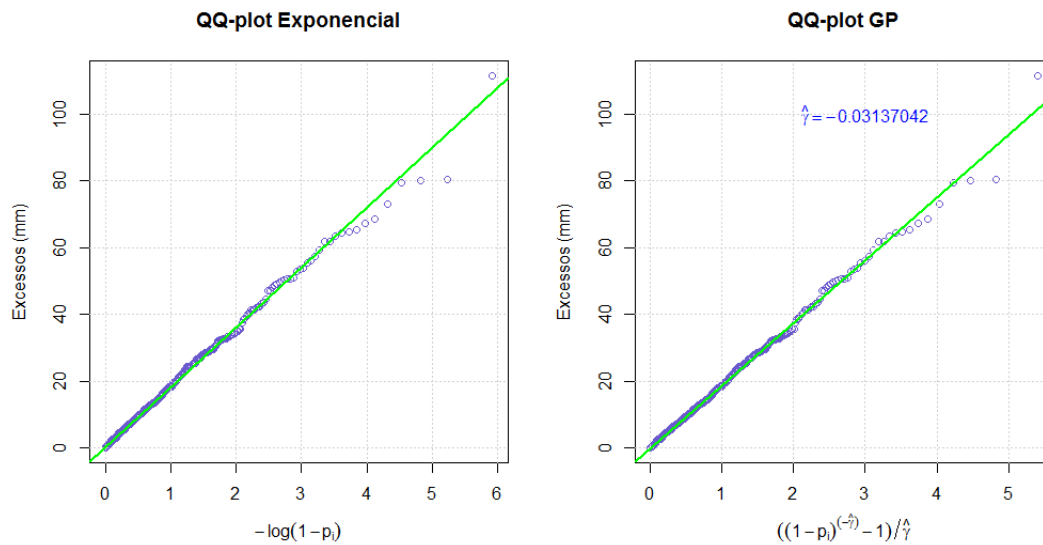


Figura 5.64: *QQ-plots GP (com $\gamma = 0$ e $\gamma \neq 0$) dos excessos, referentes à amostra de níveis pluviométricos máximos mensais*

zero. Em relação à escala, as estimativas obtidas para os casos exponencial e GP ($\gamma < 0$) foram, respetivamente, 17.99062 e 18.78516, o que não são muito diferentes. Tendo em conta o aspeto dos dois qq-plots, a proximidade à origem do parâmetro de forma e a semelhança entre os valores obtidos para o parâmetro de escala, existem fortes suspeitas de que o modelo exponencial será adequado para modelar esta amostra de excessos, o que implica que a amostra original dos níveis pluviométricos máximos mensais deverá ter uma distribuição com cauda do tipo exponencial, pertencente ao domínio max-estável Gumbel. Quando foi aplicado o MMA, chegou-se à mesma conclusão.

II) Estimação dos Parâmetros

Usando os métodos de estimação MV e MPP, poderão ser obtidas estimativas dos parâmetros do modelo GP, para os casos com $\gamma = 0$ e $\gamma \neq 0$. Assim, na tabela 5.40 poderão ser encontradas as estimativas dos parâmetros dos modelos e intervalos de 95% de confiança, construídos pelo método profile log-likelihood. Na figura 5.65 encontram-se representações gráficas das estimativas intervalares para estes parâmetros.

Quando considerado o modelo GP com $\gamma \neq 0$, os resultados obtidos pelos métodos MV e MPP são muito parecidos. As estimativas obtidas para o parâmetro de forma são negativas mas próximas de zero. O intervalo com nível de confiança de 95%, para este parâmetro compreende os valores entre -0.1515 e 0.0385, ou seja, inclui a possibilidade do EVI ser nulo e, conseqüentemente, estar-se no domínio das caudas exponenciais. Ao nível

Tabela 5.40: Estimativas dos parâmetros para os modelos Exponencial e GP ($\gamma \neq 0$), para os excessos na amostra de níveis pluviométricos máximos mensais

Parâmetros	Modelo Exponencial ($\gamma = 0$)	Modelo GP ($\gamma \neq 0$)	
	MV	MV	MPP
$\log L$	-1448.075	-1447.209	-
$\hat{\gamma}$	-	-0.07004 (-0.1515, 0.0385)	-0.07553
$\hat{\sigma}$	18.23181 (16.49, 20.23)	19.51336 (16.95092, 22.32584)	19.60878

da escala, os valores são, também, muito semelhantes, mas ligeiramente mais baixos para o modelo exponencial.

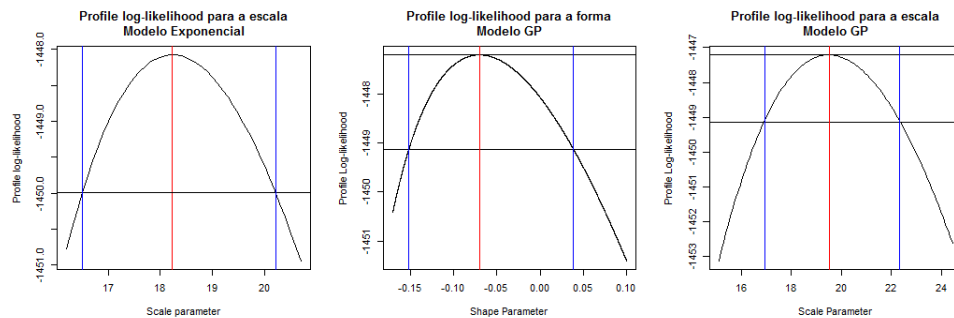


Figura 5.65: Intervalos de 95% de confiança para os parâmetros dos modelos Exponencial e GP, baseados no profile log-likelihood

A qualidade dos dois ajustamentos MV efetuados, poderá ser avaliada através dos gráficos representados nas figuras 5.66 e 5.67.

A qualidade dos modelos exponencial e GP ($\gamma < 0$) é bastante boa, como se pode observar pelos vários gráficos que constituem este diagnóstico. É de salientar a "quase perfeição" verificada no histograma sobreposto pela curva da densidade estimada, para o caso exponencial, e a linearidade nos pp-plots para os dois casos. Em relação aos qq-plots, no referente ao modelo GP, parece haver uma maior proximidade dos pontos à reta.

Usando os modelos ajustados a esta amostra de excessos, referentes à amostra de níveis pluviométricos máximos mensais, poderão ser usadas as várias expressões para estimar os quantis de probabilidades 0.95, 0.99, 0.999, 0.9999 e 0.99999, os níveis de retorno de 100, 400 e 1000 anos, o período de retorno do nível pluviométrico máximo observado, de 146 mm, a probabilidade de excedência desse nível e o limite superior do suporte, uma vez que se obteve uma estimativa negativa para o parâmetro de forma.

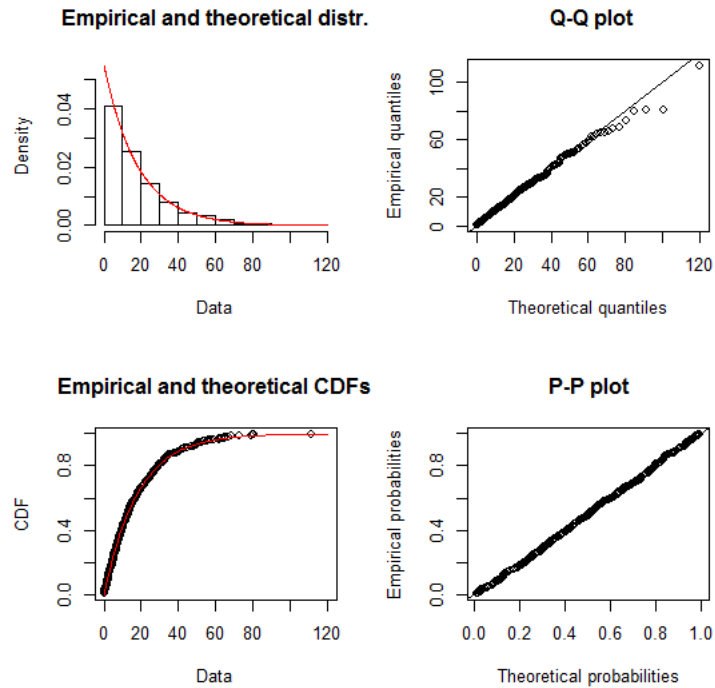


Figura 5.66: *Diagnóstico gráfico do ajustamento Exponencial aos excessos*

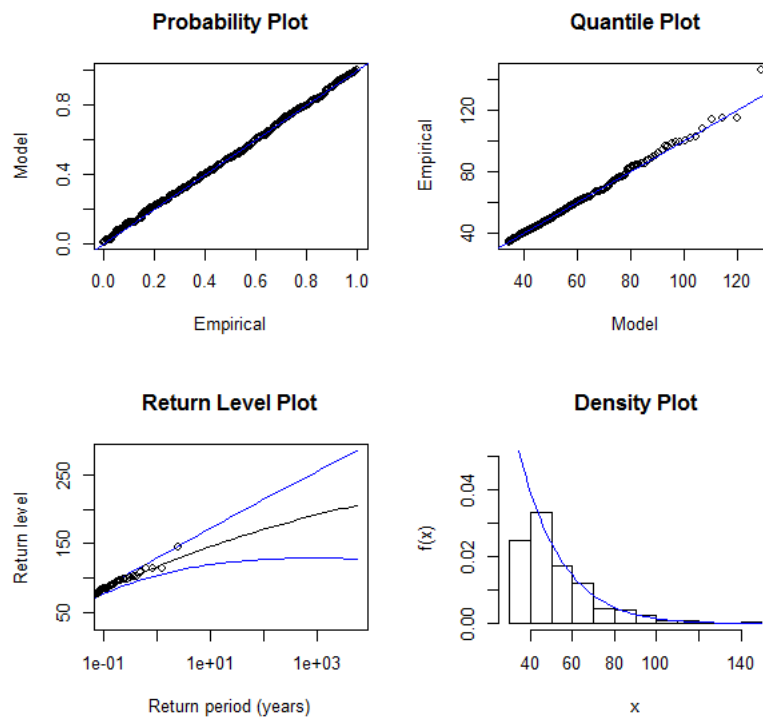


Figura 5.67: *Diagnóstico gráfico do ajustamento $GP(\gamma < 0)$ aos excessos*

Tabela 5.41: Estimativas pontuais e intervalares dos quantis extremos, para os níveis pluviométricos máximos mensais, obtidos pelo método POT

Quantis	Modelo Exponencial $\gamma = 0$	Modelo GP $\gamma \neq 0$	
	MV	MV	MPP
$\hat{q}_{0.95}$	72.75925 (68.96,76.8)	72.58147 (69.2,76.43)	72.55363
$\hat{q}_{0.99}$	102.1022 (95.5,109.36)	98.22194 (92.02,107.52)	97.91457
$\hat{q}_{0.999}$	144.08247 (133.52,155.87)	130.22634 (116.96,156.2)	129.23503
$\hat{q}_{0.9999}$	186.06274 (171.45,202.46)	157.46397 (134.5,209.7)	155.55600
$\hat{q}_{0.99999}$	228.04301 (209.5,249)	180.64480 (146.8,268.6)	177.67552

Na tabela 5.41 encontram-se os valores das estimativas para os diversos quantis extremos, bem como os respectivos intervalos de confiança. Como se percebe após leitura da tabela, existe uma grande semelhança para os primeiros quantis, nos dois modelos considerados. À medida que a probabilidade do quantil se aproxima de 1, as diferenças começam a surgir, obtendo-se valores superiores no caso exponencial, visto que o limite superior do suporte poderá ser infinito, o que não acontece no modelo GP com $\gamma < 0$, em que este será finito. Já tinham sido estimados estes quantis, para a amostra de níveis pluviométricos máximos mensais, através do MMA, e que se encontram na tabela 5.15. As estimativas MV obtidas para os quantis, usando os modelos Gumbel, no MMA, e o Exponencial, no método POT, são praticamente iguais, o que vem, mais uma vez, demonstrar a relação existente entre o modelo GP com o modelo GEV, e recorde-se que após aplicação de vários testes, o modelo Gumbel foi o selecionado para a distribuição subjacente a estas observações. Já se se comparar os quantis obtidos nos modelos GEV, com $\gamma \neq 0$, e GP, com $\gamma \neq 0$, os de probabilidade mais baixa são muito semelhantes, mas quando essa probabilidade aumenta, os quantis obtidos com a distribuição GEV são maiores. Ora tal facto é perfeitamente normal, uma vez que a estimativa MV obtida para o parâmetro de forma da GEV era positivo. Recorde-se que em Nascimento (2009) foram estimados estes quantis, tendo como base a amostra de máximos mensais, mas usando técnicas bayesianas. Esses quantis encontram-se na figura 5.32. Os valores aí obtidos através dos modelos $MGPD_1$ e $MGPD_2$ são semelhantes aos apresentados na tabela 5.41, para o caso GP com $\gamma = 0$. No entanto, a maior semelhança acontece entre os quantis do modelo $MGPD_3$, que foi considerado o melhor modelo nesse estudo, novamente com o modelo GP com $\gamma = 0$, à exceção do quantil de probabilidade 0.99999. Relembre-se que

os valores obtidos para estes quantis, no modelo $MGPD_3$, foram,

$$\hat{q}_{0.95} = 73.29, \hat{q}_{0.99} = 102.24, \hat{q}_{0.999} = 137.77, \hat{q}_{0.9999} = 184.54 \text{ e } \hat{q}_{0.99999} = 454.26.$$

Como se pode comprovar, os primeiros quatro quantis assumem, de facto, valores bastante parecidos aos obtidos no modelo exponencial, pela metodologia POT. O quantil de probabilidade 0.99999 apresenta uma grande divergência em relação a qualquer um, com a mesma probabilidade, obtido com o método POT, não fazendo parte nem dos intervalos de confiança obtidos para esta quantidade. Além disso, este valor de 454.2 mm parece, também, ser um pouco elevado quando comparado com os outros quantis obtidos no modelo $MGPD_3$.

Tabela 5.42: *Estimativas pontuais e intervalares dos níveis de retorno, para os níveis pluviométricos máximos mensais, obtidos pelo método POT*

Níveis de Retorno	Modelo Exponencial $\gamma = 0$	Modelo GP $\gamma \neq 0$	
	MV	MV	MPP
$\hat{U}(1200)$	147.4065 (136.55,159.54)	132.5468 (118.6,160.26)	131.4901
$\hat{U}(4800)$	172.6812 (159.4,187.5)	149.2538 (129.6,191.98)	147.6575
$\hat{U}(12000)$	189.3868 (174.5,206.1)	159.4388 (135.49,214.35)	157.4511

Em relação aos níveis de retorno de 100, 400 e 1000 anos que, neste caso, será equivalente a 1200, 4800 e 12000 meses, espera-se que sejam ultrapassados, em média, uma vez em cada um destes períodos, valores maiores de precipitação no caso exponencial do que na situação GP com $\gamma < 0$. Esta situação já era previsível e, facilmente, pode ser comprovada, bastando para isso, observar as estimativas pontuais obtidas. Contudo, quando se observam os limites dos intervalos de confiança para estas quantidades, existe uma maior amplitude nos intervalos obtidos pela função de profile log-likelihood do modelo GP com $\gamma < 0$. Aliás os limites superiores dos intervalos são maiores que os obtidos pelo modelo exponencial, dando assim a possibilidade de se ter um nível pluviométrico de retorno superior quando o EVI é negativo. No entanto, as estimativas pontuais obtidas, através do modelo GP com $\gamma < 0$, não são candidatas para os níveis de retorno no modelo exponencial, por serem menores que o limite inferior dos intervalos de 95% confiança. Mais uma vez se se comparar os níveis de retorno obtidos usando o MMA, existe uma grande semelhança entre os modelos Gumbel e Exponencial.

