



# ESCOLA NAVAL



talant de biefaire

Pedro Miguel Silvestre Rodrigues

Modelação de sobrelevação do nível do mar com  
origem meteorológica

*Aplicação da Teoria de Valores Extremos*

Dissertação para obtenção do Grau de Mestre em  
Ciências Militares Navais, na especialidade de Fuzileiros



Alfeite

2022





# ESCOLA NAVAL

*talant de bi-faire*



**Pedro Miguel Silvestre Rodrigues**

*Modelação de sobrelevação do nível do mar com origem meteorológica*

*Aplicação da Teoria de Valores Extremos*

Dissertação para obtenção do Grau de Mestre em  
Ciências Militares Navais, na especialidade de Fuzileiros

**Orientação de:** Dora Cristina Nunes Paulino Carinhas

**Co-orientação de:** CMG Carlos Miguel Reis Silva de Oliveira Lemos

*O Aluno Mestrando,*

*O Orientador,*

---

Pedro Rodrigues

---

Dora Carinhas

Alfeite

2022



"Respect others in their view, and demand that they respect yours. Love your life, perfect your life, beautify all things in your life. Seek to make your life long and its purpose in the service of your people"

*Chief Tecumseh*



Dedico essa dissertação de mestrado ao meu avô, exemplo de homem e de ser humano.



# Agradecimentos

Antes de dar início à exposição do trabalho, gostaria de expressar o meu agradecimento às pessoas que não só contribuíram para a realização desta dissertação de mestrado, mas que também foram fundamentais ao longo destes últimos 5 anos de percurso na Escola Naval.

O meu agradecimento à Professora Dora Carinhas, pelo acompanhamento prestado ao longo de todo o trabalho estando sempre disponível para auxiliar da melhor forma todos os desafios que foram surgindo, pelo apoio e incentivo constante em conduzir esta investigação a bom porto.

Ao CMG Oliveira e Lemos por todo o apoio prestado no aperfeiçoamento deste trabalho, pela preocupação constante de saber em que rumo estava o trabalho e pela partilha de todo o seu conhecimento, sempre mostrando-se disponível para retirar qualquer dúvida.

Ao Professor Ricardo Moura por ter aceite entrar neste trabalho numa fase mais tardia, de modo a ajudar no aperfeiçoamento da descrição do trabalho, e pela disponibilidade que demonstrou sempre que precisei de tirar uma dúvida.

Aos camaradas do curso “Capitão-de-mar-e-guerra Fernando Amor Monteiro de Barros”, por todos os momentos partilhados ao longo destes 5 anos, por todas as memórias criadas. Que o nosso futuro se cruze várias vezes e que o vosso futuro vos reserve sempre mar chão e ventos de feição.

Aos colaboradores do Instituto Hidrográfico que me acolheram na sua casa e cederam os dados necessários para o sucesso desta dissertação.

Por fim e não menos importante, aos meus familiares e amigos, pelo apoio incondicional ao longo destes 5 anos e por todos os momentos em que a sua dedicação e disponibilidade foi o meu porto de abrigo.



# Resumo

Nas últimas décadas, as sobrelevações do nível do mar, de origem meteorológica, são a causa de muitas inundações. Na costa portuguesa a combinação da sobrelevação da maré e dos efeitos da agitação marítima, origina cheias, inundações, aluimentos de terras, assim como aumento dos níveis de erosão costeira. Neste trabalho, é apresentado um estudo dos valores extremos da sobrelevação da maré, definida como a diferença entre a maré observada e a astronómica prevista, recorrendo a duas famílias de distribuições conhecidas da Teoria dos Valores Extremos. O estudo consistiu em fazer ajustamentos para determinar os parâmetros que fixam as distribuições da sobrelevação dentro de cada família (Generalizada de valores extremos e Generalizada de Pareto), e a partir daí determinar períodos de retorno associados aos níveis da sobrelevação, comparando os resultados das duas distribuições. A Teoria dos valores Extremos foi aplicada à série de dados de alturas horárias provenientes das estações maregráficas de Leixões e Sines. As séries das alturas horárias compreendem o período de 31/12/1994 a 31/12/2020 e de 06/03/1997 a 31/12/2020 registadas pelos marégrafos de Leixões e Sines, respetivamente. A distribuição que melhor se ajustou aos desvios máximos, anuais e mensais, foi a distribuição de Gumbel com parâmetros variáveis. Mantendo-se as condições atuais, é de esperar que os desvios não excedam, até ao final deste século, 1.2 e 0.79 metros para os portos de Leixões e Sines, respetivamente.

**Palavras-chave:** Desvio de maré, Distribuição Generalizada de Pareto, Distribuição Generalizada de Valores Extremos, Maré Meteorológica, Séries Temporais



# Abstract

In the last decades, the sea level rises, of meteorological origin, are the cause of many floods. On the Portuguese coast, the combination of tidal rise and the effects of sea waves, causes floods, inundations, landslides, as well as increased levels of coastal erosion. In this dissertation, a study of the extreme values of the tidal superelevation is presented, defined as the difference between the observed and the predicted astronomical tide, using two families of distributions known from the Theory of Extreme Values. The study consisted of making adjustments to determine the parameters that fix the distributions of superelevation within each family (Generalized Extreme Values and Generalized Pareto), and after that, determining return periods associated with the levels of superelevation, comparing the results of the two distributions. The Theory of Extreme Values was applied to the series of hourly height data from the tide gauge stations of Leixões and Sines. The Theory of Extreme Values was applied to the height data series from the tide gauge stations of Leixões and Sines. The series of hourly heights comprise the period from 12/31/1994 to 12/31/2020 and from 03/06/1997 to 12/31/2020 recorded by the tide gauges of Leixões and Sines, respectively. The distribution that best fit the maximum deviations, annual and monthly, was the Gumbel distribution with variable parameters. Maintaining the current conditions, it is to be hoped that the deviations will not exceed, at the end of this century, 1.2 and 0.79 meters for the ports of Leixões and Sines, respectively.

**Keywords:** Tidal Deviation, Generalized Pareto Distribution, Generalized Distribution of Extreme Values, Meteorological Tide, Time Series



# Índice

Introdução	1
<b>1 Enquadramento Histórico</b>	<b>5</b>
<b>2 Enquadramento Teórico</b>	<b>9</b>
2.1 Fundamentos de Probabilidades e Estatística . . . . .	9
2.1.1 Análise Descritiva e Exploratória de Dados . . . . .	9
2.1.2 Distribuições de Probabilidade . . . . .	11
Distribuição Normal . . . . .	11
Distribuição de Gumbel . . . . .	14
Distribuição Fréchet . . . . .	15
Distribuição de Max-Weibull . . . . .	16
2.1.3 Inferência Estatística . . . . .	17
Estimação de Parâmetros . . . . .	17
Teste de Hipóteses . . . . .	19
2.2 Séries Temporais . . . . .	21
2.2.1 Séries Temporais Estacionárias e não Estacionárias . . . . .	22
2.2.2 Tendência e Sazonalidade . . . . .	23
2.3 Teoria de Valores Extremos . . . . .	24
<b>Distribuições Limite para os Excessos - Análise POT</b>	<b>28</b>
Escolha de um <i>Threshold</i> . . . . .	29
2.4 Variância da Maré . . . . .	31
2.4.1 Maré Astronómica . . . . .	31
Força Geradora de Maré . . . . .	33
Análise Harmónica . . . . .	34
2.4.2 Maré Meteorológica . . . . .	36
Efeitos das Condições Meteorológicas . . . . .	36
<b>3 Análise de Sobrelevações na Costa Portuguesa</b>	<b>39</b>
3.1 Porto de Leixões . . . . .	40

3.1.1	Análise Preliminar das Sobrelevações no Porto de Leixões . .	41
3.1.2	Máximos Anuais . . . . .	45
	Análise GEV . . . . .	48
	Análise GP . . . . .	54
3.1.3	Máximos Mensais . . . . .	59
	Análise GEV . . . . .	62
	Análise POT . . . . .	67
3.1.4	Resultados sobre o Estudo do Porto de Leixões . . . . .	71
3.2	Porto de Sines . . . . .	72
3.2.1	Análise Preliminar das Sobrelevações no Porto de Sines . . .	73
3.2.2	Máximos Anuais do Porto de Sines . . . . .	76
	Análise GEV para os Máximos Anuais de Sines . . . . .	78
	Análise GP para os Máximos Anuais do Porto de Sines . . .	79
3.2.3	Máximos Mensais para o Porto de Sines . . . . .	81
	Análise GEV para os Máximos Mensais do Porto de Sines .	84
	Análise GP para os Máximos Mensais do Porto de Sines . .	85
3.2.4	Resultados sobre o Estudo do Porto de Sines . . . . .	88
<b>Conclusão</b>		<b>88</b>
<b>Bibliografia</b>		<b>91</b>
<b>Apêndices</b>		<b>97</b>
<b>A Script do Software R para a Análise Preliminar das Sobrelevações</b>		<b>97</b>
<b>B Script do Software R para a Análise Preliminar das Sobrelevações Máximas Anuais</b>		<b>101</b>
<b>C Script do Software R para a Análise das Sobrelevações Máximas Mensais</b>		<b>111</b>
<b>D Gráficos da Análise das Sobrelevações de maré no Porto de Sines</b>		<b>121</b>
D.1	Análise dos Máximos Anuais . . . . .	121
D.2	Análise dos Máximos Mensais . . . . .	125
<b>Anexos</b>		<b>129</b>
<b>I Esquema de Alturas do Porto de Leixões</b>		<b>129</b>

**II Participação em: 7.<sup>a</sup> Jornadas da Engenharia Hidrográfica, 2.<sup>a</sup>  
Jornadas Luso-Espanholas de Hidrografia - Resumo Alargado 135**



# Lista de Figuras

2.1	Histograma de uma Distribuição Normal . . . . .	12
2.2	QQ-Plot de uma distribuição normal . . . . .	14
2.3	Teste Bilateral de uma distribuição . . . . .	20
2.4	Intervalo de rejeição de $H_0$ para um teste bilateral . . . . .	21
2.5	Cauda direita de uma distribuição F . . . . .	26
2.6	Esquema dos sistemas anfidrômicos nos oceanos. . . . .	32
2.7	Esquema das Forças atrativas. . . . .	33
2.8	Maré meteorológica positiva. . . . .	37
2.9	Efeito de uma tempestade no nível do mar. . . . .	38
3.1	Fotografia aérea do Porto de Leixões . . . . .	40
3.2	Esquema de altura de um cais em relação ao Zero Hidrográfico de Leixões . . . . .	41
3.3	Séries temporais dos desvios em Leixões (1995-2020) . . . . .	42
3.4	Desvios ano a ano: 1995 a 2010 . . . . .	42
3.5	Diagramas de extremos e quartis dos desvios da maré por ano, entre 1995 e 2020. . . . .	43
3.6	Histograma dos desvios: Leixões 1995 a 2020 . . . . .	44
3.7	Série temporal, sem falhas, dos desvios de Leixões: 2002-2020 . . . . .	44
3.8	Desvios máximos anuais: Leixões . . . . .	45
3.9	Histograma dos Desvios máximos Anuais: Leixões . . . . .	46
3.10	Gráficos de auto-correlação e auto-correlação parcial para os máximos anuais: Leixões . . . . .	47
3.11	Diagrama de extremos de quartis correspondente aos máximos anuais: Leixões . . . . .	47
3.12	Gráfico de probabilidades GEV: Leixões . . . . .	48
3.13	Gráfico de densidade para o modelo GEV: Leixões . . . . .	49
3.14	Gráficos de probabilidade GEV com escala e localização variável: Leixões . . . . .	50
3.15	Gráficos de probabilidade e densidade de probabilidade modelo Gumbel: Leixões . . . . .	52

3.16	Gráficos de probabilidade e densidade modelo Gumbel com parâmetros variáveis: Leixões . . . . .	52
3.17	Gráfico de nível de retorno, para os máximos anuais, através do modelo Gumbel: Leixões . . . . .	54
3.18	Estimativas dos parâmetros forma e escala versus threshold para os máximos anuais de Leixões . . . . .	55
3.19	Mean Residual Life Plot para os máximos anuais de Leixões . . . . .	56
3.20	Gráfico de probabilidades dos excessos, referente à amostra nos máximos anuais de Leixões . . . . .	56
3.21	Gráfico de probabilidades dos excessos, com o parâmetro escala variável, referente à amostra nos máximos anuais de Leixões . . . . .	57
3.22	Gráfico de nível de retorno, para os máximos anuais, através do modelo GP: Leixões . . . . .	58
3.23	Máximos mensais de 2002 a 2020 para Leixões . . . . .	59
3.24	Histograma dos desvios máximos mensais para Leixões . . . . .	60
3.25	Gráficos de auto-correlação e auto-correlação parcial para os máximos mensais de Leixões . . . . .	61
3.26	Diagrama de extremos e quartis correspondente aos máximos mensais de Leixões . . . . .	61
3.27	Gráficos de probabilidade e densidade de probabilidade da amostra dos máximos mensais de Leixões para o modelo GEV . . . . .	62
3.28	Gráficos de probabilidades com os parâmetros localização e escala variável . . . . .	63
3.29	Gráficos de probabilidade e densidade de probabilidade do modelo Gumbel para os máximos mensais de Leixões . . . . .	64
3.30	Gráficos de probabilidade e densidade de probabilidade modelo Gumbel com parâmetros a variar, para os máximos mensais de Leixões . . . . .	65
3.31	Gráfico de nível de retorno, para os máximos mensais, através do modelo de Gumbel com o parâmetro localização a variar . . . . .	67
3.32	Estimativas dos parâmetros forma e escala <i>versus threshold</i> para os máximos mensais de Leixões . . . . .	68
3.33	<i>Mean Residual Life Plot</i> para os máximos mensais de Leixões . . . . .	68
3.34	Gráfico de probabilidades dos excessos, referente à amostra dos máximos mensais de Leixões . . . . .	69
3.35	Gráfico de probabilidades dos excessos, com o parâmetro escala variável, referente à amostra nos máximos mensais de Leixões . . . . .	70

3.36	Gráfico de nível de retorno, para os máximos mensais de Leixões, segundo o modelo GP com escala variável . . . . .	71
3.37	Fotografia aérea do Porto de Sines . . . . .	73
3.38	Série temporal dos desvios de Sines (1997-2020) . . . . .	73
3.39	Diagramas de extremos e quartis dos desvios da maré por ano (Sines) . . . . .	74
3.40	Histograma da série dos desvios de maré do Porto de Sines . . . . .	75
3.41	Série temporal dos desvios de maré do Porto de Sines sem falhas . . . . .	75
3.42	Desvios máximos anuais do Porto de Sines . . . . .	76
3.43	Histograma dos desvios máximos anuais do Porto de Sines . . . . .	77
3.44	Gráficos de auto-correlação e auto-correlação parcial para os máximos anuais do Porto de Sines . . . . .	77
3.45	Gráfico de nível de retorno para os máximos anuais de Sines, através do modelo Gumbel com o parâmetro localização a variar . . . . .	78
3.46	Estimativas dos parâmetros forma e escala versus threshold para os máximos anuais do Porto de Sines . . . . .	79
3.47	<i>Mean Residual Life Plot</i> para os máximos anuais do Porto de Sines . . . . .	80
3.48	Gráfico de nível de retorno, para os máximos anuais de Sines, através do modelo GP com escala a variar . . . . .	81
3.49	Série dos máximos mensais do Porto de Sines, no período de 1997 a 2020 . . . . .	82
3.50	Histograma dos desvios máximos mensais do Porto de Sines . . . . .	82
3.51	Gráficos de auto-correlação e auto-correlação parcial para os máximos mensais do Porto de Sines . . . . .	83
3.52	Diagrama de extremos e quartis correspondente aos desvios máximos mensais do Porto de Sines . . . . .	84
3.53	Gráfico de nível de retorno para os máximos mensais de Sines, através do modelo Gumbel com o parâmetro localização a variar . . . . .	85
3.54	Estimativas dos parâmetros forma e escala versus <i>threshold</i> para os máximos mensais do Porto de Sines . . . . .	86
3.55	<i>Mean Residual Life Plot</i> para os máximos mensais do Porto de Sines . . . . .	86
3.56	Gráfico de nível de retorno, para os máximos mensais de Sines, através do modelo GP com escala a variar . . . . .	87
D.1	Gráfico de probabilidades GEV para dos desvios máximos anuais de Sines . . . . .	121
D.2	Gráfico de densidade GEV para dos desvios máximos anuais de Sines . . . . .	121
D.3	Gráficos de probabilidade GEV com parâmetros localização e escala variáveis . . . . .	122

D.4	Gráficos de densidade GEV com parâmetros localização e escala variáveis . . . . .	122
D.5	Gráficos de probabilidade e densidade Gumbel . . . . .	122
D.6	Gráficos de probabilidade Gumbel com parâmetros localização e escala variáveis . . . . .	123
D.7	Gráficos de densidade Gumbel com parâmetros localização e escala variáveis . . . . .	123
D.8	Gráficos de probabilidade GP . . . . .	124
D.9	Gráficos de probabilidade GP com parâmetro escala variável . . . . .	124
D.10	Gráficos de probabilidade e densidade GEV para o máximos mensais do Porto de Sines . . . . .	125
D.11	Gráficos de probabilidade GEV com parâmetros localização e escala variáveis, para o máximos mensais do Porto de Sines . . . . .	125
D.12	Gráficos de densidade GEV com parâmetros localização e escala variáveis, para o máximos mensais do Porto de Sines . . . . .	126
D.13	Gráficos de probabilidade de densidade Gumbel com parâmetros localização e escala variáveis, para o máximos mensais do Porto de Sines . . . . .	126
D.14	Gráficos de probabilidade e densidade Gumbel, para o máximos mensais do Porto de Sines . . . . .	127
D.15	Gráfico de probabilidade GP para o máximos mensais do Porto de Sines . . . . .	127
D.16	Gráfico de probabilidade GP com parâmetro escala variável, para os máximos mensais do Porto de Sines . . . . .	128
I.1	Esquema do Quebra mar Submerso N.º 1 . . . . .	129
I.2	Esquema do Posto A . . . . .	129
I.3	Esquema do Molho Sul . . . . .	129
I.4	Esquema do Posto B . . . . .	130
I.5	Esquema do Posto C . . . . .	130
I.6	Esquema do Estaleiro para obras . . . . .	130
I.7	Esquema do Cais oeste . . . . .	131
I.8	Esquema do terminal de contentores Norte . . . . .	131
I.9	Esquema Doca 1 - Norte e Sul . . . . .	131
I.10	Esquema da Doca 2 - Norte . . . . .	132
I.11	Esquemas da Doca 2 - Norte e Sul . . . . .	132
I.12	Esquema da Doca 4 - Norte . . . . .	132
I.13	Esquema da Doca 4 - Sul e nascente . . . . .	133
I.14	Esquemas da Doca 4 - Sul e Poente . . . . .	133

I.15	Esquema das Pontes Cais . . . . .	133
I.16	Esquema da Doca 1 - Sul . . . . .	134
I.17	Esquema da Doca 1 - Norte e Sul . . . . .	134



# Lista de Tabelas

2.1	Interpretação do $p - value$ para o Teste de Hipóteses. . . . .	20
3.1	Estatísticas descritivas anuais para a série de Leixões: 1995 a 2013. . . . .	43
3.2	Medidas descritivas dos máximos anuais: Leixões. . . . .	46
3.3	Parâmetros localização, escala e forma do modelo GEV: Leixões. . . . .	51
3.4	Indicadores estatísticos AIC e BIC e teste RMSE para as distribuições GEV e Gumbel com o parâmetro localização variável. . . . .	53
3.5	Valores do nível de retorno estimado através do modelo de Gumbel, para os máximos anuais de Leixões. . . . .	54
3.6	Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GP, com o parâmetro escala variável segundo a expressão 3.1. . . . .	57
3.7	Valores do nível de retorno estimado através do modelo GP, para os máximos anuais de Leixões . . . . .	58
3.8	Medidas descritivas dos máximos mensais para Leixões. . . . .	60
3.9	Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GEV com parâmetros constantes e variáveis. . . . .	63
3.10	Parâmetros localização, escala e forma do modelo GEV para os máximos mensais de Leixões. . . . .	64
3.11	Indicadores estatísticos AIC, BIC e o teste RMSE do modelo de Gumbel para os máximos mensais de Leixões. . . . .	66
3.12	Indicadores estatísticos AIC, BIC e o teste RMSE para os modelos GEV e Gumbel, com o parâmetros localização variável, para os máximos mensais de Leixões. . . . .	66
3.13	Valores do nível de retorno estimado através do modelo de Gumbel com o parâmetro localização a variar segundo a expressão 3.2, para os máximos mensais de Leixões . . . . .	67
3.14	Indicadores estatísticos AIC, BIC e o teste RMSE para o modelo GP com parâmetros constantes e com o parâmetro escala a variar segundo a expressão 3.2. . . . .	70
3.15	Valores do nível de retorno estimado através do modelo GP com escala variável, para os máximos anuais de Leixões. . . . .	71

3.16	Estatística descritiva da série dos desvios de maré para o Porto de Sines. . . . .	74
3.17	Estatística descritiva da série dos desvios máximos anuais para o Porto de Sines. . . . .	76
3.18	Indicadores estatísticos AIC e BIC e teste RMSE para as distribuições GEV e Gumbel com o parâmetro localização variável, Porto de Sines.	78
3.19	Valores do nível de retorno para os máximos anuais de Sines, através do modelo Gumbel com o parâmetro localização a variar segundo a expressão 3.1. . . . .	79
3.20	Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GP, com o parâmetro escala constante e variável segundo a expressão 3.1. . . . .	80
3.21	Valores do nível de retorno estimado através do modelo GP com o parâmetro escala a variar segundo a expressão 3.1, para o Porto de Sines. . . . .	81
3.22	Medidas descritivas dos desvios máximos mensais do Porto de Sines.	83
3.23	Indicadores estatísticos AIC e BIC e teste RMSE para as distribuições GEV e Gumbel com o parâmetro localização variável, máximos mensais do Porto de Sines. . . . .	84
3.24	Valores do nível de retorno para os máximos mensais de Sines, através do modelo Gumbel com o parâmetro localização a variar segundo a expressão 3.2. . . . .	85
3.25	Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GP, com o parâmetro escala constante e variável segundo a expressão 3.2. . . . .	87
3.26	Valores do nível de retorno estimado através do modelo GP com o parâmetro escala a variar segundo a expressão 3.2. . . . .	88

# Lista de Equações

2.1	Condição da Função de densidade de probabilidade . . . . .	10
2.2	Função de distribuição para variável aleatória continua . . . . .	10
2.3	Equação do valor esperado para variáveis aleatórias contínuas . . . . .	10
2.4	Equação da variância . . . . .	10
2.5	Equação do desvio padrão. . . . .	11
2.6	Equação da Moda para variáveis aleatórias contínuas . . . . .	11
2.7	Equação da mediana para variáveis aleatórias contínuas . . . . .	11
2.8	Função de densidade de probabilidade para a Distribuição de Gauss .	12
2.9	Função de distribuição para a Distribuição de Gauss . . . . .	13
2.10	Função de densidade de probabilidade de Gumbel . . . . .	14
2.11	Estima de quantis elevados (distribuição de Gumbel) . . . . .	15
2.12	Estima dos período de retorno, através de aproximações (distribuição de Gumbel). . . . .	15
2.13	Função de densidade de probabilidade de Fréchet . . . . .	15
2.14	Obtenção de quantis - Distribuição de Fréchet . . . . .	16
2.15	obtenção de quantis através do Período de Retorno - Distribuição de Fréchet . . . . .	16
2.16	Função de densidade de probabilidade de Max-Weibull . . . . .	16
2.17	Obtenção de quantis - Distribuição de Max-Weibull . . . . .	17
2.18	Estima do intervalo de confiança . . . . .	18
2.19	Limites de confiança . . . . .	18
2.20	f.d.p. de um processo estocástico . . . . .	22
2.21	Estimador de tendência de uma componente sazonal . . . . .	23
2.22	Ajuste da série à tendência . . . . .	23
2.23	Função do Período de Retorno . . . . .	25
2.24	Função de distribuição geral de valores extremos . . . . .	26
2.25	Fórmula de distribuição de GEV . . . . .	27
2.26	Fórmula de distribuição de Gumbel . . . . .	27
2.27	Fórmula de distribuição de Fréchet . . . . .	27
2.28	Fórmula de distribuição de Max-Weibull . . . . .	27

2.29	Probabilidade de excedência de níveis elevados $u$ . . . . .	28
2.30	Período de retorno do nível $u$ . . . . .	28
2.31	Nível de retorno . . . . .	28
2.32	Função da Distribuição Generalizada de Pareto . . . . .	29
2.33	Função de excesso médio . . . . .	29
2.34	Função de excesso médio para o caso da Generalizada de Pareto . . . . .	30
2.35	Função empírica de excesso médio para o caso da Generalizada de Pareto . . . . .	30
2.36	Função de excesso médio, para o caso da GP . . . . .	30
2.37	Excessos para a distribuição GP . . . . .	30
2.38	Lei de Newton . . . . .	33
2.39	Aceleração da força geradora de maré . . . . .	34
2.40	Fórmula Harmónica . . . . .	35
2.41	Expressão de altura de água num determinado local . . . . .	36
3.1	Variação dos parâmetros Localização e Escala para os máximos anuais . . . . .	49
3.2	Variação dos parâmetros Localização e Escala para os máximos mensais . . . . .	62

# Lista de Abreviaturas

<b>ACF</b>	<i>AutoCorrelation Function</i> (Função de Auto-correlação)
<b>f.d.p.</b>	Função Densidade de Probabilidade
<b>f.d.</b>	Função Distribuição
<b>EFHI</b>	Extreme Flood Hazard Index
<b>GEV</b>	<i>Generalized Extreme Value</i> (Generalizada de Valores Extremos)
<b>GP</b>	Generalizada de Pareto
<b>i.i.d.</b>	Independentes e Identicamente Distribuídas
<b>IH</b>	Instituto Hidrográfico
<b>NM</b>	Nível Médio do Mar
<b>PACF</b>	<i>Partial AutoCorrelation Function</i> (Função de Auto-correlação Parcial)
<b>POT</b>	<i>Peaks Over Threshold</i>
<b>TVE</b>	Teoria de Valores Extremos
<b>v.a.</b>	Variável Aleatória
<b>ZH</b>	Zero Hidrográfico
<b>ZHL</b>	Zero Hidrográfico de Leixões



# Lista de Símbolos

$A_0(t)$	Nível Médio do Mar, no instante $t$ , num determinado ponto
$CO_2$	Dióxido de Carbono
$E(Y)$	Valor Esperado de $Y$
$Me$	Mediana
$Mo$	Moda Amostral
$\mathbb{R}$	Conjunto de Números Reais
$S(t)$	Varição meteorológica da altura de maré
$T_R$	Período de Retorno
$T(t)$	Varição astronómica da altura de maré
$u$	Nível de Retorno
$Var(Y)$	Variância de $Y$
$X(t)$	Altura da água, no instante $t$ , num dado local
$\mu$	Threshold (Nível Elevado)
$\sigma$	Desvio-padrão da população
$\Phi$	Função distribuição da Distribuição Normal
$\lambda$	Parâmetro Localização
$\delta$	Parâmetro Escala
$\rho$	Parâmetro Forma



# Introdução

Portugal é um país costeiro que faz fronteira com Oceano Atlântico, no qual possui uma costa com extensão de cerca de 987 km (IPMA, 2018). Portugal Continental tem centenas de milhares de pessoas a viver na zona costeira, com inúmeras atividades marítimas de elevado valor económico e um elevado número de infraestruturas que devem ser adaptadas e protegidas dos perigos costeiros, nomeadamente, subida do nível do mar devido a tempestades.

O Instituto Hidrográfico (IH), através da Divisão de Oceanografia, tem como objetivo contribuir para o conhecimento e previsão dos principais processos físicos que caracterizam os espaços marítimos. Este objetivo é concretizado através da observação do meio marinho, com carácter pontual ou permanente. O IH mantém um sistema de monitorização permanente que integra uma rede de boias ondógrafo (condições de agitação marítima), uma rede de marégrafos costeiros (variações do nível do mar) e uma rede de estações meteorológicas costeiras. A rede maregráfica tem como objetivos: recolher, analisar e fornecer informação de marés e níveis de água para as mais diversas finalidades, nomeadamente: segurança à navegação, estabelecimento de níveis de referência em Hidrografia, execução de projetos de engenharia costeira, investigação científica, pesca e atividades recreativas.

Através da monitorização e do estudo das marés e dos seus fenómenos extremos, é possível construir defesas, como os quebra-mares, com altura suficiente para que a probabilidade deste tipo de eventos terem impacto seja ínfima (Gomes et al., 2013).

A altura de água,  $X(t)$ , observada num determinado local, pode ser representada por uma função com a forma seguinte:

$$X(t) = A_0(t) + T(t) + S(t), \text{ onde:}$$

$A_0(t)$  – nível médio do mar no ponto de observação;

$T(t)$  – variação da maré astronómica;

$S(t)$  – variação da altura da água devido a efeitos meteorológicos (sobreelevação meteorológica).

---

O movimento regular da maré é constantemente influenciado, a um menor ou maior grau, por efeitos meteorológicos. Em certas regiões do globo, inundações costeiras intensas podem-se verificar quando há a coincidência de fenómenos meteorológicos extremos, como tempestades tropicais ou tufões, com alturas de maré astronómica elevada. Como exemplo, das consequências deste tipo de fenómenos, tem-se as cheias que ocorreram a 1 de fevereiro de 1953 no Mar do Norte, o nível do mar excedeu os 5,6 metros acima do nível previsto, tendo causado diversos estragos na costa da Holanda, Inglaterra, Bélgica, Dinamarca e França, resultando na morte de aproximadamente 2500 pessoas (Gomes et al., 2013). No caso de Portugal existem diversas áreas vulneráveis a este tipo de fenómenos, nomeadamente a Ria de Aveiro e o Estuário do Tejo (Rocha, 2016).

O estudo dos valores extremos de variáveis aleatórias tem grande importância em muitos domínios, como sejam as alterações climáticas, a sismologia ou a engenharia oceanográfica, por serem eventos extremos que têm consequências graves ou mesmo catastróficas. As condições meteorológicas são a principal causa de diferenças entre alturas de maré previstas e observadas, fazendo-se notar com maior intensidade nos períodos de Inverno (Pugh, 1987). Embora a estimativa de valores extremos seja um tópico mais difícil (e mesmo controverso) do que a estimativa de valores de tendência central e dispersão, é, ainda assim, de grande importância para a previsão de períodos de retorno de eventos suscetíveis de produzir grandes impactos (Guerreiro, 2015).

Num estudo de extremos de sobrelevação meteorológica há a preocupação em responder a questões como:

- (i) Qual o maior desvio que ocorreu?
- (ii) Qual período de retorno, estimado, destes valores extremos?
- (iii) Que impacto terão estes valores extremos quando combinados com o aumento do nível médio do mar?
- (iv) As estruturas portuárias estão dimensionadas para proteger o interior dos portos dos eventos extremos?

Um dos objetivos primordiais deste trabalho será identificar o comportamento dos registos maregráficos em estudo, verificando qual o comportamento da cauda da distribuição subjacente à amostra e, conseqüentemente, identificar qual as distribuições que melhor se ajustam aos dados. O objetivo final é estimar se até ao final do século poderá ocorrer alguma sobrelevação que ultrapasse ou fique perto de ultrapassar alturas das defesas dos portos.

---

Com vista a concretizar estes objetivos propostos, o capítulo 1 desta dissertação será dedicado à apresentação da teoria de valores extremos aplicada às sobrelevações do mar. O seguinte capítulo foca-se na apresentação da teoria de valores extremos, nomeadamente da metodologia da Distribuição Generalizada de Valores Extremos e do método de *Peaks over Threshold*, e da teoria de geração de marés. O capítulo 3 apresenta o estudo das sobrelevações com origem meteorológica nos portos de Leixões e Sines, aplicando estas duas metodologias pretende-se obter respostas sobre o comportamento das sobrelevações nestes dois portos e consequentemente ir de encontro aos objetivos propostos. Por fim, no capítulo 4, serão realizadas algumas considerações finais sobre o estudo, indicando possíveis problemas em aberto que poderão ser desenvolvidos em trabalhos futuros.



# Capítulo 1

## Enquadramento Histórico

Anualmente, em todo o mundo, acontecem fenómenos extremos que acarretam consequências graves para o homem, fenómenos tais como maremotos, terremotos, tempestades tropicais e furacões.

A aplicação da Teoria de Valores Extremos (TVE) teve impulso em 1953, quando as barragens, que protegem a Holanda do avanço do mar, se romperam e causaram a inundação de boa parte do país, provocando a morte de 1800 pessoas. Após o desastre, o governo da Holanda criou um comité que recorreu à TVE, para estabelecer a altura das barragens (Gomes, 2019).

O comportamento dos extremos pode ser modelado por uma das três distribuições: Gumbel, Fréchet e Weibull. A distribuição de Fréchet, foi introduzida em 1928 por Maurice Fréchet no livro “*Sur la loi de probabilité de l'écart maximum*” (Fréchet, 1928), a distribuição de Max-Weibull foi apresentada no livro “*A statistical distribution function of wide applicability*”, em 1951, por Waloddi Weibull (Weibull, 1951) e a distribuição de Gumbel foi introduzida pelo próprio em 1935 (Gumbel, 1935).

As distribuições anteriores podem ser representadas numa única expressão, a distribuição Generalizada de Valores Extremos (GEV). Esta distribuição de probabilidade, introduzida por Jenkinson (1955), combina as três famílias de Gumbel, Fréchet e Max-Weibull numa única expressão. No século XX, foi publicado o Teorema de Tipos Extremos por Leonard Tippett, Sir Ronald Fisher e Boris Gnedenko (Gnedenko, 1948), que permitiu concluir que é possível modelar este comportamento através de uma única expressão, a distribuição GEV.

Pickands (1975), apresentou a Distribuição Generalizada de Pareto (GP) como sendo uma distribuição para a modelação das excedências de uma amostra

acima de um nível elevado. Esta distribuição é bastante versátil, pois permite modelar observações de populações cujas distribuições subjacentes possuem caudas com diversos pesos. A GP tem grande aplicabilidade em diversos campos científicos, nomeadamente na meteorologia e ambiente (Sintra, 2017).

A importância desta temática pode ser acompanhada no livro pioneiro em "Estatística de Extremos", e nos livros de Beirlant et al. (1996), Oliveira (1997), Reiss e Thomas (1997), Embrechts et al. (1998), Kotz e Nadarajah (2000), Coles et al. (2001), Beirlant et al. (2004), Castollo et al. (2004), De Haan et al. (2006), Resnick (2007) e Markovich (2007).

Em 2009, foi publicado o artigo "*A modelling approach for estimating the frequency of sea level extremes and the impact of climate change in southeast Australia*", onde se apresentam os níveis de retorno das marés de tempestade ao longo da costa sudeste da Austrália (McInnes et al., 2009). Através da distribuição de valores extremos foram calculados os níveis de retorno da altura de maré. O impacto das alterações climáticas nos extremos de maré foi explorado adicionando diversas estimativas do aumento do nível médio do mar e através da introdução de dados de vento modificados. Este estudo conclui que, pelo menos, uma vez a cada 1 a 2 anos irá ocorrer uma sobrelevação da maré prejudicial à costa australiana, sendo que a maior sobrelevação prevista foi de 1,56 metros para 50 anos, após a data deste artigo.

Marcos et al. (2009), publicaram um artigo sobre os níveis extremos de maré no sul da Europa, "*Sea level extremes in southern Europe*". Este estudo teve como objetivo inferir os riscos, na costa causados por níveis extremos de maré. Os autores usaram três métodos para realizar o estudo: distribuição de Pareto, a *Tide-surge joint distribution* e o recurso ao cálculo dos percentis. Neste último método, não é possível estimar os períodos de retorno, apenas estudar a existência de tendências. Este trabalho concluiu que os valores mais elevados de sobrelevações de maré são observados no norte do Adriático e no Golfo de Gabes (até 1,6 m). Os valores do nível de retorno estimados indicam que é provável que para os 50 anos, após publicação do artigo, não ultrapassem os 30 cm para 80% das estações maregráficas estudadas.

Tran (2020), publicou um estudo, onde investigou, simultaneamente, mudanças temporais e espaciais nas ondas e no nível do mar através de um modelo hidrodinâmico, em particular a estimação dos níveis de retorno para os níveis extremos de maré. Este estudo foi aplicado à Baía de Port Phillip, Austrália. Tran

(2020), concluiu que os extremos de maré de tempestade ocorrem mais frequentemente na costa leste da baía, estimando que para um período de 100 anos o maior nível de retorno obtido, dentro da baía, será de 2,3m para Moondah, Austrália.

Para a costa portuguesa foram realizados diversos estudos no sentido de compreender melhor o comportamento da sobrelevações de maré, Fortunato et al. (2014) publicou o artigo "*Determination of extremes sea levels along the Portuguese coast*", que visava determinar níveis de água extremos associados a diferentes períodos de retorno, através do desenvolvimento de um modelo de circulação para marés e ondas. Neste estudo concluiu-se que a amplitude das sobrelevações aumenta de sul para norte.

Guerreiro (2015) estudou as delimitações diárias inundáveis na Ribeira da Canada, Algarve. Este estudo consistiu na simulação da ocorrência, em simultâneo, do escoamento superficial e da altura de maré, com o objetivo de determinar a distribuição que melhor explicava o comportamentos dos dados da precipitação máxima diária e da altura de maré. Neste caso, a distribuição que melhor se ajustou às amostras foi a de Gumbel, no entanto, as distribuições de Fréchet e Max-Weibull foram as que melhor se ajustaram à amostra de dados da maré astronómica. A partir dos resultados obtidos foi possível concluir que a altura de escoamento da linha de água principal é condicionada pela altura de maré até 490m da foz, o que provoca o galgamento e por sua vez inundações.

Antunes et al. (2019) estudaram inundações costeiras, em Portugal continental, causadas pela sobrelevação do nível do mar com origem meteorológica. Estes autores avaliaram o risco de inundação na costa portuguesa para os anos de 2025, 2050 e 2100, estimando, para estes anos, o nível de retorno. Antunes et al. (2019) concluíram que os distritos mais afetados por inundações serão o de Lisboa, Faro e Aveiro, com 221, 182 e 172  $Km^2$  de área inundada, respetivamente; para 2100 estes valores sobem para 250, 211 e 219  $Km^2$ , respetivamente.



# Capítulo 2

## Enquadramento Teórico

### 2.1 Fundamentos de Probabilidades e Estatística

O campo da estatística pode ser dividido em duas grandes áreas, a estatística descritiva e a inferência estatística. A estatística descritiva permite sintetizar e representar de uma forma simples a informação contida num conjunto de dados. A inferência estatística permite deduzir o que acontece na população da qual os dados foram recolhidos; a extrapolação de conclusões deve ser cautelosa, devendo sempre associar-se um determinado grau de credibilidade ou nível de confiança.

#### 2.1.1 Análise Descritiva e Exploratória de Dados

Uma variável aleatória (v.a.) é entendida como uma variável cujo valor é um resultado numérico associado ao resultado de uma experiência aleatória. Pestana e Velosa (2010), definem v.a. da seguinte maneira: “Em geral, o que contamos ou medimos resulta de uma experiência aleatória, em que o acaso intervém, pelo menos na fase de escolha da amostra, varia com o resultado da experiência. É por isso natural dizer que é uma variável aleatória”. Quando se define variável diz-se, que esta, pode ser de tipo qualitativo ou quantitativo; assim, o resultado de uma experiência aleatória não dá, necessariamente, um resultado numérico (Pestana e Velosa, 2010). No entanto, neste trabalho, estamos de um modo geral interessados em estudar resultados numéricos. A v.a. quantitativa pode ser discreta, onde os valores que esta pode tomar estão contidos num conjunto finito enumerável, ou pode ser contínua quando os valores que esta adota abrangem um intervalo de números reais.

Considere-se uma variável aleatória contínua,  $Y$ , que possa assumir qualquer valor num determinado intervalo, exemplo disto tem-se resistência de um material, a concentração de  $CO_2$  na água, a altura de maré, entre outros. Neste caso, como

se trata de uma variável aleatória contínua, resulta a função de densidade de probabilidade (f.d.p.),  $f(y)$ , que obrigatoriamente satisfaz as seguintes condições (Morais, 2020):

- $f(y) \geq 0, \forall y \in \mathbb{R}$ ,
- $\int_{-\infty}^{+\infty} f(y) dy = 1$ ,
- Sejam  $a$  e  $b$  pertencentes ao intervalo,  $-\infty < a < b < +\infty$  tem-se que:

$$P(a \leq Y \leq b) = \int_a^b f(y) dy \quad (2.1)$$

Define-se assim a função de distribuição, para uma variável aleatória contínua, é dada por (Morais, 2020):

$$F(y) = P(Y \leq y) = \int_{-\infty}^y f(y) dy \quad (2.2)$$

Existem diversas famílias de distribuições, que para cada tipo de problema, descrevem o comportamento das variáveis aleatórias. Cada uma destas distribuições é indexada por um conjunto de parâmetros, variando de família para família. Em regra, estes parâmetros são desconhecidos e têm que ser estimados a partir de amostras (Morais, 2020).

Dois parâmetros de grande importância no contexto das distribuições de probabilidade são o valor esperado,  $E(Y)$ , e a variância,  $\text{Var}(Y)$ . O valor esperado, também denominado de média ou esperança matemática, é definido por (Morais, 2020):

$$E(Y) = \int_{-\infty}^{+\infty} yf(y)dy \quad (2.3)$$

A variância de uma v.a. contínua, que considera todas as observações e mede os desvios das mesmas em relação à média, é dada por (Morais, 2020):

$$\text{Var}(Y) = E(Y - u)^2 = E(Y^2) - [E(Y)]^2 \quad (2.4)$$

O desvio padrão,  $\sigma$ , é dado pelo valor positivo da raiz quadrada da variância, conforme a equação 2.5 (Morais, 2020):

$$Var(Y) = \sigma^2 = \sqrt{Var(Y)} \quad (2.5)$$

Outras estatísticas de localização ou tendência central são a mediana ( $Me$ ) e a moda ( $Mo$ ). A moda amostral está associada ao ponto máximo da função de densidade de probabilidade, ou seja, é o valor com mais frequência (Morais, 2020):

$$Mo : f(Mo) = \max f(y) \quad (2.6)$$

Já a mediana, que representa o valor central de um conjunto de dados, é dada por (Morais, 2020):

$$Me : F(Me) = \frac{1}{2} \quad (2.7)$$

A amplitude é outra estatística de dispersão (a par da variância e do desvio padrão) que determina a diferença entre o valor máximo e o valor mínimo do conjunto de dados (Pestana e Velosa, 2010).

## 2.1.2 Distribuições de Probabilidade

A distribuição de probabilidade é definida como um modelo matemático que relaciona o valor de uma variável aleatória com uma probabilidade de ocorrência desse valor na população (Morais, 2020).

### Distribuição Normal

#### I. Caracterização

Uma das distribuições mais conhecida e utilizada é a Distribuição Normal ou de Gauss, a qual é simétrica e apresenta uma forma semelhante à de um sino, tal como representado Figura 2.1.

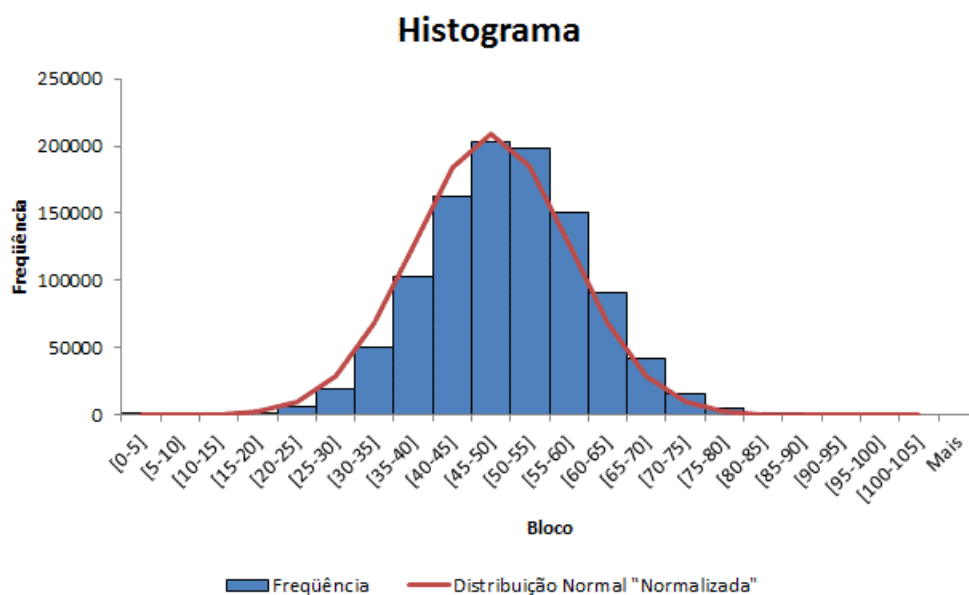


FIGURA 2.1: Histograma de uma Distribuição Normal (Fonte: <https://engenheirodoexcel.com.br/histograma-no-excel/>, 10/04/2022).

A importância da distribuição normal pode ser justificada pelo facto de que variáveis aleatórias e independentes podem ser expressas pela soma de um grande número de variáveis aleatórias e, pelo teorema do limite central, pode-se argumentar que a distribuição da soma pode ser assintoticamente normal (Guerreiro, 2015). Para se aprofundar a matéria sobre o teorema do limite central pode-se consultar o livro de Pestana e Velosa (2010).