As estimativas para o período de retorno do nível pluviométrico máximo atingido na amostra, de 146 mm, e para a probabilidade de excedência desse nível encontram-se na

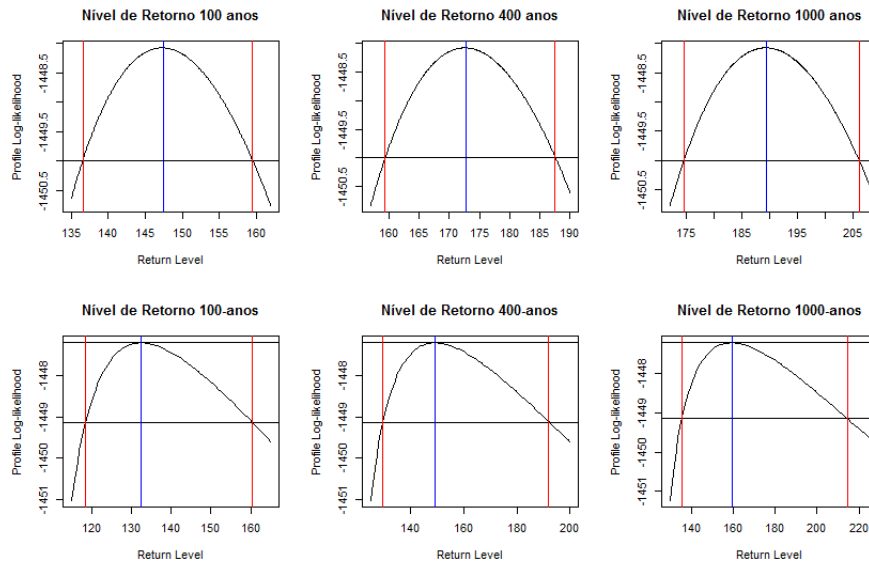


Figura 5.68: Intervalos de 95% de confiança para níveis de retorno baseados no profile log-likelihood, para os níveis pluviométricos máximos mensais, através dos modelos exponencial (parte superior) e GP (parte inferior)

Tabela 5.43: Estimativas do período de retorno e probabilidade de excedência do nível 146 mm, para os níveis pluviométricos máximos mensais, obtidos pelo método POT

Parâmetros	Modelo Exponencial $\gamma = 0$	Modelo GP $\gamma \neq 0$	
	MV	MV	MPP
$\hat{T}(146)$	1119.451 meses 93.28758 anos	3652.91 meses 304.4092 anos	4167.4 meses 347.2833 anos
$\hat{P}(Y > 146)$	0.00089	0.00027	0.00024

tabela 5.43. Existem grandes diferenças entre os valores estimados usando os modelos GP com $\gamma = 0$ e $\gamma \neq 0$. Usando o método de estimação MV, o nível pluviométricos 146 mm poderá ser excedido com uma probabilidade de 0.0009, no caso dos excessos serem modelados pela distribuição GP com $\gamma = 0$, ou de aproximadamente 0.0003, no caso da distribuição usada ser a GP com $\gamma \neq 0$. Ora isso traduz-se num período de retorno desse nível de 1119.451 e 3652.91 meses, ou equivalentemente, 93.3 e 304.4 anos, respetivamente.

Se for feita a comparação destes valores com os obtido pelo MMA, os resultados não são assim tão parecidos. As diferenças são notórias entre as estimativas obtidas através da distribuição GEV ($\gamma \neq 0$) e da distribuição GP ($\gamma \neq 0$). Se o método de estimação escolhido for o MV, obteve-se, no primeiro caso, um período de retorno de 70 anos contra os 304 anos deste último. Já se o método escolhido for o MPP, esses valores passam a 149 anos e 347 anos, respetivamente. Para o caso de $\gamma = 0$, as estimativas dadas pelos

métodos Gumbel e POT também apresentam diferenças, embora não tão acentuadas, 123 anos (estimativa MV) ou 107 anos (estimativa MPP) contra 93 anos.

Como a estimativa obtida para o parâmetro de forma do modelo GP, quer usando o método MV, quer usando o método MPP, é negativa, poderá ser estimado o limite superior do suporte, uma vez que será finito. Assim, obteve-se

Estimativa Limite superior do suporte (método MV): 313.1014;

Estimativa Limite superior do suporte (método MPP): 294.1289.

Os dois valores obtidos para o maior nível pluviométrico que alguma vez será atingido, num determinado mês, estão muito próximos entre si. Na metodologia dos máximos anuais, esse valor foi de 1100.641 mm, valor esse bastante diferente dos obtidos através do método POT.

Numa panorâmica geral, os resultados fornecidos pela metodologia POT parecem ser mais aceitáveis e, de certa forma, mais realistas, de acordo com a informação fornecida pela amostra de dados pluviométricos, em Barcelos. Veja-se, por exemplo, o período de retorno obtido quando se considera que se está num domínio de caudas do tipo exponencial. Em, pelo menos, 77 anos ocorreu uma vez o nível de precipitação de 146 mm e o valor fornecido para o período de retorno deste nível foi de 93 anos, o que parece razoável.

III) Testes ao Modelo GP

Pela análise preliminar realizada, concluiu-se que a distribuição dos excessos acima do *threshold* $u = 34.5$, poderá ser Exponencial (GP com $\gamma = 0$) ou Beta (GP com $\gamma < 0$). Com vista a fazer uma escolha acertada do melhor modelo, serão aplicados os testes apresentados na secção 3.3.6 às hipóteses:

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma \neq 0, \quad (5.17)$$

e

$$H_0 : \gamma = 0 \quad vs \quad H_1 : \gamma < 0. \quad (5.18)$$

Para as hipóteses (5.17), poderão ser encontrados na tabela 5.44 os valores da estatística de teste, o p-value e a respetiva decisão a 5%.

Não se rejeita, portanto, a hipótese nula aos níveis de significância usuais, nos dois testes. Relembre-se, mais uma vez que Marohn (2000) mostrou que, para amostras de dimensão inferior a 500, a estatística do primeiro teste apresentado na tabela 5.44 poderá ser enviesada. Esse problema poderá ser ultrapassado, aplicando-se a versão unilateral

Tabela 5.44: Resultados obtidos nos testes às hipóteses definidas em (5.17)

Teste	Valor Observado da Estatística de Teste	p-value	Decisão $\alpha = 0.05$
Teste POT (Mahron)	$t_m^* = -1.183908$	0.2364494	Não rejeitar H_0
TRV	$d^* = 1.713525$	0.1905286	Não rejeitar H_0

desse mesmo teste, por exemplo, para as hipóteses (5.18). Para essas mesmas hipóteses aplicar-se-á também o teste apresentado por Gomes and van Monfort (1986). Os resultados obtidos encontram-se resumidos na tabela 5.45.

Tabela 5.45: Resultados obtidos nos testes às hipóteses definidas em (5.18)

Teste	Valor Observado da Estatística de Teste	p-value	Decisão $\alpha = 0.05$
Teste POT unilateral (Mahron)	$t_m^* = -1.183908$	0.1182247	Não rejeitar H_0
Teste de Gomes e van Monfort	$g_m^* = -3.785687$	$7.284797e - 20$	Rejeitar H_0

Como se pode observar, também na versão unilateral do teste Marohn (2000), não se rejeita a hipótese da distribuição associada aos excessos, dos níveis pluviométricos máximos mensais, ser Exponencial. Já no segundo teste apresentado na tabela 5.45, para os níveis de significância usuais, a hipótese nula da distribuição Exponencial é rejeitada, tendo-se obtido um p-value, estranhamente, próximo de zero. Logo, este é o primeiro teste que seleciona a distribuição GP, com $\gamma < 0$, como o modelo paramétrico adequado para ser ajustado a estes excessos.

Para terminar esta secção, poderão ser aplicados os testes à qualidade do ajustamento Exponencial, com o teste de Kolmogorov-Smirnov, e GP, com os testes de Cramér-von Mises e de Anderson-Darling. Os resultados obtidos encontram-se na tabela 5.46.

Tabela 5.46: Resultados obtidos nos testes à qualidade do ajustamento do modelo GP com $\gamma = 0$

Teste	Valor Observado da Estatística de Teste	Decisão $\alpha = 0.05$
Kolmogorov-Smirnov	$d_m = 0.03273602$	Não Rejeitar H_0
Cramér-von Mises	$w_m^2 = 0.06764371$	Não Rejeitar H_0
Anderson-Darling	$a_m^2 = 0.537193$	Não Rejeitar H_0

Para o teste de Kolmogorov-Smirnov, a decisão é tomada após comparação do valor

observado da estatística com os quantis da tabela 3.4. Como a amostra é constituída por 371 excessos, deverá ser utilizada a última linha da tabela que, neste caso, se traduz nos valores,

$$\begin{array}{lll} Q_{.10} = 0.96/\sqrt{371} & Q_{.5} = 1.06/\sqrt{371} & Q_{.1} = 1.25/\sqrt{371} \\ Q_{.10} = 0.04984072 & Q_{.5} = 0.05503246 & Q_{.1} = 0.06489677. \end{array}$$

Uma vez que os quantis são superiores ao valor observado da estatística, não se rejeita a distribuição exponencial que foi ajustada a estes excessos.

Para o caso referente aos testes de Cramer-von Mises e Anderson Darling, foram usados os quantis das tabelas 3.5 e 3.6, mais em concreto os valores representados na linha referente a $\gamma = 0$, já que se pretende testar a bondade do ajustamento GP com parâmetro de forma igual a zero. Aos níveis de significância usuais, não se rejeita a hipótese nula, pois o valor observado de cada uma das estatísticas é inferior aos respetivos quantis críticos.

Portanto, a conclusão geral, será de que o modelo Exponencial, correspondente ao modelo GP com $\gamma = 0$, é adequado para explicar e prever o comportamento dos excessos dos níveis pluviométricos máximos mensais, em relação ao *threshold* de $u = 34.5$ mm. Quer isto dizer que ambas as metodologias aplicadas, a dos máximos anuais e a POT, rumam na mesma direção, ou seja, indicam que a distribuição dos níveis pluviométricos máximos mensais, em Barcelos, pertence ao domínio das caudas do tipo exponencial, isto é, ao domínio de atração Gumbel.

5.3 Análise Semi-Paramétrica dos Dados

5.3.1 Níveis Pluviométricos Máximos Anuais

I) Estimação do EVI

Como foi apresentado no capítulo 4, numa abordagem semi-paramétrica não se supõe nenhum modelo paramétrico subjacente à amostra de observações i.i.d., $F(\cdot)$. Neste caso supõe-se que $F \in \mathcal{D}(G_\gamma)$, para algum $\gamma \in \mathbb{R}$. Qualquer inferência referente à cauda de F é feita baseada nas $k + 1$ observações de topo, que são obtidas a partir de um *threshold* aleatório. Portanto, para conhecer qual o domínio de atração da distribuição subjacente aos dados, é preciso conhecer o valor do EVI, γ . Recorde-se que o caso $\gamma > 0$ corresponde ao Domínio Fréchet, caracterizado pelas distribuições de caudas pesadas e com limite superior do suporte não finito; se $\gamma < 0$ então tem-se em mão uma distribuição

pertencente ao domínio mas-estável Weibull, onde o limite superior do suporte é finito e a sua cauda direita será leve; e, por fim, se $\gamma = 0$, referente ao domínio de atração Gumbel, considerado um caso intermédio, a cauda da distribuição será do tipo exponencial, ou seja, mais pesada que as do domínio Weibull mas mais leve que as do domínio Fréchet. Nesta última situação existem distribuições com limite superior do suporte finito ou não finito.

Na secção 4.3 foram apresentados alguns estimadores semi-paramétricos comuns para o EVI, foram eles os estimadores de Hill, de Pickands, dos Momentos, Generalizado de Hill e Mixed-Moment. Serão, portanto, estes os estimadores que serão aplicados à amostra de 77 níveis pluviométricos máximos anuais, na cidade de Barcelos, recolhidos no período de 01/01/1932 a 01/05/2008, com o objetivo de compreender que valores pode assumir o EVI.

Tendo em conta toda a abordagem paramétrica já efetuada anteriormente, sabe-se que as duas grandes possibilidades para o domínio de atração da distribuição subjacente a esta amostra, são o Gumbel e o Weibull. Portanto, a possibilidade de se estar perante uma cauda pesada não faz sentido, logo não será utilizado o estimador de Hill, que apenas se aplica a esse tipo de situações. Na figura 5.69 estão representadas as estimativas para o índice de valores extremos, em função de k , obtidas com recurso aos estimadores já referidos. São também representadas as estimativas intervalares a 95% de confiança. Em cada gráfico encontra-se, também, representada uma reta horizontal, correspondente à estimativa MV, obtida pelo método POT quando considerados os 56 excessos.