Se o modelo que descreve a população fosse exato, não haveria v.a. e não seria necessário introduzir a distribuição normal. Como exemplo, as variáveis aleatórias de precipitação e o caudal anual médio de um rio seguem a distribuição normal (Guerreiro, 2015).

Uma v.a.  $Y$  segue uma distribuição normal com parâmetros  $E(Y)$  e  $\sigma^2$ ,  $Y \sim N(E, \sigma^2)$ , quando a função de densidade de probabilidade for da forma (Pishro-Nik, 2014),

$$f(y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{y - E(Y)}{\sigma}\right)^2\right), y \in \mathbb{R}, \quad (2.8)$$

Para  $Z = \frac{y - E(Y)}{\sigma}$ , tem-se a função de distribuição (Pishro-Nik, 2014):

$$f(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) dt = \Phi(z) \quad (2.9)$$

A função  $\Phi(z)$  está tabelada e pode ser consultada em Moreno e Morcillo (2019).

## II. Verificação da normalidade

A distribuição normal é, pelas suas propriedades, muitas vezes usada em análise estatística, porém, é necessário verificar os pressupostos de normalidade (Pishro-Nik, 2014). Existem diversos testes estatísticos para verificar a normalidade dos dados. Seier (2002) apresenta um estudo comparativo dos diversos métodos existentes, e o que melhor se comportou em detetar a não normalidade foi o teste W, desenvolvido por Shapiro e Wilk (1965). O valor da estatística do teste W, pode ser obtido através de softwares como o R, tornando o teste da normalidade, de um conjunto de dados, num processo simples e rápido de realizar.

Os métodos gráficos, como por exemplo, histogramas e o *Quantile Plot* (QQ-plot), são práticos e subjetivos, pelo que é aconselhável, utilizar, sempre que possível, testes estatísticos para complementar a verificação (Carinhas, 2014).

O *QQ-plot* é um método bastante prático para se avaliar a normalidade, como podemos verificar na Figura 2.2, compara-se, graficamente, no eixo dos  $xx$  representam-se os quantis teóricos (da distribuição normal), no eixo dos  $yy$  representam-se os quantis da amostra. Se as distribuições forem semelhantes então os pontos, no gráfico, vão situar-se junto da reta  $y = x$  (Torman et al., 2012).

Já o histograma, Figura 2.1, mostra o formato da distribuição e quanto mais próximo estiver da forma de sino mais indícios se tem de uma distribuição normal (Torman et al., 2012).

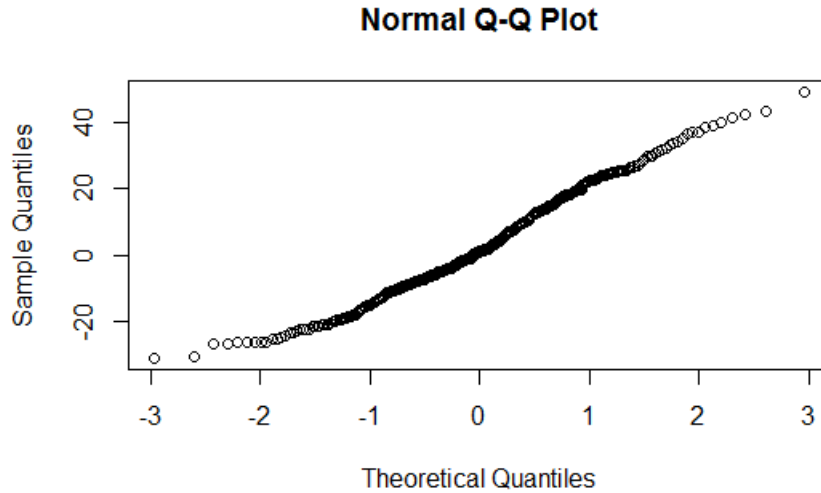


FIGURA 2.2: QQ-Plot de uma distribuição normal.(Fonte: [https://www.researchgate.net/figure/Normal-Q-Q-plot-Fit-of-the-log-normal-distribution-to-response-times-of-item-13-of-the\\_fig1\\_255730336](https://www.researchgate.net/figure/Normal-Q-Q-plot-Fit-of-the-log-normal-distribution-to-response-times-of-item-13-of-the_fig1_255730336), 05/04/2022)

### Distribuição de Gumbel

A distribuição de Gumbel é utilizada na modelação de valores máximos de ocorrência num fenómeno de interesse (Guerreiro, 2015). Nesta distribuição, os parâmetros que a indexam já não são os mesmos que na distribuição normal, onde se utilizam os parâmetros  $E(Y)$  e  $\sqrt{Var(Y)}$ . Neste caso, os parâmetros usados são a escala e a localização, onde estes determinam o local e a escala da curva da distribuição do gráfico (Guerreiro, 2015).

Uma v.a.  $Y$ , segue a distribuição de Gumbel, quando a sua f.d.p. é (Guerreiro, 2015):

$$f(\gamma|\lambda, \delta) = \frac{1}{\delta} \exp\left[-\frac{\gamma - \lambda}{\delta} - \exp\left(-\frac{\gamma - \lambda}{\delta}\right)\right], \quad -\infty < \gamma < +\infty, \lambda \in R, \delta > 0 \quad (2.10)$$

Segundo Gomes (2013), as principais características populacionais deste modelo são (Gomes et al., 2013):

$$E(Y) = \lambda + \epsilon\delta,$$

$$Var(Y) = \frac{(\pi\delta)^2}{6},$$

$$\text{Moda} = \lambda,$$

*Mediana* =  $\epsilon - \delta \log \log 2$  onde  $\epsilon = -\tau(1) = 0,5772156649\dots$  (constante de Euler).

É possível estimar, com grande facilidade, quantis elevados,  $Q(1 - p)$ , com  $p$  pequeno (Gomes et al., 2013):

$$Q(1 - p) = \hat{\lambda} - \hat{\delta} \ln(-\ln(1 - p)) \quad (2.11)$$

Na estimação do período de retorno para valores superiores ao valor elevado,  $x$ , pode ser obtido aproximadamente através da expressão (Gomes et al., 2013):

$$T_R(x) \approx \exp\left(\frac{x - \hat{\lambda}}{\hat{\delta}}\right) \quad (2.12)$$

### Distribuição Fréchet

A distribuição de Fréchet foi usada pela primeira vez em 1927, por Fréchet, na análise de frequência de caudais de cheia. Atualmente, tem sido aplicada na área de fenómenos atmosféricos, tais como temperaturas máximas e mínimas, e velocidade máxima dos ventos, bem como na área de hidrologia, no estudo de eventos extremos, nomeadamente em precipitações máximas (Guerreiro, 2015).

Uma v.a.  $Y$ , segue a distribuição de Fréchet,  $Y \sim F(\lambda, \delta, \rho)$  se a sua f.d.p. for dada pela expressão (Gomes et al., 2013):

$$(\gamma|\lambda, \delta, \rho) = \frac{\rho}{\delta} \left(\frac{\gamma - \lambda}{\delta}\right)^{-(\rho+1)} \exp\left(-\left(\frac{\gamma - \lambda}{\delta}\right)^{-\rho}\right), \quad (2.13)$$

$$\lambda < \gamma < +\infty, \lambda \in R, \delta > 0, \rho > 0$$

Neste caso  $\lambda$  é o parâmetro de localização,  $\delta$  é o parâmetro de escala e  $\rho$  é o parâmetro de forma. Segundo Gomes (2013) as principais características desta distribuição são (Gomes et al., 2013):

$$E(Y) = \lambda + \delta \tau\left(1 - \frac{1}{\alpha}\right), \quad \text{para } \alpha > 1,$$

$$Var(Y) = \delta^2 \left\{ \tau \left( 1 - \frac{2}{\alpha} \right) - \tau^2 \left( 1 - \frac{1}{\alpha} \right) \right\}, \quad \text{só existe se } \alpha > 2,$$

$$Moda = \lambda + \delta \left( 1 + \frac{1}{\alpha} \right)^{-1/\alpha},$$

$$Mediana = \lambda + \delta (\log 2)^{-1/\alpha}.$$

A obtenção dos quantis para a distribuição de Fréchet realiza-se através da f.d. inversa (Gomes et al., 2013):

$$y(F) = \lambda + (-\ln F)^{-1/\rho} \quad (2.14)$$

Ou através do período de retorno  $T_R$  (Gomes et al., 2013),

$$y(T) = \lambda + \delta \left( \ln \left( \frac{t}{T-1} \right) \right)^{-1/\rho} \quad (2.15)$$

### Distribuição de Max-Weibull

Esta distribuição foi detalhadamente descrita por Waloddi Weibull durante os seus estudos de resistência de materiais. Esta pode ser aplicada em análise de fiabilidade, pois graças à sua flexibilidade consegue-se adaptar à grande maioria das situações, permitindo assim modelar grandes variedades de dados para vários períodos de tempo (Guerreiro, 2015).

Uma v.a.  $Y$ , segue a distribuição de Max-Weibull,  $Y \sim W(\lambda, \delta, \rho)$ , quando a sua f.d.p. é (Guerreiro, 2015):

$$f(\gamma|\lambda, \delta, \rho) = \frac{\rho}{\delta} \left( \frac{\gamma - \lambda}{\delta} \right)^{\rho-1} \exp \left( - \left( \frac{\gamma - \lambda}{\delta} \right)^\rho \right), \quad (2.16)$$

$$\lambda < \gamma < +\infty, \lambda \in \mathbb{R}, \delta > 0, \rho > 0$$

De acordo com Gomes (2013), as principais características desta distribuição são:

$$E(Y) = \lambda + \delta \tau \left( 1 + \frac{1}{\alpha} \right),$$

$$Var(Y) = \delta^2 \left\{ \tau \left( 1 + \frac{2}{\alpha} \right) - \tau^2 \left( 1 + \frac{1}{\alpha} \right) \right\},$$

$$Moda = \lambda + \delta \left( 1 - \frac{1}{\alpha} \right)^{1/\alpha},$$

$$Mediana = \lambda + \delta (\log 2)^{1/\alpha}.$$

Para se calcular os quantis, realiza-se a inversa da função de distribuição (Gomes et al., 2013):

$$y(F) = \lambda + \delta (-\ln(1 - F))^{1/\rho} \quad (2.17)$$

## 2.1.3 Inferência Estatística

### Estimação de Parâmetros

#### I. Estimadores Pontuais

Os parâmetros de uma determinada distribuição são, geralmente, desconhecidos, nem são constantes pois é lógico que mudem ao longo do tempo (Requeijo e Pereira, 2012). No entanto existem métodos que permitem estimar os parâmetros de uma população e resolver outros problemas de inferência; a estimação pode ser feita recorrendo a estimadores pontuais ou a intervalos de confiança.

Apenas como referência, enumeram-se aqui quatro métodos para determinação dos estimadores pontuais: o método dos momentos, o método da máxima verosimilhança, o método de estimação linear com variância mínima e o método dos mínimos quadrados (para detalhar cada um dos métodos, podem ser consultados os livros de Murteira (1990) e Pestana e Velosa (2010)).

Os métodos referidos podem conduzir a estimadores diferentes, no entanto, um estimador deve possuir determinadas propriedades, designadamente (Requeijo e Pereira, 2012):

- não ser enviesado, ou seja, a média do estimador deve ser igual ao valor do parâmetro a estimar;
- ser eficiente, isto é, deve ter variância mínima para diminuir a variabilidade dos erros de dispersão;

- ser consistente, ou seja, quando a dimensão da amostra tende para infinito, o estimador toma o valor do parâmetro estimado.

Embora útil, a estimação pontual não tem grau de precisão associada, ou seja, não fornece informação sobre o afastamento da estimativa em relação ao verdadeiro valor do parâmetro (Requeijo e Pereira, 2012). Assim é preferível exprimir a estimativa por meio de dois valores (os limites de confiança) entre os quais se situará, com determinada probabilidade, o verdadeiro valor do parâmetro - estimação por intervalo (Pestana e Velosa, 2010).

## II. Estimação por Intervalos de Confiança

Um parâmetro pode ser estimado por um intervalo definido por duas estatísticas, no qual o verdadeiro valor do parâmetro estará contido com um determinado nível de confiança  $(1-\alpha)100\%$  (Murteira, 1990). Se  $\theta$  designar o parâmetro em causa, têm de ser definidas duas estatísticas LI e LS tais que (Murteira e Antunes, 2012):

$$P\{LI \leq \theta \leq LS\} = 1 - \alpha \quad (2.18)$$

O intervalo definido pela equação 2.18 será o intervalo de confiança a  $(1-\alpha)100\%$  para o parâmetro  $\theta$ , sendo LI e LS os limites de confiança (Murteira, 1990):

$$LI \leq \theta \leq LS \quad (2.19)$$

Se um número grande de intervalos for construído, cada um resultante de um conjunto de dados, é expectável que  $(1-\alpha)100\%$  deles contenha o verdadeiro valor de  $\theta$  (Murteira, 1990).

Para construir um intervalo de confiança é necessário definir um estimador do parâmetro, conhecer a respetiva distribuição e determinar uma estimativa pontual do parâmetro (Requeijo e Pereira, 2012). As distribuições amostrais anteriormente apresentadas desempenham aqui um papel fundamental. Para detalhar a estimação por intervalos de confiança, podem ser consultados os livros de Murteira (1990) e Pestana e Velosa (2010).

## Teste de Hipóteses

O teste de hipóteses é um método de inferência estatística que permite verificar, a partir dos resultados obtidos de um conjunto de dados, se determinada hipótese feita sobre a população deve ou não ser rejeitada (Requeijo e Pereira, 2012). Nesta secção apresenta-se, de forma genérica, a metodologia para a construção e análise de um teste de hipótese (para mais detalhe, podem ser consultados os livros de Murteira (1990) e Pestana e Velosa (2010)).

A hipótese que se pretende testar é a Hipótese Nula ( $H_0$ ), a qual contém sempre uma igualdade na sua formulação (Requeijo e Pereira, 2012). Assume-se que a Hipótese Nula é verdadeira ao longo do teste, até que haja uma evidência estatística que permita rejeitá-la; essa rejeição (ou não) é baseada numa estatística de teste adequada ao caso em questão (Requeijo e Pereira, 2012).

A Hipótese Alternativa ( $H_1$ ) é, como o nome indica, uma afirmação que constitui uma alternativa à Hipótese Nula. A rejeição de  $H_0$  reforça a validade da Hipótese Alternativa (Requeijo e Pereira, 2012). Note-se que a Hipótese Alternativa contém sempre uma desigualdade ( $<$  ou  $>$ ) na sua formulação: no primeiro caso (sinal de diferente) o teste é bilateral, enquanto que nos outros dois casos ( $<$  ou  $>$ ) será unilateral (Murteira, 1990).

Um teste hipóteses desenrola-se em 3 passos:

- Formular as hipóteses a testar;
- Calcular a estatística de teste;
- Decidir, o que consiste em rejeitar  $H_0$  se a estatística de teste (E.T.) estiver fora do intervalo definido pelos valores críticos da distribuição amostral da estatística de teste.

A Figura 2.3 ilustra o intervalo de rejeição de  $H_0$  para um teste bilateral. Note-se que se se rejeitar  $H_0$  conclui-se que  $H_0$  é falsa; porém se não se rejeitar  $H_0$ , não se pode concluir que esta hipótese é verdadeira, pode-se apenas concluir que não existe evidência (estatística) para rejeitar  $H_0$  (Maroco, 2007). O nível de significância habitual é  $\alpha=0,05$  (5%) mas outros níveis de significância usados com frequência são 0,1 e 0,01 (Maroco, 2007).

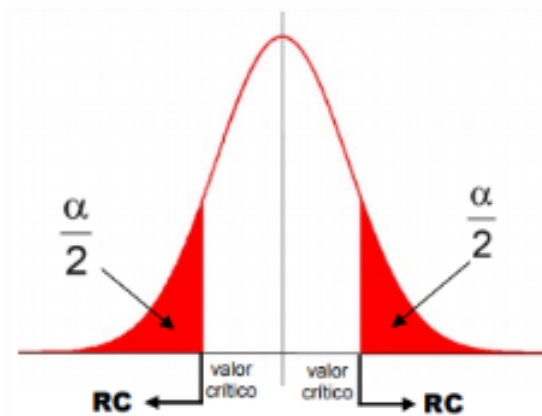


FIGURA 2.3: Teste Bilateral de uma distribuição (Fonte: [https://rstudio-pubs-static.s3.amazonaws.com/557755\\_bb5f2a0dae7b4ba7afd91de4310756ec.html](https://rstudio-pubs-static.s3.amazonaws.com/557755_bb5f2a0dae7b4ba7afd91de4310756ec.html), 04/052022).

Ao menor valor de  $\alpha$  a partir do qual se rejeita  $H_0$  chama-se probabilidade de significância ( $p$  ou  $p$ -value); a probabilidade de significância é então a probabilidade de se obter o valor da estatística de teste ou um valor mais extremo se a  $H_0$  for verdadeira (Maroco, 2007). Segundo Fisher e Bishel (1973), se o  $p$ -value for menor que o nível de significância ( $\alpha$ ) então a evidência contra a hipótese  $H_0$  é significativa. A utilização dos  $p$ -values é alvo de várias críticas (por exemplo Hubbard e Armstrong, (2006)).

Neyman e Pearson (1933), propuseram, para contrariar a interpretação subjetiva do  $p$ -value, que a rejeição ou não rejeição de  $H_0$  tem sempre associada um determinado risco ou erro. Estes conceitos são resumidos na Tabela 2.1:

TABELA 2.1: Interpretação do  $p$ -value para o Teste de Hipóteses.

Verdade	Decisão	
	Aceitar $H_0$	Rejeitar $H_0$
$H_0$ Verdadeiro	$(1 - \alpha)$	Erro Tipo I ( $\alpha$ )
$H_0$ Falso	Erro Tipo II ( $\beta$ )	$(1 - \beta)$

Assim, fixado à partida, o erro de tipo I aceitável e conhecida a distribuição amostral da estatística de teste, é possível construir uma região de rejeição que, sob a validade de  $H_0$ , tem uma probabilidade  $\alpha$  de conter o valor observado para a estatística de teste (Maroco, 2007). Conhecidas as distribuições amostrais sob a validade de  $H_0$  e de  $H_1$  é possível determinar  $\alpha$  e  $\beta$  como ilustra a Figura 2.4:

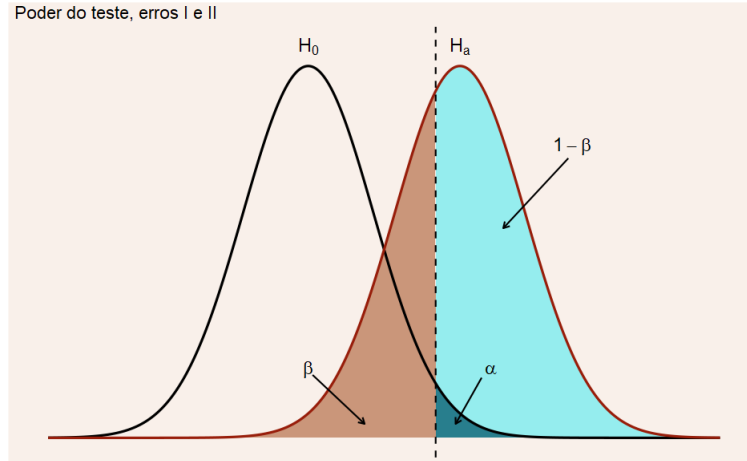


FIGURA 2.4: Intervalo de rejeição de  $H_0$  para um teste bilateral (Fonte: <https://www.inf.ufsc.br/~andre.zibetti/probabilidade/teste-de-hipoteses.html>, 04/04/2022).

Atualmente, a maioria dos softwares calcula apenas o  $p$  - *value* e para um determinado nível de significância ou probabilidade de erro de tipo I ( $\alpha$ ) deve-se rejeitar  $H_0$  se  $p$  - *value*  $\leq \alpha$  (Fisher e Bischel, 1973).

## 2.2 Séries Temporais

Uma Série Temporal é definida como um conjunto de observações sobre uma variável ordenada no tempo, como por exemplo, alturas de maré, desvios entre a maré observada e a maré astronômica. O estudo deste tipo de dados é realizado com a ambição de conhecer o fenômenos em estudo, tendo como objetivo final, se possível, realizar previsões com elevada precisão. Muitos destes fenômenos estão sujeitos a fatores aleatórios, como perturbações causadas pela meteorologia nas marés, o que torna a caracterização e previsão destes fenômenos muito complexos, exigindo o recurso à estatística (Xavier, 2016).

Desta forma pode-se dizer que estudar estatisticamente uma série temporal é: descrever o comportamento de uma série, procurar periodicidade nos dados, investigar o mecanismo que gera a série, realizar previsões, quer a curto prazo quer a longo prazo. Exemplo disto, é a análise de alturas de ondas para se descobrir como foram formadas (Morettin e Toloï, 2004).

Uma vez que as séries temporais são observações medidas ao longo do tempo, estas podem ser realizadas num intervalo de amostragem contínuo ou discreto, ou seja, ou são realizadas ao longo de um determinado período de tempo, um caso representativo disto é a medição do caudal de um rio durante 1 ano:

$$T = (t : t_1 < t < t_2), \quad T = [0, 12]$$

ou através de medições separadas por intervalos de tempo pré-definidos:

$$T = (t_1, t_2, \dots, t_n), \quad T = [01/2021, \dots, 12/2021]$$

Este tipo de séries podem ser classificadas como univariadas, quando os registos são realizados em intervalos de tempo iguais (dias, meses, anos), ou como multivariadas, quando se faz o registo de múltiplas observações, por exemplo as vendas semanais de uma firma  $Y_1(t)$  e o valor investido na publicidade  $Y_2(t)$  (Xavier, 2016).

### 2.2.1 Séries Temporais Estacionárias e não Estacionárias

Um processo estocástico é constituído, geralmente, por uma família de variáveis aleatórias. Estes processos podem ser representados sob a forma de série temporal, sendo descrito como um conjunto de v.a. indexadas no tempo, onde as variáveis não estão espaçadas no tempo de igual forma (Xavier, 2016).

Um processo estocástico  $y(t)$  é estacionário de ordem  $n$  quando a sua f.d.p., de ordem  $n$ , não varia no tempo, tal como demonstra a função 2.20 (Xavier, 2016):

$$p_{y_{t_1}, y_{t_2}, \dots, y_{t_n}}(Y_1, Y_2, \dots, Y_n) = p_{y_{t_1+k}, y_{t_2+k}, \dots, y_{t_n+k}}(Y_1, Y_2, \dots, Y_n), \forall k \quad (2.20)$$

Um processo estacionário pode ser descrito de duas formas, no sentido estrito ou no sentido débil (Xavier, 2016):

- **Sentido estrito:** quando é estacionário para qualquer valor de  $k$ .

$$F(y_i, y_j, \dots, y_k) = F(y_i, y_j, \dots, y_k)$$

- **Sentido débil:** neste caso o primeiro e o segundo dado das séries são constantes, no entanto a covariância entre as variáveis depende da sua separação temporal (Xavier, 2016):

$$E(y_t) = \mu$$

$$Var(y_t) = \sigma^2$$

$$Y(t, t - k) = y(t, t + k) = y(k) \quad \text{com } k = 0, \pm 1, \pm 2, \dots$$

Uma série temporal é não estacionária quando a variância ou a média variam ao longo do tempo, gerando piores previsões quando comparadas com as séries estacionárias. As propriedades de um conjunto finito de dados  $(y_1, y_2, \dots, y_n) \in Y_t$  são diferentes de um conjunto do tipo  $(y_1, y_2, \dots, y_{n+h}) \in Y_t$  (Xavier, 2016):

- **Serie não estacionária na média:** o nível da série não é estável ao longo do tempo, podendo crescer ou decrescer.
- **Série não estacionária na variância:** a variância varia no tempo.

### 2.2.2 Tendência e Sazonalidade

Muitas séries temporais podem ser representadas como a soma de uma tendência, uma componente sazonal, e uma componente aleatória, conforme descrito seguidamente. A tendência define o comportamento da série a longo prazo, é perceptível quando uma série demonstra crescimento ou declínios sucessivamente (Oliveira, 2009).

Considerando que uma série que não apresenta componente sazonal, onde  $a_t$  é uma v.a. com média de zero e  $T_t$  o estimador da tendência, tem-se:

$$Z_t = T_t + a_t \tag{2.21}$$

Uma vez estimada a tendência  $\hat{T}_t$  ajusta-se a série:

$$Y_t = Z_t - Z_{t-1} \tag{2.22}$$

A maioria dos testes usados para detetar tendências nas séries temporais requerem que os dados sejam independentes (Oliveira, 2009), como exemplo de testes para detetar a tendência de uma série, tem-se o teste de *Wald-Wolfowitz*, Teste de

*Cox-Stuart*, Teste de *Mann-Kendall* e Teste de *Spearman*. Para mais detalhes sobre os testes, pode-se consultar Xavier (2016).

A variação sazonal é um fenómeno que ocorre, com regularidade, ou seja, é um fenómeno que de ano para ano ocorre sempre no mesmo período, como por exemplo o aumento de vendas de bilhetes de avião durante o verão ou o aumento consumismo durante a época natalícia (Oliveira, 2009).

A sazonalidade divide-se em sazonalidade determinística, quando a série pode ser prevista com grande precisão através de dados passados, neste caso o padrão sazonal é estável no tempo, e sazonalidade estocástica, quando o padrão sazonal varia ao longo do tempo (Oliveira, 2009).

## 2.3 Teoria de Valores Extremos

A teoria de valores extremos consiste no estudo das propriedades dos dados situados nas caudas das distribuições. Esta tem grande importância na avaliação e modelação do risco de eventos raros e extremos, como por exemplo chuvas torrenciais, vagas de calor e inundações de grande dimensão. Estes eventos têm uma probabilidade de ocorrência muito baixa, porém, quando ocorrem originam estragos severos. Existem diversos métodos para se estudar estes eventos, os mais usuais são: a abordagem segundo o primeiro teorema da Teoria de Valores Extremos (Fisher e Tippett, 1928; B. Gnedenko, 1943) e a abordagem através do segundo teorema da Teoria de Valores Extremos sobre os excessos de nível (Pickands, 1975; Balkema e de Haan, 1974).

A teoria de valores extremos modela os valores extremos usando a distribuições de limite, máximo ou mínimo, ou através dos excessos de um nível pré-determinado (Neves, 2021). Neste projeto, para se estimar quais as distribuições que melhor se ajustavam, usou-se a análise através da Generalizada de Valores Extremos e a análise *Peaks-Over-Threshold*.

O estudo da distribuição de valores extremos é bastante popular na estatística, esta dedica-se ao estudo de séries estocásticas de variáveis aleatórias e independentes. Após a demonstração de Nordquist's (1945) de que os maiores terremotos na Califórnia vão ao encontro da teoria de valores extremos, esta teoria começou a ser aplicada em outras áreas, nomeadamente no estudo de sobrelevações, tendo como exemplo disto o artigo sobre *Sea level extremes in southern Europe* de Marcos (2009). Outros autores também trabalharam na teoria de valores extremos, tais como: Jenkinson (1955), e Gringorten (1963), que analisaram extremos

meteorológicos e Krumbein e Lieblein (1956) analisaram a distribuição de pedras extraordinariamente grandes em depósito de cascalho.

Sempre que se trata de cenários de catástrofe é preferível estar preparado para se enfrentar essas adversidades, para isto é necessário ter-se um previsão de quando um evento deste tipo poderá acontecer, ou seja, o seu período de retorno. O período de retorno,  $T_R$ , é o tempo médio que um acontecimento demora a realizar-se consecutivamente, tendo este acontecimento  $p$  como probabilidade (Neves, 2021).

$$T_R = \frac{1}{p} \quad (2.23)$$

Considere-se as 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.).

Se a amostra for ordenada por ordem crescente, as correspondentes estatísticas ordinais (e.o.'s) poderão ser denotadas por,  $X_{1:n}, X_{2:n}, \dots, X_{n:n}$ . 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  $X_1, X_2, \dots, X_n$  (Guerreiro, 2015). Neste trabalho, ir-se-á dar destaque aos máximos obtidos na série de alturas horárias de maré.

Através da teoria de extremos, estima-se níveis elevados, associados a um determinado acontecimento aleatório, que tenham uma probabilidade muito reduzida de se exceder o valores típicos. Neste contexto, é interessante calcular o período de retorno de um valor extremo de um dado acontecimento aleatório.

A partir do teorema de Fisher-Tippett-Gnedenko estabelece-se uma função de distribuição, equação 2.24, do máximo de amostras grandes o suficiente que pode ser aproximada por uma função de distribuição que pode ser apresentada de três formas: Weibull, Fréchet e Gumbel conhecidas por leis de máximos (Guerreiro, 2015).

A grande dificuldade em qualquer análise de valores extremos é o tamanho da amostra em estudo, o que acarreta consequências na previsão que se pretende obter (Guerreiro, 2015).

Von Mises (1936; 1954) e Jenkinson (1955), através das funções de distribuição (f.d.) de Gumbel, Fréchet e Max-Weibull (cujas distribuições foram apresentadas na subsecção 2.1.2), apresentaram a distribuição da generalizada de valores extremos, com a f.d. expressa da seguinte forma:

$$G_{\gamma}(x) = \begin{cases} \exp\left(- (1 + \gamma x)^{-\frac{1}{\gamma}}\right), & 1 + \gamma x > 0, \gamma \neq 0 \\ \exp(-\exp(-x)), & x \in \mathbb{R}, \gamma = 0 \end{cases} \quad (2.24)$$

Com  $x_+ := \max(0, x)$  e  $\gamma$ , índice de valores extremos, o parâmetro fundamental em Estatística, Figura 2.5.

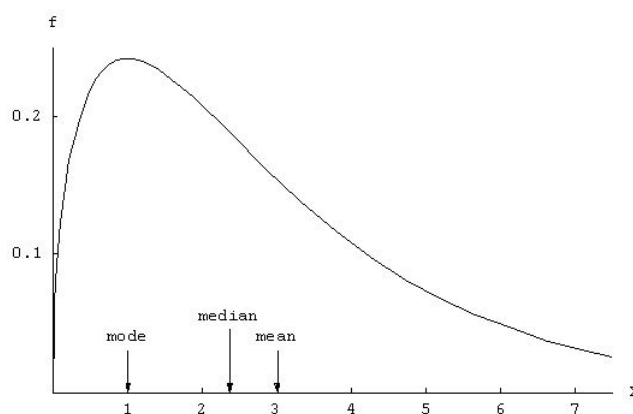


FIGURA 2.5: Cauda direita de uma distribuição F.(Fonte: <http://soniavieira.blogspot.com/2018/05/por-conta-de-umapergunta-sobre-questao3.html>, 07/05/2022)

Segundo Gomes et al. (2013), se:

- $\gamma = 0$ , obtém-se a função de distribuição de Gumbel, diz-se que se tem uma cauda exponencial;
- $\gamma > 0$ , obtém-se a função de distribuição de Fréchet, diz-se que se tem uma cauda pesada;
- $\gamma < 0$ , obtém-se a função de distribuição de Max-Weibull, diz-se que se tem uma cauda exponencial.

A partir da fórmula de distribuição de GEV, com a amostra devidamente normaliza, é possível verificar que esta engloba as três famílias de valores extremos (Gomes et al., 2013):

$$G_\gamma(x; \lambda, \delta) = \exp\left(-\left[1 + \gamma\left(\frac{x - \lambda}{\delta}\right)_+^{-1/\gamma}\right]\right) \quad (2.25)$$

**Família I** – Gumbel:

$$\wedge(x; \lambda, \delta) = \exp\left(-\exp\left(-\frac{x - \lambda}{\delta}\right)\right), x \in R \quad (\gamma = 0) \quad (2.26)$$

**Família II** – Fréchet:

$$\Phi_\alpha(x; \lambda, \delta) = \exp\left(-\left(\frac{x - \lambda}{\delta}\right)^{-\alpha}\right), x > \lambda \quad (\alpha > 0) \quad e \quad \left(\gamma = \frac{1}{\alpha} > 0\right) \quad (2.27)$$

**Família III** – Max-Weibull:

$$\Psi_\alpha(x; \lambda, \delta) = \exp\left(-\left(-\frac{x - \lambda}{\delta}\right)^\alpha\right), x < \lambda \quad (\alpha > 0) \quad e \quad \left(\gamma = -\frac{1}{\alpha} < 0\right) \quad (2.28)$$

Assim Gumbel (1958) expôs o primeiro modelo em estatística de extremos, modelo dos máximos anuais (MMA), ou modelo de GEV univariado ou modelo de Gumbel. De acordo com Gumbel, divide-se a amostra de dados,  $N$ ,  $(X_1, \dots, X_N)$ , em  $m$  sub-amostras de dimensão  $n$  ( $N = nm$ ) e ajusta-se um dos modelos de extremos (Família I, II ou III) ou o modelo de GEV a esta amostra formada pelos  $m$  máximos de cada sub-amostra, associada à variável aleatória  $Y = \max(X_1, \dots, X_N)$ .

Alguns acontecimentos raros estão associados ao ajustamento do modelo GEV para  $Y$  (Gomes et al., 2013):

- Probabilidade de excedência de níveis elevados  $u$ :

$$1 - G_\gamma(u; \lambda, \delta) = \begin{cases} 1 - \exp\left\{-\left[1 + \gamma\left(\frac{u - \lambda}{\delta}\right)\right]^{-\frac{1}{\gamma}}\right\}, & \text{se } \gamma \neq 0 \\ 1 - \exp\left\{-\exp\left[-\frac{u - \lambda}{\delta}\right]\right\}, & \text{se } \gamma = 0 \end{cases} \quad (2.29)$$

- Período de retorno do nível  $u$ :

$$T_u = \frac{1}{1 - G_\gamma(u; \lambda, \delta)} \quad (2.30)$$

- Nível de retorno em anos ( $T$ ):

$$U(T) = G_\gamma^{-1}\left(1 - \frac{1}{T}; \lambda, \delta\right)$$

$$U(T) = \begin{cases} \lambda + \frac{\delta}{\gamma} [(-\log(1-p))^{-\gamma} - 1], & \text{se } \gamma \neq 0 \\ \lambda - \delta \log(-\log(1-p)), & \text{se } \gamma = 0 \end{cases} \quad (2.31)$$

A modelação da distribuição GEV para uma amostra depende de métodos adequados na estimação dos seus parâmetros, o que, em muitas situações, não é possível. De modo a contornar este problema, realiza-se um teste para se decidir qual dos três modelos, Gumbel, Fréchet ou Max-Weibull, se adapta melhor à amostra em estudo. A partir disto ajusta-se uma das três distribuições à amostra estimando-se os parâmetros localização, escala e forma,  $(\lambda, \delta, \rho)$ , através do método da máxima verosimilhança. Para melhor compreensão deste método pode-se consultar o livro *Análise de valores extremos: Uma introdução*, de Ivette Gomes, publicado em 2013 (Gomes et al., 2013).

### Distribuições Limite para os Excessos - Análise POT

O objetivo principal da análise de valores extremos é modelar o risco de eventos raros, estimando quando estes poderão ocorrer novamente (Davison e Smith, 1990). Este tipo de análise estatística pode ser realizado através do método de *Peak-Over-Threshold* (POT).

Considere-se um conjunto de dados aleatórios e identicamente distribuídos,  $\{x_1, \dots, x_n\}$ , os acontecimentos extremos deste conjunto de dados são identificados através de um *threshold*,  $\mu$ , assim encontra um subconjunto de dados, as excedências  $(x_i : x_i > \mu)$  (Rosário, 2013).

Balkema e Haan (1974) e Pickands (1975) provaram que a distribuição Generalizada Pareto (GP) é uma aproximação para a distribuição dos excessos acima de um *threshold* elevado  $\mu$ . Assim tem-se que a função de distribuição da GP é:

$$H_\gamma(x|\mu, \sigma_\mu) = \begin{cases} 1 - \left(1 + \gamma \frac{x - \mu}{\sigma_\mu}\right)^{\frac{-1}{\gamma}}, & x \geq \mu, \quad \text{se } \gamma > 0, \\ 1 - \exp\left(-\frac{x - \mu}{\sigma_\mu}\right), & x \geq \mu, \quad \text{se } \gamma = 0, \\ 1 - \left(1 + \gamma \frac{x - \mu}{\sigma_\mu}\right)^{\frac{-1}{\gamma}}, & \mu \leq x \leq \mu - \frac{\sigma_\mu}{\gamma}, \quad \text{se } \gamma < 0 \end{cases} \quad (2.32)$$

Para  $\gamma > 0$ ,  $\gamma = 0$  e  $\gamma < 0$  a função distribuição GP reduz-se ao modelo Pareto, Exponencial e Beta, respetivamente. O *threshold*,  $\mu$ , é o parâmetro de localização,  $\sigma_\mu$  o parâmetro de escala e  $\gamma$  o parâmetro de forma (Rosário, 2013). A escolha de um *threshold* origina maior dificuldade em garantir que os dados são independentes, ao contrário da GEV onde se tem amostras com comprimento fixo, garantindo a independência (Rosário, 2013).

### Escolha de um *Threshold*

A escolha de um *threshold*,  $\mu$ , adequado é bastante importante na metodologia POT. Porém, a escolha do *threshold* é uma questão bastante sensível pois, se se escolher um valor muito pequeno para  $\mu$ , ter-se-á estimadores com menor variância, mas o seu viés será maior (Rosário, 2013). Se for escolhido um valor elevado para  $\mu$ , a variância dos estimadores será maior e o viés será menor. Assim, existe uma necessidade de se encontrar um valor para  $\mu$  que tenha estimadores com valores baixos quer para a variância e quer para o viés (Rosário, 2013). Existem 3 métodos, que me regra são aplicados, para a escolha de um *threshold* (Gong, 2012).

O método mais simples para a escolha de um *threshold* é o valor de um determinado quartil empírico de probabilidade a variar entre 0.9 e 0.97 (Rosário, 2013). O segundo método mencionado em Gong (2012) tem por base Davison e Smith (1990), onde a escolha de  $\mu$  é realizada através do estudo da função de excesso médio:

$$g(t) = E[X - t|X > t] := \frac{\int_t^{x^F} (1 - F(s))ds}{1 - F(t)}, \quad \forall t < x^F \quad (2.33)$$

No caso da Distribuição generalizada de Pareto, a função de excesso médio assume a expressão (Rosário, 2013):

$$e(\mu) = E[X - \mu | X > \mu] = E[Y | Y > 0] = \frac{\sigma_\mu + \gamma\mu}{1 - \gamma}, \quad \text{se } \gamma < 1 \quad (2.34)$$

Na prática,  $e(\mu)$  é substituída pela sua contrapartida empírica,  $\hat{e}_n(\mu)$ , expressa por (Rosário, 2013):

$$\hat{e}_n(x_{n-k:n}) = \frac{1}{k} \sum_{j=1}^k x_{n-j+1:n} - x_{n-k:n} \quad (2.35)$$

É de notar que a função de excesso médio, para o caso de distribuição GP, é linear em  $\mu$ . Deste modo, quando apresentado o ME-plot empírico, obtido através de (Rosário, 2013):

$$(x_{n-k:n}, \hat{e}_n(x_{n-k:n})), \quad 1 \leq k \leq n - 1 \quad (2.36)$$

Espera-se uma linearidade à direita de  $X_{n-k:n}$ , o *threshold* escolhido deverá ser o ponto a partir do qual se verifica essa linearidade (Rosário, 2013).

O terceiro método referido em Gong (2012), consiste na escolha de candidatos a *threshold* e para cada um deles calcularam-se os parâmetros do modelo GP. O valor adequado para  $\mu$ , é aquele a partir do qual os estimadores dos parâmetros forma e escala se tornam estáveis. Neste caso, se a distribuição GP com parâmetros de forma  $\lambda$  e de escala  $\sigma_{\mu_0}$  for a correta para os excessos acima de  $\mu_0$ , então, tem-se que, para qualquer valor de  $\mu > \mu_0$ , os novos excessos terão para a distribuição GP o mesmo parâmetro forma, mas com o parâmetro escala alterado para (Rosário, 2013):

$$\sigma_\mu = \sigma_{\mu_0} + \gamma(\mu - \mu_0) \quad (2.37)$$

Este método é bastantes útil para a escolha de um *threshold* apropriado pois, recorre a gráficos, tornando a visualização do comportamento dos parâmetros em função de  $\mu$  mais fácil (Rosário, 2013).

## 2.4 Variância da Maré

### 2.4.1 Maré Astronômica

A maré é a oscilação quase-periódica do nível da água do mar, isto devido essencialmente à atração gravítica entre os sistemas Terra-Lua e Terra-Sol, e à força centrífuga associado ao movimento de rotação da Terra (Trujillo e Thurman, 2010).

A maré observada pode ser descrita como a soma das componentes maré astronômica e maré meteorológica. A maré astronômica pode ser prevista com rigor, pois esta advém das alterações gravíticas exercidas pela Lua e o Sol sobre a Terra. O princípio de Laplace permite a maré astronômica num determinado ponto relativamente a uma maré de referência, esta denomina-se de maré de equilíbrio, que está diretamente relacionada com as forças geradoras de maré (Araújo, 2013). Já a maré meteorológica é a alteração da maré prevista, com causa meteorológica, como por exemplo variações da pressão e vento, do nível das águas (Webb, 2017).

A diferença entre a maré observada e a maré astronômica, é denominada de desvio ou maré meteorológica, pois advém essencialmente a fatores meteorológicos, como sejam as variações da pressão, da velocidade e direção do vento (Hidrográfico, 2021).

A maré astronômica é conhecida e prevista com rigor, esta nas grandes bacias oceânicas propaga-se como uma onda rotativa, influenciada pela força de Coriolis, que é uma força de deflexão devida ao movimento de rotação da Terra em torno do seu eixo (Araújo, 2013). No entanto, a onda de maré é perturbada ao longo do oceano por vários fatores, nomeadamente a diversidade dos fundos oceânicos, a profundidade e a irregularidade dos continentes (Webb, 2017).

Segundo o princípio de Laplace, a maré astronômica num dado local pode ser expressa como a soma de constituintes sinusoidais, com frequências conhecidas a partir da força geradora da maré (ou da maré de equilíbrio), porém as amplitudes e fases das constituintes têm que ser calculadas a partir de observações (Araújo, 2013).

Devido a todos estes fatores, o movimento das marés realiza-se de forma circular em torno de um ponto central denominado de ponto anfidrômico, onde a amplitude maré é nula, existem diversos pontos anfidrômicos ao longo das bacias oceânicas, constituindo, assim, um sistema anfidrômico, tal como esta ilustrado na Figura 2.6.

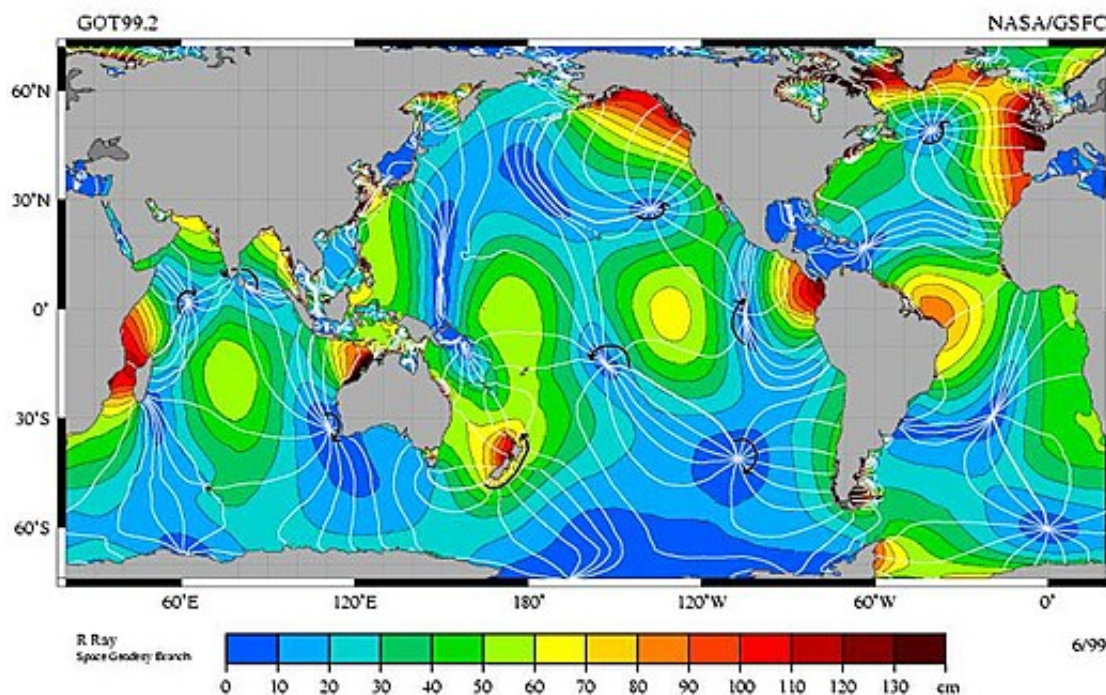


FIGURA 2.6: Esquema dos sistemas anfidrômicos nos oceanos.  
 (Fonte: [https://pt.wikipedia.org/wiki/Ponto\\_anfidr%C3%B3mico](https://pt.wikipedia.org/wiki/Ponto_anfidr%C3%B3mico)),  
 23/03/2022.

No entanto, as ondas de maré propagam-se por conservação de energia, à medida a que se aproximam de costa e a profundidade vai diminuindo, a altura da onda de maré sofre um aumento, exemplo disto é a variação de maré na Baía de Fundy onde a amplitude de maré atinge os 17 metros. O mesmo não se sucede na direção da bacia oceânica, onde a altura da onda de maré diminui com o aumento da profundidade. Em geral, devido ao Efeito de Coriolis, este movimento circular das ondas de maré dá-se no sentido contrário aos dos ponteiros do relógio no Hemisfério Norte e no sentido dos ponteiros do relógio na Hemisfério Sul (Webb, 2017).

As correntes de maré são de igual forma afetadas pelos fatores referidos anteriormente, seguem o padrão rotativo das ondas de maré, quanto maior a distância a costa mais fraca é a corrente de maré e quanto menor a distância a costa mais forte é esta corrente (Doodson e Warburg, 1946). É importante para navegação a direção e velocidade da corrente de maré, deste modo tem-se em conta qual o efeito que irá ter no movimento dos navios, sabendo que a velocidade máxima que estas atingem ocorre a meia maré (Reis, 2008).

### Força Geradora de Maré

Em 1687, no "*Philosophiae Naturalis Principia Mathematica*", Isaac Newton descreveu os fundamentos da gravitação universal, desvendando que a força Gravitica,  $F_g$ , entre dois corpos é proporcional ao produto das massas e inversamente proporcional ao quadrado da distância entre eles (Hidrográfico, 1984).

Quando aplicada ao sistema Terra-Lua, a Lei de Newton toma a forma de (Hidrográfico, 1984):

$$F_g = G \frac{m_t m_L}{R_{TL}^2} \quad (2.38)$$

Com  $G = 6.67 \times 10^{-11} \text{ N KG}^{-2} \text{ m}^2$

Uma vez que nem todos os pontos da superfície terrestre estão a mesma distância da Lua, a força gravítica varia para vários pontos da superfície terrestre. O ponto da Terra que se encontra mais próximo da Lua é o que sente maior força de atração (Zénite) e por sua vez o ponto da Terra mais distante da Lua é o que sente menor força de atração (Nadir). A força geradora de maré pode ser decomposta em forças gravíticas, exercida pelo Sol e pela LUA, e em forças centrífugas, devido ao movimento dos sistemas Terra- Lua e Terra- Sol em torno do centro de massa comum, denominado baricentro (Doodson e Warburg, 1946; Hidrográfico, 1984)

Considere-se o sistema Terra-Lua, representado na Figura 2.7.

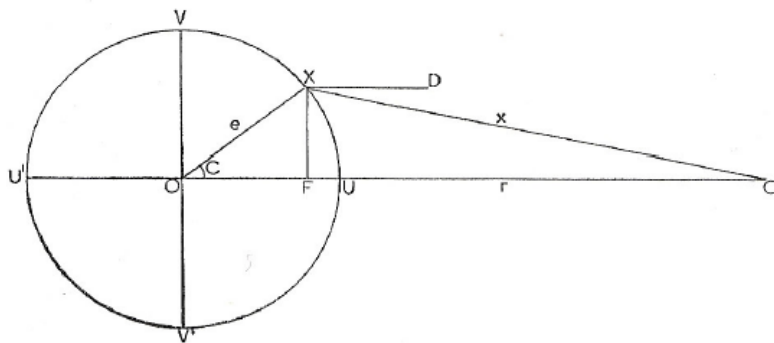


FIGURA 2.7: Esquema das Forças atrativas. (Extraído de Reis, 2008)

O, O' – Centros da Terra e da Lua, respetivamente

U, U' – Zénite e Nadir, respetivamente

X – Ponto à superfície da Terra

- V, V' – Pontos em quadratura com U, U'  
 r – Distância entre O, O'  
 e – Raio da Terra (aproximação esférica)  
 x – Distância do ponto X ao centro da Lua  
 C – Ângulo XOO'  
 $m_L$  – Massa da Lua

A força geradora de maré é traduzida pela diferença entre a força exercida num ponto á superfície da Terra e a força exercida sobre o centro da Terra. A aceleração da força geradora de maré, expressa na equação 2.39, é dada pela diferença entre a aceleração no ponto X e a aceleração no ponto O.

$$\vec{a} = \vec{a}_X - \vec{a}_O = Gm_L \left( \frac{\vec{x}}{x^3} - \frac{\vec{r}}{r^3} \right) \quad (2.39)$$

A partir da formula anterior é possível verificar que a força geradora da maré é proporcional à massa do astro atrator (Lua ou Sol) e ao raio da Terra (ou do astro sobre o qual a força geradora da maré se faz sentir), e inversamente proporcional ao cubo da distância entre os centros de massa dos dois astros (Doodson e Warburg, 1946; Hidrográfico, 1984).

### Análise Harmónica

As constantes mudanças das declinações do Sol e da Lua, as variações cíclicas da posição da Lua e do Sol em relação à Terra e as variações de longo período, como o período de 18,613 anos correspondente à regressão dos nodos lunares, são fatores que alteram a força geradora de maré. No entanto, os movimentos destes dois astros são conhecidos com precisão, pelo que é possível expressar a força geradora de maré através da soma de constituintes harmónicas simples, sendo que cada constituinte tem o seu período, fase e amplitude característicos, para uma consulta mais aprofundada ver Doodson e Warburg (1946).

O período de cada constituinte é determinado a partir do estudo astronómico, no entanto, a amplitude e a fase da resposta da maré a cada constituinte é impossível de determinar através da força geradora. Para a resolução deste problema recorre-se à análise harmónica (Hidrográfico, 1984).

A análise harmónica é um processo matemático que permite efetuar a análise e previsão da maré astronómica. Esta é usada para representar uma série temporal irregular na forma da soma das curvas sinusoides com amplitudes e fases variadas. Este processo matemático é realizado através da fórmula harmónica, expressa na equação 2.40 (Doodson e Warburg, 1946).

$$h(t) = NM + \sum_i^n f_i H_i \cos[(V_0 + u)_i + w_i t - g_i] \quad (2.40)$$

$NM$ : Nível médio das águas do mar

$f_i$ : Fator Nodal

$H_i$ : Amplitude (constante harmónica)

$W : i$ : Velocidade angular da constituinte ( $^{\circ}/h$ )

$t$ : Tempo desde o instante do primeiro valor a prever

$V_0$ : Argumento inicial da onda na maré de equilíbrio

$u$ : Correlação ao argumento inicial

$g_i$ : Fase (constante harmónica)

$n$ : Total de constituintes consideradas

A formula harmónica tem por base a teoria da maré de equilíbrio . A maré de equilíbrio é uma maré teórica usada para estudar o comportamento das marés, esta assume a Terra como uma esfera coberta por um único oceano de profundidade uniforme e que as massas de água correspondem instantaneamente às variações da força geradora de maré, desconsiderando a existência de continentes, a força de Coriolis induzida pelo movimento de rotação da terra e a topografia dos fundos oceânicos. Para um estudo mais aprofundado da teoria da maré de equilíbrio consultar Doodson e Warburg (1946).

A maré astronómica pode ser decomposta em componentes sinusoidais com frequências conhecidas, denominadas de constituintes harmónicas, obtidas a partir do conhecimento rigoroso das posições da Lua e do Sol relativamente ao centro de massa da Terra. A maré de equilíbrio é transformada em maré local quando sofre alterações devido a efeitos locais. Estes efeitos alteram a amplitude e a fase da onda de maré, que são os parâmetros que se pretendem obter através do estudo da análise harmónica. A amplitude e a fase de cada onda de maré denomina-se constantes harmónicas (Godinho, 2011). Para um estudo mais aprofundado sobre as constituintes e constantes harmónicas consultar Doodson e Warburg (1946).

A análise harmónica tem por objetivo a determinação das constantes Harmónica, ou seja, a amplitude e fase das constituintes harmónicas, para depois através

da fórmula harmónica se produzir previsões da maré para um dado local.

## 2.4.2 Maré Meteorológica

Como descrito na secção 2.4, a maré meteorológica é a variação do nível da água do mar associado a efeitos sazonais e meteorológicos, tais como as alterações do caudal de um rio, as variações de pressão, os ventos e periodicidades mal definidas (Hidrográfico, 1984). Estes fatores meteorológicos criam desvios entre a altura e o tempo da maré observada e da maré prevista.

A altura de água,  $X(t)$ , observada num determinado local, pode ser representada por uma expressão semelhante a (Pugh, 1987):

$$X(t) = A_0(t) + T(t) + S(t) \quad (2.41)$$

Onde,  $A_0(t)$  é o nível médio do mar no ponto de observação e que varia lentamente com o tempo,  $T(t)$  é a variação da maré astronómica e  $S(t)$  corresponde à variação da altura de água devido a efeitos meteorológicos. De facto, o movimento regular da maré é constantemente influenciado, a um menor ou maior grau, por efeitos meteorológicos impossíveis de prever a longo prazo. Em certas regiões do globo, inundações costeiras intensas podem-se verificar quando há a coincidência de fenómenos meteorológicos extremos, como tempestades tropicais ou tufões, com alturas de maré astronómica elevada.

A maré meteorológica, ao contrário da maré astronómica, não é prevista com elevada exatidão, pois enquanto que na maré astronómica, através das forças geradoras de maré e das observações, é possível realizar previsões precisas, na maré meteorológica tal não é possível (Hidrográfico, 2021).

### Efeitos das Condições Meteorológicas

O vento tem grande impacto, quando perto de costa, no nível da água, pois ventos que provém do oceano direcionam a água para terra o que resulta na elevação da altura de maré. O contrário também acontece quando o vento empurra a água na direção do oceano onde causa alturas de maré abaixo das previstas (Hidrográfico, 2021).

Quanto menor for a profundidade maior será o impacto do vento. O caso de a direção do vento ser paralela a costa, é o evento onde a probabilidade de acontecer

um fenómeno extremo, da altura de maré, é mais elevada, pois nestas condições existe uma maior intensidade nas forças que deslocam a água em direção a costa (Melo, 2010).

A duração do vento é um fator crucial neste fenómeno. A amplitude de maré meteorológica é proporcional ao tempo de duração do vento numa determinada direção. Ventos calmos que atuem por períodos superiores a ventos fortes podem ter mais impacto no nível da água do que ventos fortes de curta duração (Hidrográfico, 2021).

Na interação oceano-atmosfera, o vento transfere energia para a água gerando correntes superficiais com a mesma direção. No entanto, estas correntes superficiais sofrem uma deflexão, devido à força de Coriolis, na sua direção, criando o que se domina a Espiral de Ekman. Este desvio acontece na direção anticiclônica, ou seja, no sentido este no Hemisfério Norte e no sentido oeste no Hemisfério Sul (Melo, 2010).

O resultado dos vetores da Espiral de Ekman dá-se a  $90^\circ$  da direção do vento, ao volume de água transportado nesta direção denomina-se Transporte de Ekman (Melo, 2010).

Na Figura 2.8 tem-se o caso do que poderá acontecer na costa portuguesa. Representa uma alta pressão na massa continental e uma baixa pressão na massa oceânica, as setas pretas representam a direção do vento e as setas vermelhas apresentam a direção do fluxo da água.

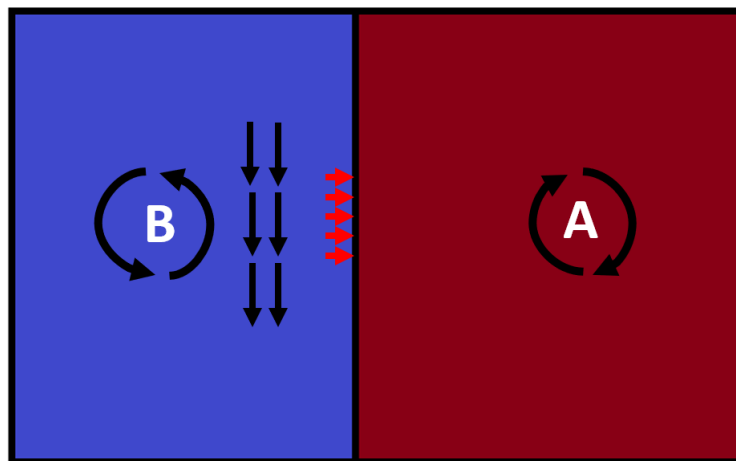


FIGURA 2.8: Maré meteorológica positiva. Nesta figura o A e o B representam uma alta e baixa pressão, respetivamente.

O efeito da pressão atmosférica no nível do mar é explicado, essencialmente, devido ao facto de a atmosfera exercer uma força (Força Gravítica) sobre a superfície da Terra, nomeadamente, sobre os oceanos (Hidrográfico, 2021). Estima-se, que uma variação de 10 milibares desta força exercida pela atmosfera sobre os oceanos produza uma variação de 0.09 m na superfície do mar (Almeida, 2008).

A Figura 2.9 ilustra a sobrelevação do mar devido a fatores meteorológicos (Almeida, 2008).

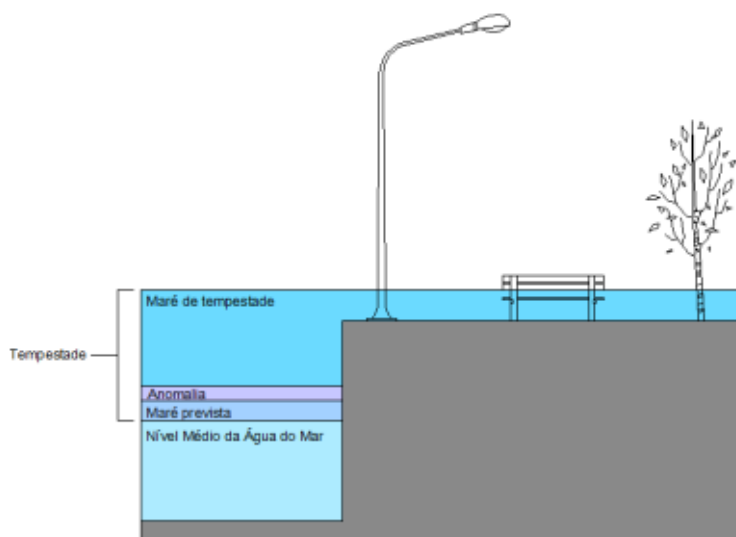


FIGURA 2.9: Efeito de uma tempestade no nível do mar (Extraído de Guerreiro, 2015).

A maré de tempestade pode ser definida como a soma da maré meteorológica com uma onda de tempestade. Este fenómeno pode ser associado à combinação dos efeitos de uma baixa pressão com o vento; esta combinação contribui para a ocorrência de marés extremas (Hidrográfico, 2021).

## Capítulo 3

# Análise de Sobrelevações na Costa Portuguesa

Com este trabalho pretende-se estudar as sobrelevações de maré na costa Portuguesa, porém, é impraticável realizar o estudo para todos os portos na costa Portuguesa. De modo a se mitigar este problema, optou-se pela seleção de marégrafos alocados em portos de grande relevância estratégica e económica e, ao mesmo tempo, geograficamente distribuídos: Porto de Leixões (Região Norte) e Porto de Sines (Região sul).

Atualmente, o Porto de Leixões é beneficiado por uma localização estratégica, permitindo uma elevada densidade industrial. Já o Porto de Sines é o grande porta de entrada na Fachada Euro-Atlântica, dotado de condições naturais perfeitas para acolher qualquer tipo de navio.

A caracterização das sobrelevações na costa portuguesa é fundamental para a tomada de medidas, por parte das entidades competentes, para que se minimize os impactos negativos desta problemática num futuro de médio a longo prazo.

Os dados, relativos às alturas horárias, utilizados neste trabalho, foram disponibilizados pela Divisão de Oceanografia do Instituto Hidrográfico.

Parte dos resultados que a seguir se apresentam, já fizeram parte dos estudos de Rodrigues et al. (2022). O estudo foi realizado tendo por base a aplicação das metodologias apresentadas no capítulo anterior através de bibliotecas do software estatístico R (R Core Team, 2021). No presente trabalho, recorreu-se às seguintes bibliotecas: “Forecast” (Hyndman et al., 2022), “DescTools” (Andri et mult. al., 2021), “Xts” (Ryan e Ulrich, 2020), “Lubridate” (Grolemund e Wickham, 2011), “TSA” (Chan e Ripley, 2020), “Ggplot2” (Wickham, 2016), “Metrics” (Hamner e Frasco,

2018), "readxl"(Wickham e Bryan, 2022), "zoo"(Zeileis e Grothendieck, 2005), "extRemes"(Gilleland e Katz, 2016), "dplyr"(Wickham et al., 2022), "POT"(Ribatet e Dutang, 2022) e "nortest"(Gross et al., 2015).

### 3.1 Porto de Leixões

O Porto de Leixões (Figura 3.1) é um dos portos mais importantes de Portugal. Construído nos finais do século XIX, na cidade do Porto, no concelho de Matosinhos, é o segundo maior porto artificial de Portugal. A sua localização estratégica que permite que seja beneficiado por uma elevada densidade industrial, sendo a maior infraestrutura portuária do Norte de Portugal.

O Porto de Leixões movimentava, anualmente, cerca de 20 milhões de toneladas de mercadorias, representando 20% do comércio externo português por via marítima. Este ainda recebe navios de cruzeiro (APP, 2022).



FIGURA 3.1: Fotografia aérea do Porto de Leixões (Fonte: <http://www.portosdeportugal.pt/app/portos/leixoes.php>), 23/4/2022.

Devido a esta grande movimentação e da sua importância no comércio português é necessário monitorizar as estruturas de defesa do porto, bem como estudar fenómenos naturais que o possam vir danificar, trazendo graves consequências para Portugal.

O Porto de Leixões possui dois molhes para defesa do Porto, o Molho Norte e o Molho Sul. O Molho Norte, molhe principal e de maior dimensão devido a ondulação na costa portuguesa ser maioritariamente proveniente de norte, tem uma

altura máxima de 15 metros, enquanto o Molhe Sul tem apenas a altura máxima de 11,2 metros.

No interior do Porto de Leixões existem diversos cais todos a uma altura de 6 metros, como é perceptível na Figura 3.2. Porém, todas as alturas são em relação ao zero hidrográfico de Leixões (ZHL), que está 0,33m acima do zero hidrográfico nacional (ZH).

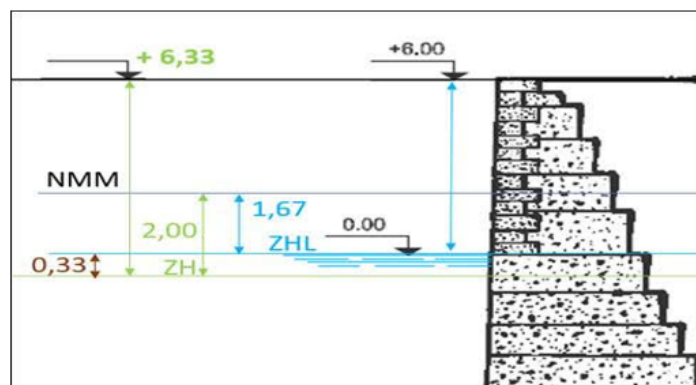


FIGURA 3.2: Esquema de altura de um cais em relação ao Zero Hidrográfico de Leixões (Fonte: ALPD).

As alturas horárias, de maré, registadas pelo marégrafo de Leixões são referidas ao ZH, e não ao ZHL.

#### 3.1.1 Análise Preliminar das Sobrelevações no Porto de Leixões

Para o caso de estudo de sobrelevações meteorológicas no porto de Leixões, usou-se uma série temporal dos desvios (maré observada - maré prevista), desde 1995 até 2020. Iniciou-se a análise com a criação de um gráfico de linhas, onde foi possível verificar que existem lacunas na série, representado na Figura 3.3. Refira-se que existe uma ausência de registos de cerca de 6 anos, de 1996 a 2002.

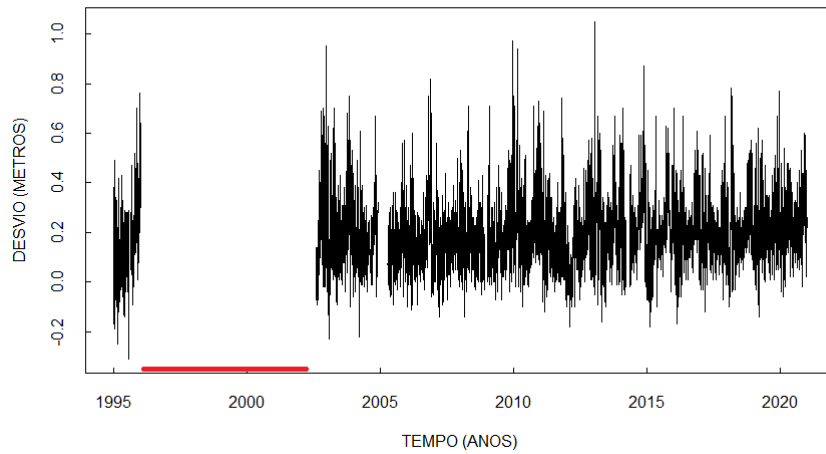


FIGURA 3.3: Séries temporais dos desvios em Leixões (1995-2020).

Para uma melhor perceção da série, efetuou-se a representação gráfica dos desvios entre a maré observada e a maré meteorológica, separados por anos (Figura 3.4). Mais ainda, pode-se verificar um aumento dos desvios nos meses de inverno e outono, apresentando um aumento da amplitude e nos meses de verão e primavera os desvios apresentam uma diminuição da amplitude.

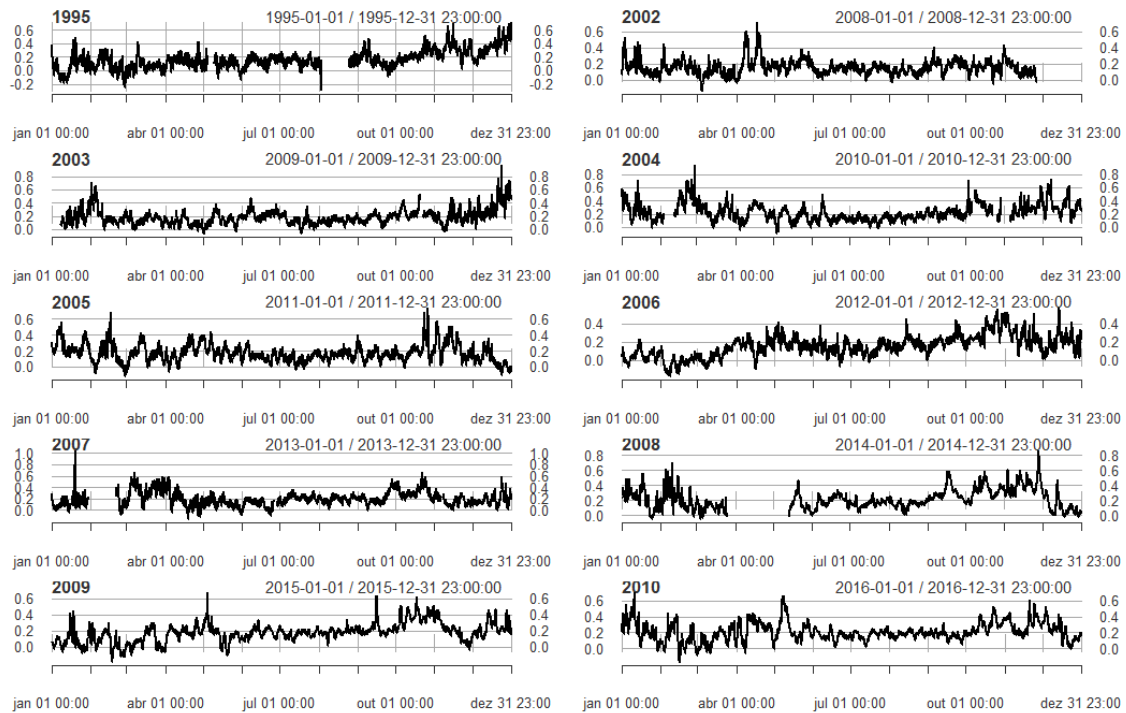


FIGURA 3.4: Série temporal ano a ano dos desvios: 1995 a 2010

A Figura 3.5 apresenta diagramas de extremos e quartis dos desvios por anos. O diagrama de extremos e quartis, permite representar os extremos, os outliers, e os quartis de uma determinada amostra (Sousa, 2019). Através da Figura 3.5, observa-se de imediato que existe uma assimetria positiva dos dados, ou seja, há uma maior concentração dos dados nas sobrelevações do que nas subelevações. Neste trabalho, somente foram estudadas as sobrelevações.

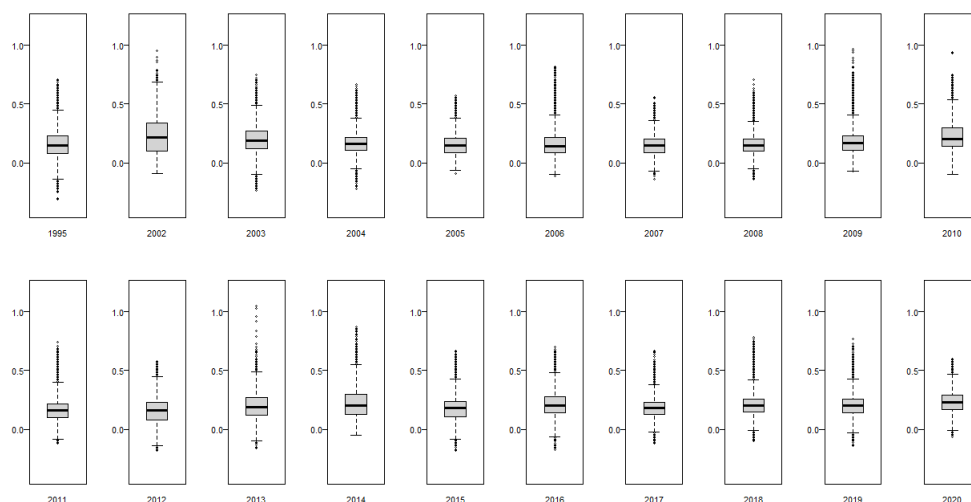


FIGURA 3.5: Diagramas de extremos e quartis dos desvios da maré por ano, entre 1995 e 2020.

Após a análise gráfica dos diagramas de extremos e quartis anuais, foram calculadas as principais estatísticas descritivas: o mínimo e máximo, a média, a mediana e 1º e 3º quartis dos desvios, expressos na Tabela 3.1. Verifica-se que a sobrelevação máxima foi de 1.05 metros, e a subelevação máxima ocorrida é, aproximadamente, 1/5 da sobrelevação máxima, 0.23 metros. A mediana e a média não apresentam uma variação muito significativa, entre 0.14 e 0.23 metros e entre 0.15 e 0.24 metros, respetivamente.

TABELA 3.1: Estatísticas descritivas anuais para a série de Leixões: 1995 a 2013.

	1995	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
<b>Mín</b>	-0.31	-0.09	-0.23	-0.22	-0.09	-0.11	-0.14	-0.14	-0.07	-0.10	-0.12	-0.18	-0.16
<b>Max</b>	0.71	0.95	0.75	0.67	0.57	0.82	0.56	0.71	0.97	0.94	0.74	0.58	1.05
<b>Média</b>	0.16	0.23	0.20	0.17	0.16	0.17	0.15	0.16	0.19	0.23	0.17	0.16	0.20
<b>Mediana</b>	0.15	0.22	0.19	0.16	0.15	0.14	0.15	0.15	0.17	0.20	0.16	0.16	0.19
<b>1º Quartil</b>	0.08	0.10	0.12	0.11	0.09	0.09	0.09	0.10	0.11	0.14	0.10	0.08	0.12
<b>3º Quartil</b>	0.23	0.34	0.27	0.22	0.21	0.22	0.20	0.20	0.23	0.30	0.22	0.23	0.27

No histograma apresentado na Figura 3.6 é perceptível a forma como estão distribuídos os desvios de maré. Verifica-se uma assimetria para a direita, indiciando um comportamento não normal da série.

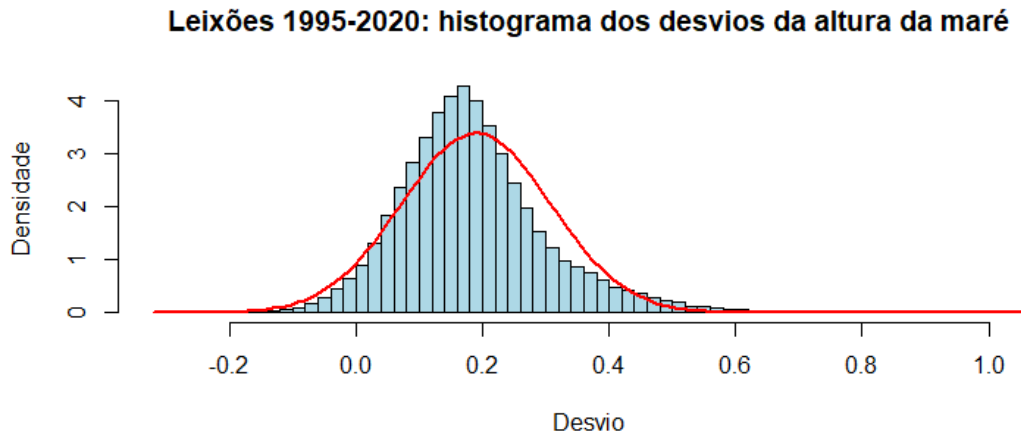


FIGURA 3.6: Histograma dos desvios: Leixões 1995 a 2020.

De modo a realizar-se o estudo de uma série temporal, esta não pode ser descontínua. Como neste caso se verifica, entre os anos de 1996 e 2002, uma ausência de registos decidiu-se realizar o estudo da série desde 2002 a 2020, ficando a série temporal apenas com 19 anos, como está representado na Figura 3.7.

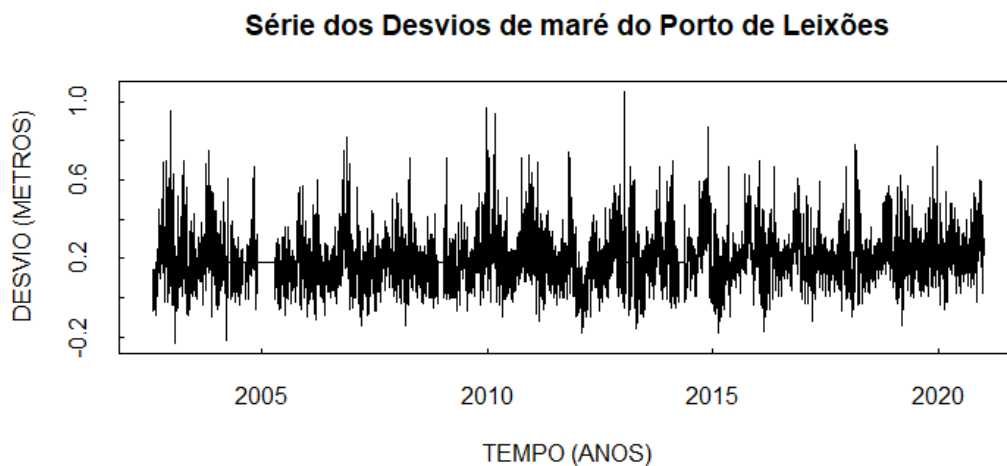


FIGURA 3.7: Série temporal, sem falhas, dos desvios de Leixões: 2002-2020.

Ao longo da série existem diversos valores omissos, pelo qual, se procedeu à imputação dos valores omissos pela mediana, 0.18m. Num estudo sobre o nível

médio mar esta imputação seria incorreta, pois afetaria a tendência central da série; Porém para o estudo de valores extremos, esta imputação não afeta o estudo pelo facto de valores extremos serem por natureza muito afastados da mediana.

Neste trabalho separaram-se os dados de duas formas distintas: Máximos Anuais e Máximos Mensais, representadas nas Figuras 3.8 e 3.23, respetivamente. Neste processo dividiu-se a série em blocos de dimensão  $k$  (mensais ou anuais) e para cada um destes blocos extraiu-se unicamente a observação de maior valor.

### 3.1.2 Máximos Anuais

Nesta abordagem recorreu-se ao Modelo dos Máximos Anuais, para se verificar qual o modelo que melhor se ajusta à amostra (GEV, Gumbel, Fréchet, Max-Weibull ou GP). Este modelo consiste em dividir os dados em blocos anuais e verificar qual dos modelos se ajusta melhor à série dos máximos anuais.

Neste caso, dispõem-se de  $k=19$  desvios máximos anuais, tal como ilustra a Figura 3.8.

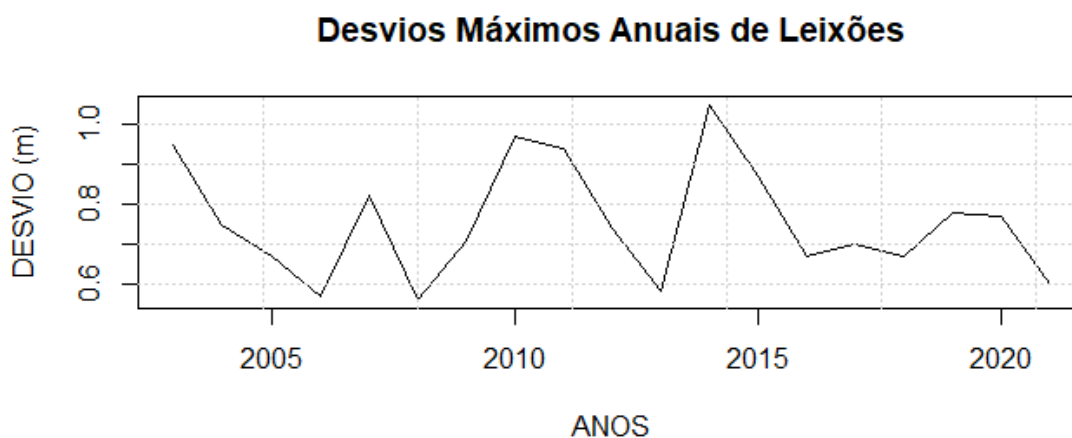


FIGURA 3.8: Desvios máximos anuais: Leixões.

Pretende-se estudar os quantis elevados, outliers e níveis de retorno,  $u$ . Deste modo recorre-se à estatística de extremos para se realizar esta análise, porém é necessário garantir que os registos são independentes, identicamente distribuídas (i.i.d.) e que se trata de uma distribuição não normal.

Sejam  $H_0$ : os registos que seguem a distribuição normal e  $H_1$  os registos que rejeitam a normalidade. Quando o valor de  $p < 0.05$  rejeita-se  $H_0$ , ou seja, não existe a evidência estatística para afirmar que os dados seguem distribuição normal.

No entanto, se  $p > 0.05$  existe a evidência estatística para que não se rejeite  $H_0$ , ou seja, os registos seguem a distribuição normal.

Apesar do histograma, Figura 3.9, sugerir uma distribuição não normal, a normalidade foi validada através do teste de *Shapiro-Wilk* ( $W = 0.94437$ ;  $p\text{-value} = 0.3155$ , esse resultado pode deferir-se ao facto da amostra ter uma dimensão reduzida (19 observações), devido a isto a conclusão da normalidade da série deve ser interpretada com cuidado.

Da análise da Figura 3.9 e da Tabela 3.2 verifica-se um enviesamento à esquerda das observações, induzindo que os modelos de Gumbel e GEV terão um bom ajustamento aos desvios máximos anuais, assim apropriados para descrever os máximos anuais relativos aos desvios da altura de maré no Porto de Leixões.

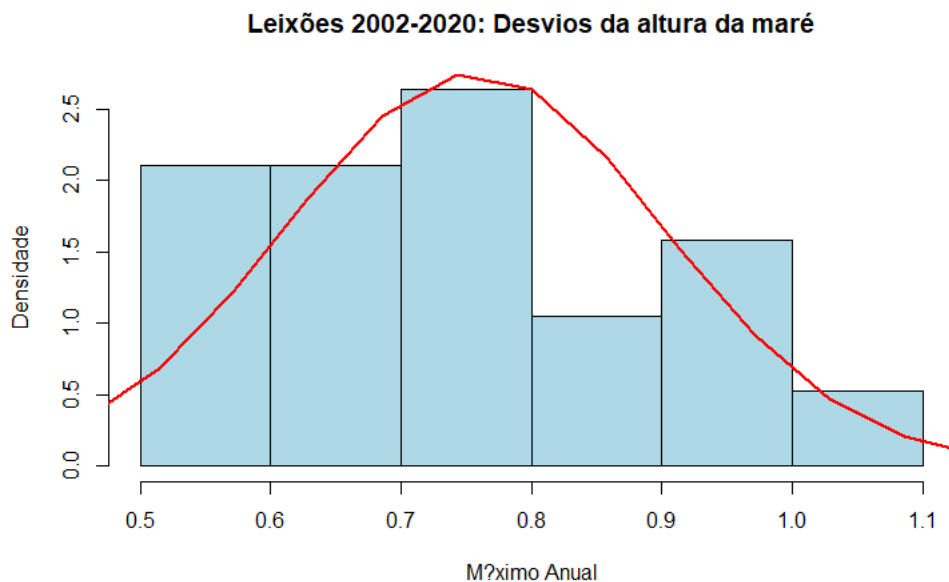


FIGURA 3.9: Histograma dos Desvios máximos Anuais: Leixões.

TABELA 3.2: Medidas descritivas dos máximos anuais: Leixões.

Nº de Obs.	Média	Mediana	Desvio Padrão	Máximo	Mínimo	1º Quartil	3º Quartil
19	0.7563	0.7400	0.144845	1.0500	0.5600	0.6700	0.8450

Para se poder analisar a série através da estatística de valores extremos, é necessário que os desvios máximos anuais sejam independentes. Através de gráficos de auto-correlação (ACF) e auto-correlação parcial (PACF) é possível testar a existência ou ausência de correlação nos desvios, e por sua vez a sua independência. A correlação determina a dependência entre duas variáveis. Para os desvios serem não

correlacionados, os valores dos gráficos de ACF e PACF não podem ultrapassar o intervalo de confiança, representado a azul na Figura 3.10.

Através da Figura 3.10, constata-se que não existe correlação, ou seja os desvios são independentes.

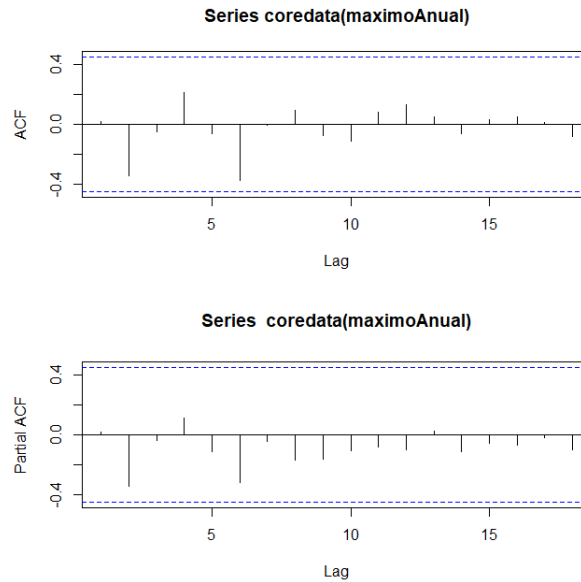


FIGURA 3.10: Gráficos de auto-correlação e auto-correlação parcial para os máximos anuais: Leixões.

De seguida construi-se um diagrama de extremos e quartis, representado na Figura 3.11, para se estudar a existência de possíveis outliers. Neste caso, verificou-se a ausência de outliers.

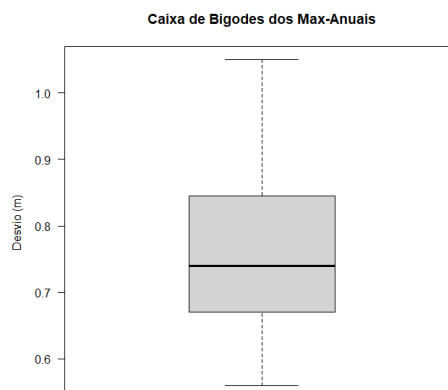


FIGURA 3.11: Diagrama de extremos de quartis correspondente aos máximos anuais: Leixões.

### Análise GEV

Uma vez provada a independência dos dados, considerou-se o modelo GEV, descrito na secção 2.3. Através do software R, criou-se um gráfico de probabilidades GEV para a amostra em estudo, conforme ilustrado na Figura 3.12. Para tal, estimaram-se os parâmetros localização ( $\lambda$ ), escala ( $\delta$ ) e forma ( $\rho$ ) usando, tal como apresentado no capítulo 2.1.3, o estimador de máxima verosimilhança. Os valores estimados dos parâmetros foram  $\hat{\lambda} = 0.693$ ,  $\hat{\delta} = 0.118$  e  $\hat{\rho} = -0.057$ .

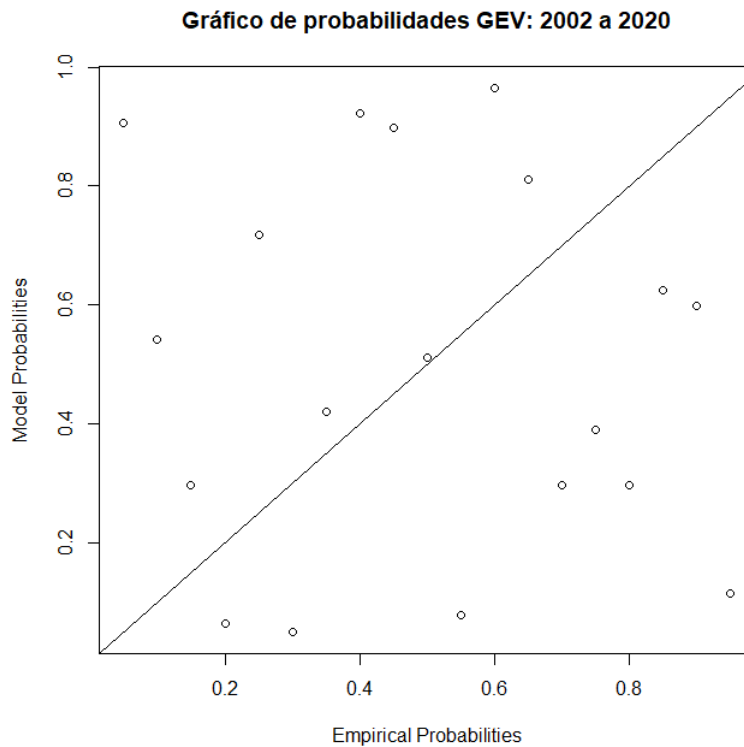


FIGURA 3.12: Gráfico de probabilidades GEV: Leixões.

A Figura 3.13 mostra que existe uma dispersão considerável dos valores observados relativamente à distribuição teórica. Para melhor compreensão do ajustamento da amostra à GEV, construiu-se um gráfico de comparação entre a amostra e o modelo teórico da GEV, Figura 3.13, indicando que o modelo terá um ajuste satisfatório.

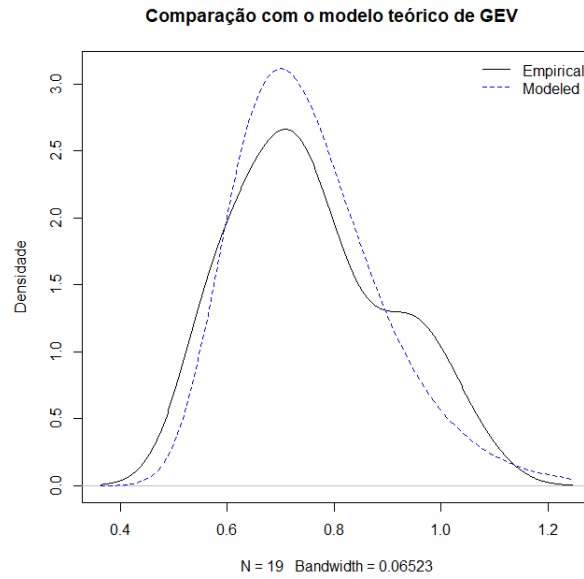


FIGURA 3.13: Gráfico de densidade para o modelo GEV: Leixões.

Como apresentado na secção 2.3, a Generalizada de Valores Extremos é composta por 3 parâmetros: localização, escala e forma. Porém o parâmetro forma não é significativo para esta distribuição, usando apenas os parâmetros localização e escala.

Com vista a melhorar o ajustamento à GEV estudou-se a amostra tendo a localização e a escala variável segundo a expressão:

$$\left(0.3x + \cos(2\pi x) + \sin(2\pi x)\right) \quad (3.1)$$

Com:

$0.3x$  representa a tendência da subida anual do nível médio do mar;

$\cos(2\pi x) + \sin(2\pi x)$  representa uma variação sazonal ao longo de um ano.

Deste modo o parâmetro varia segundo uma constante, segundo a tendência do aumento do nível do mar e segundo uma variação sazonal, obtendo-se assim um melhor ajuste da GEV à amostra.

Construiu-se um gráfico de probabilidades para a escala e localização variáveis, Figura 3.14, tornando visualmente a relação linear satisfatória, o que induz a uma validação informal do modelo GEV para a modelação dos máximos anuais dos desvios da altura de maré do porto de Leixões.