Observando a figura, o primeiro aspeto a realçar tem a ver com a maior variabilidade que existe no estimador de Pickands, que não está presente nos restantes. Em alguns deles nota-se alguma instabilidade no início e no final dos gráficos, o que é perfeitamente normal, pois os valores iniciais correspondem a valores de k pequenos, o que origina uma variância elevada no estimador, e os valores finais correspondem a um *threshold* baixo e, como tal, a k elevado o que vai aumentar o viés do estimador. Como se pode observar, pelos valores dados pelos estimadores dos momentos, generalizado de Hill e mixed-moment, existe uma grande semelhança entre estes, verificando-se duas zonas de estabilidade, a primeira para valores de k entre 10 e 25, e a segunda, na zona mais central, essencialmente para k entre 25 e 50, assumindo, predominantemente, valores negativos muito próximos de zero. Aliás, a hipótese de $\gamma = 0$ encontra-se, quase sempre, dentro dos limites de confiança, para cada um dos gráficos. Repare-se, também, que estas estimativas para γ estão muito próximas das estimativas MV obtidas pelos métodos MA, de -0.04967 , e POT, de -0.1259298 . Na figura 5.70, onde se encontram representadas as quatro trajetórias, de cada um dos estimadores, sobrepostas sobre o mesmo gráfico, poder-se-á ver com maior detalhe alguns

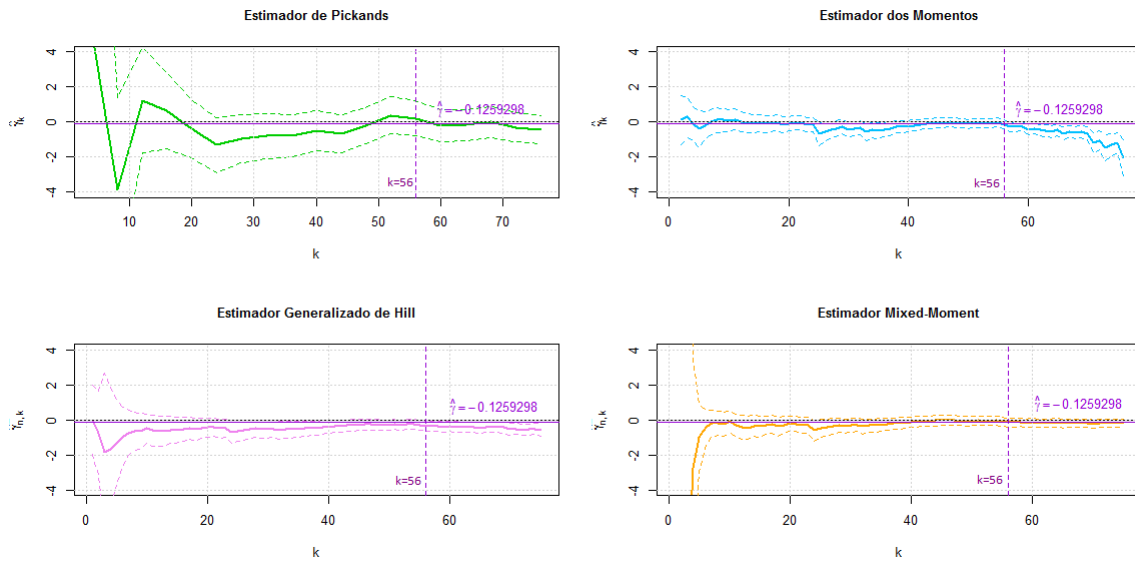


Figura 5.69: *Estimativas do EVI, para a amostra dos níveis pluviométricos máximos anuais*

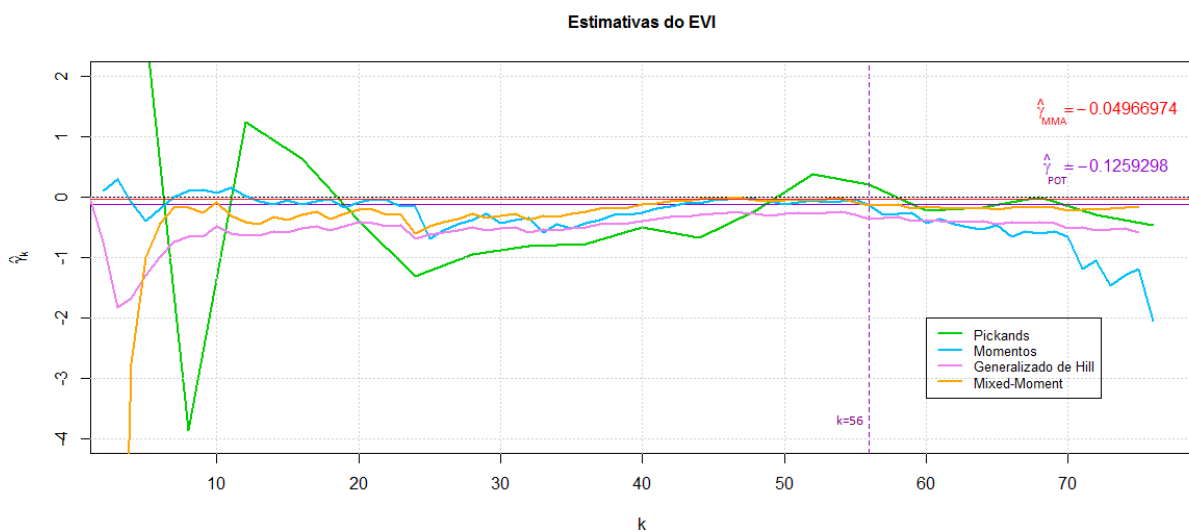


Figura 5.70: *Estimativas do EVI sobre o mesmo gráfico, para a amostra dos níveis pluviométricos máximos anuais*

dos pormenores já referidos. Nesta figura, é perceptível a grande semelhança que existe entre os estimadores dos momentos e o mixed-moment que, aliás, quando consideradas as 56 maiores observações parecem produzir uma estimativa para o EVI muito próxima da obtida pelo método MV, na abordagem POT.

Na tabela 5.47 encontram-se registadas as estimativas pontuais e intervalares para

Tabela 5.47: Estimativas pontuais e intervalares para o EVI, quando $k = 56$, para os níveis pluviométricos máximos anuais

Estimadores de γ	$k = 56$	
	Estimativa	IC a 95%
MV - POT	-0.12593	(-0.315,0.224)
Pickands	0.19931	(-0.79716,1.19578)
Momentos	-0.14173	(-0.39643,0.11297)
Generalizado de Hill	-0.35657	(-0.65869,-0.05444)
Mixed-Moment	-0.14691	(-0.40215,0.10834)

o EVI, obtidas pela abordagem semi-paramétrica, para o caso específico de $k = 56$. Embora a situação $k = 56$ possa colocar em questão as condições assintóticas, uma vez que se pressupõe $\frac{k}{n} \rightarrow 0$, será considerada para se realizar uma comparação com os resultados obtidos no método POT. Recorde-se que na abordagem paramétrica POT, retiveram-se, precisamente, 56 excessos em relação ao *threshold* $u = 62.5$, tendo-se obtido as estimativas MV para o parâmetro de forma, também, apresentadas na tabela 5.47. Como já tinha sido referido, os estimadores dos momentos e mixed-moment forneceram os resultados mais parecidos entre si e com os da metodologia POT. Ambos, assim como o de Pickands incluem a possibilidade do EVI ser nulo, com 95% de confiança, embora a maior parte da amplitude dos intervalos de confiança abranja os valores negativos, o que aponta para distribuições com caudas leves. Quer isto dizer que, a abordagem semi-paramétrica dá a indicação de que o domínio de atração, da distribuição subjacente aos níveis pluviométricos máximos anuais, tem fortes possibilidades ou de ser Gumbel ou Weibull.

II) Estimação de Outros Parâmetros

Tal como foi feito na abordagem paramétrica, também poderão ser estimados outros parâmetros de interesse em acontecimentos raros. Os estimadores para essas quantidades foram apresentados na secção 4.6, e para a sua aplicação, usar-se-á apenas o estimador dos momentos do EVI pois, foi o que forneceu resultados mais semelhantes aos obtidos na abordagem paramétrica, para esse parâmetro. Para além disso, na secção anterior constatou-se que havia a possibilidade do EVI ser nulo ou negativo, como tal serão apresentadas as trajetórias destes parâmetros considerando que $\gamma = 0$ e, o caso mais geral, $\gamma \neq 0$. Em paralelo, será feita a comparação entre as estimativas obtidas pelo método MV, na abordagem POT para os 56 excessos do nível $u = 62.5$.

Na figura 5.71 encontram-se as estimativas para os quantis extremos de probabilidades

0.95, 0.99, 0.999, 0.9999 e 0.99999. Existe uma maior estabilidade nas estimativas, quando considerado que $\gamma \neq 0$, mas que vai diminuindo à medida que aumenta a probabilidade do quantil. No entanto, a trajetória das estimativas, para cada um dos quantis, parece andar em torno da estimativa MV, obtida pelo método POT. Na tabela 5.48 poderão ser recordados os valores obtidos para cada um dos quantis, no método POT, com o método de estimação MV, e encontram-se, também, presentes as estimativas pontuais semi-paramétricas, para $k = 56$. É notória a concordância entre as valores obtidos, nas duas abordagens, quando $\gamma \neq 0$. Por exemplo, destacando os quantis correspondentes aos níveis de retorno de 100, 400 e 1000 anos, verifica-se que, pelo método POT, essas quantidades são 132.9709 mm, 148.1785 mm e 157.7551 mm, respectivamente, assumindo os valores, na abordagem semi-paramétrica, de 133.60463 mm, 148.99108 mm e 157.62201 mm.

Tabela 5.48: *Estimativas pontuais para os quantis extremais ou níveis de retorno, quando $k = 56$, para os níveis pluviométricos máximos anuais*

Quantis/ Níveis Retorno	$k = 56$			
	MV - POT		Semi-Paramétrica	
	$\gamma = 0$	$\gamma \neq 0$	$\gamma = 0$	$\gamma \neq 0$
$\hat{q}_{0.95}$	113.0336	110.8486	121.5900	111.50435
$\hat{q}_{0.99}$	143.4117	132.9709	157.7130	133.60463
$\hat{q}_{0.999}$	186.8729	157.7551	209.3932	157.62201
$\hat{q}_{0.9999}$	230.3342	176.3009	261.0734	174.95188
$\hat{q}_{0.99999}$	273.7954	190.1785	312.7536	187.45634
$\hat{U}(400)$	169.5780	148.7435	188.8276	148.99108

As estimativas referentes a estes quantis extremais, para o caso $\gamma = 0$, como pode ser observado nos gráficos representados na figura 5.71, apresentam um comportamento com uma maior variabilidade. No entanto, cada uma das linhas parece crescer e decrescer em torno de um valor, que será ligeiramente maior do que a estimativa obtida na metodologia POT, quando se assumiu que os excessos tinham um comportamento exponencial. Se se considerar a retenção das 56 maiores observações, verifica-se facilmente que as estimativas dos quantis são maiores na abordagem semi-paramétrica, ou seja, neste caso esta abordagem torna menos raro níveis elevados de precipitação, na cidade de Barcelos.

Usando as expressões (4.31) e (4.32) poderão ser obtidas estimativas para a probabilidade de excedência de um determinado nível pluviométrico elevado, consoante se considere que o EVI é diferente ou igual a zero. Como tem sido sempre feito nesta dissertação, pretende-se obter estimativas da probabilidade de excedência do nível máximo verificado, de 146 mm. Para além disso, será, também, estimado o período de retorno deste nível

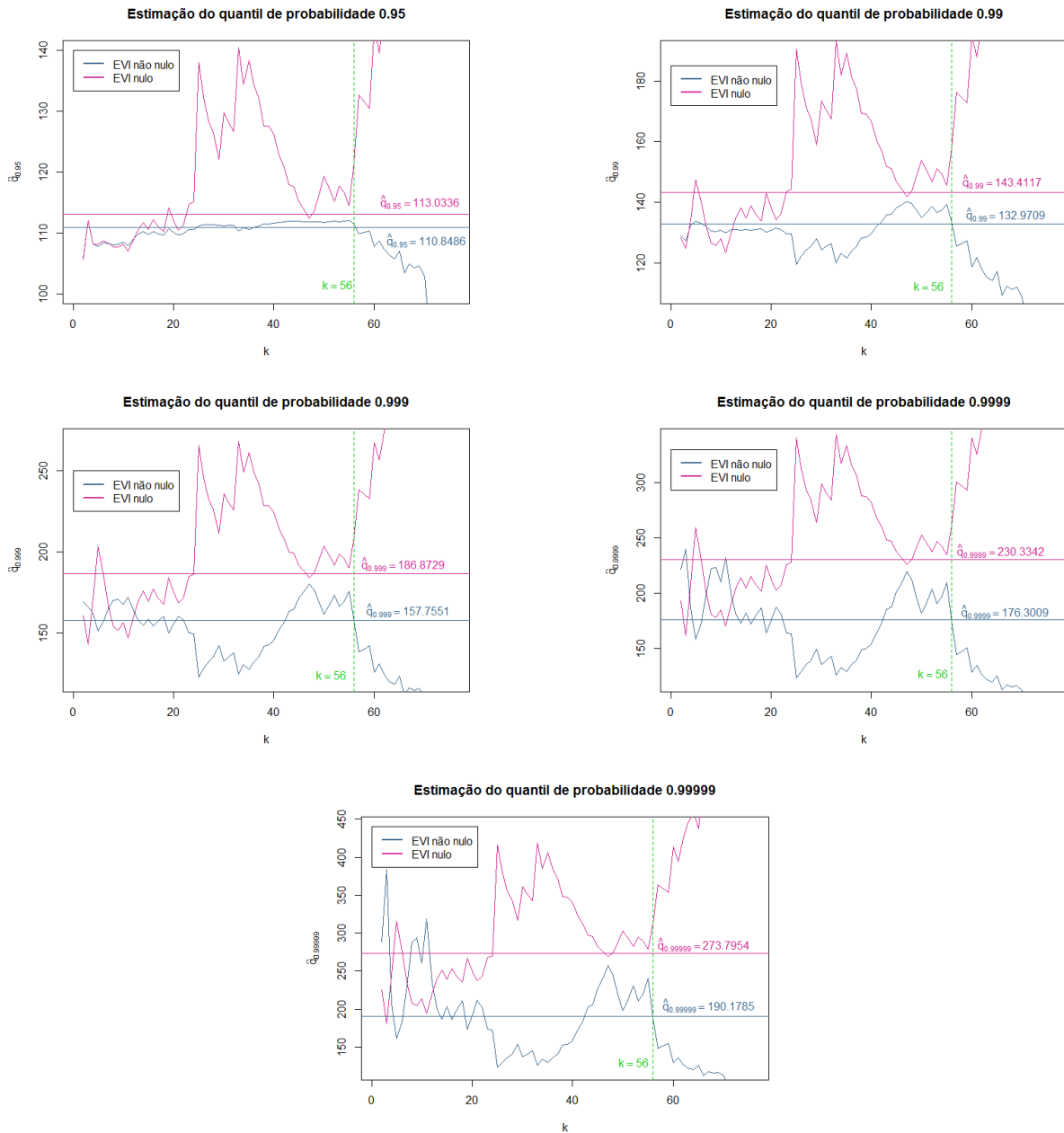


Figura 5.71: Estimativas para os quantis extremos, para a amostra dos níveis pluviométricos máximos anuais

usando a expressão (4.35).