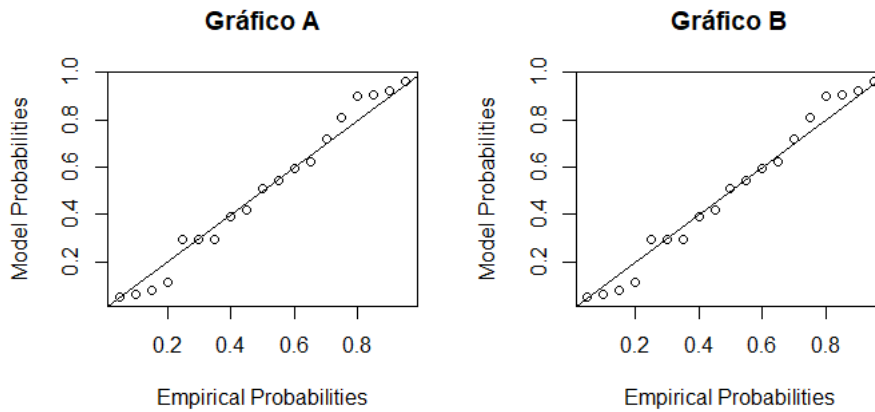


FIGURA 3.14: Gráficos de probabilidade GEV, Gráfico A e B representam o ajustamento com os parâmetros localização e escala, respetivamente, a variar segundo a expressão 3.1.

Face a este resultado, recorreu-se ao teste de Likelihood-ratio para se verificar se os parâmetros com variação produzem melhor resultado ao não. Este teste permite validar a hipótese  $H_0$ : as duas distribuições são semelhantes, para um  $p\text{-value} > 0,05$  diz-se que existe a evidência que as distribuições não são distintas e para  $p\text{-value} < 0,05$  diz-se que existe a evidência estatística que as distribuições não são semelhantes.

Comparando o ajustamento da distribuição GEV com parâmetros (Localização e Escala) constantes e variáveis, obteve-se um  $p\text{-value} > 0.9$ , pelo que existe a evidência estatística que o ajustamento à GEV como parâmetros variáveis não é distinta do ajustamento com parâmetros constantes.

De modo a se estudar qual o melhor modelo, usou-se dois indicadores estatísticos da parcimónia de modelos ou distribuições, o critério de Informação de Akaike (AIC) e o Critério de informação Bayesiano (BIC), pelo qual a distribuição que tiver os indicadores com menor valor é a que apresenta um melhor compromisso entre o ajustamento e a complexidade do modelo. Para um estudo mais aprofundado destes indicadores estatísticos consultar Emiliano et al. (2010).

Calculou-se os indicadores AIC e BIC para a GEV com parâmetros constantes, a GEV com o parâmetro localização e a GEV com o parâmetro escala variável. Obteve-se os mesmo valores para os 3 ajustamentos,  $AIC = -16.06$  e  $BIC = -13.23$ , não existindo evidências de que um modelo seja superior a outro.

De forma a testar qual dos modelos é melhor, procedeu-se ao indicador *Root Mean Square Error* (RMSE), este teste mede a diferença entre o modelo teórico e o

modelo empírico da amostra, para uma melhor compreensão deste método pode-se consultar Chai e Draxler (2014).

Os resultados obtidos foram de 0.0447 para o modelo com os parâmetros escala e localização variável e de 0.4288 para o modelo com as variáveis constantes, que os desvios máximos anuais são bem explicado através da distribuição GEV com parâmetros de localização e escala variáveis.

Os gráficos apresentados também são indicadores de que os desvios máximos anuais se podem explicar através de uma das distribuições Gumbel, Fréchet ou Max-Weibull. De forma a proceder-se a qual se ajusta melhor, calcularam-se os intervalos de confiança para os parâmetros da GEV, expressos na Tabela 3.3.

TABELA 3.3: Parâmetros localização, escala e forma do modelo GEV: Leixões.

	<b>95% IC INFERIOR</b>	<b>ESTIMADO</b>	<b>95% IC SUPERIOR</b>
<b>LOCALIZAÇÃO</b>	0.629	0.693	0.757
<b>ESCALA</b>	0.070	0.118	0.167
<b>FORMA</b>	-0.551	-0.057	0.437

Os gráficos apresentados também são indicadores de que os desvios máximos anuais se podem explicar através de uma das distribuições Gumbel, Fréchet ou Max-Weibull; e uma vez que este parâmetro tem de ser maior que 0 para as famílias de Fréchet e Max-Weibull, estas distribuições não se ajustam bem aos dados em análise. A distribuição de Gumbel é assim uma boa candidata.

Os gráficos de probabilidade e de densidade de probabilidade para a distribuição de Gumbel, representados na Figura 3.15, permitem encontrar os seguintes parâmetros:  $\lambda = 0.689$  e  $\delta = 0.116$ .

Como já foi referido anteriormente, os parâmetros escala e localização podem variar ao longo do tempo. Assim, ajustou-se aos dados a distribuição de Gumbel com parâmetros a variar de acordo com a mesma expressão anterior (3.1). Na Figura 3.16 estão representados os gráficos de probabilidade e de densidade de probabilidade, verificou-se que a distribuição de Gumbel com parâmetros variáveis se ajusta melhor do que a de parâmetros constantes.

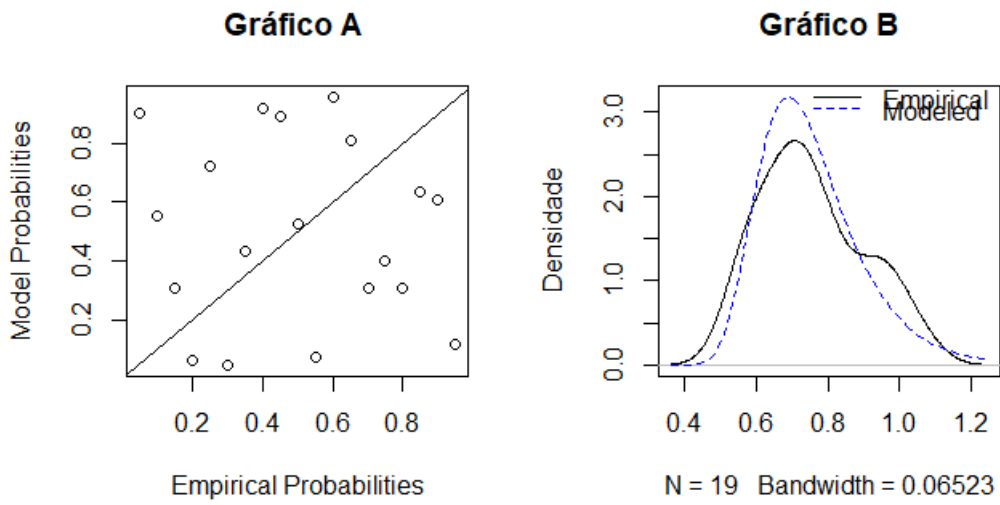


FIGURA 3.15: Gráficos de probabilidade (Gráfico A) e densidade de probabilidade (Gráfico B) para a distribuição de Gumbel: Leixões.

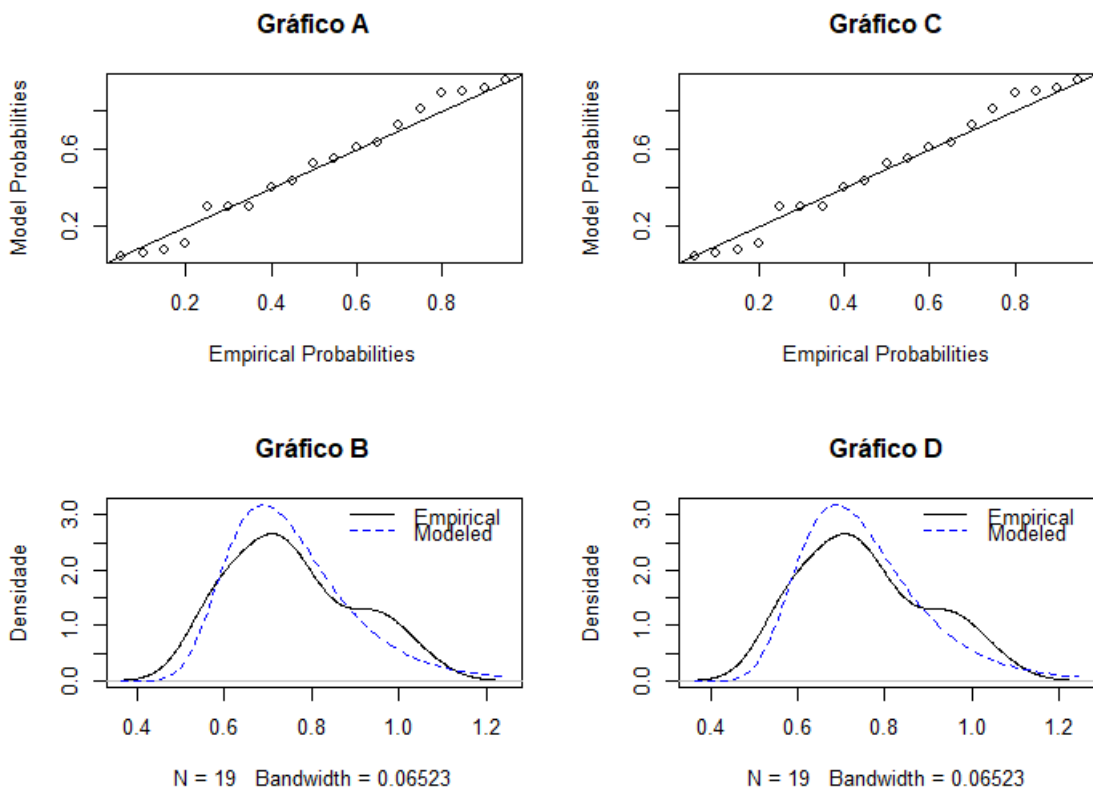


FIGURA 3.16: Gráficos de probabilidade (Gráficos A e C) e densidade (Gráficos B e D) modelo Gumbel com parâmetros variáveis segundo a expressão 3.1.

Recorreu-se ao teste de Likelihood-ratio para avaliar se, efetivamente, a distribuição de Gumbel com parâmetros variáveis é estatisticamente diferente da distribuição com parâmetros constantes. Sob  $H_0$ : as distribuições são semelhantes, obteve-se um  $p - value = 0.8242$ , pelo que existe a evidência estatística que os ajustamentos com a Gumbel com parâmetros variáveis não é distinta do ajustamento com parâmetros constantes.

De modo a se estudar qual o melhor modelo, usou-se os indicadores estatísticos AIC e BIC, para Gumbel com parâmetros constantes obteve-se os valores  $AIC = -18.01$  e  $BIC = -16.13$ , já para Gumbel com os parâmetros variáveis obteve-se os valores,  $AIC = -16.06$  e  $BIC = -13.23$ . Estes valores obtidos indiciam que o melhor modelo é o da Gumbel com parâmetros constantes.

De forma a confirmar este resultado, recorreu-se ao calculo do RMSE para os ajustamentos dos 2 modelos, os resultados obtidos foram de 0.0438 para o modelo com os parâmetros escala e localização variável e de 0.4267 para o modelo com as variáveis constantes. O que permite concluir que a distribuição de Gumbel com parâmetros variáveis se comporta melhor para a descrição dos desvios máximos anuais. Porém não existe evidências estatísticas que um dos ajustamentos, com os parâmetros localização e escala variáveis, seja melhor que o outro.

Os dados analisados permitiram concluir que a distribuição de Gumbel se ajusta melhor do que a distribuição GEV, esse resultado é apresentado na Tabela 3.4.

TABELA 3.4: Indicadores estatísticos AIC e BIC e teste RMSE para as distribuições GEV e Gumbel com o parâmetro localização variável.

	<b>AIC</b>	<b>BIC</b>	<b>RMSE</b>
<b>GEV (Localização variável)</b>	- 16.0633	-13.2300	0.0447
<b>Gumbel (Localização variável)</b>	-18.0139	-16.1251	0.0438

Escolhido o modelo de Gumbel com localização variável, construiu-se um gráfico para encontrar o nível de retorno, representado na Figura 3.17. Verifica-se que o nível de retorno tem uma tendência crescente ao longo do tempo, este poderá ser consequência das alterações climáticas e por sua vez do aumento do nível médio do mar, 0.003 metros por ano (NOAA, 2017).

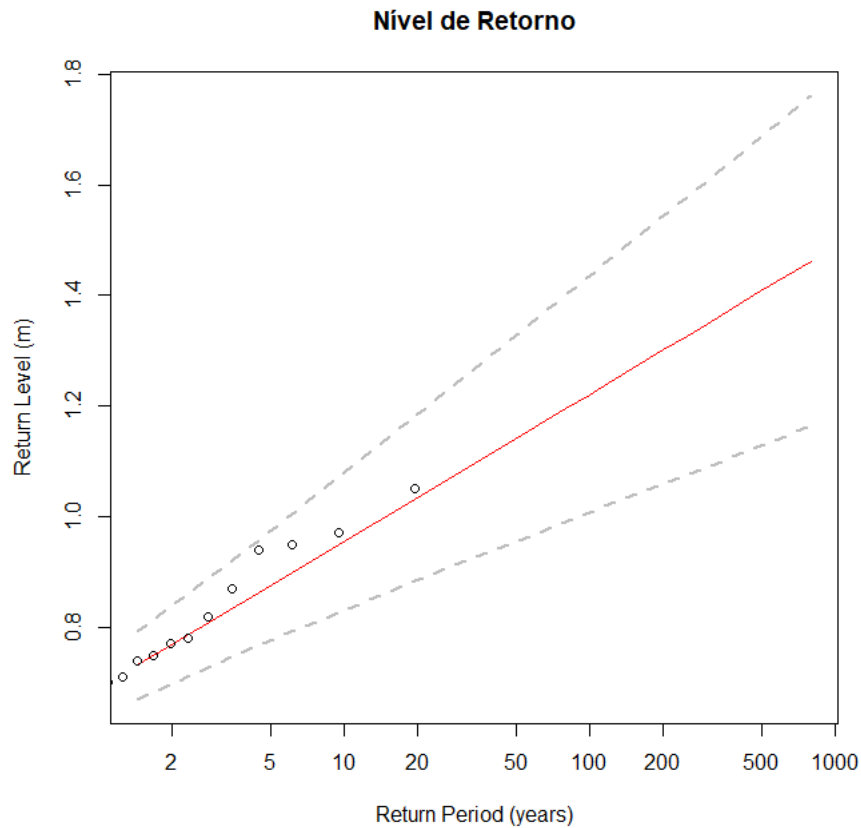


FIGURA 3.17: Gráfico de nível de retorno, para os máximos anuais, através do modelo Gumbel: Leixões.

De modo a ter uma análise mais precisa dos níveis de retorno, construiu-se a Tabela 3.5, onde estão calculados os níveis de retorno para 5, 20, 50 e 80 anos. Verificou que o maior nível de retorno previsto será de 1,2 metros para um período de retorno de 80 anos.

TABELA 3.5: Valores do nível de retorno estimado através do modelo de Gumbel, para os máximos anuais de Leixões.

TEMPO (ANOS)	5	20	50	80
<b>GUMBEL (localização variável)</b>	0.86m	1.03m	1.14m	1.2m

### Análise GP

Nesta fase da dissertação, pretende-se estudar os desvios máximos anuais de Leixões através da análise de POT, apresentada na secção 2.3, procedendo ao ajustamento de um modelo de GP para se modelar os excessos. Como descrito na

secção 2.3, a análise POT inicia-se por se delimitar um nível elevado,  $\mu$ , denominado de *threshold*.

Na secção 2.3 é apresentado um processo de escolha do *threshold* baseado na estabilidade dos estimadores dos parâmetros da distribuição GP. Na Figura 3.18 está representado os gráficos com as estimativas, pontuais e intervalares, dos parâmetros forma e escala segundo um *threshold*. A partir do 0.65m existe uma estabilidade para as estimativas, pelo qual este é a escolha adequada para *threshold*.

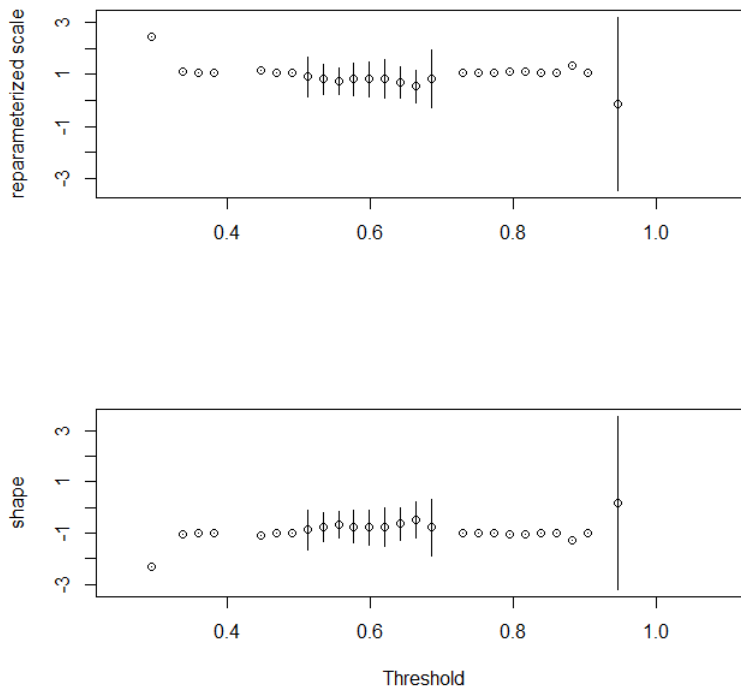


FIGURA 3.18: Estimativas dos parâmetros forma e escala versus threshold para os máximos anuais de Leixões.

De forma a validar  $\mu = 0.65m$ , procedeu-se a construção de um *Mean Residual Life Plot*, representado na Figura 3.19. Este gráfico apresenta o calculo dos excessos médios para cada intervalo de *threshold*. Os intervalos de confiança, apresentados a vermelho, são calculados com base na expressão de densidade normal para os excessos médios, tendo como objetivo encontrar o melhor *threshold* para a amostra, este encontra-se imediatamente antes do gráfico se tornar instável. Neste caso é possível validar 0.65 como o melhor candidato a *threshold*.

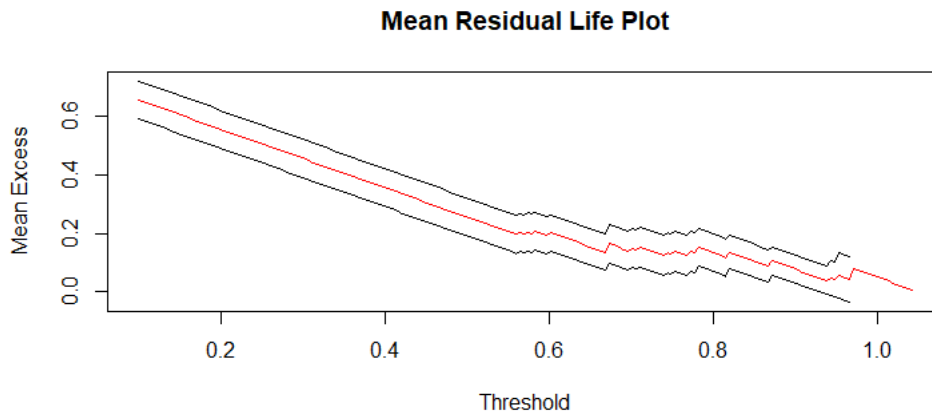


FIGURA 3.19: Mean Residual Life Plot para os máximos anuais de Leixões.

Considerando  $\mu = 0.65$  para a amostra dos desvios máximos anuais de Leixões, obteve-se 15 excedências em 19 valores, o que equivale, aproximadamente, a 79% da amostra. Na Figura 3.20 está apresentado um gráfico de probabilidades, onde se verifica dispersão considerável dos valores observados relativamente à distribuição teórica. Os valores estimados dos parâmetros foram  $\delta = 0.258$  e  $\rho = -0.593$ .

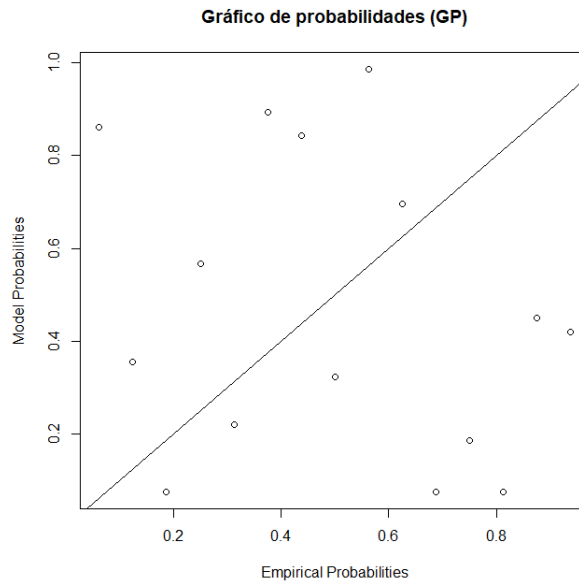


FIGURA 3.20: Gráfico de probabilidades dos excessos, referente à amostra nos máximos anuais de Leixões.

De forma a se melhorar o modelo GP às excedências, tal como na análise GEV para máximos anuais de Leixões, procedeu-se à variação do parâmetro escala

segundo a expressão 3.1. Obteve-se um gráfico de probabilidades, representado na Figura 3.21, com uma dispersão muito menor dos valores observados relativamente à distribuição teórica.

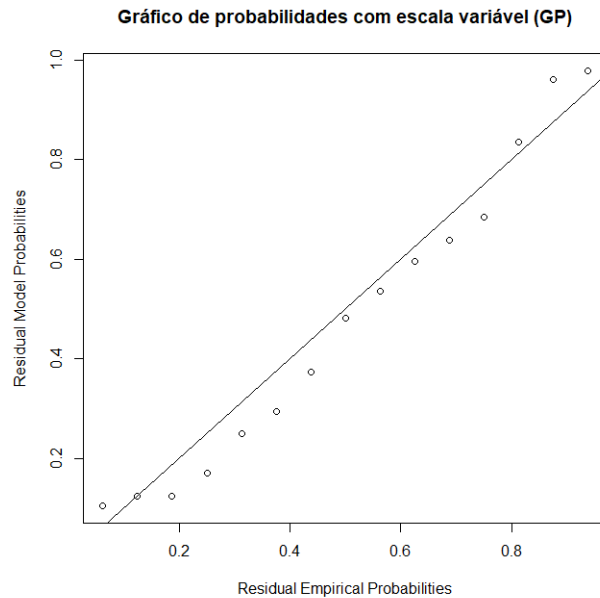


FIGURA 3.21: Gráfico de probabilidades dos excessos, com o parâmetro escala variável, referente à amostra nos máximos anuais de Leixões.

Por forma a estudar qual dos dois ajustamentos é melhor, foram efetuados os testes de Likelihood-ratio e calculados os valores AIC, BIC e RMSE. Para os testes de Likelihood-ratio obteve-se um  $p - value = 0.3359$ , pelo que existe a evidência estatística que os modelos são distintos. Através dos indicadores estatísticos expressos na tabela 3.6, é possível verificar que, os valores de BIC e AIC são muito aproximados, no entanto o RMSE indica que o modelo é melhor para a GP com escala variável, sendo que diferença entre o modelo empírico e teórico é menor para esta. Deste modo, tem-se um melhor ajuste para a GP com o parâmetro escala Variável.

TABELA 3.6: Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GP, com o parâmetro escala variável segundo a expressão 3.1.

	<b>AIC</b>	<b>BIC</b>	<b>RMSE</b>
<b>GP</b>	-24.4296	-23.0134	0.4577
<b>GP (Escala variável)</b>	-23.3556	-21.2314	0.0547

Escolhido o modelo de GP com escala variável, construiu-se um gráfico para o nível de retorno, representado na Figura 3.22. Verifica-se que o nível de retorno, tal como com o modelo de Gumbel, tem uma tendência crescente ao longo do tempo.

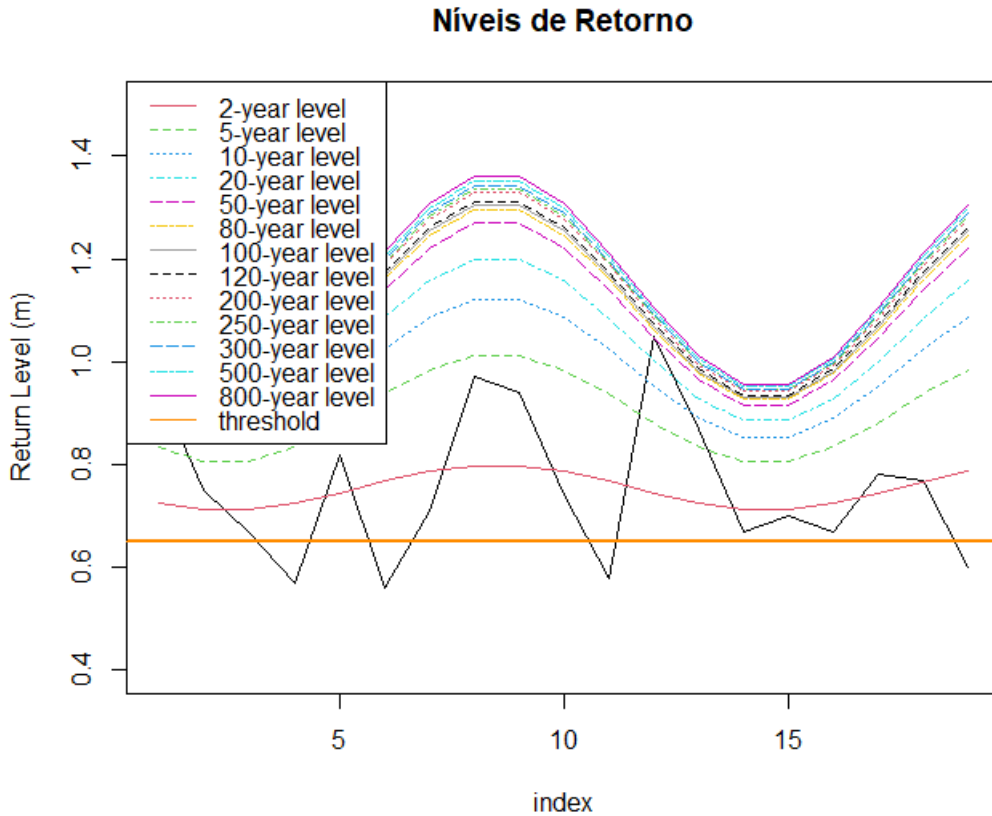


FIGURA 3.22: Gráfico de nível de retorno, para os máximos anuais, através do modelo GP: Leixões.

Para uma análise mais precisa dos níveis de retorno, construiu-se a Tabela 3.7, onde estão calculados os níveis de retorno para 5, 20, 50 e 80 anos. Através deste modelo conclui-se que o maior nível de retorno previsto será de 1,05 metros para um período de retorno de 80 anos.

TABELA 3.7: Valores do nível de retorno estimado através do modelo GP, para os máximos anuais de Leixões

Tempo (Anos)	5	20	50	80
GP	0.89m	1.00m	1.04m	1.05m

### 3.1.3 Máximos Mensais

Neste ponto do trabalho recorreu-se ao Modelo dos Máximos Mensais, de modo a se verificar qual a distribuição (GEV, Gumbel, Fréchet, Max-Weibull ou GP) que melhor se ajusta aos dados em estudo. este modelo consiste em separar os dados em blocos mensais e verificar qual dos modelos se ajusta melhor à série dos máximos mensais.

Neste caso, dividiu-se a amostra em  $k = 221$  máximos mensais, tal como ilustra a Figura 3.23.

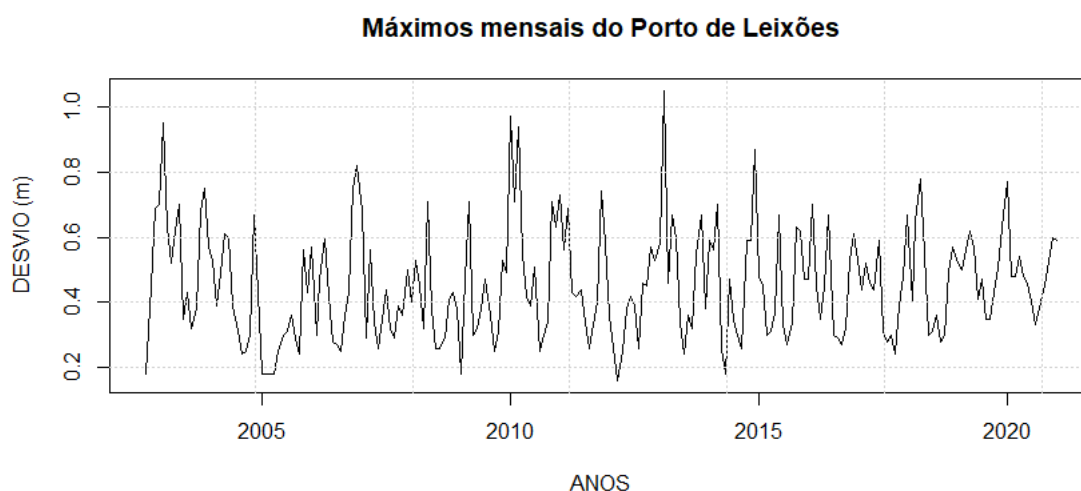


FIGURA 3.23: Máximos mensais de 2002 a 2020 para Leixões.

Pretende-se estudar os quantis elevados, outliers e níveis de retorno,  $u$ . Deste modo, tal como na análise realizada na subsecção 3.1.2, é necessário garantir que os registos são independentes e identicamente distribuídas (i.i.d.).

No histograma, apresentado na Figura 3.24, da série dos desvios mensais para Leixões verifica-se uma cauda à direita. De modo a avaliar se os dados seguem distribuição normal aplicou-se o teste de *Shapiro-Wilk*, onde se obteve  $W = 0.96052$  e  $p - value = 8.464 \times 10^{-06}$ , assim rejeita-se a hipótese de normalidade dos dados, confirmada pela Figura 3.24.

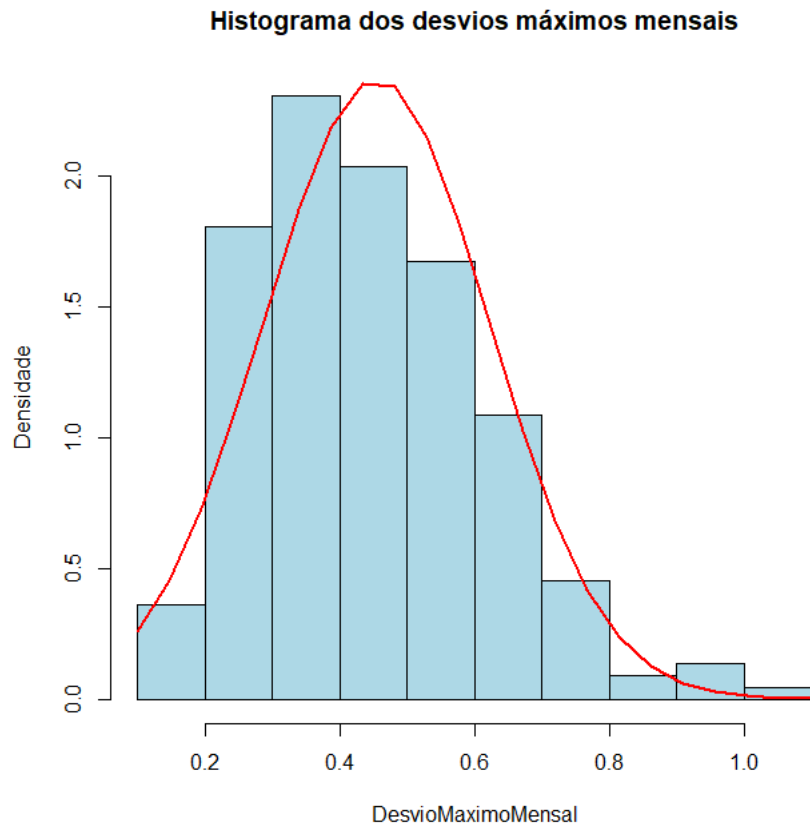


FIGURA 3.24: Histograma dos desvios máximos mensais para Leixões.

TABELA 3.8: Medidas descritivas dos máximos mensais para Leixões.

Nº de Obs.	Média	Mediana	Desvio Padrão	Máximo	Mínimo	1º Quartil	3º Quartil
221	0.4538	0.4300	0.1679689	0.5700	0.1600	0.3200	1.0500

Através da análise da Figura 3.24 e da Tabela 3.8 verifica-se um enviesamento à esquerda dos registos, induzindo que os modelos de GEV e Gumbel serão um bom ajustamento aos dados em análise e assim apropriados a modelar e descrever a série dos máximos mensais dos desvios da altura de maré no Porto de Leixões.

Através dos gráficos ACF e PACF, representado na Figura 3.25, verifica-se que só existe independência para desvios separados por 100 horas (cerca de 4 dias). Uma vez que os registos usados são os desvios máximos mensais a independência fica garantida.

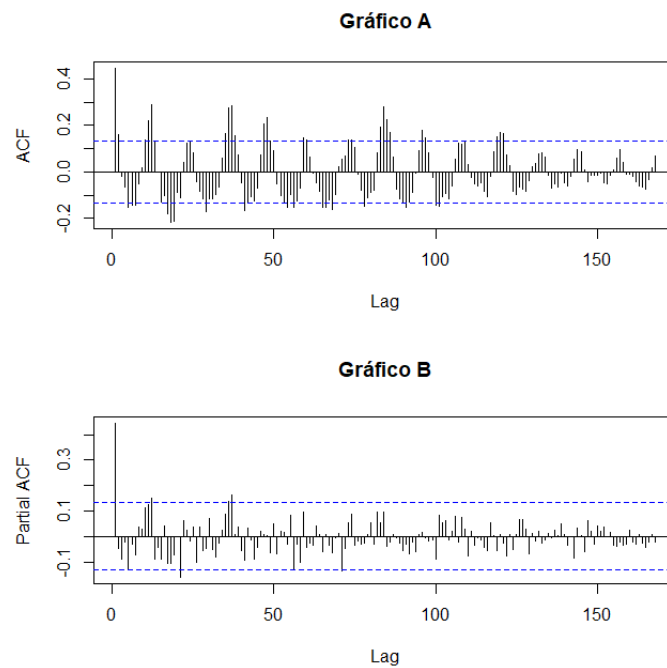


FIGURA 3.25: Gráficos de auto-correlação (Gráfico A) e auto-correlação parcial (Gráfico B) para os máximos mensais de Leixões.

Construí-se um diagrama de extremos e quartis, representado na Figura 3.26, onde se verificou a existência de outliers.

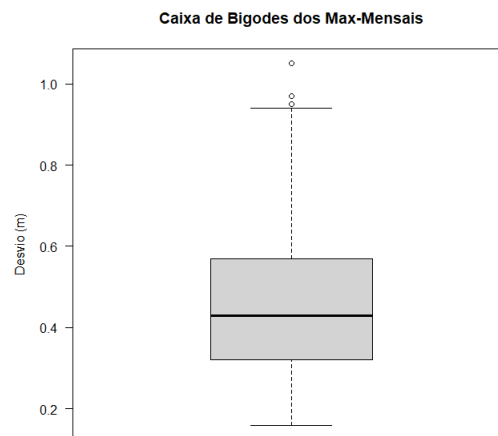


FIGURA 3.26: Diagrama de extremos e quartis correspondente aos máximos mensais de Leixões.

### Análise GEV

Uma vez aceite que os desvios máximos mensais são i.i.d. considerou-se o modelo GEV, apresentado na secção 2.3, assim, de forma a estudar o ajustamento da amostra ao modelo da GEV construíram-se os gráficos de probabilidade e densidade, representados na Figura 3.27.

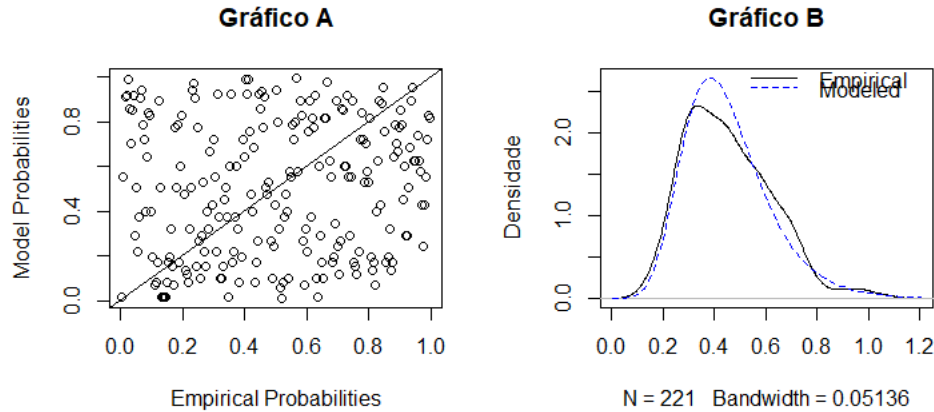


FIGURA 3.27: Gráficos de probabilidade (Gráfico A) e densidade de probabilidade (Gráfico B) da amostra dos máximos mensais de Leixões para o modelo GEV.

Através do gráfico de probabilidades, verifica-se, visualmente, que existe dispersão considerável dos valores observados relativamente à distribuição teórica. No entanto, o gráfico da densidade de probabilidade, apresenta, visualmente, um ajustamento satisfatório da distribuição teórica.

Analogamente, ao aplicado aos desvios máximos anuais, estudou-se a variação dos parâmetros escala e localização, segundo da expressão 3.2.

$$\left( 0.4 \times 10^x + \cos\left(\frac{\pi x}{12}\right) + \sin\left(\frac{\pi x}{12}\right) \right) \quad (3.2)$$

Com:

$0.4 \times 10^x$  representa a tendência da subida do nível médio do mar (NOAA, 2017);  $\cos\left(\frac{\pi x}{12}\right) + \sin\left(\frac{\pi x}{12}\right)$  representa a variação sazonal ao longo de um mês.

Deste modo, construíam-se gráficos de probabilidades com a localização e escala variável, apresentados na Figura 3.28.

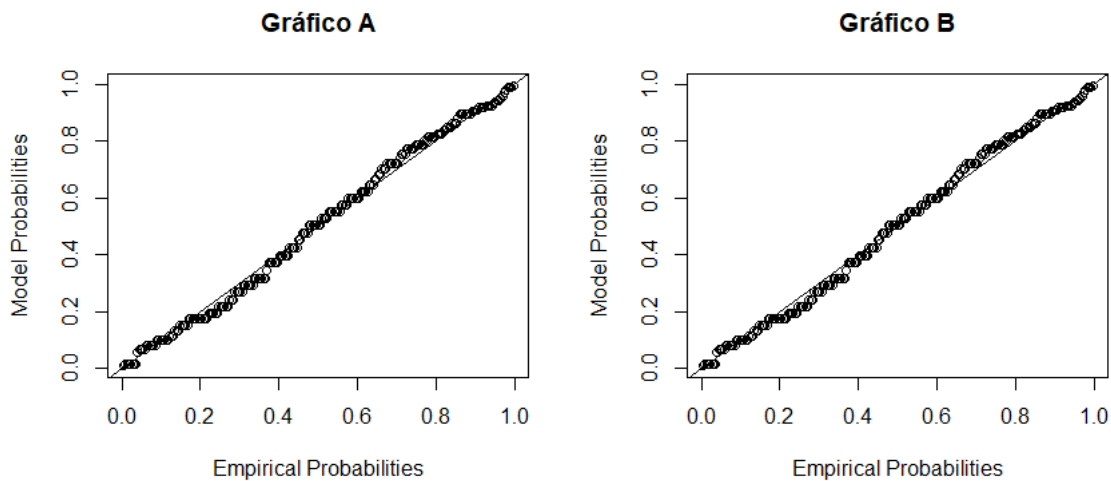


FIGURA 3.28: Gráficos de probabilidades com os parâmetros localização (Gráfico A) e escala (Gráfico B) variável, segundo a expressão 3.2.

Com uma relação linear satisfatória, realizou-se o teste de Likelihood-ratio de forma a verificar-se se os modelos (GEV com parâmetros constantes, GEV com localização variável e GEV com escala variável) são distintos. Neste caso, obteve-se um  $p - value > 0.9$ , para o teste de Likelihood-ratio, para todas as comparações, pelo que existe a evidência estatística que os modelos não são distintos.

Para se perceber qual o melhor ajustamento procedeu-se, tal como na análise GEV dos máximos anuais, ao cálculo dos indicadores estatísticos AIC, BIC e RMSE. Através da Tabela 3.9, permite verificar que a distribuição GEV com parâmetros variáveis é a melhor escolha.

TABELA 3.9: Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GEV com parâmetros constantes e variáveis.

	AIC	BIC	RMSE
<b>GEV</b>	-179.915	-169.721	0.3924
<b>GEV (Localização variável)</b>	-179.915	-169.721	0.0224
<b>GEV (Escala Variável)</b>	-179.915	-169.721	0.0224

Os gráficos anteriores uma das distribuições: Gumbel, Fréchet ou Max-Weibull podem ser bem ajustadas aos desvios máximos mensais. Desta forma, calcularam-se os intervalos de confiança para os parâmetros da GEV, expressos na Tabela 3.10.

Como é possível observar pela Tabela 3.10, o intervalo de confiança do parâmetro forma é de (-0.149; 0.067) existindo evidência de que este parâmetro não é estatisticamente diferente de zero. Assim, as distribuições de Fréchet e Max-Weibull são excluídas, pois para estes modelos o parâmetro forma tem ser positivo. No entanto, para o modelo Gumbel o parâmetro forma não é importante, pelo que esta distribuição é uma boa candidata para explicar o comportamentos dos desvios máximos mensais.

TABELA 3.10: Parâmetros localização, escala e forma do modelo GEV para os máximos mensais de Leixões.

	95% IC INFERIOR	ESTIMADO	95% IC SUPERIOR
<b>Localização</b>	0.3577804	0.37861038	0.39944037
<b>Escala</b>	0.1229408	0.13820049	0.15346017
<b>Forma</b>	-0.1491173	-0.04092003	0.06727722

Na Figura 3.29 estão representados os gráficos de probabilidade (Gráfico A) e densidade (Gráfico B) para a distribuição de Gumbel, é possível verificar, que existe uma má relação linear. Para este ajuste os parâmetros obtidos foram  $\lambda = 0.376$  e  $\delta = 0.136$ .

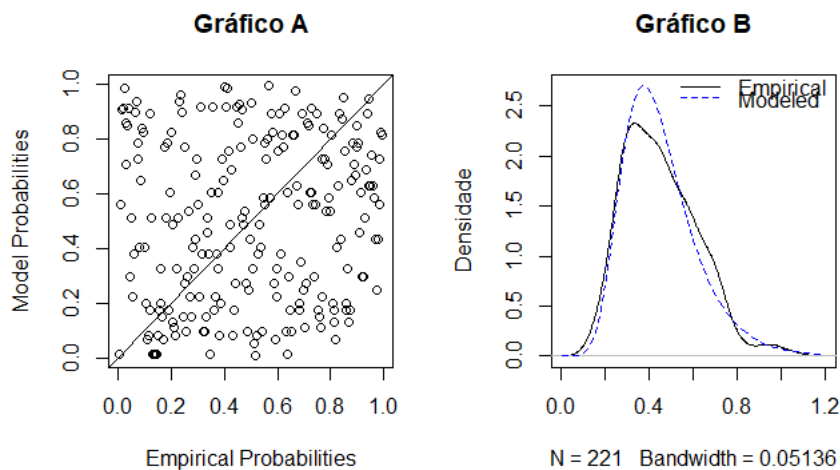


FIGURA 3.29: Gráficos de probabilidade (Gráfico A) e densidade de probabilidade (Gráfico B) do modelo Gumbel para os máximos mensais de Leixões.

Tal como na análise dos máximos anuais também se estudou a influencia de parâmetros variáveis, segundo a expressão 3.2. Desta forma, construíram-se gráficos de probabilidade e densidade de probabilidade para o modelo Gumbel com

localização variável (Gráficos A e B) e para o modelo Gumbel com escala variável (Gráficos C e D), apresentados na Figura 3.30. Verifica-se que existe um melhor ajustamento do modelo com parâmetros variáveis.

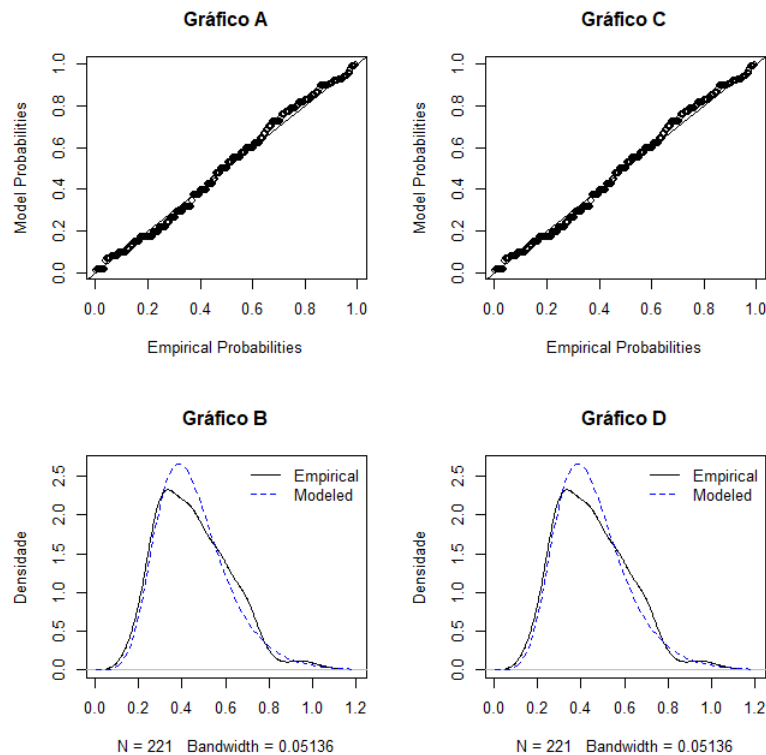


FIGURA 3.30: Gráficos de probabilidade e densidade de probabilidade, para os máximos mensais de Leixões, do modelo Gumbel com parâmetros a variar segundo a expressão 3.2. Os Gráficos A e B representam o ajuste da amostra à distribuição de Gumbel com o parâmetro localização a variar e os Gráfico C e D representam o ajuste da amostra com o parâmetro escala a variar.

Recorreu-se ao teste de *Likelihood-ratio* de forma a estudar se os modelos são distintos. Obteve-se  $Likelihood-ratio = 0$  e  $p - value > 0.9$  para a comparação entre os 3 modelos (Gumbel com parâmetros constantes, Gumbel com localização variável e Gumbel com escala variável), desta forma é possível afirmar que existe a evidência estatística que os ajustamentos dos modelos não são distintos.

De forma a perceber-se qual o melhor ajuste para o modelo de Gumbel calculou-se os valores AIC, BIC e RMSE. Através da Tabela 3.11, é possível verificar que não existe evidências que os modelos, Gumbel com localização variável e Gumbel com escala variável, sejam superiores quando comparados, assumindo-se que são

semelhantes. Assim, decidiu-se usar para futuros cálculos o ajuste Gumbel com localização variável.

TABELA 3.11: Indicadores estatísticos AIC, BIC e o teste RMSE do modelo de Gumbel para os máximos mensais de Leixões.

	<b>AIC</b>	<b>BIC</b>	<b>RMSE</b>
<b>Gumbel</b>	-181.398	-174.602	0.3909
<b>Gumbel (Localização variável)</b>	-181.398	-174.602	0.0221
<b>Gumbel (escala Variável)</b>	-181.398	-174.602	0.0221

De modo a perceber-se qual dos dois modelos (Gumbel com localização variável ou GEV com localização variável) se ajusta melhor, compararam-se os indicadores estatísticos dos mesmos. Através da tabela 3.12, é possível verificar os indicadores estatísticos têm menor valor para o modelo Gumbel com localização variável, assim sendo possível afirmar que este é o modelo que melhor a que a amostra se ajusta.

TABELA 3.12: Indicadores estatísticos AIC, BIC e o teste RMSE para os modelos GEV e Gumbel, com o parâmetros localização variável, para os máximos mensais de Leixões.

	<b>AIC</b>	<b>BIC</b>	<b>RMSE</b>
<b>GEV (Localização variável)</b>	-179.915	-169.721	0.0224
<b>Gumbel (Localização variável)</b>	-181.398	-174.602	0.0221

Escolhido o modelo de Gumbel com localização variável, avaliaram-se os níveis e períodos de retorno até ao final do século (aproximadamente daqui a 80 anos). Na Figura 3.31 está representado o gráfico de níveis de retorno, onde se verifica que o nível de retorno tem uma tendência, tal como na análise dos máximos anuais, crescente ao longo do tempo, pelo que poderá ser consequência das alterações climáticas e por sua vez da subida do nível médio do mar.

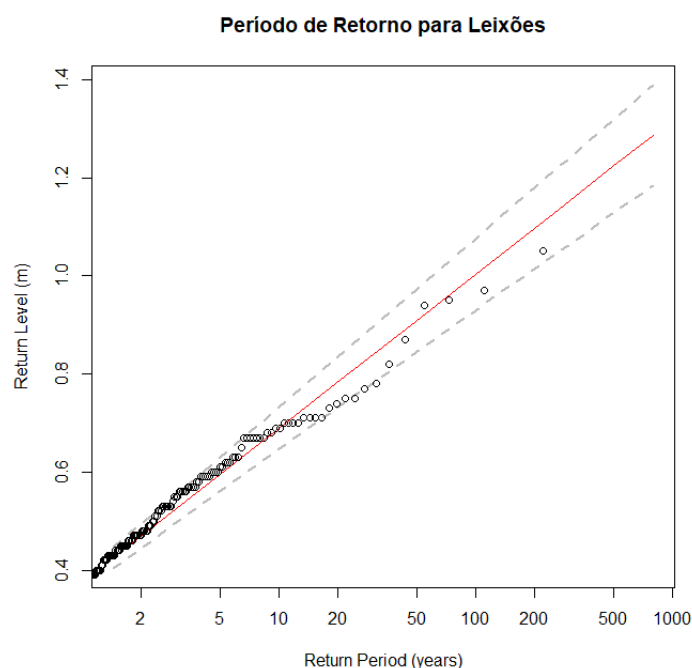


FIGURA 3.31: Gráfico de nível de retorno, para os máximos mensais, através do modelo de Gumbel com o parâmetro localização a variar segundo a expressão 3.2.

De modo a se ter uma análise mais precisa dos níveis de retorno, estudou-se estes para os períodos de retorno de 5, 20, 50 e 80 anos, expressos na Tabela 3.13. É de notar que os níveis de retorno estimados por este modelo não ultrapassaram o maior desvio registado (1.05 metros), sendo que até ao final do século seja pouco provável obter desvios superiores a 0.97 metros.

TABELA 3.13: Valores do nível de retorno estimado através do modelo de Gumbel com o parâmetro localização a variar segundo a expressão 3.2, para os máximos mensais de Leixões

Tempo (Anos)	5	20	50	80
Gumbel (Localização variável)	0.58m	0.78m	0.91m	0.97m

### Análise POT

Com esta análise pretende-se estudar os desvios máximos mensais de Leixões através do método POT, apresentado na secção 2.3, procedendo ao ajustamento de um modelo da GP para os excessos. Tal como a análise realizada para os máximos anuais de Leixões, na subsecção 3.1.2, iniciou-se esta análise pela delimitação de um *threshold*,  $\mu$ .

Na Figura 3.32 estão representados dois gráficos com as estimativas, pontuais e intervalares, dos parâmetros forma e escala segundo um determinado *threshold*. A partir do 0.55m existe uma estabilidade para as estimativas, pelo que este é um nível adequado para *threshold*.

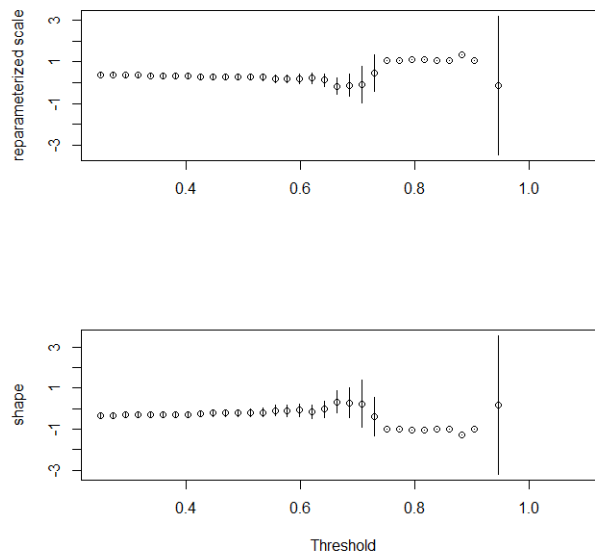


FIGURA 3.32: Estimativas dos parâmetros forma e escala *versus* *threshold* para os máximos mensais de Leixões.

De forma a corroborar  $\mu = 0.55$ , procedeu-se a construção de um *Mean Residual Life Plot*. Na Figura 3.33 é possível confirmar que 0.55m é o candidato mais adequado para *threshold*.

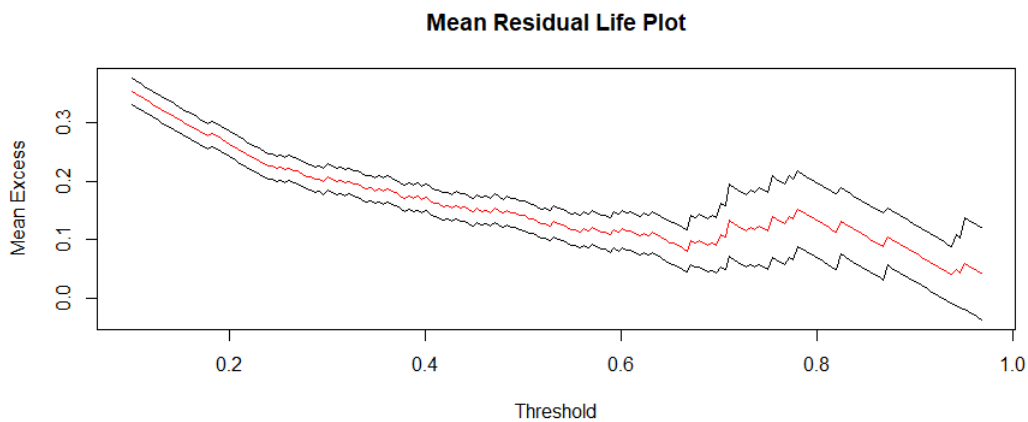


FIGURA 3.33: *Mean Residual Life Plot* para os máximos anuais de Leixões.

Considerando  $\mu = 0.55$  para a amostra dos desvios máximos mensais de Leixões, obteve-se 62 excedências em 221 valores, o que equivale, aproximadamente a 28% da amostra. De seguida realizou-se o ajustamento do modelo da Generalizada de Pareto. Na Figura 3.34 está representando um gráfico de probabilidades, onde se verifica, uma relação linear insatisfatória, o que não evidencia um bom ajustamento. Os parâmetros obtidos foram  $\delta = 0.174$  e  $\rho = -0.227$ .

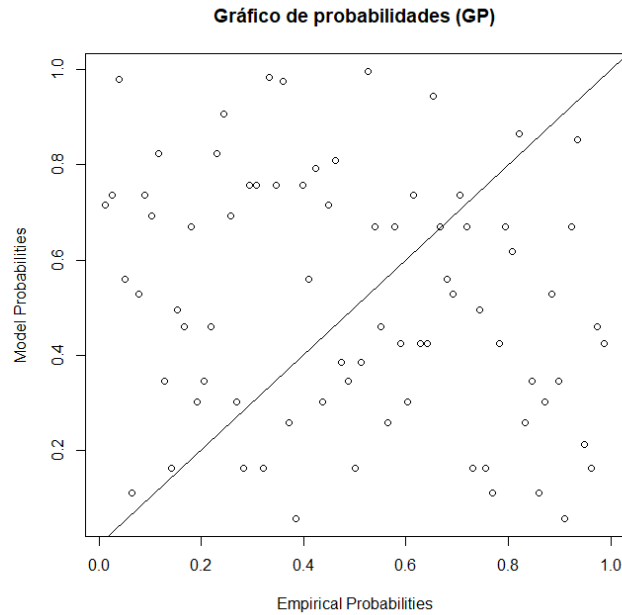


FIGURA 3.34: Gráfico de probabilidades dos excessos, referente à amostra dos máximos mensais de Leixões.

De forma a melhorar o modelo, procedeu-se, tal como na análise GP para os máximos anuais de Leixões na secção 3.1.2, à variação do parâmetro escala segundo a expressão 3.2. Na Figura 3.35 está representado o gráfico de probabilidades para o modelo GP com o parâmetro escala variável. Neste gráfico é possível verificar uma relação linear satisfatória.

De forma a verificar se os modelos são distintos realizou-se o teste de *Likelihood-ratio*, onde se obteve um  $p - value = 0.8781$ , pelo que existe evidência estatística que os ajustamentos do modelo GP não são distintos.

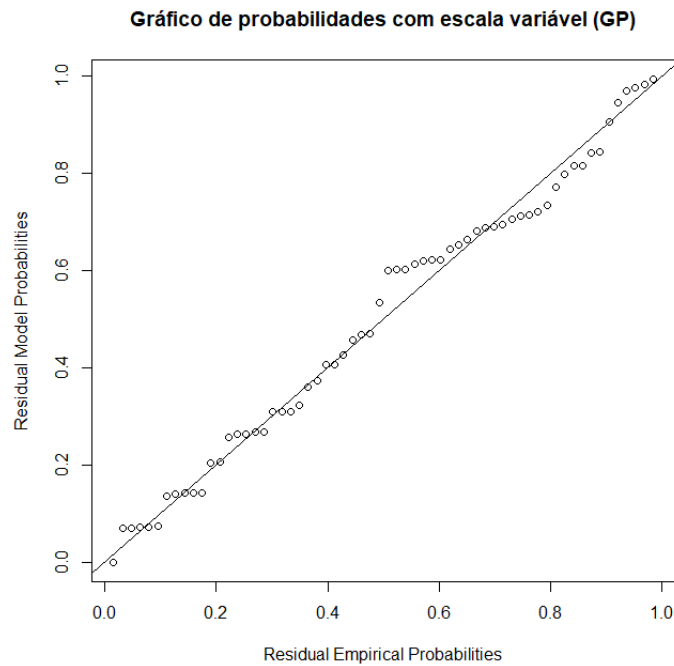


FIGURA 3.35: Gráfico de probabilidades dos excessos, com o parâmetro escala variável, referente à amostra nos máximos mensais de Leixões.

De seguida calcularam-se os indicadores estatísticos AIC, BIC e RMSE de modo a perceber qual o melhor modelo (GP com parâmetros constantes ou GP com o parâmetro escala variável) da amostra ao modelo GP. Através dos indicadores expressos na Tabela 3.14, é possível verificar que RMSE é menor para o modelo com o parâmetro escala a variar, os valores dos AIC e BIC são muito aproximados. Assim, consideramos que a distribuição que melhor se ajusta aos desvios máximos mensais é a GP com escala variável.

TABELA 3.14: Indicadores estatísticos AIC, BIC e o teste RMSE para o modelo GP com parâmetros constantes e com o parâmetro escala a variar segundo a expressão 3.2.

	<b>AIC</b>	<b>BIC</b>	<b>RMSE</b>
<b>GP</b>	-136.0506	-131.7964	0.4252
<b>GP (Escala variável)</b>	-134.0742	-127.6927	0.0310

Escolhido o modelo de GP com escala variável, procedeu-se ao estudo dos níveis e períodos de retorno. Na Figura 3.36 esta representado um gráfico com esses

níveis de retorno que, como no ajustamento através da distribuição de Gumbel com localização variável, apresenta uma tendência crescente ao longo do tempo.

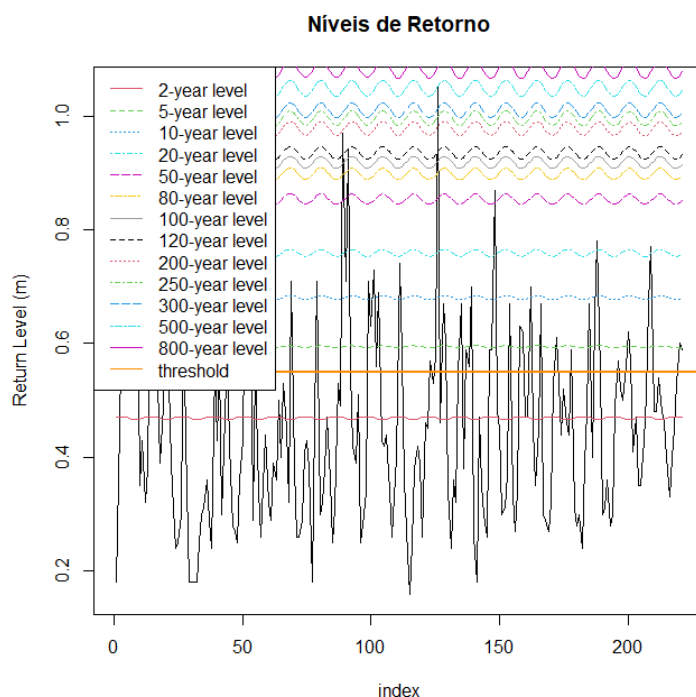


FIGURA 3.36: Gráfico de nível de retorno, para os máximos mensais de Leixões, segundo o modelo GP com escala variável segundo a expressão 3.2.

De forma a ter-se uma análise mais precisa dos níveis de retorno, calculou-se o nível de retorno para 5, 20, 50 e 80 anos. Através da Tabela 3.15 conclui-se que até ao final do século o maior nível de retorno previsto é de 0.91 metros.

TABELA 3.15: Valores do nível de retorno estimado através do modelo GP com escala variável, para os máximos anuais de Leixões.

Tempo (Anos)	5	20	50	80
GP (escala variável)	0.60m	0.77m	0.86m	0.91m

### 3.1.4 Resultados sobre o Estudo do Porto de Leixões

Nesta secção foi realizada uma análise estatística, através da Teoria de Valore Extremos aos desvios da altura de maré ocorridos no Porto de Leixões. Pretendia-se com esta análise identificar, num primeiro momento, quais a maiores sobrelevações que ocorreram na série das alturas horárias de Leixões, pelo que o maior desvio ocorrido foi no ano de 2013 atingindo uma altura de 1.05 metros.

Após isto iniciou-se a análise dos desvios máximos anuais de Leixões, onde se procurou qual a distribuição, dentro da família da Generalizada de Valores Extremos, que melhor se ajustava aos desvios. O modelo de Gumbel, com o parâmetro escala variável, foi o que melhor se ajustou aos desvios de Leixões, pelo que se procedeu ao estudo dos níveis de retorno. Para este modelo verificou-se que existe uma tendência crescente dos níveis de retorno ao longo do tempo, sendo que a maior sobre elevação prevista foi de 1.2 metros para um período de retorno de 80 anos. De seguida, estudou-se esta série aplicando o método de POT, onde se obteve para o mesmo período de retorno, um nível de retorno de 1.05 metros.

Uma vez que a amostra dos desvios máximos anuais é reduzida, com  $k = 19$ , realizou-se a mesma análise para os máximos mensais, com  $K = 221$ . Para um período de retorno de 80 anos, o modelo Gumbel com escala variável estimou um nível de retorno de 0.97 metros e o modelo POT estimou um nível de retorno de 0.91 metros.

Através da Capitania do Porto de Leixões, os cais deste porto estão a uma altura de 6,33 metros em relação ao ZH. Tendo em conta as previsões realizadas é possível concluir que, mantendo-se as condições atuais, é pouco provável que existam galgamentos neste porto. Assim, até ao final do século é pouco provável que existam fenómenos extremos que ultrapassem a altura dos cais, considerando a combinação entre a maior maré astronómica (3.99m) e o desvio máximo previsto (1.2m) tem-se uma altura de maré de 5.19m. Há ainda a considerar os efeitos da agitação marítima e da ocorrência de seichas dentro do porto, mas devido ao abrigo proporcionado pelos molhos, no caso das primeiras, e a amplitude pequena das segundas, estes efeitos não afetam de forma significativa a sobrelevação da maré.

## 3.2 Porto de Sines

O Porto de Sines, apresentado na Figura 3.37, é um porto de águas profundas que está aberto ao mar, contando com excelentes condições de acessibilidade marítima. Este porto é o que mais movimentada mercadorias a nível nacional, apresentando ainda condições para acolher todo o tipo de navios.

É o principal porto de Portugal, cujo as suas características geo-físicas contribuem para ser a principal porta de abastecimento energético do país (petróleo e gás natural), sendo também um importante porto de carga. A sua construção, em 1978, estrategicamente localizada de longe de pressões urbanas assegura capacidade de expansão a longo prazo (Cacho et al., 2020).



FIGURA 3.37: Fotografia aérea do Porto de Sines (Fonte: <https://regiao-sul.pt/2022/01/25/economia/porto-de-sines-continua-a-registar-crescimento/567014>, 03/04/2022).

Devido à importância estratégica do Porto de Sines para Portugal é necessário uma monitorização e estudo constante de fenómenos naturais que possam vir a danificar o porto, impedindo o seu normal funcionamento.

#### 3.2.1 Análise Preliminar das Sobrelevações no Porto de Sines

Para o estudo das sobrelevações com origem metrológica no Porto de Sines foi utilizada um série temporal dos desvios entre a maré astronómica e a maré observada, desde 1997 até 2020. Através do gráfico representado na Figura 3.38, é possível verificar, a existência de lacunas na série.

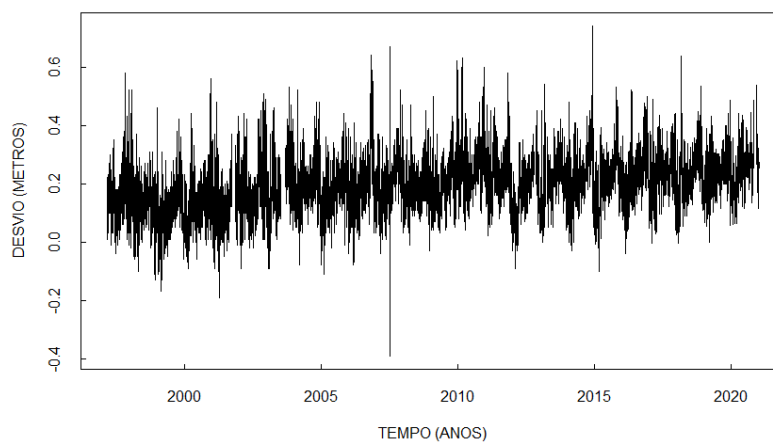


FIGURA 3.38: Série temporal dos desvios de Sines (1997-2020).

De seguida construiu-se para cada ano da série um diagrama de extremos e quartis, representados na Figura 3.39, onde se observa que existe uma assimetria positiva dos dados, ou seja, existe uma maior concentração dos dados nas sobrelevações do que nas subelevações.

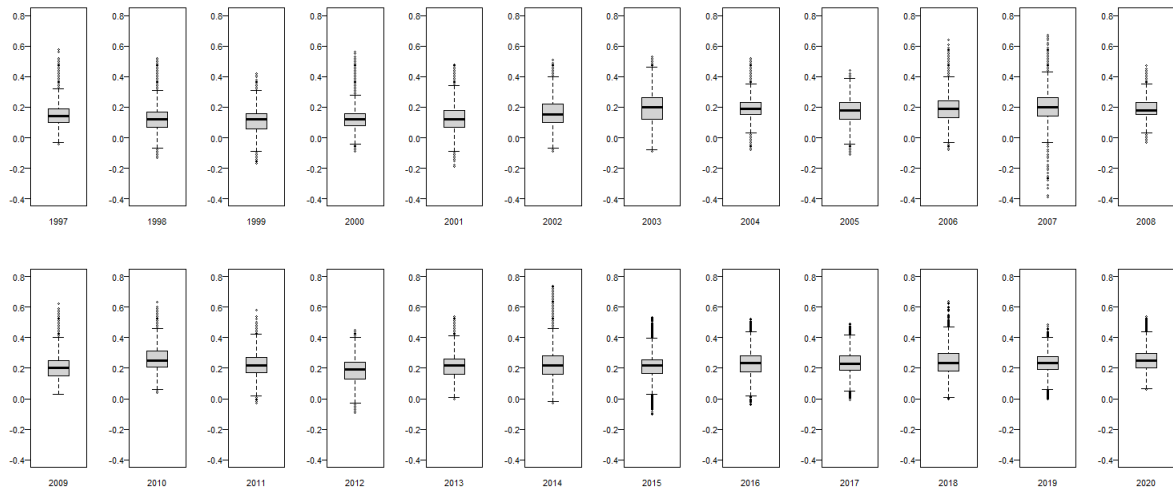


FIGURA 3.39: Diagramas de extremos e quartis dos desvios da maré por ano (Sines).

Calcularam-se as principais estatísticas descritivas: o desvio mínimo e máximo, a média, a mediana e 1º e 3º quartis. Da Tabela 3.16 é possível retirar que a sobrelevação máxima, ocorrida no porto de Sines, foi de 0.74 metros, já a subelevação máxima foi -0.19 metros.

TABELA 3.16: Estatística descritiva da série dos desvios de maré para o Porto de Sines.

Média	Mediana	Desvio Padrão	Máximo	Mínimo	1º Quartil	3º Quartil
0.194	0.190	0.09134431	0.740	-0.390	0.130	0.250

De seguida construiu-se o histograma, representado na Figura 3.40. Para avaliar se os desvios seguem uma distribuição normal, aplicou-se o teste de *Shapiro-Wilk*, obteve-se um valor da estatística de teste  $W = 0.99538$  e  $p - value = 1.671 \times 10^{-11}$ , pelo que a hipótese da série seguir uma distribuição normal é rejeitada.

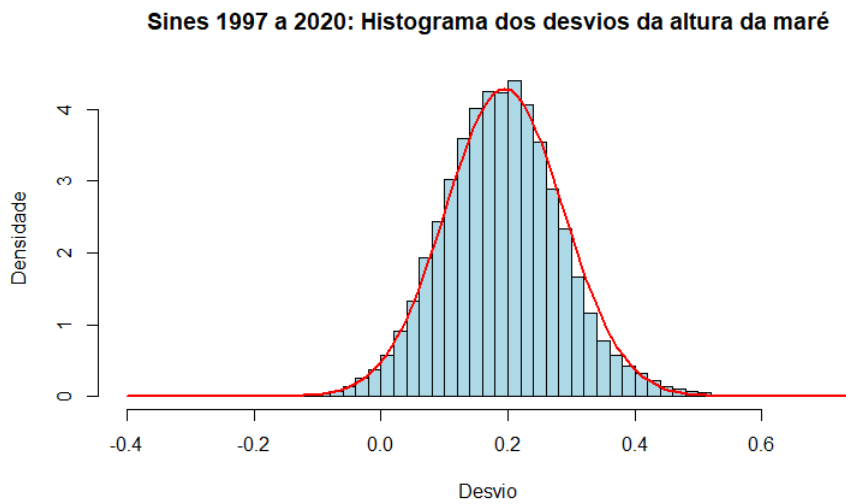


FIGURA 3.40: Histograma da série dos desvios de maré do Porto de Sines.

Repare-se que ao longo da série existem diversos valores omissos, pelo que se procedeu, tal como no estudo para Leixões, à imputação dos valores omissos pela mediana. Na Figura 3.41 está representada a série dos desvios de maré do Porto de Sines sem valores omissos.

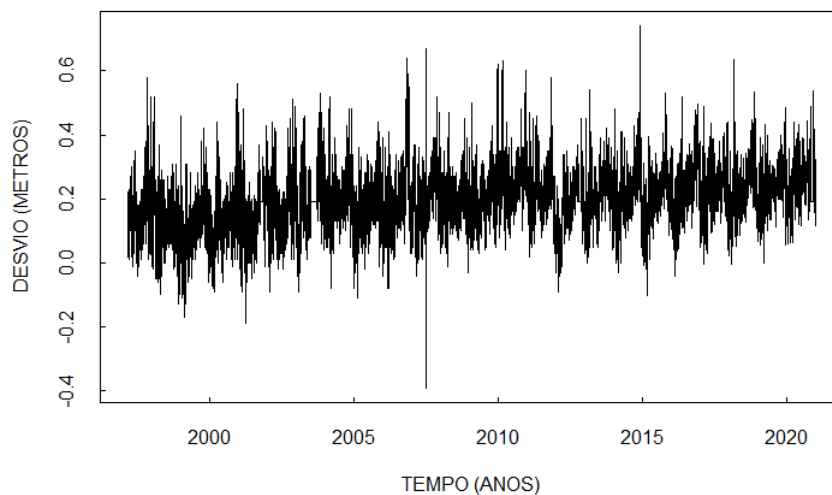


FIGURA 3.41: Série temporal dos desvios de maré do Porto de Sines sem falhas.

Tal como na análise ao Porto de Leixões, realizou-se a análise das séries dos máximos anuais e máximos mensais.

### 3.2.2 Máximos Anuais do Porto de Sines

Usando a mesma metodologia que na subsecção 3.1.2, procedeu-se ao estudo da série dos máximos anuais, de modo a verificar qual das distribuições (GEV, Gumbel, Fréchet ou Max-Weibull) se ajusta melhor. Neste caso, dispõem-se de  $K = 24$  desvios, tal como ilustra a Figura 3.42.

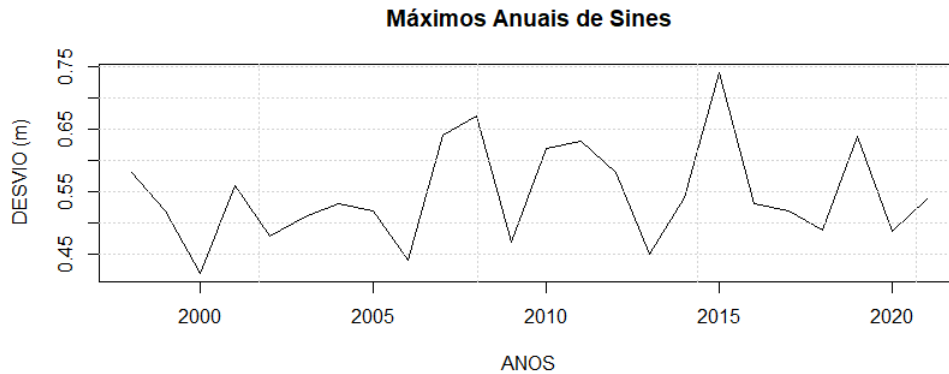


FIGURA 3.42: Desvios máximos anuais do Porto de Sines.

Nesta dissertação recorreu-se à estatística de extremos para o estudo das sobrelevações no porto de Sines, pelo que é necessário que os dados sejam independentes e identicamente distribuídos. O histograma apresentado na Figura 3.43 verifica-se, uma assimetria à esquerda dos dados, o que induz a rejeição da normalidade. De modo a avaliar se os desvios máximos anuais seguem distribuição normal recorreu-se ao teste de *Shapiro-Wilk*, onde se obteve  $W = 0.96089$  e  $p - value = 0.4566$ , assim, existe evidência estatística de que os dados em análise seguem a distribuição normal. Tal como sucedido para a análise dos desvios máximos para Leixões e este resultado pode ser explicado pela dimensão dos dados.

Através da análise da Figura 3.43 e da Tabela 3.17 verifica-se um enviesamento à esquerda dos dados, induzindo que os modelos de Gumbel e GEV terão um bom ajustamento aos dados.

TABELA 3.17: Estatística descritiva da série dos desvios máximos anuais para o Porto de Sines.

Nº de Obs.	Média	Mediana	Desvio Padrão	Máximo	Mínimo	1º Quartil	3º Quartil
24	0.5459	0.5308	0.07869498	0.7400	0.4200	0.4880	0.5900

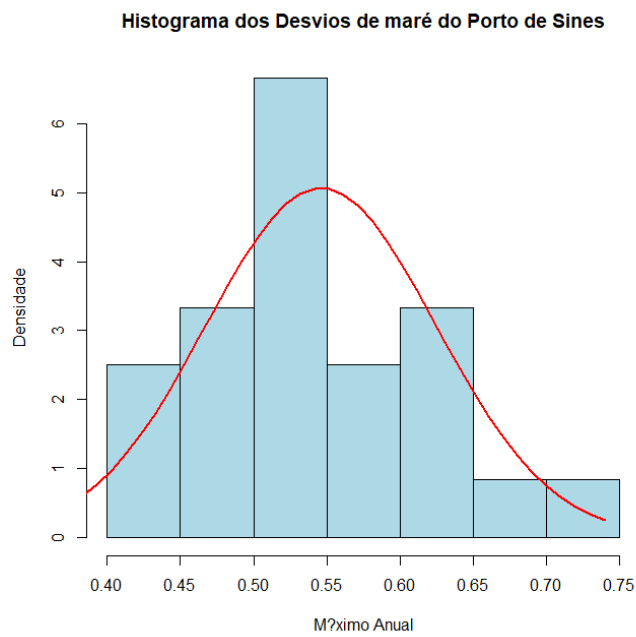


FIGURA 3.43: Histograma dos desvios máximos anuais do Porto de Sines.

De modo a testar a independência dos desvios recorreu-se aos gráficos ACF e PACF. A partir da Figura 3.44 é possível verificar que não existe correlação dos dados.

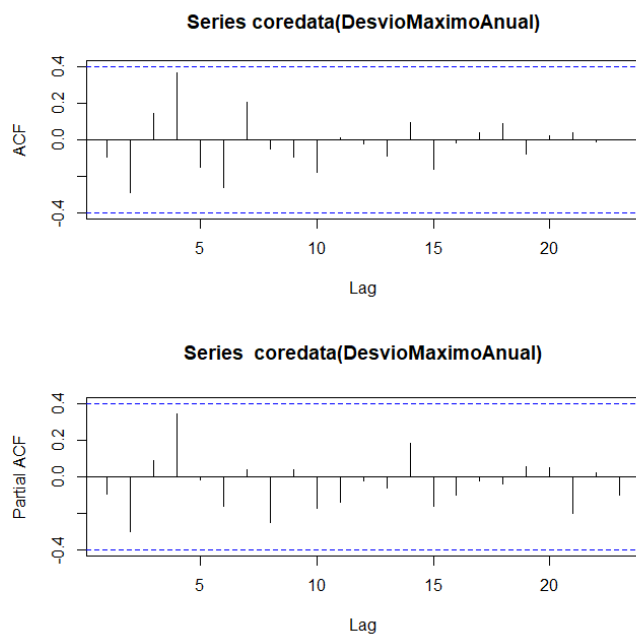


FIGURA 3.44: Gráficos de auto-correlação e auto-correlação parcial para os máximos anuais do Porto de Sines.

### Análise GEV para os Máximos Anuais de Sines

A metodologia utilizada foi a mesma que se apresentou para os desvios máximos anuais do porto de Leixões. Foram identificados os modelos GEV com parâmetros constantes, GEV com parâmetros variáveis (de acordo com a expressão 3.1), Gumbel com parâmetros constantes e Gumbel com parâmetros variáveis. A tabela 3.18 apresenta os valores de AIC, BIC e RMSE para esses modelos:

TABELA 3.18: Indicadores estatísticos AIC e BIC e teste RMSE para as distribuições GEV e Gumbel com o parâmetro localização variável, Porto de Sines.

	AIC	BIC	RMSE
<b>GEV</b>	-51.0578	-47.52364	0.3768484
<b>GEV (Localização variável)</b>	-51.0578	-47.52364	0.03748532
<b>GEV (Escala Variável)</b>	-51.0578	-47.52364	0.03748532
<b>Gumbel</b>	-52.89662	-50.54052	0.3746473
<b>Gumbel (Localização variável)</b>	-52.89662	-50.54052	0.03537697
<b>Gumbel (Escala Variável)</b>	-52.89662	-50.54052	0.03537697

Escolhido o modelo Gumbel com localização variável, estudou-se os níveis de retorno para diferentes períodos. Na Figura 3.45 verifica-se que o nível de retorno tem uma tendência crescente ao longo do tempo, o que poderá ser uma consequência do aumento do nível médio das águas do mar.

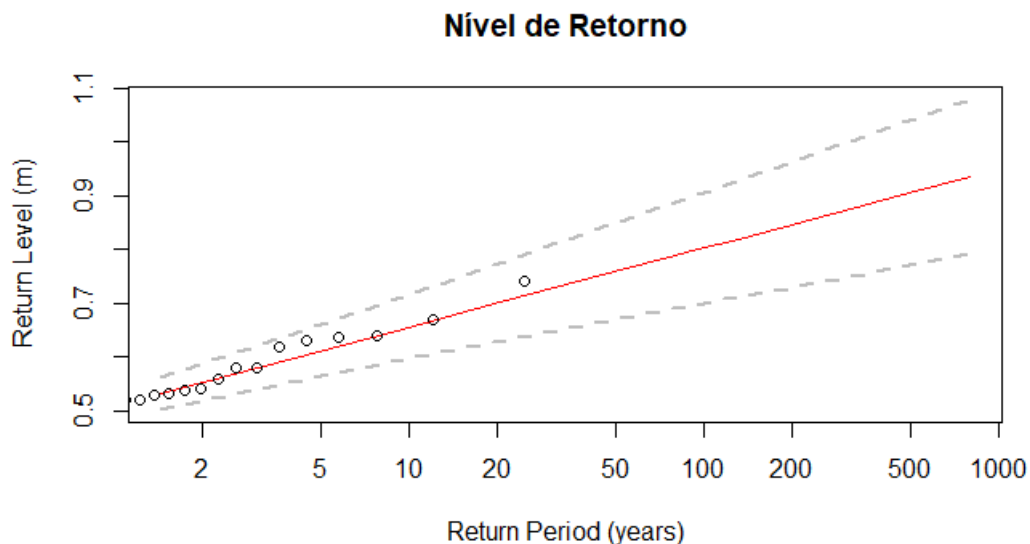


FIGURA 3.45: Gráfico de nível de retorno para os máximos anuais de Sines, através do modelo Gumbel com o parâmetro localização a variar segundo a expressão 3.1.

De modo a ter uma análise mais precisa dos níveis de retorno até ao final do século, calculou-se o nível de retorno para os períodos de 5, 20, 50 e 80 anos, expressos na Tabela 3.19, onde se verificou que o maior nível de retorno previsto será de 0.79 metros.

TABELA 3.19: Valores do nível de retorno para os máximos anuais de Sines, através do modelo Gumbel com o parâmetro localização a variar segundo a expressão 3.1.

Tempo (Anos)	5	20	50	80
Gumbel (Localização variável)	0.61 m	0.70 m	0.76 m	0.79 m

### Análise GP para os Máximos Anuais do Porto de Sines

Tal como descrito na secção 2.3, o método POT inicia-se por delimitar um nível elevado,  $\mu$ , que se denomina de *threshold*.

Na Figura 3.47 apresentam-se os gráficos com as estimativas, pontuais e intervalares, dos parâmetros forma e escala segundo um *threshold*, a partir do 0.5 existe uma estabilidade para as estimativas, pelo que este é a escolha adequada para *threshold*.

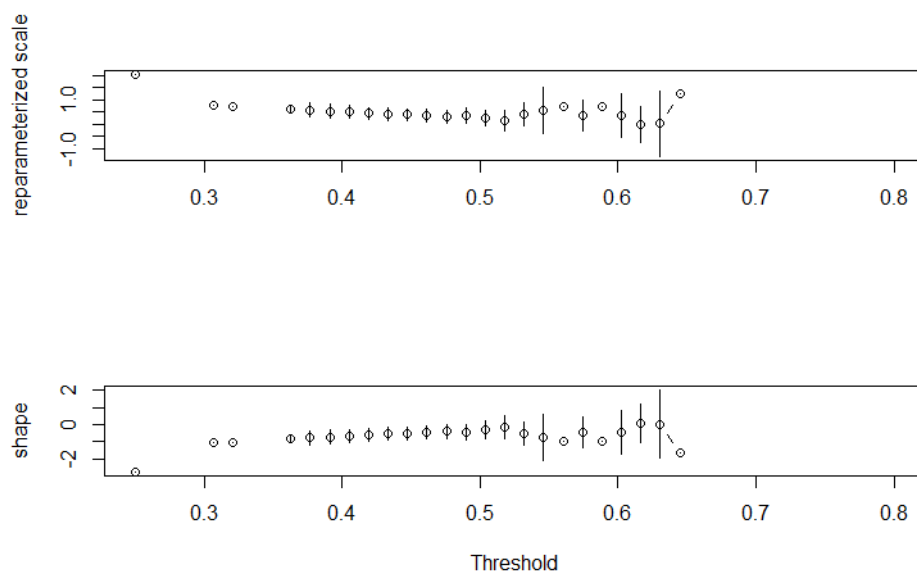


FIGURA 3.46: Estimativas dos parâmetros forma e escala versus threshold para os máximos anuais do Porto de Sines.

De forma a confirmar  $\mu = 0.5$ , procedeu-se á análise de um *Mean Residual Life Plot*, representado na Figura 3.47, onde é possível identificar 0.5m como o candidato a *threshold* mais adequado.

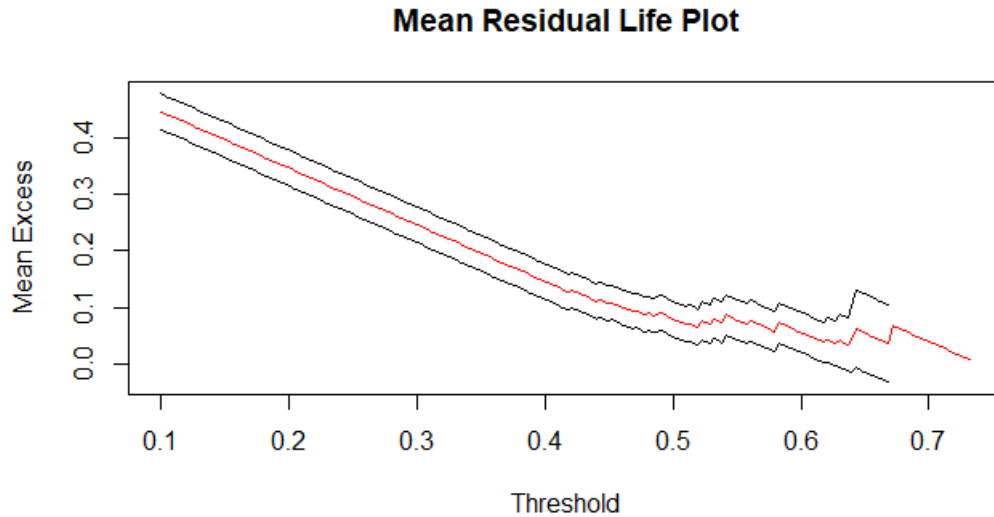


FIGURA 3.47: *Mean Residual Life Plot* para os máximos anuais do Porto de Sines.

Considerando  $\mu = 0.5$  para a amostra dos desvios anuais do Porto de Sines, obtiveram-se 17 excedências em 24 valores, o que equivale, aproximadamente, a 71% dos dados. Como na subsecção 3.1.2 realizou-se o ajuste do modelo GP com parâmetros constantes e com parâmetros variáveis, na Tabela 3.25 estão representados os valores dos indicadores estatísticos AIC e BIC e RMSE para cada modelo.

TABELA 3.20: Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GP, com o parâmetro escala constante e variável segundo a expressão 3.1.

	AIC	BIC	RMSE
<b>GP</b>	-49.1227	-47.45627	0.3589006
<b>GP (Escala variável)</b>	-47.4808	-44.98116	0.06508885

Como é possível verificar o valor do teste de RMSE é menor para o ajustamento da GP com o parâmetro escala variável, deste modo estudou-se os níveis e períodos de retorno para este modelo. Na Figura 3.48 está representado um gráfico de níveis de retorno, onde é possível verificar que o nível de retorno tem uma tendência crescente, tal como na análise feita para Leixões.

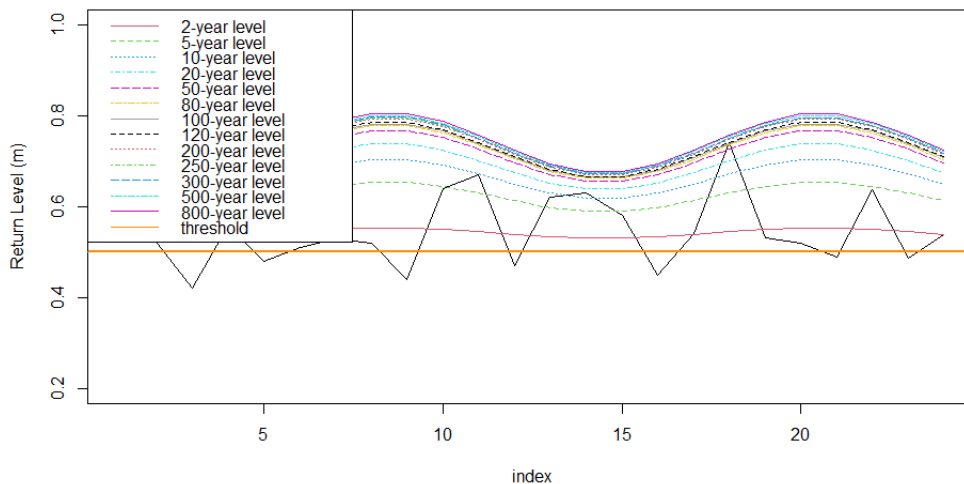


FIGURA 3.48: Gráfico de nível de retorno, para os máximos anuais de Sines, através do modelo GP com escala a variar segundo a expressão 3.1.

Para uma análise mais precisa dos níveis de retorno, construiu-se a Tabela 3.21, onde estão calculados os níveis de retorno para 5, 20, 50 e 80 anos. Através deste modelo conclui-se que o maior nível de retorno previsto até ao final do século é de 0.78 metros.

TABELA 3.21: Valores do nível de retorno estimado através do modelo GP com o parâmetro escala a variar segundo a expressão 3.1, para o Porto de Sines.

Tempo (Anos)	5	20	50	80
GP (escala variável)	0.65 m	0.74 m	0.77 m	0.78 m

### 3.2.3 Máximos Mensais para o Porto de Sines

Foi utilizada a mesma metodologia que na subsecção 3.1.3, onde se estudaram os desvios máximos mensais do Porto de Sines. Neste caso, a série dos desvios máximos mensais, representada na Figura 3.49, tem  $k = 286$ .

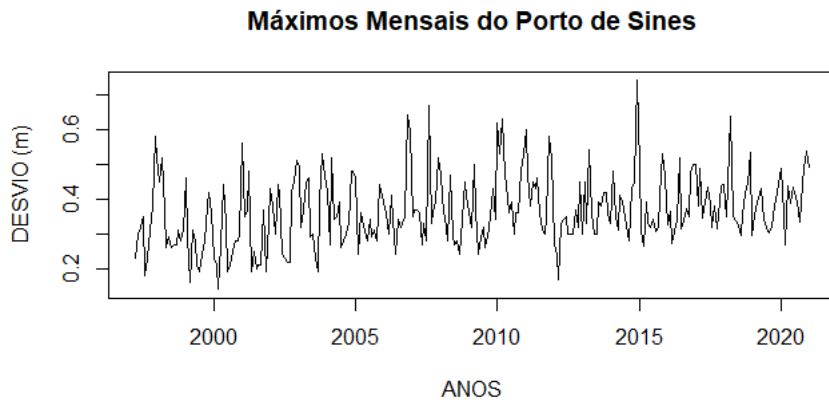


FIGURA 3.49: Representação gráfica da série dos máximos mensais do Porto de Sines, no período de 1997 a 2020.