Na figura 5.72 encontram-se representadas as estimativas referidas no parágrafo anterior. Como se pode constatar, as estimativas para a probabilidade de excedência de 146 mm apresentam uma maior estabilidade na situação $\gamma \neq 0$. Para esse caso, a estimativa MV dada pelo método POT, de 0.00324, poderá ser o valor de referência para a trajetória representada nessa figura, pois esta aparenta ter um comportamento oscilatório em torno deste valor. Para o caso $\gamma = 0$, há uma maior variabilidade e as estimativas para esta

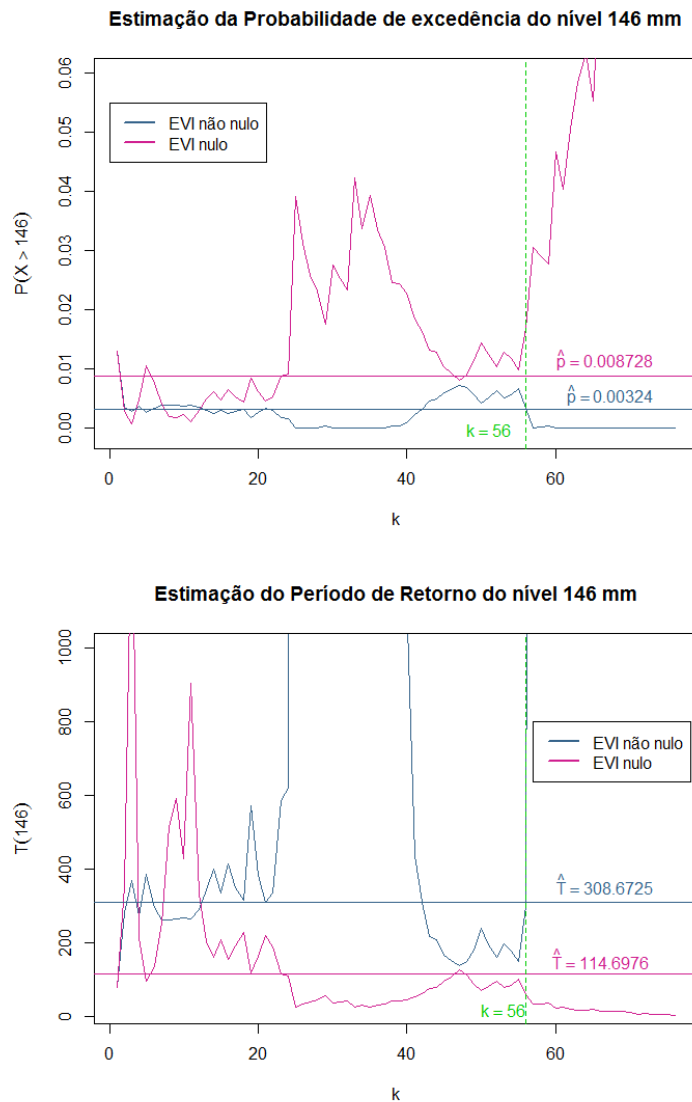


Figura 5.72: Estimativas para a probabilidade de excedência e período de retorno do nível 146 mm, para a amostra dos níveis pluviométricos máximos anuais

probabilidade assumem, naturalmente, valores superiores, visto que se está a considerar que a cauda da distribuição subjacente aos níveis pluviométricos máximos anuais é um pouco mais pesada que no caso anterior, onde o EVI é predominantemente negativo. Para além disso, a curva referente a esses valores em função de k , sugere uma estimativa para a probabilidade de se exceder o nível 146 mm, ligeiramente acima do valor 0.00872, obtido na metodologia POT. Na tabela 5.49 encontram-se as estimativas pontuais para esta quantidade obtidas nos diversos casos, quando retidas as 56 maiores observações. Mais uma vez, é de destacar a semelhança entre os valores obtidos no caso do EVI ser diferente de zero.

Relativamente ao período de retorno, por observação da figura 5.72, existe uma maior

Tabela 5.49: *Estimativas pontuais para o período de retorno e probabilidade de excedência do nível 146 mm, quando $k = 56$, para os níveis pluviométricos máximos anuais*

Período Retorno/ Prob. Excedência	$k = 56$			
	MV - POT		Semi-Paramétrica	
	$\gamma = 0$	$\gamma \neq 0$	$\gamma = 0$	$\gamma \neq 0$
$\hat{T}(146)$	114.6976	308.6725	59.3413259	298.80431
$\hat{P}(X > 146)$	0.00872	0.00324	0.01685	0.00335

estabilidade na curva referente ao caso $\gamma = 0$, e que se revela a partir de um k um pouco acima de 20. Quando comparada com a estimativa MV (método POT), a trajetória das estimativas semi-paramétricas parece assumir valores mais baixos. De facto, quando $k = 56$, espera-se que, em média, uma vez em cada 59 anos, seja ultrapassado o nível pluviométrico 146 mm, e pelo método POT esse período é de, aproximadamente, 115 anos. Para um EVI não nulo, há uma enorme variabilidade nas estimativas obtidas, assumindo valores muito grandes para k a variar entre 20 e 40. Quando $k = 56$, obtém-se a informação de que o tempo de espera, em média, até que haja uma ocorrência de precipitação igual ou superior a 146 mm é de 299 anos, valor muito próximo dos 309 anos obtido no método POT.

Recorde-se que o estimador dos momentos indicou, predominantemente estimativas negativas para o parâmetro de forma da distribuição GEV, logo, neste caso, poderá ser estimado o limite superior do suporte, pois será uma quantidade finita. Na figura 5.73 encontram-se representadas as estimativas deste valor que, como se pode observar, predominantemente são idênticas ao valor máximo observado da amostra, o nível 146 mm.

Nesta mesma figura, estão também representadas as estimativas para este parâmetro, quando considerado o estimador mixed-moment para o EVI. Ora, é visível que este fornece estimativas mais elevadas para o limite superior do suporte, essencialmente a partir de $k = 30$.

Quando foi aplicado o método POT à amostra de 56 excessos, provenientes da amostra dos níveis pluviométricos máximos anuais, a estimativa MV obtida para o limite superior do suporte foi de 231.4343 mm. Neste caso, o estimador mixed-moment fornece uma estimativa mais parecida, igual a 249.2269 mm, quando $k = 56$, embora, com 95% de confiança, o limite superior do suporte poderá assumir qualquer valor do intervalo (198.46213, 275.48835). O estimador dos momentos indica que, para esta situação, já se atingiu esse limite, ou seja, o nível 146 mm, mas podendo tomar qualquer valor entre esse nível e o nível pluviométrico 296.51965 mm, com 95% de confiança.

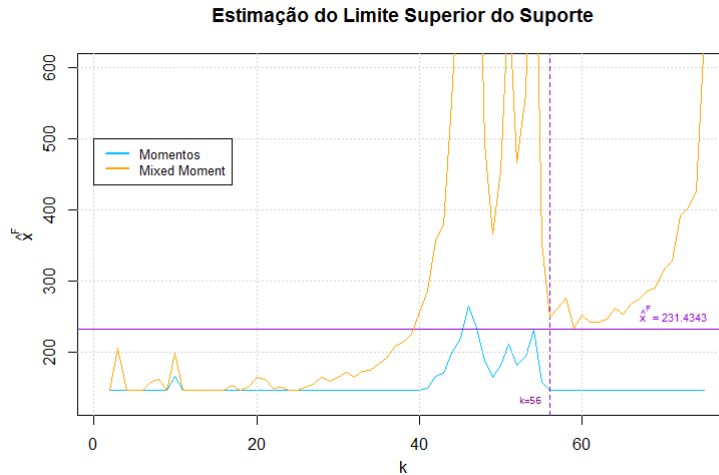


Figura 5.73: Estimativas para o limite superior do suporte, para a amostra dos níveis pluviométricos máximos anuais

III) Testes aos Domínios de Atração

Na secção onde foram apresentadas as estimativas para o EVI, foi referido que haviam dois domínios de atração possíveis para a distribuição subjacente à amostra de níveis pluviométricos máximos anuais, o domínio Gumbel, correspondente a $\gamma = 0$, e o domínio Weibull, no caso $\gamma < 0$. Assim, poder-se-á, efetuar um teste ao domínio Gumbel, recorrendo aos testes apresentados na secção 4.6.1. As hipóteses em estudo são,

$$H_0 : F \in \mathcal{D}(G_0) \quad \text{vs.} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0} \quad (5.19)$$

e poderão ser usadas as versões normalizadas das estatísticas de teste da razão, $T_n^*(k)$, de Greenwood, $R_n^*(k)$, e de Hasofer-Wang, $W_n^*(k)$.

Na figura 5.74 encontram-se representadas as referidas estatísticas, em função de k , bem como os quantis assintóticos de probabilidades 0.025 e 0.975, das distribuições standard Normal e Gumbel, sob a validade da hipótese nula $F \in \mathcal{D}(G_0)$. Recorde-se que para não haver rejeição da hipótese nula, o valor observado das estatísticas de teste deverá estar compreendido entre os quantis mencionados anteriormente. No caso das estatísticas de Greenwood, $R_n^*(k)$, e de Hasofer-Wang, $W_n^*(k)$, usam-se os quantis da distribuição Normal e para a estatística da razão, $T_n^*(k)$, os quantis da distribuição Gumbel.

Como se pode observar pelos gráficos da figura 5.74, o teste de Greenwood, que é essencialmente útil na deteção de distribuições de cauda pesada, não rejeita a hipótese nula para a maioria dos valores de k , havendo a rejeição apenas na parte final, correspondente

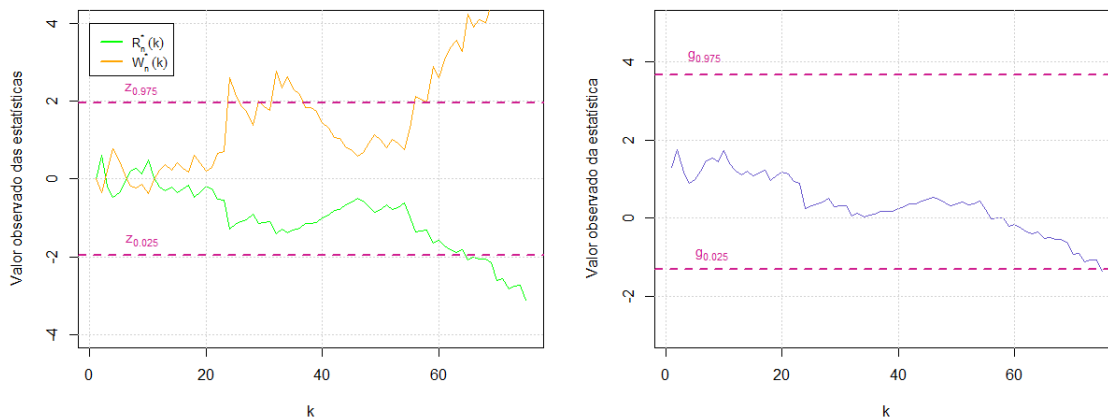


Figura 5.74: Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão bilateral

a valores elevados de k , onde se sabe que já existe algum enviesamento dos estimadores. Quanto ao teste de Hasofer-Wang, mais direcionado para o caso das caudas leves, também não rejeita a hipótese do domínio de atração ser Gumbel na maior parte dos casos. Existe a rejeição, ao nível de significância de 5%, para $k \in \{24, 25\}$, $32 \leq k \leq 36$ e na parte final do gráfico. Quanto ao teste da razão, um teste mais conservativo, a trajetória da estatística encontra-se sempre entre as retas correspondentes aos quantis críticos da Gumbel, logo não há, a 5%, a rejeição da hipótese nula. Portanto, os testes aplicados, para quase todos os valores de k , apontam para a não rejeição da hipótese de que a f.d. F subjacente aos níveis pluviométricos máximos anuais, pertence ao domínio Gumbel. Esta decisão está em conformidade com as conclusões obtidas na abordagem paramétrica realizada a esta amostra.

Apesar de não se rejeitar a hipótese do domínio de atração ser Gumbel, poderá ser efetuado o teste com hipótese alternativa unilateral e, neste caso, referente ao domínio de atração Weibull, ou seja,

$$H_0 : F \in \mathcal{D}(G_0) \quad \text{vs.} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma < 0} \quad (5.20)$$

Na figura 5.75 estão representadas os valores observados das diferentes estatísticas, para cada valor de k . No caso da estatística de Greenwood, a hipótese nula será rejeitada, ao nível de significância de 5%, se $R_n^*(k) < z_{0.05}$. Ora, por observação gráfica a rejeição apenas se dá para valores de k superiores a 60, ou seja, para um valor já muito elevado de k , tendo em conta a dimensão da amostra. Para a estatística de Hasofer-Wang, a

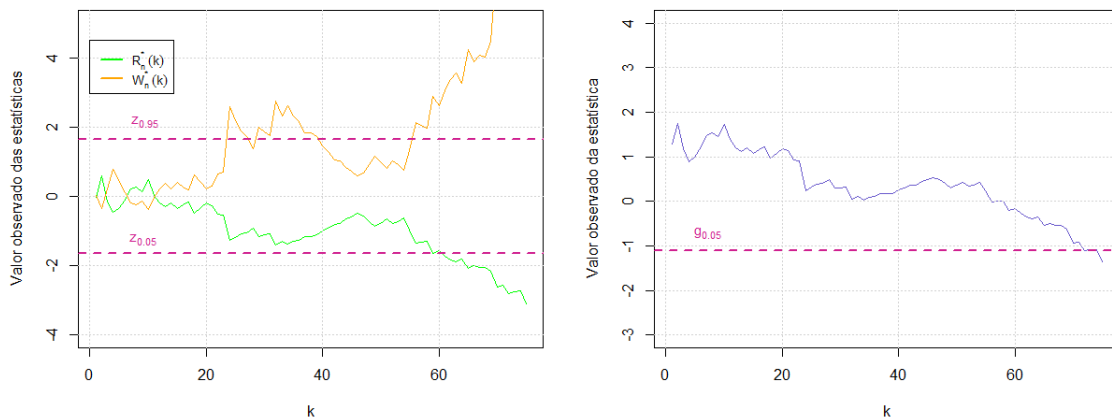


Figura 5.75: Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão unilateral

região de rejeição é dada por $W_n^*(k) > z_{0.95}$. Assim, na região central, aproximadamente para valores de k entre 24 e 39, existe rejeição da hipótese nula, e, portanto, este teste indica que a distribuição subjacente a estes dados pertence ao domínio das caudas leves, o domínio Weibull. No entanto, nesse intervalo, o valor observado da estatística não se afasta muito do valor do quantil $z_{0.95}$ e, aliás para $k = 28$ não há rejeição de H_0 . Na parte final, a trajetória da estatística volta a ultrapassar o quantil crítico, obtendo-se, novamente, a decisão de rejeição da hipótese nula. Por fim, a estatística da razão, onde haverá a rejeição da hipótese nula se $T_n^*(k) < g_{0.05}$. Como se pode observar no gráfico da direita, na figura 5.75, a trajetória da estatística encontra-se acima do quantil referido, à exceção da parte final, o que não é relevante. Sendo assim, usando esta estatística para as hipóteses (5.20), não se rejeita que se está num caso de distribuição pertencente ao domínio de atração Gumbel.

Numa panorâmica geral, na maior parte dos casos, não existe evidência estatística para não se aceitar o domínio Gumbel como o domínio max-estável ao qual pertence a distribuição subjacente à amostra dos níveis pluviométricos máximos anuais, mesmo que, na secção anterior, quando se estimaram as diversas quantidades relevantes num estudo em acontecimentos raros, haver uma maior estabilidade nos valores, para o caso $\gamma \neq 0$.

5.3.2 Níveis Pluviométricos Máximos Mensais

I) Estimação do EVI

Pretende-se, nesta secção, realizar um estudo semelhante ao do ponto anterior mas, neste caso, com a amostra de 910 níveis pluviométricos máximos mensais, registados no período de 01/01/1932 a 01/05/2008, na cidade de Barcelos.

Pelo estudo paramétrico já realizado com esta amostra, sabe-se que o domínio de atração favorável para a distribuição subjacente a estes dados é o Gumbel, mas existe, também, a possibilidade do domínio Weibull ser adequado. Ou seja, tem-se em mão uma distribuição com cauda, essencialmente, leve, sendo assim, não será usado o estimador de Hill para o cálculo do EVI, uma vez que este apenas se adequa para casos de caudas pesadas, que pertencem ao domínio Fréchet. Na figura 5.76 poderão ser observadas as estimativas para o EVI, em função do número de maiores observações retidas, k , dadas pela utilização dos estimadores de Pickands, dos momentos, generalizado de Hill e mixed-moment. Todos os gráficos parecem não deixar margem para dúvidas de que a estimativa para o EVI será zero ou algum valor muito próximo da origem. O pormenor de como as trajetórias dos quatro estimadores se encontram em torno da origem, poderá ser visível com maior detalhe na figura 5.77. Como habitual, o estimador de Pickands apresenta a maior variabilidade e os estimadores dos momentos, generalizado de Hill e mixed-moment apresentam um comportamento muito semelhante até, sensivelmente, ao valor de $k = 371$ correspondente ao número de excessos retidos aquando a abordagem POT. Após esse valor de k , as curvas correspondentes aos estimadores generalizado de Hill e mixed-moment continuam a apresentar um comportamento muito semelhante, embora os valores obtidos pelo último estimador sejam sensivelmente maiores. Quanto aos estimadores dos momentos e de Pickands, para $k > 371$, ambos apresentam um padrão de decrescimento, fornecendo estimativas negativas para o EVI, dando, assim, a sensação de que a cauda da distribuição subjacente a estes níveis pluviométricos fica mais leve.

Na tabela 5.50 são apresentadas as estimativas pontuais e os intervalos com 95% de confiança para o EVI, obtidos com os estimadores semi-paramétricos já referidos, na situação particular de $k = 371$, e os obtidos na metodologia POT. As estimativas, à exceção da de Pickands, estão muito próximas de zero, valor esse que, por sua vez, pertence a qualquer um dos intervalos de confiança. Há a realçar a grande proximidade entre as estimativas pontuais obtidas na abordagem POT e pelo estimador dos momentos, e a grande semelhança entre os limites dos intervalos de confiança para estas duas situações.

Portanto, numa perspetiva geral, o EVI assume valores muito próximos de zero, o que indica que a distribuição subjacente aos níveis pluviométricos máximos mensais, em Barcelos, tem uma forte possibilidade de pertencer ao domínio max-estável Gumbel. Mas, como na secção seguinte será usado o estimador dos momentos, para se obter outras quantidades relevantes em valores extremos, então, dando um especial destaque a este

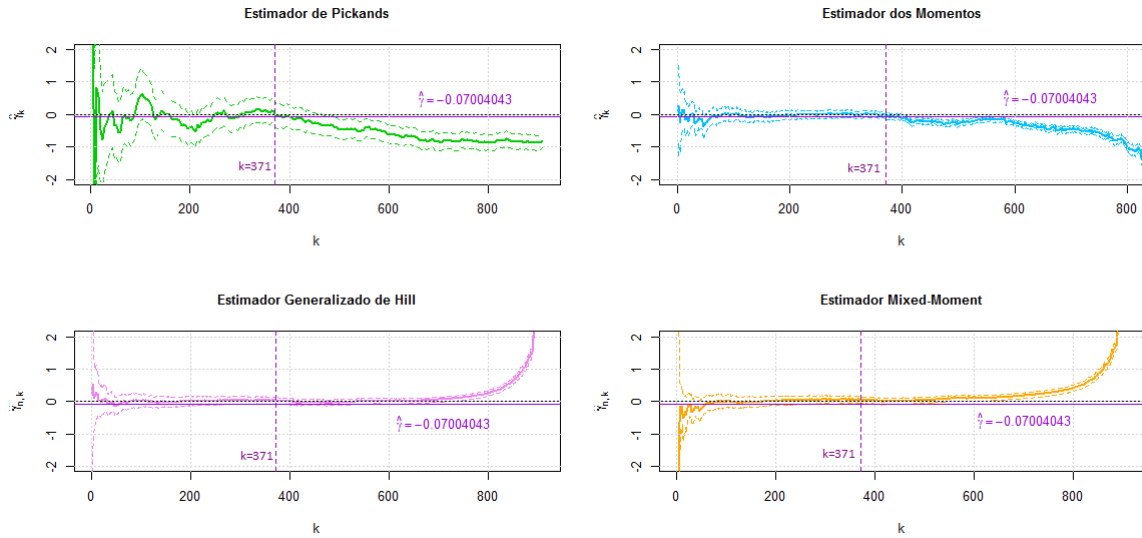


Figura 5.76: *Estimativas do EVI, para a amostra dos níveis pluviométricos máximos mensais*

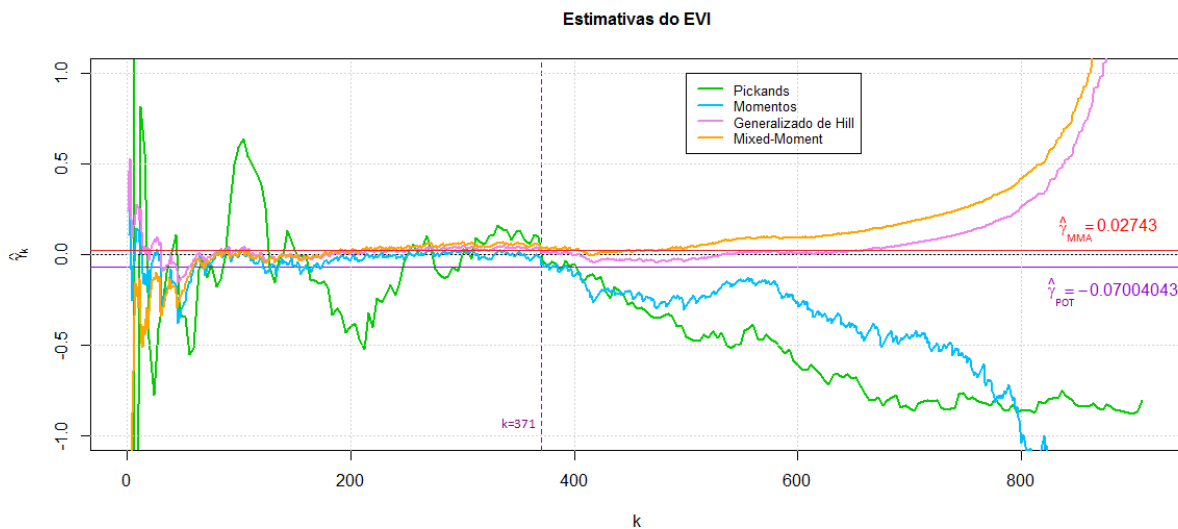


Figura 5.77: *Estimativas do EVI sobre o mesmo gráfico, para a amostra dos níveis pluviométricos máximos mensais*

estimador, o domínio de atração Weibull é também outro grande candidato. Na última secção deste capítulo, proceder-se-á à aplicação de testes estatísticos, com vista a uma escolha acertada do domínio de atração.

Tabela 5.50: Estimativas pontuais e intervalares para o EVI, quando $k = 371$, para os níveis pluviométricos máximos mensais

Estimadores de γ	$k = 371$	
	Estimativa	IC a 95%
MV - POT	-0.07004	(-0.1515,0.0385)
Pickands	-0.39939	(-0.03291,0.33356)
Momentos	-0.05994	(-0.15809,0.03820)
Generalizado de Hill	0.01279	(-0.09026,0.11586)
Mixed-Moment	0.03736	(-0.0682,0.14291)

II) Estimação de Outros Parâmetros

Como acabou de ser referido na secção anterior, será usado o estimador dos momentos para a obtenção de estimativas dos quantis extremais, níveis e períodos de retorno, probabilidade de excedência e limite superior do suporte. Serão considerados dois casos, $\gamma = 0$ e $\gamma \neq 0$, para cada uma das quantidades, à exceção, do limite superior do suporte. Tal como foi feito com a amostra dos níveis pluviométricos máximos anuais, será feita uma comparação entre os resultados obtidos nesta abordagem semi-paramétrica com os fornecidos pelo método de estimação MV, na metodologia POT.

Na figura 5.78 poderão ser observadas as estimativas para os quantis extremais de probabilidades 0.95, 0.99, 0.999, 0.9999 e 0.99999, em função do valor de k . Para valores de k compreendidos entre, aproximadamente, 200 e 371 as estimativas para cada um dos quantis, quer quando $\gamma = 0$ quer com $\gamma \neq 0$, são muito semelhantes. As trajetórias, para esses valores de k , andam em torno das estimativas MV de aproximadamente 73 mm, para o quantil de probabilidade 0.95, e para os outros quantis, em geral, encontram-se entre as duas estimativas MV, mas tendencialmente mais próximas da estimativa obtida no método POT, com $\gamma = 0$. Para $k \in (120, 170)$, acontece a situação curiosa das duas estimativas semi-paramétricas, para cada um dos quantis, estarem em consonância com as respetivas estimativas MV, obtidas no método POT.

Para a situação específica $k = 371$, correspondente ao número de excessos considerados na abordagem POT, foram registadas na tabela 5.51, as estimativas dos diversos quantis extremais. Para cada caso, $\gamma = 0$ e $\gamma \neq 0$, as estimativas obtidas nas metodologias POT e semi-paramétrica são muito semelhantes, embora haja uma maior proximidade entre os valores quando $\gamma \neq 0$. Este pormenor não vem contradizer o que foi referido no parágrafo anterior, pois, se se observar mais atentamente a figura 5.78, aproximadamente em $k = 371$, começa-se a evidenciar um afastamento das duas linhas referentes às estimativas semi-paramétricas, através de um decrescimento da trajetória para $\gamma \neq 0$ e um crescimento no

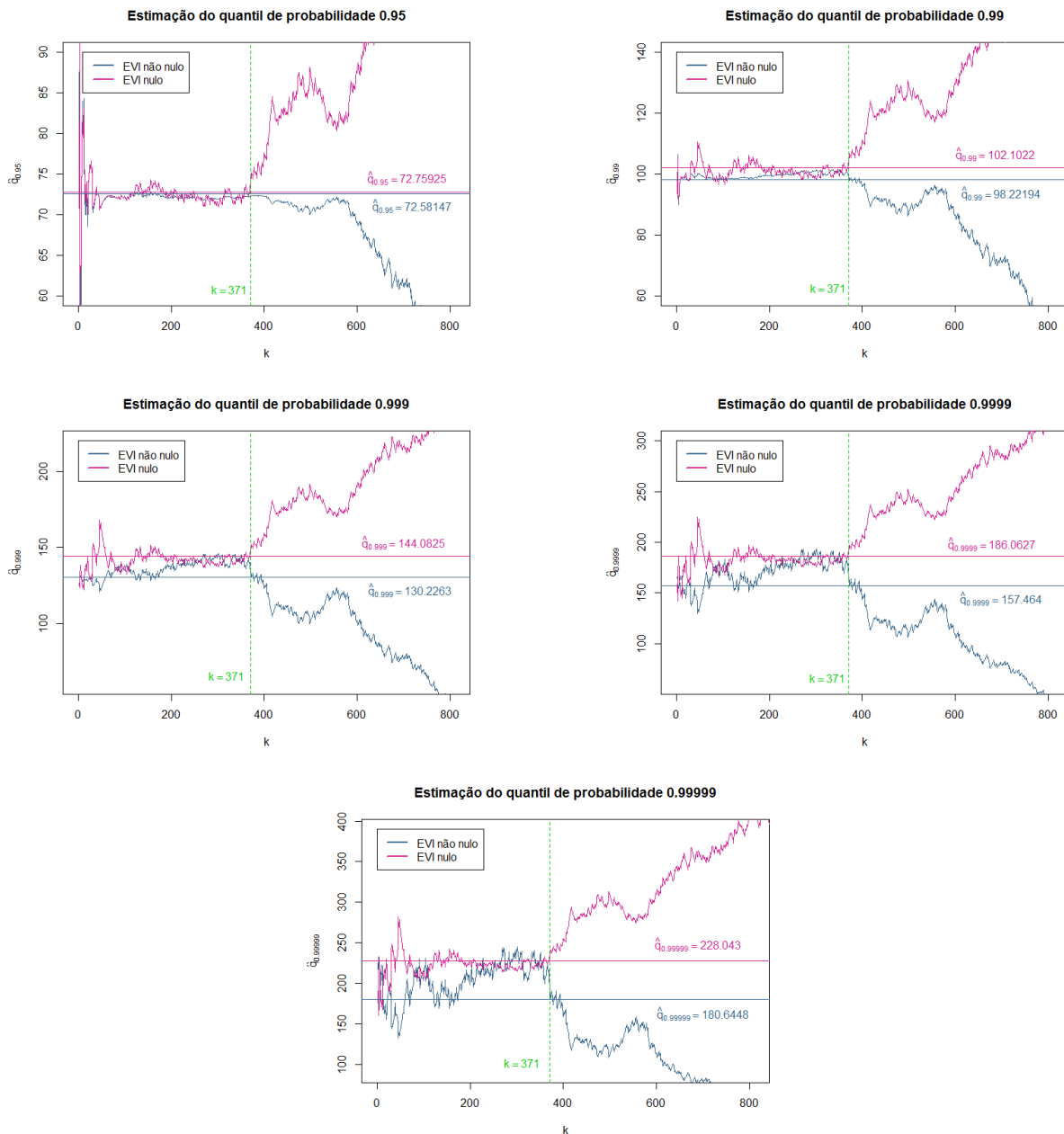


Figura 5.78: Estimativas para os quantis extremos, para a amostra dos níveis pluviométricos máximos mensais

outro caso, ou seja, quando considerado um EVI nulo.