De seguida, de modo a estudar a normalidade da série recorreu-se a um histograma, apresentado na Figura 3.50. Neste histograma verifica-se uma ligeira cauda para a direita. Por forma a avaliar a normalidade dos dados recorreu-se ao teste de *Shapiro-Wilk*, onde se obteve  $w = 0.97684$  e  $p - value = 0.0001371$ , assim, existe evidência estatística para rejeitar a normalidade dos desvios máximos mensais. Na Tabela 3.22 estão apresentados as medidas descritivas da série dos desvios de maré de Sines.

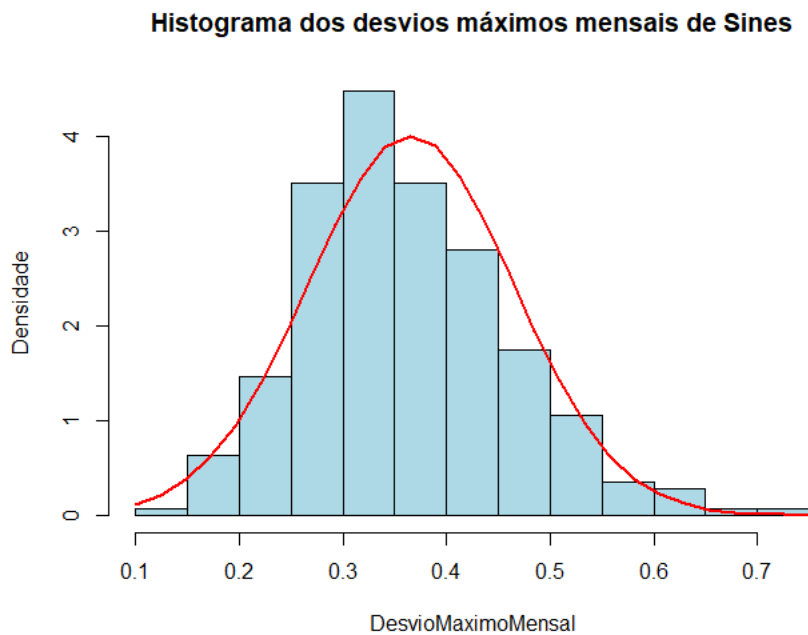


FIGURA 3.50: Histograma dos desvios máximos mensais do Porto de Sines.

TABELA 3.22: Medidas descritivas dos desvios máximos mensais do Porto de Sines.

Nº de Obs.	Média	Mediana	Desvio Padrão	Máximo	Mínimo	1º Quartil	3º Quartil
286	0.3655	0.3500	0.0996	0.7400	0.1400	0.3000	0.4314

Vamos avaliar se as observações são i.i.d. Desta forma, recorreu-se aos gráficos de ACF e PACF, apresentados na Figura 3.51; é possível verificar que só se garante a independência para desvios separados por 110 horas (4 dias e 14 horas). Uma vez que os dados usados são os desvios máximos mensais a independência fica assegurada.

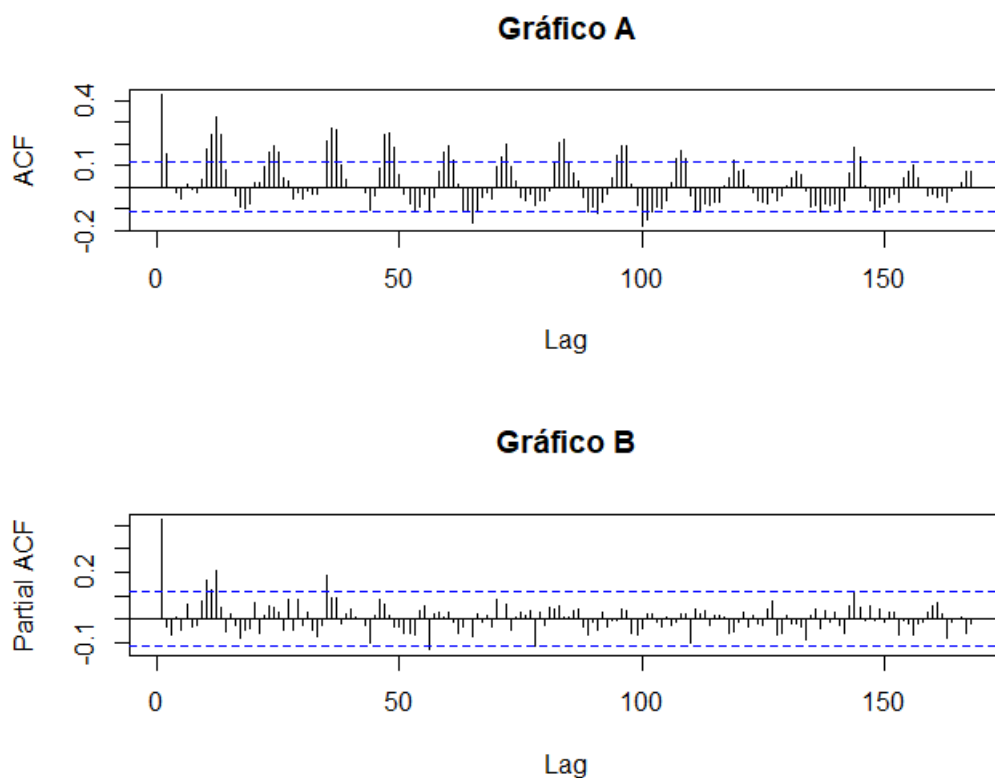


FIGURA 3.51: Gráficos de auto-correlação (Gráfico A) e auto-correlação parcial (Gráfico B) para os máximos mensais do Porto de Sines.

Construí-se um diagrama de extremos e quartis, representado na Figura 3.52, onde se verificou a existência de outliers.

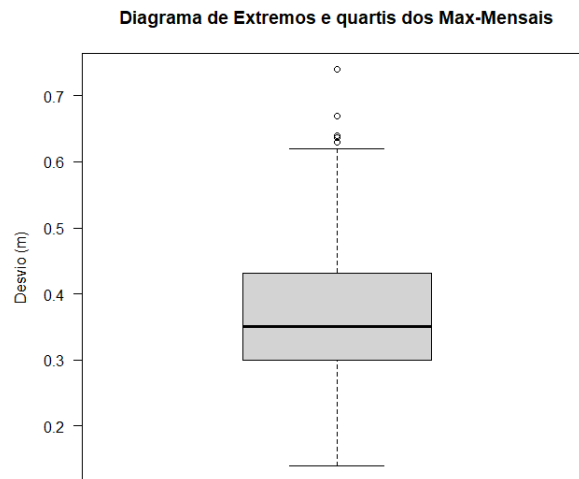


FIGURA 3.52: Diagrama de extremos e quartis correspondente aos desvios máximos mensais do Porto de Sines.

### Análise GEV para os Máximos Mensais do Porto de Sines

Foi utilizada a metodologia usada análises das subsecções 3.1.2, 3.1.3 e 3.2.3. Através dos indicadores estatísticos AIC e BIC e RMSE apresentados na Tabela 3.24, é possível identificar que a distribuição que melhor se ajusta aos dados é a Gumbel com um dos parâmetros (localização ou escala) a variar segundo a expressão 3.2.

TABELA 3.23: Indicadores estatísticos AIC e BIC e teste RMSE para as distribuições GEV e Gumbel com o parâmetro localização variável, máximos mensais do Porto de Sines.

	<b>AIC</b>	<b>BIC</b>	<b>RMSE</b>
<b>GEV</b>	-519.3069	-508.3389	0.3372146
<b>GEV (Localização variável)</b>	-519.3069	-508.3389	0.01720799
<b>GEV (escala Variável)</b>	-519.3069	-508.3389	0.01720799
<b>Gumbel</b>	-513.3489	-506.0369	0.3306266
<b>Gumbel (Localização variável)</b>	-513.3489	-506.0369	0.02084774
<b>Gumbel (escala Variável)</b>	-513.3489	-506.0369	0.02084774

Escolhido o modelo Gumbel com localização variável, estudou-se os níveis de retorno em expressão dos períodos de retorno. Na Figura 3.53 verifica-se que o nível de retorno tem uma tendência crescente ao longo do tempo, o que poderá ser uma consequência do aumento do nível médio das águas do mar.

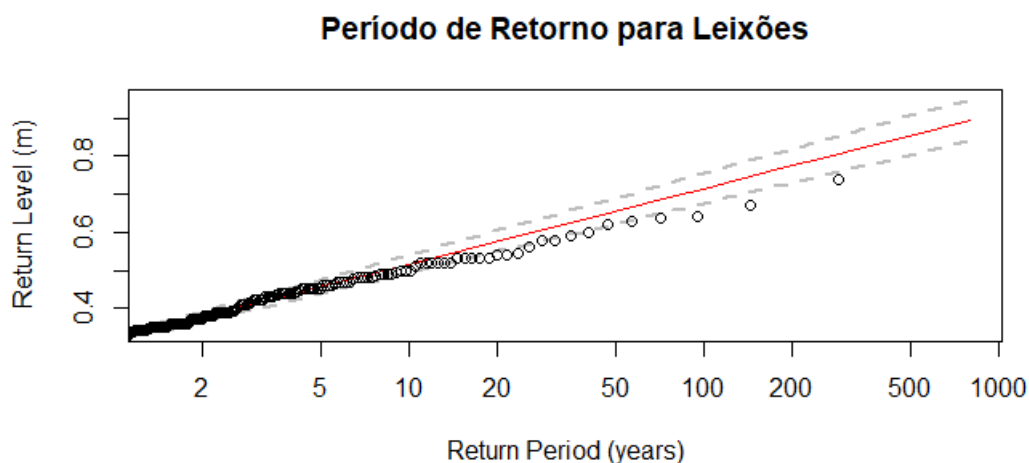


FIGURA 3.53: Gráfico de nível de retorno para os máximos mensais de Sines, através do modelo Gumbel com o parâmetro localização a variar segundo a expressão 3.2.

De modo a ter uma análise mais precisa dos níveis de retorno até ao final de século, calculou-se o nível de retorno para os períodos de 5, 20, 50 e 80 anos, expressos na Tabela 3.24, onde se verificou que o maior nível de retorno previsto será de 0.7 metros.

TABELA 3.24: Valores do nível de retorno para os máximos mensais de Sines, através do modelo Gumbel com o parâmetro localização a variar segundo a expressão 3.2.

Tempo (Anos)	5	20	50	80
Gumbel (Localização variável)	0.45 m	0.57 m	0.66 m	0.70 m

### Análise GP para os Máximos Mensais do Porto de Sines

Nesta análise, tal como na subsecção 3.2.2, pretende-se aplicar o método de POT aos desvios máximos mensais do Porto de Sines. Tal como apresentado na secção 2.3, o método de POT inicia-se pela escolha de um *threshold*,  $\mu$ .

A Figura 3.54 apresentado os gráficos com as estimativas, pontuais e intervalares, dos parâmetros forma e escala segundo um *threshold*, a partir de 0.37m existe uma estabilidade para as estimativas, pelo que este é a escolha adequada para *threshold*.

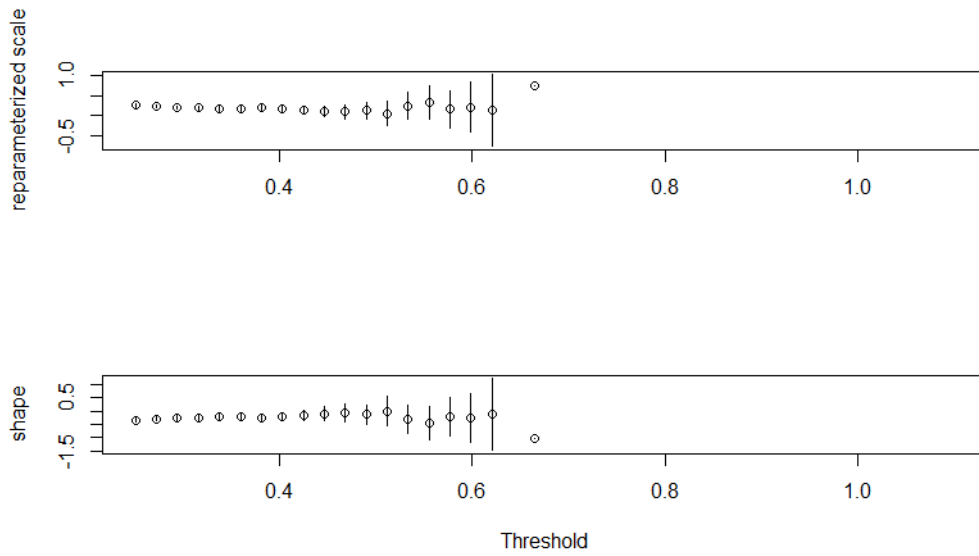


FIGURA 3.54: Estimativas dos parâmetros forma e escala *versus* *threshold* para os máximos mensais do Porto de Sines.

De forma a corroborar  $\mu = 0.37$ , procedeu-se à análise de um *Mean Residual Life Plot*, representado na Figura 3.55, onde é possível confirmar 0.37m como o candidato a *threshold* mais adequado.

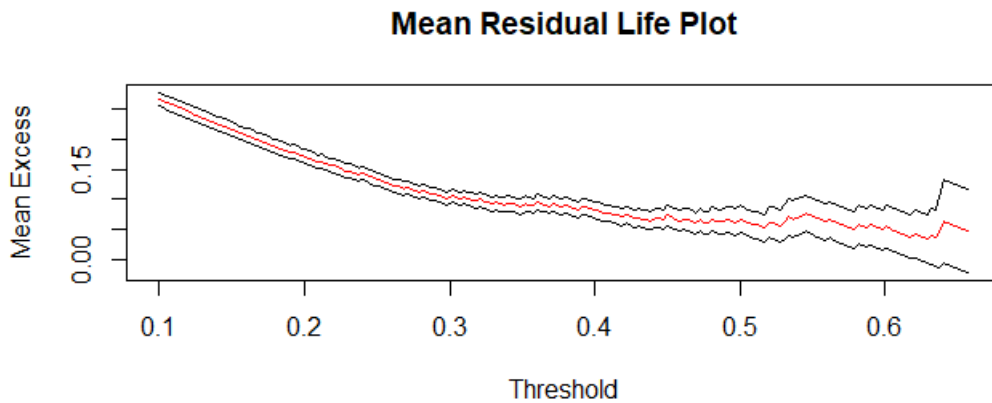


FIGURA 3.55: *Mean Residual Life Plot* para os máximos mensais do Porto de Sines.

Considerando  $\mu = 0.37$  para os desvios máximos mensais do Porto de Sines, obteve-se 119 excedências em 286 valores, o que equivale, aproximadamente, a 42% da amostra. Como na subsecção 3.2.2 realizou-se o ajuste do modelo da GP com parâmetros constantes e com parâmetros variáveis. Na Tabela 3.25 estão representados os valores dos indicadores estatísticos AIC e BIC e RMSE para cada modelo.

TABELA 3.25: Indicadores estatísticos AIC, BIC e o teste RMSE para a distribuição GP, com o parâmetro escala constante e variável segundo a expressão 3.2.

	AIC	BIC	RMSE
GP	-338.8078	-333.2495	0.3372146
GP (Escala variável)	-340.5147	-332.1773	0.03262527

Como é possível verificar o valor de RMSE é menor para o ajustamento da distribuição generalizada de Pareto com o parâmetro escala variável, deste modo estudou-se os níveis e períodos de retorno para este modelo.

Na Figura 3.56, está representado um gráfico de níveis de retorno, onde é possível retirar que o nível de retorno tem uma tendência crescente, tal como na análise feita para Leixões.

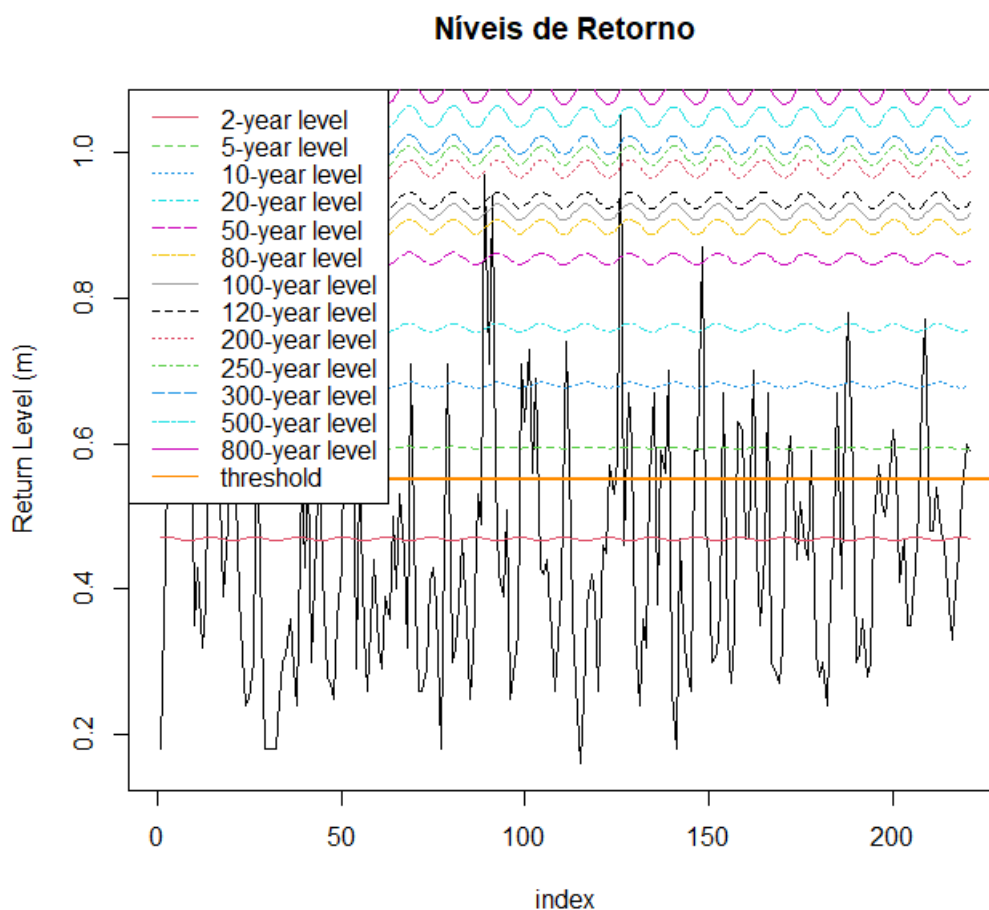


FIGURA 3.56: Gráfico de nível de retorno, para os máximos mensais de Sines, através do modelo GP com escala a variar segundo a expressão 3.2.

Para uma análise mais precisa dos níveis de retorno, construiu-se a Tabela 3.26, onde estão calculados os níveis de retorno para 5, 20, 50 e 80 anos. Através deste modelo concluiu-se que o maior nível de retorno previsto até ao final do século é de 0.77 metros.

TABELA 3.26: Valores do nível de retorno estimado através do modelo GP com o parâmetro escala a variar segundo a expressão 3.2.

<b>Tempo (Anos)</b>	<b>5</b>	<b>20</b>	<b>50</b>	<b>80</b>
<b>GP (escala variável)</b>	0.65 m	0.72 m	0.75 m	0.77 m

### 3.2.4 Resultados sobre o Estudo do Porto de Sines

Nesta secção foi realizada uma análise dos desvios, entre a maré astronómica e maré observada, do Porto de Sines, através da Teoria de Valores Extremos. Esta análise visava identificar quais as maiores sobrelevações que ocorreram na série das altura horárias do Porto de Sines, pelo que o maior desvio ocorrido foi de 0.74 metros em 2014.

De seguida, iniciou-se a análise da série dos desvios máximos anuais do Porto de Sines, onde se procurou identificar qual a distribuição, dentro da família da Generalizada de Valores Extremos (Gumbel, Fréchet ou Max-Weibull), que melhor se ajustava aos desvios. Tal como na análise dos desvios de maré do Porto de Leixões, a distribuição de Gumbel, com o parâmetro localização a variar, foi a que melhor se ajustou à série dos máximos anuais. Após isto, procedeu-se ao estudo dos níveis de retorno até ao final do século, sendo o maior desvios estimado de 0,79 metros. Em seguida, estudou-se esta série segundo o método de POT, onde se obteve para o período de retorno de 80 anos um nível de retorno de 0.78 metros.

A série dos desvios máximos anuais do Porto de Sines é reduzida, com  $k = 24$  observações, pelo que se realizou a mesma análise para a série dos máximos mensais do Porto de Sines, com  $k = 286$  observações. A distribuição que melhor se ajustou aos dados foi a de Gumbel com o parâmetro escala variável. Através desta distribuição estimou-se que até ao final do século o nível de retorno máximo é de 0.7 metros. Já para o modelo POT obteve-se um nível de retorno de 0.77 metros.

# Conclusão

Nesta dissertação realizou-se uma análise estatística aos desvios, entre a maré observada e a prevista, ocorridos nos Portos de Leixões e Sines, os dois maiores portos de Portugal Continental. Pretendia-se numa primeira fase identificar as maiores sobrelevações na zona sul do país (Porto de Sines) e na zona Norte (Porto de Leixões), de modo a estudar a variação amplitude dos desvios de maré segundo a zona geográfica onde foram registados. Fortunato et al. (2014) publicou um estudo sobre os níveis extremos de maré na costa portuguesa, afirmando que existe um aumento da amplitude dos desvios de maré sul para norte. Pela análise realizada no capítulo 3 verifica-se que os desvios para Leixões são mais elevados do que para Sines, sendo as maiores desvios observados de 1.05 e 0.74 metros, respetivamente. Deste modo, confirmaram-se as conclusões de estudos anteriores, a amplitude dos desvios entre a maré observada e a maré astronómica, nas costa portuguesa, aumentam de sul para norte.

De seguida, recorreu-se à Teoria de Valores Extremos para estudar as séries dos desvios máximos de maré dos Portos de Leixões e Sines. As séries originais eram constituídas pelas alturas horárias compreendidas nos períodos de 31/12/1994 a 31/12/2020, para o Porto de Leixões, e de 06/03/1997 a 31/12/2020, para o Porto de Sines. A partir destas séries foram construídas duas sub-séries, as sub-séries dos desvios máximos anuais para Sines e Leixões e as sub-séries dos desvios máximos mensais para Sines e Leixões. A estas sub-séries foram aplicadas duas abordagens, a metodologia GEV e a metodologia POT.

Num primeiro momento, pretendeu-se identificar qual das distribuições (GEV, Gumbel, Fréchet ou Max-Weibull) que melhor se ajusta às séries dos desvios máximos anuais e mensais, com vista a realizar o estudo dos níveis e períodos de retorno, que são fundamentais na prevenção de catástrofes associadas à ocorrência de sobrelevações do nível do mar com origem meteorológica. Neste trabalho verificou-se que a distribuição de Gumbel, com o parâmetro localização a variável, foi a que melhor se ajustou aos dados dos máximos anuais e mensais para ambos os portos. Através da análise Gumbel, com parâmetro de localização variável, para os máximos anuais conclui-se que até ao final do século o maior desvio previsto é 1.2 e 0.79 metros para

Leixões e Sines, respetivamente. Já para as séries os desvios máximos mensais a distribuição Gumbel, com parâmetro de localização variável, estimou valores menores para os desvios até ao final do século, sendo a maior previsão de 0.97 e 0.7 para Leixões e Sines, respetivamente.

Numa segunda fase, realizou-se a análise para as séries dos máximos mensais e anuais de Leixões e Sines através da metodologia POT. Esta metodologia estimou valores menores do que a metodologia GEV, até ao final do século os maiores desvios estimados foram de 1.05 e 0.78 metros para Leixões e Sines, respetivamente.

Um aspeto a ser constatado na análise efetuada é o facto de em muitas situações os dados observados serem reduzidos ou escassos, não permitindo ter a perceção de cenários extremos ocorridos nos períodos onde se verificaram falhas de dados. Exemplo disto é o facto de a sobrelevação de maré máxima para Leixões é de 1.05 metros, porém temos de ser críticos ao ponto de perceber e acreditar que sobrelevações anteriores à série usada neste trabalho poderão ter ocorrido. Deste modo, aconselha-se a monitorização constante e continua das alturas horárias da maré, para que em futuros trabalhos existam mais observações para conclusões mais precisas.

Neste tipo de estudos, de sobrelevações de maré meteorológica, deve existir uma continuidade cronológica. É recomendado que este tipo de estudo seja realizado, no máximo, de 10 em 10 anos, para que os resultados obtidos sejam atualizados com alguma frequência, permitindo prevenir os portos nacionais de danos maiores causados por estas sobrelevações.

É de elevada importância realizar este estudo para todos os portos nacionais, continentais e insulares. O estudo realizado para dois portos teve em conta a tendência da subida do nível médio no mar a nível mundial, porém, sugere-se realizar este trabalho tendo em conta a subida do nível médio do mar local, assim obtendo-se previsões mais exatas para o porto em estudo.

# Bibliografia

- Almeida, A. (2008). *Quantificação do efeito das Marés Meteorológicas na costa oeste portuguesa* (tese de mestrado). Faculdade de Engenharia Universidade do Porto.
- Andri et mult. al., S. (2021). *DescTools: Tools for Descriptive Statistics* [R package version 0.99.44].
- Antunes, C., Rocha, C. & Catita, C. (2019). Coastal flood assessment due to sea level rise and extreme storm events: A case study of the atlantic coast of Portugal's mainland. *Geosciences*, 9(5), 239.
- APP. (2022). Porto de Leixões.
- Araújo, M. (2013). Lei da gravitação universal. *Revista de Ciência Elementar*, 1(1).
- Balkema, A. A. & de Haan, L. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 2(5), 792–804.
- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications* [Pagination: 522]. Wiley.
- Beirlant, J., Teugels, J. L. & Vynckier, P. (1996). *Practical analysis of extreme values*. Leuven university press.
- Cacho, J. L., Marques, L. & Nascimento, Á. (2020). Customer-Oriented Global Supply Chains: Port Logistics in the Era of Globalization and Digitization. *Anthropological Approaches to Understanding Consumption Patterns and Consumer Behavior* (pp. 82–103). IGI Global.
- Carinhas, D. (2014). *Controlo estatístico de qualidade em ensaios laboratoriais: caso particular de nutrientes em águas salinas* (tese de mestrado). Universidade de Évora.
- Castollo, E., Hadi, A., Balakrishnan, N. & Sarabia, J. (2004). Extreme Value and Related Models with Application in Engineering and Science.
- Chai, T. & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific Model Development Discussions*, 7(1), 1525–1534.
- Chan, K.-S. & Ripley, B. (2020). *TSA: Time Series Analysis* [R package version 1.3].
- Coles, S., Bawa, J., (Berlin)., S.-V., Trenner, L. & Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

- Davison, A. C. & Smith, R. L. (1990). Models for Exceedances over High Thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(3), 393–442.
- De Haan, L., Ferreira, A. & Ferreira, A. (2006). *Extreme value theory: an introduction* (Vol. 21). Springer.
- Doodson, A. & Warburg, H. (1946). *Admiralty - Manual of Tides*. Hydrographic Department, Admiralty, London, U.K.
- Embrechts, P., Klüppelberg, C. & Mikosch, T. (1998). *Modelling extremal events: for insurance and finance* (Vol. 33). Springer Verlag.
- Emiliano, P. C., Veiga, E. P., Vivanco, M. & Menezes, F. S. (2010). Critérios de informação de akaike versus bayesiano: análise comparativa. *19<sup>o</sup> Simpósio Nacional de Probabilidade e Estatística*.
- Fisher, R. A. & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180–190.
- Fisher, R. A. & Bischel, W. (1973). The role of linear dispersion in plane-wave self-phase modulation. *Applied Physics Letters*, 23(12), 661–663. <https://doi.org/10.1063/1.1654782>
- Fortunato, A., Li, K., Bertin, X. & Rodrigues, M. (2014). Determination of extreme sea levels along the Portuguese coast. *Actas das*, 3, 151–154.
- Fréchet, M. (-1. (1928). Sur la loi de probabilité de l'écart maximum (P. T. Matematyczne, Ed.).
- Gilleland, E. & Katz, R. W. (2016). extRemes 2.0: An Extreme Value Analysis Package in R. *Journal of Statistical Software*, 72(8), 1–39. <https://doi.org/10.18637/jss.v072.i08>
- Gnedenko. (1948). On a local limit theorem of the theory of probability. *Uspekhi Matematicheskikh Nauk*, 3(3), 187–194.
- Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44(3), 423–453.
- Godinho, J. R. P. M. (2011). *Variação temporal das principais constituintes harmónicas da maré de Cascais* (tese de doutoramento).
- Gomes, M. I., Alves, M. I. F. & Neves, C. (2013). *Análise de Valores Extremos: Uma Introdução*. Sociedade Portuguesa de Estatística.
- Gomes, M. I. (2019). *Estatística e Sociedade: Controlo da Qualidade e Extremos*.
- Gong, S. (2012). *Estimation of hot and cold spells with extreme value theory* (tese de mestrado). Project Report 19, Uppsala University.

- Gringorten, I. I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research (1896-1977)*, 68(3), 813–814.
- Grolemund, G. & Wickham, H. (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25.
- Gross, J., Ligges, U. & Ligges, M. U. (2015). Package ‘nortest’. *Five omnibus tests for testing the composite hypothesis of normality*.
- Guerreiro, R. I. R. (2015). *Delimitação de áreas inundáveis – Estudo da ação combinada da maré e do escoamento superficial* (tese de mestrado). Universidade Évora. Escola de Ciências e Tecnologia - Departamento de Engenharia Rural.
- Gumbel, E. (1935). Les valeurs extrêmes des distributions statistiques. *Annales de l’institut Henri Poincaré*, 5(2), 115–158.
- Hamner, B. & Frasco, M. (2018). *Metrics: Evaluation Metrics for Machine Learning* [R package version 0.1.4].
- Hidrográfico, I. (1984). *Apontamentos sobre marés e correntes de maré* (Vol. 1,2,3). Divisão de maré e correntes de maré, Ministério da Defesa nacional - Marinha.
- Hidrográfico, I. (2021). *Tabela de marés 2022*. Instituto Hidrográfico.
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O’Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E. & Yasmeeen, F. (2022). *forecast: Forecasting functions for time series and linear models* [R package version 8.16].
- IPMA. (2018). DIRETIVA-QUADRO “ESTRATÉGIA MARINHA” - Relatório do segundo ciclo de avaliação do bom estado ambiental das águas marinhas nas subdivisões do Continente e da Plataforma Continental Estendida.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348), 158–171.
- Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions*. PUBLISHED BY IMPERIAL COLLEGE PRESS; DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO. <https://doi.org/10.1142/p191>
- Marcos, M., Tsimplis, M. N. & Shaw, A. G. P. (2009). Sea level extremes in southern Europe. *Journal of Geophysical Research: Oceans*, 114(C1). <https://doi.org/https://doi.org/10.1029/2008JC004912>
- Markovich, N. (2007). *Nonparametric analysis of univariate heavy-tailed data: research and practice*. John Wiley & Sons.
- Maroco, J. (2007). *Análise estatística com utilização do SPSS*. Lisboa, Portugal: Edições Silabo.

- McInnes, K. L., Macadam, I., Hubbert, G. D. & O'Grady, J. G. (2009). A modelling approach for estimating the frequency of sea level extremes and the impact of climate change in southeast Australia. *Natural Hazards*, 51(1), 115–137.
- Melo, R. (2010). *Análise do transporte de Ekman entre Cabo Frio-RJ e Cabo de Santa Marta-SC, através de dados de vento do satélite Quikscat* (tese de mestrado). Universidade Federal do Paraná.
- Morais, M. (2020). *Probabilidade e Estatística. Teoria, Exemplos e Exercícios*. IST Press.
- Moreno, L. & Morcillo, A. (2019). *Distribuição Normal*.
- Morettin, P. A. & Toloi, C. M. (2004). *Análise de Séries Temporais*. Edgard blcher LTDA.
- Murteira, B. & Antunes, M. (2012). *Probabilidades e Estatística”, Volume II*. Livraria Escolar Editora.
- Murteira, B. (1990). *Probabilidades e estatística*. Mcgraw-Hill, Lisboa,
- Neves, P. (2021). *A teoria de valores extremos na quantificação de precipitação elevada* (tese de mestrado). Universidade de Aveiro.
- Neyman, J. & Pearson, E. S. (1933). On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231, 289–337.
- NOAA. (2017). Tracking sea level rise ... and fall.
- Oliveira. (1997). *Statistical analysis of extremes*. Pendor.
- Oliveira. (2009). *Tendência, sazonalidade e distribuição geográfica das fraturas osteoporóticas do fémur proximal em Portugal Continental* (tese de mestrado). Universidade do Porto.
- Pestana & Velosa. (2010). "Introdução à Probabilidade e à Estatística", 4<sup>a</sup> ed. Revista.
- Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3(1), 119–131.
- Pishro-Nik, H. (2014). *Introduction to probability, statistics, and random processes*. Kappa Research LLC.
- Pugh, D. (1987). *Tides, Surges and Mean Sea Level*. Wiley.
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Reis, J. C. M. L. (2008). *Caracterização da maré no Estuário do Minho* (tese de mestrado). Universidade Évora. UNIVERSIDADE TÉCNICA DE LISBOA.

- Reiss, R. & Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhauser Verlag, Berlin.
- Requeijo, J. & Pereira, Z. (2012). *Qualidade: Planeamento e Controlo Estatístico de Processos (2ª)*. Fundação da Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.
- Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Verlag.
- Ribatet, M. & Dutang, C. (2022). *POT: Generalized Pareto Distribution and Peaks Over Threshold* [R package version 1.1-10].
- Rocha, C. S. (2016). *Estudo e análise da vulnerabilidade costeira face a cenários de subida do nível do mar e eventos extremos devido ao efeito das alterações climáticas* (tese de doutoramento).
- Rodrigues, P., Carinhas, D. & Lemos, C. (2022). Teoria de valores extremos aplicada à caracterização da sobrelevação meteorológica no porto de Leixões. *7.ª Jornadas de Engenharia Hidrográfica, 2.ª Jornadas Luso-Espanhola de Hidrografia*, 157–160.
- Rosário, P. (2013). *Análise de valores extremos para níveis pluviométricos em Barcelos* (tese de mestrado). Universidade de Lisboa.
- Ryan, J. A. & Ulrich, J. M. (2020). *xts: eXtensible Time Series* [R package version 0.12.1].
- Seier, E. (2002). Comparison of Tests for Univariate Normality. *Proc. Interstat.*
- Shapiro, S. S. & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591–611.
- Sintra, A. S. (2017). *Métodos para Estimação dos Parâmetros da Distribuição de Pareto Generalizada: novas contribuições* (tese de doutoramento).
- Sousa, Á. (2019). Os gráficos no apoio à tomada de decisões e à disseminação de informações estatísticas. *Correio dos Açores*, 14–14.
- Torman, V. B. L., Coster, R. & Riboldi, J. (2012). Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. *Clinical amp; amp; Biomedical Research*, 32(2).
- Tran, H. (2020). *Hydrodynamic modelling of a wave-dominated tidal inlet system: Port Phillip Bay, Australia* (tese de doutoramento).
- Trujillo, A. P. & Thurman, H. V. (2010). *Essentials of oceanography*. Prentice Hall.
- Webb, P. (2017). *Introduction to Oceanography*. Online OER textbook: <https://webboceanography.pressbooks.com>.

- Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of applied mechanics*.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H. & Bryan, J. (2022). *readxl: Read Excel Files* [<https://readxl.tidyverse.org>, <https://github.com/tidyverse/readxl>].
- Wickham, H., François, R., Henry, L. & Müller, K. (2022). *dplyr: A Grammar of Data Manipulation* [<https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>].
- Xavier, J. M. N. (2016). *Análise e previsão de séries temporais com modelos ARIMA e análise espectral singular* (tese de mestrado). Universidade Aberta. Mestrado em Bioestatística e Biometria.
- Zeileis, A. & Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, 14(6), 1–27. <https://doi.org/10.18637/jss.v014.i06>

# Apêndice A - Script do Software R para a Análise Preliminar das Sobrelevações

Neste SCRIPT para efeitos de compatibilidade com o programa Latex procedeu-se à substituição do Símbolo "\$" pelo símbolo "@".

```
1 #Análise Preliminar das Sobrelevações no Porto de Leixões/Sines
2
3 #Bibliotecas necessárias
4 #-----
5 library(readxl)
6 library(lubridate)
7 library(DescTools)
8 library(zoo)
9 library(xts)
10 library(extRemes)
11 library(forecast)
12 library(TSA)
13 library(ggplot2)
14 library(dplyr)
15 library(POT)
16 library(nortest)
17 library(Metrics)
18 rm(list = ls())
19
20 #-----
21 power_spec = function(y,samp.freq, ...){
22   N <- length(y)
23   fk <- fft(y)
24   fk <- fk[2:length(fk)/2+1]
25   fk <- 2*fk[seq(1, length(fk), by = 2)]/N
26   freq <- (1:(length(fk)))* samp.freq/(2*length(fk))
27   data.frame(amplitude = Mod(fk), freq = freq)
28 }
29 #-----
30 "%nin%" <- Negate("%in%")
31 #-----
32
33 setwd("C:/Users/Lenovo/OneDrive - Marinha/Ambiente de Trabalho/
34   DADOS")
35 source("./SCRIPTS/findGaps.R")
36 source("./SCRIPTS/findObserved.R")
37
38 # Ler o ficheiro com as alturas horárias, previsões e desvios
```

```

38 #-----
39 df <- read_excel("/Users/Lenovo/OneDrive - Marinha/Ambiente de
    Trabalho/DADOS/leixoes_1995_2020.xlsx",sheet = 3)
40
41 # Criar séries temporais das observações, previsões e dos desvios
42 #-----
43 tmp <- as.POSIXlt(df@hora)
44 hour <- hour(tmp)
45 min <- minute(tmp)
46 sec <- second(tmp)
47 DataHora <- as.POSIXct(paste(df@data,paste(hour,min,sec,sep = ":"),
    sep = " "),tz = "GMT")
48 Obs <- xts(df@obs, order.by = DataHora)
49 Prev <- xts(df@prev, order.by = DataHora)
50 Desvio <- xts(df@desvio, order.by = DataHora)
51 rm(tmp, hour, min, sec)
52
53 #Gráfico da série dos Desvios entre a maré observada e a meteoroló
    gica
54 #-----
55 dev.new(width = 1800, height = 1400)
56 plot(index(Desvio),coredata(Desvio),
57       xlab = "TEMPO (ANOS)",
58       ylab = "DESVIO (METROS)",
59       type = "l",
60       tck = 0.0125)
61
62 #divisão da Série dos Desvios por anos
63 #-----
64 Desvio_Ano <- split(Desvio,f = "years")
65 lacunasAnuais <- integer()
66 for (Anos in 1:length(Desvio_Ano)) {
67   if (all(is.na(coredata(Desvio_Ano[[Anos]]))))
68     lacunasAnuais <- c(lacunasAnuais,Anos)
69 }
70 Desvio_Ano <- Desvio_Ano[-lacunasAnuais]
71 Desvio_Ano[[2]] <- NULL # Ano de 1996 com muito poucas observações
72 Desvio_Ano[[1]] <- NULL #só foi realizado após descrever a série
    toda
73 str(Desvio_Ano)
74
75 # Gráficos dos desvios por anos
76 #-----
77 dev.new(width = 1500,height = 500)
78 par(mfrow = c(5,2))

```

```

79 plot(Desvio_Ano[[1]], main = "1995")
80 plot(c(Desvio_Ano[[2]]), main = "2002")
81 plot(c(Desvio_Ano[[3]]), main = "2003")
82 plot(c(Desvio_Ano[[4]]), main = "2004")
83 plot(c(Desvio_Ano[[5]]), main = "2005")
84 plot(c(Desvio_Ano[[6]]), main = "2006")
85 plot(c(Desvio_Ano[[7]]), main = "2007")
86 plot(c(Desvio_Ano[[8]]), main = "2008")
87 plot(c(Desvio_Ano[[9]]), main = "2009")
88 plot(c(Desvio_Ano[[10]]), main = "2010")
89
90 # Diagramas de Extremos e quartis dos desvios anuais
91 #-----
92 anos <- unlist(lapply(Desvio_Ano,function(x) year(first(index(x)))
93 )
94
95 dev.new(width = 1400, height = 670)
96 par(mfrow = c(2,10), mar = par()@mar + c(0,-1,-3,-1), mgp = c
97 (1,0.5,0))
98 for (i in 1:10) {
99   boxplot(Desvio_Ano[[i]], xlab = anos[i],las = 1, tck = 0.01,
100           ylim = range(pretty(range(coredata(Desvio_Ano),na.rm =
101 TRUE))))
102 }
103 for (i in 11:20) {
104   boxplot(Desvio_Ano[[i]], xlab = anos[i],las = 1, tck = 0.01,
105           ylim = range(pretty(range(coredata(Desvio_Ano),na.rm =
106 TRUE))))
107 }
108
109 # Valores numéricos
110 #-----
111 minimoAnual <- unlist(lapply(Desvio_Ano,FUN = function(x) min(x,na.
112 rm = TRUE)))
113 quantil_1 <- unlist(lapply(Desvio_Ano,FUN = function(x) quantile(x,
114 na.rm = TRUE)[2]))
115 mediaAnual <- unlist(lapply(Desvio_Ano,FUN = function(x) mean(x,na.
116 rm = TRUE)))
117 medianaAnual <- unlist(lapply(Desvio_Ano,FUN = function(x) median(x
118 ,na.rm = TRUE)))
119 quantil_3 <- unlist(lapply(Desvio_Ano,FUN = function(x) quantile(x,
120 na.rm = TRUE)[4]))
121 maximoAnual <- unlist(lapply(Desvio_Ano,FUN = function(x) max(x,na.
122 rm = TRUE)))
123

```

```

114 # Histograma dos desvios
115 #-----
116 dev.new(width = 1750, height = 1500)
117 n <- length(which(!is.na(Desvio)))
118 mu <- mean(Desvio, na.rm = TRUE)
119 sig <- sd(Desvio, na.rm = TRUE)
120
121 histDesvio <- hist(Desvio, breaks = max(2, trunc(min(n/10, 25*log(n
122     ,10))/2)),
123     freq = FALSE, col = "lightblue",
124     ylab = "Densidade",
125     main = "Leixões 1995-2020: histograma dos
126     desvios da altura da mar?")
127 x <- seq(min(histDesvio@breaks), max(histDesvio@breaks), length.out
128     = 2*length(histDesvio@breaks))
129 y <- 1/(sig*sqrt(2*pi))* exp(-0.5*((x-mu)/(sig))^2)
130 lines(x,y,col = "red",lwd = 2)
131
132 # Ler o ficheiro com as alturas horárias, previsões e desvios desde
133 # 2002 a 2020
134 #-----
135 df <- read_excel("/Users/Lenovo/OneDrive - Marinha/Ambiente de
136     Trabalho/DADOS/leixoes_2002_2020.xlsx", sheet = 3)
137
138 #medidas descritivas dos desvios
139 #-----
140 summary(Desvio)
141
142 #substituição dos valores omissos pela mediana
143 #-----
144 Desvio[which(is.na(Desvio))] <- 0.18 # Substituir lacunas pela
145     mediana
146
147 #Gráfico da série dos Desvios entre a maré observada e a meteoroló
148 # gica
149 #-----
150 dev.new(width = 1800, height = 1400)
151 plot(index(Desvio), coredata(Desvio),
152     xlab = "TEMPO (ANOS)",
153     ylab = "DESVIO (METROS)",
154     type = "l",
155     tck = 0.0125)