Na tabela 5.51 encontram-se também as estimativas, para o caso particular de $k = 371$, dos quantis referentes aos níveis de retorno de 100, 400 e 1000 anos, ou equivalentemente, 1200, 4800 e 12000 meses. A situação é análoga ao que já foi mencionado para os outros quantis e poderá ser comprovado por observação da figura 5.79, onde se apresentam os valores fornecidos pelo estimador dos momentos para estes níveis de retorno. Como já foi mencionado, sensivelmente a partir do valor $k = 371$, as duas trajetórias seguem caminhos

Tabela 5.51: Estimativas pontuais para os quantis extremais e níveis de retorno, quando $k = 371$, para os níveis pluviométricos máximos mensais

Quantis/ Níveis Retorno	$k = 371$			
	MV - POT		Semi-Paramétrica	
	$\gamma = 0$	$\gamma \neq 0$	$\gamma = 0$	$\gamma \neq 0$
$\hat{q}_{0.95}$	72.75925	72.58147	74.73951	72.29344
$\hat{q}_{0.99}$	102.1022	98.22194	105.83131	98.42925
$\hat{q}_{0.999}$	144.08247	130.22634	150.31361	131.69779
$\hat{q}_{0.9999}$	186.06274	157.46397	194.79592	160.67731
$\hat{q}_{0.99999}$	228.04301	180.64480	239.27823	185.92078
$\hat{U}(1200)$	147.4065	132.5468	153.83578	134.14104
$\hat{U}(4800)$	172.6812	149.2538	180.61680	151.87035
$\hat{U}(12000)$	189.3868	159.4388	198.31809	162.80558

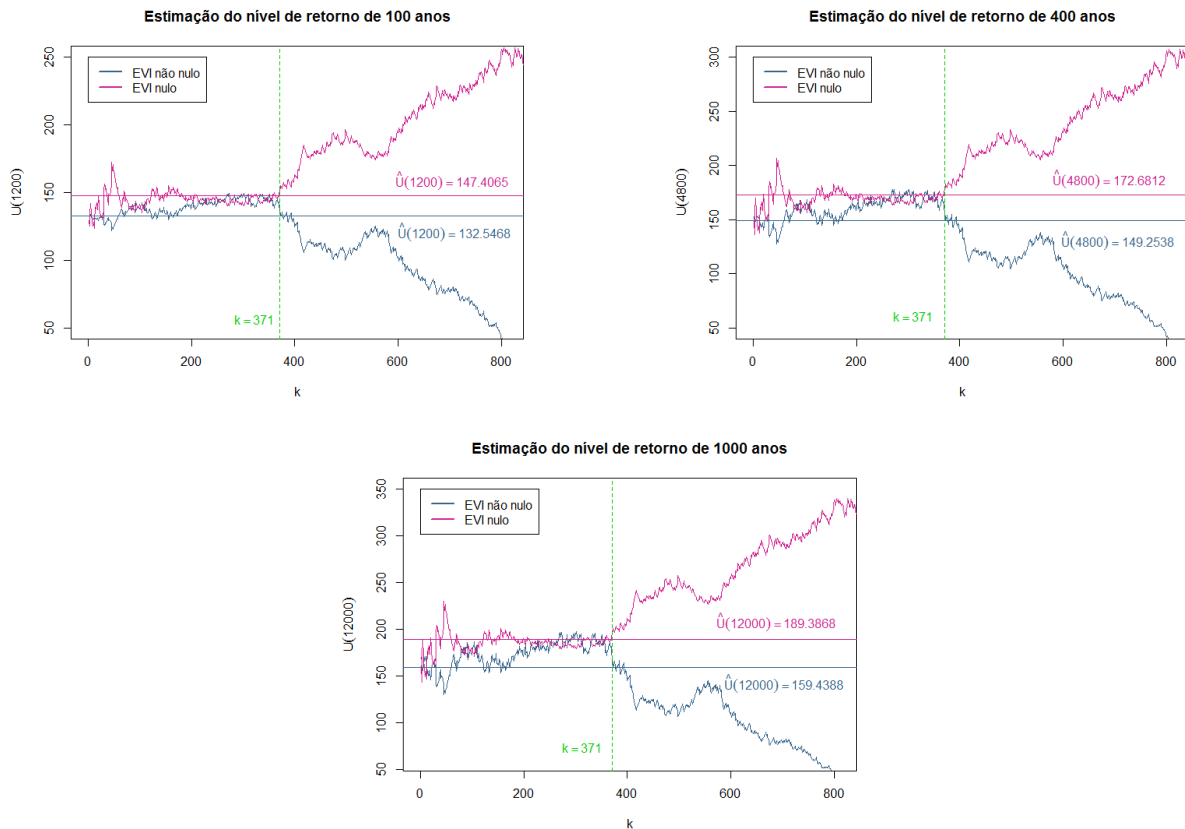


Figura 5.79: Estimativas para os níveis de retorno, para a amostra dos níveis pluviométricos máximos mensais

diferentes, uma aumentando e a outra decrescendo, para cada um dos quantis e níveis de retorno. No entanto, repare-se que para valores de k entre, aproximadamente, 400 e 600, cada uma das linhas tem um momento de quebra nesse crescimento e decrescimento, e

parece fixarem-se em torno de um determinado valor. Em particular, para os níveis de retorno de 100, 400 e 1000 anos, no caso $\gamma = 0$, esses valores estão muito próximos de 185 mm, 215 mm e 235 mm, respectivamente. Já para $\gamma \neq 0$, esses níveis de retorno são de 110 mm, 120 mm e 130 mm, respectivamente.

Na figura 5.80 estão representadas as estimativas para o período de retorno e probabilidade de excedência do nível pluviométrico 146 mm. Relativamente à probabilidade de

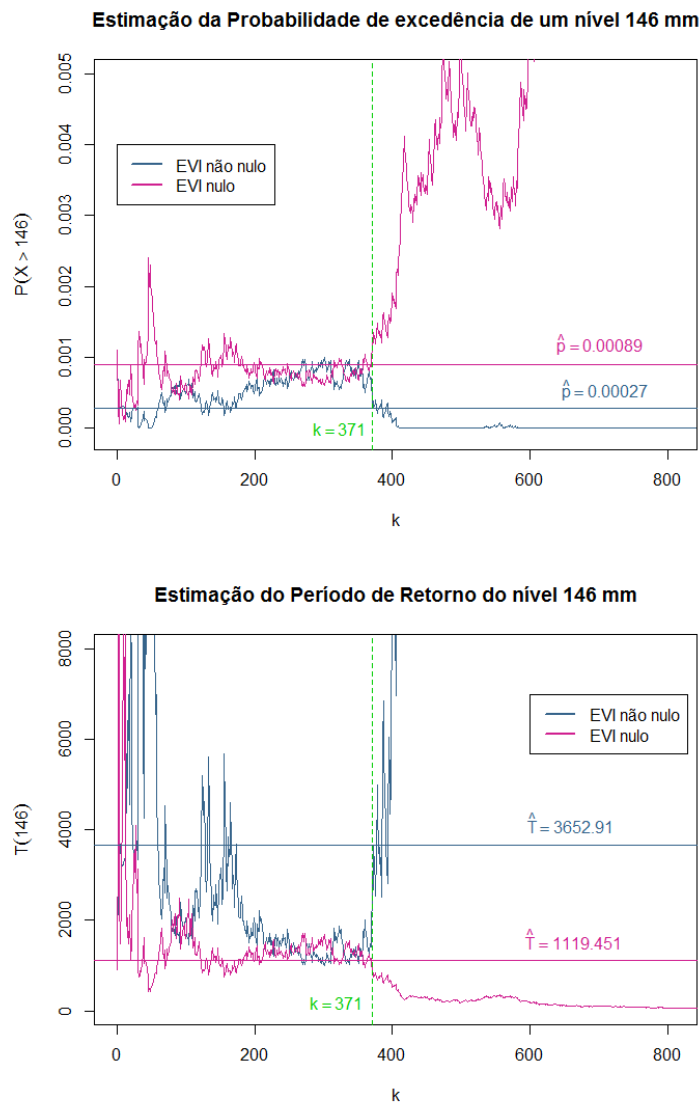


Figura 5.80: *Estimativas para a probabilidade de excedência e período de retorno do nível 146 mm, para a amostra dos níveis pluviométricos máximos mensais*

excedência do nível 146 mm, é notório um comportamento muito semelhante das trajetórias quando comparadas com as referentes aos quantis extremos. Para valores de k até 371, as estimativas para esta probabilidade encontram-se, essencialmente, entre os valores de 0.00027 e 0.00089, correspondentes às estimativas obtidas pelo método POT, quando

$\gamma \neq 0$ e $\gamma = 0$, respectivamente. No entanto, verifica-se uma maior proximidade das estimativas, para este intervalo de k , com o valor 0.00089. Para valores de k superiores a 371, as duas trajetórias seguem caminhos divergentes. A curva referente ao caso $\gamma = 0$, inicia um processo de decrescimento até ao zero e aí fica estabilizada. Quanto à linha do caso $\gamma = 0$, começa por apresentar um crescimento e, aproximadamente, para k entre 400 e 600, as estimativas parecem estabelecer uma nova probabilidade de excedência por volta do valor de 0.004. Na tabela 5.52 poderão ser observadas as estimativas pontuais da probabilidade de excedência de 146 mm, para o caso concreto de $k = 371$.

Quanto às estimativas para o período de retorno do nível 146 mm, existe uma maior regularidade na curva correspondente ao caso $\gamma = 0$, que, como se pode observar na figura 5.80, para valores de k inferiores a 371, encontra-se muito próxima da estimativa fornecida pelo método POT, de 1119.451 meses ou, equivalentemente, 93.28758 anos. Já para o caso $\gamma \neq 0$, poderão ser identificados dois momentos para $k < 371$. O primeiro, no intervalo (50, 180), a trajetória tem um percurso em volta da estimativa dada pela metodologia POT, de 3652.91 meses, que em anos corresponde a 304.4092. O segundo momento acontece para um k a variar entre 180 e 371, onde a trajetória apresenta um decrescimento, encontrando-se muito mais próxima da estimativa obtida no método POT, com $\gamma = 0$. Para valores de k acima de 371, as estimativas para o período de retorno, quando considerado o domínio Gumbel, são decrescentes e, sensivelmente, para $k \in (400, 600)$, estabilizam num valor perto de 270 meses, o correspondente a 22.5 anos. Para $\gamma \neq 0$ o gráfico das estimativas é crescente.

Tabela 5.52: *Estimativas pontuais para o período de retorno e probabilidade de excedência do nível 146 mm, quando $k = 371$, para os níveis pluviométricos máximos mensais*

Período Retorno/ Prob. Excedência	$k = 371$			
	MV - POT		Semi-Paramétrica	
	$\gamma = 0$	$\gamma \neq 0$	$\gamma = 0$	$\gamma \neq 0$
$\hat{T}(146)$	1119.451 meses 93.28758 anos	3652.91 meses 304.4092 anos	799.88266 meses 66.656888 anos	2994.3474 meses 249.52895 anos
$\hat{P}(X > 146)$	0.00089	0.00027	0.00125	0.00033

Na tabela 5.52 encontram-se as estimativas para o período de retorno do nível pluviométrico de 146 mm, na situação específica $k = 371$. Apesar de uma diferença de 26 anos, a maior semelhança acontece quando o EVI é nulo, no entanto, para ambos os casos, as estimativas semi-paramétricas indicam um menor tempo de espera para um nível de precipitação superior ou igual a 146 mm.

Na figura 5.81 estão representadas as estimativas para o limite superior do suporte.

Para k a variar, essencialmente, entre os valores 200 e um pouco antes de 371, obtêm-se estimativas, aproximadamente, entre 150 mm e 240 mm. Contudo, como pode ser observado, a maior parte da trajetória indica que o nível máximo de precipitação em Barcelos, já foi atingido com o nível de 146 mm. Este estimador não se revelou, portanto, muito eficaz. Em relação à estimativa MV obtida para este parâmetro, na metodologia POT, de 313.1014 mm, está bastante acima dos valores obtidos com o estimador semi-paramétrico dos momentos. Para $k = 371$, obtêm-se $\hat{x}^F = 146$, podendo, no entanto,

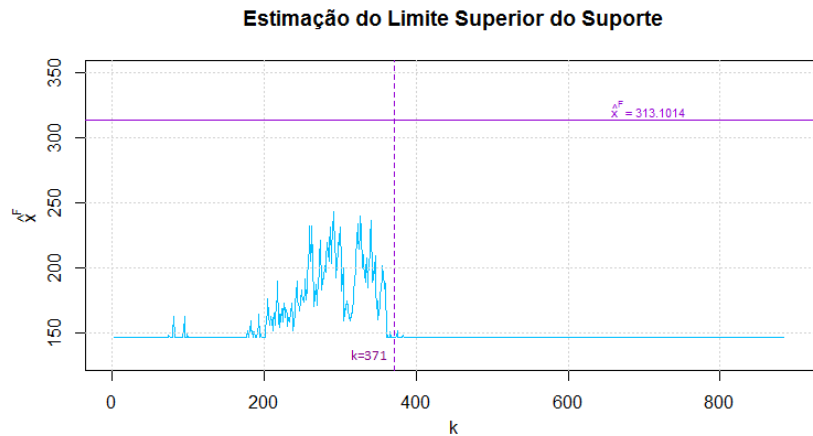


Figura 5.81: Estimativas para o limite superior do suporte, para a amostra dos níveis pluviométricos máximos mensais

atingir com 95% de confiança, qualquer valor entre 146 mm e 370.582 mm.

III) Testes aos Domínios de Atração

Quando foram calculadas as estimativas do EVI, através dos estimadores de Pickands, dos momentos, generalizado de Hill e mixed-moment, foi evidenciado que estas estavam muito próximas de zero e, por esse motivo, a distribuição dos níveis pluviométricos máximos mensais tem uma forte possibilidade de pertencer ao domínio de atração Gumbel. No entanto, as várias estimativas vão alternando entre valores positivos e negativos e, por isso, será conveniente aplicar-se um teste de hipóteses aos domínios de atração, em especial ao domínio Gumbel. Assim, para as hipóteses,

$$H_0 : F \in \mathcal{D}(G_0) \quad \text{vs.} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma \neq 0} \quad (5.21)$$

serão usadas as estatísticas normalizadas da razão, $T_n^*(k)$, de Greenwood, $R_n^*(k)$, e de Hasofer-Wang, $W_n^*(k)$. Na figura 5.82 estão representadas essas estatísticas, em função do

valor de k e os quantis críticos associados ao nível de significância de 0.05, das distribuições standard Normal e Gumbel.

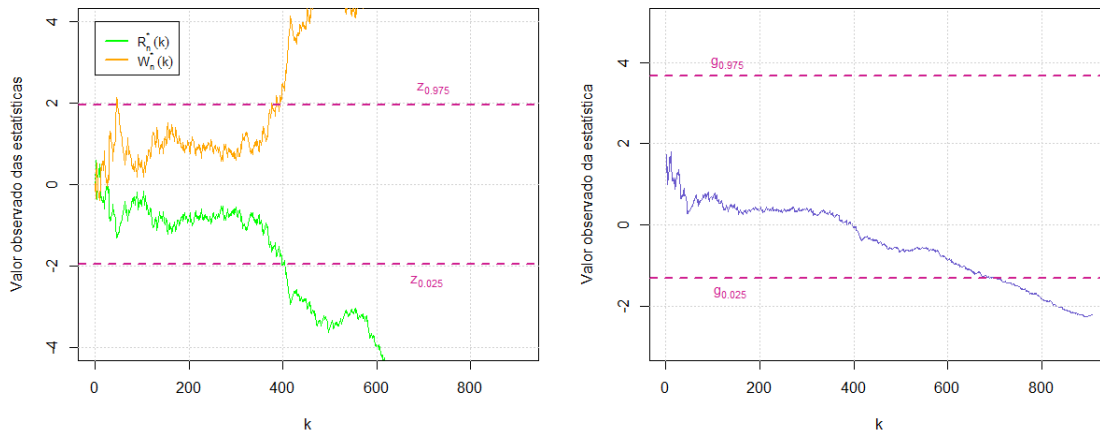


Figura 5.82: Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão bilateral

Como se pode observar, para as estatísticas de Greenwood e de Hasofer-Wang não se rejeita a hipótese nula, definida em (5.21), ao nível de significância de 5%, para valores de k inferiores a 400. Para a estatística da razão, só existe rejeição, ao mesmo nível de significância, quando $k > 700$, ou seja, já para uma quantidade elevada de observações retidas.

Portanto, com os testes de Greenwood e de Hasofer-Wang há a rejeição do Domínio Gumbel, se se considerarem mais de 400 observações da amostra original de níveis pluviométricos máximos mensais. Quer isto dizer que, para esta situação, com uma significância de 5%, existe evidência estatística de que o EVI é diferente de zero. Como, a abordagem paramétrica, efetuada a esta amostra, apontou o domínio de atração Weibull como outro potencial candidato e, já na abordagem semi-paramétrica, alguns estimadores para o EVI, mais propriamente os estimadores dos momentos e de Pickands, para $k \geq 400$, e o estimador generalizado de Hill, para $400 \leq k \leq 550$, indicaram alguma preferência por estimativas negativas para este parâmetro, serão, também, colocadas em teste as hipóteses,

$$H_0 : F \in \mathcal{D}(G_0) \quad \text{vs.} \quad H_1 : F \in \mathcal{D}(G_\gamma)_{\gamma < 0}. \quad (5.22)$$

Pela figura 5.83, para valores de k a partir, sensivelmente, de 400, o valor observado da estatística de Greenwood é inferior ao quantil da Normal standard $z_{0.05}$, e o referente à estatística de Hasofer-Wang é superior ao quantil $z_{0.95}$, logo, em ambos os casos, há

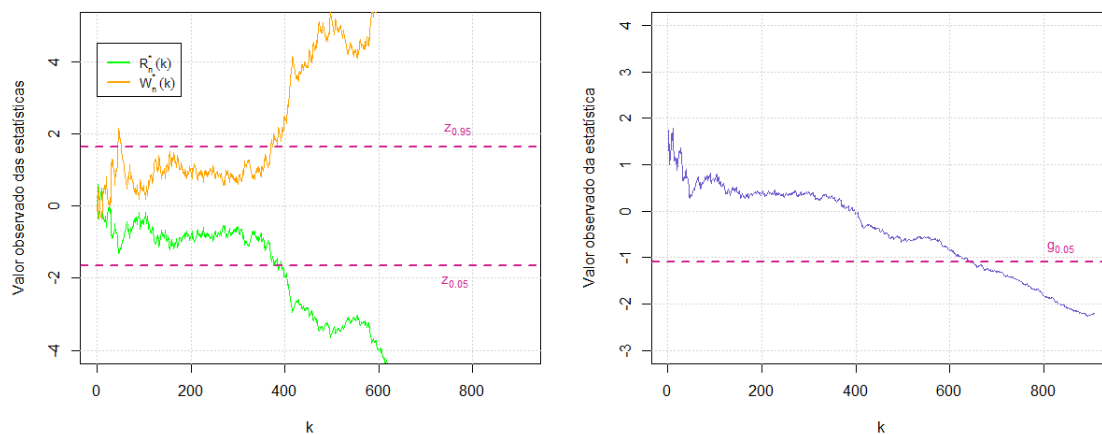


Figura 5.83: Trajetória das estatísticas de teste $R_n^*(k)$ e $W_n^*(k)$ (figura da esquerda) e $T_n^*(k)$ (figura da direita), versão unilateral

a rejeição da hipótese nula. Sendo assim, ao nível de significância de 5%, poder-se-á considerar que, quando são retidos os k maiores níveis pluviométricos, com $k \geq 400$, a distribuição subjacente aos dados pertence ao domínio de atração Weibull. Quanto à estatística da razão, a rejeição da hipótese referente ao domínio Gumbel, só acontece para valores de k superiores a 650, uma vez que o valor observado da estatística é inferior ao quantil $g_{0.05}$.

Portanto, estas conclusões vêm ao encontro dos valores obtidos anteriormente, aquando a estimação dos diversos parâmetros, em função de k . Em geral, para valores de k inferiores a 400, as estimativas semi-paramétricas apresentavam uma maior proximidade à estimativa MV, da metodologia POT, com $\gamma = 0$. De facto, lembre-se que para o *threshold* de 34.5 mm, tinham-se 371 excessos, valor esse que pertence ao sub-conjunto indicado para k . Para além disso, as estimativas dos vários parâmetros apresentaram, tendencialmente, uma maior estabilidade para $k < 371$ (que está contido em $k < 400$), logo será razoável aceitar-se que a f.d. F subjacente aos níveis pluviométricos máximos mensais, faz parte do domínio de atração Gumbel.

Assim, tendo em conta que os testes estatísticos aplicados não rejeitam a hipótese do domínio de atração ser Gumbel para, sensivelmente, $k < 400$ e os estimadores dos vários parâmetros apresentarem, na sua maioria, uma zona de estabilidade, também para $k < 400$, é sensato aceitar que os valores estimados, considerando $\gamma = 0$, terão uma maior credibilidade.

Um outro pormenor curioso e que pode ser observado ao longo das várias estimativas apresentadas, para os diversos parâmetros de interesse em acontecimentos raros, tem a

ver com o facto de existirem grandes diferenças nessas estimativas, quando considerados os casos $\gamma = 0$ e $\gamma \neq 0$, mesmo quando o valor de γ , na segunda situação, está tão próximo de zero.

Capítulo 6

Considerações finais e problemas em aberto

Nesta dissertação foi realizada uma análise estatística em valores extremos a níveis pluviométricos ocorridos na cidade de Barcelos, uma cidade portuguesa localizada numa região que, habitualmente, apresenta os mais altos registos de pluviosidade anual do país. Pretendia-se com esta análise identificar, num primeiro momento, o domínio max-estável ao qual pertence a distribuição destes níveis e, feita essa identificação, poder-se-ia, de seguida, proceder à estimação dos parâmetros do modelo em questão e, conseqüentemente, fazer inferência sobre quantidades tão importantes como os quantis extremos, os níveis de retorno de determinado número de anos, os períodos de retorno, etc, que são fundamentais na previsão e prevenção de catástrofes associadas à ocorrência de níveis pluviométricos elevados.

Recorde-se que a amostra original disponível era constituída por registos diários de precipitação, de 1932 a 2008, e, a partir daqui, foram construídas outras sub-amostras, nomeadamente, com os níveis máximos anuais, os níveis máximos mensais e os níveis máximos anuais por estação do ano, com vista à aplicação de diferentes abordagens em inferência estatística em valores extremos, como a metodologia Gumbel com os máximos extraídos por blocos, o método da maiores observações, o método POT dos excessos de um nível e a metodologia PORT dos excessos de nível aleatório.

As metodologias aplicadas e apresentadas nesta dissertação apontaram para um índice de valores extremos menor ou igual a zero, indicando, assim, que a distribuição subjacente aos níveis pluviométricos máximos anuais e mensais poderá ser Gumbel ou Weibull. Os testes de hipóteses utilizados indicaram uma preferência por um EVI nulo, remetendo, assim, a distribuição dos níveis pluviométricos para a família Gumbel. Ora tal situação

vem coincidir com o que é considerado na literatura, pois, de um modo geral, a distribuição Gumbel é o modelo mais frequentemente utilizado para tratar extremos hidrológicos.

Quando condicionados os dados à estação do ano, neste caso apenas ao outono e inverno, e aplicado o método dos máximos anuais, o modelo Gumbel revelou-se, também aqui, o mais adequado. Na presente dissertação não foram apresentadas as análises POT e semi-paramétrica, que também foram realizadas com estas amostras, para cada uma das duas estações do ano. Para os níveis pluviométricos máximos anuais no outono, a metodologia POT apontou para um EVI negativo (estimativa MV: $\hat{\gamma} = -0.67214$) e os testes aplicados foram unânimes na decisão de rejeição da hipótese nula, $\gamma = 0$, aceitando-se, assim, uma distribuição com cauda leve, pertencente ao domínio de atração Weibull. Para o inverno, a estimativa do EVI obtida está muito próxima da origem (estimativa MV: $\hat{\gamma} = -0.07152683$) e a hipótese da distribuição dos excessos ser exponencial não é rejeitada. Logo, neste caso, e mais uma vez, o domínio de atração Gumbel revelou-se o mais adequado.

Relativamente à abordagem semi-paramétrica para as duas estações, esta revelou resultados que reforçam as conclusões obtidas com a metodologia POT. Aliás, este pormenor foi, também, verificado nas duas análises semi-paramétricas realizadas e apresentadas nesta dissertação. Existe uma clara afinidade entre a abordagem semi-paramétrica e o método paramétrico POT. As várias quantidades calculadas revelaram-se bastante semelhantes, nas duas metodologias, facto esse evidenciado nas diversas figuras apresentadas, e onde era nítida a existência de regiões onde as trajetórias das estimativas semi-paramétricas, apresentavam um padrão estável em torno das estimativas obtidas pelo método POT. Perante tal situação, é evidente que tomar uma atitude rígida perante as duas abordagens e considerar que, cada uma delas, se situa num compartimento estanque, é totalmente errado! As abordagens paramétrica e semi-paramétrica, ao contrário do que alguns investigadores consideram, não estão de costas voltadas, mas sim complementam-se e quando usadas em parceria, poderão fornecer modelos probabilísticos mais próximos da realidade, que caracterizem e descrevam o comportamento dos valores extremos associados ao fenómeno em estudo. Ora isso vai permitir o aumento do conhecimento por parte da sociedade que, por sua vez, será utilizado na tomada de decisões e na avaliação dos riscos associados a esses eventos extremos.

Com esta dissertação, é, igualmente, demonstrada a utilidade da EVT na quantificação do risco de precipitação elevada, no que respeita à análise da cauda direita das distribuições, onde a abordagem tradicional, baseada no Teorema do Limite Central, se revela ineficaz. A distribuição Normal não consegue acomodar os riscos associados a eventos extremos e, como tal, justifica-se o uso da inferência estatística em acontecimentos raros.

Um outro aspeto que, também, poderá ser constatado na análise aqui efetuada, prende-se com o facto que, em muitas situações, os dados observados são escassos e não permitem ter a perceção de cenários extremos nunca antes ocorridos. Por exemplo, o nível pluviométrico máximo observado em Barcelos, no período considerado, foi de 146 mm mas, não podemos ser ingénuos ao ponto de acreditar que níveis mais elevados nunca serão atingidos. Sobre isso, Emil Gumbel sempre defendeu que,

"Il est impossible que l'improbable n'arrive jamais. Il y aura toujours une valeur qui dpassera toutes les autres."

Na área da hidrologia e climatologia essa consciência tem-se manifestado nos últimos anos. O IPCC (Intergovernmental Panel on Climate Change), um organismo fundado pela Organização Meteorológica Mundial e pelo Programa das Nações Unidas para o Meio Ambiente, afirmou que,

"... é muito provável que a frequência de chuvas intensas venha a aumentar com a subida da temperatura média global. Isto implica mudanças nos padrões de precipitação, um fator importante, entre outros, para a intensidade e frequência das inundações."

Por este motivo, um estudo sistemático de eventos extremos é de grande relevância para estas áreas ambientais, e as estimativas daí decorrentes são imprescindíveis para o planeamento e desenvolvimento das atividades sujeitas a efeitos adversos, especialmente estruturas de engenharia civil e sistemas agrícolas. O ser humano sempre estará exposto ao risco de inesperadas e desagradáveis alterações climáticas provocadas quer por causas naturais, quer por sua própria culpa. Não será possível evitá-las, contudo, com o auxílio da EVT será possível criar uma cultura de convivência com esses eventos, diminuindo, consideravelmente, os efeitos adversos provenientes dos mesmos. As estimativas, que uma análise em valores extremos fornece, deverão ser utilizadas para prever os potenciais impactos desses eventos raros e planear medidas de mitigação e adaptação.

Como já foi referido, apesar da escolha recair sobre o domínio Gumbel, as estimativas obtidas indicaram que a distribuição subjacente a estes níveis pluviométricos poderia também pertencer ao domínio de atração Weibull. Ora, a cauda direita da distribuição será do tipo exponencial mas ligeiramente mais leve, ou seja, a distribuição estará situada na fronteira entre os domínios Gumbel e Weibull. Como tal, este poderá ser um dos casos em que se tem um limite superior do suporte finito, no domínio Gumbel. Seria interessante desenvolver-se um estudo que aprofundasse mais esta ideia.

Nesta dissertação apenas foram focados os níveis pluviométricos registados pela estação de Barcelos, mas seria interessante a extensão espacial deste estudo, às restantes estações hidrológicas portuguesas. Dessa forma, poderia ser realizada uma análise em valores extremos aos níveis pluviométricos de cada uma dessas estações, com o objetivo de se proceder, de seguida, ao ajustamento dos modelos extremos mais adequados. Com base nas estimativas dos parâmetros dos modelos, poderia ser efetuada uma análise classificatória, com vista à identificação de clusters, no que toca à forma da cauda direita das distribuições destes níveis. Assim, poderiam ser definidos grupos/regiões hidro-climáticas, onde os membros de um mesmo grupo seriam regiões de Portugal, onde os níveis pluviométricos elevados teriam um comportamento semelhante, o que permitiria uma catalogação das regiões mediante o risco associado à precipitação.

Também seria interessante expandir o estudo realizado a outras variáveis climáticas, como a temperatura, a nebulosidade, a insolação e a humidade relativa, e alargar ao caso multivariado, pois, em situações ambientais existem muitas questões que estão interligadas, e que necessitam ser tomadas em consideração quando se pretende criar ferramentas de apoio à decisão.

Um outro estudo que teria bastante interesse, seria uma análise em valores extremos com uma amostra que combinasse os níveis pluviométricos e as suas durações. Desta forma poderiam ser obtidas estimativas de quantidades importantes, como, por exemplo, os níveis de retorno de precipitação para determinadas durações.