```

## Apêndice B - Script do Software R para a Análise Preliminar das Sobrelevações Máximas Anuais

Neste SCRIPT para efeitos de compatibilidade com o programa Latex procedeu-se à substituição do Símbolo "\$" pelo símbolo "@".

```
1
2 #Análise das Sobrelevações máximas anuais
3
4 #Bibliotecas necessárias
5 #-----
6 library(readxl)
7 library(lubridate)
8 library(DescTools)
9 library(zoo)
10 library(xts)
11 library(extRemes)
12 library(forecast)
13 library(TSA)
14 library(ggplot2)
15 library(dplyr)
16 library(POT)
17 library(nortest)
18 library(Metrics)
19 rm(list = ls())
20
21 #-----
22 power_spec = function(y,samp.freq, ...){
23   N <- length(y)
24   fk <- fft(y)
25   fk <- fk[2:length(fk)/2+1]
26   fk <- 2*fk[seq(1, length(fk), by = 2)]/N
27   freq <- (1:(length(fk)))* samp.freq/(2*length(fk))
28   data.frame(amplitude = Mod(fk), freq = freq)
29 }
30 #-----
31 "%nin%" <- Negate("%in%")
32 #-----
33
34 setwd("C:/Users/Lenovo/OneDrive - Marinha/Ambiente de Trabalho/
35   DADOS")
36 source("../SCRIPTS/findGaps.R")
37 source("../SCRIPTS/findObserved.R")
```

```
38 # Ler o ficheiro com as alturas horárias, previsões e desvios
39 # -----
40 df <- read_excel("/Users/Lenovo/OneDrive - Marinha/Ambiente de
    Trabalho/DADOS/leixoes_1995_2020.xlsx",sheet = 3)
41
42 # Criar séries temporais das observações, previsões e dos desvios
43 # -----
44 tmp <- as.POSIXlt(df@hora)
45 hour <- hour(tmp)
46 min <- minute(tmp)
47 sec <- second(tmp)
48 DataHora <- as.POSIXct(paste(df@data,paste(hour,min,sec,sep = ":"),
    sep = " "),tz = "GMT")
49 Obs <- xts(df@obs, order.by = DataHora)
50 Prev <- xts(df@prev, order.by = DataHora)
51 Desvio <- xts(df@desvio, order.by = DataHora)
52 rm(tmp,hour,min,sec)
53
54 # Ler o ficheiro com as alturas horárias, previsões e desvios desde
    2002 a 2020
55 # -----
56 df <- read_excel("/Users/Lenovo/OneDrive - Marinha/Ambiente de
    Trabalho/DADOS/leixoes_2002_2020.xlsx",sheet = 3)
57
58 #medidas descritivas dos desvios
59 # -----
60 summary(Desvio)
61
62
63 #substituição dos valores omissos pela mediana
64 # -----
65 Desvio[which(is.na(Desvio))] <- 0.18 # Substituir lacunas pela
    mediana de cada Porto
66
67
68 #Gráfico da série dos Desvios entre a maré observada e a meteoroló
    gica
69 # -----
70 dev.new(width = 1800, height = 1400)
71 plot(index(Desvio),coredata(Desvio),
72     xlab = "TEMPO (ANOS)",
73     ylab = "DESVIO (METROS)",
74     type = "l",
75     tck = 0.0125)
76
```

```
77 #Chamar os Desvios Máximos Anuais
78 # -----
79 DesvioMaximoAnual <- period.apply(Desvio, INDEX=endpoints(Desvio, on=
    "years"), max)
80
81 #Gráfico dos Desvios Máximos Anuais
82 # -----
83 grid(plot(index(DesvioMaximoAnual), coredata(DesvioMaximoAnual),
84         type = "l", main = "Leixões 2002 a 2022: Máximos Anuais",
85         ylab = "DESVIO (m)",
86         xlab = "ANOS" ))
87
88 #Histograma dos Desvios Máximos Anuais
89 # -----
90 dev.new(width = 1750, height = 1500)
91 n <- length(which(!is.na(DesvioMaximoAnual)))
92 mu <- mean(DesvioMaximoAnual, na.rm = TRUE)
93 sig <- sd(DesvioMaximoAnual, na.rm = TRUE)
94
95 histDesvio <- hist(DesvioMaximoAnual, breaks = max(2, trunc(min(n/
    10, 25*log(n, 10))/2)),
96                 freq = FALSE, col = "lightblue",
97                 ylab = "Densidade",
98                 main = "Leixões 1995-2020: histograma dos
    desvios da altura da mar?")
99 x <- seq(min(histDesvio@breaks), max(histDesvio@breaks), length.out
    = 2*length(histDesvio@breaks))
100 y <- 1/(sig*sqrt(2*pi))* exp(-0.5*((x-mu)/(sig))^2)
101 lines(x, y, col = "red", lwd = 2)
102
103 #Teste da normalidade: Shapiro-Wilk
104 # -----
105 st <- shapiro.test(coredata(DesvioMaximoAnual[1:19])) #mudar
    consoante o Porto (L-19; S-24)
106 st
107
108 #Medidas descritivas
109 # -----
110 summary(DesvioMaximoAnual)
111 desviopadrão <- sd(DesvioMaximoAnual)
112
113 #Testar a independência do dados: Gráficos ACF e PACF
114 # -----
115 dev.new(width = 1500, height = 1000)
116 par(mfrow = c(2, 1))
```

```
117 acf(coredata(DesvioMaximoAnual), lag.max = 168, na.action = na.pass
    )
118 pacf(coredata(DesvioMaximoAnual), lag.max = 168, na.action = na.
    pass)
119
120 #Diagrama de Extremos e Quartis
121 # -----
122 dev.new(width = 1400, height = 670)
123 boxplot(DesvioMaximoAnual, las = 1, tck = 0.01,
124         main = "Diagrama de Extremos e quartis Max-Anuais",
125         ylab = "Desvio (m)",
126         na.rm = TRUE)
127
128 # -----
129 # ----- GEV dos máximos anuais -----
130 # -----
131
132 #Ajuste      distribuição GEV
133 # -----
134 fit <- fevd(DesvioMaximoAnual, type = "GEV",
135            span = 19, units = "m", time.units = "year") #mudar
136            consoante o Porto
137 ci(fit, type = "parameter")
138
139 #Gráfico de probabilidades GEV
140 # -----
141 dev.new(width = 1000, height = 1000)
142 pc<-plot(fit, type = "prob",
143         main = "Gráfico de probabilidades GEV: 2002 a 2020")
144
145 #Gráfico de Densidade GEV
146 # -----
147 dev.new(width = 1000, height = 1000)
148 plot(fit, type = "density",
149      ylab = "Densidade",
150      xlab = "Desvio"
151      main = "Comparação com o modelo teórico de GEV ")
152
153 #Ajuste      distribuição GEV com localização variável no tempo
154 # -----
155 Y <- coredata(DesvioMaximoAnual)
156 X <- seq(0, length(Y)-1, by = 1)
157 Data <- as.data.frame(cbind(X, Y))
158 names(Data) <- c("X", "Y")
```

```
159 fitVarLocation_GEV <- fevd(Y,data = Data, type = "GEV",
160                             location.fun ~ I(1 + (0.3eX) + cos(2*pi*
161                               X) + sin(2*pi*X)),
162                             span = 19, units = "m", time.units = "
163                               year") #mudar consoante o Porto
164 ci(fitVarLocation_GEV,type = "parameter")
165
166 #Ajuste      distribuição GEV com escala variável no tempo
167 # -----
168 fitVarScale_GEV <- fevd(Y,data = Data, type = "GEV",
169                             scale.fun ~ I(1 + (0.3*X) + cos(2*pi*X) +
170                               sin(2*pi*X)),
171                             span = 19, units = "m", time.units = "year")
172 #mudar consoante o Porto
173
174 #Gráficos de probabilidade GEV para os ajuste com os parâmetros
175 # variáveis
176 # -----
177 dev.new(width = 1000, height = 1000)
178 par(mfrow = c(1,2))
179 pl<-plot(fitVarLocation_GEV,type = "prob",
180          main = "Gráfico A")
181 dev.new(width = 1000, height = 1000)
182 pe<-plot(fitVarScale_GEV,type = "prob",
183          main = "Gráfico B")
184
185 # -----
186 #----- GEV dos máximos anuais -----
187 # -----
188
189 #Ajuste      distribuição Gumbel
190 # -----
191 fit_GUMBEL <- fevd(DesvioMaximoAnual, type = "Gumbel",
192                             span = 19, units = "m", time.units = "year") #
193 #mudar consoante o Porto
194 ci(fit_GUMBEL,type = "parameter")
195
196 #Gráficos de probabilidade e densidade
197 # -----
198 dev.new(width = 1000, height = 1000)
199 par(mfrow = c(1,2))
200 g3<-plot(fit_GUMBEL,type = "prob",
201          main = "Gráfico A ")
202 plot(fit_GUMBEL,type = "density",
203      ylab = "Densidade",
```

```

198   main = "Gráfico B")
199
200 #Ajuste      distribuição Gumbel com localização variável no tempo
201 # -----
202 Y <- coredata(DesvioMaximoAnual)
203 X <- seq(0,length(Y)-1,by = 1)
204 Data <- as.data.frame(cbind(X,Y))
205 names(Data) <- c("X","Y")
206
207 fitVarLocation_GUMBEL <- fevd(Y,data = Data, type = "Gumbel",
208                               location.fun ~ I(1 + (0.3*X) + cos(2*
209                               pi*X) + sin(2*pi*X)),
210                               span = 19, units = "m", time.units = "
211                               years")#mudar consoante o Porto
212
213 #Ajuste      distribuição Gumbel com escala variável no tempo
214 # -----
215 fitVarScale_GUMBEL <- fevd(Y,data = Data, type = "Gumbel",
216                               scale.fun ~ I(1 + (0.4e-3*X) + cos(2*pi*
217                               X) + sin(2*pi*X)),
218                               span = 19, units = "m", time.units = "
219                               year")#mudar consoante o Porto
220
221 #Gráficos probabilidade e densidade Gumbel para os ajuste com os
222 # parâmetros variáveis
223 # -----
224 ev.new(width = 1000, height = 1000)
225 par(mfrow = c(2,2))
226 g1<-plot(fitVarLocation_GUMBEL,type = "prob",
227          main = "Gráfico A")
228 g2<-plot(fitVarScale_GUMBEL,type = "prob",
229          main = "Gráfico C")
230 plot(fitVarLocation_GUMBEL,type = "density",
231       ylab = "Densidade",
232       main = "Gráfico B")
233 plot(fitVarScale_GUMBEL,type = "density",
234       ylab = "Densidade",
235       main = "Gráfico D")
236
237 #Teste Likelihood-ratio
238 # -----
239 lr.test(fit,fitVarLocation_GEV)
240 lr.test(fit,fitVarScale_GEV)
241 lr.test(fit_GUMBEL,fitVarLocation_GUMBEL)
242 lr.test(fit_GUMBEL,fitVarScale_GUMBEL)

```

```
238
239 #Indicadores AIC e BIC
240 # -----
241 summary(fit)
242 summary(fitVarLocation_GEV)
243 summary(fitVarScale_GEV)
244
245 summary(fit_GUMBEL)
246 summary(fitVarLocation_GUMBEL)
247 summary(fitVarScale_GUMBEL)
248
249 #Teste RMSE
250 # -----
251 rmse(pl@model,pl@empirical)
252 rmse(pe@model,pe@empirical)
253 rmse(pc@model,pc@empirical)
254
255 rmse(g3@empirical,g3@model)
256 rmse(g1@empirical,g1@model)
257 rmse(g2@empirical,g2@model)
258
259 #Gráfico de N veis de Retorno Gumbel
260 # -----
261 dev.new(width = 1000, height = 1000)
262 plot.fevd(fitVarLocation_GUMBEL,
263           type = "rl",
264           col="red",
265           main = "N vel de Retorno" ,
266           Xlab = "Per odo de Retorno (Anos) ",
267           Ylab = "Metros")
268
269 #Valores do N vel de Retorno para os Per dos de Retorno 5, 20, 50
270 # e 80 anos
271 # -----
272 return.level(fitVarLocation_GUMBEL, return.period = c(5,20,50,80))
273 # -----
274 #----- GP dos máximos anuais -----
275 # -----
276
277 #Escolha de um Threshold: gráficos
278 # -----
279 dev.new(width = 1750,height = 700)
280 threshrange.plot(DesvioMaximoAnual,r=c(0.25,0.8),nint = 40)#mudar
281 consoante o Porto
```

```
281
282 dev.new(width = 1750,height = 700)
283 mrlplot(DesvioMaximoAnual, u.range = c(0.1, quantile(
      DesvioMaximoAnual, probs = 0.995)),
284         col = c("black", "red", "black"), nt = 200)
285
286 # Cálculo dos parâmetros da distribuição das excedências (GP)
287 # -----
288 fitgp1 <- fevd(DesvioMaximoAnual, threshold = 0.5, type = "GP",#
      mudar Threshold consoante o porto
289              span = 19, units = "m", time.units = "years")#mudar
      consoante o Porto
290 ci(fitgp1,type = "parameter")
291
292 #Gráfico de probabilidades para a distribuição GP
293 # -----
294 dev.new(width = 1000, height = 1000)
295 GP1<- plot(fitgp1,type = "prob", main = "Gráfico de probabilidades
      (GP)")
296
297 # Cálculo dos parâmetros da distribuição das excedências (GP) com a
      o parâmetro escala variável
298 # -----
299 Y <- coredata(DesvioMaximoAnual)
300 X <- seq(0,length(Y)-1,by = 1)
301 Data <- data.frame(cbind(X,Y))
302 names(Data) <- c("X","Y")
303
304
305 fitScale <- fevd(Y, data = Data, threshold = 0.5,#mudar consoante o
      Porto
306              scale.fun = ~ I(1 + (0.3eX) + cos(2*pi*X) + sin(2*
      pi*X)),
307              type = "GP", span = 19, units = "m", time.units =
      "years")#mudar consoante o Porto
308
309 #Gráfico de probabilidade GP com escala variável
310 # -----
311 dev.new(width = 1000, height = 1000)
312 Gp2 <- plot(fitScale,type = "probprob", main = "Gráfico de
      probabilidades com escala variável (GP)")
313
314 #Indicadores AIC e BIC
315 # -----
316 summary(fitgp1)
```

```
317 summary(fitScale)
318
319 #Teste RMSE
320 # -----
321 rmse(GP1@empirical,GP1@model)
322 rmse(Gp2@empirical,Gp2@model)
323
324 #Gráfico N veis de Retorno
325 # -----
326 dev.new(width = 1000, height = 1000)
327 plot(fitScale,type = "rl", ylim = c(0.2, 1), main = "")
328
329 #Valores do N vel de Retorno para os Per dos de Retorno 5, 20, 50
    e 80 anos
330 # -----
331 return.level(fitScale, return.period = c(5,20,50,80))
```



## Apêndice C - Script do Software R para a Análise das Sobrelevações Máximas Mensais

Neste SCRIPT para efeitos de compatibilidade com o programa Latex procedeu-se à substituição do Símbolo "\$" pelo símbolo "@".

```
1
2 #Análise das Sobrelevações máximas Mensais
3
4 #Bibliotecas necessárias
5 #-----
6 library(readxl)
7 library(lubridate)
8 library(DescTools)
9 library(zoo)
10 library(xts)
11 library(extRemes)
12 library(forecast)
13 library(TSA)
14 library(ggplot2)
15 library(dplyr)
16 library(POT)
17 library(nortest)
18 library(Metrics)
19 rm(list = ls())
20
21 #-----
22 power_spec = function(y,samp.freq, ...){
23   N <- length(y)
24   fk <- fft(y)
25   fk <- fk[2:length(fk)/2+1]
26   fk <- 2*fk[seq(1, length(fk), by = 2)]/N
27   freq <- (1:(length(fk)))* samp.freq/(2*length(fk))
28   data.frame(amplitude = Mod(fk), freq = freq)
29 }
30 #-----
31 "%nin%" <- Negate("%in%")
32 #-----
33
34 setwd("C:/Users/Lenovo/OneDrive - Marinha/Ambiente de Trabalho/
35   DADOS")
36 source("../SCRIPTS/findGaps.R")
37 source("../SCRIPTS/findObserved.R")
```

```
38 # Ler o ficheiro com as alturas horárias, previsões e desvios
39 # -----
40 df <- read_excel("/Users/Lenovo/OneDrive - Marinha/Ambiente de
    Trabalho/DADOS/leixoes_1995_2020.xlsx",sheet = 3)
41
42 # Criar séries temporais das observações, previsões e dos desvios
43 # -----
44 tmp <- as.POSIXlt(df@hora)
45 hour <- hour(tmp)
46 min <- minute(tmp)
47 sec <- second(tmp)
48 DataHora <- as.POSIXct(paste(df@data,paste(hour,min,sec,sep = ":"),
    sep = " "),tz = "GMT")
49 Obs <- xts(df@obs, order.by = DataHora)
50 Prev <- xts(df@prev, order.by = DataHora)
51 Desvio <- xts(df@desvio, order.by = DataHora)
52 rm(tmp,hour,min,sec)
53
54 # Ler o ficheiro com as alturas horárias, previsões e desvios desde
    2002 a 2020
55 # -----
56 df <- read_excel("/Users/Lenovo/OneDrive - Marinha/Ambiente de
    Trabalho/DADOS/leixoes_2002_2020.xlsx",sheet = 3)
57
58 #medidas descritivas dos desvios
59 # -----
60 summary(Desvio)
61
62
63 #substituição dos valores omissos pela mediana
64 # -----
65 Desvio[which(is.na(Desvio))] <- 0.18 # Substituir lacunas pela
    mediana de cada Porto
66
67
68 #Gráfico da série dos Desvios entre a maré observada e a meteoroló
    gica
69 # -----
70 dev.new(width = 1800, height = 1400)
71 plot(index(Desvio),coredata(Desvio),
72     xlab = "TEMPO (ANOS)",
73     ylab = "DESVIO (METROS)",
74     type = "l",
75     tck = 0.0125)
76
```

```
77 #Chamar os Desvios Máximos Mensais
78 # -----
79 DesvioMaximoMensal <- period.apply(Desvio, INDEX=endpoints(Desvio, on
    = "years"), max)
80
81 #Gráfico dos Desvios Máximos Mensais
82 # -----
83 grid(plot(index(DesvioMaximoMensal), coredata(DesvioMaximoMensal),
84         type = "l", main = "Leixões 2002 a 2022: Máximos Mensais"
85         ,
86         ylab = "DESVIO (m)",
87         xlab = "ANOS" ))
88
89 #Histograma dos Desvios Máximos Mensais
90 # -----
91 dev.new(width = 1750, height = 1500)
92 n <- length(which(!is.na(DesvioMaximoMensal)))
93 mu <- mean(DesvioMaximoMensal, na.rm = TRUE)
94 sig <- sd(DesvioMaximoMensal, na.rm = TRUE)
95 histDesvio <- hist(DesvioMaximoMensal, breaks = max(2, trunc(min(n/
96     10, 25*log(n, 10))/2)),
97     freq = FALSE, col = "lightblue",
98     ylab = "Densidade",
99     main = "Leixões 1995-2020: histograma dos
100     desvios da altura da mar?")
101 x <- seq(min(histDesvio@breaks), max(histDesvio@breaks), length.out
102     = 2*length(histDesvio@breaks))
103 y <- 1/(sig*sqrt(2*pi))* exp(-0.5*((x-mu)/(sig))^2)
104 lines(x, y, col = "red", lwd = 2)
105
106 #Teste da normalidade: Shapiro-Wilk
107 # -----
108 st <- shapiro.test(coredata(DesvioMaximoMensal[1:19])) #mudar
109     consoante o Porto (L-19; S-24)
110
111 #Medidas descritivas
112 # -----
113 summary(DesvioMaximoMensal)
114 desviopadrão <- sd(DesvioMaximoMensal)
115
116 #Testar a indepência do dados: Gráficos ACF e PACF
117 # -----
118 dev.new(width = 1500, height = 1000)
```

```

116 par(mfrow = c(2,1))
117 acf(coredata(DesvioMaximoMensal), lag.max = 168, na.action = na.
    pass)
118 pacf(coredata(DesvioMaximoMensal), lag.max = 168, na.action = na.
    pass)
119
120 #Diagrama de Extremos e Quartis
121 # -----
122 dev.new(width = 1400, height = 670)
123 boxplot(DesvioMaximoMensal, las = 1, tck = 0.01,
124         main = "Diagrama de Extremos e quartis Max-Mensais",
125         ylab = "Desvio (m)",
126         na.rm = TRUE)
127
128 #-----
129 #----- GEV dos máximos Mensais -----
130 #-----
131
132 #Ajuste      distribuição GEV
133 # -----
134 fit <- fevd(DesvioMaximoMensal, type = "GEV",
135            span = 19, units = "m", time.units = "year") #mudar
136            consoante o Porto
137 ci(fit,type = "parameter")
138
139 #Gráfico de probabilidades GEV
140 # -----
141 dev.new(width = 1000, height = 1000)
142 pc<-plot(fit,type = "prob",
143         main = "Gráfico de probabilidades GEV: 2002 a 2020")
144
145 #Gráfico de Densidade GEV
146 # -----
147 dev.new(width = 1000, height = 1000)
148 plot(fit,type = "density",
149      ylab = "Densidade",
150      xlab = "Desvio"
151      main = "Comparação com o modelo teórico de GEV ")
152
153 #Ajuste      distribuição GEV com localização variável no tempo
154 # -----
155 Y <- coredata(DesvioMaximoMensal)
156 X <- seq(0,length(Y)-1,by = 1)
157 Data <- as.data.frame(cbind(X,Y))
158 names(Data) <- c("X","Y")

```

```
158
159 fitVarLocation_GEV <- fevd(Y,data = Data, type = "GEV",
160                             location.fun ~ I(1 + (0.4e-3*X) + cos(2*
161                             pi*X/12) + sin(2*pi*X/12)),
162                             span = 19, units = "m", time.units = "
163                             year") #mudar consoante o Porto
164 ci(fitVarLocation_GEV,type = "parameter")
165
166 #Ajuste      distribuição GEV com escala variável no tempo
167 # -----
168 fitVarScale_GEV <- fevd(Y,data = Data, type = "GEV",
169                             scale.fun ~ I(1 + (0.4e-3*X) + cos(2*pi*X/
170                             12) + sin(2*pi*X/12)),
171                             span = 19, units = "m", time.units = "year")
172 #mudar consoante o Porto
173
174 #Gráficos de probabilidade GEV para os ajuste com os parâmetros
175 # variáveis
176 # -----
177 dev.new(width = 1000, height = 1000)
178 par(mfrow = c(1,2))
179 pl<-plot(fitVarLocation_GEV,type = "prob",
180          main = "Gráfico A")
181 dev.new(width = 1000, height = 1000)
182 pe<-plot(fitVarScale_GEV,type = "prob",
183          main = "Gráfico B")
184
185 # -----
186 # ----- GEV dos máximos Mensais -----
187 # -----
188
189 #Ajuste      distribuição Gumbel
190 # -----
191 fit_GUMBEL <- fevd(DesvioMaximoMensal, type = "Gumbel",
192                             span = 19, units = "m", time.units = "year") #
193 #mudar consoante o Porto
194 ci(fit_GUMBEL,type = "parameter")
195
196 #Gráficos de probabilidade e densidade
197 # -----
198 dev.new(width = 1000, height = 1000)
199 par(mfrow = c(1,2))
200 g3<-plot(fit_GUMBEL,type = "prob",
201          main = "Gráfico A ")
202 plot(fit_GUMBEL,type = "density",
```

```

197     ylab = "Densidade",
198     main = "Gráfico B")
199
200 #Ajuste      distribuição Gumbel com localização variável no tempo
201 # -----
202 Y <- coredata(DesvioMaximoMensal)
203 X <- seq(0,length(Y)-1,by = 1)
204 Data <- as.data.frame(cbind(X,Y))
205 names(Data) <- c("X","Y")
206
207 fitVarLocation_GUMBEL <- fevd(Y,data = Data, type = "Gumbel",
208                               location.fun ~ I(1 + (0.4e-3*X) + cos
209                               (2*pi*X/12) + sin(2*pi*X/12)),
210                               span = 19, units = "m", time.units = "
211                               years")#mudar consoante o Porto
212
213 #Ajuste      distribuição Gumbel com escala variável no tempo
214 # -----
215 fitVarScale_GUMBEL <- fevd(Y,data = Data, type = "Gumbel",
216                               scale.fun ~ I(1 + (0.4e-3*X) + cos(2*pi*
217                               X/12) + sin(2*pi*X/12)),
218                               span = 19, units = "m", time.units = "
219                               year")#mudar consoante o Porto
220
221 #Gráficos probabilidade e densidade Gumbel para os ajuste com os
222 # parâmetros variáveis
223 # -----
224 ev.new(width = 1000, height = 1000)
225 par(mfrow = c(2,2))
226 g1<-plot(fitVarLocation_GUMBEL,type = "prob",
227          main = "Gráfico A")
228 g2<-plot(fitVarScale_GUMBEL,type = "prob",
229          main = "Gráfico C")
230 plot(fitVarLocation_GUMBEL,type = "density",
231      ylab = "Densidade",
232      main = "Gráfico B")
233 plot(fitVarScale_GUMBEL,type = "density",
234      ylab = "Densidade",
235      main = "Gráfico D")
236
237 #Teste Likelihood-ratio
238 # -----
239 lr.test(fit,fitVarLocation_GEV)
240 lr.test(fit,fitVarScale_GEV)
241 lr.test(fit_GUMBEL,fitVarLocation_GUMBEL)

```

```
237 lr.test(fit_GUMBEL,fitVarScale_GUMBEL)
238
239 #Indicadores AIC e BIC
240 # -----
241 summary(fit)
242 summary(fitVarLocation_GEV)
243 summary(fitVarScale_GEV)
244
245 summary(fit_GUMBEL)
246 summary(fitVarLocation_GUMBEL)
247 summary(fitVarScale_GUMBEL)
248
249 #Teste RMSE
250 # -----
251 rmse(pl@model,pl@empirical)
252 rmse(pe@model,pe@empirical)
253 rmse(pc@model,pc@empirical)
254
255 rmse(g3@empirical,g3@model)
256 rmse(g1@empirical,g1@model)
257 rmse(g2@empirical,g2@model)
258
259 #Gráfico de N veis de Retorno Gumbel
260 # -----
261 dev.new(width = 1000, height = 1000)
262 plot.fevd(fitVarLocation_GUMBEL,
263           type = "rl",
264           col="red",
265           main = "N vel de Retorno" ,
266           Xlab = "Per odo de Retorno (Anos) ",
267           Ylab = "Metros")
268
269 #Valores do N vel de Retorno para os Per dos de Retorno 5, 20, 50
270 # e 80 anos
271 # -----
272
273 #-----
274 #----- GP dos máximos Mensais -----
275 #-----
276
277 #Escolha de um Threshold: gráficos
278 # -----
279 dev.new(width = 1750,height = 700)
```

```
280 threshrange.plot(DesvioMaximoMensal,r=c(0.25,0.8),nint = 40)#mudar
    consoante o Porto
281
282 dev.new(width = 1750,height = 700)
283 mrlplot(DesvioMaximoMensal, u.range = c(0.1, quantile(
    DesvioMaximoMensal, probs = 0.995)),
284         col = c("black", "red", "black"), nt = 200)
285
286 # Cálculo dos parâmetros da distribuição das excedências (GP)
287 # -----
288 fitgp1 <- fevd(DesvioMaximoMensal, threshold = 0.5, type = "GP",#
    mudar Threshold consoante o porto
289             span = 19, units = "m", time.units = "years")#mudar
    consoante o Porto
290 ci(fitgp1,type = "parameter")
291
292 #Gráfico de probabilidades para a distribuição GP
293 # -----
294 dev.new(width = 1000, height = 1000)
295 GP1<- plot(fitgp1,type = "prob", main = "Gráfico de probabilidades
    (GP)")
296
297 # Cálculo dos parâmetros da distribuição das excedências (GP) com a
    o parâmetro escala variável
298 # -----
299 Y <- coredata(DesvioMaximoMensal)
300 X <- seq(0,length(Y)-1,by = 1)
301 Data <- data.frame(cbind(X,Y))
302 names(Data) <- c("X","Y")
303
304
305 fitScale <- fevd(Y, data = Data, threshold = 0.5,#mudar consoante o
    Porto
306             scale.fun = ~ I(1 + (0.4e-3*X) + cos(2*pi*X/12) +
    sin(2*pi*X/12)),
307             type = "GP", span = 19, units = "m", time.units =
    "years")#mudar consoante o Porto
308
309 #Gráfico de probabilidade GP com escala variável
310 # -----
311 dev.new(width = 1000, height = 1000)
312 Gp2 <- plot(fitScale,type = "probprob", main = "Gráfico de
    probabilidades com escala variável (GP)")
313
314 #Indicadores AIC e BIC
```

```
315 # -----
316 summary(fitgp1)
317 summary(fitScale)
318
319 #Teste RMSE
320 # -----
321 rmse(GP1@empirical,GP1@model)
322 rmse(Gp2@empirical,Gp2@model)
323
324 #Gráfico N veis de Retorno
325 # -----
326 dev.new(width = 1000, height = 1000)
327 plot(fitScale,type = "rl", ylim = c(0.2, 1), main = "")
328
329 #Valores do N vel de Retorno para os Per dos de Retorno 5, 20, 50
    e 80 anos
330 # -----
331 return.level(fitScale, return.period = c(5,20,50,80))
```



# Apêndice D - Gráficos da Análise das Sobrelevações de maré no Porto de Sines

## D.1 Análise dos Máximos Anuais

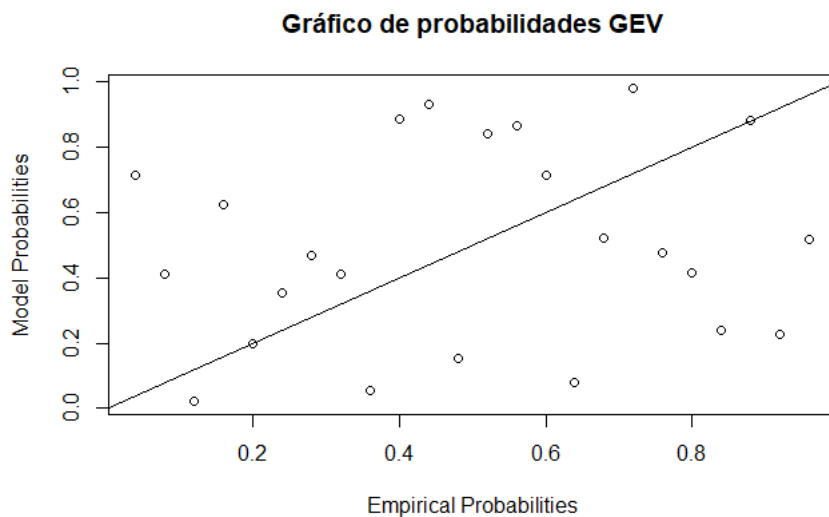


FIGURA D.1: Gráfico de probabilidades GEV para dos desvios máximos anuais do porto de Sines.

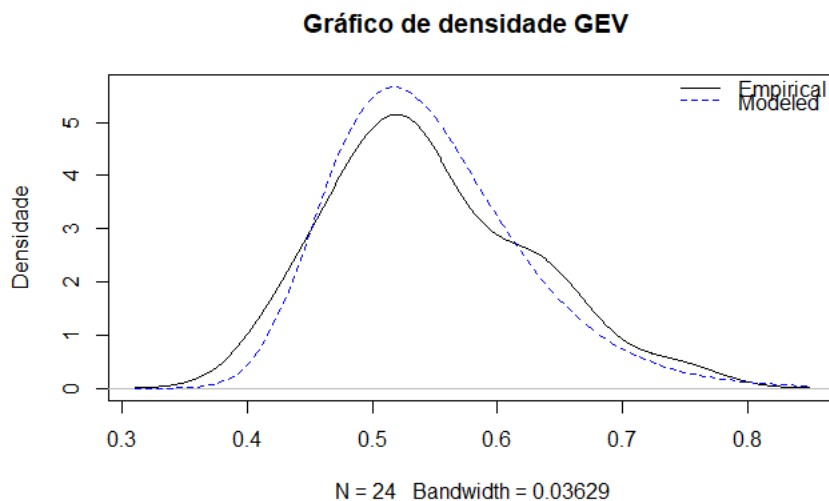


FIGURA D.2: Gráfico de densidade GEV para dos desvios máximos anuais do Porto de Sines.

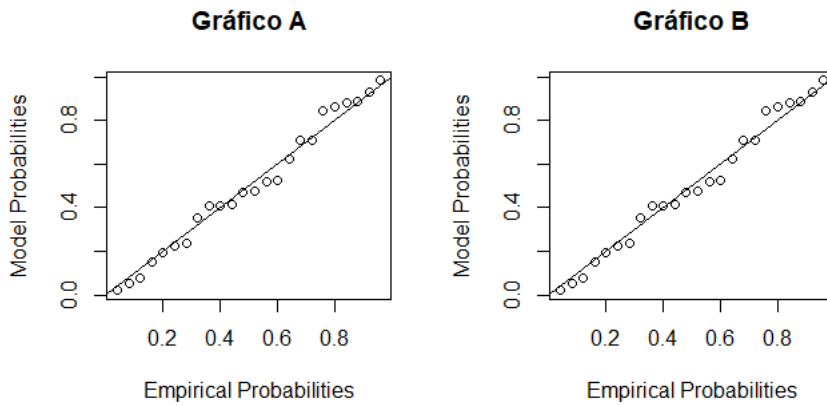


FIGURA D.3: Gráficos de probabilidade GEV com parâmetros localização e escala variáveis segundo a função 3.1.

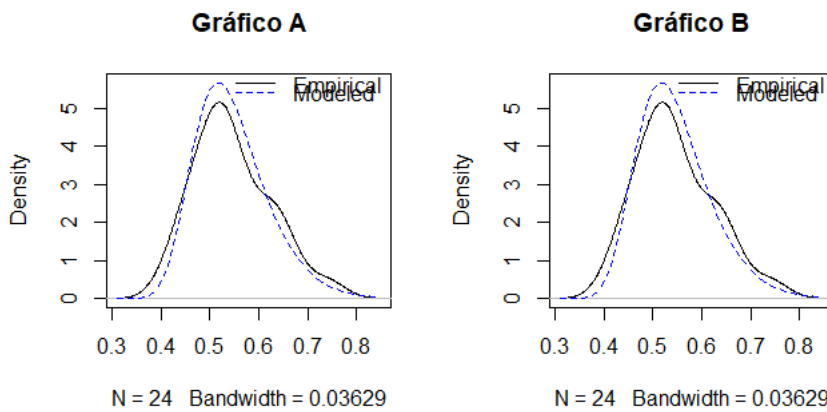


FIGURA D.4: Gráficos de densidade GEV com parâmetros localização e escala variáveis segundo a função 3.1.

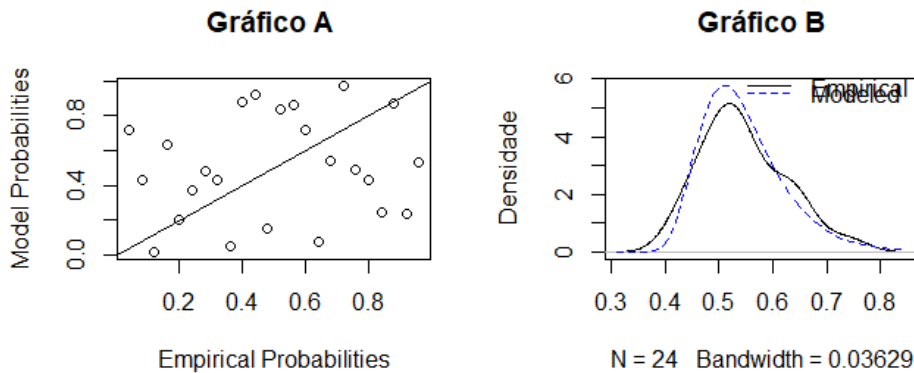


FIGURA D.5: Gráficos de probabilidade e densidade Gumbel.

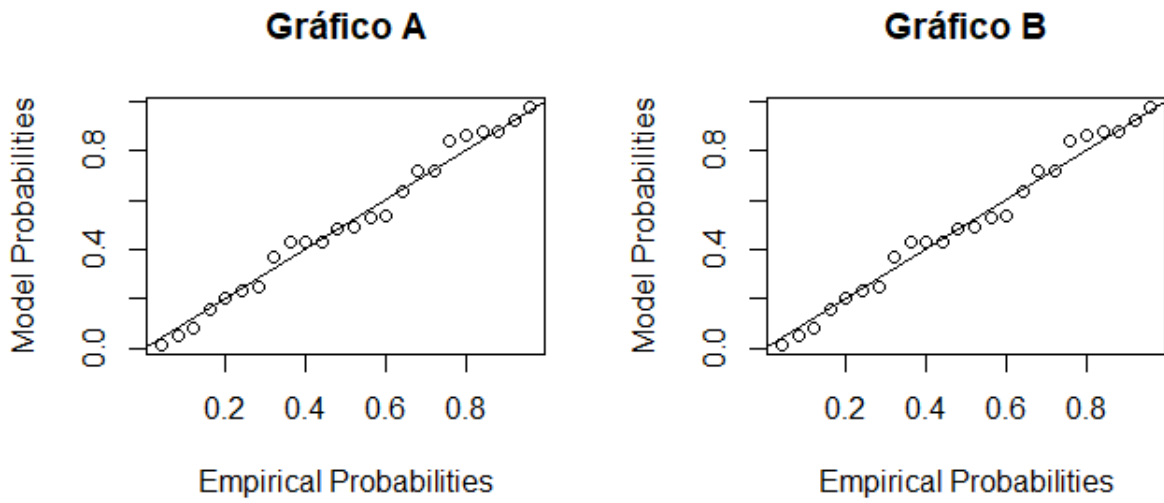


FIGURA D.6: Gráficos de probabilidade Gumbel com parâmetros localização e escala variáveis segundo a função 3.1.

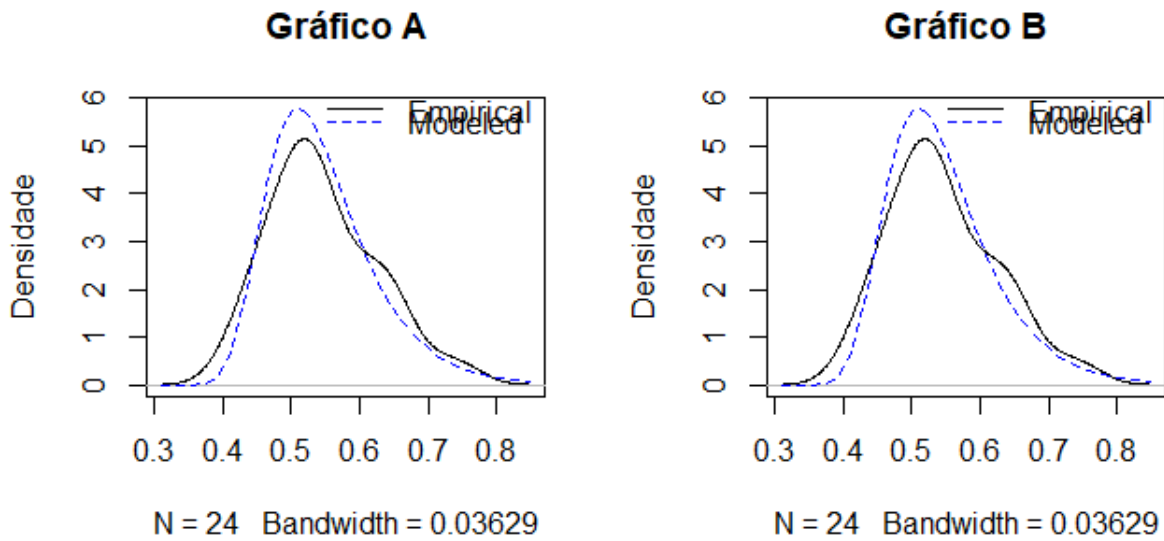


FIGURA D.7: Gráficos de densidade Gumbel com parâmetros localização e escala variáveis segundo a função 3.1.

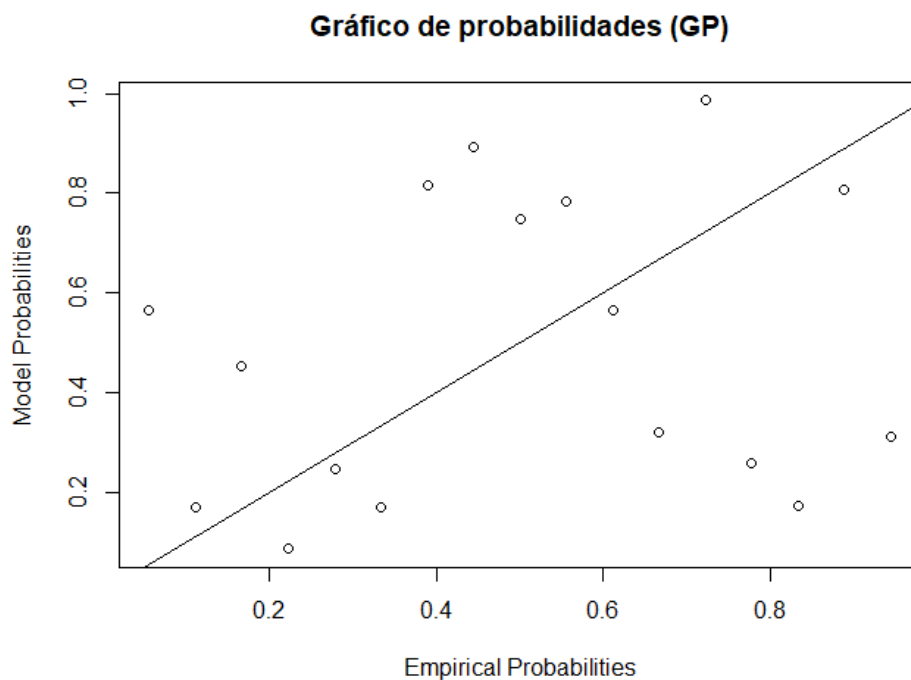


FIGURA D.8: Gráficos de probabilidade GP para os máximos anuais do Porto de Sines.

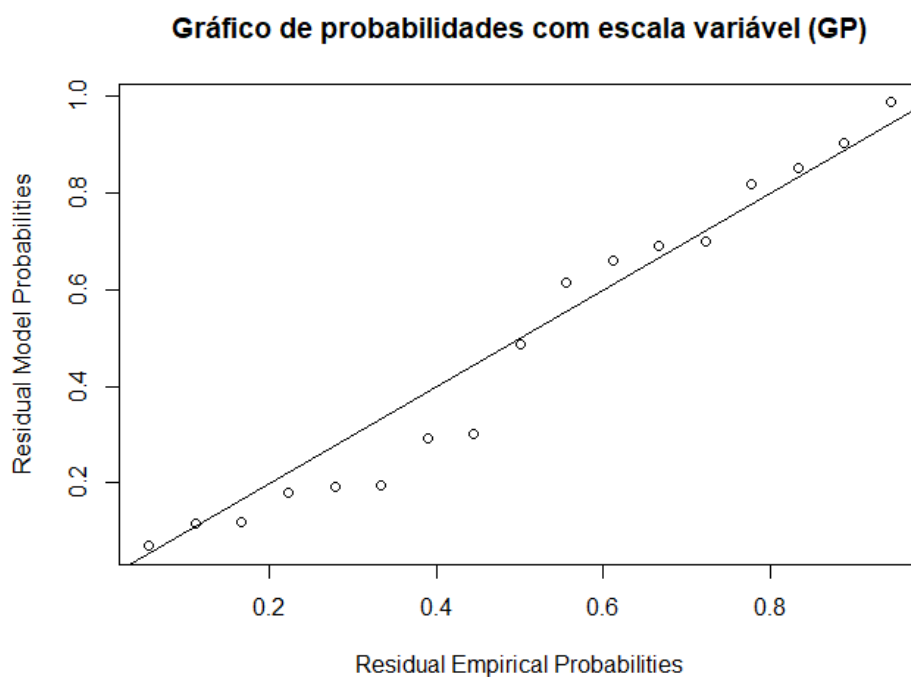


FIGURA D.9: Gráficos de probabilidade GP com parâmetro escala variável segundo a função 3.1.

## D.2 Análise dos Máximos Mensais

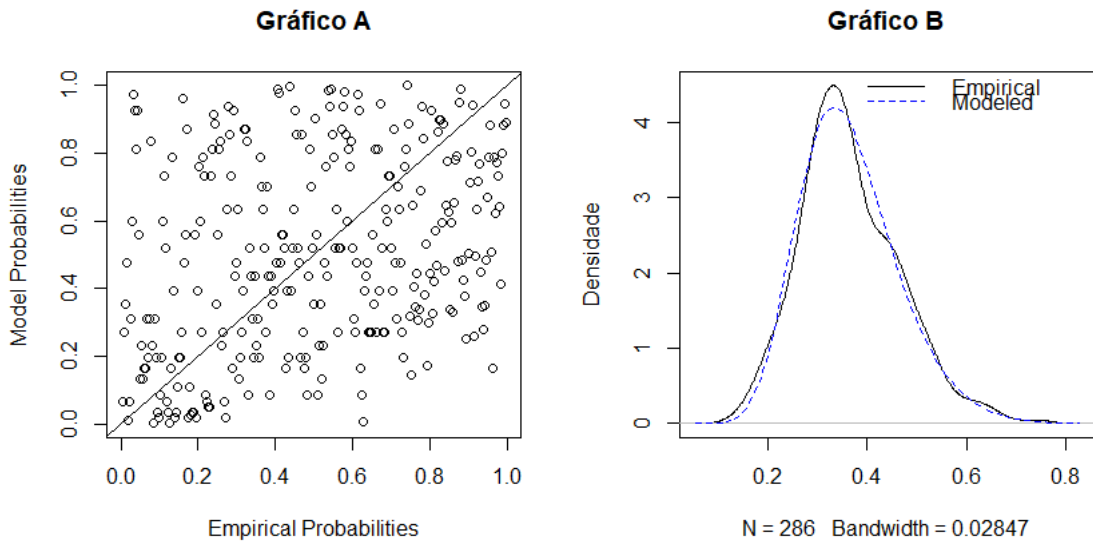


FIGURA D.10: Gráficos de probabilidade e densidade GEV para o máximos mensais do Porto de Sines.

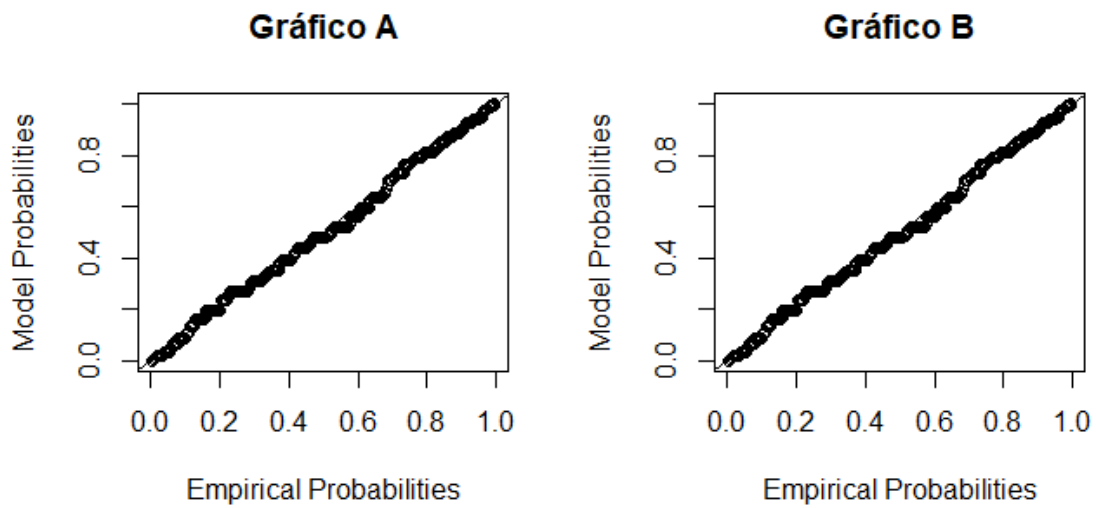


FIGURA D.11: Gráficos de probabilidade GEV com parâmetros localização e escala variáveis, para o máximos mensais do Porto de Sines.

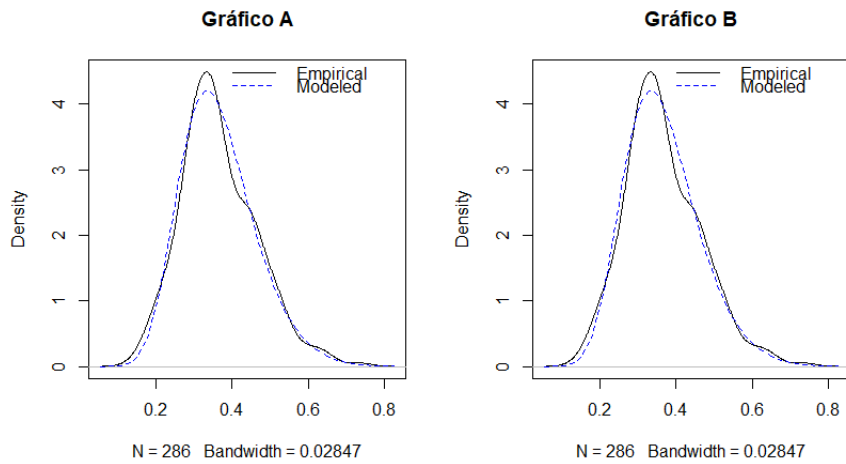


FIGURA D.12: Gráficos de densidade GEV com parâmetros localização e escala variáveis, para o máximos mensais do Porto de Sines.

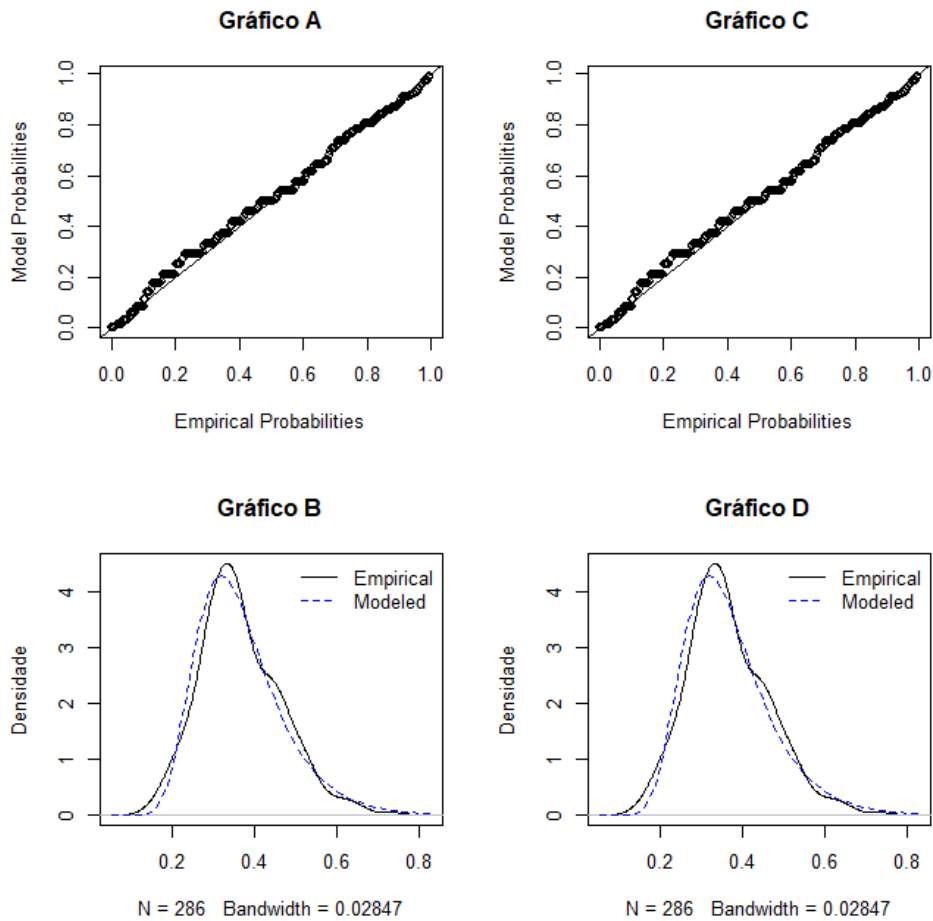


FIGURA D.13: Gráficos de probabilidade de densidade Gumbel com parâmetros localização e escala variáveis, para o máximos mensais do Porto de Sines.

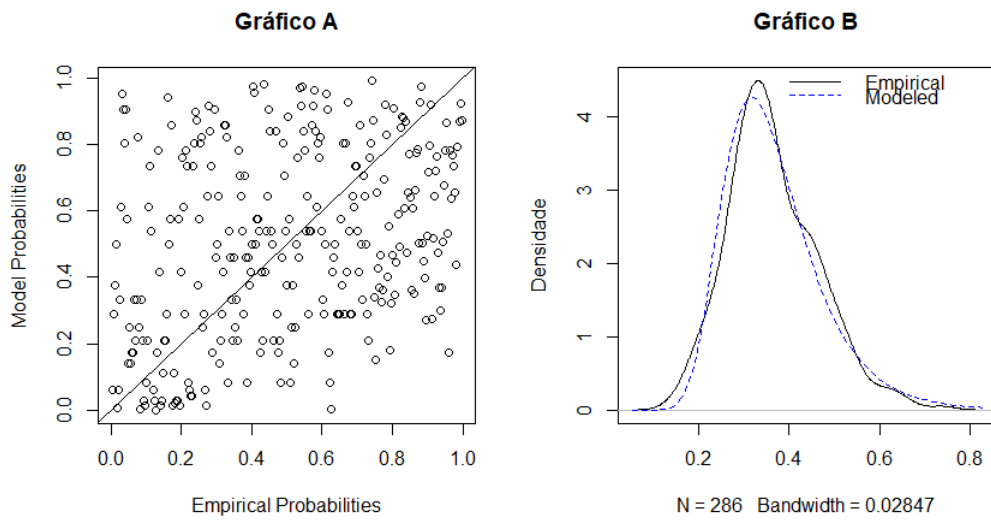


FIGURA D.14: Gráficos de probabilidade e densidade Gumbel, para o máximos mensais do Porto de Sines.

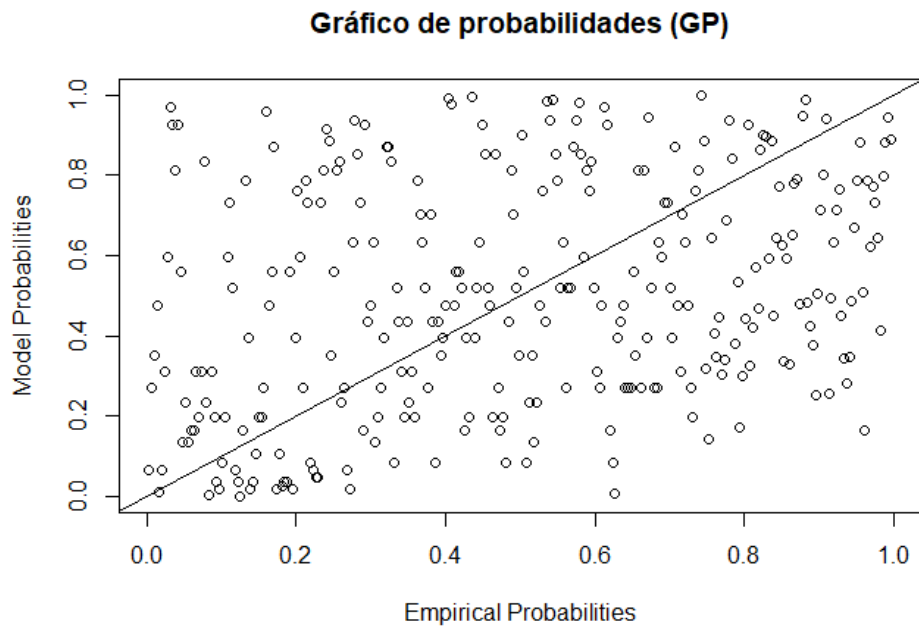


FIGURA D.15: Gráfico de probabilidade GP para o máximos mensais do Porto de Sines.

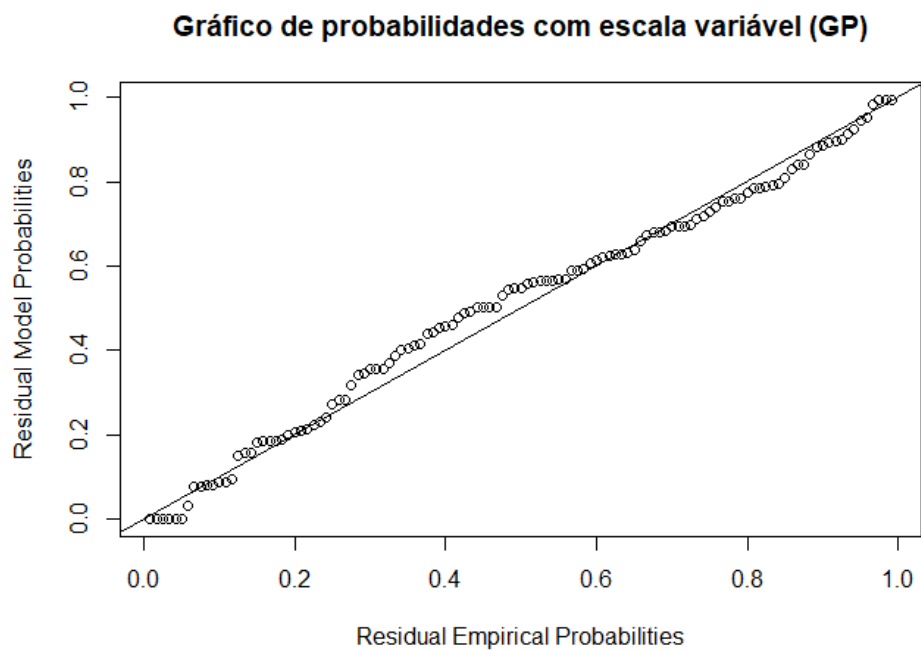


FIGURA D.16: Gráfico de probabilidade GP com parâmetro escala variável, para os máximos mensais do Porto de Sines.

# Anexo I - Esquema de Alturas do Porto de Leixões

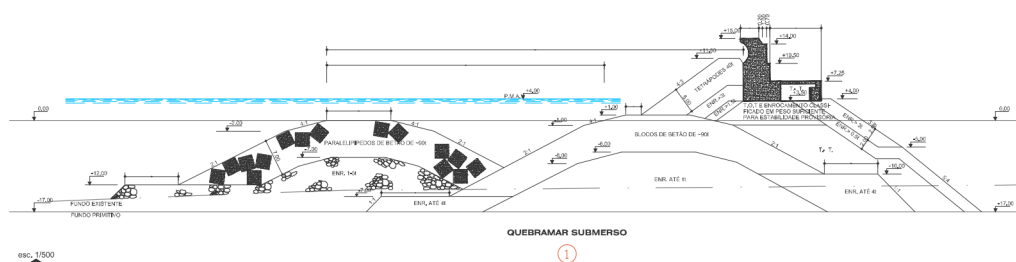


FIGURA I.1: Esquema do Quebra mar Submerso N.º 1 (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

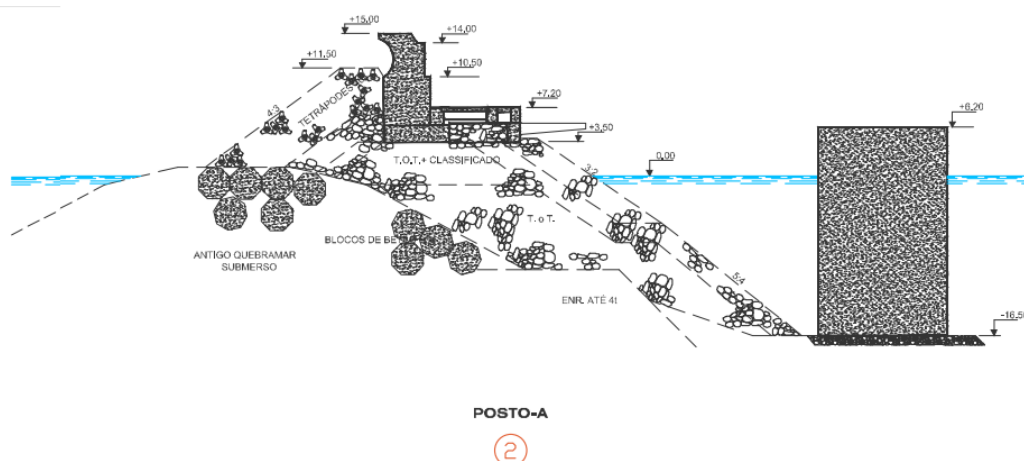


FIGURA I.2: Esquema do Posto A (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

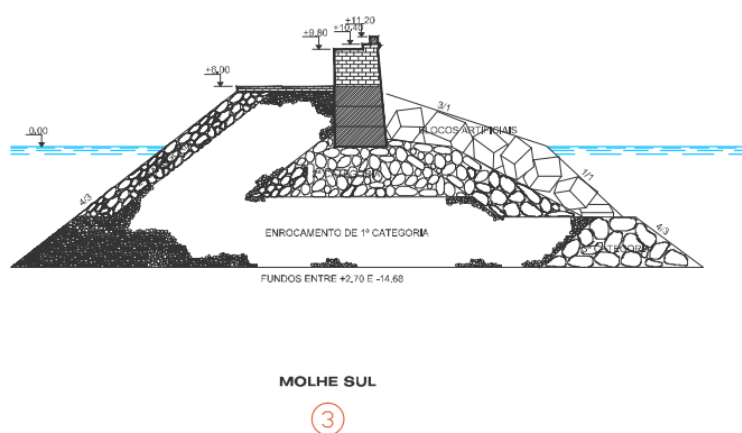


FIGURA I.3: Esquema do Molho Sul (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

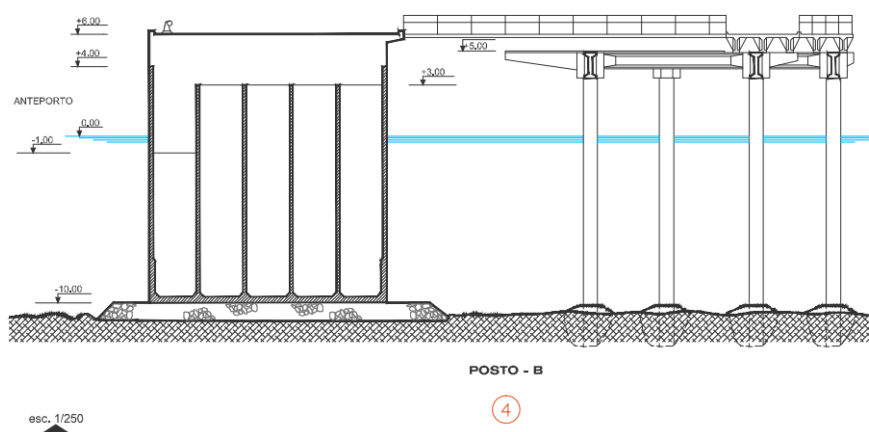


FIGURA I.4: Esquema do Posto B (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

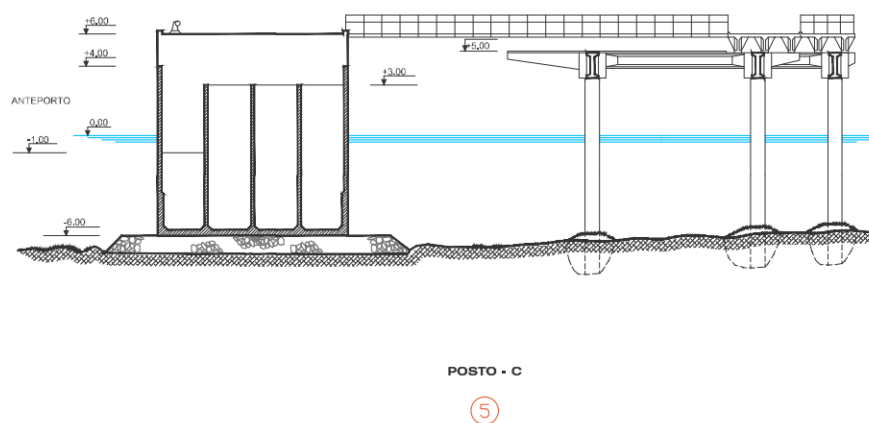


FIGURA I.5: Esquema do Posto C (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).



FIGURA I.6: Esquema do Estaleiro para obras (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

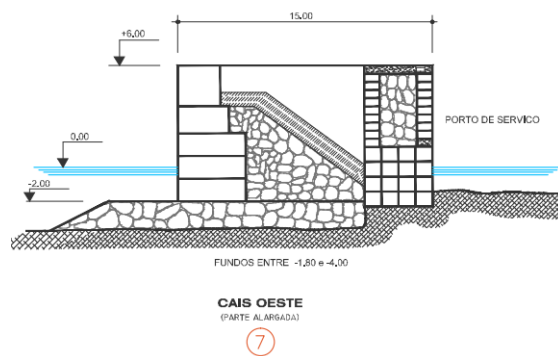


FIGURA I.7: Esquema do Cais oeste (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

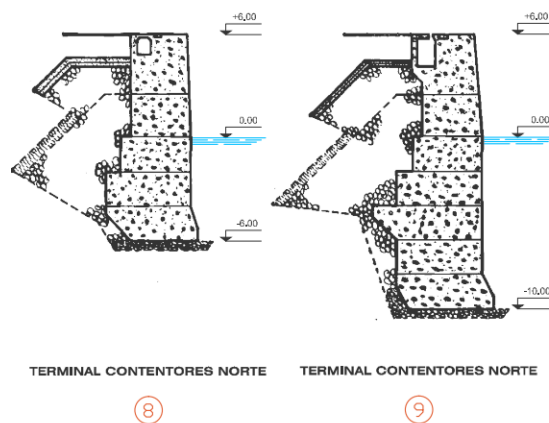


FIGURA I.8: Esquema do terminal de contentores Norte (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

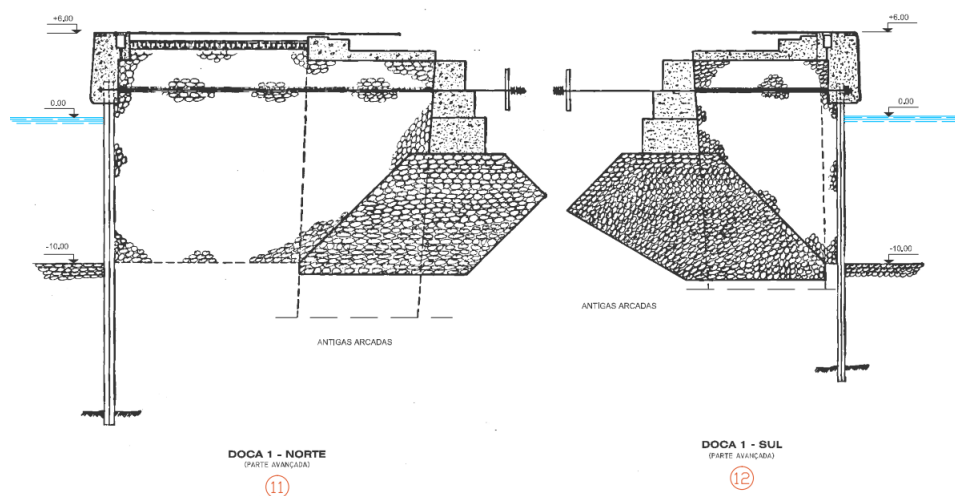


FIGURA I.9: Esquema Doca 1 - Norte e Sul (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

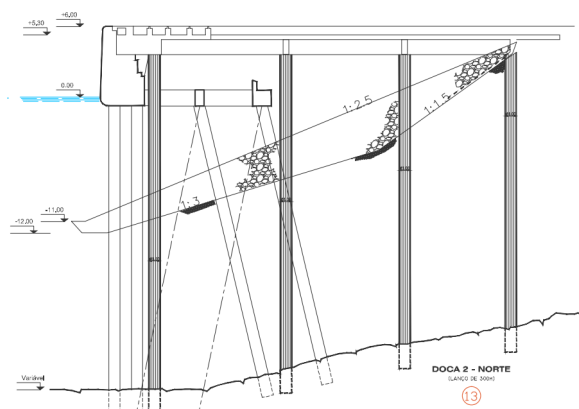


FIGURA I.10: Esquema da Doca 2 - Norte (Fonte: Administração dos Portos Douro, Leixões e Viana de Castelo).

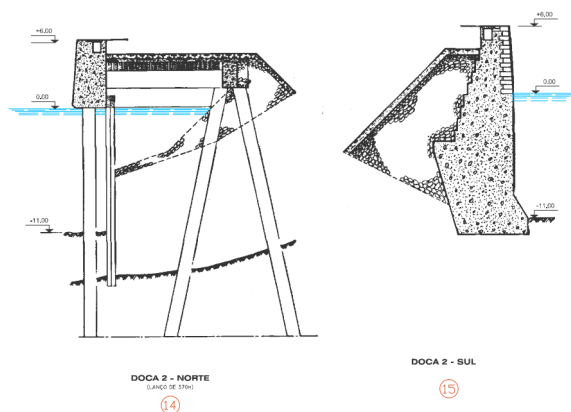


FIGURA I.11: Esquemas da Doca 2 - Norte e Sul (Fonte: Administração dos Portos Douro, Leixões e Viana de Castelo).

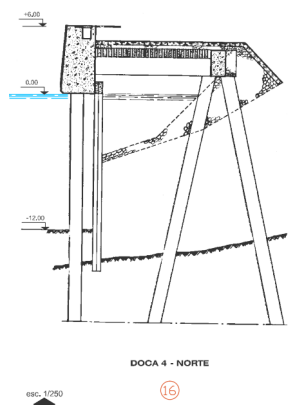


FIGURA I.12: Esquema da Doca 4 - Norte (Fonte: Administração dos Portos Douro, Leixões e Viana de Castelo).

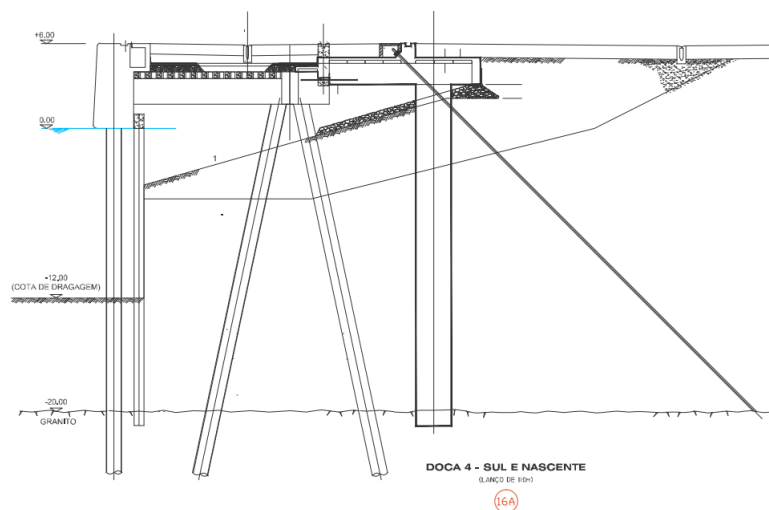


FIGURA I.13: Esquema da Doca 4 - Sul e Nascente (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

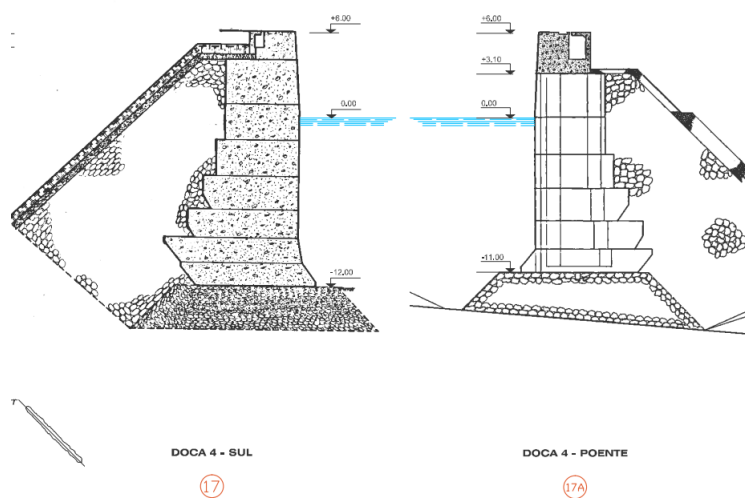


FIGURA I.14: Esquemas da Doca 4 - Sul e Poente (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

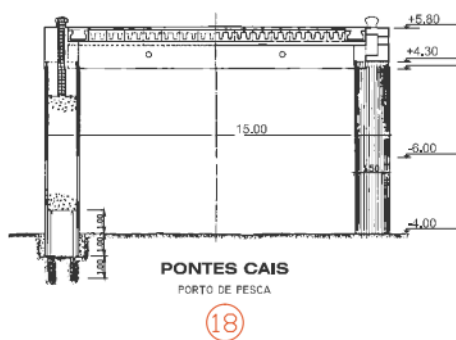


FIGURA I.15: Esquema das Pontes Cais (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

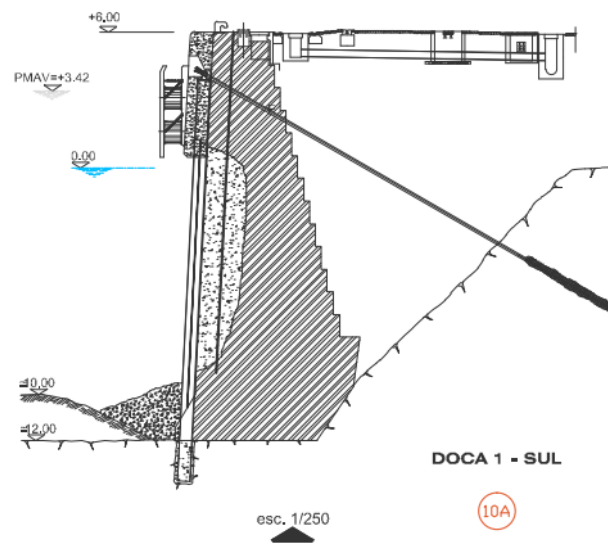


FIGURA I.16: Esquema da Doca 1 - Sul (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

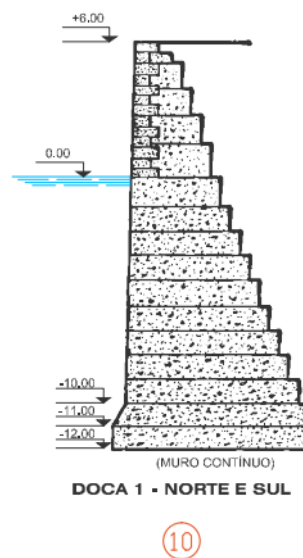


FIGURA I.17: Esquema da Doca 1 - Norte e Sul (Fonte: Administração dos Portos Douro, Leixões e Viana de castelo).

**Anexo II - Participação em: 7.<sup>a</sup> Jornadas da Engenharia Hidrográfica, 2.<sup>a</sup> Jornadas Luso-Espanholas de Hidrografia - Resumo Alargado**

# Teoria de valores extremos aplicada à caracterização da sobrelevação meteorológica no porto de Leixões

P. Rodrigues (1), D. Carinhas (2), C. Lemos (3)

- (1) Escola Naval, Instituto Universitário Militar, pedromsrodrigues98gmail.com.  
 (2) Instituto Hidrográfico, IIFA/Universidade de Évora.  
 (3) CINAV, Escola Naval, Instituto Universitário Militar, Base Naval de Lisboa – Alfeite 2810-001, Almada.

**Resumo:** Nas últimas décadas, na costa portuguesa, as sobrelevações do nível do mar, de origem meteorológica e combinadas com a agitação marítima, são a causa de muitas inundações, cheias, assim como aumento dos níveis de erosão costeira. Neste trabalho, é apresentado um estudo dos valores extremos da sobrelevação da maré, definida como a diferença entre a maré observada e a astronómica prevista, recorrendo a duas famílias de distribuições conhecidas da Teoria dos Valores Extremos (TVE), a Generalizada de Valores Extremos e Generalizada de Pareto (GEV e GP, respetivamente), com o objetivo de determinar períodos de retorno associados aos níveis da sobrelevação. A teoria dos valores extremos foi aplicada à série de dados de alturas horárias no período de 31/12/1994 a 31/12/2020 proveniente da estação maregráfica de Leixões. Mantendo-se as condições atuais, é de esperar que os desvios não excedam 1m de altura até ao final do século.

**Palavras-chave:** desvio, distribuição generalizada de Pareto, distribuição generalizada de valores extremos, marés, nível médio do mar, séries temporais.

## 1. INTRODUÇÃO

A monitorização, estudo e conhecimento da maré é de grande interesse para a zona costeira de Portugal, nomeadamente para impedir o galgamento de ondas em portos.

A altura de água,  $X(t)$ , observada num determinado local, pode ser representada por uma função com a forma seguinte:

$X(t) = Z_0(t) + T(t) + S(t)$ , onde:

$Z_0(t)$  – nível médio do mar no ponto de observação;

$T(t)$  – variação da maré astronómica;

$S(t)$  – variação da altura da água devido a efeitos meteorológicos (desvios).

O movimento regular da maré é constantemente influenciado, a um menor ou maior grau, por efeitos meteorológicos. Em certas regiões do globo, inundações costeiras intensas podem-se verificar quando há a coincidência de fenómenos meteorológicos extremos, como tempestades tropicais ou tufões, com alturas de maré astronómica elevada. As condições meteorológicas são a principal causa de diferenças entre alturas de maré previstas e observadas, fazendo-se notar com maior intensidade nos períodos de Inverno (Pugh, 1987).

Através da monitorização e do estudo das marés e dos seus fenómenos extremos, é possível construir defesas, como os quebra-mares, com altura suficiente para que a probabilidade deste tipo de eventos terem impacto seja ínfima (Gomes *et al.*, 2013).

O estudo dos valores extremos de variáveis aleatórias tem grande importância em muitos domínios, como sejam as alterações climáticas, a sismologia ou a engenharia oceanográfica, por serem os eventos extremos os que têm consequências graves ou mesmo catastróficas. Embora a estimativa de valores extremos seja um tópico mais difícil (e mesmo controverso) do que a estimativa de valores de tendência central e dispersão, é, ainda assim, de grande importância para a previsão de períodos de retorno de eventos suscetíveis de produzir grandes impactos (Guerreiro, 2015).

Num estudo de extremos de maré há a preocupação em dar resposta a questões como:

- (i) Qual o maior desvio que ocorreu?
- (ii) Qual período de retorno, estimado, destes valores extremos?
- (iii) Que impacto terão estes valores extremos quando combinados com a maior maré astronómica prevista?
- (iv) As estruturas portuárias estão dimensionadas para proteger o interior dos portos dos eventos extremos?

Este trabalho visa encontrar respostas para estas questões, tendo como caso de estudo o porto de Leixões no período de 2002 a 2020, através do ajustamento dos valores extremos das alturas horárias da maré

## 2. MATERIAIS E MÉTODOS

O teorema de Fisher-Tippett-Gnedenko estabelece que o máximo da amostra devidamente normalizado

converge em distribuição para uma variável aleatória cuja distribuição será uma das identificadas por Tipo I, II e III (Gomes et al., 2013). 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 tem os seus parâmetros de localização ( $\lambda$ ) e de escala ( $\delta$ ). As famílias de distribuições de Fréchet e Weibull têm um parâmetro de forma ( $\alpha$ ). Contudo, é possível trabalhar com uma família que abrange essas três possíveis famílias de distribuições limite. von Mises (1936) e Jenkinson (1955) (Gomes et al., 2013) unificaram as famílias Gumbel, Fréchet e Weibull na chamada distribuição Generalizada de Valores Extremos (Generalized Extreme Values - GEV) cuja expressão é a seguinte:

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

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. O parâmetro  $\gamma$  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.

Por vezes poderá ser mais útil analisar variáveis aleatórias que representem os excessos acima de um determinado nível extremo, chamado de threshold. As observações que excedem um dado threshold u são chamadas de excedências. Estas excedências têm um papel bastante importante em estudos de valores extremos. Para estas duas análises os dados da amostra em estudo, têm de ser aleatórios, independentes e identicamente distribuídos.

A distribuição Generalizada Pareto (GP) é uma aproximação para a distribuição dos excessos acima de um threshold elevado u.

A função de distribuição GP é dada por (Gomes et al., 2013):

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

Para  $\gamma > 0$ ,  $\gamma = 0$  e  $\gamma < 0$  a função distribuição GP reduz-se a um modelo Pareto, Exponencial e Beta, respetivamente. O threshold u será o parâmetro de localização,  $\sigma_u$  o parâmetro de escala e  $\gamma$  o parâmetro de forma. A escolha de um threshold origina maior dificuldade em garantir que os dados são independentes, ao contrário da GEV onde se tem

máximos de amostras com comprimento fixo, garantindo a independência. A melhor distribuição pode ser avaliada através do Critério de Informação de Akaike (AIC) que representa uma métrica avaliadora da qualidade de um modelo estatístico visando também a sua simplicidade. Fornece, portanto, uma métrica para comparação e seleção de modelos, em que menores valores de AIC representam uma maior qualidade e simplicidade (Gonçalves e Lopes, 2003).

Outro conceito muito usado na análise de valores extremos 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 não é mais do que o quantil de probabilidade  $1 - 1/T$ .

### 3. PROCESSAMENTO E ANÁLISE DOS RESULTADOS

Com vista à aplicação prática das metodologias apresentadas, a amostra em estudo compreende os desvios (diferença entre a altura da maré observada e a altura da maré astronómica prevista, devida essencialmente a fatores meteorológicos) do porto de Leixões, referentes ao período de 2002 a 2020.

Estes dados, foram disponibilizados pelo Instituto Hidrográfico. As alturas horárias são expressas em metros e calculadas em relação ao zero hidrográfico (ZH) do porto, situado a 2m abaixo do nível médio adotado (Cascais, 1938) (IH, 2021).

A série de alturas horárias tem uma dimensão de 161462 observações, o que corresponde a 19 anos de dados, havendo valores omissos (Figura 1).

As lacunas nas observações foram preenchidas com o valor da mediana da amostra. Dados que todos os valores extremos são superiores à mediana, esta forma de imputação não enviesada os resultados da estimação dos parâmetros das distribuições de extremos. Com a série completa encontraram-se os máximos mensais representados graficamente na Figura 2, obtendo se assim 221 máximos para o estudo.

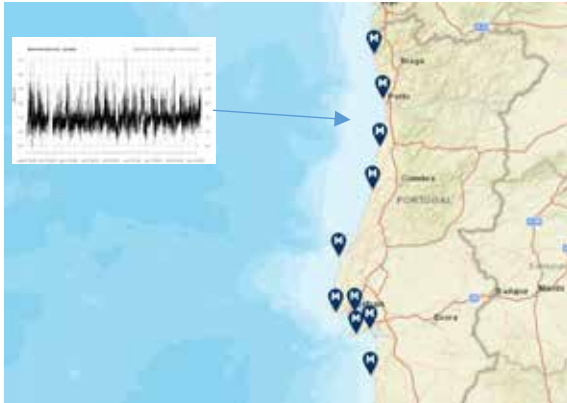


Fig. 1. Série temporal dos desvios de 2002 a 2020: Porto de Leixões (fonte: Instituto Hidrográfico).

Foi construída a série dos máximos mensais para o período em análise. Tem-se, portanto, um total de 221 meses. Estes desvios encontram-se representados na Figura 2.

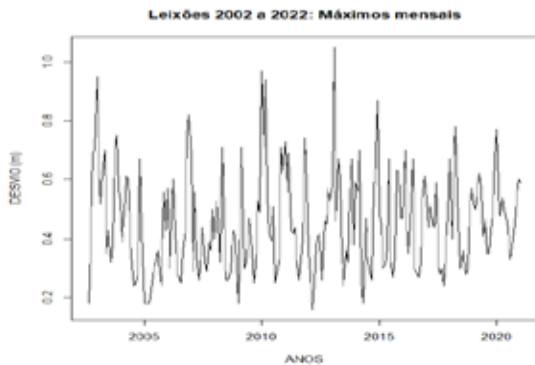


Fig. 2. Leixões 2002 a 2022: Máximos mensais.

Através do gráfico de auto correlação parcial, na Figura 3, poder-se-á avaliar a independência das observações.

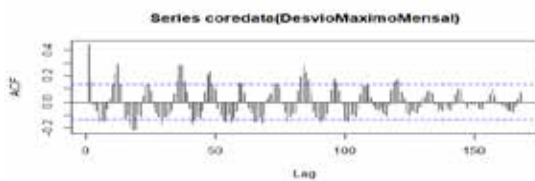


Fig. 3. Independência dos Desvios máximos mensais.

Como se pode observar, existe correlação entre as observações, pois o módulo da função de auto correlação parcial ultrapassa os limites do intervalo de confiança a 95% para a correlação nula. Através da figura, também se observa alguma sazonalidade, facto a considerar na determinação dos parâmetros de localização e/ou escala que melhor ajustam a distribuição GEV.

Nesta amostra os desvios horários variam entre os 0,23m e 1,05m. Tem-se, uma amostra com uma acentuada assimetria à direita que é visível no histograma representado na Figura 4.

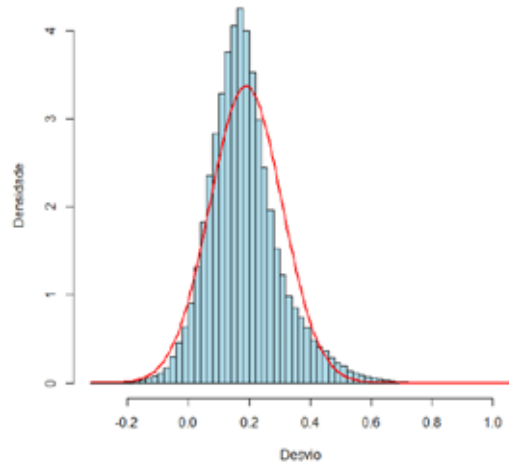


Fig. 4. Histograma dos desvios máximos mensais em Leixões.

Relativamente a cada distribuição, GEV e GP, foi encontrada uma estimativa para o parâmetro de forma  $\gamma$ ; assim como encontrado o valor do AIC (Tabela 1).

Tab. 1. Comparação entre as distribuições GEV e Gumbel

	parâmetro forma	parâmetro escala	parâmetro localização	AIC
GEV	$\gamma = -0,041$ K(95%) = [-0,149; 0,067]	$\delta = 0,138$ K(95%) = [0,122; 0,152]	$\lambda = 0,379$ K(95%) = [0,358; 0,399]	-16,06
Gumbel	-	$\delta = 0,136$ K(95%) = [0,122; 0,150]	$\lambda = 0,376$ K(95%) = [0,357; 0,395]	-18,01

Uma vez que o valor AIC da distribuição de Gumbel é menor do que o da distribuição GEV, conclui-se que a distribuição de Gumbel apresenta maior qualidade. Assim, para essa distribuição, os períodos de retorno para desvios superiores a 80 cm é de, aproximadamente, 50 anos. Até ao final do século não se esperam desvios com altura superior a 1 metro.

Seguidamente, efetuou-se o estudo da amostra através da GP. Neste caso, os blocos amostrais usados para calcular os máximos já não têm dimensão fixa, e surge a necessidade de estimar o *threshold* a usar. Se for demasiado elevado, perde-se a vantagem da distribuição GP relativamente à GEV; se for demasiado baixo, os valores podem ter correlação significativa entre si. Existe um método para estimar o *threshold*, a Figura 5 ilustra a aplicação de um estimador usado por Coles (2001). Com base nos resultados mostrados, foi escolhido o *threshold* = 0,45m.

Com a escolha deste *threshold* obteve-se uma amostra com dimensão de 103 desvios para o estudo do período de retorno.

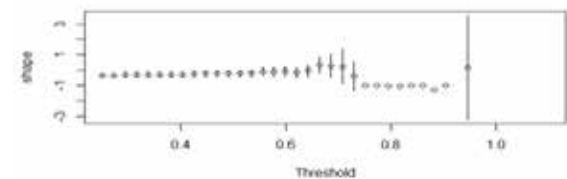


Fig. 5. Escolha do *threshold* para os máximos mensais.

Considerando a tendência do aumento do nível do mar e assumindo uma variação sazonal, é provável que em 10 anos, os desvios, sobre a maré astronómica, atinjam valores acima de 0,8m, no

entanto até ao final do século não se esperam sobrelevações superiores a 1,1 m (Figura 6).

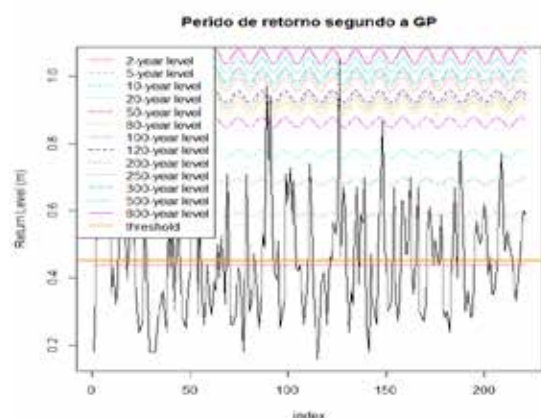


Fig. 6. Período de retorno calculado através da GP

Os desvios esperados, bem como o respetivo período de retorno são apresentados na Tabela 2.

Tab. 2. Resultados dos desvios e respetivo período de retorno para as duas distribuições

	5 Anos	10 Anos	50 Anos	80 Anos
<b>Gumbel</b>	0,60 m	0,68 m	0,88 m	0,93 m
<b>GP</b>	0,60 m	0,70 m	0,91 m	0,97 m

Através da capitania do Porto de Leixões, os cais do porto estão a uma altura de 6,33 metros em relação ao ZH. Assim, até ao final do século é pouco provável que existam fenómenos extremos que ultrapassem a altura dos cais, considerando a combinação entre a maior maré astronómica e o desvio máximo previstos (4,96 metros)

## CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentado um estudo dos valores extremos da sobrelevação devida à maré meteorológica no porto de Leixões, baseada na TVE. Através deste estudo respondeu-se às perguntas iniciais levantadas, o desvio maior ocorrido foi de 1,05 metros, prevê-se que até ao final do século o maior nível de retorno será de 0,97 (estimado através da distribuição GP), sendo que este quando combinado com a maior maré astronómica não irá ultrapassar a altura dos cais.

Tendo em conta as previsões realizadas é possível afirmar que é pouco provável que existam galgamentos no Porto de Leixões e que as suas estruturas estão dimensionadas de forma a proteger o porto destes fenómenos. Há ainda a considerar os efeitos da agitação marítima e da ocorrência de seichas dentro do porto, mas devido ao abrigo proporcionado pelos molhos, no caso das primeiras, e a amplitude pequena das segundas, estes efeitos não afetam de forma significativa a sobrelevação da maré.

## REFERÊNCIAS

- Almeida, A., Taveira-Pinto, F., Veloso-Gomes, F., Moura Paredes, G. (2009). Análise das Marés Meteorológicas em Leixões. 6as Jornadas Portuguesas de Engenharia Costeira e Portuária.
- Gomes, M. I., Alves, M. I. F., Neves, C. (2013). Análise de Valores Extremos: Uma Introdução. Sociedade Portuguesa de Estatística.
- Guerreiro, R. I. R. (2015). Delimitação de áreas inundáveis – Estudo da ação combinada da maré e do escoamento superficial. Universidade Évora. Escola de Ciências e Tecnologia - Departamento de Engenharia Rural.
- Gonçalves, E., Lopes, N. M. (2003). Estatística: Teoria e Aplicações. Escolar Editora.
- IH (2021). Tabela de Marés, Portugal-Volumel, Instituto Hidrográfico.
- Pugh, D. T. (1987). Tides, Surges and Mean Sea Level. Wiley.
- Coles, S. (2001) An Introduction to Statistical Modeling of Extreme Values. First edition. London, England: Springer series in statistics.