Em questões ambientais existem, essencialmente, duas questões fulcrais, uma delas será a compreensão dos processos físicos que originam determinados eventos, e a outra estará relacionada com a predição de realizações futuras, que foi o grande tema desta dissertação. Por este motivo, é crucial uma colaboração entre as ciências do ambiente e a Estatística, e em especial, com a área de Valores Extremos. Desta forma, os modelos ajustados que permitem fazer a tão importante inferência, quantificar os efeitos e medir o risco, como já foi referido diversas vezes, serão muito mais adequados à realidade e irão permitir dar respostas mais eficientes e em tempo útil, para ser feita a mobilização necessária dos meios que amenizem os efeitos negativos. No caso da precipitação, quando esta cai em quantidade elevada, por vezes, provoca as inundações que tantos estragos causa à Humanidade. São vidas que se perdem, produções agrícolas destruídas, casas, prédios e outras edificações que, num instante, ficam num caos, terras que desabam, estradas que desaparecem... A Teoria de Valores Extremos não impede que estes danos aconteçam, mas é um elemento primordial para que essa destruição seja menos catastrófica.

Apêndice A

Referências Bibliográficas

- Araújo Santos, P., Fraga Alves, M. I. and Gomes, M. I. (2006) Peaks Over Random Threshold Methodology for Tail Index and Quantile Estimation. *Revstat* **4**(3), 227–247.
- Balkema, A. A. and de Haan, L. (1974) Residual Life Time at Great Age. *Ann. Probab.* **2**(5), 792–804.
- Bardsley, W. E. (1977) A Test for Distinguishing Between Extreme Value Distributions. *J. Hydrology* **34**, 377–381.
- Beirlant, J., Dierckx, G. and Guillou, A. (2005) Estimation of the Extreme-Value Index and Generalized Quantile Plots. *Bernoulli* **11**(6), 949–970.
- Beirlant, J., Goegebeur, Y., Teugels, J. and Segers, J. (2004) *Statistics of Extremes: Theory and Applications*. First edition. Chichester, England: Wiley series in probability and statistics.
- Beirlant, J., Vynckier, P. and Tengels, J. (1996) Excess Functions and Estimation of the Extreme-Value Index. *Bernoulli* **2**, 293–318.
- Castillo, E., Galambos, J. and Sarabia, J. M. (1989) The selection of the domain of attraction of an extreme value distribution from a set of data. In *Extreme Value Theory, Proc. Oberwolfach, 1987*, eds J. Hüsler and R.-D. Reiss, pp. 181–190. Berlin-Heidelberg: Springer.

- Castillo, E., Hadi, A. S., Balakrishnan, N. and Sarabia, J. M. (2005) *Extreme Value and Related Models with Applications in Engineering and Science*. First edition. New Jersey, EUA: Wiley series in probability and statistics.
- Chandra, M., Singpurwalla, N. D. and Stephens, M. A. (1981) Kolmogorov Statistics for Tests of Fit for the Extreme Value and Weibull Distributions. *J. Amer. Statist. Assoc.* **76**(375), 729–731.
- Choulakian, V. and Stephens, M. A. (2001) Goodness-of-Fit Tests for the Generalized Pareto Distribution. *Technometrics* **43**(4), 478–484.
- Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*. First edition. London, England: Springer series in statistics.
- Davison, A. C. (1984) Modelling excesses over high thresholds. In *Statistical Extremes and Applications*, ed. D. Reidel, pp. 461–482. Dordrecht, Holland: .
- Davison, A. C. and Smith, R. L. (1990) Models for Exceedances Over High Thresholds. *J. Roy. Statist. Soc. Ser. B* **52**(3), 393–442.
- Dekkers, A. L. M., Einmahl, J. H. J. and de Haan, L. (1989) A Moment Estimator for the Index of an Extreme-Value Distribution. *Ann. Stat.* **17**(4), 1833–1855.
- Dekkers, A. L. M. and de Haan, L. (1989) On the Estimation of the Extreme-Value Index and Large Quantile Estimation. *Annals of Statistics* **17**, 1795–1832.
- Diebolt, J., Guillo, A., Naveau, P. and Ribereau, P. (2008) Improving Probability-Weighted Moment Methods for the Generalized Extreme Value Distribution. *REVSTAT – Statistical Journal* **6**(1), 33–50.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance*. First edition. Berlin: Springer.
- Fisher, R. A. and Tippett, L. H. C. (1928) Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proc. Cambridge Philos. Soc.* **24**, 180–190.
- Fraga Alves, M. I. (1995) Estimation of the Tail Parameter in the Domain of Attraction of an Extremal Distribution. *Extreme Value Theory and Applications (Villeneuve d’Arcq, 1992) and J. Statist. Plann. Inference* **45**(1-2), 143–173.
- Fraga Alves, M. I. (2001) A Location Invariant Hill-Type Estimator. *Extremes* **4**(3), 199–217.

-
- Fraga Alves, M. I. (2013) Max-stability at work (or not): Modeling annual or monthly maxima for daily rainfall data. In *Thematic Session 1: The Role of Statistics of Extremes in Society*. Portugal: MECC2013 - International Conference and Advanced School Planet Earth, Mathematics of Energy and Climate Change.
- Fraga Alves, M. I. and Gomes, M. I. (1996) Statistical Choice of Extreme Value Domains of Attraction - a Comparative Analysis. *Communications in Statistics - Theory and Methods* **25**(4), 789–811.
- Fraga Alves, M. I., Gomes, M. I., de Haan, L. and Neves, C. (2007) Mixed Moment Estimator and Location Invariant Alternatives. *Notas e Comunicações CEAUL* **14**.
- Furrer, E. M., Katz, R. W., Walter, M. D. and Furrer, R. (2010) Statistical Modelling of Hot Spells and Heat Waves. *Climate Research* **43**, 191–205.
- Galambos, J. (1982) A statistical test for extreme value distributions. In *Non-Parametric Statistical Inference*, eds B. V. Gnedenko, M. L. Puri and I. Vincze, pp. 221–230. Amsterdam: North-Holland.
- Gnedenko, B. (1943) Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics* **44**(3), 423–453.
- Gomes, M. (1982) A Note on Statistical Choice of Extremal Models. *Proc. IX Jornadas Mat. Hispano-Lusas* pp. 653–655.
- Gomes, M. I. (1989) Comparison of extremal models through statistical choice in multidimensional backgrounds. In *Extreme Value Theory (Oberwolfach, 1987)-Lecture Notes in Statistics*, eds J. Hausler and R.-D. Reiss, volume 51, pp. 191–203. Berlin-Heidelberg: Springer.
- Gomes, M. I. and Alpuim, M. (1986) Inference in a Multivariate Generalized Extreme Value Model - Asymptotic Properties of Two Test Statistics. *Scandinavian Journal of Statistics* **13**, 291–300.
- Gomes, M. I. and van Monfort, M. A. J. (1986) Exponentiality versus Generalized Pareto. *Proc. III Internat. Conf. Statistical Climatology* pp. 185–195.
- Gong, S. (2012) Estimation of hot and cold spells with extreme value theory. Project Report 19, Uppsala University.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979) Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form. *Water Resources Research* **15**(5), 1049–1054.

- de Haan, L. (1970) *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*. 32nd edition. Amsterdam: Mathematical Centre Tract.
- de Haan, L. (1976) Sample Extremes: an Elementary Introduction. *Stat. Neerl.* **30**, 161–172.
- de Haan, L. and Ferreira, A. (2006) *Extreme Value Theory – An Introduction*. New York: Springer.
- de Haan, L. and Stadtmüller, U. (1996) Generalized Regular Variation of Second Order. *J. Austral. Math. Soc.* **A61**, 381–395.
- Hall, P. (1982) On Some Simple Estimates of an Exponent of Regular Variation. *J. Royal Statist. Soc.* **B44**, 37–42.
- Hill, B. (1975) A Simple General Approach to Inference About the Tail of a Distribution. *Ann. Statist.* **3**, 1163–1174.
- Hosking, J. (1984) Testing Whether the Shape Parameter is Zero in the Generalized Extreme Value Distribution. *Biometrika* **71**(2), 367–374.
- Hosking, J. R. M. and Wallis, J. R. (1987) Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* **29**(3), 339–349.
- Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985) Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics* **27**(3), 251–261.
- Jenkinson, A. F. (1955) The Frequency Distribution of the Annual Maximum (or Minimum) Values of Metereological Elements. *Quarterly Journal of the royal Metereology Society* **81**, 158–171.
- Kinnison, R. (1989) Correlation Coefficient Goodness-of-Fit Test for the Extreme-Value Distribution. *Amer. Statist.* **43**(2), 98–100.
- Kozubowski, T. J., Panorska, A. K., Qeadan, F., Gershunov, A. and Rominger, D. (2009) Testing Exponentiality versus Pareto Distribution via Likelihood Ratio. *Commun. Stat.-Simul. C.* **38**(1), 118–139.
- Landwerh, J. M., Matalas, N. C. and Wallis, J. R. (1979) Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles. *Water Resour. Res.* **15**, 1055–1064.

-
- Lilliefors, H. W. (1969) On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown. *J. Amer. Statist. Assoc.* **64**(325), 387–389.
- Marohn, F. (1994) On Testing the Exponential and Gumbel Distribution. *Extreme Value Theory* **71**, 159–174.
- Marohn, F. (1997) Local Asymptotic Normality in Extreme Value Index Estimation. *Ann. Inst. Statist. Math.* **49**, 645–666.
- Marohn, F. (1998a) An Adaptive Efficient Test for Gumbel Domain of Attraction. *Scand. J. Statist.* **25**, 311–324.
- Marohn, F. (1998b) Testing the Gumbel Hypothesis via the POT-Method. *Extremes* **1**(2), 191–213.
- Marohn, F. (2000) Testing Extreme Value Models. *Extremes* **3**(4), 363–384.
- Mason, D. M. (1982) Laws of Large Numbers for Sums of Extreme Values. *Ann. Probab.* **10**, 754–764.
- Miranda, P., Valente, A., Tomé, A. R., Trigo, R., Coelho, F., Aguiar, A. and Azevedo, E. (2006) O clima em portugal nos séculos xx e xxi. In *Alterações Climáticas em Portugal. Cenários, Impactos e Medidas de Adaptação. Projecto SIAM II*, pp. 45–113. Lisboa: Gradiva.
- von Mises, R. (1936) La Distribution de la Plus Grande de n Valeurs. *Reprinted in Selected Papers Volumen II, Amer. Math. Soc., Providence, R.I.* pp. 271–294.
- Nascimento, F. F. (2009) *Abordagem Bayesiana Não-Paramétrica para Análise de Valores Extremos*. Ph.D. thesis, Universidade Federal do Rio de Janeiro, Instituto de Matemática, Departamento de Métodos Estatísticos, Rio de Janeiro, Brasil.
- Nascimento, F. F., Gamerman, D. and Lopes, H. F. (2012) A Semiparametric Bayesian Approach to Extreme Value Estimation. *Statistics and Computing* **22**(1), 661–675.
- Neves, C. and Fraga Alves, M. (2008a) The Ratio of Maximum to the Sum for Testing Super Heavy Tails. *Advances in Mathematical and Statistical Modeling* pp. 181–194. Birkhäuser, Boston.
- Neves, C. and Fraga Alves, M. I. (2007) Semi-Parametric Approach to the Hasofer-Wang and Greenwood Statistics in Extremes. *Test* **16**, 297–313.
- Neves, C. and Fraga Alves, M. I. (2008b) Testing Extreme Value Conditions – An Overview and Recent Approaches. *REVSTAT – Statistical Journal* **6**(1), 83–100.

- Neves, C., Picek, J. and Fraga Alves, M. I. (2006) The Contribution of the Maximum to the Sum of Excesses for Testing Max-Domains of Attraction. *Journal of Statistical Planning and Inference* **136**(4), 1281–1301.
- Tiago de Oliveira, J. (1981) The Selection of the Domain of Attraction of an Extreme Value Distribution from a Set of Data. *Statistical Distributions in Scientific Work* **6**, 367–387.
- Tiago de Oliveira, J. (1984) Univariate Extremes: Statistical Choice. *Statistical Extremes and Applications* pp. 91–107.
- Tiago de Oliveira, J. and Gomes, M. I. (1984) Two Test Statistics for Choice of Univariate Extreme Value Models. *Statistical Extremes and Applications* pp. 651–668.
- Otten, A. and Van Montfort, M. (1978) The Power of Two Tests on the Type of Distribution of Extremes. *J. Hydrology* **37**, 195–199.
- Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics. *Annals of Statistics* **3**(1), 119–131.
- Reiss, R.-D. and Thomas, M. (2007) *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Third edition. Basel-Boston-Berlin: Birkhäuser Verlag.
- Smith, R. L. (1985) Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika* **72**(1), 67–90.
- Smith, R. L. (1987) Estimating Tails of Probability Distributions. *Ann. Stat.* **15**(3), 1174–1207.
- Smith, R. L. (1989) Extreme Value Analysis of Environmental Time Series: an Application to Trend Detection in Ground-Level Ozone. *Statistical Science* **4**(4), 367–393.
- Stephens, M. A. (1976) Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters. *Ann. Statist.* **4**, 357–369.
- Stephens, M. A. (1977) Goodness-of-Fit for the Extreme Value Distribution. *Biometrika* **64**(3), 583–588.
- Stephens, M. A. (1986) Tests for the Exponential Distribution. *Goodness-of-Fit Techniques* **64**, 421–459.
- Stram, D. O. and Lee, J. W. (1994) Variance Components Testing in the Longitudinal Mixed-Effects Models. *Biometrics* **50**, 1171–1177.

-
- Themido Pereira, T. (1993) Second order behavior of domains of attraction and the bias of generalized pickands estimator. In *Extreme Value Theory and Applications III*, eds G. J. Lechner, J. and E. Simiu. Proc. Gaithersburg Conference: NIST special publ.
- Van Montfort, M. A. J. (1970) On Testing that the Distribution of Extremes is of Type I When Type II is the Alternative. *J. Hydrology* **11**, 421–427.
- Vicente, S. (2012) *Extreme Value Theory: an Application to Sports*. Master's thesis, Faculdade de Ciências da Universidade de Lisboa, Lisboa.
- Walshaw, D. (1991) *Statistical Analysis of Extreme Wind Speeds*. Ph.D. thesis, University of Sheffield, Sheffield.
- Walshaw, D. (1994) Getting the Most from Your Extreme Wind Data: a Step by Step Guide. *J. Res. Natl. Inst. Stand. Technol.* **99**, 399–411.
- Wang, J. Z., Cooke, P. and Li, S. (1996) Determination of Domains of Attraction Based on a Sequence of Maxima. *Austral. J. Statist.* **38**(2), 173–181.
- Weissman, I. (1978) Estimation of Parameters and Large Quantiles Based on the k Largest Observations. *J. Amer. Statist. Ass.* **73**, 812–815.
- Yun, S. (2002) On a Generalized Pickands Estimator of the Extreme Value Index. *J. Statist. Plann. Inf.* **102**, 389–409.